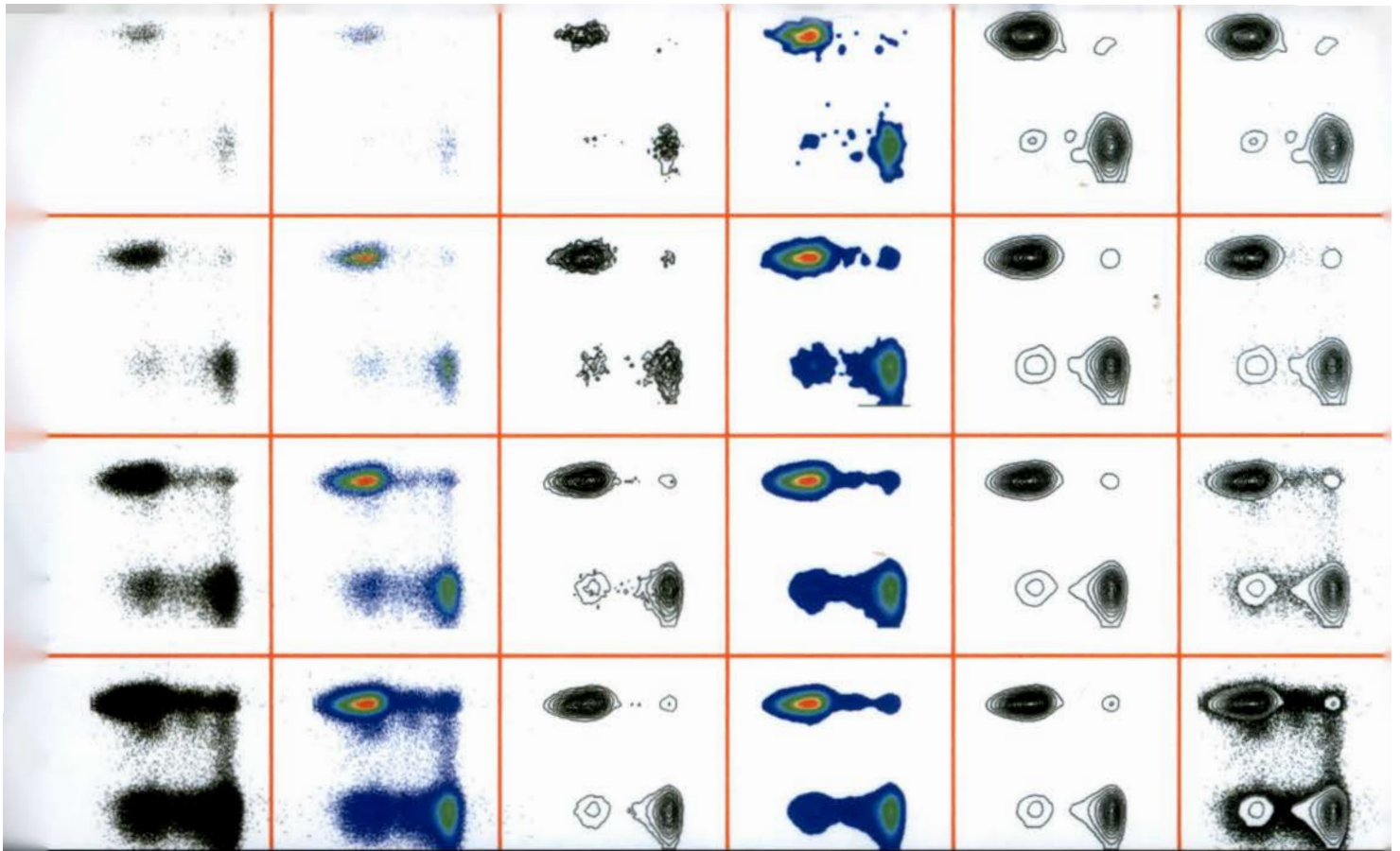




PRACTICAL FLOW CYTOMETRY

FOURTH EDITION

HOWARD M. SHAPIRO

S H A P I R O

PRACTICAL FLOW CYTOMETRY
FOURTH EDITION



WILEY-
LISS

PRACTICAL FLOW CYTOMETRY

Fourth Edition

HOWARD M. SHAPIRO

 **WILEY-LISS**

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2003 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, e-mail: permreq@wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representation or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Shapiro, Howard M. (Howard Maurice), 1941–
Practical flow cytometry / Howard M. Shapiro. — 4th ed.
Includes bibliographical references and index.
ISBN 0-471-41125-6 (alk. paper)
1. Flow cytometry.
[DNLM: 1. Flow Cytometry. QH 585.5.F56 S529p 2002] I. Title.
QH585.5.F56 S48 2002
571.6'0287—dc21 2002002969



M WEP279912 230424

DEDICATION

To the memory of Bart de Grooth, who built, and inspired his students to build, small, simple, elegant cytometers; I wish we had played more guitar duets.

To the memory of Mack Fulwyler, who gave us the cell sorter and became a biologist in the bargain, always eager to learn more and to put the knowledge to good use.

To the memory of Janis Giorgi, a very classy lady who made great advances in the immunology of HIV infection, never forgetting she was working for the patients.

And for Jacob, Benjamin, and Anna – **“לתקן עולם במלכות שדי”**

CONTENTS

LIST OF TABLES AND FIGURES.....	xxvii
PREFACE TO THE FOURTH EDITION: WHY YOU SHOULD READ THIS BOOK – OR NOT	xxxiii
FOREWORD TO THE THIRD EDITION by Leonard A. Herzenberg.....	xxxix
PREFACE TO THE THIRD EDITION	xli
PREFACE TO THE SECOND EDITION	xlvi
FOREWORD TO THE FIRST EDITION by Louis A. Kamentsky.....	xlvi
PREFACE TO THE FIRST EDITION	xlix
I. OVERTURE.....	1
1.1 What (And What Good) Is Flow Cytometry?	1
Tasks and Techniques of Cytometry	1
Some Notable Applications.....	1
What is Measured: Parameters and Probes	2
1.2 Beginnings: Microscopy And Cytometry	2
A Little Light Music	4
Making Mountains out of Molehills: Microscopy	6
Why Cytometry? Motivation and Machinery.....	9
Flow Cytometry and Sorting: Why and How.....	10
Fluorescence and Flow: Love at First Light.....	11
Conflict: Resolution	12
1.3 Problem Number One: Finding The Cell(s).....	14
Flow Cytometry: Quick on the Trigger.....	16
The Main Event	17
The Pulse Quickens, the Plot Thickens.....	17
1.4 Flow Cytometry: Problems, Parameters, Probes, and Principles	18
Counting Cells: Precision I (Mean, S.D., CV)	18
Poisson Statistics and Precision in Counting	19
Rare Event Analysis: The Fundamental Things Apply as Cells Go By.....	19
Count Constant Numbers for Constant Precision	20
Alternative Counting Aids: The Venerable Bead.....	20
And Now to See with Eye Serene the Very Pulse	
of the Machine: Display, Digitization, and Distributions	21
DNA Content Analysis: Precision II (Variance).....	21
The Normal Distribution: Does the Word “Gaussian” Ring a Bell?.....	22

1.4 Flow Cytometry: Problems, Parameters, Probes, and Principles (continued)	
Binned Data: Navigating the Channels.....	22
DNA Content: Problem, Parameter, Probes	23
One-Parameter Displays: Pulse Height Distributions	24
Mathematical Analysis of DNA Histograms: If It's Worth Doing, It's Worth Doing Well.....	25
Linear Thinking	26
Lineage Thinking: Sperm Sorting.....	26
Two-Parameter Displays: Dot Plots and Histograms	26
Multiparameter Analysis Without Computers: Gates Before Gates	27
Two-Parameter Histograms: Enter the Computer.....	29
Modern Multiparameter Analysis: List Mode	30
Three-Dimensional Displays: Can We Look at Clouds from Both Sides? No	32
Identifying Cells in Heterogeneous Populations: Lift Up Your Heads, Oh Ye Gates!.....	33
Cluster Headaches.....	34
Painting and White- (or Gray-) Washing Gates	34
The Quad Rant: Are You Positive? Negative!.....	35
Deals With the Devil: Logarithmic Amplifiers and Fluorescence Compensation	35
Evils of Axes: Truth in Labeling Cells and Plots.....	38
When Bad Flow Happens to Good Journals	40
Sorting Sorting Out	40
Parameters and Probes II: What is Measured and Why	42
Probes versus Labels	42
Living and Dyeing: Stains, Vital and Otherwise.....	43
Nucleic Acid (DNA and RNA) Stains	43
Fluorescence and Fluorescent Labels.....	44
Binary Fishin': Tracking Dyes Through Generations.....	45
Membrane Perturbation: A Matter of Life and Death?.....	46
Cytoplasmic/Mitochondrial Membrane Potential.....	46
Indicators of Cytoplasmic $[Ca^{2+}]$: Advantages of Ratiometric Measurements	47
Finding Antigen-Specific Cells Using Tetramers.....	47
Hip, Hip Arrays: Multiplexing on Slides and in Bead Suspensions	48
GFP and its Relatives: Mild Mannered Reporters	48
Beyond Positive and Negative; Putting the -Metry in Cytometry.....	48
1.5 What's In the Box: Flow Cytometer Anatomy, Physiology, and Pathology	49
Light Sources for Microscopy and Flow Cytometry.....	49
Instrument Configurations: The Orthogonal Geometry	50
Laser Beam Geometry and Illumination Optics.....	50
Flow Chamber and Forward Scatter Collection Optics.....	51
Fluorescence and Orthogonal Scatter Optics	52
Optical Filters for Spectral Separation	52
Multistation Flow Cytometers	54
Photomultipliers and Detector Electronics	54
Putting the Flow in Flow Cytometry.....	55
Signal Processing Electronics.....	57
Is It Bigger than a Breadbox?.....	57
Flow Cytometer Pathology and Diagnostics	58
1.6 Alternatives to Flow Cytometry; Cytometer Ecology	59
1.7 The Rest Of The Book	60
Lis(z)t Mode	60
2. LEARNING FLOW CYTOMETRY	61
Learning from History: Take One.....	61
Who Should Read this Book?.....	62
2.1 Information Sources and Resources	62
Books on Flow Cytometry in General	62
Books on Flow Cytometric Methodology and Protocols.....	62

2.1 Information Sources and Resources (continued)	
Clinical Flow Cytometry Books	63
Other Flow Cytometry Books	63
Flow's Golden Oldies	63
2.2 The Reader's Guide To Periodical Literature.....	64
2.3 Resources And Courses	66
Flow Cytometer Manufacturers	66
The International Society for Analytical Cytology	66
The Clinical Cytometry Society	66
The National Flow Cytometry Resource	66
"The Annual Courses" and Others.....	67
Other Societies and Programs	67
The Purdue Mailing List, Web Site, and CD-ROMs	68
2.4 Exploring The Foundations	68
Optics and Microscopy	68
Electronics.....	69
Computers: Hardware and Software	70
Digital Signal Processing.....	70
Data Presentation and Display	71
Spectroscopy, Fluorescence and Dye Chemistry	71
Cell and Molecular Biology and Immunology	71
2.5 Alternatives To Flow Cytometry	71
3. HISTORY	73
3.1 Ancient History	73
Flow Cytometry: Conception and Birth.....	73
Staining Before and After Paul Ehrlich.....	74
Origins of Modern Microscopy.....	75
Making Cytology Quantitative: Caspersson et al.....	75
Origins of Cancer Cytology: The Pap Smear.....	76
The Fluorescent Antibody Method	77
Blood Cell Counting: Theory and Practice	77
Video and Electron Microscopy	78
Optical Cell Counters and the Coulter Orifice.....	78
3.2 Classical History	79
Analytical Cytology in the 1950's	79
The Cytoanalyzer.....	79
Acridine Orange as an RNA Stain: Round One	79
How I Got Into this Mess.....	79
The Rise of Computers	80
Computers in Diagnosis: A Central Problem	80
Diagnosis and Classification: Statistical Methods	80
Cytology Automation in the 1960's	81
First Steps toward Automated Differentials	81
Pattern Recognition Tasks in Cell Identification	82
Differential Leukocyte Counting: An Early Flow Systems Approach	83
Kamentsky's Rapid Cell Spectrophotometer	84
Fulwyler's Cell Sorter.....	85
3.3 Modern History.....	85
Cell Cycle Analysis: Scanning versus Flow Systems	85
Cancer Cytology: Scanning versus Flow Cytometry	86
Early Commercial Flow Cytometers.....	87
Not Quite Commercial: The Block Projects.....	89
The Evolution of Flow Cytometers in the 1970's.....	90
Dog Days: The Genesis of Cytomutts.....	93
The 1980's: Little Things Mean a Lot.....	94

3.3 Modern History (continued)	
Measurements in the Main Stream.....	95
Immunofluorescence Comes of Age.....	95
Developments in DNA Content Analysis	96
Flow Cytometry of RNA Content	96
Measurements of Functional Parameters.....	97
Clinical Uses of Fluorescence Flow Cytometry.....	98
The End of History?	99
4. HOW FLOW CYTOMETERS WORK	101
4.1 Light and Matter	101
Introduction	101
Photometry versus Radiometry: What's in a Name?	101
Physical Measurement Units	101
Light in Different Lights	102
It's All Done With Photons	102
A Few Warm Bodies	103
Polarization and Phase; Interference.....	104
Light Meets Matter: Rayleigh and Mie Scattering.....	105
A Time for Reflection - and Refraction: Snell's Law	107
Polarization by Reflection; Brewster's Angle.....	107
Dispersion: Glass Walls May Well a Prism Make	108
Interference in Thin Films	108
Interference and Diffraction; Gratings.....	108
Optical Activity and Birefringence	109
Matter Eats Light: Absorption.....	109
Absorption: Counting the Calories.....	110
A Selective Diet.....	110
The Chance of a Lifetime.....	110
Spinning a Tale of Degeneracy.....	111
Facing Extinction: Cross Section and Optical Density	111
Unexciting Times: Emigrating from the Excited States.....	112
Fluorescence: Working the Stokes Shift.....	112
Phosphorescence	113
Fluorescence Polarization	114
Stimulated Emission	114
Resonance Energy Transfer	115
Quenching, Bleaching, and Photon Saturation.....	115
Quantum Flotsam and Jetsam.....	118
Inelastic Scattering and Doppler Measurements.....	118
Raman Scattering	118
Nonlinear Optics and Harmonic Generation.....	118
Two-Photon and Multiphoton Excitation	118
4.2 Optical Systems	119
Light Propagation and Vergence	119
Image Formation by Optical Systems: Magnification	119
Lens Types and Lens Aberrations	120
Numerical Aperture and Lens Performance	121
Gradient Index, Fresnel, and Cylindrical Lenses.....	122
The Helmholtz Invariant and Throughput.....	123
Photons in Lenses: See How They Run.....	123
Aperture and Field Stops: The f Number	124
Depth of Field and Focus and Resolution of Lenses	124
4.3 Light Sources.....	124
The Best and the Brightest	124

4.3 Light Sources (continued)	
Harc, Harc, the Arc!	126
Quartz Halogen Lamps	127
Light Emitting Diodes (LEDs)	127
Illumination Optics for Lamps and LEDs	127
Arc Source Epiillumination for Flow Cytometry	128
Lasers as Light Sources for Flow Cytometers	129
Laser Illumination: Going to Spot	130
Shedding Light on Cells: Lasers, Lamps, and LEDs	131
Lasers: The Basic Physics	133
Einstein on the Beam: Stimulated Emission	133
Look, Ma, One Cavity: Optical Resonators	133
Laser Action à la Mode	134
Pumping Ions	135
Laser Efficiency: Your Mileage May Vary	135
Mirrors and Prisms for Wavelength Selection	135
Brewster Windows for Polarized Output	135
Laser Power Regulation: Current and Light Control	136
Beam Profiles and Beam Quality	136
Puttin' on My Top Hat?	138
Harmonic Generation and Modulation	138
Lasers Used and Usable in Cytometry	138
Argon and Krypton Ion Lasers	138
Dye Lasers	141
Helium-Neon Lasers	141
Helium-Cadmium and Helium-Selenium Lasers	142
Diode Lasers: Red, Infrared, Violet, and UV	142
Solid-State Lasers: Like, YAG Me!	145
Laser and Light Source Noise and Noise Compensation	147
Fifty Ways to Lose Your Laser	148
Danger!!! Laser!!! Hazards and Haze	148
4.4 Light Collection	149
Microscope Objectives	149
Looking at the Observation Point	150
Stops versus Blockers	150
Signal versus Noise: To See or Not to See	150
Spectral Selection: Monochromators versus Filters	152
Monochromators and Polychromatic Detection	152
Interference Filters: Coatings of Many Colors	153
Absorptive Filters versus Interference Filters	153
Filter Transmission Characteristics	154
Dichroics	155
Neutral Density Filters	156
Beamsplitters: Ghosts and Ghostbusters	156
Optics for Polarization Measurements	156
Tunable Filters	157
Fiber Optics and Optical Waveguides	157
Through a Glass Darkly: Light Lost (and Found) in Optical Components	158
Collection Optics for Forward Scatter Signals	159
4.5 Detectors	160
Silicon Photodiodes	160
Photomultiplier Tubes (PMTs)	161
Sensitivity Training: Photodiode versus PMT	163
Single Photon Counting	164
Avalanche Photodiodes (APDs)	164
PMTs: Picking a Winner	165

4.5 Detectors (continued)	
Photomultipliers: Inexact Science.....	166
Charge Transfer Devices: CCDs, CIDs, Etc.....	166
4.6 Flow Systems	166
Flow System Basics	167
Gently Down the Stream: Laminar Flow	167
Flow Chambers; Backflushes, Boosts, and Burps	169
Cuvettes versus Streams for Analysis and Sorting	170
Light Collection from Streams and Cuvettes.....	171
When You've a Jet	174
Core and Sheath: Practical Details.....	175
Grace Under Pressure: Driving the Sheath and Core	175
Perfect Timing: Fluidics for Kinetic Experiments	177
Oriented and Disoriented Cells.....	178
Matchmaker, Matchmaker, Make Me a(n) Index Match!	178
Flow Unsheathed	178
Flow Systems: Garbage In, Garbage Out.....	178
4.7 Electronic Measurements	180
Electricity and Electronics 101	180
Charge Separation, Electric Fields, and Current.....	180
Resistance, Voltage, and Power; Ohm's Law.....	181
Alternating and Direct Current; Magnetism	181
Inductance, Reactance, Capacitance, Impedance	182
The Coulter Principle: Electronic Cell Sizing	182
Electrical Opacity: AC Impedance Measurement	183
4.8 Analog Signal Processing	183
Beam Geometry and Pulse Characteristics.....	183
Electronics 102: Real Live Circuits.....	184
Circuits: Current Sources and Loads.....	184
Ground Rules.....	185
Couplings, Casual and Otherwise; Transformers	186
Power Supplies	187
Active Electronics: Tubes, Transistors, ICs	188
Analog Nirvana: Operational Amplifiers.....	189
Detector Preamplifiers and Baseline Restoration	190
Analog Pulse Processing: Front Ends and Triggering.....	191
Peak Detectors	192
Pulse Integral or Area Measurements	194
Pulse Width Measurement Circuits	195
Analog Pulse Processing: The Bottom Line.....	195
Dead Times, Doublets, and Problem Pulses	196
Trigger Happy?	196
Analog Linear, Log and Ratio Circuits	197
Linear Circuits; Fluorescence Compensation	197
Logarithmic Amplifiers and Dynamic Range	199
Twin Peaks: Distributions on Linear and Log Scales.....	200
Falling Off a Log: Log Amps Behaving Badly	201
Limits to Dynamic Range.....	202
Ratio Circuits	204
4.9 Digital Signal Processing	204
Analog-to-Digital Conversion	204
Free Samples? Hold it!.....	205
Quantization: When Are Two Bits Worth a Nickel?.....	205
Analog-to-Digital Converters (ADCs) (and Digital-to-Analog Converters (DACs))	208
Digital Pulse Processing and DSP Chips	209

4.9 Digital Signal Processing (continued)	
The Screwy Decibel System.....	211
Pulse Slicing: Déjà Vu All Over Again.....	212
In Defense of de Fence.....	213
Digitization: Tying it All Together.....	214
4.10 Performance: Precision, Sensitivity, and Accuracy	214
Precision; Coefficient of Variation (CV)	214
Sensitivity I: Minimum Detectable Signal.....	215
Sensitivity II: MESF Units.....	216
Accuracy I: Linearity and Nonlinearity.....	217
Sensitivity III: What's All the Noise About?	217
Sensitivity IV: More Photons Give Better Precision	218
Sensitivity V: Background Effects	218
Sensitivity VI: Electrons Have Statistics, Too.....	218
Source Noise Fluctuations and Performance	219
I Blurred It Through the Baseline	219
Restoration Comedy: The Case of the Disappearing Leukocytes	220
Top 40 Noise Sources.....	221
Sensitivity 007: Q and B (Dye Another Day?).....	221
5. DATA ANALYSIS.....	225
5.1 Goals and Methods in Data Analysis.....	225
Cell Counting.....	225
Characterization of Pure Cell Populations.....	226
Identification of Cells in Mixed Populations	226
Characterization of Cell Subpopulations	226
Data Analysis Hardware and Software Evolve	226
5.2 Computer Systems for Flow Cytometry	227
The Beginning.....	227
The End of the Beginning.....	227
Data Rates and Data Acquisition Systems	228
PC Data Acquisition Boards.....	229
Preprocessors for Data Acquisition	230
5.3 Primary Data: Frequency Distributions	231
You Say You Want a Distribution.....	231
Gauss Out of Uniform	231
About Binomial Theorem, I'm Teeming With a Lot o' News.....	232
Distributions Have Their Moments.....	233
Statistical versus Cytometric Parameters	233
Mean, Variance, and Standard Deviation	233
With Many Cheerful Facts About the Square of the Hypotenuse: Euclidean Distance.....	234
Higher Moments; Skewness and Kurtosis.....	234
Some Features of the Normal Distribution	234
Measures of Central Tendency: Arithmetic and Geometric Means, Median, and Mode	235
Measures of Dispersion: Variance, Standard Deviation, CV, and Interquartile Range	235
Robustness in Statistics; the Robust CV.....	235
"Box-and-Whiskers" Plots of Distributions	236
Calculating and Displaying Histograms	236
Bivariate and Multivariate Distributions and Displays.....	237
Dot Plots; Correlation and Covariance.....	237
Linear Regression; Least-Squares Fits.....	237
Breaking off Undiplomatic Correlations	238
Multivariate Measures of Central Tendency and Dispersion	238
Beyond Dot Plots: Two-Parameter Histograms	238
Bivariate Distributions: Display's The Thing!.....	238
Displaying by the Numbers.....	239

5.3 Primary Data: Frequency Distributions (continued)	
Economies of Scale	240
How Green Were My (Peaks and) Valleys	241
Clouds on the Horizon: 3-Dimensional Displays	241
5.4 Compensating Without Decompensating	242
5.5 Dealing With the Data	244
Comparing and Analyzing Univariate Histograms	244
The K-S Test, Clonal Excess, and χ^2 Tests	245
"Nonparametric" Histogram Comparison	245
Cumulative (Overton) Subtraction	245
Constant CV Analysis	245
Another Approach to Histogram Comparison	246
Deconvoluting Single-Parameter Histograms	246
Analysis of Two-Parameter Data	246
Two-Parameter Gating, Bitmap and Otherwise	246
Rectangles and Quadrilaterals: Hardware and Software Gating:	246
Ellipses and Beyond: Advantages of Bitmaps	246
Storage Requirements for Bitmaps	247
Analysis of Two-Parameter Distributions	247
See You Around the Quad	247
Bivariate Cell Kinetic Analysis	247
Bivariate Karyotype Analysis	247
Smooth(ing) Operators: When are Filters Cool?	248
Bivariate Analysis in Hematology Analyzers	248
5.6 Multiparameter Data Analysis	248
Multiparameter versus Multivariate Analysis	248
Multiparameter Analysis of Leukocyte Types: 1974	248
Automated Differentials via Discriminants	249
Interactive Analysis: Finding the Training Set	249
Multiparameter Analysis of Leukocyte Types: 2002	250
Procedures for Automated Classification	250
Discriminant Functions and How They Work	250
Principal Component Analysis	252
Cluster Analysis	252
Neural Network Analysis	252
Genetic Algorithms	253
5.7 Analysis of Collected Data: How Much Is Enough/Too Much?	253
5.8 Data Analysis Odds and Ends	254
Data Storage	254
The Flow Cytometry Standard (FCS) File Format	254
Magnetic/ Optical Tumors in the Digital Attic	255
Linear and Log Scales and Ratios: Proceed with Care!	255
Ratios Only Help if Variables are Well Correlated	256
6. FLOW SORTING	257
6.1 Sort Control (Decision) Logic	257
6.2 Preselected Count Circuits and Single Cell Sorting	258
6.3 Droplet Sorting, High-Speed and Low	258
Droplet Generation	259
Drop Charging and Deflection	260
Drop Deflection Test Patterns	260
Two- and Four-Way Sorts: How Much Voltage?	260
How Many Drops Should be Charged?	261
Guilty as Charged?	261
Determining Droplet Delay Settings	262
Fractional Droplet Delays	262

6.3 Droplet Sorting (continued)	
Transducers and Transducer Drive Signals.....	263
Improving Droplet Sorting	263
Sorting Large Objects with Droplet Sorters.....	263
6.4 Fluidic Switching Cell Sorters.....	264
Sorting Large Objects Using Fluidic Switching	265
Sorting Very Small Objects: Microfluidic Switching	266
6.5 Cell Manipulation By Optical Trapping	266
6.6 Cell Damage Cell Selection ("Cell Zapping").....	266
Photodamage Cell Selection.....	266
Sorting (Zapping) Without Flow (Gasp!).....	267
Electrodamage Cell Selection in Flow	267
6.7 Measures of Cell Sorter Performance: Purity, Recovery (Yield), and Efficiency	267
Coincidence Effects on Performance.....	267
6.8 Other Considerations.....	268
Doing the Math.....	268
Speed Limits: The Reynolds Rap	269
Instrument Utilization	269
Monitoring versus Sorting for Cell Preparation	269
Collection Techniques: Life and Death Decisions	269
Dilutions of Grandeur	270
Can Getting Sorted Be Hazardous to Cells' Health?.....	270
6.9 Biohazard Control and Biosafety in Flow Cytometers and Sorters.....	271
6.10 Conclusions.....	271
7. PARAMETERS AND PROBES.....	273
7.1 Physical Parameters and Their Uses.....	273
Electrical Parameters.....	273
DC Impedance (Coulter Volume).....	273
AC Impedance (Electrical Opacity); Capacitance	273
Acoustic Measurements of Cells in Flow	274
Optical Parameters: Light Scattering	274
Scattering: The Mueller Matrix Model	274
Forward Light Scattering and Cell Size.....	275
Forward Scatter and "Viability"	276
Side Scatter and Cytoplasmic Granularity.....	276
Lymphocyte Gating: Forward Scatter Aside.....	276
Other Applications of Side Scatter.....	277
What is the Right Angle for "Right Angle" Scatter?	278
Does Side Scatter = Total Protein?.....	278
Optimizing Side Scatter: Not as Easy as It Looks.....	270
Polarized 90° Scatter Reveal Eosinophils and Malaria	
Pigment-Containing Monocytes.....	278
Multiple Wavelength Scattering Measurements	279
From Russia with Lobes	279
Optical Parameters: Absorption	281
Absorption Effects on Light Scattering.....	281
Optical Parameters: Extinction	282
Other Transmitted Light Measurements	282
Interference and Phase Measurements	282
Optical Parameters: Fluorescence	283
Fluorescence Lifetime Measurements.....	283
Fluorescence Polarization Measurements	283
Energy Transfer Measurements: Something to FRET About	283
Quenching and Energy Transfer.....	284
Measuring Fluorescence Spectra in Flow	284

Optical Parameters: Fluorescence	
Two-Photon Fluorescence Excitation in Flow	284
Bioluminescence detection in Flow?	284
7.2 Intrinsic Cellular Parameters.....	285
Cell Size	285
Mean Cell Volume: The Cellocrit as Gold Standard.....	285
Cell Volume, Area, and Diameter.....	285
Cell Sizing: Slit Scans and Pulse Widths.....	285
Size Measurements in the Submicron Range.....	288
Other Size Measurement Techniques	289
Cell Shape and Doublet Discrimination.....	289
Measurement of Intrinsic Parameters Using Absorption and Extinction Signals	290
Fluorescence Measurement of Intrinsic Parameters.....	290
Autofluorescence: Pyridine and Flavin Nucleotides.....	290
Pyridine and Flavin Nucleotides and Redox State.....	291
Pyridine and Flavin Nucleotides and Cancer	291
Bacterial Autofluorescence Measurements.....	291
Bacterial Autofluorescence Measurements.....	291
Porphyrin Fluorescence in Erythroid Cells.....	292
Other Pigments.....	292
Chlorophyll and Phycobiliproteins.....	292
Infrared Spectra and Cancer Diagnosis	293
7.3 Probes, Labels, and [Not] Protocols for Extrinsic Parameter Measurements.....	293
Probes, Labels, and Dyes.....	293
Dyes and Quality Control: Gorillas in the NIST.....	294
The Dyes are Cast: An Overview.....	295
Mechanisms of Staining by Fluorescent Dyes.....	298
Environmental Sensitivity.....	298
Metachromasia	298
Spectral Changes and Ratiometric Measurements	299
Internal Energy Transfer in Probes and Labels.....	299
“Vital” Staining.....	299
Staining and “Viability”	299
“Intact” versus “Live” Cells.....	299
Getting Dyes Into – and Out of – Intact Cells.....	300
Permeancy and Permeability.....	300
Vital Dye Toxicity and Photosensitization of Cells	301
Fixation – Why and How	302
Fixation for Biohazard Control.....	302
Fixation Mechanisms.....	302
Permeabilization versus Fixation.....	302
Fixative Effects on Scatter Signals	303
Fixation for Surface Antigen Measurements.....	303
Fixatives: Coming Out of Aldehyding Places	304
Fixation for Intracellular Antigen Measurements	305
Fixation for DNA Content Determination: Getting DNA (and Antigens) Out of a Tight Fix	305
Catch the Wave: Fixation by the (Cook) Book	305
Fixation Artifacts.....	305
Red Blood Cell Lysis: The Distilled Essence.....	306
7.4 Nucleic Acid Dyes and Their Uses	306
DNA Content Measurement.....	306
Feulgen Staining for DNA Content.....	306
DNA Staining with Ethidium and Propidium	306
Ethidium & Propidium: Ionic Strength Effects.....	307
DNA Content: Sample Preparation and Standards	307
Chromomycin A ₃ , Mithramycin, and Olivomycin.....	307

Mithramycin Plus Ethidium: Do's and Don'ts.....	308
DNA Content Measurement (continued)	
The Hoechst Dyes (33258, 33342, 34580?).....	308
Detecting BrUdR by Hoechst Dye Quenching.....	308
Hoechst Dyes Have an A-T Base Preference	308
Hoechst Dyes In – And Out Of – Living Cells: The Drug Efflux Pump Discovered.....	309
Hoechst Dye Staining Mechanisms.....	309
Hoechst 34580: Violet Time?.....	310
DAPI (and DIPI): Dyes Known for Precision.....	310
Determinants of High Precision in DNA Analysis.....	311
7-Aminoactinomycin D (7-AAD).....	311
Acridine Orange.....	312
Styryl Dyes; LDS-751	312
Cyanine Dyes I: Thiazole Orange, etc.	312
Cyanine Dyes II: TOTO and YOYO à GoGo.....	314
Cyanine Dyes III: Alphabet Soup.....	315
Seeing Red: LD700, Oxazine 750, Rhodamine 800, TO-PRO3, and DRAQ5 as DNA Stains.....	315
Miscellaneous DNA-Selective Dyes	316
What Do DNA Stains Stain?.....	316
DNA Ploidy and Aneuploidy: The DNA Index.....	317
Sample Preparation for DNA Content Analysis.....	317
DNA Base Composition	317
Chromatin Structure; Identifying Cells in Mitosis.....	319
Chromatin Structure Identifies Mitotic Cells.....	320
RNA Content.....	320
RNA/DNA Staining with Acridine Orange (AO).....	320
Cell Cycle Compartments Defined on the Basis of RNA and DNA Content.....	320
AO: Problems and Some Solutions.....	321
Pyronin Y, Oxazine 1, and Other Tricyclic Heteroaromatic Dyes as RNA Stains	322
What Does Pyronin Y Stain? Double-Stranded (Ribosomal) RNA and Sometimes Mitochondria	323
Surviving Vital Staining with Pyronin Y.....	324
Other DNA Dyes Usable with Pyronin Y.....	324
Tips on Tricyclics (Don't Get Depressed)	325
Tricyclics Gag on Mucopolysaccharides	325
Reticulocyte Counting: Cyanines Beat Tricyclics; RNA in Nucleated Cells: Cyanines Don't.....	326
Propidium Stains Double-Stranded RNA: What of Other Dyes?.....	326
7.5 Fluorescent Labels and Protein Dyes.....	326
Estimating Total and Basic Protein Content of Cells	327
Fluorescein Isothiocyanate (FITC)	327
Sulforhodamine 101 (SR101).....	327
Hematoporphyrin (HP) as a Protein Stain.....	328
Rhodamine 101 (or 640) as a Vital Protein Stain	328
Staining to Demonstrate Basic Protein	328
Covalent Labels for Antibodies and Other Molecules.....	328
Fluorescein Isothiocyanate (FITC) as a Label	329
Labeling with Lissamine Rhodamine B and Tetramethylrhodamine Isothiocyanate (TRITC).....	329
Multicolor Fluorescence I: FITC and TRITC	329
Multicolor Fluorescence II: Rhodamine 101 Dyes.....	330
Early Problems with Multicolor Fluorescence.....	331
Phycobiliproteins to the Rescue!.....	331
Phycocerythrins: R-PE, B-PE and Others.....	332
Allophycocyanin (APC) and APC-B	332
Phycocyanins	332
Phycobiliprotein Tandem Conjugates: PE-APC, PE-Texas Red, PE-Cy5, PE-Cy5.5, PE-Cy7, etc.....	333
Allophycocyanin Tandem Conjugates: APC-Cy7 and APC-Cy5.5.....	333

Phycobiliproteins to the Rescue! (continued)	
Mercy Me! PerCP!	333
Phycobiliproteins and Tandems: Dirty Little Secrets	334
Future Tandems: Heterocycles Built for Two?	335
Cyanine Dye Labels: From Cy-Fi to Hi5 for Cy5	336
Blue Notes: AMCA and Cascade Blue	337
Hey, BODIPY!	337
Alexa Dyes: Some Thoughts on Dyemographics	338
Other Organic Fluorescent Labels: A Dye Named Joe, etc.	338
Quantum Dots	339
Getting Labels Onto Molecules of Interest	340
7.6 Improving Signals from Labels: Amplification and Other Techniques	340
Limits to Sensitivity: Autofluorescence	341
Improving Sensitivity: The New Wave(length)	342
Correcting and Quenching Autofluorescence	342
Raman Scattering Effects on Sensitivity	342
Increasing Sensitivity: Amplification Techniques	343
Amplification by Indirect Staining	343
Amplification Using Labeled Particles	343
Amplification Using Enzymes as Labels: Playing the Hole CARD	344
Amplifying the Analyte: The Polymerase Chain Reaction (PCR)	344
Amplification Techniques: Pros and Cons; Fluorescent versus Nonfluorescent Labels	344
Improving Sensitivity: Time-Resolved Fluorescence	345
7.7 Measuring Cell Surface and Intracellular Antigens	345
History and Background	345
Monoclonal Antibodies for the Uninitiated	346
Cell Surface Antigens: Structure versus Function	347
Moving Toward Multicolor Immunofluorescence	347
Antibody Reagents and Staining Procedures	348
Antibody Fragments versus Antibodies as Reagents	348
Engineered Antibodies: Phage Display and scFvs	348
Molecular Probes' Zenon Antibody Labeling	348
Antibody Shelf Life and Quality Control	349
Direct Staining Using Monoclonal Antibodies for 2, 3, and More Colors Using 488 nm Excitation	349
Multicolor Work Using Multiple Lasers; Biotin-Avidin Labeling	349
Mixing Colors: Do's and Don'ts	349
Cocktails for Five: Multiplex Immunofluorescence	350
Cocktail Staining Helps Identify Rare Cells	352
Immunofluorescence Staining Procedures	352
Automated Sample Preparation	353
Fluorescence Measurements: Lurching Toward Quantitation	353
Calibration and Controls: Round One	353
Quantitative Fluorescence Cytometry: Definitions	354
Calibration Particles for QFCM	354
Defining a Window of Analysis	356
Type IIA versus Type IIIB and IIIC Standards	357
Other Aspects of Fluorescence Quantitation	358
What is "Positive"? What is "Negative"?	358
Making Weakly Fluorescent Beads and Cells: Do Try This Trick at Home!	358
Correlating Cytometry and Biochemistry: Studies of Antibody Binding Chemistry	359
Correlating Cytometry and Biochemistry: Intracellular Antigen Measurements	359
Analyzing Immunofluorescence Data	360
Estimating Antigen or Receptor Surface Density	360
Quantitative Fluorescence: Problems and Prospects	361
7.8 Nucleic Acid Sequence Detection	361
Peptide Nucleic Acid (PNA) Probes	362

7.9 Probes for Various Cell Constituents	362
Surface Sugars (Lectin Binding Sites)	362
Analysis of Total Carbohydrate Content	363
Specific Detection of Cellulose	363
A Probe for Cell Surface Aldehydes	363
Probes for Lipids and Cholesterol	364
Nile Red	364
Filipin	364
Lipid Droplet Detection Using Scatter Signals	364
Probes for Cytoskeletal Organization/Actins	364
7.10 Time as a Parameter: Kinetic Measurements	364
Sample Handling for Kinetic Measurements	365
Time as a Quality Control Parameter	366
Sloooooow Floooooowww	366
7.11 Labeled Ligand Binding	366
Labeling Strategies	367
Formal Analysis of Ligand binding	367
Labeled Ligands versus Anti-Receptor Antibodies	368
Ligand Binding Detected by Functional Changes	368
Fluorescent Ligand Binding: Some Examples	368
7.12 Functional Parameters I	369
Cell Surface Charge	369
Cell Membrane Characteristics	369
Membrane Integrity versus "Viability": Dye Exclusion Tests	369
Detecting "Dead" Cells in Fixed Samples	369
Membrane Fusion and Turnover; Cell Tracking	371
Cell Proliferation Analyzed Using Tracking Dyes	371
Membrane Organization and Fluidity/Viscosity	374
Lipid Packing Assessed with Merocyanine 540	374
Membrane Fluidity and Microviscosity: Assessment Using Fluorescence Polarization	374
Lipid Peroxidation	375
Membrane Permeability to Dyes and Drugs: The Drug Efflux Pump Revisited	376
Endocytosis of Macromolecules and Particles	377
Enzyme Activity	378
Indicators of Oxidative Metabolism I: Tetrazolium Dye Reduction	379
Indicators of Oxidative Metabolism II: 2,7-Dichlorofluorescein Diacetate (DCFH-DA), etc.	379
Indicators of Oxidative Metabolism III: Hydroethidine (Dihydroethidium)	379
Indicators of Oxidative Metabolism IV: Dihydrorhodamine 123	379
Indicators of Oxidative Metabolism V: Detection of Hypoxic Cells	380
Detection of Caspase Activity	380
Other Enzymes	380
Detection of Enzymes and Products by Antibodies	380
Enzyme Kinetics in Single Cells	380
Sulfhydryl (Thiol) Groups; Glutathione	381
7.13 Functional Probes II: Indicators of Cell Activation	381
Introduction	381
Changes in the Cellular Ionic Environment Following Activation by	
Ligand Interaction with Cell Surface Receptors	382
"Structuredness of Cytoplasmic Matrix" (SCM) and the Cercek Test for Cancer	383
Optical Probes of Cell Membrane Potential	385
Membrane Potential and Its Physicochemical Bases	385
$\Delta\psi$ Measurement Using Microelectrodes	386
Single Cell Measurements with Distributional Probes	387
Oxonol Dyes as Membrane Potential Probes	390
Possible Alternatives to Distributional Probes for Flow Cytometry of Membrane Potentials	391
Ratiometric Probes for Membrane Potential	391

Optical Probes of Cell Membrane Potential (continued)	
Using Cyanine Dyes for Flow Cytometric $\Delta\Psi$ Estimation, In Case You're Still Interested	392
Cytoplasmic Membrane Potential: Summing Up	394
Mitochondrial Membrane Potential ($\Delta\Psi_m$)	394
Mitochondrial Staining with Rhodamine 123 Is Membrane Potential-Dependent	394
The Search for Better $\Delta\Psi_m$ Probes: Round One.....	397
$\Delta\Psi_m$, JC-1, and Apoptosis.....	397
The Search for Better $\Delta\Psi_m$ Probes: Round Two	398
Bacterial Membrane Potentials	400
Ratiometric $\Delta\Psi$ Measurement in Bacteria	401
$\Delta\Psi$ Measurement: Cautions and Conclusions.....	402
Optical Probes of Intracellular Calcium	402
The Bad Old Days	402
Chlortetracycline as a Probe of "Membrane-Bound" Calcium in Cells.....	402
Probes for Free Cytoplasmic Calcium: Quin-2	403
Fura-2 and Indo-1: Ratiometric Ca^{++} Indicators.....	403
Fluo-3 and Other Visible-Excited Ca^{++} Probes	404
Flow Cytometric Probes of Intracellular pH.....	405
The Hat Trick: Multiparameter Approaches to Ion Flux Measurements in Cell Activation	407
NOsing Around for Nitric Oxide.....	408
Other Ions in the Fire	408
7.14 Reporter Genes.....	408
Somebody Cloned My Gal: Enzymes as Reporter Genes.....	408
Green Fluorescent Protein (GFP) et al	409
Minority Report(er)?.....	410
8. BUYING FLOW CYTOMETERS	411
8.1 Introduction	411
8.2 History	411
8.3 BD Biosciences	412
Background	412
The BD FACS Vantage SE™ Cell Sorter.....	413
The BD FACSCalibur™ Analyzer	414
The B-D™ LSR II Analyzer	416
The B-D FACSAria™ Cell Sorter.....	417
The B-D FACSCount.....	418
8.4 Beckman Coulter, Inc.	418
Background – and Signal-to-Background.....	418
The Beckman Coulter EPICS® ALTRA™ Cell Sorter	419
The Beckman Coulter Cytomics™ FC 500 Analyzer.....	420
The EPICS® XL and XL-MCL Analyzers.....	422
8.5 DakoCytomation	423
Background	423
The MoFlo® Cell Sorter.....	423
The CyAn™ Flow Cytometer.....	424
8.6 Cytopeia	425
The InFlux Cell Sorter.....	425
8.7 Optoflow AS.....	426
Background	426
The MICROCYTE® Flow Cytometer	426
8.8 Partec GmbH.....	427
Background	427
The CyFlow® and CyFlow® ML Flow Cytometers	427
The PAS, PAS II, and PAS III Flow Cytometers	428
PA Ploidy Analyzer and CCA Cell Counter Analyzer	429

8. BUYING FLOW CYTOMETERS (continued)	
8.9 Some Other Flow Cytometer Companies	429
Advanced Analytical Technologies, Inc. (AATI)	429
Agilent Technologies, Inc.	429
Apogee Flow Systems Ltd.	430
Bentley Instruments	430
Chemunex SA	430
CytoBuoy b.v.	430
Delta Instruments bv	431
Fluid Imaging Technologies, Inc.	431
FOSS Electric A/S	431
Guava Technologies, Inc.	431
Howard M. Shapiro, M.D., P.C.	431
iCyt – Visionary Bioscience	431
International Remote Imaging Systems	431
Luminex Corporation	431
NPE Systems, Inc.	432
Union Biometrica, Inc.	432
8.10. Hematology Instruments, Etc.	433
8.11 Little Orphan Analyzers (And Big Orphan Sorters)	434
Bio/Physics and Ortho: Cytofluorograf to Cytoron	434
HEKA Elektronik GMBH: The FLUVO II Analyzer	435
The Kratel Partograph	435
The ODAM ATC 3000	435
Also Among the Missing	436
Flow Cytometer Rehabilitation; Used Instruments	436
Following Suit	436
8.12 Third Party Software	437
8.13 The Selling of Flow Cytometers: Hype and Reality	437
8.14. Applying for a Grant for a Cytometer	438
9. BUILDING FLOW CYTOMETERS	441
9.1 Why Buy a Flow Cytometer?	441
9.2 Why Build a Flow Cytometer?	441
9.3 Learning to Build Your Own	442
10. USING FLOW CYTOMETERS: APPLICATIONS, EXTENSIONS, AND ALTERNATIVES	443
10.1 The Daily Grind	443
Keeping the Instrument Running: Diet and Exercise	443
Particulars: Drawing a Bead on Flow Cytometer Alignment, Calibration, and Standardization	444
Alignment Particles: Fearful Asymmetry	445
Reference and Calibration Particles	445
Compensation Standards	446
Cells and Nuclei as Alignment Particles	446
Rose Colored Glasses: Optical Filter Selection	446
Experimental Controls	447
Shake Well Before Using: When Controls Won't Help	447
10.2 Significant Events in the Lives of Cells	448
Taking the Census: Cell Counting	448
A Counting Alternative: Image Analysis	448
The Doubled Helix: Reproduction	448
The Cell Cycle and Cell Growth	448
DNA Content Analysis	448
Mathematical Models for DNA Analysis	449
Clinical Application of DNA Content Analysis	450

DNA Content Analysis(continued)	
DNA Content Alternatives: Static Photometry and Scanning Laser Cytometry	451
The Mummy's GC/AT: DNA Content Analysis in Anthropology and Forensic Science	451
Half a Genome is Better Than None: Sperm Sorting	452
The Widening G_0/G_1 re Detecting Mutation	453
Detecting DNA Synthesis: Cell Kinetics	453
Kinetics Before Flow Cytometry: Mitotic Indices, Doubling Times, and Radiolabel Studies	453
Labeling Index versus DNA Content	454
Early Flow Cytometric Approaches to Labeling Using BrUdR and ^3H -TdR	455
Detection of Incorporated BrUdR with Hoechst Dyes and Propidium Iodide	455
Detection of BrUdR Incorporation with Anti-BrUdR Antibodies	455
Cytochemical Detection of BrUdR Incorporation Using Difference and Ratio Signals	456
Breaking Up Is Easy To Do: SBIP, a Simpler Way to Detect BrUdR Incorporation into DNA	456
Anti-BrUdR Antibody: Seeing the Light	457
Cytochemical Detection of BrUdR: Still Around	457
Detecting RNA Synthesis Using Bromouridine	458
Generation Gaps: Tracking Dyes and Cell Kinetics	458
Cell-Cycle Related Proteins: Cyclins, Etc.	458
Detecting Mitotic Cells	462
Memento Mori: Detecting Cell Death	462
Necrosis versus Apoptosis	462
Identifying Apoptotic Cells	462
Les Feuilles Mortes (Autumn Leaves)	462
Apoptosis: Getting With the Program	463
ISNT there Light at the End of the TUNEL?	463
Apoptosis: The Case Against Flow Cytometry	463
Die Another Day: Cytometry of Telomeres	464
10.3 Identification of Cells in Mixed Populations	464
Mixed Genotypes versus Mixed Phenotypes	464
No Parameter Identifies Cancer Cells	464
Many Parameters Identify Blood Cells	464
Flow Cytometric Parameters Useful for Blood Cells	464
Specific Gene Products Identify Cell Types	465
Maturation Processes and "Missing Links": The "Ginger Root" Model	465
Practical Multiparameter Gating: Color Wars	467
Finding Rare Cells	469
One Parameter is Not Enough	469
Cocktail Staining Can Help	470
Dirt, Noise, and Rare Event Detection	470
Really, Really Rare Events: Alternatives to Flow	470
10.4 Tricks and Twists: Odd Jobs for Flow Cytometry	471
Single Molecule Detection	471
DNA Sizing, if not Sequencing, in Flow	471
Solid Phase (Bead) Assays Using Flow Cytometry	473
Cocktails for 100: Multiplexed Bead Assays	473
Cells in Gel Microdroplets and on Microspheres	474
Hanging Ten Pseudopodia?	475
10.5 Single Cell Analysis: When Flow Won't Do	475
10.6 Applications of Flow Cytometry	476
Cell Differentiation, <i>Ab Ovo</i> and <i>De Novo</i>	476
Differentiation in the Nervous System	476
Whole Embryo Sorting	476
Somatic Cell Genetics and Cell Hybridization	477
Reporter Genes Revisited	477
Isolating and Characterizing Hybrid Cells	477
Chromosome Analysis and Sorting and Flow Karyotyping	477

10.6 Applications of Flow Cytometry (continued)

Probing Details of Cellular Structures and Inter- and Intramolecular Interactions	479
Dissection of Structures Using Antibodies, Ligands, and Genetic Methods	479
Intramolecular Interactions	479
Clinical Flow Cytometry: Turf and Surf	480
Hematology	480
Clinical Application: Blood Cell Counting and Sizing	480
Red Blood Cells (Erythrocytes)	481
Clinical Application: Reticulocyte Counts	481
The Reticulocyte Maturity Index (RMI)	482
Erythrocyte Flow Cytometry: Other Clinical Uses	482
White Blood Cells (Leukocytes)	483
Clinical Application: Differential Leukocyte Counting	483
CD Characters: Leukocyte Differentiation Antigens	484
Granulocytes: Basophils	484
Basic Orange 21: The Best Basophil Stain Yet	485
Allergy Tests Using Basophil Degranulation	485
Granulocytes: Eosinophils	486
Granulocytes: Neutrophils	486
(Clinical?) Tests of Neutrophil Function	486
Neutrophil CD64 in Inflammation and Sepsis	487
When I'm Sick - CD64?	487
Platelets and Megakaryocytes	487
Hematopoietic Stem Cells	488
Clinical Application: Monitoring CD34+ Stem Cells According to the ISHAGE Protocol	488
Side Population (SP) Stem Cells: Plastic, Fantastic	489
Immunology	489
Immunologic Applications of Flow Cytometry: Still a Growth Industry	489
HIV Infection – The “Killer Application”	490
Clinical Application: T Cell Subset Analysis	490
T Cell Subsets: Alternative Technologies	491
Clinical Application: Transplantation	493
Detecting Lymphocyte Activation	494
Foundations: From PHA (the Lectin) to PHA (the Pulse Height Analyzer)	495
Functional Probes for Activation	495
DNA, RNA, and Activation Antigens	497
Mitogen Response versus Antigen Response	498
Detecting Activation by CD69 Expression	499
Cytokines: Detecting Activation and More	499
Ins and Outs of Cytokine Staining	500
Tetramer Staining: Talking the Talk; Walking the Walk?	500
Tracking Dyes: Activation and Ontogeny	501
What is “Early” Activation (Trick Question)?	501
Cancer Biology and Clinical Oncology	502
Cancer Diagnosis: Cervical Cytology	503
DNA Content Measurements Yet Again	503
Beyond DNA Content: Antigens, Oncogenes, and Receptors, and Response to Therapy	504
Immunophenotyping in Hematopathology	504
Detecting Minimal Residual Disease	505
Biological Implications of Phenotyping Results	506
Digression: A Slight Case of Cancer	507
Analysis of Sperm	508
Isolating Fetal Cells from the Maternal Circulation for Prenatal Diagnosis	509
The March of Time: Circadian Rhythms, Aging, and Atherosclerosis	510
Clinical Application: Urine Analysis	510
The Animal Kingdom	510

The Animal Kingdom (continued)	
Lions and Pumas and Clams, Oh, My!	510
Fish Story; FISH Story	511
Flow Cytometry: For the Birds?	512
Big Stuff, Vegetable, Animal or Mineral	512
Flow Cytometry of Plant Cells and Chromosomes	512
Measurement of Plant Cell DNA Content	513
Plant Chromosome Analysis and Sorting	514
Other Flow Cytometric Applications in Plants	514
Microbiology, Parasitology and Marine Biology	514
Measuring Microbes: Motivation	515
Measuring Microbes: Instrument Issues	515
Parameters Measured in Microorganisms	516
Flow Cytometric "Gram Stains"	516
Detection and Sizing: Light Scattering	517
Detection and Sizing: Electrical Impedance	517
Nucleic Acid (DNA and RNA) Staining	517
Total Protein Content: Scatter versus Stains	518
Antibodies, Etc.: Labeling Strategies	518
Ribosomal RNA-Based Species Identification	518
Functional Probes in Bacteria	518
Potential, Permeability, "Viability", and Metabolic Activity	519
Digression: A Therapeutic Approach Based on Transient Permeabilization	522
So Few Molecules; So Little Time	522
Applications in Marine Microbiology	524
Extensions: Cytometers for Marine Applications	527
References: Flow Cytometry and Oceanography	527
General Microbiology	527
Previously Noted	527
Cell Cycles and Cell Division	528
Fluorescent Protein Methods in Microbes	528
Microbial Communities: Will Flow Work?	528
Bad Guys Don't All Wear Black Hats: Microbial Detection/Identification	
in Health-Related Contexts	528
The Basic Questions	528
Detection: Intrinsic Parameters are Not Enough	529
Detection: Fluorescence Improves Accuracy	529
Detection: When the Tough Get Going	529
Identification: Too Many Broths... ..	530
Identification: Can Multiplexing Help?	531
Environmental and Sanitary Microbiology	531
Water That Made Milwaukee (and Sydney) Infamous	531
Food Microbiology	532
Bioterrorism and Bioopportunism	532
Viruses and Other Intracellular Pathogens	532
Clinical Microbiology	533
Antimicrobial Susceptibility Testing: One Size Does Not Fit All	534
Bacteria: Confusion Reigns	535
Mycobacteria: Down for the Count	535
Antifungal Susceptibility: Flow Does the Job	535
Antiviral Susceptibility by Flow Cytometry	536
Cytometry in Vaccine Development	536
Microbiology Odds and Ends	536
Parasitology	536
Pharmacology and Toxicology	537
Drugs and the Life and Death of Cells	538

Pharmacology and Toxicology (continued)	
Erythrocyte Micronucleus Assays.....	538
Toxic Waste and B Cell Proliferation	538
Radiation Dosimetry	538
Food Science	538
Somatic Cell Counts in Milk.....	538
Brewhaha	539
A Loaf of Bread, A Jug of Wine... ..	539
Seeing the Blight	539
Major Food Group: Chocolate.....	539
Biotechniques and Biotechnology	539
Protein and Gene Expression on Cells and Beads	540
Getting Big Molecules into Small Cells	540
Staying Alive, Staying Alive	540
Et Cetera	540
Alternatives: Microfluidic Cytometers, Flow and Static.....	541
Cytometry Afield	541
The Lymphocytes of the Long Distance Runner.....	541
War and Peace	541
Blood, Sweat and Tears?.....	542
Pulp Nonfiction.....	542
Flow Cytometry On the Rocks.....	542
To Boldly Go Where No Cytometer Has Gone Before.....	542
II. SOURCES OF SUPPLY.....	543
II.1 Resources, Societies, Journals	543
II.2 Optical Supply Houses.....	543
II.3 Probes and Reagents	544
II.4 Calibration Particles/Cytometry Controls.....	548
II.5 Flow Cytometers.....	549
Hematology Instruments	551
II.6 Data Analysis Software/Systems	551
Hardware and Software.....	551
Commercial Software Sources	552
Noncommercial Software Sources.....	552
II.7 Cytometer Rehabilitation/Add-ons.....	553
II.8 Flow Cytometer Parts.....	553
Flow System Plumbing.....	553
Photodetectors.....	554
DC-DC Converter Modules for HV Power Supplies	555
Power Supplies (Low Voltage)	555
Other Electronics.....	555
II.9 Lasers	555
Laser Trade Publications.....	555
Laser Manufacturers.....	555
II.10 Optical Filters.....	556
Color Glass Filters	556
Interference Filters	557
Neutral Density Filters.....	557
Polarizing Filters and Optics.....	557
Tunable Filters.....	557
II.11 Aids to Troubleshooting Flow Cytometers When All Else Fails	557
II.12 Proficiency Testing.....	557
II.13 Sex Selection	557
II.14 Alternative Technology.....	558

12. AFTERWORD	561
12.1 Dotting i's and Crossing t's	561
12.2 Late Breaking News	561
New Book	561
New Protein Stain	561
Caveat on Fluorescent Caspase Inhibitors	561
Polyamide Probes	561
Tearing Down the (Picket) Fences	562
New Instrument: The BD FACSAarray™	563
<i>Science</i> Special Section: Biological Imaging	563
Cytomics in Predictive Medicine: a <i>Clinical Cytometry</i> Special Issue and Other	
Recent Citings and Sightings	563
12.3 Analytical Biology, Such as It Isn't: Is This Any Way to Run a Science?	563
12.4 Colophon	564
12.5 Unfinished Business	565
AIDS and Infectious Disease in the Third World	565
A Center for Microbial Cytometry	565
A Nobel Prize for Herzenberg and Kamensky?	566
12.6 Flow and the Human Condition	566
There's No Business Like Flow Business	566
12.7 One More Thing	566

TABLES AND FIGURES

TABLES

1-1. Some parameters measurable by cytometry	3
3-1. A brief outline of flow cytometric history (1945-2000)	100
4-1. SI units and prefixes.....	102
4-2. Emission wavelengths of lasers	139
4-3. Cathode quantum efficiencies of diode and PMT detectors between 300 and 800 nm.....	165
4-4. Logarithmic Amplifiers: What goes in, what comes out.....	203
4-5. Characteristics of analog-to-digital converters.....	206
5-1. Some landmarks of the normal or Gaussian distribution	234
5-2. Equations (1-4) that must be solved to permit 4-color fluorescence compensation	242
5-3. Header and text portion of an FCS2.0 data file.....	254
6-1. Safe speed limits for sorting based on Reynolds number calculations.....	269
7-1. Some cellular parameters measurable by cytometry	286
7-2. Fluorescence spectral properties of a selection of reagents usable for common cytometric tasks.....	297
7-3. Tricyclic heteroaromatic compounds usable for staining DNA and/or RNA	323
7-4. Raman emission from water.....	343

FIGURES

1-1. Interaction of light with a cell	4
1-2. Transmitted light and dark field images of an unstained suspension of human peripheral blood leukocytes	7
1-3. Transmitted light microscope images of an unstained smear of human peripheral blood	8
1-4. Schematic of a fluorescence microscope.....	9
1-5. Scanned images of Feulgen-stained lymphoblastoid cells	15
1-6. One-dimensional scanning of cells deposited in a narrow line	16
1-7. Idealized plot of signal amplitude vs. time.....	16
1-8. Ideal and "real" DNA content distributions	22
1-9. Single parameter histogram displays from a multichannel pulse height analyzer.....	24
1-10. Use of a mathematical model to determine fractions of DNA-aneuploid breast cancer cells.....	25
1-11. Dot plot (cytogram) of Hoechst 33342 and fluorescein fluorescence in CCRF-CEM cells	26
1-12. Gating regions for counting or sorting set electronically, drawn on an oscilloscope display	27
1-13. Histogram of 90° (side) scatter from leukocytes in lysed whole blood.....	30
1-14. Bivariate distribution of anti-CD3 antibody fluorescence intensity vs. large angle scatter in leukocytes.....	31
1-15. The bivariate distribution of Figure 1-14 shown as an isometric or "peak-and-valley" plot	32
1-16. The two-parameter histogram of Figures 1-14 and 1-15 displayed as a contour plot.....	32
1-17. Identification of human peripheral blood T-lymphocytes bearing CD4 and CD8 antigens	34
1-18. Why fluorescence compensation is necessary	37
1-19. How compensation gets data to fit into quadrants.....	38
1-20. Fluorescence intensities of antibody stained cells and beads bearing known amounts of antibody.....	49
1-21. Schematic of the optical system of a fluorescence flow cytometer.....	51
1-22. A typical flow chamber design.....	56
1-23. FACScan Analyzer (Becton-Dickinson)	58
1-24. MoFlo High Speed Sorter (Cytomation).....	58
1-25. Microcyte Cytometer (Optoflow)	58
1-26. Estimated numbers of fluorescence flow cytometers in use worldwide, 1975-92.....	59
2-1. Growth of the flow cytometry literature, 1987-93.....	64
3-1. The first working flow cytometer (Gucker particle counter)	74
3-2. Digitized image of a neutrophil polymorphonuclear leukocyte	82
3-3. Refined and prototype versions of Kamensky's Rapid Cell Spectrophotometer (RCS).....	84
3-4. A two-parameter histogram of blood cells analyzed in the RCS	84
3-5. Louis Kamensky and the Bio/Physics Systems Cytofluorograf.....	87
3-6. The Technicon Hemalog D Differential Leukocyte Counter	88
3-7. Leonard Herzenberg with B-D's first commercial version of the Fluorescence Activated Cell Sorter (FACS)	88
3-8. A two-parameter display from the Block differential counter showing five leukocyte clusters.....	89
3-9. The author with Cytomutt and Cerberus.....	94
3-10. Cell cycle phases defined by DNA content and by DNA/RNA content	97
4-1. Radian and steradian.....	101
4-2. Light as an electromagnetic wave.....	104
4-3. Constructive and destructive interference.....	105
4-4. Circularly polarized light.....	105
4-5. Reflection and refraction of light at a surface.....	106
4-6. The Jablonski diagram of electronic energy levels, or states, and transitions	112
4-7. Fluorescence spectrum of fluorescein.....	113
4-8. Light from a point source at the focal length of a lens is collimated	119
4-9. Rays from a point source can be focused to an image point	119
4-10. Rays from many points of an object add up to make an image	120
4-11. Elements of a typical microscope lens, showing the half angle that defines the acceptance cone and N.A.	120
4-12. Showing the effect of N.A. on light collection	121
4-13. Multiple views of magnification	123

FIGURES (continued)

4-14. Output characteristics of arc, quartz-halogen, and deuterium lamps.....	125
4-15. Köhler and critical illumination	128
4-16. Optics for arc source epiillumination for fluorescence microscopy or flow cytometry	129
4-17. Use of crossed cylindrical lenses to focus a laser beam to an elliptical spot on the core stream.....	131
4-18. Energy levels involved in laser action.....	133
4-19. Schematic of a laser.....	134
4-20. Laser transverse excitation modes	134
4-21. Measured beam intensity profiles of diode, CO ₂ , and He-Ne lasers	136
4-22. Intensity profiles of a focused and defocused violet laser diode beam.....	136
4-23. Sizes, power requirements, and approximate costs of some smaller lasers used for cytometry	145
4-24. Effect of noise compensation circuitry on precision of fluorescence measurements	147
4-25. Looking at the observation point	150
4-26. Maximizing light collection: Not always a good way to do things.....	151
4-27. A prism monochromator	153
4-28. Light transmission characteristics of bandpass, short pass, and long pass interference filters.....	154
4-29. Transmission of several wavelength regions through different dichroic configurations.....	155
4-30. Cube and plate beamsplitters	156
4-31. Total internal reflection	157
4-32. A fiber optical waveguide.....	157
4-33. Optical arrangements for collection of forward scattered light.....	159
4-34. Elements of a photomultiplier tube (PMT).....	160
4-35. PMT electrode voltage supply circuits: dynode chain and Cockcroft-Walton voltage multiplier	162
4-36. Detectors and housings.....	163
4-37. Fluid flow in a flow cytometer	167
4-38. Laminar flow profile illustrated by diatoms in the "Flow CAM" imaging flow cytometer	167
4-39. Flow chamber designs.....	169
4-40. New angles on light collection in flow.....	171
4-41. Flow chamber designs used with arc source flow cytometers.....	173
4-42. Minimizing turbulence generated at the sheath inlet to flow chambers.....	175
4-43. Sheath fluid supply plumbing.....	175
4-44. The Coulter orifice	183
4-45. Effect of beam geometry on pulse shape.....	184
4-46. Some circuit elements.....	185
4-47. A line-powered DC power supply.....	187
4-48. Basic operational amplifier circuits.....	190
4-49. A photodetector preamplifier circuit	191
4-50. Waveforms in preamplifier, front end electronics, and peak detector circuits.....	192
4-51. Schematic diagram of a peak detector.....	192
4-52. Preamp and peak detector outputs (oscilloscope traces).....	193
4-53. Telling two cells from one gets harder as the cells go through closer together	196
4-54. Uncompensated and compensated fluorescence signals from FL- and PE-labeled beads.....	198
4-55. One side of a two-color compensation circuit	198
4-56. Signals from PE-labeled antibody bound to beads, on linear and 4-decade logarithmic scales.....	199
4-57. Response curves of different types of log amps as determined according to Parks, Bigos, and Moore.....	202
4-58. Continuous and sampled signals	205
4-59. Digital pulse processing: "slicing" a slightly noisy Gaussian pulse with a baseline	209
4-60. MESF threshold sensitivity determination using fluorescein-labeled beads	216
4-61. Fluorescence distributions for bead sets measured with progressively lower values of Q.....	222
4-62. Separation (or lack thereof) of CD4+ lymphocytes and unstained cells at various values of Q and B	223
5-1. Distributions of sums of uniformly distributed random numbers approach the normal distribution.....	231
5-2. Pascal's triangle	232
5-3. Binomial distributions for $n = 2, 4, 8$, and 16 , with $p = q = 0.5$	232
5-4. Binomial distributions for $n = 16$ and $p = 0.5, 0.25$, and 0.125 ; they are skewed when $p \neq 0.5$	233

FIGURES (continued)

5-5. Euclidean distance between two points, with an assist from Pythagoras	234
5-6. "Box-and-whiskers" plots showing medians and interquartile ranges of distributions, after Tukey and Tufte.....	236
5-7. Histogram display formats	237
5-8. Dot plots of computer-generated data showing various degrees of correlation.....	237
5-9. Varieties of two-parameter data display (cover figure).....	239
5-10. Density plot showing a scale indicating numbers of events	240
5-11. 3-Dimensional "cloud" plots of CD3, CD4, and CD8 antigens on blood lymphocytes and thymocytes	241
5-12. How compensation gets data to not quite fit into quadrants	243
5-13. Two generated near-normal distributions and their cumulative distributions, or integrals	245
5-14. Quadrant statistics	247
5-15. Multiparameter analysis of peripheral blood leukocyte types: partitioning in two-dimensional subspaces.....	249
5-16. Linear transformation of data to separate clusters, as is done in linear discriminant analysis.....	251
5-17. Calculation is not always necessary to reduce the dimensionality of data	254
5-18. Linear and log scales revisited.....	255
6-1. Droplet sorting	259
6-2. A droplet sorter test stream pattern	261
6-3. Mack Fulwyler with his islet sorter	264
6-4. Fluidic sorter designs	264
6-5. A microfluidic flow sorter for bacteria	265
7-1. Forward scatter does not measure particle size	275
7-2. Depolarized 90° scatter signals can be used to identify eosinophil granulocytes	278
7-3. Indicatrices (plots of intensity vs. scattering angle) of particles and cells.....	280
7-4. Erythrocytes scatter less light at a wavelength at which hemoglobin exhibits strong absorption.....	281
7-5. Side scatter signals from T2 bacteriophages	288
7-6. The principle of doublet discrimination using pulse height and pulse integral measurements	290
7-7. Fluorescence spectra of some materials implicated in mammalian cell autofluorescence	291
7-8. Zinc protoporphyrin fluorescence in human red cells.....	292
7-9. Probe fluorescence spectra and source emission wavelengths	296
7-10. A "rogues' gallery" of nucleic acid dyes.....	301
7-11. DNA content distributions in sperm from a normal ram and a ram bearing a translocation	311
7-12. Structure of symmetric cyanine dyes given the formula "DiYC _n (2m + 1)" by Sims et al.....	313
7-13. Flow karyotype of human chromosomes stained with Hoechst 33258 and chromomycin A ₃	318
7-14. Hoechst/chromomycin fluorescence signatures of bacteria with different DNA base compositions.....	318
7-15. Flow cytometry of chromatin structure identifies cells in mitosis.....	320
7-16. DNA and RNA content analysis of mitogen-stimulated lymphocytes.....	321
7-17. Chemical structures of some reactive labels	327
7-18. Multiplex immunofluorescent labeling to demonstrate multiple cell types in blood.....	351
7-19. Quantitative determination of CD4 epitopes on peripheral blood lymphocytes and monocytes	356
7-20. Flow cytometric detection of HIV-1 nucleic acids in cells after <i>in situ</i> PCR	361
7-21. Plot of cytoplasmic [Ca ²⁺] as indicated by indo-1 fluorescence ratio versus time.....	365
7-22. Time used as a quality control parameter in DNA analysis (after Watson)	366
7-23. Cell proliferation indicated by dilution of fluorescence of the tracking dye PKH26.....	372
7-24. Cell proliferation indicated by dilution of fluorescence of CFSE-labeled CD4+ lymphocytes	373
7-25. Specific staining of glutathione in cells by monobromobimane (MBB)	381
7-26. Valinomycin-induced changes in fluorescence intensity of cyanine dye-loaded red cell suspensions	387
7-27. Distributions of the fluorescence of DiOC ₂ (3) in CCRF-CEM T-lymphoblasts.....	388
7-28. Distributions of membrane potential in cells suspended in NaCl, KCl and a mixture of the two.....	389
7-29. Structures of two membrane potential probes, the cyanine DiOC ₂ (3) and the oxonol DiBAC ₁ (3)	390
7-30. Two-parameter analyses of lectin-stimulated lymphocytes: DNA vs. RNA and membrane potential	396
7-31. Demonstration of apoptotic HL-60 cells with deenergized mitochondria by JC-1 staining.....	399
7-32. Measurement of membrane potential of <i>Staphylococcus aureus</i> using a ratiometric technique	400
7-33. Emission spectra of indo-1 in solutions of increasing free calcium ion concentration.....	404

FIGURES (continued)

7-34. Emission spectra of fluo-3 in solutions of increasing free Ca^{2+} concentration	405
7-35. Estimation of cytoplasmic pH in human lymphocytes from carboxyfluorescein fluorescence ratio.....	406
7-36. The pH-dependent emission spectra of carboxy SNARF-1 excited at 488 nm.....	407
8-1. The BD FACSVantage cell sorter.....	413
8-2. The BD FACSCalibur benchtop cell sorter.....	415
8-3. The BD LSR multi-beam benchtop analyzer.....	416
8-4. BD's "Octagon" collection optics.....	416
8-5. The BD FACSAria cell sorter	417
8-6. The Beckman Coulter EPICS Altra sorter	419
8-7. The Beckman Coulter Cytomics FC 500 analyzer.....	421
8-8. The Beckman Coulter EPICS XL-MCL analyzer	422
8-9. The CyAn benchtop flow cytometer [DakoCytomation]	425
8-10. Cytopeia's InFlux cell sorter platform	426
8-11. Partec's CyFlow flow cytometer	427
8-12. The Partec PAS cytometer	429
10-1. Using DRAQ5-stained nuclei as alignment particles	446
10-2. "Equal opportunity" and "unequal opportunity" staining of T-cells.....	447
10-3. Separation of X- and Y- bull sperm	453
10-4. Detection of BrUdR incorporation using anti-BrUdR antibody and the SBIP method	457
10-5. Use of a combination of propidium and TO-PRO-3 to detect bromodeoxyuridine incorporation	458
10-6. DNA content vs. expression of cyclins B and E in exponentially growing MOLT-4 cells	459
10-7. DNA content vs. RNA content and CD71 expression in activated CD4 cells	460
10-8. Identification of mitotic cells by antibody to phosphorylated histone H3.....	462
10-9. Subsetting of human T-cells into memory classes and measurement of kinases using 11-color fluorescence.....	468
10-10. Analysis of bacteriophage lambda DNA and fragments from a digest of lambda DNA in a slow flow system.....	472
10-11. Clusters representing 100 different color-coded beads used with Luminex's system for multiplexed analysis.....	473
10-12. Growth of an encapsulated Gram-positive marine bacterium in gel microdroplets	474
10-13. Identification of occupied and unoccupied gel microdroplets.....	475
10-14. Univariate flow karyotype of human chromosomes stained with propidium iodide	477
10-15. Bivariate karyotypes of RPET001 and Daudi human cell lines.....	478
10-16. Clusters of peripheral blood leukocyte types in two-dimensional displays from a hematology analyzer	483
10-17. Side population (SP) stem cells identified by blue vs. red Hoechst 33342 fluorescence.....	489
10-18. Time course of events in T-lymphocyte activation and probes for their cytometric detection	494
10-19. Intracellular cytokine staining.....	500
10-20. ERK1/2 kinase phosphorylation in T-cells exposed to various activation stimuli	501
10-21. Gating scheme for detection of minimal residual disease in chronic lymphocytic leukemia.....	505
10-22. T-cell subset analysis in lion and puma peripheral lymphocytes.....	511
10-23. Analysis of clam cells in a hematology analyzer.....	511
10-24. Determination of nuclear genome size in diploid banana	513
10-25. DNA content of cactus nuclei showing endopolyploidy.....	513
10-26. Flow karyotype of a translocation line of broad bean.....	514
10-27. rRNA probes show differences in enteric flora between breast-fed infants and infants fed reconstituted milk ...	518
10-28. Membrane potential in dormant and resuscitated cultures of <i>Micrococcus luteus</i>	519
10-29. Combined rRNA probe and CTC staining of a genetically and metabolically complex cell mixture	520
10-30. Functional states of <i>Salmonella typhimurium</i> shown by staining with DiBAC ₄ (3), ethidium, and propidium ...	521
10-31. Effects of amoxicillin on [ratiometric] membrane potential and permeability of <i>Staphylococcus aureus</i>	521
10-32. Fluorescence profiles of viruses stained with SYBR Green I.....	523
10-33. Forward scatter, Hoechst 33342, and chlorophyll fluorescence of the marine bacterium <i>Prochlorococcus</i>	524
10-34. Side scatter and fluorescence signatures of four viruses shown in Figure 10-32.....	525
10-35. Side scatter and fluorescence signatures of viruses and bacteria in a water sample from a small Alpine pond	525
10-36. Size and DNA content of bacteria in seawater from Prince William Sound, Alaska	526
10-37. Size and DNA content of <i>Oligobacterium</i> RB1 from Resurrection Bay, Alaska	526

FIGURES (continued)

10-38. Flow cytometry <i>in vivo</i>	542
12-1. Tearing down the “picket fence” and reuniting the negatives using a BiExponential data transform	562

1. OVERTURE

This is a book about **cytometry**, in general, emphasizing **flow cytometry**, in particular. In it, I hope to tell you what cytometry is, how it works, why and how to use it, when you should favor one type of cytometry or another, and when cytometry won't solve your problem. This chapter, like the overture to an opera or a musical, presents important themes from the body of the work, but may also stand alone.

1.1 WHAT (AND WHAT GOOD) IS CYTOMETRY?

Cytometry is a process in which physical and/or chemical characteristics of **single cells**, or by extension, of other biological or nonbiological particles in roughly the same size range, are **measured**. In **flow cytometry**, the measurements are made as the cells or particles pass through the measuring apparatus, a **flow cytometer**, in a fluid stream. A **cell sorter**, or **flow sorter**, is a flow cytometer that uses electrical and/or mechanical means to divert and collect cells (or other small particles) with measured characteristics that fall within a user-selected range of values.

Neither the cells nor the apparatus are capable of putting the process of cytometry in motion; the required critical element for that is a human interested in obtaining information about a cell sample and, in the case of sorting, extracting cells of interest from the sample. At the most basic level, a cytometer might be considered to be a "black box" with cells as "inputs" and numbers as "outputs"; the outputs of a cell sorter would include both numbers and cells. However, while some modern cytometers (and some modern users) can obtain the desired results while running unattended in "black box" mode, it is fair to say that most of the applica-

tions, and all of the interesting applications, of cytometry call for some understanding and some intellectual effort on the part of the user.

Tasks and Techniques of Cytometry

From the time of van Leeuwenhoek and Hooke until the mid-20th century, determining:

- 1) whether cells were present in a specimen,
- 2) how many were there,
- 3) what kinds of cells were represented, and
- 4) what their functional characteristics might be

required that a human observer interpret a microscope image. The same tasks remain for modern cytometry.

Although electrical and acoustic properties of, and nuclear radiation emission from, single cells can be measured, it is fair to say that **optical** measurements are by far the most common in cytometry. A typical cytometer is thus a specialized **microscope**; the degree of physical resemblance is dictated by the requirements of the measurement(s) to be made, which in turn are dictated by what the user needs to know about the cell sample. In successful applications of cytometry, electro-optics, electronics, and computers are employed to improve on what could be obtained "by eye," although interpretation is required more often than not. The successful applications are many, increasing in number, and commonplace in locales as diverse as clinical laboratories and breweries.

Some Notable Applications

Cytometry is currently used to obtain the helper T lymphocyte counts needed to monitor the course and treatment of HIV infection, and to determine tumor cell DNA content

and proliferative activity, which may aid in assessing prognosis and determining treatment for patients with breast cancer and other malignant diseases. The technology has also been used to crossmatch organs for transplantation, to isolate human chromosomes for the construction of genetic libraries, to separate X- and Y-chromosome bearing sperm for sex selection in animal breeding and *in vitro* fertilization in humans, to identify the elusive hematopoietic stem cell and an expanding family of other stem cell types, and to reveal several widely distributed but previously unknown genera of marine microorganisms.

Biological particles that have been subjected to cytometric analysis range, in order of decreasing size, from multicellular organisms (e.g., *Drosophila* embryos and adult *Caenorhabditis elegans* nematodes) through cell aggregates (e.g., pancreatic islets and tumor cell spheroids), eukaryotic cells, cellular organelles (e.g., mitochondria), bacteria, liposomes, individual virus particles and immune complexes, down to the level of single molecules of proteins, nucleic acids, and organic dyes. Cytometers can also be used for sensitive chemical analyses involving the binding of suitably labeled ligands to solid substrates or to particles such as polystyrene beads.

The first practical applications of flow cytometry, beginning in the 1940's, were to counting blood cells in liquid suspension, on the one hand, and bacteria and other small particles in aerosols, on the other, based on measurements of **light scattering or electrical impedance**; these signals were also used to provide estimates of cell size.

In the early 1960's, **light absorption** measurements were used for quantitative flow cytometric analyses of cellular nucleic acid and protein. Flow cytometers in modern clinical hematology laboratories perform counts of red cells (erythrocytes), white cells (leukocytes), and platelets (thrombocytes) in blood, as well as differential leukocyte counts, using combinations of electrical impedance, light scattering, and light absorption measurements.

However, many people who know the term "flow cytometer" tend to use it – incorrectly – to describe only instruments that measure **fluorescence** as well as light scattering. The first fluorescence flow cytometers were built in the late 1960's; although there are now well over 10,000 in use in clinical and research laboratories worldwide, they are still outnumbered by impedance and scattering-based hematology analyzers. So much for fluorescence chauvinism.

What is Measured: Parameters and Probes

The novice should not be intimidated by the jargon of cytometry; there are no native speakers, and he or she can soon enough become as fluent in it as the rest of us. The term **parameter** is, unfortunately, used in several different senses in our jargon. It can refer to a **physical or chemical characteristic of a cell** (e.g., cytoplasmic granularity or nuclear DNA content) that is measurable by cytometry; it can also describe a **physical property, measured by a sensor,**

defined broadly (e.g., light scattering or fluorescence), or more narrowly (e.g., orthogonal light scattering or red fluorescence), or a **physical property of a cell-associated reagent** (e.g., propidium fluorescence). A fairly comprehensive list of measurable cellular parameters appears as Table 1-1 on the facing page.

I have characterized cellular parameters as **intrinsic** or **extrinsic**, depending upon whether they can or cannot be measured without the use of **reagents**, which are often referred to in cytometric jargon as **probes**. Some parameters can, at least in principle, be measured either with or without probes; cellular DNA content, for example, can be estimated from ultraviolet (UV) absorption at 260 nm in unstained cells, but it's much more practical to use a fluorescent dye probe such as propidium iodide. A deeper philosophical dilemma arises when considering fluorescence from *Aequorea* green fluorescent protein (GFP) or one of its genetically engineered offshoots, introduced by cloning into cells of other species to report gene expression; one could characterize this as intrinsic or extrinsic, but I lean toward the latter.

Parameters can also be defined as **structural** or **functional**, again with some ambiguity. For example, the glycoprotein efflux pump responsible for multidrug resistance in tumor cells can be detected, and the amount present in a cell quantified, using fluorescent antibodies, but such antibodies might also bind to an inactive mutant protein, and thus provide a measurement (in this case, inaccurate) based on structure. The function of the glycoprotein pump can be demonstrated by measurement of uptake or loss of fluorescent drugs or dyes by cells over periods of time.

In a **kinetic measurement** such as that just described, **time** itself can be used as a parameter. When such analyses are done by flow cytometry, the dynamic behavior of a cell population must be inferred from observations of different cells at different times, because conventional flow cytometers cannot make successive measurements of a single cell over time periods exceeding a few microseconds.

Both the novice and the expert in flow cytometry should be aware that almost every parameter that can be measured by flow cytometry can also be measured by alternative cytometric methods such as **microspectrophotometry, confocal microscopy, image analysis, and scanning cytometry**. These methods are often applicable where flow cytometric methods are not, e.g., for true kinetic analyses involving repeated examination of the same cell or cells over a period of time, or for *in situ* analyses of cells growing in aggregates attached to solid substrates. In general, the fluorescent probes used for flow cytometry can be used with alternative measurement techniques. However, most dyes and other reagents that are commonly employed in absorption microspectrophotometry are not readily usable in fluorescence flow cytometers.

1.2 BEGINNINGS: MICROSCOPY AND CYTOMETRY

It recently (i.e., since the last time I wrote an introduction to cytometry) occurred to me that the best way in

PARAMETER	MEASUREMENT METHOD AND PROBE IF USED
Intrinsic Structural Parameters (no probe)	
Cell size	Electronic (DC) impedance, extinction, small angle light scattering; image analysis
Cell shape	Pulse shape analysis (flow); image analysis
Cytoplasmic granularity	Large angle light scattering, Electronic (AC) impedance
Birefringence (e.g., of blood eosinophil granules)	Polarized light scattering, absorption
Hemoglobin, photosynthetic pigments, porphyrins	Absorption, fluorescence, multangle light scattering
Intrinsic Functional Parameter (no probe)	
Redox state	Fluorescence (endogenous pyridine and flavin nucleotides)
Extrinsic Structural Parameters (probe required)	
DNA content	Fluorescence (propidium, DAPI, Hoechst dyes)
DNA base ratio	Fluorescence (A-T and G-C preference dyes, e.g., Hoechst33258 and chromomycin A ₁)
Nucleic acid sequence	Fluorescence (labeled oligonucleotides)
Chromatin structure	Fluorescence (fluorochromes after DNA denaturation)
RNA content (single and double-stranded)	Fluorescence (acridine orange, pyronin Y)
Total protein	Fluorescence (covalent- or ionic-bonded acid dyes)
Basic protein	Fluorescence (acid dyes at high pH)
Surface/Intracellular antigens	Fluorescence; scattering (labeled antibodies)
Surface sugars (lectin binding sites)	Fluorescence (labeled lectins)
Lipids	Fluorescence (Nile red)
Extrinsic Functional Parameters (probe required)	
Surface/Intracellular receptors	Fluorescence (labeled ligands)
Surface charge	Fluorescence (labeled polyanionic molecules)
Membrane integrity (not always a sign of "viability")	Fluorescence (propidium, fluorescein diacetate [FDA]); absorption or scattering (Trypan blue)
Membrane fusion/turnover	Fluorescence (labeled long chain fatty acid derivatives)
Membrane organization (phospholipids, etc.)	Fluorescence (annexin V, merocyanine 540)
Membrane fluidity or microviscosity	Fluorescence polarization (diphenylhexatriene)
Membrane permeability (dye/drug uptake/efflux)	Fluorescence (anthracyclines, rhodamine 123, cyanines)
Endocytosis	Fluorescence (labeled microbeads or bacteria)
Generation number	Fluorescence (lipophilic or covalent-bonded tracking dyes)
Cytoskeletal organization	Fluorescence (NBD-phalloidin)
Enzyme activity	Fluorescence; absorption (fluorogenic/chromogenic substrates)
Oxidative metabolism	Fluorescence (dichlorofluorescein)
Sulfhydryl groups/glutathione	Fluorescence (bimanes)
DNA synthesis	Fluorescence (anti-BrdUrd antibodies, labeled nucleotides)
DNA degradation (as in apoptosis)	Fluorescence (labeled nucleotides)
"Structuredness of cytoplasmic matrix"	Fluorescence (fluorescein diacetate [FDA])
Cytoplasmic/mitochondrial membrane potential	Fluorescence (cyanines, rhodamine 123, oxonols)
"Membrane-bound" Ca ²⁺	Fluorescence (chlorotetracycline)
Cytoplasmic [Ca ²⁺]	Fluorescence ratio (Indo-1), fluorescence (fluo-3)
Intracellular pH	Fluorescence ratio (BCECF, SNARF-1)
Gene expression	Fluorescence (reporter proteins)

Table I-1. Some parameters measurable by cytometry.

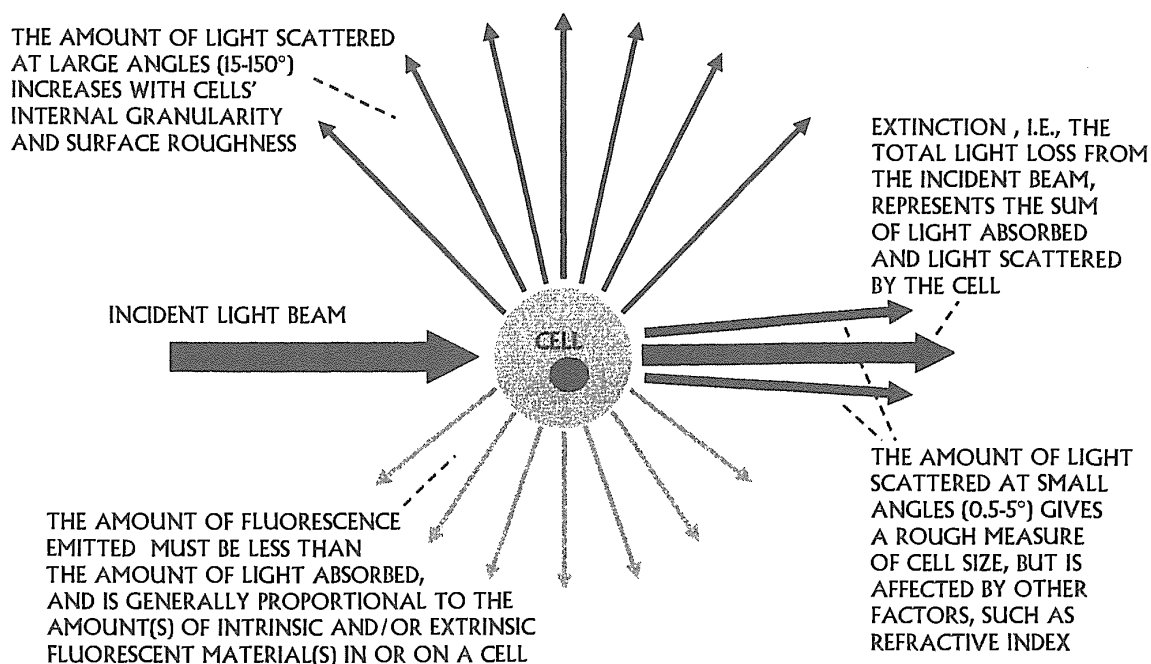


Figure 1-1. Interaction of light with a cell.

which to introduce the subject might be to consider how cytometry developed from microscopy, emphasizing both the similarities and the differences between the two, and stressing how the information gets from the cells to the user. That is what I will try to do in the remainder of this chapter. I hope this will be helpful for the uninitiated reader, but, also, that it will be equally thought-provoking, informative, and at least moderately amusing to those who have been over the terrain one or many times before.

The first order of business in both microscopy and cytometry is discriminating between the cells and whatever else is in the sample; the next is often discriminating among a number of different cell types that may be present. Optical microscopes first allowed cells to be discovered and described in the seventeenth century, and were refined in design in the eighteenth and early nineteenth, but the capacity of microscopy to discriminate among different cell types remained limited by the relative difficulty of obtaining contrast between cells and the background in microscope images.

A Little Light Music

While all the senses can provide us with pleasure and discomfort, it is predominantly vision that shapes our perception of the world around us, and, without light, our visual imagery is restricted to memories, dreams, and hallucinations. According to the Book of Genesis, the discrimination of light from darkness is the divine achievement of the first day of creation, and we humans, despite taming fire and

inventing light bulbs and lasers, remain aware of and profoundly affected by the daily difference, not least during power outages.

What most of us know as light is defined by physicists as **electromagnetic radiation** with wavelengths ranging between about 400 and about 700 nanometers (nm). Other species can detect shorter and longer wavelengths, but most lack our ability to discriminate among wavelengths, i.e., color vision, and some of us have genetic deficiencies that restrict this capacity.

When we look at the macroscopic world, most of our retinal images are formed by light that we say is **reflected** from objects around us, and an early concept of light was that of **rays** traveling in straight lines, and reflecting from a surface at the same angle at which they strike it. If we look at an object under water and attempt to grab it, we find that it is not exactly where it appears to be; this is explained by the concept of **refraction**, according to which light passing from one material medium into another is bent at an angle depending on a macroscopic property of the medium known as the **refractive index**, and on the wavelength of the light. The "white" light emitted by the sun and by incandescent and fluorescent bulbs comprises a range of visible wavelengths; objects and materials that **absorb** some, but not all, wavelengths reflect others, and thus appear colored.

As we turn our attention to smaller and smaller objects, the concepts of reflection and refraction become less and less useful, and we instead make use of the concept of **light scattering**. Figure 1-1 describes the interaction of light with a

cell in terms of **scattering**, **absorption**, and **fluorescence**. The last of these phenomena is not readily explicable in terms of either ray (geometrical) or wave optics, and can only be dealt with properly by the theory of **quantum electrodynamics**, which considers light as particles, or **photons**, which interact with **electrons** in atoms and molecules. The energy of a photon is inversely proportional to the corresponding wavelength; i.e., photons of short-wavelength, 400 nm violet light have a higher energy content than photons of long-wavelength, 700 nm red light.

Scattering, which explains both reflection and refraction, typically involves a brief interaction between a photon and an electron, in which the photon is annihilated, transferring its energy to the electron, which almost immediately releases all of the energy in the form of a new photon. Thus, light scattered by an object has the same (or almost exactly the same) wavelength, or color, as the incident light. However, the new photon does not necessarily travel in the same direction as the old one, so scattered light usually appears to be at an angle to the incident beam.

In empty space, there are, by definition, no atoms or molecules, and there are thus no electrons available to interact with photons. Although, according to quantum electrodynamics, a photon has a finite probability of going in any direction, when we actually calculate the probabilities that apply in the case of photons in empty space, we come up with what look like rays of light traveling in straight lines.

As a general rule, the density of atoms and molecules in atmospheric air is fairly low, meaning that there are few opportunities for light to be scattered as it appears to traverse distances of a few meters or tens of meters. However, we note the blue appearance of a cloudless sky, resulting from light scattering throughout the atmosphere; the color results from the fact that shorter wavelengths of light are more likely to be scattered than longer ones, with the intensity of scattering inversely proportional to the fourth power of the wavelength.

The well-known laws of reflection and refraction emerge from quantum electrodynamics applied to objects substantially bigger than the wavelength of light. Materials that appear transparent to the human eye, e.g., glass and water, still contain relatively high densities of atoms and molecules, and thus provide numerous opportunities for scattering.

Some light appears to be reflected at the interfaces between layers of different materials, with the angle of reflection equal to the angle of incidence. The total amount of light reflected is found to be a function of the thickness of the layers and the wavelength of the incident light; that is, layers of different thicknesses reflect different colors of light to different extents. This **interference** effect, explained by the theory of wave optics, accounts for the patterns of color seen in peacock feathers, butterfly wings, diffraction gratings in spectrophotometers, on credit cards, and in cheap jewelry, and in opals in somewhat more expensive jewelry. It is exploited in optical design, notably in the production of **interference filters** used to select ranges of wavelengths to be

observed and/or detected in microscopes and other optical instruments. Quantum electrodynamics comes up with the same results for interference and reflection as wave optics, even while taking into account that the phenomena are due to scattering throughout objects, not just from front and back surfaces.

The apparent bending of light striking an interface between two materials is described in classical optics with the aid of invented quantities, called **refractive indices**, which are characteristic of the materials involved. Light appears to travel more slowly through a material of higher refractive index than through a material of lower index, and a "ray" appears to "bend" toward the normal (i.e., toward a line perpendicular to the interface) when passing from a lower-index medium to a higher one, and away from the normal when passing from a higher-index medium to a lower one. The apparent velocity of light in a material is less than in empty space; the higher the refractive index, the lower the apparent velocity. Light of a shorter wavelength is "bent" more than light of a longer one, allowing a transparent object with surfaces that are not parallel (i.e., a **prism**) to **disperse** light of different wavelengths in different directions.

Armed with ray optics and the classical law of refraction, we can calculate how an object with appropriately curved surfaces, i.e., a **lens**, will "bend" light originating from two points separated in space. If the surfaces are convex, divergent "rays" coming through the lens from two points a given distance apart on the "input" side can be made to converge at two points a greater distance apart on the "output" side; this provides us with a **magnified image**. A magnifying lens is, of course, the fundamental ingredient of a microscope.

Not surprisingly, everything useful that classical optics tells us about refraction can be obtained using quantum electrodynamics. Although actually doing this usually involves a great deal of advanced mathematics, Richard Feynman, who received his Nobel Prize for work in the field, wrote a small book called *QED*⁶⁴, in which he used simple diagrams and concepts to make the subject accessible to a lay audience (which, in this context, includes me). What I am writing here paraphrases the master.

The light scattering behavior of objects of dimensions near the wavelength of light is not predictable from ray optics. For spherical particles ranging in diameter from one or two wavelengths to a few tens of wavelengths, most of the light scattering occurs at small angles (0.5° to 5°) to the incident beam; the intensity of this "**small angle**," or "**forward**," light scattering is dependent on the refractive index difference between the particle and the medium, and on particle size. However, the relationship between particle size and small angle scattering intensity is not monotonic, meaning that, although a particle 10 μm in diameter will probably produce a bigger signal than one of the same composition 5 μm in diameter, a particle 5.5 μm in diameter might produce a smaller signal than one 5 μm in diameter. It is thus wise to avoid thinking of the small angle scatter signal as an accurate measure of cell size.

Smaller particles scatter proportionally more light at larger angles (15° to about 150°) to the incident beam; the amplitude of such signals, variously described as “side,” “orthogonal,” “large angle,” “wide angle,” or “90°” light scattering, is, all other things being equal, larger for cells with internal granular structure, such as blood granulocytes, than for cells without it, such as blood lymphocytes.

Ray optics and wave optics break down when we consider the process of **light absorption**. This comes down to photons and electrons, period. Quantum theory tells us that the electrons in a given atom or molecule can exist only in discrete energy states. The lowest of these is referred to as the **ground state**, and the absorption of a photon by an electron in the ground state raises it to a higher energy **excited state**. An electron in an excited state can absorb another photon, ending up in a still higher energy excited state.

Like scattering, and all other quantum phenomena, absorption is probabilistic. We cannot say that a particular electron will absorb a particular photon; the best we can do is calculate the probability that an electron in a particular energy state will absorb a photon of a particular energy, or wavelength. This probability increases as the difference in energy between the current energy state of the electron and the next higher energy state gets closer to the energy of the photon involved.

In many molecules, the energy difference between states is greater than the energy in a photon of visible light. Such molecules may exhibit substantial absorption of higher energy, shorter wavelength photons, e.g., those with wavelengths in the **ultraviolet (UV)** region between about 200 and 400 nm. Substances made up of such molecules appear transparent to the human eye; smearing them on exposed skin decreases the likelihood that ultraviolet photons will interact with electrons in DNA and other macromolecules of dermal cells, and reduces the likelihood of sunburn (yay!) and tanning (boo!). We're not sure yet about skin cancer.

For a molecule to absorb light in the visible region, the energy differences between electronic energy states have to be rather small. This condition is satisfied in some inorganic atoms and crystals, which have unpaired electrons in *d* and *f* orbitals, in metals, which have large numbers of “free” electrons with an almost continuous range of energy states, resulting in high absorption (and high reflectance) across a wide spectral range, and in organic molecules with large systems of conjugated π orbitals, including natural products such as porphyrins and bile pigments, and synthetic dyes such as those used to stain cells.

The interaction of light with matter must obey the law of conservation of energy; the amount of light transmitted should therefore be equal to the amount of incident light minus the amount scattered and the amount absorbed. But what happens to the absorbed light? One would not expect the electrons involved in absorption to remain in the excited state indefinitely, and, indeed, they do not. In some cases, all of the absorbed electronic energy is converted to vibrational or rotational energy, and lost as heat. In others, some energy

is lost as heat, but the remainder is emitted in the form of photons of lower energy (and, therefore, longer wavelength) than those absorbed. Depending on the details of the electronic energy transitions involved, this emission can occur as **fluorescence** or as **phosphorescence**. Fluorescence emission usually occurs within a few tens of nanoseconds of absorption; phosphorescence is delayed, and may continue for seconds or longer. As is the case with absorption, fluorescence and phosphorescence are inexplicable by ray and wave optics; they can only be understood in terms of quantum mechanics.

Making Mountains out of Molehills: Microscopy

When we are not looking at luminous displays such as the one I face as I write this, most of our picture of the world around us comes from reflected light. Contrast between objects comes from differences in their reflectivities at the same and/or different wavelengths. When ambient light levels are high, we utilize our retinal cones, which give us color vision capable of prodigious feats of spectral discrimination (humans with normal vision can discriminate millions of colors), at the expense of relatively low sensitivity to incident light. The high light levels bleach the visual pigments in our more sensitive retinal rods; if the light level is decreased abruptly, it takes some time for the rod pigment to be replenished, after which we can detect small numbers of photons, sacrificing color vision in the process. Thus, while we can perceive large numbers of 450 nm photons, 550 nm photons, and 650 nm photons, respectively, as red, green, and blue light, using our cones, we cannot distinguish individual photons with different energy levels as different colors. Night vision equipment typically utilizes monochromatic green luminous displays because the rods are most sensitive to green light, but the cone system also exhibits maximum sensitivity in the green region, making the spot from a green laser pointer much more noticeable than that from a red one emitting the same amount of power.

While the spectral discrimination capabilities of the unaided human visual system are remarkable, its spatial discrimination power is somewhat limited. The largest biological cells, e.g., ova and large protists, are just barely visible, and neither the discovery of cells nor the appreciation of their central role in biology would have occurred had the light microscope not been invented and exploited.

When unstained, unpigmented cells are examined in a traditional **transmitted light**, or **bright field**, microscope, light absorption is negligible; contrast between cells and the background is due solely to scattering of light by cells and subcellular components, and the only information we can get about the cells is thus, in essence, contained in the scattered light. Some of this is scattered out of the field of view; we must therefore rely on slight differences in transmission between different regions of the image to detect and characterize cells. We are working against ourselves by presenting our eyes (or the detector(s) in a cytometer) with a large amount of light that has been transmitted by the specimen.

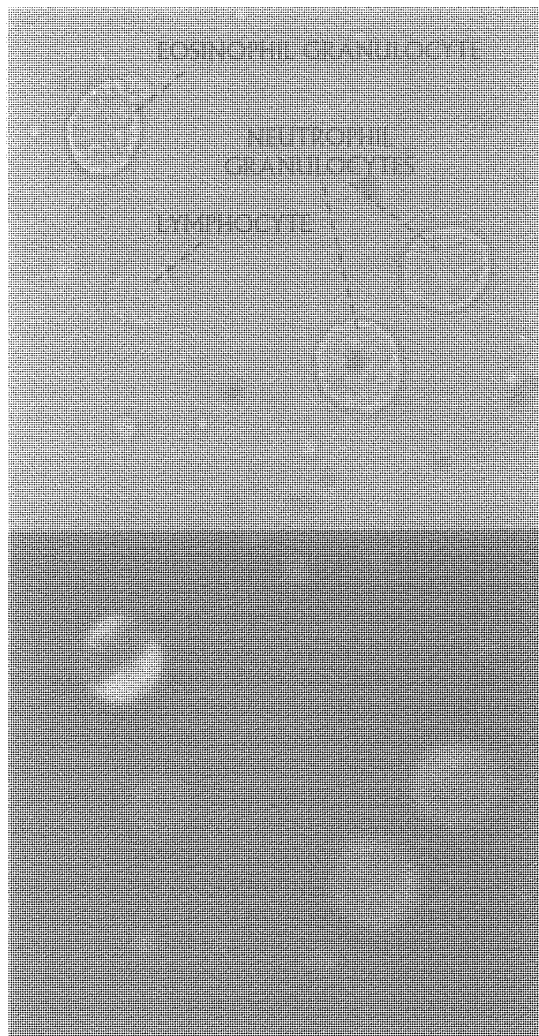


Figure 1-2. Transmitted light (bright field) (top panel) and dark field (bottom panel) images of an unstained suspension of human peripheral blood leukocytes. The objective magnification was 40 $\times$.

As it happens, the maximum spatial resolution of a microscope is achieved, i.e., the distance at which two separate objects can be distinguished as separate is minimized, when illuminating light reaches, and is collected from, the specimen at the largest possible angle. The **numerical aperture (N.A.)** of microscope condensers and objectives is a measure of the largest angle at which they can deliver or collect light. However, when the illumination and collection angles in a transmitted light microscope are large, much of the light scattered by objects in the specimen finds its way back into the microscope image, increasing resolution, but decreasing contrast. The top panel of Figure 1-2 shows a bright field microscope image of a suspension of human peripheral blood leukocytes; the condenser was stopped down to in-

crease contrast between the cells and background. The cytoplasmic granules in the eosinophil and neutrophil granulocytes are not particularly well resolved, nor is it easy to distinguish the nuclei from the cytoplasm. Increasing the level and angle of illumination might, as just mentioned, increase resolution, but this would not be useful, as contrast would not be increased.

Modern microscopy exploits both differences in **phase** and **polarization** of **transmitted light** and the phenomenon of **interference** to produce increased contrast in bright field images. However, **staining**, which came into widespread use in the late 1800's, largely due to the emergence of synthetic organic dyes, was the first generally applicable practical bright field technique for producing contrast between cells and the medium, and between different components of cells in microscope images. Paul Ehrlich, known for his later researches on chemotherapy of infectious disease, stained blood cells with mixtures of acidic and basic dyes of different colors, and identified the three major classes of blood granulocytes, the basophils, eosinophils (which he termed acidophils), and neutrophils, based on the staining properties of their cytoplasmic granules.

Stained elements of cells are visually distinguishable because of their **absorption** of incident light, even when the refractive index of the medium is adjusted to be equal or nearly equal to that of the cell. The dyed areas transmit only those wavelengths they do not absorb, resulting in a difference in spectrum, or color, between them and undyed areas or areas that take up different dyes. Absorption by pigments within cells, such as the hemoglobin in erythrocytes, also makes the cells more distinguishable from the background.

Microscopy of opaque specimens, such as samples of minerals, obviously cannot use transmitted light bright field techniques. Instead, specimens are illuminated from above, and the image is formed by light reflected (i.e., scattered) from the specimen. In **incident light bright field microscopy**, illumination comes through the objective lens, using a partially silvered mirror, or **beam splitter**, to permit light to pass between source and specimen and between specimen and eyepiece at the same time. In **dark field microscopy**, illumination is delivered at an oblique angle to the axis of the objective by a separate set of optics. The bottom panel of Figure 1-2 is a dark field image of the same cells as are shown in the top panel. In the dark field microscope, none of the illuminating light can reach the objective unless it is scattered into its field of view by objects in the specimen. The illumination geometry used in this instance ensured that the only light contributing to the dark field image was light scattered at relatively large angles to the illuminating beam. It has already been noted that this is the light represented in the side scatter signal, and it can be seen that the lymphocyte, which would have the smallest side scatter signal, appears dimmer than the neutrophil granulocytes and the eosinophil, which would have higher side scatter signals. Although the cytoplasmic granules within the granulocytes are not well resolved, the intensity of light coming from the

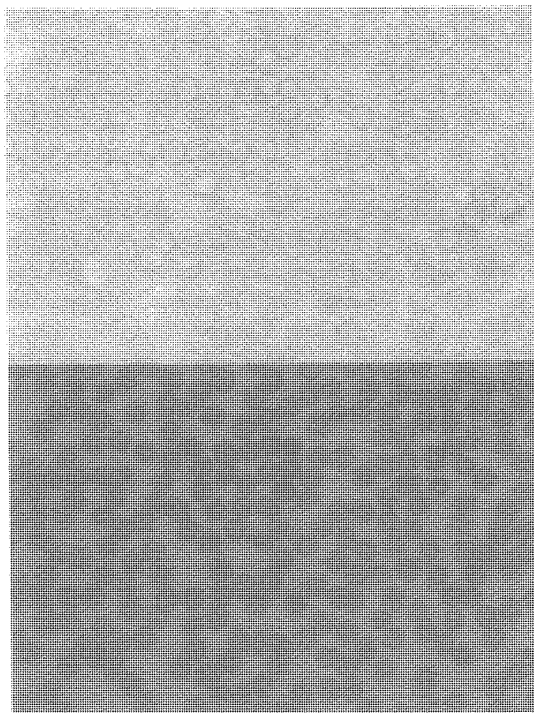


Figure 1-3. Transmitted light microscope images of an unstained smear of human peripheral blood. The picture in the top panel was taken with “white light” illumination; that in the bottom panel was taken with a violet (405 nm, 15 nm bandwidth) band pass optical filter, and demonstrates the strong absorption of intracellular hemoglobin in this wavelength region. Objective: 40 $\times$.

cytoplasm provides an indication of their presence; indeed, it is much easier to resolve nucleus from cytoplasm in the granulocytes in the dark field image than in the bright field image. Thus, we can surmise that it may be possible to get information about subcellular structures from a cytometer operating at an optical resolution that would be too low to allow them to be directly observed as discrete objects. In fact, using dark field microscopy, one can observe light scattered by, and fluorescence emitted from, particles well below the limit of resolution of an optimally aligned, high-quality optical microscope; the dark field “ultramicroscope” of the 1920’s allowed researchers to see and count viruses, although it was obviously impossible to discern any structural detail.

Absorption measurements are bright field measurements, and they work best, especially for quantification, when the absorption signal is strong. The material being looked for should have a high likelihood of absorbing incident light, as indicated by a high **molar extinction coefficient**, and there should be a lot of it in the cell. Figure 1-3 shows the absorption of hemoglobin in the cytoplasm of unstained red blood

cells. Note that the “white light” image in the top panel gives little hint of strong absorption, which is restricted to the violet region known as the Soret band; the “white” light used here, which came from a quartz-halogen lamp, contains very little violet, and the exposure time used for the picture in the bottom panel was about 100 times as long as that for the picture in the top panel.

Fluorescence microscopy is inherently a dark field technique; even in a “transmitted light” fluorescence microscope, **optical filters** are employed to restrict the spectrum of the illuminating beam to the shorter wavelengths used for fluorescence **excitation**, and also to allow only the longer-wavelength fluorescence **emission** from the specimen to reach the observer. As is the case in dark field microscopy, fluorescent cells (ideally) appear as bright objects against a dark background.

Most modern fluorescence microscopes employ the optical geometry shown in Figure 1-4. Excitation light is usually supplied by a mercury or xenon arc lamp or a quartz-halogen lamp, equipped with a lamp condenser that **collimates** the light, i.e., produces parallel “rays.” These components are not shown, but would be to the left of the excitation filter in the figure. The excitation filter passes light at the excitation wavelength, and reflects or absorbs light at other wavelengths. The excitation light is then reflected by a **dichroic mirror**, familiarly known simply as a **dichroic**, which transmits light at the emission wavelength. The microscope objective is used for both illumination of the specimen and collection of fluorescence emission, which is transmitted through both the dichroic and the emission filter.

In any microscope, a real image of the specimen is formed by the objective lens; the eyepiece and the lens of the observer’s eye then project an image of this image onto the retina of the observer. Light falling on sensitive cells in the retina produces electrical impulses that are transmitted along the optic nerves. What happens next is the province of neurology, psychology, and, possibly, psychiatry.

It has already been noted that humans are very good at color discrimination, and we also know that humans, with some training, can get pretty good at discriminating cells from other things. With more training, we can become proficient at telling at least some kinds of cells from others, usually on the basis of the size, shape, color, and texture of cells and their components in microscope images; it is not always easy to program computers to make the same distinctions on the same basis.

The human visual system can detect light intensities that vary over an intensity range of more than nine decades; in other words, the weakest light we can perceive is on the order of one-billionth the intensity of the strongest perceptible light. However, we can’t cover the entire range at once; as previously mentioned, we need dark-adapted rods to see the least intense signals, and do so only with monochromatic vision. And we aren’t very good at detecting small changes in light intensity. This has forced us to invent instruments to make precise light intensity measurements to meet the needs

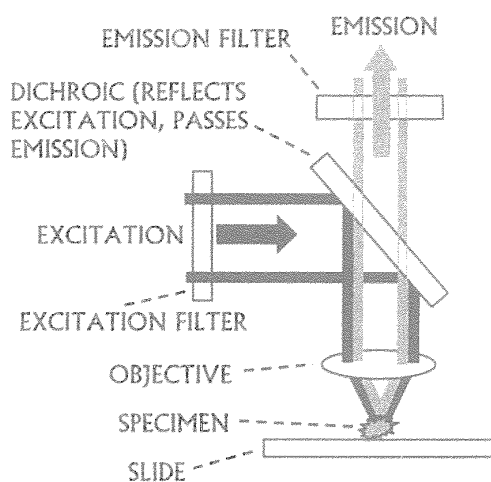


Figure 1-4. Schematic of a fluorescence microscope.

of science, technology, medicine, and/or art (remember when the light meter was not built into the camera?). It was this process that eventually got us from microscopy to cytometry.

Why Cytometry? Motivation and Machinery

In the 1930's, by which time the conventional histologic staining techniques of light microscopy had already suggested that tumors might have abnormalities in DNA and RNA content, Torbjörn Caspersson³⁴, working at the Karolinska Institute in Stockholm, began to study cellular nucleic acids and their relation to cell growth and function. He developed a series of progressively more sophisticated **microspectrophotometers**, which could make fairly precise measurements of DNA and RNA content based on the strong intrinsic UV absorption of these substances near 260 nm, and also found that UV absorption near 280 nm, due to aromatic amino acids, could be used to estimate cellular protein content. When Caspersson began working, it had not yet been established that DNA was the genetic material; he helped move others toward that conclusion by establishing, through precise measurement, that the DNA content of chromosomes doubled during cellular reproduction⁷.

A conventional optical microscope incorporates a **light source** and associated optics that are used to illuminate the specimen under observation, and an **objective** lens, which collects light transmitted through and/or scattered, reflected and/or emitted from the specimen. Some means are provided for moving the specimen and adjusting the optics so that the specimen is both properly illuminated and properly placed in the field of view of the objective. In a microscope, a mechanical stage is used to position the specimen and to bring the region of interest into focus.

A microspectrophotometer was first made by putting a small "pinhole" aperture, or **field stop**, in the image plane of

a microscope, restricting the field of view to the area of a single cell, and placing a photodetector behind the field stop. The diameter of the field stop could be calculated as the product of the magnification of the objective lens and the diameter of the area from which measurements were to be taken. If a 40× objective lens were used, measuring the transmission through, or the absorption of, a cell 10 μm in diameter would require a 400 μm diameter field stop.

Using a substantially smaller field stop, it would be possible to measure the transmission through a correspondingly smaller area of the specimen; for example, a 40 μm field stop would permit measurement of a 1 μm diameter area of the specimen. By moving the specimen in the x and y directions (i.e., in the plane of the slide) in the raster pattern now so familiar to us from television and computer displays, and recording and adding the measurements appropriately, it was possible to measure the integrated absorption of a cell, and/or to make an image of the cell with each pixel corresponding in intensity to the transmission or absorption value. This was the first, and, at the time, the only feasible approach to **scanning cytometry**.

The use of stage motion for scanning made operation extremely slow; it could take many minutes to produce a high-resolution scanned image of a single cell, and there were no computers available to capture the data. Somewhat higher speed could be achieved by using moving mirrors, driven by galvanometers, for image scanning, and limiting the tasks of the motorized stage to bringing a new field of the specimen into view and into focus; this required some primitive electronic storage capability, and made measurements susceptible to errors due to uneven illumination across the field, although this could be compensated for.

Since the late 1940's and early 1950's had already given us Howdy Doody, Milton Berle, and the Ricardos, it might be expected that somewhere around that time, someone would have tried to automate the process of looking down the microscope and counting cells using video technology. In fact, image analyzing cytometers were developed; most of them were not based on video cameras, for a number of reasons, not the least of which was the variable light sensitivity of different regions of a camera tube, which would make quantitative measurements difficult. There was also the primitive state of the computers available; multimillion dollar mainframes had a processor speed measured in tens of kilohertz, if that, and memory of only a few thousand kilobytes, and this made it difficult to acquire, store, and process the large amount of data contained even in a digitized image of a single cell.

By the 1960's, a commercial version of Caspersson's microspectrophotometer had been produced by Zeiss, and several groups of investigators were using this instrument and a variety of laboratory-built scanning systems in attempts to automate analysis of the Papanicolaou smear for cervical cancer screening, on the one hand, and the differential white blood cell count, on the other^{42,43,52,53,57-60}. It was felt that both of these tasks would require analysis of cell images with reso-

lution of 1 μm or better, to derive measures of such characteristics as cell and nuclear size and shape, cytoplasmic texture or granularity, etc., which could then be used to develop the cell classification algorithms needed to do the job. Although it was widely recognized that practical instruments for clinical use would have to be substantially faster than what was then available, this was not of immediate concern in the early stages of algorithm development, and few people even bothered to calculate the order of magnitude of improvement that might be necessary.

Flow Cytometry and Sorting: Why and How

Somewhat simpler tasks of cell or particle identification, characterization, and counting than those involved in Papanicolaou smear analysis and differential white cell counting had attracted the attention of other groups of researchers at least since the 1930's. During World War II, the United States Army became interested in developing devices that could rapidly detect bacterial biowarfare agents in aerosols; this would require processing a relatively large volume of sample in substantially less time than would have been possible using even a low-resolution scanning system. The apparatus that was built in support of this project²⁹⁻³¹ achieved the necessary rapid specimen transport by injecting the air stream containing the sample into the center of a larger (*sheath*) stream of flowing air, confining the particles of interest to a small region in the center, or *core*, of the stream, which passed through the focal point of what was essentially a dark-field microscope. Particles passing through the system would scatter light into a collection lens, eventually producing electrical signals at the output of a photodetector. The instrument could detect at least some *Bacillus* spores, objects on the order of 0.5 μm in diameter, in specimens, and is generally recognized as having been the first flow cytometer used for observation of biological cells; similar apparatus had been used previously for studies of dust particles in air and of colloidal solutions.

By the late 1940's and early 1950's, the same principles, including the use of *sheath flow*, as just described, for keeping cells in the center of a larger flowing stream of fluid, were applied to the detection and counting of red blood cells in saline solutions⁴⁸. This paved the way for automation of a diagnostic test notorious for its imprecision when performed by a human observer using a counting chamber, or *hemocytometer*, and a microscope.

Neither the bacterial counter nor the early red cell counters had any significant capacity either for discriminating different types of cells or for making quantitative measurements. Both types of instrument were measuring what we would now recognize as side scatter signals; although larger particles would, in general, produce larger signals than smaller ones composed of the same material, the correlations between sizes and signal amplitudes were not particularly strong. In the case of the bacterial counter, a substantial fraction of the spores of interest would not produce signals detectable above background; the blood cell counters had a

similar lack of sensitivity to small signals, which was advantageous in that blood platelets, which are typically much smaller than red cells, would generally not be detected. White cells, which are larger than red cells, would be counted as red cells; however, since blood normally contains only about 1/1000 as many white cells as red cells, inclusion of white cells in the red cell count would not usually introduce any significant error.

An alternative flow-based method for cell counting was developed in the 1950's by Wallace Coulter⁴⁹. Recognizing that cells, which are surrounded by a lipid membrane, are relatively poor conductors of electricity as compared to the saline solutions in which they are suspended, he devised an apparatus in which cells passed one by one through a small (< 100 μm) orifice between two chambers filled with saline. A constant electric current was maintained across the orifice; when a cell passed through, the **electrical impedance** (similar to **resistance**, which is the inverse of **conductance**) increased in proportion to the volume of the cell, causing a proportional increase in the measured voltage across the orifice. The Coulter counter was widely adopted in clinical laboratories for blood cell counting; it was soon established that it could provide more accurate measurements of cell size than had previously been available⁵⁰⁻¹.

In the early 1960's, investigators working with Leitz⁶¹ proposed development of a hematology counter in which a fluorescence measurement would be added to the light scattering measurement used in red cell counting. If a fluorescent dye such as acridine orange were added to the blood sample, white cells would be stained much more brightly than red cells; the white cell count could then be derived from the fluorescence signal, and the raw red cell count from the scatter signal, which included white cells, could, in theory, be corrected using the white cell count. It was also noted that acridine orange fluorescence could be used to discriminate mononuclear cells from granulocytes. However, it does not appear that the device, which would have represented a new level of sophistication in flow cytometry, was ever actually built.

A hardwired image analysis system developed in an attempt to automate reading of Papanicolaou smears had been tested in the late 1950's; although it was nowhere near accurate enough, let alone fast enough, for clinical use, it showed enough promise to encourage executives at the International Business Machines Corporation to look into producing an improved instrument.

Assuming this would be some kind of image analyzer, IBM gave technical responsibility for the program to Louis Kametsky, who had recently developed a successful optical character reader. He did some calculations of what would be required in the way of light sources, scanning rates, and computer storage and processing speeds to solve the problem using image analysis, and concluded it couldn't be done that way.

Having learned from pathologists in New York that cell size and nucleic acid content should provide a good indica-

tor of whether cervical cells were normal or abnormal, Kamensky traveled to Caspersson's laboratory in Stockholm and learned the principles of microspectrophotometry. He then built a flow cytometer that used a transmission measurement at visible wavelengths to estimate cell size and a 260 nm UV absorption measurement to estimate nucleic acid content^{1,65}.

Subsequent versions of this instrument, which incorporated a dedicated computer system, could measure as many as four cellular parameters⁷⁸. A brief trial on cervical cytology specimens indicated the system had some ability to discriminate normal from abnormal cells⁷⁷; it could also produce distinguishable signals from different types of cells in blood samples stained with a combination of acidic and basic dyes, suggesting that flow cytometry might be usable for differential leukocyte counting.

Although impedance (Coulter) counters and optical flow cytometers could analyze hundreds of cells/second, providing a high enough data acquisition rate to be useful for clinical use, scanning cytometers offered a significant advantage. A scanning system with computer-controlled stage motion could be programmed to reposition a cell on a slide within the field of view of the objective, allowing the cell to be identified or otherwise characterized by visual observation; it was, initially, not possible to extract cells with known measured characteristics from a flow cytometer. Until this could be done, it would be difficult to verify any cell classification arrived at using a flow cytometer, especially where the diagnosis of cervical cancer or leukemia might be involved.

This problem was solved in the mid 1960's, when both Mack Fulwyler⁶⁷, working at the Los Alamos National Laboratory, and Kamensky, at IBM⁶⁶, demonstrated cell sorters built as adjuncts to their flow cytometers. Kamensky's system used a syringe pump to extract selected cells from its relatively slow-flowing sample stream. Fulwyler's was based on ink jet printer technology then recently developed by Richard Sweet⁶⁸ at Stanford; following passage through the cytometer's measurement system (originally a Coulter orifice), the saline sample stream was broken into droplets, and those droplets that contained cells with selected measurement values were electrically charged at the droplet breakup point. The selected charged droplets were then deflected into a collection vessel by an electric field, while uncharged droplets went, as it were, down the drain.

Fluorescence and Flow: Love at First Light

Fluorescence measurement was introduced to flow cytometry in the late 1960's as a means of improving both quantitative and qualitative analyses. By that time, Van Dilla et al⁷⁹ at Los Alamos and Dittich and Göhde⁸³ in Germany had built fluorescence flow cytometers to measure cellular DNA content, facilitating analysis of abnormalities in tumor cells and of cell cycle kinetics in both neoplastic and normal cells. Kamensky had left IBM to found Bio/Physics Systems, which produced a fluorescence flow cytometer that was the first commercial product to incorporate an argon ion

laser; Göhde's instrument, built around a fluorescence microscope with arc lamp illumination, was distributed commercially by Phywe.

Leonard Herzenberg and his colleagues², at Stanford, realizing that fluorescence flow cytometry and subsequent cell sorting could provide a useful and novel method for purifying living cells for further study, developed a series of instruments. Although their original apparatus⁸², with arc lamp illumination, was not sufficiently sensitive to permit them to achieve their objective of sorting cells from the immune system, based on the presence and intensity of staining by fluorescently labeled antibodies, the second version⁸⁶, which used a water-cooled argon laser, was more than adequate. This was commercialized as the FACS in 1974 by a group at Becton-Dickinson (B-D), led by Bernard Shoor.

By 1979, B-D, Coulter, and Ortho (a division of Johnson & Johnson that bought Bio/Physics Systems) were producing flow cytometers that could measure small- and large-angle light scattering and fluorescence in at least two wavelength regions, analyzing several thousand cells per second, and with droplet deflection cell sorting capability. DNA content analysis was receiving considerable attention as a means of characterizing the aggressiveness of breast cancer and other malignancies, and monoclonal antibodies had begun to emerge as reagents for dissecting the stages of development of cells of the blood and immune system. Instruments with two lasers were used to detect staining of cells by different monoclonal antibodies conjugated with spectrally distinguishable dyes.

Image cytometers existed; they were much slower and even less user-friendly than the early flow cytometers, weren't easily adapted for immunofluorescence analysis, and couldn't sort. Meanwhile, the early publications and presentations based on flow cytometry and sorting created a large demand for cell sorters among immunologists and tumor biologists. By the early 1980's, when a mysterious new disease appeared, best characterized – using flow cytometry and monoclonal antibodies – by a precipitous drop in the numbers of circulating T-helper lymphocytes, clinicians, as well as researchers, had become anxious to obtain and use fluorescence flow cytometers – and, often, to avoid sorting!

In the decades since, confocal microscopes, scanning laser cytometers, and image analysis systems have come into use. They can do things flow cytometers cannot do; they typically have better spatial resolution and can be used to examine cells repeatedly over time, but they cannot analyze cells as rapidly, and there are many fewer of them than there are flow cytometers. They are also, unlike flow cytometers, not subject to:

*Shapiro's First Law of Flow Cytometry:
A 51 μ m Particle CLOGS a 51 μ m Orifice!*

That notwithstanding, in these first years of the 21st century, most cytometry is flow cytometry, and, for almost all

applications except clinical hematology analysis, flow cytometry involves fluorescence measurement.

Fluorescence and flow are made for each other for several reasons, but primarily because fluorescence, at least from organic materials, is a somewhat ephemeral measurement. Recall that fluorescence occurs when a photon is absorbed by an atom or molecule, raising the energy level of an associated electron to an excited state, after which a small amount of the energy is lost as heat, and the remainder is emitted in the form of a longer wavelength photon, as fluorescence. However, there is a substantial chance that a photon at the excitation wavelength will not excite fluorescence but will, instead, **photobleach** a fluorescent molecule, producing a nonfluorescent product by breaking a chemical bond. In general, you can expect to get only a finite number of cycles of excitation and emission out of each fluorescent molecule (**fluorophore**) before photobleaching occurs.

If you look at a slide of cells stained with a fluorescent dye under a fluorescence microscope, you are likely to notice that, each time you move to a new field of view, the fluorescence from the cells in the new field is more intense than the fluorescence from the field that you had been looking at immediately before, which has undergone some photobleaching. This effect makes it difficult to get precise quantitative measurements of fluorescence intensity from cells in a static or scanning cytometer if you have to find the cells by visual observation before making the measurement, because the extent of photobleaching prior to the measurement will differ from cell to cell. In a flow cytometer, each cell is exposed to excitation light only for the brief period during which it passes through the illuminating beam, usually a few microseconds, and the flow velocity is typically nearly constant for all the cells examined. These uniform conditions of measurement make it relatively easy to attain high **precision**, meaning that one can expect nearly equal measurement values for cells containing equal amounts of fluorescent material; this is especially desirable for such applications as DNA content analysis of tumors, in which the abnormal cells' DNA content may differ by only a few percent from that of normal stromal cells.

A basis for the compatibility between fluorescence measurements and cytometry in general is found in the dark field nature of fluorescence measurements. It has already been noted that precise absorption measurements are best made when the concentration of the relevant absorbing material is relatively high. When one is trying to detect a small number of molecules of some substance in or on a cell, this condition is not always easy to satisfy. In the 1930's, unsuccessful attempts were made to detect antibody binding to cellular structures by bright field microscopy of the absorption of various organic dyes bound to antibodies. In 1941, Albert Coons, Hugh Creech, and Norman Jones successfully labeled cells with an antibody containing a fluorescent organic molecule⁴⁴, enabling structures binding the antibody to be visualized clearly against a dark background. In general, fluorescence measurements, when compared to absorption meas-

measurements, offer higher **sensitivity**, meaning that they can be used to detect smaller amounts or concentrations of a relevant analyte; this is of importance in attempting to detect many cellular antigens, and also in identifying genetic sequences and/or fluorescent protein products of transfected genes present in small copy numbers.

It is also usually easier to make simultaneous measurements of a number of different substances in cells, a process referred to as **multiparameter** cytometry, by fluorescence than by absorption, and the trend in recent years in both flow and static cytometry has been toward measurement of an increasingly large number of characteristics of each cell subjected to analysis, as can be appreciated from Table 1-1, way back on page 3.

Conflict: Resolution

When I first got into cytometry in the late 1960's, and for the next twenty years or so, there was a "farmer vs. rancher" feud going on between the people who did image analysis and the people who did flow, especially in the areas of development of differential white cell counters and Pap smear analyzers.

The first automated differential counters to hit the market were, in fact, image analyzers that scanned blood smears stained with the conventional Giemsa's or Wright's stains. Most of them are gone, now; modern hematology counters, which produce total red cell, reticulocyte (immature red cell), white cell, and platelet counts and red cell and platelet size (and, in at least one case, red cell hemoglobin) distributions, in addition to the differential white cell count, are typically flow based. Various instruments may measure electrical impedance (AC as well as DC), light absorption, scattering (polarized or depolarized), extinction, and/or fluorescence. None of them uses Giemsa's or Wright's stain.

Of course, with hundreds of monoclonal antibodies available that react with cells of the blood and immune system in various stages of development, we can use fluorescence flow cytometry to count and/or classify stem cells and other normal and abnormal cells in bone marrow, peripheral blood, and specimens from patients with leukemias and lymphomas, taking on tasks in hematology that few of the pioneers seriously believed could be approached using instruments. However, while the hematology counters run in a highly automated mode and produce numbers that can go directly into a hospital chart, most of the more sophisticated fluorescence-based analyses require considerable human intervention at stages ranging from the selection of a panel of antibodies to be used to the performance of the flow cytometric analysis and the interpretation of the results. This may facilitate reimbursement for the tests, but it leaves some of us unfulfilled, although perhaps better paid.

Cytometric apparatus that facilitated the performance and interpretation of the Papanicolaou (Pap) smear reached the market much later than did differential white cell counters. The first improvements were limited to automation of sample preparation and staining; there are now several image

analysis based systems approved for clinical use in aiding screening (locating cells and displaying images of them to a human observer), and at least one approved for performing screening itself. All use the traditional Papanicolaou stain, a witches' brew of highly nonspecific acidic and basic dyes known since the 19th century and blended for its present purpose before the middle of the 20th.

Why the difference? What made the Pap smear survive the smear campaign and the Wright's stained blood smear go with the flow? The answer is simple. Both Pap smear analysis and blood smear analysis on slides depend heavily on morphologic information about the internal structure of cells. Criteria for cell identification in these tasks may include cell and nuclear size and shape, cytoplasmic granularity or texture, and, especially in dealing with abnormal blood cells, finer details such as whether nucleoli or intracellular inclusions are present.

Some of these characteristics, e.g., cytoplasmic granularity (which, as has already been noted, is a major contributor to a side scatter signal), can be determined using flow cytometers. While the fluorescent antibodies used for such tasks as leukemia and lymphoma classification using flow are highly specific (although not, in general, specific to a single cell type), most of the instruments that perform the differential leukocyte count do not need to use particularly specific reagents. In fact, it is possible, using only a combination of polarized and depolarized light scattering measurements, to do a differential white cell count with no reagents other than a diluent containing a lysing agent for red cells.

In the case of differential leukocyte counting, we have learned to substitute measurements that can be made of whole cells in flow, requiring only low-resolution optics, for those that would, if we were dealing with a stained smear, require that we make and analyze a somewhat higher-resolution image of each cell. Flow is faster, simpler, and cheaper, and, although morphologic hematologists still look at stained smears of blood and bone marrow from patients in whom abnormal cells have been found, we no longer need to look at a stained smear by eye or by machine to perform a routine white cell differential count. Although there may be combinations of low-resolution flow-based measurements that could provide a cervical cancer screening test comparable in performance to Pap smear analysis, none have yet been clinically validated; we therefore still rely on image analysis in approaches to automation of cervical cancer cytology and on visual observation where automation is not available.

Researchers face problems similar to those faced by clinicians. If you want to select and sort the 2,000 cells out of 10,000,000 cells in a transfected population that express the most green fluorescent protein (GFP), you will probably use a flow cytometer with high-speed sorting capability and settle for a low-resolution optical measurement that detects all of the GFP in or on the cell without regard to its precise location. If you have arranged for the GFP to be coexpressed with a particular structural protein involved, say, in the for-

mation of the septum in dividing bacterial cells, you will very likely want to look at images of those cells at as high a resolution as you can achieve in order to get the information you need from the cells. There are, of course, tradeoffs.

It is January 1, 2002, as I write this, and therefore particularly appropriate to continue this New Year's resolution discussion. In the age of the personal computer and the digital camera and camcorder, there is little need to introduce the concepts of **digital images** and their component **pixels** (the term originally came from "picture elements"); most of us are exposed to at least 1024×768 almost 24/7. In this instance, the familiar 1024×768 figure describes the **pixel resolution** of an image acquisition or display device, with the image made up of 768 rows, each containing 1,024 pixels (or of 1,024 columns, each containing 768 pixels). However, the pixel resolution of the device doesn't, in itself, tell us anything about the **image resolution**, i.e., the area in the specimen represented by each pixel.

This depends to a great extent on what's in the image. In an image from the Hubble Space Telescope, each pixel could be light-years across; in an image from an atomic force microscope, each pixel might only be a few tenths of a nanometer (Ångströms) across. But the image resolution also depends on the combination of hardware and software used to acquire and process the image. We are free to collect a transmitted light microscope image of a $10.24 \mu\text{m}$ by $7.68 \mu\text{m}$ rectangle (close quarters for a single lymphocyte) somewhere on a slide containing a stained smear of peripheral blood, using a digital camera chip with 1024×768 resolution, but we are not free to assume that each of the pixels in the image represents an area of approximately 0.01 by $0.01 \mu\text{m}$. In this instance, the optics of the light microscope will limit our effective resolution to somewhere between 0.25 and $0.5 \mu\text{m}$, and using a camera with a high pixel resolution won't help resolve smaller structures any more than would projecting the microscope image on the wall. Either strategy provides what microscopists have long known as **empty magnification**; the digital implementation, by allowing us to collect many more bits worth of information than we need or can use, slows down the rate at which we can process samples by a factor of at least several hundred, and is best avoided.

So what do we do when we really need high-resolution images? As it turns out, one of the physical factors that limits resolution in a conventional fluorescence microscope, in which the entire thickness of the specimen is illuminated, is fluorescence emission from out-of-focus regions of the specimen above and below the plane of what we are trying to look at. In a **confocal microscope**, the illumination and light collection optics are configured to minimize the contributions from out-of-focus regions; this provides a high-resolution image of a very thin slice of the specimen. Resolution is improved further in **multiphoton confocal microscopy**, in which fluorescence is excited by the nearly simultaneous absorption of two or more photons of lower energy than would normally be needed for excitation. The illumina-

tion in a multiphoton instrument comes from a tightly focused high-energy pulsed laser, and it is only in a very small region near the focal spot of the laser that the density of low energy photons is sufficient for multiphoton excitation to occur. This produces an extremely high-resolution image (pixel dimensions of less than $0.1\ \mu\text{m}$ are fairly readily achieved), and also minimizes bleaching of fluorescent probe molecules and photodamage to cells.

As usual, we pay a price for the higher-resolution images. We are now looking at slices of the specimen so thin that we need to construct a three-dimensional image from serial slices of the specimen to fully visualize many cellular structures. Instead of two-dimensional pixels, we must now think in terms of three-dimensional **voxels**, or volume elements. Let's go back to the single lymphocyte which, on the blood smear discussed on the previous page, was confined in a two-dimensional, $10.24\ \mu\text{m}$ by $7.68\ \mu\text{m}$, rectangular area. For three-dimensional imaging, we would prefer that the cell not be flattened out, especially if we want to look at it while it is alive, so we will assume it to be roughly spherical, and imprison it in a cube $10\ \mu\text{m}$ on a side. If we used a multiphoton microscope with each voxel representing a cube $0.1\ \mu\text{m}$ on a side, building a 3-D image of that single cell would require us to collect data from $100 \times 100 \times 100$ voxels, or 10^6 voxels, and, even if it only took one microsecond to get data from each voxel, it would take a second just to collect the data.

This is a perfectly acceptable time frame for an investigator who needs information about subcellular structures; even with the computer time required for image processing, one can examine hundreds, if not thousands, of cells in a working day. However, even this is feasible only if the experimenter and/or the hardware and software in the instrument first scan the specimen at low resolution to find the cells of interest.

1.3 PROBLEM NUMBER ONE: FINDING THE CELL(S)

Continuing with the scenario just described, suppose we have cells at a concentration of $10^6/\text{mL}$, dispersed on a slide in a layer $10\ \mu\text{m}$ thick. A $1 \times 1\ \text{cm}$ area of the slide will contain 10,000 of the $10\ \mu\text{m}$ cubicles in which we could cache a lymphocyte. Recalling that $10\ \mu\text{m}$ is $1/1,000\ \text{cm}$, and that $1\ \text{cm}^3$ is $1\ \text{mL}$, we can calculate the aggregate volume of these 10,000 little boxes as $1/1,000\ \text{mL}$. If the cell concentration is $10^6/\text{mL}$, we can only expect to find about 1,000 cells in $1/1,000\ \text{mL}$, and it would take us 16 minutes, 40 seconds to scan all of them at high resolution. However, if we adopted the brute force approach and did 3-D scans over the entire $1 \times 1\ \text{cm}$ area, instead of finding the locations of the cells and restricting the high-resolution scanning to those regions, we would waste 9,000 seconds, or 2 hours and 15 minutes, scanning unoccupied cubicles.

There's another problem; although we may arbitrarily divide the $1 \times 1\ \text{cm} \times 10\ \mu\text{m}$ volume into $10\ \mu\text{m}$ cubicles, we have not created actual physical boundaries on the slide, and we can expect the cells to be randomly distributed over

the surface, which means that parts of the same cell could lie in more than one cubicle. If we deal with a specimen thicker than $10\ \mu\text{m}$ or so, the positional uncertainty extends to a third dimension, further compounding the problem of finding the cells, which gets even more difficult if we are trying to get high-resolution images of specific cell types in a tissue section, or in a small living organism such as a *Drosophila* embryo or a *C. elegans* worm.

When I first got into the cytometry game, in the late 1960's, my colleagues and I at the National Bureau of Standards and the National Institutes of Health built a state-of-the-art computerized microscope, with stage position and focus, among other things, under computer control⁷³. The instrument could be operated in an interactive mode, which allowed an experimenter to move the stage and focus the microscope using a small console that included a keypad and a relatively primitive joystick; the actual motion remained under computer control at all times. This made it possible to scan a slide visually, find cells of interest, store their locations in the computer, and have the instrument come back and do the high-resolution scans (resolution, in this instance, was better than $0.25\ \mu\text{m}$) needed for an experiment.

We didn't have a computer algorithm for finding cells automatically; since scanning the area immediately surrounding a cell took us not one second, but two minutes, there would have been little point to automating cell finding. The actual scanning time required to collect integrated absorption measurements of the DNA content of 100 cells, stained by the Feulgen method, was 3 hours, 20 minutes. We could find the cells that interested us by eye in a few minutes; scanning the slide looking for them might have taken days.

We were able to make life a little easier for ourselves by developing an algorithm to remove objects from the periphery of an image. A typical microscope field would contain a cell of interest, which we had positioned in the center of the field, surrounded by other cells, parts of cells, or dirt and/or other junk. Since the algorithm was relatively simple-minded, our visual selection process required us to exclude cells that touched or were overlapped by other cells. Figure 1-5, on the next page, shows the results of applying the algorithm.

The figure also shows how difficult it might be to develop algorithms to find cells. Even among the few cells present in the image shown, there are substantial differences in size and shape, and there are marked inhomogeneities in staining intensity within cells. Humans get very good very fast at finding cells and at discriminating cells from junk, even when cell size, shape, and texture vary. If staining (or whatever else produces contrast between the cell and the background) were relatively uniform, recognizing a cell by computer would be fairly easy; one would only have to find an appropriately sized area of the image in which all the pixel values were above a certain threshold level. This simple approach clearly won't work with cells such as those shown in the figure.

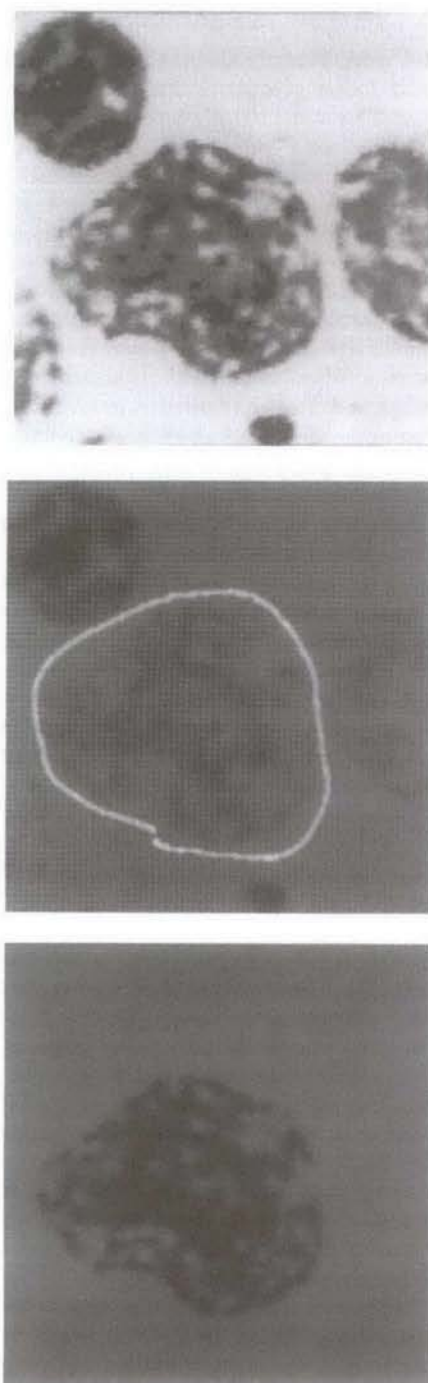


Figure 1-5. Top panel: scanned image of Feulgen-stained lymphoblastoid cells. In the middle panel, a boundary drawn around the cell of interest is shown; the bottom panel shows results of applying an algorithm to remove all objects except the cell of interest.

Before we developed the procedure for removing unwanted material from images, we had the option of picking out the cell of interest by drawing a boundary around it using a light pen, as shown in the middle panel of Figure 1-5. Many researchers working with cell images still find it convenient to locate cells and define boundaries for analysis in this fashion, and essentially the same procedure is used to draw the boundaries of regions of interest in two-parameter data displays from flow cytometers. This sidesteps the issue of automated cell finding (or of automated cluster finding, in the case of data displays). The boundary drawing is now commonly done using a personal computer and a mouse; in 1970, there were no mice, at least not the computer kind, and the interactive display and light pen we used cost tens of thousands of dollars, and had to be attached to the main-frame computer we needed to do the image processing. Very few laboratories could have afforded to duplicate our apparatus; today, you can introduce your children and grandchildren to the wonders of the microscopic world using a digital video microscope that costs less than \$100 and attaches to your computer's USB port. But, although your computer is probably hundreds of times faster than the one we used and has thousands of times the storage capacity, which could allow it to be used to implement cell finding algorithms of which we could only dream, it still takes a long time to capture high-resolution cell images, and the detail in those images makes it more difficult for those algorithms to define the boundaries of a cell or a nucleus than it would be if the images used for cell finding were of lower resolution.

A cell 10 μm in diameter occupies thousands of contiguous pixels in a high-resolution image with $0.1 \times 0.1 \mu\text{m}$ pixels, such as might be obtained from a multiphoton confocal microscope, but fewer than 100 contiguous pixels in a lower-resolution image with $1 \times 1 \mu\text{m}$ pixels, such as might be obtained from a scanning laser cytometer. The high-resolution image may contain many pixels with intensity near that of the background (as is the case with the image shown in the top panel of Figure 1-5), making it necessary to do fairly convoluted analyses of each pixel in the context of its neighbor pixels to precisely define the area of a cell or an internal organelle. However, each of the $1 \times 1 \mu\text{m}$ pixels of the lower-resolution image can be thought of as representing contributions from a hundred $0.1 \times 0.1 \mu\text{m}$ areas of the cell, and, since it is unlikely that all of these are at background intensity, it is apt to be easier to define an area as composed of contiguous pixels above a certain intensity level if one uses larger pixels.

When one is working with isolated cells, it becomes attractive to attempt to confine them to defined areas of a slide rather than to have to scan the entire surface to find cells distributed at random. By the 1960's, it had occurred to more than one group of investigators that depositing cells in a thin line on a glass or plastic tape would allow an automated cytology instrument to restrict stage motion to one dimension instead of two, potentially speeding up processing. The concept is illustrated in Figure 1-6 (next page).

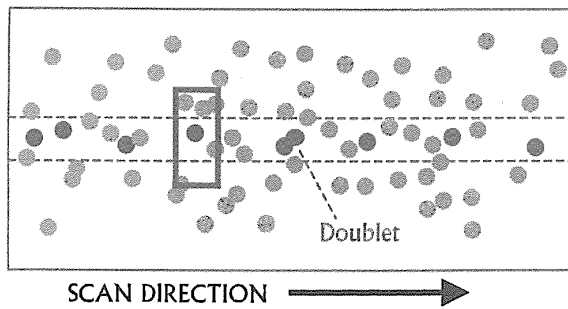


Figure 1-6. One-dimensional scanning of cells deposited in a narrow line (between dotted lines) on a slide or tape simplifies finding cells in a specimen. Black dots represent cells deposited in the line, gray dots represent cells deposited at random, and the small rectangle shows the field of view.

You can actually try this trick at home, if you happen to keep a microscope there, or in the lab, if you don't. Simulate the "cells" with dots in different colors made by a permanent marker with a fine or extra fine point; make dots in one color, corresponding to the black dots in the figure, along a straight edge placed parallel to the long edge of a slide, and make dots of another color (or enlist a [much] younger associate to do so), corresponding to the gray dots in the figure, all over the slide. Put the slide under the microscope, using a low- (10× or lower) power objective; place one of the "black" dots in the center of the field of view. Stop down the sub-stage iris diaphragm until you get a field a few times the diameter of the "cell." Then move only the horizontal stage motion control. You should note that, although the "black" cells you encounter as you scan along the slide in one dimension remain entirely in the field of view (up to a point; if the line along which you scribed wasn't exactly parallel to the edge of the slide, there will be some drift), you will almost certainly find "gray" cells cut off at the edges of the field of view. Now, looking at Figure 1-7, we can consider what a photodetector "looking" at the field of view would "see" if the slide in Figure 1-6 were scanned. We can regard this signal as a series of images, each made up of a single pixel that is considerably larger than the cells of interest.

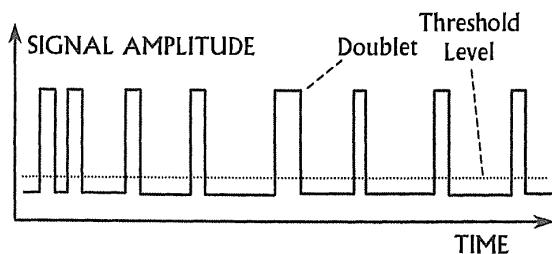


Figure 1-7. Idealized plot of signal amplitude vs. time representing a scan at constant speed along the cell deposition line of Figure 1-6; only signals corresponding to the "black" cells in that figure are shown.

For the moment, we can make believe the "gray" cells in Figure 1-6 aren't there; the simulated detector signal shown in Figure 1-7 only goes above threshold when it scans a "black" cell. This would actually happen if, for example, the slide were illuminated with blue light, and the "gray" and "black" cells were, respectively, unstained and stained with a green fluorescent dye. Figure 1-7 could then represent the electrical signal from a photodetector with a green filter in front of it.

When we look at the slide by eye, we don't scan very rapidly, and we almost never scan at uniform speed, so we don't instinctively relate what we see to the exact time at which we see it. When we scan with a cytometer, it is at least an advantage, and often an imperative, to scan at a constant speed, putting the times at which signals from objects appear at the detector output(s) in a fixed and precise relation to the positions of the objects in space.

In constructing Figure 1-7, the assumption was made that both the illumination intensity in the field of view and the scanning rate remained constant. If we look at the signal amplitude in the figure, it remains at a relatively low **baseline** level most of the time, and there are eight **pulses** during which the amplitude rises to a higher level and returns to the baseline value after a brief interval. If we glance up from Figure 1-7 to Figure 1-6, we notice that the positions of the pulses in time correspond to the positions of the black cells on the slide.

Flow Cytometry: Quick on the Trigger

The signal(s) used to detect cells' presence in the field of view (also called the **measurement point**, **region**, **station**, or **zone**, or the **analysis point**, **interrogation zone** or **point**, or **observation point**) of a cytometer is (are) called **trigger signal(s)**. The amplitude of a trigger signal must be substantially different in the cases in which a cell is and is not present at the observation point; in other words, it must be possible to define a **threshold level** above which the amplitude will invariably rise when a cell is present. If we pick a threshold level indicated by the dotted horizontal line in Figure 1-7, we see that the signal shown in the figure can serve as a trigger signal; its amplitude is well above the threshold level whenever a cell or cells are present in the field of view, and comfortably below that level when the field of view contains no cells.

Now, suppose that, instead of scanning cells deposited in a line on a slide or tape, we confine cells to the center of a flowing stream, and look at that through a microscope. We'll get rid of the gray cells this time, and only consider the black ones. And, if we want to draw a schematic picture of this, what we get is Figure 1-6, except that the gray cells aren't there, and the arrow indicates "Flow Direction" instead of "Scan Direction." Instead of defining the boundaries of the cell deposition area, the dotted lines define the diameter of the core stream containing the cells. We have sneakily built ourselves a **flow cytometer**.

Of course, if we were actually looking at the stream of cells in a flow cytometer, it would probably be flowing fast enough so that we couldn't distinguish the individual cells as they went by; remember that the visual system makes a "movie" out of images displayed at rates of 25-30/second (/s). Most photodetectors don't have this problem; they can respond to changes in light intensity that occur in nanoseconds (ns). So we could get a signal pretty much like the signal in Figure 1-7 out of a photodetector in a flow cytometer; the major difference would be in the time scale.

When scanning a slide by eye, we are apt to take at least 100 milliseconds (ms) to examine each cell; slide-scanning apparatus is substantially faster, producing pulse durations of hundreds of microseconds (μ s) or less. Flow cytometers are faster still; most current commercial instruments produce pulses with durations in the range between 0.5 and 12 μ s. Thus, the hardware and software responsible for detecting the presence of a cell need to do their job in a relatively short time, particularly in cell sorters, where the cell must be detected and analyzed, and the decision to sort it or not made and implemented, in the space of a few microseconds. If the signal in Figure 1-7 were coming from a detector in a flow cytometer, we could use it as a trigger signal.

Many of the signals of most interest to users of flow cytometers are of very low amplitude. Routine immunofluorescence measurements often require detection of only a few thousand fluorescently labeled antibody molecules bound to a cell surface. In such cases, the signal from the fluorescence detectors may be only slightly above background or baseline levels, and their use as trigger signals is likely to result in an unacceptably high level of **false triggering**, resulting in accumulation of spurious data values, due to the influences of stray light and electronic noise fluctuations. Even in cases when relatively weak fluorescence signals can be used as trigger signals to indicate the presence of stained cells, they will be of no help in detecting unstained cells. It has thus become customary to use a small-angle (forward) light scattering signal as the trigger signal when measuring immunofluorescence; all cells scatter light.

When none of the pulses from cells of interest are expected to be of high amplitude, requiring that a threshold level be set close to the baseline, discrimination of cells from background noise may be improved by using multiple triggers, requiring that two or more signals go above threshold at the same time to indicate a cell's presence. I almost always use forward light scattering and fluorescence as dual trigger signals when working with bacteria.

The Main Event

Looking back at Figures 1-6 and 1-7, though, we can see that there is another catch to triggering; it is not Catch-22, but Catch-2. Two of the black cells in Figure 1-6 are stuck together, and delineated as a "doubler" in that figure; the corresponding pulse, similarly delineated in Figure 1-7, is, though wider than the other pulses, still only a single pulse. Since cells going through a flow cytometer (or cells depos-

ited on a slide) arrive (or appear) at more or less random intervals, there is always the chance that two or more cells will be close enough in space, and their corresponding output signals close enough in time, so that they produce only a single pulse at the detector output. Note that the cells do not have to be physically stuck together for such **coincidences** to occur, they must simply be close enough so that the detector signal does not fall below the threshold value between the time the first cell enters the measurement region and the time the second (or last, if there are more than two) cell leaves it.

When we get technical about what we are really measuring in a flow cytometer (and now is one of those times), rather than saying that a pulse above threshold level represents a **cell**, we say that it represents an **event**, which might correspond to the passage of one cell, or multiple cells, or one or more pieces of noncellular junk capable of generating an equivalent optical/electronic signal, through the system, or which might result from stray light and/or electronic noise or some other glitch in the apparatus.

The Pulse Quickens; The Plot Thickens

There are ways of identifying pulses that result from coincidences; the height, width, and/or area of such pulses is/are typically different from those resulting from the transit of single cells, and, with the aid of appropriate hardware and/or software, it is possible to identify coincidences and correct counts. And now is probably an opportune time for me to confess that the pulses of Figure 1-7 are highly idealized, in that all of the pulses from single cells look pretty much the same; that definitely isn't the way things really are.

In fact, **all** of the information about a cell that can be gotten from flow cytometers is contained in, and must be extracted from, the **height**, or **amplitude**, the **area**, or **integral**, and the **width** and **shape** of the pulses produced at the detector(s) as the cell passes through the measurement region(s). Generally speaking, there isn't much point to doing flow cytometry if you expect all of the cells you analyze to look alike; the usual purpose of an experiment is the characterization of **heterogeneity** within a cell population, and the rest of this book is intended to help you make sure that the differences in pulses you see from cell to cell represent biological differences you are looking for, rather than reflecting vagaries of apparatus, reagents, and technique.

And now, at last, we have gotten our fingers on the pulse of flow cytometry. For the fact is that, while the information in scanning and imaging cytometers ultimately makes its way into the processing electronics in the form of a series of pulses, often referred to as a **pulse train**, it is only in flow cytometers and in the lowest resolution scanning devices that all of the information a detector gets about a cell (or, more accurately, an event) is contained in a single pulse. This was recognized early on as an important and distinctive characteristic of flow cytometry; before the term "flow cytometry" itself was coined in the 1970's, many workers in the field referred to it as **pulse cytophotometry**.

1.4 FLOW CYTOMETRY: PROBLEMS, PARAMETERS, PROBES, AND PRINCIPLES

Since the 1970's, it has become possible for users blissfully unconcerned with the nuts and bolts (or the atoms and bits) of instrumentation to buy flow cytometers capable of extracting more and more pulses from an increasingly diverse variety of objects, ranging downward from eukaryotic cells and microorganisms to organelles and large molecules, and upward to pancreatic islets, *C. elegans*, *Drosophila* embryos, and multicellular plankton organisms.

From reading the manufacturers' brochures and visiting their Web sites, interested researchers and clinicians can learn that it is possible to analyze and sort over a hundred thousand cells per second, to identify rare cells that represent only one of every ten million cells in mixed populations, to simultaneously measure light scattering at two or three angles and fluorescence in twelve or more spectral regions, to measure fluorescence with a precision better than one percent, and to detect and quantify a few hundred molecules of fluorescent antibody bound to a cell surface. It is somewhat harder to discern that it may be difficult or impossible to accomplish two or three of these amazing feats at once. If you're contemplating pushing the envelope, you definitely need to look at the **problem(s)** you're trying to solve, the measurement **parameters** and **probes** with which you can extract the necessary information from the cells, and the **principles** that may allow you to get your answers – or prevent you from getting them. I will take this approach in considering how the technology has gotten to its present state, starting with relatively simple problems and the relatively simple systems for solving them.

Since flow cytometers are designed to analyze single cells in suspension, it is not surprising that their development and evolution have been directed in large part by workers in the fields of hematology and immunology, who deal primarily with cells that are either in suspension, as is the case in blood samples, or relatively easy to get into suspension, as is the case when it is necessary to examine cells from bone marrow or lymphoid tissues or tumors.

In addition to being conveniently packaged, cells from the blood and immune system provide us with a number of models for fundamental biological processes. With the analysis of the genome behind us, we still need the details of differentiation that allow politically sensitive fertilized ova to develop through the politically sensitive embryonic stem cell stage into multicellular organisms who, after some years, can be dropped from the welfare rolls with the blessings of the same legislators who so staunchly defended them at smaller cell numbers. Cells in the blood and immune system develop from a single class of stem cells, which were hypothesized about and sought for years, and were finally identified with the aid of flow cytometry, and we now traffic in blood stem cells for patients' benefit as well as studying the cells' development in the interest of science. Differentiation gone wrong, with the aid of somatic mutation, produces leukemias and lymphomas, and we use flow cytometry both to

clarify the biology of neoplasia and to determine the prognosis and treatment in individual instances. The processes of clonal selection underlying both cellular and humoral immune responses provide a picture of evolution at work, as well as examples of a wide variety of mechanisms of inter- and intracellular signaling.

Counting Cells: Precision I (Mean, S.D., CV)

The simplest flow cytometers, and the first to be widely used, solved the **problem** of providing precise counts of the number of cells per unit volume of a sample, without explicitly characterizing the cells otherwise. Such instruments have only a single detector, and, because they measure an **intrinsic parameter**, typically **light scattering** or **extinction** or **electrical impedance**, do not require that the cells be treated with any reagent, or **probe**.

The **sample** used for cell counting may be taken directly from the **specimen** containing the cells, or may be an aliquot of that specimen **diluted** by a known amount, or **dilution factor**. If, for example, the specimen is diluted 1:20 to produce the sample, the dilution factor is 20.

The **principle** of operation of a cell counter is almost embarrassingly simple. An electronic counter is set to zero at the beginning of each run. Next, sample is passed through the system at a known, constant flow rate. As cells go through the measurement system, they produce pulses at the detector output; the count is increased by one whenever the output from the detector goes above the threshold level. Those cells that produce pulses with amplitudes above threshold are counted; those that do not are not. Any particle other than a cell that produces a signal above threshold is counted as a cell; any transient electrical disturbance or noise that causes the sensor output to go above threshold is also counted as a cell.

Although this sounds like a very simple-minded approach, it usually works, can be implemented using relatively primitive electronics, and can deal with thousands of cells per second. And, as will be amply illustrated later in this section, it is relatively easy to get from this point to a flow cytometer that makes one or several additional measurements of cells. The principal requirement is that, in addition to (or instead of) being used to increase the number in the counter, the trigger signal(s) initiate(s) the capture and recording of information about the height, area, and/or width of pulses from one or more detectors.

In the late 1950's and 1960's, the first optical and electronic (Coulter) cell counters reached the market. They were designed to count blood cells; I have already noted that red cell counts were done by setting a threshold high enough to prevent platelets from triggering, and that white cells were counted with red cells, but did not normally introduce significant inaccuracy into the red cell count because of their relatively low numbers. White cell counts were done on samples in which the red cells had been lysed by addition of a chemical such as saponin or one of a number of detergents to the diluent.

Before counters became available, people did cell counts by examining diluted blood (or another cell sample) in a **hemocytometer** under a microscope. A hemocytometer is a specially designed microscope slide with a ruled grid that defines square or rectangular areas, each fractions of a millimeter on a side, and with ridges on either side of the ruled area that insure that the thickness of the layer of diluted blood under the cover slip will be constant (usually 0.1 mm). For a white cell count, blood is typically diluted 1:20 with a solution that lyses red cells and stains white cells; the number of cells in four 1×1 mm squares is counted. The total volume of diluted blood counted is therefore 0.4 mm^3 , or $0.4 \text{ }\mu\text{L}$. To obtain the count of white cells/ mm^3 (the old-fashioned unit used when I was a medical student), one divides this number by 0.4 (the volume counted) and multiplies the result by 20 (the dilution factor). Because red cells are so much more numerous than white cells, blood is diluted 1:200 for red cell counts (without lysis, obviously), and a smaller area of the slide is used for counting.

Poisson Statistics and Precision in Counting

So what's wrong with hemocytometer counts, apart from the fact that they used to be done by slave labor (for which read medical students, or at least those of my generation)? The problem is with the **precision** of the counts. Precision, as was noted on p. 12, refers to the degree to which replicate measurements agree with one another. The precision of a measurement is often characterized by a statistic called the **coefficient of variation (CV)**, which, expressed as a percentage, is 100 times the **standard deviation (S.D.)** divided by the **mean** (and by **mean** I mean the **arithmetic mean**, or **average**, i.e., the sum of the individual measurements divided by the number of measurements). Well, you might say, "What mean and standard deviation? The count is only done once; how much time do you think those overworked medical students can spare?"

Enter another Student; not a 1960's medical student, this time, but a man of an earlier generation named William Sealy Gossett, who published his basic statistical works as "Student" because his employers at the Guinness Brewery worried that their competitors might improve their positions by using statistics if they discovered his identity. He showed in 1907²³¹⁷ that, if one actually counted n cells in a hemocytometer (that's before the division and multiplication steps), one should expect the standard deviation of the measurement to be the square root of n (I will use the notation $n^{1/2}$ rather than $\sqrt{n}$ for this quantity for typographic reasons), meaning that the coefficient of variation, in percent, would be $100/n^{1/2}$. We would now say that the statistics of counts conform to the **Poisson distribution**, which was described by Siméon Poisson in 1837²³¹⁸, but "Student" was apparently unaware of Poisson's work, and reached his conclusions independently. In fact, the Poisson distribution was only given that name seven years after Gossett's paper appeared²³¹⁹. We will encounter the Poisson distribution in several other contexts related to cytometry, flow and otherwise.

Now, if we consider looking at a sample with a white cell count of $5,000/\text{mm}^3$, which is in the normal range, the number of cells you would actually have counted in the hemocytometer to obtain that value would be 5,000 divided by the dilution factor (20) and multiplied by the volume (in mm^3) counted (0.4), which works out to 100 cells. The standard deviation would therefore be the square root of 100, or 10; the CV would be 10 percent. If you were dealing with an abnormally low white cell count, say one that you read as $1,250/\text{mm}^3$, you would only have counted 25 cells; the standard deviation would be 5, and the CV would be 20 percent. And all of this assumes that the counting process is perfect; we know that it isn't, and we also know that other factors, such as dilution and pipetting errors, will further decrease precision. So the precision of a hemocytometer white cell count in the normal range is barely acceptable. Getting a CV of 1 percent, which is more than respectable, would require that you count 10,000 objects, which would be 100 hemocytometers' worth if you were dealing with our original white cell count of $5,000/\text{mm}^3$. Nobody is going to sit there and do that by eye, but it's a piece of cake for an electronic or optical counter.

A typical hematology counter uses a **constant volume pump**, such as a syringe pump, to deliver sample at a constant flow rate. The flow rate is the volume of sample analyzed per unit time; dividing the number of cells counted per unit time by the flow rate gives the number of particles per unit volume. Blood specimens are usually diluted before being run in a counter, so the raw value must be multiplied by the dilution factor to get the particle count per unit volume of blood. For example, if the counter's sample flow rate is $1 \text{ }\mu\text{L/s}$, and a blood sample is diluted 1:20 (with a solution that lyses red cells) to count white cells, and running the counter for 40 seconds yields a raw count of 10,000 cells, the white cell count in the blood is:

$$\frac{10,000 \text{ (# of cells counted in 40 s)} \times 20 \text{ (dilution)}}{40 \text{ (# of }\mu\text{L counted in 40 s)}}$$

or $5,000/\mu\text{L}$. Since the raw count is 10,000, the standard deviation is 100, and the CV is 1 percent.

Rare Event Analysis: The Fundamental Things Apply as Cells Go By

Many of the tasks in modern cytometry are examples of **rare event analysis**. Examples are looking for primitive stem cells, leukemic cells or cancer cells in blood or bone marrow, for fetal cells in maternal blood, or for transfected cells present at low frequency in a culture. In comparing different samples, it is frequently necessary to determine the statistical significance of small differences between large numbers. Some people seem to think that counting hundreds of thousands or millions of cells lets them beat the Poisson statistics; what's important, however, is the number of cells of interest you count, not the total. Suppose, for example, that you find your cells of interest present at a frequency of 0.04% posi-

tives in one sample of 200,000 cells and 0.15% in another. Simple arithmetic tells you that 0.01% of 200,000 is 20 cells, so the first sample has 80 cells of interest and the second has 300. The Poisson standard deviations for the numbers of cells of interest counted would be about 9 for the 80 cells in the first sample and about 18 for the 300 cells in the second. The two values are thus separated by several standard deviations, which is to say that there is a statistically significant difference between them. However, the statistics provide no information as to the source of the difference. If the cells came from the same pot, one would suspect instrumental factors related to data collection and/or analysis, unless there is reason to believe that a process such as differential settling of the rare cell type would change the composition of a sample aliquot with time. A mild degree of paranoia is probably an asset when dealing with rare event analysis.

Poisson statistics apply to counting anything, from cells to photons and photoelectrons, and even to votes. Digressing briefly from rare event analysis to not-so-current event analysis, if 3,000,000 votes are counted, one expects a Poisson standard deviation of 1,732 votes, or roughly 6 parts in 10,000, meaning that if the vote counting process is supposedly even less reliable or accurate than Poisson statistics would predict (Florida's was said to be 99.9% reliable, or accurate to 10 parts in 10,000), neither candidate had a strong claim to having won the state's Presidential vote.

We have a little more control over cell counting than over vote counting. If you count enough cells, you can accurately discriminate between, say, .01% and .02%. If you only count 10,000 cells total, you'd expect to find one cell (and a CV of 100%) in the sample with .01% and 2 cells (CV of 70.7%) in the sample with .02%; so 10,000 cells total is too small a sample to let you discriminate. If you count 1,000,000 cells total, you end up with 100 cells in the .01% sample (10% CV) and 200 cells (7.1% CV) in the .02% sample, and this difference will be statistically significant.

Count Constant Numbers for Constant Precision

The best way to do counts, although almost nobody does them this way, is to always count the same number of cells of interest, which gives you equal precision no matter what the value is. Normally, we do absolute counts by analyzing a fixed volume of blood (or other sample) and percentage counts by analyzing a fixed number of cells. The alternative is to decide on the level of precision you want – suppose it is 5%. Then you have to count 400 cells (the square root of 400 is 20, and $100/20 = 5$). What you do is measure the volume of sample (in the case of absolute counts), or the total number of cells (in the case of percentage counts), which has to be analyzed to yield 400 of the cells of interest. If the cells of interest are at .01%, you'll have to count 4,000,000 cells total to find your 400 cells of interest; if they are at 1%, you'll only have to count 40,000 cells, but, instead of the .01% value being much less precise than the 1%

value, both will have the same 5% precision. The down side of doing things this way is that it may require some reprogramming of the apparatus, and probably uses more reagent, but, if you want good numbers, there is simply no better way to get them.

Alternative Counting Aids: The Venerable Bead

As it happens, most fluorescence flow cytometers do not use constant volume pumps for sample delivery, nor do they provide an alternative means of measuring the sample volume flow rate with sufficient precision to allow calculation of cell counts per unit volume by the method described above. Carl Stewart, being a leukocyte biologist, must have felt deprived of one of the major tools of his trade when he arrived at Los Alamos National Laboratory many years ago and discovered that the very fancy fluorescence flow cytometers built there did not provide a cell count. He and John Steinkamp solved that problem by adding fluorescent beads at known concentrations to cell samples¹⁵³⁹. If you have a bottle full of beads that contains a known number of beads per unit volume, adding a known volume of bead suspension (and it had better be well-mixed bead suspension) to a known volume of cell sample allows you to calculate the number of beads per unit volume in the sample. You can then run the sample for an arbitrary length of time, tallying the total numbers of beads and cells counted. The cell count per unit volume is then given by:

$$\frac{\# \text{ of cells counted} \times \# \text{ of beads per unit volume}}{\# \text{ of beads counted}}$$

and the number of cells per unit volume in the original material from which the cells were taken can be obtained by multiplying by the dilution factor, as in previous examples.

There are a few caveats here. If the determination of the concentration of beads per unit volume is done by a relatively imprecise method (Stewart and Steinkamp used a hemocytometer), the precision of the cell count cannot be improved by counting large numbers of cells and beads. One must also take into account the frequency of clumps and coincidences among both cells and beads, which affect the **accuracy** of the count, i.e., the degree to which the measured value agrees with the "true" value. And, of course, the cytometer must be capable of accurately identifying and counting both cells and beads.

Addition of beads to the sample is now widely practiced in the context of counting CD4 antigen-bearing (CD4-positive, or CD4+) T lymphocytes in HIV-infected individuals. The identification of these cells is most often done by staining with fluorescently labeled monoclonal anti-CD4 antibody (and, usually, at least one other monoclonal antibody labeled with a different fluorescent label). Before counting beads became available, the standard procedure was the so-called "two-platform" method, in which a hematology counter with a constant volume sample feed is used to obtain both the total white cell count per unit volume of

blood, and the differential white cell count, which includes the percentage of lymphocytes among the white cells; the number of lymphocytes per unit volume is then calculated. The fluorescence flow cytometer is used to define the lymphocyte population and the fraction of that population represented by CD4+ T-cells, allowing calculation of the number of these cells per unit volume. Using counting beads, the procedure can be done on a single platform, i.e., the fluorescence flow cytometer, and this appears to improve accuracy.

And Now to See with Eye Serene the Very Pulse of the Machine: Display, Digitization, and Distributions

In general, people who use flow cytometry want to know more about their samples than how many cells are contained in each milliliter, and that translates into getting more information about the signal pulses than whether their amplitudes exceed the threshold level. In a single-parameter electronic (Coulter) counter, the heights of pulses are proportional to the volumes of the cells passing through. However, whereas only relatively simple circuitry, triggered by the rise above threshold in the signal, is required to increment and store the cell count, more complex hardware and software are needed to capture and store measured values of the volumes of cells. Information about the measured particle may be extracted from the peak amplitude (height), the integral (area), the duration (width), and the shape of signal pulses.

The earliest electronic counters did not come equipped with the means to collect and display **distributions**, i.e., **histograms**, or **bar graphs**, of cell volumes; investigators interested in such information acquired it by feeding the pulse train from a counter into a gadget called a **pulse height analyzer**, a hardwired digital computer originally used by nuclear physicists to measure and discriminate among gamma ray energies.

The prerequisite to pulse height analysis, and to just about anything else that one might want to do in the way of data analysis in cytometry, is the conversion of information from an **analog** form, usually a **voltage** representing one of the pulse characteristics mentioned above, to a **digital** form, i.e., a **number**, using a device appropriately named an **analog-to-digital converter (ADC)**. Digital processing in the flow cytometers of the early 1970's was pretty much restricted to the use of pulse height analyzers, which had the disadvantage that their single ADCs (ADCs were expensive in those days) could only provide information on one measured quantity, or parameter, at a time. It was, however, possible to use live display and storage **oscilloscopes**, without benefit of digitization, to provide simple **dot plots** showing the interrelation of two parameters.

The pulses produced during a cell's passage through the measurement system typically last for only a few microseconds at most (making them veritable "phantoms of de light"), and, until recently, the only ADCs that could practically be used in flow cytometers required more time than this to digitize signals. As a result, it was necessary to use hybrid circuits, which combine analog and digital electron-

ics, to store the appropriate analog values for long enough to permit analog-to-digital conversion. These **peak detector**, **integrator**, and **pulse width measurement** circuits must be **reset** as each particle passes through the illuminating beam, allowing new analog signal levels to be acquired; it is then necessary to **hold** their outputs at a constant level until digitization is complete. The "reset" and "hold" signals must be delivered to the analog storage circuits at the proper times by additional hybrid "**front end**" electronics, which compare one or more trigger signal levels with preset threshold values to determine when a cell is present.

Luckily, a flow cytometer is an example of what is known as a **low duty cycle** device. Even when a sample is being run, cells pass the sensors rather infrequently; what goes by the sensors, most of the time, is the water or saline suspending medium, meaning that a certain amount of **dead time**, during which the pulse measurement circuits are occupied with data from one cell and cannot respond to signals from a second, is tolerable. Because cells arrive at random times, rather than at fixed intervals, **coincidences**, when a second cell arrives before processing of signals from the first is complete, are inevitable. The probability of coincidences can be calculated from – guess what – the **Poisson distribution**, and, while they cannot be eliminated entirely, it is possible to reduce them to acceptable levels by limiting the number of cells analyzed per unit time in accordance with the instrument's dead time.

Once held signals have been digitized, further analysis is accomplished with a digital computer, which, in modern instruments, is typically either an Intel/Microsoft-based or Apple Macintosh personal computer. The necessary software is now available from both flow cytometer manufacturers and third parties, in some cases at no cost. In recent years, as inexpensive, fast, high-resolution ADCs have become available (due largely to the needs of the consumer electronics and telecommunications markets), **digital signal processing (DSP)** hardware and software have replaced analog and hybrid circuits for peak detection, integration, and pulse width measurement, and for some other common tasks in flow cytometry, such as **fluorescence compensation** and **logarithmic conversion** of data. There will be a great deal more said about this further on in the book; for now, however, we will go back to another old problem, its old and newer solutions, and their implications for science, medicine, and society.

DNA Content Analysis: Precision II (Variance)

Most users of flow cytometers and sorters have at least a passing acquaintance with measurements of the **DNA content** of cells and chromosomes, which can be done rapidly and precisely by flow cytometry using a variety of fluorescent stains.

As a rule, all normal diploid cells (nonreplicating or G_0 cells and those in the G_1 phase of the cell cycle) in the same eukaryotic organism should have the same DNA content; this quantity is usually expressed as $2C$. DNA syn-

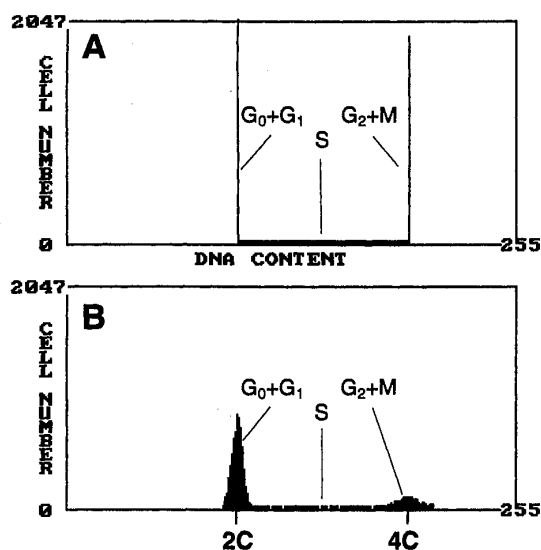


Figure 1-8. Ideal (A) and "real" (B) DNA content distributions, with the same ratios of $(G_0+G_1)/S/(G_2+M)$ cells represented in both.

thesis during the S phase of the cell cycle results in an increase in cellular DNA content, which reaches 4C at the end of S phase and remains at this value during the G_2 phase and during mitosis (M phase), at the completion of which the original cell has been replaced by two daughter cells, each of which has a DNA content of 2C. The haploid germ cells have a DNA content of approximately C; there are approximately equal populations of sperm with DNA content slightly greater than and slightly less than C due to differences in the DNA content of male and female sex chromosomes.

An idealized DNA content distribution, that is, a **bar graph** or **histogram** of values of DNA content that would be expected to be observed in a population of cells, some or all of which were progressing through the cell cycle, is shown in panel A of Figure 1-8, above. A "real" distribution, actually synthesized by a mathematical model, but more like those actually obtained from flow cytometry, appears in Panel B. Real (really real) DNA content distributions always exhibit some variance in the G_0/G_1 peak, which may be due to staining procedures, to instrumental errors, and/or to cell-to-cell differences in DNA content. The belief in the constancy of DNA content in diploid cells has been strengthened by the observation that the variances have diminished in magnitude with improvements in preparative and staining techniques and in instrumentation since the first DNA content distributions were published in the 1960's.

When I used the word "variance" in the paragraph above, I meant it, and you probably took it, to denote variability from measurement to measurement. However, the term also has a defined (and related) meaning in statistics; the **variance** of a set of measurements is the sum of squares of the differences between the individual measurements and

the **arithmetic mean**, or **average**, divided by one less than the number of measurements. In fact, the statistical **variance** is the square of the **standard deviation**, or, to put it more accurately, the standard deviation is the square root of the variance, and is calculated from it instead of the other way round. For purposes of this discussion, and in most of the rest of the book, I will try to use "variance" to mean the statistical entity unless I tell you otherwise. I may slip; word processors have spelling checkers and grammar checkers, but not intention checkers.

The Normal Distribution: Does the Word "Gaussian" Ring a Bell?

Although the number of cells counted does have some effect on the observed variance of a set of measurements, we are not dealing with Poisson statistics here; the variance of a Poisson distribution is not independent of the mean, but is always equal to it. The peak representing the G_0 and G_1 phase cells of a real DNA content distribution is generally considered to be best approximated by what statisticians define as a **normal** or **Gaussian** distribution, sometimes popularly known as a **bell curve**. The normal distribution is **symmetric**; the **arithmetic mean**, the **median** (the value separating the upper and lower halves of the distribution), and the **mode** (the highest point, or most common value) coincide. The coefficient of variation (CV) (which, you may recall, is expressed in percentage terms as 100 times the S.D. divided by the mean) remains a valid measure of precision, but there is an obvious problem in calculating the CV for a G_0/G_1 peak in a DNA content distribution. The peak falls off as one would expect on the left (low) side, but, on the right (high) side, it merges into the part of the distribution made up of S phase cells, and there isn't a convenient way to decide where the G_0/G_1 cells leave off and the S cells begin.

Because the anatomy of the normal distribution is well known and predictable, we have a statistical trick available to us. The width of the distribution between the two points on the curve at half the maximum (modal, mean, median) value, often referred to as the **full width at half maximum (FWHM)**, is 2.36 standard deviations, and the width between the two points at 0.6 times the maximum value is very nearly two standard deviations.

Binned Data: Navigating the Channels

The process of analog-to-digital conversion that occurs in a pulse height analyzer or in a modern flow cytometer's computer-based data acquisition and analysis system puts data into **bins**, to which we frequently refer as **channels**. These **binned data** are used to compile distributions of measured values of cellular parameters. The distributions in Figure 1-8 are broken into 256 channels, which, by convention, are numbered from 0 to 255. That is the number of bins, or channels, into which an 8-bit ADC distributes its output; an ADC with m bits resolution will have 2^m possible outputs, which, by convention, would be described as channels 0 to 2^{m-1} . Although the outputs of ADCs are often the

same unsigned binary numbers between 0 and 2^{n-1} that denote the channel numbers, ADCs with outputs in different binary formats are not uncommon. For our purposes, it is safest to think in terms of channel numbers, and leave the raw binary formats to the engineers and computer people.

Suppose that the maximum value, i.e., the largest number of cells, in the G_0/G_1 peak of such a distribution is 500 cells, occurring at channel 100, and that channels 97 and 103 each contain 300 cells (that is, 0.6 times the maximum number, 500). It is assumed here that each of the channels between 98 and 102 contains 300 or more cells, and that each of the channels below 97 and above 103 contains fewer than 300 cells. The width of the distribution at 0.6 times the maximum value, representing two standard deviations, is then 7 channels, one standard deviation is 3.5 channels, and the CV, expressed as a percentage, is $100 \times (3.5/100)$, or 3.5 percent. It is obviously easier for most people to calculate a CV in their heads using the width at 0.6 times maximum than it is using the width at half maximum, and a real piece of cake if you set the gain so that the maximum value ends up at channel 100, but we've all got calculators, anyway.

So what's the big deal about precision in DNA content measurement? To appreciate this, we go back to the 1960's again. The first cell counters had become available, and they were being used for counting and sizing blood cells. The 1960's also saw a great deal of progress in the field of tissue culture, resulting in substantial numbers of investigators having ready supplies of cells other than blood cells that were either in suspension or could conveniently be put into suspension. People became interested in the details of the cell cycle in cells derived from healthy tissues and from tumors, and in the effect of drugs on the cell cycle.

Once it became convenient to culture cells, it was possible to observe enough mitotic figures to establish that humans had 46 chromosomes, and not 48, as had once been believed, and to establish that cells from many tumors had more or fewer chromosomes, whereas cells from others had what appeared to be chromosomal deletions and translocations. This would mean that the amount of DNA in G_0/G_1 cells from a tumor could be different from the amount in G_0/G_1 cells from the normal stromal elements found in the tumor, potentially providing at least a means of identifying the tumor cells, and, possibly, an objective measurement with prognostic implications.

The catch here is that, as the difference you are trying to detect between two populations becomes smaller, you need better and better precision (lower CVs) in the measurement process. Generally speaking, two populations are resolvable if their means are a few standard deviations apart. If a tumor cell has one or two small chromosomes duplicated, adding, say, two percent to its G_0/G_1 DNA content, you would need a measurement process with a CV well under one percent to resolve separate G_0/G_1 peaks, although you might get a hint of the existence of two populations in a tumor specimen by observing broadening and/or **skewness (asymmetry)** in the peak of a distribution measured with a less precise process. A

triploid tumor population, with 50% more DNA than was found in stromal cells, could, of course, be resolved using very imprecise measurements.

DNA Content: Problem, Parameter, Probes

So, the **problem** became one of measuring DNA content with reasonably high precision. It was then necessary to find a suitable measurement **parameter** to solve it. Although Caspersson, in his microspectrophotometers, and Kamentzky, in his early flow cytometers, had used absorption at 260 nm for nucleic acid content measurement, the absorption measurements were difficult to make (among other things, they required special, very expensive quartz optics, because the UV wavelength used is strongly absorbed by glass), and not precise enough to detect small differences.

In the 1920's, Feulgen³⁵ developed a staining method that coupled a dye to the backbone of the DNA molecule, allowing DNA content in cells to be quantified by measuring absorption of visible light, but some fundamental problems with absorption measurements still limited the precision of DNA analysis. However, in one of the first publications describing fluorescence flow cytometry, in 1969, Van Dilla et al, at Los Alamos National Laboratory, reported the use of a modified **Feulgen procedure**, with fluorescent stains (acridavine and auramine O) and an argon laser source flow cytometer, to produce DNA content distributions with a coefficient of variation of 6% for the G_0/G_1 cell peak³⁹. The Feulgen staining procedure was relatively technically intensive, due to its requirement for fairly elaborate chemical treatment of the cells, and the search for dyes that were easier to use began almost immediately. The first step in this direction was taken in 1969, when Dittrich and Göhde published a relatively sharp DNA content distribution obtained using their arc source flow cytometer to measure the fluorescence of fixed cells stained with **ethidium bromide**⁸³. Thus, **fluorescence** became the **parameter** of choice for DNA content measurement.

Ethidium, which increases its fluorescence about thirty-fold when intercalated into double-stranded DNA or RNA, quickly replaced the fluorescent Feulgen stains as the **probe** of choice, and was then largely supplanted by a close chemical relative, **propidium**²¹⁷, which remains widely used as a DNA stain. Both dyes require that the cell be fixed, or that its membrane be permeabilized, in order to achieve good stoichiometric staining; they are frequently used to stain nuclei released from cells by treatment with one of a variety of **nonionic detergents**, such as Nonidet P-40 or Triton X-100²²³. Precise measurement of DNA in whole cells, and the best precision measurements in nuclei, require treatment of the sample with RNase to remove any residual double-stranded RNA.

Once cell sorters became available, in the 1970's, it was realized that a dye that could enter living cells and stain DNA stoichiometrically would make it possible to sort cells in different phases of the cell cycle and analyze their subsequent biological behavior and/or their chemical composi-

tion. Several *bis*-benzimidazole compounds originally synthesized as antiparasitic drugs by Hoechst AG turned out to meet these requirements²³⁸; the one that has been most widely used, by far, is **Hoechst 33342**²³⁹. This dye, like the other Hoechst dyes, has two characteristics that limit its use in some situations. Ultraviolet light is required to excite its blue fluorescence, preventing its use in the majority of fluorescence flow cytometers, which are equipped only with a 488 nm (blue-green) argon ion laser as a light source. And, although Hoechst dye staining is highly specific for DNA, the dyes, which do not intercalate but instead bind to the minor groove of the macromolecule, are selective for sequences of three **adenine-thymine (A-T)** base pairs²⁶⁸. The latter characteristic is disadvantageous for such applications as DNA content determination in plants, which is widely used as an aid in classification of species, because the Hoechst dyes would yield different results for two species having the same amount of DNA but different **base compositions**, i.e., different ratios of A-T and **G-C (guanine-cytosine)** base pairs. However, the base specificity of the Hoechst dye is an advantage in other circumstances; the combination of the A-T-selective Hoechst 33258 and G-C selective, DNA-specific dyes such as **chromomycin A₃** and **mithramycin**²³⁰, has been used to stain chromosomes from humans and other species, enabling chromosomes with similar total DNA content but different base composition to be distinguished and sorted separately²⁷⁸. **High-speed sorting** of dual-stained human chromosomes⁹⁰⁴ provided a valuable set of DNA libraries in the early phases of the Human Genome Project, but I'm getting ahead of myself. We can't get into that until we take at least a first look at one- and two-parameter data displays.

One-Parameter Displays: Pulse Height Distributions

The cells represented in Figures 1-9 are from the CCRF-CEM T-lymphoblastoid line. They were incubated with **Hoechst 33342**, which, as has already been mentioned, stains DNA stoichiometrically (neglecting, for the moment, differences in base composition). The cells were also exposed to **fluorescein diacetate (FDA)**, a nonfluorescent ester of fluorescein, which should more properly be called diacetylfluorescein but which almost never is. Both compounds are taken up by living cells; once inside cells, FDA is hydrolyzed by nonspecific esterases to **fluorescein**, which exhibits intense green fluorescence when excited with blue or blue-green light, and which, because of its anionic character, is retained in cells for minutes to hours. The cells were measured in a flow cytometer with two separated laser beams; they were first illuminated by a UV laser beam, and the blue fluorescence of Hoechst 33342 (panel A of Figure 1-9) excited by this beam was used as the trigger signal. The cells then passed through the 488 nm beam of a second laser, which provided excitation for the fluorescein fluorescence signal (panel B of Figure 1-9). The histograms of the distributions were collected at different times during a single sample run, using a multichannel pulse height analyzer. The

horizontal axis of each histogram indicates fluorescence intensity, on a 512-channel scale; the vertical axis of each histogram represents the number of cells with the corresponding fluorescence intensity. This, by the way, is not a historically informed modern performance on period instruments; the histograms are from around 1980, when one pulse height analyzer and a storage oscilloscope (see Figure 1-11) were all I had to work with for data capture and analysis.

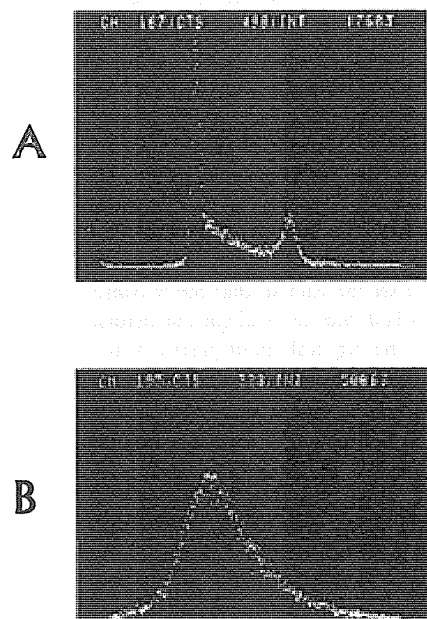


Figure 1-9. Two single parameter histogram displays from the oscilloscope screen of a multichannel pulse height analyzer. **A:** Fluorescence of the stoichiometric DNA stain Hoechst 33342. **B:** Fluorescence of intracellular fluorescein. Cells from the same sample are represented in the two histograms.

That said, the data are pretty respectable; their quality is determined primarily by the design and the state of alignment of the flow cytometer optics and fluidics. The CV of the G_0/G_1 peak of the histogram is about 3%, which is excellent for live cells stained with Hoechst 33342.

I have often described sharp peaks, such as the G_0/G_1 peak of a DNA content distribution, as being shaped like a **needle**. Such distributions are not common in flow cytometry data, unless one happens to be analyzing populations of objects that have been intentionally designed to be highly homogeneous, such as the fluorescent plastic microspheres used for instrument alignment and calibration. Although nuclei stained for DNA content, which exemplify one of the best of nature's own quality control processes, yield needles, the shapes of the distributions of most cellular parameters are closer to that of the fluorescein fluorescence distribution

in panel B of Figure 1-9, which resembles a haystack, in which it will be unlikely to find a needle.

The pulse height analyzer used to accumulate and display the histograms shown in Figure 1-9 is a specialized computer system that also incorporates some of the features of a flow cytometer's front end electronics and a peak detector. It can accept as input signal a train of pulses ranging in height from 0 to 10 V, using the input signal or another pulse train as a trigger signal, with a threshold set by the operator. Once a signal above threshold is encountered, the peak height is captured by the peak detector, and the signal is digitized by an ADC that, in this instance, produces a 9-bit output, i.e., a number between 0 and 511.

The pulse height analyzer stores its histograms in 512 memory locations. The **program**, or, more accurately, the procedure, or **algorithm**, for calculating a histogram is fairly simple. First, set the contents of all memory locations to zero. Then, every time a new numerical value emerges from the A-D converter, add one to the contents of the memory location corresponding to that numerical value. Stop when the total number of cells reaches a preset value.

This particular analyzer actually had several options on when to stop: at a preset value for the total number of cells, or for the number of cells in a single channel or memory location, or for the number of cells in a **region of interest**, a range of contiguous channels settable by the operator. It also had some refinements in its display; it would show the channel location of a cursor (CH) and the number of counts in that channel (CTS), as well as the total number of counts in the region of interest (INT). The histogram, sans numbers, could also be drawn on an X-Y plotter; several could be compared by eye in overlays using different color pens.

Pretty much the same algorithm is used for histogram computation today as was used in the analyzer. The difference is that in 1973, when the pulse height analyzer was built, a small startup company called Intel had just begun to ship samples of the first 4-bit microprocessor, and computer memory costs were on the order of 10 cents a byte. The smallest minicomputers available cost around \$10,000. The pulse height analyzer didn't have a central processing unit, couldn't process alphanumeric data, couldn't calculate a sine or a logarithm; it used special-purpose hardware to implement the algorithm, and, even at that and even then, it sold for about \$5,000. I'm not sure you can even buy a stand-alone pulse height analyzer today; instead, there are boards containing the necessary front end electronics that plug into standard personal computers. But, even if I could have afforded a second pulse height analyzer in 1980, it wouldn't have helped me do correlated analyses of two parameters.

Mathematical Analysis of DNA Histograms: If It's Worth Doing, It's Worth Doing Well

It was noted on p. 22 that, when one looks at a DNA content histogram, there isn't a convenient way to decide where the G_0/G_1 cells leave off and the S cells begin; there also isn't a convenient way to decide where the S cells leave

off and the G_2/M cells begin, or identify debris and cell aggregates in a sample, and things get worse in tumor samples.

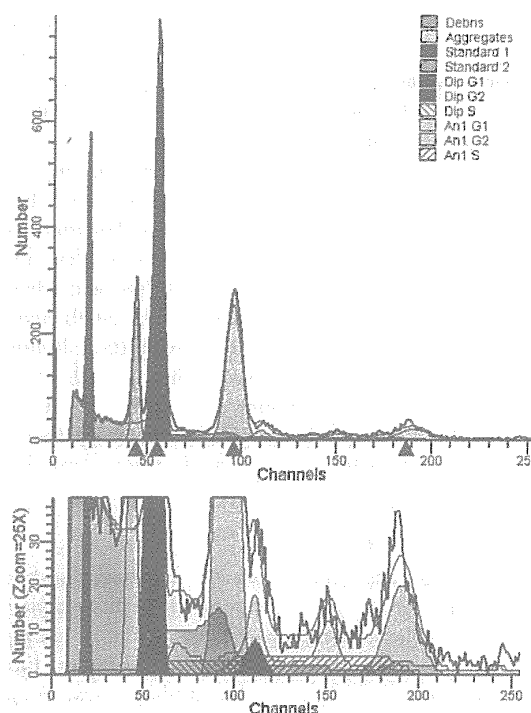


Figure 1-10. Use of a mathematical model to determine fractions of DNA-anneuploid breast cancer cells and normal stromal cells in different cell cycle phases in a sample from a tumor. Chicken and trout erythrocytes are added to the sample to provide standards with known DNA content. Contributed by Verity Software House.

Although tumor cells with abnormal numbers of chromosomes are correctly described as **aneuploid**, a tumor in which the neoplastic and stromal G_0/G_1 cells have different DNA contents is, by convention⁷⁴¹, referred to as **DNA aneuploid**. Mathematical models for DNA histogram analysis have been developed over the years, first, to estimate the fractions of cells in different cell cycle phases in an otherwise homogeneous population, and, later, to determine cell cycle distributions of both stromal cells and DNA aneuploid tumor cells. Further refinements allow for modeling of cellular debris and cell aggregates, enabling them to be largely excluded from analysis. An example of the application of one of the more sophisticated such models (ModFit LTTM, from Verity Software House) appears in Figure 1-10.

The earliest publications on fluorescence flow cytometry^{79,83} dealt with DNA analysis, and cancer researchers and clinicians began to use the technique almost immediately to attempt to establish the prognostic significance of both DNA aneuploidy and the fraction of cells in S phase in tu-

mors. The development of a method for extracting nuclei from paraffin-embedded tissue for flow cytometric analysis of DNA content^{610,891} allowed these issues to be approached by retrospective as well as by prospective studies. By the early 1990's, DNA analysis of breast cancer had come into reasonably widespread clinical use as a prognostic tool. However, in 1996, the American Society of Clinical Oncology recommended against the routine use of flow cytometry in breast cancer,²³²⁰ and the volume of specimens analyzed has declined substantially since then. Bagwell et al.²³²¹ have recently demonstrated, based on reanalysis of data from several large studies of node-negative breast cancer, that, after application of a consistent method of analysis and adjustment of some previously used criteria, DNA ploidy and S phase fraction again become strong prognostic indicators. This is not the only publication that shows that how and how well a laboratory test is done can profoundly affect its clinical significance, and that message is important whether or not flow cytometric DNA analysis comes back into vogue.

Linear Thinking

Noncycling cells with known DNA content, such as chicken and trout erythrocytes, can be added to a sample to serve as standards, as was done in the sample shown in Figure 1-10. Such standards are useful in establishing the linearity of the instrument and data acquisition system. A system is said to be linear when a proportional change in its input changes its output by the same proportion. In a simple DNA histogram, if the system is linear, and the mean or mode of the G_0/G_1 peak, representing cells with 2C DNA content, is at channel n , the mean or mode of the G_2/M peak, representing cells with 4C DNA content, will be at channel $2n$, or, because of the inherent error of ADCs, within one channel of channel $2n$. In practice, somewhat larger degrees of nonlinearity can be tolerated and corrected for, provided the nonlinearities are stable over time.

Lineage Thinking: Sperm Sorting

Since X- and Y-chromosomes in most species do not contain the same amount of DNA, one would expect a highly precise fluorescence flow cytometer to be able to distinguish them. The necessary precision has been achieved in high-speed sorters by modifications to flow chamber geometry and light collection optics, and sperm vitally stained with Hoechst 33342 have been successfully sorted by sex chromosome type and used for artificial insemination and/or *in vitro* fertilization in animals and, more recently and with a great deal more attention from the media, in humans²³²²⁻³. Gender selection in humans using sorted sperm, while still under attack from some quarters, is now deemed preferable to other methods that involve determination of the sex of preimplantation embryos. Gender selection in animals using the same methodology appears not to have generated as much controversy as has introducing a foreign gene or two into tomatoes, and may yet become big business²³²⁴.

Two-Parameter Displays: Dot Plots and Histograms

Histograms of the individual parameters do not provide any indication of correlations between Hoechst 33342 and fluorescein fluorescence values on a cell-by-cell basis. In modern flow cytometers, computer-based data acquisition and analysis systems make it trivial to capture, display, and analyze correlated multiparameter data from cells, but, until the 1980's, many instruments could only obtain correlated data on two parameters in the form of a display on an oscilloscope. Such a display was called a **cytogram** by Kamentsky and is now more commonly known as a **dot plot**. One showing both Hoechst 33342 and fluorescein fluorescence values for the cells from the same sample analyzed to produce Figure 1-9, appears in Figure 1-11, below.

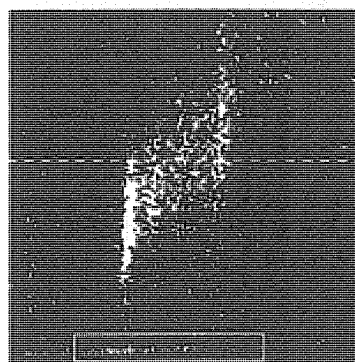


Figure 1-11. Dot plot (cytogram) of Hoechst 33342 fluorescence (x-axis) vs. fluorescein fluorescence (y-axis) for CCRF-CEM cells from the same sample shown in Figure 1-9. Cells in the box are dead; the dotted line is explained below.

Dot plots were the first, and remain the simplest, multiparameter displays in cytometry, and, as we shall presently see, tell us more than we could find out simply by looking at single-parameter histograms. In order to demonstrate this point, we should keep the histograms of Figure 1-9 in mind as we proceed.

In order to appreciate why two parameters are better than one, we need only look at the dot plot in Figure 1-11. One of the first things we notice is that cells with higher Hoechst dye fluorescence intensities, i.e., cells containing more DNA, show higher fluorescein fluorescence intensities. This shouldn't be surprising; if cells didn't get bigger during the process of reproduction, they'd eventually vanish, and it would seem logical that the amounts of FDA cells would take up, and the amounts of fluorescein they would produce and retain, would be at least roughly proportional to cell size. The horizontal dotted line across the dot plot defines two ranges of fluorescein fluorescence values that almost completely separate the diploid and tetraploid populations.

Even more significant, but less obvious to the untrained eye, are the cells represented in the box near the bottom of

the cytogram. These exhibit Hoechst 33342 fluorescence, but not fluorescein fluorescence; they are **dead cells**, or would be so defined by the criteria of a **dye exclusion test**. Such tests actually detect a breach in the cell membrane, which allows dyes such as propidium iodide and Trypan blue, which normally do not enter intact cells, to get in. In this case, the hole in the membrane allows the fluorescein produced intracellularly to leak out very rapidly. As a result, the dead cells exhibit little or no fluorescein fluorescence; their Hoechst dye fluorescence intensities remain indicative of their DNA content.

Dot plots, then, could readily generate an appetite for multiparameter data analysis capability which, given the state of instrumentation and computers in the early days of flow cytometry, was not readily satisfied. A few people could afford what were called **two-parameter pulse height analyzers**. These devices could produce distributions tabulating the number of events (cells, in this case) corresponding to each possible pair of values for two variables. They were about ten times the price of single-parameter pulse height analyzers; they also didn't have great resolution, due to the high cost of memory. Even if the two variables were digitized to only 6 bits' precision, with each yielding a number between 0 and 63, storage of the two-parameter, or **bivariate**, distribution would require 64×64 , or 4,096, memory locations. However, much of the information contained in bivariate distributions could be obtained, at much lower cost, by adding relatively simple **gating electronics** to the circuitry used to generate dot plots.

Multiparameter Analysis Without Computers: Gates Before Gates

Multiparameter analysis and **gating** may be the most important concepts in flow cytometry. Overall progress in the field was undoubtedly slowed during the 1970's and early 1980's because many of the people studying the really interesting biological problems didn't have either information about or access to the tools needed to implement even relatively simple multiparameter analysis and gating, let alone the sophisticated schemes that are now commonplace.

A dot plot, made using an oscilloscope, and demonstrating simple electronic gating, is shown in Figure 1-12. In order to understand how the gating works, we need first to consider how the dot plot is generated. An oscilloscope, like a television set, is built around a **cathode ray tube**. Electrons are accelerated toward a screen coated with a **phosphor** by the electric field generated by a high voltage applied between the cathode and the screen. The trajectory of the beam, i.e., the horizontal and vertical locations at which it will hit the screen, is determined by voltages applied to pairs of **deflection plates** inside the tube. A **modulation voltage** may be applied to control how much of the beam reaches the screen. Electrons that do reach the screen are absorbed by the phosphor, which subsequently emits some of the absorbed energy as light, by the process of **phosphorescence**.

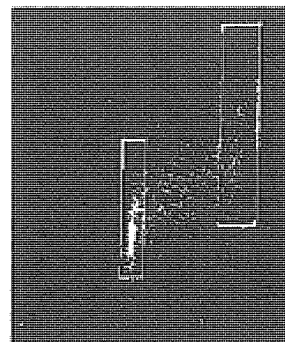


Figure 1-12. Gating regions for counting or sorting set electronically and drawn electronically on an oscilloscope display of a dot plot of DNA content (Hoechst 33342 fluorescence, shown on the x-axis) vs. RNA content (pyronin Y fluorescence, shown on the y-axis) in CCRF-CEM cells.

The dot plot above displays Hoechst 33342 fluorescence on the horizontal or x-axis, and the fluorescence of **pyronin Y**, which stains RNA, on the vertical or y-axis. To generate it, the output from the Hoechst dye fluorescence peak detector was connected to the horizontal deflection plate drive electronics, and the output from the pyronin fluorescence peak detector was connected to the vertical deflection plate drive electronics. The peak detector outputs are both **analog** signals; when applied to the deflection plates, they determine the x- and y- coordinates of the point at which the electron beam will hit the oscilloscope screen. Whether or not an intensified spot, representing the Hoechst 33342 and pyronin fluorescence values associated by the cell, is produced on the screen is determined by the oscilloscope's modulation voltage, which, in this instance, is controlled by what is called a **strobe** signal, generated by the same front end electronics that send the reset and hold signals to the peak detectors.

The strobe signal is a **digital** signal, or **logic pulse**, meaning that its output voltage values are in one of two narrow ranges, or **states**. In this case, voltages at or near about 5 volts (V) represent a "(logical) 1," or "on," or "true" output state, and voltages at or near 0 V, or **ground**, represent a "(logical) 0," or "off," or "false" output condition. The transitions between those two voltage ranges are made rapidly, which, in this instance, means within a small fraction of a microsecond; the interval required is known as the **rise time**.

Some systems use a **positive going** or **positive true** strobe signal, i.e., the strobe output is at ground when the strobe is "off" or "false" and at 5 V when the strobe is "on" or "true"; others use a **negative going** or **negative true** strobe signal, with the output at 5 V when the strobe is off and at ground when the strobe is on. The strobe signal described above is positive true.

The front end electronics are designed so that the strobe signal does not start until the analog signal value in the peak detectors, which can vary continuously between ground and 10 V, has stabilized, and the hold signal is applied to the peak detectors to keep this value from changing during the time the strobe is "on."

When a computer is used for data acquisition and analysis, the beginning, or **leading edge**, of the strobe signal is used to start analog-to-digital (A-D) conversions of the data in the peak detectors; when a dot plot is generated on an oscilloscope, no computer is used, and no digitization is done. The modulation electronics are set so that the beam will reach the screen when the strobe is on and not reach the screen when the strobe is off. Thus, every strobe signal received by the modulation electronics causes a dot to appear on the screen in a position corresponding to the values of the parameters represented on the x- and y-axes.

Early flow cytometers often used **analog storage oscilloscopes**, which incorporated special tubes, with long-persistence phosphors, and associated circuitry that could keep any region of the screen already intensified by the beam "on" until the screen was cleared, or erased, by the user. When an oscilloscope without such storage capacity was used, the dot plot could be recorded by taking a time exposure photograph of the screen.

A dot plot, whether it is recorded on an oscilloscope or using a digital computer (and today's oscilloscopes are increasingly likely to be special-purpose digital computers), does not contain as much information as a **bivariate distribution**. When you see a dot at a given position on the display, you know only that at least one cell in the sample had values of the two measured parameters corresponding to the position of the dot; and you can't get a better estimate of the actual number of cells that shared those values. That's where **gating** comes in. The strobe signal itself can be connected to a digital electronic **counter**, which will store a count and increase the count by one each time a strobe pulse is received. If the value in the counter is set to zero before analysis of a sample begins, the counter will maintain a tally of the total number of cells counted during the analysis.

Now, suppose we were interested in finding out how many of the cells in our dot plot had Hoechst dye fluorescence signals in the range between 3.5 and 4.75 V and pyronin fluorescence signals between 2.5 and 7 V. We could do this if we connected the relevant peak detector signals to an electronic circuit called a **window comparator**.

A **comparator** is a circuit element with two analog inputs, termed **positive** and **negative**, and a **digital**, or **logic level** (e.g., ground for "0" or "off"; 5V for "1" or "on") output. The digital output is on when the voltage at the positive input is higher than the voltage at the negative input, and off otherwise. Comparators are used in the analog front end circuitry of a flow cytometer to determine when the trigger signal (positive input) rises above the threshold level (negative input); one comparator is required for each trigger signal used.

A window comparator is made by connecting the logical outputs of four comparators together in a logical "AND" configuration. The inputs to the individual comparators are appropriate combinations of the two input signals and two sets of upper and lower limits such that the combined output is "on" only when both signals fall within the limits. The limits would typically be set by turning the knobs of **variable resistors**, or **potentiometers**, which are best known in their roles as volume controls in relatively unsophisticated and older radios and television sets.

Gating is accomplished by connecting the digital output of the window comparator to one input, and the digital strobe signal to the other input, of a purely digital circuit called an **AND gate**. The output of an AND gate is on only when both inputs are on; in this case, the output of the AND gate will be a pulse train containing only the strobe pulses from those cells with parameter values falling between the set limits.

While one counter, working off the strobe signal, is counting all the cells in the sample, another counter, connected to the output of the AND gate, accumulates a count of the cells falling within the gating region. The output of the AND gate can also be used as an input to the electronics that control cell sorting, allowing the cells with values within the set limits to be physically separated from the rest of the sample.

By incorporating a few other bits of analog and digital circuitry into the window comparator modules of my earliest "Cytomutt" flow cytometers, I could, at the press of a button, draw the boundaries of rectangular gating regions on the oscilloscope, as is shown in Figure 1-12; this greatly facilitated setting the upper and lower boundaries of the gating regions. Early commercial instruments had similar features. Of course, they were still limited to rectangular gating regions, and there were clearly situations when one could not separate the cells one wanted to count or sort from the unwanted cells using rectangular gates.

It was possible, by adding still more analog electronics to generate sums and differences of signals from two parameters, and feeding the sums and differences, rather than the original signals, into a window comparator, to define a gating region that corresponded to a parallelogram or other quadrilateral, rather than a rectangle, in the two-dimensional measurement space. This feature was incorporated in the instruments Kamensky built at Bio/Physics Systems in the early 1970's.

Kamensky also described, but did not put into production, a clever alternative counting/sort control circuit made by placing opaque black tape over all of an oscilloscope screen except the area corresponding to the gating region, and mounting a photodetector in front of the screen. The gating region defined in this manner could be any arbitrary shape, or even a set of disconnected arbitrary shapes, limited in size and scope only by the user's dexterity with scissors or a knife blade and black tape. Every time a cell lying within the region was encountered, the uncovered portion of the

screen was intensified, generating an output pulse at the photodetector that could be sent to a counter and/or used to initiate sorting. In the era of Bill Gates, we describe freeform gating regions of this type, implemented with mice and computers rather than blades and tape, as one type of **bit-mapped (or bitmap) gates**.

Well, most of the above is all ancient history, right? You must be wondering why I've devoted so much time to searching the souls of old machines when we do everything with computers now.

There are two reasons. The first is that the computers, in most cases, are doing the same things we did with hardwired electronics years ago, and if you understand how things worked then, you'll understand how they work now. The second is that there were, and still are, a few advantages to the old-fashioned electronics, especially for time-critical tasks.

I should mention that, then and now, nothing precluded or precludes us from defining a one-dimensional gating region, using either a simpler window comparator or a computer, and I did note that one-dimensional gating capability, allowing definition of a "region of interest," was typically built into pulse height analyzers. One-dimensional gating was widely used to control cell sorting in the earliest cell sorters, a logical choice when one considers that they typically measured only one relevant parameter.

Two-Parameter Histograms: Enter the Computer

As I wrote in 1994 for the 3rd Edition of this book, "Digital computers are extremely versatile. The same notebook computer on which I am writing this book with the aid of word processing software can be, and has been, used to acquire and analyze data from my flow cytometer. All I have to do is load and start a different program; I can even continue writing while I wait for the cytometer to get data from a new sample. Using additional telecommunications hardware and software, I can, and have, set gates on the cytometer, which is in Massachusetts, from a conference room in Maryland. However, while the computer's overall speed and its ability to switch rapidly between tasks make it appear as if it's doing many things at once, this is an illusion. About the only thing a computer can really do while it is running whatever program is occupying its attention is read or write data from or to a single source. Otherwise, digital computers do one thing at a time, even if they do that one thing really fast."

It's all still pretty much true. Of course, the notebook computer on which the 3rd Edition was written cost nearly \$5,000, weighed about seven pounds, had a 50 MHz 80486 processor, at most a couple of MB of RAM, a 500 MB disk drive, and a 640 by 480 pixel screen, and the one I use now cost about \$2,000, weighs three pounds, has a 750 MHz Pentium III processor, 256 MB of RAM, a 30 GB disk drive, and a 1024 by 768 pixel screen. The operating system and word processing software have also supposedly been improved. Last time around, my telecommunications were

limited to what I could do over standard telephone lines using a 9,600 baud modem; now, I gripe when my cable modem or DSL connections slow to even fifty times that speed. So, what I or you can do with a single computer can be done faster than what could be done eight years ago. But there's more; cytometry today can take advantage of both digital signal processing and multiprocessor systems in ways that, while obvious, were simply infeasible then.

A window comparator implemented in electronics is really making four comparisons at the same time, and they are accomplished in well under a microsecond. If you build a sorter using two window comparators to control deflection into left and right droplet streams, the two comparators work simultaneously. If you want to use a digital computer for sort control in a brute force kind of way, the computer has to fetch the value of the x-axis parameter for the left gating region, check it against the lower and upper bounds for that region, fetch the value of the y-axis parameter, check it against both bounds, and repeat the same steps for the right gating region. Obviously, the computation for a particular gating region can be stopped as soon as a parameter value is found to be out of bounds, but, if you think about it, the full four comparisons for one gating region or another have to be done for any cell that falls in either region, and, until they get done, no signal can be sent to initiate droplet deflection.

In droplet cell sorting, a sort decision has to be made within the few dozen microseconds it takes at most for a cell to get from the observation point to the point at which droplets break off from the cell stream. Up to 10 μ s may be required for the signals from the peak detectors (or integrators) to become stable. When hardwired electronics, e.g., window comparators, are used to control sorting, the sort decision signal is sent within a microsecond or so after this time. When a digital computer is used to control sorting, another time interval of at least a few μ s is required for A-D conversion before the computer can process the data. And, although the computers have gotten faster, the emergence of high-speed sorting has made it necessary for them to respond within even shorter time intervals.

Until the late 1970's, even minicomputers weren't really fast enough to be competitive with hardwired electronics for sort control. Today's much faster personal computers can easily accomplish the computations required for the window comparison described above well within the time period in which a sort decision must be made. The same computers, however, might not be able to get through a more complex computation, which, say, involved calculating four logarithms and solving quadratic equations to determine whether a cell falls in an elliptical gating region, in time to issue a sort signal, largely because while modern computer hardware is extremely fast, often requiring less than 1 ns to execute a machine instruction, the real time response of the hardware is literally slowed to a crawl by the design of the graphical user interface (GUI) based operating systems (various versions of Microsoft's Windows™, Linux with GUI

extensions, and Apple's Macintosh™ OS) now in most widespread use. The sorting problems are now solved by using some combination of external analog and digital electronics, frequently including one or more digital signal processors, or DSP chips, to implement time-critical processes such as sorting decisions, taking the load off the personal computer's central processing unit (CPU), leaving it free to do what it does best, namely, display the data informatively and attractively.

For plain old flow cytometric data analysis, in which there is no need to initiate action within a few microseconds after a cell actually goes through the beam, computers have always been better than hardwired electronics. That's why Kamensky used one in his original instrument at IBM. Computers for the rest of us only came along as we could afford them. A few lucky souls, myself included, had computers on their cytometers in the mid-1970's; they were minicomputers, and they were expensive. Now, it's virtually impossible to buy a flow cytometer that doesn't have at least one computer external to the box; most have one or more inside, as well.

Figure 1-13 is a histogram, collected, displayed, and annotated using my own competent, if ancient, MS-DOS-based 4Cyte™ data acquisition software, showing 90° (side) scatter values from a human leukocyte population. The data are plotted on a linear scale. The sample was prepared by incubating whole blood with fluorescently labeled antibodies to the CD3, CD4, and CD8 antigens, and lysing the erythrocytes by addition of an ammonium chloride solution. The "Cytomutt" cytometer used 488 nm excitation from an air-cooled argon ion laser, and measured forward and side scatter and fluorescence in 30 nm bands centered at 520 nm (green; principally fluorescence from anti-CD4 antibody labeled with fluorescein), 580 nm (yellow, principally fluorescence from anti-CD8 antibody labeled with the phyco-biliprotein, phycoerythrin), and 670 nm (red, principally fluorescence from anti-CD3 antibody labeled with a tandem conjugate of phycoerythrin and the cyanine dye Cy5). The forward scatter signal was used as the trigger signal.

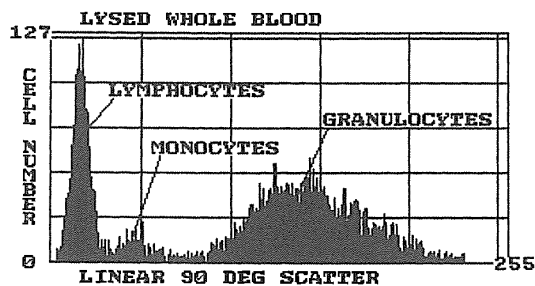


Figure 1-13. Histogram of 90° (side) scatter from leukocytes in lysed whole blood stained with fluorescent antibodies to lymphocyte antigens.

Modern Multiparameter Analysis: List Mode

The histograms and dot plots appearing in Figures 1-9, 1-11, and 1-12 are preserved for posterity only in the form of photographs. Figure 1-9A was photographed after the Hoechst dye fluorescence signal was connected to the pulse height analyzer and some 17,000 cells were run through the cytometer. The analyzer's memory was then cleared, the input was connected to the fluorescein fluorescence signal, and another 50,000 cells from the same sample were run through the instrument to generate the histogram of Figure 1-9B. The dot plots are taken from photos of the screen of an analog storage oscilloscope. I don't suppose the fact that we and a lot of other people stopped buying all of that Polaroid black-and-white film for our oscilloscope cameras loomed large in the company's eventually going bankrupt, but you never know. In the context in which we were using it, the film was a highly unsatisfactory archival medium.

The data represented in the histogram of Figure 1-13 were acquired in **list mode**, meaning that values of all parameters from all cells were stored in the computer's memory and, subsequently, on disk. List mode data acquisition doesn't preclude generating histograms, dot plots, or multivariate distributions while a sample is being run, and it does offer the user the considerable advantage of being able to reanalyze data well after they were acquired. The histogram in Figure 1-13 was generated months after the data were taken. Years ago, even after people had gotten used to having computers attached to their flow cytometers, they used to make a big fuss about acquiring data in list mode. There may have been some flimsy excuse for that attitude before mass storage media such as recordable CDs became available; today, there is simply no reason not to collect data from every run in list mode. Period. All currently available instruments have the necessary software for list mode data storage, and can write files compliant with one or another revision of the **Flow Cytometry Standard (FCS)** established by the Data File Standards Committee of the **International Society for Analytical Cytology (ISAC)**, an organization to which most serious flow cytometer users either belong or should. The standard makes it possible for analysis software from both cytometer manufacturers and third parties to read data from any conforming instrument.

As to the actual data in Figure 1-13, we notice that the histogram, like the DNA histograms in Figures 1-8B and 1-9A, is **multimodal**, meaning not that it has multiple identical maxima, but that it contains multiple peaks. Only one of these, that to the far left, would even be suspected of being a needle rather than a haystack. From the labels in Figure 1-13, it can be surmised that there is good reason to suspect that the peaks at increasingly higher values of 90° scatter represent lymphocytes, monocytes, and granulocytes; we can even go back to page 7 and look at Figure 1-2 to convince ourselves that this is the case. However, just as we can't readily separate the G_0/G_1 cells from the S cells, or the S cells from the G_2/M cells, by looking at a DNA histogram alone,

we can't readily separate the lymphocytes from the monocytes and the monocytes from the granulocytes by looking only at the histogram of 90° scatter.

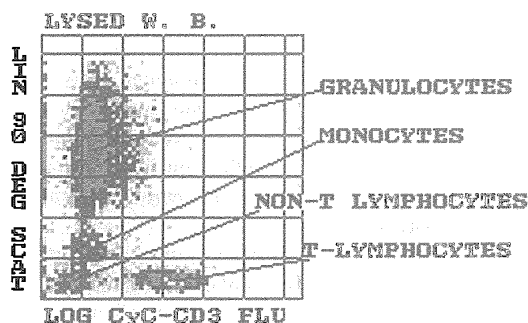


Figure 1-14. Bivariate distribution of anti-CD3 antibody fluorescence intensity vs. 90° (side) scatter for the same leukocyte population shown in Figure 1-13.

The picture gets a lot clearer when we look at the **bivariate distribution**, or **two-parameter histogram**, shown in Figure 1-14. The raw data in this distribution came from the same **list mode** file used to compute the histogram of 90° scatter shown in Figure 1-13; meaning that, thanks to the ready availability of computers and data storage media, we are able to look at the same cells from many different points of view. Figure 1-14 shows clearly identifiable **clusters** of cells; it provides a much clearer separation of lymphocytes, monocytes, and granulocytes than one could obtain using 90° scatter alone, and it also clearly separates the lymphocytes into those that bind the anti-CD3 antibody, i.e., the T cells, and those that do not, most accurately identified as "non-T" lymphocytes.

While a similar separation of cell clusters would be discernible on a dot plot, the bivariate distribution provides a more detailed picture of the relationship between two measured parameters, because the distribution provides an indication of the number of cells and/or the fraction of the cell population sharing the data values corresponding to each point in the two-dimensional measurement space, whereas the dot plot only indicates that one or more cells share the data values corresponding to a point in that space.

A bivariate distribution is computed by setting aside n^2 storage locations, where n is the number of bits of resolution desired for the data. Obviously, n cannot be greater than the number of bits of resolution available from the ADC; in practice, a lower value is typically used, for two reasons. First, the memory requirements are substantial. If each parameter has values ranging from 0 to 1,023, it is necessary to use 1,048,576 storage locations for a single distribution; this requires 2 megabytes if each location uses two bytes, or 16 bits, which would allow up to 65,535 cells or events to be tallied in any given location. If each location uses four bytes, or 32 bits, 4 megabytes of storage are required, but the

maximum number of cells that can be tallied in a location is increased to over 4 billion.

While the issue of memory requirements for distribution storage would seem moot at a time when a computer can be equipped with a gigabyte of RAM for a couple of hundred dollars, a second consideration remains. When a two-parameter histogram is computed at high resolution, it is usually necessary to include a very large number of events in order to have more than a few events in each storage location; computing at a lower resolution may actually make it easier to appreciate the structure of the data from smaller cell samples.

For a relatively long time, it was common to compute two-parameter histograms with a resolution of 64×64 ; these require 4,096 storage locations per histogram, which was a manageable amount of memory even in the early days of personal computers. Now, resolutions of 128×128 (16,384 storage locations) and 256×256 (65,536 storage locations) are widely available. The distribution displayed in Figure 1-13 has 64×64 resolution; values on a 1,024-channel scale, such as would be produced by a 10-bit ADC, would be divided by 16 to produce the appropriate value on a 64-channel scale, while the 8-bit (256 channels) values yielded by the lower-resolution converters found in older instruments would be divided by 4.

The data presentation format used in the display of Figure 1-14 is that of a **gray scale density plot**; the different shades of gray in which different points are displayed denote different numbers of cells sharing the corresponding data values. There is an alternative display format for density plots in which different frequencies of occurrence are represented by different colors instead of different shades of gray; this type of plot is described as a **chromatic** or **color density plot**. One can think of the gray levels or different colors in density plots as analogous to the scales that indicate different altitudes on topographical maps. Unfortunately, although the altitude scale is displayed on almost every published topographical map, the analogous scale of cell numbers or frequencies rarely finds its way into print alongside cytometric density plots.

Since computers now used for flow cytometric data analysis have color displays and color printers, chromatic plots are more common than gray scale plots. However, although color pictures are eye-catching and useful for presentations and posters, they can cost you money when included in publications. Those of us who run on lower budgets can almost always use a well-chosen gray scale for published displays without losing information; those lucky enough to not have to think about the cost of color plates might want to choose color scales that will not become uninformative when viewed by readers with defects in color vision.

Figure 1-15 (next page) displays the two-parameter histogram data of Figure 1-14 in an **isometric plot**, or **three-dimensional projection**, also commonly called a **peak-and-valley plot**; Figure 1-16 (next page) shows the same histogram as a **contour plot**.

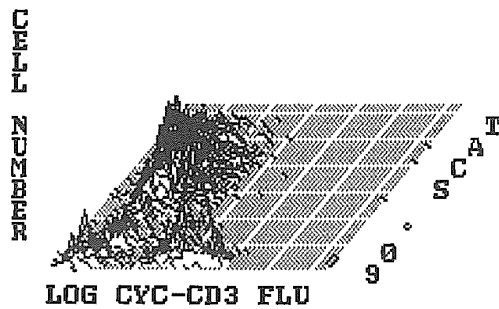


Figure I-15. The two-parameter histogram of Figure I-14 shown as an isometric or "peak-and-valley" plot.

In a peak-and-valley plot, a simulated "surface" is created; the apparent "height," or z-value, corresponding to any pair of x- and y-coordinates is made proportional to the frequency of occurrence of the corresponding paired data values in the sample. In a contour plot, a direct indication of frequency of occurrence is not given for each point in the x-y plane. Instead, a series of **contour lines**, or **isopleths**, are drawn, each of which connects points for which data values occur with equal frequency. A contour plot, like a density plot, resembles a topographic map; a peak-and valley plot is more like a relief map. The fact is, though, that there is really no more information in one type of bivariate histogram display than in another. Take this as a mantra.

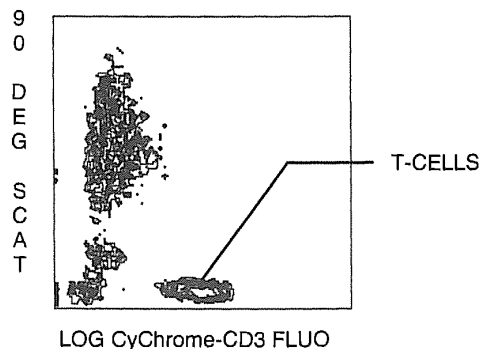


Figure I-16. The two-parameter histogram of Figures I-14 and I-15 displayed as a contour plot.

Peak-and-valley plots seem to have largely fallen out of favor; none too soon, say I. One of their major disadvantages lies in the *a priori* unpredictability of where peaks and valleys will turn up. Big peaks in the simulated "front" block the view of smaller peaks in the simulated "back," unless you use the 3-D graphics capabilities of your computer display to tilt and rotate the display. I suspect it's the dedicated computer gamers who are saving peak-and-valley plots from extinction.

Contour plots require more computation than either chromatic or peak-and-valley plots. Although contour plots may appear to have higher resolution, this appearance is deceiving, resulting as it does from the necessity to **smooth** the data, i.e., average over neighboring points, in order to get the plot to look respectable. Contour plots also do not normally show single occurrences, although these can be superimposed as dots on a contour plot; some such adaptation is essential when dealing with rare events. I think people tend to use contour plots in publications because they look more detailed than dot plots and often reproduce better than gray scale density plots.

I have always favored density plots, using chromatic plots for primary computer output and presentations, and, as a rule, gray scale plots for publication. I routinely use a **binary logarithmic intensity scale**, with one color or gray level indicating single occurrences, the next 2-3 occurrences, and subsequent colors indicating 4-7, 8-15, 16-31, 32-63, 63-127, and more than 127 occurrences. This makes it very easy to spot cells that occur with frequencies of less than one in 10,000.

Although commercially available flow cytometers are now equipped to display sixteen or more parameters (which would typically include light scattering at two angles and fluorescence in twelve spectral regions, with the balance possibly made up of different characteristics of the same pulses, such as width or height and area, and/or of ratios of the heights or areas of two signals from the same cell), almost all analysis is done using two-dimensional histograms or dot plots of two parameters in various combinations.

Three-Dimensional Displays: Can We Look at Clouds from Both Sides? No.

Humans aren't very good at visualizing spaces of more than three dimensions, but you'd certainly expect that, with everybody doing five- and six-parameter measurements in flow cytometry, three-dimensional displays would be commonplace. Some software packages produce a "three-dimensional dot plot," which I have called a **cloud plot** (see Figure 5-11, p. 241). Cloud plots have the same disadvantage as peak-and-valley plots; when one cloud gets in front of another, you have to recompute and change the viewing angle to see what's where. A few people have gone so far as to generate stereo pair images to improve the three-dimensional quality of the displays; they may be the same folks who have kept peak-and valley plots alive.

Three-parameter histograms are problematic for several reasons. First, even a $64 \times 64 \times 64$ 3-parameter histogram requires 262,144 storage locations, although there are some tricks that can reduce the storage requirements. Once you do, though, there's still a problem with how to display the data. Isometric plots would require four dimensions, which is out, and contour and chromatic plots demand x-ray vision on the part of the observer. As a result, what people have generally done when they need to represent something analogous to a 3-parameter histogram is show 2-parameter

histograms in "slices," with the resolution along the "sliced" axis often lower than that of the 2-parameter histograms, so that there might be four to sixteen 64×64 histograms rather than sixty-four. "Slicing" a histogram, if you stop to think about it, is exactly equivalent to defining a series of rectangular gating regions along the z-axis. And, speaking of gating regions, the slicing technique just mentioned is about the only practical way of setting gates in a three-dimensional space.

There's a lot of information to deal with when you're just looking at two-parameter displays; three-parameter displays could quickly get you to the point of information overload. If we are dealing with n parameters, the number of possible two-parameter displays, counting those showing the same two parameters with the x- and y-axis switched, and omitting those in which the same parameters are on both axes, is $n \times (n-1)$, while the number of possible three-parameter displays, counting those showing the same three parameters with axes switched and omitting those with the same parameter on two or three axes, is $n \times (n-1) \times (n-2)$. For the five-parameter data we have been looking at, we have 20 possible two-parameter displays and 60 possible three-parameter displays; for 16-parameter data, we would have 240 possible two-parameter displays, which is frightening enough, and a mind-boggling 3,360 three-parameter displays. It could take months to run an analysis on a single tube if we had to look at all of them. So, as usual, it is best to get our heads out of the clouds.

Identifying Cells in Heterogeneous Populations: Lift Up Your Heads, Oh Ye Gates!

Most of the interesting applications of flow cytometry involve identifying cells in heterogeneous populations; what varies from case to case is the basis of the heterogeneity. We have already noted several varieties of heterogeneity in our brief examination of DNA content analysis. Cells in a presumably pure, clonally derived, unsynchronized culture will contain different amounts of DNA because they are in different stages of the cell cycle. A DNA aneuploid tumor contains stromal and tumor cells with different G_0/G_1 DNA contents, and both stromal and tumor cells may be in different cell cycle phases. Sperm differ in DNA content depending on which sex chromosome is present. Heterogeneous populations of microorganisms such as are encountered in seawater contain many different genera and species, each with its characteristic genome size. In all of the above cases, it is possible to identify cell subpopulations based on differences in DNA content.

In the widely studied heterogeneous cell populations that comprise blood, the majority of cells are neither DNA aneuploid nor progressing through the cell cycle. Thus, when the **problem** is the identification of different cell types in blood, DNA content is generally not a **parameter** of choice. Figure 1-17, on the next page, illustrates the use of several better suited parameters and of multiple gating methods in one of the most common clinically relevant applications of

flow cytometry, the identification of T lymphocytes bearing CD4 and CD8 antigens in human peripheral blood.

If we simply stained cells with a combination of differently colored acidic and basic dyes, as Paul Ehrlich, who developed the basic technique, did in the late 1800's, we would be able to use transmitted light microscopy (with relatively strongly absorbing dyes at high concentrations) or fluorescence microscopy (with fluorescent dyes, probably at lower concentrations) to do a classical differential white blood cell count. The presence or absence of cytoplasmic granules would let us distinguish the granulocytes from the mononuclear cells (monocytes and lymphocytes). The relative amount of staining of those granules by the acidic and the basic dye would allow us to identify eosinophilic (acidophilic to Ehrlich), basophilic, and neutrophilic granulocytes. The size of the cells, amount of cytoplasm, and nuclear shape would allow us to distinguish most of the monocytes from most of the lymphocytes. But that's about as far as we would get. A typical peripheral blood lymphocyte is a small, round cell with a relatively thin rim of cytoplasm surrounding a compact, round nucleus. The nucleus, like the nuclei of all cells, stains predominantly with the basic dye (one of the methylene azure dyes in a typical Giemsa or Wright's stain), which is attracted to the acidic phosphate groups of the nuclear DNA. The basic cytoplasmic proteins attract some of the acidic dye (eosin in the mixtures commonly used for staining blood), but RNA in the cytoplasm also attracts the basic dye. And the staining pattern of most peripheral blood lymphocytes is pretty much the same, whether they are B lymphocytes or T lymphocytes, and, if T lymphocytes, whether they bear the CD4 or the CD8 antigen (although both antigens are present on developing T lymphocytes in the thymus, almost all of the T lymphocytes present in the peripheral blood have lost one or the other).

The optical flow cytometers used for differential white cell counting in hematology laboratories, which typically measure forward and side scatter, can distinguish lymphocytes from monocytes and granulocytes using these measurements alone, but cannot thereby distinguish different types of lymphocytes. However we have already seen from Figures 1-13 through 1-16 that the combination of side scatter measurements and measurements of fluorescence of cell-bound antibodies allows us to distinguish T lymphocytes from other lymphocytes. It should therefore come as no surprise that the **probes**, or reagents, that allow us to define lymphocyte subpopulations, and most other subclasses of cells in the blood, bone marrow, and organs of the immune system, are **antibodies**, and that we detect antibodies bound to cells by the **fluorescence** of labels attached, usually covalently, to the antibody molecules. Flow cytometry greatly facilitated the development of monoclonal antibody reagents, and flow cytometry has since been indispensable for defining the specificities of these reagents and, thereby, allowing their routine use for cell classification in clinical and research laboratories.

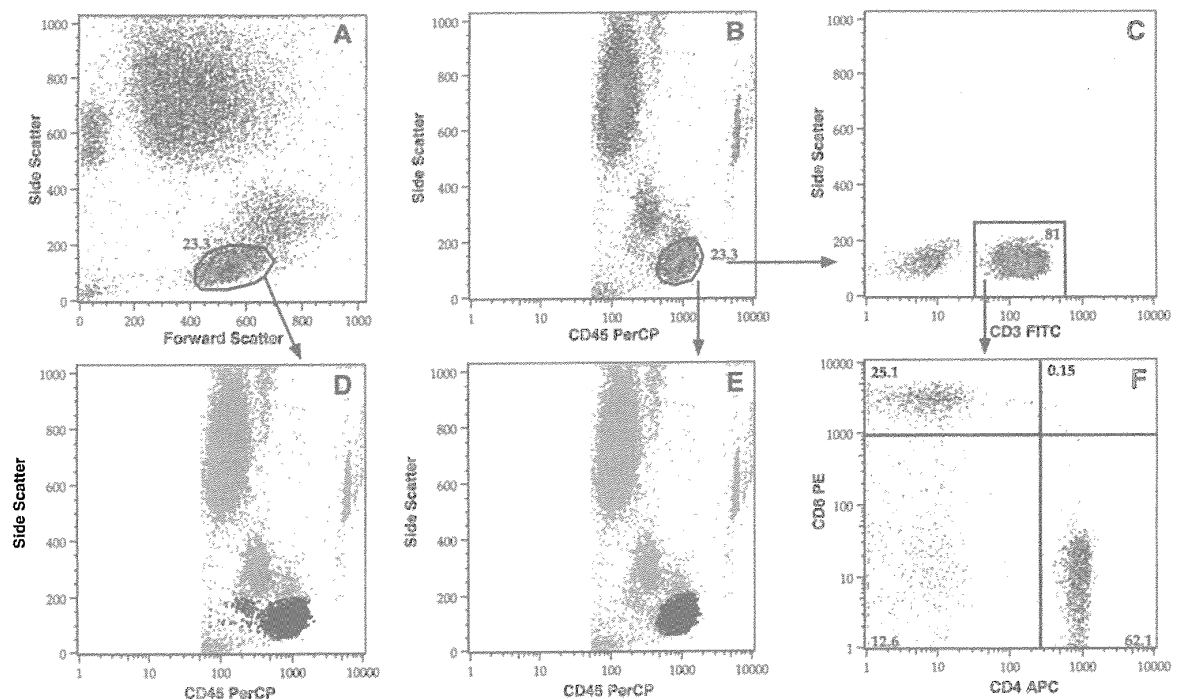


Figure 1-17. Identification of human peripheral blood T lymphocytes bearing CD4 and CD8 antigens. Data provided by Frank Mandy; analysis and displays done by Jennifer Wilshire using FlowJo software (Tree Star, Inc.).

The displays in Figure 1-17 show an older and a newer gating method for defining a lymphocyte population. The gates are drawn with the aid of a mouse or other pointing device. Flow cytometric software typically provides for several types of **bitmap gating**, which allows the user to define more or less arbitrarily shaped gating regions on a dot plot or two-parameter histogram. Almost all programs allow the user to draw **polygons** to define the boundaries of gating regions; most also allow definition of regions bounded by **rectangles**, **ellipses**, or **free-form curves**. While most cell sorters make use of no more than four gating regions at any given time, data analysis software typically provides for a larger number, to facilitate deriving counts of a reasonable number of cell subpopulations in heterogeneous samples such as are obtained from blood.

Cluster Headaches

The objective of gating is the isolation, in the measurement space, of a **cluster** of cells. The term **cluster** is used in flow cytometry (and in multivariate data analysis in general) to denote any relatively discernible, reasonably contiguous region of points in a bivariate display; that may sound imprecise, but there isn't any more precise definition. You're supposed to know a cluster when you see one.

In panel A of Figure 1-17, a polygonal gate is drawn around a cluster of cells with intermediate values of forward scatter and low values of side scatter; it was established by sorting experiments in the 1970's that most of the cells in

such a cluster were lymphocytes¹⁵⁷, and lymphocyte gating was incorporated into analysis of lymphocyte subsets at a fairly early stage in the game¹⁷⁵⁻⁶. However, there was some concern that cells other than lymphocytes might be found in the gate. If one were interested only in T lymphocytes, it would be possible, as Mandy et al demonstrated in the early 1990's¹⁰²⁷, to define a well isolated cluster of these cells on a display of anti-CD3 antibody fluorescence vs. side scatter (look back at Figures 1-13 to 1-16). This did not satisfy the HIV immunologists, who wanted to know not only the absolute number of CD4-bearing T cells per unit volume of blood, but also what percentage of total lymphocytes the CD4-bearing T cells represented. The current practice for defining a lymphocyte cluster uses a two-dimensional display of anti-CD45 antibody fluorescence vs. side scatter, as shown in panel B of Figure 1-17, taking advantage of the fact that lymphocytes have more CD45 antigen accessible on their surfaces than do monocytes and granulocytes¹²⁵¹.

Painting and White- (or Gray-) Washing Gates

The gates in panels A and B of Figure 1-17 have been painstakingly drawn so that each includes 23.3 percent of the total number of events (where events include cells, doublets, debris, and the counting beads added to the sample). We have decided to accept the CD45/side scatter gate in panel B as the "true" lymphocyte gate; the question then comes up as to whether the forward scatter/side scatter gate in Panel A contains cells that would not fit into this gate.

In order to answer this question, we need a way of finding the cells in the forward scatter/side scatter gate on a display of CD45 vs. side scatter. Most modern data analysis programs incorporate the means to do this; the user can associate a different color with each gate set, thus allowing cells falling within that gate to be distinguished on plots of parameters other than those used to set the gates. Becton-Dickinson's "Paint-A-Gate" program was one of the first to provide this facility.

When you are working on a low budget, and restricted to monochrome displays, you can always emulate Whistler and use shades of gray instead of colors, as has been done in panels D and E of Figure 1-17. In this instance, the cells from the gates in panels A and B have, respectively, been shown in black in panels D and E; all of the other cells appear in light gray. The panel D and E displays also use a convenient feature found in the FlowJo program; the dots can be, and here are, made larger. This can be useful when one tries to show very small subpopulations in dot plots, and, indeed, we see that there are a few cells from the forward scatter/side scatter gate of panel A that show up in panel D outside the "true" lymphocyte gate as defined using CD45/side in panel B. Of course, the cells from the gate in panel B remain in the same positions in panel E; since we started out assuming that the gate in panel B was the true gate, you can't really call that a whitewash.

Moving right along, in this case to panel C, we will look only at the cells from the lymphocyte gate defined in panel B, on a plot of anti-CD3 antibody fluorescence vs. side scatter. We can now draw a rectangular gate around those that bear the CD3 antigen; these are the T cells.

The Quad Rant: Are You Positive? Negative!

In panel F, the T cells defined by the gate in panel C are shown on a plot of anti-CD4 antibody fluorescence vs. anti-CD8 antibody fluorescence. This plot is broken into **quadrants**, i.e., four rectangular gating regions that intersect at a single central point. The percentages of events that fall in each of the quadrants are indicated. There are clear clusters of events with high levels of CD8 and low levels of CD4 and of events with high levels of CD4 and low levels of CD8, a small but respectable number of events with low levels of both, and a few events with high levels of both. At first, it seems as if all's right with the world. But maybe there's a problem with our world view.

Dividing measurement spaces into quadrants is, in part, a throwback to the old days of flow cytometry without computers, when gates were implemented using hardware, and it was much easier to make them rectangular than it was to make them any other shape. Quadrants work best when the data fall neatly into rectangular regions, and when cells either have a lot of a particular antigen or other marker, making them **positive**, or very little or none, making them **negative**. The CD4-CD8 distribution of peripheral blood lymphocytes is about as good an example of this situation as can be found, but, even here, we see that, while the events divide

clearly into positives and negatives on the CD4 axis, there are some events with intermediate levels of CD8.

If we were looking at cells from the thymus, quadrants wouldn't work well at all, because there are a lot of immature T cells in the thymus that have both CD4 and CD8, some of which are acquiring the antigens and some of which are losing them, and where one draws the quadrant boundaries is pretty much arbitrary. But problems with immunofluorescence data go beyond that, and beyond quadrants.

Deals with the Devil: Logarithmic Amplifiers and Fluorescence Compensation

The need and desire to measure immunofluorescence have motivated much of the development of modern flow cytometry. However, two problems associated with immunofluorescence measurement, and the less than satisfactory techniques applied to their solution, have been frustrating to beginners and experts alike.

The first problem is that of making and representing the results of measurements encompassing a large **dynamic range**. The first flow cytometers used to make immunofluorescence measurements weren't very sensitive. The green fluorescent dye **fluorescein** was used to label antibodies, and fluorescence was measured through color glass long pass filters, which, in addition to fluorescein fluorescence, let through **cellular autofluorescence**, probably due primarily to intracellular **flavins**. The filters themselves also emitted some fluorescence when struck by stray laser light. This made it impossible to detect fewer than several thousand antibody-bound fluorescein molecules on an unstained cell. However, since the maximum number of antibody-bound fluorescein molecules present on a cell might be a million or more, it was desirable, even before sensitivity was increased, to have some useful way of expressing results that varied over the three decade range between 1,000 and 1,000,000.

One obvious technique was to report and display results on a **logarithmic scale**. You can see examples of this in Figure 1-17, if you look at the numeric values and the positions of the tick marks on the axes of panels B, C, D, E, and F. Although the linear scales shown for forward and side scatter measurements (as in panel A) are accurate, the logarithmic scales may only be approximate. When analog data are digitized to relatively high resolution (16 to 20 bits), it is possible to convert signals accurately from a linear to a four decade (range 1 to 10,000) logarithmic scale and back using a digital computer; some modern cytometers employ this technique. However, in the 1970's and 1980's, the high-resolution ADCs needed to implement this procedure simply weren't available. The stopgap solution, which is still in use by some manufacturers, was to employ **logarithmic amplifiers**, commonly if not affectionately known as **log amps**.

A log amp is an analog electronic circuit that, in principle, puts out a voltage or current proportional to the voltage or current at its input. So far, so good. The bad news is that the proportionality constant may vary with time, temperature, input voltage, and, I suspect, the experimenter's astro-

logical sign. A log amp isn't a log table, or even approximately like one. The worse news is that nobody much cared how bad log amps were until the late 1980's, when people got interested in trying to make quantitative measurements of immunofluorescence and got really screwed up trying to convert from logarithmic to linear scales and back. We can expect the trend toward digital processing will continue, allowing logarithmic amplifiers to be replaced or, alternatively, monitored and calibrated; either approach should result in increased accuracy of representation of measurements on logarithmic scales.

A different set of complications was introduced by the development of antibody labels that enabled flow cytometers with a single illuminating beam (488 nm) to be used to make simultaneous measurements of immunofluorescence from several cell-bound antibodies. The first of these labels was the yellow fluorescent **phycoerythrin (PE)**, a protein found in the photosynthetic apparatus of algae. By attaching dyes to this molecule, making what are called **tandem conjugates**, it is possible to obtain fluorescence emission at longer wavelengths. When the rhodamine dye **Texas red** is attached to phycoerythrin, the resulting conjugate emits in the orange spectral region (620 nm); phycoerythrin with the **indodicarbocyanine** dye **Cy5** attached emits in the red (670 nm). Tandem conjugates of phycoerythrin with the cyanine dyes **Cy5.5** and **Cy7** emit even farther in the red or near infrared, at 700 and 770 nm. Some flow cytometers now in commercial production can be used to make simultaneous measurements of cells labeled with fluorescein, phycoerythrin, and all of the tandem conjugates just mentioned; most allow fluorescence in at least three spectral regions to be measured. The raw measurements, however, will not leave us in a state of conjugate bliss; we still have to contend with the problem of **compensation for fluorescence emission spectral overlap** between the labels, which only gets worse as the number of labels excited at a single wavelength increases. Figures 1-18 and 1-19 (pages 37 and 38) should provide some understanding of the problem and its solution.

Most fluorescent materials emit over a fairly broad range of wavelengths. When we describe fluorescein as green fluorescent, what we really mean is that if you look at it under a fluorescence microscope, the fluorescence looks green, and that if you measure the spectrum in a spectrofluorometer, the emission maximum is in the green spectral region. When we try to measure fluorescein fluorescence in a flow cytometer, we typically use a detector fitted with a green filter that passes wavelengths between 515 and 545 nm. However, as can be seen in Figure 1-18, the emission spectrum of fluorescein doesn't start abruptly at 515 nm and stop abruptly at 545 nm; it extends out well beyond 600 nm, although the fluorescence at the longer wavelengths is considerably less intense. There's quite a bit of emission from fluorescein in the 560-590 nm spectral region that we call yellow, and in which the emission maximum of phycoerythrin lies. That means that if we were to stain cells or other particles with fluorescein and nothing else, and measure them in a flow

cytometer with both green and yellow detectors, we'd pick up a strong signal in the green detector, and also detect some signal in the yellow detector.

The phycoerythrin emission spectrum also extends well beyond the 560-590 nm yellow wavelength range we use for measurements of phycoerythrin fluorescence. There is some emission below 560 nm, in the 515-545 nm green region, and considerably more above 580 nm. If we put cells stained with phycoerythrin and nothing else into the cytometer, we'd get the strongest signals from the yellow detector, and some signals from the green, orange, and possibly the red detectors as well. The same argument holds for the orange and red fluorescent tandem conjugates; each of these will definitely be detectable in the detector intended to measure the other, and signals from the orange conjugate will show up at the yellow detector as well, and possibly also in the green one.

If we put a cell sample stained with antibodies labeled with fluorescein, phycoerythrin, and the orange and red PE-Texas red and PE-Cy5 conjugates into the machine, the signal we get from the green detector is going to be comprised mostly of fluorescein fluorescence, with a smaller contribution from phycoerythrin fluorescence, possibly a wee bit from the orange conjugate, and some from cellular autofluorescence. The signal from the yellow detector will represent mostly phycoerythrin fluorescence, with substantial contributions from fluorescein and the orange conjugate, possibly some from the red conjugate, and some from autofluorescence. And so on for the signals from the orange (600-620 nm) and red (660-680 nm) detectors. Now, how much is "some," "a substantial contribution," or "a wee bit"?

That will depend on the gain settings used for the various detectors. Once these are set, it is fairly simple to quantify the degree of spectral overlap. For example, suppose we measure cells or beads stained only with fluorescein, and they produce signals with a mean intensity (peak height or area) of 5 V from the green detector and signals with a mean intensity of 1 V from the orange detector. If we were then to measure cells stained with fluorescein and phycoerythrin, and we wanted to remove the fluorescein contribution from the orange detector signal, we could subtract 1/5 of the green signal intensity. If a doubly stained cell yielded a signal of 1 V from the green detector, we'd subtract 0.2 V from the orange signal, no matter what the value of the orange signal was; if the cell yielded a 4 V signal from the green detector, we'd subtract 0.8 V from the orange signal, and so on. Well, actually, we'd also have to do the reciprocal calculations to figure out how much of the orange signal to subtract from the green signal to remove the contribution due to phycoerythrin. In principle, though, we could figure out the whole business using high school algebra, by solving **simultaneous linear equations**. Linear equations... aye, there's the rub.

All of the operations involved in fluorescence compensation must be performed on **linear** signals. You have to make the measurements on a linear scale to determine the fractions of fluorescence signal at each detector due to each fluores-

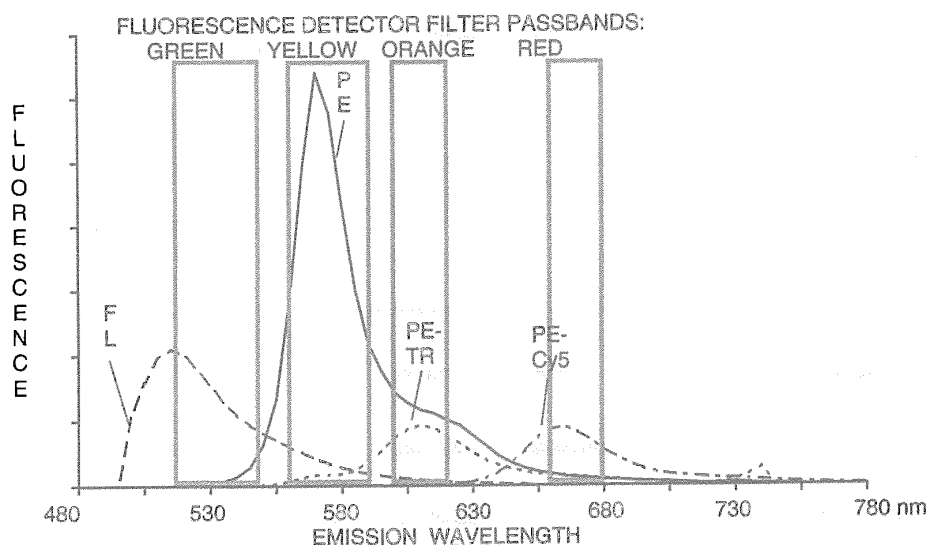


Figure 1-18. Why fluorescence compensation is necessary. The emission spectra shown are for equal concentrations (mg antibody/mL) of mouse anti-human IgG directly conjugated with fluorescein (FL), phycoerythrin (PE), and tandem conjugates of phycoerythrin with Texas red (PE-TR) and Cy5 (PE-Cy5), with excitation at 490 nm. Boxes demarcate the passbands of the green (530 nm), yellow (575 nm), orange (610 nm), and red (670 nm) filters used on the fluorescence detectors. Spectra are corrected for PMT responsivity differences at different wavelengths.

cent label, and the subtractions needed to make the necessary corrections also have to be done in the linear domain. But, as you remember, we usually tend to feed signals from immunofluorescence measurements through logarithmic amplifiers. How, then, do we introduce the fluorescence overlap compensation?

What happens in most older flow cytometers is that yet another analog circuit is built in between the preamplifier outputs and the log amp inputs. The circuit is something like an audio mixer, except that it subtracts signals instead of adding them; the operator adjusts one knob to determine the amount of green signal to subtract from yellow, another to determine the amount of yellow signal to subtract from green, etc. For two colors, this isn't all that hard to do. For three colors, you need six knobs, although you can get away with four if you ignore the green-orange and orange-green interactions. For four colors, you should have twelve knobs, though you might get away with eight. Each knob, of course, is attached to a **potentiometer**, or **variable resistor**, which, as was noted in the discussion of window comparators on p. 28, is basically a volume control. That starts to add up to a lot of electronic circuitry. Things may look neater if you let a computer control the compensation using **digital-to-analog (D-A) converters**, but you still end up with a lot of electronics at the input of your log amps.

Now, the whole reason we bother using log amps is to get a large dynamic range for our measurements. If we want a four decade dynamic range, with the top of the highest decade at 10 V, we end up with the top of the next highest

at 1 V, the top of the next highest at 100 mV, and the lowest decade encompassing signals between 1 and 10 mV. If you want to process signals between 1 and 10 mV, you have to keep the noise level below 1 mV. I've measured noise in a number of older flow cytometers from a number of manufacturers, and I haven't run across one with noise below 1 mV at the preamplifier outputs. The more electronic components you stick in the circuit, the more opportunities there are to increase the noise level, and my considered opinion is that it is unlikely that a system that implements four-color compensation in electronics will be able to maintain the low noise level needed to insure a true four decade dynamic range.

Quite aside from all that, though, most people can't solve simultaneous linear equations in their heads, and those few who can probably can't manage to solve equations and twiddle knobs on compensation circuitry at the same time. You have a reasonable chance of getting two-color compensation close to right by eye; three-color compensation gets a little tougher, and you're kidding yourself if you think you can do four-color compensation correctly without solving equations. As far as I know, the manufacturers have capitulated completely on the subject of compensation for more than four colors; the knobs are gone.

There was really no choice. If you keep all the electronic measurements linear, using an A-D converter with 16 or more bits' precision, you can dispense with 1) all of the knobs and their associated electronics, 2) all of the log amps, and 3) the semiempirical process of knob twiddling for fluo-

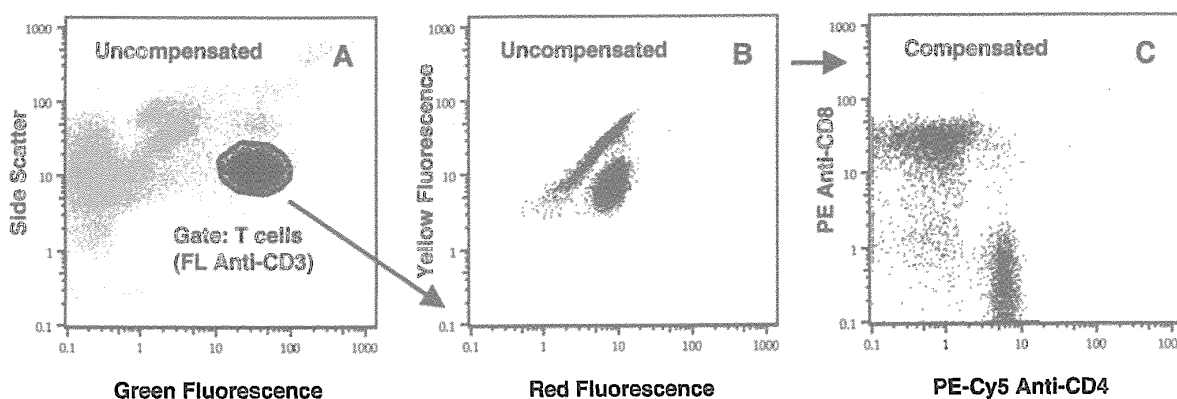


Figure 1-19. How compensation gets data to fit into quadrants.

rescence compensation. The simultaneous linear equations can be solved using digital computation, which can also do highly accurate conversions to a logarithmic scale. Once you have access to high-resolution digitized data, the logarithmic scale is only really needed for display, anyway; any statistical calculations that need to be done can be done on the linear data.

A significant advantage of high-resolution digitization of data is that you can go back to data that were not properly compensated when they were collected, transform them from a log scale to a linear one, if necessary, and redo the compensation. There are flow cytometry software packages that will let you play this game with log scale data that were digitized using 8- or 10-bit ADCs, but you end up with plots that have “holes” in the clusters due to the substantially larger, unavoidable digitization errors associated with lower resolution converters. The plots can be, and usually are, made more lovely to look at by **dithering**, adding random numbers to the data values. This technically degrades the quality of the data, but not by that much. I used to disapprove of it; I am now willing to accept it as yet another of the many deals with the devil that have to be made at the current state of the art. Within a few more years, almost all of the instruments in use will have higher resolution data analysis, and the plots of flow cytometric data will look pretty without benefit of dithering and diddling.

Evils of Axes: Truth in Labeling Cells and Plots

Mislabeling of axes, usually unintentional (I hope), is seen all too often in plots of flow cytometric data. Beginners and old-timers do it, and the mislabeling gets by journal reviewers, editors, and proofreaders. With the aid of Figure 1-19, which illustrates the effects of compensation, we can consider why mislabeling may occur and how to avoid it.

The data in Figure 1-19 were taken from a sample of whole blood stained with fluorescein anti-CD3 antibody, phycoerythrin anti-CD8 antibody, and phycoerythrin-Cy5 anti-CD4 antibody. Erythrocytes in the sample were lysed, and the sample was fixed with a low concentration of for-

maldehyde, before analysis. Panel A shows a dot plot of green fluorescence vs. side scatter, with both parameters displayed on a 4-decade logarithmic scale. A polygonal gate is drawn around a cluster I claim are T cells; the cells (events, if we want to be more precise) in this gate are plotted in black, while the remainder of the population is plotted in light gray.

If you look at Figure 1-14 (p. 31), you will notice that it is also a plot, in this case, a two-dimensional histogram, with anti-CD3 fluorescence on the x-axis and side scatter on the y-axis. In Figure 1-14, the y-axis is explicitly labeled as linear, and the x-axis as log, since one cannot tell whether the scale is log or linear simply by looking at the superimposed grid. The logarithmic scales on the axes of the panels in Figure 1-19 provide us with tick marks that would tell us that the scale was logarithmic even without the associated numbers, which are simply arbitrary indicators of intensity.

However, the x-axis of Figure 1-14 is labeled as “Log CyC-CD3 Flu,” which means that this axis represents the intensity of fluorescence, on a logarithmic scale, of an anti-CD3 antibody, labeled in this case with PE-Cy5, with CyC being an abbreviation for one of the trademarked versions of this tandem conjugate label. The x-axis in panel A of Figure 1-19 is labeled “Green Fluorescence.” What’s the difference?

The difference is that fluorescence compensation has been applied to the data in Figure 1-14, but not to the data in panel A (or panel B) of Figure 1-19. So what is displayed on the x-axis in panel A is really green fluorescence, most of which is from the fluorescein label on the anti-CD3 antibody, but some of which is from the PE anti-CD8 and PE-Cy5 anti-CD4 antibodies. And some is probably from cellular autofluorescence, but we’ll neglect that for the time being. We can get away with drawing a T cell gate using the uncompensated data because the fluorescein fluorescence pretty much dominates the uncompensated signal.

The situation is quite different when we look at panel B of Figure 1-19. The cells in this dot plot are only those with side scatter and fluorescence values falling within the T-cell gate shown in panel A. The axes of panel B are labeled as

showing red and yellow fluorescence, both on logarithmic scales, and I labeled them that way because the data are not compensated. There are two major clusters of cells/events visible in panel B, but points in each display significant intensities of both red and yellow fluorescence.

Panel C of Figure 1-19 shows the same cells, i.e., those in the original T cell gate, after compensation has been applied. What compensation has done is solve three linear equations in three unknowns; this gives us the fluorescence intensities of the fluorescein anti-CD3, PE anti-CD8, and PE-Cy5 anti-CD4 antibodies, which can now be plotted as such, allowing the x- and y-axes of panel C to be labeled "PE-Cy5 anti-CD4" and "PE anti-CD8." The major clusters of cells, representing CD-4 bearing T lymphocytes (often described as CD3⁺CD4⁺ cells, where the superscript "+" denotes positive) and CD-8 bearing T lymphocytes (CD3⁺CD8⁺ cells), are clearly visible, and could be fit nicely into quadrants.

Now, it would probably be perfectly legitimate to label the x-axis of panel C as "PE-Cy5 CD4," or even just "CD4," and the y-axis as "PE CD8," or just "CD8." However, if you want to be picky, what you are looking at is antibody bound to the cells. There's little doubt that almost all of the anti-CD4 antibody bound to T cells is bound to CD4 antigen on the cell surfaces, or that almost all of the anti-CD8 antibody bound to T cells is bound to cell surface CD8 antigen. On the other hand, both Figure 1-14 and panel A of Figure 1-19 show apparent binding of anti-CD3 antibody to monocytes and granulocytes; this is almost certainly **nonspecific binding**, which can occur via a variety of different mechanisms, and if we haven't got "truth in labeling" for the cells, we won't have it for the axes.

Some labels for axes should get the axe right away. The first candidates on my hit list are "FL1," "FL2," "FL3," etc., which usually mean green (515-545 nm), yellow (564-606 nm), and red (635 to about 720 nm, by my guess, limited by the characteristics of the 650 nm long pass filter at the short end and by the fading response of the detector at the long end) fluorescence. These were the fluorescence measurement ranges in the Becton-Dickinson FACScan, the first really popular benchtop 3-color fluorescence flow cytometer. The fluorescence filters in this instrument could not be changed, so at least FL1, FL2, and FL3 always meant the same thing – to FACScan users. However, in the B-D FACSCalibur, which has replaced the FACScan, while FL1 and FL2 still represent the same wavelength ranges, FL3 is different for 3- and 4-color instrument setups (650 long pass for 3-color; 670 long pass for 4-color). I think it's perfectly appropriate to use, for example, "green fluorescence," "515-545 nm fluorescence" (indicating the approximate range), or "530 nm fluorescence" (indicating the center wavelength), or even "Green (530-545 nm) fluorescence," but let's lose FL1, FL2, FL3, etc. If you're using a long pass filter, then say, for example, ">650 nm fluorescence." If the data come from a flow cytometer with multiple excitation beams, then you might want a label like "UV-excited blue fluorescence,"

or "355→450 nm fluorescence." And also remember that the fluorescence color designation or bandwidth range is only really appropriate if you're displaying or talking about uncompensated data; the whole point of compensation is to get you a new set of variables that represent the amounts of probes or labels in or on the cells, rather than the measurement ranges in the cytometer.

I've already been through the labels once, in the discussion in the previous column about whether to use the antigen name or the antibody name as an axis label. However, I will return to this area to skewer the next victim on my hit list, which is "FITC." Almost everybody uses this; I have done so myself, but I have seen the error of my ways. "FITC" is a perfectly valid abbreviation for fluorescein isothiocyanate, which is the most popular reactive fluorescein derivative used to attach a fluorescein label to antibodies and other probe molecules. Once the FITC reacts with the antibody, it isn't FITC anymore, and one typically dialyzes the fluorescent antibody conjugate, or runs it over a column, in order to remove free fluorescein (the FITC is pretty much all hydrolyzed by the time you finish, anyway). Oh, yes, FITC can also be applied directly to cells, to stain proteins; once again, what you end up with bound to the proteins is fluorescein, not FITC. It would seem simple enough to use "FL" as an abbreviation for fluorescein, the way we use "PE" for phycoerythrin. I guess the problem here is that nobody wants to describe a fluorescent antibody as, say, "FL anti-CD3," rather than "FITC anti-CD3," because that might get it confused with "FL1," "FL2," "FL3," etc. Well, after I take over the world, we won't have that problem.

Then there are the scatter signals. "Forward Scatter," "Small Angle Scatter," "FALS," and "FSC" are all acceptable as axis labels; however, unless you have calibrated your measurement channel and have derived a cell size measurement from forward scatter, "Cell Size" is really inappropriate. In the same vein, I'd use "Side Scatter," "Large Angle Scatter," "90° scatter," "RALS," or "SSC" without much hesitation, but avoid "Granularity." People knowledgeable about flow will know what you are measuring; if your audience is uninitiated, you should provide a brief explanation.

It's also about time that people stopped referring to data collected with flow cytometers as "FACS data" instead of "flow cytometry data." "FACS" is the abbreviation for "Fluorescence-Activated Cell Sorter (or Fluorescence-Activated Cell sorting)," originally used by Herzenberg et al, and has been a Becton-Dickinson trade name since B-D commercialized their instrument in the 1970's. All FACSees are flow cytometers, but not all flow cytometers are FACSees, and some FACSees, such as FACScans and FACSCounts, aren't even Fluorescence-Activated Cell Sorters.

And, finally, as long as I'm ticked off, I should remind you that the tick marks on the log scale will almost certainly not represent the real scale if the instrument uses log amps without compensating for their deviations from ideal response.

When Bad Flow Happens to Good Journals

Well, you might ask, does it really matter that much whether the axis labels are absolutely correct? Won't the more egregious mistakes be picked up before manuscripts get accepted and published? Unfortunately not; there has been a great deal of weeping and wailing in the cytometry community of late about this issue, because we see a lot of bad cytometry data presentation in a lot of the more prestigious general interest and cell biology journals, and even in some of the tonier titles in hematology and immunology.

To be sure, flow cytometry may not be the only technical area in which there are such problems. A typical paper with ten or more authors might include data from gel electrophoresis, gene array scanning, confocal microscopy, etc., as well as flow data. It will probably have been reviewed by no more than three people, and they can't know all of the methodology in detail. There may be gel curmudgeons and array curmudgeons out there grumbling at least as loudly as the flow curmudgeons and the confocal curmudgeons.

In preparing this edition of *Practical Flow Cytometry*, I asked several people to send me corrected versions of data displays that appeared in papers dealing with significant refinements in technology that were critical to the biological or medical applications discussed. The referees didn't pick up the original mistakes; neither did the authors, who were good sports about responding to my requests.

Most of the time, bad flow data presentations, or even minor errors in interpretation, don't invalidate the principal conclusion(s) of a paper. When they do, the obvious remedy is for the original authors to correct their errors, or for some other people to produce another paper using better technique to reach the right conclusion. But it's much better all around if the mistakes are corrected before the manuscripts get sent in.

Meanwhile, it is incumbent upon us all to maintain a certain level of vigilance, not only when preparing cytometric data for presentation and publication, but when looking at data that others have presented or published. If it's important to you to know the details of an experiment, either because you want to duplicate it and/or adopt the methodology or because its conclusions form part of the foundation for something you want to do, work through the details. These days, it's not that uncommon to find multiparameter flow data in a paper in which little details such as the source of the antibodies used, or even which antibodies had which labels, are omitted from the "Materials and Methods" section.

Now, in an ideal world, in which everything has been done correctly, it shouldn't matter that much; I've already come out in favor of simple axis labels such as "anti-CD4" or "CD4," and, assuming that the reagents and cell preparation, initial measurements, gating, and compensation were not flawed, it shouldn't matter which antibody or label was used in an experiment. But it does. If the details you need aren't in the published paper, contact the author. That's why

the e-mail address, and the snail mail address, are there. There is also an increasing likelihood that there will be another option; the journal and/or the authors may maintain a web site from which you can get technical details that were omitted from the published work.

Sorting Sorting Out

Flow sorting extends gated analysis to isolate pure populations of viable cells with more homogeneous characteristics than could be obtained by any other means. If you can get the cells that interest you into a gate in your multiparameter measurement space, you can get them into a test tube, or into the wells of a multiwell plate. Flow sorting is especially useful in circumstances in which further characterization of the selected cells requires short- or long-term maintenance in culture or analytical procedures that cannot be accomplished by flow cytometry.

A flow cytometer is equipped for sorting by the addition of a mechanism for diverting cells from the sample stream and of electronics and/or computer hardware and software that can determine, within a few microseconds after a cell passes by the cytometer's sensors, whether the values of one or more measurement parameters fall within a range or ranges (called a **sort region**, or **sort gate**) preset by the experimenter, and generate a signal that activates the sorting mechanism. The selected cells can then be subjected to further biochemical analysis, observed in short- or long-term culture, or reintroduced into another biological system (as was mentioned on p. 26, a substantial number of animals and more than a few babies have been conceived from sorted sperm).

The range of particles that can be sorted has been extended substantially in recent years; laboratory-built²³²⁵ and commercially available instruments are now in routine use for sorting *C. elegans* nematodes and *Drosophila* embryos, while laboratory-built microfluidic apparatus has been used to sort bacteria²³²⁶ and could, in principle, sort DNA fragments, other macromolecules, or viruses²³²⁷. Sorting of beads, rather than cells, has also come into use for various applications of combinatorial chemistry; the work of Brenner et al.²³²⁸⁻⁹ on gene expression analysis presents a good example.

The first generation of practical sorters accomplished cell separation by breaking the sample stream up into **droplets**, applying an electric charge to the droplets containing the selected cells, and passing the stream through an electric field, which would divert the charged droplets into an appropriate collecting vessel. A few older, and some newer, instruments use mechanical actuators to collect cells from a continuous fluid stream; while such mechanical sorters operate at lower rates (hundreds versus thousands of cells/s) than droplet sorters, their closed fluidic systems are better adapted for work with potentially infectious or otherwise hazardous materials that might be dispersed in the aerosols inevitably generated by droplet sorters. Large-particle sorters are typically mechanical, but not all of them have closed fluidic systems.

In general, sorting larger objects limits you to lower sorting speeds. If you're sorting lymphocytes, or something smaller, in a droplet sorter, you can use a 50 μm orifice, and generate droplets at rates of 100,000 droplets/s. If you're sorting pancreatic islets, which may be a few hundred μm in diameter, you'll need a 400 μm orifice, and you probably won't be able to go much above 1 kHz for a droplet generation frequency. If you're sorting *Drosophila* embryos, using a mechanical sorter (they're probably a little too big for a droplet sorter), you can measure your sort rate in dozens per second, rather than thousands.

Since cells arrive at the observation point at random times, at least approximately following Poisson statistics, there is always some probability of coincidences, which, as was noted on pp. 17 and 20-21, can pose some problem in flow cytometric analysis. Coincidences pose a fairly obvious problem in sorting, as well; they can result in your getting cells you don't want in the same droplet/well/tube as cells you do want. If the sorter is operated in the so-called **coincidence abort** mode, in which a wanted cell accompanied by an unwanted cell is not sorted, the **purity** of sorted cells is maintained, but the **yield** is decreased, while if wanted cells coincident with unwanted ones are sorted, yield is maintained at the expense of purity. All other things being equal, working at higher cell analysis rates ultimately ends up increasing the likelihood of coincidences, but there may be times when the best strategy is to sort twice, first for enrichment of a rare subpopulation, and then to increase purity of the cells recovered during the first sort.

In many cases in which flow sorting comes to mind as an obvious way of answering questions about a cell subpopulation, multiparameter analysis may allow the desired information to be obtained expeditiously without physically isolating the cells. Since the 1990's, most flow cytometry is multiparameter flow cytometry, as should be obvious from the content of the past dozen or so pages. Things were different in the bad old days.

In the 1970's, a method that was likely to come to mind for determining the distribution of DNA content in a lymphocyte subpopulation defined by the presence of a particular cell surface antigen involved staining cells with the appropriate fluorescent antibody, and then flow sorting to isolate those cells bearing the surface antigen. The sorted cells would subsequently be stained with a DNA fluorochrome such as propidium iodide; the restained sorted cells could then be run through the flow cytometer once more to determine the DNA content distribution.

This procedure was actually followed when Ellis Reinherz and Stuart Schlossman wanted to know whether there was any difference in DNA synthetic patterns between CD4- (then T4-) and CD8- (then T8-) bearing T cells; cells were stained with fluorescein-labeled monoclonal antibodies, sorted on a Becton-Dickinson FACS fluorescence-activated cell sorter, then sent to my lab, stained with propidium iodide, and analyzed on my recently built flow cytometer, which, at that time, wasn't sensitive enough to measure im-

munofluorescence. The chart recorder attached to my "Cytomutt" duly produced histograms of DNA content for the CD4-positive and CD8-positive cells and the antigen-negative cells, which had also been sorted.

The technically demanding and tedious exercise just described, which required at least an hour's combined use of the two instruments, did get the desired results. However, it would have been much easier to stain the entire cell population with both the fluorescent antibody and the DNA fluorochrome, making correlated multiparameter measurements of antibody fluorescence and DNA fluorescence in each cell, and using gated analysis to compile the DNA content distributions of antibody-positive and antibody-negative cells, eliminating the sorting. There was even an instrument available to us that could have done the job.

To be fair, most immunologists, faced with the same problem today, would instinctively look toward multiparameter measurement for the solution. When some colleagues and I recently had occasion to revisit the issue of DNA content of peripheral blood CD4-positive and CD8-positive T cells in the context of HIV infection and response to multidrug therapy, it was reasonably simple to deal with cells simultaneously stained for CD3, CD4 or CD8, and DNA (and RNA) content²³³⁰.

However, those of us who have been in the flow cytometry and sorting business for a long time are likely to experience a sense of *déjà vu* when the cell and molecular biologists and geneticists bring in samples to be sorted on the basis of expression of *Aequorea* green fluorescent protein (GFP) or, more likely, one of its variants. When I wrote the previous edition of this book, Martin Chalfie et al¹⁶⁴⁸ had just demonstrated the use of GFP as a reporter of gene expression; as far as he or I knew, nobody had yet done flow cytometry on cells transfected with GFP. Most cell sorting involved selection of cells bearing one or more surface antigens. Today, people who run sorting facilities tell me that a substantial amount of their time is now spent sorting samples for cells expressing GFP or its relatives. And they also mention that the people who bring in those samples often initially contemplate sorting the cells, staining them again to measure some other parameter, and reanalyzing them.

So, although multiparameter cytometry is now old hat for the immunologists, there are some other folks out there who haven't made it that far along the learning curve. I hope the above cautionary tale, and the lengthy discussion of multiparameter cytometry that has preceded it in this chapter, will help prevent unnecessary sorting. When in doubt, work with your sorter operator and facility manager.

The nuts and bolts details of sorting will be covered at length in Chapter 6; I'll devote the rest of this discussion to what is probably the most important step in designing a sorting experiment: doing the math. A lot of people think they know that state-of-the-art high-speed cell sorters can analyze at least 16 parameters and sort (into four streams) at rates of 100,000 cells/s. However, when I polled a select group of people who actually run state-of-the-art high-speed

sorters in various labs at universities, medical facilities, and biotech and pharmaceutical companies in the Boston area, I found that nobody had done more than 8-parameter analysis, and that, while a few people had run 40,000 cells/s on occasion, 20,000 cells/s was a more typical analysis rate. Ger van den Engh, who has played and continues to play an important role in high-speed sorter development, recommends that experimenters assume analysis rates no higher than 10,000 cells/s when assessing the feasibility of proposed experiments.

Now, a lot of people want to use sorting to isolate cells that make up a very small fraction of the population being analyzed. Gross et al.²³³¹ showed that it was possible to detect and sort cells from a human breast cancer line seeded into peripheral blood mononuclear cells at frequencies ranging from 1 cell in 10^5 to 1 cell in 10^7 ; they reported 40% yield and 22% purity for the sorts of cells at the lowest frequency. The raw numbers may be more impressive; a sample of 1.2×10^8 cells, which should have contained 12 cancer cells, was analyzed, giving rise to 23 sort decisions, of which 5 yielded cancer cells identifiable as such by microscopy. That sounds encouraging; even at 10,000 cells/s, it would only take about 3 hours to get 5 cells. Or about 6 hours to get 10 cells. And if you wanted to get 1,000 cells, you'd have to sort for about 25 days, 24/7.

You may have noticed that, when you're looking for cells present at low frequencies, while it is advantageous to be able to analyze at high speeds, there isn't much need for a high-speed sorting mechanism. In the above example, the sort frequency was 8/hr. There are a lot of people taking up time on very expensive, multiparameter high-speed sorters doing low frequency sorts based on one- or two-parameter measurements; sooner or later, somebody is going to make money selling simpler instruments for those jobs. Of course, if there is a method of enriching the population for the cells of interest before you start sorting – immunomagnetic separation, for example – you should take advantage of it.

A surprisingly large number of folks seem not to be doing the math before they write and submit grant applications involving sorting, which, for example, propose to isolate 10^6 cells initially present at a frequency of 1 cell/ 10^7 . Even if you had a 100% yield, that would require analysis of 10^{13} cells *in toto*, and, even if you ran the high-speed sorter at 10^5 cells/s, it would take 10^8 seconds, or a little over three years, to do the sort. And, amazing though it seems, some of these cockamamie proposals actually get funded. A grant application is typically reviewed by a few more people than review a manuscript, but, if there are enough other high-tech gimmicks in the application, there may not be a reviewer who knows enough about sorting to ask the right questions. So, do the math. Whether as an applicant or as a reviewer, you could save the taxpayers some money.

Parameters and Probes II: What is Measured and Why

Most flow cytometers used for research, and the majority of such instruments used in clinical immunology applica-

tions, measure only three physical parameters, namely, forward (or small angle) and side (or large angle) light scattering and fluorescence, even if they measure 16 colors of fluorescence using excitation from four separate light sources. A few instruments can also measure light loss (extinction), or sense electronic impedance to measure cell volume. The remainder of the discussion of parameters and probes in this chapter will deal only with scatter and fluorescence measurements; Chapter 7 is more ecumenical and more comprehensive.

In the course of introducing cytometry in general and flow cytometry in particular, I have already covered DNA content determination using various fluorescent dyes and the identification of cells in mixed populations using fluorescently labeled antibodies. If you will flip back to Table 1-1 (p. 3), you will see that there are a great many parameters and probes about which I have, thus far, said nothing at all. However, DNA stains, on the one hand, and labeled antibodies, on the other, do represent two fundamentally different types of probes.

Probes versus Labels

The chemical properties of the DNA dyes themselves determine the nature and specificity of their interactions with the target molecule. The nature and specificity of interactions of labeled antibodies with their targets is, ideally, determined solely by the structure of their combining sites; labels are added to facilitate detection and quantification of the amount of bound antibody based on the amount of fluorescence measured from the label. Under various circumstances, the labels themselves may decrease the specificity of antibody binding; this is always at least slightly disadvantageous and may be intolerable. DNA dyes can fairly be classified as probes; molecules such as fluorescein more often serve as labels. But, as usual, there are gray areas.

Fluorescein diacetate (FDA), actually diacetylfluorescein, was discussed on pp. 24-27; this is an example of a **fluorogenic enzyme substrate**. The nonfluorescent, uncharged FDA diester freely crosses intact cell membranes; once inside cells, it is hydrolyzed by nonspecific esterases to produce the fluorescein anion, which is highly fluorescent and which leaves intact cells slowly. Since most cells contain nonspecific esterases, FDA is not terribly useful as an indicator of enzyme activity; other nonfluorescent fluorescein derivatives can be used as probes for the activity of more interesting enzymes, such as beta-galactosidase. Different derivatives of fluorescein and other dyes can be introduced into cells and cleaved by esterases to produce indicators of pH, oxidation-reduction (redox) state, and the concentration of sulfhydryl groups or of ions such as calcium and potassium. So the best I can do to clarify the status of fluorescein is to say that it is a label when it is used covalently bound to a relatively large molecule such as an antibody, oligonucleotide, or protein ligand for a cellular receptor, and a probe when introduced into cells in a slightly chemically modified, low molecular weight form. The detailed discussion of probes in Chapter 7 provides examples of when this distinction breaks down.

We will now embark on a quick tour of selected parameters and probes for their measurement. Details and spectra appear in Chapter 7. It is appropriate to mention that the single most useful reference on fluorescent probes is the *Handbook of Fluorescent Probes and Research Products*²³³², edited by Richard P. Haugland; this is the catalog of Molecular Probes, Inc. (Eugene, OR). The latest printed version is the 9th Edition, which appeared in 2002. A CD-ROM version is available as well, and all the information in the handbook, and more, with updates, can also be found at Molecular Probes' Web site (www.probes.com).

Living and Dyeing: Stains, Vital and Otherwise

Before getting down to specific (and not-so-specific) stains, it's probably a good idea to define some terms relevant to staining cells and what does or does not have to be done to the cells in order to get them to stain. A dye or other chemical that can cross the intact cytoplasmic membranes of cells is said to be **membrane-permeant**, or, more simply, **permeant**; a chemical that is excluded by intact cytoplasmic membranes is described as **membrane-impermeant**, or just **impermeant**. Because permeant dyes stain living cells, they (the dyes) are also described as **vital dyes**, or **vital stains**. You will occasionally find an opposite, incorrect definition of a vital stain as a stain that does not stain living cells; don't believe it. This seems to be one of the few urban legends of cytometry.

There are numerous transport proteins that concentrate certain chemicals in, or extrude other chemicals from, cells. Many commonly used dyes, including Hoechst 33342, serve as substrates for the glycoprotein pump associated with multidrug resistance in tumor cells, and may not readily stain cells in which this pump is active; the general lesson is that the action of transporters may make it appear that a permeant compound that is efficiently extruded is impermeant. Microorganisms may have a broader range of transporters than do mammalian cells, making it risky to assume that they will handle dyes in the same way.

Staining cells with impermeant dyes requires that the membrane be **permeabilized**. This can be accomplished in the context of **fixation** of the cells. "Fixation" originally described a process that made tissue tough enough to section for microscopy and prevented it from being autolyzed by internal hydrolytic enzymes and/or chewed up by contaminating microorganisms. Most fixatives act either by denaturing proteins (e.g., ethanol and methanol) or by cross-linking them (e.g., formaldehyde and glutaraldehyde); since this is likely to change the structure of cell-associated antigens, it is common practice to stain with fluorescent antibodies before fixing cells. In general, the fixation procedures used for flow cytometry are relatively mild; one principal objective is to kill HIV and other viruses that may be present in specimens, and another is to allow samples to be kept for several days before being analyzed. In recent years, the real pathologists have been using microwave radiation as a fixative or adjunct; I have not run across reports of its use for flow cytometry.

Permeabilization without fixation can be accomplished using agents such as the nonionic detergents Triton X-100 and Nonidet P-40; permeabilizing agents may also be added to a mixture of one or more fixatives to make cytoplasmic membranes permeable to fluorescent antibodies while retaining cellular constituents, allowing staining of intracellular antigens. Several proprietary mixtures, some of which include red cell lysing agents, are available from manufacturers and distributors of antibodies.

Most sorting is done with the intention of retrieving living cells, so fixation is not an option. However, there are procedures, such as lysolecithin treatment and electroporation, which can transiently permeabilize living cells, allowing otherwise impermeant reagents to enter while preserving viability of at least some of the cells in a sample. In this context, it is important to remember that a permeant "vital" stain may eventually damage or kill cells. It is always advisable to establish that measurement conditions do not themselves perturb what one is attempting to measure.

Nucleic Acid (DNA and RNA) Stains

Although a large number of fluorescent dyes can be used to stain DNA and/or RNA, relatively few of them are specific for DNA, and most of these are sensitive to base composition (A-T/G-C ratio). DAPI (4', 6-diamidino-2-phenylindole), Hoechst 33258, and Hoechst 33342 increase fluorescence approximately 100 times when bound to A-T triplets in DNA. All these dyes are excited by UV light (325-395 nm), and emit in the blue spectral region with maxima between 450 and 500 nm.

Chromomycin A₃ and mithramycin exhibit increased fluorescence on binding to G-C pairs in DNA; they are excited by violet or blue-violet light (400-460 nm) and emit in the green between 525 and 550 nm. The combination of Hoechst 33258 and chromomycin A₃ has been used with dual excitation-beam flow cytometers to discriminate the majority of human chromosomes based on differences in DNA base composition, and to demonstrate differences in base composition among bacterial species. 7-amino-actinomycin D (7-AAD⁷³⁵) also enhances fluorescence (maximum around 670 nm) on binding to G-C pairs in DNA; although it is best excited by green light (500-580 nm), it can be excited at 488 nm.

Dyes such as ethidium bromide (EB) and propidium iodide (PI), both excitable over a range from 325 to 568 nm and emitting near 610 nm, increase fluorescence on binding to double-stranded nucleic acid, whether DNA or RNA, and the latter property is shared by a large number of asymmetric cyanine nucleic acid stains (e.g., the TO-PRO- and TOTO- series (impermeant), SYTO-series (permeant), Pico Green, etc.) introduced by Molecular Probes. These dyes can be used to stain total nucleic acid in cells; specific staining of DNA requires RNase treatment. Many of the cyanine nucleic acid dyes increase fluorescence several thousandfold; they have been used for detection of DNA fragments^{1144,2327,2333-4} and viruses²³³⁵⁻⁷.

Until recently, Hoechst 33342 was the only dye that could be used reliably to determine DNA content in living cells. However, in 1999 and 2000, Smith et al²³³⁸⁻⁹ reported that **DRAQ5**, an anthraquinone dye with an excitation maximum around 650 nm and an emission maximum near 700 nm when bound to DNA, could also provide a reasonably good DNA content histogram. DRAQ5 can also be excited at 488 nm, albeit somewhat inefficiently.

DRAQ5 does not increase fluorescence significantly on binding to DNA; it stains nuclei because it is present in higher concentrations in association with nuclear DNA than elsewhere in the cell, and the quality of staining is thus relatively more dependent on relative concentrations of dye and cells than is the case for most other DNA dyes. **Acridine orange (AO)**, like DRAQ5, does not increase fluorescence on binding to either DNA or RNA, but stains by virtue of its concentration on the macromolecules.

Darzynkiewicz et al showed, beginning in the mid-1970's, that, after cell membrane permeabilization and acid treatment, AO could be used for stoichiometric staining of DNA and RNA in cells^{262,3,525,1348-9}. On excitation with blue light (488 nm is eminently suitable), the DNA-bound monomer fluoresces green (about 520 nm); the RNA-bound dye forms red (>650 nm) fluorescent aggregates. The combination of DNA and RNA staining allows the cell cycle to be subdivided into stages that are not distinguishable on the basis of DNA content alone, permitting discrimination between G₀ and G₁ cells.

Relatively specific staining of double-stranded (predominantly ribosomal) RNA in cells can be achieved using a combination of **pyronin Y** (excitable at 488 nm with emission in the yellow around 575 nm), which stains RNA, with one of the Hoechst dyes, which binds to DNA and prevent DNA staining by pyronin Y. In a dual excitation-beam instrument (UV and 488 nm), DNA and RNA content in living cells can be estimated simultaneously from pyronin Y and Hoechst 33342 dye fluorescence, providing information that is substantially equivalent to what could be obtained using AO (Fig. 1-2, p. 27) without requiring that the cells be sacrificed¹¹³. Cells stained with this dye combination have been sorted with retention of viability²³⁴⁰⁻².

Toba et al²³⁴³⁻⁵ found that DNA and RNA could be measured in permeabilized cells using the combination of 7-AAD and pyronin Y in a system with a single 488 nm excitation beam; Schmid et al modified the staining conditions and reported improved precision and reproducibility²³⁴⁶.

Fluorescence and Fluorescent Labels

Because the fluorescent label on a probe is usually not intended to interact directly with the structure to which the probe binds, labels are developed and/or synthesized predominantly for their desirable spectral characteristics.

In order for an atom or molecule – or part of a molecule; the all-inclusive term would be **fluorophore** – to emit fluorescence, it must first absorb light at a wavelength shorter than or equal to the wavelength of the emitted light, raising

an electron to an excited state. Absorption requires only about a femtosecond. In order to have a high likelihood of fluorescing, a material must have a high likelihood of absorbing the excitation light; the likelihood that a molecule will absorb is quantified as the **absorption cross-section** or the **molar extinction coefficient**.

Fluorescence results from the loss of at least some of the absorbed energy by light emission. The period between absorption and emission is known as the **fluorescence lifetime**; for organic compounds, this is typically a few nanoseconds. Some of the absorbed energy is almost always lost nonradiatively, i.e., unaccompanied by emission, by transitions from higher to lower vibrational energy levels of the electronic excited state. The fluorescence emission will then be less than the energy absorbed; in other words, emission will occur at a wavelength longer than the excitation wavelength. The difference between the absorption and emission maxima is known as the **Stokes shift**, honoring George Stokes, who first described fluorescence in the mid-1800's. Stokes shifts are typically only a few tens of nanometers.

Fluorescence is an intrinsically quantum mechanical process; the absorbed and emitted energy are in the form of photons. The **quantum yield** and **quantum efficiency** of fluorescence are, respectively, the number and percentage of photons emitted per photon absorbed; they typically increase with the cross section and extinction coefficient, but are also dependent on the relative likelihoods of the excited molecule losing energy via fluorescence emission and nonradiative mechanisms. The quantum yields of some dyes used in cytometry are quite high, above 0.5, but it is important to note that quantum yield, particularly for organic fluorophores, is affected by the chemical environment (i.e., the pH, solvent polarity, etc.) in which the molecule finds itself. If an excited molecule that might otherwise fluoresce instead loses energy nonradiatively, for example, by collision with solvent molecules, it is said to be **quenched**; once returned to the electronic ground state, it can be reexcited. However, there is usually a finite probability that light absorption will be followed by a change in molecular structure, making further cycles of fluorescence excitation and emission impossible; this is called **(photo)bleaching**.

In principle, increasing the illumination intensity can increase the intensity of light scattering signals without limit. However, this is not even theoretically possible for fluorescence signals, because, at some level of illumination, all the available molecules will be in excited states, leaving no more to be excited if illumination intensity is further increased. This condition of **photon saturation** is often reached in cytometers which use laser powers of 100 mW or more; bleaching, which may also make the dependence of emission intensity on excitation intensity less than linear, is noticeable at power levels of tens of milliwatts. Saturation and bleaching are discussed at length by van den Engh and Farmer¹¹³⁰.

When an excited fluorophore is in close proximity (typically no more than a few nanometers) to another fluorophore, nonradiative energy transfer (**fluorescence resonance**

energy transfer, or **FRET**) from the excited (donor) molecule to the nearby acceptor molecule may occur, followed by fluorescence emission from the acceptor in its emission region. The probability of energy transfer increases with the degree of overlap between the absorption spectrum of the second fluorophore and the emission spectrum of the first. I have said "fluorophore" rather than "molecule" here because energy transfer can occur between different structures within the same molecule. An accessible review of FRET is provided by Szöllösi et al.²³⁴⁷

In the intact photosynthetic apparatus of algae and cyanobacteria, absorbed blue-green and green light is utilized for photosynthesis by a series of intra- and intermolecular energy transfers via **phycobiliproteins** to **chlorophyll**, without subsequent emission. In 1982, Oi, Glazer, and Stryer¹¹⁴ reported that extracted algal phycobiliproteins could be used as highly efficient fluorescent labels with large Stokes' shifts. As you might have noticed from the extensive previous discussion, it has become common practice to attempt to improve on nature by conjugating dyes to phycobiliproteins to add an additional phase of energy transfer and further shift the emission spectrum of the **tandem conjugates**. The first such tandem conjugate, described by Glazer and Stryer in 1983³⁶⁶, was made by linking **phycoerythrin** (PE) to **allophycocyanin** (APC), a phycobiliprotein which absorbs relatively efficiently, although not maximally, at phycoerythrin's yellow (575 nm) emission wavelength and which emits maximally in the red at 660 nm.

Until both flow cytometers and monoclonal antibodies became widely available in the early 1980's, the most widely used fluorescent label was **fluorescein**, usually conjugated to proteins as the **isothiocyanate** (FITC); second labels were only infrequently needed. Fluorescein is nearly optimally excited at 488 nm, and emits in the green near 525 nm. While rhodamine dyes had been used for two-color immunofluorescence analysis by microscopy, they were not suitable for 488 nm excitation. A small number of studies were done with yellow-excited dyes, which needed a second excitation beam, making flow cytometers substantially more expensive. Phycoerythrin (PE), which emits in the yellow near 575 nm, is maximally excited by green light but absorbs reasonably well at 488 nm. Its extinction coefficient is high enough to make the fluorescence signal from PE-labeled antibody substantially higher than that from an equivalent amount of fluorescein-labeled antibody (Fig. 1-18, p. 37).

We have already encountered tandem conjugates of PE suitable for 488 nm excitation (PE-Texas red, emitting near 610 nm; PE-Cy5, near 670 nm; PE-Cy5.5, near 700 nm; PE-Cy-7, near 770 nm). Allophycocyanin absorbs maximally in the red near 650 nm, and is well excited by red diode (635-640 nm) and He-Ne (633 nm) lasers. Tandem conjugates of APC with Cy5.5 and Cy7 emit in the far red and near infrared, as do the PE conjugates with the same dyes.

A principal disadvantage of phycobiliproteins as fluorescent labels is their large size; with a molecular weight near

240,000, PE binding increases the molecular weight of an immunoglobulin G antibody by about 150 percent. This may not be an issue when labeled antibodies or lectins are used to stain cell surface structures, but becomes one when it is necessary to use labeled reagents to demonstrate intracellular constituents. A number of lower molecular weight labels have been developed for this purpose. The symmetric cyanines¹³⁶¹⁻⁴ include Cy5, Cy5.5, and Cy7, and their shorter wavelength absorbing cousins, e.g., Cy3, which can be excited at 488 nm and emits in the same region as PE; we have already run across them as acceptors in tandem conjugates. Molecular Probes has recently developed the Alexa series of dyes²³⁴⁸ (also see the Molecular Probes handbook/Web site²³²²); different members of this series are excitable at wavelengths ranging from the UV to the near infrared. Alexa dyes, used alone or as acceptors in tandem conjugates, are reported to have better fluorescence yields and photostability (resistance to bleaching) than more commonly used labels with similar spectral characteristics, and seem to be coming into wider use. Low, rather than high, molecular weight labels are almost always used on oligonucleotide probes, which allow demonstration and quantification of specific nucleic acid sequences in cells or on beads or solid substrates (e.g., in gene arrays).

As was mentioned previously, it is the probe, not the label, that confers specificity; dyes must be derivatized into forms that contain a functional group, such as an isothiocyanate or sulfonyl chloride, that will allow the **reactive dye** to bind covalently to the probe. FITC, applied to cells, will stain accessible proteins. Staining of intact cells will be limited to the cell surface; in fixed or permeabilized cells, both surface and intracellular proteins will be stained.

Binary Fishin': Tracking Dyes Through Generations

Otherwise nonspecific, but persistent fluorescent staining of cellular proteins or lipids has recently been put to good use in studying cell proliferation. Since cellular proteins and lipids are apportioned more or less equally to each daughter cell during cell division, analysis of the fluorescence of cells after staining with a so-called **tracking dye** should allow determination of how many cycles of division have occurred since its ancestor was stained. The dye first widely used for such studies was **PKH26**¹⁵⁵¹⁻³, a yellow fluorescent cyanine dye with long alkyl side chains that incorporates itself tightly enough into lipid bilayers that it is not readily lost from cells. It was called a tracking dye because it could also be used to follow cells that had been removed from animals, labeled, and reinjected. Estimation of the numbers of cells in various daughter generations after PKH26 labeling requires application of a mathematical model¹⁵⁵⁵.

An alternative to PKH26, **carboxyfluorescein diacetate succinimidyl ester** (CFSE)²³⁴⁹, is a nonfluorescent fluorescein ester that enters cells and is hydrolyzed to a reactive dye by nonspecific esters; the end result is that fluorescein molecules are bound covalently to intracellular protein. Distributions of CFSE fluorescence in proliferating populations

usually show peaks indicating the positions of cells in different daughter generations; these can be analyzed with mathematical models, but it is also possible to combine sorting with CFSE labeling to isolate cells from different generations²³⁵⁰, which cannot be done reliably when PKH26 is used as a tracking dye.

Membrane Perturbation: A Matter of Life and Death?

The integrity of the cytoplasmic membrane is essential to cell function. Although at least some cells can survive transient small breaches of the membrane, longer-term and/or larger defects may deprive the cell of materials it would normally accumulate, and may also expose it to toxins it would normally exclude. Thus, we tend to think that cells with a demonstrable loss of membrane integrity are dead.

Trypan blue has been the preferred probe for a **dye exclusion test for "viability,"** i.e., retention of membrane integrity, performed by visual inspection of cells under the microscope; the Bio/Physics Systems Cytograf, made in the early 1970's, measured extinction and scattering using a red He-Ne laser source, and could detect trypan blue uptake by cells. These days, people who want to do dye exclusion testing by flow cytometry typically use impermeant nucleic acid dyes such as **propidium iodide** or **7-aminoactinomycin D**, both excitable at 488 nm, and, emitting, respectively, at about 620 and about 670 nm, or the red-excited dye **TO-PRO-3**, emitting at about 670 nm. Cells that take up the dye and become fluorescent are considered to be nonviable.

Fluorescein is anionic, and, therefore, relatively impermeant; when produced intracellularly by hydrolysis of fluorescein diacetate (FDA), it leaves cells slowly, giving us a **dye retention test for "viability."** Cells with intact membranes accumulate and retain fluorescein after exposure to FDA and become (green) fluorescent; cells with membrane damage do not retain fluorescein and do not fluoresce. The fluorescein derivative **calcein**, produced in cells by esterase action after exposure to the **acetoxymethyl ester, calcein-AM**, is retained much more effectively than fluorescein and is now preferred for dye retention tests.

The problem with dye exclusion and retention tests is that, while the methodology works well for cells that are killed by freezing or heat or by interaction with cytotoxic T or NK cells, all of which inflict early and usually lethal damage on the cytoplasmic membrane, cells that are killed by other means, e.g., those rendered reproductively nonviable by such agents as ionizing radiation, may retain membrane integrity for days after exposure. Uptake of impermeant dyes is therefore a better indicator of nonviability than retention is of viability, but there are situations in which impermeant dyes can end up in viable cells²³⁵¹.

One can, of course, combine dyes, for example, propidium iodide and calcein-AM, which will result in cells with intact membranes exhibiting green cytoplasmic fluorescence while cells with damaged membranes show red nuclear fluorescence, but this does not solve the basic problem. And,

in part thanks to cytometry, we can now distinguish one kind of death (**necrosis**) from another (**apoptosis**), making the issue of viability assays even more contentious. Darzynkiewicz et al have discussed the cytometry of cell necrobiology in detail²³⁵². Disturbances in membrane organization in apoptosis, resulting in the exposure of phosphatidylserine, are usually looked for using fluorescently labeled **annexin V**²³⁵³⁻⁴.

When viability is not an issue, measurements of fluorescence of cells over time after exposure to fluorescent dyes, drugs, or labeled drug analogs can be useful in detecting the presence of various transport proteins. Uptake or efflux kinetics in themselves can only suggest a mechanism; when the transporter or pump being investigated has been well characterized, establishing that known substrates and inhibitors affect fluorescence kinetics as predicted is critical for confirmation of the initial hypothesis.

Cytoplasmic/Mitochondrial Membrane Potential

Electrical potential differences are present across the cytoplasmic membranes of most living prokaryotic and eukaryotic cells, and also between the cytosol and the interior of organelles such as chloroplasts and mitochondria. Membrane potential ($\Delta\Psi$) is generated and maintained by transmembrane concentration gradients of ions such as sodium, potassium, chloride, and hydrogen.

Changes in cytoplasmic $\Delta\Psi$ play a role in transmembrane signaling in the course of surface receptor-mediated processes related to the development, function, and pathology of many cell types. Cytoplasmic $\Delta\Psi$ is reduced to zero when the membrane is ruptured by chemical or physical agents; mitochondrial $\Delta\Psi$ is reduced when energy metabolism is disrupted, notably in apoptosis. In bacteria, $\Delta\Psi$ reflects both the state of energy metabolism and the physical integrity of the cytoplasmic membrane.

Flow cytometry can be used to estimate membrane potential in eukaryotic cells, mitochondria *in situ*, isolated mitochondria, and bacteria^{424,2355}. Older methods, using lipophilic cationic dyes such as the symmetric cyanines **dihexyloxacarboxyanine** [**DiOC₆(3)**] and **hexamethylindodicarbocyanine** [**DiIC₁(5)**] or **rhodamine 123**, or lipophilic anionic dyes such as **bis (1,3-dibutyl-barbituric acid) trimethine oxonol** [**DiBAC₄(3)**] (which is often, incorrectly, referred to as **bis-oxonol**), can detect relatively large changes in $\Delta\Psi$, and identify heterogeneity of response in subpopulations comprising substantial fractions of a cell population. All of the dyes just mentioned can be excited at 488 nm and emit green fluorescence, with the exception of **DiIC₁(5)**, which is red-excited and emits near 670 nm. Newer techniques that use energy transfer and/or ratios of fluorescence emission at different wavelengths allow precise measurement of $\Delta\Psi$ to within 10 mV or less²³⁵⁶⁻⁷.

Since, in most eukaryotic cells, $\Delta\Psi$ across mitochondrial membranes is larger than $\Delta\Psi$ across cytoplasmic membranes, exposure of cells to lipophilic cationic dyes results in higher concentrations of dye in the cells than in the suspending

medium, and higher concentrations in mitochondria than in the cytosol. If cells are washed after being loaded with dye, staining of the cytosol may be minimized while mitochondrial staining persists. This is the basis for the use of DiOC₆(3), DiIC₆(5), rhodamine 123, and other cationic dyes to estimate mitochondrial $\Delta\Psi$; the procedure has become commonplace for studies of apoptosis, in which early increases in mitochondrial membrane permeability result in loss of $\Delta\Psi$. JC-1, a cyanine, exhibits green fluorescence in monomeric form and red fluorescence when aggregated at higher concentrations¹⁶⁸¹⁻², and has become popular for work on mitochondria in apoptosis.

Among other factors, action of efflux pumps, changes in membrane structure, and changes in protein or lipid concentration in the medium in which cells are suspended can produce changes in cellular fluorescence which may be interpreted erroneously as changes in $\Delta\Psi$. For example, it was observed in the 1980's that hematopoietic stem cells were not stained by rhodamine 123, and some people concluded that this reflected low mitochondrial $\Delta\Psi$; it was later found that the dye was being actively extruded by a glycoprotein pump. Getting good results from cytometric techniques for estimation and measurement of $\Delta\Psi$ demands careful control of cell and reagent concentrations and incubation times and selection of appropriate controls.

Indicators of Cytoplasmic [Ca²⁺]: Advantages of Ratiometric Measurements

The importance of calcium fluxes in cell signaling was appreciated when flow cytometry was in a relatively early stage of development, but it was not until some years later that suitable probes became available²³⁵⁸. The first probes exhibited differences in the intensity of fluorescence in the presence of low and high intracellular [Ca²⁺], but did not change either their fluorescence excitation or emission spectral characteristics to a significant degree. Since the distribution of fluorescence intensity from cells loaded with the probes was typically quite broad (a problem also associated with membrane potential probes), it was possible to appreciate large changes in cytoplasmic [Ca²⁺] affecting all or most of the cells in a population, which would shift the entire distribution substantially, but not to detect even a large change in cytoplasmic [Ca²⁺] involving only a small subpopulation of cells. This came as a disappointment to immunobiologists who hoped to use flow cytometry to detect calcium responses associated with activation of lymphocytes by specific antigens.

Roger Tsien and his colleagues, who had developed some of the earlier calcium probes, came to the rescue in 1985 with Indo-1⁸⁵⁸. This, like other probes, is a selective calcium chelator, but does not significantly perturb cellular calcium metabolism. Its fluorescence is excited by UV light; wavelengths between 325 and about 365 nm, which pretty well covers the range of UV sources available for flow cytometry, are suitable. Indo-1's attraction, however, is due primarily to the fact that there are substantial differences in emission

spectra between the free dye, which shows maximum emission at about 480 nm, and the calcium chelate, which emits maximally at about 405 nm. The ratio of emission intensities at 405 and 480 nm in cells loaded with Indo-1 [it is introduced as an acetoxymethyl (AM) ester] can, therefore, provide an indication of cytoplasmic [Ca²⁺]. The ratiometric measurement cancels out many extraneous factors, most notably including the effect of cell-to-cell variations in dye content, which plague older techniques for calcium measurement and for measurement of $\Delta\Psi$. Effects of uneven illumination and of light source noise also are eliminated by virtue of their equal influences on the numerator and denominator of the ratio. This advantage, it should be noted, is common to other ratiometric measurements (e.g., of $\Delta\Psi$ and of intracellular pH) in which both parameters used in the ratio are measured at the same time in the same beam.

If aliquots of loaded cells are placed in solutions with various known Ca²⁺ concentrations and treated with a calcium ionophore such as A23187 or ionomycin, it is possible to calibrate the fluorescence ratio measurement to yield accurate molar values of cytoplasmic [Ca²⁺]. Indo-1 is widely used, at least by people with UV excitation sources in their flow cytometers^{862, 1714-8}.

Since there are probably more than 10,000 fluorescence flow cytometers out there that don't have UV sources, that's small comfort. Luckily, there are alternatives. In 1989, Tsien and his collaborators described a series of fluorescein- and rhodamine-based calcium indicators suitable for use with 488 nm excitation¹⁷¹⁹. The most widely used of these is Fluo-3, which has the spectral characteristics of fluorescein, but which is almost nonfluorescent unless bound to calcium. Unlike Indo-1, Fluo-3 does not exhibit a spectral shift with changes in calcium concentration. A Fluo-3 fluorescence distribution is a haystack; if you're stimulating a cell population, the haystack moves to the right when the cytoplasmic [Ca²⁺] goes up and back to the left when it goes back down. However, there is another dye, Fura red, also suitable for 488 nm excitation, which exhibits high fluorescence when free in solution (or cytosol) and low fluorescence when bound to calcium; a Fura red haystack moves in the opposite direction from a Fluo-3 haystack with changes in cytoplasmic [Ca²⁺]. More to the point, the ratio of fluo-3 to Fura red fluorescence provides a precise, calibratable indicator of cytoplasmic [Ca²⁺] that can be used in the majority of fluorescence flow cytometers²³⁵⁸. Both Fluo-3 and Fura red, like Indo-1, are loaded into cells as AM esters.

Finding Antigen-Specific Cells Using Tetramers

While ratiometric probes did improve the precision of intracellular calcium measurements, they did not get them quite to the point of being able to detect specific responses of very small numbers of lymphocytes to antigens. As it turned out, a more direct approach was to succeed. In 1996, Altman et al²³⁵⁹ described identification of antigen-specific cytotoxic (CD3⁺CD8⁺) T cells using a fluorescently labeled complex containing four each of 1) a class I major histo-

compatibility complex (MHC) α chain, 2) β_2 -microglobulin, and 3) an antigenic peptide. Since that gets to be a lot to write or say, the probes are now universally described as **tetramers**.

Antigen presentation to T cells requires binding of antigenic peptides associated with HLA proteins (class I proteins for cytotoxic [CD3⁺CD8⁺] T cells, class II proteins for helper [CD3⁺CD4⁺] T cells) on the antigen presenting cell to the T cell receptor; attempts to bind a labeled monomeric complex (1 each) of α chain, β_2 -microglobulin, and peptide to cytotoxic T cells were unsuccessful because the binding affinity of the monomers was too low. Tetramers did the trick, and have come into wide use since they were originally described²³⁶⁰⁻². We now have not only **class I tetramers**, reactive with cytotoxic T cells, but also **class II tetramers**, which contain MHC class II proteins, and react in an antigen-specific fashion with helper T cells²³⁶³⁻⁴. They're not available at the corner store quite yet, but rumor has it that it was tetramers and their possibilities that made one of the major instrument companies decide to stay in the fluorescence flow cytometry business.

Hip, Hip Arrays: Multiplexing on Slides and in Bead Suspensions

If you have been keeping up with biology at all over the past few years, it's unlikely that you have not run across gene array technology²³⁶⁵⁻⁷, which allows the expression of hundreds or thousands of genes to be studied by, for example, hybridizing different colors of labeled cDNA derived from the same cells grown under different circumstances to a slide on which the requisite genetic sequences have been synthesized or deposited in small spots. The slides are then scanned, allowing differences in expression to be detected by color differences resulting from the presence of different amounts of the cDNAs on each spot. The array concept has taken off; we have gene arrays, protein arrays, cell arrays, and even tissue microarrays, which allow high-throughput molecular profiling of tumors²³⁶⁸.

Multiplex analysis allows flow cytometry to accomplish some of the same tasks for which gene arrays are now used. It occurred to various people in the mid-1980's¹⁸²⁰⁻³⁴ that various types of ligand binding assays could be done in a flow cytometer by using fluorescence measurements to quantify binding to appropriately coated beads. By using a different size and/or color bead for each of a number of assays, it would be possible to perform all of them at once on a single sample in a single tube²³⁶⁹.

The latest incarnation of multiplex analysis uses a small, dedicated flow cytometer capable of identifying as many as 100 different colors of beads, and has been applied successfully to both protein²³⁷⁰⁻¹ and nucleic acid²³⁷²⁻⁴ analysis. In a study monitoring multiple pathogenesis-related genes simultaneously in chemical-treated and control *Arabidopsis* samples, Yang et al²³⁷⁵ reported that a multiplexed flow cytometric assay they developed yielded results comparable to those obtained from a slide-based gene array.

GFP and Its Relatives: Mild-Mannered Reporters

The 1994 report by Chalfie et al¹⁶⁴⁸ on the use of *Aequorea* green fluorescent protein (GFP) as a reporter of gene expression quickly spawned a growth industry. GFP mutants are now available with cyan, green, and yellow fluorescence and with excitation characteristics far better suited to flow cytometry (and imaging, confocal microscopy, etc.) than the wild type protein. Moreover, GFP variants have been engineered to behave as sensors of such functional parameters as intracellular (or intracompartmental) pH, [Ca²⁺], etc., and, using energy transfer between molecules with different spectra, for quantitative measurements of protein-protein interactions²³⁷⁶⁻⁸. The mild-mannered reporter has shed glasses and business suit and emerged from the phone booth as Supermolecule. I have already mentioned that sorting for fluorescent protein expression now seems to account for a significant amount of flow facilities' time; this trend can be expected to continue.

Beyond Positive and Negative: Putting the -Metry in Cytometry

If you spend most of your flow cytometer time doing immunofluorescence analysis, you can pick up some bad habits. Given an instrument that often costs upwards of a hundred thousand dollars, is full of fancy electronics, has its own computer attached to it, and can probably detect a few hundred molecules of fluorescent dye in or on a cell, it does seem that we underutilize its capacities when we report the results of highly sensitive and precise fluorescence measurements as "positive" and "negative."

To be sure, sometimes "positive" and "negative" are good enough to get the job done. In the previous examples of counting various types of T lymphocytes in human peripheral blood, we defined the subpopulation of T cells by their scattering characteristics and by the presence of the CD3, CD4, or CD8, cell surface antigens, and, in general, the cells we're looking at either have a substantial amount of the antigen or have little or none. When we look at our "CD3-positive" cells, they either do or do not have substantial amounts of the CD4 and CD8 antigens. We don't need to be experienced in flow cytometry to know "positive" and "negative" when we see them in these contexts, and, using these concepts, we can obtain a satisfactory answer to the question, "What are the relative proportions of (CD3⁺CD8⁺) and (CD3⁺CD4⁺) T cells in this blood sample?"

However, if the question we are asking is, "What proportion of (CD3⁺CD4⁺) T cells are activated?" we may need to extend our conceptual framework somewhat, both in terms of biology and in terms of cytometry. "What is an activated lymphocyte?" "What is a cancer cell?" and "What is a dead cell?" are major quasitheological questions guaranteed to provoke debate among analytical cytologists for a while to come. But let's suppose we have decided to define activation in terms of expression of the CD25 antigen, which is the cell surface receptor for the cytokine interleukin-2. Well, then,

we can just gate the T cells, further gate the CD4-positive cells, and then count the CD25-positive and negative cells, right? Unfortunately not. The number of CD25 molecules on an inducer T cell seems to range from hundreds or less to many thousands; the problem in defining "positive" and "negative" is that there is no clear breakpoint.

Well, then, perhaps we could say that a cell with more than 5,000, or 10,000, or some other seemingly arbitrary number of molecules of CD25 on its surface is activated. That might work, provided we had a way of determining the number of molecules from the immunofluorescence measurement. As it turns out, this can be done, but it isn't always as easy as it looks.

The hematology counters are ahead of the fluorescence flow cytometers in this department. They all report **red cell indices**, including erythrocytes' **mean corpuscular volume (MCV)** in femtoliters and **mean corpuscular hemoglobin (MCH)** in picograms. Every instrument in every lab everywhere uses the same units. Way back in 1977, I suggested that we should have "**white cell indices**," which didn't go over resoundingly well in the Dark Ages of polyclonal antisera. The proposition has been better received of late, for several reasons. The need is more apparent, our apparatus and reagents are better, and there are people interested in developing and testing standardized materials that will make it possible for everyday users of flow cytometry to do quantitative immunofluorescence measurements. Figure 1-20 illustrates one technique, which uses beads with known numbers of antibody binding sites as standards.

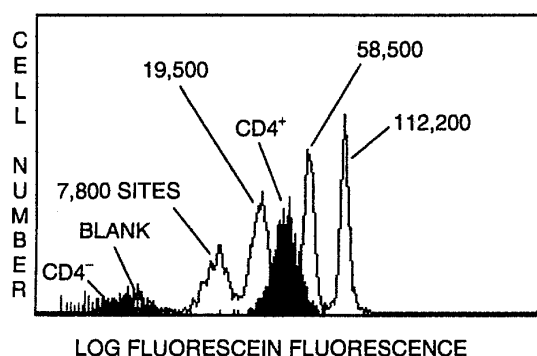


Figure 1-20. Fluorescence intensities of CD4-positive and negative cells (plotted as bars) compared with intensities of beads bearing known numbers of antibody binding sites, stained with the same fluorescein-anti-CD4 antibody as was used to stain the cells.

What Figure 1-20 shows is that most of the CD4-positive cells in the same lysed whole blood sample as is depicted in Figures 1-13 through 1-16 exhibit fluorescein fluorescence intensities consistent with there being somewhere between 19,500 and 58,500 antibodies bound to the cell surface. Is that a good number? Actually, it's probably a

little low; people who've done the experiments carefully seem to come up with an average of about 50,000 molecules of CD4 per CD4-positive cell. I may have come up with the lower number because there wasn't enough antibody added to the blood sample to bind to all of the available CD4 molecules; I didn't **titrate** the antibody, i.e., determine whether adding more antibody would have increased the cells' fluorescence intensities. So, as I said, it isn't always as easy as it looks.

However, there has been a great deal of work done on improving quantitative fluorescence measurement since the last edition of this book was written; for now, it's probably enough to mention that an entire issue of the journal *Cytometry* was devoted to the topic in October, 1998²³⁷⁹.

1.5 WHAT'S IN THE BOX: FLOW CYTOMETER ANATOMY, PHYSIOLOGY, AND PATHOLOGY

It may have occurred to you that I have spent a great deal of time dealing with history, data analysis, parameters, and probes without getting into the details of how a flow cytometer works. That fits in with my idea that what we should be concerned with, first and foremost, is what information we want to get out of the cells and what we have to do to the cells to get it. It is now fairly clear that, although we can derive some information about cell size and morphology from light scattering signals, getting the details about biochemistry and physiology will require treating the cells with one or more fluorescent probes. We are now ready to consider more of the details of how the fluorescence of those probes is measured.

Light Sources for Microscopy and Flow Cytometry

There are substantial differences in time scale between flow cytometry and microscopy. A human observer at a microscope moves different cells into and out of the field of view at a rate that is, under any circumstances, much slower than the rate at which cells are transported through the observation region (or, if you prefer, past the "interrogation point," which always seems to me to describe a "?" of a flow cytometer. The response time of the human observer is pretty long, i.e., hundredths of seconds, or tens of thousands of microseconds. That's why movies and television work; changing the picture a few dozen times a second produces the illusion of continuous motion. In flow cytometry, a cell passing through the apparatus is typically illuminated for somewhere between one and ten microseconds. This disparity in observation times means, among other things, that flow cytometers need more intense light sources than are commonly used in microscopes.

Both the **sensitivity** (i.e., how much light can be detected) and **precision** (i.e., how reproducibly this can be done) of light measurements are functions of the **amount of light**, i.e., the **number of photons**, reaching the detector. The human eye is an extremely sensitive photodetector; when properly dark-adapted, a person with good eyesight may well perceive single photons emitted from weakly fluo-

rescent or luminescent objects. The quantum nature of light obviously does not allow for any improvement upon this level of sensitivity in the electro-optical photodetectors used in flow cytometers.

Therefore, to make a flow cytometer comparable in sensitivity to a human observer, we would expect to have to get approximately the same amount of light from the observation region of the flow cytometer in a few microseconds as is collected by the observer at the microscope in a few milliseconds. Since the amount of light collected is, in general, directly dependent on the intensity of illumination, a cytometer needs a light source approximately a thousand times as bright as would be needed in the microscope.

The term **brightness**, when used in a technical sense, denotes the amount of light emitted from or through a unit surface area or solid angle, rather than the total amount of light emitted from a source. By this criterion, the 800 μW laser in a supermarket bar code scanner is brighter than the sun, and practically any laser can potentially be used as a light source for flow cytometry. The requisite brightness is also found in some kinds of **arc lamps** (high-pressure mercury and xenon lamps, sometimes specified as “short arc” lamps).

The majority of fluorescence flow cytometers now in use are benchtop models with a single blue-green (488 nm) illuminating beam, derived from an air-cooled argon ion laser. If a benchtop apparatus has a second illuminating beam, it is usually red (nominally 635 nm), coming from a diode laser. Larger instruments, such as high-speed sorters, use water-cooled argon and krypton ion lasers, which can be tuned to produce emission at a variety of UV (350-364 nm) and visible wavelengths; some systems obtain UV emission at 325 nm from an air-cooled helium-cadmium laser. Typical laser powers range from 10 to 25 mW in benchtop cytometers and up to hundreds of milliwatts in larger systems.

Instrument Configurations: The Orthogonal Geometry

Flow cytometers using arc lamp sources have been and still may be built around upright or inverted microscopes, simply by placing the **flow cell** or **flow chamber** in which cells are observed where the slide would normally go. Most modern fluorescence flow cytometers, however, use laser sources, and employ a different optical geometry, which is shown schematically in the intimidating but informative Figure 1-21 (the uncaptioned color version of the figure on the back cover may be helpful). The cytometer shown in the figure is designed to measure light scattering at small and large angles and fluorescence in four spectral regions.

The figure is a top view. If you look carefully along the left side, about halfway up from the bottom, you'll see the cell, which is, or at least should be, the *raison d'être* for the instrument and for our mutual efforts. The **core** or **sample stream** of cells would pass through the system in a direction perpendicular to the plane of the drawing, and the axes of the sample stream, the focused laser beam used for illumination, and the lens used to collect orthogonal scatter signals

are all at right angles to one another, which is why the cytometer is described as having an **orthogonal** geometry. For the time being, we won't go into the details of how the cell gets into the center of the rectangular quartz cuvette in which the measurements are made.

Laser Beam Geometry and Illumination Optics

The beam coming out of the laser is radially symmetric, but the intensity varies with distance from the axis of the beam. If you plotted intensity versus distance from the axis, you'd come up with the familiar bell-shaped **Gaussian** or **normal** distribution.

It helps our cause to illuminate the cell and as little of the region surrounding it as possible. Most cells that are subjected to flow cytometry are less than 20 μm in diameter, so it would be advantageous to focus the illuminating beam to a spot not much bigger than this. This could be done using a single convex spherical lens. However, problems arise due to the Gaussian intensity profile of the laser beam and to the vagaries of fluid flow.

In order to measure scatter and fluorescence signals from cells with a precision of a few percent, it is necessary that illumination be uniform within that same few percent over the entire width of the sample or **core stream**. As long as the sample is flowing, we know that cells will get through the plane, defined by the intersection of the axes of the illuminating beam and the collection lens, in which the observation point lies. However, while, under ideal conditions, we'd like to have the cells strung out along the axis of flow like beads on a string, in practice, there's apt to be some variation in lateral position of cells in the core stream. If the beam is focused to a very small spot, the variation in intensity of illumination reaching cells at different positions will be too high to permit precise measurements.

Calculations show that if the diameter of the focused beam is about 100 μm , there will be only about 2% variation in intensity over the width of a 20 μm sample stream. There are, however, good reasons not to use a 100 μm round spot. If cells travel through the apparatus at velocities in the range of 2-5 m/sec, it will take 20-50 μs for a cell to traverse a 100 μm beam. During this time, most of the beam will be illuminating things other than the cell, and any scatter and fluorescence signals from these things will increase background noise levels.

Since variations in intensity over the Gaussian profile of the laser beam along the axis of fluid flow aren't a problem, because each cell goes through the whole beam, it makes sense to use a relatively small focal spot dimension in the direction of the axis of flow. A spot size of 20 μm allows cells to traverse the beam in 4-10 μs , increasing illumination of the cells during their **dwel time** in the beam and decreasing background as well. If the spot is made smaller than a cell diameter, say 5 μm , cells of different sizes spend different lengths of time in the beam – everybody isn't famous for the same number of microseconds – and **pulse width** can be used to measure cell size.

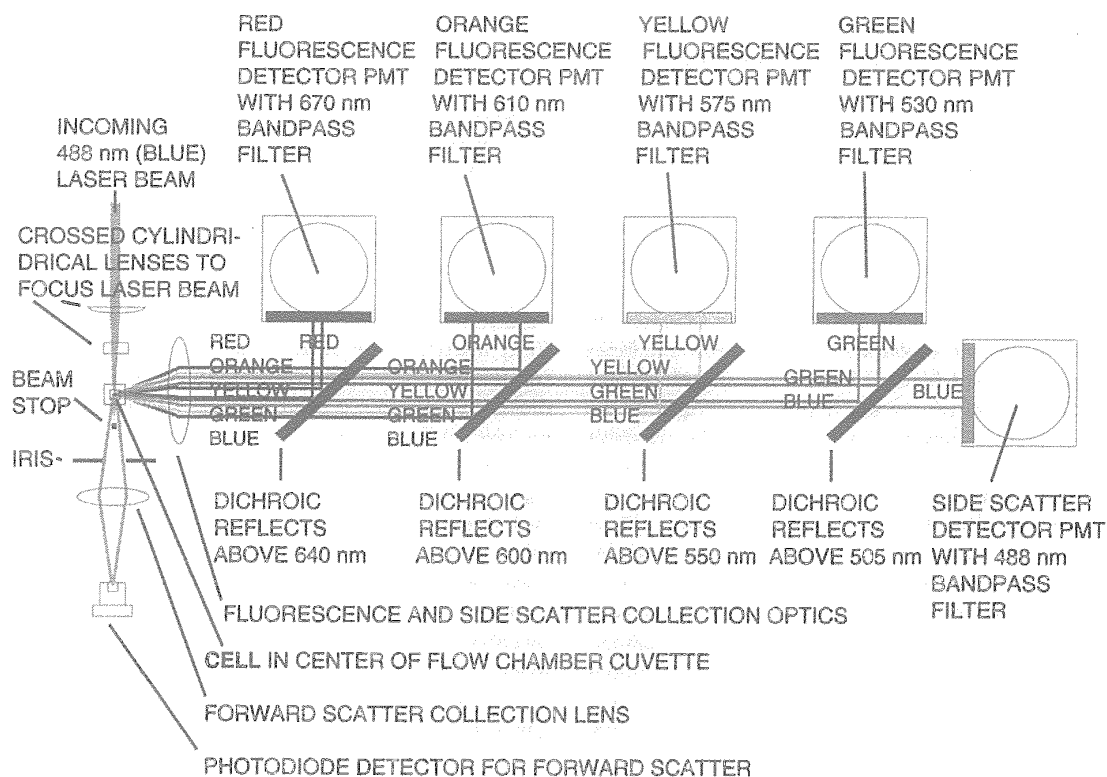


Figure 1-21. Schematic of the optical system of a fluorescence flow cytometer.

Using a really small spot, say 2 μm , you can extract a substantial amount of information about cell shape and structure by digitizing the signal at very high rates. Until recently, the processing electronics required for this technique, which is called **slit-scanning** flow cytometry, were too complex and expensive to be widely used, but the hardware and software are now more accessible should a compelling application come along. Current conventional instruments settle for **elliptical focal spots** 5-20 μm high and about 100 μm wide; these are obtained using **crossed cylindrical lenses** of different focal lengths, each of which focuses the beam in only one dimension. The crossed cylindrical lenses are shown at the left of Figure 1-21, above the cuvette. The lens closest to the cuvette is placed one focal length away from the sample stream, and focuses the beam in the dimension perpendicular to the plane, which is why you can't see the lens's curvature. The other lens, in this diagram, is placed so that its focal point is at the **beam stop**, which is a component of the forward scatter collection optics.

Flow Chamber and Forward Scatter Collection Optics

Earlier instruments examined cells in cylindrical quartz capillaries, or in a cylindrical **stream in air** following passage of fluid through a round orifice; the observation point in

most cell sorters is still in a stream in air. However, in the benchtop instruments that are most widely used, observation is done in **flat-sided quartz cuvettes** with a square or rectangular cross section. The internal dimensions of the cuvettes are typically 100-200 by 200-400 μm ; they are essentially small spectrophotometer cells and are, not surprisingly, produced for the flow cytometer manufacturers by the same companies that make spectrophotometer cells for other purposes. Cylindrical capillaries or streams in air themselves act like cylindrical lenses, and refract substantial amounts of light from the illuminating beam, which greatly increases the background noise level in scatter measurements and may also interfere with fluorescence measurements. Flat-sided cuvettes scatter relatively little of the incident light, minimizing such interferences.

The **beam stop** in the cytometer shown here is a vertical bar; we're looking at its cross section in the top view. What a beam stop needs to do is block the illuminating beam, once the beam has traversed the cuvette, so that as little of the beam as possible will reach the forward scatter detector and interfere with the measurement of light scattered by the cell at small angles to the beam. In an instrument in which observation is done in a round capillary or in a stream in air, the beam stop has to be horizontal, to block light refracted by the capillary or stream; the forward scattered light that is

detected is light scattered "up and down," i.e., out of the plane of Figure 1-21. The laws of physics that govern focused laser beams end up dictating that we can collect light scattered at smaller angles using a flat-sided cuvette and a vertical beam stop than we can using a capillary or round stream and a horizontal beam stop.

The actual range of angles over which small-angle or forward scatter signals are collected varies considerably from instrument to instrument. The lower end of the range is set by the placement and dimensions of the beam stop; in many flow cytometers, the upper end of the range is adjustable by manipulating an **iris diaphragm**, shown below the beam stop at the left of the figure. The light that gets around the beam stop and through the diaphragm is converged by the **forward scatter collection lens**, which, in the apparatus shown in the figure, is bringing the light to a focus at the **forward scatter detector**.

The detector illustrated here is a **photodiode**, a silicon solid-state device that takes photons in and puts electrons out, usually at the rate of about 5 electrons out for every 10 photons in, giving it a quantum efficiency of 50 percent. The actual sensing area of the detector is in the neighborhood of 1 mm². When you make the same kind of silicon chip with a larger surface area, you can get some fairly serious electric currents out of the resulting **solar cell**. The photodiodes used as forward scatter detectors in most flow cytometers typically have output currents of a few microamperes, not because they're smaller than solar cells, but because there aren't enough photons, even in the relatively strong forward scatter signal, to produce higher currents. When you're trying to measure forward scatter signals from relatively small particles, e.g., bacteria, a photodiode may not be up to the job, and it may be better to use a more sensitive detector, such as a **photomultiplier tube (PMT)**. These are used for side scatter and fluorescence detection, but are larger, more complicated, and - probably most important from the commercial point of view - more expensive than photodiodes. In an ideal world, the flow cytometer manufacturers would offer a high-sensitivity PMT forward scatter detector option on all models; turn on the news if you still think ours is an ideal world.

Fluorescence and Side Scatter Optics

The really hairy part of Figure 1-21, and of the average flow cytometer, is the part that deals with the collection of fluorescence and side scatter signals and the diversion of light in different spectral regions to the appropriate photomultiplier tube detectors. The first task is to collect the light. I have shown a single, simple **collection lens** for fluorescence and side scatter, but the optics actually used are somewhat more complicated.

As was noted in Figure 1-1, light is scattered, and fluorescence emitted, in all directions, i.e., over a solid angle corresponding to the entire surface of a sphere. In principle, we'd like the lens to collect light over as large a solid angle as possible, so we can collect as much of the fluorescence as

possible. One way to do this is to use a high-N.A. microscope lens to collect the light; this is done in many instruments, some of which even use a functional equivalent of oil immersion to get the highest possible N.A. Another is to place the collection lens at its focal distance from the sample stream. Various experimenters have used parabolic or ellipsoidal reflectors and high-N.A. fiber optics for light collection in attempts to increase the total amount of light collected.

As has already been suggested in the discussion of forward scatter detectors, ideal solutions are hard to come by. Every decision made in the design of a flow cytometer involves tradeoffs. In the case of light collection optics, the problem we run into is usually that, as we collect more light, we have less control over where we collect it from. What we really need to do is collect as much light from the immediate region of the cell, and as little from elsewhere, as possible, because any light we collect from elsewhere will only contribute to the background or noise. Thus, the all-important **signal-to-noise ratio** will decrease, even though the signal itself increases. Flow cytometer designs using ellipsoidal or parabolic reflectors or fiber optics for light collection have, so far, run into this problem.

The simple collection lens shown in the figure is illustrated as producing a **collimated beam** of light, i.e., one in which rays entering the lens at all angles come out parallel, with a so-called "focus at infinity". In most real flow cytometers, the light collected from the collection lens is either not collimated or is converged by a second lens, and then passes through a small aperture, or **field stop** (see p. 9), that lets most of the light collected from the region near the cell through and blocks most of the light collected from elsewhere. Some instruments incorporate an additional lens behind the field stop to recollimate the collected light, because there is some advantage in presenting a collimated beam to the **dichroics** and **optical filters** used direct light collected at different wavelengths to different detectors.

Optical Filters for Spectral Separation

The lens that collects the fluorescence emitted from, and the light scattered at large angles by, cells transmits light encompassing a range of wavelengths. Most of the light is scattered laser light, at 488 nm; much of the rest should be fluorescence from the cells, which will of necessity be at wavelengths above 488 nm. The choice of wavelength regions for fluorescence measurements is based on the fluorescence emission spectral characteristics of the available fluorescence probes or labels that can be excited at 488 nm.

The apparatus illustrated in Figure 1-21 is designed to detect fluorescence in four spectral regions, which we call **green (515-545 nm)**, **yellow (560-590 nm)**, **orange (600-620 nm)**, and **red (660-680 nm)**. It also detects scattered light at the excitation wavelength, 488 nm. Each of the detectors is a photomultiplier tube, and all of the detectors are fitted with **bandpass optical filters** that transmit light in the appropriate wavelength ranges.

There are basically two kinds of optical filters that can be used for wavelength selection; they are **color glass**, or **absorptive**, filters and **dielectric**, or **interference**, filters. Color glass filters are made of glass or plastic impregnated with **dyes** that absorb light in the unwanted wavelength regions and transmit most of the light in the desired regions. Dielectric filters are made by depositing thin layers of dielectric materials on a glass or quartz substrate; within some wavelength range, which is determined by the thickness of these layers, there will be **destructive interference**, resulting in light of these wavelengths being reflected from, rather than transmitted through, the filter.

Filters can be made with several kinds of transmission characteristics. There are **edge** filters, which may be either **long pass** or **short pass** types; long pass filters block shorter and transmit longer wavelengths and short pass filters block longer and transmit shorter wavelengths. Long pass and short pass filters are usually specified by the wavelength at which their transmission is either 50% of the incident light or 50% of their maximum transmission. There are **bandpass** filters, which block wavelengths above and below the desired region of transmission; they are specified by the wavelength of maximum transmission and by the **bandwidth**, which defines the range of transmission, usually expressed as the range between the points below and above the peak at which transmission is 50% of maximum. There are also **notch** filters, which are designed to exclude a narrow range of wavelengths.

Absorptive filters can be very effective at getting rid of light outside their desired **passbands**, i.e., those regions in which they transmit light (many transmit less than 0.01% outside the passband), and can also be made to have good (>90%) light transmission in the passband. However, the dyes incorporated into the filter to absorb the unwanted light may fluoresce; this phenomenon can (and did, in the earlier fluorescence flow cytometers) interfere with the detection of weak fluorescence signals from cells. As a result, most modern instruments now use interference filters, which reflect rather than absorb unwanted light.

Real interference filters used as long pass or bandpass filters frequently incorporate an absorptive layer behind the dielectric layers to get rid of the last little bit of unwanted light, because it's difficult to get rid of more than 99% of it by interference and reflection alone. Fluorescence in these filters is not a big problem because the interference layers get rid of most of the light that might excite fluorescence before it hits the absorptive layer — provided, that is, that you mount the filter **shiny side out**, that is, with the interference layers facing where the light's coming from and the colored absorptive side facing where it's going.

Dichroics, also called **dichroic mirrors** or **dichroic beamsplitters**, are interference filters, usually without an added absorptive layer. They can be made with either **long reflect** (i.e., short pass) or **short reflect** (i.e., long pass) characteristics, and both kinds are used in flow cytometers. As is the case with other types of interference filters, it's easier to

make a filter that reflects 97% of unwanted light than it is to make one that transmits 90% of wanted light. When flow cytometers measured fluorescence in only two spectral regions, they only needed one dichroic (maybe two, if you count one to reflect blue (488 nm) light to the orthogonal scatter detector and keep it away from the fluorescence collection optics). When you start measuring fluorescence in three or four regions, it becomes advisable to do careful calculations to make sure you don't lose a lot of the light you want in the dichroics. The Devil, as we all know, is in the details, and more deals with the Devil are made in the details of dichroics and filters than in most other areas of flow cytometer design.

The layout shown in Figure 1-21 assumes that the strongest signal, or the one with the most light we can waste, is the blue orthogonal scatter signal, and that the green, yellow, orange, and red fluorescence signals are progressively weaker. Even if all of the dichroics transmit 90% of the incident blue light, only 65% of the light coming through the collection lens will reach the filter in front of the orthogonal scatter detector PMT. About 70% of the green fluorescence will make it to the filter in front of the green detector PMT, while 77% of the yellow, 86% of the orange, and 96% of the red fluorescence will get to the filters in front of the detectors for those spectral regions. We therefore lose the least light from the weakest signal.

There are other ways to improve light transmission; one is to ditch the in-line arrangement of PMTs shown in the figure, instead first splitting the red/orange and the blue/green/yellow regions, so that the green fluorescence signal passes through two dichroics and the others through only a single dichroic. Another, which I routinely use in the "Cytomutt" flow cytometers I build, is to place a second fluorescence collection lens at 180° from the first one, so that each lens collects light for at most three detectors.

The spacing between the dielectric layers of interference filters and dichroics determines the wavelengths at which interference will occur, and, therefore, the wavelengths that will be transmitted or reflected by these components. The distance between the layers changes with the angle at which light hits the filter (remember trigonometry?), and, as a result, the passband of the filter changes with the angle of incidence of the light. In theory, light should be collimated before it gets to the dichroics and filters; this is generally not done because the light coming from the collection lens is contained within a fairly small solid angle. Problems with dichroics and filters are more likely to result from using the wrong filters or from mounting filters incorrectly. Dielectric filters also degrade over time, as moisture gets in between the dielectric layers, but, when this occurs, the filters tend to look ugly enough so that you'd think about ordering new ones.

I hope, by now, to have conveyed the impression that dichroics and filters are among the most critical parts of a flow cytometer; not surprisingly, the right — or wrong — selection of dichroics and filters can also make a big difference when

you're doing fluorescence microscopy, by eye or with image analyzers, etc. A few hundred dollars spent on good filters may dissuade you from smashing tens or hundreds of thousands of dollars worth of instrument to smithereens out of frustration.

Multistation Flow Cytometers

Before going on to a discussion of detectors and electronics, I will point out that, whereas most flow cytometers have a single excitation beam, and you can have any color you want as long as it's 488 nm, there are systems available that offer a wider choice of excitation wavelengths. Some of these can use two or more illumination beams, separated by a small distance in space. A good way to conceptualize such a **multistation flow cytometer** might be to imagine two or more copies of Figure 1-21 stacked one on top of another. Because the beams in a multistation instrument are separated by a short distance, it takes a short time for cells to travel from one beam to another, and the signals are therefore separated in time. Since the velocity of cells through the system is approximately constant, the time interval between signals from different beams is also approximately constant.

In flow cytometers that form an image of the sample stream, as most now do, it is customary to form separate images of the intersections of two or more beams with the sample stream, and divert light from each observation point to the appropriate detectors. In instruments in which no image is formed, and in which light from multiple observation points reaches all the detectors, a **time-gated amplifier** is used. This allows signals from the detectors that measure events at the downstream observation point to reach the signal processing electronics only at a set time interval after signals are detected at the upstream observation point.

Multistation instruments have also been built that incorporate electronic volume sensors as well as laser or arc lamp illumination; cell sorters are also multistation instruments, as are cell "zappers" or **photodamage cell sorters**. These use a high energy pulsed laser beam downstream from the measurement beam and switch the beam on to destroy cells with selected characteristics.

Flow cytometers with multiple illumination beams are used primarily for multiparameter measurements involving probes that cannot be excited at the same wavelength. For example, sorting human chromosomes stained with combinations of dyes that preferentially stain A-T and G-C rich regions of DNA requires separated ultraviolet (325-363 nm) and blue-violet (436-457 nm) illuminating beams. Other applications use ultraviolet and 488 nm beams and 488 and red (633 or 635 nm) beams; as many as five beams have been used in a single apparatus. The current trend is toward multiple illumination beams, even in benchtop instruments.

Photomultipliers and Detector Electronics

A **photomultiplier tube (PMT)**, like a photodiode, takes in photons and puts out electrons. However, whereas a plain photodiode never does much better than 7 electrons

out for every 10 photons in, a PMT may get as many as a few hundred thousand electrons out for each photon that reaches its **photocathode**. PMTs, like cathode ray television tubes and the tubes favored by audiophiles and rock musicians who can't see the trees for DeForest, are among the last survivors of the vacuum tube era. They incorporate a photocathode, which is placed behind a glass or quartz window so light can reach it, a series of intermediate electrodes, or **dynodes**, and another electrode called the **anode**. A voltage is applied to each electrode; the photocathode is at the lowest voltage, with each dynode at a successively more positive voltage and the anode at the most positive voltage of them all – which is usually ground, because the photocathode is generally a few hundred to a couple of thousand volts negative.

Photons hitting the photocathode result in **photoelectrons** being emitted from the photocathode, and accelerated toward the first dynode by the electric field resulting from the difference in electric potential (voltage) between these electrodes. The electrons acquire energy during this trip, so, when they whack into the dynode, they dislodge more electrons from it, which are accelerated toward the next dynode, and so on. The bigger the difference in potential, i.e., applied voltage, between stages, the more energy is imparted to the electrons at each stage, and the more electrons are released from the receiving electrode. This gives the PMT a mechanism for **current gain** that is relatively noise-free. The PMTs used in most flow cytometers have current gains as high as 10^6 . However, the **quantum efficiency** of PMT photocathodes is typically lower than that of photodiodes, with peak values of 25% (i.e. 25 electrons out for 100 photons in) in the blue spectral region, and, usually, much lower values in the red. Detector quantum efficiency is important because the sensitivity and precision with which fluorescence (or any other optical signal) can be measured ultimately depend on the number of electrons emitted from the detector photocathode.

Why is it that at detectors, we measure success one electron at a time? Because detection is subject to the same Poisson statistics we ran into on p. 19. When you count (or detect) n of anything, including photoelectrons, there is an associated standard deviation of $n^{1/2}$. When you detect 10,000 photoelectrons, the standard deviation is $10,000^{1/2}$, or 100, and the coefficient of variation (CV) is $100 \times (100/10,000)$, or 1%. When you detect 10 photoelectrons, the standard deviation is $10^{1/2}$, or about 3.16, and the CV is $100 \times (3.16/10)$, or 31.6%. I am talking about photoelectrons, rather than photons, here, because, while the detector, whether diode or PMT, "sees" photons, if you will, all the electronics lets us "see" is electrons.

If we had reliable low-noise amplifiers with gains of several million, we'd always be better off with the 50-70 electrons we could get out of the photodiode for every 100 photons hitting it than we would with the 8-25 electrons emitted from the PMT cathode under the same conditions; all the gain in the PMT doesn't get around the imprecision

introduced by the lower number of electrons it starts with and, in fact, there is also a statistical aspect to the PMT's gain mechanism.

Unfortunately, the high-gain, low noise amplifiers we'd need to use photodiodes as sensitive fluorescence detectors don't exist. There are, however, solid-state devices called **avalanche photodiodes (APDs)**, which combine high quantum efficiency with a mechanism that can produce gains as high as a few thousand when a voltage is applied across the diode. While APDs are now used for both scatter and fluorescence detection in some commercial flow cytometers, they do not match the sensitivity of PMTs.

The photodetectors we have been talking about are sources of **electric current**. A **preamplifier**, which is the first stage in the **analog signal processing electronics**, converts the current output from its associated detector to a voltage. The preamplifier also accomplishes the important task of **DC baseline restoration**.

An ideal flow cytometer is something like an ideal dark field microscope; when there's no cell in the observation region, the detector shouldn't be collecting any light at all. In practice, there's always some small amount of light coming in. In the case of the scatter detectors, most of this light is stray scattered light from the illuminating beam; in the case of the fluorescence detectors, the light background may come from fluorescence excited in various optical elements such as the flow chamber, lenses, and filters, from fluorescence due to the presence of fluorescent materials in the medium in which cells are flowing, and from **Raman scattering**, which produces light at frequencies corresponding to the difference between the illumination frequency and the frequencies at which absorption changes molecular vibrational states. In flow cytometry, the major interference due to Raman scattering results from scattering by water; when 488 nm illumination is used, this scattering occurs at about 590 nm, and may interfere with detection of signals from probes labeled with phycoerythrin, which fluoresces near this wavelength.

The net result of the presence of all of the abovementioned stray light sources is that there are some photons reaching the detectors in a flow cytometer even when there isn't a cell at the observation point, producing some current at the detector outputs. There may also be some contribution from the so-called **dark current** of the detector, which results from the occasional electron breaking loose from the cathode due to thermal agitation. There are some situations in which performance of photodetectors is improved by refrigerating them to reduce dark current; flow cytometry in the contexts we're discussing isn't one of them. Even with the detectors in liquid nitrogen, we'd have to deal with the background light, which will contribute a signal with an average value above zero to whatever signal we collect from the cells.

The background signal can be considered as the sum of a constant **direct current (DC)** component and a variable **alternating current (AC)** component, representing the fluc-

tuations due to photon statistics and to other sources of variation in the amount of stray light reaching the detector. One important source of such variation may be **light source noise**, i.e., fluctuations in the light output of the laser or lamp used for illumination; in some circumstances, particularly scatter measurements of small particles, source noise can be the major factor limiting sensitivity.

What we'd like to measure when a cell does pass by the observation station is the amount of light coming from the cell, not this amount plus the background light. We can do this, to a first approximation, by incorporating an electronic circuit that monitors the output of the detector and uses negative feedback to subtract the slowly varying component of the output from the input, thereby eliminating most of the DC background signal, and restoring the **baseline** value of the preamplifier output to ground.

In practice, baseline restorers will keep their voltage outputs within a few millivolts of ground when no cells are coming by. When a cell does arrive, it will scatter and probably emit small amounts of light, which will be collected and routed to the various detectors, producing transient increases, or **pulses**, in their output currents, which will result in voltage **pulses** at the preamplifier outputs. At this point, as was noted on p. 17, all of the information we wanted to get from the cell resides in the heights, areas, widths, and shapes of those pulses; we will ultimately convert these to digital values, in which form they can be dealt with by the computers that are almost universally used for data analysis in flow cytometry. However, before we get into the details of how pulse information is processed, we ought to consider the only element of Figure 1-21 that has been neglected to this point, namely, the cell flowing through the apparatus, and how it gets there.

Putting the Flow in Flow Cytometry

Figure 1-21 describes the cell as being in the center of the cuvette, and I have already talked about a **core** or sample stream of cells that is about 20 μm wide, while mentioning that the internal dimensions of the cuvette are on the order of 200 by 200 μm . The space between the core and the inner walls of the cuvette is occupied by another stream of flowing fluid, called the **sheath**. How the core and sheath get where they are can be appreciated from a look at Figure 1-22.

Fluid mechanics tells us that, if one smoothly flowing stream of fluid (i.e., the core stream) is injected into the center of another smoothly flowing stream of fluid (i.e., the sheath stream), the two streams will maintain their relative positions and not mix much, a condition called **laminar flow**. There are generally differences in fluid flow velocity from the inside to the outside of the combined stream, but the transitions are even. If the velocities of the two streams are initially the same, and the cross-sectional area of the vessel in which they are flowing is reduced, the cross-sectional areas of both streams will, obviously, be reduced, but they will maintain the same ratio of cross-sectional areas they had

at the injection point. If the sheath stream is flowing faster than the core stream at the injection point, the sheath stream will impinge on the core stream, reducing its cross-sectional area. In the flow chamber of a flow cytometer, both mechanisms of constricting the diameter of the core stream may be operative.

The core stream, which contains the cell sample, is injected into the flowing water or saline sheath stream at the top of a conical tapered region that, in the flow chamber shown in the figure, is ground into the cuvette. The areas of both streams are reduced as they flow through the tapered region and enter the flat-sided region in which cells are observed. Core and sheath streams may be driven either by gas pressure (air or nitrogen), by vacuum, or by pumps; most instruments use air pressure. **Constant volume pumps**, e.g., syringe pumps, which, if properly designed, deliver a known volume of sample per unit time with minimum pulsation, provide finer control over the sample flow rate. Since knowing the sample flow rate makes it easy to derive counts of cells per unit volume, flow cytometric hematology analyzers incorporate constant volume pumps; why fluorescence flow cytometers, in some cases made by the same manufacturers, do not remain something of a mystery.

The overall velocity of flow through the chamber is generally determined by the pressure or pump setting used to drive the sheath. If the sheath flow rate is increased with no change in the core flow rate, the core diameter becomes smaller and the cells move faster; if the sheath flow rate is decreased under the same circumstances, the core diameter becomes larger and the cells move more slowly. In some circumstances, it is desirable to adjust sheath flow rates; if cells move more slowly, they spend more time in the illuminating beam, receive proportionally more illumination, and they therefore scatter and emit proportionally more light. If the amount of light being collected from cells is the limiting factor determining sensitivity, slowing the flow rate can improve sensitivity, allowing weaker signals to be measured.

This aside, it is generally preferable to be able to control the core diameter, and therefore the volume of sample and number of cells analyzed per unit time, without changing the velocity at which cells flow through the system. This is done by leaving the sheath flow rate constant and changing the driving pressure or pump speed for the core fluid. More drive for the core results in a larger core diameter; more cells can be analyzed in a given time, but precision is likely to be decreased because the illumination from a Gaussian beam is less uniform over a larger diameter core. Less drive for the core gives a smaller core diameter and a slower analysis rate, but precision is typically higher. When the cytometer is being used to measure DNA content, precision is important; when it is being used for immunofluorescence measurement, precision is usually of much less concern.

The use of **sheath flow** as just described has proven essential in making flow cytometry practical. Without sheath flow, the only way of confining 10 μm cells within a 20 μm diameter stream would be to observe them in a 20 μm di-

ameter capillary or in a stream in air produced by ejecting the cells through a 20 μm diameter orifice. This would very quickly run afoul of Shapiro's First Law (p. 11). As a matter of fact, even with sheath flow, Shapiro's First Law frequently came into play when cell sorters were typically equipped with 50 μm orifices. That orifice size was fine for analyzing and sorting carefully prepared mouse lymphocytes, but people interested in analyzing things like disaggregated solid tumors might encounter mean intervals between clogs of two minutes or so. With the larger cross-sectional areas of the flow chambers now used in most flow cytometers, clogs are not nearly the problem they once were.

Clogs, however, are not the only things that can disturb the laminar flow pattern in the flow chamber. Air bubbles perturb flow, as do objects stuck inside the chamber but not large enough to completely obstruct it. In the first commercial cell sorters, the standard method for getting rid of air bubbles was to remove the chamber from its mount while the apparatus was running, and turn it upside down; the bubble would rise to the top and emerge from the nozzle along with a stream of sheath and sample fluid that would spray all over the lab. This technique became inappropriate with the emergence of AIDS in the 1980's. Now, even droplet sorters incorporate an air outlet (which I have referred to elsewhere as a "burp line") for getting rid of bubbles. In some flow cytometers with closed fluidic systems, the air bubble problem is minimized by having the sample flow in

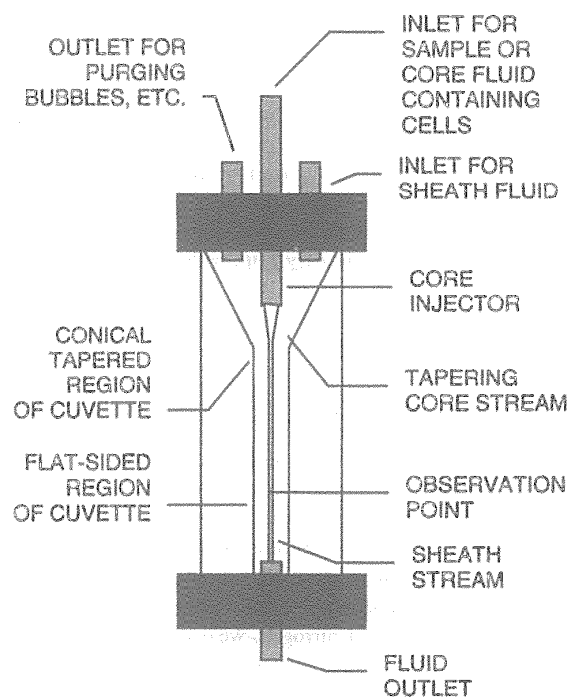


Figure I-22. A typical flow chamber design.

at the bottom and out at the top, essentially turning Figure 1-22 upside down; bubbles are more or less naturally carried out of the flow chamber.

Disturbances in laminar flow, whether due to bubbles or junk, often result in the core stream deviating from its central position in the flow chamber and in differences in velocity between different cells at different points within the core. Turbulent fluid flow is now described mathematically using chaos theory; you can recognize turbulent flow in the flow chamber by the chaos in your data.

For the present, we will assume the flow is laminar, the optics are aligned, and the preamplifiers are putting out pulses with their baselines restored, and consider the next step along the way toward getting results you can put into prestigious journals and/or successful grant applications.

Signal Processing Electronics

We have already mentioned that a cell is going to pass through the focused illuminating beam in a flow cytometer in something under 10 μ s, during which time the detectors will produce brief current pulses, which will be converted into voltage pulses by the preamplifiers. Using analog peak detectors, integrators, and/or pulse width measurement circuits, followed by analog-to-digital conversion, or, alternatively, rapid A-D conversion followed by digital pulse processing (p. 21), we will reduce pulse height, area, and width to numbers, at least some of which will, in turn, be proportional to the amounts of material in or on the cell that are scattering or emitting light. But which numbers?

First, let's tackle the case in which the focal spot, in its shorter dimension, along the axis of flow, is larger than the cell, meaning that there is some time during the cell's transit through the beam at which the whole cell is in the beam. Because the beam is Gaussian, the whole cell may not be uniformly illuminated at any given time, but intuition tells us that when the center of the cell goes through the center of the beam, we should be getting the most light to the cell and the most light out of it. The preamplifier output signal, after baseline restoration, is going to be roughly at ground before the cell starts on its way through the beam, and rise as the cell passes through, reaching its **peak value** or **height** when the center of the cell is in the center of the beam, and then decreasing as the cell makes its way out of the beam. Since the whole cell is in the beam when the pulse reaches its peak value, this value should be proportional to the total amount of scattering or fluorescent material in or on the cell.

Things get a little more complicated when the beam is the size of the cell, or smaller. In essence, different pieces of the cell are illuminated at different times as the cell travels through the beam. In order to come up with a value representing the signal for the whole cell, we have to take the **area**, or **integral**, rather than the height of the pulse. There are two ways to do this with analog electronics. One is to change the frequency response characteristics of the preamplifier, slowing it down so that it behaves as an integrator, in the sense that the height of the pulse coming out of the

slowed-down preamplifier is proportional to the area or integral of the pulse that would come out of the original fast preamplifier. Putting the slowed pulse into a peak detector then gives us an output proportional to the area or integral we're trying to measure. Alternatively, we can keep the fast preamplifier, and feed its output into an analog **integrator** instead of a peak detector.

If we decide to do digital pulse processing, we have to digitize the pulse trains from the preamplifier outputs rapidly enough so that we have multiple samples or "slices" of each pulse. We can then add the values of a number of slices from the middle of the pulse to get an approximation of the area, or integral; eight slices will do, but sixteen are better. This works pretty well. However, if we're only taking eight or sixteen slices of a pulse, we may not get as accurate a peak value or a pulse width value as we could using analog electronics.

The peak value we get from digital processing is simply the largest of our eight or sixteen slices. These provide us with only a fairly crude connect-the-dots "cartoon" of the pulse, thus, while there is a substantial likelihood that the largest digitized slice is near the peak value, there is a relatively low probability that the digitization will occur exactly when the peak value is reached.

Similarly, if we estimate pulse width from the number of contiguous slices above a set threshold value, we will have a fairly coarse measurement; if the digitization rate gives us at most sixteen slices, our range of pulse widths runs from 1 to 16, with each increment representing at least a 6 percent change over the previous value. If we had fast enough analog-to-digital converters to be able to take a few hundred slices of each pulse, and fast enough DSP chips to process the data, we could get rid of analog peak detectors and pulse width measurement circuits, but we're not there yet. The digital integrals are already good enough to have been incorporated into commercial instruments.

Is It Bigger than a Breadbox?

I have been referring to benchtop flow cytometers and big sorters, but I haven't shown you any pictures. Now's the time to fix that.

Figure 1-23, on the next page, shows the Becton-Dickinson FACScan, the first really successful benchtop flow cytometer, introduced in the mid-1980's. It uses a single 488 nm illuminating beam from an air-cooled argon ion laser, and measures forward and side scatter and fluorescence at 530 and 585 and above 650 nm. The data analysis system is an Apple Macintosh personal computer, shown in front of the operator.

Figure 1-24 (courtesy of Cytomation) shows that company's MoFlo high-speed sorter. The optical components, including two water-cooled ion lasers and a large air-cooled helium-neon laser, are on an optical table in front of the operator. Most of the processing electronics are in the rack to the operator's left; the two monitors to her right display data from an Intel/Microsoft type personal computer.



Figure 1-23. FACSscan Analyzer (Becton-Dickinson)

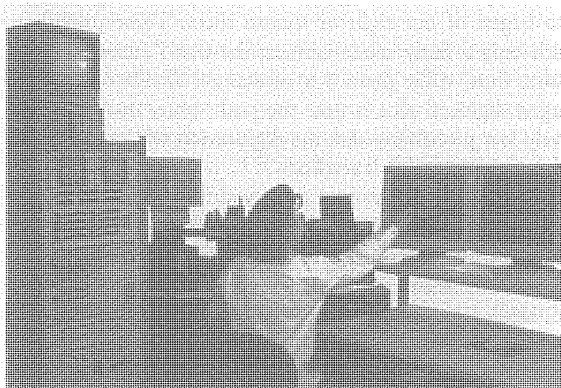


Figure 1-24. MoFlo High-Speed Sorter (Cytomation)

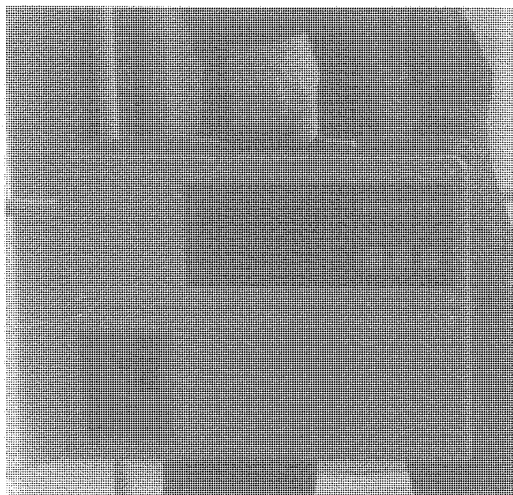


Figure 1-25. Microcyte Cytometer (Optoflow)

Neither the FACSscan nor the MoFlo risks being mistaken for a breadbox. However, the Microcyte analyzer shown in Figure 1-25 (photo courtesy of Optoflow AS) comes close. It is a two-parameter instrument with a red diode laser source, and measures medium angle scatter using a photodiode and fluorescence using an avalanche photodiode. As you might guess, it can be run on batteries.

Flow Cytometer Pathology and Diagnostics

As the benchtop flow cytometer starts to look more and more like a “black box” (okay, a “beige box,” “gray box,” or whatever from some manufacturers), it becomes increasingly important for a user to know how to verify that the instrument is running properly. It is, of course, equally important to know when a big sorter is and is not running properly, but the larger instruments tend to make their operators aware of problems.

An instrument in proper alignment, running particles through an unobstructed flow system at a rate within the manufacturer’s specifications, should get nearly identical measurements from nearly identical particles. There are now several companies producing nearly identical particles in the form of plastic microspheres, i.e., beads, impregnated with fluorescent dyes. If everything’s right, one ought to be able to make scatter and fluorescence measurements of such particles with high precision, meaning coefficients of variation no higher than a few percent. The only biological objects that are likely to yield CVs in that range are noncycling cells, such as peripheral blood lymphocytes, stained with a fluorescent DNA stain; most people stick with beads.

An instrument in which optical alignment is adjustable by the operator will typically yield the lowest measurement CVs at the point at which signal amplitudes are maximized. However, optical misalignment is not the only potential cause of poor measurement precision. Fluctuations in the power output of the light source will decrease precision, as will the presence of cell aggregates, large pieces of debris, and/or gas bubbles in the flowing stream. These create turbulence, resulting in the measured particles being distributed over an excessively large portion of the stream and/or traveling at different velocities; under these conditions, nearly identical particles will obviously not produce nearly identical signals.

Sensitivity, which, in the context of flow cytometry, basically means the degree to which fluorescence distributions from dimly stained cells (or beads) can be discriminated from distributions from unstained (control) cells or (blank) beads will usually be degraded if precision falls substantially short of the mark. Loss of sensitivity may also be due to degradation and/or incorrect choice or installation of optical filters.

Precision of instruments should always be determined using beads carrying fairly large amounts of dye, to minimize the contribution to variance from photoelectron statistics. Determination of instrument sensitivity virtually demands that at least some of the test objects used produce low-

intensity signals. Beads used for sensitivity testing typically come in sets containing an undyed or blank bead and beads loaded with four or more different levels of fluorescent dye.

Flow cytometer manufacturers and third parties also supply beads that can be used to optimize fluorescence compensation settings, and, as was previously noted in the discussion of quantitative fluorescence measurements, beads that allow the scale of the instrument to be calibrated in terms of numbers of molecules of a particular probe or label.

1.6 ALTERNATIVES TO FLOW CYTOMETRY; CYTOMETER ECOLOGY

In order to use flow cytometry to study characteristics of **intact cells** from solid tissues or tumors, or of cultured cells that grow attached to one another and/or to a solid substrate, various methods are used to prepare **single cell suspensions** from the starting material. Flow cytometry itself can provide a good indication of the efficacy of such preparative procedures. In a similar fashion, the technique can be very useful in **monitoring bulk methods** for purifying cell subpopulations, e.g., **sedimentation** and **centrifugation** techniques and **affinity-based separations**. If large cell yields are more important than high purity, bulk separation with flow cytometric monitoring may be preferable to sorting as a preparative method.

We have learned and can probably continue to learn a great deal by dissociating tissues and even organisms into suspensions of intact cells that can be characterized in flow cytometers, sorted, and subsequently studied in culture. However, the procedures used for cell dissociation, by nature, have to remove most of what holds the cells together. Since such **adhesion molecules** are probably as important as anything else for our understanding of cells' behavior, it is inevitable that there will come a point at which we won't be able to answer critical questions using cells stripped of these essential components.

It will make sense, at that point, to find instrumental alternatives to flow cytometry in a new generation of image analyzers and scanning cytometers, designed with an emphasis on preserving cell viability, which allow us to use the armamentarium of analytical techniques and reagents, in the development of which flow cytometry has played a major role, to study cells in organized groups.

We may also, of course, run up against the limits of flow cytometry simply by developing a desire to measure something repeatedly in one cell over an interval greater than a few hundred microseconds. This can be accomplished by combining static cytometry with **kinetic analysis techniques**, such as **flow injection analysis**, adapted from analytical chemistry.

I am reminded that one of the Mayo brothers said that a good surgeon had to know when to stop cutting and when not to cut; a good analytical cytologist will have to know when to put aside flow cytometry. Not now, though. Keep reading.

When we do consider the alternatives to flow cytometry, and even the availability of different types of flow cytometers, we run into something of an information gap. In the last edition of this book, I described flow cytometry as having been a growth industry since about 1985, based on census data compiled by Kit Snow of Beckman Coulter Corporation and shown in Figure 1-26.

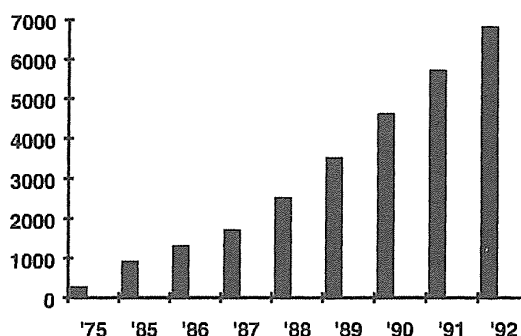


Figure 1-26. Estimated numbers of fluorescence flow cytometers in use worldwide, 1975-1992.

I tried to get updates on these numbers from various manufacturers, and even expanded my search to look for data about scanning laser cytometers, confocal microscopes, etc. Nobody's talking. The best I could do was come up with numbers that nobody would say were way too high or way too low. So here goes.

The great majority of fluorescence flow cytometers now in use are benchtop models similar to the one shown in Figure 1-23; they use low-power, air-cooled argon ion laser light sources operating at a fixed emission wavelength of 488 nm, and measure forward and orthogonal light scattering and fluorescence in three or four (green, yellow and/or orange, and red) spectral regions. Most of these systems have been designed for ease of use, with the needs of the clinical laboratory market foremost in mind. Newer instruments in the same class have added features such as a second (red) laser and closed fluidic sorting systems. The estimate is that there are somewhere between 12,000 and 20,000 such flow cytometers in use worldwide.

There are also probably around 2,000 larger, more elaborate fluorescence flow cytometers, which may use one or more air-cooled or water-cooled laser sources, can be equipped to measure eight or more parameters, and offer droplet sorting capability. These instruments are typically used in research laboratories rather than in clinical settings. Then, there are several hundred commercially produced fluorescence flow cytometers using arc lamp rather than laser sources, at least an equal number of instruments designed for multiplexed assays on beads, and one to two hundred laboratory-built flow cytometers.

The confocal microscopy folks seemed happy with the estimate that there are between 3,000 and 5,000 confocal systems worldwide; only two or three hundred of these are equipped for multiphoton excitation.

The area of relatively low-resolution scanning laser cytometry²³⁸⁰ has gotten more active in recent years. The CompuCyte Laser Scanning Cytometer (LSC), developed by Lou Kamensky²⁰⁴⁷, is generating an increasing number of interesting publications²³⁸¹⁻², about which I will say more later. I estimate that there are 100 to 250 LSCs now in circulation, and probably a similar number of volumetric capillary cytometers¹³⁶⁵, built by Biometric Imaging, now part of B-D. And there is at least one promising scanning system that hasn't yet made it into production but is worth watching²³⁸³.

1.7 THE REST OF THE BOOK

In Chapter 2, I will point you toward some sources of information that may be of use to you in learning more of the details of cytometry, flow and otherwise, discussion of which began in this Chapter and will continue in Chapter 4. Chapter 2 will also provide brief descriptions of a bunch of books on cytometry and related topics that have appeared since the last edition of this tome.

I have devoted the intermediate Chapter 3 to the history of flow cytometry, because I think that an appreciation of how things came to be as they are is as important to further progress as is an understanding of the science and technology. Chapters 5 and 6, respectively, provide additional material on data analysis and flow sorting.

Parameters and the probes used for their measurement are discussed in Chapter 7, which also presents some basic applications of flow cytometry and of some alternative methods. Chapter 8 considers flow cytometers, software, and related accessories now available from commercial manufacturers, and criteria that may influence buying decisions.

Chapter 9 briefly discusses the option of building flow cytometers; although the details on the construction, care and feeding of "Cytomutts" featured in the earlier editions have been omitted, some material that may help users understand their apparatus better has been retained.

Current and proposed applications of, and alternatives to, flow cytometry in biomedical research and laboratory medicine are considered in Chapter 10. Chapter 11 lists "Sources of Supply," while Chapter 12 is an Afterword, containing afterthoughts, aftershocks, and late breaking news. That's all I wrote. Well, almost.

Lis(z)t Mode

When cells are in such altered states
You don't know where to set the gates,
It's best to minimize the risk
And store them all on your hard disk.
If there's a clog before you're done,
You'll save some data from a run,
And, thus, you may stay out of jams
You'd get in with live histograms.

List mode, just work in list mode;
When you consider all the options, it's the only thing to do.
This mode, and only this mode,
Lets you make sense of samples that, at first, leave you without a clue.

Once we're in list mode, anyway,
With prices as they are today,
It isn't putting on the Ritz
To digitize to sixteen [or more] bits.
It's clear that, once we've made this change,
We'll have enough dynamic range
To transform data digitally,
So log amps will be history.

List mode, we'll work in list mode,
And go from linear to log and back without the log amps' ills.
Once we've got list mode, our only pissed mode
May be when we try pinning down which agencies will pay the bills.

List mode can help us analyze
How many molecules of dyes
And antibodies will be found
On each cell type to which they're bound.
At long last, different labs can see
Results compared objectively,
Advancing science as a whole
And aiding quality control.

List mode, by using list mode,
We'll all get heightened sensitivities and much reduce the fears
And trepidation of calibration,
Although the folks who make the particles may have us by the spheres.

From East to West, from South to North,
We'll send our data back and forth,
Why, we'll soon have it in our reach
To run our samples from the beach.
But, unless they've been well prepared,
When they are run, we'll run them scared,
List mode or not, there's still no doubt
That garbage in gives garbage out.

List mode, we all need list mode,
Though there are ends for which list mode itself can never be the means.
Even with list mode, there won't exist code
That gets good data from bad samples and/or misaligned machines.

("List Mode" © Howard Shapiro; used by permission. The music is derived from Liszt's Hungarian Rhapsody No. 2.)

4. HOW FLOW CYTOMETERS WORK

4.1 LIGHT AND MATTER

Introduction

In contemporary flow cytometry, measurements of light scattered, emitted, or absorbed by cells provide the values of almost all measurable parameters. Other techniques of analytical cytology, such as confocal microscopy and scanning cytometry, are just as dependent on optical principles and measurements as is flow cytometry. To understand cytometry, flow or otherwise, you need to know some basic facts about light and its interactions with matter.

Photometry versus Radiometry: What's in a Name?

When I just said that most of the information we get about cells from flow cytometry is derived from measurements of light, you probably knew what I meant. From a physicist's point of view, however, I was obviously in error. According to the precise definitions of physics, **light**, which is measured by **photometry**, is electromagnetic radiation perceptible to the human eye. The eye is most sensitive to wavelengths around 550 nm, and, in most of us, incapable of seeing much below 400 or above about 750 nm. Since, in our cytometric peregrinations, we may delve into the ultraviolet and/or the infrared, we are really measuring **radiant energy** or **radiation**, i.e., doing **radiometry**.

Lucky us. The International System of Units (SI Units)²⁴³⁸ deals with radiant energy in **joules** and with **radiant flux** (energy per unit time) in **watts** (1 watt equals 1 joule/second), units with which we are apt to have at least some familiarity. If we were, instead, forced into using the physical units related to light, we'd be up to our eyeballs in candelas, lumens, lamberts, nits, and apostilbs. Under these circumstances, we had best not make light of radiation. We should, however, get the **quantities** and **units** we will be

using on the table. The table, in this case, is Table 4-1, on the next page.

Physical Measurement Units

Of the quantities in this table, the one that is probably least familiar to you is the **solid angle**, measured in **steradians**. You may recall from high school geometry that the circumference of a circle of radius r is $2\pi r$, and that that angle which intercepts an arc of length r along the circumference is defined as one **radian**, which is approximately 57.3 degrees. A sphere of radius r has a surface area of $4\pi r^2$; one **steradian** is defined as that solid angle which intercepts an area equal to r^2 on the surface of the sphere. Figure 4-1 illustrates these concepts.

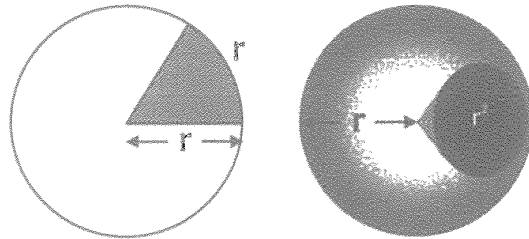


Figure 4-1. Radian and steradian.

The "pie wedge" (or is it a pi wedge?) with its apex at the center of the circle on the left side of the figure subtends a plane angle of 1 radian. The cone with its apex at the center of the sphere on the right side of the figure subtends a solid angle of 1 steradian. A "fisheye" lens of the type customarily described as having a 180-degree field of view collects light over a solid angle of 2π steradians, i.e., a hemisphere. The rest of the tabulated units are likely to be old acquaintances.

Then, there are a few other primary and derived quantities we will need.

Next stop is high school chemistry, where we first encounter the mole. One mole of an element or compound contains **Avogadro's number**, or 6.02×10^{23} molecules, of the substance, and has a mass in grams equal to the

molecular weight of the material. You may notice from the table of prefixes that 1 yoctomole, or 10^{-24} mol, comes out to less than one molecule (it would be about 0.602 molecule); as far as I know, there is not yet an international convention for which corner to start from when slicing off a yoctomole. And, while we're down at the atomic level, we should mention that the **charge** of a single electron is about 1.6×10^{-19} **coulombs**. This means that an **electric current** of 1 **ampere** (A), which is 1 coulomb/second, represents a "flow" of 6.25×10^{18} electrons/second in the direction opposite to the direction in which the current is said to "flow." The electrons don't really travel very much, but I won't get into that now.

If you want definitive information on SI Units, you can get all you'll ever need from the National Institute of Science and Technology (NIST)'s Web site:

<http://physics.nist.gov/cuu/Units/index.html>

Light in Different Lights

Since physicists from Newton's time to the middle of this century noted that light resembles waves in some aspects of its behavior and particles in others, both practitioners and teachers of physics have found it convenient to deal with light in whichever guise was more suitable to the context at hand. When I took high school physics, both the wave and the particle aspects of light were avoided in favor of an introduction to **geometrical optics**, which deals with the laws of reflection and refraction and with image formation by lenses and mirrors. In college physics, which had as a prerequisite the mathematics required to appreciate the properties of waves, the weightier topics of **physical optics** crept in, and interference, diffraction, and polarization were approached in terms of electromagnetic fields and waves, after which lip service was paid to such things as the photoelectric effect and the particle aspects of light. Further discussion of the concepts of **quantum mechanics**, and the interactions of light particles or **photons** with matter, was left for advanced courses in physics and/or chemistry.

It's All Done With Photons

In some respects, it made sense to deal with light in this schizophrenic fashion, provided you knew when to shift gears. In relation to flow cytometry, we can treat the light coming out of a laser as a "beam," and confidently predict what will happen when we bounce it off a plane mirror or two. When we want to focus that "beam" to a small spot to illuminate the cell stream, however, we have to go beyond geometrical optics to diffraction theory to calculate the focal length(s) of the lens(es) needed to achieve the desired spot size and geometry. Even after the light and the cells interact, we can cover most aspects of light scattering using a wave model of light. Once we get to fluorescence, however, we're forced to confront photons, whether we like it or not, and we find that photons behave in ways which seem strange to people who have learned "classical" physics. We also quickly discover that problems associated with the detection and

QUANTITY		UNIT
area (A)		square meter (m ²)
volume (V)		cubic meter (m ³)
mass (m)		kilogram (kg)
time (t)		second (s)
angle (θ)		radian (rad)
solid angle (Ω)		steradian (sr)
frequency (ν)		hertz (Hz)
wavelength (λ)		meter (m)
charge (q)		coulomb (C)
FACTOR		PREFIX
		SYMBOL
10 ²⁴ (= (10 ³) ⁸)	yotta	Y
10 ²¹ (= (10 ³) ⁷)	zetta	Z
10 ¹⁸ (= (10 ³) ⁶)	exa	E
10 ¹⁵ (= (10 ³) ⁵)	peta	P
10 ¹² (= (10 ³) ⁴)	tera	T
10 ⁹ (= (10 ³) ³)	giga	G
10 ⁶ (= (10 ³) ²)	mega	M
10 ³	kilo	k
10 ²	hecto	h
10 ¹	deka	da
10 ⁰	deci	d
10 ⁻¹	centi	c
10 ⁻²	milli	m
10 ⁻³ (= (10 ⁻³) ¹)	micro	μ
10 ⁻⁶ (= (10 ⁻³) ²)	nano	n
10 ⁻⁹ (= (10 ⁻³) ³)	pico	p
10 ⁻¹² (= (10 ⁻³) ⁴)	femto	f
10 ⁻¹⁵ (= (10 ⁻³) ⁵)	atto	a
10 ⁻¹⁸ (= (10 ⁻³) ⁶)	zepto	z
10 ⁻²¹ (= (10 ⁻³) ⁷)	yocto	y

Table 4-1. SI units and prefixes.

measurement of light are often best treated in terms of photons.

The truth is that light isn't sometimes waves and sometimes photons, it's sometimes waves and always photons. Geometrical optics and wave formulas work because photons, in large numbers, and over large (compared to atomic dimensions) distances, behave in ways that are, on the average, predictable and well modeled by equations describing waves. However, when you take "wave" phenomena such as interference to their limits, and set up an experiment in which half of your light should go one way and half the other, once you get down to detecting very small amounts of light, one photon at a time, you find that photons don't split in half; they either go one way or the other. So, you might ask, why don't we forget about all this wave stuff?

If you'd really like an answer to that question, you should, as the English Lit instructors say, compare and contrast two books by Richard Feynman. In Volume I of *The Feynman Lectures on Physics*⁶⁰, written for undergraduates at CalTech, he gives elegant descriptions of the behavior of light as a wave in a language that might best be described as mathematics with English subtitles. Mathematics is, of course, a language in which any serious student of physics must become fluent, and it is a useful language for the description of physical phenomena because a few lines of succinct formulae tell the whole story – to the fluent reader. The material on light in the *Lectures* is hard going, even for the budding physicists, who will have to keep studying math through their undergraduate and graduate careers just to be able to keep learning physics, which gets mathematically more difficult as it gets more advanced. It looks a lot like physics, but, as Feynman confides in some relatively non-mathematical asides, it doesn't play at the single-photon level; what does is the theory of **quantum electrodynamics**, the mathematics of which are far too complex to be taught to undergraduate physics students.

In the second book, a small and remarkable work called *QED*⁶¹, which I invoked on pp. 4-6, Feynman provides equally elegant descriptions in diagrams and English, intelligible (and entertaining) to interested laymen, rather than in mathematical terms, of photons and electrons and their interactions, which, as explained by quantum electrodynamics, account for most of what happens in the physical world, excluding gravitation and radioactivity but specifically including all of chemistry and biochemistry. This book may or may not help physicists. They have to plow through the mathematics of classical physical optics, which provide a good enough approximation to much of what goes on in the real world to be useful for everyday work; then, if they want to work in the areas in which only quantum electrodynamics gives them the right answers, they have to go into that area in a mathematically rigorous way. On the other hand, a physicist I know told me that even physicists appreciated Feynman's habit of explaining complex physical phenomena without resorting to complicated math. Unfortunately for those of us who would just like to get a little bit more of a

handle on what's going on in the instruments we use in our biomedical work, most books about physics are written by people who teach physics, and most people who teach physics see their primary mission in life as teaching physicists, and don't write books like *QED*. We have to take what we can get; for now, what we will get is back to photons.

Photons are particles that, unlike electrons, protons, and neutrons, have no **rest mass**; they are composed of pure electromagnetic energy, and the absorption and emission of photons by atoms and molecules is the only mechanism by which the atoms and molecules can gain or lose energy. Absorption and emission are **quantized**, that is, each discrete process by which an atom or molecule gains or loses energy is always associated with the same energy gain or loss, and therefore involves a photon of the same energy every time it occurs.

Aggregates of photons are detectable as **electromagnetic radiation**, which behaves like a wave traveling at the **speed of light** (c). The speed of light is approximately 3×10^8 meters per second in a vacuum, and less in materials. Some physicists in the Boston area have recently succeeded in slowing light to a speed I can easily beat on my bicycle; this is not yet of much practical interest, but it makes me feel as if I'm in great physical shape.

From the **frequency** (ν), in hertz (Hz) (formerly called cycles per second), or from the **wavelength** (λ), in meters, of an electromagnetic "wave," we can calculate the **energy** [E , in joules (J)] of a single photon, using the formulas

$$E = h\nu \quad \text{and} \quad E = hc/\lambda.$$

In these, h is **Planck's constant**, which is roughly 6.63×10^{-34} joule-seconds. A single photon coming out of an argon ion laser emitting at a wavelength of 488 nm has an energy of approximately 4.07×10^{-19} J. To get a whole joule out of a 488 nm laser, you'd need 2.45×10^{18} photons. Since 1 watt (W) is equal to 1 J/s (that's joule/second), a laser emitting 10 mW at 488 nm is putting out 2.45×10^{16} photons per second. Photon energies are higher at shorter and lower at longer wavelengths; a 325 nm (UV) photon from a helium-cadmium laser has an energy of 6.12×10^{-19} J, so 1 J of photons at this wavelength contains only 1.63×10^{18} photons, while, at the 633 nm (red) emission wavelength of a helium-neon laser, the energy of a single photon is 3.14×10^{-19} J, and there are 3.18×10^{18} photons/joule.

A Few Warm Bodies

The photons most readily accessible to man have, throughout history, been derived from **incandescent** sources, that is, objects that produce light solely by virtue of their temperature. The distribution of energies of photons emitted from an incandescent object shifts toward higher modal values as the temperature increases. At relatively low temperatures, emission in the infrared predominates; this is perceptible as **heat** rather than as light. At about 1000 degrees Kelvin (K) (or "1000 Kelvins" in SI), the object ap-

pears dull red; at about 1750 K, yellow. By the time the surface temperature of the sun, about 6000 K, is reached, the color is what we have become conditioned to as "white." Stars hotter than the sun appear bluish white to our eyes.

The physical theory that has been developed to deal with incandescence describes the behavior of an ideal radiating source called a **black body**, which absorbs all radiation falling on it and emits radiation at all frequencies with 100 percent efficiency. According to the **Wien displacement law**, the wavelength λ_m (in μm) at which maximum power is radiated from a black body at temperature T is given by $\lambda_m = 2898/T$. The related wavelength λ'_m , at which the maximal number of photons are emitted per unit time (remember, there are more photons per joule at longer wavelengths) is given by $\lambda'_m = 3670/T$. Working out the numbers for the sun, we get maximum power output at 0.483 μm , or 483 nm; that's blue-green, but the maximum photon output is at 612 nm, which is orange. For the cooler but more accessible 3400 K photoflood lamp, we come up with maximum power output at 852 nm and maximum photon output at 1079 nm; both of those wavelengths are in the infrared. In other words, incandescent sources of any kind are going to give us a lot of heat along with our light, and we're going to have to use a pretty hot incandescent source, such as an arc lamp, if we want to get much light out in the blue, violet, and ultraviolet.

The attractive alternative, in our modern world, is to dispose of the warm bodies and use a **laser**, which does not depend upon incandescence to produce its light. I said on p. 50 that a fairly puny laser is actually brighter than the sun; the numbers to follow will prove it. The total solar power reaching the earth (actually, reaching the outer layer of the atmosphere) is about 1400 W/m^2 ; the 0.8 mW (800 μW) lasers in bar code scanners have a beam diameter of 0.8 mm and, therefore, a beam area of $5 \times 10^{-7} \text{ m}^2$, and thus produce 1600 W/m^2 . Will wonders never cease?

Polarization and Phase; Interference

Two characteristics of light that are most easily understood in terms of the wave model are **polarization** and **phase**. Light behaves as a **transverse** wave or, more precisely, as two perpendicular transverse **electric** and **magnetic** waves, both perpendicular to the direction of propagation. We can start trying to clarify this with Figure 4-2.

Wave motion is **periodic**; it repeats itself in both space and time. An electromagnetic wave has associated with it periodic changes in both the **electric field** and the **magnetic field**. Electric and magnetic fields are **vector fields**; that is, each has both a **magnitude** and a **direction**. In a light wave, the electric field vector **E** and the magnetic field vector **B** are mutually perpendicular, or orthogonal, and both are also perpendicular or orthogonal to the axis of propagation.

Using a Cartesian coordinate system, with the light propagating along the z-axis, we have the electric field vector oscillating along the y-axis and the magnetic field vector oscillating along the x-axis; this keeps the electric and mag-

netic waves perpendicular to one another and perpendicular to the axis of propagation. This arrangement has actually overspecified the system somewhat; all of the **E** vectors are lined up in a neat row in a single plane, as are all of the **B** vectors. This is the way things are in what we call **plane** or **linearly polarized** light.

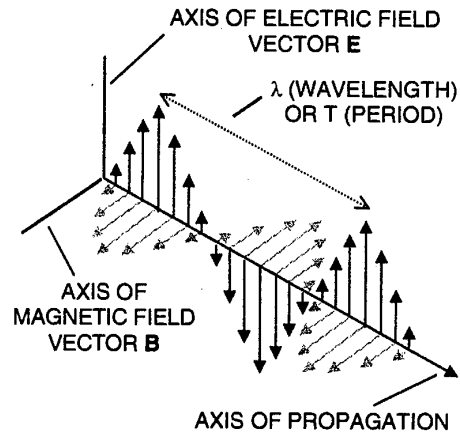


Figure 4-2. Light as an electromagnetic wave.

Since we know that the magnetic field vector is perpendicular to the electric field vector, and since most of the interesting things that happen when light interacts with matter relate directly to the electric field vector, we will, from this point on, keep track only of what's happening to the electric field or **E** vector to simplify things.

If you look at the axis of propagation in Figure 4-2, you might wonder why I identified the same dimension as both the **wavelength** and the **period**. I did because the wave can be considered as propagating in **space** or as propagating in **time**. Looking at the axis as a time axis, and the figure as representing successive values of the amplitudes of the **E** and **B** vectors at a fixed point in space, we can define the time interval, in seconds, between two successive occurrences of the maximum amplitude of the **E** vector as the **period** of the wave. The inverse of the period is the **frequency**, in Hertz, which indicates the number of complete **cycles** from maximum value to maximum value that occur in one second.

If you look at the axis of propagation as a space axis, the figure represents values of the amplitudes of the **E** and **B** vectors at different points in space at a fixed point in time. The distance between two points in space at which the vector is at maximum amplitude is the **wavelength**, measured in meters, or, more often, in **micrometers** ($1 \mu\text{m} = 10^{-6} \text{ m}$; they used to be called **microns**), **nanometers** ($1 \text{ nm} = 10^{-9} \text{ m}$), or **Angstrom units** ($1 \text{ \AA} = 10^{-10} \text{ m}$). The inverse of wavelength is also referred to as **frequency**; this, however, is a spatial, rather than a temporal frequency, and is measured in cycles per unit distance, rather than in cycles per second. The customary unit is the **wavenumber**; it denotes the number of cycles per centimeter, and is abbreviated cm^{-1} . Spatial frequency is often given the same symbol, ν , as is

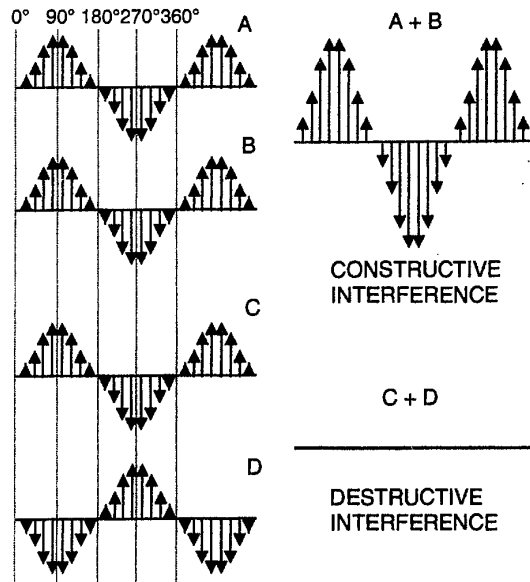


Figure 4-3. Constructive and destructive interference.

used for temporal frequency; as your high school physics teacher must have told you, check your units.

Each **cycle** of a wave can be represented as a full circle; a cycle doesn't have to be measured from peak to peak, but could be measured from trough to trough, or between any two places at which the amplitude and slope of the waveform are equal. The full circle corresponds to an angle of 360° , or 2π radians. Two waves of the same frequency, displaced from one another along the axis of propagation, are said to differ in **phase**. The phase difference is conveniently expressed as an **angle**, given in degrees or radians; phase differences are therefore restricted to values between 0° and 360° (or between -180° and $+180^\circ$).

Two waves of identical amplitude with a phase difference of 0° (or 360°) are superimposable and add to one another, producing a wave with twice the amplitude; this is called **constructive interference**. Two waves with a phase difference of 180° , or π radians, cancel one another; this is **destructive interference**. The principle is illustrated in Figure 4-3; waves A and B are in phase, that is, they have a phase difference of 0° , and add; waves C and D have a phase difference of 180° , and cancel. Both **phase contrast** and **interference contrast** microscopy are based on interference.

Suppose we had two plane polarized waves of the same frequency and phase, but with different, orthogonal planes of polarization, such that the E vector of one wave oscillates along the x-axis, while the E vector of the other oscillates along the y-axis. The two E vectors would then add, as vectors add, to give us a new wave, with an E vector at a 45° angle ($\pi/4$ radians) to both the x- and y-axes. Note that this situation is different from the case illustrated in Figure 4-3, in which the waves were polarized in the same plane. If the

two waves were equal in amplitude, but different in phase by 90° or $\pi/2$ radians, they would add to produce a wave with an E vector that would rotate around the z-axis; this resultant wave would be **circularly polarized**. Circularly polarized light is somewhat crudely illustrated in Figure 4-4.

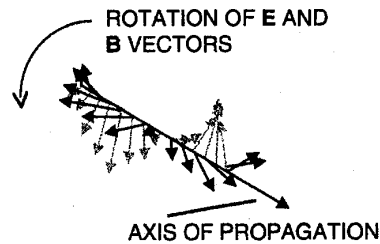


Figure 4-4. Circularly polarized light.

Two waves with perpendicular E-vectors that are different in amplitude and out of phase by 90° produce **elliptically polarized** light. Circular and elliptical polarization can be either **left-** or **right-handed**, depending upon the direction of rotation of the vectors. As long as the E and B vectors remain perpendicular to one another, they can rotate freely in their mutual plane. So, we could have a situation in which both the electric and the magnetic field vector, while remaining perpendicular to one another in the x-y plane, moved through a 360° range of angles with the x- and y-axes; this would give us **unpolarized** light.

Light Meets Matter: Rayleigh and Mie Scattering

In a vacuum, light, whether we're thinking of it as waves or photons, travels in straight lines at a velocity, c , of approximately 3×10^8 m/s, carrying with it its oscillating electric field. Just get some matter, even a few atoms worth, into the picture, though, and things change. Even in unperturbed atoms and molecules, there is some separation of positively charged protons and negatively charged electrons. When exposed to an oscillating electric field, these positively and negatively charged atomic constituents move in opposite directions in response to the field, becoming alternately closer together and farther apart, and thus giving rise to oscillating fields, i.e., electromagnetic waves, of their own, with the same frequency as the light which initiated the process, but not necessarily either of the same phase or propagating in the same direction. The resulting phenomenon is described as **light scattering**, although the "scattered" light is actually new photons, rather than old ones that have changed direction (p. 5). So far, so good, but where does the quantization I mentioned before fit in? The best answer is that it is conspicuous by its absence.

Among the interactions between light and matter, scattering is perhaps the most common and certainly the most casual. In more intimate encounters, in which photons are absorbed, atoms and molecules tend to be finickier about the energies of the photons involved, and quantization is more obvious. Even in these cases, however, we are dealing with

probabilities. If we look at the absorption spectrum of a compound, we find one or more peaks corresponding to wavelengths, or frequencies, or photon energies at which the molecule involved is most likely to absorb photons. Although absorption of photons of other energies is less likely, it is not impossible, and the likelihood of absorption increases as the photon energy approaches regions near absorption peaks. The peaks can also be dealt with in a mechanical model as representing **resonant frequencies** of the molecule; transfer of energy to the molecule from an incident wave becomes more efficient as the frequency of the wave approaches a resonant frequency.

Scattering exhibits this characteristic as well. Materials scatter light at wavelengths at which they do not absorb. Atoms and small molecules (dimensions less than $1/10$ wavelength) scatter, but do not absorb light in or near the visible region (let's consider this wavelength range to be 350-850 nm), and therefore appear transparent to the eye; they typically have absorption bands in the ultraviolet below 300 nm. As the wavelength of incident light decreases, approaching these absorption bands, the amplitude of oscillation induced in the intramolecular dipoles formed by charge separation increases, which means that the intensity of light scattering increases. The intensity of such **Rayleigh scattering** is directly proportional to a property of the scattering molecules called **molecular polarizability** (the term polarizability as used here relates to electric dipole formation, not to polarized light), and inversely proportional to the fourth power of the wavelength of the incident light. The cloudless sky appears blue in sunlight because gas molecules in the atmosphere scatter more light at shorter (i.e., blue and violet) than at longer wavelengths.

Polarization can be produced by Rayleigh scattering of unpolarized light. The E vectors of both the incident and the scattered light oscillate in planes perpendicular to the axis of propagation of the incident light. Light scattered at 90° to the incident light, however, itself propagates in a plane perpendicular to the axis of the incident light, i.e., in the same plane in which its E vector oscillates. If you are having trouble visualizing this, we can go back to Cartesian coordinates, and you can lend me a hand, let's say your right one. Cock your thumb and finger at right angles (go ahead, make my day), and then extend your middle finger perpendicular to the plane defined by your thumb and forefinger (the old hand has gotten a little stiff since high school physics, eh?).

Suppose the incident light propagates along the z-axis, represented by your middle finger. The E vectors of the incident and of the scattered light are oscillating in the plane defined by the your thumb and index finger, which respectively represent the x- and y-axes. Now imagine your thumb and index finger alternately lengthening and contracting, or move them back and forth, to represent the oscillation of the E vectors. Looking in the direction of the incident beam, i.e., end-on at your middle finger, you could readily perceive changes in length of either your thumb or index finger.

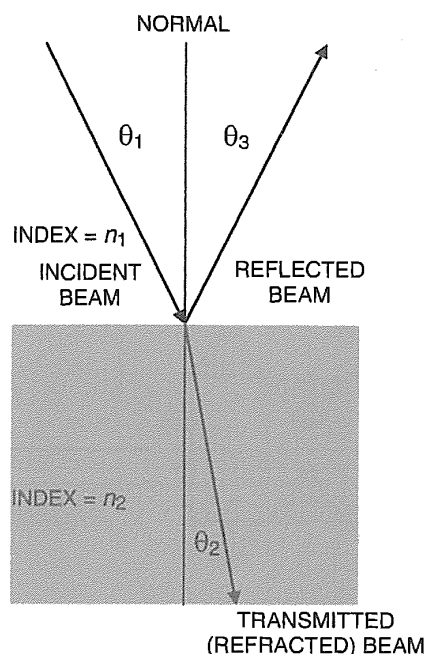


Figure 4-5. Reflection and refraction of light at a surface.

However, if you looked in the direction of light scattered at 90° from the incident beam, say, along the x-axis, or end-on at your thumb, you would be able to perceive a change in length of your index finger much more readily than a change in length of your thumb.

This is a handy way of getting you to appreciate one of the basic rules (of thumb?) of this silly game, which is that oscillations of the E vector in the direction of propagation don't count as light. In our coordinate system, with incident light propagating along the z-axis, when we look at 90° scattered light along the x-axis, we only see, or detect, light associated with oscillations of the E vector in the y-direction. This light, according to the definition in the previous section, is therefore linearly polarized. If you don't trust your right hand, you can see real polarization occurring via this mechanism by looking through a polarizing filter or polarized sunglasses at the blue sky at angles of 90° and 180° from the sun. At 90° , rotating the filter produces a noticeable change in brightness, because much of the light from this direction is polarized; at 180° , little or no effect is seen.

In Rayleigh scattering, the intensities of forward (near 0°) and back (near 180°) scattered light are nearly equal, because the scattering particles are so small that there can be only small differences in phase between light scattered from any two points within a particle. For larger particles, substantial phase differences may exist, leading to **interference**. Over a range of particle sizes from about $1/4$ wavelength to tens of wavelengths, increasing amounts of light are scattered in the forward direction; this is **Mie scattering**, named for

Gustav Mie, who worked out the relevant theory. Flow cytometric forward scattering measurements for cell "sizing" are based on Mie's analysis; Mie scattering itself is a complex phenomenon influenced by numerous factors in addition to particle size, and we will discuss it and them in detail later.

A Time for Reflection – and Refraction: Snell's Law

Both **reflection** and **refraction** of light at surfaces result from light scattering, and both, like Mie scattering, involve interference. These phenomena can be described by either wave or photon models. The reflection and refraction of light at a plane surface are illustrated in Figure 4-5.

Incident light strikes the surface at an angle θ_1 to the normal. The **angle of reflection**, θ_2 , is equal to the **angle of incidence**, θ_1 , regardless of the material of which the surface is made. The angle θ_2 at which light is **transmitted** through the surface, does, however depend upon the composition of the material. According to **Snell's law of refraction**,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2,$$

where n_1 and n_2 are the **refractive indices**, respectively, of the materials through which the incident and transmitted beams pass. In the figure, the incident beam is shown in air, which has a refractive index very close to 1 (about 1.0003).

According to wave models, the light reflected from a surface is light scattered backwards by the layers of molecules of the scattering material nearest the surface. Reflected light is observed at the angle $\theta_2 = \theta_1$ because light scattered at all other angles is removed by destructive interference. Light scattered forward, i.e., in the direction of transmission, by the surface layers is parallel to the incident beam, but lags behind the incident beam in phase, and adds to the now somewhat attenuated incident beam to produce a resultant wave with a slight phase lag. This wave is scattered from the next deeper layer of molecules, with the result that the light transmitted through this layer is still further retarded in phase, etc. The cumulative phase lags result in the light wave appearing to travel through the material at a velocity lower than the speed of light in a vacuum, c ($\approx 3 \times 10^8$ m/s); the velocity of light in a material of refractive index n is c/n . Fine, but why does the transmitted light travel at an angle different from the angle of incidence?

The wave model gives us both highly complicated and simple explanations, which come down to the same thing. The complicated explanation states that Maxwell's equations, which describe electromagnetic radiation beautifully if you're fluent in vector calculus, have to be satisfied at the surface as well as on either side of it; you then wade through three pages of formulas and find out that the transmitted light has to change direction. The simple explanation considers a wave entering the surface of the material.

As before, we'll have the wave coming in at the angle θ_1 . This is a transverse wave; its "crests," or wavefronts, can be treated as planes that are parallel to one another and perpendicular to the direction in which the light is traveling. Sup-

pose that the wavelength, i.e., the perpendicular distance between the wavefronts, is 600 nm in air (refractive index ≈ 1) and that the material has a refractive index of 1.2. The frequencies of the incident wave in air and the transmitted and scattered light in the material will be the same. The velocity of a wave is the product of the frequency and the wavelength. In air, the velocity is close to c , and the wavelength is 600 nm. In the material, the velocity is c/n , $= c/(1.2)$. Since the frequency remains constant, the wavelength in the material must be $600/(1.2)$ or 500 nm.

As the wave enters the surface, then, the wavefronts have to be 600 nm apart on the outside and 500 nm apart on the inside. Since the wave enters the surface, and is transmitted, at angles, the distance between wavefronts **along the surface, on the outside**, will be greater than 600 nm; it will be $(600/\sin \theta_1)$ nm. The distance between wavefronts **along the surface, on the inside**, will be $(500/\sin \theta_2)$ nm. Since these two distances must be equal, the angles of incidence and transmission will be unequal unless both are zero. We can bend the wavefronts, but we can't break them.

The explanations of reflection and refraction based on the photon model (see *QED*⁶⁴ for details) also, incidentally, strip some of the mystery from a well-known, but less well understood, law of optics known as **Fermat's principle of least time**. This principle can be used, without benefit of photons, to find the directions in which reflected and refracted light go, based on the light always picking the path that takes least time to traverse. There are two problems with Fermat's principle; one is that, just when you least expect it, light will turn out to take the path that takes the most time to traverse, the other is that it seems to require the light to have road maps or some other advance information.

When we're looking at light as photons, we can't say any given photon will go any given way, we can only compute the probability that a photon will go in any particular direction. As it turns out, when we consider reflection and refraction, we end up with a substantial probability that photons will go in the directions we have otherwise found reflected and refracted light to prefer, and a very low probability that they will go in other directions. The neat thing about this is that probabilities for most trajectories cancel out, leaving nonzero probabilities only for trajectories that take approximately the same time to traverse. This time is usually the shortest possible time, but occasionally the longest, in accordance with Fermat's principle. The photons don't need road maps; they just roll dice at every intersection and get there anyway.

Polarization by Reflection; Brewster's Angle

Light reflected from a surface is actually scattered from the material beneath the surface as well as from material at or near the surface. This light is linearly polarized, for the same reason that Rayleigh scattered light at 90° from the incident beam is linearly polarized; that is, oscillations of the E vector of the reflected light are only detectable in the direction perpendicular to the plane defined by the axis of propagation of

the incident light and the normal to the surface. The maximum polarization occurs when the angle between the reflected light and the transmitted light is 90° . At that point, $\theta_3 + \theta_2 = 90^\circ$, and, since $\sin(90^\circ - x) = \cos x$, Snell's law gives us $(\sin \theta_3 / \cos \theta_3) = n_2 / n_1$, making $\theta_3 = \tan^{-1}(n_2 / n_1)$. This value of θ_3 is known as **Brewster's angle**. Each time light strikes a surface at Brewster's angle (remember that $\theta_1 = \theta_3$), the reflected light is linearly polarized normal to the plane of incidence, and the transmitted light therefore contains a greater proportion of light polarized parallel to the plane of incidence. The end windows of a laser plasma tube are typically affixed at Brewster's angle to the optical axis of the tube. Since light emitted from the laser has effectively been reflected many times by the mirrors at opposite ends of the laser cavity, thus passing through these **Brewster windows** many times, the laser light is highly polarized.

Dispersion: Glass Walls May Well a Prism Make

Like the atmosphere, the "transparent" materials such as glass and quartz out of which we make optical gadgets do not absorb much visible light, but scatter increasing amounts of incident light as the wavelength gets nearer their absorption peaks in the ultraviolet. As the intensity of scattering increases, more phase lag is introduced into the transmitted beam; light thus appears to travel more slowly through the material. The refractive index therefore is higher at shorter wavelengths than at longer ones, meaning that the angles at which light of different wavelengths or colors are transmitted will differ, with blue and violet light more strongly deviated than orange and red. The change in refractive index with wavelength is called **dispersion**; it is the mechanism by which a **prism** forms a spectrum from incident white light. The degree of dispersion is different for different types of glass; high-dispersion glasses are used to make prisms. Dispersion is also the basis for **chromatic aberration**, an undesirable characteristic of lenses arising from different colors of light being focused at different distances from the lens.

Interference in Thin Films

A small fraction of the incident light is reflected from any interface between two materials of different refractive indices. When we pass monochromatic light through slabs of different thickness made from the same material, and measure the amount of light reflected, we find considerable variation with thickness, because the reflections from the front and back surfaces will, depending upon the thickness, interfere constructively or destructively. The difference in thickness over which a change from constructive to destructive interference occurs is less than a wavelength, so the effect is usually much more noticeable in thin layers of material than in thick slabs. A layer of material of a given thickness will reflect colors selectively, because there will be constructive interference at some wavelengths and destructive interference at others. Which wavelengths are maximally reflected will also depend upon the angle of incidence of the light, because this will determine the angle at which light

reflected from the back surface is transmitted through the medium, and, therefore, the distance traveled by this reflected light.

Interference effects produce the iridescent colors reflected from thin films of oil on water; under much more carefully controlled conditions, thin layers of dielectric material can be deposited on optical components to produce **interference filters**, **antireflection coatings**, and **mirrors**. A typical antireflection coating for a camera lens, for example, is made by depositing a material such as magnesium fluoride, with a refractive index intermediate between the indices of air and glass, on the lens surface. Destructive interference at the wavelength λ will occur if the thickness of the coating is $\lambda/4$; the coating thickness used is generally about 125 nm, which is chosen for maximum efficacy against reflections in the green region of the spectrum, where the eye is most sensitive. The coating appears purple since, by design, it reflects very little green light (instead allowing it to be transmitted; energy is conserved) but does reflect somewhat more light at the violet and red ends of the spectrum.

Interference filters are usually made of several "sandwiches" of dielectric material separated by spacers; the structure of the filter is determined by the location and width of the **passband** and the degree of transmission of wanted light and reflection of unwanted light required. Dielectric thin films can also be used to make highly reflective **mirrors**, such as those used in lasers. Operation of lasers at some wavelengths requires that the mirrors reflect more than 99.5 percent of incident light. While such high reflectivity can be achieved fairly readily, it is generally possible to maintain it over only a small portion of the spectrum, and thus necessary to use different sets of mirrors when a laser is operated in different spectral regions. A spectrally selective mirror is said to be **dichroic**; in cytometry, the term is generally reserved for a spectrally selective mirror used at or near a 45° angle of incidence to separate light into two spectral bands.

Interference and Diffraction; Gratings

Diffraction, in its broadest sense, describes any departure from the predictions of geometric optics. Using geometric optics, for example, we would predict a sharp edge illuminated by a point source to cast a sharp shadow; instead, we see alternating bright and dark fringes at the periphery of the shadow of such an object. Light passing through one or more slits or apertures also produces fringe patterns. Diffraction results from interference, and it is diffraction that ultimately limits the resolution of optical systems; we will have more to say on this subject later.

We have already noted that selective transmission and reflection of different wavelengths can be achieved using thin layers of material. Similar effects can be obtained by passing light through an array of small slits or by reflecting light from a surface containing closely spaced (i.e., separated by a few wavelengths) grooves. Such structures are referred to as **gratings**; they are useful because they can provide greater spectral separation than can be obtained using prisms and

because the spectra they produce are linear, whereas those produced by prisms are not. Precisely ruled gratings find application in spectrophotometers, while less precisely ruled ones are used in costume jewelry. Although both old-fashioned phonograph records and compact discs can act as gratings to disperse light, their primary esthetic appeal is, or should be, auditory.

The "smooth" surface of a crystal does not behave as a grating when illuminated with visible light, because the periodic structure of atoms or molecules in a crystal has dimensions much smaller than the wavelength of the light. Crystal-line materials, however, will produce diffraction patterns in response to electromagnetic radiation at appropriately shorter wavelengths, e.g., x-rays, and this provides the basis for x-ray crystallographic determination of molecular structures.

Optical Activity and Birefringence

Although molecular dimensions are small compared to the wavelengths of visible and near-visible light, the scattering of light at these wavelengths is highly dependent upon structural characteristics of the molecules. The presence of asymmetric carbon atoms and/or helical structure in a small or large molecule, and the presence of oriented asymmetric macromolecules in a material, typically make the molecule or material scatter light of different polarizations differently.

These molecular properties account for **optical rotation**, a situation in which the plane of polarized light transmitted by a material is rotated with respect to the plane of polarization of incident light. They are also the basis for **birefringence**; a birefringent medium exhibits different values of refractive index for different polarizations of incident light. Both optical rotation and birefringence may, when measured in whole cells, provide some information about internal macromolecular structure.

Some molecules exhibit little or no birefringence when randomly oriented but become strongly birefringent when aligned, either by mechanical forces, e.g., in a flowing fluid, or by an electric field. This has provided a basis for a number of practical inventions, ranging from Edwin Land's original sheet polarizer (whence the Polaroid Corporation's name) to the ubiquitous liquid crystal display.

Matter Eats Light: Absorption

As you are now aware, a lot of optical phenomena can occur without benefit of absorption. Some of those in which we are most interested, however, such as fluorescence, cannot. Wave models don't add much to discussions of either absorption or fluorescence, both of which are conveniently understood only in terms of relations between molecules and photons.

According to quantum mechanics, molecules can only absorb energy as quanta, or photons, and any given type of molecule can only absorb photons with energies in specific ranges. Thus, a molecule can only exist in a countable number of discrete **energy states** (I said countable rather than

finite because the number is potentially infinite, but we generally don't have to worry about more than a few hundred). Absorption of a photon by a molecule in a minimum-energy **ground state** raises the molecule to a higher-energy **excited state**. A change from one state to another is called a **transition**.

There is a hierarchy of energy states. The total energy content of a molecule is the sum of **electronic, vibrational, rotational, translational, electron spin orientation, and nuclear spin orientation** energies; the energies of photons associated with these different types of transitions differ by orders of magnitude. Electronic transitions involve absorption in the near infrared, visible, and ultraviolet, with photon energies ranging from about 2×10^{-19} J at 1 μm to about 10×10^{-19} J at 200 nm. Changes in vibrational energy states accompany absorption in the infrared; the photon energies range from about 2×10^{-21} J at 100 μm to about 7×10^{-20} J at 3 μm . Thus, the energies involved in vibrational transitions are 1-2 orders of magnitude smaller than those involved in electronic transitions.

Rotational transitions result from absorption of microwave radiation, with photon energies 2 to 3 orders of magnitude lower than those involved in vibrational transitions. Molecular translation requires even lower energies. We associate molecular motion with heat; while many of us now use quantized absorption of microwave energy to increase the motion of water molecules, many more of us rely on thermal energy to heat food. Even in microwave ovens, some of the cooking is accomplished by the transfer of thermal energy from water molecules to other molecules that do not absorb microwave energy.

A substantial amount of molecular motion can be produced solely by thermal agitation even at room temperature. At any given Kelvin or absolute temperature T , the ratio of the numbers of molecules, n_{upper} and n_{lower} , in upper and lower energy states associated with an energy difference ΔE can be calculated from the **Boltzmann distribution law**,

$$n_{\text{upper}}/n_{\text{lower}} = e^{-\Delta E/kT},$$

where k is **Boltzmann's constant** (1.38×10^{-23} J/K). Let's plug in some figures for different kinds of transitions. My microwave oven runs at 2450 MHz, or $2.45 \times 10^9/\text{s}$; the corresponding photon energy (p. 103) is 1.62×10^{-24} J. At a room temperature of 300 K, $n_{\text{upper}}/n_{\text{lower}}$ for rotational transitions of this energy (i.e., $\Delta E = 1.62 \times 10^{-24}$ J) turns out to be 0.9996. In other words, heat energy at this ambient temperature raises almost as many molecules to the upper state as remain in the lower state.

Next, we'll consider a vibrational transition produced by another high-tech device sometimes used for heating water, in this case a carbon dioxide laser emitting at a wavelength of 10.6 μm . The photon energy is 1.88×10^{-20} J, and $n_{\text{upper}}/n_{\text{lower}}$ at 300 K is .0107. Thus, the thermal agitation present at room temperature is sufficient to put only about one percent of molecules into an upper vibrational state.

Absorption of visible light produces electronic transitions. A (red) He-Ne laser operating at 633 nm emits photons with an energy of 3.14×10^{-19} J. At 300 K, the ratio $n_{\text{upper}}/n_{\text{lower}}$ for this transition energy is 1.2×10^{-33} . In other words, at room temperature, it is close to impossible for thermal energy to result in electronic excitation. At 6000 K, the ratio is still only 0.023. Among other things, this means that if you're trying to get things to light up by heating them, you have to get them a lot hotter than room temperature. More to the point, however, it means that if we want to detect and quantify molecules based on their absorption and/or emission of light, we don't have to worry too much about the molecules getting into excited states due to thermal agitation.

Absorption: Counting the Calories

Since we've been talking about heating water, we'll resurrect the **calorie** (just call it cal), sometimes known as the **small calorie**, which is the amount of heat energy required to raise the temperature of 1 g (or 1 ml) of water by 1 K. There are 4.184 J in a calorie. For those of us who diet, there are no small calories, only large ones; the calories we count are really **kilocalories** (kcal), each of which is equal to 4184 J. This digression serves as more than a coffee break; it gives us the numbers we need to be able to relate energies in joules per photon to the kilocalories per mole commonly used to describe energies of chemical reactions.

There are 6.02×10^{23} molecules in 1 mol, and 4184 J in 1 kcal, so the energy change, in kcal/mol, resulting from the absorption of 6.02×10^{23} photons (a "mole" of photons, officially known as 1 **einstein**) by 1 mol of a substance can be calculated by multiplying the photon energy (in J) by $(6.02 \times 10^{23})/4184$, or 1.44×10^{20} .

The photon energy in J corresponding to an energy change given in kcal/mol can be found by multiplying by the reciprocal figure, 6.95×10^{-21} .

From this, we find that absorption in the microwave oven described above is producing an energy change of only 2.33×10^{-4} kcal/mol; that's low calorie cooking indeed. Absorption of 10.6 μm radiation from the CO_2 laser is good for 2.71 kcal/mol. How do you handle a hungry molecule? Try some visible light; at 633 nm we end up with 45.2 kcal/mol. We all live by converting ADP to ATP and vice versa, which only takes 7.3 kcal/mol; but we all live off plants, which accomplish this by photosynthesis, in which chlorophyll absorbs red (680 nm) light, good for about 42 kcal/mol, and, ever obedient to the laws of thermodynamics, wastes a chunk of it while providing us with our meal tickets.

A Selective Diet

Molecules are finicky eaters; those photon energies or wavelengths at which a given molecule will absorb are determined by the structure of the molecule, according to **selection rules** dictated by quantum mechanics. The electrons we're trying to excite reside in **orbitals**. Any given molecular orbital may contain up to two electrons, but, according to

the **Pauli Exclusion Principle**, the two electrons must have opposite spins. Thus, an electron cannot go from a low-energy orbital ground state to a higher-energy orbital excited state if there is another electron of the same spin already present there.

Molecular symmetry, or the lack thereof, can affect absorption. Because transitions require a precise spatial relationship between the molecular dipole moment and the E vector of the light that is being absorbed, asymmetric or oriented molecules may show **dichroism**, absorbing light of different polarizations differently. Circular and linear dichroism in absorption are analogous to optical rotation and birefringence in scattering.

Speaking of scattering, in the discussion of that phenomenon, I mentioned that most small molecules absorb light in the ultraviolet. How, then, do we make molecules absorb at visible frequencies? Obviously, we have to arrange for electronic transitions to involve smaller changes of energy. One of the more tried and true methods involves making large systems of conjugated double bonds, typically including some atoms other than carbon. The π orbitals in such molecular structures are formed by hybridization of p orbitals from the individual constituent atoms; one consequence of this is that the energy difference between electronic excitation states becomes smaller, shifting absorption to longer wavelengths. Most of the dyes we use, fluorescent or otherwise, contain condensed and/or multiple heterocyclic rings. The people who design dyes have gotten quite good at tailoring their spectral characteristics.

Of course, if a material contained a whole lot of electrons free to run around loose, it would be apt to exhibit strong absorption over a broad range of wavelengths. This type of electronic structure is found in metals (and semiconductors), and accounts for both their broadband absorption and their high reflectivity; very high electric field intensities are produced in the surface layers of molecules of a metal object, resulting in intense scattering. Materials with discrete absorption bands also show changes in scattering characteristics near their absorption bands; this classically was referred to as **anomalous dispersion**.

The Chance of a Lifetime

In addition to being finicky eaters, molecules are fast eaters; absorption associated with electronic transitions occurs in about 1 **femtosecond**, i.e., 10^{-15} s. According to the **Franck-Condon Principle**, this time is too short for any nuclear displacement to occur. What does occur is charge displacement, or a change in the molecular dipole moment. Once absorption has occurred, things take on several aspects of a Roman orgy, because, depending upon how you look at it, molecules either don't keep down what they've absorbed or don't remain in excited states for long. The length of the average **lifetime** for a molecule in the excited state is dependent upon the processes by which the molecule can lose the absorbed energy, a subject we will consider further in our discussion of fluorescence.

The **uncertainty principle** of quantum mechanics tells us, among other things, that the more rapidly the energy of a system is changing, the less precisely we can define its energy. Thus, **excited states with long lifetimes are associated with narrow absorption peaks, whereas excited states with short lifetimes are associated with broad absorption peaks.** We will presently find out that this aspect of the uncertainty principle will limit how much we can do in the way of designing multiple fluorescent probes with minimal spectral overlap.

Spinning a Tale of Degeneracy

Discrete states of a molecule with the same energy content are said to be **degenerate**. In some cases, an external influence, such as an applied electric or magnetic field, can **resolve degeneracy**, producing **splitting**, i.e., the appearance of two or more spectral lines where there was only one before. The production of splitting by an applied magnetic field is known as the **Zeeman effect**. Applied magnetic fields increase the energy change associated with changes in orientation of electron and nuclear spins. Even in strong magnetic fields, however, we're talking about pretty low energies, down in the radio frequency bands between a few dozen and a few hundred MHz. This is the region in which **nuclear magnetic resonance (NMR)** and **electron spin resonance (ESR)** spectroscopy, which measure absorption related to spin orientation changes, are done.

Since the energies involved are so low, there are nearly equal numbers of molecules present in upper and lower states (consider the calculation for microwave absorption on the previous page). Thus, techniques such as NMR, which measure absorption in this energy region, can only detect absorption by a few molecules among many billions, and thus are relatively low in sensitivity. This is why NMR, which has become indispensable as a tool for qualitative analysis and structure determination, is not usable for the detection of very low concentrations of substances.

Facing Extinction: Cross Section and Optical Density

At this point, you are probably thinking that what you really needed to know about absorption was how to use it to measure things, and that all this business about molecules sucking up photons leaves you pretty far from that knowledge. It's closer than you may think. Suppose we have a slab of material, the thickness of which is d cm, containing n molecules/cm³ of some absorbing substance. Moving closer to reality, we could make this slab a cuvette d cm across, filled with a solution of the absorber at a concentration of a mol/dm³ (that's the the official SI unit for moles/liter; we'll use M). We would then have $n = Na/10^3$, where N is Avogadro's number.

Beer's law, also called the **Beer-Lambert law**, relates the intensity of light entering the cuvette, I_0 , and the intensity of light transmitted through the cuvette, I , by the formula

$$\ln(I_0/I) = \sigma nd,$$

where σ is a molecular property of the absorber called its **absorption cross section**. The cross section is expressed as an area (in this case, in cm²); this area is not the actual physical area of the molecule but, rather, the area over which it will act as a barrier to the passage of light. Beer's law can also be stated as

$$\ln(I_0/I) = \alpha Cd,$$

where C is the concentration and α is the **absorption coefficient**, which, like the cross section, reflects the capacity of the absorbing substance to absorb light.

The units here get a little sticky. If we have n in molecules/cm³ and d in cm, we have to have σ in cm² to get things to work out. Since α is often represented in cm²/mol, C must be in mol/cm³ rather than in M (mol/dm³), meaning that $C = a/10^3$. Converting to decimal logarithms gives

$$\log_{10}(I_0/I) = \epsilon ad = A.$$

Here, A is the **absorbance**, or **optical density (O.D.)**, a dimensionless quantity, and ϵ is the **decadic molar extinction coefficient**, represented in units of dm³mol⁻¹cm⁻¹. The preferred SI units for the extinction coefficient, concentration, and thickness are m²/mol, mol/dm³, and mm, but a confusing variety of units are used in the literature, and, if you're not careful, your calculations may go off by powers of ten.

Absorbance or O.D. units are convenient to use because, being logarithmic, they add instead of multiplying. An object with an O.D. of 1.0 absorbs 90% of the incident light. If another object with an O.D. of 1.0 is placed in the path of the 10% of light transmitted by the first object, 10% of this light, or 1% of the light incident on the first object, is transmitted by the second object. Thus, two objects with O.D. 1.0 are equivalent to one object with O.D. 2.0. The additive nature of absorbance units also makes it possible to express the absorbance of a mixture of substances at any given wavelength as the sum of the absorbances of the components of the mixture at that wavelength.

Absorption cross sections and absorption or extinction coefficients are, obviously, functions of wavelength, and are calculated from spectrophotometric measurements of known concentrations of absorbing materials in cuvettes of known path lengths. Note, however, that spectral curves often show the percentage of light absorbed, or the percentage of light transmitted, on a linear scale, rather than absorbance on a logarithmic scale.

The cross section can be explicitly related to the extinction coefficient by combining the several versions of Beer's law. Since $\ln x = 2.303 \log_{10} x$, we get

$$\sigma = 2.303\epsilon a/n = 3.82 \times 10^{-21}\epsilon,$$

which allows us to calculate the cross section from the extinction coefficient and vice versa. We can take an example

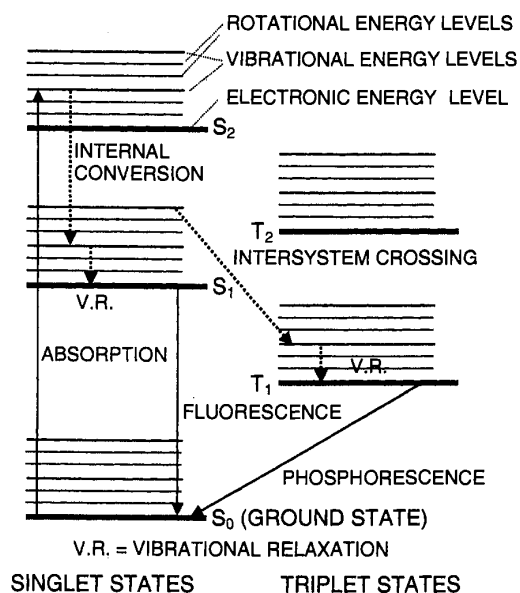


Figure 4-6. The Jablonski diagram of electronic energy levels, or states, and transitions. Solid arrows show radiative and dotted arrows nonradiative transitions.

from Lakowicz²³, who calculates, given that the value of ϵ for anthracene at 253 nm is $160,000 \text{ dm}^3\text{mol}^{-1}\text{cm}^{-1}$, that the absorption cross section at this wavelength is $6.1 \times 10^{-16} \text{ cm}^2$ or, if you prefer, $6.1 \text{ \AA}^2$; recall that $1 \text{ \AA} = 10^{-10} \text{ m} = 10^{-8} \text{ cm}$. The actual area of an anthracene molecule is about $12 \text{ \AA}^2$; thus, at 253 nm, an anthracene molecule absorbs about half the photons that come its way.

Unexciting Times: Emigrating from the Excited States

Absorption of UV or visible light leaves us with molecules in electronic excited states. There are a number of mechanisms by which they can get out of those excited states; they are often illustrated diagrammatically as was first done by Jablonski in 1935¹¹²⁹. A Jablonski diagram is shown in Figure 4-6. At a minimum, absorption involves a transition from the electronic ground state, known as S_0 , to the first electronic excited state, S_1 , but there's usually more to it than that. We have previously (pp. 109-10) discovered the hierarchy of energy states, and noted that vibrational transitions only require 1/100 to 1/10 as much energy as electronic transitions. One consequence of this is that each electronic state has associated with it a set of vibrational states. Immediately following absorption, the molecules involved are likely to be in higher vibrational excited states associated with S_1 or, if they have ordered New York cut photons, with S_2 (some of the molecules absorb enough energy to make it to the second electronic excited state).

In general, the excited molecules "shake off" their excess vibrational energy and, if they are in S_2 , their excess electronic energy, in about 1 picosecond (10^{-12} s) by mechanisms

called **internal conversion** and **vibrational relaxation**, which are **nonradiative transitions**, i.e., changes in energy level that are not accompanied by the emission of photons. The excess energy is lost either to other vibrational modes of the excited molecule or, through collision or radiationless transfer, to other molecules, thus ultimately being converted to heat.

This leaves the excited molecules in the lowest vibrational energy level of the first electronic excited state, S_1 , trying to book passage back to the ground state S_0 , which can be reached by several alternate routes. From our point of view, the route taken will determine whether getting there is all the fun, half the fun, or no fun at all.

What we would generally like the molecules to do is fluoresce, emitting all or some of their remaining excitation energy as photons that we can detect in our instruments. Fluorescence, however, is only one of several mechanisms by which **emission** can occur. **Luminescence** encompasses **fluorescence** and **phosphorescence**, both of which are types of **spontaneous emission**. There is also **stimulated emission**, which is the basis for laser operation. Then, there are nonradiative mechanisms through which the energy can be lost, including **internal conversion**, **resonance energy transfer**, various types of **quenching**, and **bleaching**.

Fluorescence: Working the Stokes Shift

If we get lucky, molecules that have returned, by internal conversion and vibrational relaxation, to the lowest vibrational state of S_1 will get themselves back to the ground state S_0 in relatively short order by purging themselves of photons, a process we have come to know and love as **fluorescence**. In almost all cases, the energy content of the photons emitted will be lower than the energy content of the photons originally absorbed, for two reasons. First, as we have just mentioned, some of the absorbed energy is lost by internal conversion before fluorescence occurs. Second, although fluorescence emission will get the molecules back to S_0 , they will often be in higher vibrational states of S_0 , and will lose a little bit more energy by vibrational relaxation in the 10^{-12} seconds following fluorescence emission, arriving back at the lowest vibrational state of S_0 . Thus, fluorescence emission will occur at longer wavelengths than the absorption that preceded it.

The wavelength difference between the absorption or excitation maximum and the emission maximum is known as the **Stokes shift**, after George Stokes, who described and named fluorescence in the mid-1800's. The term fluorescence was derived from fluorspar, which is fluorescent, by analogy to opalescence, another optical effect named for a mineral⁶⁴². The material Stokes observed was a solution of quinine; the light source was sunlight, the excitation filter a dark blue stained glass window, and the emission filter a glass of white wine. I am not making this up.

If we look at a fluorescence spectrum, such as that shown in Figure 4-7, we generally see that the emission spectrum is shaped like a mirror image of the absorption (or excitation)

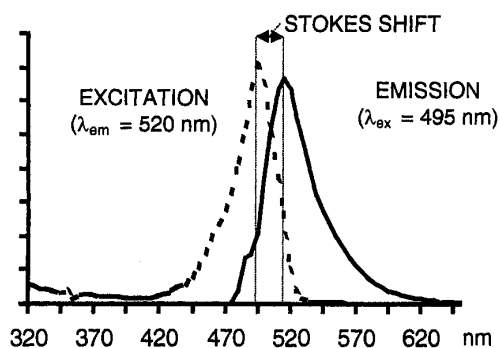


Figure 4-7. Fluorescence spectrum of fluorescein.

spectrum. Also, in general, the shape of the emission spectrum remains the same, irrespective of whether the material is excited at shorter or longer wavelengths within the absorption region. This is so because the energy difference between the longer and shorter wavelength photons is dissipated by nonradiative mechanisms following absorption.

Okay, you might say, but the excitation spectrum overlaps the emission spectrum. If the shape of the emission spectrum remains the same, no matter what the excitation wavelength is, wouldn't that mean that we could get 500 nm emission from 510 nm excitation, seemingly violating the Law of Conservation of Energy? Well, we could get 500 nm emission from 510 nm excitation, if the molecule in question was already in a vibrational excited state when the 510 nm excitation photon arrived. The cost of electronic excitation remains the same, but the molecule itself is coming up with some of the money.

The width and fine structure of the excitation spectrum reflect the energy differences between vibrational states of the electronic excited states; absorption peaks are typically skewed with long tails toward shorter wavelengths, because the transitions with the highest probabilities of occurrence are those to the first few vibrational levels of S_1 . The width and fine structure of the emission spectrum reflect energy differences between the vibrational states of S_0 ; emission spectra usually have long tails toward longer wavelengths, because the most probable transitions in emission are those to the lowest few vibrational states of S_0 .

In fluorescence, exit from the excited state generally obeys first order exponential kinetics. If fluorescence were the only process by which molecules returned to the ground state, the mean lifetime of molecules in the excited state would be $1/k_f$, where k_f is the rate constant for fluorescence emission. This quantity is defined as the **intrinsic lifetime**, τ_0 . In reality, there are other processes competing with fluorescence to deexcite the molecule; the actual **excited state lifetime**, τ , will therefore be shorter. The actual mean excited state lifetimes for most fluorescent molecules are on the order of 10^{-8} s, or 10 ns. Since the decay kinetics are exponential, the mean lifetime is not the median lifetime; the fraction $1/e$, or about 37 percent, rather than half, of the

molecules will remain in the excited state at $t = \tau$, while about 63 percent will have left the excited state by then.

It is often possible to use differences in fluorescence lifetimes between molecules to advantage for analytical purposes. The lanthanide rare earth elements, such as europium and terbium, exhibit atomic fluorescence due to emission from electrons in *f* orbitals; the excited state lifetime of such fluorescence is typically several microseconds. When rare earth chelates are used as fluorescent labels, it is possible to detect their fluorescence in the presence of much higher, but shorter-lived, background fluorescence by using a pulsed source such as a nitrogen laser and electronically gating the fluorescence measurement system so that 50 nanoseconds or so elapse before the measurement is made. This is an example of **time-resolved fluorescence spectroscopy**.

The **quantum efficiency**, or **quantum yield**, of fluorescence, Φ , is equal to the ratio of the number of photons emitted to the number of photons absorbed, and can also be represented as

$$\Phi = \tau/\tau_0 = k_f/(k_f + \sum k_i),$$

where $\sum k_i$ is the sum of rate constants for the competing nonradiative deexcitation processes. The total fluorescence emission obtained from a fluorescent material is the product of the number of photons or amount of light absorbed and the quantum efficiency. If you want to have a lot of light emitted, you've got to get a lot of light absorbed first; thus, good fluorescent probes need to have high extinction coefficients.

Phosphorescence

The S in S_0 , S_1 , etc. stands for **singlet**. In absorption, an electron is raised from a low-energy orbital to a higher-energy orbital, typically leaving behind another electron in the low-energy orbital. According to the Pauli exclusion principle, two electrons in the same orbital must have opposite spins, which are described as paired. A singlet excited state is one in which the electrons in the high- and low-energy orbitals have paired spins; absorption processes which result in formation of singlet excited states have relatively high transition probabilities, and emission leaving both electrons in the low-energy orbital also has a high probability of occurrence. Such transitions are described as **allowed**.

A **triplet** excited state is one in which the electrons in high- and low-energy orbitals have the same spin. Transitions to and from triplet states require changes of spin and therefore have a low probability of occurrence; they are called **forbidden**. The first triplet excited state, T_1 , generally has a lower energy than the first singlet excited state, S_1 . Following absorption, molecules can relax via a nonradiative transition to the T_1 , rather than the S_1 state; this is called **intersystem crossing**. The transition to the ground state S_0 from T_1 is forbidden. This doesn't mean it can't happen; like a lot of other supposedly forbidden things, it does happen, but it takes longer to occur because it has a lower probability

and because the energy must be passed through secret Swiss bank accounts. The associated emission is called **phosphorescence**. Lifetimes for phosphorescence are much longer than for fluorescence, typically milliseconds to seconds, thanks to which we have television and watch dials that glow in the dark. Since the energy difference between T_1 and the ground state S_0 is usually smaller than the energy difference between S_1 and S_0 , phosphorescence typically occurs at longer wavelengths than fluorescence.

Fluorescence Polarization

I have already mentioned that, in order for absorption to occur, the **E** vector of the incident light must be aligned with the dipole moment of the absorbing molecule. If the incident light is linearly polarized, only those molecules that happen to be oriented properly with respect to the plane of polarization will absorb light. Since only those molecules absorb the incident light, only they are capable of fluorescence emission. Like absorption, emission can only occur in a direction determined by the orientation of the molecule.

Absorption occurs so rapidly (10^{-15} s) that the absorbing molecules have no time to move during the process. If fluorescence emission occurred as rapidly as absorption, or if the molecules involved were completely immobilized, the fluorescence emission occurring following excitation by linearly polarized light would be linearly polarized, although not necessarily polarized in the same plane as the exciting light. By now, we know that that isn't the way the world works; fluorescence is going to occur over a period of nanoseconds following absorption, and it's a cinch that at least some of the molecules are going to change their orientations (i.e., rotate) before they emit. This means that some **fluorescence depolarization** will occur. The more motion there is before emission, the more depolarization we can expect.

We can make use of this effect to determine the relative rotational freedom of fluorescent molecules, or the **fluidity** of their microenvironment; or, looking at the other side of the coin, we can determine the extent to which molecular movement is restricted, or the **viscosity** of the microenvironment. This is done by using appropriate polarization optics, which may be as simple as Polaroid filters in different orientations, and making measurements of fluorescence intensities polarized in the planes parallel and perpendicular to the plane of polarization of the excitation. These intensities are, respectively, denoted by $I_{\parallel}$ and $I_{\perp}$.

From the intensities, we can compute either the **fluorescence polarization**, p , as

$$p = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + I_{\perp}),$$

or the **fluorescence emission anisotropy**, r , where

$$r = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2 I_{\perp}).$$

Values of both polarization and anisotropy increase as molecular rotation is increasingly restricted. The use of fluores-

cence polarization and anisotropy measurements to measure rotational diffusion of molecules in membranes and the cytosol will be discussed further in the chapter on parameters and probes.

As a general rule, when we are not trying to measure anisotropy or polarization in our cytometers, we pay little or no attention to the degree or direction of polarization of fluorescence. Most of the time, we get away with it. However, an article published in 2000 by Asbury, Uy, and van den Engh²⁴³⁹ suggests that polarization effects may represent a fair-sized skeleton in our cytometric closet.

Since the light emitted by most lasers used as light sources in cytometry is polarized, both scattered light and fluorescence emission are typically polarized to some degree. This makes the intensity of detected signals more dependent on the angle and direction at which they are detected than would otherwise be the case. Differences from instrument to instrument in optical geometry, and in the polarization response of optical elements such as lenses, dichroics, and filters, may therefore lead to otherwise inexplicable differences in the intensities of signals measured from supposedly identical cells or particles. Further complications may be introduced by the fact that different fluorescent probes exhibit differing degrees of fluorescence polarization, some intrinsic to the molecular structure of the probes, and some dependent on binding to macromolecules and on other environmental characteristics.

The bottom line for most users is that polarization-related differences in the response of different instruments may interfere with the standardization of quantitative fluorescence measurement. The bottom line for those of us who develop and manufacture instruments is that we need to determine the nature and extent of those differences, in hopes of reconciling results from existing systems and improving the design of future systems. A simple solution was suggested by Asbury, Uy, and van den Engh; placing a polarizer at the so-called "**magic angle**" (54.7° for linearly polarized source emission) in the light path of each fluorescence detector removes the dependence of intensity measurements on polarization, with only a modest loss of overall signal intensity.

Stimulated Emission

One of the stranger things photons can do is make more photons just like themselves. I don't mean from nothing; there has to be some energy input to start with, but it's still pretty remarkable. It took Einstein to figure it out. We already know that a photon is likely to be absorbed by a molecule if the energy difference between the molecule's ground and excited states is equal to the energy of the photon. It turns out that just having photons of that energy around also increases the likelihood that molecules already excited will emit identical photons.

That, of course, is the catch. In general, when we're talking electronic excitation, there are a lot fewer excited molecules than molecules in the ground state, as our exercises

with the Boltzmann distribution showed. On the other hand, we also calculated that, for less energetic transitions, we could end up with nearly equal numbers of molecules in the upper and lower energy levels. So, what happens if we put a whole lot of energy and some well-chosen photons into a small volume of molecules?

There are, as it turns out, a number of ways of doing this. One of them is to confine an incandescent ionized gas, or **plasma**, using a magnetic field, to the central portion of a tube placed in a mechanically rigid structure in which mirrors at both ends efficiently reflect light in some spectral region in which components of the plasma emit, with one of the mirrors transmitting a tiny bit of light. The plasma will contain a higher proportion of ions or molecules in an excited state than in a corresponding lower energy state, or a **population inversion**. With time, more and more light will bounce back and forth between the mirrors, and, wonder of wonders, the photons in the light transmitted out one end will all be of the same wavelength, and in the same phase, and going in the same direction. In other words, this light output will be **coherent**. We start with a little light of a given wavelength, and **stimulated emission** gives us more, so we are getting **gain** at that wavelength, or, to recoin a phrase, **Light Amplification by Stimulated Emission of Radiation**. In other words, we've made ourselves a **laser**.

Since actually building working lasers involves substantial amounts of both high tech and black art, I am tempted to say that reproduction of photons is an unnatural act. As it turns out, though, there is at least one known example of a natural laser; there is some gain at carbon dioxide emission lines in the atmosphere of Mars. Assuming we don't put enough laser light through flow cytometers to get the dyes in cells to lase (and I wouldn't swear we don't), there should not be competition between fluorescence and stimulated emission in flow cytometry.

Resonance Energy Transfer

There are a number of ways in which a molecule in an excited state can transfer energy nonradiatively to other molecules; collision, for example, is inelegant but effective. The transfer of energy, however, is more efficient when the energy differences between the S_0 and S_1 states of the donor and acceptor molecules are the same, in which case there is said to be a resonance between them. This condition holds when an emission peak of the donor species and an absorption peak of the acceptor species overlap substantially. Even then, however, significant **resonance energy transfer** can only occur when donor and acceptor molecules are within about 60 Å of one another in space, because, as Förster showed in the 1950's, the amount of energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor. As will be discussed further in the chapter on parameters and probes, this sensitivity to intermolecular distances makes it possible to use measurements of the extent of energy transfer between molecular species as a "spectroscopic ruler" to measure those distances^{315, 2347}.

Energy transfer is **nonradiative**; that is, the donor is not emitting a photon that is then absorbed by the acceptor. Nonradiative transfer of absorbed light energy through one or more phycobiliproteins to chlorophyll is widespread in nature and essential to life; while these molecules are fluorescent to some extent in their natural environment, they are more fluorescent when separated from it. Energy transfer can be used for spectral shifting of fluorescence, since emission from the acceptor can result from absorption by the donor. This trick is commonly employed in scintillation counting, and is, as was noted in Chapter 1, also the basis for tandem conjugate labels.

Quenching, Bleaching, and Photon Saturation

More often than not, cytometry uses fluorescence as its medium of exchange for information; we use fluorescent probes or dyes to measure the amounts of various substances of interest on or in cells. We must therefore be concerned with processes that may, as it were, change the rate of exchange, making the fluorescence signal no longer proportional to the amount of probe or label used. Quenching, bleaching, and photon saturation can all, in some circumstances, interfere with our measurements; in other circumstances, we may be able to use these effects to advantage.

Quenching results when excited molecules relax to the ground state via any nonradiative pathways that provide alternatives to fluorescence. Loss of energy by vibration and collision, by energy transfer, and by intersystem crossing may all account for quenching. Molecular oxygen and paramagnetic molecules and heavy ions such as iodide quench by increasing the probability of intersystem crossing. Polar solvents such as water generally quench fluorescence to some extent because such molecules reorient around excited state dipoles. This may explain the popularity of cold showers as a means of getting out of excited states.

Bleaching occurs when the structures of fluorescent molecules are altered sufficiently to render them nonfluorescent. Bleaching can result directly from the action of light on the fluorescent molecules or from chemical reactions between the fluorescent species and other reactive molecules, such as oxidizing agents present in solution or free radicals produced by photochemical reactions.

In fluorescence excitation, one is shooting at a finite number of targets. We have a certain number of dye molecules, in each of which electrons can be raised to excited states, usually returning to the ground state after a few nanoseconds. However, there is always some chance that absorption of one or more photons will alter molecular electronic structure enough to break a chemical bond, yielding a non-fluorescent product(s); the process is called **photolysis**.

Bleaching can occur in flow cytometers at surprisingly low laser power levels, as was shown by Pinkel et al.¹⁵⁶. These authors set up a dual-beam flow cytometer, with excitation from two separated argon ion laser beams, both at 488 nm, each focused to an elliptical spot approximately 100 µm wide (perpendicular to flow) and 15 µm high (parallel to

flow) at its intersection with the cell stream. The two observation points thus produced were separated in space by 10–100 μm . The power in the second “probe” beam was kept constant at 150 mW; power in the first beam was varied between zero (i.e., the beam was blocked) and 400 mW. This instrument was used to examine sperm cells stained for DNA by the acriflavine Feulgen method; distributions obtained from fluorescence measurements made in the probe beam were compared using different powers in the first beam. If passage through the first beam resulted in bleaching, i.e., destruction of dye molecules, fewer dye molecules would be present by the time the cells reached the probe beam, and the intensity of fluorescence detected in the probe beam would therefore be lower.

A slight decrease in fluorescence intensity was noticeable when the power in the first beam was set at 8 milliwatts (mW); when power was 400 mW, fluorescence intensity measured in the probe beam was decreased by about 75 percent, compared to values obtained with the first beam off. This indicated that most of the dye was destroyed, or at least temporarily converted to a nonfluorescent form, by exposure to argon ion laser emission at 488 nm at power levels commonly used for flow cytometry. Pinkel et al also showed that increasing excitation power above 15 mW in a single beam instrument did not result in a proportional increase in the peak channel number of the observed fluorescence distribution; this provided further evidence of significant bleaching. While most dyes used in flow cytometry are less sensitive to bleaching than is acriflavine, the message is clear; there is always a power level above which “throwing more photons at the problem” may be counterproductive.

My colleagues at Block and I were not at all surprised by these results. Tomas Hirschfeld had concluded, some years before the Pinkel paper came out, that one might get a better signal-to-noise ratio in some kinds of immunofluorescence analysis by taking advantage of bleaching. What we had been trying to do was to detect single viruses in serum isolates, using antibodies tagged with a macromolecule (polyethylenimine) to which several hundred molecules of fluorescein had previously been conjugated.

We encountered two major problems; the serum was highly fluorescent, due to its content of flavins and other naturally fluorescent molecules, and the highly fluoresceinated antibodies weren't fluorescent, because the fluorescein molecules were close enough to quench one another by energy transfer. This was an example of a common phenomenon called **concentration quenching**; one normally runs fluorescence spectra on micromolar or submicromolar concentrations of dyes to avoid it.

It occurred to Tomas that if he had several hundred fluorescein molecules quenching one another, it meant that they couldn't be spending a lot of time in the excited state, i.e., **quenching shortened the excited state lifetime**. He then reasoned that the quenched molecules would be less likely than unquenched molecules to bleach if he threw a lot more photons at them.

The interfering background substances weren't present in enough concentration to be quenched to anything like the extent to which the fluorescein molecules were. Therefore, clobbering the sample with a lot of laser power should rapidly bleach the molecules responsible for the background fluorescence. From that point on, photons would be squeezed out of the fluoresceins at a leisurely pace until they went to their reward. As long as the relative probabilities of fluorescence and bleaching remain the same, one can expect to get a certain number of photons out of a molecule before it gets bleached. Tomas concluded, and found, that he could get the same number of fluorescent photons out of a quenched fluorescein over a long time period than would be obtained from an unquenched fluorescein in a shorter time¹⁰⁸.

“Bleaching out the background” before making a cytometric measurement is analogous to letting the background fluorescence fade out when you are looking at cells under the fluorescence microscope (p. 12). It only works when the fluorescent material you want to measure is quenched to a significantly greater degree than are interfering fluorescent substances in the background. We had planned to use this strategy in the gadget we built to count single viruses⁹⁴, and Mike Hercher and others had done fairly extensive calculations of how much laser power we'd need. It turned out that, with 100 mW of excitation at 488 nm, we'd be able to see a few dozen molecules of fluorescein, and we wouldn't gain anything by increasing the power because we would have zapped every last fluorescein molecule with the 100 mW. This made us curiouiser and curiouiser about why almost everybody else who was building flow cytometers was using big lasers to get a watt or so of excitation power. It seemed that they wouldn't need that much power unless they were throwing away photons right and left everywhere in their instruments; to make a long story short, they were. But there are some justifications for the use of high laser power.

As the intensity of the incident light goes up, the probability of hitting dye molecules with photons increases. Up to a point, this will result in increased fluorescence emission. Once the number of dye molecules in the excited state becomes equal to the number of dye molecules in the ground state, however, a state of **photon saturation** is reached. Further net transition between the ground state and the first excited state is impossible; throwing more photons at the sample will not result in increased fluorescence.

A 1992 study by van den Engh and Farmer¹¹³⁰ added considerably to our understanding of how of saturation and bleaching affect measurements made in flow cytometers. The theoretical section of this paper calculated the average time interval τ_p at which photons hit a dye molecule; this quantity is equal to $1/\sigma I$, where σ is the molecule's absorption cross section, in units of area, and I is the intensity of incident light, in units of photons per unit area per unit time. A typical dye might have an extinction coefficient of $3 \times 10^4 \text{ mol}^{-1}\text{cm}^{-1}$; the formula on p. 111 would convert this to

an absorption cross section of $1.15 \times 10^{-20} \text{ m}^2$. A 1 W, 488 nm laser beam, focused to a 20 μm spot, provides an average illumination intensity, or **photon flux**, of 7.87×10^{27} photons $\text{m}^{-2}\text{s}^{-1}$; at the center of the Gaussian beam, the intensity is twice this. A dye molecule in the center of the beam will thus encounter a photon about every 6 ns. This time τ_p is of the same order of magnitude as the excited state lifetime τ .

Additional fluorescence cannot result from the interaction of photons with dye molecules that are already in the excited state; as van den Engh and Farmer put it, such molecules have an effective cross section of zero. The probability of a photon producing excitation of a molecule is $\tau_p/(\tau + \tau_p)$; this reaches a maximum value when $\tau = \tau_p$, and the excitation rate becomes $1/\tau$. Increasing excitation intensity beyond this point simply decreases the efficiency of excitation, and does not increase the total amount of emitted fluorescence. Van den Engh and Farmer's data also showed that bleaching is almost entirely due to absorption of single photons by molecules in the ground state, rather than to absorption of second photons by molecules in excited states.

A molecule is unlikely to go through an infinite number of cycles of excitation and emission; the probability of its undergoing bleaching while in an excited state is constant, and, therefore, as the number of cycles increases, the likelihood that the molecule will be destroyed increases. A molecule of Hoechst 33258 bound to DNA was found to have an effective life of about 100 excitation-emission cycles, whereas a molecule of DNA-bound propidium iodide would last for just over 200 cycles. The critical point here is that these figures are independent of excitation intensity. You get the same number of photons from a dye molecule whether you excite it at low intensity for a long time or high intensity for a short time, and you get the maximum number of photons out only by bleaching all of the dye molecules.

For any given observation period, there is a lower range of intensities at which neither bleaching nor saturation are significant; in this range, usually corresponding to laser power levels of no more than a few tens of milliwatts, fluorescence emission increases linearly with excitation intensity. Above this range, the relative increase in emission is less than the relative increase in intensity, and above the intensity at which $\tau = \tau_p$, there is no relative increase in emission. Thus, at least for dyes of the type studied by van den Engh and Farmer (i.e., DNA stains), there is some advantage to using very high excitation powers, i.e., hundreds of milliwatts.

Since a typical laser beam has a Gaussian profile, focusing the beam to a 20 μm spot produces a large variation in illumination intensity over the width of even a small (e.g., 2 μm) core stream. At low laser powers, this will result in unacceptably large CV's for fluorescence measurements. However, if the power is high enough, the intensity variation over the width of the core will produce only small variations in fluorescence emission, with minimal effects on the measurement CV; effects of fluctuations in laser output that would otherwise decrease precision will be minimized, as well.

One downside to measurements at high excitation intensity relates to the photochemical effects of dye bleaching on the biological and/or biochemical properties of the sample. Kissane et al.²⁵⁶ reported that UV laser powers above 100 mW substantially decreased viability of sorted cells stained with Hoechst dyes. It is likely that this is due to formation of photoadducts between the dye and DNA that are not repairable by normal mechanisms, and it seems probable that chromosomal DNA, or sperm, sorted after staining with Hoechst dye and exposure to very high laser powers, might be similarly damaged.

There are also other situations in which the use of high power levels may be counterproductive. Doornbos, de Grooth, and Greve²⁴⁴⁰ carried out theoretical and experimental studies of bleaching and saturation in a flow cytometer, considering the behavior of fluorescein (FL), phycoerythrin (PE), and allophycocyanin (APC), and using a wide range of both laser powers and observation times. While APC fluorescence could be accurately fit to a model in which exit from the excited singlet state could only proceed via fluorescence emission, nonradiative decay, bleaching, and conversion to the triplet state, the fluorescence of FL and PE could not, indicating the involvement of other processes. For APC, it was found that the best signal-to-noise ratio for detection could be obtained with relatively low excitation power (30 mW), and a much longer observation time (1 ms) than is commonly used in conventional flow cytometers. Under these conditions, Doornbos, de Grooth, and Greve calculated it should be possible to detect a single molecule of cell-associated APC. In fact, a scanning system built by their group²³⁸³, using an 8 mW, 635 nm laser diode for excitation, with an observation time of approximately 1 ms, resolves beads bearing small numbers of APC molecules better than do current commercial flow cytometers. Long observation times, and very slow flow rates, have also been used to advantage in flow cytometers specially designed for observation of DNA molecules^{660,1144,2327,2333-4}, viruses⁹⁴, and bacteria²³²⁶.

Here's a riddle: when can a lifetime be a dead time? As was noted on p. 111, the uncertainty principle results in an inverse relationship between the widths of absorption peaks and excited state lifetimes. For most fluorophores, the excitation spectrum and the absorption spectrum are the same, or very nearly so, and emission spectra tend to be mirror images of excitation spectra. Thus, one would expect a fluorophore with a narrow emission peak to have a long excited state lifetime. The emission peaks of the lanthanide elements mentioned above are quite narrow, and, as would be predicted, their excited state lifetimes are very long, meaning microseconds. You would therefore not want to use a lanthanide label in a high-speed sorter, in which a cell spends less than a microsecond in the excitation beam; you wouldn't be able to excite each label molecule more than once. Well, you'd also have to look downstream some distance to detect the emission, and make the measurements over a long time period (it has been done¹³⁸⁵⁻⁶), but that's another problem.

But let's say we have two dyes, one with a 3 ns lifetime and one with a 9 ns lifetime, and the excitation intensity is high enough so that a photon capable of exciting either comes by every 3 ns. You can get three excitation-emission cycles out of the dye with the 3 ns lifetime for every one you get out of the dye with the 9 ns lifetime. In other words, the 3 ns dye is just at photon saturation; the 9 ns dye is well past it, and its lifetime, i.e., the time it spends in the excited state, has become a "dead time" during which it cannot be excited, no matter how many eligible photons pay it court.

This may explain why some probes with short lifetimes yield higher intensity signals from flow cytometers than do other probes, with longer lifetimes that, according to spectrofluorimetric measurements, have higher absorption at the excitation wavelength and higher quantum efficiency. The average spectrofluorometer doesn't get its specimens anywhere near photon saturation, meaning that the time between excitations is very long, and the time molecules spend in the excited state is negligible, whereas photon saturation is not infrequently approached or reached in flow cytometers.

I was thinking that a cover blurb describing this book as being about excited states, vibrational relaxation, and stimulated emission might boost sales; "no emission without a quantum" seemed like a useful slogan. However, good taste won out, at least for the cover. Before I start to improve my image, I should cover a few odds and ends to complete this discussion of the interaction of light and matter complete.

Quantum Flotsam and Jetsam

Inelastic Scattering and Doppler Measurements

When we talked about scattering, I described what is more precisely defined as **elastic scattering**, which results in emission of photons at the same energy (or frequency, or wavelength) as the incident light. **Inelastic scattering** occurs when the scattering object is moving; the **Doppler shift** results in scattering at a higher frequency than that of the incident light if the object is moving toward the source and in scattering at a lower frequency if the object is moving away. The faster the object is moving, the larger the frequency difference between incident and scattered light. Since relatively low energies are involved in molecular motion, the frequency differences between the incident and the scattered radiation are relatively small. Inelastic scattering of microwave radiation has been widely and profitably employed by law enforcement agencies. Until the advent of lasers, it was difficult to use light to measure particle velocities, because small frequency differences were hard to detect. Now, **laser Doppler velocimetry** can be used for such purposes, and the cops can tell whether your particle is going over 65 before you can take any countermeasures.

Raman Scattering

There is a small probability that a molecule will undergo a vibrational transition at precisely the time at which scattering occurs, resulting in the emission of a photon differing in

energy from the energy of the incident photon by the amount of energy involved in the vibrational transition. This is **Raman scattering**. The vibrational event involved in Raman scattering can be either a transition to a higher vibrational level or a transition to a lower level. In the former case, which is the likelier of two rather improbable events, the Raman emission is at a wavelength longer than that of the incident light, and is described as **Stokes Raman emission**. In the latter case, the Raman emission is at a wavelength shorter than that of the incident light, and is described as **Anti-Stokes Raman emission**. Nobody is getting something for nothing in either case. Until lasers came along, Raman spectroscopy was next to impossible, because the intensity of Raman scattering is only about 1/1000 the intensity of Rayleigh scattering. It is now fairly easy to do, and can provide useful information about molecular vibrations. Stokes Raman emission can also be a significant source of interference with some flow cytometric fluorescence measurements, notably that of phycoerythrin fluorescence, since the Raman emission from water illuminated at 488 nm is at about 590 nm, and at least some filters used for phycoerythrin measurement transmit this wavelength.

Nonlinear Optics and Harmonic Generation

At the high radiant flux densities achievable with lasers, electric field intensities are extremely high. Many physical effects are described by formulae containing terms that include higher powers of field intensity as well as the first power; the higher-order terms are usually negligible and what we observe are those effects that depend linearly on field intensity. Lasers have facilitated the study and application of **nonlinear optics**. One common application is in **harmonic generation**, in which nonlinear effects in crystals result in generation of light at two or more times the frequency of the incident light. Second harmonic generation, or **frequency doubling** of the 1064 nm YAG (yttrium aluminum garnet) laser line, for example, produces 532 nm, while third harmonic generation, or frequency tripling, produces 355 nm. Frequency-doubled YAG lasers have been used as sources in cytometry, and may be more widely used in the future as they become competitive in price with argon lasers. In the long run, solid-state lasers doubled from the 980 nm range to the 490 nm range will probably replace argon lasers in most fluorescence flow cytometers; Coherent introduced such a solid-state laser in 2001²⁴⁴.

Two-Photon and Multiphoton Excitation

At very high photon fluxes, it is possible to use two or more low-energy photons to induce a transition to an excited state that would normally require a single photon with two or more times the energy content. Denk, Strickler, and Webb¹¹³¹ accomplished this in the context of scanning fluorescence microscopy, using a tightly focused pulsed red laser beam to excite UV-absorbing fluorophores. Fluorescence emission increases quadratically with excitation intensity; since the depth of focus of the excitation beam is very small,

the technique allows detection of fluorescence from a thin "slice" of the specimen. There are now at least a hundred multiphoton confocal microscopes in operation; they typically use very expensive pulsed Ti-sapphire lasers for excitation, although it has been claimed that there are cheaper ways to play the game. Multiphoton excitation for flow cytometry poses problems because of the small size of the excitation spot, which makes it difficult to get signals from more than a small region of a particle the size of a typical cell during transit through the instrument²⁴⁴².

4.2 OPTICAL SYSTEMS

Having sated our voyeuristic curiosities about the private lives of photons, we can now turn our attention to the building blocks of both the peeping Tom's and the analytical cytologist's hardware, i.e., the lenses and other elements that make up **optical systems**. Optical systems have only one basic function, which is diverting light from where it wants to go to where the user wants it to go. In order to learn how this is done, we must first consider where light wants to go to start with. This gets us away from quanta and back into the domain of geometric optics.

Light Propagation and Vergence

We usually consider light as originating from a mythical object called a **point source**, from which we imagine **rays** emanating in all directions, i.e., over a solid angle of 4π steradians. Rays go in straight lines unless reflected or refracted. Reflection is simple; the angle of incidence equals the angle of reflection. Refraction involves an important optical property of the material, i.e., the **index of refraction**. **Snell's law** and the indices of refraction of two materials allow us to determine the angle at which a ray will be "bent" at the interface between them, and provide much of what we need to model the behavior of optical systems (Fig. 4-5, p. 106, and pp. 107-8).

Rays from a point source get farther apart as they go along; they are **divergent**, or have **negative vergence**. Parallel rays have zero vergence; light consisting of parallel rays is said to be **collimated**. Rays that **converge**, or have **positive vergence**, get closer together as they go along. Convergent and divergent light in three-dimensional space can also be thought of as **waves** with spherical **wavefronts**; imagine something like an ice cream cone, with the tip of the cone being the point of origin or convergence of the rays and the hemispherical surface of the ice cream representing the wavefront. The wavefronts in collimated light are planar.

At a distance D from a point of origin (a point source; D here is, by convention, negative) or a point of convergence (in this case D is, by convention, positive), the **reduced vergence**, $V = n/D$. Vergence is measured in **diopters**, where 1 diopter = 1 m^{-1} .

If a ray travels a distance L through a medium with a refractive index n , we say that it traverses an **optical path length** $S = Ln$. The optical path length reflects the distance light would have traveled in a vacuum in the same time it

took to go the distance L in the medium. If many rays from one point travel the same optical path length to end up at another point, they will have the same phase (the peaks of their waves will coincide) and they will reinforce in such a way as to form a **real image**.

Image Formation by Optical Systems: Magnification

Images can be formed by mirrors, but we are concerned here with lenses. Lenses are pieces of material (usually glass, plastic, or quartz) with a shape such that all rays traversing them reach a distant point having traveled the same optical path. In Figure 4-8, that point is at infinity. The rays from the point source all emerge parallel to each other, so we say they are **collimated**. The distance from the point source to the lens is the **focal length**, f , of the lens.

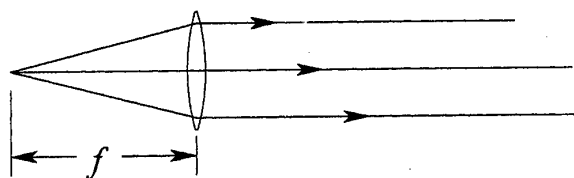


Figure 4-8 Light from a point source at the focal length of a lens is collimated. This defines the focal length, f .

Light rays are entirely reversible, so the arrow heads on the rays of Figure 4-8 could be reversed, bringing light from, say, a distant star to a focus at the focal length f of the lens. That's a lot easier way to measure a focal length than going a long way away to see if the light is really collimated. The best version of a real point source is a distant star, and since it is very far away, all the rays from it that go through the lens are parallel. That means we can evaluate our lens by looking at the focus to see if the rays all converge on one point, forming an image of the distant point source. If they do, there will be little rings around the point image, due to the wave nature of light. Those rings are **diffraction rings**, and a lens that can show them is said to be **diffraction limited** (as opposed to **aberration limited**).

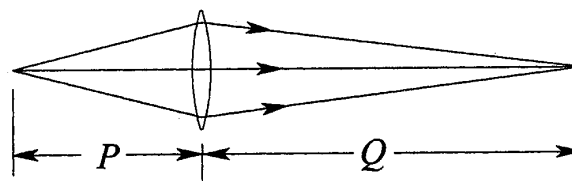


Figure 4-9. Rays from a point source can be focused to an image point.

In Figure 4-9, we see the formation of an image point that is not at infinity. If we put the point source, which we

now refer to as the **object**, at a distance P , greater than the focal length f , from the lens, an image of the object will be formed at the distance Q on the other side of the lens; the distances P and Q and f are related by the **Lens Formula**,

$$1/P + 1/Q = 1/f .$$

Real point sources are hard to come by, and what we really want to make images of are things that are small, but bigger than points. So we use the lens of Figure 4-9 to make a real image from point images of multiple points in a real object, as shown in Figure 4-10 below.

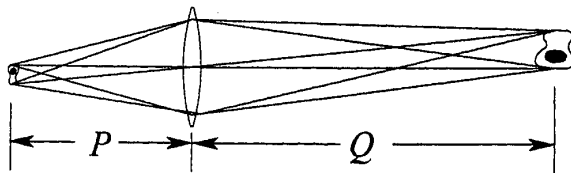


Figure 4-10. Rays from many points of an object add up to make an image.

We generally use lenses to make an image larger or smaller than the object. The **transverse**, or **lateral**, **magnification** of the lens shown in Figure 4-10 is M , where $M = P/Q$. But look closely at the object and the image in the figure, and you will see that their proportions are not the same; the ratio of width to height is larger for the image than for the object. This change of proportions has nothing to do with the fact that the image is inverted; it arises because the **axial magnification**, M_{axial} , of the lens is not the same as the lateral magnification, M ; in fact, $M_{\text{axial}} = M^2$.

Lens Types and Lens Aberrations

The lens in Figures 4-8 to 4-10 is a **simple lens**, or **singlet**, made of a single material. Both surfaces are curved outward, or **convex**, making the lens a **bi-convex** or **double convex** lens. A lens with one flat surface and one surface curved outward is called **plano-convex**. Bi-convex and plano-convex lenses converge light, and are termed **positive** lenses. A lens with both surfaces curved inward is said to be **bi-concave** or **double concave**; a **plano-concave** lens has one flat side and one curved inward. Bi-concave and plano-concave lenses diverge light, and are termed **negative** lenses.

A **concavo-convex** lens, with greater curvature on the convex than on the concave side, converges light and is also known as a **converging**, or **positive**, **meniscus lens**. A **convexo-concave** lens, with greater curvature on the concave than on the convex side, diverges light and is therefore called a **diverging**, or **negative**, **meniscus lens**.

If the object is at some distance between f and $2f$ from a convex lens, a magnified, real, inverted image is formed at a distance between $2f$ and infinity. If the object is at $2f$, the image, again real and inverted, is at $2f$ and the same size as

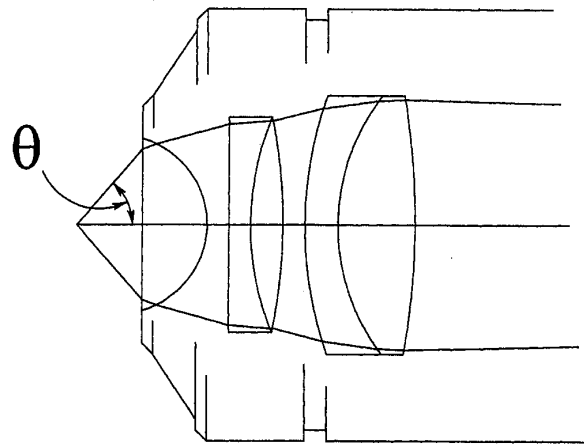


Figure 4-11. Elements of a typical microscope lens, showing the half angle θ that defines the acceptance cone and the numerical aperture (N.A.) of the lens.

the object. If the object is between $2f$ and infinity, a mini-fied, real, inverted image is formed between f and $2f$.

When the object is at a distance less than f , a magnified, erect, **virtual image** appears. When real images are formed, light diverging from the object converges in the image plane; a real image can be seen on a screen placed in the image plane. When a virtual image is formed, light from the object appears to be diverging from the image plane, although neither the object nor a real image is located there. The eye contains the necessary optics to convert virtual images to real retinal images, and thus to make use of simple convex lenses as magnifying glasses and of mirrors for a variety of purposes. However, the detectors in our instruments need to be presented with real images. Simple lenses do not make the best images, due to the presence of **aberrations**.

Most lenses are **spherical**, that is, their curved surfaces are the surfaces of portions of a sphere. **Spherical aberration** results from rays passing through the outer portions of a lens coming to a focus at a different point from rays passing through the central portion. **Chromatic aberration** is a consequence of dispersion (p. 108); since the refractive index of a material is higher at shorter wavelengths, rays of light of different wavelengths come to foci at different distances from the lens. Other aberrations include **astigmatism**, resulting in different lateral magnifications in the vertical and horizontal directions perpendicular to the lens axis, **curvature of field**, resulting in the image of a rectangular object taking on a "barrel" or "hourglass" shape, and **coma**, resulting in images with comet-like "tails." Spherical aberration may be corrected by using **aspheric** rather than spherical surfaces in lenses. Although it is difficult for lens grinders to produce aspheric elements one at a time, it is simple and inexpensive to mold aspheric lenses once a prototype is

made. Chromatic aberration is corrected by making lenses incorporating two or more elements made of materials with different dispersion characteristics. The simplest such lenses incorporate two elements, and are called **achromatic doublets**, or just **achromats**; they bring blue and red light to a focus at the same point. **Apochromatic** lenses have at least three elements, and focus blue, green, and red light to the same point. The **planapochromatic** lenses used in good microscopes also incorporate elements to correct spherical aberration, and to correct for curvature of field, providing an image that is in focus over most of the field of view, or flat, whence the "plan" in planapochromatic.

Numerical Aperture (N.A.) and Lens Performance

When a lens needs to collect rays over a very wide angle, multiple elements are needed to achieve diffraction-limited performance. Figure 4-11 shows a cross section of a typical microscope lens. The **half angle** θ shown in the figure is a measure of the **acceptance cone** of the lens. The **numerical aperture (N.A.)** of the lens is given by $N.A. = n \sin \theta$, where n is the refractive index of the medium between the subject and the front element of the lens. The N.A. of a lens is important in determining its resolution. The numerical aperture is important also because it measures the **resolution** of the lens; two self-luminous points can be resolved if they are a distance Δx apart, where $\Delta x = 0.61\lambda/(N.A.)$, with λ representing the wavelength of the illuminating light in a vacuum. The wavelength in vacuum is specified here, because, as was noted on p. 107, the wavelength of light in material media is shorter than the wavelength in vacuum; in a medium of index n , $\lambda_{med} = \lambda_{vac}/n$. One consequence of this shortening is that cells in water ($n = 1.33$) can be resolved almost exactly 1.33 times better than cells in air ($n = 1.0003$). The use of **immersion lenses**, with water or an appropriate oil or gel with an index higher than that of air, increases both resolution and light collection.

Even the most exquisitely engineered microscope lenses impose some restraints on the user. For instance, they are designed to minimize aberrations when used at specified distances. Usually that means that Q , which, in this context, is called the **tube length**, is on the order of 150 mm; in many newer lenses, however, $Q = \text{infinity}$. Departure from the specified **conjugates**, i.e., the tube length and the **working distance**, which is the intended distance of the object from the front element of the lens, degrades the image quality. Also, most high-N.A. microscope objective lenses have large magnifications at their specified conjugates. That may mean that the axial magnification is so large that the objects in the field of view must all be in the same plane to be in focus. To meet the needs of such application areas as scanning microscopy, in which it is desirable to have a relatively large depth of focus, objective lenses with low magnification but relatively high N.A. have become available.

To be completely frank, though it is essential to have lenses with high resolution if you are doing confocal microscopy, or anything fancy in the way of scanning or image

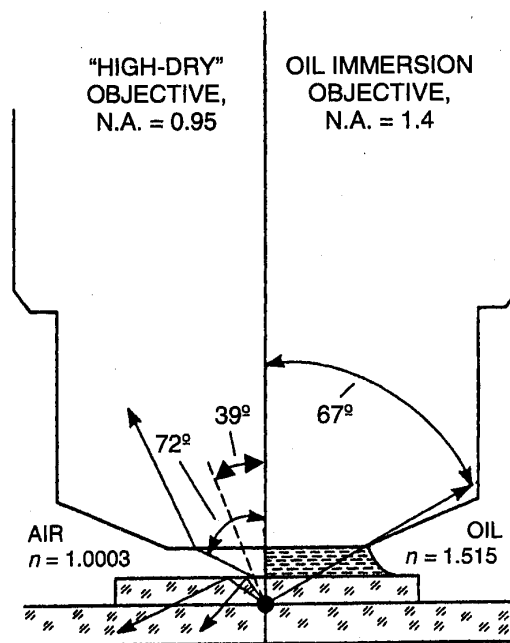


Figure 4-12. Showing the effect of N.A. on light collection. Modified from Murphy DB: *Fundamentals of Light Microscopy and Electronic Imaging*²⁴⁰⁷, Copyright 2001 by Wiley-Liss, Inc. Used by permission.

analysis, or even just trying to look at a blood smear by eye, neither flow cytometry nor low-resolution scanning laser cytometry requires high-resolution optics. What we do need are optics that collect as much light as possible. That means a large acceptance cone, i.e., high N.A. However, there are a lot of high-N.A. lenses that have been optimized for high resolution in transmitted light microscopy by putting in a lot of lens elements, and, even with antireflection coatings on the elements, there is likely to be enough light lost at all the interfaces to reduce the light transmission of such lenses. The effects of N.A. on light gathering power can be appreciated from Figure 4-12.

The figure, modified from one in Murphy's admirable *Fundamentals of Light Microscopy and Electronic Imaging*²⁴⁰⁷, shows the acceptance angles of a "high-dry" objective (N.A. = 0.95) and an oil-immersion objective (N.A. = 1.4); it is split to show the high-dry objective on the left and the immersion objective on the right. Rays of light defining the acceptance cone are shown coming from a "cell" depicted as a black dot between a slide and cover slip, with the line defining the split between the two objectives passing through the center of the cell and perpendicular to the slide, cover slip, and front surface of the lenses. We will assume that the cell is mounted in a medium with the same index as both the glass in these elements and the immersion oil, $n = 1.515$. We will first examine the high-dry objective.

Since the medium closest to the front element of this lens is air, with $n = 1.0003$, N.A. could not be any higher

than this value. In fact, about the best one can do with a high-dry lens is N.A. 0.95. Let us consider the refraction of light coming from the cell at the interface between the cover slip and the layer of air between the cover slip and the front element of the lens. Snell's law tells us that light rays leaving the cell at angles larger than 41° to the perpendicular will not emerge from the cover slip at all; they will either be transmitted along the surface of the cover slip by **total internal reflection**, or be reflected back from the cover slip toward the slide. In practice, the half angle of the largest cone of light that will make it from the cell into the objective is 39° ; this is the angle between the dotted line and the perpendicular through the cell. However, a ray of light coming from the cell at 39° and hitting the outer surface of the cover slip will be refracted according to Snell's law and emerge at an angle of 72° ($1.515 \sin 39^\circ \approx 1.0003 \sin 72^\circ \approx 0.95$, which is the N.A. of the lens). This is interesting, because it shows us that the quantity $n \sin \theta$, which defines N.A., remains the same from one medium to another. However, what does not remain the same is the half angle. We are trying to get light out of a cell and into our lens, and the biggest cone of light we can capture coming out of the cell has a half angle of 39° . So what fraction of the total amount of the light coming out of the cell does that cone represent?

Assuming that we don't have any anisotropy or other directional effects due to polarization, etc., the light should be coming out of the cell uniformly in all directions, i.e., over a solid angle of 4π steradians. We need to calculate the solid angle subtended by a 39° cone. It just so happens that the solid angle, in steradians, subtended by a cone with half angle θ is $2\pi(1 - \cos \theta)$; a 39° cone therefore subtends a solid angle of 1.40 steradians. Since 4π steradians = 12.57 steradians, we can only collect about 11 percent of the light from the cell using the N. A. 0.95 objective. Now let's turn our attention to the oil immersion objective on the right.

On this side, we have immersion oil, matching the index of refraction of the slide, mounting medium, cover slip, and front element of the lens, in between the cover slip and the lens. There is therefore no change in refractive index between the cell and the lens, so the half angle of the cone of light collected from the cell remains at 67° all the way along. This subtends a solid angle of 3.83 steradians; the oil immersion lens could, therefore, collect just over 30 percent of the light coming from the cell, or almost three times as much as the high-dry lens.

It is generally stated that the light gathering power of lenses increases as the square of N.A.; according to this formula, the oil immersion lens should collect 2.17 times as much as the high-dry lens. Of course, if we were using the high-dry lens to look at a cell on the surface of a slide, without the refraction at the surface of the cover slip, we would be able to collect light over a somewhat larger half angle. However, since, in flow cytometry, we are almost always looking at cells in either a stream of water in air or a stream of water in a quartz cuvette, we can expect to have the accep-

tance cones of lenses used without a coupling medium, such as immersion oil or a gel, restricted somewhat. When you can't afford to throw photons away, it pays to use the most efficient light collection system available, and the manufacturers of cytometric instruments have increasingly taken this lesson to heart.

Gradient Index, Fresnel, and Cylindrical Lenses

In a convex spherical or aspheric lens, the optical path length is made equal for rays passing through the center and edges of the lens by making the lens thicker at the center than at the edges. In a **gradient index**, or "**grin**" lens (the trade name "Selfoc" often used to describe such lenses properly refers only to those made by Nippon Sheet Glass), the thickness of the lens is constant, but the refractive index changes from the center to the outside. Such lenses can be made by diffusing various chemicals, which will change the index, into a cylinder of glass from the outside, and then heating the glass and drawing it out until a desired smaller diameter is reached. Small slices of the drawn material can then be cut and the ends polished, allowing large numbers of small lenses to be produced without the need for grinding curved surfaces. The process is virtually identical to that used in the production of fiber optics, and, in fact, grin lenses less than a millimeter in diameter and only a few millimeters long are often used in fiber optic systems.

A **Fresnel lens** is a flat surface with concentric trapezoidal grooves cut or, more commonly, molded into one surface; the angles of the grooves vary with the distance from the center, and, thus, rays entering the lens at different distances from the center are reflected and refracted at different angles. A Fresnel lens superficially resembles a grating, and you might guess it worked by diffraction, but it's all straight geometric optics. Just to confuse matters, there is also something called a Fresnel zone plate, which can focus light, but which does work by diffraction. Fresnel lenses with very high light gathering power can be made cheaply; they are most likely to be seen in overhead projectors, where they are used as condenser lenses.

Cylindrical lenses converge (if they are convex) or diverge (if they are concave) light in only one plane. Whereas a convex spherical lens will focus a circular collimated beam to a point, a convex cylindrical lens will focus the same beam to a line. Crossed convex cylindrical lenses with different focal lengths are typically used in the illumination optics of flow cytometers to focus a laser beam with a circular cross section to an elliptical spot. In a stream-in-air flow system such as is found in most high-speed cell sorters, the stream itself acts as a relatively strong cylindrical lens. **Toric lenses** have cylindrical surfaces with different curvatures in two perpendicular planes, each of which includes the optical axis. A simple toric lens is a piece of glass or plastic with a spherical surface on one face and a cylindrical surface on the other. Opticians use toric lenses to correct the astigmatism that results from the eye's own lens being somewhat toric.

The Helmholtz Invariant and Throughput

We can do some remarkable things with optics, but we can't beat the second law of thermodynamics and the law of conservation of energy. There are a few concepts that can help us keep out of trouble in this regard, and one of them is **throughput**. Figure 4-13 will get us started with this.

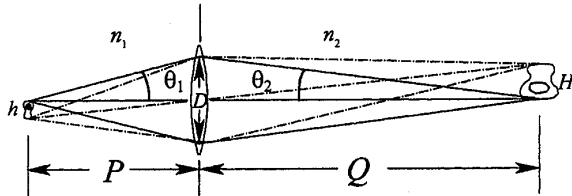


Fig 4-13. Multiple views of magnification.

Figure 4-13 largely reproduces Figure 4-10 (p. 120), which introduced us to magnification. We have provided for media of different indices, n_1 and n_2 , on the object and image sides of the lens. We already know that the lateral magnification of the lens in the figure is M , and that $M = Q/P$. There are two similar triangles with apices at the center of the lens, one including perpendicular sides with the dimensions h (the height of the object) and P , and the other including perpendicular sides with the dimensions H (the height of the image) and Q . Because these triangles are similar, $H/h = Q/P = M$; this is the **magnification equation**. We can also see that $h/P = H/Q$, and, multiplying both sides of this equation by the lens diameter, D , we get $hD/P = HD/Q$.

The **Abbe sine condition**, derivable from geometric optics and also from thermodynamics²⁴⁴⁵, tells us that

$$n_1 h \sin \theta_1 = n_2 H \sin \theta_2.$$

We already know that the magnification, M , is equal to H/h ; rearranging the terms in the equation thus tells us that

$$H/h = n_1 \sin \theta_1 / n_2 \sin \theta_2 = M.$$

But $n_1 \sin \theta_1$ and $n_2 \sin \theta_2$, respectively, are the numerical apertures (N.A.'s) of the lens from the object side and the image side, respectively. So we can express magnification not only in terms of the ratios of image size (height) to object height, H/h , and image distance to object distance, Q/P , but also in terms of the ratio of the N.A. for light collection and the N.A. for imaging. And we see that, for any given optical system, the product of the linear dimension of the object or image and the N.A. on the appropriate side of the system remains constant, or invariant. In the two-dimensional model we have just considered, the constant $h n \sin \theta$ is sometimes referred to as the **Helmholtz invariant** and sometimes as the **Lagrange** or **optical invariant**.

We can find a similar invariant quantity for the three-dimensional case. We consider an object, or source, of area

A_1 in a medium of index n_1 , with the lens collecting a cone of light in a cone with a half angle θ_1 . The lens forms an image of area A_2 in a medium of index n_2 , and the half angle of the cone on the image side is θ_2 . We can actually use Figure 4-13 to play this game if we just assume that the areas of the object (with height h) and the image (with height H) in the figure are A_1 and A_2 . The solid angles, Ω_1 and Ω_2 , subtended by the cones with half angles θ_1 and θ_2 can be calculated from the formula $\Omega = 2\pi(1 - \cos \theta)$. And we end up with

$$n_1^2 A_1 \Omega_1 = n_2^2 A_2 \Omega_2 = \Theta.$$

The invariant quantity here, Θ , is variously known as the **optical extent**, **étendue**, or **throughput** and, as the last of these terms suggests, it tells you how much light can be transmitted through the system.

It is not the case that throughput can't be increased; we do it all the time when we switch from the high-dry to the oil immersion lens while using a microscope. However, once you've picked your optical components, you've defined the throughput of your system, and, while you can lose light, you can't get more than there was in the first place. The bottom line on all of this is that you get the most light into an optical system by collecting over the largest possible angle, or through the largest possible aperture. We'll get to a practical application of throughput a little further on, when we talk about light sources.

Photons in Lenses: See How They Run

To return to the peregrinations of photons for a moment, we might consider how lenses get photons to alter course. Let's keep it simple and think of photons leaving an object at a distance $2f$ from our simple convex lens. If the lens weren't there, they'd just go off in all directions. However, the formulation of quantum electrodynamics is that photons have the highest probability of going along paths from which any given deviation results in the smallest possible change in the time taken to traverse the path. A photon headed along the optical axis would normally get to the image plane a lot faster than a photon headed off toward the rim of the lens. That's why we make a convex lens convex. The lens is made of a material of refractive index greater than that of air, and light travels more slowly in such a material than in air. We make the center of the lens thickest, and make it thinner toward the rim, so that getting from the object plane to the image plane takes about the same time regardless of the angle at which the photon originally started.

If we really wanted to get the photons synchronized, we'd have to shape the lens so that its curvature was not spherical; in practice, spherical surfaces are easier to grind, so we usually put up with the **spherical aberration** that results from the lens not being quite the right shape. For some applications, molded or ground **aspheric** lenses are preferable.

The refractive index n for air is 1.0003, and $\sin \theta \leq 1$; therefore, the highest value of N.A. that could theoretically be achieved for an optical system in air would be 1.0003.

Higher N.A., and thus increased light gathering capacity, can, as Figure 4-12 shows, be achieved by filling the space between the optical system and the object with a medium of higher refractive index, such as **immersion oil**; the system must, of course, be designed with this in mind.

Aperture and Field Stops; The f Number

If you are familiar with photography, or, in this day and age, with video, you may recall that the **f number** ($f/\#$) of a lens indicates the light gathering capacity of the lens. The iris diaphragm in a camera lens is an **aperture stop**; that is, it limits the cross-section and solid angle through which light can be collected. The f number is the ratio of focal length to aperture stop diameter; for small values of x ,

$$f/\# \approx 1/[2 \times (\text{N.A.})].$$

Unlike an aperture stop, a **field stop** limits the field of view, but does not affect the angle over which light is collected. Field stops are typically located in the image plane, and some of us (*mea culpa*) refer to them as image plane apertures. Field stops, usually in the form of "pinholes," are frequently used in flow cytometers to limit the area from which light can reach the detectors to the region immediately surrounding the observation point.

Depth of Field and Focus and Resolution of Lenses

It seems logical that as an optical system deals with light collected over an increasing range of angles, it becomes harder to get the light precisely where we want it. Thus, systems with high light gathering power should have relatively small depths of field and focus, while decreasing the aperture should increase depth of field and focus at the expense of getting less light through the system. This is exactly the way things work; you probably know that stopping down a camera lens increases depth of field, and you may remember that cheap cameras with $f/11$ lenses don't need a focus adjustment. You may also have been disappointed by the poor visual results you got using a high N.A. microscope lens because its depth of field and depth of focus were so small.

Depth of field, Δ , denotes the longitudinal range of object distances over which a "sharp" image of an object is obtained at a fixed point in the image plane, where the sharpness criterion is defined by the acceptable diameter of the "blur circle" in the image of an axial point object. For typical microscope objectives working at a wavelength λ , Δ is given by

$$\Delta = \lambda[(n^2 - (\text{N.A.})^2)^{1/2} / (\text{N.A.})^2].$$

Depth of focus, Δx , denotes the longitudinal range of distances in the image plane over which the image of an object at a fixed distance remains sharp; the relation between depth of field and depth of focus is expressed by $\Delta x = M\Delta$.

To determine the **limit of resolution** of a lens, we must take into account the phenomenon of **diffraction**. The diffracted image of a point shows a bright central spot, or **Airy**

disk, surrounded by alternating dark and light circles. At a wavelength λ , the radius of the Airy disk is 0.61λ . The images of two points are just resolved when the distance between the images is equal to this radius; this is the case when the distance between the two points is $0.61\lambda/(\text{N.A.})$.

4.3 LIGHT SOURCES

The Best and the Brightest

If we intend to do fluorescence measurements, the light will undoubtedly come from either a laser or a mercury arc lamp. Why are we restricted to these two sources? It's pretty obvious that the laser can readily be induced to put more photons than we really need through a very small area; we used to burn little holes through pieces of paper to impress visitors. Why can't we use a high-intensity quartz-halogen lamp? The microscope companies, after all, sell quartz-halogen illuminators for fluorescence microscopy, and even promote them for immunofluorescence work.

The characteristic we are looking for in a light source is commonly thought of as brightness, which was the old photometric term for it. What we're after is the analogous radiometric quantity, called **radiance** or **sterance**. We can describe radiant energy coming from anywhere, be it a source, an element of an optical system, a wall, etc., in terms of radiant **areance** or **exitance**, which measures the **power**, or **radiant flux** (ϕ) emitted per unit area (A), in units of W/m^2 . A true point source can be described in terms of its **intensity** or **pointance**, which measures power per unit solid angle (Ω), in units of W/sr . There are no real point sources; arcs, filaments, **light emitting diodes** (LEDs), and the good old sun are, instead, what are called **extended sources**. **Radiance** (L) measures power emitted from, transmitted through, or reflected by a surface, per unit of its area, per unit solid angle; that is, $L = \phi/A\Omega$. The units of radiance are $\text{Wm}^{-2}\text{sr}^{-1}$. It is this quantity that is used to compare light outputs of extended sources.

We have already noted that the throughput of an optical system cannot be increased at any point in the system. Since throughput, $\Theta = n^2 A\Omega$, for a system in air, with $n \approx 1$, $\Theta = \phi/L$. We are starting with whatever throughput we get to work with **at the source**, and, as we go through the system, things can only get worse. We can't take light from a system with low radiance and "squeeze it down" through a small cross-section. If we could, we'd probably be able to start a fire by holding a convex lens up to a lightning bug's rear end.

The optical system that transmits illumination light from an extended source to a specimen must, at a minimum, include a lens to collect light from the source and another to converge the collected light on the specimen. Both lenses are generally referred to as **condensers**; for purposes of this discussion, they will be distinguished as the **lamp condenser** and the **microscope condenser**. To add to the confusion, it should be noted that, in the epiillumination systems generally used in fluorescence microscopes, the microscope objec-

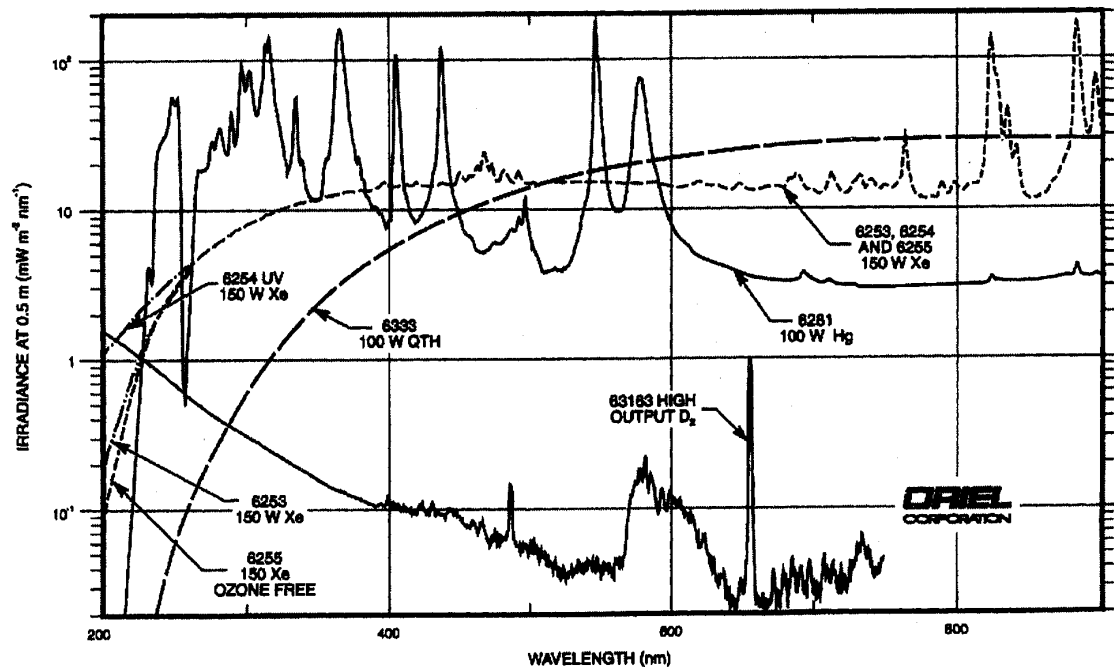


Figure 4-14. Output characteristics of arc (Hg, Xe), quartz-halogen (QTH), and deuterium (D_2) lamps (courtesy of Oriel Corp.).

tive, which collects the emitted fluorescence, also serves as the microscope condenser.

The most efficient illumination, i.e., that which produces maximum radiant flux through the specimen, can be obtained by making an image of some portion of the emitting surface in the plane of the specimen. In order to maximize light collection and specimen illumination, the condenser lenses should be of high N.A., to make the collection and illumination angles as large as possible. Consider, for the moment, that a single convex lens serves as both lamp condenser and microscope condenser. If the emitting surface of the lamp is located two focal lengths from the lens on one side, and the specimen plane is located two focal lengths away on the other, the lens will form a 1:1 image of the source in the specimen plane, and the solid angles for collection and illumination will be the same. If a magnified image of the source is to be formed, the lamp must be placed closer to the lens, and the specimen moved further away; the lens can then collect light over a larger solid angle, but the radiant flux per unit area in the specimen plane will not be increased. If a minified image of the source is formed, the lamp must be placed more than two focal lengths away from the source, and will therefore collect light over a smaller solid angle; again, the radiant flux through the specimen will not be increased.

The largest possible collection angle obtainable using a convex collecting lens will be achieved if the lens is placed one focal length from the source; it will then collect a collimated beam of light, requiring that a second lens be used to

converge light on the specimen. It is also possible to collect light from a substantial fraction of the emitting surface of a source using ellipsoidal or parabolic reflectors. However, while this can allow substantial amounts of power to be directed through surface areas larger than the emitting surface of the source, it is still not possible to put more radiant flux per unit area of the specimen than can be collected per unit area from the source. I'll say it again: **you can't put any more light through a given area of the specimen than comes from the same area of an extended source.** And, if you don't collect all the light you can from the source, you can't get it to the specimen.

Thus, in illumination optics, as was the case for collection optics, it is obviously a good idea to use the **fastest** (highest N.A. or lowest $f/\#$) condenser lens you can get to collect light from a lamp. The condenser lens doesn't have to produce a high quality image; it just has to be well enough corrected to do a decent job of "collimating" the light. Manufacturers of arc lamp systems, slide and movie projectors, and microscopes generally use aspheric lenses for this light collection job; it would otherwise require a lens shaped something like a marble, which would have intolerable spherical aberration. We can't improve on things, by the way, by using an ellipsoidal collector with an arc lamp; this will let you make a brighter image on a movie screen but not in a microscope or flow cytometer; it's a throughput thing. One might be able to collect and refocus the largest fraction of source emission from an arc using paraboloids, a notion suggested by Mike Hercher, to whom I am indebted for

many explanations of throughput. The available parabolooids may not be good enough, and you would have to use humungous interference filters, but it might be possible. You still couldn't get around throughput; you could only collect more of the light to start with. Many people, including some who do optics for manufacturers of flow cytometers, do not seem to appreciate throughput; I kept hounding Mike, and also Rob Webb, until I got explanations in English, but I'll take the rap for this version. Whatever else you do, however, if you want to shine lots of light on a cell, you need a high-radiance source.

Harc, Harc, the Arc!

Now that we are all aware of the desirability of using the brightest lamp source available for flow cytometry, it would be nice to know which of the lamps available is the brightest. May I have the envelope, please? The lucky winner of the brightness sweepstakes is a **100 watt mercury short arc lamp**, for example, the Osram HBO 100W/2; the characteristics of this and some other lamps are shown in Figure 4-14. A 100 W Hg arc lamp has an average (photometric) brightness four times that of a 75 W xenon lamp, eleven times that of a 150 W xenon lamp, almost six times that of a 200 W mercury lamp, and about 100 times the brightness of a quartz-halogen lamp. The radiances go in pretty much the same order. The 100 W lamp has the highest radiance because the arc itself is smaller in size (about 0.25 by 0.25 mm, or 250 × 250 μm) than the arcs in more powerful lamps, and the difference in size more than offsets the lower total flux.

There is one lamp with higher average brightness than a 100 W mercury arc; this is a 500 W xenon arc lamp. However, the mercury lamp, as I calculate things, has a higher radiance in the region of its strong spectral lines at 365, 405, 436, 546, and 578 nm than does the xenon lamp, which has no strong lines in the near UV or visible spectrum. The 500 W xenon lamp, in other words, won't give you more usable power than a mercury lamp in those regions of the spectrum in which you're apt to be interested, and it has the undesirable characteristic of self-destructing without warning, doing a creditable imitation of a hand grenade in the process. This may have adverse consequences for your mental and physical health and for the physical health of any optical elements in the path of the fragments. If you still think you need a 500 W xenon lamp, you probably have space for it in the garage next to the assault weapons.

Is it just coincidental that a better and more peaceful approach to increasing excitation power in arc lamp flow cytometers has come from Norway? Steen and Sørensen¹¹³² modified the front end electronics and arc lamp power supply in a flow cytometer to increase the lamp current by a factor of ten for a few microseconds after the trigger signal goes above its threshold value, thus substantially increasing the excitation power while the cell is still in the observation region. The modification is inexpensive, increases the sensitivity of the instrument, does not substantially decrease the

life of the lamp, and does not appear to increase the likelihood that the lamp will explode; whether anyone will attempt this trick in a commercial instrument remains to be seen.

Why, you may ask, does anyone bother using a 50 or a 200 W mercury lamp, a xenon lamp, or a quartz-halogen lamp in a fluorescence microscope if the 100 W mercury lamp is so good? There are several reasons. The radiance issue is germane to flow cytometry, where we want to illuminate a relatively small region of space. The 250 μm by 250 μm arc in the 100 W lamp is ideal for illuminating an area its own size (or smaller, using a field stop) in a flow cytometer, but we would benefit from the use of a larger arc size if we wanted to get uniform illumination over the entire field of view of a fluorescence microscope or an imaging cytometer, especially when using the lower magnifications. The other mercury lamps have larger arcs; in the 200 W lamp, arc dimensions are 600 by 2200 μm. The other lamps also have longer average lifetimes, e.g. 400 hours for a 200 W mercury and 1000 hours for a 200 W mercury-xenon lamp as opposed to 200 hours for the HBO100W/2.

Xenon lamps are generally chosen for their spectral characteristics, or lack thereof. Fluorescence microscopy pretty much grew up with mercury lamps. At first, only the ultraviolet emission from these lamps, principally the strong 365 nm line, was used for excitation. As a result, the dyes that were first found useful for fluorescence microscopy were those that could be excited in the near UV, either because they had absorption maxima in that spectral region or because they had high enough quantum efficiencies to emit appreciably when excited at wavelengths far from their absorption maxima.

Xenon lamps do not have strong spectral peaks in the visible region; they are preferred to mercury lamps as sources for spectrofluorometers for that reason, and are similarly useful for microspectrofluorometry. In principle, they allow use of a wider range of dyes; in practice, however, the continuum between the strong lines of a mercury arc is apt to be as bright a source in those spectral regions as is a xenon arc, and the mercury lamp is better near its strong UV, blue, and green lines.

A publication by Koper et al¹¹⁰⁷ on the addition of arc lamp illumination to a B-D FACS reported that ILC Technology (Sunnyvale, CA) had made xenon lamps with a zinc iodide additive, producing strong emission lines between 460 and 480 nm. Such lamps should be useful for excitation of fluorescein. ILC Technology told me they were no longer available, but some more were apparently made for use in the B-D FACS Analyzer; special lamps are also now available from Partec, which manufactures arc source cytometers.

Fluorescein and acridine orange, both of which have absorption maxima between 450 and 500 nm, in a region in which a mercury lamp has no strong lines, first became useful in fluorescence microscopy at a time when only UV excitation was used. These dyes are better excited by the blue mercury line at 436 nm than by the UV line, and thus be-

came more useful when filters permitting use of blue excitation became widely available. Both fluorescein and acridine orange are ideally suited for excitation by the 488 nm line of an argon ion laser; this characteristic has made these dyes particularly useful for flow cytometry. There are other dyes that are easily used in fluorescence microscopy, but see only limited use in flow cytometry, e.g., brilliant sulfaflavine, which has an excitation maximum at about 420 nm. This dye is well excited by the UV and blue lines of a mercury lamp, but very poorly excited at 488 nm.

Arc lamps require relatively elaborate power supplies, including circuitry to generate a high-voltage pulse or RF (radio frequency) pulse train to ionize the gas, producing a conductive medium in which an arc can be started and maintained. Although arc lamps intended for spectrophotometry and microscopy are sold with power supplies that are supposed to provide light output regulated to better than 1%, I have measured 6-10% variations in light intensity, synchronized with the power line frequency, in arc lamp systems made by several manufacturers. A lamp, or any other light source, with such intensity fluctuations is apt to be unacceptable for cytometry.

Several of the largest manufacturers of lamps have, since the late 1980's, undertaken major efforts at developing arc lamps for use as automobile headlights, primarily to provide the stylists with headlights less than 1" high, and the status-conscious automobile buyer with yet another option to covet. It seems to have worked; I see a lot of cars with what the industry calls "discharge headlights" on the road these days, but I haven't heard of anybody using one in a flow cytometer yet. O Gucker, where art thou?

Peters¹⁰⁶ compared laser and mercury arc lamp illumination for flow cytometry, and found that arc lamps offered some advantages over ion lasers as UV sources. At the present state of the art, I would favor UV helium-cadmium lasers and/or violet diode lasers²⁴⁴ over both arc lamps and ion lasers for most applications; I will have more to say about this later.

Quartz-Halogen Lamps

In comparison to arc lamps, quartz-halogen lamps are less expensive and can be operated from simpler power sources; they employ incandescent filaments, but produce more green and blue light than conventional filament lamps. They can thus be used for fluorescence excitation in microscopy. The Technicon Hemalog D differential counter^{84,85} successfully used a quartz-halogen lamp for flow cytometric absorption and scattering measurements; the newer hematology systems from Bayer Diagnostics, which absorbed Technicon, add a red laser, but keep the lamp as a source for absorption measurements.

I have detected brightly stained fluorescent objects (beads and cells stained with DNA fluorochromes) using a microscope-based flow cytometer with a quartz-halogen source. However, the quartz-halogen lamp is at best less than a tenth as bright as an arc lamp, and its low photon output

severely limits the precision with which fluorescence can be measured at conventional flow rates. LEDs (see below) are better.

Light Emitting Diodes (LEDs)

LEDs are everywhere these days, not just on electronic gadgets. My bicycle tail light, and various auxiliary rear lights on automobiles, use red LEDs, and green LEDs are replacing incandescent sources in traffic lights. What has made this possible is the development of "high-brightness" LEDs by companies including Cree, LumiLeds (a joint venture of Agilent, formerly part of Hewlett-Packard, and Philips), and Nichia. Although an LED, like a lamp, is an extended source, it is not an incandescent source; when current is applied to the device, photons are emitted as electrons traverse the energy "band gap" in a semiconductor material. The precise composition of the material determines the energy range of these transitions, and emission is typically restricted to a relatively narrow spectral region (tens of nanometers). The lower end of the range of available LED emission wavelengths has moved steadily downward; there are now UV and violet LEDs in addition to blue, green, yellow, red, and infrared devices. Bob Hoffman and Eric Chase presented a poster at the 2000 ISAC meeting describing the use of UV and blue LEDs as illumination sources in flow cytometers²⁴⁵. They measured DNA content in DAPI-stained calf thymocyte nuclei, with a CV of 2 percent, using a UV LED light source, and, using a blue LED source, could readily discriminate different levels of immunofluorescent staining of leukocytes by a PerCP-antiCD45 or a PE-antiCD4 antibody. I have recently compared a 100 W mercury arc lamp, a 50 W quartz-halogen lamp, and a high-brightness blue (460 nm) LED as illumination sources in a fluorescence microscope; the LED is at least as good as the quartz-halogen lamp in terms of its ability to produce visually detectable images of dimly stained objects, while consuming only a few hundred milliwatts of electrical power, rather than tens of watts.

Illumination Optics for Lamps and LEDs

Figure 1-21 and the related discussion on p. 51 show that two simple cylindrical lenses are all you need to get proper illumination for a flow cytometer using one (or more) laser sources; the monochromatic nature of laser light is, in general, all that is needed to define the excitation wavelength range.

The optics required for lamp illumination in flow cytometry are more complicated. Even when the same microscope objective is used for illumination and collection, as in an epiilluminated fluorescence microscope or a typical arc source flow cytometer, we also have to have a condenser lens for the lamp, and possibly a few additional lenses and diaphragms, in addition to some optical filters and a dichroic mirror (Figure 1-4, p. 9).

What we will consider first, however, is the plain, old-fashioned **transmitted light microscope** and its illumina-

tion optics. Separate lenses, i.e., the **condenser** and **objective**, are used for illumination and light collection. This configuration has also been used in flow cytometers, beginning with Kamentsky's⁶⁵; the original Block system⁸⁸ used two opposed pairs of lenses to permit absorption and scatter measurements, and Technicon's Hemalog D differential counter^{84,85} and the Steen/Lindmo instrument¹⁰⁰⁻¹⁰³, now being resurrected by Apogee Flow Systems, also used or can use opposed lenses.

In transmitted light microscopy, one of the goals (if I said objectives that would just confuse things) is generally high spatial resolution. Resolution increases with increasing N.A., which is a function of the angle over which light is collected. Now, if we are concerned with scattered light or fluorescence, we can essentially treat our specimens as what optical theorists call **self-luminous objects**. It doesn't matter whether we illuminate the specimen over a large angle or a small one; light will be scattered, and fluorescence emitted, in all directions, i.e., over a solid angle of 4π steradians.

If what we are looking at, however, is light transmitted by the specimen, the situation is quite different. In order for the light to come out of the specimen at large angles, it has to go into the specimen at large angles. Thus, the condenser has to have an N.A. equal to that of the objective. It's amazing to me how many people either don't know that or have forgotten it. The N.A. value engraved on a lens simply defines the highest N.A. you can get; actually getting it is up to you. The immersion objectives on good microscopes typically have an N.A. of 1.3-1.4, as do the condensers. However, recall that the n in the formula that defines N.A. as $n \sin \theta$ is the refractive index, which is only 1.0003 for air. In other words, if the condenser is not immersed, you might as well not bother immersing the objective, because the effective N.A. can't be higher than 1. If the condenser is immersed, but its aperture diaphragm is stopped down, restricting the region through which light enters the condenser, you also won't be able to achieve the full N.A. of the microscope objective.

The condenser lens illuminates the specimen with light derived from the source. There are two basic arrangements used for specimen illumination; they are shown in Figure 4-15. In **Köhler illumination**, an image of the light source is formed in the back focal plane of the condenser lens. Thus, light rays from the source (image) leave the condenser lens and pass through the specimen as parallel bundles. Each point in the specimen plane is illuminated by light coming from all points of the source image. The alternative illumination setup is called **critical illumination**, for which an image of the source is formed in the specimen plane.

The position of the arc in an arc lamp will change with time; this phenomenon is referred to as **arc wander**, and is more of a problem with some lamps than with others. Significant arc wander may make it preferable to use Köhler illumination, which will average out intensity fluctuations at various points in the arc image. However, the overall level of illumination obtained from Köhler illumination is typically

lower than that achieved when critical illumination is used. Critical illumination can be used satisfactorily with an arc source if a large enough image of the arc is formed in the specimen plane to keep the intensity distribution homogeneous over the observation region, even in the presence of arc wander.

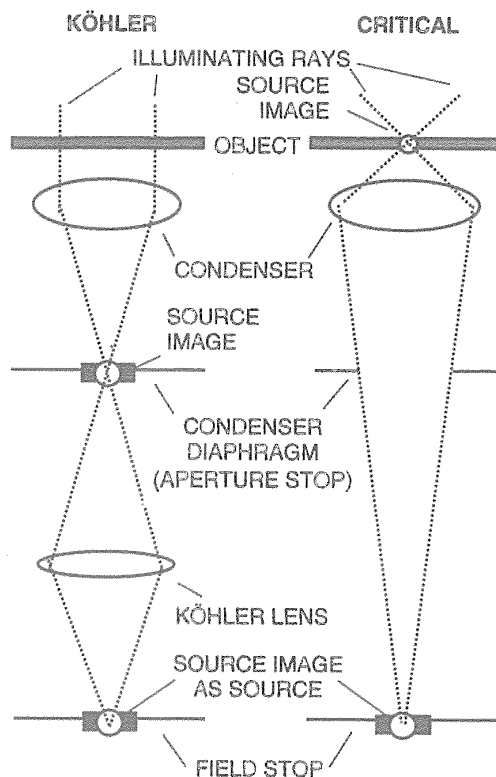


Figure 4-15. Köhler and critical illumination.

When Köhler illumination is employed, maximal illumination is achieved when the source image fills the entire aperture of the condenser in the condenser's back focal plane. The use of intermediate **relay lenses** relaxes the rather severe constraint on the physical geometry of the system that would be imposed by the use of a single lens to collect light from the source and image the source in the back focal plane of the condenser. In a typical transmitted light microscope with a substage illuminator, one or two lenses are used to collect light from the source, forming an image of the source in the plane of the substage diaphragm, which serves as a field stop. In critical illumination, the condenser is used to form an image of this image in the specimen plane; in Köhler illumination, an additional **Köhler lens** forms an image of this image in the back focal plane of the condenser.

Arc Source Epiillumination for Flow Cytometry

An optical system typical of those used in an arc source flow cytometer or an epiilluminated fluorescence microscope

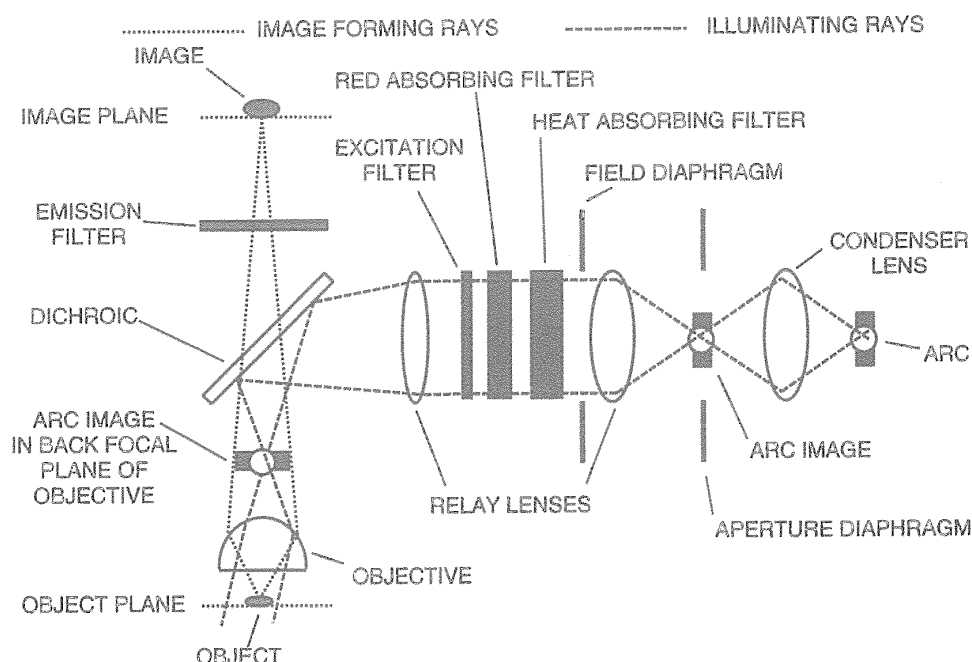


Figure 4-16. Optics for arc source epiillumination for fluorescence microscopy or flow cytometry.

is diagrammed in Figure 4-16. Infrared (heat) and red light are removed from the arc source by glass **heat absorbing** and **red absorbing filters**; since these are color glass rather than dielectric or interference filters, their transmission is not affected by angle of incidence. Their placement is not critical, but putting them between the source and the **excitation filter** used to select excitation wavelength reduces the heat load on that filter. In the past, color glass excitation filters, such as the UV-transmitting Schott UG 1, were generally used for UV and violet and sometimes used for blue illumination; interference filters were preferred for blue-green and green illumination. These days, whether you're doing fluorescence microscopy, image cytometry, or flow cytometry, it's more likely that you'll be using interference filters, and getting the right filter characteristics is critical to the quality of results.

In some instruments, a collimated beam is formed in the excitation optics to allow placement of interference filters normal to the beam; since the spectral transmission of interference filters changes with the angle of incidence, this arrangement maximizes the chance of the filter performing as specified. It is obviously difficult to avoid having converging beams impinging on the dichroic from both the illumination and collection sides of an epiilluminated fluorescence microscope, and the fluorescence collected by the objective will also pass through the **emission filter(s)** as a converging beam. However, this beam, being on the image side of a high-N.A. lens, forms a cone with a relatively small half angle; it is unlikely that any of its constituent rays strikes the emission filter at an angle far enough from normal to cause

problems. It is also possible to reject even unwanted light that hits the filter at somewhat larger angle by keeping the passbands of excitation and emission filters well separated.

If critical illumination is used, images of both the arc and the object are formed in the image plane; with Köhler illumination, light from the arc is diverging as it passes through the image plane. The latter arrangement should result in less stray light from the arc getting through the filters and reaching the detector. However, if the emission filters contain a colored (absorptive) backing, filter fluorescence is more likely to occur and, because the light is diverging, one cannot simply use a field stop to eliminate the filter fluorescence, as is easily done in a flow cytometer with a laser source and orthogonal geometry.

Figure 4-16 shows a single emission filter, as would be used in a fluorescence microscope; a microscope-based flow cytometer making multicolor fluorescence measurements would use a series of filters and dichroics, similar to those shown in Figure 1-21 (p. 51), to separate the desired wavelength regions.

Lasers as Light Sources for Flow Cytometers

If arcs work for fluorescence microscopy and flow cytometry, why does anybody use lasers? Looking back over flow cytometric history, we see that Kamensky's first instruments^{65,66,78}, the original Stanford sorter⁸², and the ICP⁸³ all used arc sources. We have already seen that collecting the light from an arc source and putting into a small volume such as the observation region of a flow cytometer is a non-trivial task, just in terms of selecting lenses. We haven't

really paid any attention to another major problem area, i.e., the selection of filters or other optical elements to define the excitation and emission wavelength regions that will be used. It is obviously much easier to get laser beams to go where you want them to than it is to ride herd on arc lamp emissions, and one doesn't have to tweak and/or realign laser focusing optics nearly as frequently as is required with arc lamp optics. Two-thirds of the optical filter problems associated with arc lamps are essentially eliminated when using laser sources, because you dispense with the excitation filter and dichroic, and only have to worry about emission filters for the detectors. The monochromaticity of laser light often permits you to relax the specifications on the emission filters, as well. Lasers are undeniably brighter than arc lamps; finally, before they became ubiquitous in CD players and as pointers, they were trendy and sexy and some people got (and may still get) a feeling of power from having to have the whole building replumbed and rewired so they could/can connect their instruments.

The Los Alamos group, which first reported the use of a laser for fluorescence flow cytometry⁷⁹, probably was influenced by the high radiance and the resulting ease with which an apparatus could be designed and built; since high technology tends to come to Los Alamos before it gets to a lot of other places, people there had the first crack at using lasers anyway. The use of an argon laser instead of an arc lamp in the second version of the Stanford sorter⁸⁶ was motivated by the improvement in the quality of weak (immuno)-fluorescence measurements possible using the brighter source.

There is no doubt (L. Kamensky, personal communication) that the novelty and trendiness of lasers strongly influenced the decision in the late 1960's to use them as sources in the production Cytograf and Cytofluorograf. The latter was the first commercial product of any kind to incorporate an argon ion laser; its light source was a small, air-cooled device very much like those now used in most commercial flow cytometers, with a power output of about 10 mW. The laser machismo which subsequently developed among commercial and noncommercial builders of flow cytometers made it difficult, during the 1970's and early 1980's, to appreciate that bigger was not necessarily better. Fortunately, times have changed.

Laser Illumination: Going to Spot

Laser sources differ radically from extended sources in several respects; one is that the emission from lasers is confined to a very small solid angle, so it is generally possible (neglecting transmission losses) to focus all of the energy in the beam to a circular or elliptical spot. Most lasers used for flow cytometry emit a so-called TEM_{00} beam, in which the energy distribution is Gaussian. Spot "diameters," in the case of a circular spot, or "width" and "height," in the case of an elliptical spot, define the " $1/e^2$ points," at which intensity is $1/e^2$ (or 0.135) times the value on-axis. Approximately 87.5 percent, or the fraction $[1 - (1/e^2)]$, of the total

emission is contained in the region within the $1/e^2$ points; the area of this region corresponds to the area of a central elliptical region of a bivariate (two-dimensional) Gaussian (normal) distribution within two standard deviations of the bivariate mean.

Let's consider a typical laser illumination setup for a flow cytometer. The laser is assumed to be emitting a beam with a diameter of D mm. We will assume that the beam is collimated, i.e., not diverging at all, because its actual divergence is pretty small. If we took a convex spherical lens of focal length f mm and placed it in the beam, with its axis coincident with the beam axis, the beam would be focused to a round spot with diameter d μ m at the focal distance f mm from the lens. The formula generally used for estimating the spot diameter d is

$$d \cong (4/\pi)(\lambda f/D) \cong 1.27(\lambda f/D),$$

with λ being the wavelength (in μ m, **not** nm) of the laser emission. This formula neglects lens aberrations and assumes that the beam is focused to a **diffraction limited spot**; the formula is not, and cannot be, derived from geometric optics.

Loken and Stall calculated the spot size normally obtained in the original B-D FACS, which focused the beam with a 125 mm focal length spherical lens, as 55 μ m with 515 nm illumination and 61 μ m with 458 nm illumination⁶, based upon laser beam diameters of 1.5 mm and 1.2 mm, respectively, at these wavelengths. Let us assume the center of a 20 μ m wide core stream runs along a diameter of a 60 μ m spot, and determine the variation in illumination over the width of the core. A table of the Gaussian distribution shows that the $1/e^2$ points, 30 μ m off-axis, represent distances of 2 standard deviations (S.D.) on either side of the mean. A point 10 μ m off-axis is thus 0.67 S.D. from the mean, and receives only 0.79 times peak beam irradiance.

What this means is that a small particle traveling near the outside of a 20 μ m core stream receives only 79 percent as much illumination as an identical particle traveling down the core axis. If fluorescence emission is a linear function of illumination intensity, two identical particles following these different trajectories will produce signals differing in amplitude by over 20 percent. This does not make for low coefficients of variation. If one observes a low CV in a fluorescence distribution measured with this beam geometry and core size, the only possible explanation is that the illumination power used is at a level at which the change in fluorescence intensity for a given change in illumination intensity is relatively small (p. 117), minimizing the effects of uneven illumination. If you have enough laser power, you can be sloppy about illumination optics.

If the laser power level is lower, there are two obvious ways to improve performance of the system just described. One is to make the core smaller. Suppose we keep the 60 μ m round spot and go to a 10 μ m core diameter; the edges of the core are now 5 μ m off axis, at which point the irradiance is almost 95 percent of the peak irradiance. That's bet-

ter than 80 percent but it still isn't where we want to be. We'd have to have a 5 μm core to keep illumination variations to within 1 percent of the peak value with a 60 μm spot. This is acceptable if you're not processing 10 μm cells; if you want to look at bacteria, or chromosomes, you can probably get to a 5 μm core and keep the round spot as is. In fact, the unmodified B-D FACS didn't do badly at chromosome analysis¹⁰⁹, although high laser power, rather than smaller particle or core size, was probably responsible for the low CVs.

It's not that easy to use a really small core if we're looking at cells, and easiest way to get more even illumination over the width of the core is to form an elliptical focal spot, using **crossed cylindrical lenses**, as shown in Figure 4-17, decreasing the 60 μm dimension along the direction of flow and increasing the dimension perpendicular to the direction of flow. Going back to the Gaussian distribution, we find that we can keep illumination at over 98 percent of the peak value within the central 10 percent of the beam width. With a 100 μm beam width, we can handle a 10 μm core. By keeping the beam height a few times as large as a cell diameter, we avoid the "slit-scan" effect discussed on p. 51, and can use peak detectors instead of integrators, making the signal processing simpler, but losing the ability to derive information about cell size from pulse width or to discriminate doublets by their pulse height vs. pulse area. If we do want the additional information, we can simply use a shorter focal length lens, and add the extra electronics.

I should note that, in some of its instruments, B-D uses a wedge, or prism, and a spherical lens, rather than crossed cylindrical lenses, to form an elliptical spot. Either way, lasers make optical design much easier than do lamps.

If you want to focus down to extremely small spots, e.g. 5 μm or less, it may be advisable to expand the laser beam first. This is not usually done in commercial laser source flow cytometers. I have tried beam expansion several times, and find that it causes as many problems as it cures, so I've given up on it; commercial instruments, by and large, don't use beam expansion.

Suppose you want to use two or more laser beams? My approach, and that used in many commercial instruments, has generally been to use mirrors to get the unfocused beams lined up in space and then to put them through a single set of crossed cylindrical lenses. Others, such as the groups at Livermore and Los Alamos, have gone to more elaborate arrangements with the beams at small angles to one another going through different sets of lenses, primarily to get very precise control over focus and spot size for high-precision analyses of chromosomes. If there isn't a lot of power to spare, and/or if the beam diameters and wavelengths used are different enough so that one lens pair won't give you acceptable dimensions for the spots from both lasers, it may be easiest to place the lasers and their focusing optics on opposite sides of the flow cell; instruments from both Cytomation and Luminex follow this practice, and I've been doing it in my Cytomutts, when necessary, for years.

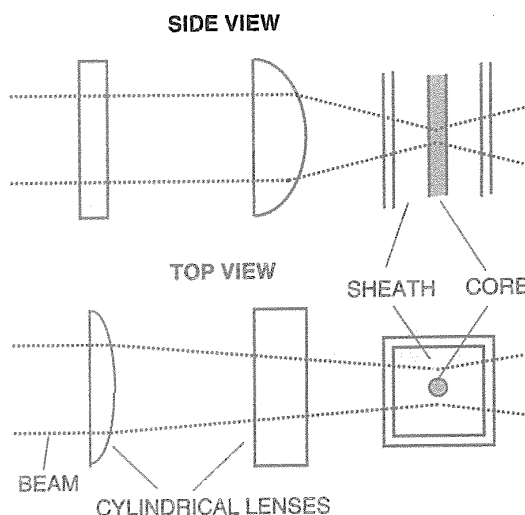


Figure 4-17. Use of crossed cylindrical lenses to focus a laser beam to an elliptical spot on the core stream.

Shedding Light on Cells: Lasers, Lamps, and LEDs

By now, it should be clear that lasers and arc lamps are usually the most realistic choices as light sources for flow cytometers. However, I have mentioned that LEDs have been used successfully as sources for two common types of flow cytometric analysis, namely, DNA content determination and immunofluorescence measurements. So what's the bottom line? How much light can each of these sources get through a cell? Let's do the math and find out.

If a laser beam were focused to an elliptical spot 100 μm wide by 20 μm high, a not unreasonable size to use in a flow cytometer, it can be calculated, from characteristics of the bivariate Gaussian distribution, that about 10 percent of the total power in the beam would illuminate a cell 10 μm in diameter at or near the center of the focal spot. That percentage figure will hold for any laser wavelength or power. If the laser in question is an air-cooled, 488 nm argon ion laser, emitting about 20 mW, probably the most widely used source in fluorescence flow cytometers at present, we get all of 2 mW through our 10 μm diameter cell. That is an illumination intensity of $2.55 \times 10^7 \text{ W/m}^2$; since a 488 nm photon has an energy of $4.07 \times 10^{-19} \text{ J}$, the figure is equivalent to $6.27 \times 10^{25} \text{ photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (photons per square meter per second), and the photon flux through the 10 μm cell is $4.9 \times 10^{15} \text{ photons}\cdot\text{s}^{-1}$ (photons per second).

There would be less power and fewer photons from a focused laser beam impinging on smaller targets. Only about 1 percent of the light (roughly $5 \times 10^{14} \text{ photons}\cdot\text{s}^{-1}$) would hit a 1 μm diameter bacterium; about 0.002 percent of the light ($10^{12} \text{ photons}\cdot\text{s}^{-1}$) would hit a 0.1 μm virus particle.

The laser calculation is relatively easy; to get the corresponding numbers for lamps and LEDs, we have to dig up data on radiance from sources such as Figure 4-14 (p. 125). The figure originally appeared in an optics catalog from Oriel Corporation, now Thermo Oriel (Stratford, CT; www.oriel.com), a major supplier of light sources. It compares the **spectral irradiance** of mercury (Hg) and xenon (Xe) arc and quartz tungsten halogen (QTH) filament lamps. The irradiance values are given in $\text{mW}\cdot\text{m}^{-2}\cdot\text{nm}^{-1}$ (milliwatts per square meter area per nanometer wavelength) at a distance of 0.5 m; the corresponding radiance is calculated by determining the solid angle represented by a square meter at this distance and by integrating over an appropriate wavelength region.

The 100W Hg arc lamp has an irradiance of about $120 \text{ mW}\cdot\text{m}^{-2}\cdot\text{nm}^{-1}$ in a 10 nm band around the strong ultraviolet (UV) line at 366 nm, so total irradiance in this spectral band is approximately $1.2 \text{ W}\cdot\text{m}^{-2}$. The surface area ($4\pi r^2$) of a sphere with a radius of 0.5 m is $4\pi(0.25)^2 \text{ m}^2$. Since this is the surface area subtended by a solid angle of $4\pi \text{ sr}$ (steradians), 0.25 m^2 is the surface area occupied by 1 sr; the radiance is thus about $0.3 \text{ W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}$, or $300 \text{ mW}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}$.

The arc in a 100W Hg lamp is 0.25 mm (250 μm) in diameter; its surface area is $4\pi(0.000125)^2 \text{ m}^2$; therefore the surface area of a 1 sr segment of the arc is $(.000125)^2 \text{ m}^2$, or $15625 \mu\text{m}^2$. The power radiated through 1 sr is 300 mW. A spot 10 μm in diameter on the surface has an area of $\pi(25)^2 \mu\text{m}^2$, or $78.5 \mu\text{m}^2$; thus, about 1.5 mW $[(78.5/15625)\cdot 300]$ of UV power in the bandwidth discussed would be emitted through this surface. This represents the maximum amount of power collected from the arc which can be directed back through the same area, thus, if the arc is used to illuminate a cell 10 μm in diameter, no more than 1.5 mW can impinge on the cell at any given time, no matter how efficiently light is collected from the arc and transmitted through the optics.

Mercury arc lamps have strong emission lines at several wavelengths other than 366 nm; peak radiance near that of the 366 nm line is also obtainable at 313 nm, farther in the UV, at 405 nm (violet), 436 nm (blue-violet), 546 nm (green), and 578 nm (yellow). Under the best conditions, neglecting transmission losses in lenses and in the filters used for wavelength selection, one would expect to be able to direct at most 1-1.5 mW in any of these wavelength regions from an Hg arc lamp through a 10 μm cell at any time. Since we have already figured out that we can get 2 mW (10 percent of the total beam power) through a cell with relatively little effort if we use a 20 mW laser source, we can calculate that the 100W Hg arc lamp will, at best, provide cytometrically usable light output equivalent to that from a 15 mW laser. That's more than adequate to do a lot of cytometry, especially when oil immersion lenses are used for illumination and light collection, but it's suboptimal for applications such as high speed sorting.

Illumination intensities obtainable from Hg arc lamps in the real world may be even lower than the above calculations suggest. Unger et al.¹²⁴⁶ measured an illumination intensity of

$4.8 \times 10^5 \text{ W}\cdot\text{m}^{-2}$ in a fluorescence microscope using green (546 nm) illumination from a 100W Hg arc lamp, with a N.A. 1.4, 60 $\times$ oil immersion objective serving as condenser and collector lens. At 546 nm, a photon has an energy of $3.64 \times 10^{-19} \text{ J}$; the photon flux through a 10 μm cell in this setup would therefore be about 10^{14} photons $\cdot\text{s}^{-1}$, or about 1/50 of what one could expect to get from a 20 mW laser. However, this level of illumination proved more than adequate to detect fluorescence from single molecules of a tetramethylrhodamine conjugate using a cooled CCD camera and an observation time of 100 ms.

A xenon arc lamp, as can be seen from Figure 4-14, has a relatively flat emission spectrum between the near ultraviolet and the near infrared (350-750 nm); this makes the Xe arc a desirable illumination source for spectrophotometers and spectrofluorometers. However, since the radiance of the Xe arc over this range is only about 1/10 the radiance of the Hg arc at its strong emission lines, the Xe arc is less desirable as a source for cytometry, except possibly in the region between 450 and 500 nm, where its radiance is slightly higher than that of Hg arc lamps. In any given 10 nm wavelength band, a Xe arc lamp probably won't put any more photons through a 10 μm cell than would a 200 μW laser.

The radiance of a quartz tungsten halogen filament lamp is substantially lower than that of a xenon arc lamp below 600 nm, and slightly higher between 600 and 800 nm. However, the area of the emitting surface of the filament lamp is much larger than the area of the arc in an arc lamp; thus, much less power – probably only a few tens of microwatts at most – can be collected from and directed through a small area. This makes filament lamps poorly suited for fluorescence excitation in flow cytometry, although they have been used quite successfully in flow cytometers which measure absorption and light scattering. For these purposes, they have the advantage that it is relatively easy to achieve precise regulation of output power.

What about LEDs? The current brightness champion among blue LEDs is Cree, Inc.'s XBright 470 nm device, which emits 150 mW from an $810 \times 810 \mu\text{m}$ surface area. The emission from the $78.5 \mu\text{m}^2$ area corresponding to the area of a cell 10 μm in diameter is about 18 μW , or less than 1/100 what one could get from a 20 mW laser. However, measurements made on my fluorescence microscope show that a blue LED should deliver about twice as much light to a cell as can be obtained from a quartz halogen lamp.

I'm sure LEDs will get brighter, but I don't think they will ever get hundreds of times brighter, so I don't see them replacing lasers for many flow cytometry applications. However, I will mention that Agilent is now producing a simple flow cytometer module that uses a red diode laser and a blue LED as light sources. It was designed for a restricted range of applications, and it appears to do the jobs it was designed to do. I have also seen a very impressive, inexpensive fluorescence imaging cytometer prototype with a blue LED source. The next few years should be interesting, but, for now, we need to focus, and focus on, lasers.

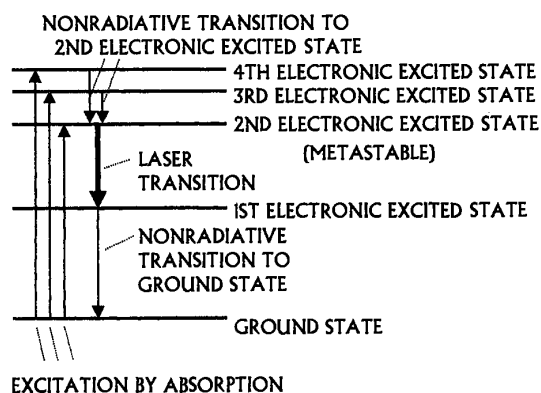


Figure 4-18. Energy levels involved in laser action.

Lasers: The Basic Physics

Einstein on the Beam: Stimulated Emission

The word “**laser**,” as previously noted, is an acronym for “**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation.” The physical process behind all lasers is **stimulated** or **induced emission**, described by Einstein in the early 1900’s. In order for any kind of light emission to occur, the prospective emitter, an atom, ion, or molecule, must be excited by absorption of a photon, raising an electron to a higher energy level. After a brief period of time, the molecule typically returns to a lower energy state by emitting a photon with energy less than or equal to that of the absorbed photon. Under most circumstances, only a small fraction of the molecules in a material are in excited states, and the photons emitted from different excited molecules are different in wavelength, phase, and polarization; such emission as occurs is called **spontaneous emission**. However, as Einstein showed, once a molecule (or atom or ion) has been excited by absorption, the mere presence of a photon or photons of a particular energy in its vicinity increases the probability that it will emit a photon of the same energy (frequency or wavelength), phase, and polarization. Thus, photons can **induce** or **stimulate** the emission of like photons, and the light generated by stimulated emission is **monochromatic**, and **coherent**, i.e., the emitted radiation is at the same wavelength, in phase with, and propagating in the same direction as the stimulating radiation. No other mechanism can generate light with such uniform characteristics.

Stimulated emission becomes more likely as the fraction of the molecules in excited states increases, and can become self-sustaining when there is a **population inversion**, i.e., when the excited molecules outnumber those in the lower energy state. In general, it is difficult to create population inversions for energy transitions between the lowest excited state and the ground state of a molecule, because the ground state is more favorable on thermodynamic grounds according to the Boltzmann law (p. 109). Many practical

lasers emit at a wavelength corresponding to the energy of a transition between a **metastable** higher energy excited state, i.e., one with a relatively long lifetime, and a lower energy excited state. The lasing medium is excited, or **pumped**, by electrical energy or by a high-intensity light source, causing the molecules in the medium to undergo transitions to excited states with energies equal to or higher than that of the metastable state; those at higher energies subsequently drop to the metastable state nonradiatively. Initially, spontaneous emission occurs at a particular laser wavelength as molecules drop from the metastable state to the lower excited state; thereafter, spontaneously emitted photons stimulate the emission of additional photons at that laser wavelength and the process continues. The population inversion required to sustain stimulated emission is maintained because molecules rapidly leave the lower energy state of the laser transition by thermodynamically favorable transitions to excited states of still lower energy or to the ground state. A diagram of the energy levels typically involved in laser action appears as Figure 4-18.

Although the acronym “**laser**” stands for “**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation,” an operating laser is more like an amplifier that has been driven into oscillation by application of positive feedback. One can drive an audio amplifier into oscillation in this fashion by placing a microphone in front of the speaker; the **resonant frequency** of oscillation is a function of the distance between the microphone and speaker.

Look, Ma, One Cavity: Optical Resonators

The initiation of stimulated emission in a volume of a suitable material will not in and of itself produce the concentrated, low-divergence light beams that characterize lasers and on which so much of their utility depends; it will, instead, result in light emission in all directions, i.e., over a solid angle of 4π steradians. This is so because, whereas the photons produced by stimulated emission travel in more or less the same direction as the stimulating photons, the spontaneously emitted photons responsible for the first round of stimulation do not have any directional preference. It is therefore necessary to perform some geometrical and optical manipulations in order to make a usable laser.

First, the volume of **lasing medium** in which stimulated emission occurs is shaped to produce some directionality of emission. As was just mentioned, spontaneously emitted photons are equally likely to be emitted in any given direction, and photons produced by stimulated emission, which follow the paths of these stimulating photons, will therefore also be equally likely to be emitted in any given direction. The probability that one photon will stimulate emission of others in the medium is proportional to the length of the path of the photon in the medium. If the medium were formed into a spherical shape, this average path length would be the same in all directions. The **gain** of the lasing medium, i.e., the number of stimulated photons emitted per unit distance per incident photon, is

predominantly dependent on the quantum mechanical properties, i.e., energy levels and transition probabilities, of the medium. If the gain is high enough, and the intensity of excitation of the medium is sufficient, stimulated emission may be sustained in a spherical volume, but emission will be neither directional nor coherent.

In gas, ion, and solid-state lasers, the lasing medium is shaped into a long, thin cylinder or rod; photons emitted parallel to or at small angles to the axis of this cylinder are more likely to stimulate emission than photons emitted along or near the radius, because the path of the axial photons is substantially longer. Thus, the geometry of the medium will favor emission along the axis. Making the medium longer will, in general, increase the amount of power that can be obtained.

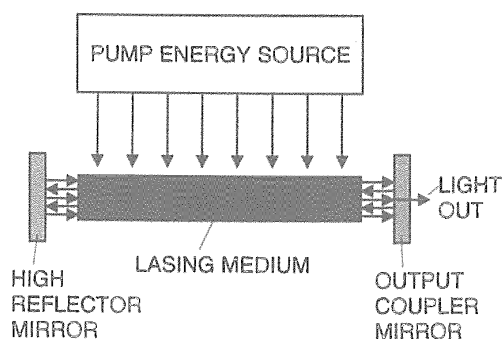


Figure 4-19. Schematic of a laser. The cavity is the region between the mirrors.

The directional property achieved by shaping the lasing medium is augmented by placing the medium inside a relatively rigid structure, called an **optical resonator**, with precisely aligned and spaced mirrors, highly reflective at the desired output wavelengths, mounted at opposite ends. Light emitted along the axis of the resonator is reflected back along the same path again and again; light at increasingly larger angles to the axis is less and less efficiently reflected back through the medium. Since light produced by stimulated emission is identical in wavelength, phase, and direction to the stimulating light, most of the emission confined within the **laser cavity**, i.e., the space between the mirrors, will be concentrated along or very near its axis.

Laser output is produced by making one of the mirrors, called the **output coupler**, able to transmit a small fraction of incident light; the amount of transmission permissible varies with the gain of the medium, which must be high enough to make up for the light lost by transmission outside the cavity and the light lost by absorption within the cavity. The mirror opposite the output coupler, called the **high reflector**, is made to reflect as much light as possible. The spacing between the mirrors is critical. If they are an even number of wavelengths apart, there will be constructive

interference between the rays incident on and those reflected from the mirrors, maximizing output; if not, there will be destructive interference, which may be enough to prevent laser action entirely. A schematic of a laser is shown in Figure 4-19.

Laser Action à la Mode

The resonator can be thought of as analogous to an organ pipe; the length of the pipe, and the effective distance between the mirrors of the resonator, determine the frequency of the standing wave sustained by the structure. In the case of the resonator, this characterizes what is known as the **longitudinal mode** of the laser.

The energy profile of the beam itself, or the **transverse electromagnetic mode (TEM)** of the laser, is determined by the geometry of the medium as well as by the geometric optics of the mirrors. If stimulated emission is confined to a volume close to the axis of the resonator, the laser will operate in what is called TEM_{00} (pronounced "tee-ee-em-zero-zero"); the intensity profile associated with this mode is Gaussian. As the effective cross section of the medium increases, other transverse excitation modes, cartooned in Figure 4-20, are superimposed on TEM_{00} . These modes are individually undesirable in lasers designed for use in cytometry because they are, in general, not radially symmetric, but, rather, multilobed, and, therefore, are likely to produce nonuniform illumination.

Diode lasers are very different in structure from gas, ion, metal-vapor, and most solid state lasers. A diode laser is basically a light emitting diode fabricated so that polished facets on the semiconductor material itself or adjacent structures of differing refractive index reflect emitted light back into the active region of the diode, favoring stimulated emission and directional propagation. In the common edge

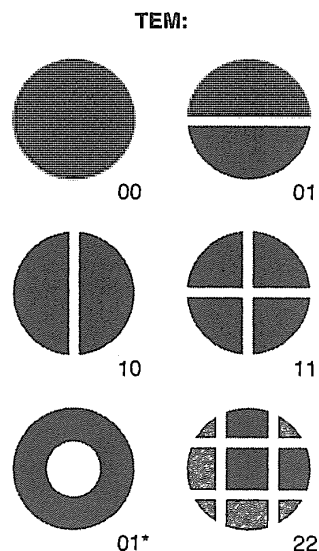


Figure 4-20. Laser transverse electromagnetic modes.

emitting diode laser designs, the emitting surface is a stripe about 1 μm high and 3 to 5 μm wide; the emission mode structure (see Figures 4-21 and 4-22) is substantially different from those shown in Figure 4-20.

Pumping Ions

To produce and maintain the population inversion necessary to sustain stimulated emission and laser action, energy must be injected from the outside. The method by which this **pumping** is done varies with the lasing medium used.

In gas [e.g., **helium-neon (He-Ne)**], ion (e.g., **argon** and **krypton**), and **metal vapor** [e.g., **helium-cadmium (He-Cd)**] lasers, an electric current is used to produce the **plasma** which serves as the lasing medium; in some of these lasers, particularly larger ion lasers, a magnetic field is used to confine the plasma to a region near the axis of the **plasma tube** in which the medium is contained.

Pulsed dye lasers, in which the lasing medium is a solution of fluorescent dye, and **pulsed solid-state** lasers, in which the lasing medium is a rod made of a material such as **ruby** or **yttrium aluminum garnet (YAG)**, are typically **optically pumped**; light from a flash lamp is often the pump energy source. In a **CW (continuous wave)**, as opposed to pulsed) dye or solid-state laser, the **CW output** of a **pump laser**, typically an ion laser in the former case and a diode laser in the latter, is used. **Diode lasers** themselves are pumped by input of electric current. Since a substantial power density of excitation is typically necessary to produce a population inversion, all laser types have a **threshold level** of pump power, below which laser action cannot be achieved.

Laser Efficiency: Your Mileage May Vary

The **efficiency** of lasers varies greatly. An argon ion laser emitting a watt or so of light typically consumes about ten kilowatts of electrical power while in operation; the overall efficiency of this system is therefore on the order of 0.01 percent. A CW dye laser, optically pumped with the 1-watt output of the argon laser, might emit a few hundred milliwatts; the efficiency of the dye laser, neglecting the power consumption of the pump laser, is typically 20-30 percent. Diode lasers are also relatively efficient.

Less efficient lasers are more likely to require forced-air or water cooling, particularly when high power outputs are needed; this increases their size, complexity, and cost. Efficiency is strongly dependent on the gain of the laser, which may vary substantially for different laser lines. In the example of the argon laser given above, the same power input that could produce 1 watt of visible output might produce only 10-20 mW of UV output; efficiency would then drop to 0.00001-0.00002 percent.

Mirrors and Prisms for Wavelength Selection

In media that support laser action at different wavelengths, the **gain** at different wavelengths varies,

profoundly affecting efficiency. When gain is low, higher mirror reflectivity is needed to maintain laser action, thus, less light can be allowed to pass through the output coupler mirror; the lower the transmission of this mirror, the lower the laser's output. Because of this, and because the combination of high reflectivity and controlled transmission of laser mirrors is achieved through the use of dielectric coatings, it is not generally possible to produce mirrors which will reflect and transmit appropriately in all of the spectral regions in which media such as argon and krypton exhibit laser action. It is thus generally necessary to use **interchangeable mirror sets**, each designed for emission over a wavelength range of no more than 100 nm. The mirrors with the broadest bandwidth are those typically installed in the krypton and argon-krypton ion lasers used for light shows, which allow simultaneous emission of light in the blue, green, yellow, and red spectral regions between 460 and 680 nm. In some instances, there is **competition** between two lasing processes during multiline operation; the power levels of the yellow and red lines in krypton lasers frequently exhibit seesaw behavior on this basis.

In medium and high power lasers, wavelength selection within the spectral range attainable with a single set of mirrors is generally done by insertion of a **Littrow prism** in the cavity between the mirrors. The **dispersion** of the prism results in light of different wavelengths being refracted at different angles on passage through the prism. At any given position of the prism, only a relatively narrow range of possible emission wavelengths will be reflected along the axis of the laser cavity between the high reflector and output coupler mirrors. Gain in this selected wavelength range will be sufficient to maintain laser action; gain at wavelengths above and below the selected range will not. The emission wavelength is changed by changing the orientation of the prism; this usually involves a vertical angular adjustment.

Dispersive elements other than prisms, e.g., **gratings**, can also be used for wavelength selection; it is also possible to insert an **optical filter** or another interference-based component, an **etalon**, in the laser cavity to restrict the range of emission wavelengths by reducing transmission, and, therefore, gain, outside of the desired narrow wavelength range. Although low power argon lasers of the type most commonly used in benchtop flow cytometers can, like their larger counterparts in sorters, be equipped with Littrow prisms, most are instead fitted with fixed narrow bandwidth mirrors that confine output to 488 nm. The He-Ne and diode lasers used for cytometry also do not make provision for the emission wavelength to be changed.

Brewster Windows for Polarized Output

In many types of lasers, **polarization** is introduced into the beam by putting windows between the medium and the end mirrors. The windows are placed at **Brewster's angle** (pp. 107-8) to the axis of the system. At this angle (about 57 degrees for glass), reflection from the window surface is minimized for light of one polarization, while a small

percentage of light of the perpendicular polarization is reflected out of the cavity. The slight difference in transmissions of the two polarizations is magnified many times by the feedback characteristic of the optical resonator structure, with the result that the laser output in a system with such **Brewster windows** is highly polarized, typically in a ratio of at least 500:1.

Laser Power Regulation: Current and Light Control

Ion lasers, small or large, can generally be operated in either a **current control mode** or a **light control mode**. In the current control mode, the laser power supply is regulated to deliver a constant current; if the mechanical and optical characteristics of the laser do not change during operation, light output remains constant. If things change, e.g., if a mirror becomes slightly misaligned, light output decreases even though power supply current remains the same. In the light control mode, power supply output is regulated by a feedback circuit that samples the energy in the beam and adjusts the laser current to maintain constant light output. This works well when the laser is emitting at a single wavelength. When emission of several lines occurs simultaneously, it is more difficult to keep power constant, particularly if the gains differ considerably and/or if there is competition between lines.

The air-cooled argon lasers in benchtop flow cytometers are operated in the light control mode; so are diode lasers, which are usually built with a light-sensing photodiode in the same package. In the case of diode lasers, the incorporation of a light control feedback loop into the power supply is almost essential to prevent the laser from frying itself when it is turned on. He-Ne and He-Cd lasers typically do not incorporate light control circuits.

Beam Profiles and Beam Quality

The ion lasers widely used for cytometry are usually operated in the radially symmetric TEM_{00} or Gaussian mode, discussed on p. 134. He-Cd lasers emitting at 325 nm, used as UV sources in low-power systems, often emit in the TEM_{01} , or “donut” mode, so-called because it produces a radially symmetric spot with a dark center. These and other modes were sketched in Figure 4-20 (p. 134). Actual Gaussian, “donut” and “yecchh” (loads of modes with nodes) intensity profiles, measured from He-Ne, CO_2 , and diode lasers, are illustrated in Figure 4-21.

The intensity profiles in Figure 4-21 are drawn as isometric “peak and valley” plots; the shading lines in this figure were in different colors in the original, reflecting the tendency of laser beam profiling software to use overkill (chromatic *and* isometric representation) in data display. You can find some of that in flow cytometry software, too.

The Gaussian profile of the He-Ne laser beam in Figure 4-21 is almost as smooth as a computer-generated curve. The “donut” from the CO_2 laser is closer to a Bundt cake than a bagel, suggesting that the beam is not of the highest quality. Then we have the diode laser, which looks really

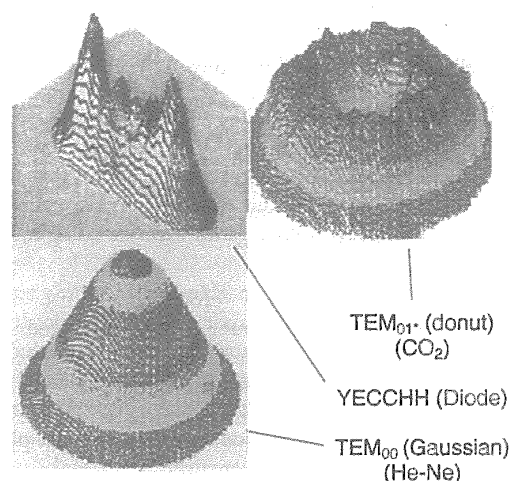


Figure 4-21. Beam intensity profiles of a CO_2 laser, a diode laser, and a He-Ne laser (courtesy of Excitech, Ltd.)

ugly, as most of them tend to do; the structure of diode lasers, as was mentioned on p. 135, is quite different from that of gas, ion, and most solid state lasers, and the ugly beam comes with the territory. But there is good news; it is possible, using fairly simple optics, to clean up a diode laser beam to a point at which it looks almost Gaussian, as shown in Figure 4-22.

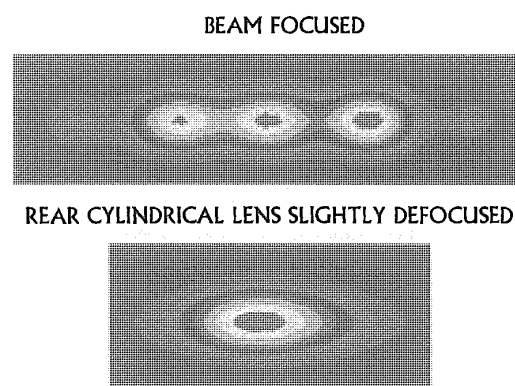


Figure 4-22. Profiles of the beam from a 397 nm (violet) laser diode, focused through crossed cylindrical lenses.

The displays of Figure 4-22 show chromatic plots of intensity at different points in the focused beam of a 397 nm violet diode laser. The top panel shows the profile obtained at the point of focus of both crossed cylindrical lenses; there are multiple modes similar to those that appear in the diode laser profile in Figure 4-21. The bottom panel of Figure 4-22 shows a much smoother, near Gaussian profile, achieved by defocusing the rear cylindrical lens slightly. We were shooting for this smooth beam profile in order to be able to

do fluorescence measurements with reasonably good precision; with only 4 mW of available power, it was unlikely that we'd get help lowering the CV's from bleaching and saturation effects. And our optical manipulation worked; we measured DNA content in DAPI-stained nuclei of cultured lymphoblasts, with a 1.7% CV for the G_0/G_1 peak²⁴⁴⁴.

Several manufacturers now offer violet, red, and infrared diode laser systems incorporating optics that produce a circular beam that is very close to Gaussian in its intensity profile. In Blue Sky Research's CircuLaserTM modules, this is accomplished by putting a high-N.A. (0.7) cylindrical microlens into the case housing the laser diode, capturing the entire beam and reducing its relatively large divergence in the vertical direction to match the smaller divergence in the horizontal direction. Although a circular beam can also be obtained using lenses external to the diode package, this generally involves the loss of more of the laser's output power.

A Gaussian beam should be focusable to a diffraction-limited spot; an approximation formula for calculating spot diameter was given on p. 130. Actually getting a spot of the size predicted by the formula would require both an ideal Gaussian beam and a lens that was diffraction limited, rather than limited by aberrations. There is also a throughput issue associated with focusing laser beams; to put it most simply, **the smaller the spot, the smaller the distance over which it stays small.**

Compared to just about anything else, the beam that comes out of a laser looks collimated; the narrow beam that emerges from your laser pointer makes a small spot on the screen on which your PowerPoint presentation is being projected. But, if you actually bothered to measure the spot diameters at the pointer output and the screen, you'd see that the spot on the screen was bigger. The beam coming out of the laser is diverging; it just isn't diverging very much.

In actuality, the output beam of a laser has been brought to a focus, or **waist**, somewhere near its point of exit from the laser package, courtesy of the laser mirrors. A typical beam diameter at this point is somewhere between 0.5 and 2.0 mm, and the beam divergence angle is usually less than 1 milliradian. If you put the beam through a **beam expander**, which works like a telescope in reverse, what comes out is a beam with a larger diameter and a smaller divergence. If you use the telescope in an observatory, you can produce a beam a few feet in diameter with a divergence so low that the beam will be less than 100 feet across when it is reflected from a mirror structure left by the Apollo astronauts on the moon in 1969. Ever since then, astrophysicists have been able to use the transit and return time of pulsed laser beams to measure the distance from the earth to the moon.

Biophysicists, or at least the cells they look at, need smaller beam diameters, and getting smaller beam diameters means putting up with higher beam divergences. If W_0 is the radius of the beam at the waist, and λ is the wavelength, the **divergence angle** $\theta_0 = \lambda/\pi W_0$. The **depth of focus**, or

confocal parameter, of a Gaussian beam, is defined as the distance between the points at which the beam area is twice the area at the waist. The depth of focus is usually expressed as $2z_0$; the quantity z_0 is known as the **Rayleigh range**, and is equal to $\pi W_0^2/\lambda$. The divergence and depth of focus formulas also apply to the axis dimensions (height and width) of elliptical beams; an elliptical beam diverges more rapidly in the plane of its shorter axis than in the plane of its longer axis.

It is generally impractical to attempt to use a laser beam to fill the field of view of a microscope objective; even if the illumination profile were a near-ideal Gaussian, intensity over the field would vary unacceptably. A bigger problem arises from the coherent nature of laser light; interference effects produce an intolerable "speckle" in the image. For this reason, practical use of laser light sources for cell imaging applications generally involves scanning the field of view with a relatively small-diameter beam.

Scanning laser cytometers, which produce low-resolution images, generally use focal spots 5 μm or larger in diameter, permitting uniform illumination through the thickness of a layer of cells on a slide. At 488 nm (0.488 μm), a 5 μm focal spot has a depth of focus of 80 μm ; a 10 μm focal spot has a depth of focus of 320 μm . A 2.5 μm spot, used at the highest magnifications in scanning laser cytometry, has a depth of focus of 20 μm .

In confocal and multiphoton confocal microscopy, it is advantageous to have spot diameters of 1 μm or less; this concentrates excitation light in a narrow layer of the specimen, improving resolution of fluorescence images. A 0.5 μm , 488 nm focal spot has a depth of focus of only 0.8 μm .

If we're trying to do a "slit-scan" (p. 51) of cells or particles to attempt to measure cell size (diameter) from pulse width, calculating the depth of focus, $2z_0$, won't really tell us what we need to know, because it gives us the points at which the beam area is twice the minimum area, meaning that the beam radius is $\sqrt{2}$ times W_0 , or $(1.414) W_0$. If we want the pulse width measurement to be precise, we need to be sure that the beam radius W is constant to within a small percentage at any point in the core stream. There is a formula that can help us find out. The beam radius $W(z)$ at a distance z from the beam waist is

$$W(z) = W_0 [1 + (z/z_0)^2]^{1/2}.$$

From this, we can calculate values of (z/z_0) , and, therefore, z , for which $W(z)$ is within any given percentage of W_0 . For example, if we want to keep variation to within 1 percent, we set $W(z) = (1.01)W_0$, and find that $z = 0.142z_0$. The corresponding values of z for 2 percent and 5 percent variation are, respectively, $0.201z_0$ and $0.32z_0$.

This tells us that it would not be very practical to use a beam 2 μm high at the waist for slit-scanning; even if we are willing to tolerate as much as 5 percent variation in beam height over the core, we'd have to work with a core diameter (and a particle diameter) of less than 4 μm . If we use a 5 μm

beam waist, the distance over which variation in beam height is less than 5 percent is just over 25 μm , which is a reasonably manageable number, meaning that it is larger than either a typical cell diameter or a typical core diameter.

The divergence angle for an ideal Gaussian beam focused to a given spot size can be calculated from the formula given on p. 137. We can also measure both the spot size and the divergence angle of an actual laser beam. For a given spot size, the ratio of the observed to the calculated divergence angle, called M^2 , serves as a measure of beam quality. A value of 1.2 or less is considered good. However, some manufacturers (and some knowledgeable users) put more stock in M^2 than do others.

Puttin' on My Top Hat?

While scanning and confocal microscopy put all of the energy in the laser beam through the cell of interest, most of the elliptical beam in a flow cytometer, by design, does not illuminate the core stream containing the cells. If we're not using a high power laser, the Gaussian beam profile in the TEM_{00} mode forces us to widen the focal spot to get uniform enough illumination over the width of the core stream to permit high precision in fluorescence measurement (p. 131). This means that much of the beam is wasted. We could get much more usable energy through the cells if we had a "top hat" beam profile with steep shoulders and a flat top. Unfortunately, it isn't that easy to get such a profile.

There are a few ways to flatten the top of the beam somewhat; one can widen the bore of the lasing medium to produce either a mixture of TEM_{00} and TEM_{01} modes or true multimode output, which adds in some of the higher, lobed modes as well. Alternatively, the beam can be put through an aperture and defocused slightly. This permits use of smaller focal spots, wasting less of the energy in the beam. It is possible, although tricky, to get a somewhat flat-topped profile out of a diode laser; as a general rule, however, you also end up with ugly pulse shapes, which can cause problems in signal processing.

The newest approach to getting uniform illumination from lasers involves the use of arrays of microlenses, which produce arrays of overlapping, small, focused beams. The profiles I have seen generated using this technique have fairly steep shoulders, but there is a lot of intensity variation across the top of the top hat. I'm not sure how much of this can be smoothed out, and I'm also not sure how much of the beam energy is lost in playing this game. I'm sure we'll find out in the next few years.

Harmonic Generation and Modulation

The light produced by lasers, like all other light, has associated electric and magnetic fields, and, since the radiance of a laser beam is substantially higher than that of an incandescent source, the associated electric field intensity may be high enough to produce nonlinear responses in certain materials. One notable application of these phenomena is in **harmonic generation**, in which nonlinear

effects in crystals result in generation of light at two or more times the frequency of the incident light. **Second harmonic generation**, or **frequency doubling** of the 1064 nm YAG laser line, for example, produces 532 nm; **third harmonic generation**, or **frequency tripling**, produces 355 nm. The same nonlinear crystals may also be used to produce output at the sum and/or difference of the frequencies of two incident beams; in the case of YAG lasers, frequency summing can produce emission at 473 nm. Although a reasonably broad range of crystalline materials capable of harmonic generation is available, the range of wavelengths at which continuous (CW) output can be obtained is restricted.

It is sometimes desirable to vary the output power of a laser more rapidly than can be accomplished by adjustments to the power supply. **Modulation** at frequencies up to several hundred megahertz is possible using **electro-optic modulators**, which incorporate crystals that change their refractive index as a function of an applied voltage. **Acousto-optic modulators**, which use sound waves to produce changes in density that affect the light transmission characteristics of a substrate, work at lower frequencies, generally below 100 MHz. A light sensor and a modulator, connected by a feedback circuit, placed in the output path of a laser, can be used as a "noise eater," providing light regulated output, albeit at a relatively high price.

Lasers Used and Usable in Cytometry

NOTE: Discrete laser emission wavelengths are given to fractions of a nanometer in Table 4-2, on the next page; in most of the rest of the book, I'll stick to whole numbers with inconsistent rounding off, which is what everybody does most of the time.

Argon and Krypton Ion Lasers

The most popular lasers for fluorescence flow cytometry are **argon ion lasers**; they are usually operated at 488 nm, a wavelength useful for excitation of fluorescein, phycoerythrin and its tandem conjugates, propidium iodide or ethidium bromide, acridine orange, pyronin Y, various rhodamine and cyanine dyes, anthracycline drugs such as adriamycin, and various GFP variants. Argon ion lasers (and **krypton ion lasers**, which are less friendly but work on the same principles) are available from several manufacturers. At present, Coherent seems to be the most popular source for the large, water-cooled systems; smaller, air-cooled lasers are likely to come from Melles Griot (which absorbed Omnicrome), Spectra-Physics, and Uniphase.

The lasing medium in an ion laser is a plasma, which, in the larger lasers, is confined in a strong magnetic field generated by a solenoid. A high-voltage pulse is used to ionize the gas to start the plasma, a procedure similar to that used in starting an arc lamp. Ion lasers require a high current to maintain the plasma discharge; the bigger ones also put a high current through the solenoid, generally increasing power consumption enough to necessitate water-cooling.

Ar Ion	Kr Ion	He-Cd	He-Ne	Solid State
275.4				
300.3				
302.4				
305.5				
		325.0		
333.6				
	337.4			
	350.7			
351.1				
		354.5		355 (Nd:YAG x 3)
	356.4			
363.8				
	406.7			
	413.1			
	415.4			
				430 (Cr:LiSAF)
		441.6		
454.5				
				457 (Nd:YVO ₄ x 2)
457.9				
				460 (semi)
	468.0			
465.8				
472.7				
	476.2			
476.5				
	482.5			
488.0				488.0 (semi)
496.5				
501.7				
514.5				514.5 (Yb:YAG x 2)
	520.8			
528.7				
	530.9			
				532 (Nd:YAG x 2)
		533.7		
		537.8		
			543.5	
	568.2			
			594.1	
			611.9	
			632.8	
		635.5		
		636.0		
	647.1			
	676.4			
	752.5			
	799.3			

Table 4-2. Emission wavelengths of argon and krypton ion, helium-cadmium, helium-neon, and solid-state lasers that emit at discrete wavelengths. Diode, dye, and some solid-state lasers are tunable over ranges of wavelengths. See text for details.

In addition to the strong blue-green and green lines at 488.0 and 514.5 nm, argon ion lasers emit at 454.5, 457.9

(violet-blue), 465.8, (blue) 472.7, 476.5, (blue-green) 496.5, and 501.7 (green) nm. Emission can also be obtained in the ultraviolet at 333.6, 351.1 and 363.8 nm, and in the green at 528.7 nm, using specially coated mirrors. In addition, the largest high power argon ion lasers can produce some deep ultraviolet lines between 275.4 and 305.5 nm. An infrared argon laser line at 1090 nm is not likely to be useful for cytometry in the near term.

Ion lasers are made out of relatively esoteric materials and are complex enough in their construction so that the relationship between price and power output is highly nonlinear. An argon laser that puts out 25 mW at 488 nm costs around \$6,000; for less than ten times that amount, you can almost certainly get more than fifty times the power, from whichever manufacturer you like. The bigger laser is not necessarily a better investment. The best reason to use an argon ion laser is a heavy commitment to measurements of weak fluorescence, e.g. immunofluorescence, adriamycin uptake, etc.; the 15-25 mW air-cooled argon lasers now most common in flow cytometers, which have plasma tube lifetimes of 6,000 hours or more, are more than adequate for the job in benchtop systems with efficient light collection optics. Even for larger instruments, with less efficient light collection, one could still consider using an air-cooled argon laser; systems with power outputs as high as a few hundred mW are available, generally for less than \$10,000.

Despite their many glamorous aspects, large ion lasers are basically big, expensive light bulbs. The hotter you run them, the faster they burn out. In order to be capable of putting out 100 mW in the UV, an argon laser generally has to be capable of 5-6 watts "all lines" power in the visible, and 1.3-2.0 watts at 488 nm. To get 100 mW of UV out of it, you need to run it near maximum rated current, and you're apt to need a plasma tube replacement at least once every couple of years. Run the same laser, or the next smaller model, at 200 mW output at 488 nm and you're close to idle current; the plasma tube is apt to last for several years. If you build your own instruments, you can use a beamsplitter to get two 100 mW 488 nm beams, and another one to get two 50 mW beams from one of the 100 mW beams, and run three flow cytometers from a single argon laser (as in "Cerberus"), if you have a large optical table on which to put it¹⁰⁵. However, it may make more sense to use smaller, air-cooled argon lasers at a ratio of one laser per cytometer.

There are several reasons why you need big argon lasers to get UV output. You need high current, because the UV lines are emitted from a higher ionization state of argon than is needed to emit the visible wavelengths, and you need large size, because gain is relatively low (low-gain lines benefit from longer cavity lengths). So, don't expect that there will be any little air-cooled UV argon lasers for \$6,000, soon or ever.

It is possible⁶ to produce mirrors for big argon ion lasers that permit simultaneous emission at 351/363 nm in the ultraviolet and at 488 nm and other visible wavelengths; this provides the dual-wavelength source needed, for example, to

do simultaneous analysis of DNA, using Hoechst 33342, and surface antigens, using antibodies labeled with fluorescein, phycoerythrin, and phycoerythrin-based tandem conjugates. In order to emit in the ultraviolet, the laser has to be run at very high current; since the laser is much more efficient at 488 nm than in the UV, power output at 488 nm is quite high. If the sensor in the light output regulator circuit responds to both UV and visible light, or to light at 488 nm alone, there may be considerable fluctuation in UV power output. If the sensor's optical bandwidth is restricted so that 488 nm light is blocked, the sensor, and the regulation electronics, then respond to fluctuations in UV power output. The relatively large changes in current that may be necessary to keep UV output stable are then likely to result in large fluctuations in power output at 488 nm. It is probably preferable to use separate sources for the UV and the visible wavelengths, and, at present, I would favor helium-cadmium over ion lasers as UV sources.

Coherent's "Enterprise" argon laser system is capable of simultaneous output of over 100 mW at 488 nm and a few dozen mW in the UV. This laser is water-cooled by a closed, recirculating system, and is fairly widely used in cell sorters. Since many localities now have environmental regulations requiring that conventional water-cooled lasers be equipped with recirculating systems, this makes some sense; personally, I'd rather stick to air-cooled lasers. My name is Howard, and I've been dry for over fourteen years...

Why, you might wonder, did so many people spend so much money a few years back buying instruments equipped with large argon and krypton lasers? The krypton laser is a rude beast whose hour came round at last for light shows and two-color immunofluorescence measurements. Unlike argon lasers, which have high gain lines and low gain lines, krypton lasers have low gain lines and lower gain lines. Take the same size laser that would put out five watts in all-visible-lines mode with an argon tube, put in a krypton tube instead, and you're lucky if you get one watt of visible light out. But what visible light! Blue-green, at 468.0, 476.2, and 482.5 nm, green, at 520.8 and 530.9 nm, yellow, at 568.2 nm, and red, at 647.1 (the strongest krypton line) and 676.4 nm, all at once, explaining the popularity of krypton lasers for light shows, and also explaining why the multicolor shows tend to be indoors; the outdoor spectacles need the more powerful blue-green and green argon lasers. With different mirror sets, and appropriate adjustment of solenoid magnetic fields, krypton lasers can also emit multiple ultraviolet (337.4/350.7/356.4 nm), violet (406.7/413.1/415.4 nm) and infrared (752.5/799.3 nm) lines.

If all the colors of light are the good news, the bad news is that the optimum values for gas pressure and solenoid magnetic field for krypton laser operation are different for different lines; since the gains at all lines are low, these parameters need to be controlled. Then, if the alignment is really good and the optics are really clean, you'll get laser output. You can literally spit on the mirrors of an argon laser running at 488 nm and still get laser output, albeit less than

before. Krypton laser (and UV argon laser) mirrors need to be squeaky clean or you might as well not bother trying to get the laser to run. If the optical alignment isn't near-perfect in the visible, as indicated by maximum all-lines power output near the manufacturer's specs, you'll never get an ion laser to work in the ultraviolet. Low gain at all lines means that krypton laser plasma tubes run hotter and die sooner than argon laser tubes. Last but not least, there is, as I mentioned previously, **competition** among the visible krypton lines; if, for example, you attempt to run yellow and red simultaneously, in the light regulated mode, you'll probably find that, while total power stays constant, the yellow and red power outputs alternately go up and down.

Krypton lasers got the call for flow cytometry primarily because, way back in 1979, when monoclonal reagents were beginning to make staining with two antibodies practical, but before there were phycobiliprotein labels, immunologists found it difficult to do studies of two surface antigens with argon lasers. Fluorescein was and is the obvious label for demonstrating one antigen. The dye most widely used as a second label was tetramethylrhodamine, which is poorly excited at 488 nm, absorbing maximally near 550 nm.

While it was possible to resolve signals from antibodies conjugated with fluorescein and tetramethylrhodamine by using the strong line at 515 nm for excitation, provided one went to offbeat optical filter combinations and introduced an electronic compensation circuit to take out the crosstalk between the signals¹¹⁵, this didn't always work well.

XRITC and Texas Red^{116,117}, reactive derivatives of rhodamine 101 that have absorption and emission maxima at wavelengths 30 or 40 nm longer than those of tetramethylrhodamine, cannot be excited satisfactorily by any of the argon laser lines, but can be used with either a krypton laser at 568 nm or a CW dye laser operating near the absorption maximum of rhodamine 101 at about 590 nm¹¹⁸. In 1979, when these dyes and the related instrument modifications were unveiled, it seemed easier to use krypton lasers than dye lasers; thus, these were pushed by the manufacturers and you simply had to have a krypton laser in order to keep up to the state of the art, even if you never actually did two-color immunofluorescence.

I once put a prism in an all-visible-lines or "light show mode" krypton laser beam to break the beam up into seven beams of different colors, and directed them all through a flow cytometer, just to preserve the Cytomat/Cytomutt tradition of working with more beams than anybody else had. I actually could make measurements in three or four of the beams; unfortunately, power output in individual beams jumped all over the place. The krypton laser may be used to excite porphyrin fluorescence in blood cells (C. Stewart, personal communication) or in tumor cells labeled with porphyrin derivatives, using the violet lines. I have used the blue krypton lines to excite fluorescein and propidium, but only when my argon laser was out for repairs.

Keeping the optics of an ion laser clean and in alignment is a challenge best met by following the manufacturer's

instructions. Mirrors and Brewster windows are generally cleaned with methanol; the slightest contamination with grease, from fingers or elsewhere, will cause your solvent to deposit a film of crud on optics, which can cripple or kill your laser output. Acetone is a good cleaner for some mirrors and destroys others; don't use it unless the manufacturer says so. Electronic or HPLC grade solvents are generally free enough of contaminants to be safe to use for cleaning optics.

Changing mirrors and getting the laser to lase again is a tedious procedure that must be done a little bit differently for each manufacturer's lasers; it is learned by doing and remains something of a black art to even experienced laser jockeys. The first time you put the UV optics in by yourself, you should set aside a day. When you become confident you can do it in one hour, you will budget one hour on a busy day and find the task takes three hours. This is another good reason for using smaller, simpler, single wavelength lasers with sealed optics, and, as of 2002, the major cytometer manufacturers have gotten this message.

Dye Lasers

With the advent of phycoerythrin^{6,114} as a second label, it became practical to do two-color immunofluorescence measurements with single-beam excitation at 488 nm, using fluorescein- and phycoerythrin-labeled antibodies. The krypton laser was no longer necessary for two-color immunofluorescence, and was not useful for three-color measurements adding Texas red or XRITC as a third label because the 568 nm yellow line from the krypton laser, while it could excite the third dye, lay smack in the middle of the emission spectrum of phycoerythrin. At this point, interest turned to CW dye lasers as excitation sources.

The lasing medium in a dye laser is a fluorescent dye, usually dissolved in an organic solvent such as ethanol or ethylene glycol. Which dye is actually used depends on the wavelengths at which operation is desired; there are dyes now available for use with blue-green/green (457-515 nm) argon ion pump lasers that permit operation at wavelengths extending from 540 to over 900 nm. While ion lasers lase only at a few discrete wavelengths, a dye laser can emit at any of a wide range of wavelengths in the emission spectrum of the dye used. Wavelength selection is usually done using a wedge, grating, or filter rather than a prism.

The dye in a CW dye laser circulates through a nozzle, producing a flat-walled stream; continuous circulation is necessary to minimize bleaching of the dye and to allow for cooling by a heat exchanger. CW dye lasers require minimal electrical power, most of which is used to operate the circulator pump. In pulsed dye lasers, especially those with low duty cycles, the medium does not either heat or bleach very rapidly, so the dye need not be circulated; a pulsed dye laser has been made using Jell-O as the lasing medium.

Some dyes bleach faster than others, but they all need to be replaced after a few months' operation. This, and the fact that CW dye lasers are relatively hard to keep tweaked, and do not maintain output power as stably as do ion lasers,

originally restricted their use as sources for flow cytometry to people who do relatively esoteric things with flow systems¹¹⁹ and will put up with idiosyncratic apparatus. Improvements in dye laser design, and a perceived market for sources usable for excitation of Texas red or XRITC and allophycocyanin (which, while its maximum absorption is in the red, is reasonably well excited near 600 nm), led manufacturers to promote and users to buy flow cytometers with dye laser sources. Using a single argon laser as a 488 nm source and as a pump laser allowed multicolor immunofluorescence work to be done using only one large laser.

The dye most commonly employed as a lasing medium in flow cytometry is rhodamine 6G. The threshold power required from the pump laser to achieve output from a rhodamine 6G dye laser is usually about 700 mW. With this power input, over 100 mW of light can typically be obtained from rhodamine 6G at wavelengths between about 570 and 620 nm. For Texas red excitation, the dye laser is usually operated at 594 nm; for excitation of Texas red and allophycocyanin, a longer wavelength, 605-610 nm, is often used. The disadvantage of using dye lasers in dual-laser systems lies in the resulting inability to use either the dye laser or the pump laser as a UV source. In principle, one could get around this using dual-wavelength UV/visible mirrors in the argon pump laser, but this solution requires great technical skill on the part of the user and also is apt to involve frequent plasma tube replacements. In most cases, there are better alternatives. If you must use a dye laser, you're now probably best off using a 532 nm frequency-doubled YAG laser, which runs on house current and needs no water cooling, as a pump, rather than a big argon laser.

Helium-Neon Lasers

If you want to meet a nice laser, try a **helium-neon (He-Ne) laser**. They plug into the wall, they don't need water cooling, they're relatively small and very stable, the mirrors aren't even adjustable, and the plasma tubes last for years. Besides, they're relatively cheap. The most common He-Ne lasers emit red light at 632.8 nm; they are available with power outputs ranging from less than 1 mW to about 50 mW; other visible wavelengths at which He-Ne lasers are now available include 543.5 (green), 594.1 (yellow), and 611.9 (orange-red) nm.

Red He-Ne lasers were used at Los Alamos, beginning in the late 1960's, for scatter measurements at various^{110,111} angles, and, in the 1970's, were incorporated into Ortho's instruments and Technicon's hematology systems for scatter and axial extinction measurements. At Los Alamos, they used a 5 mW laser; the 0.8 mW He-Ne supplied with the Ortho systems provided less than optimal fluorescence excitation, but it had very low noise and was therefore quite good for extinction measurements. I started using 5-7 mW red He-Ne lasers around 1980, to excite fluorescence of cyanine dye probes of membrane potential¹¹², for RNA measurements using oxazine 1¹¹³, and for immunofluorescence excitation, using antibodies conjugated to

allophycocyanin^{8,114}. Mike Loken's group at B-D⁶⁴⁴ and Bob Hoffman's at Ortho^{645,646} examined red He-Ne lasers as sources for one- and two-color immunofluorescence measurements using phycocyanin and/or allophycocyanin as labels; there was general agreement that red He-Ne lasers were well suited for these applications, and they began to be offered in commercial flow cytometers. In recent years, some manufacturers have substituted diode lasers emitting in the 635 nm region for red He-Ne lasers.

Until 1985, 633 nm was the only visible wavelength available from commercial He-Ne lasers, and the 633 nm He-Ne lasers have been more widely used than the 543 nm, 594 nm, and 611 nm varieties. The 543 nm green He-Ne laser is useful for excitation of immunofluorescence from antibodies labeled with phycoerythrin or its tandem conjugates, and can also be used to excite DNA stains such as propidium²⁴⁷¹. However, the gain of the green He-Ne line is quite low, and available power levels remain below 2 mW. This power level is sufficient to permit immunofluorescence measurement in an efficiently designed flow cytometer; B-D uses a green He-Ne laser in the FACSCount, a dedicated instrument for counting CD4- and CD8-positive T lymphocytes based upon staining with phycoerythrin and tandem-conjugate or PerCP-labeled antibodies²⁴⁷².

The 594 nm yellow He-Ne laser is usable for excitation of Texas red and of the Texas red/allophycocyanin combination; it operates at the same wavelength at which dye lasers are often used for these purposes. Maximum power output is still only about 2 mW, adequate for immunofluorescence work if a well-designed optical system is used. The major disadvantage of the 594 nm laser is the proximity of its lasing wavelength to the emission region of phycoerythrin; the availability of large numbers of labels excitable by 488 nm and 633 or 635 nm lasers has decreased demand for Texas red and other yellow-excited labels.

Helium-Cadmium and Helium-Selenium Lasers

Helium-cadmium (He-Cd) lasers, which can emit 5-200 mW in the blue (441.6 nm) and 1-100 mW in the UV (325.0 nm), depending on size, are relatively practical sources for flow cytometry. Lower power output in the UV at 354 nm is also available in some lasers. Like He-Ne lasers, He-Cd lasers plug into the wall and do not require water cooling; they need few or no adjustments and have relatively long plasma tube lifetimes. They are also cheaper than most ion lasers, but not by much. The Block Cytomat instruments used a single He-Cd laser with mirrors designed to permit simultaneous emission at 325 and 441 nm. Similar dual-wavelength models are now commercially available; there is little or no competition between the UV and blue wavelengths. The lasing medium is cadmium vapor; the pressure of both it and helium and the temperature of the medium must be carefully controlled to assure stable operation. Melles Griot has absorbed Liconix and Omnicrome, the American He-Cd manufacturers; Kimmon, in Japan, also makes He-Cd lasers.

I have used the 441 nm blue He-Cd line (at 10-100 mW) for ratiometric intracellular pH measurement using fluorescein derivatives⁸ and for excitation of the chromomycin family of DNA stains (olivomycin, chromomycin A₃, and mithramycin), and the 325 nm UV line (at 1-35 mW) for excitation of the fluorescence of the DNA stains DAPI and Hoechst 33342 and of the calcium probes quin-2 and indo-1. I have found that 10 mW at 325 nm and 40 mW at 441 nm produce strong fluorescence signals from bacteria stained with a mixture of DAPI or Hoechst 33342 and mithramycin or olivomycin, and I¹¹³³ and others¹¹³⁴⁻⁵ have found similar or slightly higher power levels usable for chromosome analysis, using Hoechst 33342 or DAPI and chromomycin or mithramycin. Separate UV and blue He-Cd lasers or a single dual-wavelength laser should also work for measurement of BrUdR incorporation by a relatively simple, antibody-free staining technique developed at Los Alamos⁶⁴⁷, using Hoechst 33342 and mithramycin; this is discussed in detail on p. 456.

He-Cd laser technology was originally pursued because a hollow cathode design, different from that commonly used, permits simultaneous emission of UV or blue, green (533.7 and 537.8 nm) and red (635.5 and 636.0 nm) beams. The "white-light" multiline He-Cd laser was intended for the graphics arts industry. Such lasers are now available from Cooke; their power levels are in the 10 mW range for each of the spectral regions. On the plus side, they are less noisy than conventional He-Cd lasers, but their size and cost are disadvantages; these days, combinations of diode and solid-state lasers may be a better bet.

The major disadvantage of He-Cd lasers is optical noise (light output fluctuations) at frequencies around 300 kHz; as plasma tubes age, helium pressure tends to build up, increasing the amplitude of these fluctuations and decreasing measurement precision. I will have more to say about this, and about light source noise in general, a little later on.

Helium-selenium lasers represent another type of metal vapor laser, similar to He-Cd lasers in construction but less tractable and with much shorter lifetimes (I heard of them dying after a few days in operation). They emit many lines ranging from blue-green through green, yellow, orange, and red, usually at very low power. They have not, as far as I know, been looked at as light sources for flow cytometry. I stopped trying to lay hands on one when the green and yellow He-Ne lasers first appeared.

Diode Lasers: Red, Infrared, Violet, and UV

Diode lasers began to tantalize flow cytometer designers at about the time commercial CD players, which were the first large-volume commercial products to incorporate diode lasers, appeared on the market. However, until the late 1980's, practical CW diode lasers had only been made to work in the near infrared (IR) (780 nm and above), and there were few fluorescent probes suitable for excitation at the available laser wavelengths.

Diode lasers are, like transistors, made of materials classed as **semiconductors**. The light emission from semiconductors is not from excited atoms or ions, as is the case in ion, He-Ne, and He-Cd lasers, and not, strictly speaking, from excited molecules, as is the case in dye lasers. The electrons that are excited in a diode laser are "free" in a crystalline material. Such "free electrons" also occur in metals; the nuclei in a metal are packed relatively close together, and the electrons in the outermost shells are not tightly held by any given nucleus, and may be excited from the so-called "valence band" to the so-called "conduction band" by ambient thermal energy. Energy transfer among electrons in a metal can occur fairly readily; this is what makes metals good conductors of electricity.

Semiconductors are so named because, though they do not conduct electricity well when in an unperturbed state, they may become conductive in the presence of an applied electric field or of incident light. Their electronic structure differs critically from that of metals in that there is a substantial energy difference, or "bandgap," between the valence and conduction bands, with almost all of the electrons lying in the valence band under normal conditions. The application of an electrical current to an appropriately configured semiconductor can result in light emission as electrons relax from the conduction to the valence band; this type of spontaneous emission is what occurs in light emitting diodes (LEDs).

A laser diode is basically an LED with its geometry tailored to provide a resonator structure that will support stimulated emission. The active regions of diode lasers typically have dimensions on the order of a few micrometers; they use either polished facets on the semiconductor material itself or adjacent structures of differing refractive index to perform the function of the mirrors used in larger lasers. Because the efficiency of diode lasers is extremely high, typically on the order of 20-30 percent, high reflectivity is not needed. Diode lasers are much less expensive than any other type of laser, because hundreds of lasers, if not more, can be produced from slices of a single semiconductor wafer. The down side of this is that small differences in semiconductor composition and dimensions of the finished chips affect the output wavelength; wavelength variation also occurs with changes in temperature. That's why Table 4-2 doesn't have a column listing diode laser emission wavelengths to a tenth of a nanometer.

The first practical diode lasers were, and many diode lasers still are, made of gallium aluminum arsenide (GaAlAs); the emission wavelength, other things being equal, is varied by changing the ratio of gallium to aluminum in the semiconductor material. The emission wavelengths theoretically achievable with GaAlAs lasers range from about 650 nm, at which point the material is almost pure AlAs, to about 900 nm, at which point the material is almost pure GaAs. However, GaAlAs lasers that emit below 750 nm are typically unstable, and immolate themselves within minutes to hours. The shortest

wavelength GaAlAs lasers available in quantity emit at 750-780 nm; millions of them go into CD players, laser printers, and CD-ROM drives, and they cost at most a few dollars.

Gallium indium phosphide (GaInP) lasers go down to around 670 nm, providing up to 50 mW emission from small devices and hundreds of milliwatts from larger ones and arrays. The 635-640 nm diodes now used as alternatives to 633 nm He-Ne lasers in cytometric applications are made of aluminum gallium indium phosphide (AlGaInP); they are now available with output powers ranging from a few hundred microwatts to 35 mW.

The problem with long wavelength sources in flow cytometry stems primarily from the fact that most of the fluorescent dyes that are of any use as probes for cellular parameters require excitation at or below 650 nm. Indeed, many of the dyes usable with red excitation are, or are closely related to, substances long known and used as stains for transmitted light microscopy. It would probably be at least as difficult to devise new red-excited stains, especially specific ones, as to develop new types of diode lasers. The difference between 635 nm operation and 670 nm operation is, therefore highly significant.

Before 635 nm diode lasers became available, I managed to do DNA content analysis using rhodamine 800⁷³⁷ with a 670 nm diode laser as a light source; this wavelength is also usable for excitation of dibenzodicyanobenzene dyes, used either alone as membrane potential probes or in reactive form (e.g., Cy5.5) as antibody labels, and for some aluminum phthalocyanine dyes which have also been investigated as antibody labels.

Working at 635 nm adds rhodamine 700, oxazine 750, Molecular Probes' TOTO-3 and TO-PRO-3 and some dyes of the SYTO series, and DRAQ5²³⁵⁸⁻⁹ to the list of DNA dyes, Cy5 and allophycocyanin and its tandem conjugates to the antibody labels, oxazine 1 for RNA, and assorted cyanine and oxonol membrane potential probes. Doornbos et al¹¹³⁶ were first to report using a 635 nm diode laser for DNA measurements with TO-PRO-3 and immunofluorescence measurements with allophycocyanin.

In an earlier paper, Doornbos et al¹¹³⁷ discussed the use of 670 nm and 780 nm diode lasers as light sources for forward and right angle scatter measurements; they were able to discriminate lymphocytes, monocytes, and granulocytes in flow cytometers using either wavelength. However, CV's of fluorescence measurements were disappointingly large. My guess is that this may have had something to do with the typically ugly beam intensity profiles of diode lasers.

At this point, those of us who build flow cytometers are a lot farther along the learning curve; general- and special-purpose flow cytometers using 635 nm diode lasers, alone or in combination with other light sources, are now available from a number of manufacturers. There is a 635 nm diode laser in the lunchbox-sized OptoFlow Microcyte (Figure 1-25, p. 58), which is very likely the smallest production model flow cytometer, and you can put one on the Cytomation MoFlo sorter (Figure 1-24, p. 58), which is

probably the largest. And, for many routine applications, e.g., counting CD4⁺ T cells, an instrument with only a red diode laser source can do just as good a job as a conventional benchtop flow cytometer with a 488 nm laser²⁴⁴⁷.

Diode lasers are used for laser printing and for compact disc recording and playback; in these applications, achieving a smaller focal spot size allows more information to be stored in and/or retrieved from the same area, either as more dots per inch on a page or, more importantly, as more bits per unit area on a disc. Recordable and CDs, written and read with 780 nm diode lasers, top out at around 700 MB per disc. A CD "burner" costs more than a reader because it needs a higher power diode laser to "burn" the CD by photobleaching spots in an IR-absorbing dye layer.

Using a shorter wavelength (650-660 nm) red diode laser instead of an IR laser allows a DVD to store 5.7 GB on a side. Getting 30 GB worth of data, music, or (where the money is) video onto one side of a disc would require a "blue" laser (matching some of the video). This prospect attracted the interest of some very large companies, with proportionally deep pockets, but the company that first succeeded was a smaller one, Nichia, in Japan, where Shuji Nakamura succeeded in making the first "blue" laser diodes in the late 1990's²⁴⁴⁸. These lasers are now available with emission in the range from 370 nm (UV) to 445 nm (blue-violet); while the newer devices at the far ends of the range still have short operating lifetimes (perhaps 3,000 hr), the 5 mW and 30 mW diodes with output between 395 and 415 nm, which have been in production for some time, are quoted as having lifetimes in excess of 15,000 hr.

Violet laser diodes are good excitation sources for a lot of materials that otherwise would require a krypton laser for excitation. The list²⁴⁴⁴ includes Molecular Probes' labels Cascade Blue and Cascade Yellow, monobromo- and monochlorobimane, both used for detection of intracellular glutathione¹⁶⁵³, ECFP, the cyan-fluorescent reporter protein, and the DNA dyes mithramycin and chromomycin A₃. Laser diodes operating at the short (370-400 nm) end of the range are effective excitation sources for DAPI (38% of maximum excitation at 395 nm) and usable with the Hoechst dyes; they cannot be used with the calcium probe indo-1, but it may be possible to synthesize a similar calcium probe that would work with 370 nm excitation.

Diodes emitting at 405-425 nm could be used to measure intrinsic absorption and fluorescence of porphyrins in cells. The strong absorption of hemoglobin in erythrocytes in this wavelength region (the Soret band) influences the cells' light scattering properties, enabling discrimination of leukocytes from erythrocytes in dilute, unstained whole blood²⁴⁴⁹; zinc protoporphyrin fluorescence in erythrocytes may provide clinically relevant information in cases of iron-deficiency anemia and lead poisoning¹²⁶⁶. Violet diode lasers could be used in practical clinical instruments; violet krypton lasers cannot.

When I wrote the last edition of this book, in 1994, 635 nm laser diodes were limited in power to about 5 mW and

the diodes themselves cost several hundred dollars; in 2002, a packaged 635 nm, 35 mW laser system with a power regulator, temperature control to increase output stability, and beam circularizing optics costs less than a 633 nm He-Ne laser with the same power output. Although violet diode laser systems available now cost several thousand dollars, the major cost is the cost of the diodes, which can't stay high if millions of them are made for optical storage devices.

I mentioned the relative variability of diode laser emission wavelengths; this has advantages and disadvantages. Small variations in device dimensions and composition result in manufacturers' inventories including lasers with a range of emission wavelengths; for a price, the buyer (meaning the cytometer manufacturer or the laser system manufacturer; making a diode laser system is not a trick to try at home) can specify a selected wavelength range. If you're buying a diode laser system, you can also specify a wavelength range; again, it will cost you.

A red He-Ne laser will emit at 632.8 nm, not 632.85 or 632.75 nm, for all of its useful life. When cytometer manufacturers buy red diodes, they typically specify an emission range no narrower than, say, 635 to 640 nm; a diode in your instrument could emit anywhere in that range. If you measure fluorescence at 660 nm, your detector filter probably has a higher transmission at 640 nm than at 635 nm. More stray laser light leaking through the filter means higher fluorescence background, so which end of the range the diode emits at will affect measurement sensitivity.

The emission wavelength of a diode laser varies with temperature; this doesn't happen uniformly, but, rather, in jumps, and each wavelength change represents a different longitudinal mode of the laser. The resultant "**mode hopping**" may affect the stability of the intensity profile, even though the light control feedback circuitry needed to keep the laser from burning itself out keeps the total power output relatively constant. The best way to eliminate this problem is to control the temperature of the diode; this is absolutely essential for the violet diodes, and becoming more common in red diode systems designed for critical applications such as cytometry.

A diode laser is basically an LED, and it will emit a small amount of incoherent light, sometimes referred to as **LED glow**. You may not be all aglow yourself if you encounter this light, which is usually at wavelengths longer than the laser wavelength, meaning it may be able to get through the filters on one or more of your fluorescence detectors. Because it is incoherent, LED glow diverges fairly rapidly, but, if the laser diode is close to the flow chamber, enough LED glow may get in and be scattered from cells to show up as increased fluorescence background. The workaround for this problem is fairly simple; a bandpass or short pass filter is mounted between the laser diode and the flow chamber.

As I mentioned previously, diode lasers must be operated in a light control mode; as a result, their optical noise levels are typically substantially lower than those of most ion or He-Ne lasers. Diode lasers therefore offer an advantage

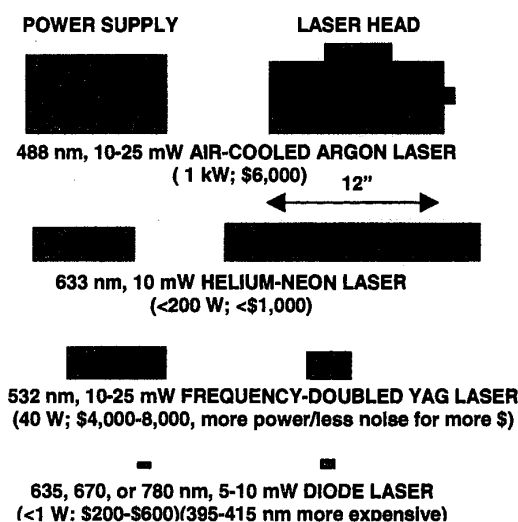


Figure 4-23. Sizes, power requirements, and approximate costs of some smaller lasers used for cytometry.

over these other sources for extinction measurements, which are more sensitive to optical noise than are scatter and fluorescence measurements. When putting a diode laser into an optical system, it is critical to position optical elements in such a way as to avoid reflecting light back into the laser, which can interfere with the operation of the light control circuitry and result in fluctuations in power output.

As Figure 4-23 shows, red diode lasers are 1/1000 the size (volume) of, consume 1/1000 as much power as, and cost less than 1/10 as much as, air-cooled argon lasers. This commends the diodes for use in portable equipment, and in applications and locations where cost and power consumption are an issue (e.g., CD4⁺ cell counting in Africa, or (at least in mid-2001) California²⁴⁴⁷). Violet diode lasers, and solid-state lasers, are also small and energy-efficient, if not inexpensive; the violet lasers should get considerably cheaper, and some of the solid-state sources may (see below). If it is ever possible to produce flow or image cytometers for a tenth of today's prices, the instruments will be likely to use laser diodes, or even LED's, as light sources.

Solid-State Lasers: Like, YAG Me!

Neodymium-YAG (yttrium aluminum garnet) lasers, in which the lasing medium is a solid rod of crystalline material pumped by a flashlamp or a diode laser, can produce power outputs of tens of watts at 1064 nm. Doubling or tripling the output of a diode-pumped YAG laser, using a crystal within the laser cavity, can yield green light (CW or pulsed) at 532 nm or UV light (pulsed only) at 355 nm. Frequency-doubled diode-pumped YAG lasers (green YAG lasers from here on) can be used for excitation of ethidium and propidium, rhodamine, phycoerythrin and its tandem conjugates, and the reporter protein dsRed²⁴⁵⁰, among other things. And the green YAG laser does a better

job of exciting most or all of these than does a 488 nm argon ion laser, and, unlike the argon laser, excites very little autofluorescence in most mammalian cell types.

In 1994, miniaturized green YAG lasers emitting 2 mW at 532 nm were offered for sale for \$4,000. This was about twice the price of a 2 mW green (543 nm) He-Ne laser, which does an even better job of exciting phycoerythrin. Today, you can buy that 2 mW green YAG laser in the form of a laser pointer for about \$200. Unfortunately, cheap green YAG lasers tend to be noisy, and when I say noisy, I mean they may turn themselves on and off. Most of these cheap lasers, by the way, come from Russia and China. A so-called single longitudinal mode green YAG laser, which may or may not actually have only one longitudinal mode, but which will have relatively low optical noise, can be had from several sources (I can speak for products from Coherent and Uniphase), but will set you back more than the cost of an argon laser. The small size and low power consumption are attractions; Guava and Luminex, among others, use green YAG lasers with about 10 mW power output in their instruments, which can be operated from batteries. If you'd like, you can buy a 10 W green YAG laser, and it will still run off house current. You're not likely to need to put that kind of power through cells, but, if you're still using a dye laser, a multiwatt 532 nm YAG makes a much more efficient pump than does a multiwatt ion laser.

The tripled UV lasers could be useful for excitation of DNA stains and calcium probes. They are at the right wavelength (355 nm), but they only operate in pulsed modes. However, a technique called **mode-locking** allows the lasers to emit regularly spaced pulses at 80-100 MHz; this repetition rate is high enough so that the laser behaves more or less as if it was a continuous light source. Both Lightwave Electronics and Spectra-Physics have introduced mode-locked 355 nm YAG lasers, and BD Biosciences is offering them as sources in both benchtop systems and sorters. The lasers are expensive, possibly even more expensive than the UV ion lasers that represent the only alternative sources offering power levels of tens to hundreds of milliwatts. However, mode-locked YAG lasers share the modest power and cooling requirements of their CW cousins, and therefore do not require the expensive infrastructure needed to support UV ion lasers.

An attempt to develop a solid-state CW UV laser for flow cytometry is ongoing at Light Age, a company that makes **alexandrite** lasers. Alexandrite, which is beryllium aluminum oxide containing chromium, emits between 700 and 850 nm, and can be pumped by 635-670 nm diode lasers; UV emission of 10-15 mW at approximately 370-380 nm has been achieved from a frequency-doubled, diode-pumped alexandrite laser; one of Light Age's prototypes was run for several months, and maintained a reasonably low optical noise level. The tunability of Alexandrite lasers is another point in their favor. The lasers have been expensive, although probably not any more expensive than UV ion lasers or mode-locked UV YAG lasers.

Alexandrite is an example of a **vibronic** crystalline laser material. Such materials behave similarly to laser dyes; laser transitions may occur between an excited electronic state and any vibrational state associated with a lower electronic energy state, allowing output to be tuned over a broad range. **Titanium-doped sapphire** lasers can operate in pulsed or CW mode between 660 and 1180 nm. Pump energy can be supplied by a high-power green YAG laser. Pulsed Ti:sapphire lasers are used for multiple-photon excitation of fluorescent dyes and intrinsically fluorescent cellular constituents; their high cost (typically over \$100,000) has, to date, limited their use. **Cr:LiSAF** lasers, in which the laser transitions occur in Cr^{3+} ions in a matrix of LiSrAlF_6 , have been doubled to 430 nm; a laser based on this technology was, and may still be, offered for sale by Melles Griot, but violet diodes may be a better bet for this wavelength range.

Melles Griot also offers 457 nm CW lasers made by frequency doubling the output of **neodymium yttrium vanadate** (Nd:YVO_4). The biggest of these runs on house current, drawing 75 W at the plug, and puts out 400 mW; I am not clear why people aren't using these lasers instead of water-cooled argon lasers in chromosome sorters. Admittedly, the water-cooled lasers do put out other wavelengths besides 457.9 nm, but the solid-state jobs are about half the price and let you do without the plumbing and the cooling arrangements.

Although green Nd:YAG lasers are good excitation sources for a lot of dyes and labels used in cytometry, they can't quite reach fluorescein, which is not only still very popular as a label, but which also has been derivatized into probes for a large number of structural and functional parameters. Fluorescein can be excited at wavelengths as high as 515 nm; in fact, the first work on two-color immunofluorescence flow cytometry¹¹⁵ was done using fluorescein- and rhodamine-labeled antibodies excited by the 514.5 nm line of an argon ion laser. As it happens, the primary emission wavelength of **ytterbium YAG** (**Yb:YAG**) lasers is 1029 nm; they could be doubled to 514.5 nm and used for fluorescein excitation. Yb:YAG lasers have some advantages over Nd:YAG in terms of ease of pumping and stability, and this suggests that green Yb:YAG lasers might have fewer problems than green Nd:YAG lasers, but the emergence of 488 nm solid-state lasers (see below) has probably discouraged development of a green Yb:YAG for cytometry, and is also likely to set back development of 473 nm, frequency-summed Nd:YAG lasers, which have been offered in some cytometers.

Some years back, Uniphase, a major supplier of argon and He-Ne lasers as well as green YAG lasers, described a 544 nm solid-state laser using an erbium-doped fluoride glass fiber as the lasing medium¹³⁸. This is not a frequency-doubled laser; pumping to the excited state by 971 nm light is accomplished by two-photon absorption (see pp. 118-9). The 544 nm wavelength is ideal for excitation of phycoerythrin and its tandem conjugates; since power levels of 10 mW should have readily been attainable, erbium fiber

lasers might have been preferable to both green He-Ne and frequency-doubled YAG lasers as sources for flow cytometry. Uniphase's recent dramatic up-and-down performance in the financial markets suggests that the company's development programs may not be running at the level originally intended. Business aside, fiber lasers can be made to operate at other visible wavelengths; they also have excellent **pointing stability**, meaning the beam doesn't have as much of a tendency to undergo slight changes in direction as do beams from other types of lasers. This would also be advantageous for flow cytometry. Maybe Uniphase should have sold Enron moral fiber lasers.

But, getting back to (or away from) business, the big story in lasers for cytometry in recent years is about **frequency-doubled semiconductor (diode) lasers**. Doubling a semiconductor laser operating at 750-1000 nm will (do the math) yield UV, violet, or blue light. The process is inefficient; a laser capable of producing several hundred milliwatts is required to get a few milliwatts of visible light. The resulting laser system is typically considerably more complex and more expensive than a diode laser, because other components, notably, an external mirror and a crystal of the material used for harmonic generation, must be incorporated into the system. Before 2000, there were a few doubled-diode lasers on the market, but they were neither powerful nor cheap. Things have since changed, attracting much attention.

Coherent's "Sapphire" laser, apparently (and confusingly) named for its blue output, was introduced to the cytometry community at the May, 2000 ISAC meeting²⁴¹. The Sapphire isn't a Ti:sapphire laser and doesn't, as far as I know, have any sapphire in it at all; it is a frequency-doubled, diode-pumped semiconductor laser; Coherent currently sells a 10 mW, 460 nm model and both 20 mW and 200 mW models emitting at 488 nm; output wavelengths are guaranteed to ± 2 nm. BD Biosciences and DakoCytomation, among others, offer the 20 mW, 488 nm Sapphire in benchtop instruments; iCyt-Visionary Bioscience is retrofitting the 200 mW, 488 nm Sapphire to sorters. These lasers are not inexpensive, but they are small in size, run on small amounts of house current, and, reportedly, feature long lifetimes (>10,000 hr).

Another laser company, Novalux, has recently been making the rounds with another frequency-doubled semiconductor laser, the Protera; this will also deliver 10- 20 mW at 488 nm, and is priced competitively with argon lasers in the same power range.

The semiconductor lasers in both Coherent's Sapphire and Novalux's Protera are **surface emitting lasers**, and, unlike typical edge-emitting diodes, produce circular, near-Gaussian beams. The technology could yield relatively inexpensive laser sources at wavelengths ranging from UV to yellow. I would not be surprised to see the cytometer manufacturers and/or third parties produce drop-in solid-state replacements for the air-cooled 488 nm argon lasers now installed in most instruments, and soon.

Laser and Light Source Noise and Noise Compensation

Laser noise may originate from several sources. Poor power supply regulation and/or design typically result in light output fluctuations at the frequency of the line current used to run the power supply or at a multiple thereof; if a high-frequency switching circuit is used in the power supply, light noise due to power supply problems may also occur at the switching frequency. In some lasers, particularly He-Ne and He-Cd lasers, noise may be found at frequencies of a few hundred kilohertz, due either to radio frequency energy used to pump the lasing medium or to fluctuations in the medium itself.

In most circumstances, the level of light source noise determines the minimum detectable signal level for scatter measurement channels; although preamplifier baseline restoration circuitry (p. 55) removes the steady DC component of the background noise produced by stray illuminating light, the AC component of the background, representing fluctuations around the DC level, is amplified along with the signals produced by particles passing through the illumination beam. To be detectable, the signal from a particle must be substantially above the level of the fluctuations; therefore, the use of a light source with lower noise allows detection of smaller particles.

Noise also has a direct effect on the precision of both scatter and fluorescence measurements. Noise is usually specified as an average, RMS (Root Mean Square) value, and/or as a **peak-to-peak** value; both are given as percentages. The RMS noise value is actually the coefficient of variation (CV) of the power output level. If noise is normally distributed, the RMS value can be estimated from the peak-to-peak value as follows. Since more than 99 percent of the area of a normal distribution lies within three standard deviations of the mean, dividing the peak-to-peak noise percentage by 6 provides an approximation to the RMS noise percentage. RMS noise levels are generally on the order of 0.05% or less for diode and high-quality solid-state lasers, 0.2% to 0.5% for water-cooled ion lasers, and 0.5% to 1% for air-cooled argon and He-Ne lasers.

In an otherwise perfect cytometer, measuring particles that were absolutely identical, the CVs of scatter measurements could be no lower than the RMS noise of the light source. The same restriction on precision would hold for fluorescence measurements using low power illumination, where a given percentage change in excitation power produces the same percentage change in emission intensity; at higher power levels (p. 117), precision might be less affected.

It should be remembered that operation in the light control mode does not guarantee protection against noise. If the light output drops substantially enough for the supply current to rise to its maximum value, any further mechanical or optical deterioration cannot be compensated for by either the light control or the current control circuitry; thus, a change in light output will occur.

One solution to source noise problems is the incorporation of circuitry in the signal processing electronics that senses variations in illumination and corrects measured signal levels accordingly. This was done in Kamensky's RCS, in the Block apparatus, and in several laboratory-built instruments, including mine. While commercial fluorescence flow cytometers do not now offer source noise compensation, it is relatively easy to implement using digital signal processing, and could thus become common in the future; this could help tame cheap, noisy YAG's.

The principal problem with He-Cd lasers is plasma noise, at frequencies between 300 and 400 kHz. It is difficult to keep RMS noise levels much below 1.5% even when the laser is new. Noise levels tend to increase thereafter, especially if the laser is left idle for long periods, because this leads to an irreversible buildup of helium in the plasma tube, which increases noise. This effect can be minimized, but not eliminated, by running the laser for at least several hours a week. That's not much help if the problem has already developed; I have had several He-Cd lasers become unusable in this fashion.

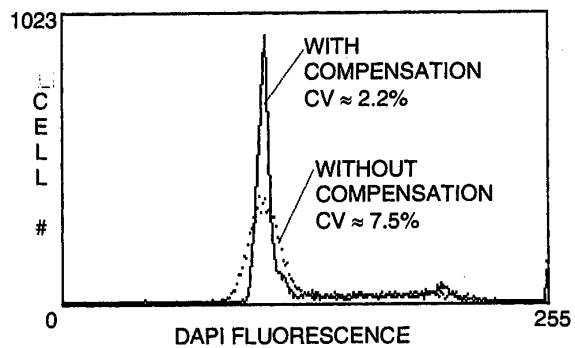


Figure 4-24. Effect of noise compensation circuitry on precision of fluorescence measurements using a noisy He-Cd laser for excitation.

If I measure DNA content with DAPI using these lasers and my standard electronics, I get atrocious precision; the corresponding distribution shown in Figure 4-24 has a CV of about 7.5%, because the illumination intensity reaching any given cell at any given time varies over such a large range due to noise. Things get better when I use the noise compensation circuit. A beam splitter reflects about 5% of the laser power to a photodiode detector, providing a reference signal. The fluorescence of dye in the cells is detected by the cytometer's PMT, and an analog circuit is used to divide the PMT signal by the reference signal, yielding a compensated fluorescence value. The DNA content distribution measured using the compensation circuit has a CV $\approx 2.2\%$. Incidentally, if I were making a pH or calcium measurement, using a ratio of fluorescence emission from one dye at two wavelengths, I wouldn't need to monitor laser output; taking the ratio compensates for the

fluctuations in source power output because the numerator and denominator are equally affected by the noise.

There are other ways of compensating for laser noise. "Noise eaters" that you can put between the laser and the rest of the world sense laser output, and use the signal to control acousto-optic or electro-optic modulators, which adjust their light transmission rapidly in response to an applied voltage. These gadgets, however, cost thousands of dollars, as opposed to tens of dollars for a noise compensation circuit, and may require that the laser beam be polarized, which many He-Cd beams are not.

Some pretty sophisticated flow cytometry can be done by using a 325 nm He-Cd laser, a low-power 488 nm argon laser, and a 633 nm He-Ne laser. B-D's LSR uses those sources, although they haven't put in noise compensation. To get UV excitation for Hoechst dyes and 488 nm for immunofluorescence, it has always made more sense to me (and others^{107,643}) to use an He-Cd laser or an arc lamp (and, these days, maybe a diode or a solid state UV laser) for the UV than to get UV and 488 nm out of one big, one big and one small, or two big ion lasers. In 2002, even manufacturers who have stuck with the big lasers for years are prepared to change. There are both flow and scanning system that use violet/UV and red diodes and green YAG or solid-state 488 nm lasers. I have seen the future, and it is air-cooled. To paraphrase an old standard, with apologies to Paul Simon:

Fifty Ways to Lose Your Laser

If you have thought of doing flow cytometry
With dyes requiring excitation by UV,
Using an ion laser makes it plain to see
There must be fifty ways to lose your laser.

The plasma tube goes, Mose,
The power supply's fried, Clyde,
The optics burn, Vern,
And the solenoid melts,
And, when they're replaced,
A hose comes unlaced,
So you want to cuss, Gus,
And try something else.

You can pick up an arc lamp, as has long been done
In places where a great big laser's hard to run,
But lining up your optics won't be much more fun
Than learning fifty ways to lose your laser.

Don't throw in the towel, Raul,
There's help on the way, Jay;
Just send the arc back, Jack,
And listen to me.
What will do the job, Bob,
Plugs into the wall, Paul;
An He-Cd, Lee,
Will get you UV.

Kit Snow, Scott Cram, Tom Frey et al and I have shown
You can sort chromosomes with HeCd light alone,
This might help when your budget's cut back
to the bone,
And there are fifty ways to lose your laser.

We've got diodes, too, Lou,
And alexandrite, Dwight,
Mode-locked YAG as well, Mel,
So listen to me.
With power costs high, Guy,
You'll shell out too much, Dutch,
Argon's not the way, Ray,
In this century.

Why pay to replumb and rewire? Don't be dense.
Small lasers can't be used for national defense,
But, for cytometry, they make a lot of sense,
And you'll find fifty ways to use your laser.

Danger!!! Laser!!! Hazards and Haze

Any laser that emits more than 800 μW anywhere from the UV to the near IR is classified as hazardous by the U.S. Bureau of Radiological Health. Once you get past the "Star Wars" notion of a laser as a handy tool for vaporizing enemies and competitors, you start to think of laser hazards primarily in terms of eye damage. This makes sense, up to a point, but there have only been a few dozen accidental injuries reported resulting from undiverged or focused beams getting into peoples' eyes.

Laser damage to the eye, or any other tissue, is caused when the absorption of light by the tissue causes enough heating to kill cells. The power density of the laser beam and the absorption coefficient of the tissue are the primary determinants of how much damage will be done. Argon lasers emitting a watt or so are routinely used for "spot welding" detached retinas, but you can shine a focused beam at lower power (say about 80 μW) into your eyeball without hurting your retina. In fact, Rob Webb, who taught me a lot about building and using flow cytometers, developed a highly useful gadget called the scanning laser ophthalmoscope, which produces a wide-field, high-resolution video image of the retina by scanning across it with a low-power laser. Incidentally, UV light doesn't penetrate into the eye very well; this means that it is unlikely to damage the retina. It also means we can't measure the UV-excited fluorescence of pyridine nucleotides in the retina, which might have given some useful information about metabolism in diabetes and other conditions. We can, however, measure flavin fluorescence with blue light from lamps or He-Cd or argon lasers.

Argon lasers are also used by dermatologists to treat pigmented lesions such as hemangiomas. These blood vessel tumors, which are reddish-purple in color, absorb more blue-green light than the surrounding normal tissue, and

thus get hotter faster. The tolerance of normal skin for laser radiation is also a function of pigmentation; I am fair-skinned, and experience has taught me that I won't burn my finger by putting it into a 200 mW, 488 nm beam 2 mm in diameter, while colleagues of mine with darker complexions will quickly withdraw their fingers from the same beam. If I accidentally get my fingers into the beam at a point where it is converging toward the flow system, I will get a slight burn. Black construction paper, however, will catch fire if held in either the 2 mm or the focused beam.

People tend to worry most about invisible laser beams, meaning those in the UV and IR wavelength ranges, because it's harder to avoid what you can't see. UV light is of special concern because the photon energy levels are high enough to cause ionization; this can lead to mutation and carcinogenesis in exposed tissue. The probability of mutation increases considerably if UV-excited dyes, e.g., Hoechst dyes, are bound to DNA; mutation rates can also be increased by irradiation of acridine orange-treated cells with blue light or by irradiation of methylene blue-treated cells with red light. That said, I must point out that I know a lot of sun worshippers and former sun worshippers who haven't gotten skin cancer, and I've never heard of anybody picking up a tan in a flow lab, regardless of the state of undress of arc lamps, lasers, and (heh, heh!) people.

At the power levels that are, or should be, used in flow cytometry (a few hundred mW or less), widely diverged laser beams and diffuse reflections should not represent major hazards to users' eyes or skin. The potential problems come from concentrated (undiverged or converging) beams. It is fairly easy for users to identify the paths of beams from lasers emitting at visible wavelengths, and thus to avoid exposure. Most flow cytometers don't use either UV or IR beams, which might otherwise be more dangerous. Flow cytometers, like other systems that incorporate lasers, are required by law to be equipped with light shielding, and operators are required to be provided with safety goggles that prevent laser light from reaching the eye. Unfortunately, adjustments to the laser and illumination optics are best made, and often only possible, when one can see the beam, which requires removing the shields and the goggles.

Even if you were running one of the larger ion lasers used in flow cytometry at full power, you probably wouldn't be able to kill yourself directly using the laser beam; you'd be more likely to die from complications of the hernia you got trying to lift the power supply. You could very easily kill yourself, however, while fiddling with the optics of an ion laser by allowing a careless finger to brush past one of the many open electrical connections carrying several hundred volts at over fifty amperes. There have been a few dozen fatal accidents associated with laser use; in all cases, the victims were electrocuted. Many of them were far from novices at working with lasers. I got a very fast trip across my lab once while tweaking optics with sweaty fingers; since then, I have made it a point always to have somebody else, preferably trained in CPR, around the lab when working on a big laser.

I have already made mention of the fact that almost none of the applications of flow cytometry require use of more than about 50 mW of laser power when an efficiently designed optical system is used. Most of the small lasers run on house current and don't require the user to get into the laser head and tinker with the optics, and their lower-power beams pose less of a hazard to users' eyes and virtually no hazard to skin. Small lasers are therefore considerably safer than large ones, a point that probably helped speed the adoption of the former in commercial flow cytometers.

What about other possible effects of lasers that are, at present, inexplicable by science? There is a provocative body of literature on effects of low-power laser irradiation on cells and tissues; try doing a search on "laser acupuncture" in the MEDLINE database some time if you're interested. It has been claimed that shining a few hundred microwatts of red He-Ne laser light on the appropriate acupuncture points stimulates immune response, reduces gastroesophageal reflux (not in me, unfortunately), and relieves symptoms of arthritis, diabetes, and gallbladder disease, to name a few. Maybe a little laser exposure is good for you.

To sum up, big lasers, which few of us use anymore, are associated with big hazards; one should respect them. Most people who use benchtop flow cytometers with small visible lasers are in greater danger from their reagents and their samples than from their lasers. But watch out for the laser pointers; many of them put out more than the legal 800 μ W, especially the green ones.

4.4 LIGHT COLLECTION

We now know what we need to do to illuminate cells, whether they are in a flow system or on a microscope slide. We next turn to the task of collecting at least some of the light emitted and/or scattered by the cells, and directing as much of it as possible, and as little other light as possible, to our photodetectors.

The optical elements of microscopes have been designed to do this job. The lenses provide **spatial resolution**, enabling us to collect a great deal of the light coming from a very small region of space and relatively little of the light coming from other regions a very small distance away. Optical filters provide **spectral resolution**, allowing discrimination between scattered, fluorescent, and background light.

Microscope Objectives

A microscope objective is designed to form a real, magnified image of a small object at a fixed distance, referred to as the **tube length**, from its mount. It is customary to describe objectives in terms of two principal parameters, magnification and N.A., and also to specify the working distance between the front element of the lens and the object. When placed at the working distance from the object, the objective forms a real image, magnified at its stated magnification, in a plane (the **image plane**) approximately one tube length behind the "shoulder" of the

objective (that's where it screws into the nosepiece; you or I might have picked a different anatomical reference).

Looking at the Observation Point

In order to select a collection lens, we ought to look at a magnified image of the observation point in a flow cytometer; I've sketched one as Figure 4-25. What I've drawn is what you see if you use a decent microscope lens and make an image of the intersection of the laser beam with a round capillary flow chamber or with a stream in air. You find out where the core is and how big it is by running a fluorescent dye through the system; if your excitation is anywhere between UV and blue-green, Mercurochrome, which is a fluorescein derivative, is fine.

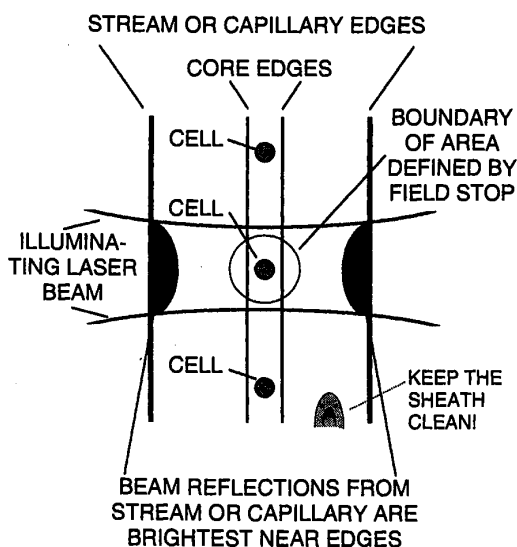


Figure 4-25. Looking at the observation point.

If the image is big enough, you'll see that, while there is a lot of reflection of the excitation beam off the capillary or stream, most of the reflection is from the edges, as long as the stream or capillary is, say, at least 50% wider than the beam. Since we want to measure light coming from the core, we almost instinctively think of using a field stop, i.e., a pinhole or slit in the image plane, so that the only light that gets through is from the region of the core. I've shown the region we want to look at in the figure.

What happens if we don't use a field stop? That depends upon what we're trying to measure. You might figure that you don't have to worry about all of the reflected light at 488 nm if you're measuring fluorescence above 515 nm, because you can get rid of it with optical filters. This is true, to some extent; however, as we will presently see, this approach places a considerably heavier burden on the filters, in terms of the performance required, and increases the

likelihood of noise due to light leakage and/or filter fluorescence. If we're trying to measure light scattered at large angles to the excitation beam, we've got a problem; all of that 488 nm light is going to clobber our scatter signal, which is at the same wavelength as the excitation beam. In this case, we're forced into using some type of light barrier to insure that we collect light predominantly from the cells in the core rather than from elsewhere.

Stops vs. Blockers

I mentioned that what I've shown in the figure is what we see when the stream or capillary is wider than the beam by 50% or more. This is usually not the case in cell sorters in which measurements are made of cells flowing in a stream in air. If, for example, we have a 70 μm stream and a beam 80 μm wide, we will see a "ring" of light from the illuminating beam around the entire circumference of the stream, and we simply won't have the option of using a field stop to get rid of stray scattered light.

The alternative typically in stream-in-air cell sorters is the inclusion of an **obscuration** or **blocker bar**, the effect of which you can simulate by placing a finger over the drawing along the path of the illuminating beam. Particles in the core will emit fluorescence and scatter light in all directions; the obscuration bar will block a substantial amount of fluorescence and scattered light, but it will allow some of both to get into the collection system, while preventing almost all of the light reflected from the stream from going into the collection system.

Up to a point, this is a good idea. What we want to be able to do in a flow cytometer is to detect weak signals above background, and anything which increases the ratio of (photons collected from cells) to (photons collected from other sources) is helpful, unless it diminishes the number of signal photons to fewer than can be reliably detected above noise based upon sampling statistics. The use of a field stop in the image plane, which masks light coming from everywhere but the immediate vicinity of the intersection of the core stream with the illuminating beam, accomplishes much the same purpose as the use of an obscuration bar, but allows more signal photons to get through. While this generally isn't done in stream-in-air flow cytometers because of the relative sizes of beam and stream, it is done in flow cytometers in which observation occurs within a capillary or cuvette. When the stream in a stream-in-air system is wider than the beam, use of a field stop instead of an obscuration bar increases the signal intensity, and the signal-to-noise ratio is improved¹⁰⁵. If the observation point is in a flat-sided cuvette, reflections of the laser beam off the cuvette walls are minimal; however, it still helps to use a field stop to reduce problems such as interference from Raman scattering.

Signal versus Noise: To See or Not to See

We have already found that we can only extract a certain number of photons from the sample, no matter how much excitation power we use, so it seems like a good idea to

collect as many of those photons as we can get. We have just found out, however, that increasing light collection from the sample doesn't help us much if it is accompanied by a proportionate increase in light collection from the background.

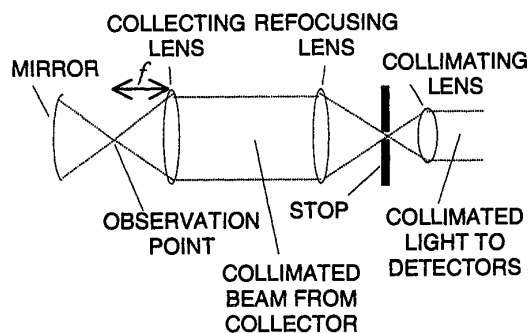


Figure 4-26. Maximizing light collection: Not always a good way to do things.

One can collect the largest possible amount of light by putting the collection lens at a distance from the observation region equal to its focal length, f . In such an arrangement, illustrated in Figure 4-26, no image is formed; the light rays emerge from the collector lens in parallel, in a collimated beam. The lenses used for collection in this fashion are typically not microscope objectives, but somewhat longer focal length, high-aperture aspheric lenses of the type normally used as condensers. The beam collected in this manner usually contains a substantial amount of light at the excitation wavelength. The burden of excluding this light from the fluorescence detector falls entirely on the optical filters used to limit detector bandwidth, unless the collimated beam of collected light is brought to a focus in a plane containing a pinhole as a field stop before being routed to the detector filters. A refocusing lens and field stop are shown in Figure 4-17. Behind them is an additional collimating lens; this is needed to reconverge the beam in order to relay light to the filters and detectors.

The setup in Figure 4-26 also illustrates another trick for increasing light collection; a **concave spherical mirror**, with a focus at the observation point, is placed opposite the lens, allowing the lens to capture some of the light emitted in this direction, which would otherwise be lost. Mirrors can converge and diverge beams as can lenses; a concave surface converges a beam, and a convex surface diverges one. Since refraction is not involved, mirrors, unlike lenses, do not exhibit wavelength-dependent variations in focal length. Most flow cytometers do not use the collection arrangement shown in Figure 4-26, but, instead, make a real image of the observation point with a high-N.A. lens, and employ field stops to prevent light from outside the observation region from reaching the detector. However, since the lens has to be farther than its focal length (between 1 and 2 focal

lengths) away from the observation point in order to make an image, it must necessarily collect light over a smaller solid angle, and therefore get less light from the specimen. Are we missing a free lunch somewhere?

Of course not. In the first place, the imaging lens generally produces a fairly highly (20-50 $\times$) magnified image of the observation point. If placed two focal lengths from the observation point, it would produce an image the same size as the observation region, let's say about 50 by 50 μm . Geometric optics tells us that to produce a highly magnified image, the lens is going to be not much more than one focal length from the observation point, so it won't collect all that much less light than it would if it were exactly one focal length away.

Since the lens is almost one focal length from the observation point, the collected light emerging from it is going to converge to an image plane some distance between two focal lengths and infinity behind it; the closer the lens is to one focal length from the observation point, the closer the image is to infinity. In this case, "close to infinity" is about one tube length, or about 6 inches. The light is "close to collimated"; throughput considerations are at work here, and they tell us (p. 123) that, when we collect light over a large solid angle to make a magnified image, the image forming rays occupy a smaller solid angle in proportion to the amount of magnification. Thus, if we make a magnified image of the observation region, we don't need either a refocusing or recollimating lens, and we eliminate any transmission losses that might be associated with these optical elements. The dichroics used for spectral separation, which might have trouble with the divergent beam emerging from behind the stop in Figure 4-26, can easily handle the beam from the imaging lens. Also, because axial magnification goes as the square of lateral magnification, we collect much less interfering light from planes closer to and farther away from the observation region as we increase the magnification of the image.

Now, what about the stop in Figure 4-26 and the field stops used with imaging optics? If you're looking at a 50 by 50 μm observation region magnified 50 times, the corresponding region of the image is 2.5 by 2.5 mm, which is pretty macroscopic, and your field stop can be slightly larger than this and do a good job of blocking light collected from outside the observation region. With the optics in Figure 4-26, the refocusing lens is forming something pretty much equivalent to a 1:1 image in the plane of the stop. In order to block light from outside the observation region as effectively as could be done with an imaging system, you'd need a stop with an opening less than 100 μm across. These are available, but, in practice, you'd probably have to use a substantially bigger opening, because it would be difficult to keep the small one from drifting slightly out of alignment and blocking all the light. This means you let in more light from outside the immediate vicinity of the observation region. When the cell is in the observation region, most of what's in the neighborhood is water, and, with too big a

stop, you'll collect signals from whatever is in the water. If you're using dyes with high background fluorescence, you may overload your baseline restoration circuitry. If you're using 488 nm excitation and trying to detect phycoerythrin fluorescence, you'll end up with higher background noise and lower fluorescence sensitivity due to Raman scattering from water than you would have if you had made a magnified image and used an appropriately sized field stop.

In principle, light can be collected over extremely large solid angles by **ellipsoidal**¹²⁰ or **parabolic**⁶⁴⁸ reflectors; in practice, use of such elements for fluorescence collection in a flow cytometer requires higher performance optical filters than would be necessary in a system with better spatial resolution, and orthogonal scatter signals may be masked by noise. There are also the same problems with stops as were just discussed.

Flow cytometers built around microscopes, which use high-N.A. objectives^{83,88,100-103}, often show better precision than laser source systems, particularly in measurements of asymmetric cells, because the wider angles of illumination and collection in the former minimize effects of differences in particle orientation¹²¹. However, instrument performance is not necessarily improved simply by increasing the N.A. of the collecting lens. In order to make effective use of a high-N.A. objective, it may be necessary to compensate for the concomitant small depth of focus. In the Impulscytometer^{9,83}, this was done by directing sample flow along the axis of the objective, so that portions of cells passed sequentially through the focal plane. In the Steen/Lindmo instrument¹⁰⁰⁻¹⁰³, the sheathed sample is directed against a flat surface, resulting in reduction of core thickness and in orientation of asymmetric cells, allowing most cells in the sample to be kept in focus even by a high-N.A. lens.

The spatial resolution required in flow cytometry is only that which is necessary to define and isolate the image of the intersection of the illuminating beam and the core. People tend to want to use the "best" microscope objectives; estimating quality from price, they choose an apochromat or planapochromat. These lenses do have high N.A.; without it, they would not be capable of high resolution. However, the resolution increase with increasing N.A. is not due to the collection of more light, but to the collection of light over a wider angle. A planapochromat is designed to produce flatness of field over a relatively large viewing area, and high resolution, for observation and photomicrography at high magnification. Its light gathering power may actually be sacrificed to achieve better performance in these areas. Although equal amounts of light go into the front elements of a simple achromat and a planapochromat with the same N.A., the light transmission of the achromat is likely to be higher, because the planapochromat achieves its flat field and high resolution through the use of more optical elements than are used in the achromat, and light is lost in the extra elements, even with the best of antireflection coatings. For flow cytometry, the achromat is almost always the better choice; it will certainly be cheaper, and very

probably have a longer working distance and greater total light gathering capacity than will the more highly corrected lens. In recent years, I have been using, and recommending, inexpensive, single-element molded aspheric lenses with relatively high N. A. (0.62-0.68) as collection lenses, and these have found their way into some commercial instruments. And, after having had problems with low light transmission in some very expensive microscope lenses, I make it a practice to measure transmission of a low-power laser beam through any lens before I put down the money.

Spectral Selection: Monochromators versus Filters

With few exceptions, the definition of the spectral response of the detectors in flow cytometers is accomplished by the use of **optical filters**, although more sophisticated methods have been examined. Alternatives to filters may be desirable if it is necessary to make measurements in many wavelength regions. In the late 1970's and early 1980's, there were reports on the measurement of emission spectra in flow systems using polychromatic detectors¹²² and on the use of monochromators to replace filters (Stöhr discussed this at a 1981 meeting, but apparently didn't publish on it). In the mid-1980's, Buican⁶⁴⁹ described the use of Fourier transform methods for flow cytometric measurement of fluorescence in multiple spectral regions using a single detector PMT.

Monochromators and Polychromatic Detection

Spectral selection using **prism** or **grating monochromators** is accomplished by diverting different wavelengths of light in different directions and then positioning an aperture or slit between the prism or grating and the detector. The basic elements of a prism monochromator are shown in Figure 4-27. Incoming light is focused on an **entrance slit** at the focus of a **collimating lens**, which delivers a collimated beam to the prism, from which different wavelengths emerge at different angles. The **focusing lens** forms an image of the dispersed spectrum in the plane of the **exit slit**; the wavelength of light emerging from the exit slit is selected by rotating the prism about an axis perpendicular to the plane of the diagram.

The angular dispersion of a prism is not a linear function of wavelength; the angular dispersion of a grating is. Monochromators are more often built with gratings than with prisms as the dispersing element for this reason; the linear behavior makes it easier to motorize wavelength selection. Gratings also have higher dispersion than prisms, making monochromators more compact, but gratings are also less selective. A **spectrophotometer** typically uses a single monochromator to select the illumination wavelength for absorption measurements; a **spectrofluorometer** uses a monochromator to select excitation wavelength and one or two more monochromators to select emission wavelength.

Polychromatic detection is accomplished by placing multiple detectors in several regions of a spectral image. If a linear array of diode or CCD (**charge coupled device**)

detectors were placed where the exit slit appears in Figure 4-27, each element of the detector array would respond to radiation in a different spectral region. It would be more difficult to use conventional PMTs for polychromatic detection because of their size; an optical path length of several feet would be required to achieve the required spacing between desired spectral regions. Newer, multianode PMTs might be usable in this application. Experimental work with polychromatic detection using diode and CCD arrays in flow cytometry continues¹¹³⁹, although this technology is not used in commercial systems.

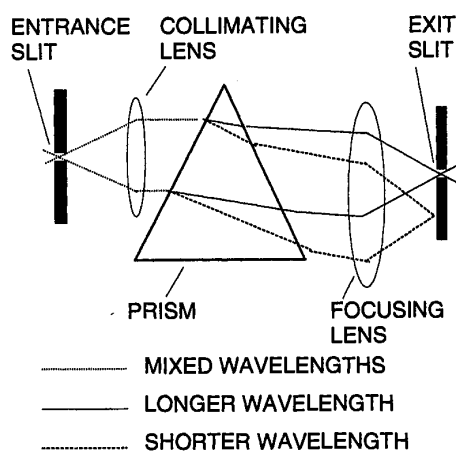


Figure 4-27. A prism monochromator. Output wavelength selection is accomplished by rotating the prism. Polychromatic detection could be done with a linear diode or CCD array at the position of the exit slit.

Buican^{649,1140} built a flow cytometer in which fluorescence emission in eight spectral regions could be measured using a detector incorporating a Fourier transform or interferometric spectrometer. Such a detector system, however, is an order of magnitude more expensive than monochromators, which are themselves several times more expensive than a set of filters. The Fourier transform method has been widely adopted for infrared absorption spectroscopy because it offers some improvement (the so-called Fellgett advantage) in signal-to-noise ratio when detector noise is a limiting factor. This is not generally the case in flow cytometry; thus, it is not clear that interferometric detection offers either theoretical or practical advantages for most applications.

In general, the cell sample analyzed in a flow cytometer contains no more than three or four fluorescent materials, with well-known spectral characteristics, selected for efficient excitation by the illumination wavelength(s) used. Even in work with eight or more colors, optical filters remain the most cost-effective means for detector bandwidth limitation.

Interference Filters: Coatings of Many Colors

An **interference filter**, also called a **dichroic**, **dielectric**, or **reflective filter**, is composed of a transparent glass or

quartz substrate(s) on which multiple thin layers of dielectric material, sometimes separated by spacer layers, are deposited. Constructive and destructive interference occur between reflections from the various layers, with the wavelength range(s) transmitted being determined primarily by the thickness of the layers and the selectivity by the number of layers. The unwanted wavelengths are reflected.

The layers of dielectric material on a substrate are collectively referred to as an **interference coating**, or, more commonly, simply as a **coating**. Coatings are used in optical elements other than filters; lenses, for example, are commonly coated to reduce reflections at their surfaces. Since, in the absence of a coating, about 5 percent of light is lost at each surface, the light transmission of the highly corrected multielement lenses used in microscopy and photography would be unacceptably low if antireflection coatings were not used.

Coatings are described as **hard** or **soft**; hard coatings are durable and moisture-resistant, but somewhat more difficult and more expensive to manufacture. They are typically used on surfaces that may be exposed to air and/or cleaned with tissues or swabs and solvents. Laser mirrors and dichroics are usually hard coated; most interference filters are made with soft coatings. The individual layers of coated substrates must be glued together, and a seal applied around the edges of the filter to prevent ambient moisture from degrading the coating.

Absorptive Filters versus Interference Filters

In an **absorptive filter**, such as a **color glass filter**, unwanted light is disposed of by absorption. It is possible to remove much more of the unwanted light by absorption than by interference; light transmission outside the **passband**, i.e., the region in which the filter is designed to transmit light, is commonly a few percent for a purely dielectric filter and only a few hundredths of a percent or less for a color glass filter. Transmission in the passband, however, is usually lower for interference filters than for color glass filters. Interference filters, as a general rule, allow sharper transitions between rejected regions and the passband(s), and permit greater selectivity; filters with a **FWHM (Full Width at Half Maximum [transmission])** bandwidth of 3 nm are readily available.

In order to measure weak fluorescence, e.g., immunofluorescence, we really need to get rid of stray excitation light, and of excitation light scattered within the collection angle of the fluorescence collection optics. Letting 1 percent of that light through isn't good enough; we therefore can't use just a pure interference filter.

However, if we put a long pass color glass filter in a laser beam, we see that some of the absorption responsible for removal of excitation and scattered light is associated with fluorescence emission from the filter; this emission from the filter is at wavelengths that will pass through the filter, thus creating noise at the detector. So what do we do?

Well, we should, of course, use a field stop, as discussed in the preceding pages, to eliminate as much of the extraneous excitation light as possible, but we still need to specify what kind of filter goes between the field stop and the detector. And the winner is: both. The “interference filters” we use are made with an absorptive layer behind the coated layers. You can usually tell which side the absorptive layer is on because it appears more strongly colored; the coated side is more reflective, i.e., shinier. When you mount the filter in your instrument, the shiny side of the filter should face the collection lens; the other side should face the detector. When in doubt, reflect light at the unwanted wavelength from the filter onto a piece of paper (or a power meter); the side of the filter that reflects the most light should face the collection lens.

Interference filters specifically designed for use in cytometry frequently have low enough transmission outside the passband so that additional filters are not needed. They also have little arrows telling you which way to insert them in case you forget the “shiny side out” rule. If there is a need to further restrict bandwidth, the field stop (and the blocker bar, if one is used) and the interference filter get rid of enough of the stray excitation light to permit placing a color glass filter between them and the detector without worrying too much about filter fluorescence. This doesn’t mean it can’t happen; if it does happen, the absorptive layer of the interference filter should not be neglected as a source of such fluorescence. Loken and Stall discussed the evaluation of filters in some detail⁶.

We generally neglect the fact that absorption results in the conversion of light into heat. A laser beam hitting a color glass filter may thus destroy dye molecules in that filter, changing its absorption, fluorescence, and transmission characteristics. Those dye molecules may fade more slowly during exposure to sunlight, arc lamps, etc. If one keeps it out of strong light, however, a color glass filter should have a relatively long lifetime. If you drop one and break it, the transmission characteristics of the pieces are the same as those of the intact filter; you can cover a lot of detectors with a 1" square filter if you don’t mind getting glass splinters.

The companies that make interference filters have ways of cutting them; don’t attempt this stunt at home. Interference filters have finite lifetimes even when they’re not dropped; moisture eventually gets through their seals and causes the soft coatings to degrade. This is sometimes referred to as “delamination,” which more properly describes the filter coming unglued, but the end result is the same: “Old interference filters never dye, they just lose their layers.” This typically happens after a few years; if you hold the filter up to the light and see mottling around the edges, it’s time to get a new filter. The old-timers didn’t discover the problem until their cohorts of new filters aged.

Filter Transmission Characteristics

Figure 4-28 shows idealized transmission curves of bandpass, short pass, and long pass interference filters. The

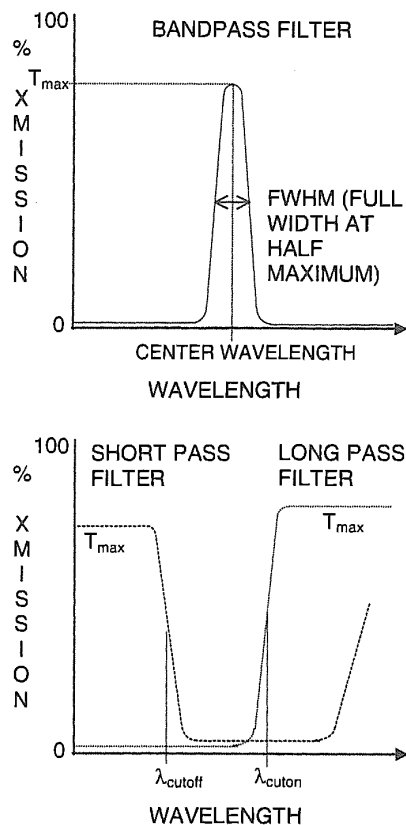


Figure 4-28. Light transmission characteristics of bandpass, short pass, and long pass interference filters.

pertinent characteristics of a bandpass filter are its **center wavelength**, usually specified in nanometers (nm), its **maximum transmission**, T_{max} , and its **bandwidth**, usually given by the FWHM in nanometers. The bandpass filters used in flow cytometry typically have a FWHM of 10-50 nm and a T_{max} of 65-80%.

If you turned the curve for the bandpass filter upside down, you’d have the transmission curve of a **notch filter**, which is not included in the figure. One of the more common notch filters is the magenta filter used for photographic color separation. A green bandpass filter blocks the red and blue/violet ends of the spectrum and transmits the green in the middle; a magenta notch filter passes red and blue/violet but blocks green. Narrow notch filters specifically designed to block laser lines may be useful for high-sensitivity fluorescence measurements.

Long pass filters are characterized by their T_{max} and by their **cut on wavelength**, λ_{cutoff} , at which light transmission is 50% of T_{max} . Long pass color glass filter transmission curves resemble those of long pass interference filters, but typically show much lower transmission outside the passband.

Short pass filters are characterized by a T_{max} and a **cutoff wavelength**, λ_{cutoff} , at which transmission is 50% of T_{max} .

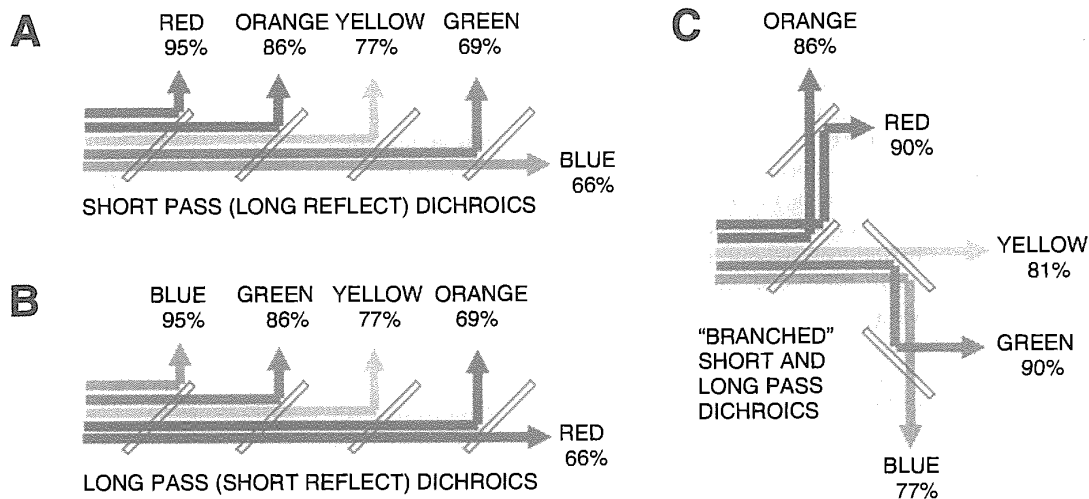


Figure 4-29. Transmission of several wavelength regions through different dichroic configurations.

They can be tricky to work with. The transmission of most short pass filters falls off sharply below about 420 nm due to absorption by some of the coating materials. However, the big problem is at the long end; they're really notch filters. A 480 nm short pass filter passes blue light but also passes a lot of red light. To get rid of this, you can use either a color glass red absorbing filter or a longer wavelength (e.g., 580 nm) short pass filter, which blocks out to the near infrared. If you use short pass filters, keep your guard up.

You can measure transmission and reflectance characteristics of optical filters using a spectrophotometer⁸⁸², but, since most spectrophotometers don't reliably measure transmission below 1 percent, you can't tell whether a filter is truly "blocked" to 0.01% outside its passband. I have a few mW available from various laser sources at wavelengths from the UV to the near infrared, and a power meter that can detect a few nanowatts; this lets me measure transmission down to 0.01 percent reliably, in a well-darkened room. The filter manufacturers can do the same.

The transmission characteristics of an interference filter are determined by the distance incident light must travel between the layers of the filter; this is shortest for rays that are normal, or perpendicular, to the filter, and increases for rays at progressively larger angles to the normal. The transmission characteristics thus change with the angle of incidence of light, so users and manufacturers must specify the angle of incidence at which a filter is designed to work. If you hold a bandpass filter up to a white light source and vary the angle at which you look through it, you will see a change in the color of the transmitted light. Absorptive filters do not exhibit this angular dependence.

Dichroics

Epiilluminated fluorescence microscopes, and arc source flow cytometers, make use of what are termed **dichroics**,

dichroic mirrors, or **dichroic beam splitters**; these are also used in light collection systems to direct light in different spectral regions to different detectors. Dichroics are interference filters; those used in fluorescence microscopy are usually the long pass type. Excitation light at a short wavelength, outside the passband, is reflected from a long pass dichroic; fluorescence emitted at longer wavelengths is transmitted through it. A short pass dichroic reflects longer wavelengths and transmits shorter ones. It is desirable to minimize the fluorescence of a dichroic, because any fluorescence excited in it by light at the illumination wavelength may be transmitted to the detector. Dichroics are, therefore, almost always made without the absorptive layers typically used in "interference" filters.

Filter sets for fluorescence microscopy have been developed that incorporate excitation filters and dichroics that, respectively, transmit and reflect in several discrete wavelength regions. This allows simultaneous viewing of (for example) blue, green, and red fluorescence excited by UV, blue-green, and yellow light. While these filter sets provide lovely visual images, their real purpose is to eliminate changing filter blocks during multicolor imaging experiments, minimizing the likelihood that the specimen will move slightly during that mechanical operation.

Managing light loss in dichroics was mentioned on p. 53; we now return to it with the aid of Figure 4-29. In a flow cytometer with 488 nm excitation, the blue side scatter signal is strongest, green and yellow (fluorescein and phycoerythrin) fluorescence signals are fairly strong, and orange and red (tandem conjugate) signals are weakest. We want to save as much of the weaker signals as possible.

It is reasonably easy to make a dichroic that reflects at least 95% of light in some region outside its passband, and relatively difficult to make a dichroic, or any other interference filter, with 95% transmission in its passband.

Thus, weak signals should be reflected from, rather than transmitted through, dichroics. Figure 4-29 assumes 95% reflectance and 90% transmission from the dichroics.

The configuration illustrated in (A) in Figure 4-29 is that shown in Figure 1-21 (p. 51); a linear array of short pass (long reflect) dichroics preserves better than 85% of the weaker red and orange signals, although 31% of the green signal is lost. This is preferable to the configuration in (B), which loses over 30% of both the orange and red signals. If the transmission of the filters were lower, say 85%, only 52% of the red signal would get through. For optimal light transmission, it may be best to have the dichroics and detectors in a branched configuration, such as that in (C), in which no signal must pass through more than two dichroics.

Most flow cytometers use a single lens, with its axis orthogonal to both the axis of flow and the axis of the illuminating laser beam, to collect fluorescence and side scatter signals. In some instruments, a spherical mirror is placed on the opposite side of the flow chamber from this lens, to reflect additional fluorescence emission and scattered light back into it. One can also get more light to the detectors by using two separate collection lenses on opposite sides of the chamber. I do this in most of my instruments; it allows collection of four-color fluorescence with none of the fluorescence signals being transmitted through more than one dichroic, reducing the light losses associated with multiple dichroics. One could also use relay lenses and/or fiber optics (see pp. 157-9) to direct light from both collection lenses to the same detector(s).

Neutral Density Filters

It is sometimes desirable to attenuate light without discriminating wavelengths; this can be done with **neutral density (N.D.) filters**. A neutral density filter can work by reflection or by absorption. A partially "silvered" (i.e., coated) mirror will reflect some portion of the light reaching it and transmit the remainder. Absorptive neutral density filters are color filters incorporating substances that absorb over the entire visible spectrum and which therefore appear gray or black in color; they are often seen on beaches. Neutral density filters are often used in front of the forward scatter detectors in flow cytometers, particularly in instruments that use high-power laser sources.

Beamsplitters; Ghosts and Ghostbusters

A reflective neutral density filter can serve as a **beamsplitter**, diverting the reflected portion of the incident beam in one direction and transmitting the remainder in another. Uncoated glass surfaces typically reflect about 5% of incident light; common cover slips thus can be used as beamsplitters to divert a small portion of an incident beam, e.g., for a side scatter measurement. To send 50% of the light one way and 50% of the light the other way, you need a reflective coated surface.

Beamsplitters (Figure 4-30) are frequently made in the form of cubes; the cube is formed by fusing two right

triangular prisms with the coating along the "hypotenuse" face of one prism. Light therefore enters and leaves a cube beamsplitter normal to the surface, minimizing undesirable "ghost" reflections that are shown coming from the rear face of the plate beamsplitter.

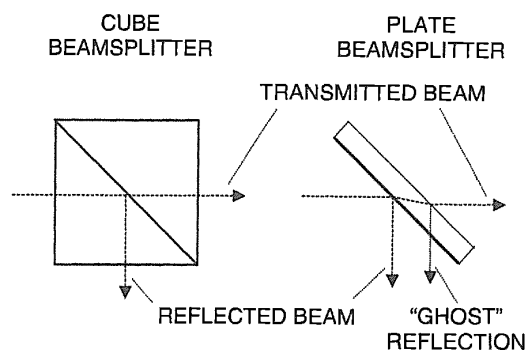


Figure 4-30. Cube and plate beamsplitters.

The more colors you measure in a flow cytometer, the more dichroics you need; and they also produce ghost reflections. The problem starts to get noticeable when there are a lot of them, so who(m) do you call? Well, you may have to call your optical filter supplier, and get dichroics and filters that you can place at angles other than 45 and 90 degrees, or you may need to use wedges to move the ghost reflections further apart so they don't go where you don't want them. But you probably won't have to call anybody; the manufacturer of your flow cytometer should take care of the problem for you. Until, that is, you decide to measure a different set of colors; then, you'll have to call both the manufacturer and the filter supplier.

It should be obvious that absorptive neutral density filters cannot be used as beamsplitters, because they are designed to absorb light, not reflect it. They, or at least the relatively inexpensive plastic ones you can buy in sheets in real camera stores, also tend to melt and/or burn when you try to use them to attenuate a few hundred milliwatts worth of laser beam.

Optics for Polarization Measurements

It is occasionally of interest to measure fluorescence or scattered light in two planes of polarization, respectively perpendicular and parallel to the plane of polarization of the excitation beam. This requires that light of different polarizations be separated and diverted to different detectors. Such measurements are relatively easy to do in flow cytometers with argon ion laser sources, because the laser source produces highly polarized illumination; you do have to be able to get at the detector filters. Two approaches can be used. The first makes use of a **polarizing beamsplitter**; this is an optical element that looks more or less like those shown in Figure 4-30, but which transmits light of one polarization and reflects light of the other.

Most beamsplitters polarize light to some extent; Asbury et al, in their extensive study of polarization in flow cytometry²⁴³⁹, showed that beamsplitters caused strong, orientation-dependent polarization artifacts. Polarizing beamsplitters, which polarize light in well-defined and predictable ways, must be selected for the appropriate wavelength and tend to be fairly expensive.

If what you'd like to do is discriminate eosinophils from other granulocytes by their higher depolarized scatter signals^{710,986,1137}, you can use a simpler, less expensive approach employing **polarizing filters**. These are typically made of a plastic in which are embedded optically active molecules that have been oriented in one direction, e.g., by application of an electric field. This results in differential absorption of light of different polarizations. If you have a pair of polarized sunglasses, you have probably noticed that rotating one of the lenses about its axis changes its transmission, making a hazy blue sky appear darker blue.

Polarizing filters decompose the electric field vectors of incident light into parallel and perpendicular components; what comes out is light of one polarization. If you put two polarizing filters together, and rotate one, there will be one position at which light transmission through the pair is at a maximum, and another, at 90° to the first, at which no light gets through, because the second filter is then positioned to transmit only the polarization which has been excluded by the first filter. There is generally an orientation mark on such filters to facilitate finding the appropriate positions.

Setting up a cytometer for quick and dirty scatter polarization measurements requires, first, that two detectors be positioned and equipped with proper wavelength selection filters to measure scatter in the same spectral region. Using beads to provide a standard signal, the detector gains are adjusted to provide equal signal strengths. Then, a polarizing filter is placed in front of each detector. The filter on the "parallel" detector is rotated to produce maximum signal strength from the beads; the filter on the "perpendicular" detector is rotated to produce minimum signal strength. Gains are then readjusted.

If you're trying to measure fluorescence emission anisotropy or polarization against some standard, it may be necessary to play additional games, such as changing the laser polarization with a retardation plate, and you probably want to use higher quality components. See p. 114, then read Asbury et al²⁴³⁹ and proceed from there.

Tunable Filters

It is now possible to make **tunable filters**, in which an electrical input changes the passband. These devices don't have conventional filters in them; some of them use acousto-optic modulators and gratings; others use liquid crystals. They can respond in milliseconds, and have been used in multispectral imaging and confocal microscopy, but not for flow cytometry. Because they polarize light, tunable filters have relatively low (35% or so) transmission; they are also on the pricey side, at least at present.

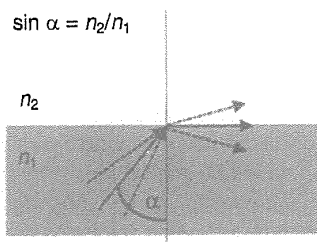


Figure 4-31. Total internal reflection.

Fiber Optics and Optical Waveguides

Lenses and mirrors aren't the only components useful for getting light from point A to point B; the communications industry, as the ads on TV remind us, gets a lot of light from one end of the country to another through **fiber optics**. These, and other **optical waveguides**, work by **total internal reflection**; the principle is illustrated in Figure 4-31. This shows three rays of light traveling in the direction of the boundary between regions of two different refractive indices n_1 and n_2 , where $n_1 > n_2$. Light is refracted according to Snell's law (pp. 106-7); light striking the boundary at the **critical angle**, $\alpha = \sin^{-1} n_2/n_1$, will be refracted along the boundary, while light striking the boundary at larger angles will be reflected back into the medium from which it came. Light striking the boundary at angles smaller than the critical angle will be refracted at the boundary and emerge into the medium with the lower refractive index.

An optical fiber consists of a plastic, glass, or silica **core** surrounded by a **cladding** with a lower refractive index. The refractive indices of the core and cladding determine the **numerical aperture (N.A.)** of the fiber; this defines the maximum half angle θ at which light entering the fiber will propagate through it, and also the maximum angle at which light will emerge from the other end. The fiber illustrated in Figure 4-32, below, is a **step index fiber**; there are also **gradient index fibers**, in which refractive index decreases continuously with distance from the fiber axis.

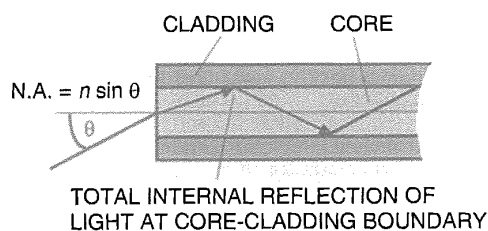


Figure 4-32. A fiber optical waveguide.

Single fibers used for optical communication typically have core diameters between 5 and 120 μm and outside di-

ameters of 125-200 μm . They are thus, at least in principle, well suited to delivering light to, and collecting light from, regions of space roughly the size of the observation volume of a flow cytometer. In the mid-1980's, Mike Hercher and I⁸⁷⁷ built a fluorescence flow cytometer in which fiber optics were used for both illumination and collection; the only lens in the whole thing was the one we used to focus the illuminating laser beam into the fiber optic that delivered light to the round capillary flow chamber.

This prompted a brief episode of euphoria in which we envisioned "flow cytometers on a chip," with the flow chamber, diode laser and detectors, and all on a single substrate, incorporating optical waveguides to move light to and from the sample stream. Waveguides of various geometries can be made on a substrate by depositing layers of material of the appropriate refractive index, or by doping the surface to change the refractive index. However, I now feel honor bound to point out that waveguide-based cytometers present a few problems without obvious solutions.

Light collection is limited by the N.A. of the fiber or waveguide. If you use a silica core, the maximum N.A. you can get is 0.4; for fluorescence collection, you'd be better off with a higher N.A. Plastic fibers with N.A. of 0.54 and glass fibers with N.A.'s of 0.66 or higher would be suitable, but, unfortunately, waveguide flow cytometers tend to have high levels of stray scattered light, and this induces fluorescence in plastic and glass fibers, which decreases measurement sensitivity in a big way. Beam shaping is at best hard to do, and collecting light selectively from the observation point is a lot easier to do using lenses and field stops. Finally, if you put the source, detectors, and all on one chip, you have to throw out the whole thing when it clogs. If waveguide flow cytometers ever do reach maturity, it will be because somebody finds an application for which their disadvantages don't disqualify them. I've been waiting for about 18 years now.

The Ortho System 30 and System 50 of the late 1970's used plastic fiber optics to relay light to the detectors; one end of the fiber was placed in an image plane, serving as a field stop; the other end went to the PMT housing. The fluorescence detector filters were placed in front of the fibers, forestalling problems with fluorescence that might have been induced in the fibers by stray laser light. BD Biosciences' new LSR II and FACSAria (Chapter 8) also use relay fibers.

Through a Glass Darkly: Light Lost (and Found) in Optical Components

I just mentioned problems with waveguide cytometers due to fluorescence induced in the optics by stray illuminating light; you should also know about some other instances in which the materials of which optical elements are made can cause severe problems.

The first set of headaches comes from the UV transmission characteristics of commonly used glasses such as BK-7 and optical crown glass; both lose very little light at wavelengths between 400 and 800 nm, but transmission falls off noticeably below 400 nm and rapidly below 350 nm. If you

look in something like the Oriel Optics Catalog, you'll see transmission is usually tabulated for fixed thicknesses of glass, e.g., 10 mm. Obviously, the thicker the glass element through which light must pass, the lower the transmission.

The "long" UV wavelengths used for fluorescence excitation range from the 350 nm krypton laser line to the 366 nm mercury arc lamp line; a 10 mm thickness of BK-7 transmits over 85% of incident light in this wavelength range, but 10 mm of crown glass transmits only 72% of incident light. If you're not using a lot of thick lenses or cube beamsplitters in your system, you can generally live with these figures and get away with glass illumination optics instead of quartz or silica, which have high transmission even at 260 nm. You should also be okay with glass optics if you use a violet diode laser. If you want to use a UV He-Cd laser, which emits at 325 nm, you'll need quartz or silica optics, despite their premium prices; the transmission of 10 mm of crown glass at 325 nm is less than 30 percent.

When you do use quartz or silica optics for UV, you should avoid antireflection coatings designed for visible light, because these coatings have very high absorption in the UV, particularly down at 325 nm. The mirrors that are used to bounce laser beams around in cytometers can also pose problems here; they typically have coatings with very high reflectivity (>95%) above 420 nm or so and very low reflectivity (<50%) in the UV. Bounce an illuminating beam off two mirrors like that and you've lost 75% of your power. I tend to use uncoated lenses and minimize the number of mirrors in the system when I work with UV; where I do need a mirror, I use the same coating as is applied to UV laser mirrors. As a last resort in cases where I need reflectivity in the UV and at long wavelengths, I'll use an aluminum mirror, which has about 85% reflectivity across the board.

Since glass can absorb in the UV and violet, it should come as little surprise to you that glass can fluoresce when illuminated in that wavelength region. With illumination at 488 nm and longer wavelengths, you can probably use glass flow chambers; with UV illumination, and even with illumination at 441 nm, however, considerable fluorescence is observable in most types of optical glass. This does not generally interfere with measurements of strong fluorescence but is likely to impair or preclude measurements of weak fluorescence. In arc lamp systems, when the same lens is used for illumination and light collection, it is wise to use fluorite lenses because they have high UV transmission and low fluorescence, both of which are desirable.

Plastic optical elements, e.g., Fresnel lenses and fiber optics, are worse offenders than glass ones and tend to fluoresce noticeably even when illuminated at wavelengths of 500 nm and above. Problems with fluorescence from optical elements can generally be avoided by minimizing the amount of stray excitation light which gets into the collection optics, using field stops, blocker bars, and/or coatings.

I noted above that some Ortho instruments used plastic fiber optics to transmit light between the fluorescence emission filters and the PMTs; fiber fluorescence was not a prob-

lem. When Ed Luther and I (unpublished) used the same fibers to collect light from an argon laser source cytometer, collection was highly efficient, i.e., we got the same signal intensities as we would get from a lens of the same N.A. as the fiber, but fluorescence induced in the fibers limited the sensitivity with which we could measure immunofluorescence. More recently, when I was trying to measure some very weak autofluorescence with UV and blue excitation, I found problems with fluorescence in the lens; I'll have more to say about this later.

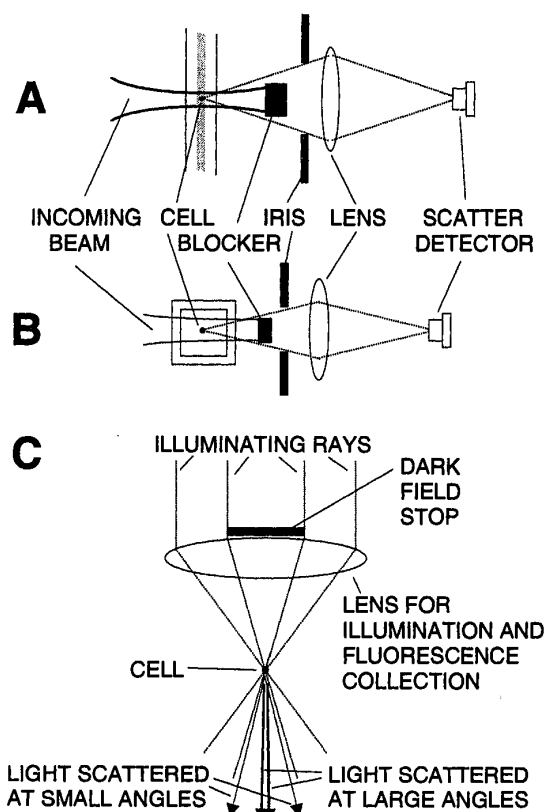


Figure 4-33. Optical arrangements for collection of forward scattered light.

Collection Optics for Forward Scatter Signals

Three basic optical arrangements are used for collection of forward scattered light; they are illustrated in Figure 4-33. Panel A of the figure shows a side view of the configuration used in flow cytometers in which cells are observed in a stream in air or in a round capillary. The stream and capillary walls act as cylindrical lenses, refracting light from the illuminating beam in the plane perpendicular to the axis of flow; in order to avoid interference from this source, it is necessary to collect light scattered at angles above and/or below this plane. A beam dump or blocker bar is placed with

its long axis perpendicular to the plane of the drawing, i.e., perpendicular to the direction of flow. The illuminating beam diverges from its focal point on the core stream; the blocker must be large enough to intercept the diverging beam, and its height determines the minimum angle to the illumination beam at which forward scattered light can be collected. The maximum angle is determined by an iris diaphragm in front of the lens, which brings the collected light to a focus at the detector.

Panel B of the figure shows a top view of the configuration used in flow cytometers in which observation is done in flat-sided cuvettes. The arrangement is similar to that shown in A, except that the axis of the blocker bar is parallel to the direction of flow. The cylindrical lens effects of a round stream and/capillary are eliminated when a flat-sided flow chamber is used; it is therefore possible to measure light scattered in the plane of the long axis of the elliptical focal spot. This is advantageous because the divergence of the illuminating beam is lower in this plane, allowing the blocker bar to be made narrower, and permitting collection of light scattered at smaller angles to the beam.

Since crossed cylindrical lenses are used to produce the elliptical focal spot, it is possible to place the back focusing lens in such a way that the beam is brought to a focus on the blocker bar, rather than on the core stream; this is shown in the drawing, and its effect is to further decrease the minimum angle at which scattered light can be collected. The iris and lens function as in panel A. Ortho replaced the conventional blocker bar with a reflective strip, which, with a lens, diverted the transmitted beam to an extinction detector.

Panel C illustrates dark field illumination as used with arc or incandescent lamps; this configuration, used in the original Gucker apparatus²⁹ (see Figure 3-1, p. 74), was also adopted by Steen^{102,1142}, who added a valuable wrinkle.

In dark field illumination, a circular stop, coaxial with the illuminating lens, results in the illuminating rays forming hollow cones of light with central shadows, converging to and diverging from the plane of the specimen. It is customary to collect light scattered by the specimen into the shadow region on the side opposite the illuminating lens.

Steen¹¹⁴² noted that light scattered at larger angles emerges near the optical axis of the system, while light scattered at smaller angles emerges nearer the edge of the shadow cone. Diverting light from these two regions to different detectors allows separate measurement of small- and large-angle scatter signals. The characteristics of "orthogonal," "90°," or "side" scatter signals dominate scatter signals collected at angles as small as 18°, but the amplitudes are higher at the smaller angle. Steen's approach, utilized in instruments produced by Skatron of Norway, also variously made and distributed commercially by Leitz, Bruker, Ortho, and Bio-Rad, and currently being revived by Apogee, has, among other things, facilitated sensitive multiangle scatter measurements of bacteria.

In experimental systems, fiber optics have been used successfully for forward scatter signal collection, but noise levels

tend to be high. My experience in recent years has been that forward scatter measurement sensitivity can be improved by using a relatively high N.A. collection lens, but I suspect I'm not collecting light at the really small angles.

4.5 DETECTORS

Once we have collected light from various directions, and selected the spectral region(s) from which we will allow light to reach the detector(s), we reach the point in our cytometer at which, Presto!, the light changes back into photons. The numbers of photons we collect from a cell vary over a wide range, depending upon whether we are measuring absorption or extinction (too many), light scattering (lots for forward scatter, enough for orthogonal scatter), strong fluorescence, e.g., of dyes bound to DNA (usually more than enough), or weak fluorescence, e.g., immunofluorescence (please, Sir, I want some more). The **detectors** we generally use are those best suited to the numbers of photons with which they will have to deal.

It is common practice to use **photodiodes** as detectors for absorption, extinction, and forward scatter signals and **photomultiplier tubes (PMTs)** as detectors for orthogonal scatter and fluorescence signals. In flow cytometers, whether the source is a laser or a lamp, extinction signals are detected by sampling the illumination beam on-axis. Since most light is transmitted through the sample, the light levels reaching the detector are high, and may be high enough to damage the detector unless attenuated by placing a neutral density (N.D.) filter in front of the detector. This filter should be reflective rather than absorptive so it doesn't melt or catch fire (no, I'm not kidding). The sensitivity of a PMT is rarely needed to detect extinction signals.

Objects the size of cells scatter light approximately as predicted by Gustave Mie's theory; the lion's share of the scattered light goes in the forward direction, at relatively small angles to the beam axis. This light can be detected by a photodiode if there is enough of it, i.e., if the excitation power is sufficient, and if it is in a spectral region to which a photodiode is sensitive.

Silicon Photodiodes

As was mentioned previously, a silicon photodiode produces current when photons impinge upon it; solar cells are silicon photodiodes with a large enough aggregate surface area to provide enough current to do something useful. A photodiode does not require an external power source in order to operate. The peak sensitivity of silicon photodiodes is at about 900 nm; at this wavelength, the **responsivity** of the devices, i.e., the current produced per unit of incident radiant power, is about 0.5 amperes/watt (A/W). At 500 nm, even in a "blue-enhanced" or "UV-enhanced" photodiode, responsivity is only about 0.28 A/W; at 350 nm, the figures are about 0.12 A/W for a UV-enhanced diode, 0.05 A/W for a blue-enhanced, and half that for the garden variety. What manufacturers' diodespeak calls "blue-enhanced" and "UV-enhanced" devices, incidentally, actually end up

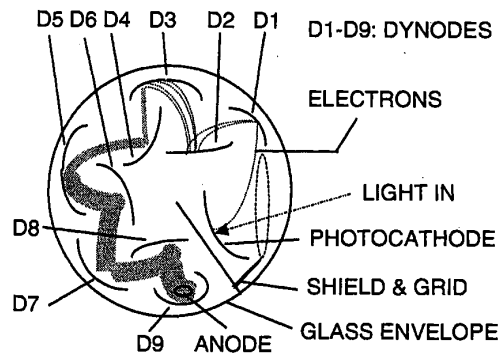


Figure 4-34. Elements of a photomultiplier tube (PMT).

with improved performance over most of the visible spectrum compared to "standard" photodiodes.

To relate responsivity in A/W into what goes on at the photon level, we need to deal with photodiodes, and with other detectors, in terms of **quantum efficiency** (Φ); for detectors, or for the photosensitive elements of detectors, this is defined as

$$\Phi (\%) = 100 \times (\text{electrons out})/(\text{photons in}).$$

By definition, $1 \text{ A} = 1 \text{ coulomb/s} = 6.2 \times 10^{18} \text{ electrons/s}$. Also by definition, $1 \text{ W} = 1 \text{ J/s}$. However, although the number of electrons in a coulomb remains constant, the number of photons in a joule varies. At 350 nm, there are 1.76×10^{18} photons/J, at 500 nm, there are 2.52×10^{18} photons/J, and at 800 nm, there are 4.03×10^{18} photons/J. Thus, the responsivity in A/W at a quantum efficiency of 100 percent (one electron out for each photon in) will vary with wavelength. At 350 nm, this responsivity would be about 0.28 A/W, at 500 nm, about 0.41 A/W, and at 800 nm, about 0.65 A/W. The actual responsivities for a typical UV-enhanced silicon photovoltaic photodiode at these wavelengths are, respectively, about 0.12, 0.28, and 0.47 A/W; we can therefore calculate the respective quantum efficiencies as 44, 68, and 72 percent. And the quantum efficiencies are what's important.

Photodiodes are usually operated in the **photovoltaic** mode, in which no external voltage is applied across the diode; they can also be run in the **photoconductive** mode, with a **bias voltage** applied. This does not increase responsivity or quantum efficiency, but speeds up response time of the device, usually with some increase in noise. The response time of photodiodes increases with their intrinsic capacitance and therefore with their size; devices with an active area of 1 mm^2 or less usually have fast enough rise times for most cytometric applications.

Photovoltaic and photoconductive photodiodes have no gain. The small currents they produce in response to incident light must be amplified electronically. This is generally

done by an active electronic preamplifier circuit that converts small input currents to proportional, but much larger, output voltages. The voltages may or may not be amplified further along in the signal processing electronics. The preamplifier and amplifier circuits I use for forward scatter measurement are probably fairly typical; they have an overall gain of about 5 million volts/ampere.

Knowing this, I know that if the 5 million V/A preamplifier's output pulse has a peak amplitude of 5 V, which is fairly typical when I measure light scattered by 2 μm plastic microspheres, using 20 mW excitation at 488 nm, the photocurrent generated in the detector was 1 μA .

Given the diode's responsivity of 0.25 A/W at 488 nm, this would mean that 4 μW of scattered light reached the detector. If I used a He-Ne laser at 633 nm, where the diode responsivity is 0.4 A/W, I'd need less laser power; when I use a big argon laser (200 mW at 488 nm) for excitation, I have to put a 1.0 neutral density filter in front of the detector to keep the signals on scale.

Photomultiplier Tubes (PMTs)

Like photodiodes, **photomultiplier tubes** produce current at their anodes when photons impinge upon their light-sensitive **photocathodes**. Unlike most photodiodes, PMTs do require external power sources; they also incorporate gain, which can be quite high (10^7 or more electrons out for each photon in). The gain mechanism of PMTs is one of the closest things in the physical world to a free lunch, which probably explains why PMTs are among the few types of **vacuum tubes** (without the vacuum, they're down the tubes) that survive in this solid state era. RCA, which developed the PMT, produced an excellent handbook on PMTs¹²³ which can be used should the reader want to flesh out the bare bones I have provided on page 54 and in what follows; Hamamatsu, which now supplies most of the PMTs used in flow cytometry, also has a handbook²⁴⁵¹.

A schematic top view of the inside of a side-window PMT is shown in Figure 4-34. The photocathode is the part of a PMT that responds directly to light; when photons hit it, it emits electrons. The photocathode, like a photodiode, therefore exemplifies the **photoelectric effect**, for studies of which Einstein was awarded his Nobel Prize in physics (relativity was regarded as too radical back then). The PMT also contains a series of electrodes called **dynodes**, to each of which is applied a potential slightly more positive than that on its neighbor dynode nearer the photocathode. There is also an **anode**, which is kept at a more positive potential than any of the other electrodes. The electrodes are arranged in space so that **photoelectrons**, i.e., electrons emitted from the photocathode in response to incident photons, are accelerated toward the first dynode by the electric field set up by the potential difference between these two electrodes.

When they hit a dynode, electrons cause **secondary emission** of more electrons from the dynode surface; these are accelerated toward the next dynode by the electrode field between the dynodes, and then produce further secondary

emission, and so on. The higher the applied potential between dynodes, the more energy is imparted to the electrons between stages; this results in emission of increasing numbers of secondary electrons from each successive dynode surface and increases the gain of the tube. Anywhere between 10^3 and 10^8 electrons may reach the anode for every electron that left the cathode, and the gain mechanism is termed "noise-free" because the dynodes only emit electrons when struck by electrons. As I mentioned on page 54, in practice, the anode is usually kept near ground potential; the photocathode is at minus a few hundred to a couple of thousand volts, and the dynodes (D1-D9 in the diagram) are at progressively higher voltages. The shield and grid shown in the diagram prevent interactions between the anode and the photocathode.

There is some noise in PMTs; most of it results from **thermionic** (i.e., due to temperature alone) **emission** of electrons from the photocathode in the absence of incident light, and from DC leakage. These generate a **dark current** that, with photoelectron and secondary electron statistics, determines the signal-to-noise ratio obtainable from the PMT. When extremely low light levels must be detected, as in single photon counting (p. 164), it is common practice to refrigerate the detector to decrease thermionic emission, thereby reducing dark current. In flow cytometry, extraneous light reaching the detector, rather than dark current, is usually the principal noise source^{8,124}; in this case, cooling the PMT won't help.

The **spectral response** of a PMT is determined by the composition of the photocathode. Tubes with **bialkali** photocathodes (some of the standard compositions are referred to as S-4 and S-5) have peak photocathode responsivity of about 40 mA/W at about 400 nm; their responsivity falls off sharply above 550 nm and is low enough at 600-650 nm to make these tubes largely unusable in this region.

Multialkali photo-cathodes, such as the S-20 type, extend the usable wavelength range to beyond 750 nm; even longer wavelength response is obtained using **gallium arsenide** (GaAs) cathodes, which provide high responsivity (50 mA/W) and a relatively flat responsivity curve between 300 and 850 nm, making them useful in spectrophotometers and spectrofluorometers. Tubes with GaAs photocathodes, however, tend to cost much more than other PMTs, and also typically have lower gain; in general, unless you are looking at fluorescence or orthogonal scatter beyond 800 nm, you shouldn't need one.

For fluorescence above 550 nm, you need multialkali tubes; they are noticeably better than bialkali tubes even for measurement of fluorescein fluorescence at 520 nm. However, you can get by with a bialkali tube for scatter measurements at 633 nm because scatter signals are so much stronger than fluorescence signals. For fluorescence or scatter measurements in the blue or blue-green, bialkali tubes, which are the cheapest, are just dandy. I'll go into more details shortly.

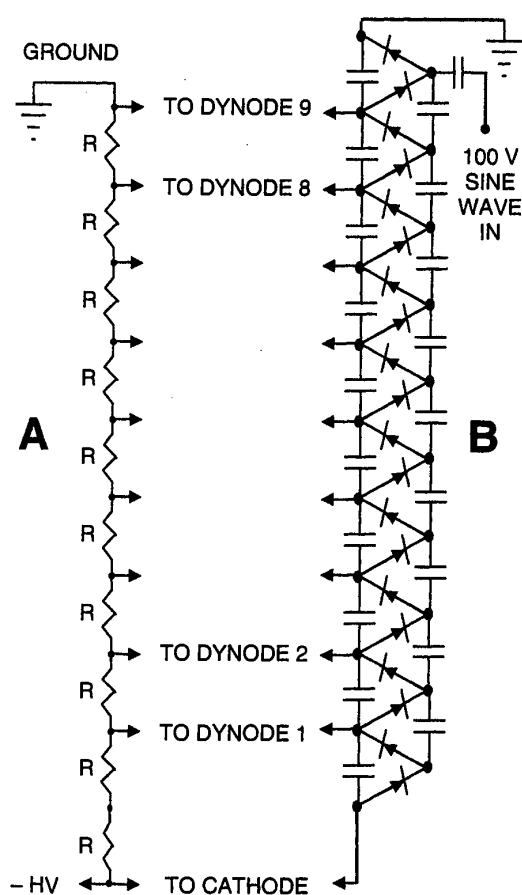


Figure 4-35. PMT electrode voltage supply circuits. Panel A: resistive dynode chain; Panel B: Cockcroft-Walton voltage multiplier.

To get signals out of a PMT, you have to plug it into a socket, which provides a way of connecting the photocathode, anode, and dynodes to the voltages required for each. Until recently, the customary way of powering a PMT involved connecting a very well regulated high voltage (300-2000 V, depending on the tube type and the gain desired) power supply to a network of resistors connected in series, known as a **dynode chain**; such a circuit is shown in panel A of Figure 4-35. The photocathode is connected to the end at the lowest potential, here indicated by $-HV$, and the dynodes are connected at every junction of resistors in between, up to the highest potential, which here is ground (0 V). Good regulation of PMT power supply voltage is essential because the relation between applied voltage and PMT gain is nonlinear (it is close to logarithmic); therefore, a change of one or two percent in applied voltage may change gain substantially.

The total resistance in the dynode chain is generally about 1 megohm (the value of the resistance R in this case is 100 K Ω), meaning the chain draws about 1 mA from a 1000

V supply. To keep PMT response linear within 1%, the anode current must be no greater than a small fraction of the current drawn by the dynode chain, say 10 μA or less. A 1994 publication¹¹⁴² from Hamamatsu pointed out that an alternative circuit, the Cockcroft-Walton voltage multiplier shown in panel B of Figure 4-35, offers several advantages over conventional dynode chains. This circuit was originally used to generate high voltages for particle accelerators, work with which earned Cockcroft and Walton the 1951 Nobel Prize in physics; it uses a network of diodes and capacitors to convert a sine wave input with amplitude of about 100 V into successively higher voltages. The multiplier circuit generates less heat than a resistive dynode chain, which results in lower noise; it also provides good linearity at anode currents well over 100 μA , and draws less input current because no resistive elements are used. Small detector modules incorporating a miniature PMT, shielded housing, and a power supply utilizing a Cockcroft-Walton multiplier are now available; one is shown in Figure 4-36 (F).

PMTs require **shielding** from stray light and magnetic fields; commercially available PMT housings provide both and may also incorporate a dynode chain, power supply, and even a preamplifier. When not in use, PMTs should be kept in the dark; exposure of an unconnected PMT to room light will affect the photocathode, increasing dark current, while exposure of a PMT to the same amount of room light while a power supply is connected will make the tube unusable, at least for purposes of cytometry.

PMT designs use a variety of physical layouts. The first PMTs were so-called **side-window** tubes, shown in Figures 4-34 and 4-36; their descendants are living and well today. A side-window tube usually has a wire grid, which blocks a small fraction of light, in front of the cathode; this **shadow effect** sometimes caused trouble in microspectrophotometry, and a lot of old hands in analytical cytology preferred **end-window** tubes (A in Figure 4-36), which don't have the obstruction. Grid shadows should not pose a problem in flow cytometry; there is no practical reason not to use side-window tubes, which tend to be less expensive than other types, and have largely replaced the end-window tubes used in earlier commercial and laboratory-built instruments. The primary selection criteria for PMTs are high gain and red sensitivity; the 1 1/8 inch diameter side-window R928, R1477, and R3896 tubes, all from Hamamatsu, are now widely used, and the last of these, which has very good red sensitivity, is made without a grid, eliminating any possibility of a grid shadow problem.

The commercial history of PMTs is enlightening. They were invented in the 1930's at RCA (the Radio Corporation of America, of which only the logo remains), and produced in small quantities until World War II, when it was discovered that the "white noise" produced by amplifying the dark current of a PMT was useful for jamming radar. This led to mass production of PMTs, making them affordable to scientific instrument builders after the war. They have been widely used in apparatus for nuclear medicine, and deeply

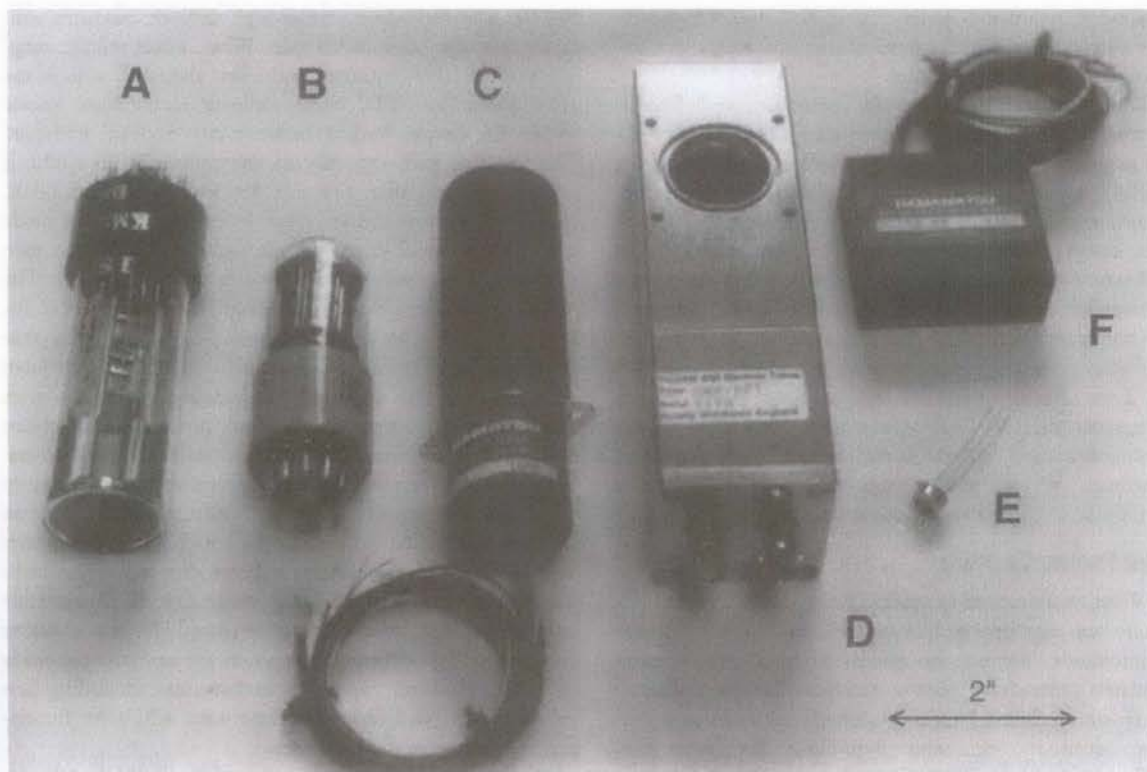


Figure 4-36. Detectors and housings. A: End-window PMT. B: 1 1/8" side-window PMT. C: 1 1/8" side-window PMT, with a magnetic shield, in a socket with a voltage multiplier. D: RF shielded housing for a side-window PMT, with a dynode chain. E: Silicon photodiode. F: Detector module with small side-window PMT and voltage multiplier power supply.

(underground) used by physicists hunting for neutrinos and various kinds of quarks. We're not the only ones analyzing rare events.

When RCA was phagocytosed by General Electric, the RCA division which made PMTs was bought out by its management, and became Burle Industries; Burle has chosen to focus on the relatively large market for PMTs for scintillation cameras and does not now make affordable red-sensitive PMTs, but it has at least kept the *Photomultiplier Handbook*¹⁴³ alive (it can now be found at www.burle.com).

More recently, the PMT division of the British company Thorn EMI, which supplied the end-window tubes in the older B-D FACS instruments, was bought out by its management, becoming Electron Tubes, Ltd.; this may have staved off extinction but didn't do much for fluorescence. Luckily, Hamamatsu, which has become pretty much the only game in town, especially for red-sensitive tubes, appears committed to manufacturing PMTs usable for cytometry. The industry probably can't sell more than a few thousand tubes a year for flow cytometry, although other fluorescence instrumentation has the same requirements and expands the market somewhat; some larger companies obviously haven't thought the sales volume was worth the effort. If everybody thinks that way, we're in trouble.

Sensitivity Training: Photodiode versus PMT

The following arithmetic should provide you with some feel for why PMTs are used as detectors for relatively weak signals, e.g. fluorescence and orthogonal scatter. When I ran my big laser source instrument using 200 mW of excitation at 488 nm, with about 300 volts applied to a fluorescence detector PMT, and associated electronics with an overall gain of 1 million volts/ampere, I got 10 volt pulses from 3.7 μ m fluorescent polystyrene spheres; the output (anode) current from the photomultiplier was therefore 10 μ A. To find how much light was detected, I have to factor in the gain of the PMT. Operating at 300 V, the gain is about 1000; a PMT anode current of 10 μ A thus corresponds to a cathode current of 10 nA. Since the PMT photocathode responsivity is 30 mA/W, this indicates that 0.333 μ W of light reached the PMT cathode. This, incidentally, is about 8 percent of the forward scatter signal intensity calculated on p. 161.

While a photodiode is much more efficient at converting light to current than is a PMT cathode, I would only get about 80 nA of output current from a photodiode, even when it responded to the exceptionally bright fluorescence signals produced by the microspheres. I could certainly get away with using a photodiode as a detector in this case, and

I and a few manufacturers have used diodes with high-gain electronics to detect orthogonal scatter and bright fluorescence, as well as forward scatter.

Diodes are much harder to use to measure weak fluorescence signals, e.g. immunofluorescence. These signals typically require at least 100 times as much electron gain from the PMT to yield a 10 V output signal as do the bright fluorescence signals just discussed. Since a conventional photodiode doesn't have gain available, I'd have to work with <800 pA output from the diode if I wanted to use it as a fluorescence detector. That's practically impossible, even with the most meticulous electronics design; there are too many stray currents running around which will swamp 800 pA. Even dealing with 80 nA requires pretty careful electronics design and construction practice; while photodiodes cost less than PMTs and their associated electronics, if you add enough fancy electronics to a photodiode, it may be easier and cheaper to use a PMT.

Single Photon Counting

When you do need to squeeze every last photon out of a sample, you may have to bite the bullet and do the job with a photodiode, because the quality of your measurements ultimately depends on photon statistics. The more photons you detect, the better your measurement precision, signal-to-noise, sensitivity, etc., and photodiodes are better than PMTs at converting photons into electrons.

The ultimate light detection task is **single photon counting**, where you literally want to pick up the little buggers one at a time. This technique is used for bioluminescence measurements; it has also been employed^{660,1144} by the Los Alamos group in detection of single phycoerythrin molecules and very small (≤ 5 kilobase) fragments of DNA labeled with fluorescent dyes. In single photon counting, each photoelectron released from the detector cathode generates a current pulse; sensitivity is limited by dark current, and is improved by refrigerating the detector, which decreases thermionic emission.

In flow cytometry, all cells of the same size spend, or at least should spend, the same amount of time in the illuminating beam; different numbers of fluorescence photons will therefore be collected from cells bearing different amounts of fluorescent dyes. Since the precision of any individual measurement is a function of the number of photons counted, measurements of lower fluorescence intensities are less precise than measurements of higher intensities. In a static cytometer, it is feasible to measure fluorescence with uniform precision by illuminating each cell until a preset number of photons have been counted; fluorescence intensity can be determined from the time required to accumulate the preset photon count. This technique has been used by Deutsch and Weinreb¹¹⁴⁵ for fluorescence polarization measurements.

Avalanche Photodiodes (APDs)

The **avalanche photodiode (APD)** is a detector which combines some desirable properties of photodiodes and

PMT's; like the former, it has high cathode quantum efficiency, like the latter, it has gain. When a bias voltage ranging between a few hundred and a few thousand volts is applied across an APD, some electron acceleration occurs within the device, leading to secondary electron emission. The resulting gain, typically on the order of 100-1,000, is considerably less than one can get with photomultipliers. However, since the cathode quantum efficiency of the diode is roughly an order of magnitude higher, an APD at a gain of 1,000 is as good as a PMT at a gain of 10,000. The APD's requirement for a bias voltage is something of a disadvantage, but the current required is lower than that needed by a PMT. APDs also suffer from relatively high dark currents, and their gain fluctuates considerably with slight variations in temperature. Both of these problems can be dealt with effectively by controlled cooling of the devices. However, once you add the bias supply and cooling circuitry to an already expensive APD, you end up paying about as much as you would for a PMT with its housing and power supply. Thus far, calculations indicate that the higher cathode quantum efficiency could make the APD assembly competitive with, or even superior to, a PMT as a sensitive detector for flow cytometry; experiments have not yet made the case. However, several manufacturers, including Luminex, Partec, and Optoflow, have used APDs for fluorescence detection in their products.

The "avalanche" in avalanche photodiode describes a phenomenon that occurs at relatively high applied bias voltages; liberation of a single photoelectron at the cathode leads to a massive electronic catharsis, followed by a refractory period. The response of the APD in this "Geiger" mode is markedly nonlinear, but well suited to single photon counting. The "Geiger" metaphor refers to the Geiger counter, in which ionization of gas by radiation produces a similar electrical breakdown phenomenon.

RCA had a division in Vaudreuil, Quebec, Canada (a suburb of Montreal) that developed single photon counting modules incorporating cooled APDs. These modules were and are used in the high sensitivity, low flow rate flow cytometers built at Los Alamos for sizing DNA fragments¹¹⁴⁴. The output of the devices is digital; detection of a photon results in output of a logic pulse, and the subsequent data processing hardware and software used at Los Alamos are substantially different from those used in conventional flow cytometers. Agronskaia et al²⁴⁵² described a circuit that allowed them to convert signals from an APD photon counting module to analog pulses, facilitating a comparison of conventional detection using a PMT and photon counting with an APD module for nucleic acid sizing in a slow flow system. Photon counting, as one might expect, gave better resolution of low intensity peaks.

Using APDs in "Geiger" mode for photon counting will generally not provide the dynamic range available from a PMT, because the APD itself becomes insensitive to photons for a brief period after each pulse is generated. There may be workarounds for this. If you're interested in laying out a few

	λ (nm)	300	350	400	450	500	550	600	650	700	750	800
Photons/J $\times 10^{-8}$		1.51	1.76	2.01	2.27	2.52	2.77	3.02	3.27	3.53	3.77	4.0
mA/W @ $\Phi = 100\%$		244	284	324	366	406	447	487	527	569	608	650
DETECTOR	GAIN	QUANTUM EFFICIENCY (Φ) (PERCENT) AT TABULATED WAVELENGTH										
UV-enhanced diode	1.0	41	44	46	61	68	73	77	76	75	74	72
931B PMT	8.0×10^4	3	18	18	14	10	6	2	0.2	—	—	—
9798B PMT	1.3×10^4	15	17	16	13	11	8	6	3	2	1	—
R928 PMT	1.0×10^7	25	22	21	17	15	12	9	7	6	4	2
R4457 PMT	1.2×10^7	22	20	19	17	15	13	11	9	3	0.6	0.1
R1477 PMT	1.0×10^7	22	22	22	21	20	16	13	11	9	6	4
R3896 PMT	9.5×10^4	29	28	26	23	21	19	15	12	10	8	6
R6357 PMT	4.0×10^4	32	30	29	27	25	22	18	15	11	7	4
R636 PMT	2.7×10^4	22	21	16	13	11	9	8	7	6	5	5
C 972 Channel PMT	2.0×10^8	3	4	4	5	6	7	8	7	5	3	1.5

Table 4-3. Cathode quantum efficiencies of diode and PMT detectors between 300 and 800 nm. The R928, R4457, R1477, R3896, R6357, and R636 are made by Hamamatsu. The 9798B is by Electron Tubes; the C972 is made by PerkinElmer Optoelectronics. Burle, Electron Tubes, and Hamamatsu all make the "industry standard" 931B.

thousand bucks for one of the APD modules from Vaudreuil, you'll need to contact PerkinElmer Optoelectronics; RCA was bought by GE, which sold the Vaudreuil division to E G & G, which changed its name to Perkin-Elmer after an acquisition and dropped the hyphen some time later.

PMTs: Picking a Winner

Table 4-3 shows quantum efficiencies at various wavelengths between 300 and 800 nm for a typical UV-enhanced silicon photovoltaic photodiode and for the photocathodes of a variety of PMTs; maximum available gains are also tabulated. The number of photons/J and the responsivity in mA/W for a quantum efficiency of 100% are also tabulated. I had similar tables in the second and third editions; some of the PMTs in them have become irrelevant, but there are some new and interesting varieties on the market.

The venerable 931B was developed by RCA, and is now made by Burle, Electron Tubes, Hamamatsu, and possibly others. It's a 1 1/8" side-window type with a bialkali photocathode and costs under \$100. Thorn EMI's 9798B is an end-window type which was used as the red fluorescence detector in older B-D FACSeS. The R928, R1477, and R3896, all from Hamamatsu, are 1 1/8" side-window PMTs with multialkali cathodes; the R3896, as noted on p. 162, does not have a grid in front of the photocathode. The Hamamatsu R4457 and R6357 are older and newer miniature (1/2") multialkali side-window tubes used in detector modules such as that shown in Figure 4-36(F); the R6357 is a gridless design. The 1 1/8" side-window Hamamatsu R636 has a gallium arsenide photocathode.

The C 972 channel photomultiplier is made by Perkin-Elmer Optoelectronics (sans hyphen; see left). It has a photocathode, but, in place of a series of dynodes, it substitutes a narrow semiconductive channel, across which a high voltage is applied, between the cathode and anode. Photoelectrons liberated from the cathode hit the walls of this channel, releasing secondary electrons, etc., etc., with all the free electrons getting accelerated toward the anode. The net effect is that the device behaves as if it had a whole bunch of dynodes; gains can be as high as 2×10^8 . The concept is interesting, but, as can be seen from Table 4-3, the cathode quantum efficiency is pretty low, and you can't amplify photoelectrons unless you generate them at the cathode first. Noise-free gain is a great concept, but 10^6 times zero is still zero. I'll probably have to try a channel photomultiplier to satisfy my curiosity.

The table shows that a diode's quantum efficiency is only about twice that of the best PMT photocathodes at 300-400 nm, about 4 times at 550 nm, and over 10 times beyond 700 nm. At this point, the quantum efficiency of bialkali tubes, such as the 931B, has dropped below 0.1%.

The widely used R928 costs about \$400; the souped-up R1477, a selected R928, doesn't cost much more, and is therefore preferable. Hamamatsu now pushes the R3896, a stellar performer; its initial equally stellar (about \$1,000) price has come down since the last edition of the book came out, and I'd recommend it, particularly for measurements of really long wavelength labels, such as PE-Cy7 and APC-Cy7. Out in their territory, at 800 nm, APDs actually give PMTs more of a run for the money than is the case at shorter wavelengths.

The gallium arsenide R636 is useful in spectrofluorometers because it has a relatively flat response curve, but its maximum gain is quite low; unless you need to work at 900 nm, the R1477 and R3896 are better choices. For most work at or below 500 nm, or for scatter measurements out as far as 633/635 nm, the bargain-priced 931B will do a fine job, but, for measuring fluorescence anywhere above 500 nm, a tube with a higher quantum efficiency is worth its price. I found, when I excited propidium with less than 5 mW at 488 nm, that the extra quantum efficiency of an R928 (vs. a 931B) helped lower measurement CVs.

I have already mentioned the compact Hamamatsu detector modules that incorporate 1/2" PMTs. The current H7710-03 features an R6357 PMT; other, less expensive modules in the series are made with less spectacular tubes, which will probably be fine at 550 nm and shorter wavelengths.

Hamamatsu has also gone in some other interesting directions in PMT development. They have made ultraminiature PMTs that fit into the 16 mm diameter, 12 mm long TO-8 "can" package normally used for transistors and diodes. The first generation of these tubes had neither high gain nor high fluorescence sensitivity, but the newest offerings, the R7400U series, include at least one tube with high red sensitivity; however, while the quantum efficiency of this tube is competitive, the gain (5×10^5) is still on the low side. These PMTs are also available in modules; I have been told that neither the tubes nor the modules are significantly cheaper than the larger varieties.

The other notable Hamamatsu offering is a **multianode** PMT, with a square or linear array of anodes and fine mesh dynodes. The different anodes respond to light impinging on different areas of the cathode, at least up to a point. The linear array multianode PMT can receive the light dispersed from a grating, with the outputs from the different anodes then providing spectral information. Zeiss has apparently used a multianode PMT in a spectral detector for its Meta confocal microscope system; I have also heard of one being used in an experimental flow cytometer.

Photomultipliers: Inexact Science

After all this discussion of PMT sensitivity, I am obliged to let you in on one of the dirty little secrets of electro-optics; the tabulated values are a rough guide. There is a lot of variation from device to device in most of the important parameters; cathode sensitivity and gain for a given applied voltage will vary over at least a 2:1 range, and individual variations in photocathode composition make for individual deviations from the spectral response curves of Table 4-3. The good news is that manufacturers test the sensitivity of individual PMTs and provide the results to the buyer. So, if you acquire two R3896's, you probably want to use the "hotter" one at the longer wavelength. My impression is that plain silicon photodiodes don't vary nearly as much as PMTs, although avalanche diodes may.

Charge Transfer Devices: CCDs, CIDs, Etc.

You are, by now, likely to have encountered the **charge coupled device**, or **CCD**, either in its low-cost form in your camcorder or digital camera, or in its rarer, cooled, more esoteric and expensive guise in imaging cytometers designed for low light level measurements. CCDs are one of a class of photodetectors described as **charge transfer devices**; there are also, for example, **charge injection devices**, or **CIDs**.

In all of these, exposure to light causes accumulation of electric charge in individual elements that are usually arranged in a linear or rectangular array; attached electronic circuitry senses the amount of stored charge in each element at regular intervals. Charge transfer devices are well suited for imaging; because they integrate over time, they are useful for measurement of low light intensities, especially when cooled. However, they tend to be relatively slow, and, on that account, they have not been widely used in flow cytometry. Newer, faster arrays may be useful in polychromatic detection for measurement of emission spectra in flow¹¹³⁹.

I hear that there are now ways of getting gain out of CCDs, but I don't have either details or confirmation. New **CMOS image sensors** are starting to give CCDs a run for their money in the commercial camera markets; whether they will make inroads in science remains to be seen.

Intel, which joined forces with Mattel to produce the QX3 Computer Microscope, a cute toy that uses a CCD to provide 320 by 200 pixel images, decided in late 2001 to stop making the gadgets; they came on the market at \$119.95, and I've snapped up a few for \$49.95. There is still time to introduce your kids or grandkids to microscopy and cytometry via this route.

4.6 FLOW SYSTEMS

It has probably not escaped your notice that, to this point, in this book on flow cytometry, I have gone into great detail about light, optics, light sources, lenses, filters, and detectors and said very little indeed about **flow systems**, without which flow cytometry wouldn't be flow cytometry. I claim there has been a method to this madness. All the other stuff doesn't change just because you work in a flow system, and all the other stuff works when you don't work in a flow system. You can use the same light sources, and the same lenses, and the same filters and detectors, to illuminate and collect and detect light from cells on slides, or in culture dishes, or in microtiter (or nanotiter) plates, or in small capillaries, as you use to do the same jobs for cells in flow systems. A few chapters from now, we will discuss parameters and probes, virtually all of which can be measured in or applied to cells in static as well as in flow cytometers. There are some cytometric tasks for which flow cytometry is preferable, and some for which it is not, but most of the fundamentals of flow cytometry are the fundamentals of cytometry in general. Among those that are not are the theoretical and practical details of **fluid mechanics** or **hydrodynamics** and flow systems, to which we now turn.

In a flow cytometer, it is the task of the flow system to transport cells in the sample to and through the measurement station(s). In "static" microspectrophotometers or image analysis systems, the same job is usually delegated to precisely made and well-controlled mechanical hardware. However, while the mechanical transport system in a static cytometer may be inactive while actual measurements are being made, the flow system in a flow cytometer is continually active, and must move the entire cohort of cells in a sample past the measurement station(s) along almost identical trajectories at almost identical velocities if satisfactory data are to be obtained from the measurement process. This requires that a stable flow pattern be achieved and maintained, and both designers and users of flow cytometers must play active roles in this process.

Flow System Basics

The design of flow systems and the underlying physical principles have been discussed at length by Pinkel and Stovel⁶⁵⁰ and by Kachel, Fellner-Feldegg, and Menke¹¹⁴⁶. If you feel a strong urge to design your own flow cytometer, you will probably want to refer to one or both of those publications. If you're willing to put up with what the manufacturers give you, and/or to do things my way, stick with me, and I will expand on the brief discussion of flow systems that appeared on pp. 55-57, hoping to hit the high points of the references just cited.

Almost all modern optical flow cytometer designs make use of **sheath flow**, or **hydrodynamic focusing**, to confine the sample or **core** fluid containing the cells to the central portion of a flowing stream of cell-free **sheath** fluid. Sheath flow improves the precision with which the cell sample can be positioned in the **observation region** of the cytometer by restricting cells to the central region of the stream, and reduces the likelihood of obstruction of the flow system. Stable, unobstructed flow minimizes variations in the position and velocity of the core stream; when flow becomes unstable or **turbulent**, due to obstruction or other causes, measurements are likely to become imprecise and inaccurate.

Figure 4-37 illustrates some aspects of fluid flow in a flow cytometer. Core (sample) and sheath inlet tubes are shown near the top of the flow chamber; near where these enter, there may also be a third port which can be connected to vacuum, allowing easy removal of air bubbles and back suction to clear clogs out of the orifice. Application of vacuum is more likely to be successful for the first purpose than for the second.

Gently Down the Stream: Laminar Flow

Flow must be stable from the region of the core injector tip downward if core velocity and position are to be maintained well enough to allow good measurements to be made. We want **stationary** or **streamline flow**, a condition characterized by the constancy over time of flow velocity at any given point in the system. Since the law of conservation of mass dictates that the same volume of fluid must pass in the

same time through the narrow and wide portions of the capillary, the flow velocities at different points in the system will be different, i.e., higher in the narrow portions than in the wider ones. In fact, the product of cross-sectional area, A , and average flow velocity, v , remains constant and equal to the volume flow rate, Q , at any point along the flow system. But why are we talking about "average" velocity?

Water, which is the major component of both the sheath and core fluids and which therefore determines their flow characteristics, is not what physicists call an ideal liquid; it exhibits **viscosity**, which, in physical terms, means that some work must be done on a volume of fluid to get it to change its shape. While the everyday definition of viscosity conjures up fluids such as glycerin, which has a viscosity about 1000 times that of water, the effects of the viscosity of water on its pattern of flow are noticeable enough. In particular, we observe that the stationary flow of water through small tubes is **laminar**. If we look at a cylindrical tube of radius R containing flowing water, we find that the velocity of water at different distances from the axis or center of the tube varies. Velocity is highest along the axis; at the walls of the tube, there is actually a thin **boundary layer** of water that is not moving (i.e., it has zero velocity). At any intermediate point a distance r along the radius, the velocity is proportional to $(R - r)^2$. This produces a so-called **parabolic profile** of flow velocities, as if the water were broken up into thin cylindrical layers (**laminae** in Latin) that were sliding over one an-

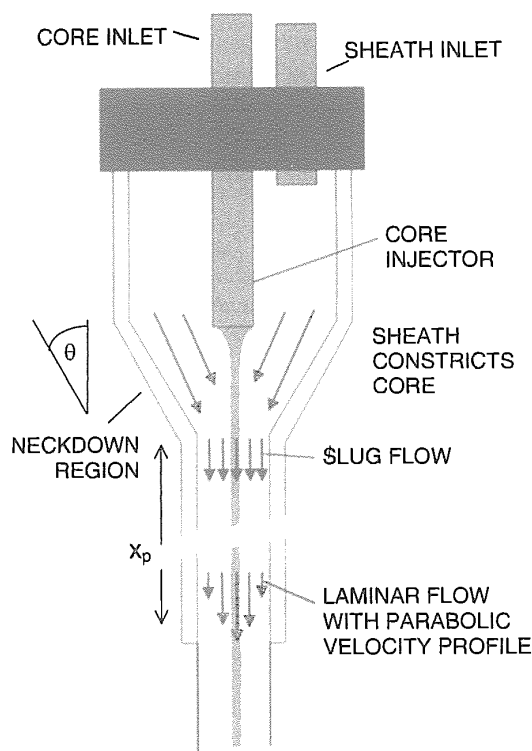


Figure 4-37. Fluid flow in a flow cytometer.

other. So, we can't assume that the velocity of the fluid will be constant across the entire cross section of the tube, but we can do our calculations based on average velocity and use the conservation law.

I can't think of a better illustration of a laminar flow profile than Figure 4-38. This shows a chain of diatoms (*Thalassiosira spp.*) in water flowing at 10 mL/min through the 3 mm wide, 300 μ m deep flow chamber of the "Flow CAM," an apparatus developed by Sieracki et al.²⁴⁵³ at the Bigelow Laboratory for Ocean Sciences, Boothbay Harbor, ME, and now available commercially from Fluid Imaging Technologies. The instrument stores a digital image of each particle measured, in addition to fluorescence and size data.

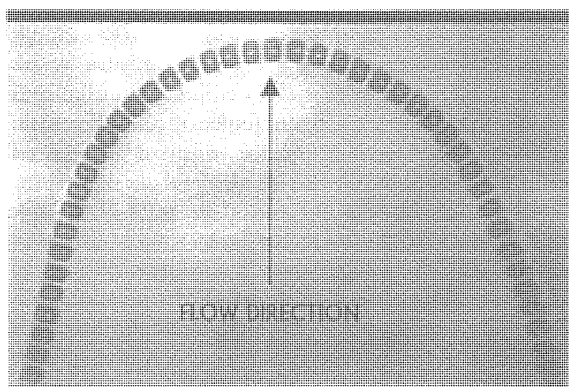


Figure 4-38. Laminar flow profile illustrated by diatoms in the "Flow CAM" imaging flow cytometer. Contributed by Chris Sieracki, Bigelow Laboratory for Ocean Sciences.

Returning to Figure 4-37, we see that at some point near the core injector tip, the cross section of the flow chamber is gradually decreased; the length of the chamber over which this happens has been called the "neckdown region." It's a grotesque name, with vaguely Rabelaisian overtones, but it will do. As the cross section decreases, the flow velocity increases; also, the ratio of core cross section (or diameter) to sheath cross section (or diameter) may be changed, depending upon the relative **volume flow rates** of sheath and core. What we are aiming for is a core of small enough diameter so that cells generally pass through the observation region one at a time; what we most want to avoid in the neckdown region is anything that will generate **turbulence**. Sharp edges and/or sudden changes in diameter will do that, and are to be avoided. In terms of design, a neckdown region with a gentle conical taper ($\theta \approx 30^\circ$) is good. People have introduced various dodges such as tapered and/or eccentric injectors in order to orient asymmetric cells in flow^{650,651}; for now, we will stick to the basics.

The figure shows an extension of the capillary past the tip of the neckdown region; this is the configuration used in Ortho's early Cytofluorographs and in my Cytomutts, with round capillaries, and in most modern benchtop instruments, with rectangular cuvettes. In stream-in-air systems,

such as the B-D FACS and Coulter EPICS series sorters, the stream emerges through an orifice placed at the end of the neckdown region.

When the cross section of a tube through which a viscous fluid is flowing decreases, the velocity profile at the point of entrance to the constricted region is nearly constant across almost the entire cross section; this is referred to as **slug flow**. The fluid must flow for some distance x_p through the constricted portion of the tube before the parabolic flow profile reestablishes itself. Pinkel and Stovel⁶⁵⁰ state that, for water at 20° C, this distance, in mm, is

$$x_p = 6 \times 10^{-5} d^2 v,$$

where d , in μ m, is the diameter of the constricted portion of the tube, and v , in m/s, is the average fluid velocity. For a flow velocity of 10 m/s, which is a common value in flow cytometers, values of x_p for (constricted) tube diameters of 70, 100, and 250 μ m, are, respectively, 3, 6, and 38 mm. This means that in a system in which cells are observed in a stream in air after exiting a 70 μ m orifice, the parabolic velocity profile will be established at the observation point if it is 3 mm below the orifice. In an instrument in which cells are observed inside a 250 μ m square quartz cuvette less than 38 mm long, the observation point is necessarily in a region where the flow velocity profile is not parabolic. This can be advantageous, because velocity differences between cells at different distances from the core axis will be minimized.

In laminar flow, the flow in the region of the boundary layer, i.e., the region near the tube walls, is inherently unstable, as a result of which any irregularities of the walls, or adherent particles or bubbles, may produce turbulence. If you happen to have a flow system handy, you can actually observe the flow pattern (sheath flow, I hope) by running an aqueous solution of a dye such as methylene blue through the core injector. By manipulating the core injection rate, you should be able to produce stable cores of varying diameters. If you see the core wiggling, or diffusing into the sheath, your flow system needs work. It is critical, by the way, to use an aqueous solution of dye for this exercise; watching the behavior of an ethanol or methanol core and an aqueous sheath can make you think your flow system needs an exorcist when absolutely nothing is wrong.

The crew I worked with at Block Engineering probably developed an appreciation of the importance of flow stability for good instrument performance earlier than did most other people, for the simple reason that we were always looking at measurements made at two or more observation points. Using an oscilloscope, we would observe waveforms from the detectors at two observation points, using the waveform from the first (upstream) detector to trigger the display. When the velocity was constant, the interval between pulses from the two detectors was constant; when flow was disturbed, the second pulse moved with respect to the first (the effect is called **jitter**), because cells took different times to traverse the distance between observation points. We built injector assemblies to allow changing the injector position

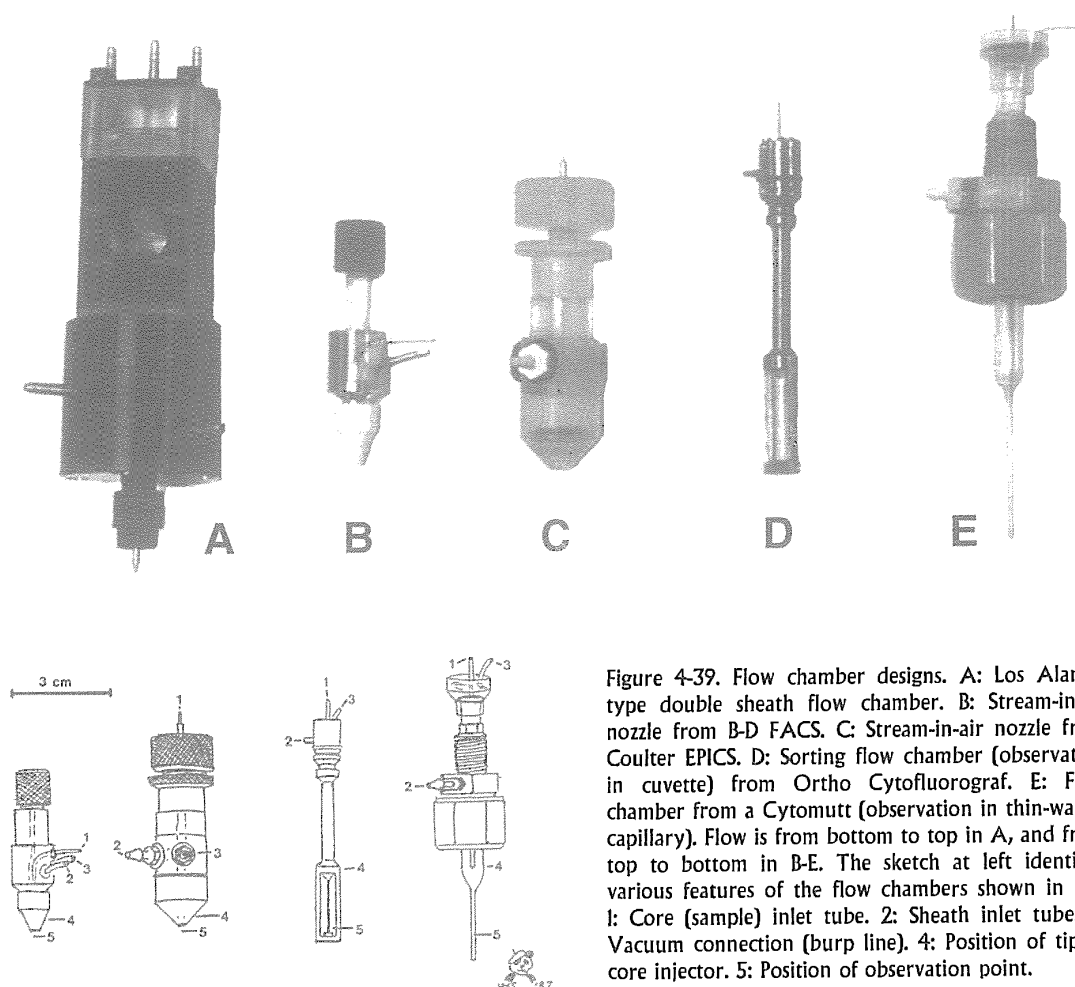


Figure 4-39. Flow chamber designs. A: Los Alamos type double sheath flow chamber. B: Stream-in-air nozzle from B-D FACS. C: Stream-in-air nozzle from Coulter EPICS. D: Sorting flow chamber (observation in cuvette) from Ortho Cytofluorograf. E: Flow chamber from a Cytomutt (observation in thin-walled capillary). Flow is from bottom to top in A, and from top to bottom in B-E. The sketch at left identifies various features of the flow chambers shown in B-E. 1: Core (sample) inlet tube. 2: Sheath inlet tube. 3: Vacuum connection (burp line). 4: Position of tip of core injector. 5: Position of observation point.

(core steering), allowing us to place the core in the position that gave the stablest flow pattern. This was not always in the geometric or optical center of the round capillary used in our flow chambers!

The adjustment mechanisms in the instruments I myself have built are cruder, although they do not get to the level of needles stuck through corks¹⁰². This makes it harder to arrive at the best core position, and harder to maintain it, but, when you are there, you can get good measurement precision (coefficients of variation 1.5 percent or less) even with needles stuck through corks (O. Bakke, personal communication; D. Pinkel, personal communication; L. Scherr, personal communication; H. Steen, personal communication).

Flow Chambers; Backflushes, Boosts, and Burps

Flow chambers, or flow cells (the latter term invites confusion with biological cells and I will avoid it), used in various instruments are shown in a photograph and scale drawing in Figure 4-39. The flow chambers shown were or

are used only with laser source flow cytometers, except for the Cytomutt chamber, which has been used with both laser and arc lamp light sources. The chamber shown at A is similar to those built for apparatus at Los Alamos. Observation is done in a 1 cm diameter photometer cuvette; a second sheath, moving at very low velocity, is used to fill much of the volume of the flow chamber.

The B-D FACS and Coulter EPICS flow chamber designs at B and C are used to generate streams that are observed in air after they emerge from an orifice; both the flow chamber and the orifice are sometimes referred to as **nozzles**.

In the EPICS flow chambers and older B-D nozzles, the orifice is formed by a sapphire watch jewel; the shape of such an orifice is not a cylinder, but a truncated cone. The proper orientation for a watch jewel in a flow chamber is with the larger diameter toward the outside; this substantially reduces surface tension that would otherwise cause the emerging stream to spread out. The original Stanford apparatus used a quartz orifice; the B-D FACS flow chamber shown uses a

ceramic orifice. It is also possible to sort cells emerging from thin-walled capillaries such as are used in the Cytomutt flow chamber shown at E. Sorting flow chambers are typically mounted in a bracket that also holds the transducer used to stabilize the droplet breakoff pattern.

A major difference between the stream-in-air nozzles and the other flow chambers shown in the figure is the much shorter distance between the injector tip and the stream exit from the flow chamber in the former. During sample changes, after a sample is removed, it is customary to let sheath fluid flow back through the sample inlet tube, or, in a system using a syringe pump for sample feed, to suck back on the feed syringe. These **backflushing** maneuvers will clear the sample line of cells from that sample, but will not remove any cells that had already left the injector but had not yet emerged from the chamber when the sample was removed. It is thus common practice to drive some of the new sample through the system at a higher than normal rate, either by briefly increasing the sample drive pressure or, if a syringe pump is used for sample feed, by increasing its delivery rate; the accelerated sample delivery is called a **boost**.

Both backflushing and boosting can introduce turbulence, i.e. disrupt laminar flow, which, among other things, may result in some of the cells from the old sample becoming mixed with the cells in the new sample. The shorter the distance between the injector tip and the exit or observation point, the less such **sample carryover** should occur. Stream-in-air flow chamber designs are therefore well suited for high throughput operations. Graves et al, who have been working for some time in this area, recently examined the relationships of various nozzle design parameters on the time taken for flow to stabilize between samples²⁴⁵⁴. They found that flow stabilized more rapidly when a large (150-200 μm) orifice diameter was used, because, when a smaller (50 μm) orifice was used, the boost generated higher back pressure, causing turbulence.

The original B-D sorting nozzle design lacked a connection through which air bubbles might be bled out of the system; removing bubbles required demounting the flow chamber and inverting it, which generally sprayed sample all over the place. In the HIV era, this doesn't even play in Peoria. Most newer flow chamber designs incorporate a bleed or "burp" line that can be connected to a vacuum line or a syringe.

Most stream-in-air designs have no provision for core steering, i.e., changing the core position laterally with respect to the sheath stream. People who used them told me that some of the older B-D nozzles consistently gave better measurement precision than others; I would take this to be an indication of how well the core was centered.

The Ortho flow chamber shown at D was used for observation of cells within a quartz cuvette with a square cross section; the channel is approximately 200 μm across. A 75 or 100 μm orifice was generally used for sorting; the observation point is positioned just above the orifice jewel, which is held, with the cuvette, in a steel cage. The flat surfaces in the

cuvette minimize scatter of excitation light from the flow chamber walls; this reduces noise in scatter measurements.

The Cytomutt flow chamber shown at E is based on the designs used in the Block Cytomat systems^{88,92}. Observation is done in a thin-walled quartz capillary, typically between 150 and 175 μm in inside diameter, which is held in a standard compression tube fitting. The same capillaries can be used for both analysis and sorting; for sorting, the observation point is usually moved down near the tip of the capillary to minimize the distance between it and the droplet breakoff point. The sample and sheath injection ports are fit into the tube fitting and sealed with epoxy glue. The core injector is a piece of 27 gauge stainless hypodermic tubing glued into the bottom portion of the barrel of a plastic Luerlok syringe; core position is adjusted by twisting this fitting. Bubbles are removed through a vacuum connection. The large bore of Cytomutt flow chambers makes them less likely to clog than most sorter nozzles, and clogs (or bubbles) are usually readily dislodged by running a piece of stainless steel wire down through the injector, without removing the flow chamber. This largely eliminates the need to realign the system after a problem has been dealt with.

Cuvettes vs. Streams for Analysis and Sorting

Observation in a stream in air rather than in a cuvette offers potential advantages. There are two fewer interfaces from which light may be scattered in a stream in air system, and there are no cuvette or capillary walls that might get scratched or dirty. Proponents of stream-in-air interrogation for systems in which sorting is to be done argue that the greater distance between the observation point and the orifice in a design such as Ortho's leads to less accurate sorting. This need not be so; observation within the walls of the flow chamber allows more power to be used to drive the transducer than would be acceptable in a stream-in-air system, in which high power levels distort the stream at the observation point, degrading its optical quality and decreasing measurement precision. Sorting accuracy depends on maintaining the relative positions of the interrogation point and the droplet breakoff point; increased transducer drive may allow better control of this factor.

When the objective is high speed sorting, stream-in-air systems have been preferable, because they are compatible with higher analysis rates. The flow rate of sample through the observation point in the Ortho chamber was relatively low; velocity through a 75 μm orifice was 10 m/s, meaning that flow velocity in the 200 μm square cross section cuvette was only about 1.1 m/s. The Ortho system used a beam focused to a spot less than 10 μm high, allowing useful pulse width information to be derived, but cell transit time through the beam was around 9 μs , making it essentially impossible to process 100,000 cells/s. Beckman Coulter's Altra and BD Biosciences' FACSAria, newer sorters utilizing cuvettes, provide shorter transit times and can process tens of thousands of cells/s. However, the transit time through the beam in modern high speed stream-in-air sorters can be

less than 1 μs , still potentially yielding a performance advantage for the stream-in-air systems.

On the minus side, stream-in-air systems generate aerosols, a concern when specimens contain potentially hazardous materials; the optical characteristics of the stream are also subject to change when the sheath is turned off and on. Light collection from streams in air is also usually less efficient than is light collection from flat-sided cuvettes. Some cytometer manufacturers allow the user to make the tradeoff.

Flow cuvettes with square or rectangular cross sections, designed for analysis rather than sorting, are now standard in commercial benchtop fluorescence flow cytometers. Such chambers were first used in Technicon's Hemalog D differential leukocyte counter. In the mid-1970's, Ortho introduced the FC-200 Cytofluorograf, featuring a chamber with a 200 μm square cross section; a spherical mirror on the side of the cuvette opposite to the collector lens for fluorescence and orthogonal scatter signals could be used in this system to increase light collection. Similar designs were used in Ortho's System 30, Spectrum, and Cytoron analyzers. Coulter offered a closed system analytical flow chamber for the EPICS sorters; the large cross section allowed cells to be analyzed at lower flow velocities for time-of-flight cell sizing measurements¹⁰⁴.

Observation in square cuvettes is standard in the Beckman Coulter Elite and Altra sorters and EPICS XL analyzers; the cuvettes have the front element of the collection lens built in. The new DakoCytomation CyAn analyzer uses a square cuvette without an integral lens. The flow chambers used in the new BD LSR series analyzers, and in the older FACSCalibur, FACScan, FACSsort, FACSTrak, and FACSCount, have a rectangular cross section, roughly 180 by 400 μm , permitting use of a high-N.A. (1.2) lens, which increases light collection substantially. BD's new FACSaria sorter and the new Beckman Coulter FC500 analyzer also have rectangular cuvettes with coupled high-N.A. lenses.

Dan Pinkel, then at Lawrence Livermore Laboratory, designed a sorting flow chamber with an observation portion of square or rectangular cross section; the quartz cuvette (made by NSG Precision Cells) fits into a holder that allows it to be mounted to a B-D transducer mount. Although this chamber initially produced good results in analysis, there were some problems fitting watch jewels to it for sorting; these were eventually resolved⁶². A design similar in appearance to Pinkel's was offered by Coulter for the EPICS; this utilizes a cuvette only a few millimeters long with a square channel 76 μm across, attached where the tip of the nozzle is shown in Figure 4-39. This chamber design decreases noise in orthogonal scatter measurements, as compared to a stream-in-air system, and still permits sorting to be done at reasonably high speeds. The Beckman Coulter Elite and Altra sorters can be fitted with this cuvette, with 150 or 250 μm square flow cuvettes, or with stream-in-air nozzles. DakoCytomation sorters restrict observation to streams in air; all three manufacturers provide the same range of orifice diameters, i.e., 50 to 400 μm .

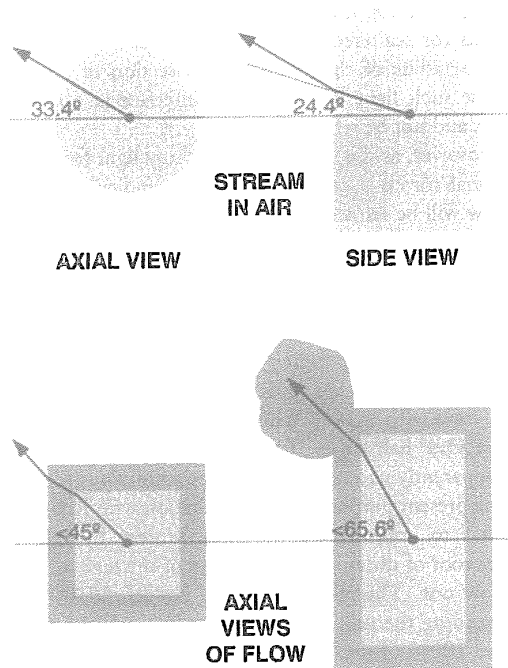


Figure 4-40. New angles on light collection in flow.

Light Collection from Streams and Cuvettes

To understand how the configuration and composition of a stream in air or a flow cuvette determine how much light can be collected, you don't need to know all the angles, but you do need to know a few of them; the old ones are shown in Figure 4-12 and discussed on pp. 121-2, and the new ones appear in Figure 4-40. You might also want to look back at p. 150 and Figure 4-25.

In flow cytometry, we are generally looking at particles in an aqueous medium, which I will assume for the purposes of the discussion here is water. Water has a refractive index n of 1.33 at 20 °C. Light coming from a particle in the stream will be refracted at the air-water interface in a stream-in-air system, and, assuming the cuvette is made of quartz ($n = 1.46$), at the water-quartz interface and the quartz-air interface in a system using a flow cuvette without an optically coupled lens. The angle that is of importance to us is the half angle of the cone of light coming from the particle in water that we can manage to coax into our collecting lens after all of the refractions have taken place.

The upper part of Figure 4-40 shows axial and side views of a stream in air. In most high-speed sorters, particles are interrogated in a stream in air; the lenses used for side scatter and fluorescence collection in these instruments are usually "ultra" long working distance (10-13 mm from the stream, so they won't get wet) microscope objectives, with an N.A. of 0.55. The formula $\text{N.A.} = n \sin \theta$ tells us that the largest angle θ at which an N.A. 0.55 lens, working in air, will collect light is 33.4°. If we look at the axial view of the stream

in air, we see that there is no problem collecting light that is emitted (or scattered) at that angle from a central particle along a radius of the circular cross section of the stream, because such light will strike the interface at normal incidence, and not be refracted.

However, as the side view shows, any light from the particle that travels outside the plane perpendicular to the axis of flow will be refracted at the air-water interface. Light that reaches the lens in air at an angle of 33.4° will have come from the particle at an angle of 24.4° in water. So the half angle θ of the cone from which we can collect light using an N.A. 0.55 lens is somewhere between 24.4° and 33.4° . The solid angle Ω corresponding to θ can be calculated from the formula $\Omega = 2\pi(1 - \cos \theta)$; it would be 0.562 sr (steradians) for a cone with a 24.4° half angle and 1.035 sr for a cone with a 33.4° half angle. To calculate the "right" solid angle, it is apparently necessary to use elliptic functions (don't ask); for the present purpose, after consultation with Rob Webb, I decided to approximate the right answer by taking the square root of the sum of the squares of the high answer and the low one. This yields 0.869 sr. By definition, a sphere surrounding the cell, representing 4π sr, or 12.57 sr, would receive 100% of the light from the particle; 0.869 sr represents a relatively unimpressive 6.9% of the total. In this calculation, and those that follow in this section, we neglect the fact that transmission of the lens is always less than 100%.

Up to a point, we could improve light collection from the stream by using a lens with a higher N.A., say 0.68 or 0.7. An N.A. 0.7 lens gets us to a 44.4° half angle for the radial rays and a 31.8° half angle for the refracted rays; the back-of-the-envelope calculation says it collects over a 1.529 sr solid angle, or about 12.2% of the total light, a 77% improvement over the N.A. 0.55 lens. An N.A. substantially higher than 0.7 won't help, because, as it turns out, 48.8° is the critical angle for an air-water interface, and light hitting the interface at that angle and larger angles is subject to total internal reflection (see pp. 121-2 and 157-8 and the figures on those pages). This means that most of the extra light that, say, an N.A. 0.95 lens would collect from a specimen in air will travel (gently?) down the stream as if the stream were a light pipe. In fact, Mariella et al.²⁴⁵⁵⁻⁶ have demonstrated that a fiber optic sharpened to a conical tip and inserted into the stream along the axis of flow downstream from the observation point will do a pretty good job of collecting light scattered at large angles. Of course, this precludes using the stream in air for sorting.

If you're using a stream in air because you want to sort at the absolute maximum rate, you'll probably need to use a 50 or 70 μm orifice to allow you to get a droplet rate of at least 50,000/s. Even if your beam is focused to a 20 μm spot, there is likely to be enough reflection from the stream (see p. 150) to make it necessary to use an obscuration bar in front of the fluorescence/side scatter collection lens. This will result in the loss of about 30% of the light that would otherwise be collected by the lens.

While the square and rectangular cuvettes used in flow cytometers minimize noise due to stray scattered light, allowing obscuration bars to be dispensed with and thereby facilitating high-precision measurements using relatively low-powered light sources, some problems with light collection arise from the refractive properties of the cuvettes. The bottom half of Figure 4-40 shows axial views of a square and a rectangular cuvette. The geometry of a square, as shown at the left, restricts collection to a half angle of less than 45° ; any light leaving a particle centered in the flow stream at that angle will emerge from the cuvette into air at an angle of no more than 72° , limiting the maximum N.A. of a collection lens usable with a square cuvette to 0.95. In practice, it is difficult to obtain lenses with an N.A. that high and the long working distance of a few millimeters necessary for observation in cuvettes. An N.A. 0.65 "high dry" objective with a long working distance, or an aspheric lens with an N.A. between 0.62 and 0.68, would be a more realistic choice as a collection lens. The N.A. 0.65 lens would collect light emitted by a particle in water in a cone with a half angle of 29.3° , corresponding to a solid angle of only 0.801 sr, or about 6.4% of the total, a little less than is collected from a stream in air by an N.A. 0.55 lens.

An N.A. substantially higher than 1.0 can only be achieved if the lens is optically coupled to the specimen, in this case the cuvette, by a medium with a relatively high refractive index, matched to those of the materials of which the cuvette and lens are made. It is also necessary to use a rectangular cross section for the cuvette rather than a square one, as can be seen from the picture on the bottom right of Figure 4-40. Light leaving the target particle in water at an angle above 65.6° will be subject to total internal reflection; the maximum usable N.A., even for a coupled lens, is thus limited to 1.21 (recall from p. 122 that N.A. is invariant from medium to medium). The gray "blob" shown surrounding the ray emerging from the rectangular cuvette represents the coupling medium, which could be immersion oil if the chamber is horizontal, a gel if it is vertical (as in the B-D FACScan, etc.), or the substance of a lens attached directly to or built into the chamber. The solid angle over which an N.A. 1.2 lens collects light from a particle in the rectangular flow cuvette is 2.655 sr, or just over 21% of the total; this is more than three times as much light as is collected from a stream in air or a square cuvette by the lenses normally used with them.

Various authors^{653-4,1147,2457} have described flow chamber designs incorporating optical elements that, in theory, permit the collection of two to six times as much light as can be collected from a square cuvette or stream in air. Goodwin et al.²⁴⁵⁸ measured the actual light collected by an N.A. 0.55 and an N.A. 0.85 objective as about 4% and about 9% of total emission, respectively. Considering that the transmission of the lenses was just above 80%, these figures are in reasonable agreement with the calculations just worked through.

In studies done during the course of designing the Cytomutt family¹⁰⁵, my colleagues and I found we could obtain

essentially identical sensitivity and precision in fluorescence measurements using a stream-in-air system, a flat-walled flow chamber, or a round-walled chamber, provided that a stable flow pattern was achieved and maintained and that the beam and stream or capillary dimensions permitted use of a field stop, rather than an obscuration bar, to decrease collection of stray scattered light. The key to performance then resided in the design of illumination and collection optics, rather than in the geometry of the flow chamber.

For some time, I had better luck with my round capillary systems than with stream-in-air systems, which led me to conclude that observation in an enclosed space was inherently better. What I actually compared, however, were capillaries with outer diameters of about 300 μm and inner diameters of about 150 μm , on the one hand, and streams in air with diameters between 70 and 100 μm , on the other. Round streams, and round capillaries, behave as cylindrical lenses; the smaller the capillary, the stronger the lens. Therefore, the optical properties of the capillary and stream-in-air systems that I had compared were quite different.

Inspired by the large stream issuing from a broken capillary, Dick Adams and I (unpublished) did a brief comparison, and found that large ($\geq 300 \mu\text{m}$ diameter) streams in air gave us fluorescence measurements at least as good as those obtained using round capillaries of the same outside diameter, and that scatter measurements made using the stream-in-air system were noticeably less noisy than those made using a capillary. We could detect forward scatter signals from small bacteria using a laser emitting less than 0.5 mW; when a capillary was substituted for the stream in air, signals were barely detectable above noise. There are practical problems with large streams in air; one has to use large volumes of sheath fluid per unit time, and streams in air in general, as mentioned previously, raise biohazard safety issues. Both of these objections can be eliminated by using a system in which the sheath flows through standing water and light is collected by a water immersion lens, with no interfaces between the lens and the core stream; such a flow chamber was used in the Block Engineering apparatus with which scatter measurements were made of single virus particles⁹⁴.

If there is a bottom line here, it is that you can't have it both ways. If you want to sort at the absolute maximum rate, you'll have to put up with inefficient light collection; if you want to hoard photons, you'll have to slow down. And, speaking of slowing down, Figure 4-41 shows flow chamber designs used in several flow cytometers with arc lamp sources. These instruments typically have highly efficient illumination and light collection optics, but, as we found on pp. 131-2, their lamps don't put out anywhere near as much usable light as do lasers, and both sample flow rates and cell acquisition rates are, accordingly, lower than in most laser-based instruments.

Panel A illustrates the arrangement in the Ditttrich/Göhde Impulscytophotometer (ICP)^{9,83}; this is an axial flow system, in which cells confined in a sheath are observed as they emerge from a tube oriented along the axis of a fluores-

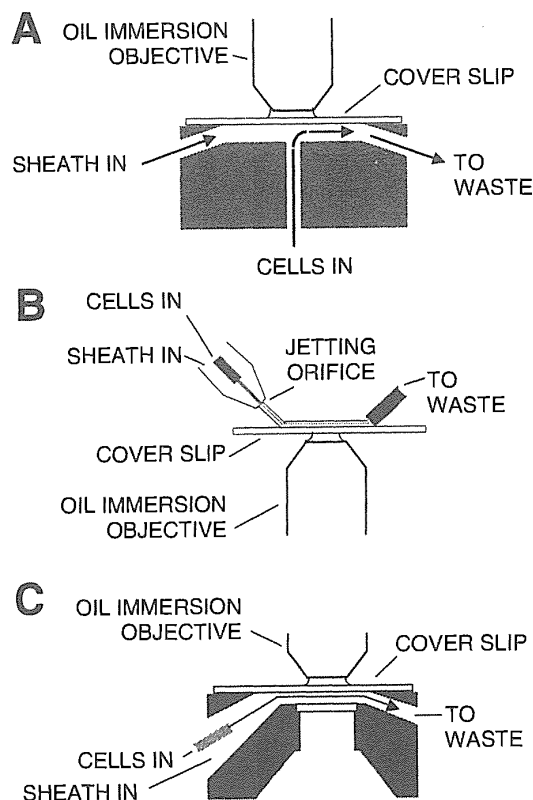


Figure 4-41. Flow chamber designs used with arc source flow cytometers. A: Impulscytophotometer. B: Lindmo/Steen system (Skatron). C: Closed system similar to some used by Bio-Rad, Heka, and Partec.

cence microscope objective and make a 90° turn into a flowing stream. The Phywe ICP used suction to draw cells through the system; the later Ortho version forced them through with gas or air pressure. Wolfgang Göhde felt that axial flow of cells through the objective focal plane was an important determinant of the high precision obtained with the ICP under the best conditions. However, Göhde's newer flow system designs for Partec achieve equivalent precision with what amounts to a modified orthogonal geometry.

So does the instrument developed by Lindmo and Steen¹⁰⁰⁻¹⁰³ (originally sold as the Leitz MPV-Flow; later by Skatron, Ortho (Europe), Bruker, Bio-Rad and (in the near future) Apogee). In this cytometer, exemplified in Panel B, cells confined in a sheath are ejected from a nozzle as they would be in a stream-in-air system. The stream then impinges on a cover slip at a relatively acute (about 20°) angle, and is observed with the high N.A. oil immersion lens of an epiilluminated fluorescence microscope downstream from the point of impact; fluid is removed from the cover slip by suction still further downstream.

If the flow pattern isn't stable before the cells leave the nozzle, it won't be stable after they hit the cover slip, and the excellent precision reported won't be achieved. If the flow

pattern is stable, some improvement results from a further stabilization of the existing laminar sheath flow pattern by forces between the fluid and the cover slip. The stream flattens out, a phenomenon my colleagues and I noted when using a pen nib design to deposit cells from a flow cytometer onto movie film⁹⁵. This tends to orient asymmetric cells and also to confine them to a narrow region of space, allowing a high aperture objective, with its small depth of focus, to be used without compromising performance.

A closed flow system was/will be available on the Bio-Rad/Apogee versions of the Lindmo/Steen instrument; it resembles that shown in Panel C. The Fluvo II, designed by Kachel⁶⁶, the newer Partec flow cytometers, designed by Göhde, and Thomas's RATCOM and NPE instruments are similar. All are basically orthogonal flow systems, although they are not symmetric about the axis of flow; some allow observation of the stream from two or more sides. The Partec instruments offer the options of multistation illumination, with two lamps or a lamp and a laser, and fluidic sorting in a closed system.

Since the same microscope objective serves as condenser lens and fluorescence collector in most arc source cytometers, increasing the N.A. of the lens improves both illumination and collection. Temsch et al²⁴⁵⁹ reported that substitution of a gel-coupled N.A. 1.25 lens for an N.A. 0.8 lens in an arc source instrument increased fluorescence signal intensities approximately fourfold.

The flow chamber in B-D's FACS analyzer, an arc source instrument introduced in the early 1980's and soon eclipsed by the FACScan, was essentially a thick-walled round capillary, with an hourglass profile that allowed its use for electronic (Coulter) volume measurement; a second sheath was used largely to provide some fluid flow on both sides of the volume sensing orifice. Some of the old Los Alamos flow chambers (about the size and shape shown in Panel A of Figure 4-39), used with laser source cytometers, also incorporated electronic volume measurement capability⁶⁵; the combination of electronic volume measurements with optical measurements in cytometry has probably been pursued for the longest period of time by Bob Leif and his colleagues^{654,1108-9,2457}. The NPE/RATCOM instruments, and a variety of clinical hematology analyzers (including Beckman Coulter's), incorporate both optical and electronic volume measurement capabilities; an electronic volume measurement option is also offered by Partec.

When You've a Jet...

While observation of cells in a stream in air instead of in a closed system eliminates some of the complexity, and may lower the cost, of a flow chamber, more constraints on the hydrodynamics are, in general, associated with stream-in-air systems. Pinkel and Stovel⁶⁵⁰ have covered some of the details.

One of the interesting things they mention is that the exit velocity v (in m/s) of fluid from an orifice or nozzle is, to a first approximation, directly proportional to the square

root of the pressure used to propel the fluid through the orifice; i.e., the sheath pressure. If the sheath pressure, ΔP , is given in pounds per square inch (PSI),

$$v = 3.7 (\Delta P)^{1/2};$$

if the pressure is given in atmospheres, the constant is 14, and if in pascals (N/m²), it is 0.044. This relationship applies to typical stream-in-air nozzles, in which the orifice is placed immediately below the neckdown region; it does not hold for flow systems in which there is a capillary tube of any appreciable length, because of the effects of viscous drag.

A dimensionless constant called the **Reynolds number**, Re , is often used to characterize the stability of fluid flow. If d is the diameter of the stream, in cm, ρ is the density of the fluid, in gm/cm³, η is the fluid viscosity, in poise (1 poise = 1 g/cm·s), and v_{av} is the average fluid flow velocity, in cm/s,

$$Re = \rho v_{av} d / \eta.$$

Laminar flow is possible for values of $Re < 2300$; it may be possible at higher values, but turbulence is much more likely to occur. Thus, in flow cytometry, it makes sense to keep Re below 2300. For water at 20 °C, a temperature at which its density is very nearly 1 gm/cm³ and its viscosity is 0.01 poise, we can relate stream diameter d , in μm , not cm, and v_{av} , in m/s, not cm/s, by

$$Re = d v_{av}.$$

In other words, if v_{av} is 10, and d 230 or below, a person won't get turbulent flow. Unless you're trying to sort really big objects in a droplet sorter (the principle works for stream diameters as large as 1 mm (1000 μm)), you can work at 10 m/s. What you may have to worry about under more typical conditions, however, are **changes in temperature**; the viscosity of water is highly temperature dependent.

Turbulence is apt to be a problem at the inlets, rather than the outlets, of flow chambers, stream-in-air and otherwise. This happens because the sheath inlet is usually a relatively small diameter tube going fairly abruptly into a larger diameter chamber. If you are trying to squeeze the last few tenths of a percent out of an already very good measurement CV, as is the case in chromosome sorting, which stimulated Pinkel and Stovel's efforts, you need to devote some attention to this possible source of flow instability.

Figure 4-42 illustrates three configurations of core and sheath inlets in otherwise identical flow chambers. The "bad" configuration, at left, has the small sheath inlet entering the chamber at a right angle. The "better" configuration represents a compromise reached in many commercial systems; the sheath inlet tube is gently curved, and a perforated "spider" is put in to break up the flow pattern; empirically, this facilitates faster establishment of laminar flow downstream. The "best" configuration, at right, gently sneaks the core into a large bore sheath flow. I've seen chambers like this in action, and they give good CVs; I haven't yet bitten the bullet and made a Cytomutt flow chamber with a more hydrodynamically correct core and sheath inlet.

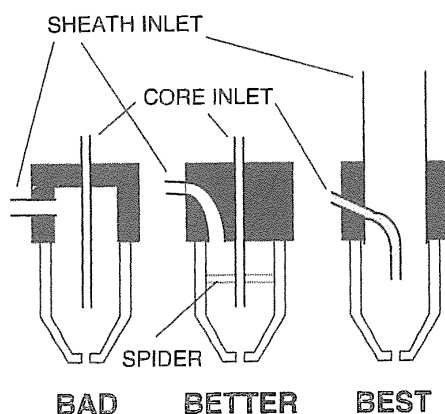


Figure 4-42. Minimizing turbulence generated at the sheath inlet to flow chambers.

Core and Sheath: Practical Details

If you looked at pp. 50-1 and 130-1, or otherwise have some acquaintance with flow cytometry, you've run across the notion that core sizes affect measurement precision. It thus may be useful to know how big a core you're actually running. Let's look at a numerical example. Suppose we have a stream-in-air system with an orifice 80 μm in diameter, through which we want cells to pass at a velocity of 10 m/s. When things are working properly, running the system for one second will squirt out a volume of fluid that would fill a cylinder 80 μm in diameter and 10 m high. The volume of this cylinder is π times the square of the radius times the height. Converting the radius and height measurements to cm, which will give us volume in cm^3 , or mL, we find that, in one second, we run a total fluid volume of

$$\{\pi \times [(4 \times 10^{-3})^2] \times 1000\} \text{ mL} \approx 0.05 \text{ mL}.$$

In one minute, or 60 seconds, we'd get about 3 mL through the orifice. Knowing any two of the three variables, orifice diameter, flow velocity, and flow rate, we can calculate the third. If you have a commercial system, the manufacturer has provided you with the internal dimensions of the flow chamber (if you built your own, you should know!), and you can measure the amount of fluid which goes through the system per unit time; this will let you calculate the flow velocity.

Now, how can we figure out the core diameter or radius? If the instrument has a syringe pump, or some other means for measuring the sample or core volume flow rate, it's fairly easy; however, most flow cytometers are not so equipped. We therefore resort to subterfuge, and possibly centrifuge as well, and get and run a sample of cells, or beads, at a known concentration, determined by an automated counter, or, as a last resort, with a hemocytometer.

Suppose the concentration is 10^6 particles/mL, and we find we're counting 1000 particles/s (for the moment, as-

sume we count every one that comes through). The count rate divided by the cell concentration gives us the rate of flow of the core fluid, which is 10^{-3} mL/s. The velocity is 10 m/s; from the core flow rate and velocity, we calculate the volume of fluid that would fill a cylinder with its height equal to 10 m and its radius equal to the core radius. The core radius is

$$[(\text{core flow rate})/(\pi \times \text{velocity})]^{1/2}.$$

The formula gives us a core radius of 5.64 μm ; the core diameter is therefore 11.28 μm .

As was mentioned earlier, a parabolic velocity profile is established within a relatively short distance of the orifice in a stream-in-air system. Pinkel and Stovel report that, near the exit of an 80 μm nozzle, with flow at 10 m/s, fluid elements on axis and 10 μm off axis will differ in velocity by 1 m/s, or 10 percent. Particles traveling off-axis at lower velocities will, on that account, spend more time in the laser beam; this should result in higher signal intensities. However, we also have the Gaussian intensity distribution in the beam to consider here; the intensity of illumination diminishes with distance from the axis. In real stream-in-air instruments, the two effects undoubtedly counteract one another to some extent; the off-axis particles spend more time in a lower-intensity beam, and thus should receive, emit, and scatter more nearly the same amount of light than would be the case if the velocity profile were more sluglike than parabolic. In flow systems in which cells are observed in capillaries, the observation point may be at a level at which slug flow persists; widening the core might then be expected to degrade precision more noticeably.

Grace Under Pressure: Driving the Sheath and Core

The pressure-driven sheath used in most flow cytometers operates on the same principle as the wash bottle you used in

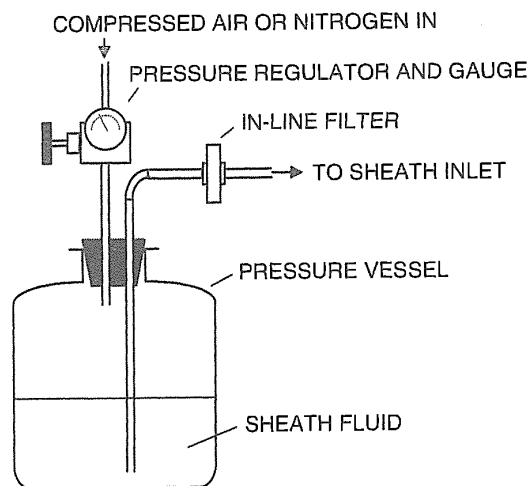


Figure 4-43. Sheath fluid supply plumbing.

chemistry if you took it as long ago as I did. The pressure bottle in the original Becton-Dickinson FACS looked exactly like one of those old wash bottles; a sheath bottle differs, however, in being connected to a pressurized gas supply with a regulator and a gauge so you don't have to blow into it. The setup is shown in Figure 4-43.

The range of drive pressures used for sheath fluid in various instruments is from about 3 PSI on the older Ortho Cytofluorographs to about 30 PSI in the Lindmo/Steen apparatus; the high speed sorters at Livermore and Los Alamos could go up to several hundred PSI, but the current commercial high speed sorters seem to top out near 100 PSI. For the lower pressures used in most commercial and lab-built systems, a variety of plastic bottles available from many laboratory equipment suppliers will do just fine; Ortho favored a plastic bottle held in a metal box to limit its expansion under pressure. I, and most of the manufacturers, now use stainless steel pressure vessels, originating at Alloy Products Corporation and acquired through as few intermediaries and with as few markups as possible. These are relatively expensive if you don't eliminate the middlemen when you buy them, but they are very unlikely to shatter. They typically have plastic on the outside around the bottom, which prevents them from making electrical contact with whatever they're sitting on; be advised that steel vessels containing saline can give you a nasty shock if you touch them while the drop charging circuit of your sorter is on.

The most common cause of turbulent flow in a previously "healthy" flow system is complete or partial **obstruction**. **Complete obstruction** of flow is relatively easy to diagnose, if not always equally easy to correct. **Partial obstruction** can be much more troublesome, because sample flow continues, but turbulence, caused by the obstruction, results in a broader range of cell trajectories and velocities, which typically alters pulse shapes and broadens distributions of parameter values. Inexperienced operators would frequently note the increase in CVs and promptly make unnecessary adjustments to the optics and electronics in an attempt to improve the quality of measurements; things then went from bad to worse, often requiring a visit from service personnel may be required to get signals to reappear. Since most flow cytometers in use these days don't permit the operator to make the adjustments, the misalignment problem has largely disappeared; the obstruction problem has not.

Most flow cytometer users don't get the opportunity to redesign their flow systems to minimize obstructions, and must therefore do what they can solely in terms of keeping apparatus, sheath fluids and samples as clean as possible. **Filtration** of sheath fluids and all diluents used in sample preparation through 0.22 or 0.47 μm filters will remove large particles and also minimize bacterial contamination. Even pharmaceutical grade stuff may have some particles in it; if you wouldn't inject it intravenously, don't use it as sheath fluid.

It's good practice to keep a 0.22 or 0.47 μm filter in the line between the reservoir and the flow chamber, to clean

out any residual junk. However, if the sheath hasn't been filtered previously, such an in-line filter may clog rapidly enough to decrease the flow rate. This can happen when you follow the nearly universal practice of driving the sheath with gas (air or nitrogen) blown into the reservoir at constant pressure, which means that as the resistance of the filter goes up, the flow rate goes down. In theory, if the sheath were driven by a robust constant volume pump, instead of by pressure, flow velocity could be maintained. Don't bet on it. If you are working with really small particles, such as bacteria, it is advisable to prefilter sheath through a 0.11 μm filter; you probably don't want a filter with pores that small in your sheath line because it is likely to offer substantial resistance to fluid flow.

Generally speaking, flow cytometric analysis, as opposed to sorting, can be done using water instead of a saline solution as a sheath fluid; a prominent exception occurs in the case of electronic (Coulter orifice) cell volume measurements, which, like droplet cell sorting, require that cells be suspended in an electrically conductive medium. This usually means an isotonic sodium chloride solution, with or without a little buffering. Why not use saline solutions all the time? Well, all other things being equal, water is cheaper than saline. Water also doesn't leave salt deposits all over everything. For analysis, even using live cells, the interval between injection of the core (which does have to be isotonic to keep cells osmotically happy) into the sheath and the measurement is so brief that the cells won't have time to object to being in a yucky hypotonic medium until well after they've passed the observation point, unless there's an obstruction downstream which causes water to back up into the sample. For sorting and Coulter volume measurements, which require ionic solutions, normal saline for injection, which is fairly clean and sterile, is readily available in liter bottles at reasonable prices. "Injectable" or not, it does have to be filtered.

Nitrogen supposedly produces fewer bubbles than does compressed air when used as a sheath propellant. When house air supplies are non-existent or inadequate, and gas from tanks must be used for sheath drive, nitrogen has the advantage of being cheaper than air is. It remains to be determined whether substitution of nitrogen for air affects cell physiologic parameters such as membrane potential; when in doubt, use air and isotonic solutions.

When gas pressure is used to drive the core, or sample, as well as the sheath, the relative sheath and core flow rates are dependent upon the difference in pressures between the sheath container and the sample container. This determines the core size. If the pressure drive system for the sheath is a giant wash bottle, that for the core is a micro wash bottle; the pressure vessel is frequently the sample tube itself, with a volume of only a few mL. Separate regulators for core and sheath pressure are advisable; pressure in the sample vessel is adjusted to change the core flow rate. A **differential pressure gauge**, monitoring the difference in pressure between

sheath and sample vessels, provides the best operator feedback for this process.

Unless valves in the fluid lines are closed, if there is no tube mounted for sample feed, sheath pressure forces sheath fluid back through the sample inlet tube; this is a relatively easy way of clearing out the residue of the last sample before running the next one, but it can also be a way of diluting the next sample with sheath fluid when you'd prefer not to have diluted it.

On the other side of that coin, most pressure feed systems also have switchable inlets for drive gas at both the regulated pressure and a higher boost pressure; when a tube is first put on the system, the higher pressure is applied to drive cells into the sample tubing. The input is then switched to the regulated pressure. Flow should then be allowed to stabilize before any measurements are made.

In order to avoid the consequences of Shapiro's First Law (p. 11), we'd like to have a filter in the sample path to prevent particles greater in diameter than the orifice (cell clumps and other junk) from clogging it. B-D used to use a small filter made of a bundle of glass microcapillary tubes, each with a 40 μm internal diameter; the filter was placed at the very tip of an aspirator assembly which was inserted into the tube containing the sample. Ortho used a hollow fiber filter that was much larger physically; most of the people I know replaced it with a few pieces of nylon mesh with a 40 μm pore size held over the aspirator tip with a piece of plastic tubing. You can now buy sample filters from various sources; Partec makes some nice ones. You can also buy nylon mesh and roll your own.

Just as is the case with a pressure-driven sheath, when a filter is used with a pressure-driven core, as the filter clogs, the core flow rate decreases. This can be a real problem when the cells being analyzed are treated with an equilibrium stain such as acridine orange or one of the cyanine dyes, because relative flow rates of core and sheath affect dye diffusion rates from core to sheath, and changes in these flow rates may affect staining intensity. Changes in flow rate and core size may also affect measurement precision.

If a pump, instead of gas pressure, is used to drive the sheath, the volume flow rate of the sample is directly adjustable, and less susceptible to the sources of variation just mentioned. The pump used to drive the core should be a **positive displacement pump**, e.g., a **syringe pump**, which produces a (relatively) constant volume flow rate even when driving a varying resistance. Dick Sweet's piece on flow sorters in reference 9, circa 1980, contained statements to the effect that syringe pumps weren't good for sorting because they produced pulsatile flow; opinions have changed. I have almost always used syringe pumps for sample feed; others are beginning to find them advantageous^{5,100-103,649-650}, and some newer commercial systems use them.

When really smooth drive is necessary, it can be obtained from a syringe pump that uses a feed screw instead of the more common rack and pinion drive. Pulsations can also be reduced by putting a small air bubble between the syringe

piston and the sample; this gets away from true constant-volume operation, but remains closer to this ideal than would a constant pressure drive in the presence of increasing resistance.

If you're good enough at designing fluidics, you can even manage to make a peristaltic pump produce a stable flow pattern in a flow cytometer; I can't do it, but the people at OptoFlow seem to have managed, and I have heard of a variety of other pump types being used successfully.

Gravity, all by itself, can be used to drive fluids through flow systems; it can also effect whatever else you're using. Changes in the height at which the sheath tank is placed and/or in the liquid level in the sheath tank can affect flow. Gravity is often used to advantage when very slow flow rates are required, e.g., for the outer sheath of double-sheath flow systems, or for systems designed for molecular analysis.

In hematology analyzers, which are the prototypical clinical flow cytometers, sample and sheath may both be propelled by syringe pumps, which, in combination with valves and other mechanical and electronic components, allow blood samples to be withdrawn from the tubes in which they were collected, diluted, treated with appropriate reagents, and analyzed. This is a good way to handle samples that should all be subjected to the same treatment, and which can be analyzed in a minute or so. On the other hand, if you wanted to do a four-hour preparative sort on a droplet sorter, you'd need a syringe the size of a Bugatti Royale piston for the sheath; pressure drive from a big reservoir is the only practical way to go. You'd also probably have to use pressure feed for the sample, keeping it a container with appropriate temperature regulation and a magnetic stirrer or other mechanical device to keep the cells in suspension.

Figure 4-39 shows that the fluid connections in flow chambers are readily accessible; up until recently, it was fairly easy to get at those connections, at least on instruments designed for research purposes, and hook up whatever core and sheath drive mechanisms you might want. Third parties such as Cytel Development provide such gadgets, optimized for such tasks as kinetic experiments with calcium probes. However, if you're in the market for flow cytometer and contemplate playing games with the fluidics, it's a good idea to get confirmation from the manufacturer that the changes can be made without rendering the cytometer inoperative.

Perfect Timing: Fluidics for Kinetic Experiments

Kinetic experiments, in which the addition of one or more reagents to a cell sample and the introduction of the sample into the core stream must all be carried out at precisely timed intervals, are likely to need more elaborate external plumbing than the flow cytometer manufacturers normally supply.

The adaptation of the analytical chemical technique of **flow injection analysis**¹¹⁴⁷⁻⁸ to cytometry was described by Lindberg, Ruzicka, and Christian¹¹⁴⁹ shortly before the last edition of this book appeared. In addition to allowing rapid

and flexible, automated, sample preparation and solution handling, flow injection analysis offers methods for analysis of data from samples in which chemical equilibrium has not been reached, which may be useful in both static and flow cytometric applications.

More recently, John Nolan, Larry Sklar, and their colleagues have developed instrumental approaches to both rapid kinetic experiments and high throughput flow cytometry^{2454,2460-3}; these are discussed on pp. 364-6.

Oriented and Disoriented Cells

When velocities are different at different points in a flowing stream, an object in the stream is subjected to unequal forces at different points on its surface. These forces cause cells, particularly asymmetric cells, to tumble in the stream, and, in some cases, to assume one orientation rather than another. Kachel, Fellner-Feldegg, and Menke¹¹⁴⁶ show photographs of the behavior of erythrocytes in flow, demonstrating the orienting effect of a rectangular core injector.

Johnson and Pinkel⁶⁵¹ made modifications to a commercial flow cytometer to permit high-resolution DNA analysis of sperm. Beveling the round core injector to a chisel point produced a ribbonlike stream; asymmetric sperm heads were oriented in the plane of the stream, and CVs of DNA measurements were substantially reduced. Orienting sperm and monitoring orientation are critical to selecting x- and y-chromosome enriched fractions by sorting, as will be discussed further in Chapter 10.

Matchmaker, Matchmaker, Make Me a(n) Index Match!

Differences in chemical composition between the core and sheath fluids (as in the instances when, for example, you use a water sheath and a saline core, or a saline sheath and a saline core with added protein) may be reflected in refractive index differences between the core and sheath. The refractive index mismatch usually isn't a problem for eukaryotic cells, but definitely will be for bacteria or really small (0.5 μm) beads and may be for platelets. The noise is predominantly in the scatter channels; fluorescence measurements should not be affected significantly although variable scattering of excitation and emission at the interface might be expected to increase CVs slightly. However, don't be too quick to assign blame to a refractive index mismatch when the noise is due to particles in the sheath fluid. The easiest way to tell the difference is to run the sheath without the core; if the noise persists, it can't be due to an index mismatch. Refractive indices of natural water samples may differ; Cucci and Sieracki have demonstrated the effects of differences in sheath and sample salinities on forward scatter signals from a variety of small marine organisms²⁴⁶⁴.

Flow Unsheathed

Kamentsky's original Rapid Cell Spectrophotometer⁶⁵, and many of the early hematology counters that preceded it, did not use sheath flow, and there are still some flow cytometers that don't, including a fluidic sorter for very large

particles (*Drosophila* embryos)²³²⁵ and a microfluidic sorter for very small ones (bacteria; DNA fragments)²³²⁶⁻⁷. Among commercial systems, we have the Flow CAM, an imaging flow system for analysis of fairly large marine organisms²⁴⁵³, and the Guava PCTM, designed to count and analyze cells smaller than 60 μm ²⁴⁶⁵. The principal problems with not using a sheath are that there are likely to be differences in flow velocities between particles traveling at different distances from the center of the stream and that there is a somewhat greater likelihood of the flow system becoming obstructed by large particles or aggregates. It would probably be difficult to obtain a distribution with a CV of 1% if you measured DNA content of nuclei in a flow cytometer without sheath flow, but it is evidently not a problem to get measurement CVs of 5-10%; if these are adequate for the application, you can safely dispense with the sheath.

Flow Systems: Garbage In, Garbage Out...

It should be obvious that spending a few hundred thousand bucks on an instrument won't do anything to relieve you of the burden of designing experiments. Flow cytometers are like computers in many respects; both are high-tech, often high-budget gadgets which can very rapidly and very precisely do exactly what you've told them to do, and neither has any capacity for doing what you meant instead of what you said.

If, for example, you are interested in finding cells so rare that they represent only one cell in every 10^8 cells in your sample, you should bear in mind that, even if your flow cytometer could zip through 100,000 cells/s, and could identify your cells of interest flawlessly, you would only encounter 3 or 4 cells of interest an hour. If all you wanted to do was get a count of your population of interest, with a precision of 10 percent, you'd have to count 100 cells to get the contribution to variance due to sample size down into the acceptable range. This would take 25 to 30 hours, not counting time to change samples, replenish sheath fluid, etc. It is all but impossible to do a single experiment of this kind by flow cytometry; it is beyond impossible and heading for irrational to contemplate doing such analyses for some routine purpose.

On the other hand, if you're looking for one cell in 10^5 , you could get a good count in a few minutes; whether or not it is feasible to do so depends on the efficacy of your cytometric identification procedure. It's easy to do if you can stain your cells of interest with one or, even better, two very bright fluorescent dyes; it's nearly impossible to do if you are relying on something like a dim immunofluorescent stain as the sole identification criterion. In general, you have to decide what measurement(s) you need to make of cells to answer the question(s) you're asking about them.

Recall that a flow cytometer, by itself, doesn't even have an intuitive way of telling what is a cell and what isn't; you basically provide a definition for it by picking a trigger channel or channels, setting a threshold value or values and, optionally, by defining selection regions in your measure-

ment space using hardware and/or software gating. If no other gating is used, the machine treats every particle for which the amplitude of the trigger signal is above the set threshold value as a cell.

When you analyze clean samples from relatively homogeneous cell populations, you don't go too far wrong in using a forward light scatter measurement as a trigger parameter; immunofluorescence measurements have traditionally been done in this way, with the fluorescence measurement gated by the volume or scatter signal. In samples containing cell populations with greater variance in characteristics, thresholds and gates set on a fluorescence channel indicating the presence and amount of cellular DNA may provide more stringent criteria for distinguishing cells of interest from debris, on the one hand, and from aggregates, on the other. At present, the quality of both instruments and antibody reagents is high enough for it to be feasible to trigger on immunofluorescence signals to identify cells in relatively messy samples.

Given the prices so many people have to pay for flow cytometer time, it's wise to have some idea of what your sample looks like before you put it in the machine. The best way to get some idea is to look at the sample under a phase contrast and/or fluorescence microscope. It is true that the flow cytometer can pick up more subtle differences than you can discern by visual microscopic observation, but I'm not talking about subtle differences. If your immunofluorescent stained cells are lying on the slide in one long strand, you can cancel your appointment with the machine. If, unbeknownst to you, the cells are sitting in the tube in one long strand, putting them in the cytometer is apt to cancel other people's appointments as well. This can lead to stress, if it's your cytometer, and to physical harm, if it's someone else's.

Although in-line filtration is an obvious method of preventing large particles from entering and obstructing the flow system, filters may exacerbate the problem of sample carryover, which is encountered even in cytometers without them. Cells may accumulate in various places in the flow system during turbulent flow while samples are being changed, and subsequently pass into the measurement system.

Sample carryover is tolerable, if only because it is unnoticeable, in a lot of routine flow cytometry. If you are running a whole bunch of similar samples, each with approximately the same cell concentration, you probably won't know or care if 1 or 2 percent of the cells in the $(n+1)$ st sample were really stragglers from the n th sample. Carryover can really kill you, however, when you are looking for rare cell subpopulations and/or when you run cell samples with very low cell densities after cell samples with very high densities.

Let's suppose we have a 1 percent carryover rate between two samples, each of which contains 10^6 cells/ml, and each of which is run at a rate of 1 μ l/s, with data from 10,000 cells being collected. When the second sample is run, 1,010 cells will come through the machine each second; 1,000 of

these will be from sample 2, and 10 will represent carryover from sample 1 (on the average). If we were now to put on a third sample with only 10^5 cells/ml, and run at the same sample flow rate, carryover, now from sample 2, would still give us 10 cells/s from sample 2, while we would only collect 100 cells/s from sample 3; for the third sample, carryover would be greater than 9 percent instead of 1 percent. You can generally get a good sense of the extent of carryover in your system from the persistence of fluorescent plastic beads, with which most of us start our day's runs, in subsequent cell samples. Reducing carryover may require some combination of backflushing, filter change or removal, and vigorous cleaning of the flow system.

Particles from the last sample aren't the only foreign agents which may introduce disinformation into data; microorganisms, particularly fungi, may grow in flow systems and in the sheath fluid supply, and surprisingly large amounts of hydrophobic dyes (acridine orange, the cyanines, and propidium are notorious in this regard) in stained samples may adhere to even short lengths of plastic sample inlet tubing, from which they are readily transferred to cells in subsequent samples. Chlorine bleach usually clears out both microorganisms and dyes; you do, however, have to be sure you've washed the bleach out thoroughly before you put samples back into the instrument, or you're apt to end up with yet another class of artifacts. I usually follow the bleach with water, then 70% ethanol, and then rinse well with water. If there's any liquid in the tubing and the flow cell when I shut down the system, I want it to be clean water.

In a posting (28 August 2001) to the Purdue Cytometry Mailing List, Mario Roederer recommended running 0.1 N NaOH (made in clean water and filtered through a small pore filter), CoulterCleanse solution, and distilled, deionized filtered water, in that order, through a stream-in-air sorter to solubilize any residual DNA, RNA, cells, and microorganisms. Mario also uses 70% ethanol for sterilization. I (and he) caution that it is probably a good idea to find out from the manufacturer whether there is any part of your instrument's fluidic system that won't stand up to the NaOH solution - or to any other solution or solvent you plan to run through the system.

The statistical distributions of cell arrival times and of the time intervals between successive cells' transits of the observation point may provide information about measurement quality. Lindmo and Fundingsrud³⁵¹ examined the latter distribution and its relation to various aspects of sample preparation. In some cases, even when samples did not contain cell clumps, the distribution did not fit the expected Poisson statistics; the authors hypothesized that cells that had stuck to the sample tubing were dislodged when hit by other cells passing through.

Watson⁸¹⁹ later extended this type of observation to permit the use of **time measurements for quality control** within individual sample runs. The interval between cell arrival or acquisition times is used to derive a real-time measurement of **sample flow rate**, which is displayed

against time; data collected during periods of rapid fluctuation of flow rate are discarded. Gross et al.¹¹⁵⁰, using a combination of a computer algorithm for excluding such "burst" data, a rigorous cleaning procedure for the flow system, and staining and analysis techniques which minimized contributions from instrument noise and nonspecific fluorescence, reliably detected rare cells at frequencies of 1 per million using a benchtop flow cytometer; getting the garbage out, one way or another, definitely helps.

4.7 ELECTRONIC MEASUREMENTS

The ambiguous heading above will allow me to discuss electronic measurements of cells, i.e., **electronic** or **Coulter volume measurement**, which is based on **DC impedance**, and **electrical opacity**, which is based on **AC impedance**, and also to introduce a little bit about electronics in general.

Electricity and Electronics 101

Electrons and protons possess equal and opposite charge. The SI unit of charge (q), the **coulomb (C)**, is approximately 6×10^{18} times the charge on an electron, which is about 1.6×10^{-19} C. The electrons in an atom hang around the general vicinity of the protons, but the degree of mutual attraction varies, and may change with electronic energy levels. Electrons are, by convention, said to bear a **negative** charge; protons bear a **positive** charge. Particles of like charge repel one another; particles of opposite charge attract one another. **Coulomb's Law** states that the **electrostatic force** between two charges is proportional to the product of their magnitudes and inversely proportional to the square of the distance between them. The electrostatic force repelling two electrons spaced 1 mm apart is 10^{43} times as strong as the gravitational force attracting them.

Charge Separation, Electric Fields, and Current

According to the **Law of Conservation of Charge**, charge cannot be created, but it can be redistributed. If the electrons in a material are held relatively loosely, it is possible for charge to be transferred from one point in the material to another, and the material is called a **conductor**. If the electrons are more tightly held, charge transfer is not possible, and the material is called an **insulator**. There are also materials that normally behave as insulators, but which will conduct when some of their electrons are raised to appropriate energy levels; these materials are **semiconductors**. Atoms or molecules can lose or gain electrons to become positively or negatively charged **ions**; the Law of Conservation of Charge demands that ions be formed in pairs, with opposite charges. Ionic solutions and **plasmas**, which are essentially ionic gases, conduct electricity.

Separation of charges produces an **electric field**, which exerts a force on any charged particle in the field; the strength of the field is given in terms of force per unit charge. The **potential energy** of the particle varies with its position in the field. For example, if an electric field exists (never mind how it gets there) between two metal plates

bearing opposite charges, an electron placed between the plates will be repelled by the negatively charged plate and attracted toward the positively charged one; the electron therefore has **potential energy**, which will decrease as it is accelerated toward the positive plate, being converted to **kinetic energy**. The potential energy of a charged particle in an electric field is proportional to its charge; the difference in potential energy, or **electrical potential difference**, between two points in the field at which the potential energy of a 1 coulomb charge changes by 1 joule, is defined as 1 **volt (V)**. The amount of potential energy lost, or kinetic energy acquired, as a single electron (charge 1.6×10^{-19} C) moves through a 1 volt potential difference, is called an **electron volt (eV)**; it is about 1.6×10^{-19} joules.

The mutual repulsion of like charges will cause redistribution of charge in a conductor, such that any excess charges are uniformly distributed on the surface of the conductor. The electric field "inside" the conductor is zero. Surrounding a volume with a conductor therefore produces **electrostatic shielding**, which is why low-level electrical signals are generally transmitted through **coaxial cable**, familiarly known as "coax" (pronounced "co-axe"). The signal is, in theory, carried in the central conductor, which is surrounded by an insulator, which is surrounded by braided or twisted wire and/or metal or metal coated foil. In principle, this outer conductive layer, which is connected to "ground," carries no current, and shields the inner conductor from electrical interference. In practice, the "shield" serves as a **return path** for the signal, and does carry current, which makes it less than perfect as a shield.

This has come to the attention of manufacturers of audio cables, who are now happy to sell you cables in which the signal and return signal are carried by a twisted pair of insulated wires surrounded by a foil shield that is only connected to ground at the signal source. This shield cannot carry current, and therefore works as advertised; the twisted pair configuration of the signal wires also reduces interference. Shielded twisted pair cables are probably not a necessity for a home audio system; I have put them in mine more as a matter of principle than because I think they'll make an audible difference.

The transfer of charge through conductors produces what is called an **electric current**; one **ampere (A)** of current represents the transfer of one coulomb of charge per second. It is totally incorrect to think of electric current "flow" in conductors in terms of an electron going into one end of a wire and coming out at the other end. An electric field produced at the "positive" end of the conductor accelerates a free electron toward that end, temporarily creating an excess positive charge "one atom over"; this generates a local field that accelerates a free electron from the vicinity of the next atom, etc., the process being repeated until an electron enters the "negative" end of the conductor. There is some analogy to the science toy made of a series of suspended metal balls, which is used to demonstrate transfer of momentum. You pick up the ball at one end, and let it

swing into its static neighbor. The neighbor doesn't move, nor do any of the other balls except the one at the other end of the chain, which flies away from its neighbor to about the distance from which you released the first ball. In this instance, it is the local electric field which propagates along the conductor, moving at the speed of light; the current is, by convention, said to "flow" in the direction in which the field moves, which is opposite to that in which the electrons "move". The "speed of light" at which the field propagates in a conductor is not, by the way, the old familiar 3×10^8 m/s (186,000 miles/s) at which light travels in a vacuum; electricity, like light, moves more slowly through material media.

Resistance, Voltage, and Power; Ohm's Law

You could more or less guess, from the description of current flow in a conductor, that, just as friction and other real world effects prevent the metal balls from bouncing back and forth in perpetual motion, there must be some losses involved in all of those electrons jumping around. With the exception of a few esoteric materials that display the property of **superconductivity** at very low temperatures, there aren't any perfect conductors. Every material offers some **resistance** to the flow of electric current. According to **Ohm's Law**, the flow of a current of I amperes through a material with a resistance of R ohms (Ω) produces a drop in electrical potential, or a voltage difference, of E volts, across the resistance; that is:

$$E = IR.$$

It's amazing how much you can get done using just that formula.

Just as friction losses dissipate mechanical energy as heat, resistance losses dissipate electrical energy as heat. The amount of energy, in joules, lost per second is EI (remember that volts are joules/coulomb and amperes are coulombs/second), or I^2R . Joules per second, however, are **watts (W)**. A current flow of 1 A through a 1 Ω resistor produces a potential, or voltage, drop of 1 V across the resistor, and dissipates 1 W. The resistor gets hot, but not very; 4.184 J are equal to a "small" calorie, which is the amount of heat needed to raise the temperature of 1 gm of water by 1 °C.

What's a resistor made of? Generally speaking, resistors are made of conductive material, but not much of it. Again, as you'd probably expect, the less conductor you have, the harder it is to get current through it. It's much harder to move the same amount of traffic down a two-lane road than to move it down an eight-lane highway; similarly, the resistance of a given length of thin wire is higher than the resistance of the same length of thicker wire made of the same material. Many resistors are intentionally made to get hot, but, with exceptions, they are designed to dissipate the heat rather than to be destroyed by it.

Put 10 A through a 12 Ω resistor, which is what happens when you connect it to 120 V, and you dissipate 1200 W; the resistor is apt to get red hot, at least until the bell rings

and the toast pops out. The filament of a 60 W light bulb is a 240 Ω resistor that gets white hot with 120 V across it, at which point it draws 0.5 A; the bulb must be filled with inert gas to prevent the filament from burning up. The active element in a **fuse** is a thin metal strip or wire that will melt once the current flowing through it gets above a certain value. And I oversimplified in all of the above cases; the resistance of materials typically increases with temperature. So, looking more carefully at the 60 W light bulb, we would find that the resistance of the cold filament is only about 1 Ω . The filament initially, and very briefly, draws a current much higher than 0.5 A (that's why bulbs are more likely to die just as they are turned on), increasing its resistance to 240 Ω as it reaches operating temperature.

Alternating and Direct Current; Magnetism

Where does current come from? We have already considered one current source in our discussion of the photo-voltaic photodiode, or solar cell, which is a photoelectric source of current. The more light hits it, the more current the cell generates; however, the polarity of the current remains the same. The positive end of the cell remains positive. Current sources with this characteristic are said to generate **direct current**, or **DC**. DC is also produced by batteries, which convert chemical energy into electrical energy, and by some types of electromechanical generators.

Most of the current generated by electric utilities worldwide is not DC, but **alternating current**, or **AC**. The generators that produce either DC or AC make use of **magnetic fields** to convert mechanical energy into electrical energy. Magnetic fields result in part from the motion of charges; motion of a conductor in a magnetic field generates a current in the conductor. Passage of current through a conductor in a magnetic field creates a force on the conductor that may make it move if it is free to do so.

A microphone can be made from a very small coil of wire attached to a very small diaphragm and suspended in the field of a small permanent magnet; small currents are generated in the coil as the diaphragm is moved back and forth by interaction with sound waves in the air. These currents, electronically amplified, can be fed through a larger coil of wire, attached to a larger diaphragm and suspended in the field of a larger magnet; this device, a loudspeaker, converts the applied electrical energy back to mechanical energy, in the form of sound. The process of **transduction** between electrical and mechanical energy can work in both directions; a loudspeaker can function as a microphone, although its high mass makes it relatively insensitive, and a microphone can function as a loudspeaker, or at least as an earphone, although such a delicate device is likely to be destroyed by application of even a small amount of current.

The current generated in a microphone is an alternating current. No current is generated in the coil when the diaphragm is at rest; motion of the diaphragm away from its rest position in either direction causes an electrical potential difference across the coil, but the **polarity**, i.e., which side of

the coil is positive and which negative, changes with the direction of motion. Alternating current is characterized by its **frequency**, f , measured in **hertz** (Hz), which used to be called **cycles per second**. Two alternating currents of the same frequency can additionally be related by their difference in **phase**; this is expressed by the same angular measure described in connection with the discussion of light on pp. 104-5.

Electric generators and electric motors are similar in their mechanical construction; they are comprised of substantial coils of wire suspended in the fields of relatively strong magnets. The magnets themselves are often electromagnets; passage of current through a coil of wire will create a magnetic field, and the field intensity is higher when the coil of wire is wrapped around a magnetic material such as an iron alloy. When a source of mechanical energy such as a steam turbine or internal combustion engine is used to move the coil, electric current generated in the coil can be used to operate other electrical devices. When current is applied to the coil, the coil moves, providing mechanical energy.

The alternating current output of generators is frequently in the form of a **sine wave**; that is, the voltage, $V(t)$, at any time, t , is related to the maximum voltage, $V_{\max}$, by the formula

$$V(t) = V_{\max} \sin(2\pi ft).$$

The mathematical technique of **Fourier Analysis** often makes it convenient to deal with more complex alternating current waveforms as sums of series of sine (or cosine) waves.

Inductance, Reactance, Capacitance, Impedance

A conductor exhibits a property called **inductance** (L), and responds to alternating current in a frequency-dependent fashion. The alternating current produces a changing magnetic field, which generates a voltage opposite in polarity to the applied voltage. In an inductance of 1 **henry** (H), a voltage of 1 volt is induced by a current changing at the rate of 1 ampere/second. An inductor therefore has a property called **reactance**; reactance, like resistance, provides an impediment to the flow of current, but, unlike resistance, is dependent on the frequency of the current. The **inductive reactance** of an inductor, X_L , in ohms, is:

$$X_L = 2\pi fL,$$

where f is the frequency in Hz and L is the inductance in henries. Inductive reactance increases with increasing frequency; an inductor conducts better at lower frequencies.

If the positive and negative sides of a DC current source are applied to a **capacitor**, a device consisting of two conductors separated by an insulator, there is a transient current flow, which is opposed by the accumulation of charges of opposite polarities on the conductors, and which stops when the potential difference between the conductors is equal to the potential of the source. If the current source is then removed, the potential difference between the conductors remains. The stored charge can be extracted as current. The

amount of charge that can be stored in a capacitor increases with the surface area of the opposed conductors, and also varies with the **dielectric constant** of the insulator used. The **capacitance**, measured in **farads** (F), is equal to the amount of charge on either electrode, in coulombs, divided by the potential difference between the electrodes, in volts; 1 farad equals one coulomb/volt. Real capacitors have capacitance values ranging from picofarads (pF) to thousand of microfarads (μ F); some capacitance exists between any two nearby conductors.

Direct current will not flow "through" a capacitor; alternating current will, because the capacitor exhibits **capacitive reactance** (X_C), which like inductive reactance, is a function of the frequency of the applied current. For X_C in ohms (Ω), C in farads, and f in Hz,

$$X_C = 1/(2\pi fC).$$

A 1000 pF [or 1 nanofarad (nF)] capacitor has a reactance of about 159,000 Ω (159 k Ω) at 1,000 Hz (1 kHz), 15,900,000 Ω (15.9 M Ω) at 10 Hz, and 1,590 Ω (1.59 k Ω) at 100,000 Hz (100 kHz). Thus, a capacitor conducts better at higher frequencies.

When alternating current is applied to a resistor, the voltage and current vary together, in phase. In a capacitor, current and voltage are out of phase, with current 90° ahead of voltage. In an inductor, current and voltage are also out of phase, with current 90° behind voltage. Thus, in a circuit that contains both inductance and capacitance, the effects of one tend to cancel those of the other. The combined effect of resistance, inductive reactance, and capacitive reactance is referred to as the **impedance** (Z) of the circuit; this is, obviously, a frequency-dependent quantity.

Impedance is not simply the sum of resistance and reactance; it is expressed as

$$Z = [R^2 + (X_L - X_C)^2]^{1/2}.$$

It may seem as if we've gone a long way to get to a definition of impedance simply to explain the measurement made in a Coulter, or impedance, counter; the foregoing will, however, come in handy when we start talking about electronic circuits later on.

The Coulter Principle: Electronic Cell Sizing

Much of the stuff of which cells are made doesn't conduct electricity all that well. While intracellular fluid is a conductive ionic solution, the movement of most ions across the cell membrane, which is largely composed of lipid, an insulator, is restricted. Cells are therefore relatively poor conductors.

Blood is a suspension of cells in plasma, which is a reasonably good conductor. It was found in the late 19th century that the fraction of blood made up by the cells could be estimated from the conductance (conductance, measured in mhos, is the reciprocal of resistance) of blood; as the ratio of cells to plasma increases, the conductance of a given volume of blood decreases.

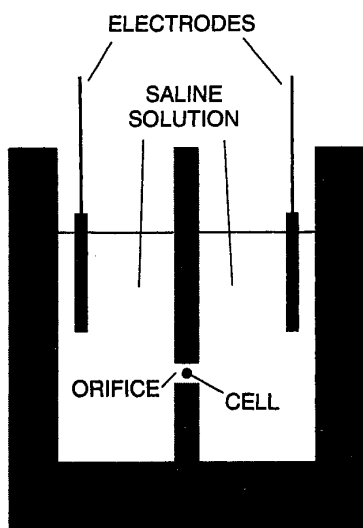


Figure 4-44. The Coulter orifice.

In the late 1940's, Wallace Coulter extended this type of measurement to the single cell level with his development of an electronic method for detecting, counting, and sizing cells based on their relatively low conductance. The principle of the Coulter orifice is illustrated in Figure 4-44.

Two chambers filled with a conductive saline solution are separated by a barrier containing a small orifice (typically 100 μm or less in diameter and no more than a few hundred μm in length) that provides the only fluid (and electrical) connection between the chambers. Most of the resistance, or impedance, in this arrangement is in the orifice. The electric circuit is analogous to that made up of two railroad rails connected by a thin wire; my friends and I used to enjoy demonstrating where the resistance was in that setup by dropping unwound wire coat hangers from an overpass onto the local subway tracks, briefly creating a conductive path between the third rail and one of the grounded rails on which the trains traveled. Coulter's analogue, while generating considerably less heat, light, and sound, has been vastly more useful and profitable.

A DC power supply that provides a constant current is connected to electrodes in each of the two chambers. As long as the orifice is entirely filled with saline, its impedance remains constant; we know from Ohm's law that, with a constant current flowing in the circuit, the voltage applied across the electrodes must also remain constant.

In operation, a stream of cells is passed through the orifice. A cell in the orifice displaces an equivalent volume of saline solution; the impedance of the orifice is therefore increased during the cell's transit, with the extent of the increase dependent on the volume of the cell. In order to keep the current through the system constant, the power supply must transiently apply a higher voltage between the electrodes. If the voltage output of the power supply is continu-

ously monitored, a temporary increase in voltage output, or a **voltage pulse**, is observed whenever a cell passes through the orifice.

The details, which are discussed at length by Kachel¹¹⁵¹, are quite a bit more complicated. The electric field intensity varies considerably with distance from the axis of the orifice; as a result, the widths, amplitudes, and shapes of pulses produced by a cell's passage vary depending on its position in the stream. In order to deal with this problem, some instruments analyze individual pulse shapes, and reject pulses that appear to be due to cells close to the wall of the orifice. Some also employ **hydrodynamic focusing**, i.e., sheath flow, to confine the cells to a region of the stream near the axis of the orifice. Once the necessary corrections are made, cell volume can be derived from pulse amplitude.

Electrical Opacity: AC Impedance Measurement

Coulter and Hogg⁷¹⁵ established that the Coulter orifice could be operated using an AC, rather than a DC power supply. When the frequency used is in the radio frequency range, the cellular parameter measured is referred to as **electrical opacity**; this reflects the AC impedance of cells, and is more dependent on cellular structure and less dependent on size than is DC impedance. Some of Beckman Coulter's hematology instruments (e.g., the STK-S) and some from other manufacturers incorporate opacity measurements.

4.8 ANALOG SIGNAL PROCESSING

In optical flow cytometry, some light falls on the scatter and fluorescence detectors even when a particle is not passing through the observation region, producing a **background** current output from the detector; this background current fluctuates about some **baseline** value above zero. As a particle passes through the observation region, it produces a temporary increase in current output, i.e., a **current pulse**, at each detector. This is why flow cytometry has also been called "pulse cytophotometry." As is the case with Coulter volume and opacity measurements, information about the cells is derived from characteristics of the pulse.

Beam Geometry and Pulse Characteristics

Depending upon the illumination geometry, the **peak** amplitude, or **height**, and/or the **integral**, or **area**, of the fluorescence pulse will be proportional to the total amount of fluorescent material contained in the cell or particle. To understand this better, we need to look at Figure 4-45.

The figure illustrates what happens when cells of different sizes pass through focused beams of different sizes. Suppose the 20 μm diameter particle contains twice as much dye as the 10 μm diameter particle. Since the volume of the 20 μm particle is eight times the volume of the 10 μm particle, the amount of dye per unit volume is higher in the smaller particle; if we looked at them under a fluorescence microscope, we would see a small, bright particle and a larger, dimmer one. At the top of the figure, we see the small particle in the middle of a 30 μm beam waist, and the larger par-

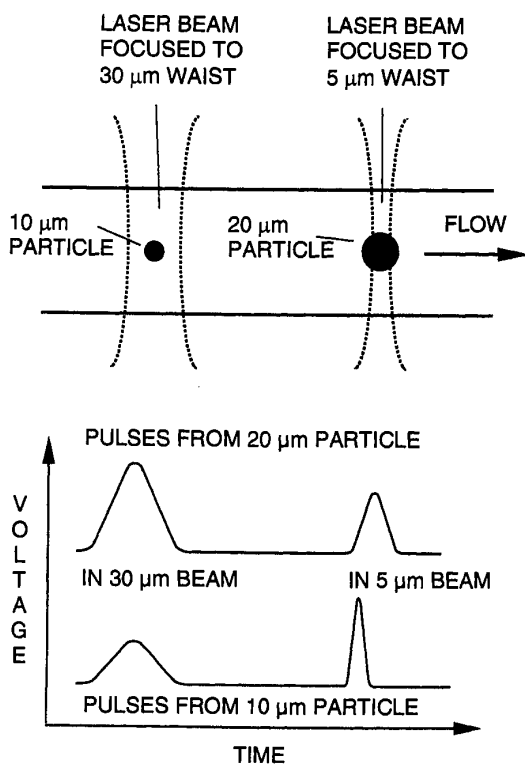


Figure 4-45. Effect of beam geometry on pulse shape.

particle en route through a 5 μm beam waist. Note that we assume here that the power in the laser beams is low enough to prevent substantial bleaching of dye molecules, say 5-10 mW; this keeps the relationship between illumination and emission intensities linear.

The bottom portion of the figure illustrates the pulse shapes we may expect at the fluorescence detector preamplifier output as the particles traverse first the 30 μm and then the 5 μm beam. The scale is approximate. The pulses produced by both particles as they pass through the 30 μm beam will be Gaussian, because they both pass through the Gaussian intensity profile of the beam; the amplitude will reach a peak when the particles pass through the center of the beam.

Since the beam diameter is substantially larger than the particle diameter, the particles will receive approximately equal illumination, despite their different sizes. Thus, the peak amplitudes will be proportional to the amounts of dye contained in the two particles; the 20 μm particle, containing twice as much dye, produces a peak twice as high as the 10 μm particle.

The situation is different at the 5 μm beam. If both particles are traveling at the same velocity, the 20 μm particle is going to be illuminated for a longer time than the 10 μm particle. Since the beam waist is smaller than the diameter of either particle, neither particle will be entirely contained

within the illuminated region at any one time. The larger particle will produce a longer pulse; the peak amplitude will be lower than the peak amplitude of the shorter pulse produced by the smaller particle because the smaller particle contains more molecules of dye per unit volume, i.e., it is brighter. Making the beam waist small compared to the particle diameter produces a **slit-scan**; one can get several types of information from pulses thus produced which cannot be derived using a larger beam waist. The **duration** of the pulse, or **pulse width**, yields size information^{97,98}. The **peak height** gives information about **brightness**, or **fluorescence density**. In order to get a measure of the total amount of fluorescent material contained in the particles, however, you need the **integral**, or **area**, of the pulse, rather than the peak amplitude or pulse height. This makes the electronics somewhat more complicated.

Electronics 102: Real Live Circuits

As they come from the detectors, the current pulses we deal with in cytometry don't give us much to work with. The output current from the anode of a PMT or photodiode is only a few microamperes; let's say 10 μA (10^{-5}A) at most. To get the signal to the point where we can comfortably deal with it, we need an **amplifier**, something that will provide higher levels of current and/or voltage output than are available direct from the detector.

Circuits: Current Sources and Loads

It is customary to speak of electronic **circuits**; at a minimum, a circuit is composed of a **current source**, which supplies current, and a **load**, through which the current passes, eventually returning to the source, or completing the circuit. A circuit is not a perpetual motion machine; energy from the source is transferred to the load. In one of the simplest circuits, a resistance (which might be a light bulb) connected to a battery, the chemical energy in the battery is gradually lost by conversion into heat and light (an aside: only about 3% of the total wattage dissipated in an incandescent bulb goes into light; the rest is lost as heat).

Both current sources and loads have associated **impedances**. Ohm's law holds for impedances, that is, $E = IZ$; this means that, at a constant voltage, a low impedance load draws more current than a higher impedance load. As you may have noticed, the amount of power stereo amplifiers are rated to deliver is often specified separately for speaker impedances of 4 or 8 Ω ; that's why. A low impedance source can supply more current than a high impedance source for a given voltage. PMTs and photodiodes are high impedance sources.

Figure 4-46 shows some circuit elements and the basic patterns in which they are interconnected; it features resistors and capacitors, which are known as **passive elements**. We'll leave inductors out of this discussion. Elements connected end to end are said to be **in series**; elements that have their corresponding terminals connected to the same point, or **node**, are said to be in parallel.

The resistances of resistors in series add, which makes sense. If we connect the series resistors R_1 and R_2 to a source, the same current, I , passes through both resistors. If E_1 is the voltage across R_1 and E_2 the voltage across R_2 , Ohm's law gives us $E_1 = IR_1$ and $E_2 = IR_2$; the total voltage across the pair, $E = E_1 + E_2$. Resistors (and reactances) in series can therefore be used as **voltage dividers**; the voltage across a 10 k Ω resistor in series with a 90 k Ω resistor is 1/10 the voltage across the pair. A **variable resistor**, or **potentiometer**, contains a resistive element with a third connection which can be physically moved to make contact at any point between the two end terminals, providing a variable voltage divider usable as a light dimmer or volume control, among other things.

According to **Kirchhoff's Current Law**, the sum of currents entering any node in a circuit equals the sum of currents leaving that node. Circuit elements in parallel carry current in inverse proportion to their impedances; the voltage drops across any two elements in parallel are equal.

The charge storage capacity of a capacitor, all other things being equal, is a function of the surface area of its elements or plates; intuition correctly tells us that the capacitance of two capacitors in parallel is the sum of their individual capacitances. It is a little less obvious that two capacitors in series can store less charge than can either alone, but that's the way it works.

Ground Rules

If you are at all familiar with the electrical wiring of cars or bicycles, you've probably noticed that a lot of the loads have only a single wire connected to them. How is the circuit completed? The answer is that the other end of the load, and the other end of the battery providing the current, are both connected to the metal frame of the car or bicycle; they are said to be **grounded**, or, in British, **earthed**. The planet itself is conceived as representing the mother of all reference voltages, and, in fact, most power wiring is electrically connected at some point to a real ground, usually through the medium of a fairly substantial copper rod.

By convention, the "zero voltage" point in a circuit is defined as **ground**; this is represented in circuit diagrams by the symbol shown at the upper left in Figure 4-46. All points shown as connected to ground in a circuit diagram are electrically connected to one another; this is sometimes accomplished by a wire connection and sometimes accomplished by connecting the points to a metal (or otherwise conductive) enclosure in which the circuit is housed. Since the electric field inside a conductor is zero, such an enclosure provides **shielding** from stray electric fields.

In the real world, all grounds are not created equal. Let's go back to the car, where the positive terminal of the battery is connected to one terminal of various loads, and the negative terminal is connected to the car's chassis, as are the other ends of the loads. The current from a load is, therefore, being conducted back to the battery through the chassis, which has a low, but finite, resistance, which - Ohm's law again -

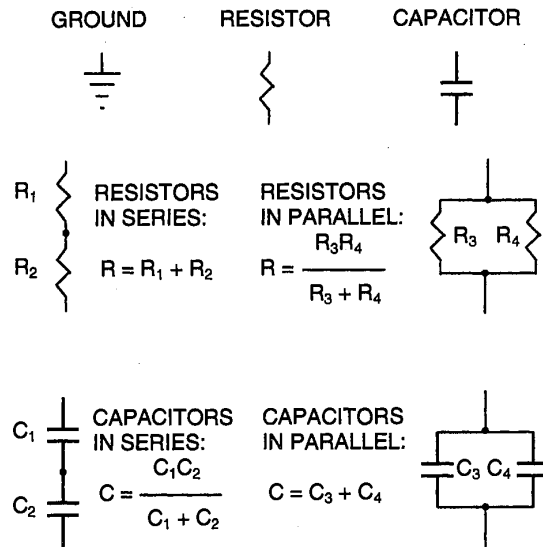


Figure 4-46. Some circuit elements.

means there is a potential difference between the point on the chassis at which the load makes contact and the point at which the battery is connected.

Most household AC power wiring in the United States delivers electricity to loads through two wires. One is called "hot," generally color-coded black (an unfortunate choice, since electronikers frequently use black as a color code for ground); the second, generally color-coded white, is called "neutral," and theoretically represents ground. Well, it is connected to ground, at the power station, at the transformer outside the house, and perhaps at points in between.

The nominal "110-115 V" value for household wiring is actually an RMS average; the peak values are higher. The voltage on the hot wire fluctuates (at 60 Hz), reaching extremes at 156 V above and 156 V below the voltage on the neutral wire. Inside the house, the wiring is divided up into parallel circuits, each with its own hot and neutral wires, and each with a fuse or circuit breaker limiting current in that circuit, generally to 15 or 20 A. If nothing at all is plugged into one of these parallel circuits, there should be no connection between its hot and neutral wires, and the neutral wire should, in theory, be at ground potential. You could touch it with wet bare hands without ill effect, again in theory. **DO NOT ATTEMPT THIS TRICK AT HOME, IN THE LAB, OR ANYWHERE ELSE!!!** The neutral wire is never at ground, and is usually hot enough to kill you, because there's usually current flowing in the circuits.

That's why modern wiring has a third, ground wire, color-coded green. This is connected to ground, either through the metal conduit and boxes in which the electric circuits are contained, or through a separate wire. You can safely connect yourself to the ground connection on a three-wire plug; in fact, it's a good idea to do so when you're

working with solid-state electronics, which will be susceptible to static damage if you're not grounded. In a three-wire circuit, current goes from the hot wire through the load and back to the generating station via the neutral wire; the ground wire isn't carrying current. If the metal case of an electric or electronic appliance is connected to the third wire ground, and a **short circuit**, for example, an electrical connection between the hot or neutral wire and the case, develops due to mechanical damage, the ground connection provides a return path for the current. If the case isn't connected to ground, and you're holding it when the short develops, you become the ground connection, after which you may become the underground connection.

While the third-wire ground of power wiring is good enough to protect people from many of the potential lethal effects of 110 V AC, it may leave something to be desired as a "zero volt" reference. In many buildings in which AC wiring has been added over time, there may be substantial differences between the "ground" voltages at the third wire in different circuits. This seems to happen particularly often in hospitals, and seems to be exacerbated by the power surges and spikes which inevitably occur in such places. This makes it necessary to exercise extreme caution when plugging in electronic equipment. I and several people I know have had the experience of connecting a computer system to a flow cytometer, only to have some electronic components in one, the other, or both go up in smoke. It is often said that electronic devices work because they contain a small amount of smoke; when the smoke comes out, they stop working. In this instance, the proximate cause of the smoke coming out has been a measurable potential difference of tens of volts between the third wire grounds of two different circuits to which different components were connected.

There are several things you can do to minimize the likelihood of such disasters. If possible, plug any electronics you're connecting to your flow cytometer into the same circuit from which the cytometer gets power, i.e., into an outlet connected to the same circuit breaker. If you can't, use a voltmeter and/or the differential input of an oscilloscope to measure potential differences (AC and DC) between the grounds of the circuits to which you have access and which you might need to use in combination. If you don't see a big difference between "ground" potentials, you may be okay, but the problems tend to come from voltage transients. For this reason, you should always use surge suppressors with three-way protection (they block surges between hot and neutral and between either of those wires and the ground wire) and noise filtering on all of your electronics. Maybe now you know why you have had to replace all those boards in your instrument.

As I mentioned previously, low-level signals are typically transmitted through coaxial cable; the center conductor, in principle, carries the signal, while the outer conductor, usually connected to ground at least one point, provides shielding. By now, I hope you're becoming accustomed to the idea that, in a circuit, there has to be a conductive **return path**

from the load back to the source as well as a path from the source to the load. Where's the return path in the coaxial cable? In the outer conductor, or shield. But wait a minute; doesn't that mean there's current flowing from the "ground" end of the load to the "ground" end of the source, which, as Ohm's law tells us, implies there is a potential difference between these two points? Yup. It probably doesn't amount to more than a few dozen microvolts, but there is one. When you start trying to measure signal levels around 1 millivolt, which is the low end of the bottom decade of a 4-decade log amp's input range, the details of ground connections become significant. In general, low-noise circuit design separates **power ground**, i.e., the point at which the return paths from the electronics to the power supplies converge, from **signal ground**, the "zero voltage" reference for signals. Grounds for analog signals are often separated from ground for digital signals, which tend to be noisier. The trick is to prevent what are called **ground loops**, i.e., current paths in the ground wiring which increase noise. This, as might be expected, requires both science and art; when you try to hunt down noise in circuits, you find your prey is not easily run to ground.

Couplings, Casual and Otherwise; Transformers

If you have any experience with audio equipment, you're probably familiar with hum at power line frequencies and multiples thereof (60 and 120 Hz, in the U.S.) as a major component of noise. How does this get into audio equipment? Primarily by **capacitive** and **inductive coupling**. I previously described a capacitor as being composed of two conductors separated by an insulator, and mentioned that AC could get from one side of a capacitor to another, while DC could not.

Maybe you thought, at that time, that the definition of a capacitor was too broad; it could for example, fit any two unshielded wires. So, indeed it could, and does; there is a small capacitance between any two conductors, and the closer they are together (while not actually in electrical contact) the higher the capacitance gets. A little bit of an AC signal in one can, and will, be transmitted to the other. This capacitive coupling is an electrostatic field effect, and is usually guarded against by **shielding**, which, as has already been mentioned, surrounds the conductor to be protected with another conductor, usually connected to ground.

Inductive coupling involves interactions between the magnetic fields surrounding two inductances; just as any old wire can be part of a capacitor, any old wire can be an inductor. The simplest defense against inductive coupling of an AC current to the outside world is to have the signal path and return path running through a **twisted pair** of wires; the associated magnetic fields pretty much cancel one another out. Putting a grounded shield around the twisted pair (p. 180) placates the capacitive demons, as well.

Inductive coupling is used intentionally and effectively in **transformers**, by which I mean not the metamorphosing toys, but the heavier and uglier devices used to convert low

voltage to high voltage AC and *vice versa*. Suppose we had a coil with ten turns of wire placed next to a coil with twenty turns of wire, with a 10 V peak-to-peak AC signal running through the first coil; this would induce a 20 V peak-to-peak signal in the second coil, not only in principle, but pretty much in practice, too. Whether we could do anything effective with the 20 V signal, however, would depend upon the amount of current we were able to draw. If the coupling efficiency were 100%, putting 10 A at 10 V through the first coil, or **primary**, would allow us to get 5 A output current at 20 V from the **secondary**. Note that the products of voltage and current on both sides here are equal; in practice, there is always some loss. Coupling efficiency of transformers at relatively low frequencies, say below 100 kHz, is increased by winding both coils on a core of magnetic material, which is what makes transformers, particularly those used in power supplies, heavy.

Power Supplies

Resistors, capacitors, inductors, and transformers are all **passive devices**; you can't get any more power or wattage out of them than you put into them. You can use a transformer to convert a low-voltage AC signal to a higher voltage signal, but you'll get proportionally less output current. Doing useful things with signals usually requires one or more stages of **power amplification**, in which the output voltage and/or current are increased while the temporal characteristics of the signal, or **waveform**, are preserved. This requires **active electronics**, incorporating devices such as transistors and vacuum tubes, and these need **power supplies**, which, for reasons we will get into shortly, must provide direct current. In principle, we could run our active electronics on battery power, which provides DC, but it's much more convenient to be able to plug things into the wall.

Edison originally envisioned household electricity as being provided by local generating stations as direct current. The drawback with this scheme is that the current has to be carried between the generating station and the users at a fixed voltage, say 110 V. If the ten houses on a block used 50 A apiece, the wire supplying the current just to that block would have to provide 500 A, which is a lot of current; you'd need a conductor more than a half inch in diameter to carry it without appreciable voltage drops and losses due to heating. If, instead, power is supplied as AC, as suggested by Westinghouse, the voltage on transmission lines can be thousands of volts, reducing the current transmitted, and allowing the use of thinner, cheaper conductors. A series of transformers placed at local power stations and in neighborhoods reduce voltage to the nominal 110 VAC required for household use, incidentally providing electromagnetic fields about which we have to decide whether to worry.

An active device, e.g., a power amplifier, has to be able to provide a (relatively) high-voltage, high-current replica of the input waveform. It doesn't know in advance how that waveform varies in time; it therefore needs to be able to call on its power supply for maximum power at any time, which

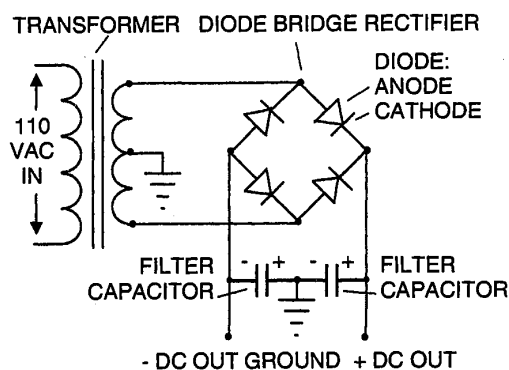


Figure 4-47. A line-powered DC power supply.

means it needs DC power. As it turns out, most active devices run on voltages which are either substantially lower or substantially higher than 110 V; transistors typically require 25 V or less, while vacuum tubes need hundreds or, in some cases (e.g., TV picture tubes), thousands of volts.

A typical DC power supply, operated from AC line voltage, is shown in Figure 4-47. The first element in a power supply, after the on-off switch and the fuse or circuit breaker, which aren't shown here, is a transformer. This converts the 110 V AC from the power outlet into an AC voltage in the neighborhood of the DC voltage required by the electronics. The circuit symbol for a transformer is actually made up of back-to-back symbols for two inductors; the two vertical lines between them represent the metal core of the transformer.

The next element in the power supply is a **rectifier**, which converts the AC into DC; the rectifier shown here is composed of four **diodes** in what is known as a **bridge** configuration. A diode, as its name suggests, is a two-terminal electronic device; its two electrodes are a **cathode** and an **anode**. Current will flow from the anode to the cathode, but not from the cathode to the anode; the circuit symbol lets you know the direction in which current will flow. When current does flow from the anode to the cathode, there is a voltage drop across the diode; this is more or less fixed and its value depends on the materials of which the diode is made.

In the olden days, when we used vacuum tubes, which ran at high voltages and low currents, typical rectifiers were vacuum (tube) diodes, which are similar to PMTs without the dynodes. Instead of depending on light hitting the cathode to break electrons loose, which would have provided an extremely inefficient source of solar power, a high-resistance wire, or filament, running on relatively low-voltage AC, was used to heat the cathode, resulting in thermionic emission of electrons. Thermionic emission is a nuisance in a PMT, where it is the source of the unwanted dark current; in some applications, the PMT is cooled to reduce thermionic emis-

sion. In rectifiers and most other vacuum tubes, thermionic emission is an absolute necessity, hence the heated cathode.

Recall that in a PMT, the anode, thanks to a high-voltage DC power supply, is kept at a higher potential than the cathode, as are the dynodes; electrons accelerate from an electrode at one potential toward an electrode at a higher potential. When a diode is used as a rectifier, an AC potential is applied between the anode and cathode. When the anode is at a higher potential, electrons emitted from the cathode are accelerated toward the anode; current flows from anode to cathode. When the anode is at a lower potential than the cathode, there is no current flow. The diodes used in most power supplies these days are **semiconductor diodes**, usually made out of silicon; they generally run at lower voltages than vacuum tube diodes, but can handle larger currents, which is what the transistors in most modern electronic equipment want. Semiconductor diodes, while they do get warm due to their internal voltage drops, don't need heated filaments, which eventually burn out, nor do they need to be placed in evacuated glass bulbs; they have replaced vacuum diodes for almost all applications.

For this discussion, I will consider the diodes in the bridge as located at 1, 4, 7, and 10 o'clock. The AC voltage in the transformer secondary coil fluctuates; the top is alternately positive and negative with respect to the bottom. When the top is positive, the diodes at 1 and 7 conduct; current can flow out of the diode bridge through the wire connected at 3 o'clock, and return through the wire connected at 9. When the bottom of the secondary is positive, the diodes at 4 and 10 conduct, and the 3 o'clock point remains positive with respect to 9 o'clock, so current continues to flow in the same direction as before.

What's coming out of the rectifier at this point, however, isn't a constant DC voltage; it is, instead, the top halves of sine waves, one after the other, and the output voltage, while constant in polarity, fluctuates all the way down to zero. This isn't usable for running electronic circuits; we need to provide a DC output that remains at or above some fixed level. In order to do this, we put in the **filter capacitors** shown in the circuit diagram. A capacitor stores charge, which can later be extracted as current; the voltage across the capacitor will go as high as the voltage applied across its terminals, and then drop as the capacitor discharges. As the voltage difference between the 3 and 9 o'clock outputs of the rectifier increases from zero to its maximum value, the capacitor charges, approaching the maximum voltage; as the voltage returns toward zero, the capacitor discharges, maintaining some DC voltage across the rectifier output terminals. In order to supply high output currents, it is necessary that filter capacitors have high capacitances, typically thousands of microfarads. The capacitors used generally incorporate an electrolyte in their construction in order to achieve the high capacitance values required; such **electrolytic capacitors** are electrically polarized, and connecting them the wrong way will destroy them.

Once a capacitor is charged, completing a circuit between its positive and negative terminals will discharge it; the rate of discharge, and the discharge current (amperes = coulombs/second) depend on the resistance of the circuit. Short out a charged 10,000 μF capacitor with a screwdriver and you may melt the metal where it makes contact. On a more practical note, it used to be common practice to put a **bleeder resistor** across the output terminals of DC power supplies to gradually discharge the filter capacitors after power was turned off, thus preventing people from getting zapped while working on the electronics. These days, power supplies incorporate additional active devices called **voltage regulators**, which reduce the DC output fluctuations to a much lower level than could be accomplished by a Godzilla-sized filter capacitor.

In the power supply shown, the middle of the transformer secondary is shown connected to ground. The rectifier outputs at 3 and 9 o'clock provide DC voltages above and below ground. Power supplies with such symmetric positive and negative DC outputs, usually ± 12 or ± 15 V, are typically required for analog electronic circuits. Digital or logic circuits more often require a unipolar power supply, typically +5 V and ground. The electronic circuit of such a supply would omit the ground connection at the transformer secondary center tap; the 9 o'clock output would be grounded, the positive output would remain at 3 o'clock, and only a single filter capacitor would be needed.

The power supply of Figure 4-47 is what is known as a **linear power supply**. The current entering the transformer is 60 Hz line current. Many electronic devices, e.g., most computers, now use **switching power supplies**, in which a transformer input current at a frequency ranging from tens to hundreds of kilohertz is generated electronically. Inductive coupling is more efficient at these higher frequencies, making it possible to use smaller and lighter transformers and generate less heat for a given output power. Switching supplies, however, because of their high operating frequencies, are much more likely to generate noise which will interfere significantly with signal processing (at least in the context of flow cytometry) than are linear supplies.

Active Electronics: Tubes, Transistors, ICs

I don't know why I feel compelled to bring up vacuum tubes in a transistorized world. I know I'm not alone; there are rock musicians and audio nuts who insist that vacuum tube amplifiers sound better than transistor amplifiers. They have kept the vacuum tube industry alive, but I'm not sure I believe their claims. I do believe that it is somewhat easier to conceptualize the way vacuum tubes work than it is to do the same for transistors.

Let's go back to the vacuum diode. Following in the footsteps of Lee DeForest, we'll make a similar tube, but we'll put a third electrode between the anode and cathode, making it a wire **grid**, which won't mechanically prevent electrons from traveling from cathode to anode. We now have a three-electrode tube, or a **triode**. We'll heat the fila-

ment, and apply a fixed DC voltage between the anode and cathode, with the anode positive; a fixed current should then flow between the anode and cathode. Now, suppose we apply an AC signal to the grid. The area of the grid is small, so it won't act like a dynode; most of its effects will be exerted by its electric field. Depending on the potential difference between cathode and grid, the grid will either accelerate or decelerate electrons heading toward the anode from the cathode, or, equivalently, increase or decrease current flow from anode to cathode. Thus, there will be an AC current waveform analogous to that on the grid superimposed on the DC anode current. The AC anode current, however, is not derived from the grid input current; it comes from the DC power supply, and its voltage and current can be higher than the voltage and current applied to the grid. Thus, the triode can act as an **amplifier**. This is what made radio and television transmission practical, and what got rid of the big horns required by early phonographs and the even bigger horns needed for the acoustic phonograph recording process.

Vacuum tubes were improved on; further refinements were achieved by adding extra electrodes to make tetrodes, pentodes, etc., but they all used hot filaments and required sealed evacuated envelopes. When one of these components failed, as was inevitable, the tube ceased to function. By the 1950's, there were problems in making electronic computers larger than a certain size; even if it were possible to meet the enormous power and cooling requirements of systems containing thousands of tubes, the tubes' intrinsic failure rate would result in the computers' breaking down at progressively shorter intervals.

Working at Bell Laboratories in the 1940's, John Bardeen, Walter Brattain, and William Shockley developed a solid-state alternative to the vacuum tube, the **transistor**. The first transistors were made of germanium, which, like carbon and silicon, has four valence electrons. "Doping" germanium by adding impurities such as aluminum, with three valence electrons, produces a **p-type** (p for positive) semiconductor, a crystal structure with "holes," or relative deficiencies of electrons. Doping with impurities such as phosphorus, with five valence electrons, produces an **n-type** (n for negative) semiconductor, with extra electrons. A **p-n junction** between p- and n-type materials conducts electricity when a potential is applied to make the p-type material positive with respect to the n-type, but not when the potential is reversed; it therefore acts as a diode.

When fabricated with the appropriate geometry, a three-terminal semiconductor device containing two junctions (p-n-p or n-p-n) can be made to operate in such a way that fluctuations in a small current flowing across one junction are reproduced in the larger current flowing across the other; the device, a transistor, can, like the vacuum triode, be used as an amplifier. I told you it wasn't as easy to explain as a vacuum tube; maybe that's why its inventors got the Nobel Prize. For a lucid nonmathematical explanation of transistors and almost everything else, see Rodney Cotterill's *Cambridge Guide to the Material World*¹⁵².

Vacuum tube electronics wasn't exactly intuitive; most engineers who had mastered it found transistor electronics even more difficult, because currents, rather than voltages, are what count in transistors, while tube electronics dealt primarily with voltages. I won't get into details. What is called a **bipolar transistor** has three electrodes, called the **emitter, base, and collector**; current input to the base modulates current flow between emitter and collector. A newer device, the **field effect transistor (FET)** has analogous electrodes called the **source, gate, and drain**; FETs are best dealt with in terms of voltages; the voltage applied to the gate controls the circuit between the source and drain.

Bipolar transistors have low impedances, and therefore tend to draw relatively high currents, although they don't get as hot as tubes. FETs have very high impedances; much of today's action in electronics depends on **complementary metal oxide silicon (CMOS)** devices, which incorporate FETs. These draw very little current, although their current requirements increase with operating frequency, which is why a microprocessor runs warmer at 2.2 GHz than at 1.1 GHz.

The microprocessor, however, is, as should be obvious, not a single transistor; it is an **integrated circuit (IC)**, fabricated on a single silicon substrate, containing millions of transistors. Transistors carrying small amounts of current don't need to be very large; they could, in theory, be reduced to molecular dimensions, and, in practice, the individual devices in a complex circuit such as a microprocessor may have conductive paths a fraction of a micrometer wide. Integrated circuits were first developed in the 1960's, also eventually garnering their inventors Nobel Prizes, and electronics hasn't been the same since. Except where a lot of power must be handled, almost all analog and digital signal processing is accomplished using integrated circuits; even some power devices, such as audio amplifiers, are now more likely to be built as monolithic structures than assembled from individual transistors.

Analog Nirvana: Operational Amplifiers

As I mentioned before, it was difficult to design real circuits using vacuum tubes, and, for me at least, harder to do so using transistors. If you were designing a tube amplifier, you generally wanted a **voltage gain**; for example, a 1 V peak-to-peak signal out for a 10 mV peak-to-peak signal in, which represents a voltage gain of 100. Transistors, however, were specified in terms of their **current gain**, a parameter known as beta (β). Like a lot of people, I never mastered beta. However, some people who did made it possible for the rest of us to do most of our own analog circuit design by inventing a truly wondrous device called an **operational amplifier, or op amp**.

An op amp is a three-terminal electronic device (neglecting, for the moment, its connections to the power supply and some adjustment points which may be added). It has two inputs, a **noninverting, or plus input**, and an **inverting, or minus input**, and one output. An ideal op amp has

infinite input impedance, that is, it draws no current from a signal source, and zero output impedance, that is, it can supply as much voltage and current as are necessary. It also is capable of infinite voltage gain. However, the op amp is normally operated with its output connected to its minus input through a **feedback network**, which can be anything ranging from a simple wire to a complex circuit which itself contains other op amps. When the feedback network is connected, the output will do whatever it takes to minimize the voltage difference between the minus and plus inputs.

A real op amp may not have infinite input impedance, but those with FET inputs have impedances of about $10^{12} \Omega$, which is close enough. Running on a ± 12 V power supply, a real op amp can put out voltages between -10 and +10 V at currents of at least 10 mA; this is all that's needed for many purposes, and there are ways of boosting both voltage and current outputs to the range of hundreds of volts and tens of amperes. At low frequencies, the gain of real op amps may be in the range of 100,000; many newer devices can give you a voltage gain between 10 and 100 at frequencies of over 100 MHz.

Figure 4-48 shows some basic op amp circuits. The simplest is the **follower**, in which the output is connected directly to the minus input; the output voltage equals the input voltage. Followers are used with high impedance sources, such as pH electrodes; the very high input impedance of the op amp draws almost no current, while the output can provide the same voltage while boosting the current sufficiently to drive a chart recorder, A-D converter, etc.

In the **follower with gain**, the op amp has to make the voltage at the minus input equal to that at the plus input; the output is driving a voltage divider made up of the two resistors, and Ohm's law pretty much gets us the rest. The **inverting amplifier** is also relatively easy to figure out using Ohm's law. The plus input is grounded; the output therefore has to keep the minus input at ground, as well. In an inverting amplifier circuit, the input impedance is not the op amp's impedance, but R_1 , which is the resistance between the input voltage and the **virtual ground** at the minus input.

The **differential amplifier**, as shown, subtracts one input voltage from another; it can also be built in fancier guises, to do sums and differences of more than two voltage inputs, with an additional input resistor required for each input. Gain can also be obtained. Differential amplifiers are useful in several contexts. I mentioned previously that ground isn't always ground, especially when you're trying to look at very low voltage signals. If you run such signals over a shielded twisted pair into a differential amplifier, you end up "bringing ground with you," so to speak. The sum and difference capability is used in audio mixers, and also in the analog fluorescence compensation circuits incorporated in most flow cytometers.

Remember how we got into this discussion? Trying to figure out how to convert the current output from a PMT or photodiode detector in a flow cytometer into a voltage?

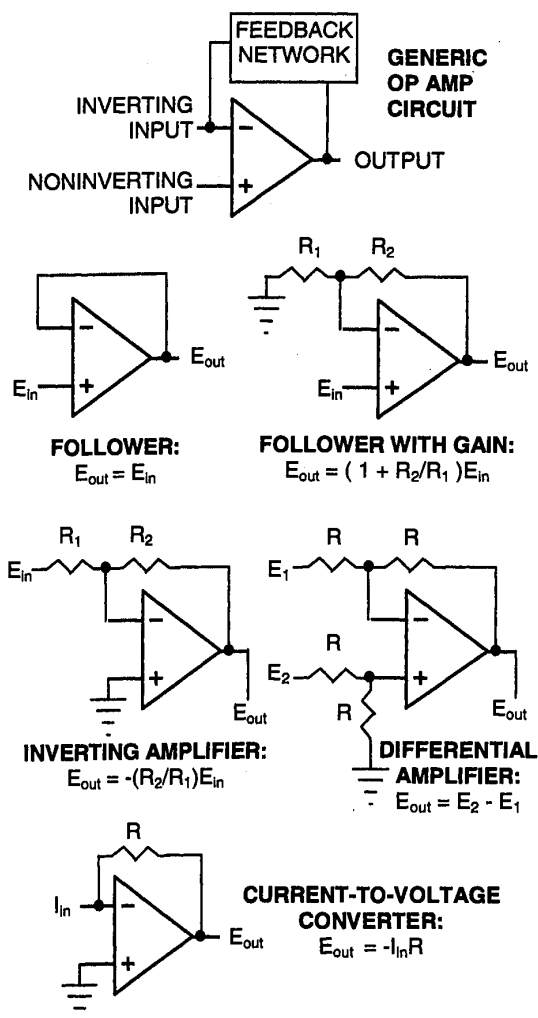


Figure 4-48. Basic operational amplifier circuits.

Guess what. I'm finally getting to the point. We use a **current-to-voltage converter** circuit, also known as a **transresistance amplifier** or a **transimpedance amplifier**. This is pretty much an inverting op amp circuit, minus the input resistor. But hold on a second. Doesn't that give us an input impedance of zero ohms, and an infinite voltage gain? Yes and no. The current is going into a virtual ground, but what we're concerned with is an input current, not an input voltage. In order to maintain the minus input at ground, the op amp has to produce an output voltage that draws as much current as is being put in, so there will be no current flow into or out of the minus input. Ohm's law tells it how to do that. So, on to how we process the signals from our photodetectors.

Detector Preamplifiers and Baseline Restoration

In a real flow cytometer, the PMT anode is generally near ground; the cathode is at minus a few hundred volts.

The current flow is from anode to cathode, so the PMT output will actually draw current from the minus input of the current-to-voltage converter. For reasons discussed previously, we'd like to keep the PMT output current below $100\ \mu\text{A}$ ($10^{-4}\ \text{A}$); we also want to keep the op amp output below 10 V. If we use a 100 K resistor, we'll get 10 V out for $100\ \mu\text{A}$ in. The first stage of our photodetector preamplifier, then, will be a current-to-voltage converter. This would, in fact, be all we needed if it weren't for the background light, which produces a fluctuating DC baseline on which the pulses we're trying to measure are superimposed. Figure 4-49 shows a photodetector preamplifier circuit incorporating baseline restoration.

The first stage of the circuit is the current-to-voltage converter just discussed. The 5 pF capacitor in parallel with the 100 K feedback resistor removes high frequency noise and prevents oscillation. The preamplifier output is taken from the output of the current-to-voltage converter, but this output also goes into a larger feedback network incorporating two additional op amps and associated circuitry. The first stage of this circuit is an inverting op amp with a gain of 2.2, but note that diodes of reversed polarities are connected in parallel with the 22 K resistor in the op amp feedback loop. Wherever the op amp output does, one or the other diode will conduct; they therefore **limit** or **clip** the voltage across the resistor, insuring it will be no more than the diode drop, which for silicon diodes is about 0.6 V. The net effect of the second op amp is that any pulses, either above or below the baseline, get "sliced off"; the amplified baseline goes into the second stage of the baseline restorer.

In this stage, there is an op amp set up as if it were an inverting amplifier, but there is a capacitor instead of a resistor between the output and the minus input. Such a circuit acts as an **integrator**; it can also be conceived of as a **low pass filter**, since the capacitor's reactance decreases with increasing frequency, reducing the effective gain of the circuit. The signal at the output of the third op amp, then, represents the baseline value, averaged over time. Since it has been through two inverting amplifiers, it has the same sign it started out with. We then complete the feedback network of the entire circuit by putting the amplified baseline through a 10K resistor into the minus input of the first stage current-to-voltage converter, using the summing property of the op amp illustrated in the differential amplifier circuit.

What we end up doing, overall, is subtracting a current proportional to the baseline current from the PMT current at the input stage. Feedback is a wonderful thing; you can vary the component values in the baseline restorer circuit over a substantial range without changing the performance characteristics very much. Trust me; I've built a lot of these circuits, and, under virtually all circumstances, I find that the DC output level when no cells are coming through is generally within a few millivolts of ground.

While the circuit illustrated isn't the only one that can be used for baseline restoration, it is reasonably typical, and has been employed in Cytomutts and in some commercial

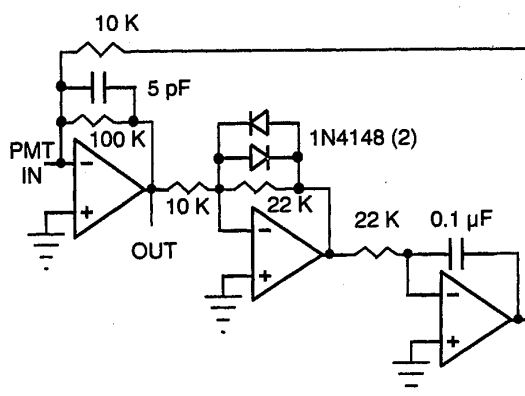


Figure 4-49. A photodetector preamplifier circuit.

systems. Phil Stein and I have fiddled around with resistor and capacitor values and different op amps and tweaked performance some, but what's in the figure still works.

Analog Pulse Processing: Front Ends and Triggering

The duration of the pulses produced by cells' passage through the illuminating beam is quite short. Velocities of 1-10 m/s are commonly used in flow cytometry; at 10 m/s, a particle traverses a $50\ \mu\text{m}$ beam in $5\ \mu\text{s}$ and a $5\ \mu\text{m}$ beam in 500 nsec. Before the option of digital pulse processing became available, it was necessary to use analog **peak detectors**, **integrators**, and **pulse width measurement circuits** to provide short-term (microseconds to tens of microseconds) storage of the appropriate analog values for long enough to permit analog-to-digital conversion of the data.

Before the height, integral, or width of a pulse from a cell can be captured by the appropriate analog circuit, it is necessary to establish that a cell is present in the measurement system; this is the function of a **front end** or **trigger** circuit, which operates on analog inputs derived from one or more detectors. Figure 4-50 gives some idea of what happens in the front end circuit. The baseline-restored preamplifier output signal at the top of the figure, which we will assume is the **trigger signal**, includes a pulse. The trigger signal is fed to the plus input of a **comparator**; a device we have encountered, in company with some of the concepts of digital circuitry, on pp. 27-8. The minus input of the comparator is connected to a circuit that sets a **threshold voltage**, V_{th} . The threshold voltage is usually set with a **variable resistor** or **potentiometer**; one could also set it from a computer using a **digital-to-analog converter (DAC)**, especially considering that good DACs and the associated microprocessor interface circuitry now cost less than precision potentiometers, and have even displaced potentiometers from their traditional roles as audio volume and tone controls.

A comparator is related to an op amp; the signals applied to its inputs are **analog signals**, which can vary in voltage over at least a $\pm 10\ \text{V}$ range. The output produced by the comparator is a **digital signal** or **logic level**. When the

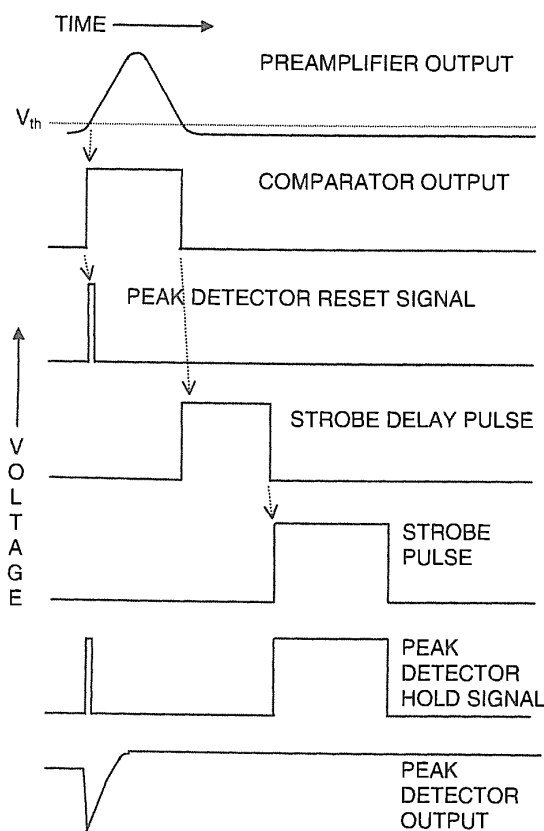


Figure 4-50. Waveforms in preamplifier, front end electronics, and peak detector circuits.

signal level on the plus input is less than or equal to the signal level on the minus input, the output is **logical zero** (typically 0 V, or ground). While the signal level on the plus input is greater than the signal level on the minus input, the output is **logical one** (typically 3.5 to 5 V).

When a cell is not passing through the observation point, the trigger signal is at baseline level, below V_{th} , and the comparator output is at logical zero. As a cell passes through the observation point, a pulse appears on the trigger signal, and, once the pulse voltage rises above V_{th} , the comparator output changes to logical one, remaining there until the trailing edge of the pulse, when the trigger signal again drops below threshold. In some instruments, it is possible to use logical or Boolean combinations (AND, OR, etc.) of the outputs of comparators connected to two or more input signals for triggering. When the comparator signal goes positive, the front end "knows a pulse is there."

Comparators take a finite time to **change output states**, i.e., to change their output levels from logical zero to logical one, after the input signal goes above V_{th} ; the fastest ones may respond in 10 nsec, but they tend to be somewhat unstable, particularly with noisy input signals. The LM311

comparator, which is widely used, takes about 300 nsec to change state. We want to use this change of state to tell a peak detector circuit to start looking for a peak value to "memorize" and/or to tell an integrator and/or a pulse width measurement circuit to get on the job. Figure 4-50 deals only with a peak detector.

Peak Detectors

Peak detector, integrator, and pulse width measurement circuits all store their acquired signal values in **capacitors**. The circuits incorporate operational amplifiers and other active electronic components that prevent the capacitor from discharging unless a digital logic "reset" signal is applied, and allow the input to the capacitor to be disconnected, holding the voltage on the capacitor approximately constant, when a digital logic "hold" signal is applied.

The **peak detector reset signal**, **strobe delay pulse**, and **strobe pulse** shown in Figure 4-50 are all, at least in my "Cytomutt" front end electronics, generated by digital circuits called **monostable multivibrators**, or "**one-shots**." These circuits can generate a logic pulse in response to either the rising edge or the falling edge of another digital signal. The duration of the output pulse of a one-shot is set by the values of an external resistor and capacitor attached to the device; a potentiometer can be used as the external resistor to provide a variable pulse duration.

In the configuration shown in Figure 4-50, the leading edge of the comparator output triggers a brief (< 200 ns) reset pulse, which, when applied to the peak detector, discharges its storage capacitor. During the reset pulse, a hold pulse is also applied to the peak detector, isolating its input and preventing the capacitor from charging. When a hold pulse is not present, the peak detector receives an input signal, which, in this case, is the same preamplifier output shown in the top trace of Figure 4-50. The peak detector's storage capacitor charges, following the input voltage as long as this voltage is increasing. When the input voltage falls below its maximum value, a diode in the peak detector circuit prevents the capacitor from discharging. The schematic diagram of a peak detector is shown in Figure 4-51.

In an ideal circuit, the voltage across the storage capacitor would remain that to which the capacitor was originally

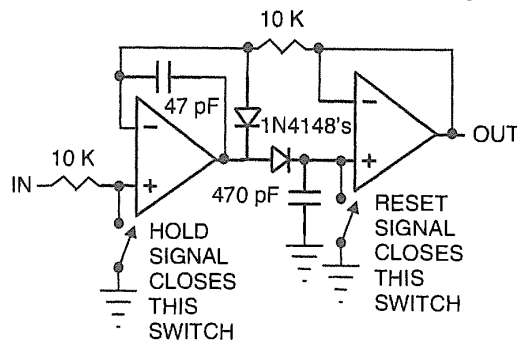


Figure 4-51. Schematic diagram of a peak detector.

charged, as in Figure 4-50; in the real world, charge is lost, resulting in a decrease in voltage, or **droop**, with time. This is observable in the output of a real peak detector, shown in Figure 4-52.

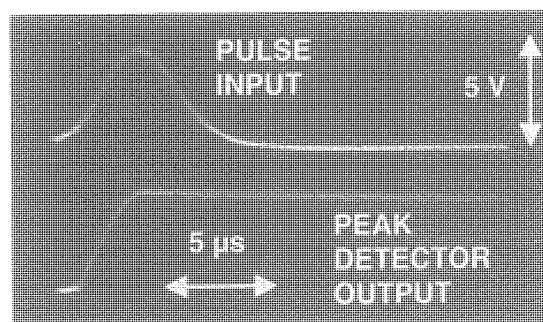


Figure 4-52. TOP: Preamplifier (upper trace) and peak detector (lower trace) outputs from a sample of propidium iodide-stained cell nuclei.

Peak detector performance is affected by the capacitance value chosen for the hold capacitor. Larger capacitors, other things being equal (some types of capacitors are “leakier” than others), decrease the droop, since they take longer to discharge, but this also makes the circuit take longer to reset, decreasing its fidelity of response to rapidly changing signals. Since the peak detectors used in flow cytometers only need to hold signal values for tens of microseconds, at most, i.e., until the values are transferred to a computer, to sort logic, or to other data analysis circuitry, it is generally possible to use capacitors of no more than a few hundred pF.

The reset and hold “switches” shown in the peak detector schematic of Figure 4-51 are electronic switches. The devices used for the task are usually field effect transistors, either discrete or incorporated in integrated circuits. It takes on the order of 100 nsec for the switches to respond, and about the same time for the capacitor to discharge through the resistance of the switch, which is about 100 ohms.

Reset and hold signals, which reach levels of at least 5 V in a matter of a few nanoseconds, are applied to circuit components in close physical proximity to those which carry the analog signals in peak detectors, and a small amount of crosstalk is inevitable. This results in the injection of small amounts of additional charge into the hold capacitor, introducing some inaccuracy into the held signal value. Charge injection errors are relatively inconsequential for signal levels above 100 mV, but seriously compromise the function of peak detectors at lower signal levels.

The strobe delay logic pulse shown in Figure 4-50 is, at least in my circuitry, adjustable in length. It is not strictly necessary in a single-beam flow cytometer, but is needed in a multistation instrument, where it allows the same reset pulse to reset peak detectors connected to signals coming from two or more illuminating beams separated in space. The strobe

delay pulse is triggered by the falling edge of the comparator signal.

The falling edge of the strobe delay pulse triggers the strobe, or “data ready,” logic pulse, which indicates that the outputs of peak detectors, integrators, and/or pulse width measurement circuits have reached stable values after the passage of a cell through the measurement system. Generally speaking, we know, or can find out with the aid of an oscilloscope monitoring the detector signals from the first and last observation points, how much time elapses between these signals. We can then set the length of the strobe delay so that the strobe signal will not be sent until the data from all peak detectors are valid. The strobe signal is transmitted, with the peak detector outputs, to the digital data acquisition system, where the rising edge of the strobe pulse is used to initiate the analog-to-digital conversion process.

The hold signal shown in Figure 4-50 is not generated by a one-shot; it is derived from the output of a digital circuit called an **OR gate**, with the reset and strobe signals connected to the inputs. The output of an OR gate is high when either or both inputs are high. When the hold signal is at logical one, the input to the peak detectors is connected to ground, preventing the output value from changing (neglecting the effects of droop). We have already noted that the hold signal is kept at logical one during the reset pulse, allowing the peak detector capacitor to discharge all the way to ground. Keeping the hold signal at logical one during the strobe pulse insures that the value being digitized by the data acquisition system will not change in the middle of the digitization process.

In the example of Figure 4-50, the interval between the point at which the preamplifier output pulse signal first rises above threshold and the point at which the strobe pulse returns to logical zero represents the minimum time required to process a pulse; this varies from instrument to instrument, but is typically between a few microseconds and a few tens of microseconds. The arrival of a second cell in the measurement region during this interval represents a **coincidence**, and the signal from the second cell can potentially interfere with that from the original cell to produce values of peak height, integral, and/or pulse width not accurately representative of either cell. Different instruments incorporate different circuit arrangements for dealing with coincidences; in simpler systems, a coincidence results in signals from both cells being aborted, while, in more sophisticated apparatus, “pipeline” processing of pulses allows resolution of almost all but the most closely separated cells. If you make the electronics more complicated, you can track as many cells as you want through as many beams as you want; book chapters by Hiebert and Sweet⁶³⁹ and Hiebert¹¹⁵³ provide additional details.

In the actual peak detector circuit of Figure 4-51, the output signal from the preamplifier goes to the positive input of the first op amp in the peak detector through a 10 KΩ resistor, to insure that the peak detector input impedance will be at least that high when the hold switch is closed.

When the trigger signal goes above threshold, the peak detector is reset; the voltage on the 470 pF hold capacitor drops to zero. The output of the first op amp goes to the ungrounded side of the hold capacitor through a diode; as long as the op amp output is higher than the voltage on the capacitor, the diode will conduct, charging the capacitor and increasing its voltage. This happens on the up side of the pulse.

The capacitor is also connected to the plus input of the second op amp, which is connected as a follower; its output is therefore the same voltage as is on the hold capacitor. The second op amp is typically a very high impedance FET-input type, and thus draws only minuscule current from the capacitor. The output is also connected, via a 10 K Ω resistor, to the minus input of the first op amp; the inner feedback loop of the first op amp contains a diode and a small (47 pF) capacitor. In this instance, the diode is connected so it will only conduct when the voltage at the minus input of the op amp is higher than the voltage at the op amp output. This happens after the pulse amplitude reaches its peak and begins falling back toward the baseline. The combined feedback loops compensate for the diode drop; the first op amp puts out extra voltage, so that the peak voltage on the hold capacitor is equal to the peak voltage of the input signal.

Pulse Integral or Area Measurements

When the beam dimension along the axis of flow is larger than a cell diameter, the peak height of a fluorescence pulse should be proportional to the total amount of fluorescent material in a cell. When this dimension is smaller, it is necessary to measure the integral, or area, of the pulse (Figure 4-45; p. 184). There are two basic approaches taken to this measurement in flow cytometers.

The first, and simpler, approach uses the preamplifier itself as an integrator. We have already seen that the pulses we're looking at last for only a few microseconds; the analog electronics used for processing such pulses therefore have to respond as rapidly as they would have to process periodic signals with periods similar to the pulse duration. Thus, the frequency response required of the preamplifier should extend at least to a few hundred kHz and possibly to above 1 MHz. This is an order of magnitude out of the league of stereo; we're right up in the frequency range of AM radio.

The electronics we use in this frequency range tend to become unstable and oscillate unless we add a little bit of capacitance to the circuit. This keeps the system stable, but decreases the high frequency response; what we have done is to insert a **low pass filter** (electronic, not optical) into the circuit. If we add more capacitance, we can slow the preamplifier down to the point at which the output pulse rises and falls at a considerably slower rate than does the input pulse; at this point, the preamplifier is behaving like an integrator. The peak value of the output pulse no longer faithfully represents the peak value of the input pulse, but approximates the area under the pulse. We can then use a peak detector to capture this peak value; its output will be the integral. This is a quick solution, although it gives us

gral. This is a quick solution, although it gives us slow electronics that produce what are known as **long-tailed pulses**, decreasing the rate at which we can process cells without adding further circuitry or a farmer's wife. Because integration using low pass filters changes the shape of the pulse, it is also called integration by **pulse shaping**.

Integration by pulse shaping is generally necessary when logarithmic amplifiers (log amps) are used, because the output signal we need in that situation is the log of the integral of the pulse. If the output of a log amp is routed to an integrator, the integrator output will be the integral of the log, which is not the same as the log of the integral. However, if the input to the log amp is a low pass filtered signal, with the peak proportional to the integral of the original pulse, a peak detector operating on the log amp output will acquire the right signal, i.e., one proportional to the log of the integral.

What I will call the *de rigueur* electronic approach to integration, a circuit called a **gated integrator**, also uses a capacitor to store the integral value. In this circuit, however, speed need not be, and usually is not, sacrificed. The input signal to the integrator is turned on at the beginning of the pulse and off at the end, by a **gating signal** operating the same type of electronic switch that is used to reset a peak detector. The integrator also needs a reset switch to discharge the capacitor when it's time to process the next pulse.

Well, then, can we use the peak detector reset circuit of Figure 4-50 to reset the integrator and/or the comparator output to gate it? No, not if we want accurate values. The timing of reset signals for integrators and pulse width measurement circuits is more critical than the timing for peak detectors. In the sequence illustrated in Figure 4-50, the reset signal is not generated until the comparator changes output state, but this does not occur until the signal level exceeds threshold, which occurs some way up the leading edge of the pulse. An integrator reset at this time would "miss" the first portion of the pulse.

The problem is typically solved with the aid of an additional circuit element called an **analog delay line** to delay the input signal to the integrator by anywhere from half a microsecond up. What is a delay line? Very simple. Remember that current flows through a conductor at the speed of light, which, while it is slower in a conductor than in a vacuum, is pretty fast. A 1 μ s delay line is basically a few hundred feet of wire. There are some tricks involved in dealing with the inductance created by packing all that wire into a reasonably sized package, but the finished product works. The real-time trigger signal is applied to the comparator; and the resultant reset pulse reaches each integrator just as the delayed pulse at its input "begins."

If hold signals for integrators are not applied precisely at the end of a pulse, some error is introduced into the output. A peak detector signal, however, can be held for an arbitrary time, because it reaches its maximum value in the middle of the pulse. Thus, as is also the case for reset signals, the timing of hold signals is more critical for integrators than for peak detectors.

Integrators, like peak detectors, store analog signal values in capacitors, and their dynamic characteristics, such as response time and output droop rate, are similarly affected by the capacitance of the hold capacitor. Charge injection introduces inaccuracies into the output signals of integrators in the same manner as occurs in peak detectors.

Pulse Width Measurement Circuits

A pulse width measurement circuit is essentially a timer that is turned on when the pulse starts, or rises above a certain threshold, and is held at whatever level it eventually reaches once the pulse amplitude falls back below threshold. In the classical analog implementation, the circuit is an integrator; its storage capacitor is charged not by the input signal, but by the input of a constant voltage that is applied after the circuit is reset and disconnected when the hold signal is applied. While it is charging, the voltage on the storage capacitor represents a linear ramp, increasing linearly with time; the voltage on the capacitor at the end of the measurement is therefore proportional to the duration of the pulse.

Pulse width measurement circuits, like integrators and unlike peak detectors, require precision timing to be accurate. A constant threshold pulse width measurement can be timed using a comparator output, with the rising edge of the comparator signal triggering reset and start signals, and the falling edge triggering a hold signal. Timing becomes trickier when the objective is to measure constant fraction pulse width, i.e., the interval between the times the pulse reaches a constant fraction of its eventual peak height.

A pulse width measurement circuit with a direct digital output can be made by replacing the integrator with a digital counter, to which the input is a clock signal. A clock signal is a logic signal that goes from positive to negative at a constant, known frequency; it is usually produced by a crystal-controlled oscillator. The reset signal sets the counter to zero and connects the clock circuit to its input; the value in the counter will then increase linearly with time until the hold signal disconnects the clock, leaving a digital value proportional to pulse duration in the counter.

This sounds like a great idea, but there is a problem. The pulse durations we are interested in measuring are only a few microseconds. In order to measure pulses with a precision of better than one percent, the counter has to end up with a stored value of several hundred. In order to get the counter to count to that value in a few microseconds, we need a fairly fast clock, operating at a frequency of around 100 MHz. It is not at all impossible to obtain clocks that fast, but we could get equivalent or better precision using the linear ramp analog circuit and digitizing the held value; even an inexpensive 12-bit ADC would give us a range of pulse widths between 0 and 4,096, allowing for greater precision than we would get with a digital circuit using a 100 MHz counter.

If all that is required is a relatively crude measure of pulse width, say, one good enough to tell the difference between a

single cell and two or more coming through the system in close proximity, it may be feasible to do without a pulse width measurement circuit. In general, the area of a geometric figure is proportional to both its width and its height; the ratio of the peak and integral of a pulse can therefore be used as an approximation of pulse width.

Analog Pulse Processing: The Bottom Line

Peak detectors, integrators, and pulse width measurement circuits operate on baseline-restored signals, which may be on a linear or a logarithmic scale, with or without fluorescence compensation applied. The held signal outputs of peak detectors, integrators, and pulse width measurement circuits are sometimes referred to as stretched pulses, a term borrowed from nuclear instrumentation, which involves a lot of pulse processing. You could also call them aroused pulses, because they stay up, but let's not. They are also properly referred to as flat-topped pulses, and the flat-topped portions are known as quasi-D.C. signals.

Although we don't use these features explicitly in flow cytometry, pulses in other contexts are often characterized by their rise times and fall times, respectively defined as the times taken to get from 10% to 90% of peak value and from 90% to 10% of peak value. One can make more rectangular flat-topped pulses, i.e., pulses with shorter rise times, by using the strobe pulse to electronically switch the input of an amplifier between ground and the peak detector (or integrator, etc.) output. The output of that amplifier will then be at ground until the strobe pulse comes on, at which point it will go rapidly to the peak (or integral) value; when the strobe pulse goes off, the amplifier output will go back down to ground. The output of such a circuit may have a shorter fall time than would the output of a peak detector following a reset pulse, because the circuit has no hold capacitor to discharge. I used to use circuits of this type to generate real-time dot plots on oscilloscopes; I long ago abandoned both the oscilloscopes and the circuits in favor of digital computation.

In the real world, we have to be careful about choosing which of our measured parameters we use as trigger signals, particularly for integrators, when we are looking at low-level signals such as those obtained from immunofluorescence measurements. Figure 4-50 shows a robust trigger signal pulse rising above a baseline that is as straight a line as the laser printer can produce. Such pulses (and such quiet baselines) are easy to get from forward light scatter or DNA fluorescence measurements of most eukaryotic cells. However, if we look instead at immunofluorescence, the signals are considerably weaker, and the baseline may get noisy. If we try to use a noisy, weak signal as a trigger signal, lowering the threshold in an attempt to capture signals from cells generating low-amplitude pulses, the comparator starts to trigger on the baseline fluctuations and we start to measure background noise that we, or, more to the point, the instrument may count as cells. We also have to be careful, especially in the presence of a noisy baseline, to turn the integrator off (place

it in a **hold** mode, not reset it) after the pulse ends; otherwise, we just keep adding noise to the integral.

Different flow cytometers have different triggering and thresholding schemes; some require that the operator choose a single signal as the trigger signal, while others allow thresholds to be set for all measured parameters. In the latter case, setting the threshold at zero for a particular parameter essentially eliminates that parameter's influence on triggering. I should mention that instruments that make measurements in multiple beams and use analog pulse processing generally require that the trigger signal be obtained from a measurement made in the first beam used. This should be obvious, but every so often I and other people I know have tried to trigger on a signal from a second or third beam and wondered why we weren't picking up any measurements from the first beam. Of course, if there's some pressing reason to trigger on signals from the second beam, it can be done without a time machine by running signals from the first beam through delay lines.

Dead Times, Doublets, and Problem Pulses

The foregoing discussion of pulse height, area, and width measurement touched briefly on the subject of coincidences, but glossed over situations in which it may be difficult to tell that a coincidence is a coincidence. Figure 4-53 gives you the idea. Every flow cytometer, at least so far, has a **dead time**; i.e., a period after the arrival of a cell at the first measurement station during which the apparatus cannot successfully deal with a second cell. This is the hardware equivalent, if you will, of a neuron's refractory period following propagation of a nerve impulse. For the front end electronics described on the previous pages and in Figure 4-50, this period begins when the comparator output goes high and lasts until the end of the strobe pulse. In my single-beam instruments, this is around 50 μ s; when I use two or more beams, it may be as long as 200 μ s.

The dead time is frequently taken as indicating the number of cells per second that can be analyzed; e.g., with a 50 μ s dead time, it should be possible to analyze 20,000 cells/s. It should be, provided the cells are well trained in synchronized swimming and arrive at 50 μ s intervals. Real cells don't; they arrive at random times that generally fit a Poisson distribution. If you measure cell arrival times, and fit such a distribution, you can calculate the fraction of coincidences expected for any analysis rate, and select an analysis rate at which the fraction of coincidences will be no higher than you're willing to accept.

It is generally not a problem to determine that a coincidence has occurred when cells arrive separated enough in space and time so that the signal returns to the baseline. Neither panel A nor panel B in Figure 4-53 presents much of a problem in that regard.

Panel C shows the pulses that might be obtained from two cells passing through in rapid succession; the signal may not get all the way back to the baseline, but it does drop below the threshold level indicated by the dotted line; the

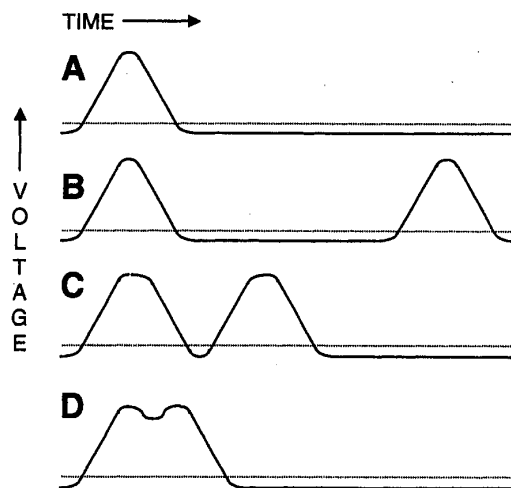


Figure 4-53. Telling two cells from one gets harder as the cells go through closer together.

comparator should go low and then high again. This won't happen when pulses are almost superimposed, as in panel D, which is one pattern which may be observed when a **doublet**, i.e., two cells either stuck together or very close in space, goes through the beam. If you get fancy, it is possible to pick up such occurrences in the front end electronics by measuring widths of comparator output pulses; the more conventional way of dealing with doublets is to wait until data are digitized and compare peak and integral values of signals. Wersto et al have a good recent paper²⁴⁶.

Trigger Happy?

In most flow cytometers, only one signal is used as a trigger signal; the important characteristic is that it have a relatively low background noise level. In immunofluorescence work, it has long been the custom to use forward scatter as a trigger signal, while in DNA analysis, it's as likely to be the fluorescence signal from the DNA stain, e.g., propidium. There are alternatives.

These days, signal-to-noise ratios of immunofluorescence signals are much better than they used to be, and it is often possible to trigger on an immunofluorescence signal. I have, for example, analyzed T-lymphocyte subsets in whole blood, without lysing the red cells, by triggering on fluorescence of a PE-Cy5 tandem-labeled anti-CD3 antibody, and immunofluorescence triggering is featured in at least one commercial instrument intended for subset analysis (The B-D FACSCount). It is advisable to use a log-amplified signal if the raw signal intensity is much below 100 mV, because comparators themselves are likely to behave unpredictably when their input voltages get too low.

In immunofluorescence analysis under other circumstances, I have found it advantageous to trigger on a nuclear (DNA) fluorescence signal; this is also a good way to deal with blood or marrow without red cell lysis, and, in addi-

tion, it allows you to discriminate single cells from doublets. Triggering on nuclear fluorescence isn't new; it was used in the original Cytofluorograf and the Block apparatus for differential leukocyte counting without red cell lysis, and later revived for the same purpose by Terstappen and Loken¹¹⁵⁴. This represents a simple but effective approach to rare event analysis. Nucleated cells account for approximately 1 of every 1,000 cells in blood; if you're interested in nucleated cells, and you trigger on scatter, running 10,000 cells/s, you'll only acquire data from 10 nucleated cells in that time. If you trigger on nuclear fluorescence, you can increase your analysis rate to several hundred nucleated cells a second, because, as far as the instrument is concerned, the other few hundred thousand cells aren't there. It's not difficult to detect populations accounting for a few tenths of a percent of the nucleated cells; these are a few cells per million in the input sample. There are some drawbacks; forward scatter signals are often messy enough to be unusable under these circumstances, and side scatter signals may not be as clean as you would expect, but fluorescence signals are not unacceptably compromised.

There are some circumstances under which there isn't a clean signal usable as a trigger; one of these is analysis of microorganisms. The instrument I use for bacterial analysis employs a technique known as **coincidence detection**. The "coincidence" here does not refer to two cells coming through the instrument in close proximity; what is detected is the coincidence of signals above threshold at two (or more) detectors. In this instance, fluorescence and forward scatter signals, both with relatively high background noise levels, are fed to separate comparators, with independently settable threshold levels; the logical AND of the comparators when both signals are above threshold initiates the reset pulse and other front end signals.

The logic game could be carried further; one might want to detect "live" cells stained green with calcein and "dead" cells recognizable by their red propidium-stained nuclei; in this case, it would be appropriate to use the logical OR of the signals from dual comparators to initiate front end response. Some commercial flow cytometers can implement coincidence detection and logical OR multiparameter triggering; some cannot (see Chapter 8).

Analog Linear, Log, and Ratio Circuits

The signal processing electronics in a flow cytometer have finished their job when they send the data analysis system flat-topped or held pulses, of amplitude 0-10 V or thereabouts, representing pulse peak or integral amplitudes or widths, and an accompanying logic pulse or strobe signal to indicate to the data analysis system that valid data are available.

Certain algebraic manipulations of either raw pulses or held levels are commonly accomplished using analog or hybrid analog and digital electronics, before signals are routed to the data analysis system, rather than digital computation thereafter. The electronic components em-

ployed include **linear sum and difference circuits**, used for fluorescence compensation, **logarithmic amplifiers**, used to facilitate analysis and display of data with values spanning a large dynamic range, and **ratio circuits**, used in fluorescence measurements of certain physiologic probes, compensation for power fluctuations in the light source, and calculation of quantities such as antigen surface density.

Linear Circuits; Fluorescence Compensation

Linear circuits generate outputs that are proportional to sums and/or differences of their inputs; they typically incorporate op amps in the differential amplifier configuration shown in Figure 4-48. They are most commonly used for **fluorescence compensation**¹¹⁵, a practice made necessary by the overlap of emission spectra of fluorescent antibody labels, and previously discussed on pp. 36-8, with the aid of Figures 1-18 and 1-19. You might want to flip back there for a brief refresher.

Fluorescence compensation is, of necessity, an empirical process. The filter passbands shown indicate the range of wavelengths transmitted by each filter; the percentage of light transmitted varies with wavelength within the passband of each filter, and there may be noticeable variation between filters with the same overall specifications. There is some variation in emission spectra of labels; for example, fluorescein emission intensity, when excitation is at or near 488 nm, increases as pH rises from 7.0 toward 8.0. There are also variations from PMT to PMT in responsivity at different wavelengths. In short, there is no way to determine *a priori* what fraction of each signal must be subtracted from the others in order to yield reliable measurements of the emission intensities of each antibody-bound fluorescent label.

Figures 4-54 and 4-55, respectively, illustrate a simple case in which hardware compensation is used, and the hardware used for the compensation. When measurements are made of cells or, in this instance, of beads, bearing both fluorescein (FL)- (green) and phycoerythrin (PE)- (yellow) labeled antibodies, the substantial emission of FL in the yellow spectral region is detected by the yellow fluorescence detector intended for PE measurement. Similarly, the emission of PE in the green spectral region is detected by the green fluorescence detector intended for FL measurement. The spectral overlaps are evident in the left panel of Figure 4-54. The resulting positions of clusters in the dot plot make it impossible to use lines parallel and perpendicular to the axes to separate the display into quadrants containing what we would call dual negative, FL-positive/PE-negative, FL-negative/PE-positive, and dual-positive beads (or cells). Our partitioning of the display, of course, is for the purpose of setting gates that will allow us to count and/or sort the cells in each region. There is no law of nature that demands that the gating regions be rectangular; having rectangular gating regions does, however, let us use somewhat simpler hardware to set gates, and, for better or worse, it seems to be much easier for people to think in terms of rectangular gates.

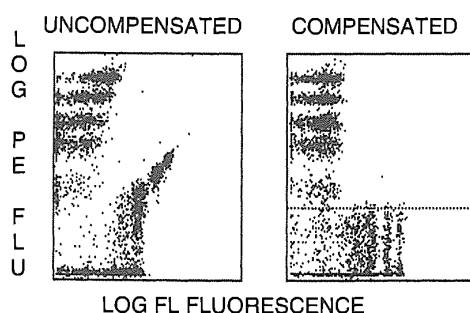


Figure 4-54. Uncompensated and compensated fluorescence signals from FL- and PE-labeled beads.

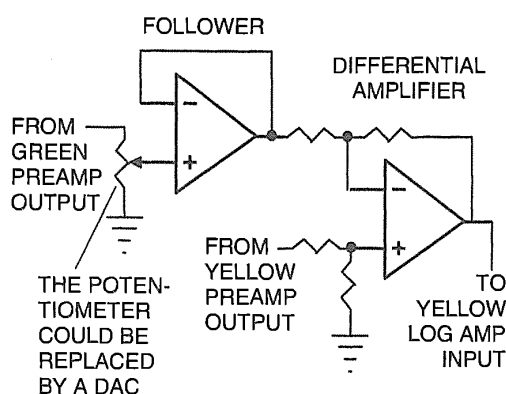


Figure 4-55. One side of the two-color compensation circuit used to generate the compensated data in Figure 4-54.

The data points on the left side of Figure 4-54 could be thought of as clustered along axes which are at an acute angle to one another; what we'd like to do is make the axes orthogonal. There are a number of well-known statistical procedures for doing this, including **factor analysis** and **principal component analysis**. They involve **linear transformation** of the data, which means computing values of a and b such that a data point that would be at (x, y) in the old coordinate system is at (x', y') , where $x' = x - ay$ and $y' = y - bx$, in the new coordinate system.

In the case of the yellow and green fluorescence, we have the old yellow signal, composed of a contribution from PE and a contribution from FL, and the old green signal, composed of a contribution from FL and a contribution from PE, and what we want to do is new yellow and green signals, respectively representing "pure" PE fluorescence and "pure" FL fluorescence.

We make the assumption that the yellow fluorescence from FL is a linear function of the amount of bound FL, which can be estimated from the green fluorescence signal, and the green fluorescence from PE is, similarly, a linear function of the amount of bound PE, which can be estimated from the yellow fluorescence signal. There are some

cases in which this might not be true, for example, if there is substantial energy transfer from FL-labeled to PE-labeled antibodies, but we neglect this possibility.

The output from the green detector preamplifier is routed through the circuit of Figure 4-55, which subtracts a fraction of the output from the green detector from the yellow detector signal. You determine the amount subtracted by turning the knob on the variable resistor, or potentiometer, connected at the input to the follower; if compensation is mediated by the computer, a digital-to-analog converter (DAC) is used instead of a potentiometer to determine how much of the green signal is subtracted from the yellow signal. Another copy of the same circuit, and another potentiometer or DAC, are used to subtract a fraction of the yellow signal from the green signal.

The result is shown in the right-hand display of Figure 4-54. The operator, working by eye, twiddles knobs until the data clusters lie more or less parallel to the axes, so that the display can be partitioned into quadrants in the conventional manner. Since most people find twiddling knobs easier than doing algebra, there aren't many loud complaints about two-color compensation.

Things get somewhat more difficult when three colors, green, yellow, and red, are involved. In theory, it is necessary to subtract a little of both the yellow and the orange signals from the green signal, a little of the green and red from the yellow, and a little of the green and yellow from the red. There are now three circuits, each of which needs two potentiometers or DACs, for a total of six adjustments. Many instrument designs, with some justification provided by the spectra shown in Figure 1-18 (p. 37), omit subtracting red from green and green from red, dropping the number of adjustments to four.

For four colors, green, yellow, orange, and red, full compensation would require four circuits, each with three potentiometers or DACs. The circuit in the yellow channel, for example, would need to have some green, some orange, and some red subtracted; a total of twelve adjustments would be required. Looking to Figure 1-18 again for corners to cut, we find that the green channel should only need one adjustment, for correction from the yellow channel, while the yellow and orange channels would each need three. The red channel would have two adjustments, reducing the total number required from twelve to nine. Anybody who could actually get hired to operate a flow cytometer would be likely to find it easier to do the algebra required to solve the compensation problem mathematically than to adjust nine knobs to get a well-orthogonalized display. Bagwell and Adams¹¹⁵⁵ have described the algebra, software, and control samples required for the mathematical approach, about which considerably more will be said in Chapter 5. To eliminate some of the suspense, let me say here that **neither hardware nor software compensation can fit clusters into quadrants across the full intensity range**; we'll see why later.

There is a more compelling reason than the anxiety induced by large numbers of knobs for switching from hard-

ware to software compensation. Hardware compensation must be applied at the level of the linear signals coming out of the detector preamplifiers; the circuitry, as Figure 4-55 shows, is placed between the preamp outputs and the inputs to the logarithmic amplifiers that were almost universally used in immunofluorescence measurements. As was mentioned on pp. 35-37, a four-decade logarithmic amplifier must deal with inputs covering the voltage range between 1 mV and 10 V; the more circuitry there is connected to its inputs, the less likely it is that the noise level there will remain below 1 mV. If the noise level rises to just over 3 mV, a half-decade of dynamic range is lost; if it reaches 10 mV, a full decade of dynamic range is lost. In short, complex compensation circuitry almost inevitably decreases dynamic range. Recognizing this, Coulter, in designing its EPICS XL four-color, single-laser instrument, eliminated both linear compensation circuits and logarithmic amplifiers in favor of high-resolution analog-to-digital converters and software; other manufacturers have since followed suit.

In addition to their use in hardware compensation, linear circuits have been employed by the Los Alamos group for source noise compensation in extinction measurements⁶⁵⁹ and in a cytochemical method for detection of BrUdR incorporation into DNA without the use of antibodies⁶⁴⁷.

Logarithmic Amplifiers and Dynamic Range

In many applications of flow cytometry, the values of the measured variable span a restricted range. In DNA content analysis, for example, in the absence of clumping and aneuploidy, the brightest cells are twice as bright as the dimmest ones. In other cases, e.g., immunofluorescence analysis, the brightest cells may be hundreds of times brighter than the dimmer ones. In these circumstances, it has often made sense to use an electronic component called a **logarithmic amplifier** (familiarly, if not always contemptuously, known as a **log amp**) somewhere in the signal path, to increase **dynamic range**.

As its name suggests, a logarithmic amplifier produces an output signal, the amplitude of which is proportional to the logarithm of the input signal amplitude. The components typically used in flow cytometers have a range of at least three **decades**, i.e., they can accommodate signals varying in amplitude by at least a factor of 1000. Figure 4-56 illustrates the effect of the logarithmic transformation. The figure shows distributions of fluorescence from a population of beads (Quantum Simply Cellular® beads, Flow Cytometry Standards Corp., now available from Bangs Laboratories) bearing known numbers of molecules of phycoerythrin-labeled antibodies. The top plot shows data on a 256-channel linear scale; the bottom plot shows data on a 4-decade log scale mapped into a 256-channel linear scale. The logarithmic transformation was not, in this instance, done using a log amp; instead, data were digitized to 16 bits precision and the logarithmic values were obtained by digital computation. We couldn't have done any better using an

ideal log amp; we might have done a lot worse using some real ones, for reasons that should shortly become clear.

Flow cytometers generally follow the standard practice of amplifying signals to values ranging between 0 and 10 V before they are converted to digital form. We will therefore assume that 10 V is the maximum output available from our

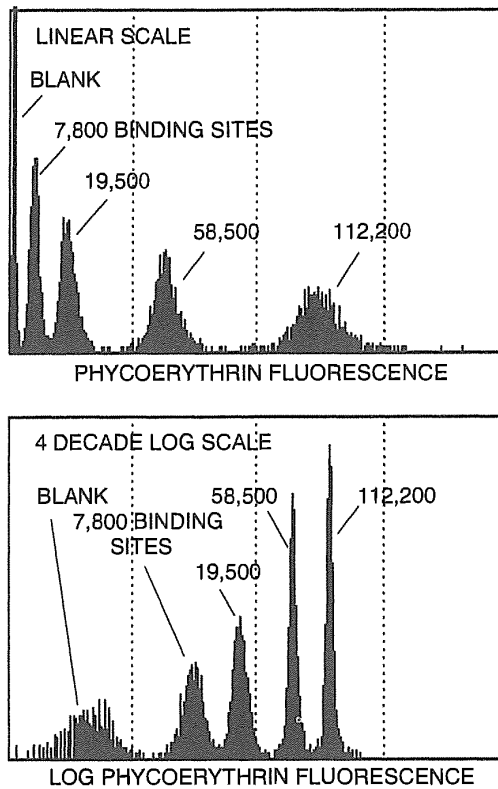


Figure 4-56. Signals from phycoerythrin-labeled anti-CD8 antibody bound to beads, displayed on 256-channel linear and 4-decade logarithmic scales.

preamplifier. If we want to keep our brightest cells on scale, we'll want them to produce pulses with a maximum height, after amplification, of 10 V. Cells which are dimmer by factors of 10, 100, 1,000 or 10,000 will, therefore, produce pulses with heights of 1 V, 100 mV, 10 mV, and 1 mV, respectively.

If we convert linear analog signals to digital signals, using an 8-bit analog-to-digital converter (ADC), we will have a 256 channel range (channels 0-255) corresponding to the range between 0 and 10 V; the difference between channels is just under 40 mV. A 10 V signal will be in channel 255, a 1 V signal in channel 25 or 26, a 100 mV signal in channel 2 or 3, and signals between 1 and 10 mV, or, for that matter, between 0 and 40 mV, will show up in channel 0. For most practical purposes, signals below 100 mV, which would correspond to "weakly positive" cells in an immunofluorescence analysis, will not be distinguishable from signals from "negative" cells.

If we run the linear signal through a logarithmic amplifier, however, we can manage to get everything on scale. Let's imagine a log amp for which 10 V in gives us 10 V out, and let's make the log amp's response 2.5 V/decade, i.e., for every tenfold increase in the input signal, we will get an increase of 2.5 V in the output signal. This should give us four decades worth of dynamic range. For a 10 V signal in, we get a 10 V signal out, for a 1 V signal in, 7.5 volts out, for a 100 mV signal in, 5 V out, for a 10 mV signal in, 2.5 V out, and for a 1 mV signal in, 0 V out. Under ideal conditions, such a log amp should, given the same linear input data, produce the log scale distribution shown in Figure 4-56.

If an 8-bit ADC is used to digitize the log amp output, signals with amplitudes of 1 mV, 10 mV, 100 mV, 1V, and 10 V, which, on a linear scale, would, respectively, be in channels 0, 0, 2 or 3, 25 or 26, and 255, end up in channels 0, 63, 127, 193, and 255. Sounds great. In fact, it is great, except that you may have more than a millivolt of noise running around in the electronics, which would prevent you from getting as many decades of log amplifier dynamic range as you might have paid for (three-and-a-half to four decades are generally provided in flow cytometers). Notice that the data in Figure 4-56 reflect a noise level fluctuating around 3 mV, which is around channel 31 on the 256-channel 4-decade log scale.

Twin Peaks: Distributions on Linear and Log Scales

When you use a linear intensity scale, and you obtain a distribution that has peaks near the low end and near the high end of the scale, the higher peak doesn't look as much like a peak as does the lower one, even when there are approximately the same number of cells in each. When the same data are displayed on a logarithmic scale, it's often easier to perceive the two peaks and to assess the relative sizes of the populations in them.

Figure 4-56 illustrates this point. There are 1,045 events in the peak representing beads with 58,500 antibody binding sites, and 1,180 events in the peak representing beads with 112,200 binding sites. It is not at all easy to guess this from the display on the linear scale; the relationship is more apparent when the data are displayed on a log scale. We can consider each peak in a multimodal distribution such as those in Figure 4-56 as a separate distribution. Suppose we had two peaks centered at channels 100-101 and channels 200-201, with the **coefficient of variation (CV)**, i.e., the standard deviation divided by the mean, 2% for each distribution. Assuming these peaks to be Gaussian, we should find 99% of the events in each within 3 standard deviations of the mean. Looking at the peak at channels 100-101, then, we should find 99% of the events in it between channels 95 and 106, because, if the mean is at channel 100, and the CV is 2%, one standard deviation is two channels. The peak at channels 200-201 is lower, and wider, because, if the mean is at channel 200, one standard deviation is four channels instead of two channels. Thus, to find 99% of the events in this peak, we have to count everything between channels

189 and 212. It's not easy to tell by eye whether there are the same number of events in the lower and the higher peak.

When the data are plotted on a 256-channel linear scale, each channel represents 1/256 of the range of values. When the data are plotted on a 4-decade log scale, each decade occupies 64 channels. Since $\log(a \times b) = \log a + \log b$, we can calculate that the linear value corresponding to channel $(n + 1)$ is $10^{1/64}$ (about 1.0366) times the linear value corresponding to channel n . On a logarithmic scale, distributions with the same CV occupy the same number of channels, regardless of mean channel position.

This property makes log scales particularly well suited, as Brian Mayall has pointed out (not, as far as I know, in print, but in conversations and at meetings), for looking at flow karyotypes, i.e., histograms obtained from flow cytometry of chromosomes. One of the first things you'd like to find out from a flow karyotype is whether there are extra or missing chromosomes. Counting chromosomes is not that hard to do when you look at slides, because all the chromosomes from a particular metaphase stay close to the same place. However, when you do a flow karyotype, you mix up the chromosomes from a few thousand metaphases, stain them with a DNA fluorochrome, and run them through the system to generate a histogram, leaving you with no way of keeping track of how many or which chromosomes came from each cell.

In a real flow karyotype, in which there are a couple of dozen peaks, it is difficult to compare them when they are displayed on a linear scale. On a log scale, the standard deviation ends up being the same number of channels wherever the mean may be, which means that peaks containing the same number of chromosomes should be the same height as well as the same width. This enables you to spot a trisomy or the lack of a chromosome (assuming it occurs in a substantial fraction of the cell population) at a glance.

The real world isn't always so kind. Looking back at the log scale display in Figure 4-56, we notice that the peak representing beads with 19,500 binding sites, which contains 1,002 events, is considerably shorter and wider than that representing beads with 58,500 sites, which contains 1,045 events; the peak representing beads with 7,800 binding sites, which contains 912 events, is shorter and wider still. If peaks with the same CVs and roughly the same numbers of events should be roughly the same height and width when displayed on a log scale, what's the problem?

Simple. The peaks don't have the same heights and widths because they don't have the same CVs. When you're measuring signals varying in intensity over a range of decades, it's almost a sure bet that the weaker signals will be associated with larger CVs. In the case of weak fluorescence signals, photon statistics may be involved, but noise also has its effects. If there are 3 mV of noise on a 10 V signal, or even a 1 V signal, the noise accounts for less than 1% of the signal value. The same 3 mV of noise on a 100 mV signal represents 3% of the signal value, while, on a 10 mV signal, 3 mV noise represents 30% of the signal value.

If you can measure the brightest and dimmest chromosomes with roughly the same CV (and you need a CV around 1% to be able to resolve the different peaks), Brian's suggestion of displaying data on a log scale to spot aneuploidy works. However, even when your CV varies, as in immunofluorescence measurements, log scales make peaks easier to recognize on displays. Unfortunately, you can't always assume that your log amp is generating an accurate log scale.

Falling Off a Log: Log Amps Behaving Badly

What are log amps made of? In the classical log amp, the central component is an op amp, with an element in the feedback circuit generating the log response characteristic. As it happens, the voltage drop across the base-emitter junction of a transistor is proportional to the log of the collector current; at 20 °C, the voltage changes by about 60 mV for each tenfold change in current. The simplest log amp is an op amp with a transistor and a few other components in its feedback loop; its output ranges between -100 and -400 mV.

As you may have guessed, the voltage change across the transistor junction used to generate logarithmic response varies with factors other than current, notably ambient temperature. In order to compensate for this, many log amps incorporate another op amp circuit, with another transistor; the two transistors are typically a matched pair made on a single substrate, and, sharing the same environment, respond identically to changes in ambient temperature. Some of the fancier log amp modules have built-in temperature regulation. With all of that, it is still difficult to get a log amp module incorporating a single op amp-based log circuit to respond accurately over a range of more than 3 decades. For this reason, several manufacturers make log amp modules which incorporate four or five stages of amplification; each stage is optimized for response in a restricted input range, and the sum of the outputs produces a reasonably accurate logarithmic response.

Whether a single-stage or multistage log amp module is used as the nucleus of a log amp, it is generally necessary to amplify the output to get signals into the 0-10 V range prior to digitization, and often necessary to condition the signal by adding or subtracting an offset voltage and/or putting a baseline restorer into the op amp design. Because the logarithm of zero is undefined (or "minus infinity"), even those log amp modules that respond to input signals that are more than a few mV on either side of ground act unpredictably at really low input levels. Other modules need added circuitry to restrict the input range to one polarity. Most log amp circuits incorporate several adjustments for trim, offset, and gain, and require some time on the bench with a skilled technician before they work properly; they don't always stay adjusted.

In 1983, Muirhead, Schmitt, and Muirhead¹¹⁵⁶ described a procedure for measuring the response of log amps and converting data to linear values. Either test pulses from a

light-emitting diode (LED) or fluorescence signals from calibration beads could be used as fluorescence standards. The log amp studied was a component of a commercial flow cytometer; while it was reasonably accurate at some points in its operating range, deviations of 10% or more from true logarithmic response were noted in other regions. Schmid, Schmid, and Giorgi¹¹⁵⁷ reported similar results in 1988, using a calibration procedure in which channel differences between multiple peaks were plotted as PMT gain was varied to place the upper peak at various points in a 256-channel range. They found the characteristic response patterns of individual log amps to be relatively stable over time, allowing the calibration curve to be used to make corrections when transforming logarithmic to linear data.

Figure 4-57 illustrates typical response curves of different types of log amps as determined using a procedure described in 1988 by Parks, Bigos, and Moore¹¹⁵⁸. Measurements are made of bright fluorescent beads; the ratio of fluorescence intensities is determined on a linear scale. A ratio between 1.5 and 2 is desirable. Then, a series of fluorescence measurements are made with the log amp, with an acquisition gate for single beads set using scatter parameters. The gain is initially adjusted to place the median channel of the peak representing the brighter beads at or near the top channel of the ADC (255 in the illustration), and the number of channels' difference between the medians of the bright and dim peaks is recorded. The PMT gain is then lowered to move the median channel of the upper (bright) peak down 5-10 channels, and the number of channels between the peaks is again recorded. The procedure is continued, moving the peaks to lower channels. If PMT voltage must be lowered below 300 V to place the peaks at the lower end of the range, a neutral density filter should, if possible, be placed in front of the PMT, because PMT response may itself be nonlinear at low voltages. The response curve is a plot of median channel of the upper peak, along the horizontal axis, vs. the number of channels' difference between the medians of the bright and dim peaks, along the vertical axis. Since the ratio of intensities of the beads is constant, the difference between the medians of the two peaks, which is proportional to the logarithm of this ratio, should also be constant (log bright = log dim + log ratio). The ideal response curve, indicated by dotted lines in Figure 4-57, is a line parallel to the horizontal axis.

The first real amplifier response curve, at the top, is characteristic of the log amps built in the Herzenberg lab at Stanford¹¹⁵⁹; these incorporate a temperature-stabilized single stage module (Model 2531A, Optical Electronics, Inc., Tucson, AZ). I have also used this module in Cytomutt log amps, and have observed similar response. There is typically a substantial deviation from the ideal at high input signal values, then a response reasonably close to ideal over about a 3-decade range, followed by an abrupt turn up or down at the low end as the noise catches up with the input signal. The amplifier is usable over a 4-decade range.

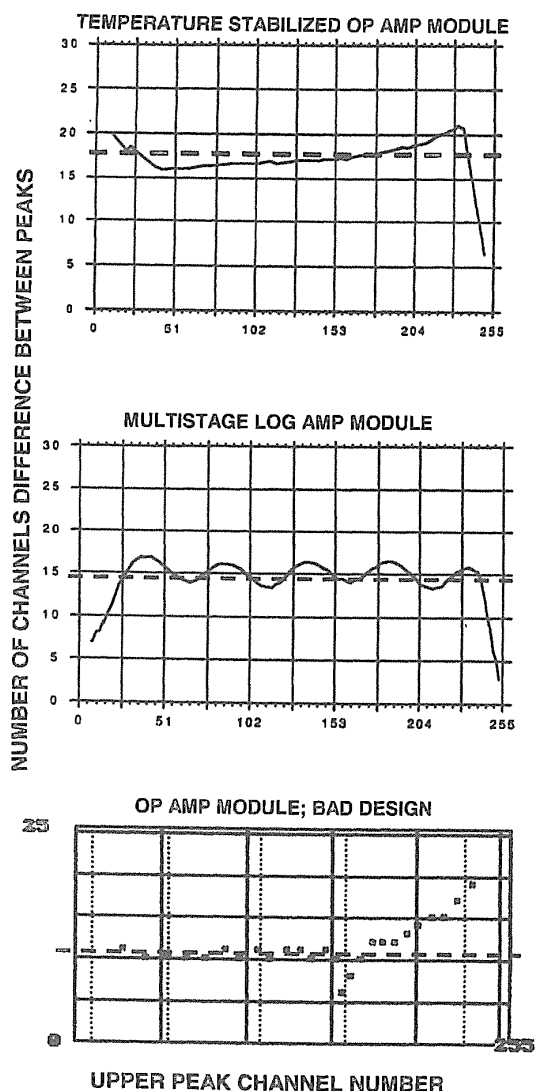


Figure 4-57. Response curves of different types of log amps as determined according to Parks, Bigos, and Moore¹⁵³. The horizontal dotted lines show ideal response.

The middle curve is characteristic of log amps built with multistage modules such as the Texas Instruments (Dallas, TX) TL441, used in some Becton-Dickinson instruments, or the Analog Devices (Norwood, MA) AD640 and its relatives, used in a design described by De Grooth et al¹¹⁶⁹ in 1991. The bumps reflect the response characteristics of the individual stages in the module; there are relatively large deviations at high and low input signal values, with the latter due primarily to noise. Like the single-stage module above, the multistage amplifier illustrated is usable over a range of almost 4 decades.

The curve shown in the bottom panel, generated in my lab, illustrates some highly undesirable characteristics of a

log amp, which were confirmed by Dave Parks and Marty Bigos when I sent them one of my early design efforts to test. The curve deviates wildly from ideal at the high end, and includes a relatively sharp notch in the third decade. This log amp would occasionally yield distributions with crimps in them; when it didn't, however, it seemed to be pretty good, because it was accurate at the low end and could handle an input range that appeared to be larger than 4 decades. However, it would have been an extremely bad idea to try to do log-to-linear conversions with this beast; it wasn't even stable over time. Bad log amps have also, unfortunately, turned up in some commercial instruments.

The take-home message here is: calibrate your log amps, even if you're not trying to do quantitative immunofluorescence, just so you know. And hope that the next flow cytometer you buy will have replaced the #@%\$&^% log amps with trustworthy linear circuits and high-resolution ADCs. It will make flow cytometry in the rest of this century much easier.

Limits to Dynamic Range

If we had a perfect 5-decade log amp, it might not help much in extending the dynamic range of flow cytometric fluorescence measurements, for reasons best explained with the aid of Table 4-4. In previous discussions of PMTs and preamps, we established that we usually don't want to get more than 100 μ A out of the PMT, and that the maximum signal we can put into the log amp is 10 V. Suppose our baseline-restored preamp gets 10 V out for 100 μ A in, and drives a 5-decade log amp with an output slope of 2V/decade, and we measure 10 μ s fluorescence pulses at 520 nm, with an R928 PMT. The tube's cathode quantum efficiency at 520 nm is 14%; the cathode current is 59 mA/W. We assume a PMT gain of 10,000.

The anode current of 100 μ A yielding a 10 V preamp output at the top of the highest (fifth) decade results from amplification of a cathode current of 10 nA by the PMT gain of 10,000. Since a 1 A current is about 6.2×10^{18} electrons/s, 10 nA in a 10 μ s pulse represents 620,000 photoelectrons. Dividing by the quantum efficiency, 0.14, we see that 4,430,000 photons reach the cathode during the pulse. The contribution to measurement CV from Poisson photoelectron statistics is $[(620,000)^{1/2}/620,000]$, or 0.13%, based on the number of photoelectrons emitted from the cathode. We use the number of photoelectrons because they are what we are actually counting; we are only inferring that the photons were there. If we used a 16-bit (65,536 channel) linear ADC, which has a difference between channels of about out 153 μ V, the 10V signal would be in the highest channel, i.e., channel 65,535.

The top of the next (fourth) decade is a 1 V signal, corresponding to channel 6,553; the contribution to CV, from 62,000 electrons, would be 0.4%. The top of the third decade, at 100 mV, or channel 655, corresponds to 6,200 photoelectrons; the contribution to CV is 1.3%. The top of the fourth decade, at 10 mV, or channel 66, represents 620 elec-

	LOG AMP OUTPUT	LOG AMP INPUT	ANODE CURRENT	CATHODE CURRENT	PHOTONS/ 10 μ s	ELECTRONS/ 10 μ s	CV (%)	CHANNEL (16 BIT ADC)
5	10 V	10 V	100 μ A	10 nA	4,430,000	620,000	0.13	65,535
4	8 V	1 V	10 μ A	1 nA	443,000	62,000	0.4	6,554
3	6 V	100 mV	1 μ A	100 pA	44,300	6,200	1.3	656
2	4 V	10 mV	100 nA	10 pA	4,430	620	4.0	66
1	2 V	1 mV	10 nA	1 pA	443	62	12.9	7
DECADE	0 V	100 nV	1 nA	100 fA	44	6	40.8	0

Table 4-4. Logarithmic amplifiers: What goes in, what comes out.

trons; the contribution to CV increases to 4%. At the bottom of the fourth decade, 1 mV, or channel 7, there are 62 electrons in the pulse, with the contribution to CV 12.9%, or almost one channel width ($1/7 \approx 14\%$). The bottom of the fifth decade, a pulse from 6 photoelectrons, ends up at channel 0, the contribution to CV would be 40.8%.

These are actually best case calculations. The numbers of electrons are calculated for 10 μ s rectangular pulses; real pulses from a flow cytometer are apt to be shorter in duration and Gaussian in shape, and would probably represent something less than half the numbers of photons and photoelectrons used in the table. The PMT gain of 10,000 used in the calculations is conservative; it is typically attained with PMT voltages of 400-500 V. However, if the calculations are repeated assuming higher PMT gains, we end up with even fewer photons at the cathode and fewer photoelectrons emitted, so things only get worse. If the gain is increased to 100,000, the lowest decade gives us only 6 photoelectrons in 10 pulses.

In my experience, displays of flow cytometric immunofluorescence data on 4-decade log scales that place the brightest stained cells near the top of the top decade almost always have the unstained cells occupying at least half of the bottom decade, while, in displays that place the unstained cells near the bottom of the bottom decade, the brightest stained cells fall below the midpoint of the top decade. This suggests that the real dynamic range of most immunofluorescence data is closer to 3 1/2 than to 4 decades. We have established that 3 1/2 decades of dynamic range are fairly easy to get; with a 3 mV noise level, we can't do any better than that. If we reduce the noise, another 1/2 decade is

probably within reach, especially if we settle for lower accuracy. Five-decade response from a single detector approaches the realm of fantasy.

We end up at pretty much the same point if we stop thinking about photons and electrons for a bit, and think instead about numbers of antibody molecules detectable on cell surfaces. If the high end of the top decade represents 1,000,000 molecules, the low end of the fourth decade is 101 molecules, while the fifth decade encompasses the range between 11 and 100 molecules. Calculations of the minimum number of fluorescent molecules detectable by conventional flow cytometers generally end up with figures above 100 molecules. Such calculations, which will be discussed in detail later on, are typically done using data collected with log amps, and may in themselves be suspect. I have produced differences of several hundred molecules in calculated detection limits, simply by interchanging log amps with slightly different DC offset levels, and therefore believe that some of the lowest reported threshold values may well be due to artifacts. This issue should be resolved within the next few years, but the only practical way to resolve it will be to determine sensitivity using linear amplification and high-resolution analog-to-digital conversion instead of log amps, to which good riddance.

None of the above should be taken as preventing you from making flow cytometric measurements over a dynamic range larger than 4 decades; you just can't do it with one detector and keep everything on scale. If you use one detector, at low gain, for the bright stuff, and another, at high gain, for the dim stuff, you can easily cover almost 8 decades. Marine biologists probably analyze samples encompass-

ing the largest dynamic range; with five detectors, they could go from marine bacteria (10^{12} g) to whales (10^8 g). Even if the largest objects in the sample are only krill, the krill (and Shapiro's First Law), rather than detector dynamic range limitations, are the main obstacles to success.

Ratio Circuits

Ratio circuits are analog circuits that produce an output proportional to the ratio of two input signals. They are usually made from modules called **analog multipliers**, which can also be connected to produce signals proportional to the product, square, or square root of input signals. One application of ratio circuits has been in the derivation of signals proportional to **surface density of antigenic or receptor sites**, which can be calculated by dividing the number of bound ligand molecules by the cell surface area. An accurate approximation to cell surface area can be obtained from digital computation of the $2/3$ power of volume¹¹. Use of this technique is obviously feasible only with flow cytometers capable of making volume measurements. However, since both cell surface area and the "cell size" measurement obtained from small angle light scattering are proportional to the square of cell diameter, the ratio of immunofluorescence and scatter signals can serve acceptably as a parameter representative of antigen surface density for making comparisons^{657,658}.

Compensation for light source intensity fluctuations using a ratio circuit was discussed on pp.147-8 and illustrated in Figure 4-24. If the output of the light source increases or decreases, fluorescence and scatter signal amplitudes from identical particles will increase or decrease proportionally. If the source intensity is continuously monitored by a dedicated detector, the signal produced by dividing the raw fluorescence and scatter signals (i.e., the signals before baseline restoration has been applied) by the signal from this detector will be free of variations due to source fluctuations. Steen¹⁰² has published a schematic for a source noise compensation circuit.

Ratio circuits are also useful for measurements of probes such as BCECF and indo-1, which change spectral characteristics with pH, cytoplasmic $[Ca^{++}]$, etc. In some cases, a single excitation wavelength is used and the ratio of emission intensities at two wavelengths is computed; in others, two excitation beams are used and the ratio of emission intensities at one wavelength is computed. Ratiometric measurements have already been alluded to on p. 47; they will be discussed further in Chapters 7 (Parameters and Probes) and 10 (Using Flow Cytometers).

4.9 DIGITAL SIGNAL PROCESSING

The analog signal processing techniques we have just discussed leave the information from cells in the form of **stretched pulses**, i.e., slowly varying (constant in the absence of droop) voltage signals, stored in the short-term analog memories of peak detectors, integrators, and/or pulse width measurement circuits. In some cases, additional elec-

tronic circuitry may have been used to form linear combinations of signals, as in fluorescence compensation, to convert signals from a linear to a logarithmic scale, or to derive a signal representing the ratio of two other signals. The front end electronics, which include some rudimentary digital circuits, generate a digital logic signal, the strobe pulse, to inform a conventional data analysis system that all the analog voltage signals from a cell (or, more accurately, an event) are ready to be **digitized**, i.e., converted into numbers that can be further manipulated by the digital computer used for data analysis. We have already had a first look at digitization, and at the **analog-to-digital (A-D or A-to-D) converter (ADC)**, which is the hardware responsible for the task, in Chapter 1.

Most commercial and laboratory-built flow cytometers still use analog and hybrid circuits for pulse processing, but, over the years, more and more builders and manufacturers of instruments have introduced digital processing at progressively earlier stages. The Coulter EPICS XL¹¹⁸³ and some of my newer Cytomutts²⁴⁶, for example, dispense with log amps, and use high-fidelity integrators or peak detectors and high-resolution ADCs for conversion between linear and logarithmic scales. This allows fluorescence compensation and ratio calculations to be done in software, eliminating the need for analog compensation and ratio circuits.

Analog-to-Digital Conversion

An analog-to-digital converter, as the name implies, has an **analog** input, usually a voltage in the range 0-10 V, 0-10.24 V, or something like that, and a **digital** output, which is a binary number with 0's and 1's represented by different logic voltage levels. For this and future discussions of logic circuits, I will stick to the ground-for-0, 5 V-for-1 "positive logic" convention which roughly represents the voltage levels encountered in older TTL and CMOS digital circuits. Just so you know, these days, microprocessor chips and their associated logic run on lower DC voltages, from below 2 V to 3.3 V, and even analog circuits tend to run on ± 5 V supplies, with ADC input voltage ranges typically ± 1 or 2 V. It doesn't really matter; the whole point of A-to-D conversion is to leave voltages and currents behind and convert the data to numbers.

Analog signal processing is pretty much instantaneous, and definitely continuous. The output of a current-to-voltage converter circuit, or an analog log amp, changes as the input changes. Some digital circuits also change outputs whenever their inputs change. By contrast, analog-to-digital converters, like digital computers, do not, by and large, change their outputs "instantaneously"; they take a finite time to do what they do, and the process, more often than not, involves a sequence of steps.

The rate at which the devices execute that sequence of steps is controlled by an internal or external **clock** signal. Clock signals were mentioned on p. 195, in the discussion of digital approaches to pulse width measurement; in the world of microprocessors and digital logic, clocks, overclocking,

and clock envy are ubiquitous. You have undoubtedly noticed that AMD and Intel regularly compete for the fastest clock speed, with the current loser always claiming its processor is faster even if its clock is slower. And the fastest ADCs are no slouches in the clock department; some convert data at rates of several GHz. The ADCs we use in flow cytometry, even for digital pulse processing, are quite a bit slower, at least for now, but, as will soon become obvious, speed isn't everything.

Free Samples? Hold it!

If an ADC takes a finite time to operate, you might well ask how we can feed it a continuous input signal. The answer is that we can't. Instead, we take a series of **samples** of the input signal. In what now seem like ancient times, this required the use of a completely separate circuit called a **sample-and-hold**. A sample-and-hold operates in a fashion similar to a peak detector (pp. 192-4; Figures 4-51 and 4-52). It is controlled by "sample" and "hold" logic signals; the "sample" signal allows the input signal to charge a capacitor (called a "hold" capacitor), and the "hold" signal disconnects the input signal from the capacitor. The output of the circuit is the capacitor voltage, buffered by a high-impedance op amp.

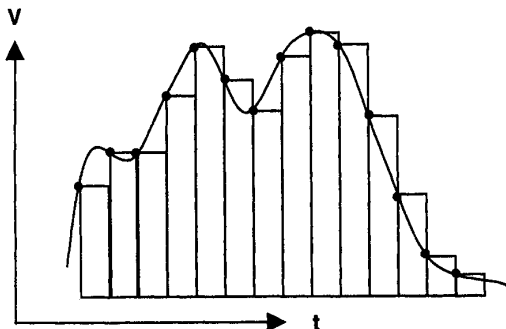


Figure 4-58. Continuous and sampled signals. The filled circles represent the sample points.

Figure 4-58 illustrates the effect of sampling an arbitrary voltage waveform. The dark filled circles represent sample points, taken at regular intervals; each sample point is an accurate representation of the value of the waveform at the point at which the sample was taken. If we used a sample-and-hold in the sampling process, the output of the sample and hold would, in the absence of droop, remain at the value of the last sample and then jump rapidly to the value of the next sample. Rapidly, not instantaneously, because it takes time for the hold capacitor to charge or discharge to the new sample voltage. But can we "trust" the series of samples to faithfully represent the waveform?

The short answer is yes, provided we take samples at short enough intervals. Using **Fourier analysis**, it is possible to represent any arbitrary waveform as the sum of sine

and/or cosine waves of discrete frequencies. According to the **Nyquist sampling theorem**, it is possible to completely reconstruct a waveform from samples taken at twice the frequency of the highest frequency component of the waveform. If we look at the waveform in Figure 4-58, we see some sine wave-like "ups and downs," but the sample points appear to be taken frequently enough to reproduce those fluctuations. If the waveform had changed more rapidly than the one shown in the figure, or if we had taken samples less frequently, that might not have been the case; imagine what the "connect the dots" line, or the "staircase" sample and hold output, would look like if we only used every other sample point, or every third sample point.

The process of reconstructing a waveform from samples also typically involves Fourier analysis; it generates coefficients for sine and/or cosine waveforms at the various frequencies that make up the original waveform. If we attempt to reconstruct a waveform from samples taken at a rate below the **Nyquist frequency**, which is twice the frequency of the highest component of the waveform, the computer we're using won't suddenly emit sparks and smoke; instead, it will give us an initially reasonable-looking set of coefficients that gives substantial weight to frequencies that weren't actually of significance in the original waveform. This is described as **aliasing**, or **sample aliasing**, and, since it leads to totally incorrect results, should be avoided. But how?

The necessary fix is an **antialiasing filter**, a low pass (electronic, not optical) filter that is inserted into the signal path somewhere ahead of the sample-and-hold. That sounds easy enough, and, because engineers have devoted a great deal of time and effort over the years to the design of filters with appropriately smooth responses in their passbands and appropriately sharp high frequency cutoff rates, we can take it for granted that a suitable filter will be available. In fact, the filter, and the sample-and-hold, may well be built into the same module as the analog-to-digital converter itself.

The bottom line on sampling, then, is that samples are "free"; as long as we take enough of them, the discrete samples will contain all the information that was found in the original continuous signal. However, we will have to pay the piper when we actually do our analog-to-digital conversion.

Quantization: When Are Two Bits Worth a Nickel?

Since we now live in a digital world, we are likely to have opinions on how many bits we need for this or that. The 16-bit digital audio recording that seemed so awesome in the early days of compact discs is now looked at with disdain; even music hobbyists can now afford computer-based hardware with 24-bit ADCs. In deciding how many bits we need for acquisition and analysis of data from cytometry, we have to consider what we want to achieve in terms of precision and dynamic range, and determine how each of these will be affected by **quantization error**, which is an essential, necessary, unavoidable characteristic of the analog-to-digital conversion process.

The ADCs used to date for flow cytometry generally convert to somewhere between 8 and 20 bits' resolution. An n -bit converter has 2^n channels, and we typically consider its output in terms of channel numbers between 0 and $(2^n - 1)$. The corresponding channel numbers, and the voltages represented by the **least significant bit (LSB)**, i.e., the voltage difference between channel n and channel $(n + 1)$, assuming 10 V full scale, are shown below in Table 4-5.

Number of Bits	# of Channels	Voltage/LSB
8	256	39.1 mV
10	1,024	9.77 mV
12	4,096	2.44 mV
14	16,384	610 μ V
16	65,536	153 μ V
18	262,144	38.1 μ V
20	1,048,576	9.54 μ V
22	4,194,304	2.38 μ V
24	16,777,216	596 nV

Table 4-5. Characteristics of analog-to-digital converters.

For an 8-bit converter, the lowest voltage that can make the output of an ADC register as channel 1 instead of channel 0 is just over 39.1 mV. Any signal level between 0 V (ground) and 39.1 mV at the input would produce a digital output of 0, and any signal greater than 39.1 mV and less than or equal to 78.2 mV would produce a digital output of 1. At the top end, a signal level greater than 9.961 V would produce a digital output of 255.

By convention, the voltage value representing a given channel is taken as the middle of the range of voltages that would be converted to that channel value. In the case of the 8-bit, 10 V full scale ADC, the voltage values would be 19.55 mV for channel 0, 58.65 mV for channel 1, 9.98045 V for channel 255, etc. For any given channel, the **quantization error** is constant in magnitude at one-half the channel width, or $\pm (1/2)$ LSB. In effect, digitizing the signal has the same effect as adding a small amount of **quantization noise**, with the noise having a mean of 0 and a uniform, or rectangular, distribution between $-(1/2)$ LSB and $+(1/2)$ LSB. The standard deviation of the noise is 0.29 LSB; this emerges from the properties of rectangular distributions, rather than from my hat. So, while sampling a signal does not degrade it, digitizing a signal does.

While its magnitude is constant, the relative importance of quantization error is quite different at opposite ends of the measurement scale. At the top end of the range, channel 255, the error is small; for the 8-bit converter, the peak-to-peak range of quantization noise is only 0.391% of full scale. However, at the bottom end of the range, the error is large; for channel 0, the peak-to-peak range of quantization noise represents 100% of the channel value for an 8-bit converter - or, for that matter, for a converter with any number of bits, since the channel value for channel 0 is always $(1/2)$ LSB.

Suppose we are measuring DNA content in a tumor specimen, and want to be able to clearly resolve peaks from

tumor and stromal cells differing in G_0/G_1 DNA content by 1% or less. If we use a linear scale, we need to place the G_0/G_1 peak of a DNA content distribution at a value less than one-half of full scale in order to get both the G_0/G_1 and G_2/M peaks on scale. If we use an 8-bit converter, and the G_0/G_1 peak of the stromal cells appears at channel 100, the corresponding voltage is 3.91 V. The peak-to-peak quantization noise at channel 100 is 1% of the signal value; the standard deviation of the noise is 0.29 percent of the signal value, and, if there are any other contributions to the variance of the measurement, other than quantization noise, we will probably not be able to resolve the tumor and stromal cell G_0/G_1 peaks.

As I have pointed out elsewhere⁸, DNA is the only substance present in cells in which cell-to-cell variations in content are so small that measurement to a precision of 1 percent is warranted. Distributions of cellular DNA content represent most of the few "needles" among the many "haystacks" found in biology. Sure, if you run plastic beads through the instrument, you can get scatter distributions with CVs less than 1 percent, but you don't need anything like that kind of precision to measure light scattering from biological particles, which give distributions with CVs no less than 7-10%. In fact, the major contributors to variance in measurements of almost all cellular parameters are biology and photoelectron statistics.

The first data acquisition systems for flow cytometers dealt only with data reduced to stretched pulses representing peak heights, integrals, etc., and used 8-bit ADCs quite successfully. Most instruments weren't measuring DNA content with high (CVs of 1% or less) precision. Measurements of low intensity fluorescence, which had high CVs due to both photoelectron statistics and the biological characteristics of samples, were made using log amps, which placed signal levels high enough up on the 256-channel scale so that quantization error was not a concern. Almost all of the more recent instruments use 10-bit ADCs, which are more than adequate for high-precision DNA content measurements on a linear scale.

We run into problems, however, when we try to take logarithmically transformed data digitized with a 10-bit converter and interconvert between log values and linear values. This becomes necessary when we want to revise compensation settings after the fact using software. Plots of data that have been thus manipulated tend to have a "grainy" quality, although some software packages smooth out the transformed data by **dithering**, i.e., adding small amounts of additional noise in the form of random numbers. This subterfuge inevitably further degrades data, although usually not significantly, and intuition suggests that we could eliminate the need for it by using higher resolution ADCs. But how many bits do we need? Would 16 bits be enough?

A look back at Figure 4-56 (p. 199) tells us "almost, but not quite." The figure shows fluorescence distributions of beads bearing different, known amounts of fluorescent antibodies; data are displayed on a 256-channel (8-bit) linear

scale in the top panel and on a 256-channel log scale in the bottom panel. However, the raw data were taken from one of my "Cytomutts"²⁴⁶⁷, using a 16-bit (65,536-channel) ADC; conversion to a 256-channel log scale was done by a digital computer using a look-up table.

We note that the peak representing the blank beads in the log scale display shown in the bottom panel of Figure 4-56 looks different from the other peaks in the histogram; the lower half of the peak has what has lately come to be known as a "picket fence" appearance, with no events appearing in roughly every other channel. This tells us that we will "run out of bits" when we try to fit 16-bit data onto an 8-bit, 4-decade logarithmic scale.

The look-up table we use for the logarithmic conversion is a table of 65,535 bytes, with "addresses" 0-65,535. The byte at address n contains an 8-bit number between 0 and 255 representing the log value corresponding to the linear value n . When we compare data displayed on linear and log scales, which we can do with another backward glance at Figure 4-56, we see that the log scale makes peaks at the high end of the scale narrower and peaks at the low end of the scale wider than they appear on a linear scale. It is therefore not a surprise to find, when we calculate the numbers in the look-up table, that, at the high end, one value on the logarithmic scale represents a large range of linear values.

In fact, 2,316 linear values, between 63,220 and 65,535, all map to log value 255. And 2,234 values, between 60,986 and 63,219, map to log value 254. As we move toward the lower end of the scale, fewer and fewer linear values map to a given log value. By the time we get down to channel 22 linear (channel 34 log), there is only a single linear value for each log value. And, for linear values lower than 22, we actually have a choice of two or more log values; since there can only be one value at any address in a look-up table, it is necessary to select from the alternatives.

That is what I had to do when I worked out the look-up table used to generate the log values shown in Figure 4-56. For linear values of 18, where the choice of log values was 28 or 27, I chose 28, so there were no points plotted for channel 27. For linear values of 17, where the choice was 26 or 25, I chose 26; there were no points plotted for Channel 25. That's how the "picket fence" got built; I can whitewash it somewhat by pointing out that my instrument noise level was right around channel 21 linear, anyway, and there wasn't any point in trying to make data at that level and below look better. However, I would have been able to eliminate the picket fence effect using higher resolution ADCs. The problem I had was that, when I built my 16-bit data acquisition system, there pretty much weren't any higher resolution ADCs fast enough to handle the data at the rate at which they were generated.

When Bob Auer was designing the electronics for the Coulter EPICS XL¹¹⁸³, in which the goal was to eliminate log amps and allow data to be displayed on a 10-bit (1,024-channel) log scale, he calculated that he would need 20-bit

linear data to obtain a display without a picket fence. And he faced the same problem I did; there weren't any usable 20-bit ADCs. So he used a "divide-and-conquer" approach usually known in technical jargon as **subranging**.

The XL splits held signals from each of its channels; one is fed directly into a 15-bit converter, and the other is amplified, using analog circuitry, by a factor of 32, and then fed into a 15-bit converter. For the smallest signals, the unamplified signal converts to a digital value of zero; the digital value used for data analysis is the value obtained from the amplified signal. For the large signals, the digital value obtained from the unamplified signal is shifted left by five bits (equivalent to multiplication by 32), and this 20-bit value is used for data analysis. The large signals have 20-bit values, in which the lower 5 bits are all zeros, but they still have 15-bit precision; the 15-bit data from the small signals are treated as 20-bit values in which the upper 5 bits are all zeros. For the signals on the borderline, a switch controlled by a comparator is used to determine whether the digital value used will be from the amplified or the unamplified signal, and a small amount of noise is fed into the comparator to produce slight changes in the position of the crossover point; this avoids the generation of "spikes" or "notches" in histograms, which would otherwise be likely to occur at the crossover point because of slight differences in the gain of the analog and digital signal amplification. Most important, there are no picket fences to be seen. The design works.

Both my quick-and-dirty 16-bit data analysis system and the more elegant 20-bit data analysis system in the EPICS XL eliminate the need to use log amps, and allow fluorescence compensation and other procedures involving log-to-linear and linear-to-log conversion to be done by software without making the data look ratty. However, they still operate on held analog signals, and that poses a problem. If we want a 4-decade dynamic range, we have to be able to get accurate integrals or peak heights of pulses ranging in amplitude from 1 mV to 10 V. It is not at all easy to make peak detectors and/or integrators respond over such a large dynamic range. Designing the integrators for the EPICS XL to meet this specification was (R. Auer, personal communication) harder than designing the data acquisition and analysis system; I gave up on trying to design peak detectors with the necessary performance for my "Cytomutt" and, instead, bought them (ironically enough, from the same company from which I used to buy log amp modules). And the peak detectors only work in my 16-bit systems when they are performing well beyond the manufacturer's spec; the last few I bought are up to spec but not good enough to use.

The obvious solution to the problem would be to eliminate peak detectors, integrators, etc., in favor of **digital pulse processing**. It sounds easy enough; after all, that's the way we process all of our audio signals these days. In order to understand why digital pulse processing has taken as long as it has to find its way into commercial flow cytometers, we have to take a closer look at the innards of ADCs.

Analog-to-Digital Converters (ADCs) (and Digital-to-Analog Converters (DACs))

In the pulse height analyzer era, raising the issue of using general purpose computers for flow cytometric data analysis was likely to get the hardware freaks upset about a characteristic of ADCs known as **differential nonlinearity**. There are several different kinds of ADCs; you can see Horowitz and Hill¹⁰⁶, newer books on digital audio²⁴⁶⁷⁻⁸, and the websites of manufacturers such as Analog Devices (www.analog.com) for more thorough descriptions than I am about to give). Some types of ADCs have more differential nonlinearity than others.

Pulse height analyzers incorporated what are known as **Wilkinson** converters, which are a type of **dual slope** ADC. The dual slope ADC was conceived as an improvement on – what else? – the **single slope** ADC. A single slope ADC compares its input voltage with a **reference voltage** produced by a circuit called a **linear ramp generator** (see p. 195); this voltage starts at the low end of the range (say 0 V, or ground, or even a little bit below ground), and increases linearly with time. When a single slope ADC is used for flow cytometry, the **strobe pulse** from the flow cytometer (pp. 191-3) is used to generate a **start convert** signal, a logic pulse which resets the ramp generator, and also resets a **binary counter**, which is a digital circuit capable of counting up to, and storing, a value of $(2^n - 1)$, where n is the number of bits to which the ADC converts. The reset leaves a value of zero in the counter.

When the ramp voltage goes above zero, a comparator turns on, or **enables**, the counter, which starts counting pulses from a **clock** circuit, i.e., an oscillator operating at a precisely controlled frequency, which puts out logic pulses at regular intervals. When the ramp voltage reaches the input voltage, another comparator stops the counter, which holds the count it reached. Since the ramp voltage is linear with time, and the clock operates at a constant frequency, the number stored in the counter at this point is proportional to the input voltage. The Wilkinson and other double-slope ADCs work by charging a capacitor to the input voltage, and then discharging it with a linear current drain, counting the time it takes for the voltage to reach zero, but it's the same basic principle.

The absolute accuracy of single- and dual-slope ADCs is not all that good, but, because of the linear characteristics of the reference voltage generator or current drain, they have very good **differential linearity**, that is, the voltage differences between any two adjacent channels are very nearly equal. These ADCs thus produce relatively smooth histograms. On the minus side, single-slope ADCs also incorporate a lot of parts, which tends to make them expensive, and they're fairly slow, with speed determined by the clock frequency. The **conversion time**, i.e., the time required to generate valid digital output data following a start convert pulse, is also a function of the input voltage. For example, if the ADC uses a 10 MHz clock, and the input signal corre-

sponds to an output value of 127, the conversion takes some small setup time – a microsecond or so – plus 127 clock ticks, or 12.7 μ s. If the digital signal value is 10, everything's done in the setup time plus 1 μ sec. If it's 255, the conversion time is the setup time plus 25.5 μ s, and so on. Even with a 100 MHz clock, converting a value of 1023 takes 10.23 μ s plus the setup time.

Most of the ADCs available these days are faster, cheaper, and less complex than Wilkinson converters. Many of these are the so-called **successive approximation** type; they generate comparison voltages using a **digital-to-analog converter (DAC)**, which takes a digital input and generates an analog (voltage or current) output. A classical DAC incorporates a **voltage divider** made up of a string of resistors in series, so that points between any two resistors will be at 1/2, 1/4, 1/8, and successive fractions of the output voltage. This divider is the primary source of differential nonlinearity in a successive approximation ADC.

The reason for this is that, in order to get the voltage outputs needed from the divider, the resistors must have values R , $R/2$, $R/4$, etc. Resistors are real-world components. Most of the resistors used in electronic circuits are only specified to be within 5% or 10% of their stated resistance value; for slightly more money, you can get resistors made to a tolerance of 1%, and for substantially more than that, parts made to a tolerance of .01% or better. Let's suppose the smallest resistor in the voltage divider of a successive approximation ADC, i.e., the resistor in the "1's" bit, has a precision of 1%. In order to maintain the same voltage accuracy (not precision) in the resistor in the "128's" bit, the resistor has to be precise to better than .01%. That was a tough order when converter modules were fabricated from individual components. In the modern era of large scale integrated circuits, resistors in the DAC incorporated into a typical successive approximation ADC are "trimmed" to the necessary precision during fabrication of the circuit, usually by laser ablation of resistive material. The quality of the trim determines the differential linearity of the converter. In the interest of full disclosure, I will mention that DACs can now use calibrated current sources instead of calibrated resistors, and that many use combinations of the two; the state of the art of circuit fabrication allows differential linearity to be well controlled either way.

The differential linearity spec quoted for most successive approximation ADCs is 1/2 LSB, which means that, in an 8-bit ADC, the "widths" of channels, nominally 39.1mV, could vary between about 20 mV and about 60 mV. The differential nonlinearities in an 8-bit single- or dual-slope ADC arise from the slight deviations from true linear response in the ramp generator due to causes such as temperature fluctuations, and are very small; the channel widths generally vary by less than 1 mV. If we tried to generate histograms from the same data using an 8-bit Wilkinson ADC and an 8-bit successive approximation ADC, the latter would be apt to produce a pretty ragged looking display, because what would come out of a Wilkinson ADC as, say,

60 cells in channel 41 and 60 cells in channel 42 could come out of a successive approximation ADC as 90 cells in 41 and 30 in 42, or 30 and 90, or anywhere in between. Not what you'd like to send in to the journals.

As Tom Sharpless pointed out in the original Big Yellow Flow Cytometry Book¹²⁸, there's a quick fix for this problem. What we do is get a 12-bit successive approximation ADC, with output range 0-4095, and 2.5 mV between channels. This will be more expensive than an 8-bit ADC, in part because it has to have a much better trimmed resistive divider. If we use the upper (most significant) 8 bits' output of this converter, we can reduce the apparent differential nonlinearity. The nonlinearity will still be 1/2 LSB, but we are now taking the converter's 16's bit as our 1's bit, and thus, from our point of view, differential nonlinearity is reduced by a factor of 16.

Without doing statistical analyses on the data, you can't tell the difference between data taken with an 8-bit Wilkinson ADC and data taken to 8 bits' resolution using a 12-bit successive approximation ADC, and it's hard to tell the difference between data from the 8-bit Wilkinson ADC and data taken to 8 bits' resolution using a 10-bit successive approximation ADC. Successive approximation converters, which have a fixed conversion time, are also considerably faster than single- and dual-slope ADCs, another point in favor of the successive approximation devices.

When ADCs were more expensive, people would **multiplex** them, switching several different analog signals in succession to the input of a single ADC, and storing the output data in sequence. Some of the data acquisition boards available for personal computers, and used with older flow cytometers, still follow this practice. Multiplexing several input signals into one ADC obviously takes longer than using several ADCs in parallel, one for each signal, because you have to convert the first signal, store it, convert the second signal, store that, etc. By the time the previous edition of this book appeared, 16-bit successive approximation ADCs with conversion times of 10 μ s were available for less than \$25, and 12-bit converters were even cheaper, leading most of the cytometer manufacturers toward using a separate ADC for each data channel. This made it relatively simple to keep total conversion times for all channels (and few instruments had more than five) under about 20 μ s, theoretically allowing collection of data on as many as 50,000 events/s.

The recent development of ADCs has, to a large extent, been driven by the needs of the digital audio and telecommunications markets. Newer ADCs incorporate technologies other than successive approximation, and achieve high-resolution conversion with better differential linearity.

The fastest converters, appropriately called **flash ADCs**, incorporate a voltage divider and a large number of comparators [$(2^n - 1)$ comparators for an n -bit conversion], but, unlike the ADCs we have discussed previously, convert in a single step. **Digital storage oscilloscopes**, which are rapidly replacing the analog variety, use 8-bit or 10-bit flash converters with conversion rates of several GHz. While pure

flash converters are typically relatively low-resolution devices (10 bits or less), hybrids that incorporate several flash stages and other circuitry have achieved higher (14-16 bits) resolution. The fastest 16-bit successive approximation converter requires 1 μ s for a conversion; hybrid 16-bit converters can convert at 5-10 MHz rates (100-200 ns conversion time). BD Biosciences' FACSDiVa digital pulse processing electronics use 14-bit, 10 MHz converters, which, when the system was designed, were the fastest available devices with that resolution. Analog Devices, the manufacturer of those ADCs, now offers 14-bit, 105 MHz converters.

Digital Pulse Processing and DSP Chips

As long as the ADCs used in flow cytometry only have to digitize held analog signals such as those stored in peak detectors and integrators, conversion times of several μ s are perfectly acceptable. Once we start doing digital pulse processing, conversion rates must be higher.

We have noted (p. 205) that, according to the Nyquist theorem, faithful reconstruction of a waveform in which the highest frequency component is at a frequency f requires that we sample and digitize the signal at a rate of at least $2f$. Now, let us imagine that the pulses we measure in flow cytometry are not Gaussian, but are, instead, the similarly shaped upper halves of sine waves. A 10 μ s pulse would then be half of a sine wave with a period of 20 μ s, and a corresponding frequency of 50 kHz. So, by the Nyquist theorem, we ought to be able to reconstruct the waveform if we sample at a rate of 100 kHz, meaning one sample every 10 μ s. At this point, we should be questioning the validity of the conceptual "model" we have just made of a pulse.

The Nyquist theorem is perfectly valid, provided we are prepared to use Fourier analysis to reconstruct the waveform; this, unfortunately, is likely to require more computation, and more time, than we are likely to have available if we are trying to digitize and process eight or more channels worth of pulse trains coming out of a flow cytometer. If we want to get decent pulse information from a fast and relatively simple computational process, we have to take more samples of our pulses than the Nyquist theorem would suggest. Figure 4-59 may be of help.

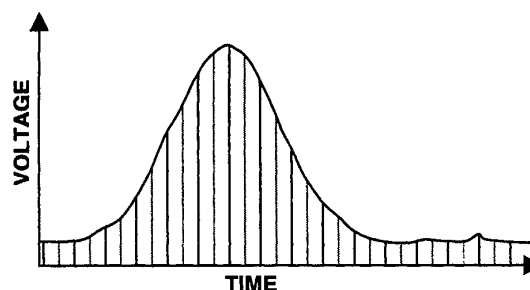


Figure 4-59. Digital pulse processing: "slicing" a slightly noisy Gaussian pulse with a baseline.

The trace in the figure represents a slightly noisy Gaussian pulse, with a baseline with DC and AC components. The vertical lines represent samples, or "slices," taken at regular intervals; I have shown alternate samples in black and gray.

At first glance, it appears that we might get a more than adequate characterization of the pulse by using every other sample, i.e., either all of the black lines or all of the gray lines. If you "connected the dots" at the point at which each set of lines intersects the trace, you'd get two curves that would follow the trace pretty closely. However, what we need to do by way of pulse processing is find the height, the width, and the area, or integral, of the pulse, and, if we start with the height, we see that the gray samples and the black samples aren't equivalent. As it happens, the sample that hits the trace closest to the true peak value is one of the black samples; the values represented by gray samples on either side of it appear to be at least 4 or 5 per cent lower than the peak value. On the other hand, if we're trying to find the pulse width, or where the pulse "begins" and "ends" by rising above the baseline, the gray samples may give us a more accurate idea than the black ones; as was the case with the peak height estimate, this is no more than the "luck of the draw." But we do get the sense that, if we tried to estimate the area or integral of the pulse, by taking the sum of a number of black or gray samples nearest the peak value, we'd end up with better agreement between the values from the black samples and the values from the gray samples than we had in our peak height and pulse width estimates.

Calculation confirms what intuition suggests; Phil Stein, Nancy Perlmutter, and I spent a couple of days working up a model, in a Microsoft Excel spreadsheet, that simulates digitization of Gaussian pulses with various random noise levels. Since we had simulated the pulses, we knew their true areas, peak heights, and widths; we then examined the accuracy and precision of 12-, 14-, 16-, and 18-bit measurements of peak height, width, and area, and compared measurements obtained by examining 8, 16, or 32 samples of a pulse. We found that area could be estimated to better than 1% using only 8 samples of the pulse; taking 16 samples increased precision considerably, while taking 32 samples yielded only a small additional improvement. As was expected, peak height and width measurement were less precise, although values were within a few percentage points of the "true" value; taking more samples substantially reduced the measurement error, again as expected.

These days, digital pulse processing is typically accomplished using a **Digital Signal Processor**, or **DSP Chip**. A DSP chip is a digital computer designed specifically for signal processing; it typically has a smaller instruction set than a general-purpose microprocessor, with emphasis on fast fixed-and/or floating-point arithmetic operations. Unlike general-purpose computers, a DSP chip is likely to have separate memories for program and data storage, and facilities to allow multiple computers and multiple analog-to-digital converters and other data sources to be interconnected. In

particular, a DSP chip can very rapidly process a data stream, keeping track of a **running sum** of any given number of data values. Thus, for example, the chip could output a series of numbers representing the sum of input data points 1 through 8, the sum of data points 2 through 9, the sum of data points 3 through 10, etc. This would be an 8-point **unweighted** running sum; it would also be possible to multiply different data points in the input by different coefficients, producing a **weighted** running sum. In the case where the output stream of the DSP chip is an unweighted, 8-point running sum of input data, we say that an **8-point moving average filter** has been applied to the data.

Because each point in the output data stream of a moving average filter represents contributions from multiple points in the input data stream, the maximum change in values between two consecutive points in the output stream will be less than the maximum change in values between two consecutive points in the input stream. The highest frequency information in both the input and output streams is predominantly represented by the changes in values between consecutive points; the high frequency information in the output stream is therefore attenuated in comparison to the high frequency information in the input stream. Thus, the moving average filter is a low pass filter, and, like an analog low pass filter, behaves as an integrator.

When we try to capture the integral of a pulse using analog or hybrid electronic integrators, we have to select a trigger signal that will generate appropriately timed signals for the electronics that will hold the signal values. In digital pulse processing, the input signal streams are sampled at a relatively high rate, and the digital values are kept in a **buffer memory** until pulses have been identified and their areas and/or other characteristics have been determined. When we use a DSP chip to compute a running sum, we can use the points at which this sum reaches local maxima to represent the approximate centers of pulses, and the values of the sum at those maxima to represent the areas of those pulses. With a fast DSP chip and enough buffer memory, it is possible to identify a pulse in a signal from any detector at any of several observation points, and, if necessary, to work backward through the input data stream in the buffer to find and characterize low-level signals. The flexibility of this digital "triggering after the fact" is advantageous.

In principle, we could use DSP to eliminate analog baseline restoration by monitoring the baseline signal digitally and subtracting an averaged value from all signal values; this is the digital version of the analog computation performed by an analog baseline restorer. However, unless the DC baseline signal is very small, this approach may lose dynamic range. In a fluorescence channel, DC baseline is low, e.g., 1/1,000 of full scale, corresponding to a digital value of 65 on a 16-bit linear scale. If data are to be converted to an 8-bit or even a 10-bit logarithmic scale, there is ample room in the top channel of the scale, which encompasses many linear values (see p. 207). However, in a forward scatter channel, DC baseline is likely to be much higher, on the order of half

of full scale, corresponding to a digital value of 32,767. Subtracting this value from the raw digitized signal values leaves only 32,767 channels available to describe pulse amplitudes, effectively turning the 16-bit ADC into a 15-bit ADC. This is not the direction in which we would like to go; we already know (p. 207 again) that we need 20-bit data to eliminate the picket fences that occasionally make bad neighbors in our cozy cottage industry.

For digital pulse processing in anything approaching a conventional flow cytometer, ADC conversion rates, as previously mentioned, must be at least 1 MHz; high-speed sorting and data processing demand higher conversion rates. As noted (p. 209), BD Biosciences elected to use a 14-bit (16,384-channel), 10 MHz ADC in their FACSDiVa system digital pulse processing system, because a higher-resolution part with the same conversion rate was not available. For the record, as of late 2002, you can get 14-bit converters with conversion rates above 100 MHz, but the fastest 16-bit converter I have seen advertised is a 5 MHz part, and the fastest 18-bit converter, introduced in mid-2002, has an 800 kHz conversion rate. The electronics required for high-speed digital pulse processing are way too complex to make subranging (as used in the Beckman Coulter XL, etc.; see p. 207) a viable option, and there are some tradeoffs involved in using 14-bit converters, the most obvious of which is a "picket fence" effect at the low end of the logarithmic scale.

A brief digression on audio recording may help us understand the tradeoffs currently involved in digital pulse processing. There is, in fact, a distinct possibility that slower, high-precision, high-resolution cytometric data analysis without digital pulse processing could make effective use of the **oversampling delta-sigma** ADCs now used for digital audio recording²⁴⁶⁸⁻⁹. These converters continuously sample signals at rates of a few MHz, and put out 24-bit data at rates as high as 192 kHz. You can find an easily understood audiovisual introduction, which also covers some of the basics of ADCs and sampling presented in previous pages, if you download the "Short Course in Digital Audio" from Syntrillium Software, at (<http://support.syntrillium.com/tutorials.html>).

The Screwy Decibel System

Modern audio components are likely to have a specified **signal-to-noise (S/N) ratio** of better than 80 **decibels (dB)**. Decibels (the **bel** derives from, and honors, Alexander Graham Bell) measure the ratio of signal amplitudes on a logarithmic scale, with a few tricks involved. When you're listening to speakers, or any other sound sources, your hearing is responding to the acoustic **power** delivered to your ears. The power that drives a loudspeaker with an impedance of Z ohms comes from an amplifier that puts a current I through the speaker, producing a voltage difference E across the load. Ohm's law tells us that $E = IZ$; the power, P , is EI , or E^2/Z .

The ratio, in decibels, of two **power** levels, P_1 and P_2 , is

$$\text{dB} = 10 \log_{10} P_1/P_2.$$

However, we have already seen that power varies as the square of voltage; therefore, the ratio in decibels of the two **voltage** levels, E_1 and E_2 , corresponding to the power levels P_1 and P_2 , is

$$\text{dB} = 20 \log_{10} E_1/E_2.$$

In our discussions of signal processing, we are generally interested in the ratios of signal voltages, or signal and noise voltages, not in powers, so we use the second formula.

If we say that the (voltage) **signal-to-noise ratio (S/N)** of a circuit or system is 80 dB, it means that the ratio of (signal voltage amplitude)/(noise voltage amplitude) is 10,000:1. If the output is 10 V, the noise level is 1 mV. To those of us who grew up in the era of vinyl records and analog tape recording, a S/N of 80 dB looks pretty good, but, having previously looked at dynamic range in cytometry, we are now (I hope) aware that an 80 dB range just barely encompasses 4 decades.

I might add that the frequency response spec of 20-20,000 Hz $\pm$ 0.5 dB, which also sounds pretty good when it is hung on audio equipment, allows for about 6 percent variation in output level over the frequency range, and that $\pm$ 0.5 dB is about as close to ideal as the best log amp responses get. So there is plenty of motivation for moving toward digital signal processing and high-resolution analog-to-digital conversion in cytometry.

A stereo audio CD contains 16-bit data for each channel, which is played back at the rate of 44,100 data points per channel per second; it obviously must be recorded using a sampling rate of at least 44.1 kHz and 16-bit ADCs. However, most professional audio recording is now done at higher sampling rates, using ADCs with more bits, for a number of reasons.

One is that music, or at least most of the music I listen to, has a fairly wide dynamic range. The **SPL** or **Sound Pressure Level** scale used to describe audible sound defines the level of 0 dB SPL as the intensity of a 1 kHz tone barely audible to the normal human ear; at a rock concert, the level frequently reaches 120 dB SPL, which, before there were rock concerts, was observed to make most people uncomfortable, and thus came to be known as the **threshold of pain**. The SPL scale, incidentally, describes pressure levels rather than audio power levels; two sounds differing by a factor of 10 in amplitude are thus 20 dB apart, rather than 10 dB apart, on this scale. Microphones and their associated amplifiers convert pressures into voltages, which, as noted at left, also differ by 20 dB, rather than 10 dB, for each factor of 10 difference in amplitude.

Classical music covers a range from 40 dB SPL (*ppp*, very soft) to 100 dB SPL (*fff*, very loud). When recording it, we need to keep the *fff* signal at or below the top of the range of our ADC to avoid **clipping**, which would distort the sound when it was played back (this isn't usually a problem in rock music, in which clipping is frequently introduced intentionally). But back to Symphony Hall, where we have adjusted the gain of our microphone preamplifiers so that the loudest

part of the piece, at 100 dB SPL, comes out at the high end of a 16-bit converter, i.e., channel 65,535. The softest (*ppp*) passages, at 40 dB SPL, with 1/1,000 the amplitude, will therefore end up at channel 65, and much of the music will come out somewhere in between, probably at channel numbers below 1,000.

Now, suppose we want to do our audio mixing and editing digitally. This is going to involve some multiplications and subtractions, and, the more of them we do, the more likely we are to end up with some "picket fence" effects at the low end of the digital scale in the output data. There are a number of fixes for this. One of the things we can do is use "comparers" (compressor-expanders) in the electronic circuitry, transforming amplitudes according to a nonlinear curve so that sounds at the low and middle ends of the range get converted to higher channel values than they would otherwise.

The same trick has long been used in cytometry; the nonlinear curve with which we are already familiar is a logarithmic curve, and the input side of a compander acts pretty much like a log amp. And, yes, it may only be accurate to 6 percent or so, but that is within the ± 0.5 dB tolerance usually specified for high quality audio.

We could also add a little random noise to the signal, getting rid of the "picket fences," albeit also degrading signal quality slightly; this maneuver is also used in cytometry, as when data have been converted from relatively low resolution (10 bits or less) log scales to linear and back during after-the-fact software compensation.

The diehard digital audio purists, and most of the professionals in the recording business, even those who record rock, won't have any of this. They can now get 24-bit converters that sample at 192 kHz, giving them enough dynamic range to do all of the editing and add all of the effects they need, and then interpolate and round off to convert the finished product to a 16-bit, 44.1 kHz sample rate CD-quality audio file. Or they can leave it as a 24-bit, 192 kHz DVD-quality audio file.

You can now buy an 8- or 10-channel, 24-bit, 192 kHz digital audio recording accessory for your computer for under \$2,000; the 24-bit converters used in such systems cost less than \$10 apiece. Table 4-5 (p. 206) tells us that 1 LSB in a 24-bit converter with a 10 V full scale input range amounts to 596 nV; since the noise level in even the most carefully designed electronics is likely to be tens of microvolts, we can guess that we won't capture a real 24-bit signal. However, we are likely to be able to get 20 bits' worth of useful information, and that should eliminate the picket fence effect and the inaccuracy at the low end of the standard 4-decade scale. I speculate that one could feed the outputs of peak detectors and integrators into such a gadget, and get good data from cells at rates of a few thousand/s. That would be perfectly adequate for high-sensitivity, high precision fluorescence flow cytometry, which is likely to require relatively slow flow rates and thus not process tens of thousands of cells per second, anyway. I've already got the

audio recorder; playing with it (for audio as well as for flow data) will have to wait until I finish writing this book. At worst, it should be useful for the sequel to my audio opus, *Songs for the Jaundiced Ear*²⁴⁷⁰.

Pulse Slicing: Déjà Vu All Over Again

Moving from CD players back toward CD antigens, it seems inevitable that the detector signals in most flow cytometers will eventually be handled by DSP techniques, eliminating the analog and hybrid baseline restorers, threshold sensing and front end electronics, integrators, peak detectors, and pulse width measurement circuits on which we have just spent so much time. A fast enough DSP pulse processing system would even be able to sample light source power output and perform digital compensation for noise fluctuations, almost certainly performing better than the analog source noise compensation circuit I described on pp. 147-8. That's the future of digital pulse processing in cytometry; however, we should not forget that it also has a distinguished past.

Slit-scanning flow cytometers^{171-4, 690-7, 1111-2, 1166-72} were pioneered by Leon Wheelless and his colleagues beginning in the 1970's, and investigated by a number of others then and since. A slit-scanning cytometer uses an extremely narrow region of illumination; beam heights under 5 μm are the norm. The resulting signals from cells can reveal morphologic information in the form of intensity variations along the length of a pulse. The original work in this field used relatively primitive computers for signal analysis, and was done at a time when it was only feasible to use 8- and 10-bit ADCs for data acquisition. This limited the dynamic range of data collection; however, many of the early applications of slit-scanning were to analysis of fairly strong fluorescence, low dynamic range linear signals from cells or chromosomes stained with nucleic acid dyes, and investigators' primary motivation for studying the technique was the desire to pursue applications of pulse shape information.

Other experimental systems^{1173-5, 2473-4} digitized detector signals at rates up to tens of MHz in order to collect enough points (in some cases, over 100) to provide detailed information on the shapes of pulses as well as the more conventional measures of pulse height, width, and area. The RATCOM Personal Cytometer, the predecessor of today's NPE instruments (see Chapter 8), was probably the first commercial instrument to implement digital pulse processing, using 8-bit ADCs, around 1990; the objective in this case was to provide a highly accurate integral for DNA analysis, rather than high dynamic range. Earlier attempts at digital pulse processing used specially fabricated **application-specific integrated circuits (ASICs)** to perform the function of DSP chips, which were not yet available with the required speeds and processing capabilities; newer digital pulse processing systems may use both DSP chips and ASICs.

Those of us who design flow cytometers today typically want to be able to process sixteen or more detector signals, captured in three or more illumination regions, from tens of

thousands of cells per second. We expect that the data will have a large dynamic range, that is, at least four decades, and that there will be substantial spectral overlap between signals, requiring fast software compensation and conversion between linear and logarithmic scales, and we want to do all of the processing in a short enough time to permit us to make sort decisions before the cells of interest are lost to us.

Getting pulse shape information is not a major concern, at least for the time being, and we can usually settle for relatively crude indicators of pulse width and height, which will be used primarily for discriminating between single cells and multiplets. What we care most about is getting accurate integral values representative of the amounts of fluorescence measured from each cell in each spectral region.

In theory, taking a large number of digital samples of each pulse should provide a very large dynamic range for the value of the integral. If a pulse is at least 6.4 μ s long, an 8-bit converter sampling at 40 MHz (1 sample every 25 ns), will acquire at least 256 samples of the pulse, each with a value between 0 and 255. The integral, which is the sum of the 256 values, can theoretically range between 0 and 65,280; this appears to yield the same dynamic range as a 16-bit ADC. However, we are still dealing with an 8-bit converter, which, if it operates over a range of 0-10 V, has channels spaced 39 mV apart (Table 4-5, p.206). Unless the pulse amplitude exceeds 39 mV at some point, the converter output will remain at 0, and the integral will remain 0. The converter won't even notice 10 mV or 1 mV pulses. We're stuck in the upper 2 1/2 decades of a 4-decade dynamic range.

If we move to a 12-bit converter, with channels spaced about 2.4 mV apart, we won't have a problem with 10 mV pulses, or even 3 mV pulses; this gives us a 3 1/2 decade range, although the 1 mV pulses at the bottom of the fourth decade remain invisible. Since the 12-bit converter output ranges between 0 and 4,095, we could, in theory, get away with 16 samples of a pulse, dropping the sampling rate from 40 MHz to 2.5 MHz. With a 14-bit converter, output ranges between 0 and 16,383, and the channels are 600 μ V apart. If we kept the sampling rate at 2.5 MHz, we'd get 16 samples of a 6.4 μ s pulse, the values of the integral could range between 0 and 262,128, and the converter would respond to a 1 mV pulse. And, if we sampled at 10 MHz with the 14-bit ADC, as is done in BD's FACSDiVa electronics, we could get 16 slices of a 1.6 μ s pulse, again giving us a maximum possible value of 262,128 for the integral. However, if we think about it, we will quickly find that we won't, and don't want to, ever get actual values close to that maximum. Here's why.

We expect to be looking at pulses that are more or less Gaussian in shape. In order to get a pulse integral value of 262,128, we have to record a value of 16,383 for each of the 16 slices of the pulse in question. Such a pulse would be rectangular, not Gaussian; if we encountered a real pulse that gave us this integral value, we could be sure that we were not capturing an accurate value of the integral. We

would, presumably, be looking at the lower portion of a Gaussian pulse with amplitude exceeding the maximum acceptable input value for the ADC, and our signal would be **clipped**. Clipping (p. 211) is a well-known phenomenon in audio recording; if you overdrive preamplifiers and/or amplifiers, or put too large a signal into a digital recording device, you turn sine waves into roughly square waves, producing audible distortion.

In audio applications, clipping is often done intentionally, without penalty, to produce a distinctive sound effect; in digital pulse processing and football, there are penalties. On our playing field, the only way to avoid them is to keep the pulse amplitude sufficiently low so that no more than one point – the peak of the Gaussian pulse – produces the maximum output from the ADC. The other samples of the pulse will have lower values; we can easily calculate what they will be using a table of the Gaussian or normal distribution, and this will tell us what the practical attainable maximum value for the pulse integral should be.

In Chapter 5, we will delve into the anatomy of the Gaussian distribution in some detail. For the present, we will simply consider an integral of the Gaussian pulse computed by taking 16 14-bit slices evenly spaced over an interval in which theory tells us we will find a little more than 99 percent of the area of the pulse. The maximum value for the 16-slice integral, obtained when the true peak value is 16,383 (corresponding to a 10 V pulse if the full scale range is 0-10 V), is about 102,500. If 10 V represents the top of a 4-decade dynamic range, 1 mV represents the bottom; the digital value for the 16-slice integral of a 1 mV pulse is 10, that for the integral of a 10 mV pulse is 102, etc.

In theory, a digital value of 1 should correspond to the integral of a 100 nV pulse, allowing an integral of 16 14-bit slices to represent data spanning a five decade dynamic range. However, Table 4-5 (p. 206) tells us that the 14-bit converter will not produce an output above 0 unless the input signal level exceeds 610 nV. That means that there is little point in attempting to display integral data from a system such as the FACSDiVa on a five-decade logarithmic scale. In fact, since a four-decade scale would encompass digital values between 10 and 102,500, and since we know (p. 207) that we need 20-bit integral values (ranging between 104 and 1,048,575) to generate a smooth four-decade logarithmic scale occupying 1,024 channels, we should not be surprised when we see some "picket fence" effects at the lower end of the FACSDiVa's displays. But how much does it matter?

In Defense of de Fence

I feel obliged to make the same apologies for the FACSDiVa's picket fences as I made on p. 207 for my own, which result from my use of a 16-bit ADC, rather than a 20-bit converter, to convert held peak signal values. As already noted, digital pulse processing makes it possible to use almost any combination of signals from any illuminating beam(s) for triggering, and provides digital integral values

covering a sufficiently large range to eliminate the need for logarithmic amplifiers and to permit fluorescence compensation to be done rapidly enough (by yet another DSP chip) for sort decisions to be based on compensated data values. That's the good news; the bad news is that, if we want to get 16-slice integrals from tens of thousands of cells per second, we're pretty much stuck with using 14-bit converters, and picket fences appear in the bottom half of the bottom decade on a four-decade logarithmic scale.

If we define that scale in terms of numbers of molecules of some substance measured in or on a cell, we can cover the range from 100 molecules to 1,000,000 molecules. Most of the CD antigens we detect on cells from the blood and immune system do not reach the level of 1,000,000 molecules per cell, and most of the flow cytometers we use to detect these antigens cannot reliably detect 100 molecules of antigen above background. So, we might want to regard the picket fences in our displays as if they bore the sign "beware of the data," and not try to play with what is behind them. I'll have more to say about this later (pp. 244 and 562-3).

Digitization: Tying It All Together

A few years back, a good friend of mine sent me a book called *The 85 Ways to Tie a Tie*²⁴⁷⁵. I knew three of them, and have added a fourth to my repertoire because it is useful for woolen and woven neckties. I don't think there are 85 ways to get high-resolution digital data out of a flow cytometer, but there are a couple of approaches I haven't mentioned, and now is the time to summarize existing methods.

As noted above, digital pulse processing is unique in offering highly flexible triggering schemes and the potential for really effective source noise compensation; if 18-bit, 20 MHz ADC's were available, there would be almost no downside, other than the complexity of the circuitry involved.

Digitizing held peak and/or integral values to 20 bits' precision, as is done in the Beckman Coulter EPICS XL, requires a high level of performance from peak detectors and integrators, and also, in its present form, introduces the complexity required for subranging analog-to-digital conversion. The electronics can be made fast enough to deal with tens of thousands of cells per second; although Beckman Coulter's current sorters still use log amps and hardware compensation, it seems likely that a switch will be made to high-resolution digitization to keep the product line competitive. There is a possibility that 24-bit ADCs designed for audio recording could be used for high-resolution capture of flow cytometric data from a few thousand cells per second; this approach might also permit the use of somewhat lower-performance peak detectors and integrators.

DakoCytomation's MoFlo line of high-speed sorters uses log amps, and digitizes log scale held peak and integral signal values to 16 bits' precision. Calibration curves for the log amps are stored in memory, and the digitized log data are converted to a linear scale to allow real-time software compensation to be accomplished by a DSP chip. The 16-bit

digitization of the original log data provides enough resolution to allow compensated data to be displayed on a four-decade log scale with no picket fence effects.

The manufacturers' marketing and sales people are a good deal more partisan about the virtues of their respective methodologies than are the people who design and build the hardware and develop the firmware and software. At the present state of the art, while all of the approaches to high-resolution digital flow cytometric data processing involve tradeoffs, any of them is preferable to doing things the old-fashioned way with log amps and hardware compensation, especially when more than three fluorescence measurements are made of each cell.

4.10 PERFORMANCE: PRECISION, SENSITIVITY, AND ACCURACY

The performance of flow cytometers is generally discussed in terms of **precision**, **sensitivity**, and **accuracy**, often in that order. It is, perhaps not coincidentally, easiest to characterize precision, and hardest to characterize accuracy.

Precision describes the extent to which identical values are obtained from measurements of identical particles; the measure of precision most commonly used is the **coefficient of variation** or **CV**.

The definition of **sensitivity** has changed somewhat as cytometry has evolved; the measures originally proposed were the minimum size of particles or amount of a substance detectable above background by the apparatus, inferred from measurements of scatter or fluorescence signal intensities. Current practice aims to quantify the extent to which populations with very small signal intensities can be resolved from one another.

Accuracy describes the degree to which the measurement results produced by the flow cytometer conform to "true" values. Until recently, considerably more attention was paid to precision and to sensitivity than to accuracy, with good reason; until precision and sensitivity have been established, it is difficult to define a framework in which to characterize accuracy.

Precision; Coefficient of Variation (CV)

As was noted in the introductory discussion on pp. 18-22, the measurement precision of a cytometer is routinely characterized by accumulating a distribution of measured values of fluorescence or light scattering intensities from "nearly identical" particles, and computing the **coefficient of variation (CV)**, which, expressed as a percentage, is 100 times the **standard deviation (S.D.)** divided by the **arithmetic mean**, or **average**.

The particles may be artificial, e.g., fluorescent polystyrene microspheres, or biological, e.g., fixed, stained cell nuclei. Among cellular parameters measurable by flow cytometry, only DNA content is so precisely regulated as to vary by less than 2% from cell to cell in homogeneous, nondividing populations. Nuclei from quiescent or largely quiescent cell populations such as avian erythrocytes or mammalian pe-

peripheral blood lymphocytes, stained with appropriate fluorochromes, can therefore be used to test the precision of fluorescence measurements. When this is done, the CV is calculated for the G_0/G_1 or diploid peak of the DNA fluorescence distribution. Because inclusion of early S phase cells may affect calculations done by computation of the mean and variance, it is usually convenient to estimate the standard deviation as one-half the width of the peak at 60% of its maximum height or as $(1/2.36)$ the **full width at half maximum height (FWHM)**.

When brightly stained fluorescent beads and cell nuclei stained for DNA are used to assess instrument precision, enough photons are typically collected from specimens so that photon statistics are not a major contributor to the variance of fluorescence distributions. In most cases, i.e., for most flow cytometric measurements, CVs below 5% are acceptable; for DNA measurements, CVs below 3% are preferable. Most modern instruments can achieve CVs between 1% and 2% in scatter and fluorescence measurements of beads.

Although beads are now more commonly used than nuclei to measure instrument precision, the intrinsic CV of nuclei is likely to be lower; the most precise instruments, using a highly DNA-selective stain such as DAPI or a mixture of mithramycin and ethidium, can achieve CVs on the order of 0.5 percent. The stains just mentioned, respectively, require UV and blue-violet excitation; DNA fluorochromes excitable at the 488 nm wavelength most commonly used in flow cytometers are less DNA-selective, and require that nuclei be treated with RNase to eliminate fluorescence variations due to RNA staining.

Precision may be decreased by fluctuations in light source output, by poor alignment of optics, and by disturbances of fluid flow. If the optics of a flow cytometer are adjustable, it is customary to make such adjustments as are needed to maximize the signal from the fluorescent particles used for precision determination before making a definitive estimate of precision; brightly stained beads with a low intrinsic CV are often referred to as **alignment beads** because they are used for this purpose. In modern instruments that normally do not need alignment, fluid flow problems should be the primary suspect when a decrease in precision (i.e., an increase in CV) is noted, and the flow system should be checked and flushed and cleaned, if necessary, before a service call is initiated. A consistent decrease in laser output power points to laser light noise as the source of an otherwise explained loss of precision.

It should be obvious that it is pointless to attempt to determine either the sensitivity of an instrument or its measurement accuracy until it can be demonstrated that performance is up to the manufacturer's specifications in terms of precision.

Sensitivity I: Minimum Detectable Signal

The sensitivity of a flow cytometer may be defined in several different ways. Measurement of what should properly

be called **instrument sensitivity** was originally based on determination of the minimum amount of fluorescent material, or minimum number of photons from a test light source, detectable by the instrument when no cells are present. This was originally done by determining the minimum concentration of dye which, when run through the system, increased the fluorescence detector signal above background, and calculating the number of dye molecules in the observation volume; detection limit experiments were also done by Loken and Herzenberg using cells stained with antibodies bearing both fluorescent and radioactive labels¹⁵⁴. The Stanford and B-D FACS instruments of the mid-1970's were found to be capable of detecting about 3,000 molecules of fluorescein. By the late 1970's, the Block apparatus used for analyses of viruses³⁴ could detect a few dozen molecules of fluorescein; in the 1980's, work by the Los Alamos group culminated in the detection of single molecules of phycoerythrin in solution, using single photon counting techniques on a slow flow system⁶⁶⁰.

In applications of flow cytometry such as immunofluorescence measurement, the paramount issue is generally how many molecules of a label such as fluorescein, attached to a ligand, are detectable on a cell surface; this provides a more practical definition of sensitivity, which, in this context, is more likely to be limited by cellular characteristics, in particular, autofluorescence, than by instrumental factors.

For example, the instrument sensitivity of 3,000 fluorescein molecules mentioned above for the 1975 vintage FACS instruments was not attainable in immunofluorescence measurements of lymphocytes, because the fluorescence of unstained lymphocytes, as then measured with those instruments, was broadly distributed in the range equivalent to the fluorescence of 5,000-10,000 fluorescein molecules. It was therefore necessary to have several times that amount of fluorescein bound to antibody on a cell surface before one could be sure that the cell was stained.

A third measure of sensitivity is useful in evaluating the efficiency of flow cytometer designs. If measurements are made of unstained cells (autofluorescence) and of weakly stained cells, two peaks representing brighter and dimmer cells will appear in the fluorescence distribution. Sequential measurements are made of a bimodal population, starting at low laser power; the ratio of the linear values representing the bright and dim cell peaks is plotted against laser power and power is increased and measurements repeated until the ratio reaches a plateau value; the laser power used is then noted. By this criterion, i.e., the amount of laser power at 488 nm needed to achieve maximum fluorescein fluorescence measurement sensitivity, the B-D FACScan (plateau at 8-10 mW) and Coulter Elite are among the most sensitive apparatus; the stream-in-air sorters (50 mW for the B-D FACStar, up to a few hundred mW for older B-D and Coulter EPICS sorters) are least sensitive. However, by the same criterion, the old arc source B-D FACS Analyzer, using 546 nm light to excite phycoerythrin-labeled antibodies, was approximately ten times as sensitive as any instrument using

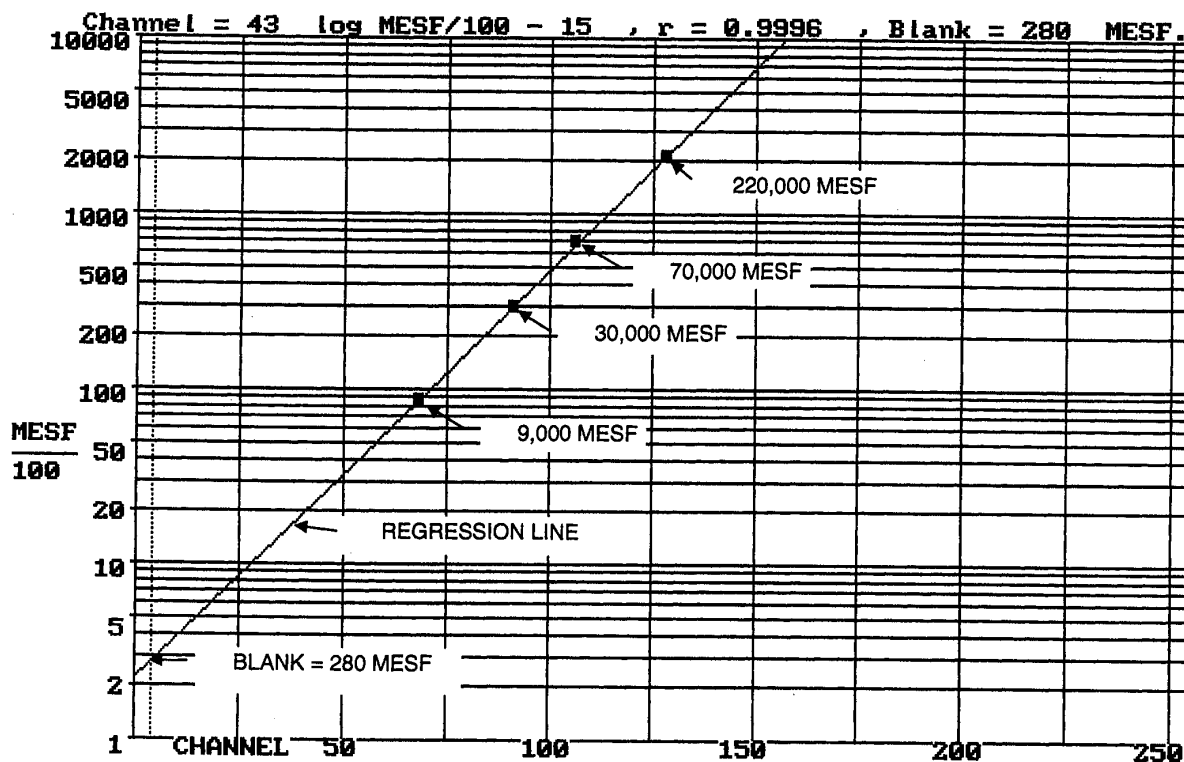


Figure 4-60. MESF threshold sensitivity determination using fluorescein-labeled beads.

488 nm light and fluorescein antibodies, in part because the highly efficient phycoerythrin is optimally excited at 546 nm and in part because autofluorescence, which in most cells is due predominantly to flavins, is much less with 546 nm excitation than with 488 nm excitation.

Sensitivity II: MESF Units

A technique still widely used for determination of flow cytometer sensitivity, originally developed by Dr. Abraham Schwartz¹⁶¹, expresses detection limits in terms of **Molecules of Equivalent Soluble Fluorochrome**, or **MESF Units**. The method is illustrated in Figure 4-60; it uses a mixture of beads, including some labeled with several different, known, numbers of molecules of the fluorescent dye of interest and some undyed blank beads. The expression of fluorescence in terms of equivalent soluble fluorochrome is necessary because the amount of dye on the beads is measured by comparing the bulk fluorescence of a bead suspension and a dye solution; in general, the quantum efficiency of a dye is decreased following its covalent linkage to a bead or other substrate.

The bead mixture is analyzed in the flow cytometer; a histogram of fluorescence from single beads, on a log scale, is acquired using a forward/orthogonal scatter gate. This histogram yields the channel numbers corresponding to the median values of the peaks representing the dyed and blank

beads; the peak locations of the dyed beads and the mean numbers of dye molecules on each dyed bead population are used to derive the equation of a least-squares regression line relating the channel number to the log of the number of MESF units per bead. The resulting equation is then used to calculate the number of MESF units corresponding to the channel number of the blank bead peak. This **MESF threshold** represents the minimum number of molecules of the fluorochrome detectable by the instrument; in essence, it tells you how many molecules the cytometer sees when there aren't any molecules there. Figure 4-60 represents data from beads covalently labeled with fluorescein; the median fluorescence of the blank beads was calculated to be equivalent to the fluorescence of 280 molecules of soluble fluorescein. Surveys of several hundred laboratories¹⁶¹ in 1989-90 found median fluorescein detection thresholds near 900 MESF.

Abe Schwartz later suggested that the MESF threshold determination be modified by using the first channel number above the peak representing the blank beads, rather than the median channel of the peak; the calculation then yields the minimum number of dye molecules detectable above the background of blank particles. This number, obtained using the data shown in Figure 4-60, is 620 MESF.

When unstained cells are added to the bead mixture used for sensitivity determination, the level of cellular autofluorescence can be expressed in MESF units. The 1989-90 mul-

laboratory survey conducted by Schwartz and Fernández-Repollet¹⁶¹ found the lymphocyte autofluorescence peak corresponded to 657 ± 270 MESF of fluorescein; the 98th percentile value was 2552 ± 1748 MESF. Monocytes and granulocytes had somewhat higher autofluorescence. In comparing the value of 657 MESF for lymphocytes to the 5,000 fluorescein equivalent value reported by Loken and Herzenberg in 1975¹⁵⁴, we should note that the older measurement was made over a much wider spectral range; a 530 nm long pass filter was used, rather than the 520 or 525 nm bandpass filters now favored for fluorescein measurement. Since the peak of cellular autofluorescence emission is at around 550-560 nm, the larger bandwidth of the long pass filter will increase the autofluorescence signal to a greater extent than the fluorescein signal. In addition, it is likely that the older measurement was affected by both filter fluorescence and relatively high background light levels.

The major sources of error in sensitivity determination and autofluorescence measurement in MESF units as just described appear to be offsets and deviations from true logarithmic response in log amp circuits. Both of these have a much greater effect on the lowest decade, in which the blank generally lies, than on the higher decades; as a result, the fact that the labeled beads lie very close to the regression line provides little assurance that the line can be accurately extrapolated.

The problems likely to be introduced by log amps may be avoided by measuring the beads at high gain on a linear scale. Start with a low enough gain so you can see the brightest beads, then, increase the gain so that all but the one or two dimmest labeled beads and the blank go off scale. You can then calibrate your channel numbers in MESF units. If, for example, the peak representing the beads with 9,000 MESF is placed at channel 200, then each channel represents 45 MESF. If the peak representing the unlabeled beads is at channel 10, that corresponds to 450 MESF.

I should add that when you're looking for very small signals, e.g., weak immunofluorescence or nucleic acid probes, you're likely to be better off using a linear scale for two reasons. The dynamic range with which you have to deal under those circumstances is usually not very large, so you can keep everything on scale, and not have to deal with inaccuracies that log amps might introduce at the low end. Also, in order to be accurate, you should express the fluorescence intensity of stained cells in terms of their difference from unstained cells, and you can't do the subtraction on a log scale. However, whether you work on a linear or a log scale, you do need to establish the linearity of your instrument, which leads me to digress briefly from sensitivity to accuracy.

Accuracy I: Linearity (and Nonlinearity)

The accuracy of flow cytometers, i.e., the degree to which the measured values approach the true values of the variables measured, is less often discussed than the precision or sensitivity. Numerous instrumental factors can affect accuracy; one of those most often encountered is **nonlinearity**.

One of the more common examples of nonlinearity is the situation in which G_2/S phase cells, which we know have twice the amount of DNA in G_1/G_1 phase cells, appear to have substantially more or less than that. This can be due to substantial offsets in the electronics, in which case dealing with the problem may require anything from turning a knob or trim potentiometer to redesigning the instrument. However, in older instruments, nonlinearity in DNA content measurements was more likely to be caused by the use of pulse height instead of pulse area or integral measurements with beam geometries in which the pulse height was not proportional to the integral.

Perhaps the easiest way to assess the linearity of a fluorescence channel in a flow cytometer is to analyze high-intensity fluorescent plastic beads, measuring the integral or area rather than the height of pulses on a linear scale. There will invariably be a small population in the fluorescence distribution representing doublets; the channel number at which the peak for this population is found should, ideally, be twice the channel number at which the main peak representing the singlet population is found. If the ratio of the doublet and singlet peak channel numbers is more than 2.02, or less than 1.98, it is probably advisable to adjust the instrument. For critical work, the linearity assessment should be repeated at different PMT gains to cover as much of the measurement range as possible.

Accuracy is also often estimated by adding internal standards to specimens, e.g., avian and/or fish erythrocytes, which contain known amounts of DNA, to cell nuclei stained for DNA, or by interspersing controls, such as beads bearing known amounts of fluorescent antibody, among immunofluorescence samples. These and other techniques for measuring and maintaining accuracy are essential for both quantitative measurements and quality control, and will be discussed in Chapter 10.

Sensitivity III: What's All the Noise About?

We are now approaching the end of our long strange trip (drip?) through the hardware of the flow cytometer. Like medical students, who develop the symptoms of each new disease they study, we are now acutely aware that things can go wrong with most parts of the instrument, and suspicious that, according to Murphy's Law, they will. And, like medical students, who are told that, under most circumstances, when they hear hoofbeats, the hooves are more likely horses' than zebras', we have to try to put what we've learned into perspective, and figure out what's likely to go wrong enough to bother us.

Most of what goes wrong enough to seriously compromise the sensitivity and precision of flow cytometric measurements falls under the heading of **noise**; since the early days of flow cytometry, the sources and effects of noise have been studied in detail from both theoretical and experimental viewpoints. In the 1970's, Holm and Cram¹⁶² and McCutcheon and Miller¹²⁴ established that fluorescence measurement precision decreased, i.e., that CV increased, as

lower fluorescence intensities were measured; this was what would be expected on the basis of photon statistics.

Sensitivity IV: More Photons Give Better Precision

In 1983, Pinkel and Steen¹¹⁶³ described a method for determination of sensitivity based on measurement of the CV of fluorescence distributions of pulses from a light-emitting diode (LED) placed in the light collection path. PMT voltage and amplifier gain were adjusted to keep the pulses in the same position as the intensity of the pulses was varied. Using LED pulses instead of test beads allowed sensitivity to be determined with the flow cytometer's laser turned off, eliminating most of the effects of background light; the magnitude of these effects was then estimated by repeating the sensitivity determination with the laser on. CV's obtained from plastic particles can be compared to standard curves obtained using the LED to determine whether photon statistics contribute significantly to variance, in which case precision may be increased by using higher laser powers.

Ubezio and Andreoni¹¹⁶⁴ also investigated the relative contributions of photon statistics and instrumental factors to CV, by analysis of measurements made of propidium-stained nuclei at different laser power levels. Between 10 and 60 mW, the channel number of the diploid peak was exactly proportional to excitation power, i.e., results were not significantly affected by bleaching or saturation. A linear relationship was observed between $(CV)^2$ and the reciprocal of excitation power, with a fixed offset due to instrumental factors (contributions to CV from different sources add in a "root-mean-square" (RMS) manner; the total CV is the square root of the sum of the squares from individual contributions).

Sensitivity V: Background Effects

A 1992 article¹¹⁶⁵ by Steen considered effects of signal strength, background, and detector quantum efficiency on sensitivity and precision; a pulsed LED was again used for sensitivity determination. When only photon statistics and background light noise are considered, maintaining instrument fluorescence sensitivity in the face of increasing background noise requires increasing the excitation intensity. When background noise is significant, sensitivity increases with the square root of excitation intensity; cutting the detection limit in half requires a fourfold increase in power. When noise is negligible, the same increase in sensitivity requires only a twofold increase in power. The increased power needed to maintain sensitivity as noise increases has an effect that, at first, seems paradoxical; given two instruments with equivalent sensitivity, the precision is higher in the instrument with the higher background.

This is so because the effect of background on precision diminishes at increasing signal levels. At the detection limit, the signal-to-noise ratio, i.e., the ratio of the number of photons coming from the sample particle to the number coming from the background, is 1. The more photons come from the background, the more are needed from the particle in

order to detect it; increasing the excitation power gets more photons. All other things being equal, however, CV decreases, and precision increases, as the number of photons collected from the sample particle increases.

Near the detection limit, CV is near 100% for any instrument; well above the limit, CV is lower, and precision is higher, for the instrument with higher excitation power. This is not to say that you will improve performance of your instrument by taking the cover off and running it at the beach on a sunny day; decreasing the background while keeping excitation power constant will increase sensitivity and also improve precision at least slightly. Also bear in mind that virtually all dyes are susceptible to photodamage at some power level, and when you get past that level (which can be as low as 10 mW for a label such as PerCP), increasing excitation power only makes things worse.

A concrete illustration of the effect of background can be obtained from sensitivity values for two cytometers. Steen¹¹⁶⁵ found the fluorescein fluorescence detection limit of a B-D FACScan to be 826 MESF; in the absence of background, the limit would have been 285 MESF. The corresponding figures for his arc source instrument^{100,2}, the prototype of the Skatron Argus/Bio-Rad Bryte HS, were 1800 MESF with background and 500 MESF without background.

In the arc source instrument, which uses the epiillumination system of a fluorescence microscope, much of the fluorescence background is due to fluorescence induced in the cover slip and lens, and possibly in the dichroic, by excitation light; this becomes a much more severe problem when using UV (366 nm) or blue (436 nm) excitation than when excitation is at higher wavelengths. The use of quartz or silica in place of glass in optical components can reduce, but probably not eliminate, the fluorescence background; Partec, the major manufacturer of arc source flow cytometers, supplies quartz lenses and filters.

Sensitivity VI: Electrons Have Statistics, Too

In a PMT, the random processes don't stop after photons hit the cathode; Steen¹¹⁶⁵ considers the effects of electron statistics in the PMT at various gain levels on precision. PMT gain results from secondary electron emission from successive dynodes; as the voltage between dynodes is increased, electrons acquire more energy before striking the next dynode, and release more secondary electrons. Electron statistics increase the contribution to CV; for an R928 PMT at 300 V, this factor is almost 1.4; with 1000 V applied to the tube, the factor is less than 1.1.

Most flow cytometers provide several stages of electronic gain, allowing the preamplifier output to be amplified by a selectable factor before signals are digitized. Steen's calculations and data provide excellent reasons for not using such electronic amplification; an amplifier gain of 10 and a low PMT voltage result in a larger contribution to CV from the PMT and 10 times as much amplifier noise as an amplifier gain of 1 and a higher PMT voltage that increases PMT gain by a factor of 10. If your preamp can get a full-scale output

signal from a 10-100 μ A input, you can avoid using additional gain stages; check with your manufacturer.

Source Noise Fluctuations and Performance

The background light reaching a flow cytometer's detectors, which are shielded from outside light, is dominated by light derived from the source. In the case of scatter detectors, background light levels are relatively high, and due primarily to stray illumination light scattered by interfaces in the flow chamber (or, in stream-in-air systems, by the air/stream interface) and by optical components. In the case of fluorescence detectors, the background light level is typically lower. Fluorescence from materials in the core and sheath and from optical elements, including flow chamber components, lenses, and filters, predominates; there may also be some Raman scattering from water molecules. The levels of stray light exhibit the same fluctuations as the source. The background light can be considered as composed of a constant (DC) component, which is, ideally, removed by baseline restoration, and a variable (AC) component, which is not.

I Blurred It Through the Baseline

Periodic fluctuations in the output of the light source result from such factors as power supply and laser plasma noise and arc wander. After baseline restoration, these fluctuations show up as noise or "ripple" on the baseline; higher source noise levels result in higher baseline noise. At illumination levels that do not produce photon saturation, the CV of fluorescence and scatter measurements cannot be lower than the RMS noise level of the source (which is the CV of source intensity) unless a source noise compensation circuit (pp. 147-8) is used. In the absence of background, source noise affects precision to at least this extent. When background is present, it also exerts proportionally larger effects on smaller signals.

Suppose the light source is an air-cooled argon ion laser, with 1% RMS noise. If no background is present, and the flow cytometer is otherwise ideal, measurements of completely uniform fluorescent particles will have a CV representing contributions from the source variation and from photon statistics. From the photon statistical CV's given for large signals in Table 4-3, and the contributions from electron statistics in the PMT previously discussed, we can guesstimate a contribution of 0.3%. The final CV, in percent, will be $[(1)^2 + (0.3)^2]^{1/2} = (1.09)^{1/2} \cong 1.044\%$.

If the same light source is used, with background increased to produce 1 μ A at the PMT output, or 100 mV at the output of the 100,000 V/A current-to-voltage converter in the preamp, the 100 mV DC average will be removed by baseline restoration. The AC representing the 1% RMS noise, will show up as 1% of 1 V, or 1 mV, RMS, centered on the 0 V baseline. Assuming the noise is Gaussian, I use the quick-and-dirty rule of thumb given on p. 147 to estimate peak-to-peak voltage as 6 times RMS voltage. In this instance, with 1 mV RMS noise, the noise waveform should be within ± 3 mV of zero at least 99% of the time.

If you're looking at the highest decade (1-10 V), 3 mV is negligible at the top, and represents only 0.3% of signal at the bottom. In the next decade, when you get down to 100 mV, the 3 mV is 3% of signal value. By the bottom of the third decade, 3 mV looms very large on top of a 10 mV signal, and the 3 mV noise level prevents you from getting into the bottom half of the fourth decade. Even with perfectly identical 100 nA pulses going into the preamp, which should give you identical 10 mV pulses out, you'd see a very much broadened distribution due to the background noise. Thus, in the presence of background light, source noise fluctuations will decrease precision; the lower the signal value, the worse it gets. Fluctuations will also compromise sensitivity; with 3 mV noise, signals below 3 mV become undetectable. More background with the same RMS noise, or, alternatively, the same background with more RMS noise, will make things proportionally worse.

Good flow cytometer design should keep both the background and the source noise level low. Observation in flat-sided cuvettes instead of round streams or capillaries has been the principal means by which background has been reduced; background in fluorescence channels has also been reduced by the use of better designed filters, which do not themselves fluoresce and which transmit virtually no light at the excitation wavelength.

In stream-in-air instruments, in which some of the illuminating laser light is scattered in the direction of the fluorescence collection lens, an obscuration bar is typically used to keep most of this light from reaching the lens. Although fluorescence signal levels can be increased when the bar is removed, sensitivity may actually be decreased if the excitation light used induces substantial fluorescence in the lens elements. UV and blue-violet light are the worst offenders here; I have found fluorescence due to UV excitation light even in a quartz lens, although fluorescence in glass lenses is much higher. Where maximum light collection is the goal, it may be necessary to use catadioptric reflective lenses for light collection to eliminate fluorescence.

Since scatter channels operate at the illuminating wavelength, they cannot include filters that block this wavelength; background is therefore higher than in fluorescence channels. Baseline noise generally determines the sensitivity of a scatter channel; excessive noise may impede or prevent use of the scatter signal for triggering. A water-immersed flow system, in which light scattering interfaces are several millimeters away from the sample stream and well out of the field of view of the collection lens, provides the lowest scatter background. Block's instrument for virus detection²⁴, using such a flow system, could trigger on scatter signals from particles smaller than 0.1 μ m, using a 100 mW water-cooled argon laser source, with 0.2% RMS noise. It was also possible to detect the same particles, with a predictably lower signal-to-noise ratio, using a 10 mW air-cooled argon laser with a similar RMS noise level, while such particles could not be detected in either a round capillary or a stream in air using either laser, due to the higher background. Illustrating

the same principle, I found I could detect and trigger on scatter signals from 0.3 μm particles in a stream-in-air system using a He-Cd laser when the source noise was reduced to 3% peak-to-peak by an electro-optic modulator, but not when the same laser was running at slightly higher power with 15% peak-to-peak noise.

Source noise compensation (pp. 147-8; Figure 4-15) can be used to improve precision and, to some extent, sensitivity of measurements, although it is difficult to use with scatter detectors, because of the relatively high background noise levels involved. The minimum measurement CV calculated previously for measurements with an air-cooled argon laser with 1% RMS noise was 1.044%; instruments using such light sources achieve CVs in the range of 1.5%, suggesting that source noise may be the major contributor to variance and that CVs might be improved by noise compensation.

Restoration Comedy: The Case of the Disappearing Leukocytes

Baseline restoration is great stuff, up to a point, but we should remember that backgrounds, DC or AC, are never desirable. A 1 V DC background on the signal coming out of the preamp first stage, means that 10 μA of the PMT output current represented DC background; that only leaves us 90 μA to work with if we want to restrict output current to 100 μA to prevent nonlinear PMT response.

Under some circumstances, fluorescence backgrounds can get very high; with a 10 V background, even if you have a 10 V signal, you won't see it, because the op amps in the preamp and baseline restorer stop working like op amps when they reach **saturation**, i.e., get to the voltage or current limits imposed by their power supplies. Restorer failure on this basis was the culprit in one of my most puzzling cases, one I have called "The Case of the Disappearing Leukocytes" (I know, I should have gotten this into Dr. Watson's book¹⁰³⁰).

My colleagues at Block and I used the first multiple excitation beam flow cytometers^{88,92} to count blood cells stained with a mixture of three dyes^{89,90,91,93}. Ethidium bromide, now well known in flow cytometry, produced red nuclear staining. Brilliant sulfaflavine (BSF), excited at 420 nm, stained eosinophil granules green, and LN ("Long Name"), excited by UV light, stained neutrophil and eosinophil granules blue. BSF and LN, both sulfonated acid dyes, produced intense staining, but there was a tremendous amount of background fluorescence in cell suspensions. The instruments^{88,92} incorporated baseline restorers and noise compensation, and, in addition, used a very small core stream to minimize background fluorescence.

I was somewhat surprised when John Steinkamp (personal communication) told me in 1978 that he was not able to detect LN and BSF signals in blood cells stained with the dye mixture, when he examined the cells in Los Alamos' dual laser source instrument¹²⁵; the multiwatt ion lasers used there for UV and violet illumination should have produced very strong fluorescence signals from the cells.

Shortly thereafter, I was able to run a sample of leukocytes stained according to our protocol on one of Ortho's ICP instruments^{9,83}, using UV illumination to excite LN and blue illumination to excite BSF. I was unable to get signals from cells that I had seen were brightly stained, having examined the sample under the fluorescence microscope.

This represented a clear violation of:

*Shapiro's Second Law
of Flow Cytometry
[Flip] Wilson's Rule¹²⁶:
What You See
Is What You Get!*

The provision in the ICP for observation of the sample stream allowed resolution of the mystery, if not of the cells. The background fluorescence from dye in the core stream was very strong.

Under staining conditions that result in high dye background, the cells may be brighter than the core, per unit volume, but the volume of material from which the fluorescence signal is derived, which is larger than the cell volume, becomes critical. Taking typical numbers, suppose we have a 20 μm core illuminated by a beam 20 μm high. The **observation volume** thus defined will be roughly the volume of a cylinder with diameter equal to the core diameter and height equal to the beam height; this is $\pi \times 10^2 \times 20$ femtoliters (fL), or 6284 fL, where a femtoliter is 10^{-15} liters, or what used to be called a cubic micron (1 μm^3). The volume of a neutrophil is about 400 fL. Suppose the cell is 5 times as bright as the core; since the observation volume is about 15 times the cell volume, we can expect the fluorescence signal obtained during the passage of a cell through the observation volume to be only 1 1/3 times the background signal from an observation volume containing no cells.

Even when the ratio of core fluorescence to cell fluorescence is low and the observation volume is small, we can be prevented from resolving cell fluorescence over background (core) fluorescence if the background fluorescence is bright enough to drive some stage of the electronics into saturation. The instruments originally used to detect BSF and LN fluorescence^{88,92} were designed to deal with the high background fluorescence associated with these dyes; a 5 μm core and 20 μm beam height were used, giving an observation volume of <400 fL, and, in the electronics, feedback was applied to the first stage of the preamplifier to subtract the DC baseline. The Los Alamos instrument¹²⁵ and the ICP had larger observation volumes and thus had to deal with more background fluorescence; while I am not certain about the resulting effects on the electronics in the Los Alamos system, it was clear that the ICP's electronics could not cope with the high background fluorescence. Thus, the mystery of the disappearing leukocytes was solved, proving that what you see is not always what you get, and that the flow cytometer is not always quicker than the eye.

Top 40 Noise Sources

To this point, I haven't said much about noise and related problems originating in flow cytometer electronics, although I have mentioned that there always seemed to be a few millivolts of noise about. Flow cytometer design, like everything else, faces economic constraints, and it doesn't pay to design components with better performance than you need. As long as people were using 8-bit A-D converters, in which the distance between channels is 39 mV, any noise level or offset below 39 mV was acceptable, especially when weak signals went into log amps. If you measure the outputs of older commercial flow cytometers, you'll find that "zero" can be anywhere within 30 mV of ground, and that peak detectors and integrators may be inaccurate in the range below a few hundred millivolts. This means you can pretty much forget about upgrading an old instrument to 4-decade digital log performance simply by hanging a 16-bit or higher resolution ADC on its outputs.

The manufacturers have fixed things in the design of newer instruments; recent work on the subject, to be discussed below, indicates that electronic noise is no longer a limiting factor for sensitivity, at least in commercial systems. In the course of trying to wring the last few millivolts of noise out of one of my CytoMutts²⁴⁶⁷, Phil Stein and I hung an electronic (not optical) spectrum analyzer on the output of the prototype low noise electronics to see where the noise was coming from; in addition to the usual power line noise, the most significant contributions came from the horizontal sweep generator in the computer monitor and the power supply in a He-Ne laser. Things got noticeably quieter when we moved the offending laser supply further away from the electronics and used a notebook computer with an active-matrix screen to run the instrument, eliminating the monitor (this consideration later justified buying LCD displays for the lab long before they got as cheap as they are now).

The laser supply and monitor noise were in the frequency range below 100 kHz; when we looked up in the 100 MHz range, we found detectable noise spikes from most of the local FM stations. Just when we think we've gotten the grunge out of the flow system, it shows up in the electronics.

Sensitivity 007: Q and B (Dye Another Day?)

After the third edition of this book appeared, with its lengthy discussion of source noise fluctuation effects on sensitivity, Harald Steen took me to task; he didn't think source noise was nearly as important as photoelectron statistics as a limiting factor in sensitivity. I was offended until I realized that his light sources are quieter than mine. The dominant current approach to the definition and measurement of sensitivity is based on an amplification of Harald's earlier work¹¹⁶³⁻⁵ by Jim Wood, formerly of Beckman Coulter, and Bob Hoffman, of BD Biosciences²⁴⁷⁶⁻⁸⁰, and I again acknowledge their help in putting this section together. The topic is

obviously important if rival manufacturers have cooperated over a period of years to make better sense of things.

Q and B are neither mysterious secret agents nor (yet) Jim and Bob's new nicknames. Q is a measure of detection efficiency, with units of photoelectrons (or fractions of photoelectrons) emitted from the detector photocathode per molecule-equivalent (MESF) of fluorochrome measured. It is a function of many variables, potentially being affected by optical misalignment and disturbances of fluid flow, but more directly and obviously influenced by which fluorochrome is measured, by the power of the excitation source and by the efficiency of the light collection optics and the quantum efficiency of the detector photocathode.

B is a measure of background noise, usually expressed in MESF units as the number of fluorochrome molecule equivalents required to produce that level of background. B includes contributions from photoelectrons generated by photons emitted by fluorochrome molecules not associated with the particle (e.g., dye in solution), by photons from fluorochromes other than the one of interest (crosstalk due to spectral overlap), by photons emitted by cellular constituents (autofluorescence), by photons at the excitation wavelength from the light source that have passed through the detector filter, and by photons resulting from Raman scattering of the illuminating light by water. Electrons produced by thermionic emission from the photocathode (dark current) are also included in B, but usually not significant except in the far red and near infrared. The quantity F, also usually expressed in MESF units, denotes the number of particle-associated fluorochrome molecules.

It is possible to determine Q and B using only sets of beads. Measurements are made on the flow cytometer using a linear, rather than a logarithmic, scale. As a first step, the scale of the instrument must be calibrated in terms of MESF units of the fluorochrome of interest, using beads labeled with known amounts of that fluorochrome. Bangs Laboratories makes fluorescein-, phycoerythrin-, and PE-Cy5-labeled Quantum MESF beads, and BD Biosciences provides QuantiBRITE beads bearing known numbers of molecules of phycoerythrin²⁴⁸¹⁻²; such beads can be used for a single-point MESF calibration, assuming the linearity of the instrument is within specification.

The next step uses beads from sets bearing various amounts of fluorescent dyes that emit in the same spectral region as the fluorochrome(s) of interest; these are available from a number of manufacturers, including Bangs Laboratories, Molecular Probes, and Spherotech. Dimly fluorescent or unstained beads will emit relatively few fluorescence photons, and the variances (i.e., the squares of the standard deviations, see pp. 19-22 and Chapter 5) of fluorescence distributions from such beads will therefore be dominated by contributions from photoelectron statistics, but will also contain "basis" contributions resulting from variations in bead size and dye loading and from instrumental factors. Since the brightest bead in such a bead set typically carries at

least 100,000 molecules of dye, the variance of its fluorescence distribution will be dominated by the “basis” factors. The bright bead is first measured, and its fluorescence variance calculated. The measurement need not be made on the calibrated scale; in fact, it will usually be necessary to lower the gain to get the fluorescence distribution from the bright bead on scale. The gain is next (re)adjusted, if necessary, to the value used for MESF calibration, and fluorescence from three dim beads is measured; variances are calculated for each of the resulting bead peaks, while MESF units per bead are determined from the positions of the peaks.

“Photoelectron variances” for each of the three dim beads are calculated by subtracting the measured variance of the brightest bead from the measured variances of the dim beads. Next, the photoelectron variances are plotted (as the dependent variable, i.e., on the vertical axis) against the MESF values for the dim beads, and the equation of a regression line best fitting through the points is determined using the method of least squares (a spreadsheet for the calculation has been developed by Bob Hoffman and Jim Wood, and should be available on the Wiley web site associated with this book). The slope of the line will be $1/Q$, and it will intersect the vertical axis at B/Q , allowing B and Q to be determined.

B and Q have been calculated for a reasonable number of working flow cytometers. Before I give you the details, however, I will ask you to look back at Table 4-4, on p. 203, in which I tabulated signal levels that might be expected to come out of a PMT in a flow cytometer. If we assume that the 5-decade scale used in constructing that table represents signals from between 10 and 1,000,000 cell-associated MESF, we can calculate Q without beads; using the numbers of photoelectrons in the sixth column of the table we get $Q = 0.62$ photoelectrons/MESF. I suggested that the calculations in the table might be optimistic, and they were. Hoffman and Kuhlmann²⁴⁸² observed that typical values for Q in BD FACScan and FACSCalibur cytometers were 0.25 photoelectrons/MESF for phycoerythrin (PE) measurements and only 0.012 photoelectrons/MESF for fluorescein measurements. They found it unusual to see a value of Q much higher than 0.4, even for PE, on these optically efficient benchtop systems.

B varies considerably more than does Q with operating conditions. Once the instrument is properly aligned, assuming there is no substantial variation in laser output, one would expect Q to remain relatively constant. Some factors that contribute to B , such as Raman scattering and illumination light leakage through filters, will also remain relatively constant; however, others, such as fluorescence from dye in solution and from other fluorochromes with spectral overlap, will vary. Typical values of B associated with bead measurements as just described were 2,000 MESF for FITC and 300 MESF for PE. However, background contributed by unbound fluorescent antibody in no-wash immunofluorescence analyses was 4,000 fluorescein MESF and 1,600 PE MESF. Spectral overlap from fluorescein-labeled anti-CD45 anti-

body on lymphocytes generated a 7,800 MESF background level in the detector used for PE.

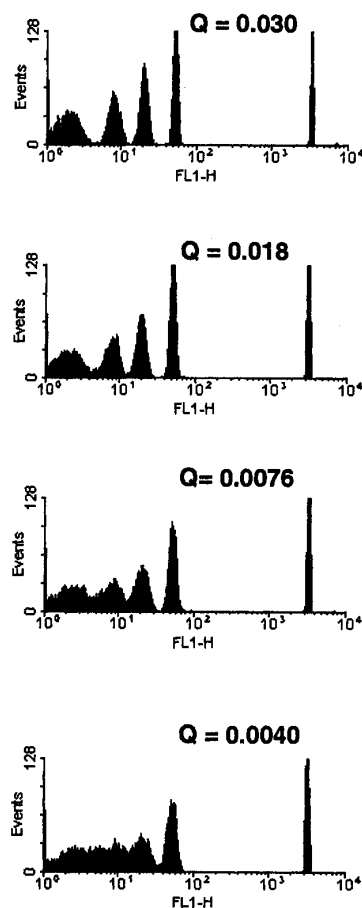


Figure 4-61. 488 → 525 nm fluorescence distributions for the four dimmest beads and the brightest bead of an 8-bead dyed bead set (Spherotech Rainbow Beads) measured with progressively lower values of Q . Courtesy of Bob Hoffman, BD Biosciences.

It's easy enough to figure out why Q is important in determining sensitivity. Figure 4-61 shows distributions of green fluorescence (the “fluorescein channel,” FL1 in BD parlance) of a mixture of five of the eight beads in a Spherotech Rainbow Bead set. The measurements were all made in the same cytometer; Q was progressively decreased by lowering laser power, while B remained constant. Note that the peaks representing the dimmer beads become progressively broader as Q decreases, to the point at which the second dimmest peak is not clearly separable from either the dimmest or the third dimmest when $Q = 0.0040$. That makes sense; when fewer photoelectrons contribute to each measurement, the CV of the peak increases, and the peak gets wider, although its center (median or mode) remains in

pretty much the same place, meaning that the degree of overlap between neighboring peaks keeps increasing as Q drops. Not incidentally, since it has already been noted that the typical value of Q for a fluorescein measurement channel is 0.012^{2482} , we can conclude that the top two curves show better separation of dim peaks than is likely to be observed in an average instrument, and hope that we will do better than the bottom two.

While decreasing Q has the effect of broadening all of the distributions, increasing B has the primary effect of broadening the dimmer peaks and pushing the distributions up the scale, and can be an equally effective way of preventing us from telling dim objects from dimmer ones, as can be seen from Figure 4-62.

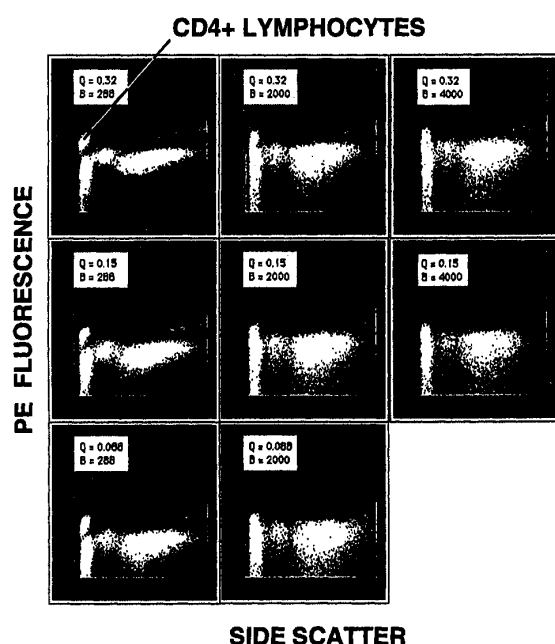


Figure 4-62. Separation (or lack thereof) of dimly stained CD4+ lymphocytes from unstained cells as observed with various values of Q and B . Courtesy of Bob Hoffman, BD Biosciences.

The cells shown in the figure were stained with a sub-saturating concentration of PE-labeled anti-CD4 antibody in the presence of a saturating concentration of PerCP-labeled antibody, which exhibits little or no spectral overlap with the PE-antibody. CD4+ lymphocytes normally bind about 50,000 anti-CD4 antibody molecules per cell; it is probable that the CD4+ cells shown here are binding at most only a few thousand molecules of PE-labeled antibody. The CD4+ lymphocyte population (identified in the top left panel of the figure, with low side scatter and higher PE fluorescence than the other cell clusters) is readily distinguishable when background is low ($B = 288$ MESF), although separation from unstained lymphocytes gets worse as Q drops from 0.32 to 0.088. However, when B is 2000 or 4000, only the samples for which $Q = 0.32$ show even a hint of separation between CD4+ cells and other lymphocytes.

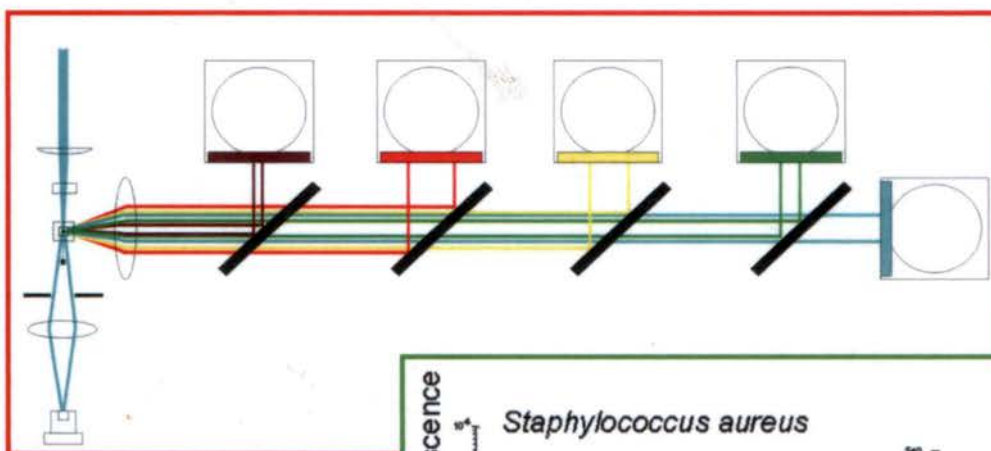
A "separation parameter," SP , where

$$SP = (Q \times F)^{1/2} / [1 + 2(B/F)]^{1/2},$$

may be calculated²⁴⁷⁷; it defines the difference, in normalized standard deviation units (see Chapter 5), between a stained population, with mean intensity F MESF units, and an unstained or blank population. As it turns out, decreasing B , if possible, can allow a constant degree of separation between populations to be maintained in the face of a decrease in Q . To look at things in another light, as it were, the sensitivity of the system, meaning its ability to separate dim fluorescence peaks, cannot be predicted from either Q or B alone; values of both must be known, and it is important to remember that B can change dramatically with experimental conditions, e.g., whether or not samples are washed before immunofluorescence measurements are made.

You may not want to get involved in the measurement of Q and B , but it is a good idea to keep sets of beads on hand and run them every now and then; if you find that you are less able to resolve dimly stained populations than you once were, it's time to adjust or fix the instrument.

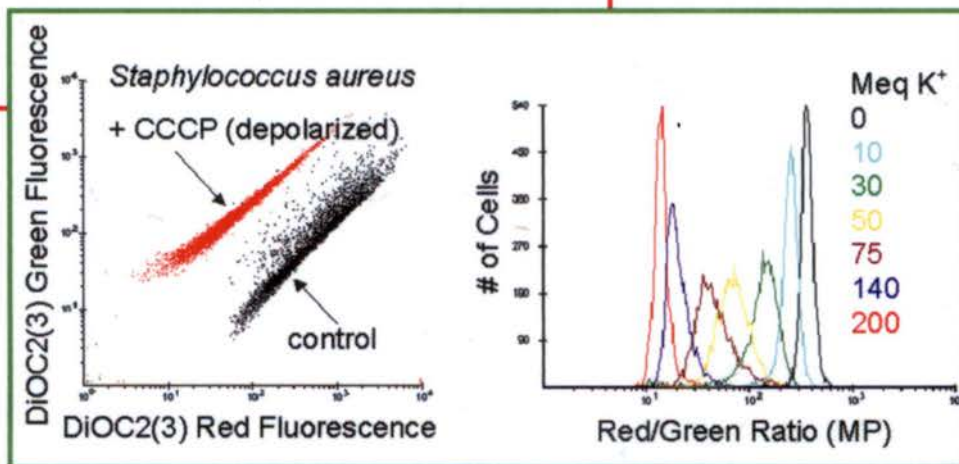
That's it for now on how and how well flow cytometers work, except for a final message for you to take along while you read the next chapter on data analysis: Both the flow system and the data analysis system of a flow cytometer can give you garbage out in response to garbage in.



Acclaim for the
Third Edition:

"The standard reference
for anyone interested in
understanding flow
cytometry technology."

—*American Journal of
Clinical Oncology*

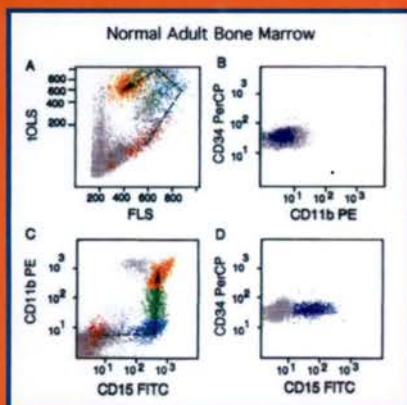


PRACTICAL FLOW CYTOMETRY

FOURTH EDITION

HOWARD M. SHAPIRO

Practical Flow Cytometry provides the most readable and comprehensive resource covering the technology, principles, and biomedical applications of analytical cytology using laser-activated detection and cell sorting. This *Fourth Edition* of the discipline's benchmark text is completely updated with several added features, including:



- A completely revised introduction, emphasizing connections between flow cytometry, static cytometry, and microscopy
- Expanded coverage of applications in biotechnology, cell biology, genetics, hematology, immunology, microbiology, and oncology
- Discussion of new commercial instruments
- Review of alternatives to flow sorting
- Updated and enhanced references and

Cover Design: Denise Papania

WILEY-LISS
wiley.com

This book is required reading for anyone in cell and molecular biology, microbiology, oncology, and hematology.

