ATTACHMENT K-2

LINDA HALL
DEC 31 1986
LIBRARY

Journal of APPLIED PHYSICS

Volume 61 | 1 January 1987 | Number 1

a publication of the American Institute of Physics

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 626

CODEN: JAPIAU

ISSN: 0021-897

Journal of APPLIED PHYSICS

Lester Guttman, Editor

Robert E. Holland, Associate Editor

Catherine M. Krippel, Assistant to the Editor

Editorial Board

Term ending 31 December 1987	Term ending 31 December 1988	Term ending 31 December 1989
John Clarke	Rolf Landauer	John S. Blakemore
George D. Cody	C. Kumar N. Patel	Karl Hess
Y. R. Shen	James J. Rhyne	George I. Stegeman
Takuo Sugano		

AIP EDITORIAL AND COMPOSITION STAFF: Janis Bennett, *Managing Editor;* Elizabeth Belmont, *Chief Copy Editor;* Anna Damaskos, *Senior Copy Editor;* Leon Benedict, *Composition Manager;* Dianne Longi, *Production Control Supervisor;* Mark Green and Marilyn Schrage, *Composition;* Barbara Maki, *Keyboard Supervisor;* Honey West, *Pasteup Supervisor*

The *Journal of Applied Physics* is published semimonthly by the American Institute of Physics (AIP) with the cooperation of The American Physical Society. It is devoted to general physics and its applications to other sciences, to engineering, and to industry. (Prior to 1937, the Journal carried the name *Physics*.) The Editor welcomes manuscripts describing significant new experimental or theoretical results in applied physics or manuscripts concerning important new applications of physics to other branches of science and engineering.

Information for Contributors

Submit manuscripts to Lester Guttman, Editor, *Journal of Applied Physics,* Argonne National Laboratory, P.O. Box 8296, Argonne, IL 60439-8296. Submission is a representation that the manuscript has not been published previously nor currently submitted for publication elsewhere. The manuscript should be accompanied by a statement transferring copyright from the authors (or their employers—whoever holds the copyright) to AIP; a suitable form for copyright transfer is occasionally printed in the back of this journal and is available from the Editor's office or AIP. This written transfer of copyright, which previously was assumed to be implicit in the act of submitting a manuscript, is necessary under the 1978 U.S. copyright law in order for AIP to continue disseminating physics research results as widely as possible. Further information may be obtained from AIP.

Publication Charge: To support the cost of wide dissemination of research results through publication of journal pages and production of a data base of articles, the author's institution is requested to pay a *page charge* of $45 per page (with a one-page minimum) and an *article charge* of $20 per article. The page charge (if honored) entitles the author to 100 free reprints. For **Errata** the minimum page charge is $10, with no article charge and no free reprints. Nonpayment of the publication charge may lead to delays in publication.

Two categories of manuscripts are acceptable, **full-length articles** and **Communications.** The latter are short contributions not generally exceeding in length nine double-spaced typewritten pages or approximately 2500 words of text reduced, however, by allowances for equations, tables, and figures. Abstracts are required for manuscripts of both types. Being short, Communications are often published more quickly than full-length articles. Instructions for preparation of **manuscripts** will be found in the January issue of this journal and occasionally in later issues. In brief and incomplete summary: Two copies of the manuscript, including clear copies of the figures, are required, double-spaced throughout on one side of 21.6×28 cm (8½×11 in.) paper. The original should be retained by the author until it is requested, after review. The *abstract, references, figure captions,* and *tables* are similarly to be double-spaced each on its separate sheets. The original *figures* should be original line drawings or high-contrast glossy prints not larger than 21.6×28 cm and should be included with the original submission. *References* should conform to the style used in AIP journals including this one. Unusual mathematical symbols should be clarified with marginal notes.

AIP's **Physics Auxiliary Publication Service (PAPS)** is a low-cost depository for material which is part of and supplementary to a published paper, but is too long to be included in the journal; inquire of the Editor.

Proofs and all correspondence concerning papers in the process of publication should be addressed to: Editorial Supervisor, *The Journal of Applied Physics,* American Institute of Physics, 335 East 45th Street, New York, NY 10017. In all correspondence reference should be made to title, author, journal, and scheduled date of issue. A limited number of **alterations in proof** are unavoidable, but the cost of making extensive alterations after the article has been set in type will be charged to the author.

Copyright 1987, American Institute of Physics. Individual teachers, students, researchers, and libraries acting for them are permitted to make copies of articles in this journal for their own use in research or teaching, including multiple copies for classroom or library reserve use, provided such copies are not sold. Copying for sale is subject to payment of copying fees. (See "Copying Fees" paragraph elsewhere in this journal.) Permission is granted to quote from this journal with the customary acknowledgment of the source. To reprint a figure, table, or other excerpt requires in addition the consent of one of the original authors and notification to AIP. Reproduction for advertising or promotional purposes, or republication in any form, is permitted only under license from AIP, which will normally require that the permission of one of the authors also be obtained. Direct inquiries to Office of Rights and Permissions, American Institute of Physics, 335 East 45th Street, New York, NY 10017.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 627

Second-class postage rates paid at Woodbury, N.Y. and at additional mailing offices

Journal of **APPLIED PHYSICS** Vol. 61, No. 1, 1 January 1987

GENERAL PHYSICS: Nuclear, Atomic, and Molecular (PACS 01–36)

1 Superconducting Sn/Sn-oxide/Sn tunneling junctions as high-resolution x-ray detectors — D. Twerenbold, A. Zehnder

8 A new frequency-domain fluorometer for the rapid determination of picosecond rotational-correlation times — Frank V. Bright, Curtis A. Monnig, Gary M. Hieftje

CLASSICAL PHENOMENOLOGY: Electricity, Magnetism, Optics, Acoustics, Heat, Mechanics (PACS 41–52)

12 Guiding of an electron beam from a rf accelerator by a laser-ionized channel — R. L. Carlson, S. W. Downey, D. C. Moir

20 Beam deflection phenomena in a Pulselac post-acceleration gap — D. J. Johnson, T. R. Lockner

36 Effects of cathode surface roughness on the quality of electron beams — Y. Y. Lau

45 Periodic permanent magnet field-transported relativistic electron beams — Hidenori Matsuzawa, Tetsuya Akitsu, Yoshihiko Suzuki, Satoshi Fuki

52 Ion exchange process for fabrication of waveguide couplers for fiber optic sensor applications — S. Honkanen, A. Tervonen, H. von Bagh, M. Leppihalme

57 Spike-mode oscillations in flashlamp pumped Nd lasers — Lars Lindgren, Lennart Lundgren

64 Proton-exchanged, lithium niobate planar-optical waveguides: Chemical and optical properties and room-temperature hydrogen isotopic exchange reactions — A. Loni, R. M. De La Rue, J. M. Winfield

68 Singlet oxygen generator using a porous pipe — K. Takehisa, N. Shimizu, T. Uchiyama

74 Comparison of transient thermal conduction in tellurium and organic dye based digital optical storage media — A. N. Burgess, K. E. Evans, M. Mackay, S. J. Abbott

81 Large-signal time-domain modeling of low-pressure rf glow discharges — Michael S. Barnes, Tina J. Colter, Michael E. Elta

(Continued)

Subscription Prices (1987)	U.S.A. & Poss.	Foreign (incl. Can. & Mex.)	Optional air freight Europe, Mideast, N. Africa	Optional air freight Asia & Oceania	Microfiche (1st class/airmail)
Members*	$80.00	$140.00	$180.00	$230.00	$80.00
Reg. rate	$580.00	$640.00	$680.00	$730.00	$580.00

*AIP Member and Affiliated Societies.

Back-Number Prices. Single copies from Vol. 24 (1953) through Vol. 55 (1984): $35.00; Vol. 56 (1984) and thereafter: $25.00. Special Supplements: $50.00. Volumes 1–23 (1931–1952) are available only on microfilm.

The *Journal of Applied Physics* (ISSN: 0021-8979) is published semimonthly by the American Institute of Physics, 500 Sunnyside Blvd., Woodbury, NY 11797. Second-class postage paid at Woodbury, NY, and additional mailing offices. POSTMASTER: Send address changes to *Journal of Applied Physics*, 500 Sunnyside Blvd., Woodbury, NY 11797.

Subscriptions, renewals, address changes, and single-copy orders should be addressed to *AIP Subscription Fulfillment Division, 335 East 45th St., New York, NY 10017.* Allow at least six weeks advance notice. For address changes please send both old and new addresses and, if possible, include a mailing label from the wrapper of a recent issue. For your convenience a **change of address form is included in every issue of *Physics Today;* please use it.** Requests from subscribers for missing journal issues will be honored without charge only if received within six months of the issue's actual date of publication; otherwise, the issue may be purchased at the single-copy price. (AIP Headquarters are located at 335 East 45th St., New York, NY 10017; Subscription Fulfillment offices are located at 500 Sunnyside Blvd., Woodbury, NY 11797.)

Reprints of individual articles in this journal may be ordered singly at $10.00 per article copy (postage included) for articles up to 20 pages. Beyond 20 pages there is a surcharge of $0.20 per page. Air mail delivery is available. Orders are filled within one week of receipt or of the date of publication, whichever is later. Send orders to the American Institute of Physics, *Current Physics Reprints,* 335 East 45th St., New York, NY 10017.

Copying Fees: The code that appears on the first page of articles in this journal gives the fee for each copy of the article made beyond the free copying permitted by AIP. (See statement under "Copyright" elsewhere in this journal.) If no code appears, no fee applies. The fee for pre-1978 articles is $0.25 per copy. With the exception of copying for advertising and promotional purposes, the express permission of AIP is not required provided the fee is paid through the *Copyright Clearance Center, Inc. (CCC), 21 Congress St., Salem, MA 01970.* Contact the CCC for information on how to report copying and remit payment.

Microfilm subscriptions of complete volumes of the *Journal of Applied Physics* are available on 16 mm and 35 mm. The *Journal of Applied Physics* also appears on a monthly basis in *Current Physics Microform* (CPM) Section 1 along with 25 other journals published by the American Institute of Physics and its member societies. A *Microfilm Catalog* is available on request. The *Journal of Applied Physics* is indexed quarterly in *Current Physics Index,* a subject and author index (with abstracts) to all journals published by AIP and its member societies.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 628

90	Spark propagation mechanisms in ambient air at the surface of a charged dielectric. I. Experimental: The main stages of the discharge	S. Larigaldie
102	Mechanisms of spark propagation in ambient air at the surface of a charged dielectric. II. Theoretical modeling	S. Larigaldie
109	Influence of electrical resonance on the interpretation of optogalvanic data	Seong-Poong Lee, Erhard W. Rothe, Gene P. Reck
113	Theoretical and experimental study of rotation in a vacuum-arc centrifuge	Rahul R. Prasad, Mahadevan Krishnan
120	Monte Carlo calculations on the influence of attachment on the thermalization of secondary electrons in an electron-beam-sustained discharge	G. Schaefer, G. F. Reinking, K. H. Schoenbach
126	Spectroscopic studies of an argon plasma produced in a relativistic electron beam gas puff Z pinch	R. E. Stewart, D. D. Dietrich, P. O. Egan, R. J. Fortner, R. J. Dukart
131	Flash x radiography of laser-accelerated foils	R. R. Whitlock, S. P. Obenschain, J. Grun, J. A. Stamper, J. A. Sprague, B. V. Sweeney, B. H. Ripin, R. S. Craxton
142	Atomic chlorine concentration and gas temperature measurements in a plasma etching reactor	J. Wormhoudt, A. C. Stanton, A. D. Richards, H. H. Sawin

CONDENSED MATTER: Structure, Mechanical, and Thermal Properties (PACS 61–68)

149	The influence of oxygen and other impurities on the crystallization of $NiZr_2$ and related metallic glasses	Z. Altounian, E. Batalla, J. O. Strom-Olsen, J. L. Walter
156	Defect-state generation in Czochralski-grown (100) silicon rapidly annealed with incoherent light	D. Barbier, M. Remram, J. F. Joly, A. Laugier
161	Strain in self-implanted silicon	A. C. Chami, E. Ligeon, R. Danielou, J. Fontenille, R. Eymery
166	Microstructural characterization of nitrogen-implanted silicon-on-insulator	P.-H. Chang, C. Slawinski, B.-Y. Mao, H. W. Lam
175	High-energy-density pulsed ion-beam irradiation of Co/Si, Pt/Si, and Au/Si	R. Fastow, J. W. Mayer
182	Refinements in the method of moments for analysis of multiexponential capacitance transients in deep-level transient spectroscopy	K. Ikossi-Anastasiou, K. P. Roenker
191	Theory of optical beam induced current images of defects in semiconductors	T. Wilson, E. M. McCabe
196	Shock compression of powder compacts of CaF_2	T. Taniguchi, K. Kondo, A. Sawaoka
201	PtSi contact metallurgy formed by three-temperature annealing sequences and short annealing time	Chin-An Chang, Armin Segmüller
206	Preparation and structural analysis of a PbSe-SnSe strained-layer superlattice	Z. Hiroi, N. Nakayama, Y. Bando
215	Investigation of crystalline and optical properties of $Al_{0.48}In_{0.52}As$ grown by molecular-beam expitaxy	J. P. Praseuth, L. Goldstein, P. Hénoc, J. Primot, G. Danan
220	Thermal stability of WSi_x/GaAs interface	S. Takatani, N. Matsuoka, J. Shigeta, N. Hashimoto, H. Nakashima
225	An investigation of the thermal stability of the interfacial oxide in polycrystalline silicon emitter bipolar transistors by comparing device results with high-resolution electron microscopy observations	G. R. Wolstenholme, N. Jorgensen, P. Ashburn, G. R. Booker

CONDENSED MATTER: Electrical and Magnetic Properties (PACS 71–76)

234	Transient photoconductive response of InP:Fe	A. Evan Iverson, D. L. Smith, N. G. Paulter, R. B. Hammond
240	Temporal behavior of modulated optical reflectance in silicon	Jon Opsal, Michael W. Taylor, W. Lee Smith, Allan Rosencwaig
249	Negative resistance in a finite superlattice owing to quantum mechanical reflection	Asuo Aishima, Yoshifumi Fukushima

(Continued)

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 629

257	Arsenic passivation of InP surface for metal-insulator-semiconductor devices based on both ultra-high vacuum technique and chemical procedure	J. Chave, A. Choujaa, C. Santinelli, R. Blanchet, P. Viktorovitch
261	Electrical conduction mechanism in semi-insulating polycrystalline silicon films	R. B. Comizzoli, R. L. Opila
271	A small-signal equivalent-circuit model for GaAs-$Al_xGa_{1-x}As$ resonant tunneling heterostructures at microwave frequencies	J. M. Gering, D. A. Crim, D. G. Morgan, P. D. Coleman, W. Kopp, H. Morkoç
277	Temperature dependence of electron mobility for inversion-mode InP metal-insulator-semiconductor field-effect transistors	Y. Hirota, M. Okamura, T. Hisaki, E. Yamaguchi
284	Characteristics of thermal SiO_2 films during nitridation	Paihung Pan
294	Electronic subbands for $Al_xGa_{1-x}As$/GaAs multilayer and superlattice structures	P. P. Ruden, D. C. Engelhardt, J. K. Abrokwah
299	A simple technique for measuring the interface-state density of the Schottky barrier diodes using the current-voltage characteristics	Hsun-Hua Tseng, Ching-Yuan Wu
305	Compositional distribution in the cross section of chemical-vapor-deposited Nb_3Ge tape conductors and its correlation with superconducting properties	Y. Nakagawa, H. Yamasaki, Y. Kimura
311	Rotational hysteresis of elongated-single-domain particles by chain-of-spheres model	Y. Ishii, M. Sato
319	Aging kinetics of magnetic losses in annealed $Fe_{78}B_{13}Si_9$ amorphous-metallic alloy	H. H. Liebermann
325	Rare-earth substitution in $(BiYCa)_3(FeSiGe)_5O_{12}$ bubble films	L. C. Luther, S. E. G. Slusky, C. D. Brandle, M. P. Norelli

CONDENSED MATTER: Dielectric and Optical Properties (PACS 77–79)

328	Complex refractive indices of AlGaAs at high temperatures measured by *in situ* reflectometry during growth by metalorganic chemical vapor deposition	Hiroji Kawai, Shunji Imanaga, Kunio Kaneko, Naozo Watanabe
333	Effect of elastic strain on the energy band gap in heteroepitaxially grown ZnSe	J. E. Potts, H. Cheng, S. Mohapatra, T. L. Smith
337	Photoluminescence study of the Si-implanted and rapid thermal annealed InP:Fe	Mulpuri V. Rao
342	Refractive index, n, and dispersion, $-dn/d\lambda$, of GaAs at 2 K determined from Fabry–Perot cavity oscillations	D. C. Reynolds, K. K. Bajaj, C. W. Litton, G. Peters, P. W. Yu, J. D. Parsons
346	Alignment of thermal donors in Si by uniaxial stress	P. Wagner, H. Gottschalk, J. Trombetta, G. D. Watkins
354	Characterization of the free-carrier concentrations in doped β-SiC crystals by Raman scattering	H. Yugami, S. Nakashima, A. Mitsuishi, A. Uemoto, M. Shigeta, K. Furukawa, A. Suzuki, S. Nakajima

CROSS-DISCIPLINARY PHYSICS (PACS 81–98)

359	Elimination of dislocations in bulk GaAs crystals grown by liquid-phase electroepitaxy	C. F. Boucher, Jr., O. Ueda, T. Bryskiewicz, J. Lagowski, H. C. Gatos
365	Correlation between dislocation effects on carrier concentration and growth conditions for molecular-beam epitaxially grown GaAs	Masanori Shinohara
372	Ultraviolet laser ablation of polyimide films	R. Srinivasan, B. Braren, R. W. Dreyfus
377	A coagulation equation with separable solutions	S. Simons
381	Performance and analysis of amorphous silicon *p-i-n* solar cells made by chemical-vapor deposition from disilane	Steven S. Hegedus, R. E. Rocheleau, W. Buchanan, B. N. Baron
390	Analytical electron microscopy of Al/TiN contacts on silicon for applications to very large scale integrated devices	A. Armigliato, G. Valdré
397	Varistors based on barium titanate-graphite-rubber composites	K. Lalkishore, K. Ramkumar, M. Satyam

(Continued)

iv

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 630

404 Protective coating on the *p-n* junction of $In_{0.53}Ga_{0.47}As/InP$ *p-i-n* planar diodes by vacuum-evaporated glass — Yusuke Ota, Patrick H-S. Hu, Charles W. Seabury, Michael G. Brown

411 Preparation and characteristics of pyroelectric infrared sensors made of *c*-axis oriented La-modified $PbTi0_3$ thin films — Ryoichi Takayama, Yoshihiro Tomita, Kenji Iijima, Ichiro Ueda

416 Experimental study of microwave magnetostatic waves in thin-film yttrium-iron-garnet with a new magnetic-induction probe — Nickolas P. Vlannes

COMMUNICATIONS

427 Formation of biaxial anisotropy in nickel by plastic deformation — M. Tejedor, J. Carrizo, B. Hernando, A. Fernández

429 Radiation hardness of zinc oxide varistors — J. F. Cordaro, C. E. Shipway, J. T. Schott

431 Excimer laser-induced deposition of silicon nitride thin films — J. M. Jasinski, B. S. Meyerson, T. N. Nguyen

434 Hydrostatic pressure dependence of the exponential optical absorption in $CdInGaS_4$ — T. Toyoda, H. Nakanishi, S. Endo, T. Irie

436 Novel highly sensitive coincidence detection method for resonance ionization signals — Tatsuya Minowa, Hidetsugu Katsuragawa, Koichi Nishiyama, Mitsuyoshi Shimazu, Takashi Inamura

438 Isotope separation during diffusion along an edge dislocation in a simple cubic crystal — Gary Clark, J. P. Stark

440 High power cw operation over 400 mW on five-stripe phase-locked laser arrays assembled by new junction down mounting — Y. Seiwa, T. Aoyagi, S. Hinata, T. Kadowaki, N. Kaneno, K. Ikeda, W. Susaki

442 Temperature effects on carbon and zinc incorporations in GaAs grown by low-pressure metalorganic chemical vapor deposition — L. P. Chen, C. Y. Chang, C. H. Wu

444 Additional considerations in the analysis of mixed conductivity in GaAs — J. J. Winter, H. A. Leupold, A. Ballato, T. Monahan

446 Effect of argon dilution of silane on persistent photoconductivity in single-layered hydrogenated amorphous silicon — Suk-Ho Choi, Choochon Lee, Jin Jang

449 Photoacoustic response of condenser microphones — Tamás Diószeghy

451 Comment on "Small-angle neutron scattering evidence for the absence of long-range magnetic order in amorphous $Fe_{91}Zr_9$"[J. Appl. Phys. **57**, 3407 (1985)] — S. N. Kaul

454 Reply to "Comment on 'Small-angle neutron scattering evidence for the absence of long-range magnetic order in amorphous $Fe_{91}Zr_9$' " [J. Appl. Phys. **57**, 3407 (1985)] — G. E. Fish, J. J. Rhyne

ANNOUNCEMENT

456 Seventh International Conference on Electronic Properties of Two-Dimensional Systems, Santa Fe, New Mexico, 27–31 July 1987

457 CUMULATIVE AUTHOR INDEX

v

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 631

An investigation of the thermal stability of the interfacial oxide in polycrystalline silicon emitter bipolar transistors by comparing device results with high-resolution electron microscopy observations

G. R. Wolstenholme
Department of Electronics and Information Engineering, University of Southampton, Southampton, England

N. Jorgensen
Department of Metallurgy and Science of Materials, University of Oxford, Oxford, England

P. Ashburn
Department of Electronics and Information Engineering, University of Southampton, Southampton, England

G. R. Booker
Department of Metallurgy and Science of Materials, University of Oxford, Oxford, England

(Received 6 May 1986; accepted for publication 9 September 1986)

A comparison is made between the results of high-resolution electron microscope observations and the electrical characteristics of polycrystalline silicon emitter bipolar transistors. Devices are fabricated with and without a deliberately grown interfacial oxide layer, and the thermal stability of this oxide layer is investigated by carrying out a preanneal at temperatures between 800 and 1100 °C after polysilicon deposition, but prior to emitter implant and 900 °C drive-in. The electron microscope observations show that the deliberately grown interfacial oxide is of uniform thickness ~14 Å, but breaks up when annealed at ~950 °C and above, with "balling-up" occurring at ~1100 °C. This correlates with a transistor gain that decreases from ~1400 to ~40. The electron microscopy also shows that a thin interfacial oxide layer is present even when not deliberately grown. This oxide breaks up when annealed at ~900 °C and above, with "balling-up" occurring at ~1000 °C. This correlates with a transistor gain that decreases from ~240 to ~50. Calculations of the effect that such interfacial oxide layers will have on the characteristics of polysilicon emitter bipolar transistors are made, and these predictions correlate well with the measured characteristics.

I. INTRODUCTION

Polycrystalline silicon ("polysilicon") is now widely used in bipolar technology as a means of making contact to the emitter and base of the transistor[1] and this can lead to a dramatic improvement in both switching speed and packing density.[2,3] An additional advantage of using a polysilicon emitter is that an improvement in gain by a factor of between 3 and 30 can be obtained compared with conventional transistors.[4] This improvement in gain can be partly attributed to the transport properties of the polysilicon[5] although this mechanism can only explain an improvement by a factor of ~3. The much larger gain improvements are obtained when a thin interfacial oxide layer is deliberately introduced between the polysilicon and single-crystal silicon.[4,6,7] In this case the gain improvement can be attributed to tunneling through this oxide layer.[8] Eltoukhy and Roulston,[9,10] and later Yu *et al.*,[11] have published unified theories that incorporate both these mechanisms, the former giving a solution suitable for numerical implementation, the latter a versatile analytical solution.

If high-gain polysilicon emitter transistors incorporating an interfacial oxide layer are to be used in commercial integrated circuit processes, it is important that the oxide layer is fully characterized in terms of thickness and uniformity, and that it is stable during subsequent high-temperature processing. Our previous structural studies[12–16] showed that the oxide layer progressively breaks up during high-temperature heat treatments. In order to investigate this behavior in more detail and to relate it directly to device performance, we have carried out a detailed comparison of electrical results on polysilicon emitter transistors with structural results from high-resolution transmission electron microscopy. The thermal stability of the oxide layer was investigated by carrying out preanneals at temperatures between 800 and 1100 °C immediately after polysilicon deposition but prior to emitter implant and drive-in at 900 °C. In this way, the oxide layer could be subjected to a high-temperature heat treatment without significantly affecting the rest of the emitter structure, and therefore changes in gain could be correlated directly with changes in the structure of the oxide layer.

II. EXPERIMENT

Two types of polysilicon emitter bipolar transistor were fabricated using identical standard processing up to the surface treatment of the (100) silicon slices prior to polysilicon deposition. For the first type (RCA treatment), the slice was given a dip in buffered HF followed by an RCA clean[4] (NH_4OH:H_2O_2:H_2O as 1:1:5 for 10 min; HCl:H_2O_2:H_2O as 1:1:6 for 10 min) to produce a surface oxide layer of nominal thickness ~15 Å. For the second type (HF treatment), the slice was given a dip in buffered HF with the intention of

 0021-8979/87/010225-09$02.40 © 1986 American Institute of Physics

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 632

removing any native oxide which might have been present.[4] These surface layer treatments were immediately followed by the deposition of a ~0.4-μm-thick layer of undoped LPCVD polysilicon at ~610 °C. One of each type of slice was given a 10 min preanneal in dry nitrogen at one of several temperatures in the range 800–1100 °C in order to breakup the interfacial oxide layer to different extents. The emitters of the transistors were then formed by implanting 1×10^{16}-cm^{-2}, 70-keV As^+ into the polysilicon and driving it in at 900 °C in wet oxygen for 30 min. The active emitter area was 20 μm by 18 μm. The aluminium contact to the emitter was remote from the active emitter area, being made a via polysilicon track 60 μm in length.

Detailed electrical measurements were carried out on the transistors including measurements of collector and base current as a function of base/emitter voltage. Unpatterned check slices which had been given identical base and emitter processing were used for the TEM studies. Cross sections were prepared corresponding to the $(01\bar{1})$ plane of the (100) single-crystal silicon slice by grinding, polishing, and argon ion-beam thinning. These were examined using a JEOL 200CX TEM with the electron beam aligned parallel to the $[01\bar{1}]$ direction of the single-crystal silicon. High-resolution electron microscopy (HREM) was performed to give lattice images using conventional procedures. All of the specimens examined (except for that of Fig. 6) corresponded to the final processed slices and so the observed TEM structures could be directly correlated with the device results. Large numbers of micrographs were taken of each specimen in order to determine its characteristic structure. The micrographs of Figs. 1–6 are typical of those obtained, but do not show all the features described in the text.

III. TEM RESULTS

For specimens given in an RCA treatment, a 900 °C preanneal and full processing (As^+ implant and 900 °C drive-in), cross-section micrographs of the interface region showed the single-crystal silicon with the characteristic (111), $(\bar{1}11)$, and (200) lattice planes clearly delineated (Fig. 1). The individual polysilicon grains only showed lattice planes when the grains were favorably oriented with respect to the electron beam, e.g., as at A. When such planes were observed in the individual polysilicon grains, they were in general not aligned with the planes present in the silicon.

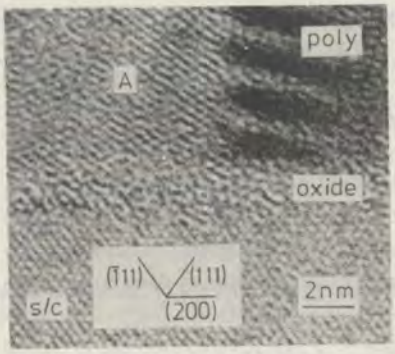


FIG. 1. HREM micrograph of a fully processed specimen given an RCA treatment and a 900 °C preanneal. Note the interfacial oxide layer and the polysilicon grain A. s/c—single-crystal silicon; oxide—interfacial oxide layer; poly—polysilicon layer.

The interfacial oxide layer was seen as a thin bright band separating the silicon from the polysilicon. The band exhibited a random wormlike contrast in the lattice-image micrographs, this being the characteristic behavior for an amorphous material. The oxide layer was continuous and uniform with a thickness of ~14 ± 1 Å and it followed the undulations of the surface of the silicon slice with an "amplitude" of ~6–9 Å (perpendicular to the interface) and a "wavelength" of ~200–400 Å (parallel to the interface). The good uniformity of the oxide layer indicates that its structure had been little changed by the subsequent 900 °C preanneal and 900 °C drive-in, and so Fig. 1 shows it in effect as it was when initially produced by the RCA treatment. The directly measured oxide layer thickness of 14 Å is in good agreement with the values of 13–15 Å obtained from ellipsometry,[17] and 10–15 Å obtained from Auger spectroscopy profiles,[18] for silicon slices given an RCA treatment. The TEM results are summarized in Table I.

For specimens given an RCA treatment, a 950 °C preanneal and full processing, the interfacial oxide layer was nonuniform (Fig. 2). Some regions had increased in thickness to a maximum of 19 Å, some regions were thinner, and in some regions holes (B) were present in the layer. Small areas (C) of the polysilicon had been regrown through such holes to become single-crystal material with the same crystallographic orientation as the underlying silicon, this being deduced from the micrograph contrast and the alignment of

TABLE I. Summary of results.

Silicon slice surface treatment	Preanneal temperature (°C)	Interfacial oxide layer: Geometry	Interfacial oxide layer: Maximum thickness (Å)	Polysilicon: Fraction of interface regrown	Polysilicon: Grain size (Å) Surface	Polysilicon: Grain size (Å) Bulk	Transistor gain
RCA	900	continuous	~14	0	1200	400	1400
RCA	950	broken up	~19	0.3	1200	600	350
RCA	1000	broken up	~22	0.7	1200	1000	90
RCA	1100	balled up	~55	0.97	2000	1000	40
HF	None	broken up	~8	0.2	1200	400	240
HF	1000	balled up	~50	0.95	1200	1000	50

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 633

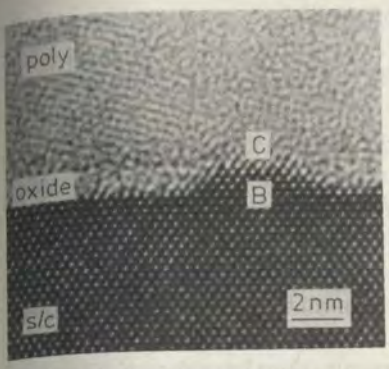


FIG. 2. HREM micrograph of a fully processed specimen given an RCA treatment and a 950 °C preanneal. Note the hole B in the interfacial oxide layer and the regrown polysilicon area C.

the lattice planes. The regrown areas had often spread out laterally in the polysilicon and so were sometimes wider than the holes in the oxide layer through which they had been grown. Individual areas extended typically 35 Å along the interface and 35 Å into the polysilicon. When the individual lengths of the regrown areas of the polysilicon were measured along the interface and added together, this represented ~0.3 of the total length of the interface.

For the specimens given an RCA treatment, a 1000 °C preanneal and full processing (micrograph not included), the breakup of the interfacial oxide layer was more pronounced. Some regions had increased in thickness to a maximum of 22 Å, and more holes were present in the layer. However, no substantial "balling-up" of the layer had occurred. Individual regrown areas of the polysilicon extended typically 120 Å along the interface and 40 Å into the polysilicon. The fraction of the length of the interface that had regrown was ~0.7.

For specimens given an RCA treatment, a 1100 °C preanneal and full processing, the breakup of the interfacial oxide layer was almost complete (Fig. 3), the layer now consisting mainly of balls of oxide (D) up to 55 Å in diameter or worms ~30×100 Å across. Individual regrown areas of the polysilicon overlapped to give almost complete coverage of the interface with the regrowth extending typically 170 Å into the polysilicon. The large difference between the



FIG. 3. HREM micrograph of a fully processed specimen given an RCA treatment and an 1100 °C preanneal. Note the balled-up oxide D.

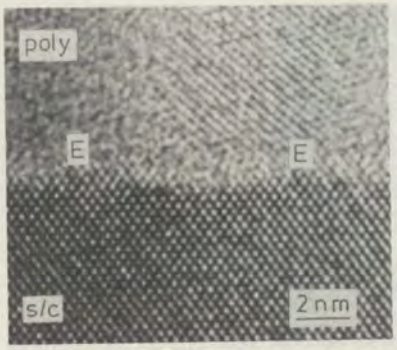


FIG. 4. HREM micrograph of a fully processed specimen given an HF treatment and no preanneal. Note the regrown polysilicon areas E.

oxide layer in the specimens given the 950, 1000, and 1100 °C preanneals, and the specimen given the 900 °C preanneal, indicates that the breakup of the oxide in the 950, 1000, and 1100 °C specimens occurred during the preanneals.

For specimens given an HF treatment, no preanneal and full processing, an interfacial oxide layer was present but it was not continuous (Fig. 4). In some regions it was up to a maximum thickness of 8 Å, in other regions it was thinner, and in some regions there were holes. Small areas (E) of the polysilicon had regrown through the holes, individual areas extending typically 25 Å along the interface and 25 Å into the polysilicon. The fraction of the interface that had been regrown was ~0.2.

For specimens given an HF treatment, a 1000 °C preanneal and full processing, the breakup of the interfacial oxide layer was severe, the layer consisting mainly of balls of oxide up to 50 Å diameter (Fig. 5). Individual areas (F) of regrown polysilicon extended typically 300 Å along the interface and 150 Å into the polysilicon, each area often covering several oxide balls (G) and holes (H). The fraction of the length of the interface that had regrown was ~0.95. The

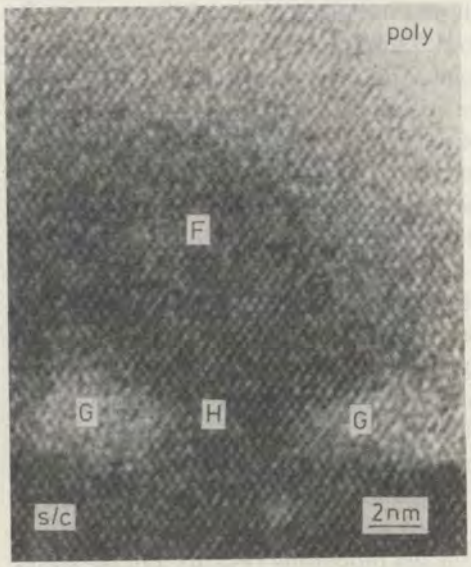


FIG. 5. HREM micrograph of a fully processed specimen given an HF treatment and a 1000 °C preanneal. Note the regrown polysilicon area F, the balled-up oxide G, and the holes in between H.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 634

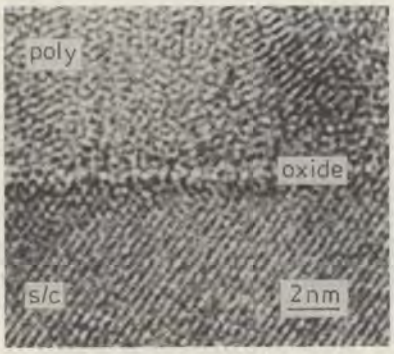


FIG. 6. HREM micrograph of a specimen given an HF treatment, but no preanneal, implant, or drive-in.

large difference between the oxide layers in the specimen given the 1000 °C preanneal, and the specimen given no preanneal, indicates that the "balling-up" of the oxide in the 1000 °C specimen mainly occurred during the preanneal.

In order to obtain further information concerning the interfacial oxide layer that was present after an HF treatment, additional specimens were examined that corresponded to an HF treatment and polysilicon deposition, but no further processing. A white band was seen at the interface (Fig. 6) but this was extremely thin, and so the oxide thickness could not be directly measured accurately. The interface was very uniform, and in a few places adjacent to the interface small areas of the polysilicon were of the same crystallographic orientation as the single-crystal silicon. These areas could have arisen by either direct epitaxial growth through holes initially present in the oxide layer or regrowth through holes created initially present in the oxide layer during the deposition of the polysilicon at 610 °C. The results indicate that an extremely thin and mainly continuous oxide layer was present immediately after the polysilicon deposition. The results also indicate that for the specimens that were given an HF treatment, no preanneal and full processing, the significant breakup of the oxide layer that was observed occurred during the 900 °C drive-in.

In our previous TEM studies[13] of the breakup of an analogous interfacial oxide layer formed by a sulfuric acid/hydrogen peroxide surface chemical treatment, it was shown that the volume of the oxide present after complete balling-up of the layer corresponded closely to the volume of the ~10-Å-thick uniform oxide layer that was present initially. This suggested that the breakup of the oxide layer occurred mainly by a local diffusion mechanism. Application of this principle to the present work gives the following results. For the specimens given an HF treatment, a 1000 °C preanneal, and full processing, the balled-up oxide corresponded to an approximately planar array of spheres of mean diameter 40 Å and mean spacing 100 Å. Calculation shows that this is equivalent to an initial uniform oxide layer of thickness ~4 Å, a result which is consistent with the TEM observations for the specimens given an HF treatment and polysilicon deposition, but no further processing (Fig. 6). For the specimens given an RCA treatment, an 1100 °C preanneal, and full processing, a similar calculation showed that the balled-up oxide layer is equivalent to an initial uniform oxide layer of thickness ~15 Å. This result is in good agreement with the TEM value of ~14 Å for the specimens given an RCA treatment, preannealed at 900 °C, and fully processed (Fig. 1).

Present evidence indicates that there are genuine differences in the thickness of oxide layers formed on silicon slices by the HF treatments of different organizations. These differences may depend on the HF solution, the dip procedure, or the elapsed time prior to the subsequent deposition of the polysilicon. Our TEM examinations of silicon specimens given nominally the same HF treatment, and subsequently processed, by different organizations have shown differences in the behavior of the oxide layer which we attribute to differences in its initial thickness.[19] Furthermore, measurements by different organizations of the thickness of oxide layers present on silicon slices after HF treatments have given markedly different values, e.g., 8–9 Å from ellipsometry[17] and ~3 Å from Auger spectroscopy.[18]

The above TEM results show that the interfacial oxide layers break up faster when they are heated at higher temperatures, and when they are initially thinner. Thus, for the oxide layers arising from the HF treatment, which were initially ~4 Å thick, breakup had already occurred during a heat treatment at 900 °C (drive-in), while for oxide layers arising from the RCA treatment, which were initially ~14 Å thick, breakup only began to occur during a treatment at 950 °C (preanneal). These present results are consistent with our previous TEM results[14,16] which indicated that an initially ~10-Å-thick oxide layer was just beginning to break up at 900 °C. The results are also consistent with the TEM results of Schaber *et al.*[20] which showed that for an oxide layer ~30 Å thick, breakup did not occur at 950 °C. However, some caution is required when making comparisons of such results obtained by different organizations because the precise behavior depends on a number of factors, e.g., it has been shown that such oxide layers break up more rapidly when the arsenic concentration is higher.[15,16]

Grain sizes for the polysilicon layers in fully processed specimens were determined from TEM cross-section micrographs taken at lower magnifications. For individual specimens, the grain size near the polysilicon layer surface was in general larger than in the remainder of the polysilicon layer. This was because the upper portion of the polysilicon layer had been made amorphous by the arsenic implant. Consequently, during the drive-in at 900 °C, larger grains had formed in this portion of the layer by regrowth from amorphous material than in the remainder of the layer by regrowth from fine grains. For the grains in the main portion of the layer, i.e., away from the layer surface, the results showed that the mean grain size was typically 400, 600, 1000, and 1000 Å for preanneal temperatures of 900, 950, 1000, and 1100 °C, respectively. Conversely, for the grains near the surface, the mean grain size was typically 1200 Å for all of these preanneal temperatures, except for 1100 °C when it was typically 2000 Å.

IV. ELECTRICAL RESULTS

Figures 7 and 8 show Gummel plots for the RCA- and HF-treated devices, respectively, that were given preanneals

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 635

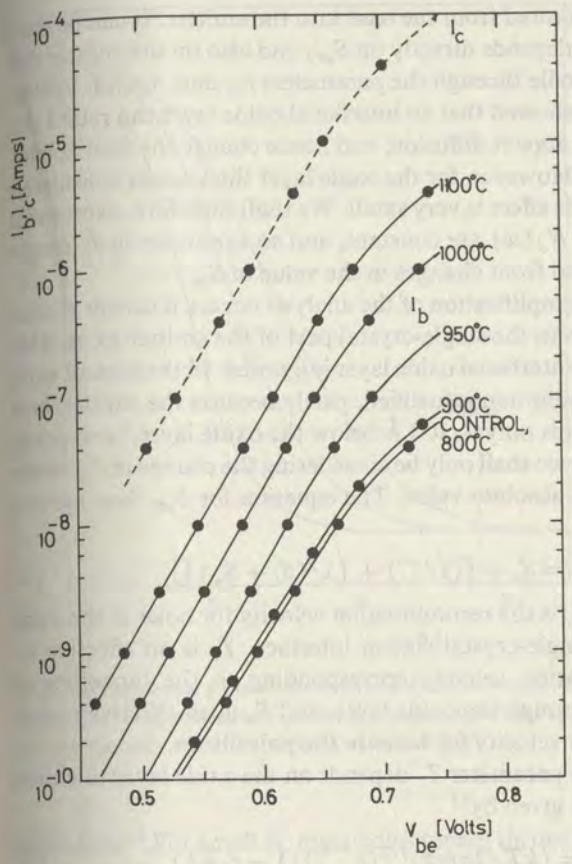


FIG. 7. Gummel plots of devices given an RCA treatment and preanneals in the range 800–1100 °C.

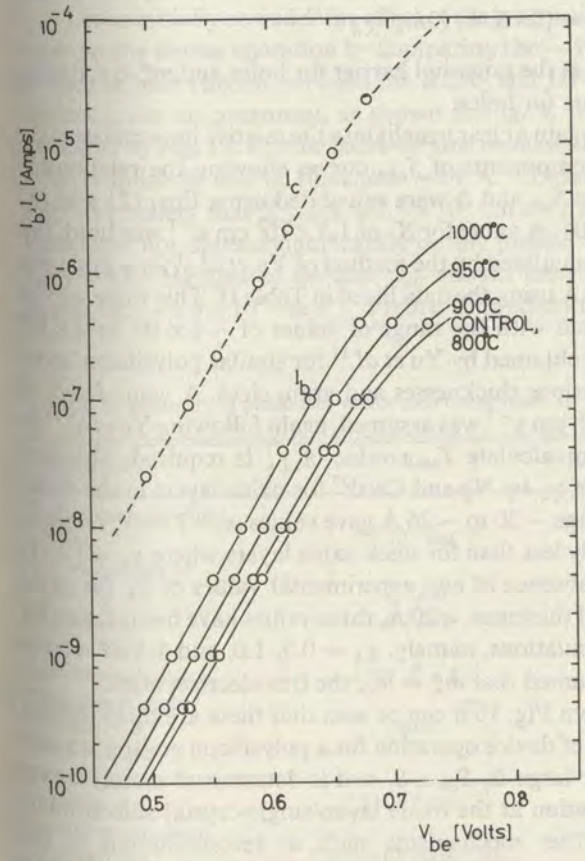


FIG. 8. Gummel plots of devices given an HF treatment and preanneals in the range 800–1000 °C.

between 800 and 1100 °C. The collector characteristics for both the HF- and RCA-treated devices are all ideal over at least three orders of collector current (60 mV/decade). Similarly the base characteristics of the HF-treated devices are also ideal, although the base characteristics of the RCA-treated devices are slightly nonideal (70 mV/decade) corresponding to an ideality parameter of $n = 1.12$. To show the effect of the preanneal on the electrical characteristics of the RCA- and HF-treated devices, I_{b0} and I_{c0} have been plotted against the preanneal temperature in Fig. 9. Because of the nonideality of the base current characteristics of the RCA-treated devices, the values of I_{b0} were calculated from the ideal collector characteristics and maximum current gain β_{max} using the relationship[21] $I_{b0} = I_{c0}/\beta_{max}$.

For the RCA-treated devices, Figs. 7 and 9 show that the collector current is relatively unaffected by a preanneal. Conversely, the base current is strongly affected with the electrical results falling into two distinct regimes. Below ~900 °C, the base current is approximately constant. In this regime, the 900 °C drive-in is the dominant process as far as the interfacial oxide layer is concerned. Above ~900 °C, the base current increases rapidly with preanneal temperature. In this regime, the preanneal is the dominant process as far as the oxide layer is concerned. For example, the base current increases after a 950 °C preanneal by ~6×, after a 1000 °C preanneal by ~28×, and after a 1100 °C preanneal by ~100×. This trend correlates well with the TEM results. Thus, after a 900 °C preanneal, the oxide layer is uniform and continuous (Fig. 1) and the gain is not significantly affected by the preanneal. After a 950 °C preanneal, the oxide layer begins to break up (Fig. 2) and a large increase in

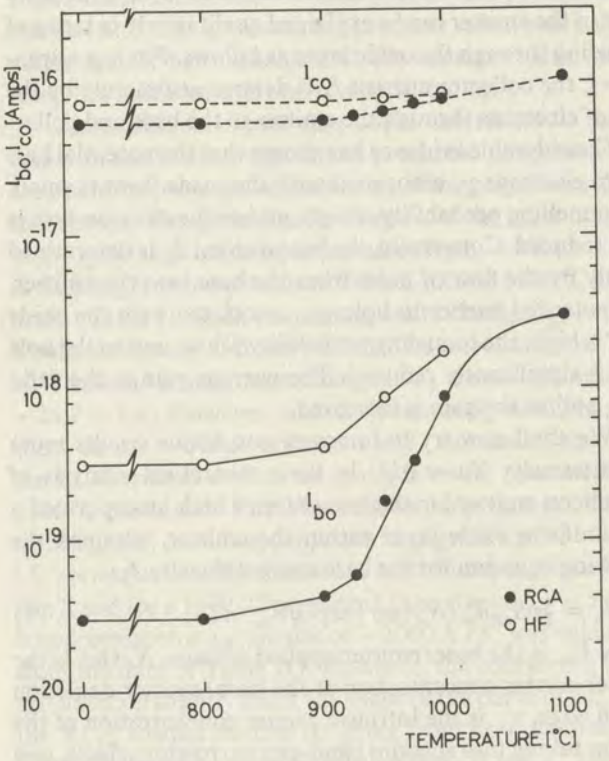


FIG. 9. I_{b0} and I_{c0} vs preanneal temperature for devices given RCA and HF treatments.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 636

the base current occurs. As the preanneal temperature is further increased, the breaking up of the oxide layer continues (Fig. 3) and this corresponds to a continuing increase in the base current. Over the preanneal temperature range of 800–1100 °C, the transistor gain decreases from ~1800 to ~40.

For the HF-treated devices, Figs. 8 and 9 show a similar but less pronounced trend to that of the RCA-treated devices. Thus, the collector current is relatively unaffected by a preanneal. For the base current, after a preanneal below ~900 °C, no significant increase occurs. After a 950 °C preanneal it increases ~3×, and after a 1000 °C preanneal it increases ~5.5×. This increase in the base current again correlates with a progressive increase in the breaking up of the interfacial oxide layer (Figs. 4 and 5), although it should be noted that some polysilicon grain growth also occurs (Table I). Over the pre-anneal temperature range of 800–1000 °C, the transistor gain decreases from ~240 to ~50.

A comparison of the RCA- and HF-treated devices shows the following. After a preanneal below ~900 °C, the base current for the RCA-treated devices is ~10× less than for the HF-treated devices. This difference correlates with the former devices having a relatively thick, continuous oxide layer, and the latter devices having a thin broken-up oxide layer. After a preanneal at 1000 °C, the base current for the RCA-treated devices is ~2× less than for the HF-treated devices. This difference correlates with the former devices having a broken-up oxide layer, and the latter devices having a completely balled-up oxide layer.

V. INTERPRETATION OF ELECTRICAL RESULTS

The larger gain that occurs for polysilicon emitter bipolar transistors when a thin interfacial oxide layer is incorporated in the emitter can be explained qualitatively in terms of tunneling through the oxide layer as follows. For *n-p-n* transistors, the collector current I_c is determined mainly by the flow of electrons through the emitter to the base and collector. Considerable evidence has shown that the potential barrier to electrons χ_e associated with the oxide layer is small, the tunnelling probability is high, and so the electron flow is little reduced. Conversely, the base current I_b is determined mainly by the flow of holes from the base into the emitter. The potential barrier to holes χ_h associated with the oxide layer is high, the tunneling probability is low, and so the hole flow is significantly reduced. The current gain is the ratio I_c/I_b and so the gain is increased.

We shall now try to interpret our device results more quantitatively. Yu *et al.*[11] in their theoretical analysis of polysilicon emitter bipolar transistors which incorporated a thin uniform oxide layer within the emitter, obtained the following equation for the base current density J_b:

$$J_b = [qS_{\text{eff}} n_{ie}^2/N_d(w)]\exp(qV_{be}/kT), \tag{1}$$

where V_{be} is the base/emitter applied voltage, $N_d(w)$ is the emitter carrier concentration at the base/emitter depletion region edge, n_{ie} is the intrinsic carrier concentration of the emitter taking into account band-gap narrowing effects, and S_{eff} is an effective recombination velocity which incorporates the various recombination processes occurring for the holes injected from the base into the emitter. It can be seen that J_b depends directly on S_{eff}, and also on the emitter dopant profile through the parameters n_{ie} and $N_d(w)$. Earlier work[22] showed that an interfacial oxide layer can retard the emitter dopant diffusion, and hence change the final dopant profile. However, for the oxide layer thicknesses considered here, this effect is very small. We shall therefore assume that n_{ie} and $N_d(w)$ are constant, and so variations in J_b can be predicted from changes in the value of S_{eff}.

A simplification of the analysis occurs if carrier recombination in the single-crystal part of the emitter located below the interfacial oxide layer is ignored. In the present work this assumption is justified, partly because the emitter/base junction is only ~100 Å below the oxide layer,[1] and partly because we shall only be considering the change in J_b, rather than its absolute value. The equation for S_{eff} then reduces to[11]

$$S_{\text{eff}} = S_i + \{(1/T_i) + [1/(S_i + S_p)]\}^{-1}, \tag{2}$$

where S_i is the recombination velocity for holes at the oxide layer/single-crystal silicon interface, T_i is an effective recombination velocity corresponding to the tunnelling of holes through the oxide layer, and S_p is an effective recombination velocity for holes in the polysilicon.

The parameter T_i depends on the oxide layer thickness Δ and is given by[11]

$$T_i = (kT/2\pi m_h^*)^{1/2}(e^{-b_h})(1 - c_h kT)^{-1}, \tag{3}$$

where

$$b_h = (4\pi\Delta/h)(2m_h^*\chi_h)^{1/2},$$
$$c_h = (2\pi\Delta/h)(2m_h^*/\chi_h)^{1/2},$$

and χ_h is the potential barrier for holes and m_h^* is the effective mass for holes.

To gain a clear insight into the relative importance of the three components of S_{eff}, curves showing the relationship between S_{eff} and Δ were calculated using Eqs. (2) and (3) (Fig. 10). A value for S_p of 1.5×10^5 cm s^{-1} was used, this being calculated by the method of Yu *et al.*[11] for a grain size of 1000 Å using the data listed in Table II. This value may be compared with the range of values of $\sim8\times10^4$ to 2×10^5 cm s^{-1} obtained by Yu *et al.*[11] for similar polysilicon layers with various thicknesses and grain sizes. A value for S_i of 1.5×10^3 cm s^{-1} was assumed, again following Yu *et al.*[11] In order to calculate T_i, a value of χ_h is required. Measurements of χ_h by Ng and Card[27] for oxide layers in the thickness range ~20 to ~26 Å gave values of ~1.0 eV, i.e., considerably less than for thick oxide layers where $\chi_h \simeq 3.0$ eV. In the absence of any experimental values of χ_h for oxide layers of thickness < 20 Å, three values have been chosen for our calculations, namely, $\chi_h = 0.5$, 1.0, and 1.5 eV. It was also assumed that $m_h^* = m_0$, the free electron mass.

From Fig. 10 it can be seen that there are three distinct regions of device operation for a polysilicon emitter transistor. For large Δ, $S_{\text{eff}} \simeq S_i$ and is determined mainly by recombination at the oxide layer/single-crystal silicon interface. Other mechanisms such as recombination in the single-crystal part of the emitter and recombination in the base may also contribute to the base current in this region of

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 637

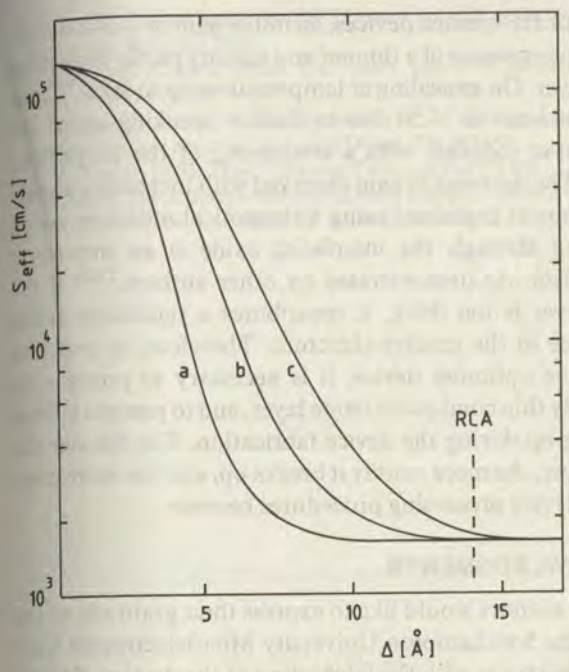


FIG. 10. Theoretical plots of effective recombination velocity (S_{eff}) vs interfacial oxide layer thickness (Δ) for hole potential barriers (χ_h) associated with the oxide of (a) 1.5, (b) 1.0, and (c) 0.5 eV.

operation.[22] For small Δ, most holes tunnel through the oxide layer, $S_{eff} \simeq S_p$ and is determined mainly by transport and recombination in the polysilicon. For intermediate Δ, tunneling through the oxide layer is the dominant mechanism, and S_{eff} is extremely sensitive to small changes in Δ.

Let us initially consider the effect of the interfacial oxide layer on the device operation by comparing the $\sim 10\times$ difference in base current between the RCA- and HF-treated devices given no preanneal, as shown in Fig. 9, with that predicted by Fig. 10. For the RCA-treated devices, the oxide layer is uniform and of thickness ~ 14 Å. Consequently, Fig. 10 predicts that $S_{eff} \simeq S_i \simeq 1.5\times 10^3$ cm s^{-1}, and this value does not depend significantly on the precise value of χ_h. For the HF-treated devices, S_{eff} should then be $10\times$ larger, i.e., $\sim 1.5\times 10^4$ cm s^{-1}. Figure 10 predicts that this would correspond to a uniform oxide layer of thickness ~ 4 to ~ 8 Å, the range arising because of the uncertainty in the value of χ_h. This prediction is in reasonable agreement with the observation of a uniform oxide layer of thickness ~ 4 Å immediately after polysilicon deposition, and a broken up oxide layer of maximum thickness ~ 8 Å after no preanneal, but a 900 °C drive-in.

TABLE II. Summary of parameter values used to calculate S_p.

Parameter	Value	Reference
Doping concentration $N_d(w)$	5×10^{19} cm^{-3}	1
Diffusion length of holes, L_p	0.7 μm	21
Grain size	0.1 μm	
Hole mobility at grain boundary, μ_{gb}	2 cm^2 V^{-1} s^{-1}	23
Thickness of grain boundary, δ	10 Å	24
Grain boundary interface recombination velocity, S_{gb}	7.5×10^4 cm s^{-1}	25,26
Length of polysilicon emitter track	Large with respect to L_p	

Let us now consider the effect of the interfacial oxide layer on the devices by comparing the $\sim 100\times$ difference in base current between the RCA-treated devices given no preanneal with those given a 1100 °C preanneal. For no preanneal, the oxide layer is continuous and uniform, and $S_{eff} \simeq S_i \simeq 1.5\times 10^3$ cm s^{-1}. For a 1100 °C preanneal, the oxide layer is balled-up. If it is assumed that such a balled-up oxide layer has little effect on the flow of holes through the interface, then S_{eff} will correspond to the value for $\Delta = 0$, i.e., to $S_{eff} \simeq S_p \simeq 1.5\times 10^5$ cm s^{-1} (Fig. 10). Hence, an increase in base current of $\sim 100\times$ is predicted, in good agreement with the experimental value of $\sim 100\times$ (Fig. 9).

With regard to the effect of the interfacial oxide layer on the RCA-treated devices as the preanneal temperature is progressively increased, comparison of the device results and TEM observations suggest the following behaviors. On going from no preanneal to ~ 800 °C, I_b is constant and is determined mainly by recombination of holes at the oxide-layer/single-crystal silicon interface (mechanism A). From ~ 800 to ~ 925 °C, I_b increases slowly and is determined by mechanism A, together with tunnelling of holes through the thinner portions of the oxide layer (mechanism B). From ~ 925 to ~ 975 °C, I_b increases rapidly and is determined by mechanism B, together with an increasing contribution from the flow of holes through the penetrated regions of the oxide layer and the transport properties in the polysilicon layer (mechanism C). From ~ 975 to ~ 1100 °C, I_b continues to increase and then approaches a constant value, and is determined mainly by mechanism C, the constant value arising when the oxide layer is completely balled up.

An additional factor that could contribute to the increase in the base current as the preanneal temperature increases is the progressive coarsening of the polysilicon grains that occurs. The grain size for the part of the layer adjacent to the interfacial oxide layer increases from ~ 400 to ~ 1000 Å on going from no preanneal to both 1000 to 1100 °C preanneals (Table I). Such coarsening could change S_p, and this would change S_{eff} if Δ were small. The calculations of Yu *et al.*[11] indicate a maximum change in S_p due to this effect of ~ 2 or $\sim 3\times$. However, such a change in S_p would have no effect on the $\sim 100\times$ difference in I_b calculated above for the RCA-treated devices on going from no preanneal to a 1100 °C preanneal. This is because for no preanneal, the value of $S_{eff} \simeq S_i$ used is appropriate for a grain size of ~ 400 Å (S_i corresponds to mechanism A and is independent of grain size), and for a 1100 °C preanneal, the value of $S_{eff} \simeq S_p$ used is appropriate for a grain size of ~ 1000 Å (S_p was calculated from the data of Table II). However, the coarsening of the grains could slightly affect the shape of the curve of Fig. 9 for the RCA-treated devices on going from no preanneal to a 1100 °C preanneal because mechanisms B and C both depend on the transport properties in the polysilicon layer.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 638

Let us now consider the effect of the interfacial oxide layer on the devices by comparing the $\sim 5.5\times$ difference in base current between the HF-treated devices given no preanneal with those given a 1000 °C preanneal. It is more difficult to calculate the behavior in this case because the oxide layer is already broken up for no preanneal. Comparison of the device results and TEM observations suggests that on going from no preanneal to ~ 900 °C, I_b is determined by mechanisms B and C; and on going from ~ 900 to ~ 1000 °C, I_b is determined mainly by mechanism C.

In this case the progressive coarsening of the grains could make a contribution to the increase in I_b of $\sim 5.5\times$. This is because over the whole range from no preanneal to a 1000 °C preanneal, the transport properties in the polysilicon layer play a role. The experimental results of Ning and Isaac,[5] performed on what were in effect HF-treated devices, showed that on going from bipolar silicon transistors with polysilicon emitters to analogous devices with single-crystal emitters, the device gain decreased by $\sim 3\times$. These results suggest that in our case, i.e., for polysilicon emitter bipolar transistors given an HF treatment, some of the increse in I_b of $\sim 5.5\times$ is due to the coarsening of the grains and some to the continued breaking up of the oxide layer.

From the above results it might be expected that all devices with oxide layers thicker than ~ 10 Å would give the same best performance. However, although all such devices would possess the same lowest value of S_{eff}, and as a result J_b, the oxide layers introduce a series emitter resistance which increases rapidly with increasing oxide thickness.[11,28] This is detrimental to device and circuit performance, and so in practice a compromise between increased gain and increased emitter resistance needs to be made.

As far as the breakup and eventual balling-up of the thin interfacial oxide layer is concerned, the main driving force is considered to be a significant reduction in the total amount of energy associated with the oxide layer/silicon interface[13,15] (analogous to the well-known Oswald ripening effect that occurs in metallurgical systems). Thus, on going from a continuous uniform oxide layer to a balled-up oxide layer, the interfacial area decreases by typically ~ 5 to $\sim 10\times$ (deduced from TEM examinations). However, the presence of the As dopant can also have an effect. For example, in our work[15,16] on similar specimens post-annealed at 900 °C, the breakup of the oxide layer was found to proceed more rapidly when the As^+ implantation dose was increased from 5×10^{15} to 2×10^{16} cm^{-2}. This result suggests[15] that the As dopant can either enhance the local oxide diffusion rate or change the oxide/silicon interfacial energy.

VI. CONCLUSIONS

A detailed comparison of the electrical and TEM results has revealed that there is a strong correlation between the high gain of polysilicon emitter bipolar transistors and the existence of a continuous, thin, interfacial oxide layer between the single-crystal silicon and the polysilicon. For RCA-treated devices, an initial high gain of ~ 1400 occurs due to the presence of an ~ 14-Å-thick oxide layer. On annealing at temperatures up to 1100 °C, the gain progressively decreased to ~ 40 due mainly to the breaking up of the oxide layer. For HF-treated devices, an initial gain of ~ 240 occurs due to the presence of a thinner and already partly broken up oxide layer. On annealing at temperatures up to 1000 °C, the gain decreases to ~ 50 due to further breaking up of the oxide layer together with a coarsening of the polysilicon grains. The decrease in gain observed with increasing anneal temperture is explained using a theoretical model in which tunneling through the interfacial oxide is an important mechanism. As demonstrated by other authors,[11,28] if the oxide layer is too thick, it contributes a significant series resistance to the emitter structure. Therefore, in order to obtain the optimum device, it is necessary to produce an extremely thin continuous oxide layer, and to prevent it from breaking up during the device fabrication. The thinner the oxide layer, the more readily it breaks up, and the more critical the device processing procedures become.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the staff of the Southampton University Microelectronics Centre for assistance with the fabrication of the devices. Special thanks are due to Dr. A. Cuthbertson and Dr. E. F. Chor of Southampton University and Dr. P. C. Hunt and Dr. M. C. Wilson of Plessey, Caswell for many stimulating and helpful discussions. The work was supported by the Science and Engineering Research Council, UK, and the Alvey Directorate, UK.

[1] A. Cuthbertson and P. Ashburn, Int. Electron Devices Meeting Techical Digest, 749 (1984).
[2] T. H. Ning, R. D. Isaac, P. M. Solomon, D. D. Tang, H. N. Yu, G. C. Feth, and S. K. Wiedmann, IEEE Trans. Electron Devices **ED-28**, 1010 (1981).
[3] T. Sakai, Y. Yamamoto, Y. Kobayashi, K. Kawarada, V. Inabe, T. Hayashi, and H. Miyanaga, IEEE J. Solid State Circuits **SC-16**, 424 (1981).
[4] B. Soerowirdjo and P. Ashburn, Solid State Electron. **26**, 495 (1983).
[5] T. H. Ning and R. D. Isaac, IEEE Trans. Electron Devices **ED-27**, 2051 (1980).
[6] B. Soerowirdjo, P. Ashburn, and A. Cuthbertson, International Electron Devices Meeting Technical Digest, 688 (1982).
[7] P. Ashburn and B. Soerowirdjo, IEEE Trans. Electron Devices **ED-31**, 853 (1984).
[8] H. C. de Graff and J. C. de Groot, IEEE Trans. Electron Devices **ED-26**, 1771 (1979).
[9] A. A. Eltoukhy and D. J. Roulston, IEEE Trans. Electron Devices **ED-29**, 961 (1982).
[10] A. A. Eltoukhy and D. J. Roulston, IEEE Trans. Electron Devices **ED-29**, 1862 (1982).
[11] Z. Yu, B. Ricco and R. W. Dutton, IEEE Trans. Electron Devices **ED-31**, 773 (1984).
[12] M. C. Wilson, P. Ashburn, B. Soerowirdjo, G. R. Booker, and P. Ward, J. Phys. (Paris) Colloq. **43**, Suppl. 10, C1-253 (1982).
[13] A. Albu-Yaron, J. C. Barry, and G. R. Booker, in *Proceedings of the 8th European Congress on Electron Microscopy*, Budapest (Programme Committee, Congress Bureau MOTESZ, Budapest, Hungary, 1984), Vol. 1, p. 521.
[14] A. Albu-Yaron, J. C. Barry, N. Jorgensen, G. R. Booker, P. Ashburn, and P. Ward, British Association for Crystal Growth Annual Conference, Lancaster (1984) (unpublished).
[15] N. Jorgensen, J. C. Barry, G. R. Booker, P. Ashburn, G. R. Wolstenholme, M. C. Wilson, and P. C. Hunt, Inst. Phys. Conf. Ser. **76**, 471 (1985).
[16] M. C. Wilson, N. Jorgensen, G. R. Booker, and P. C. Hunt, *Proceedings of the 5th Symposium of VLSI Technology*, Kobe, Japan (IEEE and Japanese Society of Applied Physics, Japan, 1985), p. 46.
[17] R. C. Henderson, J. Electrochem. Soc. **119**, 772 (1982).

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 639

[18]J. E. Duffill, *Proceedings of the 12th European Solid State Device Research Conference,* Munich (1982).
[19]N. Jorgensen, A. Albu-Yaron, J. C. Barry, and G. R. Booker (unpublished results) (1984).
[20]H. Schaber, B. Benno, L. Treikaguer and A. W. Wieder, International Electron Devices Meeting Technical Digest, 738 (1984).
[21]A. W. Wieder, IEEE Trans. Electron Devices **ED-27**, 1402 (1980).
[22]P. Ashburn and B. Soerowirdjo, Solid State Electron. **24**, 475 (1981).
[23]D. M. Kim, A. N. Khondker, R. R. Shah, and D. C. Crosthwaite Jr., IEEE Electron Device Lett. **EDL-3**, 141 (1982).
[24]M. M. Mandurah, K. C. Saraswat, and T. I. Kamins, IEEE Trans. Electron Devices **ED-28**, 1163 (1981).
[25]J. Oualid, H. Anzil, J. P. Crest, J. Dugas, G. Mathian, M. Zemt, and S. Martinazzi, J. Phys. (Paris) Colloq. **43**, Suppl. 10, C1-313 (1982).
[26]C. H. Seager, J. Appl. Phys. **53**, 5968 (1982).
[27]K. K. Ng and H. C. Card, J. Appl. Phys. **51**, 2153 (1980).
[28]E. F. Chor, P. Ashburn, and A. Brunnschweiler, IEEE Electron Device Lett. **EDL-6**, 516 (1985).

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 640

ATTACHMENT K-3

Journal of Applied Physics

RESEARCH ARTICLE | JANUARY 01 1987

An investigation of the thermal stability of the interfacial oxide in polycrystalline silicon emitter bipolar transistors by comparing device results with high-resolution electron microscopy observations

G. R. Wolstenholme; N. Jorgensen; P. Ashburn; G. R. Booker

Check for updates

J. Appl. Phys. 61, 225–233 (1987)
https://doi.org/10.1063/1.338861

View Online Export Citation

Articles You May Be Interested In

Effect of surface treatment on dopant diffusion in polycrystalline silicon capped shallow junction bipolar transistors

Appl. Phys. Lett. (November 1985)

Investigations of evaporated silicon *p-n* junctions and their application to junction field-effect transistors

J. Appl. Phys. (February 1990)

Early stage evolution kinetics of the polysilicon/single-crystal silicon interfacial oxide upon annealing

J. Appl. Phys. (January 1991)

04 August 2025 00:02:09



AIP Publishing

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

An investigation of the thermal stability of the interfacial oxide in polycrystalline silicon emitter bipolar transistors by comparing device results with high-resolution electron microscopy observations

G. R. Wolstenholme
Department of Electronics and Information Engineering, University of Southampton, Southampton, England

N. Jorgensen
Department of Metallurgy and Science of Materials, University of Oxford, Oxford, England

P. Ashburn
Department of Electronics and Information Engineering, University of Southampton, Southampton, England

G. R. Booker
Department of Metallurgy and Science of Materials, University of Oxford, Oxford, England

(Received 6 May 1986; accepted for publication 9 September 1986)

A comparison is made between the results of high-resolution electron microscope observations and the electrical characteristics of polycrystalline silicon emitter bipolar transistors. Devices are fabricated with and without a deliberately grown interfacial oxide layer, and the thermal stability of this oxide layer is investigated by carrying out a preanneal at temperatures between 800 and 1100 °C after polysilicon deposition, but prior to emitter implant and 900 °C drive-in. The electron microscope observations show that the deliberately grown interfacial oxide is of uniform thickness ~14 Å, but breaks up when annealed at ~950 °C and above, with "balling-up" occurring at ~1100 °C. This correlates with a transistor gain that decreases from ~1400 to ~40. The electron microscopy also shows that a thin interfacial oxide layer is present even when not deliberately grown. This oxide breaks up when annealed at ~900 °C and above, with "balling-up" occurring at ~1000 °C. This correlates with a transistor gain that decreases from ~240 to ~50. Calculations of the effect that such interfacial oxide layers will have on the characteristics of polysilicon emitter bipolar transistors are made, and these predictions correlate well with the measured characteristics.

04 August 2025 00:02:09

I. INTRODUCTION

Polycrystalline silicon ("polysilicon") is now widely used in bipolar technology as a means of making contact to the emitter and base of the transistor[1] and this can lead to a dramatic improvement in both switching speed and packing density.[2,3] An additional advantage of using a polysilicon emitter is that an improvement in gain by a factor of between 3 and 30 can be obtained compared with conventional transistors.[4] This improvement in gain can be partly attributed to the transport properties of the polysilicon[5] although this mechanism can only explain an improvement by a factor of ~3. The much larger gain improvements are obtained when a thin interfacial oxide layer is deliberately introduced between the polysilicon and single-crystal silicon.[4,6,7] In this case the gain improvement can be attributed to tunneling through this oxide layer.[8] Eltoukhy and Roulston,[9,10] and later Yu *et al.*,[11] have published unified theories that incorporate both these mechanisms, the former giving a solution suitable for numerical implementation, the latter a versatile analytical solution.

If high-gain polysilicon emitter transistors incorporating an interfacial oxide layer are to be used in commercial integrated circuit processes, it is important that the oxide layer is fully characterized in terms of thickness and uniformity, and that it is stable during subsequent high-temperature processing. Our previous structural studies[12–16] showed that the oxide layer progressively breaks up during high-temperature heat treatments. In order to investigate this behavior in more detail and to relate it directly to device performance, we have carried out a detailed comparison of electrical results on polysilicon emitter transistors with structural results from high-resolution transmission electron microscopy. The thermal stability of the oxide layer was investigated by carrying out preanneals at temperatures between 800 and 1100 °C immediately after polysilicon deposition but prior to emitter implant and drive-in at 900 °C. In this way, the oxide layer could be subjected to a high-temperature heat treatment without significantly affecting the rest of the emitter structure, and therefore changes in gain could be correlated directly with changes in the structure of the oxide layer.

II. EXPERIMENT

Two types of polysilicon emitter bipolar transistor were fabricated using identical standard processing up to the surface treatment of the (100) silicon slices prior to polysilicon deposition. For the first type (RCA treatment), the slice was given a dip in buffered HF followed by an RCA clean[4] (NH_4OH:H_2O_2:H_2O as 1:1:5 for 10 min; HCl:H_2O_2:H_2O as 1:1:6 for 10 min) to produce a surface oxide layer of nominal thickness ~15 Å. For the second type (HF treatment), the slice was given a dip in buffered HF with the intention of

 0021-8979/87/010225-09$02.40 © 1986 American Institute of Physics

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 643
2 of 10 8/3/2025, 8:02 PM

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

removing any native oxide which might have been present.[4] These surface layer treatments were immediately followed by the deposition of a ~0.4-μm-thick layer of undoped LPCVD polysilicon at ~610 °C. One of each type of slice was given a 10 min preanneal in dry nitrogen at one of several temperatures in the range 800–1100 °C in order to breakup the interfacial oxide layer to different extents. The emitters of the transistors were then formed by implanting 1×10^{16}-cm^{-2}, 70-keV As^+ into the polysilicon and driving it in at 900 °C in wet oxygen for 30 min. The active emitter area was 20 μm by 18 μm. The aluminium contact to the emitter was remote from the active emitter area, being made a via polysilicon track 60 μm in length.

Detailed electrical measurements were carried out on the transistors including measurements of collector and base current as a function of base/emitter voltage. Unpatterned check slices which had been given identical base and emitter processing were used for the TEM studies. Cross sections were prepared corresponding to the $(01\bar{1})$ plane of the (100) single-crystal silicon slice by grinding, polishing, and argon ion-beam thinning. These were examined using a JEOL 200CX TEM with the electron beam aligned parallel to the $[01\bar{1}]$ direction of the single-crystal silicon. High-resolution electron microscopy (HREM) was performed to give lattice images using conventional procedures. All of the specimens examined (except for that of Fig. 6) corresponded to the final processed slices and so the observed TEM structures could be directly correlated with the device results. Large numbers of micrographs were taken of each specimen in order to determine its characteristic structure. The micrographs of Figs. 1–6 are typical of those obtained, but do not show all the features described in the text.

III. TEM RESULTS

For specimens given in an RCA treatment, a 900 °C preanneal and full processing (As^+ implant and 900 °C drive-in), cross-section micrographs of the interface region showed the single-crystal silicon with the characteristic (111), $(\bar{1}11)$, and (200) lattice planes clearly delineated (Fig. 1). The individual polysilicon grains only showed lattice planes when the grains were favorably oriented with respect to the electron beam, e.g., as at A. When such planes were observed in the individual polysilicon grains, they were in general not aligned with the planes present in the silicon.

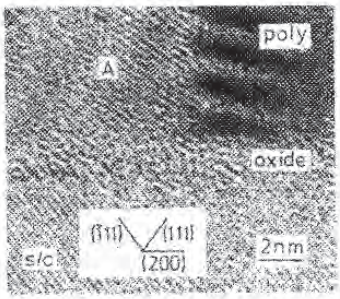


FIG. 1. HREM micrograph of a fully processed specimen given an RCA treatment and a 900 °C preanneal. Note the interfacial oxide layer and the polysilicon grain A. s/c—single-crystal silicon; oxide—interfacial oxide layer; poly—polysilicon layer.

The interfacial oxide layer was seen as a thin bright band separating the silicon from the polysilicon. The band exhibited a random wormlike contrast in the lattice-image micrographs, this being the characteristic behavior for an amorphous material. The oxide layer was continuous and uniform with a thickness of ~14 ± 1 Å and it followed the undulations of the surface of the silicon slice with an "amplitude" of ~6–9 Å (perpendicular to the interface) and a "wavelength" of ~200–400 Å (parallel to the interface). The good uniformity of the oxide layer indicates that its structure had been little changed by the subsequent 900 °C preanneal and 900 °C drive-in, and so Fig. 1 shows it in effect as it was when initially produced by the RCA treatment. The directly measured oxide layer thickness of 14 Å is in good agreement with the values of 13–15 Å obtained from ellipsometry,[17] and 10–15 Å obtained from Auger spectroscopy profiles,[18] for silicon slices given an RCA treatment. The TEM results are summarized in Table I.

For specimens given an RCA treatment, a 950 °C preanneal and full processing, the interfacial oxide layer was nonuniform (Fig. 2). Some regions had increased in thickness to a maximum of 19 Å, some regions were thinner, and in some regions holes (B) were present in the layer. Small areas (C) of the polysilicon had been regrown through such holes to become single-crystal material with the same crystallographic orientation as the underlying silicon, this being deduced from the micrograph contrast and the alignment of

TABLE I. Summary of results.

Silicon slice surface treatment	Preanneal temperature (°C)	Interfacial oxide layer: Geometry	Interfacial oxide layer: Maximum thickness (Å)	Polysilicon: Fraction of interface regrown	Polysilicon: Grain size (Å) Surface	Polysilicon: Grain size (Å) Bulk	Transistor gain
RCA	900	continuous	~14	0	1200	400	1400
RCA	950	broken up	~19	0.3	1200	600	350
RCA	1000	broken up	~22	0.7	1200	1000	90
RCA	1100	balled up	~55	0.97	2000	1000	40
HF	None	broken up	~8	0.2	1200	400	240
HF	1000	balled up	~50	0.95	1200	1000	50

04 August 2025 00:02:09

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 644

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

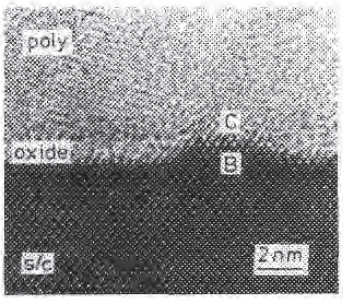


FIG. 2. HREM micrograph of a fully processed specimen given an RCA treatment and a 950 °C preanneal. Note the hole B in the interfacial oxide layer and the regrown polysilicon area C.

the lattice planes. The regrown areas had often spread out laterally in the polysilicon and so were sometimes wider than the holes in the oxide layer through which they had been grown. Individual areas extended typically 35 Å along the interface and 35 Å into the polysilicon. When the individual lengths of the regrown areas of the polysilicon were measured along the interface and added together, this represented ~0.3 of the total length of the interface.

For the specimens given an RCA treatment, a 1000 °C preanneal and full processing (micrograph not included), the breakup of the interfacial oxide layer was more pronounced. Some regions had increased in thickness to a maximum of 22 Å, and more holes were present in the layer. However, no substantial "balling-up" of the layer had occurred. Individual regrown areas of the polysilicon extended typically 120 Å along the interface and 40 Å into the polysilicon. The fraction of the length of the interface that had regrown was ~0.7.

For specimens given an RCA treatment, a 1100 °C preanneal and full processing, the breakup of the interfacial oxide layer was almost complete (Fig. 3), the layer now consisting mainly of balls of oxide (D) up to 55 Å in diameter or worms ~30×100 Å across. Individual regrown areas of the polysilicon overlapped to give almost complete coverage of the interface with the regrowth extending typically 170 Å into the polysilicon. The large difference between the

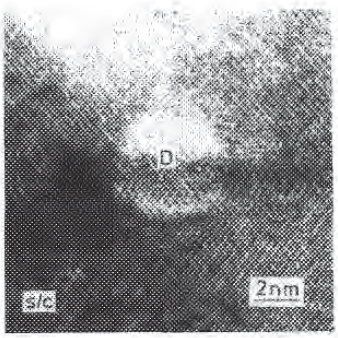


FIG. 3. HREM micrograph of a fully processed specimen given an RCA treatment and an 1100 °C preanneal. Note the balled-up oxide D.

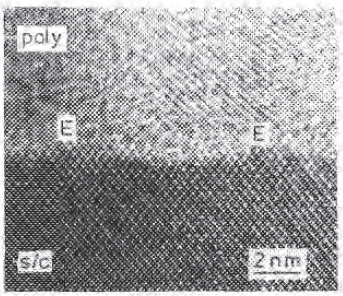


FIG. 4. HREM micrograph of a fully processed specimen given an HF treatment and no preanneal. Note the regrown polysilicon areas E.

oxide layer in the specimens given the 950, 1000, and 1100 °C preanneals, and the specimen given the 900 °C preanneal, indicates that the breakup of the oxide in the 950, 1000, and 1100 °C specimens occurred during the preanneals.

For specimens given an HF treatment, no preanneal and full processing, an interfacial oxide layer was present but it was not continuous (Fig. 4). In some regions it was up to a maximum thickness of 8 Å, in other regions it was thinner, and in some regions there were holes. Small areas (E) of the polysilicon had regrown through the holes, individual areas extending typically 25 Å along the interface and 25 Å into the polysilicon. The fraction of the interface that had been regrown was ~0.2.

For specimens given an HF treatment, a 1000 °C preanneal and full processing, the breakup of the interfacial oxide layer was severe, the layer consisting mainly of balls of oxide up to 50 Å diameter (Fig. 5). Individual areas (F) of regrown polysilicon extended typically 300 Å along the interface and 150 Å into the polysilicon, each area often covering several oxide balls (G) and holes (H). The fraction of the length of the interface that had regrown was ~0.95. The

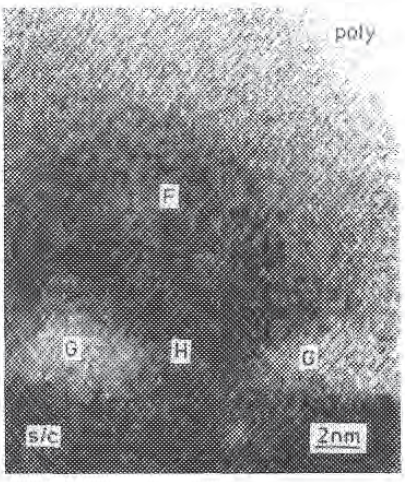


FIG. 5. HREM micrograph of a fully processed specimen given an HF treatment and a 1000 °C preanneal. Note the regrown polysilicon area F, the balled-up oxide G, and the holes in between H.

04 August 2025 00:02:09

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 645
4 of 10 8/3/2025, 8:02 PM

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

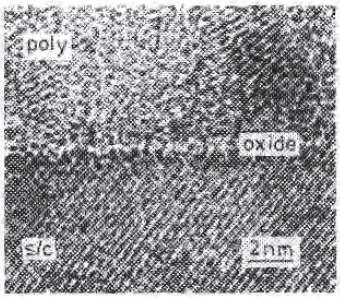


FIG. 6. HREM micrograph of a specimen given an HF treatment, but no preanneal, implant, or drive-in.

large difference between the oxide layers in the specimen given the 1000 °C preanneal, and the specimen given no preanneal, indicates that the "balling-up" of the oxide in the 1000 °C specimen mainly occurred during the preanneal.

In order to obtain further information concerning the interfacial oxide layer that was present after an HF treatment, additional specimens were examined that corresponded to an HF treatment and polysilicon deposition, but no further processing. A white band was seen at the interface (Fig. 6) but this was extremely thin, and so the oxide thickness could not be directly measured accurately. The interface was very uniform, and in a few places adjacent to the interface small areas of the polysilicon were of the same crystallographic orientation as the single-crystal silicon. These areas could have arisen by either direct epitaxial growth through holes initially present in the oxide layer or regrowth through holes created initially present in the oxide layer during the deposition of the polysilicon at 610 °C. The results indicate that an extremely thin and mainly continuous oxide layer was present immediately after the polysilicon deposition. The results also indicate that for the specimens that were given an HF treatment, no preanneal and full processing, the significant breakup of the oxide layer that was observed occurred during the 900 °C drive-in.

In our previous TEM studies[13] of the breakup of an analogous interfacial oxide layer formed by a sulfuric acid/hydrogen peroxide surface chemical treatment, it was shown that the volume of the oxide present after complete balling-up of the layer corresponded closely to the volume of the ~10-Å-thick uniform oxide layer that was present initially. This suggested that the breakup of the oxide layer occurred mainly by a local diffusion mechanism. Application of this principle to the present work gives the following results. For the specimens given an HF treatment, a 1000 °C preanneal, and full processing, the balled-up oxide corresponded to an approximately planar array of spheres of mean diameter 40 Å and mean spacing 100 Å. Calculation shows that this is equivalent to an initial uniform oxide layer of thickness ~4 Å, a result which is consistent with the TEM observations for the specimens given an HF treatment and polysilicon deposition, but no further processing (Fig. 6). For the specimens given an RCA treatment, an 1100 °C preanneal, and full processing, a similar calculation showed that the balled-up oxide layer is equivalent to an initial uniform oxide layer of thickness ~15 Å. This result is in good agreement with the TEM value of ~14 Å for the specimens given an RCA treatment, preannealed at 900 °C, and fully processed (Fig. 1).

Present evidence indicates that there are genuine differences in the thickness of oxide layers formed on silicon slices by the HF treatments of different organizations. These differences may depend on the HF solution, the dip procedure, or the elapsed time prior to the subsequent deposition of the polysilicon. Our TEM examinations of silicon specimens given nominally the same HF treatment, and subsequently processed, by different organizations have shown differences in the behavior of the oxide layer which we attribute to differences in its initial thickness.[19] Furthermore, measurements by different organizations of the thickness of oxide layers present on silicon slices after HF treatments have given markedly different values, e.g., 8–9 Å from ellipsometry[17] and ~3 Å from Auger spectroscopy.[18]

The above TEM results show that the interfacial oxide layers break up faster when they are heated at higher temperatures, and when they are initially thinner. Thus, for the oxide layers arising from the HF treatment, which were initially ~4 Å thick, breakup had already occurred during a heat treatment at 900 °C (drive-in), while for oxide layers arising from the RCA treatment, which were initially ~14 Å thick, breakup only began to occur during a treatment at 950 °C (preanneal). These present results are consistent with our previous TEM results[14,16] which indicated that an initially ~10-Å-thick oxide layer was just beginning to break up at 900 °C. The results are also consistent with the TEM results of Schaber *et al.*[20] which showed that for an oxide layer ~30 Å thick, breakup did not occur at 950 °C. However, some caution is required when making comparisons of such results obtained by different organizations because the precise behavior depends on a number of factors, e.g., it has been shown that such oxide layers break up more rapidly when the arsenic concentration is higher.[15,16]

Grain sizes for the polysilicon layers in fully processed specimens were determined from TEM cross-section micrographs taken at lower magnifications. For individual specimens, the grain size near the polysilicon layer surface was in general larger than in the remainder of the polysilicon layer. This was because the upper portion of the polysilicon layer had been made amorphous by the arsenic implant. Consequently, during the drive-in at 900 °C, larger grains had formed in this portion of the layer by regrowth from amorphous material than in the remainder of the layer by regrowth from fine grains. For the grains in the main portion of the layer, i.e., away from the layer surface, the results showed that the mean grain size was typically 400, 600, 1000, and 1000 Å for preanneal temperatures of 900, 950, 1000, and 1100 °C, respectively. Conversely, for the grains near the surface, the mean grain size was typically 1200 Å for all of these preanneal temperatures, except for 1100 °C when it was typically 2000 Å.

IV. ELECTRICAL RESULTS

Figures 7 and 8 show Gummel plots for the RCA- and HF-treated devices, respectively, that were given preanneals

04 August 2025 00:02:09

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 646
5 of 10 8/3/2025, 8:02 PM

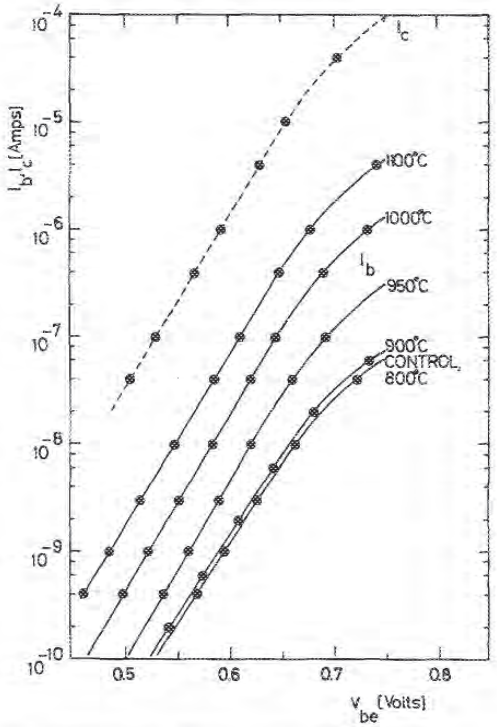


FIG. 7. Gummel plots of devices given an RCA treatment and preanneals in the range 800–1100 °C.

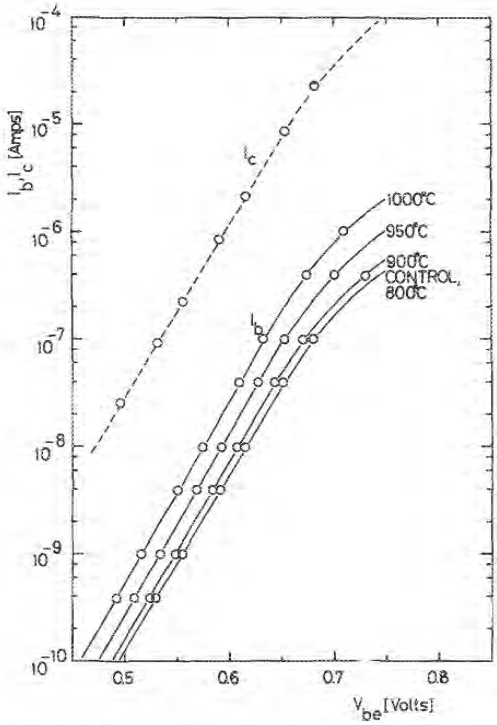


FIG. 8. Gummel plots of devices given an HF treatment and preanneals in the range 800–1000 °C.

between 800 and 1100 °C. The collector characteristics for both the HF- and RCA-treated devices are all ideal over at least three orders of collector current (60 mV/decade). Similarly the base characteristics of the HF-treated devices are also ideal, although the base characteristics of the RCA-treated devices are slightly nonideal (70 mV/decade) corresponding to an ideality parameter of $n = 1.12$. To show the effect of the preanneal on the electrical characteristics of the RCA- and HF-treated devices, I_{b0} and I_{c0} have been plotted against the preanneal temperature in Fig. 9. Because of the nonideality of the base current characteristics of the RCA-treated devices, the values of I_{b0} were calculated from the ideal collector characteristics and maximum current gain β_{max} using the relationship[21] $I_{b0} = I_{c0}/\beta_{max}$.

For the RCA-treated devices, Figs. 7 and 9 show that the collector current is relatively unaffected by a preanneal. Conversely, the base current is strongly affected with the electrical results falling into two distinct regimes. Below ~900 °C, the base current is approximately constant. In this regime, the 900 °C drive-in is the dominant process as far as the interfacial oxide layer is concerned. Above ~900 °C, the base current increases rapidly with preanneal temperature. In this regime, the preanneal is the dominant process as far as the oxide layer is concerned. For example, the base current increases after a 950 °C preanneal by ~6×, after a 1000 °C preanneal by ~28×, and after a 1100 °C preanneal by ~100×. This trend correlates well with the TEM results. Thus, after a 900 °C preanneal, the oxide layer is uniform and continuous (Fig. 1) and the gain is not significantly affected by the preanneal. After a 950 °C preanneal, the oxide layer begins to break up (Fig. 2) and a large increase in

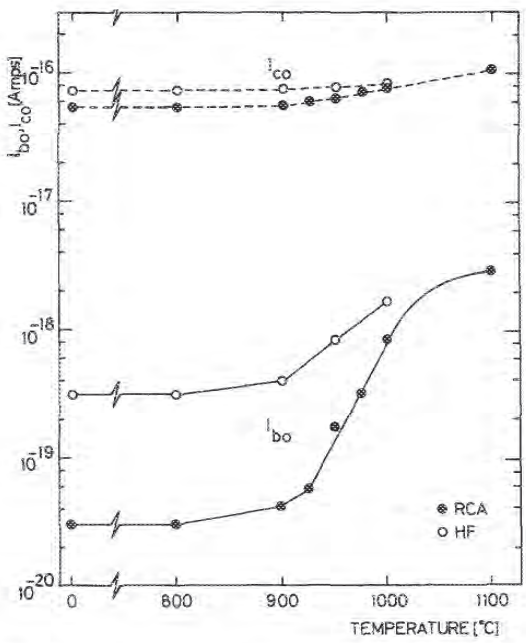


FIG. 9. I_{b0} and I_{c0} vs preanneal temperature for devices given RCA and HF treatments.

04 August 2025 00:02:09

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

the base current occurs. As the preanneal temperature is further increased, the breaking up of the oxide layer continues (Fig. 3) and this corresponds to a continuing increase in the base current. Over the preanneal temperature range of 800–1100 °C, the transistor gain decreases from ~1800 to ~40.

For the HF-treated devices, Figs. 8 and 9 show a similar but less pronounced trend to that of the RCA-treated devices. Thus, the collector current is relatively unaffected by a preanneal. For the base current, after a preanneal below ~900 °C, no significant increase occurs. After a 950 °C preanneal it increases ~3×, and after a 1000 °C preanneal it increases ~5.5×. This increase in the base current again correlates with a progressive increase in the breaking up of the interfacial oxide layer (Figs. 4 and 5), although it should be noted that some polysilicon grain growth also occurs (Table I). Over the pre-anneal temperature range of 800–1000 °C, the transistor gain decreases from ~240 to ~50.

A comparison of the RCA- and HF-treated devices shows the following. After a preanneal below ~900 °C, the base current for the RCA-treated devices is ~10× less than for the HF-treated devices. This difference correlates with the former devices having a relatively thick, continuous oxide layer, and the latter devices having a thin broken-up oxide layer. After a preanneal at 1000 °C, the base current for the RCA-treated devices is ~2× less than for the HF-treated devices. This difference correlates with the former devices having a broken-up oxide layer, and the latter devices having a completely balled-up oxide layer.

V. INTERPRETATION OF ELECTRICAL RESULTS

The larger gain that occurs for polysilicon emitter bipolar transistors when a thin interfacial oxide layer is incorporated in the emitter can be explained qualitatively in terms of tunneling through the oxide layer as follows. For *n-p-n* transistors, the collector current I_c is determined mainly by the flow of electrons through the emitter to the base and collector. Considerable evidence has shown that the potential barrier to electrons χ_e associated with the oxide layer is small, the tunnelling probability is high, and so the electron flow is little reduced. Conversely, the base current I_b is determined mainly by the flow of holes from the base into the emitter. The potential barrier to holes χ_h associated with the oxide layer is high, the tunneling probability is low, and so the hole flow is significantly reduced. The current gain is the ratio I_c/I_b and so the gain is increased.

We shall now try to interpret our device results more quantitatively. Yu *et al.*[11] in their theoretical analysis of polysilicon emitter bipolar transistors which incorporated a thin uniform oxide layer within the emitter, obtained the following equation for the base current density J_b:

$$J_b = [qS_{\text{eff}} n_{ie}^2 / N_d(w)] \exp(qV_{be}/kT), \tag{1}$$

where V_{be} is the base/emitter applied voltage, $N_d(w)$ is the emitter carrier concentration at the base/emitter depletion region edge, n_{ie} is the intrinsic carrier concentration of the emitter taking into account band-gap narrowing effects, and S_{eff} is an effective recombination velocity which incorporates the various recombination processes occurring for the holes injected from the base into the emitter. It can be seen that J_b depends directly on S_{eff}, and also on the emitter dopant profile through the parameters n_{ie} and $N_d(w)$. Earlier work[22] showed that an interfacial oxide layer can retard the emitter dopant diffusion, and hence change the final dopant profile. However, for the oxide layer thicknesses considered here, this effect is very small. We shall therefore assume that n_{ie} and $N_d(w)$ are constant, and so variations in J_b can be predicted from changes in the value of S_{eff}.

A simplification of the analysis occurs if carrier recombination in the single-crystal part of the emitter located below the interfacial oxide layer is ignored. In the present work this assumption is justified, partly because the emitter/base junction is only ~100 Å below the oxide layer,[1] and partly because we shall only be considering the change in J_b, rather than its absolute value. The equation for S_{eff} then reduces to[11]

$$S_{\text{eff}} = S_i + \{(1/T_i) + [1/(S_i + S_p)]\}^{-1}, \tag{2}$$

where S_i is the recombination velocity for holes at the oxide layer/single-crystal silicon interface, T_i is an effective recombination velocity corresponding to the tunnelling of holes through the oxide layer, and S_p is an effective recombination velocity for holes in the polysilicon.

The parameter T_i depends on the oxide layer thickness Δ and is given by[11]

$$T_i = (kT/2\pi m_h^*)^{1/2} (e^{-b_h})(1 - c_h kT)^{-1}, \tag{3}$$

where

$$b_h = (4\pi\Delta/h)(2m_h^* \chi_h)^{1/2},$$

$$c_h = (2\pi\Delta/h)(2m_h^*/\chi_h)^{1/2},$$

and χ_h is the potential barrier for holes and m_h^* is the effective mass for holes.

To gain a clear insight into the relative importance of the three components of S_{eff}, curves showing the relationship between S_{eff} and Δ were calculated using Eqs. (2) and (3) (Fig. 10). A value for S_p of 1.5×10^5 cm s^{-1} was used, this being calculated by the method of Yu *et al.*[11] for a grain size of 1000 Å using the data listed in Table II. This value may be compared with the range of values of $\sim 8\times10^4$ to 2×10^5 cm s^{-1} obtained by Yu *et al.*[11] for similar polysilicon layers with various thicknesses and grain sizes. A value for S_i of 1.5×10^3 cm s^{-1} was assumed, again following Yu *et al.*[11] In order to calculate T_i, a value of χ_h is required. Measurements of χ_h by Ng and Card[27] for oxide layers in the thickness range ~20 to ~26 Å gave values of ~1.0 eV, i.e., considerably less than for thick oxide layers where $\chi_h \simeq 3.0$ eV. In the absence of any experimental values of χ_h for oxide layers of thickness <20 Å, three values have been chosen for our calculations, namely, $\chi_h = 0.5$, 1.0, and 1.5 eV. It was also assumed that $m_h^* = m_0$, the free electron mass.

From Fig. 10 it can be seen that there are three distinct regions of device operation for a polysilicon emitter transistor. For large Δ, $S_{\text{eff}} \simeq S_i$ and is determined mainly by recombination at the oxide layer/single-crystal silicon interface. Other mechanisms such as recombination in the single-crystal part of the emitter and recombination in the base may also contribute to the base current in this region of

04 August 2025 00:02:09

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 648

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

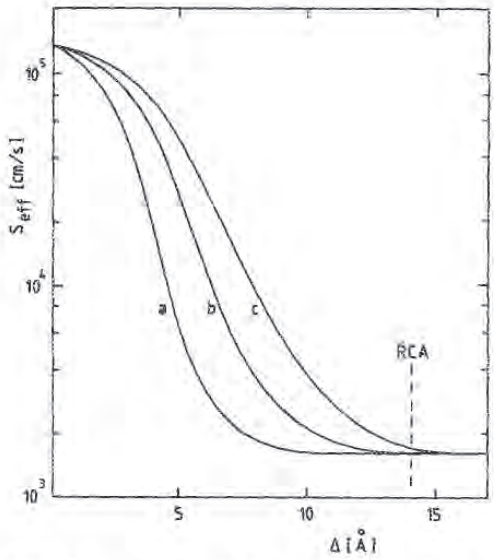


FIG. 10. Theoretical plots of effective recombination velocity (S_{eff}) vs interfacial oxide layer thickness (Δ) for hole potential barriers (χ_h) associated with the oxide of (a) 1.5, (b) 1.0, and (c) 0.5 eV.

operation.[22] For small Δ, most holes tunnel through the oxide layer, $S_{eff} \simeq S_p$ and is determined mainly by transport and recombination in the polysilicon. For intermediate Δ, tunneling through the oxide layer is the dominant mechanism, and S_{eff} is extremely sensitive to small changes in Δ.

Let us initially consider the effect of the interfacial oxide layer on the device operation by comparing the $\sim 10\times$ difference in base current between the RCA- and HF-treated devices given no preanneal, as shown in Fig. 9, with that predicted by Fig. 10. For the RCA-treated devices, the oxide layer is uniform and of thickness ~ 14 Å. Consequently, Fig. 10 predicts that $S_{eff} \simeq S_i \simeq 1.5\times 10^3$ cm s^{-1}, and this value does not depend significantly on the precise value of χ_h. For the HF-treated devices, S_{eff} should then be $10\times$ larger, i.e., $\sim 1.5\times 10^4$ cm s^{-1}. Figure 10 predicts that this would correspond to a uniform oxide layer of thickness ~ 4 to ~ 8 Å, the range arising because of the uncertainty in the value of χ_h. This prediction is in reasonable agreement with the observation of a uniform oxide layer of thickness ~ 4 Å immediately after polysilicon deposition, and a broken up oxide layer of maximum thickness ~ 8 Å after no preanneal, but a 900 °C drive-in.

TABLE II. Summary of parameter values used to calculate S_p.

Parameter	Value	Reference
Doping concentration $N_d(w)$	5×10^{19} cm^{-3}	1
Diffusion length of holes, L_p	0.7 μm	21
Grain size	0.1 μm	
Hole mobility at grain boundary, μ_{gb}	2 cm^2 V^{-1} s^{-1}	23
Thickness of grain boundary, δ	10 Å	24
Grain boundary interface recombination velocity, S_{gb}	7.5×10^4 cm s^{-1}	25,26
Length of polysilicon emitter track	Large with respect to L_p	

Let us now consider the effect of the interfacial oxide layer on the devices by comparing the $\sim 100\times$ difference in base current between the RCA-treated devices given no preanneal with those given a 1100 °C preanneal. For no preanneal, the oxide layer is continuous and uniform, and $S_{eff} \simeq S_i \simeq 1.5\times 10^3$ cm s^{-1}. For a 1100 °C preanneal, the oxide layer is balled-up. If it is assumed that such a balled-up oxide layer has little effect on the flow of holes through the interface, then S_{eff} will correspond to the value for $\Delta = 0$, i.e., to $S_{eff} \simeq S_p \simeq 1.5\times 10^5$ cm s^{-1} (Fig. 10). Hence, an increase in base current of $\sim 100\times$ is predicted, in good agreement with the experimental value of $\sim 100\times$ (Fig. 9).

With regard to the effect of the interfacial oxide layer on the RCA-treated devices as the preanneal temperature is progressively increased, comparison of the device results and TEM observations suggest the following behaviors. On going from no preanneal to ~ 800 °C, I_b is constant and is determined mainly by recombination of holes at the oxide-layer/single-crystal silicon interface (mechanism A). From ~ 800 to ~ 925 °C, I_b increases slowly and is determined by mechanism A, together with tunnelling of holes through the thinner portions of the oxide layer (mechanism B). From ~ 925 to ~ 975 °C, I_b increases rapidly and is determined by mechanism B, together with an increasing contribution from the flow of holes through the penetrated regions of the oxide layer and the transport properties in the polysilicon layer (mechanism C). From ~ 975 to ~ 1100 °C, I_b continues to increase and then approaches a constant value, and is determined mainly by mechanism C, the constant value arising when the oxide layer is completely balled up.

An additional factor that could contribute to the increase in the base current as the preanneal temperature increases is the progressive coarsening of the polysilicon grains that occurs. The grain size for the part of the layer adjacent to the interfacial oxide layer increases from ~ 400 to ~ 1000 Å on going from no preanneal to both 1000 to 1100 °C preanneals (Table I). Such coarsening could change S_p, and this would change S_{eff} if Δ were small. The calculations of Yu *et al.*[11] indicate a maximum change in S_p due to this effect of ~ 2 or $\sim 3\times$. However, such a change in S_p would have no effect on the $\sim 100\times$ difference in I_b calculated above for the RCA-treated devices on going from no preanneal to a 1100 °C preanneal. This is because for no preanneal, the value of $S_{eff} \simeq S_i$ used is appropriate for a grain size of ~ 400 Å (S_i corresponds to mechanism A and is independent of grain size), and for a 1100 °C preanneal, the value of $S_{eff} \simeq S_p$ used is appropriate for a grain size of ~ 1000 Å (S_p was calculated from the data of Table II). However, the coarsening of the grains could slightly affect the shape of the curve of Fig. 9 for the RCA-treated devices on going from no preanneal to a 1100 °C preanneal because mechanisms B and C both depend on the transport properties in the polysilicon layer.

04 August 2025 00:02:09

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 649
8 of 10 8/3/2025, 8:02 PM

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

Let us now consider the effect of the interfacial oxide layer on the devices by comparing the $\sim 5.5\times$ difference in base current between the HF-treated devices given no preanneal with those given a 1000 °C preanneal. It is more difficult to calculate the behavior in this case because the oxide layer is already broken up for no preanneal. Comparison of the device results and TEM observations suggests that on going from no preanneal to ~ 900 °C, I_b is determined by mechanisms B and C; and on going from ~ 900 to ~ 1000 °C, I_b is determined mainly by mechanism C.

In this case the progressive coarsening of the grains could make a contribution to the increase in I_b of $\sim 5.5\times$. This is because over the whole range from no preanneal to a 1000 °C preanneal, the transport properties in the polysilicon layer play a role. The experimental results of Ning and Isaac,[5] performed on what were in effect HF-treated devices, showed that on going from bipolar silicon transistors with polysilicon emitters to analogous devices with single-crystal emitters, the device gain decreased by $\sim 3\times$. These results suggest that in our case, i.e., for polysilicon emitter bipolar transistors given an HF treatment, some of the increse in I_b of $\sim 5.5\times$ is due to the coarsening of the grains and some to the continued breaking up of the oxide layer.

From the above results it might be expected that all devices with oxide layers thicker than ~ 10 Å would give the same best performance. However, although all such devices would possess the same lowest value of S_{eff}, and as a result J_b, the oxide layers introduce a series emitter resistance which increases rapidly with increasing oxide thickness.[11,28] This is detrimental to device and circuit performance, and so in practice a compromise between increased gain and increased emitter resistance needs to be made.

As far as the breakup and eventual balling-up of the thin interfacial oxide layer is concerned, the main driving force is considered to be a significant reduction in the total amount of energy associated with the oxide layer/silicon interface[13,15] (analogous to the well-known Oswald ripening effect that occurs in metallurgical systems). Thus, on going from a continuous uniform oxide layer to a balled-up oxide layer, the interfacial area decreases by typically ~ 5 to $\sim 10\times$ (deduced from TEM examinations). However, the presence of the As dopant can also have an effect. For example, in our work[15,16] on similar specimens post-annealed at 900 °C, the breakup of the oxide layer was found to proceed more rapidly when the As^+ implantation dose was increased from 5×10^{15} to 2×10^{16} cm^{-2}. This result suggests[15] that the As dopant can either enhance the local oxide diffusion rate or change the oxide/silicon interfacial energy.

VI. CONCLUSIONS

A detailed comparison of the electrical and TEM results has revealed that there is a strong correlation between the high gain of polysilicon emitter bipolar transistors and the existence of a continuous, thin, interfacial oxide layer between the single-crystal silicon and the polysilicon. For RCA-treated devices, an initial high gain of ~ 1400 occurs due to the presence of an ~ 14-Å-thick oxide layer. On annealing at temperatures up to 1100 °C, the gain progressively decreased to ~ 40 due mainly to the breaking up of the oxide layer. For HF-treated devices, an initial gain of ~ 240 occurs due to the presence of a thinner and already partly broken up oxide layer. On annealing at temperatures up to 1000 °C, the gain decreases to ~ 50 due to further breaking up of the oxide layer together with a coarsening of the polysilicon grains. The decrease in gain observed with increasing anneal temperture is explained using a theoretical model in which tunneling through the interfacial oxide is an important mechanism. As demonstrated by other authors,[11,28] if the oxide layer is too thick, it contributes a significant series resistance to the emitter structure. Therefore, in order to obtain the optimum device, it is necessary to produce an extremely thin continuous oxide layer, and to prevent it from breaking up during the device fabrication. The thinner the oxide layer, the more readily it breaks up, and the more critical the device processing procedures become.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the staff of the Southampton University Microelectronics Centre for assistance with the fabrication of the devices. Special thanks are due to Dr. A. Cuthbertson and Dr. E. F. Chor of Southampton University and Dr. P. C. Hunt and Dr. M. C. Wilson of Plessey, Caswell for many stimulating and helpful discussions. The work was supported by the Science and Engineering Research Council, UK, and the Alvey Directorate, UK.

[1] A. Cuthbertson and P. Ashburn, Int. Electron Devices Meeting Techical Digest, 749 (1984).
[2] T. H. Ning, R. D. Isaac, P. M. Solomon, D. D. Tang, H. N. Yu, G. C. Feth, and S. K. Wiedmann, IEEE Trans. Electron Devices **ED-28**, 1010 (1981).
[3] T. Sakai, Y. Yamamoto, Y. Kobayashi, K. Kawarada, V. Inabe, T. Hayashi, and H. Miyanaga, IEEE J. Solid State Circuits **SC-16**, 424 (1981).
[4] B. Soerowirdjo and P. Ashburn, Solid State Electron. **26**, 495 (1983).
[5] T. H. Ning and R. D. Isaac, IEEE Trans. Electron Devices **ED-27**, 2051 (1980).
[6] B. Soerowirdjo, P. Ashburn, and A. Cuthbertson, International Electron Devices Meeting Technical Digest, 688 (1982).
[7] P. Ashburn and B. Soerowirdjo, IEEE Trans. Electron Devices **ED-31**, 853 (1984).
[8] H. C. de Graff and J. C. de Groot, IEEE Trans. Electron Devices **ED-26**, 1771 (1979).
[9] A. A. Eltoukhy and D. J. Roulston, IEEE Trans. Electron Devices **ED-29**, 961 (1982).
[10] A. A. Eltoukhy and D. J. Roulston, IEEE Trans. Electron Devices **ED-29**, 1862 (1982).
[11] Z. Yu, B. Ricco and R. W. Dutton, IEEE Trans. Electron Devices **ED-31**, 773 (1984).
[12] M. C. Wilson, P. Ashburn, B. Soerowirdjo, G. R. Booker, and P. Ward, J. Phys. (Paris) Colloq. **43**, Suppl. 10, C1-253 (1982).
[13] A. Albu-Yaron, J. C. Barry, and G. R. Booker, in *Proceedings of the 8th European Congress on Electron Microscopy*, Budapest (Programme Committee, Congress Bureau MOTESZ, Budapest, Hungary, 1984), Vol. 1, p. 521.
[14] A. Albu-Yaron, J. C. Barry, N. Jorgensen, G. R. Booker, P. Ashburn, and P. Ward, British Association for Crystal Growth Annual Conference, Lancaster (1984) (unpublished).
[15] N. Jorgensen, J. C. Barry, G. R. Booker, P. Ashburn, G. R. Wolstenholme, M. C. Wilson, and P. C. Hunt, Inst. Phys. Conf. Ser. **76**, 471 (1985).
[16] M. C. Wilson, N. Jorgensen, G. R. Booker, and P. C. Hunt, *Proceedings of the 5th Symposium of VLSI Technology*, Kobe, Japan (IEEE and Japanese Society of Applied Physics, Japan, 1985), p. 46.
[17] R. C. Henderson, J. Electrochem. Soc. **119**, 772 (1982).

04 August 2025 00:02:09

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 650
9 of 10 8/3/2025, 8:02 PM

225_1_online.pdf https://watermark.silverchair.com/225_1_online.pdf?token=AQECA...

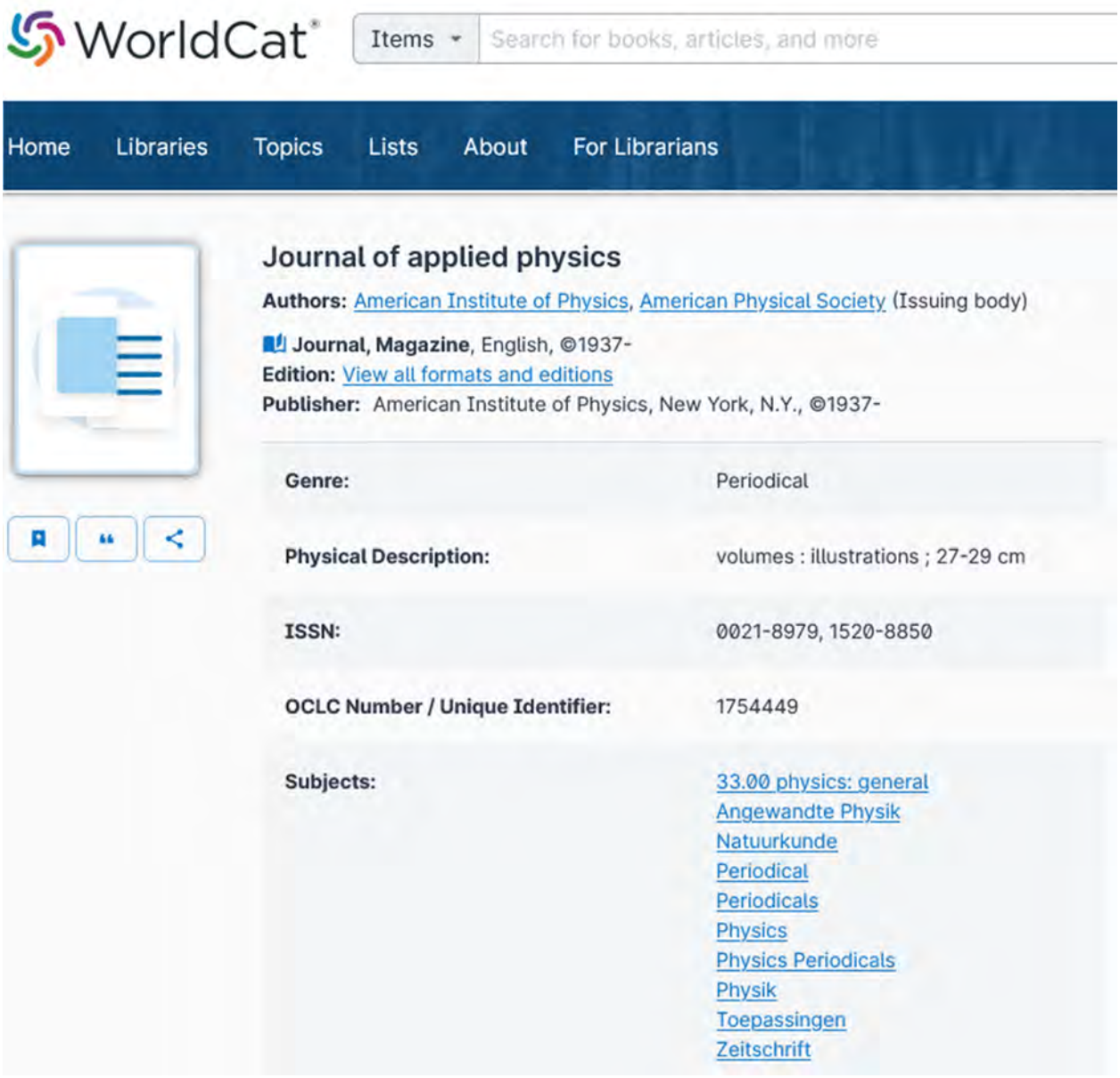
[18]J. E. Duffill, *Proceedings of the 12th European Solid State Device Research Conference*, Munich (1982).
[19]N. Jorgensen, A. Albu-Yaron, J. C. Barry, and G. R. Booker (unpublished results) (1984).
[20]H. Schaber, B. Benno, L. Treikaguer and A. W. Wieder, International Electron Devices Meeting Technical Digest, 738 (1984).
[21]A. W. Wieder, IEEE Trans. Electron Devices **ED-27**, 1402 (1980).
[22]P. Ashburn and B. Soerowirdjo, Solid State Electron. **24**, 475 (1981).
[23]D. M. Kim, A. N. Khondker, R. R. Shah, and D. C. Crosthwaite Jr., IEEE Electron Device Lett. **EDL-3**, 141 (1982).
[24]M. M. Mandurah, K. C. Saraswat, and T. I. Kamins, IEEE Trans. Electron Devices **ED-28**, 1163 (1981).
[25]J. Oualid, H. Anzil, J. P. Crest, J. Dugas, G. Mathian, M. Zemt, and S. Martinazzi, J. Phys. (Paris) Colloq. **43**, Suppl. 10, C1-313 (1982).
[26]C. H. Seager, J. Appl. Phys. **53**, 5968 (1982).
[27]K. K. Ng and H. C. Card, J. Appl. Phys. **51**, 2153 (1980).
[28]E. F. Chor, P. Ashburn, and A. Brunnschweiler, IEEE Electron Device Lett. **EDL-6**, 516 (1985).

04 August 2025 00:02:09

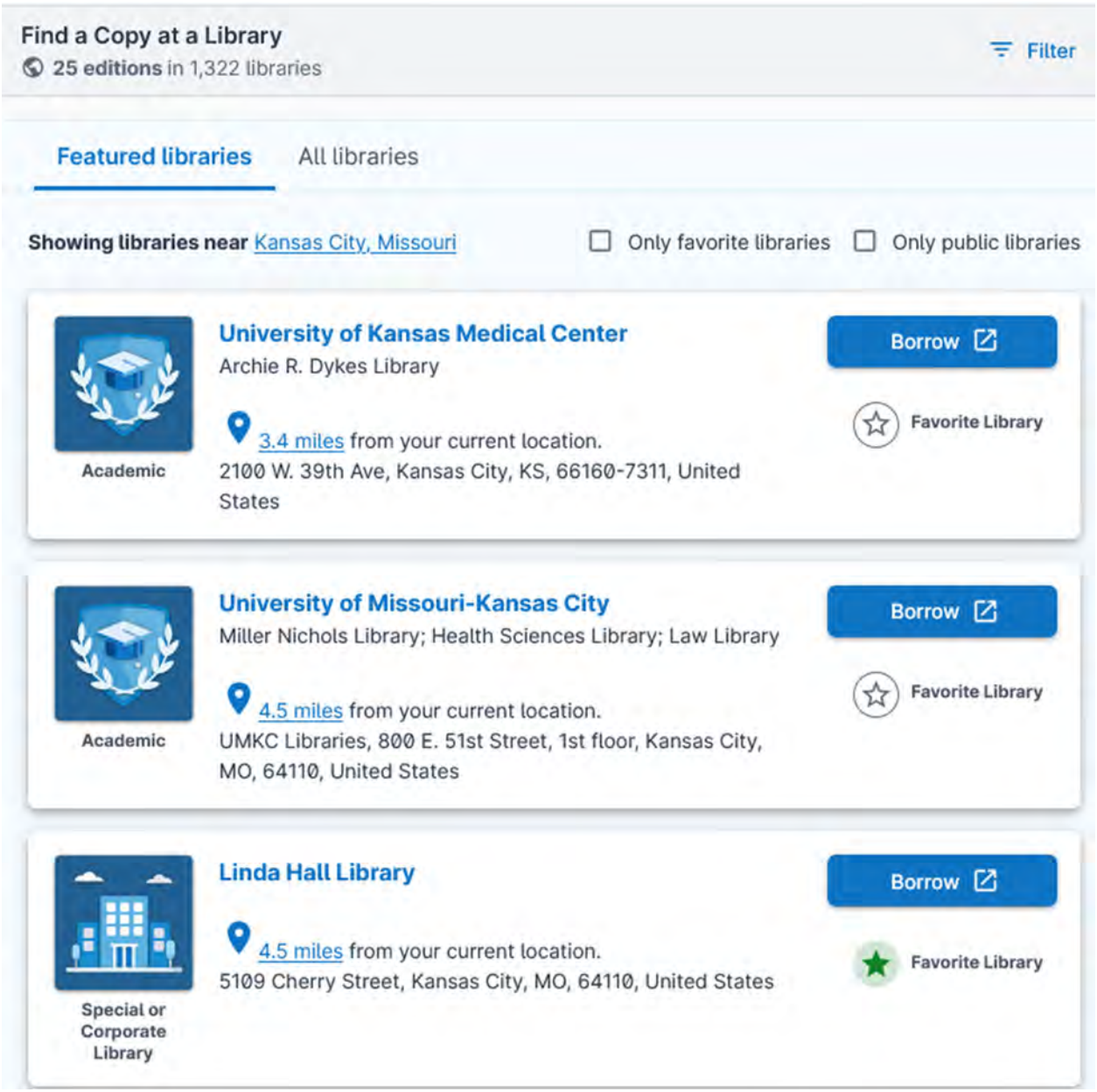
Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 651
10 of 10 8/3/2025, 8:02 PM

ATTACHMENT K-4

Screen capture created by J L Mullins from WorldCat on August 12, 2025 at URL:
https://search.worldcat.org/title/1754449



Screen capture created by J L Mullins from WorldCat on August 12, 2025 at URL: https://search.worldcat.org/title/1754449



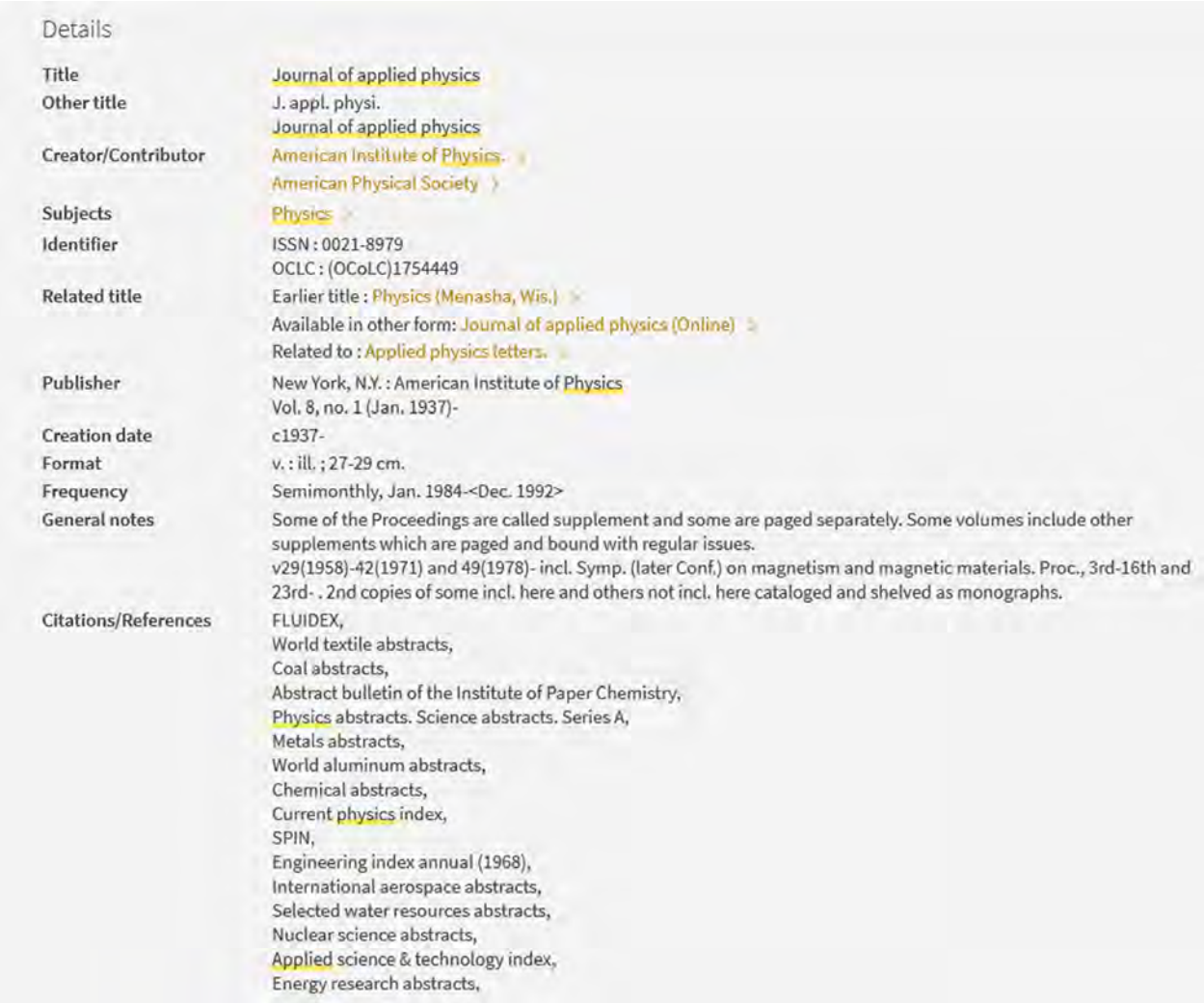
ATTACHMENT K-5

Screen capture created by J L Mullins from Linda Hall Library OPAC on August 12, 2025 at URL:
https://catalog.lindahall.org/discovery/fulldisplay?docid=alma992698823405961&context=L&vid=01LINDAHALL_INST:LHL&lang=en&search_scope=MyInstitution&adaptor=Local%20Search%20Engine&tab=LibraryCatalog&query=title,contains,Journal%20of%20applied%20physics,AND&sortby=date_d&facet=frbrgroupid,include,9031824872188245755&mode=advanced&offset=0





Screen capture created by J L Mullins from Linda Hall Library OPAC on August 12, 2025 at URL:
https://catalog.lindahall.org/discovery/fulldisplay?docid=alma992698823405961&context=L&vid=01LINDAHALL_INST:LHL&lang=en&search_scope=MyInstitution&adaptor=Local%20Search%20Engine&tab=LibraryCatalog&query=title,contains,Journal%20of%20applied%20physics,AND&sortby=date_d&facet=frbrgroupid,include,9031824872188245755&mode=advanced&offset=0



Details

Title	Journal of applied physics
Other title	J. appl. physi. Journal of applied physics
Creator/Contributor	American Institute of Physics. American Physical Society
Subjects	Physics
Identifier	ISSN : 0021-8979 OCLC : (OCoLC)1754449
Related title	Earlier title : Physics (Menasha, Wis.) Available in other form: Journal of applied physics (Online) Related to : Applied physics letters.
Publisher	New York, N.Y. : American Institute of Physics Vol. 8, no. 1 (Jan. 1937)-
Creation date	c1937-
Format	v. : ill. ; 27-29 cm.
Frequency	Semimonthly, Jan. 1984-<Dec. 1992>
General notes	Some of the Proceedings are called supplement and some are paged separately. Some volumes include other supplements which are paged and bound with regular issues. v29(1958)-42(1971) and 49(1978)- incl. Symp. (later Conf.) on magnetism and magnetic materials. Proc., 3rd-16th and 23rd- . 2nd copies of some incl. here and others not incl. here cataloged and shelved as monographs.
Citations/References	FLUIDEX, World textile abstracts, Coal abstracts, Abstract bulletin of the Institute of Paper Chemistry, Physics abstracts. Science abstracts. Series A, Metals abstracts, World aluminum abstracts, Chemical abstracts, Current physics index, SPIN, Engineering index annual (1968), International aerospace abstracts, Selected water resources abstracts, Nuclear science abstracts, Applied science & technology index, Energy research abstracts,

ATTACHMENT L-1

Semiconductor Manufacturing Technology



Michael Quirk
Julian Serda

SEMICONDUCTOR MANUFACTURING TECHNOLOGY

MICHAEL QUIRK
STAKTEK CORPORATION

JULIAN SERDA
ADVANCED MICRO DEVICES



Upper Saddle River, New Jersey
Columbus, Ohio

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 660

Library of Congress Cataloging-in-Publication Data
Quirk, Michael
Semiconductor manufacturing technology / Michael Quirk, Julian Serda.
p. cm.
ISBN 0-13-081520-9
1. Electronic industries. 2. Semiconductors—Design and construction. I. Serda, Julian.
II. Title.

TK7836.Q27 2001

00-041653

Vice President and Publisher: Dave Garza
Editor in Chief: Stephen Helba
Assistant Vice President and Publisher: Charles E. Stewart, Jr.
Production Editor: Alexandrina Benedicto Wolf
Production Coordination: Lithokraft II
Design Coordinator: Robin G. Chukes
Cover Designer: Linda Fares
Cover Image: John Foxx
Production Manager: Matthew Ottenweller
Marketing Manager: Barbara Rose

This book was set in Times Roman by Lithokraft II. It was printed and bound by R. R. Donnelley & Sons Company. The cover was printed by Phoenix Color Corp.

Copyright © 2001 by Prentice-Hall, Inc., Upper Saddle River, New Jersey 07458. All rights reserved. Printed in the United States of America. This publication is protected by Copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permission(s), write to: Rights and Permissions Department.



10 9 8 7 6 5 4 3 2
ISBN 0-13-081520-9

CONTENTS

CHAPTER 1 INTRODUCTION TO THE SEMICONDUCTOR INDUSTRY

OBJECTIVES 1
INTRODUCTION 2
DEVELOPMENT OF AN INDUSTRY 2
- Industry Roots 2
- The Solid State 3

CIRCUIT INTEGRATION 4
- Integration Eras 5

IC FABRICATION 6
- Wafer Fab 7
- Stages of IC Fabrication 7

SEMICONDUCTOR TRENDS 9
- Increase in Chip Performance 9
- Increase in Chip Reliability 12
- Reduction in Chip Price 12

THE ELECTRONIC ERA 14
- The 1950s: Transistor Technology 14
- The 1960s: Process Technology 14
- The 1970s: Competition 14
- The 1980s: Automation 15
- The 1990s: Volume Production 15

CAREERS IN SEMICONDUCTOR MANUFACTURING 16
- Technician 16
- Job Descriptions 18

SUMMARY 19
KEY TERMS 19
REVIEW QUESTIONS 20
SELECTED INDUSTRY WEB SITES 20
REFERENCES 20

CHAPTER 2 CHARACTERISTICS OF SEMICONDUCTOR MATERIALS

OBJECTIVES 21
INTRODUCTION 21
ATOMIC STRUCTURE 22
- Electrons 22

THE PERIODIC TABLE 24
- Ionic Bonds 26
- Covalent Bonds 28

CLASSIFYING MATERIALS 28
- Conductors 29
- Insulators 30
- Semiconductors 32

SILICON 33
- Pure Silicon 33
- Why Silicon? 33
- Doped Silicon 34
- pn Junctions 39

ALTERNATIVE SEMICONDUCTOR MATERIALS 39
- Gallium Arsenide (GaAs) 40

SUMMARY 41
KEY TERMS 41
REVIEW QUESTIONS 42
REFERENCES 42

CHAPTER 3 DEVICE TECHNOLOGIES

OBJECTIVES 43
INTRODUCTION 43
CIRCUIT TYPES 44
- Analog Circuits 44
- Digital Circuits 44

PASSIVE COMPONENT STRUCTURES 45
- IC Resistor Structures 45
- IC Capacitor Structures 46

ACTIVE COMPONENT STRUCTURES 46
- The pn Junction Diode 46
- The Bipolar Junction Transistor 49
- Schottky Diode 51
- Bipolar IC Technology 52
- CMOS IC Technology 52

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 662

Enhancement and Depletion-Mode MOSFETs 58
LATCHUP IN CMOS DEVICES 59
INTEGRATED CIRCUIT PRODUCTS 59
Linear IC Product Types 60
Digital IC Product Types 60
SUMMARY 62
KEY TERMS 62
REVIEW QUESTIONS 63
IC MANUFACTURERS' WEB SITES 64
REFERENCES 65

CHAPTER 4 SILICON AND WAFER PREPARATION

OBJECTIVES 67
INTRODUCTION 67
SEMICONDUCTOR-GRADE SILICON 68
CRYSTAL STRUCTURE 68
Amorphous Materials 69
Unit Cells 69
Polycrystal and Monocrystal Structures 70
CRYSTAL ORIENTATION 71
MONOCRYSTAL SILICON GROWTH 72
CZ Method 72
Float-Zone Method 75
Reasons for Larger Ingot Diameters 76
CRYSTAL DEFECTS IN SILICON 78
Point Defects 78
Dislocations 79
Gross Defects 80
WAFER PREPARATION 80
Shaping Operations 81
Wafer Slicing 82
Wafer Lapping and Edge Contour 82
Etching 83
Polishing 83
Cleaning 84
Wafer Evaluation 84
Packaging 84
QUALITY MEASURES 84
Physical Dimensions 85
Flatness 86
Microroughness 86
Oxygen Content 86
Crystal Defects 87
Particles 87
Bulk Resistivity 87
EPITAXIAL LAYER 87
SUMMARY 88
KEY TERMS 88
REVIEW QUESTIONS 89
SELECTED INDUSTRY WEB SITES 90
REFERENCES 90

CHAPTER 5 CHEMICALS IN SEMICONDUCTOR FABRICATION

OBJECTIVES 91
INTRODUCTION 91
STATES OF MATTER 91
PROPERTIES OF MATERIALS 92
Chemical Properties for Semiconductor Manufacturing 93
PROCESS CHEMICALS 99
Liquids 99
Gases 103
SUMMARY 109
KEY TERMS 109
REVIEW QUESTIONS 110
CHEMICAL SUPPLIERS' WEB SITES 110
REFERENCES 111

CHAPTER 6 CONTAMINATION CONTROL IN WAFER FABS

OBJECTIVES 113
INTRODUCTION 113
Clean Background 114
TYPES OF CONTAMINATION 114
Particles 115
Metallic Impurities 116
Organic Contamination 117
Native Oxides 118
Electrostatic Discharge 119
SOURCES AND CONTROL OF CONTAMINATION 120
Air 120
Humans 121
Facility 123
Water 126
Process Chemicals 130

Production Equipment 131
Workstation Design 131
WAFER WET CLEANING 135
Wet-Cleaning Overview 135
Wet-Clean Equipment 138
Alternatives to RCA Clean 142
SUMMARY 143
KEY TERMS 143
REVIEW QUESTIONS 144
CHEMICAL AND EQUIPMENT SUPPLIERS' WEB SITES 145
REFERENCES 146

CHAPTER 7 METROLOGY AND DEFECT INSPECTION

OBJECTIVES 149
INTRODUCTION 149
IC METROLOGY 150
Measurement Equipment 151
Yield 151
Data Management 152
QUALITY MEASURES 152
Film Thickness 153
Film Stress 158
Refractive Index 158
Dopant Concentration 159
Unpatterned Surface Defects 161
Patterned Surface Defects 164
Critical Dimension (CD) 165
Step Coverage 167
Overlay Registration 167
Capacitance-Voltage (C-V) Test 168
Contact Angle 171
ANALYTICAL EQUIPMENT 171
Secondary-Ion Mass Spectrometry (SIMS) 171
Atomic Force Microscope (AFM) 173
Auger Electron Spectroscopy (AES) 174
X-Ray Photoelectron Spectroscopy (XPS) 174
Transmission Electron Microscope (TEM) 175
Energy- and Wavelength-Dispersive Spectrometer (EDX and WDX) 176
Focused Ion Beam (FIB) 176
SUMMARY 177
KEY TERMS 178
REVIEW QUESTIONS 178
METROLOGY EQUIPMENT SUPPLIERS' WEB SITES 179
REFERENCES 180

CHAPTER 8 GAS CONTROL IN PROCESS CHAMBERS

OBJECTIVES 181
INTRODUCTION 181
VACUUM 183
Vacuum Ranges 183
Mean Free Path 184
VACUUM PUMPS 184
Roughing Pump 185
High Vacuum Pump 186
Vacuum in Integrated Tools 188
PROCESS CHAMBER GAS FLOW 189
Mass Flow Controllers 189
RESIDUAL GAS ANALYZER (RGA) 190
RGA Basics 190
RGA as Real-Time Monitor 191
PLASMA 192
Glow Discharge 193
PROCESS CHAMBER CONTAMINATION 195
SUMMARY 196
KEY TERMS 196
REVIEW QUESTIONS 196
VACUUM EQUIPMENT SUPPLIERS' WEB SITES 197
REFERENCES 197

CHAPTER 9 IC FABRICATION PROCESS OVERVIEW

OBJECTIVES 199
INTRODUCTION 199
CMOS PROCESS FLOW 199
Overview of Areas in a Wafer Fab 200
CMOS MANUFACTURING STEPS 205
1. Twin Well Process 205
2. Shallow Trench Isolation Process 207
3. Poly Gate Structural Process 210
4. Lightly Doped Drain (LDD) Implant Process 211
5. Sidewall Spacer Formation 212
6. Source/Drain (S/D) Implant Processes 213

7. Contact Formation 214
8. Local Interconnect (LI) Process 214
9. Via-1 and Plug-1 Formation 216
10. Metal-1 Interconnect Formation 217
11. Via-2 and Plug-1 Formation 218
12. Metal-2 Interconnect Formation 219
13. Metal-3 to Pad Etch and Alloy 220
14. Parametric Testing 221

SUMMARY 222
KEY TERMS 223
REVIEW QUESTIONS 223
REFERENCES 224

CHAPTER 10 OXIDATION

OBJECTIVES 225
INTRODUCTION 225
OXIDE FILM 226
- Nature of Oxide Film 226
- Uses of Oxide Film 227

THERMAL OXIDATION GROWTH 231
- Chemical Reaction for Oxidation 231
- Oxidation Growth Model 232

FURNACE EQUIPMENT 239
HORIZONTAL VERSUS VERTICAL FURNACES 240
- Vertical Furnace 241
- Fast Ramp Vertical Furnace 245
- Rapid Thermal Processor 246

OXIDATION PROCESS 248
- Pre Oxidation Cleaning 248
- Oxidation Process Recipe 249

QUALITY MEASUREMENTS 250
OXIDATION TROUBLESHOOTING 252
SUMMARY 252
KEY TERMS 253
REVIEW QUESTIONS 253
FURNACE AND RTP EQUIPMENT SUPPLIERS' WEB SITES 254
REFERENCES 255

CHAPTER 11 DEPOSITION

OBJECTIVES 257
INTRODUCTION 257
- Film Layering Terminology 258

FILM DEPOSITION 260
- Thin-Film Characteristics 260
- Film Growth 263
- Film Deposition Techniques 264

CHEMICAL VAPOR DEPOSITION 265
- CVD Chemical Processes 265
- CVD Reaction 266

CVD DEPOSITION SYSTEMS 269
- CVD Equipment Design 270
- APCVD (Atmospheric Pressure CVD) 271
- LPCVD (Low Pressure CVD) 273
- Plasma-Assisted CVD 277

DIELECTRICS AND PERFORMANCE 282
- Dielectric Constant 282
- Device Isolation 286

SPIN-ON-DIELECTRICS 287
- Spin-On-Glass (SOG) 287
- Spin-On-Dielectric (SOD) 288

EPITAXY 289
- Epitaxy Growth Methods 290

CVD QUALITY MEASURES 292
CVD TROUBLESHOOTING 292
SUMMARY 294
KEY TERMS 295
REVIEW QUESTIONS 295
DEPOSITION EQUIPMENT SUPPLIERS' WEB SITES 296
REFERENCES 296

CHAPTER 12 METALLIZATION

OBJECTIVES 299
INTRODUCTION 300
TYPES OF METALS 302
- Aluminum 302
- Aluminum-Copper Alloys 305
- Copper 305
- Barrier Metals 307
- Silicides 309
- Metal Plugs 312

METAL DEPOSITION SYSTEMS 313
- Evaporation 313
- Sputtering 314
- Metal CVD 320
- Copper Electroplate 323

METALLIZATION SCHEMES 325
- Traditional Aluminum Structure 325

Copper Damascene Structure 326
METALLIZATION QUALITY MEASURES 329
METALLIZATION TROUBLESHOOTING 330
SUMMARY 331
KEY TERMS 331
REVIEW QUESTIONS 332
METALLIZATION EQUIPMENT AND MATERIALS SUPPLIERS' WEB SITES 332
REFERENCES 333

CHAPTER 13 PHOTOLITHOGRAPHY: VAPOR PRIME TO SOFT BAKE

OBJECTIVES 335
INTRODUCTION 336
Photolithography Concepts 336
PHOTOLITHOGRAPHY PROCESSES 339
Negative Lithography 339
Positive Lithography 340
EIGHT BASIC STEPS OF PHOTOLITHOGRAPHY 342
Step 1: Vapor Prime 343
Step 2: Spin Coat 343
Step 3: Soft Bake 344
Step 4: Alignment and Exposure 344
Step 5: Post-Exposure Bake (PEB) 344
Step 6: Develop 344
Step 7: Hard Bake 344
Step 8: Develop Inspect 345
VAPOR PRIME 345
Wafer Cleaning 345
Dehydration Bake 346
Wafer Priming 346
SPIN COAT 348
Photoresist 348
Photoresist Physical Properties 349
Conventional I-Line Photoresists 351
Deep UV (DUV) Photoresists 354
Photoresist Dispensing Methods 357
SOFT BAKE 360
Soft Bake Equipment 360
Process Characterization 361
PHOTORESIST QUALITY MEASURES 362
PHOTORESIST TROUBLESHOOTING 363
SUMMARY 364
KEY TERMS 364
REVIEW QUESTIONS 364
PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES 365
REFERENCES 366

CHAPTER 14 PHOTOLITHOGRAPHY: ALIGNMENT AND EXPOSURE

OBJECTIVES 367
INTRODUCTION 367
Importance of Alignment and Exposure 368
OPTICAL LITHOGRAPHY 370
Light 370
Exposure Sources 372
Optics 376
Resolution 385
PHOTOLITHOGRAPHY EQUIPMENT 388
Contact Aligner 388
Proximity Aligner 389
Scanning Projection Aligner 390
Step-and-Repeat Aligner (Stepper) 390
Step-and-Scan System 393
Reticles 394
Optical Enhancement Techniques 398
Alignment 400
Environmental Conditions 404
Comparison of Photo Tools 405
MIX AND MATCH 406
ALIGNMENT AND EXPOSURE QUALITY MEASURES 407
ALIGNMENT AND EXPOSURE TROUBLESHOOTING 408
SUMMARY 409
KEY TERMS 410
REVIEW QUESTIONS 410
PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES 411
REFERENCES 412

CHAPTER 15 PHOTOLITHOGRAPHY: PHOTORESIST DEVELOPMENT AND ADVANCED LITHOGRAPHY

OBJECTIVES 413
INTRODUCTION 413
Advanced Lithography 414
POST-EXPOSURE BAKE 414
DUV Post-Exposure Bake (PEB) 415
Conventional I-Line PEB 416
DEVELOP 416
Negative Resist 417
Positive Resist 418
Development Methods 419
Resist Development Parameters 421
HARD BAKE 422
DEVELOP INSPECT 422
ADVANCED LITHOGRAPHY 424
Next-Generation Lithography 424
Advanced Resist Processing 428
DEVELOP QUALITY MEASURES 429
DEVELOP TROUBLESHOOTING 431
SUMMARY 432
KEY TERMS 433
REVIEW QUESTIONS 433
PHOTOLITHOGRAPHY MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES 433
REFERENCES 434

CHAPTER 16 ETCH

OBJECTIVES 435
INTRODUCTION 435
Etch Processes 436
ETCH PARAMETERS 437
Etch Rate 437
Etch Profile 438
Etch Bias 439
Selectivity 440
Uniformity 441
Residues 441
Polymer Formation 442
Plasma-Induced Damage 442
Particle Contamination 443
DRY ETCH 443
Etching Action 444
Potential Distribution 444
PLASMA ETCH REACTORS 446
Barrel Plasma Etcher 446
Parallel Plate (Planar) Reactor 447
Downstream Etch Systems 448
Triode Planar Reactor 448
Ion Beam Milling 448
Reactive Ion Etch (RIE) 450
High-Density Plasma Etchers 450
Etch System Review 453
Endpoint Detection 455
Vacuum for Etch Chambers 456
DRY ETCH APPLICATIONS 456
Dielectric Dry Etch 457
Silicon Dry Etch 459
Metal Dry Etch 462
WET ETCH 464
Types of Wet Etch 465
HISTORICAL PERSPECTIVE 466
PHOTORESIST REMOVAL 466
Plasma Ashing 466
ETCH INSPECTION 469
ETCH INSPECTION QUALITY MEASURES 469
DRY ETCH TROUBLESHOOTING 470
SUMMARY 471
KEY TERMS 472
REVIEW QUESTIONS 472
ETCH EQUIPMENT SUPPLIERS' WEB SITES 473
REFERENCES 474

CHAPTER 17 ION IMPLANT

OBJECTIVES 475
INTRODUCTION 475
Doped Regions 477
DIFFUSION 479
Diffusion Principles 479
Diffusion Process 481
ION IMPLANTATION 482
Overview 483
Ion Implant Parameters 484
ION IMPLANTERS 488
Ion Source 488
Extraction and Ion Analyzer 490
Acceleration Column 491

Scanning System 494
Process Chamber 498
Annealing 499
Channeling 501
Particles 502
ION IMPLANT TRENDS IN PROCESS INTEGRATION 503
Deep Buried Layers 503
Retrograde Wells 503
Punchthrough Stoppers 504
Threshold Voltage Adjustment 504
Lightly Doped Drain 505
Source/Drain Implants 505
Polysilicon Gate 506
Trench Capacitor 506
Ultrashallow Junctions 506
Silicon-On-Insulator (SOI) 507
ION IMPLANT QUALITY MEASURES 508
ION IMPLANT TROUBLESHOOTING 509
SUMMARY 510
KEY TERMS 511
REVIEW QUESTIONS 512
ION IMPLANTER EQUIPMENT SUPPLIERS' WEB SITES 513
REFERENCES 513

CHAPTER 18 CHEMICAL MECHANICAL PLANARIZATION

OBJECTIVES 515
INTRODUCTION 515
TRADITIONAL PLANARIZATION 518
Etchback 518
Glass Reflow 519
Spin-On Films 520
CHEMICAL MECHANICAL PLANARIZATION 520
CMP Planarity 521
Advantages of CMP 522
CMP Mechanisms 523
CMP Slurry and Pad 526
CMP Equipment 529
CMP Clean 532
CMP Equipment Manufacturers 534
CMP APPLICATIONS 535
STI Oxide Polish 535
LI Oxide Polish 535
LI Tungsten Polish 536
ILD Oxide Polish 536
Tungsten Plug Polish 536
Dual-Damascene Copper Polish 537
CMP QUALITY MEASURES 538
CMP TROUBLESHOOTING 540
SUMMARY 541
KEY TERMS 541
REVIEW QUESTIONS 542
CMP EQUIPMENT SUPPLIERS' WEB SITES 542
REFERENCES 543

CHAPTER 19 WAFER TEST

OBJECTIVES 545
INTRODUCTION 545
IC Electrical Tests 546
WAFER TEST 547
In-Line Parametric Test 547
Wafer Sort 555
Yield 560
Wafer Sort Yield Models 563
TEST QUALITY MEASURES 565
TEST TROUBLESHOOTING 567
SUMMARY 567
KEY TERMS 568
REVIEW QUESTIONS 568
TEST AND PROBER EQUIPMENT SUPPLIERS' WEB SITES 569
REFERENCES 569

CHAPTER 20 ASSEMBLY AND PACKAGING

OBJECTIVES 571
INTRODUCTION 571
Packaging Levels 573
TRADITIONAL ASSEMBLY 574
Backgrind 574
Die Separation 575
Die Attach 575
Wirebonding 577
TRADITIONAL PACKAGING 580
Plastic Packaging 581
Ceramic Packaging 584
Final Test 586

ADVANCED ASSEMBLY AND PACKAGING 586
- Flip Chip 587
- Ball Grid Array (BGA) 589
- Chip on Board (COB) 590
- Tape Automated Bonding (TAB) 591
- Multichip Module (MCM) 591
- Chip Scale Packaging (CSP) 592
- Wafer-Level Packaging 592

ASSEMBLY AND PACKAGING QUALITY MEASURES 596

IC PACKAGING TROUBLESHOOTING 597

SUMMARY 597

KEY TERMS 598

REVIEW QUESTIONS 598

ASSEMBLY AND PACKAGING SUPPLIERS' WEB SITES 599

REFERENCES 600

APPENDICES

APPENDIX A CHEMICALS AND SAFETY 601

APPENDIX B CONTAMINATION CONTROLS IN CLEANROOMS 610

APPENDIX C UNITS 614

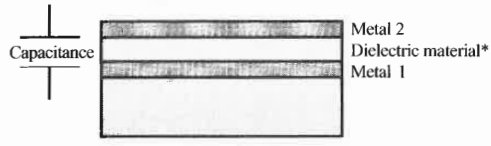
APPENDIX D COLOR AS A FUNCTION OF OXIDE THICKNESS 617

APPENDIX E OVERVIEW OF PHOTORESIST CHEMISTRY 618

APPENDIX F ETCH CHEMISTRY 622

GLOSSARY 624

INDEX 648



** Low k dielectric reduces capacitance between metal layers.*

FIGURE 2.17 Low-*k* Dielectric Material

action occurs when a semiconductor is heated, thus improving conductivity with an increase in temperature (which is the opposite condition for a conductor).

The most important semiconductor material in wafer fabrication is silicon, used as the semiconductor wafer substrate in over 85% of all chips.[3] This text will only address silicon because of its predominant use in the semiconductor industry. A short review of some of the other semiconductor materials and their applications is given at the end of this chapter.

SILICON

Silicon is an elemental semiconductor material because it has four valence-shell electrons, along with the other elements located in Group IVA of the periodic table (see Figure 2.18 on page 34). The number of valence-shell electrons in silicon places it midway between the valence-shell condition of a good conductor (one valence electron) and an insulator (eight valence electrons).

Silicon is not found pure in nature. It must be refined and purified to become pure for semiconductor manufacturing (see Chapter 4). It usually occurs as silica (silicon dioxide, or SiO_2) and other silicates. Silica is sand and is a primary ingredient of glass. Other forms of SiO_2 are rock crystal, quartz, agate, and opal.

The melting point of silicon is 1412°C. Silicon is a hard and brittle material that fractures easily if deformed, much like a piece of glass. It can be polished to a mirrorlike finish. Silicon exhibits many of the same characteristics as metals and at the same time exhibits characteristics of nonmetals. This is why silicon is classified as a semiconductor, midway between the conductors (metals) and insulators (nonmetals) on the periodic table.

Pure Silicon

Pure silicon is referred to as *intrinsic silicon,* with no contaminants or impurities from other substances. The silicon atoms in pure silicon bond together through covalent bonds to share electrons and complete their valence shells (see Figure 2.19 on page 34).

Many of silicon's properties result from this strong covalent bonding. The covalent bonds in pure silicon hold the atoms together to form a solid, electrically stable material that is an insulator. Pure silicon is a poor conductor because all valence shells are fully occupied in covalent bonds. In this pure form, silicon is not useful as a semiconductor.

The solid material formed when two or more silicon atoms bond together in this set, repeatable pattern, as the one shown in Figure 2.19 on page 34, is referred to as a *crystal.* A crystal is a smooth, glassy solid that forms a lattice structure with a three-dimensional shape. An example of a crystal material is window glass. We will study more about crystals in Chapter 4 when we discuss silicon wafer preparation.

Why Silicon?

Germanium was the first material used for semiconductors in the 1940s and early 1950s, but as we learned in Chapter 1, it was soon replaced by silicon. Why was silicon chosen as the predominant semiconductor material? There are four main reasons for the selection of silicon as the primary semiconductor material:

- Abundance of silicon
- Higher melting temperature for wider processing range

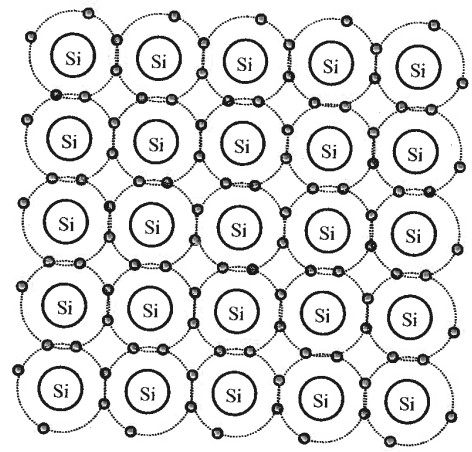
Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 670

- Wider temperature range of operation
- Natural growth of silicon dioxide

Semiconductors

Group IVA	
C, Carbon	6
Si, Silicon	14
Ge, Germanium	32
Sn, Tin	50
Pb, Lead	82

FIGURE 2.18 Group IVA Elemental Semiconductors



Silicon atoms share valence electrons to form insulator-like bonds.

FIGURE 2.19 Covalent Bonding of Pure Silicon

Silicon is the second most abundant element on earth and makes up about 25% of the earth's crust. If processed properly, silicon can be refined into ample quantities of the very pure form necessary for semiconductor fabrication which leads to lower costs. Silicon's melting temperature of 1412°C is much higher than the melting temperature for germanium which is 937°C. This higher melting temperature permits silicon to withstand high-temperature processing. Another advantage to using silicon is that a semiconductor device made from silicon can function over a wider temperature range than germanium, which increases the semiconductor's applications and reliability.

Finally, an important reason for using silicon as a semiconductor material is the ability to naturally grow *silicon dioxide* (SiO_2) on the silicon surface (see Figure 2.20). SiO_2 is a high-quality, stable electrical insulator material that also serves as a good chemical barrier to protect silicon from external contaminants.[4] Electrical stability is important to avoid leakage between adjacent conductors in an IC. The ability to grow stable, thin SiO_2 material is fundamental to the fabrication of high-performance metal-oxide semiconductor (MOS) devices. SiO_2 has mechanical properties similar to silicon, which allows high temperature processing without excessive wafer warpage.

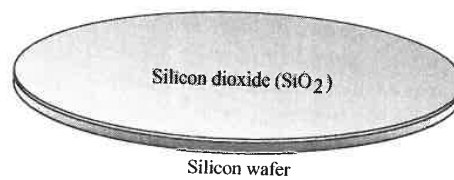


FIGURE 2.20 SiO_2 on Silicon Wafer

Doped Silicon

Silicon in its pure state has little use in semiconductor technology. However, the structure of silicon can be altered to greatly enhance its conductivity by adding small amounts of other elements to the material through a process known as doping. *Doping* is the process of adding certain elements to pure silicon to improve the conductivity of the semiconductor (see Figure 2.21). For example, the electrical resistivity (ρ) of pure silicon is approximately 2.5×10^5 ohms-cm. If only one in

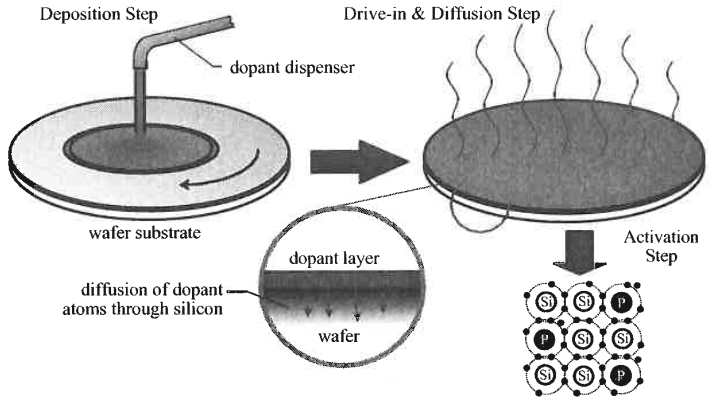


FIGURE 2.21 Doping of Silicon
(Used with permission from Advanced Micro Devices)

every one million silicon atoms is replaced by one atom of arsenic, resistivity will drop to 0.2 ohms-cm.[5] This is an improvement in conductivity of 1,250,000 times.

The elements added during doping are referred to as *dopants* or *impurities* because the silicon is no longer pure. In other words, we dope the silicon with an impurity so that it will conduct electricity. The more impurity added then the higher the conductivity (or the lower the resistivity). Note that we use the term impurity to indicate that we have added another element to the silicon. We intentionally add the impurity to increase the conductivity of silicon. Doped silicon is also known as *extrinsic silicon.*

The concept of doping the silicon with an impurity to improve electrical conductivity is a critical aspect of semiconductor fabrication. If we can introduce impurities that alter the silicon's resistivity and then control when the silicon performs as an insulator or as a conductor, then we have the essence of solid-state technology.

Dopant Materials ■ Silicon is located in Group IVA of the periodic table and has four valence electrons. Elements from the two adjacent groups are commonly used for doping: Group IIIA and Group VA (see Figure 2.22). Group IIIA elements are referred to as *trivalent* because of three valence electrons, whereas elements from group VA are *pentavalent* because of five valence electrons. The trivalent dopant increases the number of holes (positive doping or p-type), whereas the pentavalent will increase the number of free electrons (negative doping or n-type).

Acceptor Impurities		Semiconductor		Donor Impurities	
Group III (p-type)		**Group IV**		**Group V (n-type)**	
Boron	**5**	Carbon	6	Nitrogen	7
Aluminum	13	**Silicon**	**14**	**Phosphorus**	**15**
Gallium	31	Germanium	32	**Arsenic**	**33**
Indium	49	Tin	50	**Antimony**	**51**

* *Elements underlined are the most commonly used in silicon-based IC manufacturing.*

FIGURE 2.22 Silicon Dopants

When trivalent dopant atoms are added to silicon, the resulting material is called a *p-type silicon.* The trivalent dopants are referred to as *acceptors* (they accept an extra mobile electron), with the most common acceptor element being boron. When a pentavalent element is added to pure silicon, the resulting material is referred to as an *n-type silicon.* Pentavalent dopants are referred to as

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 672

donors (they donate an extra mobile electron), and are typically either phosphorus, arsenic, or antimony. The different doping elements are listed in Table 2.2.

TABLE 2.2 Dopant Elements Commonly Used in Semiconductor Manufacturing

Trivalent Dopants (p-type, positive doping, acceptor)	Pentavalent Dopants (n-type, negative doping, donor)
Boron (B)	Phosphorus (P)
	Arsenic (As)
	Antimony (Sb)

n-Type Silicon. For an n-type silicon, there are more electrons than valence-band holes. This is shown in Figure 2.23 for silicon with a pentavalent phosphorous dopant atom.

The silicon atoms will form covalent bonds with the donor phosphorous atom, each sharing one phosphorous electron. However, the fifth phosphorous electron is not bound to any of the surrounding silicon atoms. Because of this, the fifth phosphorous electron needs little energy to break away and enter the conduction band. For n-type silicon, conduction-band free electrons are the majority carriers, of which the material has an abundance. There are also many fewer minority carrier valence-band holes.

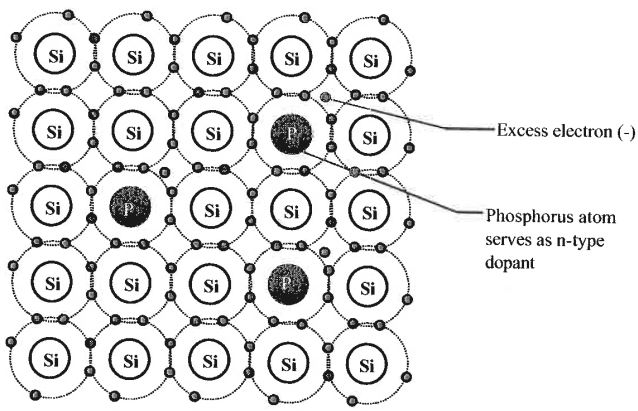


FIGURE 2.23 Electrons in n-Type Silicon with Phosphorous Dopant

One electron in the conduction band is not significant for conduction. However, when we dope silicon with a dopant, we add literally millions of dopant atoms, producing many electrons that are not part of the covalent bonds. There is a high amount of movement between electrons and holes. Negative electrons are attracted to the positive holes. Electrons enter the conduction band relatively easily. If a voltage is applied to this material, the electrons can be made to flow through the material as electric current (see Figure 2.24).

Note that the doped silicon is still electrically neutral (this is true for n-type or p-type silicon). In the case of n-type silicon, this is because each phosphorous atom still has the same number of protons as electrons, as do the silicon atoms. Thus the overall number of protons and electrons in the semiconductor is still equal and the result is a net charge of zero. What is not equal is that there are many more conduction-band electrons (majority carriers) than there are valence-band holes (minority carriers).

p-Type Silicon. In the p-type silicon shown in Figure 2.25, the boron atom is a p-type acceptor dopant that forms a covalent bond with four adjacent silicon atoms. The boron acceptor atoms provide a deficiency of electrons caused by the lack of a fourth electron in the boron

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 673

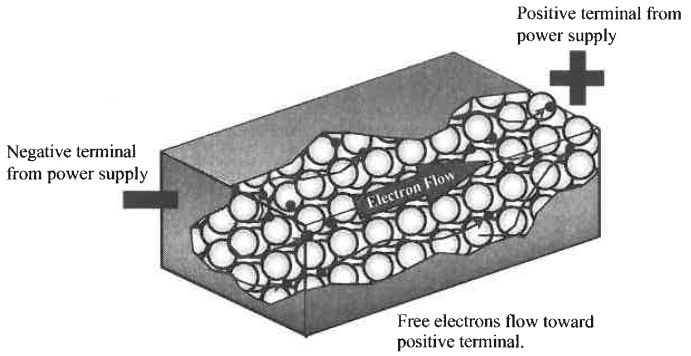


FIGURE 2.24 Flow of Free Electrons in n-Type Silicon

atom, thus creating the p-type silicon. There is an excess of holes (absence of electrons) in the valence shell.

Since there are many more valence-band holes than conduction-band electrons, the holes are the major current carriers in p-type material. The holes are referred to as *majority carriers* and electrons are the *minority carriers.* If a dc voltage is applied across a p-type semiconductor, then the large number of holes in the material will attract electrons from the negative terminal of the voltage source into the p-type semiconductor. This is current flow in a p-type material (see Figure 2.26 on page 38). The holes appear to move because each time an electron moves into a hole it serves to create a hole in its previous position. The holes seemingly move in the opposite direction as the electrons.

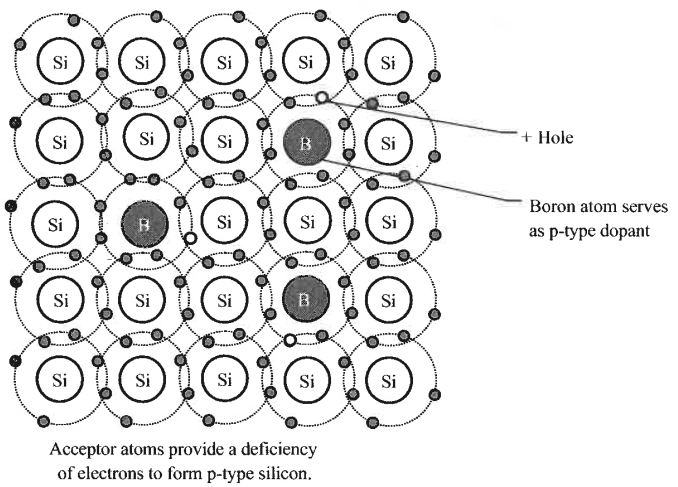


FIGURE 2.25 Holes in p-Type Silicon with Boron Dopant

Resistivity of Doped Silicon. We obtain precise control of silicon's resistivity by introducing dopants into the silicon crystal structure. The concentration of the dopant atom placed in the silicon crystal defines how well the material will conduct electrical current. Pure silicon has a resistivity of about 250,000 Ω-cm and is an insulator. Compare this to copper, a good conductor, with a resistivity of 1.7 μΩ-cm (0.0000017 Ω-cm). By adding the proper type and concentration of dopants to pure silicon, the doped silicon's resistivity is decreased and conductivity is improved (see Figure 2.27 on page 38). For a given resistivity there is less concentration of n-type dopants than p-type dopants. This is because it takes less energy to move an electron than to move a hole.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 674

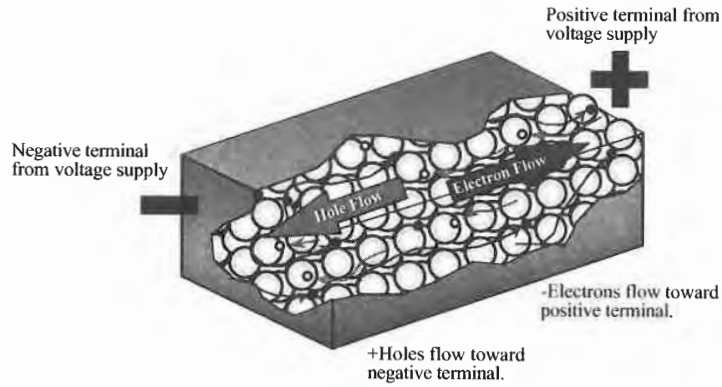


FIGURE 2.26 Flow of +Holes in p-Type Silicon

It only takes a small amount of dopant (from as little as 0.000001% to 0.1%) to make silicon a useful conductor. This is important for the fabrication of semiconductor devices on wafers. The amount of dopant in the silicon, or concentration, is carefully controlled during semiconductor fabrication to attain a precise resistivity. The ion implantation process for adding dopants to silicon will be covered in Chapter 17.

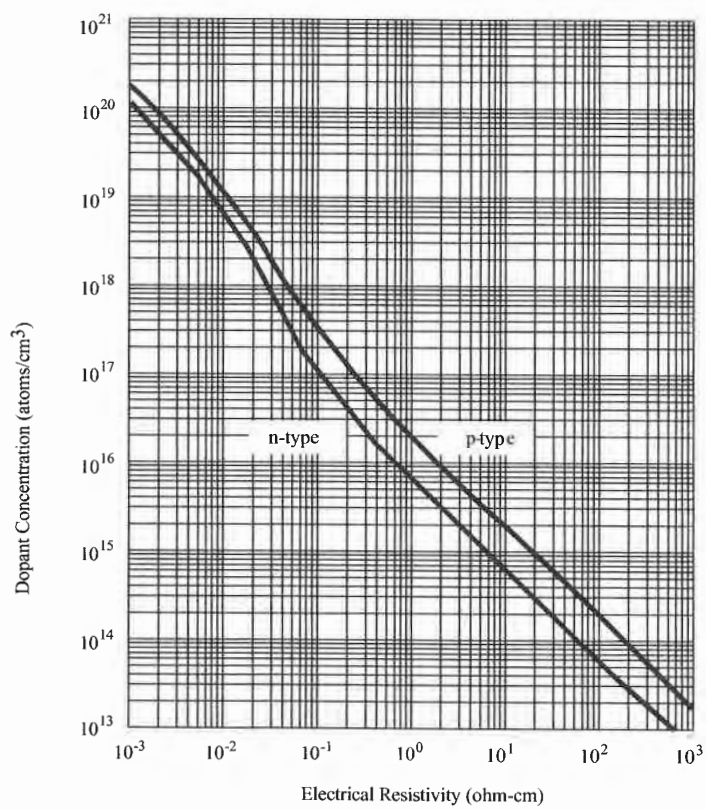


FIGURE 2.27 Silicon Resistivity Versus Dopant Concentration
Redrawn from S. Ghandi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* John Wiley & Sons, 1994.

pn Junctions

We dope pure silicon with pentavalent or trivalent elements to obtain either an n-type or p-type semiconductor. The type and concentration of the dopant element determines whether conduction occurs by hole or by electron and determines the final value of the silicon's resistivity. The number of carriers, and therefore the resistivity, is determined by the net donor and acceptor atom concentration in the silicon crystal. Because of this, an n-dopant can be implanted or diffused into a p-region to convert that region into an n-region (or vice versa). To achieve this, the concentration of the n-type donor dopant must exceed that of the p-type acceptor dopants in the region.

The ability to incorporate both n-type and p-type regions in the silicon crystal is beneficial because semiconductor devices require both of these regions in order to function as a useful electronic device. It is the junction between the n-type and p-type regions that is important and creates the useful characteristics of silicon as a semiconductor. This junction is referred to as a *pn junction* (see Figure 2.28).

The pn junction is the essence of solid-state electronics and is the basis for how semiconductor wafers can achieve their unique ability to act as an insulator and a conductor, depending on the bias voltage applied to the junction. The specifics of how a pn junction functions as a useful electronic device are discussed in Chapter 3. The creation of a pn junction is almost universally done in wafer fabrication with ion implantation (see Chapter 17).

Remember that a pn junction is formed between two pieces of essentially the same material. The p-type and the n-type materials scarcely differ except for a minute amount of a dopant. The n-type material has extra mobile electrons from the donor impurity, whereas the p-type material has extra mobile holes. It is deceptive to speak of one material meeting the other.[6] The junction is so intimate that the n-type and p-type materials are formed from one continuous solid. The silicon crystal with a pn junction continues to look like and in most ways behave like the solid crystal material that it is.

The depth and definition of this junction is critical in semiconductor manufacturing. As device critical dimensions are reduced, the ability to precisely control both the pn junction in silicon (e.g., junction depth) and the concentration of dopants is becoming a major challenge for semiconductor chip fabrication.

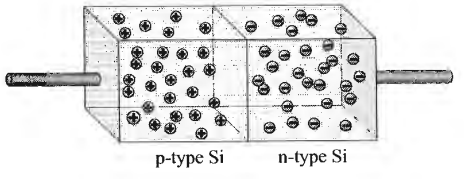


FIGURE 2.28 Cross Section of Planar pn Junction

ALTERNATIVE SEMICONDUCTOR MATERIALS

Germanium and silicon are the two elemental semiconductor materials from Group IVA with four valence electrons. We have learned that germanium was the first semiconductor material used for transistor fabrication and was replaced by silicon in the 1950s for process and performance reasons.

There are alternative semiconductor materials used for specific market applications. These are primarily the *compound semiconductors*. A major class of compound semiconductors are formed from Group IIIA and Group VA of the periodic table (often referred to as III-V compounds). An example is gallium arsenide (GaAs). In addition, other semiconductor compounds come from elements found in Groups IIA and VIA, and are referred to as II-VI compounds.

The need for alternative semiconductor materials is driven by the growth of markets that require IC performance beyond that capable by silicon semiconductors. The key IC performance factor is speed. Wireless and high-speed digital communications, space applications, and consumer markets such as the automotive industry, are developing special niche semiconductor markets that

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 676

Oxide growth is a natural phenomenon that occurs by exposing the silicon wafer to oxygen at an elevated temperature. The ability to grow an oxide film on silicon is often listed as a significant reason for the use of silicon as the most common semiconductor substrate material (another major reason is silicon's relatively high melting temperature). We use the term "grow" to indicate that temperature is used to cause the oxide to grow out of the silicon semiconductor material, actually consuming silicon in the process.

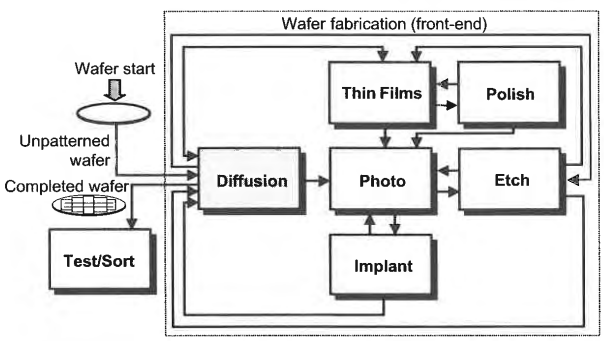


FIGURE 10.1 Diffusion Area of Wafer Fabrication
(Used with permission from Advanced Micro Devices)

The amount of thermal exposure (i.e., temperature multiplied by time) for a wafer during processing is referred to as the *thermal budget.* The thermal budget requirement for wafer fabrication is rapidly dropping.[2] A goal throughout semiconductor processing is to minimize the amount of thermal (heat) exposure to the wafer. As the critical dimension on device structures decreases to 0.18 μm and beyond, scaling requirements dictate shallower junction depths. To minimize the unacceptable diffusion of dopants out of a shallow junction region, the thermal budget must decrease accordingly. Another problem caused by excessive thermal budget is the increase in ohmic contact resistance of metal-layer interconnects (see Chapter 12), which increases the overall resistance of the conductive paths. A factor in determining the process conditions for many wafer fabrication process steps is the ability to minimize the thermal budget, either by reduced temperature or by minimizing the time at temperature.

OXIDE FILM

An oxide layer is grown on silicon typically at thermal oxidation temperatures between 750°C to 1100°C. An oxide layer grown on silicon is referred to as *thermal oxide* or *thermal silicon dioxide* (SiO_2). Since silicon dioxide is an oxide material, the two terms are used interchangeably. Another term for silicon dioxide is *glass*. Silicon dioxide is a dielectric material and will not conduct electricity.

Nature of Oxide Film

When a silicon surface is exposed to oxygen, an amorphous silicon dioxide film grows immediately. Although undoped silicon is a semiconductor material, SiO_2 is an insulator. The atomic structure of this SiO_2 film consists of a silicon atom surrounded by four oxygen atoms (see Figure 10.2). We can refer to this atomic structure as a silicon dioxide *tetrahedron cell.* Amorphous SiO_2 has no long-range periodic crystal order at the atomic level. Long-range order is absent because the tetrahedra are not arranged in a regular, three-dimensional array as in a crystal.[3]

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 677

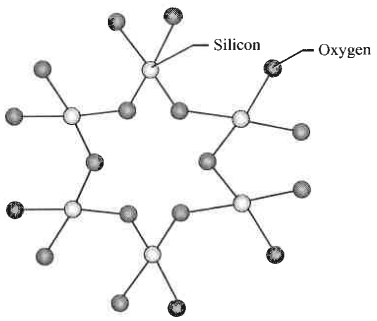


FIGURE 10.2 Atomic Structure of Silicon Dioxide
(Used with permission from International SEMATECH)

Silicon dioxide is a form of intrinsic (pure) glass with a melting temperature of 1732°C.[4] Thermally grown SiO_2 has strong adhesion to silicon and exhibits excellent dielectric properties. A layer of SiO_2 always covers the wafer surface because silicon oxidizes immediately upon exposure to air with a few monolayers of native oxide (a monolayer is one layer of atoms on a surface). Over time, the native oxide layer will thicken to an upper limit of about 40 Å, even at 25°C room temperature.[5] This type of oxide is nonuniform and usually considered a contaminant, with little use in semiconductor fabrication. There are some applications for use of native oxide in wafer fabrication, such as part of a dielectric film stack in memory cells. Native oxide grown from air is contaminated, but native oxide grown in chemical baths using high-purity chemicals is very clean.

Uses of Oxide Film

Oxide is important for silicon semiconductor fabrication because of its ease of formation and its excellent interface with the underlying silicon material. It is the most common layer used in semiconductor fabrication. Different ways in which an oxide layer is used to fabricate a microchip are:

- Device scratch protection and contaminant isolation
- Field isolation to confine charged carriers (termed surface passivation)
- Dielectric material in the gate oxide or memory cell structures
- Impurity-mask barrier during doping
- Dielectric layer between metal conductor layers

Device Protection and Isolation ■ Silicon dioxide grown on the surface of the wafer serves as an effective barrier to isolate and protect sensitive devices in the silicon. SiO_2 physically protects devices because it is a very hard and nonporous (dense) material that effectively insulates active devices in the silicon surface. The hard SiO_2 layer will protect the silicon from scratches and processing damage that might occur during fabrication. Transistors have traditionally been electrically isolated by thermally growing thick SiO_2 in the region between them using the LOCOS process (discussed later in this chapter). However, this process is unacceptable for sub-0.25 μm technology and has been replaced by shallow trench isolation (STI). The STI process uses deposited oxide as the main dielectric (see Chapters 9 and 11).

Surface Passivation ■ A major benefit from thermally grown SiO_2 is the reduction in the surface-state density of silicon that occurs by tying up dangling silicon bonds. This action is referred to as *surface passivation,* which prevents the deterioration of the electronic properties and reduces current-leakage paths of the semiconductor caused by moisture, ions or other external contaminants.[6] The hard SiO_2 layer will protect the silicon from scratches and processing damage that might occur during backend final fabrication. SiO_2 growth on the silicon surface can serve to tie up electrically active contaminants (mobile ionic contaminants) in the oxide layer away from the active surface of the silicon. Passivation is also important for controlling the leakage current of

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 678

junction devices and growing a stable gate oxide.[7] For the oxide layer to function as a good passivation layer, it has a quality requirement to be a uniform thickness and not have pinholes or voids in the material.

Another factor in using the oxide layer to passify the wafer surface is the oxide thickness. Adequate oxide thickness is necessary to prevent electrical charging in a metal layer caused by a charge buildup in the silicon wafer surface, much like charge storage and breakdown characteristics of ordinary capacitors. This charging would lead to electrical shorting and other undesirable electrical effects. A thick layer of oxide that inhibits charge buildup from metal layers is the *field oxide layer,* which typically is between 2,500 Å to 15,000 Å thick (see Figure 10.3).

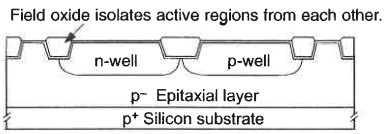


FIGURE 10.3 Field Oxide Layer

Silicon dioxide also has a coefficient of thermal expansion very similar to that of silicon. During high-temperature processing, the wafer will expand and then contract during cooling. The SiO_2 will expand and contract at close to the same rate as silicon, thus minimizing wafer warpage during the thermal excursions. This action also prevents film stress from separating the oxide film from the silicon substrate.

Gate Oxide Dielectric ■ An extremely thin layer of oxide is used as the dielectric material for the important *gate oxide structure* common in MOS technology (see Figure 10.4). The gate oxide is thermally grown because of its high quality and stability with the underlying semiconductor silicon. SiO_2 has a high dielectric strength of 10^7 volts/cm and high resistivity of approximately 10^{17} ohms-cm.[8]

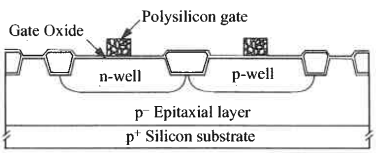


FIGURE 10.4 Gate Oxide Dielectric

The thickness of the gate oxide is chosen specifically for the scaling requirements of the device technology. For 0.18-μm generation technology, a typical gate oxide thickness is 20 ± 1.5 Å. The gate oxide has a specific thickness so that it scales properly with the entire gate structure to permit induction of a charge in the silicon wafer under the oxide. The widespread use of MOS technology in the ULSI era has made the formation of the gate oxide a primary concern in process development. A critical aspect of device reliability is *gate oxide integrity.* The gate structure in a MOS device is the part that controls the flow of electrical current through the device. Because this oxide is fundamental to the proper function of microchips based on field-effect technology, it is essential that it have high quality, with excellent film-thickness uniformity and no contaminants. Any contamination that degrades the proper functioning (integrity) of the gate oxide structure must be scrupulously controlled.

Dopant Barrier ■ Silicon dioxide can be used as an effective barrier to selectively introduce dopants into the silicon surface (see Figure 10.5). Once an oxide layer is grown on the silicon surface, then mask openings are etched into the SiO_2 to create windows where dopant materials can enter the wafer. The oxide protects the surface of silicon from impurity diffusion where there are

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 679

no openings and therefore allows selective impurity doping. Dopants have a slow rate of movement through SiO_2 when compared to silicon, requiring only a thin oxide layer to block dopants (note that this rate is temperature dependent).

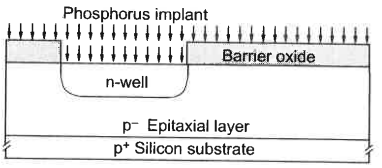


FIGURE 10.5 Oxide Layer Dopant Barrier

A thin oxide layer (e.g., 150 Å) is also grown in areas where ion implant occurs. This oxide screen serves to reduce damage to the silicon wafer and obtain better control over the depth that the dopant is implanted into the silicon by reducing channeling effects (see Chapter 17). Once implantation is done, HF (hydrofluoric acid) can be used to selectively remove the oxide, resulting in a planar silicon surface again.

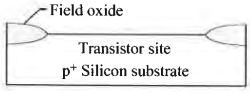
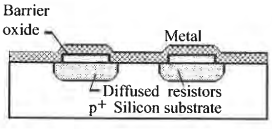
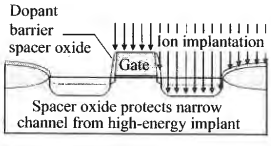
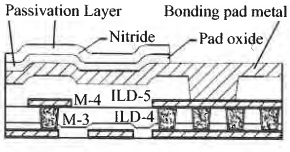
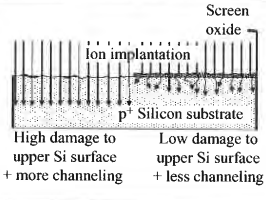
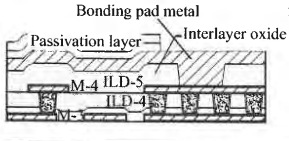
Dielectric Between Metal Layers ■ Under normal conditions, silicon dioxide will not conduct electricity. For this reason, SiO_2 is an effective insulator between metal layers on the microchip. Silicon dioxide prevents shorting of an upper metal layer to the underlying metal layer, just as the insulator material on an electrical cord prevents an electrical short. The oxide quality is critical with no holes or voids allowed. This oxide is often doped to produce more effective flow properties and to better minimize the diffusion of contaminants (i.e., it acts as a getter). It is typically deposited using chemical vapor deposition (not thermally grown) and will be covered in more detail in Chapter 11.

A summary of different types of oxides and their uses in semiconductor fabrication is shown in Table 10.1.[9]

TABLE 10.1 Oxide Applications

Application	Purpose	Structure	Comments
Native Oxide	This oxide is a contaminant and is generally undesirable. Sometimes used in memory storage or film passivation.	Silicon dioxide (oxide) / p^+ Silicon substrate	Growth rate at room temperature is 15 Å per hour up to a maximum thickness of about 40 Å.
Gate Oxide	Serves as a dielectric between the gate and source-drain parts of MOS transistor.	Gate oxide / Gate / Source / Drain / Transistor site / p^+ Silicon substrate	Common gate oxide film thickness range from about 20 Å to several hundred Å. Dry thermal oxidation is the preferred growth method.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 680

Application	Purpose	Structure	Comments
Field Oxide	Serves as an isolation barrier between individual transistors to isolate them from each other.		Common field oxide film thickness ranges from 2,500 Å to 15,000 Å. Wet oxidation is the preferred growth method.
Barrier Oxide	Protects active devices and silicon from follow-on processing.		Thermally grown to several hundred angstroms thickness.
Dopant Barrier	Masking material for depositing or implanting dopants into wafer.		Dopants diffuse into unmasked areas of silicon by selective diffusion.
Pad Oxide	Provides stress reduction for silicon nitride (Si_3N_4).		Thermally grown and very thin.
Implant Screen Oxide	Used to reduce implant channeling and damage.		Thermally grown.
Insulating Barrier Between Metal Layers	Serves as a protective layer between metal lines.		This oxide is not thermally grown, but is deposited.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 681

CHAPTER 17

ION IMPLANT

Intrinsic silicon crystal structure is formed by silicon covalent bonds. As explained in Chapter 2, intrinsic silicon is a poor conductor. Silicon only becomes useful as a semiconductor when its structure and conductivity are altered by adding small amounts of impurities called dopants. This process is referred to as doping. The doping of silicon is fundamental to pn junction fabrication in semiconductor devices. Ion implantation is the primary method for doping wafers.

Doping is widely used throughout wafer fabrication. Dopants are precisely and uniformly placed into the device to alter its electrical performance. Silicon chips require doping with Group IIIA and Group VA dopants for many different reasons, which will be discussed in this chapter.

The continual shrinking of chip critical dimensions and the subsequent increase in chip packing density force the scaling of all device elements. An example of this is the reduction of the polysilicon gate length to obtain a narrower channel region. A shorter channel length requires shallower junction depths for the source and drain doped regions. Junction depths are ultrashallow for chips designed at the 0.18 μm and below technology nodes. The junction depth requirements decrease to 30 ± 10 nm for devices with a CD of 0.1 μm.[1]

OBJECTIVES

After studying the material in this chapter, you will be able to:

1. Explain the purpose and applications for doping in wafer fabrication.
2. Discuss the principles and process of dopant diffusion.
3. Provide an overview of ion implantation, including its advantages and disadvantages.
4. Discuss the importance of dose and range in ion implant.
5. List and describe the five major subsystems for an ion implanter.
6. Explain annealing and channeling in ion implantation.
7. Describe different applications of ion implantation.

INTRODUCTION

Doping is the introduction of a dopant into the crystal structure of a semiconductor material to modify its electronic properties (e.g., electrical resistivity). As discussed in Chapter 2, certain Group IIIA and VA elements are used as dopants in wafer fabrication. For review, Table 17.1 on page 476 highlights the four principal dopants used in semiconductor manufacturing.

Dopant species are also referred to as impurities but should not be confused with contaminating impurities. Dopants are introduced into the silicon and other fabrication materials for a variety of reasons. For instance, dopants such as boron and phosphorus are used to form the majority carriers in silicon devices, to form conductive layers in the wafer, or to alter material properties (e.g., doping SiO_2 to form borophosphosilicate glass, or BPSG). The polysilicon gate electrode is doped for conductivity.

There are two techniques in wafer fabrication for introducing dopant elements to the wafer materials: thermal diffusion and ion implantation. *Thermal diffusion* uses high temperature to move the dopant through the silicon lattice structure. This method is dependent on time and temperature. *Ion implantation* introduces dopants into the substrate through a high-voltage ion bombardment. The dopants are implanted into the wafer through high-energy collisions at the atomic level. In the beginning of semiconductor manufacturing, thermal diffusion was the primary method for doping the wafer. However, due to shrinking critical dimensions and the corresponding scaling requirements, nearly all dopant processes in modern wafer fabrication are now done

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 682

TABLE 17.1 Common Dopants Used in Semiconductor Manufacturing

Acceptor Dopant Group IIIA (p-Type)		Semiconductor Group IVA		Donor Dopant Group VA (n-Type)	
Element	**Atomic Number**	**Element**	**Atomic Number**	**Element**	**Atomic Number**
Boron (B)	5	Carbon	6	Nitrogen	7
Aluminum	13	Silicon (Si)	14	Phosphorus (P)	15
Gallium	31	Germanium	32	Arsenic (As)	33
Indium	49	Tin	50	Antimony	51

Note: Common dopants are shaded.

with ion implantation (see Figure 17.1). Table 17.2 outlines some critical processes commonly doped with ion implantation versus the limited doping done by diffusion. Figure 17.1 shows these doped regions on a cross section drawing of a CMOS inverter. Note the dopant type (p or n) and concentration (− or +) are also shown. The different doped regions labeled A to O are defined in Table 17.2.

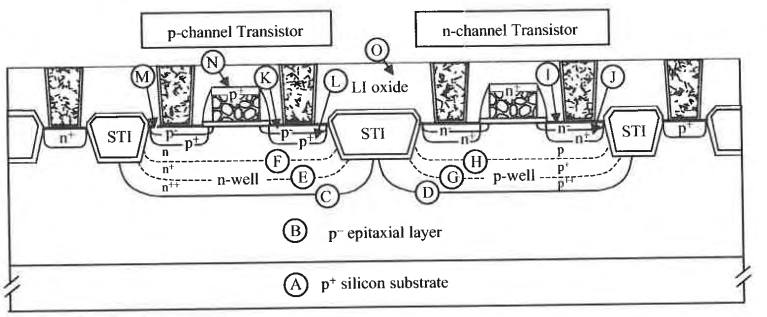


FIGURE 17.1 CMOS Structure with Doped Regions

TABLE 17.2* Common Dopant Processes CMOS Fabrication

Process Step	Common Dopant Species	Ion Implant or Diffusion	Remarks
A. p^+ Silicon Substrate	B	Diffusion	The monosilicon crystal is doped during the crystal growing process.
B. p^- Epitaxial Layer	B	Diffusion	The epitaxial layer is doped by diffusion during epitaxial silicon growth.
C. Retrograde n-well	P	Ion Implant	Retrograde well has dopant profile with peak concentration buried at a certain depth. The dopant concentration decreases as it approaches the surface.
D. Retrograde p-well	B	Ion Implant	Retrograde well has dopant profile with peak concentration buried at a certain depth.
E. p-Channel Punchthrough	P	Ion Implant	Phosphorus implanted more laterally than source/drain implants to keep electric field of drain from punching through p-channel and reaching source region.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 683

Process Step	Common Dopant Species	Ion Implant or Diffusion	Remarks
F. p-Channel Threshold Voltage (V_T) Adjust	P	Ion Implant	Implant phosphorus to adjust MOS V_T threshold voltage.
G. n-Channel Punchthrough	B	Ion Implant	Boron is implanted more laterally than source/drain implants to keep electric field of drain from punching through n-channel and reaching source region.
H. n-Channel V_T Adjust	B	Ion Implant	Implant boron through oxide layer to adjust MOS threshold voltage.
I. n-Channel Lightly Doped Drain (LDD)	As	Ion Implant	Low-dose implant of arsenic in region adjacent to the n-channel to improve electrical performance by reducing the peak electrical field and therefore hot carriers, which minimizes interface charges trapped in the gate oxide.
J. n-Channel Source/Drain (S/D)	As	Ion Implant	Higher implant dose of arsenic to form n-channel source and drain.
K. p-Channel LDD	BF_2	Ion Implant	Low-dose implant of boron in region adjacent to the p-channel to improve electrical performance between drain and channel region.
L. p-Channel S/D	BF_2	Ion Implant	Higher implant dose of boron to form p-channel source and drain.
M. Silicon	Si	Ion Implant	Implant with nondopant atom to amorphize silicon to reduce transient enhanced diffusion (TED) and channeling effects.
N. Doped Polysilicon	P or B	Ion Implant or Diffusion	Doping of polysilicon gate electrode to make conductive.
O. Doped SiO_2	P or B	Ion Implant or Diffusion	Doped oxide to obtain material benefits (e.g., create better flow and serve as getter).

*Adapted from E. Rimini, *Ion Implantation: Basics to Device Fabrication,* (Boston: Kluwer Academic Publishers, 1995).

Ion implantation typically follows photolithography in the general process flow (see Figure 17.2 on page 478); although, there may be some processes that call for the use of an etched-oxide layer to serve as a mask during either diffusion doping or ion implant. In this case, wafers could flow from etch to diffusion or ion implant. Ion implantation plays a crucial role in chip performance improvement because of the need for smaller dimensions. Transistor performance is increasingly dependent on precise doping profiles in the silicon that are achievable only with ion implant.

Doped Regions

Recall from Chapter 4 that a wafer is uniformly doped with dopant atoms during the crystal growth process to create a p-type or n-type wafer. The type of dopant depends on the manufacturer's specification. During wafer fabrication, dopants are selectively introduced to create the devices on the individual wafers. The selective introduction of dopants into masked openings on the wafer forms a *doped region* where dopants reside within the silicon crystal structure (see Figure 17.3 on page 478). The doped region is characterized by its *dopant profile,* which determines the amount of dopant actually added in the doped region as a function of depth.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 684

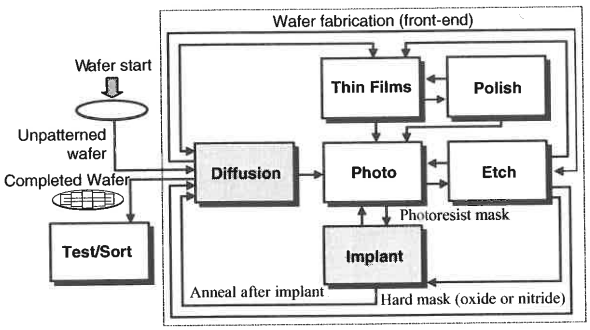


FIGURE 17.2 Ion Implant in Wafer Process Flow
(Used with permission from Advanced Micro Devices)

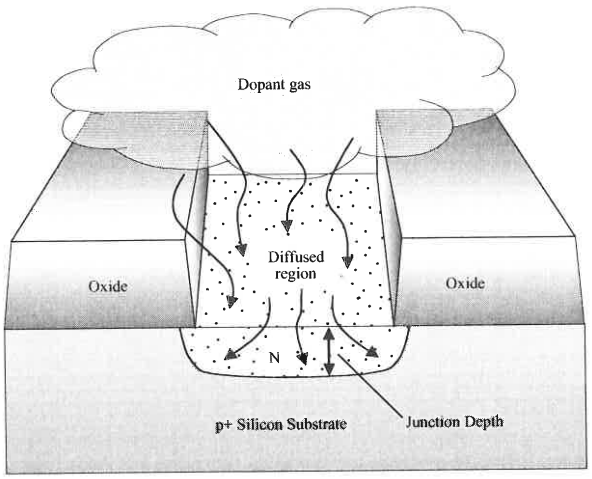


FIGURE 17.3 Doped Region in a Silicon Wafer

The dopant atoms enter the solid silicon through an oxide mask opening. In ion implantation, this mask is often photoresist, but a variety of masks can be used since it is a low-temperature process. Doping by diffusion requires an oxide or nitride mask because this technique is done in a high-temperature furnace.

The doped region can be of the opposite type as the silicon substrate, such as an n-type doped region in a p-type wafer, or it can be the same type with a different dopant concentration (e.g., a heavily doped p^+ region in a p-type wafer). The location where the conductivity of the doped region makes a change from p-type to n-type, or vice versa, is the pn junction. The exact depth in the wafer where the p-type dopant meets the n-type dopant is the *junction depth*, x_j. At the precise location of the junction depth, the concentration of electrons equals the concentration of holes. Another way of looking at this feature is the net dopant concentration at x_j is zero.

The ultimate interest in wafer doping is related to the individual dopant profiles for the devices after all processing is complete. A wafer undergoes many thermal excursions during processing, such as oxide growth, CVD, and so on, all contributing to undesirable dopant diffusion through the silicon. Dopant diffusion can alter critical parameters (e.g., junction depth and concentration) of the initial doped region and affect device performance. This is an important reason for minimizing the wafer thermal budget.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 685

The main focus of this chapter will be doping by ion implantation. We will first review diffusion to analyze basic concepts and explain how intentional and unintentional diffusion occurs in wafer fabrication.

DIFFUSION

Diffusion is a basic property of matter that describes the movement of one material through another. Diffusion occurs from a region of relatively higher concentration into a region of lower concentration, resulting from the motion of atoms, molecules, or ions. In semiconductor fabrication, high-temperature diffusion is the mechanism used to move a dopant through the silicon crystal lattice. Diffusion can occur as a gas-state, liquid-state or solid-state. An example of liquid-state diffusion is a drop of food colorant in a glass of water. The highly concentrated drop of colorant immediately starts to diffuse into the glass of water and will continue until the entire volume of water is the same color as the drop of colorant. The same type of diffusion occurs in the solid-state when a high concentration of dopant material is introduced to a wafer. The dopant material diffuses through the silicon atoms of the solid crystal lattice.

Diffusion Principles

Thermal diffusion of a solid-state dopant in silicon requires three steps: predeposition, drive-in, and activation. During *predeposition,* or *predep,* wafers are loaded into a high-temperature diffusion furnace and a quantity of dopant atoms is transferred from the source cabinet to the diffusion furnace. The furnace temperature setting is typically 800 to 1100°C for 10 to 30 minutes. The dopants are introduced into a very thin layer of the wafer while maintaining a constant concentration of dopants at the surface. A thin oxide layer, referred to as a *cap oxide,* is grown on the surface to prevent dopant atoms from diffusing out of the silicon. The total number of dopant atoms deposited is referred to as *Q*. Note that for ion implantation (discussed later in this chapter), Q is referred to as the dose and also represents the number of dopant atoms implanted.

The predeposition step establishes the concentration gradient for the entire diffusion process. The concentration of dopants is highest at the surface and reduces when moving away from the surface, forming a gradient (or slope). This slope establishes the dopant profile, which can be mapped using four point probes (see Chapter 7). *Fick's laws,* first postulated in 1855, mathematically describe the diffusion process. These laws state in part that the number of particles moving through a cross-sectional area is proportional to the concentration gradient.[2] Fick's laws are used to predict the dopant concentration at some distance, *x*, from the surface.

The second part of thermal diffusion is the *drive-in* step. This is a high-temperature process (1000 to 1250°C) used to move the deposited dopants through the silicon crystal to the desired junction depth in the wafer. No additional dopants are added to the wafer. However, the wafer surface oxidizes in the high-temperature environment, which affects the diffusion of the dopant atoms during drive-in. Some dopants, such as boron, tend to move into the growing oxide layer while others, such as phosphorus, are pushed away from the SiO_2. This modification of dopant concentration due to the silicon surface oxidation is termed *redistribution.*

The third part of thermal diffusion is the *activation* step. Here the temperature of the furnace is raised slightly higher to enable the dopant atoms to bond with the silicon atoms in the lattice structure. This action activates the dopant atoms, which changes the conductivity of silicon.

Dopant Movement ■ Each dopant has a particular *diffusivity* in silicon, which represents the rate (or speed) that the dopant moves in the wafer. The higher the diffusivity, the faster the dopant moves. Diffusivity increases with temperature, which is reflected in a term known as the *diffusion coefficient, D,* of the dopant being diffused. Elements have different diffusion coefficients within the temperature range of the drive-in step that reflect whether they are fast or slow diffusers in silicon. The diffusion coefficient is an equation variable in Fick's laws used to predict the final dopant concentration in silicon.

Within the wafer, dopant atoms move by two different mechanisms: interstitional and substitutional (see Figure 17.4 on page 480). Dopants with high diffusivity, such as gold (Au), copper (Cu), and nickel (Ni), use primarily *interstitional movement* to move between the interstitial space

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 686

between regular crystal sites of the silicon lattice. Slower moving dopants, such as arsenic and phosphorus that are common in semiconductor doping, typically use *substitutional movement* where atoms fill empty crystal positions in the lattice.

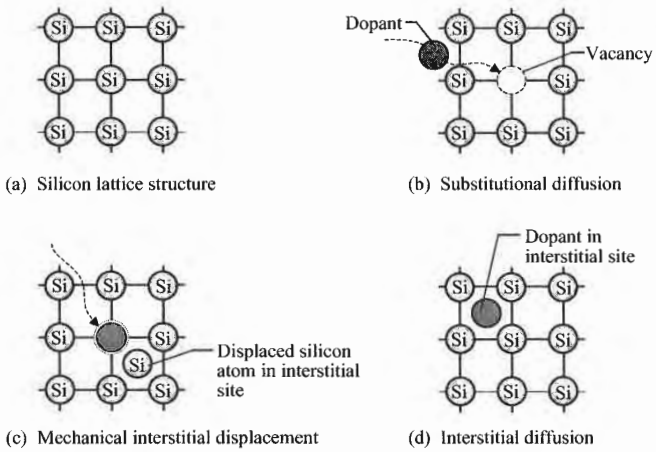


FIGURE 17.4 Dopant Diffusion in Silicon

Dopants are useful to form semiconductor silicon only if they are *activated dopants,* meaning they are part of the silicon lattice structure. An activated dopant can act as a donor or accepter of electrons and is an n-type or p-type dopant with respect to silicon. If the dopant occupies interstitial space, then it is not activated and is ineffective as a dopant. Thermal energy moves dopants into the regular crystal sites, a process known as *crystal activation.* Activation is a regular part of diffusion because it occurs at a high temperature. For ion implant, crystal activation is done with an anneal step.

Solid Solubility ■ At a given temperature, there is a limit to how much dopant can be absorbed by the silicon. This is referred to as the *solid solubility limit.* This is true for most materials, such as pouring salt into a container of water and watching it dissolve. If salt is continually added, at some point the solubility of the salt in the water reaches a limit where it no longer dissolves. Each particular dopant has a solid solubility limit (see Table 17.3). Note that only a fraction of the dopants in silicon are actually activated and contribute electrons or holes for conduction (about 3 to 5%).[3] Most dopants remain in interstitial sites and are electrically inactive.

TABLE 17.3* Solid Solubility Limits in Silicon at 1100°C

Dopant	Solubility Limit (atoms/cm^3)
Arsenic (As)	1.7×10^{21}
Phosphorus (P)	1.1×10^{21}
Boron (B)	2.2×10^{20}
Antimony (Sb)	5.0×10^{19}
Aluminum (Al)	1.8×10^{19}

*SEMATECH "Diffusion Processes," *Furnace Processes and Related Topics* (Austin, TX: SEMATECH, 1994), p. 15.

Lateral Diffusion ■ The diffusion mask is either silicon dioxide (SiO_2) or silicon nitride (Si_3N_4) since a photoresist mask could not withstand the high-temperature process. When atoms diffuse into the wafer, they move in all directions: down into the silicon, laterally, and back out of the wafer. *Lateral diffusion* occurs when dopant atoms travel in a direction along the surface of the wafer. Typically, lateral diffusion during a thermal diffusion process is 75 to 85% of the vertical junction

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 687

depth.[4] Lateral diffusion is undesirable in advanced MOS circuits because it can cause a reduction in the channel length, affecting device density and performance.

Diffusion Process

The objective of the diffusion process is to bring the diffusing impurity in contact with a wafer and maintain the specified time and temperature for diffusion to occur. Diffusion takes place in high-temperature diffusion furnaces, discussed in Chapter 10. Typically, one furnace is used for the predeposition, drive-in, and activation steps. Quartz and other furnace parts used in diffusion processes should be segregated from other furnace processes to avoid cross-contamination.

Diffusion should produce reproducible results from run-to-run and wafer-to-wafer. The eight steps required to properly perform diffusion in wafer fabrication are:

1. Run qualification test to ensure the tool meets production quality criteria.
2. Verify the wafer properties using a lot control system.
3. Download the process recipe with the desired diffusion parameters.
4. Set up the furnace, including a temperature profile.
5. Clean the wafers and dip them in hydrofluoric (HF) to remove native oxide.
6. Perform predeposition: load wafers into the deposition furnace and diffuse the dopant.
7. Perform drive-in: increase the furnace temperature to drive-in and activate the dopant bonds, then unload wafers.
8. Measure, evaluate, and record the junction depth and sheet resistivity.

The first six steps are done to prepare wafers for diffusion in the drive-in operation of step 7. A qualification test (or "qual") is performed on the furnace. This test determines the tool's production worthiness in terms of particles and other specified criteria. Once the tool is proven production worthy, the appropriate wafer lots are identified and the proper process recipe is downloaded to the furnace. A thermal profile should always be done to verify furnace temperatures are correct for the process to be performed. In a production environment, a given furnace is dedicated to a specific process step and is maintained at the precise temperature profile.

Wafer Cleaning ■ Wafer cleaning is critical because contaminants can block the diffusion of the dopant atoms into the wafer. Wafers should be cleaned immediately prior to loading the wafers into the furnace to minimize wafer contamination (e.g., native oxide). Cleaning typically includes an immersion in an acid and oxidizer, followed by an etch in an HF solution to remove any remaining oxides (see Chapter 6 for wet cleaning). The use of a dilute HF solution is referred to as a *deglaze* step to remove the poor quality native oxide on the wafer surface.

Dopant Sources ■ Although they may have been used in the early days of the semiconductor industry, it is impractical to use pure elements as a dopant source for submicron IC manufacturing. For instance, boron and phosphorus are solids at room temperature with low vapor pressures, making them difficult to melt or vaporize. Dopant concentration is also difficult to control when using a solid dopant source. Dopants are typically supplied as a gaseous or liquid source in the form of a compound. Some of the most common dopants are listed in Table 17.4 on page 482.

A liquid source (bubbler) has a carrier gas (e.g., nitrogen) bubbled through it and is delivered to the furnace tube as a vapor. Oxygen is also supplied for the dopant source to react and form oxides. Using boron as an example, the boric oxide undergoes a second reaction with silicon to form a boron-rich silicon dioxide layer on top of the wafer. This layer serves as the local source of boron for the predeposition step. The following two reaction equations illustrate how the diborane source is converted into a boron dopant:[5]

1st Reaction:	B_2H_6 (gas) *diborane*	+	$3O_2$ (gas) *oxygen*	→	B_2O_3 (solid) *boric oxide*	+	$3H_2O$ (liquid) *water*
2nd Reaction:	$2B_2O_3$ (solid) *boric oxide*	+	3Si (solid) *silicon*	→	4B (solid) *boron (dopant)*	+	$3SiO_2$ (solid) *silicon dioxide*

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 688

TABLE 17.4* Typical Dopant Sources for Diffusion

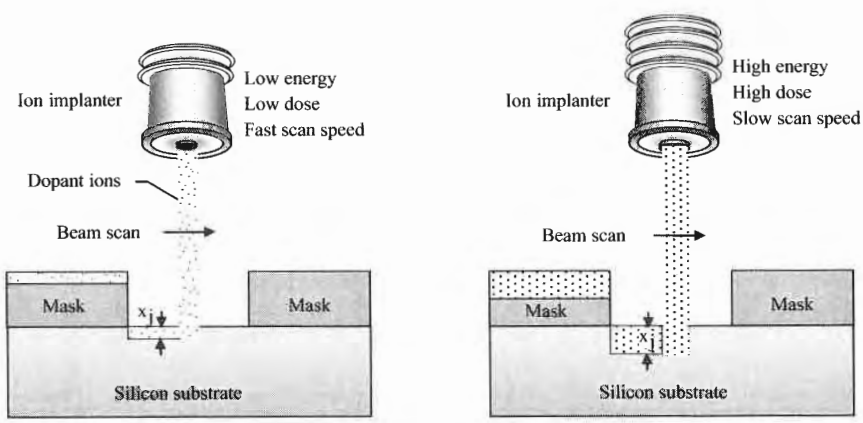
Dopant	Formula of Source	Chemical Name
Arsenic (As)	AsH_3	Arsine (gas)
Phosphorus (P)	PH_3	Phosphine (gas)
Phosphorus (P)	$POCl_3$	Phosphorus oxychloride (liquid)
Boron (B)	B_2H_6	Diborane (gas)
Boron (B)	BF_3	Boron tri-fluoride (gas)
Boron (B)	BBr_3	Boron tri-bromide (liquid)
Antimony (Sb)	$SbCl_5$	Antimony pentachloride (solid)

*SEMATECH "Diffusion Processes," *Furnace Processes and Related Topics,* (Austin, TX: SEMATECH, 1994), p. 7.

Gaseous sources are delivered to the furnace directly from the gas cylinder through a mass flow controller. The dopant is diluted in the cylinder with an inert gas to help prevent system corrosion and control the flow of the gas. A point-of-use filter should always be used to prevent fine particulates from getting into the furnace. Some diffusion sources are highly toxic, especially the gaseous sources. The most toxic of the commonly used dopant sources are arsine (AsH_3) and diborane (B_2H_6). Proper storage and usage is required at all times, especially to avoid a gas leak.

ION IMPLANTATION

Ion implantation is a method for introducing controlled amounts of dopants into the silicon substrate to change its electronic properties. It is a physical process—there are no chemical reactions. A number of applications use ion implantation in modern wafer fabrication (see Table 17.2 on page 476). The main application of ion implantation is the doping of semiconductor materials. Each doping application has specific requirements for the concentration of the dopant material and its depth. Ion implantation is preferred over diffusion in nearly all applications because of its ability to repeatedly control dopant concentration and depth. It has become a critical process to meet the challenges of wafer fabrication with sub-quarter micron feature sizes and larger diameter wafers (see Figure 17.5).



(a) Low dopant concentration (n^-, p^-) and shallow junction (x_j)

(b) High dopant concentration (n^+, p^+) and deep junction (x_j)

FIGURE 17.5 Controlling Dopant Concentration and Depth

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 689

until reaching high vacuum. At this point, an isolation valve opens and a robotic mechanism moves the cassette into the main target chamber. When the cassette is in the process chamber, a robotic wafer handler system moves the wafers from the cassette and places them on the scanning disk. The wafer usually undergoes initial alignments using the wafer flat or notch to ensure proper wafer orientation in the disk.

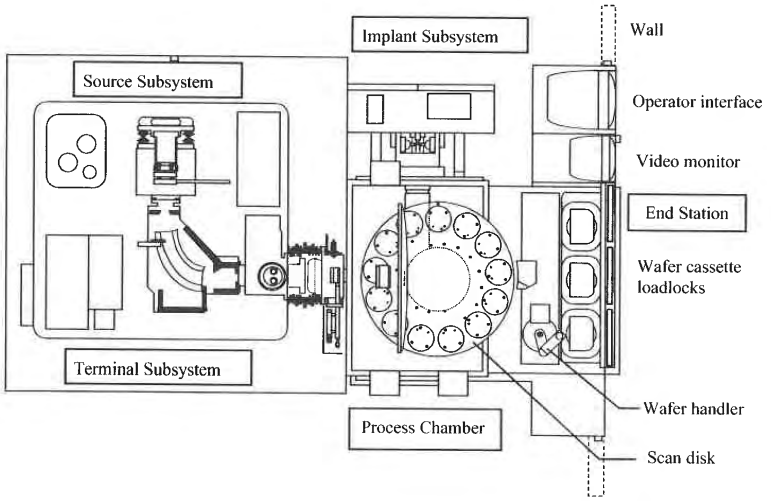


FIGURE 17.25 Wafer Handler for an Implant Process Chamber
(Used with permission from Varian Semiconductor Equipment, VIISion 200 Ion Implanter)

All subsystems are accessed by the technician through a central computer system. System status and information are obtained through the computer, including diagnostics and maintenance information.

Dose Control ■ In ion implanters, real-time dose control is done by measuring the ion beam impinging upon the wafers. The beam current is measured using a sensor called the *Faraday cup.* In a simple Faraday system, a current sensing target is placed in the ion beam path to measure the beam current. However, this placement poses a problem because the ion beam interacts with the target and creates secondary electrons, which give a false beam current reading. The Faraday system uses a design that suppresses the secondary electrons using either an electrical bias (see Figure 17.26 on page 500) or magnetic field to obtain a true ion beam current reading. The measured beam current from the Faraday cup is input into the electronic *dose controller,* which functions as a current integrator (it continuously sums the measured ion beam current). The dose controller correlates the summed beam current to the elapsed implant time and calculates the total time needed to reach a set dose value.

Annealing

Ion implantation damages the silicon lattice by knocking atoms out of the lattice structure. With high doses, the implanted layer actually becomes amorphous. Furthermore, the implanted ions rarely enter the lattice structure sites of the silicon, instead stopping in interstitial sites outside the lattice. These interstitial dopants are electrically inactive until activated by a high-temperature *annealing* step. Annealing heats the implanted silicon substrate to repair crystalline damage and electrically activate dopants by moving the atoms into crystal lattice sites (see Figure 17.27 on page 500). Crystal damage repair is done at about 500°C and dopant activation occurs at about 950°C. Electrical activation of dopants occurs as a function of time and temperature, with longer

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 690

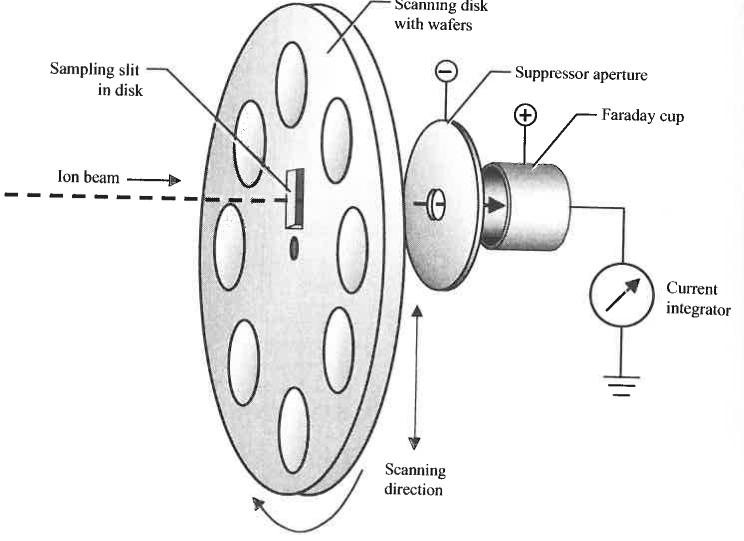


FIGURE 17.26 Faraday Cup Beam Current Measurement
Redrawn from S. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd ed., (New York: Wiley, 1994), p. 417.

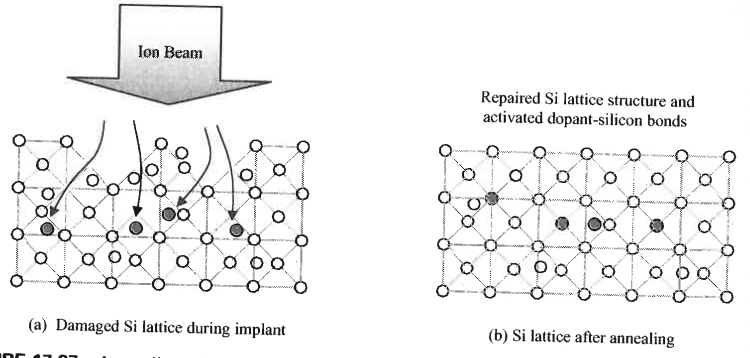


FIGURE 17.27 Annealing of Silicon Crystal

times and higher temperatures increasing the dopant activation. There are two basic methods for heating the implanted wafer for anneal:

- Furnace anneal
- Rapid thermal anneal (RTA)

Furnace Anneal ■ *Furnace anneal* is the traditional method of annealing by heating the wafer in a hot-wall furnace in the range of 800 to 1000°C for approximately 30 minutes. At this temperature, the silicon atoms move back into the lattice sites and the dopant atoms enter substitutional sites in the lattice. However, annealing cycles at this time and temperature can cause extensive dopant diffusion and is undesirable for advanced IC wafer fabrication.

Rapid Thermal Anneal (RTA) ■ *Rapid thermal anneal (RTA)* anneals the wafer by using an extremely fast ramp and short dwell time at the target temperature (typically 1000°C). Annealing of

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 691

implanted wafers is usually done in a single-wafer rapid thermal processor (RTP) with argon (Ar) or nitrogen (N_2) flowing into the chamber. A fast ramp rate and short dwell time is optimized to anneal the wafer to restore lattice damage and electrically activate the dopant while minimizing dopant diffusion in the silicon. RTA also minimizes a phenomenon known as transient enhanced diffusion (see the following section). RTA is the optimal anneal method to achieve acceptable junction depth control in shallow implants.

Investigations have been done into spike anneals for RTA. The ramp-up rate is very fast at up to 150°C/sec. This speed permits a soak time reduction to about 1 second. Uniform ramp rates during a spike anneal are critical because much of the thermal exposure occurs during the ramp.[22]

The RTA equipment design is similar to the rapid thermal processor previously discussed in Chapter 10. There are different types of heating, with a common method being tungsten halogen lamps positioned on both sides of the wafer. Multiple optical pyrometers are often used to measure temperature across the wafer to ensure uniformity.

Transient Enhanced Diffusion. There is a recent discovery of a dopant diffusion referred to as *transient enhanced diffusion* (*TED*) that occurs in the anneal step after implantation. It is caused by extra interstitial atoms from the dopant ion implanted in the silicon and is not directly related to silicon damage. It has significance when forming ultrashallow junctions because of the need to minimize all dopant diffusion.

Channeling

Single-crystal silicon atoms exhibit long range order, lying in rows and planes throughout the crystal. *Channeling* occurs when the implanted ions are not slowed by collisions with silicon atoms and instead pass through the interstitial areas of the crystal (see Figure 17.28). The undesirable aspect of channeling is that the expected projected range has a very large spread, especially for shallow implants at low ion beam energies. Dopant ions passing through lattice channels do not encounter electronic or nuclear stopping (i.e., have little energy loss) and therefore penetrate deeper than those ions that are stopped by the lattice atoms. There are four ways that channeling is controlled during implant: (1) wafer tilt, (2) screen oxide layer, (3) preamorphization of the silicon, and (4) using dopants with greater amu's (atomic mass units).

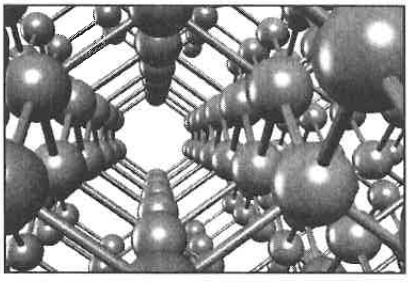


FIGURE 17.28 Silicon Lattice Viewed Along <110> Axis
(Used with permission from Edgard Torres Designs)

Wafer Tilt. The most widely used method to reduce channeling during ion implantation is *wafer tilt.* This method orients the wafer at an angle relative to the incident ion beam. A common angle for (100) wafers is 7° from the perpendicular and ensures that a collision occurs within a short distance of the dopant ion entering the silicon. The tilt angle is set during the scanning process. In this manner, the crystal lattice presents a dense orientation to the incident beam to obtain better control of the projected range on the implanted ion (see Figure 17.29 on page 502). Note that channeling behavior is different at low energies (<1 keV) for ultrashallow junctions, where tilt has little effect on reducing channeling.[23] Furthermore, tilt angles increase shadowing effects and may lead to asymmetric device behavior. It is also necessary sometimes to twist (or rotate in the vertical plane) the wafer surface. For the <110> crystal orientation, twisting the wafer 15° to 35° can help reduce channeling.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 692

ATTACHMENT L-2

S&B RR

Semiconductor Manufacturing Technology



Michael Quirk
Julian Serda

Semiconductor Manufacturing Technology
Quirk
Serda

TK 7836
.Q27
2001
COPY 1

Prentice Hall

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 694

SEMICONDUCTOR MANUFACTURING TECHNOLOGY

MICHAEL QUIRK
STAKTEK CORPORATION

JULIAN SERDA
ADVANCED MICRO DEVICES



Upper Saddle River, New Jersey
Columbus, Ohio

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 695

SaRR

Library of Congress Cataloging-in-Publication Data
Quirk, Michael
Semiconductor manufacturing technology / Michael Quirk, Julian Serda.
p. cm.
ISBN 0-13-081520-9
1. Electronic industries. 2. Semiconductors—Design and construction. I. Serda, Julian.
II. Title.

TK7836.Q27 2001

00-041653

Vice President and Publisher: Dave Garza
Editor in Chief: Stephen Helba
Assistant Vice President and Publisher: Charles E. Stewart, Jr.
Production Editor: Alexandrina Benedicto Wolf
Production Coordination: Lithokraft II
Design Coordinator: Robin G. Chukes
Cover Designer: Linda Fares
Cover Image: John Foxx
Production Manager: Matthew Ottenweller
Marketing Manager: Barbara Rose

This book was set in Times Roman by Lithokraft II. It was printed and bound by R. R. Donnelley & Sons Company. The cover was printed by Phoenix Color Corp.

LIBRARY OF CONGRESS
JAN 0 3 2001
COPYRIGHT OFFICE

Copyright © 2001 by Prentice-Hall, Inc., Upper Saddle River, New Jersey 07458. All rights reserved. Printed in the United States of America. This publication is protected by Copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permission(s), write to: Rights and Permissions Department.



10 9 8 7 6 5 4 3 2 1
ISBN 0-13-081520-9

CONTENTS

CHAPTER 1 INTRODUCTION TO THE SEMICONDUCTOR INDUSTRY 1

OBJECTIVES 1
INTRODUCTION 2
DEVELOPMENT OF AN INDUSTRY 2
- Industry Roots 2
- The Solid State 3

CIRCUIT INTEGRATION 4
- Integration Eras 5

IC FABRICATION 6
- Wafer Fab 7
- Stages of IC Fabrication 7

SEMICONDUCTOR TRENDS 9
- Increase in Chip Performance 9
- Increase in Chip Reliability 12
- Reduction in Chip Price 12

THE ELECTRONIC ERA 14
- The 1950s: Transistor Technology 14
- The 1960s: Process Technology 14
- The 1970s: Competition 14
- The 1980s: Automation 15
- The 1990s: Volume Production 15

CAREERS IN SEMICONDUCTOR MANUFACTURING 16
- Technician 16
- Job Descriptions 18

SUMMARY 19
KEY TERMS 19
REVIEW QUESTIONS 20
SELECTED INDUSTRY WEB SITES 20
REFERENCES 20

CHAPTER 2 CHARACTERISTICS OF SEMICONDUCTOR MATERIALS

OBJECTIVES 21
INTRODUCTION 21
ATOMIC STRUCTURE 22
- Electrons 22

THE PERIODIC TABLE 24
- Ionic Bonds 26
- Covalent Bonds 28

CLASSIFYING MATERIALS 28
- Conductors 29
- Insulators 30
- Semiconductors 32

SILICON 33
- Pure Silicon 33
- Why Silicon? 33
- Doped Silicon 34
- pn Junctions 39

ALTERNATIVE SEMICONDUCTOR MATERIALS 39
- Gallium Arsenide (GaAs) 40

SUMMARY 41
KEY TERMS 41
REVIEW QUESTIONS 42
REFERENCES 42

CHAPTER 3 DEVICE TECHNOLOGIES

OBJECTIVES 43
INTRODUCTION 43
CIRCUIT TYPES 44
- Analog Circuits 44
- Digital Circuits 44

PASSIVE COMPONENT STRUCTURES 45
- IC Resistor Structures 45
- IC Capacitor Structures 46

ACTIVE COMPONENT STRUCTURES 46
- The pn Junction Diode 46
- The Bipolar Junction Transistor 49
- Schottky Diode 51
- Bipolar IC Technology 52
- CMOS IC Technology 52

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 697

Enhancement and Depletion-Mode MOSFETs 58
LATCHUP IN CMOS DEVICES 59
INTEGRATED CIRCUIT PRODUCTS 59
Linear IC Product Types 60
Digital IC Product Types 60
SUMMARY 62
KEY TERMS 62
REVIEW QUESTIONS 63
IC MANUFACTURERS' WEB SITES 64
REFERENCES 65

CHAPTER 4 SILICON AND WAFER PREPARATION

OBJECTIVES 67
INTRODUCTION 67
SEMICONDUCTOR-GRADE SILICON 68
CRYSTAL STRUCTURE 68
Amorphous Materials 69
Unit Cells 69
Polycrystal and Monocrystal Structures 70
CRYSTAL ORIENTATION 71
MONOCRYSTAL SILICON GROWTH 72
CZ Method 72
Float-Zone Method 75
Reasons for Larger Ingot Diameters 76
CRYSTAL DEFECTS IN SILICON 78
Point Defects 78
Dislocations 79
Gross Defects 80
WAFER PREPARATION 80
Shaping Operations 81
Wafer Slicing 82
Wafer Lapping and Edge Contour 82
Etching 83
Polishing 83
Cleaning 84
Wafer Evaluation 84
Packaging 84
QUALITY MEASURES 84
Physical Dimensions 85
Flatness 86
Microroughness 86
Oxygen Content 86
Crystal Defects 87
Particles 87
Bulk Resistivity 87
EPITAXIAL LAYER 87
SUMMARY 88
KEY TERMS 88
REVIEW QUESTIONS 89
SELECTED INDUSTRY WEB SITES 90
REFERENCES 90

CHAPTER 5 CHEMICALS IN SEMICONDUCTOR FABRICATION

OBJECTIVES 91
INTRODUCTION 91
STATES OF MATTER 91
PROPERTIES OF MATERIALS 92
Chemical Properties for Semiconductor Manufacturing 93
PROCESS CHEMICALS 99
Liquids 99
Gases 103
SUMMARY 109
KEY TERMS 109
REVIEW QUESTIONS 110
CHEMICAL SUPPLIERS' WEB SITES 110
REFERENCES 111

CHAPTER 6 CONTAMINATION CONTROL IN WAFER FABS

OBJECTIVES 113
INTRODUCTION 113
Clean Background 114
TYPES OF CONTAMINATION 114
Particles 115
Metallic Impurities 116
Organic Contamination 117
Native Oxides 118
Electrostatic Discharge 119
SOURCES AND CONTROL OF CONTAMINATION 120
Air 120
Humans 121
Facility 123
Water 126
Process Chemicals 130

Production Equipment 131
Workstation Design 131
WAFER WET CLEANING 135
Wet-Cleaning Overview 135
Wet-Clean Equipment 138
Alternatives to RCA Clean 142
SUMMARY 143
KEY TERMS 143
REVIEW QUESTIONS 144
CHEMICAL AND EQUIPMENT SUPPLIERS' WEB SITES 145
REFERENCES 146

CHAPTER 7 METROLOGY AND DEFECT INSPECTION

OBJECTIVES 149
INTRODUCTION 149
IC METROLOGY 150
Measurement Equipment 151
Yield 151
Data Management 152
QUALITY MEASURES 152
Film Thickness 153
Film Stress 158
Refractive Index 158
Dopant Concentration 159
Unpatterned Surface Defects 161
Patterned Surface Defects 164
Critical Dimension (CD) 165
Step Coverage 167
Overlay Registration 167
Capacitance-Voltage (C-V) Test 168
Contact Angle 171
ANALYTICAL EQUIPMENT 171
Secondary-Ion Mass Spectrometry (SIMS) 171
Atomic Force Microscope (AFM) 173
Auger Electron Spectroscopy (AES) 174
X-Ray Photoelectron Spectroscopy (XPS) 174
Transmission Electron Microscope (TEM) 175
Energy- and Wavelength-Dispersive Spectrometer (EDX and WDX) 176
Focused Ion Beam (FIB) 176
SUMMARY 177
KEY TERMS 178
REVIEW QUESTIONS 178
METROLOGY EQUIPMENT SUPPLIERS' WEB SITES 179
REFERENCES 180

CHAPTER 8 GAS CONTROL IN PROCESS CHAMBERS

OBJECTIVES 181
INTRODUCTION 181
VACUUM 183
Vacuum Ranges 183
Mean Free Path 184
VACUUM PUMPS 184
Roughing Pump 185
High Vacuum Pump 186
Vacuum in Integrated Tools 188
PROCESS CHAMBER GAS FLOW 189
Mass Flow Controllers 189
RESIDUAL GAS ANALYZER (RGA) 190
RGA Basics 190
RGA as Real-Time Monitor 191
PLASMA 192
Glow Discharge 193
PROCESS CHAMBER CONTAMINATION 195
SUMMARY 196
KEY TERMS 196
REVIEW QUESTIONS 196
VACUUM EQUIPMENT SUPPLIERS' WEB SITES 197
REFERENCES 197

CHAPTER 9 IC FABRICATION PROCESS OVERVIEW

OBJECTIVES 199
INTRODUCTION 199
CMOS PROCESS FLOW 199
Overview of Areas in a Wafer Fab 200
CMOS MANUFACTURING STEPS 205
1. Twin Well Process 205
2. Shallow Trench Isolation Process 207
3. Poly Gate Structural Process 210
4. Lightly Doped Drain (LDD) Implant Process 211
5. Sidewall Spacer Formation 212
6. Source/Drain (S/D) Implant Processes 213

7. Contact Formation 214
8. Local Interconnect (LI) Process 214
9. Via-1 and Plug-1 Formation 216
10. Metal-1 Interconnect Formation 217
11. Via-2 and Plug-1 Formation 218
12. Metal-2 Interconnect Formation 219
13. Metal-3 to Pad Etch and Alloy 220
14. Parametric Testing 221

SUMMARY 222
KEY TERMS 223
REVIEW QUESTIONS 223
REFERENCES 224

CHAPTER 10 OXIDATION

OBJECTIVES 225
INTRODUCTION 225
OXIDE FILM 226
Nature of Oxide Film 226
Uses of Oxide Film 227
THERMAL OXIDATION GROWTH 231
Chemical Reaction for Oxidation 231
Oxidation Growth Model 232
FURNACE EQUIPMENT 239
HORIZONTAL VERSUS VERTICAL FURNACES 240
Vertical Furnace 241
Fast Ramp Vertical Furnace 245
Rapid Thermal Processor 246
OXIDATION PROCESS 248
Pre Oxidation Cleaning 248
Oxidation Process Recipe 249
QUALITY MEASUREMENTS 250
OXIDATION TROUBLESHOOTING 252
SUMMARY 252
KEY TERMS 253
REVIEW QUESTIONS 253
FURNACE AND RTP EQUIPMENT SUPPLIERS' WEB SITES 254
REFERENCES 255

CHAPTER 11 DEPOSITION

OBJECTIVES 257
INTRODUCTION 257
Film Layering Terminology 258
FILM DEPOSITION 260
Thin-Film Characteristics 260
Film Growth 263
Film Deposition Techniques 264
CHEMICAL VAPOR DEPOSITION 265
CVD Chemical Processes 265
CVD Reaction 266
CVD DEPOSITION SYSTEMS 269
CVD Equipment Design 270
APCVD (Atmospheric Pressure CVD) 271
LPCVD (Low Pressure CVD) 273
Plasma-Assisted CVD 277
DIELECTRICS AND PERFORMANCE 282
Dielectric Constant 282
Device Isolation 286
SPIN-ON-DIELECTRICS 287
Spin-On-Glass (SOG) 287
Spin-On-Dielectric (SOD) 288
EPITAXY 289
Epitaxy Growth Methods 290
CVD QUALITY MEASURES 292
CVD TROUBLESHOOTING 292
SUMMARY 294
KEY TERMS 295
REVIEW QUESTIONS 295
DEPOSITION EQUIPMENT SUPPLIERS' WEB SITES 296
REFERENCES 296

CHAPTER 12 METALLIZATION

OBJECTIVES 299
INTRODUCTION 300
TYPES OF METALS 302
Aluminum 302
Aluminum-Copper Alloys 305
Copper 305
Barrier Metals 307
Silicides 309
Metal Plugs 312
METAL DEPOSITION SYSTEMS 313
Evaporation 313
Sputtering 314
Metal CVD 320
Copper Electroplate 323
METALLIZATION SCHEMES 325
Traditional Aluminum Structure 325

Copper Damascene Structure 326
METALLIZATION QUALITY MEASURES 329
METALLIZATION TROUBLESHOOTING 330
SUMMARY 331
KEY TERMS 331
REVIEW QUESTIONS 332
METALLIZATION EQUIPMENT AND MATERIALS SUPPLIERS' WEB SITES 332
REFERENCES 333

CHAPTER 13 PHOTOLITHOGRAPHY: VAPOR PRIME TO SOFT BAKE

OBJECTIVES 335
INTRODUCTION 336
Photolithography Concepts 336
PHOTOLITHOGRAPHY PROCESSES 339
Negative Lithography 339
Positive Lithography 340
EIGHT BASIC STEPS OF PHOTOLITHOGRAPHY 342
Step 1: Vapor Prime 343
Step 2: Spin Coat 343
Step 3: Soft Bake 344
Step 4: Alignment and Exposure 344
Step 5: Post-Exposure Bake (PEB) 344
Step 6: Develop 344
Step 7: Hard Bake 344
Step 8: Develop Inspect 345
VAPOR PRIME 345
Wafer Cleaning 345
Dehydration Bake 346
Wafer Priming 346
SPIN COAT 348
Photoresist 348
Photoresist Physical Properties 349
Conventional I-Line Photoresists 351
Deep UV (DUV) Photoresists 354
Photoresist Dispensing Methods 357
SOFT BAKE 360
Soft Bake Equipment 360
Process Characterization 361
PHOTORESIST QUALITY MEASURES 362
PHOTORESIST TROUBLESHOOTING 363
SUMMARY 364
KEY TERMS 364
REVIEW QUESTIONS 364
PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES 365
REFERENCES 366

CHAPTER 14 PHOTOLITHOGRAPHY: ALIGNMENT AND EXPOSURE

OBJECTIVES 367
INTRODUCTION 367
Importance of Alignment and Exposure 368
OPTICAL LITHOGRAPHY 370
Light 370
Exposure Sources 372
Optics 376
Resolution 385
PHOTOLITHOGRAPHY EQUIPMENT 388
Contact Aligner 388
Proximity Aligner 389
Scanning Projection Aligner 390
Step-and-Repeat Aligner (Stepper) 390
Step-and-Scan System 393
Reticles 394
Optical Enhancement Techniques 398
Alignment 400
Environmental Conditions 404
Comparison of Photo Tools 405
MIX AND MATCH 406
ALIGNMENT AND EXPOSURE QUALITY MEASURES 407
ALIGNMENT AND EXPOSURE TROUBLESHOOTING 408
SUMMARY 409
KEY TERMS 410
REVIEW QUESTIONS 410
PHOTORESIST MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES 411
REFERENCES 412

CHAPTER 15 PHOTOLITHOGRAPHY: PHOTORESIST DEVELOPMENT AND ADVANCED LITHOGRAPHY

OBJECTIVES 413
INTRODUCTION 413
Advanced Lithography 414
POST-EXPOSURE BAKE 414
DUV Post-Exposure Bake (PEB) 415
Conventional I-Line PEB 416
DEVELOP 416
Negative Resist 417
Positive Resist 418
Development Methods 419
Resist Development Parameters 421
HARD BAKE 422
DEVELOP INSPECT 422
ADVANCED LITHOGRAPHY 424
Next-Generation Lithography 424
Advanced Resist Processing 428
DEVELOP QUALITY MEASURES 429
DEVELOP TROUBLESHOOTING 431
SUMMARY 432
KEY TERMS 433
REVIEW QUESTIONS 433
PHOTOLITHOGRAPHY MATERIALS AND EQUIPMENT SUPPLIERS' WEB SITES 433
REFERENCES 434

CHAPTER 16 ETCH

OBJECTIVES 435
INTRODUCTION 435
Etch Processes 436
ETCH PARAMETERS 437
Etch Rate 437
Etch Profile 438
Etch Bias 439
Selectivity 440
Uniformity 441
Residues 441
Polymer Formation 442
Plasma-Induced Damage 442
Particle Contamination 443
DRY ETCH 443
Etching Action 444
Potential Distribution 444
PLASMA ETCH REACTORS 446
Barrel Plasma Etcher 446
Parallel Plate (Planar) Reactor 447
Downstream Etch Systems 448
Triode Planar Reactor 448
Ion Beam Milling 448
Reactive Ion Etch (RIE) 450
High-Density Plasma Etchers 450
Etch System Review 453
Endpoint Detection 455
Vacuum for Etch Chambers 456
DRY ETCH APPLICATIONS 456
Dielectric Dry Etch 457
Silicon Dry Etch 459
Metal Dry Etch 462
WET ETCH 464
Types of Wet Etch 465
HISTORICAL PERSPECTIVE 466
PHOTORESIST REMOVAL 466
Plasma Ashing 466
ETCH INSPECTION 469
ETCH INSPECTION QUALITY MEASURES 469
DRY ETCH TROUBLESHOOTING 470
SUMMARY 471
KEY TERMS 472
REVIEW QUESTIONS 472
ETCH EQUIPMENT SUPPLIERS' WEB SITES 473
REFERENCES 474

CHAPTER 17 ION IMPLANT

OBJECTIVES 475
INTRODUCTION 475
Doped Regions 477
DIFFUSION 479
Diffusion Principles 479
Diffusion Process 481
ION IMPLANTATION 482
Overview 483
Ion Implant Parameters 484
ION IMPLANTERS 488
Ion Source 488
Extraction and Ion Analyzer 490
Acceleration Column 491

Scanning System 494
Process Chamber 498
Annealing 499
Channeling 501
Particles 502
ION IMPLANT TRENDS IN PROCESS INTEGRATION 503
Deep Buried Layers 503
Retrograde Wells 503
Punchthrough Stoppers 504
Threshold Voltage Adjustment 504
Lightly Doped Drain 505
Source/Drain Implants 505
Polysilicon Gate 506
Trench Capacitor 506
Ultrashallow Junctions 506
Silicon-On-Insulator (SOI) 507
ION IMPLANT QUALITY MEASURES 508
ION IMPLANT TROUBLESHOOTING 509
SUMMARY 510
KEY TERMS 511
REVIEW QUESTIONS 512
ION IMPLANTER EQUIPMENT SUPPLIERS' WEB SITES 513
REFERENCES 513

CHAPTER 18 CHEMICAL MECHANICAL PLANARIZATION

OBJECTIVES 515
INTRODUCTION 515
TRADITIONAL PLANARIZATION 518
Etchback 518
Glass Reflow 519
Spin-On Films 520
CHEMICAL MECHANICAL PLANARIZATION 520
CMP Planarity 521
Advantages of CMP 522
CMP Mechanisms 523
CMP Slurry and Pad 526
CMP Equipment 529
CMP Clean 532
CMP Equipment Manufacturers 534
CMP APPLICATIONS 535
STI Oxide Polish 535
LI Oxide Polish 535
LI Tungsten Polish 536
ILD Oxide Polish 536
Tungsten Plug Polish 536
Dual-Damascene Copper Polish 537
CMP QUALITY MEASURES 538
CMP TROUBLESHOOTING 540
SUMMARY 541
KEY TERMS 541
REVIEW QUESTIONS 542
CMP EQUIPMENT SUPPLIERS' WEB SITES 542
REFERENCES 543

CHAPTER 19 WAFER TEST

OBJECTIVES 545
INTRODUCTION 545
IC Electrical Tests 546
WAFER TEST 547
In-Line Parametric Test 547
Wafer Sort 555
Yield 560
Wafer Sort Yield Models 563
TEST QUALITY MEASURES 565
TEST TROUBLESHOOTING 567
SUMMARY 567
KEY TERMS 568
REVIEW QUESTIONS 568
TEST AND PROBER EQUIPMENT SUPPLIERS' WEB SITES 569
REFERENCES 569

CHAPTER 20 ASSEMBLY AND PACKAGING

OBJECTIVES 571
INTRODUCTION 571
Packaging Levels 573
TRADITIONAL ASSEMBLY 574
Backgrind 574
Die Separation 575
Die Attach 575
Wirebonding 577
TRADITIONAL PACKAGING 580
Plastic Packaging 581
Ceramic Packaging 584
Final Test 586

ADVANCED ASSEMBLY AND PACKAGING 586
Flip Chip 587
Ball Grid Array (BGA) 589
Chip on Board (COB) 590
Tape Automated Bonding (TAB) 591
Multichip Module (MCM) 591
Chip Scale Packaging (CSP) 592
Wafer-Level Packaging 592
ASSEMBLY AND PACKAGING QUALITY MEASURES 596
IC PACKAGING TROUBLESHOOTING 597
SUMMARY 597
KEY TERMS 598
REVIEW QUESTIONS 598
ASSEMBLY AND PACKAGING SUPPLIERS' WEB SITES 599
REFERENCES 600

APPENDICES

APPENDIX A CHEMICALS AND SAFETY 601
APPENDIX B CONTAMINATION CONTROLS IN CLEANROOMS 610
APPENDIX C UNITS 614
APPENDIX D COLOR AS A FUNCTION OF OXIDE THICKNESS 617
APPENDIX E OVERVIEW OF PHOTORESIST CHEMISTRY 618
APPENDIX F ETCH CHEMISTRY 622

GLOSSARY 624

INDEX 648

ATTACHMENT L-3

Screen capture created by J L Mullins from WorldCat on August 22, 2024 at URL:
https://search.worldcat.org/title/44162548

WorldCat Items

Home Libraries Topics Lists About For Librarians



Semiconductor manufacturing technology

★★★★☆ 1 reviews

Authors: Michael Quirk, Julian Serda

Show more

Summary: "This book is written for students in two- and four-year technology programs at community colleges and universities. Chapters are organized around the broad

Show more

Print Book, English, 2001
Edition: View all formats and editions
Publisher: Prentice Hall, Upper Saddle River, NJ, 2001

Physical Description:	xiii, 666 pages : illustrations ; 29 cm
ISBN:	9780130815200, 0130815209
OCLC Number / Unique Identifier:	44162548
Subjects:	Electronic industries Halbleitertechnologie Industries électroniques Semiconductors Design and construction

Screen capture created by J L Mullins from WorldCat on August 22, 2024 at URL:
https://search.worldcat.org/title/44162548

Contents:

1. Introduction to the Semiconductor Industry. 2. Characteristics of Semiconductor Materials. 3. Device Technologies. 4. Silicon and Wafer Preparation. 5. Chemicals in Semiconductor Fabrication. 6. Contamination Controls in Wafer Fabs. 7. Metrology and Defect Inspection. 8. Gas Control in Process Chambers. 9. IC Fabrication Process Overview. 10. Oxidation. 11. Deposition. 12. Metallization. 13. Photolithography: Vapor Prime to Soft Bake. 14. Photolithography: Alignment and Exposure. 15. Photolithography: Photoresist Development and Advanced Lithography. 16. Etch. 17. Ion Implant. 18. Chemical Mechanical Planarization. 19. Wafer Test. 20. Assembly and Packaging. Appendices. Glossary. Index.

5 editions in 159 libraries

Featured libraries | All libraries

Showing 84 libraries near Washington, District of Columbia

☐ Only public libraries



Regional or National

Library of Congress

2.2 miles from your current location.
101 Independence Avenue, SE, Washington, DC, 20540, United States

Borrow

Favorite Library

ATTACHMENT L-4

BOOK

Semiconductor manufacturing technology

Full Record MARC Tags

Personal name

Quirk, Michael, 1953-

Main title

Semiconductor manufacturing technology / Michael Quirk, Julian Serda.

Published/Created

Upper Saddle River, NJ : Prentice Hall, 2001.

Request this Item LC Find It Item Availability

LCCN Permalink	https://lccn.loc.gov/00041653
Description	xiii, 666 p. : ill. ; 29 cm.
ISBN	0130815209
LC classification	TK7836 .Q27 2001
Related names	Serda, Julian.
LC Subjects	Electronic industries. Semiconductors--Design and construction.
Browse by shelf order	TK7836
Notes	Includes bibliographical references (p. 647) and index.

LCCN	00041653
Dewey class no.	621.3815/2
Type of material	Book

Item Availability

CALL NUMBER	TK7836 .Q27 2001 Copy 2
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

CALL NUMBER	TK7836 .Q27 2001 Copy 1
Request in	Reference - Science & Business Reading Rm (Adams, 5th Floor)
Status	Not Charged

ATTACHMENT L-5

LIBRARY | CATALOG

Search Navigation < 3 of 3 >

BOOK

Handbook of photovoltaic science and engineering

Full Record MARC Tags

Main title

Handbook of photovoltaic science and engineering / edited by Antonio Luque and Steven Hegedus.

Published/Created

Hoboken, NJ : Wiley, c2003.

Request this Item | LC Find It

Links

Links	Table of contents http://www.loc.gov/catdir/toc/wiley041/2002191033.html Publisher description http://www.loc.gov/catdir/description/wiley036/2002191033.html Book review (E-STREAMS) http://www.e-streams.com/es0706/es0706_3291.html

More Information

LCCN Permalink	https://lccn.loc.gov/2002191033
Description	xxvii, 1138 p. : ill. ; 26 cm.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 712

ISBN	0471491969 (alk. paper)
LC classification	TK8322 .H33 2003
Related names	Luque, A. (Antonio) Hegedus, Steven.
LC Subjects	Photovoltaic cells--Handbooks, manuals, etc. Photovoltaic power generation--Handbooks, manuals, etc.
Browse by shelf order	TK8322
Notes	Includes bibliographical references and index.
LCCN	2002191033
Dewey class no.	621.31/244
Type of material	Book

Item Availability

CALL NUMBER	TK8322 .H33 2003 Copy 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 713

ATTACHMENT M-1

Handbook of Photovoltaic Science and Engineering

Edited by

Antonio Luque
Instituto de Energía Solar, Universidad Politécnica de Madrid, Spain

and

Steven Hegedus
Institute of Energy Conversion, University of Delaware, USA



Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 715

Copyright © 2003
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

All Rights Reserved. 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 under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

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

Library of Congress Cataloging-in-Publication Data

Handbook of photovoltaic science and engineering / edited by Antonio Luque and Steven Hegedus.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-49196-9 (alk. paper)
1. Photovoltaic cells. 2. Photovoltaic power generation. I. Luque, A. (Antonio) II. Hegedus, Steven.

TK8322 .H33 2003
621.31′244 – dc21

2002191033

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-471-49196-9

Typeset in 10/12 Times by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

We dedicate this book to all those who have worked so hard for half a century to bring solar electricity to where it is today, and to our colleagues present and future who must work even harder in the next half century to make sure that it fulfills its potential as a widely available clean energy source.

The editors also owe much appreciation to the authors of the chapters contained in this book. Their long hours spent writing the best possible chapter covering their field of expertise, and then suffering through a storm of editorial criticisms, has hopefully made this a high-quality publication of lasting value.

Finally, we want to express our gratitude to our loved ones (Carmen, Ignacio, Sofía, Victoria, Inés, and Debbie, Jordan, Ariel) for the many hours stolen from family life while working on this book.

AL & SH
December 2, 2002

List of Contributors

Jesús Alonso
Departamento de I+D
ISOFOTON
C/Caleta de Velez, 52
Pol. Ind. Santa Teresa
29006 Malaga
Spain
Phone: +3495 224 3790
Fax: +3495 224 3449
email: *j.alonso@isofoton.es*

Hironori Arakawa
National Institute of Advanced
Industrial Science and Technology
(AIST)
1-1-1 Higashi, Tsukuba, Ibaraki
305-8565, Japan
Phone: 29-861-4410
Fax: 29-856-3445
email: *h.arakawa@aist.go.jp*

Sheila Bailey
NASA Lewis Research Center
MS 302-1, 21000 Brookpark Road
Cleveland, OH 44135
USA
Phone: +1 216 433 2228
Fax: +1 216 433 6106
email: *Sheila.bailey@lerc.nasa.gov*

Carlos del Cañizo
Instituto de Energía Solar
Universidad Politécnica de Madrid
E.T.S.I. Telecomunicación
28040 Madrid
Spain
Phone: +34 91 544 1060
Fax: +34 91 544 6341
email: *canizo@ies-def.upm.es*

Bruno Ceccaroli
Silicon Technologies AS
P.O. Box 8309 Vaagsbygd
N-4676 Kristiansand
Norway
Phone: +47 38 08 58 81
Fax: +47 38 11 99 61
email: *br-c@online.no*

Xunming Deng
Department of Physics and
Astronomy
University of Toledo
Toledo, OH 43606
USA
Phone: +1 419 530 4782
Fax: +1 419 530 2723
email: *dengx@physics.utoledo.edu*

Michael T. Eckhart
Solar Bank Program
Solar International
Management Inc.
1825 I Street, NW, Suite 400
Washington, DC 20006 USA
USA

Phone: +1 202-429-2030
Fax: +1 202-429-5532
email: *eckhart@solarbank.com*

Keith Emery
NREL
1617 Cole Boulevard
Golden, CO 80401-3393
USA
Phone: +1 303 384 6632
Fax: +1 303 384 6604
email: *keith_emery@nrel.gov*

Arthur Endrös
Corporate R&D department
Siemens and Shell Solar GmbH
Siemens AG
Munich, Germany

Dieter Franke
ACCESS e.V.
Aachen
Germany

D. J. Friedman
NREL
1617 Cole Boulevard
Golden, CO 80401-3393
USA

Jeffery L. Gray
Purdue University
West Lafayette
Indiana
USA
email: *grayj@ecn.purdue.edu*

Lalith Gunaratne
Solar Power & Light Co, Ltd
338 TB Jayah Mawatha
Colombo 10
Sri Lanka
Phone: +94 014 818395
Fax: + 94 014 810824
email: *laithq@sri.lanka.net*

Christian Haessler
Central Research Physics
Bayer AG Krefeld
Germany
email: *christian.haessler@bayerpolymers.com*

Steven S. Hegedus
Institute of Energy Conversion
University of Delaware
Newark DE 19716
USA
email: *ssh@udel.edu*

Jorge Huacuz
Unidad de Energías no Convencionales
Instituto de Investigaciones Eléctricas
P.O. Box 1-475
Cuernavaca, Morelos
62490 Mexico
Phone/Fax: +52 73 182 436
email: *jhuacuz@iie.org.mx*

J. A. Hutchby
Semiconductor Research Corporation
P.O. Box 12053
Research Triangle Park
North Carolina 27709
USA

S. A. Johnston
P.O. Box 12194
Research Triangle Park
North Carolina 27709
USA

Juris Kalejs
RWE Schott Solar Inc.
4 Suburban Park Drive
Billerica, MA 01821 USA
Phone: 978-947-5993
Fax: 978-663-2868
email: *jkalejs@asepv.com*

Wolfgang Koch
Central Research, Physics
(ZF-FPM), Photonic Materials
Chemicals-Bayer Solar, (CH-BS),
Projects
Bayer AG
Geb.R82, PF111107
D-47812 Krefeld
Germany
Phone: +492151-883370
Fax: +492151-887503
email: *wolfgang.koch.wk2@bayer-ag.de*

Hara Kohjiro
National Institute of Advanced
Industrial Science and Technology
(AIST)
1-1-1 Higashi, Tsukuba, Ibaraki
305-8565, Japan
Phone: 29-861-4494
Fax: 29-861-6771
email: *k-hara@aist.go.jp*

Sarah Kurtz
NREL
1617 Cole Boulevard
Golden, CO 80401-3393
USA
Phone: +1 303 384 6475
Fax: +1 303 384 6531
email: *sarah_kurtz@nrel.gov*

Otto Lohne
Norwegian University of Science
and Technology
Department of Materials
Technology
N-7491 Trondheim
Norway
Phone: +47 73 59 27 94
Fax: +47 43 59 48 89
email: *Otto.Lohne@sintef.no*

Eduardo Lorenzo
Instituto de Energía Solar
Universidad Politécnica de Madrid
E.T.S.I. Telecomunicación
Ciudad Universitaria
28040 Madrid
Spain
Phone: +3491 366 7228
Fax: +3491 544 6341
email: *lorenzo@ies-def.upm.es*

Antonio Luque
Instituto de Energía Solar
Universidad Politécnica de Madrid
E.T.S.I. Telecomunicación
28040 Madrid
Spain
Phone: +34 91 336 7229
Fax: +34 91 544 6341
email: *luque@ies-def.upm.es*

Joachim Luther
Fraunhofer Institute for Solar
Energy Systems ISE
Heidenhofstrasse 2
79110 Freiburg
Germany
Phone: +49 (0) 761 4588-5120
Fax: +49 (0) 761 4588-9120
email: *luther@ise.fhg.de*

Antonio Martí
Instituto de Energía Solar
Universidad Politécnica de Madrid
E.T.S.I. Telecomunicación

28040 Madrid
Spain
Phone: +34 91 544 1060
Fax: +34 91 544 6341
email: *amarti@etsit.upm.es*

Brian McCandless
Institute of Energy Conversion
University of Delaware
Newark, DE 19716
USA
Phone: +1 302 831 6240
Fax: +1 302 831 6226
email: *bem@udel.edu*

H. J. Moeller
Institut für Experimentelle Physik
TU Bergakademie Freiberg
Silbermannstr. 1
09599 Freiberg
Germany
Phone: +493731-392896
Fax: +493731-394314
email: *moeller@physik.tu-freiberg.de*

J. M. Olson
NREL
1617 Cole Boulevard
Golden, CO 80401-3393
USA

Klaus Preiser
Produktion Energie badenova
AG & Co. KG
Tullastraße 61
79108 Freiburg i.Br.
Telefon 0761/279-2207
Telefax 0761/279-2731
Mobil 0160/7154879
email: *klaus.preiser@badenova.de*
www.badenova.de

Ryne Raffaelle
Rochester Institute of Technology
84 Lomb Memorial Drive
Rochester, NY 14623-5603
USA

Tjerk Reijenga
BEAR Architecten
Gravin Beatrixstraat 34
NL 2805 PJ Gouda
The Netherlands
Phone: +31 182 529 899
Fax: +31 182 582 599
email: *Tjerk@bear.nl*

Keith Rutledge
Renewable Energy Development
Institute
Willits, CA 95490
USA

Dirk Uwe Sauer
Electrical Energy Systems -
Storage Systems
Fraunhofer Institut für Solare
Energiesysteme ISE
Heidenhofstrasse 2
D-79110 Freiburg
Germany
Phone: +49 761 4588 5219
Fax: +49 761 4588 9217
email: *sauer@ise.fhg.de*

Eric A. Schiff
Department of Physics
Syracuse University
Syracuse, New York 13244-1130
USA
http://physics.syr.edu/∼schiff

Jürgen Schmid
ISET−Institut für Solare
Energieversorgungstechnik e.V.,
Universität Kassel
Königstor 59

34119 Kassel
Germany
Phone: +49 (0)5 61/72 94-3 45
Fax: +49 (0)5 61/72 94-3 00
email: *jschmid@iset.uni-kassel.de*

Heribert Schmidt
Fraunhofer Institut für Solare
Energiesysteme ISE, Freiburg
Heidenhofstr. 2
79110 Freiburg
Germany
Phone: +49 (0)7 61/45 88-52 26
Fax: +49 (0)7 61/45 88-92 26
email: *heri@ise.fhg.de*

William Shafarman
Institute of Energy Conversion
University of Delaware
Newark, DE 19716
USA
Phone: 1 302 831 6215
Fax: 1 302 831 6226
email: *wns@udel.edu*

James Sites
Department of Physics
Colorado State University
Fort Collins, CO 80523-1875
USA
Phone: +1 970 491 5850
Fax: +1 970 491 7947
email: *sites@lamar.colostate.edu*

Bushan Sopori
NREL
1617 Cole Boulevard
Golden, CO 80401-3393
USA
Phone: +1 303 384 6683
Fax: +1 303 384 6684
email: *bsopori@nrel.gov*

Lars Stolt
Ångström Solar Center
Uppsala University
P.O. Box 534
SE-751 21 Uppsala
Sweden
Phone: +46 18 471 3039
Fax: +46 18 555 095
email: *Lars.Stolt@angstrom.uu.se*

Jack L. Stone
NREL
1617 Cole Boulevard
Golden, CO 80401-3393
USA

Richard Swanson
SUNPOWER Corporation
435 Indio Way
Sunnyvale, CA 94086
USA
Phone: +1 408 991 0900
Fax: +1 408 739 7713
email: *Rswanson@sunpowercorp.com*

Ignacio Tobías
Instituto de Energía Solar
Universidad Politécnica de Madrid
ETSI Telecomunicación
Ciudad Universitaria
28040 Madrid
Spain
Phone: +3491 5475700-282
Fax: +3491 5446341
email: *Tobias@ies-def.upm.es*

Richard A. Whisnant
Parameters, Inc.
1505 Primrose Lane
Cary, NC 27511
(919) 467-8710 (phone, fax)
(919) 523-0456 (cell phone)

Contents

List of Contributors **xxiii**

1 Status, Trends, Challenges and the Bright Future of Solar Electricity from Photovoltaics **1**
Steven S. Hegedus and Antonio Luque
1.1 The Big Picture 1
1.2 What Is Photovoltaics? 3
1.3 Six Myths of Photovoltaics 5
1.4 History of Photovoltaics 11
1.5 PV Costs, Markets and Forecasts 15
1.6 What Are the Goals of Today's PV Research and Manufacturing? 19
1.7 Global Trends in Performance and Applications 20
1.8 Crystalline Silicon Progress and Challenges 23
1.9 Thin Film Progress and Challenges 27
1.10 Concentration PV Systems 31
1.11 Balance of Systems 32
1.12 Future of Emerging PV Technologies 37
1.13 Conclusions 39
References 41

2 Motivation for Photovoltaic Application and Development **45**
Joachim Luther
2.1 Characteristics of Photovoltaic Energy Conversion 45
2.2 A Long-term Substitute for Today's Conventional Electricity Production – The Ecological Dimension of Photovoltaics 48
2.2.1 In Summary 54
2.3 A Technological Basis for Off-grid Electricity Supply – The Development Dimension of Photovoltaics 54
2.3.1 In Summary 57
2.4 Power Supply for Industrial Systems and Products – The Professional Low Power Dimension 57
2.5 Power for Spacecraft and Satellites – the Extraterrestrial Dimension of Photovoltaics 59
References 60

3 The Physics of the Solar Cell **61**
Jeffery L. Gray
3.1 Introduction 61
3.2 Fundamental Properties of Semiconductors 64
3.2.1 Crystal Structure 64
3.2.2 Energy Band Structure 65
3.2.3 Conduction-band and Valence-band Densities of State 66
3.2.4 Equilibrium Carrier Concentrations 67
3.2.5 Light Absorption 70
3.2.6 Recombination 74
3.2.7 Carrier Transport 78
3.2.8 Semiconductor Equations 81
3.2.9 Minority-carrier Diffusion Equation 82
3.3 *PN*-Junction Diode Electrostatics 83
3.4 Solar Cell Fundamentals 87
3.4.1 Solar Cell Boundary Conditions 87
3.4.2 Generation Rate 89
3.4.3 Solution of the Minority-carrier Diffusion Equation 89
3.4.4 Terminal Characteristics 89
3.4.5 Solar Cell $I-V$ Characteristics 92
3.4.6 Properties of Efficient Solar Cells 95
3.4.7 Lifetime and Surface Recombination Effects 96
3.4.8 An Analogy for Understanding Solar Cell Operation: A Partial Summary 98
3.5 Additional Topics 99
3.5.1 Efficiency and Band gap 99
3.5.2 Spectral Response 100
3.5.3 Parasitic Resistance Effects 102
3.5.4 Temperature Effects 104
3.5.5 Concentrator Solar Cells 106
3.5.6 High-level Injection 107
3.5.7 *p-i-n* Solar Cells 109
3.5.8 Detailed Numerical Modeling 109
3.6 Summary 110
References 111

4 Theoretical Limits of Photovoltaic Conversion **113**
Antonio Luque and Antonio Martí
4.1 Introduction 113
4.2 Thermodynamic Background 114
4.2.1 Basic Relationships 114
4.2.2 The Two Laws of Thermodynamics 116
4.2.3 Local Entropy Production 116
4.2.4 An Integral View 117
4.2.5 Thermodynamic Functions of Radiation 117
4.2.6 Thermodynamic Functions of Electrons 119
4.3 Photovoltaic Converters 120

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 724

4.3.1 The Balance Equation of a PV Converter 120
4.3.2 The Monochromatic Cell 124
4.3.3 Thermodynamic Consistence of the Shockley–Queisser Photovoltaic Cell 126
4.3.4 Entropy Production in the Whole Shockley–Queisser Solar Cell 129
4.4 The Technical Efficiency Limit for Solar Converters 131
4.5 Very High Efficiency Concepts 132
4.5.1 Multijunction Solar Cells 132
4.5.2 Thermophotovoltaic Converters 135
4.5.3 Thermophotonic Converters 136
4.5.4 Higher-than-one Quantum Efficiency Solar Cells 140
4.5.5 Hot Electron Solar Cells 141
4.5.6 Intermediate Band Solar Cell 144
4.6 Conclusions 148
References 149

5 Solar Grade Silicon Feedstock 153
Bruno Ceccaroli and Otto Lohne
5.1 Introduction 153
5.2 Silicon 154
5.2.1 Physical Properties of Silicon Relevant to Photovoltaics 154
5.2.2 Chemical Properties Relevant to Photovoltaics 156
5.2.3 Health Factors 156
5.2.4 History and Applications of Silicon 157
5.3 Production of Metallurgical Grade Silicon 161
5.3.1 The Carbothermic Reduction of Silica 161
5.3.2 Refining 163
5.3.3 Casting and Crushing 166
5.3.4 Economics 167
5.4 Production of Semiconductor Grade Silicon (Polysilicon) 167
5.4.1 The Siemens Process 168
5.4.2 The Union Carbide Process 172
5.4.3 The Ethyl Corporation Process 173
5.4.4 Economics and Business 175
5.5 Current Silicon Feedstock to Solar Cells 175
5.6 Requirements of Silicon for Crystalline Solar Cells 179
5.6.1 Solidification 179
5.6.2 Effect of Crystal Imperfections 182
5.6.3 Effect of Various Impurities 186
5.7 Routes to Solar Grade Silicon 193
5.7.1 Crystallisation 193
5.7.2 Upgrading Purity of the Metallurgical Silicon Route 194
5.7.3 Simplification of the Polysilicon Process 198
5.7.4 Other Methods 201
5.8 Conclusions 201
References 202

6 Bulk Crystal Growth and Wafering for PV **205**
W. Koch, A. L. Endrös, D. Franke, C. Häßler, J. P. Kalejs and H. J. Möller
6.1 Introduction 205
6.2 Bulk Monocrystalline Material 206
6.2.1 Cz Growth of Single-crystal Silicon 207
6.2.2 Tri-crystalline Silicon 211
6.3 Bulk Multicrystalline Silicon 214
6.3.1 Ingot Fabrication 214
6.3.2 Doping 216
6.3.3 Crystal Defects 217
6.3.4 Impurities 219
6.4 Wafering 223
6.4.1 Multi-wire Wafering Technique 224
6.4.2 Microscopic Process of Wafering 226
6.4.3 Wafer Quality and Saw Damage 229
6.4.4 Cost and Size Considerations 230
6.5 Silicon Ribbon and Foil Production 230
6.5.1 Process Description 232
6.5.2 Productivity Comparisons 238
6.5.3 Manufacturing Technology 239
6.5.4 Ribbon Material Properties and Solar Cells 240
6.5.5 Ribbon/Foil Technology – Future Directions 243
6.6 Numerical Simulations of Crystal Growth Techniques 244
6.6.1 Simulation Tools 245
6.6.2 Thermal Modelling of Silicon Crystallisation Techniques 245
6.6.3 Simulation of Bulk Silicon Crystallisation 247
6.6.4 Simulation of Silicon Ribbon Growth 249
6.7 Conclusions 251
6.8 Acknowledgement 252
References 252

7 Crystalline Silicon Solar Cells and Modules **255**
Ignacio Tobías, Carlos del Cañizo and Jesús Alonso
7.1 Introduction 255
7.2 Crystalline Silicon as a Photovoltaic Material 257
7.2.1 Bulk Properties 257
7.2.2 Surfaces 257
7.3 Crystalline Silicon Solar Cells 259
7.3.1 Cell Structure 259
7.3.2 Substrate 260
7.3.3 The Front Surface 263
7.3.4 The Back Surface 266
7.3.5 Size Effects 266
7.3.6 Cell Optics 268
7.3.7 Performance Comparison 270

7.4 Manufacturing Process 271
7.4.1 Process Flow 271
7.4.2 Screen-printing Technology 276
7.4.3 Throughput and Yield 279
7.5 Variations to the Basic Process 280
7.5.1 Thin Wafers 280
7.5.2 Back Surface Passivation 281
7.5.3 Improvements to the Front Emitter 281
7.5.4 Rapid Thermal Processes 282
7.6 Multicrystalline Cells 283
7.6.1 Gettering in mc Solar Cells 283
7.6.2 Passivation with Hydrogen 283
7.6.3 Optical Confinement 285
7.7 Other Industrial Approaches 288
7.7.1 Silicon Ribbons 288
7.7.2 Heterojunction with Intrinsic Thin Layer 288
7.7.3 Buried Contact Technology 289
7.8 Crystalline Silicon Photovoltaic Modules 291
7.8.1 Cell Matrix 291
7.8.2 The Layers of the Module 292
7.8.3 Lamination and Curing 293
7.8.4 Postlamination Steps 294
7.8.5 Special Modules 294
7.9 Electrical and Optical Performance of Modules 295
7.9.1 Electrical and Thermal Characteristics 295
7.9.2 Fabrication Spread and Mismatch Losses 297
7.9.3 Local Shading and Hot Spot Formation 297
7.9.4 Optical Properties 300
7.10 Field Performance of Modules 301
7.10.1 Lifetime 301
7.10.2 Qualification 301
7.11 Conclusions 302
References 303

8 Thin-film Silicon Solar Cells 307
Bhushan Sopori
8.1 Introduction 307
8.2 A Review of Current Thin-film Si Cells 310
8.2.1 Single-crystal Films Using Single-crystal Si Substrates 317
8.2.2 Multicrystalline-Si Substrates 320
8.2.3 Non-Si Substrates 321
8.3 Design Concepts of TF-Si Solar Cells 324
8.3.1 Light-trapping in Thin Si Solar Cells 326
8.3.2 Description of PV Optics 327
8.3.3 Electronic Modeling 333
8.3.4 Methods of Making Thin-Si Films for Solar Cells 341

8.3.5 Methods of Grain Enhancement of a-Si/μc-Si Thin Films 343
8.3.6 Processing Considerations for TF-Si Solar Cell Fabrication 350
8.4 Conclusion 353
References 354

9 High-Efficiency III-V Multijunction Solar Cells 359
J. M. Olson, D. J. Friedman and Sarah Kurtz
9.1 Introduction 359
9.2 Applications 363
9.2.1 Space Solar Cells 363
9.2.2 Terrestrial Energy Production 363
9.3 Physics of III-V Multijunction and Single-junction Solar Cells 363
9.3.1 Wavelength Dependence of Photon Conversion Efficiency 363
9.3.2 Theoretical Limits to Multijunction Efficiencies 364
9.3.3 Spectrum Splitting 364
9.4 Cell Configuration 365
9.4.1 Four-terminal 365
9.4.2 Three-terminal Voltage-matched Interconnections 366
9.4.3 Two-terminal Series-connected (Current Matched) 366
9.5 Computation of Series-Connected Device Performance 366
9.5.1 Overview 366
9.5.2 Top and Bottom Subcell QE and J_{SC} 367
9.5.3 Multijunction $J-V$ Curves 368
9.5.4 Efficiency versus Band Gap 370
9.5.5 Top-cell Thinning 372
9.5.6 Current-matching Effect on Fill Factor and V_{OC} 373
9.5.7 Spectral Effects 374
9.5.8 AR Coating Effects 375
9.5.9 Concentration 376
9.5.10 Temperature Dependence 380
9.6 Materials Issues Related to GaInP/GaAs/Ge Solar Cells 382
9.6.1 Overview 382
9.6.2 MOCVD 382
9.6.3 GaInP Solar Cells 383
9.6.4 GaAs Cells 393
9.6.5 Ge Cells 395
9.6.6 Tunnel-junction Interconnects 396
9.6.7 Chemical Etchants 397
9.6.8 Materials Availability 398
9.7 Troubleshooting 398
9.7.1 Characterization of Epilayers 398
9.7.2 Transmission Line Measurements 400
9.7.3 I-V Measurements of Multijunction Cells 400
9.7.4 Evaluation of Morphological Defects 401
9.7.5 Device Diagnosis 401

9.8 Future-generation Solar Cells 403
9.8.1 Refinements to the GaInP/GaAs/Ge Cell 403
9.8.2 Mechanical Stacks 404
9.8.3 Growth on Other Substrates 405
9.8.4 Spectrum Splitting 406
9.9 Implementation into Terrestrial Systems 406
9.9.1 Economic Issues 406
9.9.2 Concentrator Systems 406
9.9.3 Terrestrial Spectrum 407
References 407

10 Space Solar Cells and Arrays 413
Sheila Bailey and Ryne Raffaelle
10.1 The History of Space Solar Cells 413
10.1.1 Vanguard I to Deep Space I 413
10.2 The Challenge for Space Solar Cells 416
10.2.1 The Space Environment 417
10.2.2 Thermal Environment 420
10.2.3 Solar Cell Calibration and Measurement 424
10.3 Silicon Solar Cells 425
10.4 III-V Solar Cells 426
10.4.1 Thin-film Solar Cells 428
10.5 Space Solar Arrays 431
10.5.1 Body-mounted Arrays 432
10.5.2 Rigid Panel Planar Arrays 432
10.5.3 Flexible Fold-out Arrays 433
10.5.4 Thin-film or Flexible Roll-out Arrays 435
10.5.5 Concentrating Arrays 436
10.5.6 High-temperature/Intensity Arrays 438
10.5.7 Electrostatically Clean Arrays 439
10.5.8 Mars Solar Arrays 440
10.5.9 Power Management and Distribution (PMAD) 441
10.6 Future Cell and Array Possibilities 441
10.6.1 Low Intensity Low Temperature (LILT) Cells 441
10.6.2 Quantum Dot Solar Cells 442
10.6.3 Integrated Power Systems 442
10.6.4 High Specific Power Arrays 443
10.6.5 High-radiation Environment Solar Arrays 443
10.7 Power System Figures of Merit 444
References 446

11 Photovoltaic Concentrators 449
Richard M. Swanson
11.1 Introduction 449
11.1.1 The Concentrator Dilemma 450
11.2 Basic Types of Concentrators 452
11.2.1 Types of Optics 452
11.2.2 Concentration Ratio 455

11.2.3 Types of Tracking 456
11.2.4 Static Concentrators 456
11.3 Historical Overview 460
11.3.1 The Sandia National Laboratories Concentrator Program (1976 to 1993) 461
11.3.2 The Martin Marietta Point-focus Fresnel System 462
11.3.3 The Entech Linear-focus Fresnel System 463
11.3.4 Other Sandia Projects 465
11.3.5 The Concentrator Initiative 465
11.3.6 Early Demonstration Projects 466
11.3.7 The EPRI High-concentration Program 467
11.3.8 Other Concentrator Programs 471
11.3.9 History of Performance Improvements 472
11.4 Optics of Concentrators 474
11.4.1 Basics 474
11.4.2 Reflection and Refraction 478
11.4.3 The Parabolic Concentrator 479
11.4.4 The Compound Parabolic Concentrator 482
11.4.5 The V-trough Concentrator 483
11.4.6 Refractive Lenses 485
11.4.7 Secondary Optics 489
11.4.8 Static Concentrators 491
11.4.9 Innovative Concentrators 492
11.4.10 Issues in Concentrator Optics 494
11.5 Current Concentrator Activities 495
11.5.1 Amonix 496
11.5.2 Australian National University 496
11.5.3 BP Solar and the Polytechnical University of Madrid 496
11.5.4 Entech 497
11.5.5 Fraunhofer-Institut fur Solare Energiesysteme 497
11.5.6 Ioffe Physical-Technical Institute 498
11.5.7 National Renewable Energy Laboratory 498
11.5.8 Polytechnical University of Madrid 498
11.5.9 Solar Research Corporation 499
11.5.10 Spectrolab 499
11.5.11 SunPower Corporation 499
11.5.12 University of Reading 500
11.5.13 Tokyo A&T University 500
11.5.14 Zentrum fur Sonnenenergie und Wasserstoff Forschung Baden Wurttenberg (ZSW) 500
References 500

12 Amorphous Silicon–based Solar Cells 505
Xunming Deng and Eric A. Schiff
12.1 Overview 505
12.1.1 Amorphous Silicon: The First Bipolar Amorphous Semiconductor 505

12.1.2 Designs for Amorphous Silicon Solar Cells: A Guided Tour 508
12.1.3 Staebler–Wronski Effect 511
12.1.4 Synopsis of this Chapter 512
12.2 Atomic and Electronic Structure of Hydrogenated Amorphous Silicon 513
12.2.1 Atomic Structure 513
12.2.2 Defects and Metastability 514
12.2.3 Electronic Density-of-states 515
12.2.4 Bandtails, Bandedges, and Band Gaps 516
12.2.5 Defects and Gap States 517
12.2.6 Doping 518
12.2.7 Alloying and Optical Properties 518
12.3 Depositing Amorphous Silicon 520
12.3.1 Survey of Deposition Techniques 520
12.3.2 RF Glow Discharge Deposition 521
12.3.3 Glow Discharge Deposition at Different Frequencies 523
12.3.4 Hot-wire Chemical Vapor Deposition 525
12.3.5 Other Deposition Methods 526
12.3.6 Hydrogen Dilution 526
12.3.7 Alloys and Doping 528
12.4 Understanding a-Si *pin* Cells 528
12.4.1 Electronic Structure of a *pin* Device 528
12.4.2 Photocarrier Drift in Absorber Layers 530
12.4.3 Absorber Layer Design of a *pin* Solar Cell 533
12.4.4 The Open-circuit Voltage 534
12.4.5 Optical Design of a-Si:H Solar Cells 537
12.4.6 Cells under Solar Illumination 540
12.4.7 Light-soaking Effects 541
12.5 Multiple-Junction Solar Cells 542
12.5.1 Advantages of Multiple-junction Solar Cells 542
12.5.2 Using Alloys for Cells with Different Band Gaps 544
12.5.3 a-Si/a-SiGe Tandem and a-Si/a-SiGe/a-SiGe Triple-junction Solar Cells 546
12.5.4 Microcrystalline Silicon Solar Cells 551
12.5.5 Micromorph and Other μc-Si-based Multijunction Cells 552
12.6 Module Manufacturing 553
12.6.1 Continuous Roll-to-roll Manufacturing on Stainless Steel Substrates 553
12.6.2 a-Si Module Production on Glass Superstrate 555
12.6.3 Manufacturing Cost, Safety, and Other Issues 556
12.6.4 Module Performance 557
12.7 Conclusions and Future Projections 558
12.7.1 Status and Competitiveness of a-Si Photovoltaics 558
12.7.2 Critical Issues for Further Enhancement and Future Potential 559
12.8 Acknowledgments 559
References 560

13 Cu(InGa)Se_2 Solar Cells **567**
William N. Shafarman and Lars Stolt
13.1 Introduction 567
13.2 Material Properties 570
13.2.1 Structure and Composition 571
13.2.2 Optical Properties 574
13.2.3 Electrical Properties 574
13.2.4 The Surface and Grain Boundaries 576
13.2.5 Substrate Effects 578
13.3 Deposition Methods 578
13.3.1 Substrates 579
13.3.2 Back Contact 580
13.3.3 Coevaporation of Cu(InGa)Se_2 580
13.3.4 Two-step Processes 583
13.3.5 Other Deposition Approaches 584
13.4 Junction and Device Formation 584
13.4.1 Chemical Bath Deposition 585
13.4.2 Interface Effects 586
13.4.3 Other Deposition Methods 587
13.4.4 Alternative Buffer Layers 588
13.4.5 Transparent Contacts 590
13.4.6 Buffer Layers 591
13.4.7 Device Completion 592
13.5 Device Operation 592
13.5.1 Light-generated Current 593
13.5.2 Recombination 595
13.5.3 The Cu(InGa)Se_2/CdS Interface 599
13.5.4 Wide and Graded Band Gap Devices 600
13.6 Manufacturing Issues 602
13.6.1 Processes and Equipment 602
13.6.2 Module Fabrication 604
13.6.3 Module Performance 604
13.6.4 Production Costs 607
13.6.5 Environmental Concerns 608
13.7 The Cu(InGa)Se_2 Outlook 609
References 611

14 Cadmium Telluride Solar Cells **617**
Brian E. McCandless and James R. Sites
14.1 Introduction 617
14.2 CdTe Properties and Thin-film Fabrication Methods 621
14.2.1 Condensation/Reaction of Cd and Te_2 Vapors on a Surface 628
14.2.2 Galvanic Reduction of Cd and Te Ions at a Surface 629
14.2.3 Precursor Reaction at a Surface 630
14.3 CdTe Thin-Film Solar Cells 631
14.3.1 Window Layers 631
14.3.2 CdTe Absorber Layer and $CdCl_2$ Treatment 633

14.3.3 CdS/CdTe Intermixing 637
14.3.4 Back Contact 642
14.3.5 Solar Cell Characterization 644
14.3.6 Summary of CdTe-cell Status 650
14.4 CdTe Modules 651
14.5 The Future of CdTe-based Solar Cells 653
14.6 Acknowledgments 657
References 657

15 Dye-sensitized Solar Cells 663
Kohjiro Hara and Hironori Arakawa
15.1 Introduction to Dye-Sensitized Solar Cells (DSSC) 663
15.1.1 Background 663
15.1.2 Structure and Materials 664
15.1.3 Mechanism 670
15.1.4 Charge-transfer Kinetics 673
15.1.5 Characteristics 678
15.2 DSSC Fabrication ($\eta = 8\%$) 678
15.2.1 Preparation of TiO_2 Colloid 678
15.2.2 Preparation of the TiO_2 Electrode 679
15.2.3 Dye Fixation onto the TiO_2 Film 680
15.2.4 Redox Electrolyte 681
15.2.5 Counter Electrode 681
15.2.6 Assembling the Cell and Cell Performance 681
15.3 New Developments 682
15.3.1 New Oxide Semiconductor Film Photoelectrodes 683
15.3.2 New Dye Photosensitizers 683
15.3.3 New Electrolytes 688
15.3.4 Quasi-solid-state and Solid-state DSSCs 689
15.4 Approach to Commercialization 691
15.4.1 Stability of the DSSC 691
15.4.2 Module Fabrication and Other Subjects for Commercialization 694
15.5 Summary and Prospects 695
References 696

16 Measurement and Characterization of Solar Cells and Modules 701
Keith Emery
16.1 Introduction 701
16.2 Rating PV Performance 701
16.2.1 Standard Reporting Conditions 702
16.2.2 Alternative Peak Power Ratings 715
16.2.3 Energy-based Performance Rating Methods 716
16.2.4 Translation Equations to Reference Conditions 719
16.3 Current Versus Voltage Measurements 721
16.3.1 Measurement of Irradiance 721
16.3.2 Simulator-based $I-V$ Measurements: Theory 722

16.3.3 Primary Reference Cell Calibration Methods 723
16.3.4 Uncertainty Estimates in Reference Cell Calibration Procedures 726
16.3.5 Intercomparison of Reference Cell Calibration Procedures 727
16.3.6 Multijunction Cell Measurement Procedures 728
16.3.7 Cell and Module $I-V$ Systems 731
16.3.8 Solar Simulators 736
16.4 Spectral Responsivity Measurements 738
16.4.1 Filter-based Systems 739
16.4.2 Grating-based Systems 741
16.4.3 Spectral Responsivity Measurement Uncertainty 742
16.5 Module Qualification and Certification 745
Acknowledgements 746
References 747

17 Photovoltaic Systems 753
Klaus Preiser
17.1 Introduction to PV Systems and Various Forms of Application 753
17.2 Principles of photovoltaic Power System Configuration and their Application 755
17.2.1 Grid-independent Photovoltaic Systems for Small Devices and Appliances 755
17.2.2 Photovoltaic Systems for Remote Consumers of Medium and Large Size 761
17.2.3 Decentralised Grid-connected Photovoltaic Systems 774
17.2.4 Central Grid-connected Photovoltaic Systems 779
17.2.5 Space Application 780
17.3 Components for PV Systems 784
17.3.1 Battery Storage 784
17.3.2 Charge Controller 787
17.3.3 Inverters 788
17.3.4 Auxiliary Generators 790
17.3.5 System Sizing 791
17.3.6 Energy-saving Domestic Appliances 793
17.4 Future Developments in Photovoltaic System Technology 794
17.4.1 Future Developments in Off-grid Power Supply with Photovoltaics 794
17.4.2 Future Developments in Grid-connected Photovoltaic Systems 796
References 797

18 Electrochemical Storage for Photovoltaics 799
Dirk Uwe Sauer
18.1 Introduction 799
18.2 General Concept of Electrochemical Batteries 801
18.2.1 Fundamentals of Electrochemical Cells 801

18.2.2 Batteries with Internal and External Storage 807
18.2.3 Commonly Used Technical Terms and Definitions 809
18.2.4 Definitions of Capacity and State of Charge 811
18.3 Typical Operation Conditions of Batteries in PV Applications 812
18.3.1 An Example of an Energy Flow Analysis 812
18.3.2 Classification of Battery-operating Conditions in PV Systems 813
18.4 Secondary Electrochemical Accumulators with Internal Storage 817
18.4.1 Overview 817
18.4.2 NiCd Batteries 818
18.4.3 Nickel-metal Hydride (NiMH) Batteries 821
18.4.4 Rechargeable Alkali Mangan (RAM) Batteries 822
18.4.5 Lithium-ion and Lithium-polymer Batteries 822
18.4.6 Double-layer Capacitors 824
18.4.7 The Lead Acid Battery 826
18.5 Secondary Electrochemical Battery Systems with External Storage 849
18.5.1 Redox-flow Batteries 850
18.5.2 Hydrogen/Oxygen Storage Systems 852
18.6 Investment and Lifetime Cost Considerations 857
18.7 Conclusion 859
References 860

19 Power Conditioning for Photovoltaic Power Systems 863
Jürgen Schmid, Heribert Schmidt
19.1 Charge Controllers and Monitoring Systems for Batteries in PV Power Systems 864
19.1.1 Charge Controllers 864
19.1.2 Charge Equaliser for Long Battery Strings 877
19.2 Inverters 881
19.2.1 General Characteristics of PV Inverters 881
19.2.2 Inverters for Grid-connected Systems 881
19.2.3 Inverters for Stand-alone Operation 883
19.2.4 Inverter Principles 884
19.2.5 Power Quality of Inverters 896
19.2.6 Active Quality Control in the Grid 900
19.2.7 Safety Aspects with Grid-connected Inverters 900
19.3 Acknowledgement 902
References 902

20 Energy Collected and Delivered by PV Modules 905
Eduardo Lorenzo
20.1 Introduction 905
20.2 Movement between Sun and Earth 906
20.3 Solar Radiation Components 912
20.4 Solar Radiation Data and Uncertainty 915
20.4.1 Clearness Index 920
20.5 Radiation on Inclined Surfaces 920

20.5.1 Estimation of the Direct and Diffuse Components of Horizontal Radiation, Given the Global Radiation 920
20.5.2 Estimation of the Hourly Irradiation from the Daily Irradiation 925
20.5.3 Estimation of the Radiation on Surfaces on Arbitrary Orientation, Given the Components Falling on a Horizontal Surface 927
20.6 Diurnal Variations of the Ambient Temperature 933
20.7 Effects of the Angle of Incidence and of the Dirt 934
20.8 Some Calculation Tools 937
20.8.1 Generation of Daily Radiation Sequences 937
20.8.2 The Reference Year 937
20.8.3 Shadows and Trajectory Maps 939
20.9 Irradiation on Most Widely Studied Surfaces 940
20.9.1 Fixed Surfaces 943
20.9.2 Sun-tracking Surfaces 945
20.9.3 Concentrators 946
20.10 PV Generator Behaviour under Real Operation Conditions 947
20.10.1 The Selected Methodology 949
20.10.2 Second-order Effects 953
20.11 Reliability and Sizing of Stand-alone PV Systems 956
20.12 The Case of Solar Home Systems 962
20.13 Energy Yield of Grid-connected PV Systems 964
20.14 Conclusions 966
Acknowledgements 967
References 967

21 Economic Analysis and Environmental Aspects of Photovoltaic Systems 971
Richard A. Whisnant, Stephen A. Johnston and James H. Hutchby
21.1 Background 972
21.2 Economic Analysis 973
21.2.1 Key Concepts 973
21.2.2 General Methodology 980
21.2.3 Case Studies 984
21.3 Energy Payback and Air Pollution Reduction 997
21.4 Prospects for the Future 999
References 1003

22 PV in Architecture 1005
Tjerk H. Reijenga
22.1 Introduction 1005
22.1.1 Photovoltaics (PV) as a Challenge for Architects and Engineers 1005
22.1.2 Definition of Building Integration 1006

22.2 PV in Architecture 1008
22.2.1 Architectural Functions of PV Modules 1008
22.2.2 PV as Part of "Green Design" 1011
22.2.3 PV Integrated as Roofing Louvres, Facades and Shading 1011
22.2.4 Well-integrated Systems 1014
22.2.5 Integration of PV Modules in Architecture 1019
22.2.6 Brundtland Centre, Toftlund (DK) – a Case Study 1022
22.3 BIPV Basics 1026
22.3.1 Categories and Type of Buildings 1026
22.3.2 Cells and Modules 1029
22.4 Steps in the Design Process with PV 1036
22.4.1 Urban Aspects 1036
22.4.2 Practical Rules for Integration 1037
22.4.3 Step-by-step Design 1038
22.4.4 Design Process: Strategic Planning 1039
22.5 Conclusions 1040
References 1041
Further Reading 1042

23 Photovoltaics and Development 1043
Jorge M. Huacuz and Lalith Gunaratne
23.1 Electricity and Development 1043
23.1.1 Energy and the Early Man 1043
23.1.2 Let There be Electricity 1044
23.1.3 One Third of Humanity Still in Darkness 1044
23.1.4 The Centralized Electrical System 1045
23.1.5 Rural Electrification 1045
23.1.6 The Rural Energy Scene 1046
23.2 Breaking the Chains of Underdevelopment 1046
23.2.1 Electricity Applications in the Rural Setting 1046
23.2.2 Basic Sources of Electricity 1047
23.3 The PV Alternative 1048
23.3.1 PV Systems for Rural Applications 1049
23.3.2 Barriers to PV Implementation 1051
23.3.3 Technical Barriers 1052
23.3.4 Nontechnical Issues 1055
23.3.5 Trained Human Resources 1059
23.4 Four Examples of PV Rural Electrification 1061
23.4.1 Argentina 1061
23.4.2 Bolivia 1061
23.4.3 Brazil 1063
23.4.4 Mexico 1064
23.4.5 Sri Lanka 1065
23.4.6 Water Pumping in the Sahel 1067
23.5 Toward a New Paradigm for Rural Electrification 1068
References 1069

24 Financing PV Growth **1073**
Michael T. Eckhart, Jack L. Stone and Keith Rutledge
24.1 Historical Development of PV Financing 1073
24.2 Capital Requirements 1075
24.2.1 Market Drivers 1075
24.2.2 Growth Outlook 1075
24.2.3 Capital Requirements 1076
24.3 Financial Characteristics of PV 1077
24.4 Financing PV for Grid-connected Residences 1079
24.4.1 Impact of Loan Terms on End-user Cost 1079
24.4.2 Types of Residential Financing 1080
24.4.3 Lender's Issues 1081
24.4.4 Borrowers' Experience 1081
24.4.5 Example Calculation 1082
24.4.6 Improving the Financing of Residential PV 1082
24.5 Financing PV in Rural Areas of Developing Countries 1083
24.5.1 Rural Applications 1083
24.5.2 Impact of Financing on Market Demand 1084
24.5.3 Examples of PV Financing in Rural Areas 1085
24.6 Sources of International Financing 1086
24.6.1 International Aid and Donor Funding 1086
24.6.2 United Nations 1087
24.6.3 World Bank Solar Home System Projects 1088
24.6.4 International Finance Corporation (IFC) 1089
24.6.5 Global Environment Facility 1089
24.7 Financing the PV Industry 1091
24.7.1 Financing Working Capital in the Distribution Channels 1092
24.8 Government Incentives and Programs 1092
24.8.1 Potential Impact of Financing as a Government Policy Option 1092
24.8.2 Direct Subsidies ("Buy-downs") 1094
24.8.3 Soft Loans (Interest Subsidies) 1095
24.8.4 Income Tax Deductions and Credits 1096
24.9 Funding Government Research and Development 1096
24.9.1 PV Programs in the United States 1096
24.9.2 PV Programs in Japan 1097
24.9.3 PV Programs in Europe 1097
24.9.4 Future PV R&D Programs 1099
24.9.5 Sources of R&D Funding 1099
Annex 1100
References 1114

Index **1117**

1

Status, Trends, Challenges and the Bright Future of Solar Electricity from Photovoltaics

Steven S. Hegedus[1] and Antonio Luque[2]

[1]*Institute of Energy Conversion, University of Delaware, Newark, Delaware, USA,* [2]*Universidad Politecnica de Madrid, Madrid, Spain*

1.1 THE BIG PICTURE

Congratulations! You are reading a book about a technology that has changed the way we think about energy. Solar electricity, also known as photovoltaics (PV), has shown since the 1970s that the human race can get a substantial portion of its electrical power without burning fossil fuels (coal, oil or natural gas) or creating nuclear fission reactions. Photovoltaics helps us avoid most of the threats associated with our present techniques of electricity production and also has many other benefits. Photovoltaics has shown that it can generate electricity for the human race for a wide range of applications, scales, climates, and geographic locations. Photovoltaics can bring electricity to a rural homemaker who lives 100 kilometers and 100 years away from the nearest electric grid connection in her country, thus allowing her family to have clean, electric lights instead of kerosene lamps, to listen to a radio, and to run a sewing machine for additional income. Or, photovoltaics can provide electricity to remote transmitter stations in the mountains allowing better communication without building a road to deliver diesel fuel for its generator. It can help a major electric utility in Los Angeles, Tokyo, or Madrid to meet its peak load on hot summer afternoons when air conditioners are working full time. It allows homes and businesses a new level of guaranteed energy availability and security, and photovoltaics has been powering satellites orbiting the Earth or flying to Mars for over 30 years.

Photovoltaics is an empowering technology that allows us to do totally new things, as well as, do old things better. It allows us to look at whole new modes of supplying

Handbook of Photovoltaic Science and Engineering. Edited by A. Luque and S. Hegedus
© 2003 John Wiley & Sons, Ltd ISBN: 0-471-49196-9

electricity to different markets around the world and out of the world (in outer space). It also allows us to do what we already do (generate electricity, which is distributed over the transmission grid) but to do it in a sustainable, pollution-free, equitable fashion. Why is photovoltaics equitable? Because nearly every one has access to sunlight!

Electricity is the most versatile form of energy we have. It is what allows citizens of the developed countries to have nearly universal lighting on demand, refrigeration, hygiene, interior climate control in their homes, businesses and schools, and widespread access to various electronic and electromagnetic media. Access to and consumption of electricity is closely correlated with quality of life. Figure 1.1 shows the Human Development Index (HDI) for over 60 countries, which includes over 90% of the Earth's population, versus the annual per capita electricity use (adapted from ref 1). The HDI is compiled by the UN and calculated on the basis of life expectancy, educational achievement, and per capita Gross Domestic Product. To improve the quality of life in many countries, as measured by their HDI, will require increasing their electricity consumption by factors of 10 or more, from a few hundred to a few thousand kilowatt-hrs (kWh) per year. How will we do it? Our choices are to continue applying the answers of the last century such as burning more fossil fuels (and releasing megatons of CO_2, SO_2, and NO_2) or building more nuclear plants (despite having no method of safely disposing of the high-level radioactive waste) or to apply the new millennium's answer of renewable, sustainable, nonpolluting, widely available clean energy like photovoltaics and wind. (Wind presently generates over a thousand times more electricity than photovoltaics but it is very site-specific, whereas photovoltaics is generally applicable to most locations.)

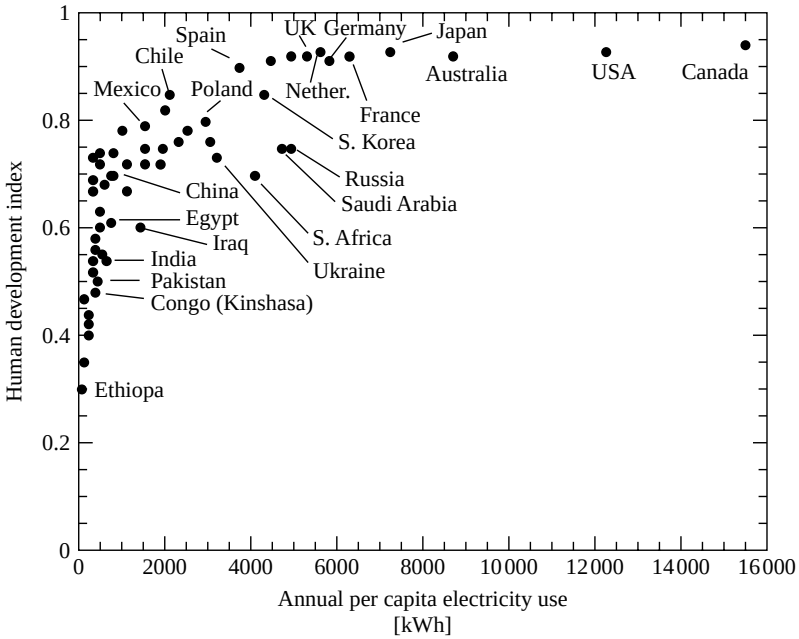


Figure 1.1 Human development index (HDI) vs. per capita kW usage [1]

1.2 WHAT IS PHOTOVOLTAICS?

Photovoltaics is the technology that generates direct current (DC) electrical power measured in Watts (W) or kiloWatts (kW) from semiconductors when they are illuminated by photons. As long as light is shining on the solar cell (the name for the individual PV element), it generates electrical power. When the light stops, the electricity stops. Solar cells never need recharging like a battery. Some have been in continuous outdoor operation on Earth or in space for over 30 years.

Table 1.1 lists some of the advantages and disadvantages of photovoltaics. Note, that they include both technical and nontechnical issues. Often, the advantages and disadvantages of photovoltaics are almost completely opposite of conventional fossil-fuel power plants. For example, fossil-fuel plants have disadvantages of: a wide range of environmentally hazardous emissions, parts which wear out, steadily increasing fuel costs, they are not modular (deployable in small increments), and they suffer low public opinion (no one wants a coal burning power plant in their neighborhood). Photovoltaics suffers none of these problems. The two common traits are that both PV and fossil fueled power plants are very reliable but lack the advantage of storage.

Notice that several of the disadvantages are nontechnical but relate to economics and infrastructure. They are partially compensated for by a very high public acceptance and awareness of the environmental benefits. During the late 1990s, the average growth rate of PV production was over 33% per annum.

What is the physical basis of PV operation? Solar cells are made of materials called semiconductors, which have weakly bonded electrons occupying a band of energy

Table 1.1 Advantages and disadvantages of photovoltaics

Advantages of photovoltaics	Disadvantages of photovoltaics
Fuel source is vast and essentially infinite	Fuel source is diffuse (sunlight is a relatively low-density energy)
No emissions, no combustion or radioactive fuel for disposal (does not contribute perceptibly to global climate change or pollution)	
Low operating costs (no fuel)	High installation costs
No moving parts (no wear)	
Ambient temperature operation (no high temperature corrosion or safety issues)	
High reliability in modules (>20 years)	Poorer reliability of auxiliary (balance of system) elements including storage
Modular (small or large increments)	
Quick installation	
Can be integrated into new or existing building structures	
Can be installed at nearly any point-of-use	Lack of widespread commercially available system integration and installation so far
Daily output peak may match local demand	Lack of economical efficient energy storage
High public acceptance	
Excellent safety record	

called the *valence band*. When energy exceeding a certain threshold, called the *band gap energy*, is applied to a valence electron, the bonds are broken and the electron is somewhat "free" to move around in a new energy band called the *conduction band* where it can "conduct" electricity through the material. Thus, the free electrons in the conduction band are separated from the valence band by the band gap (measured in units of electron volts or eV). This energy needed to free the electron can be supplied by photons, which are particles of light. Figure 1.2 shows the idealized relation between energy (vertical axis) and the spatial boundaries (horizontal axis). When the solar cell is exposed to sunlight, photons hit valence electrons, breaking the bonds and pumping them to the conduction band. There, a specially made selective contact that collects conduction-band electrons drives such electrons to the external circuit. The electrons lose their energy by doing work in the external circuit such as pumping water, spinning a fan, powering a sewing machine motor, a light bulb, or a computer. They are restored to the solar cell by the return loop of the circuit via a second selective contact, which returns them to the valence band with the same energy that they started with. The movement of these electrons in the external circuit and contacts is called the *electric current*. The potential at which the electrons are delivered to the external world is slightly less than the threshold energy that excited the electrons; that is, the band gap. Thus, in a material with a 1 eV band gap, electrons excited by a 2 eV photon or by a 3 eV photon will both still have a potential of slightly less than 1 V (i.e. the electrons are delivered with an energy of 1 eV). The electric power produced is the product of the current times the voltage; that is, power is the number of free electrons times their potential. Chapter 3 delves into the physics of solar cells in much greater detail.

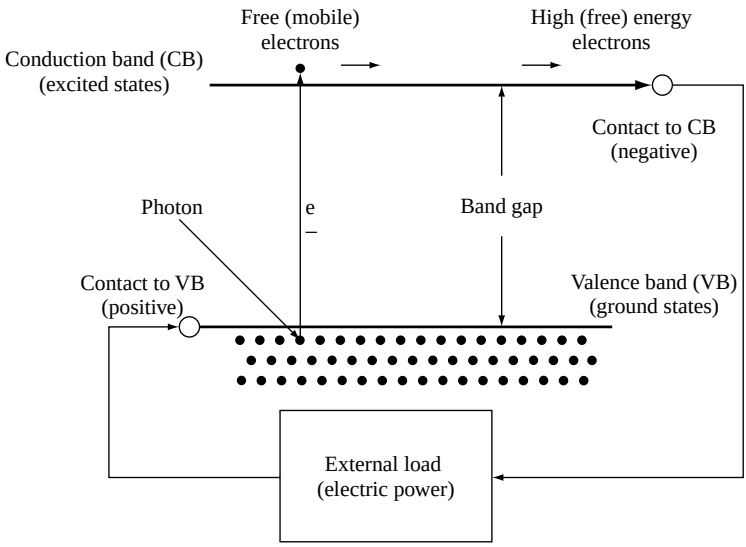


Figure 1.2 Schematic of a solar cell. Electrons are pumped by photons from the valence band to the conduction band. There they are extracted by a contact selective to the conduction band (an *n*-doped semiconductor) at a higher (free) energy and delivered to the outside world via wires, where they do some useful work, then are returned to the valence band at a lower (free) energy by a contact selective to the valence band (a *p*-type semiconductor)

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 742

Sunlight is a spectrum of photons distributed over a range of energy. Photons whose energy is greater than the band gap energy (the threshold energy) can excite electrons from the valence to conduction band where they can exit the device and generate electrical power. Photons with energy less than the energy gap fail to excite free electrons. Instead, that energy travels through the solar cell and is absorbed at the rear as heat. Solar cells in direct sunlight can be somewhat (20–30°C) warmer than the ambient air temperature. Thus, PV cells can produce electricity without operating at high temperature and without mobile parts. These are the salient characteristics of photovoltaics that explain safe, simple, and reliable operation.

At the heart of any solar cell is the *pn* junction. Modeling and understanding is very much simplified by using the *pn* junction concept. This *pn* junction results from the "doping" that produces conduction-band or valence-band selective contacts with one becoming the *n*-side (lots of negative charge), the other the *p*-side (lots of positive charge). The role of the *pn* junction and of the selective contacts will be explained in detail in Chapters 3 and 4. Here, *pn* junctions are mentioned because this term is often present when talking of solar cells, and is used occasionally in this chapter.

Silicon (Si), one of the most abundant materials in the Earth's crust, is the semiconductor used in crystalline form (c-Si) for 90% of the PV applications today (Chapter 5). Surprisingly, other semiconductors are better suited to absorb the solar energy spectrum. This puzzle will be explained further in Section 1.10. These other materials are in development or initial commercialization today. Some are called thin-film semiconductors, of which amorphous silicon (a-Si) (Chapter 12), copper indium gallium diselenide ($Cu(InGa)Se_2$ or CIGS) (Chapter 13), and cadmium telluride (CdTe) (Chapter 14) receive most of the attention. Solar cells may operate under concentrated sunlight (Chapter 11) using lenses or mirrors as concentrators allowing a small solar cell area to be illuminated with the light from larger area. This saves the expensive semiconductor but adds complexity to the system, since it requires tracking mechanisms to keep the light focused on the solar cells when the sun moves in the sky. Silicon and III-V semiconductors (Chapter 9), made from compounds such as gallium arsenide (GaAs) and gallium indium phosphide (GaInP) are the materials used in concentrator technology that is still in its demonstration stage.

For practical applications, a large number of solar cells are interconnected and encapsulated into units called PV modules, which is the product usually sold to the customer. They produce DC current that is typically transformed into the more useful AC current by an electronic device called *an inverter*. The inverter, the rechargeable batteries (when storage is needed), the mechanical structure to mount and aim (when aiming is necessary) the modules, and any other elements necessary to build a PV system are called the *balance of the system* (BOS). These BOS elements are presented in Chapters 17 to 19.

1.3 SIX MYTHS OF PHOTOVOLTAICS

Borrowing a format for discussing photovoltaics from Kazmerski [2], in this section, we will briefly present and then dispel six common myths about photovoltaics. In the following sections, we identify serious challenges that remain despite 40 years of progress in photovoltaics.

The six myths are as follows:

1. *Photovoltaics will require too much land area to ever meet significant fraction of world needs*:

Solar radiation is a rather diffuse energy source. What area of PV modules is needed to produce some useful amounts of power? Let's make some very rough estimates to give answers that will be accurate within a factor of 2. Using methods described in detail in Chapter 20 (especially equations 20.50 and 20.51 and Table 20.5), one can calculate how much sunlight falls on a square meter, anywhere in the world, over an average day or a year. We will use an average value of 4 kilowatt-hrs (kWh) per m^2 per day to represent a conservative worldwide average. Now, a typical PV module is approximately 10% efficient in converting the sunlight into electricity, so every square meter of PV module produces, on average, $4 \times 0.1 = 0.4$ kWh of electrical energy per day. We can calculate the area in m^2 needed for a given amount of electrical energy E in kWh by dividing E by 0.4 kWh/m^2. (Chapter 20 contains much more detailed methods to calculate the incident sunlight and the PV module output as a function of time of day, month of year, etc.)

Let us consider three different-sized PV applications: a family's house in an industrialized country, replacing a 1000 MW (megawatt) coal or nuclear powered generating plant, or providing all the electricity used in the USA.

First, for a typical family, let us assume that there are four people in the house. Figure 1.1 shows a range of electricity usage for the industrialized countries. Let us use 6000 kWh/person/year as an average. But, this includes all their electrical needs including at work, at school, as well as the electricity needed for manufacturing the products they buy, powering their street lights, pumping water to their homes, and so on. Since people spend about a third of the day awake in their home, let us assume that a third of their electrical needs are to be supplied in their home, or 2000 kWh/person/year. Dividing this by 365 days in a year gives about 5 kWh/person/day, or 20 kWh/day per family of four. This is consistent with household data from various sources for the US and Europe. Thus, they would need 20 kWh/0.4 kWh/m^2 or 50 m^2 of solar modules to provide their electrical power needs over the year. Thus, a rectangular area of solar modules of 5 by 10 meters will be sufficient. In fact, many roofs are about this size, and many homes have sunny areas of this size around them, so it is possible for a family of four, with all the conveniences of a typical modern home, to provide all their power from PV modules on their house or in their yard.

Next, how much land would it take to replace a 1000 MW coal or nuclear power plant that operates 24 hours/day and might power a large city? This would require 10^6 kW $\times$ 24 hr/(0.4 kWh/m^2) or 6×10^7 m^2. So, with 60 km^2 (or 24 square miles) of photovoltaics we could replace one of last century's power plants with one of this century's power plants. This is a square 8 km (or 5 miles) on a side. For the same electricity production, this is equivalent to the area for coal mining during the coal powered plant's life cycle, if it is surface mining, or three times the area for a nuclear plant, counting the uranium mining area [3]. This is also the same area required to build a 600 km (373 miles) long highway (using a 100 m wide strip of land).

Finally, we can calculate how much land is needed to power the entire US with photovoltaics (neglecting the storage issue). The US used about 3.6×10^{12} kWh of electricity in 2000. This could be met with 2×10^{10} m^2. If we compare with the area of paved roads across the country, of about 3.6×10^6 km and assume an average width

of 10 m this leads to 3.6×10^{10} m^2. It is to be concluded that all the electricity needed in the US can be met by covering the paved roads with PV modules. Of course, no one is seriously proposing this action. We use the road analogy to show that if society wanted, it could establish land use priorities favorable to photovoltaics just as it has done to accommodate the ubiquitous automobile. We are certain that each state could find areas of unused land around airports, parking spaces, rooftops, highway dividing strips, or desert land that could be used for photovoltaics.

These simplistic "back-of-the-envelope" calculations show that having enough area for PV modules is not a limit for a homeowner or a large city. Certainly, there are sunny places in every country that could be used for generating significant amounts of PV power. As will be evident in other chapters, it is the initial cost of the photovoltaics, not the amount of land that is the primary barrier to be overcome.

2. *Photovoltaics can meet all of the world's needs today if we would just pass laws requiring photovoltaics and halting all fossil and nuclear plants*:

Besides the difficulty of convincing the people's representatives to pass such a law, the first technical problem faced would be the intermittent nature of the solar radiation, available only during the day and strongly reduced in overcast skies. Energy storage would solve this problem but no cheap storage method appears on the horizon. Nevertheless, well-developed electric grids may accept large amounts of PV electricity by turning off some conventional power plants when PV plants are delivering power. Adequate grid management would allow up to 20 to 30% of the electric production to be intermittent [4].

But now for a dose of reality. The cumulative production of PV modules up to the year 2002 is about 2000 MW. Thus, if you took all of the PV panels that were ever made up to and including the year 2002, and put them all in the same sunny place at the same time, they would generate enough electricity to displace about one of last century's 500 MW smoke- or radioactive-waste–producing power plants. (This assumes that the solar plant would operate at full output for an equivalent of six hours per day owing to the daily variation in sunlight). Clearly, if we want photovoltaics to make any meaningful contribution to the world's energy supply, very massive increases in manufacturing capacity are needed. Additionally, PV electricity is very expensive, presently between 5 to 10 times more expensive than conventional alternatives. Mass use of PV electricity today could produce significant negative distortion of the economic system.

Thus, requesting the immediate and exclusive use of photovoltaics is not feasible technically or, probably, economically. It would also be socially unacceptable.

3. *Photovoltaics cannot meet any significant fraction of world needs. It will remain a small-scale "cottage" industry that will only meet the needs of specialty markets like remote homes in developing countries or space satellites*:

Figure 1.3 shows the evolution of markets associated with different applications [5]. Some used to be considered as *specialty* markets, for example, the category of "world off-grid power" which is trying to supply power to the $\sim$1/3 of the world's citizens who lack it. The grid-connected market, whose growth has been meteoric in the past decade, is by no means a small market. Ironically it is the large-scale (recently awakened) centralized power plant market which is the smallest "specialty" application in today's world. Thus, evidence from the recent past tends to refute

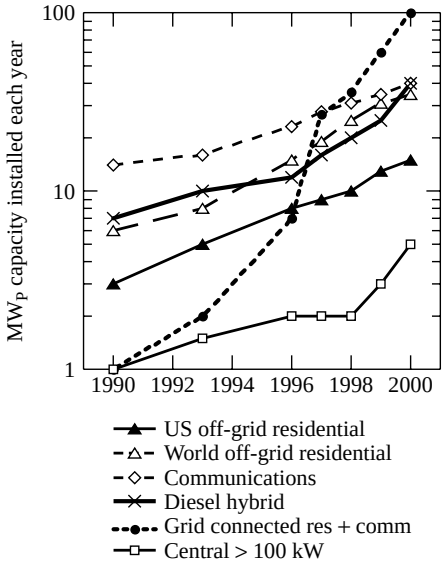


Figure 1.3 Trend in worldwide PV applications (From Reference [5] Maycock P, *Renewable Energy World* **3**, 59–74 (2000)

the modest forecasts that some attribute to photovoltaics. We shall come to this point again.

4. *No more R/D is needed since PV technology has demonstrated the technical capability to perform, so we should stop all public funding and let the economic markets decide if it is worthwhile*:

The present cost of photovoltaics is affordable for certain markets but it is still too high to actually compete with conventional electricity. If PV technology is to be promoted for environmental or social reasons, public subsidy to R&D and to installation will be necessary to stimulate production and thereby reduce costs. Without continued subsidies, photovoltaics will probably remain as a specialty cottage industry for the next half century.

Public support for photovoltaics is one of the major factors compelling politicians to fund R&D. This funding had been comparable to PV sales in the 1980s, as shown in Figure 1.4. Private funding has doubled this public support so that PV companies themselves have also heavily supported the development of photovoltaics. After two decades of constant investment in a promising market that was slow to actually start, the market finally awoke and became one of the fastest growing in the world by the beginning of the twentieth century, with sales now greatly exceeding public investment.

But, this fast growing market is still dependent on public/government funding. As with many goods and services (e.g. military hardware, commercial air travel), photovoltaics is partly publicly financed. In Germany or in Japan, for instance, significant public support is being given to grid-connected installations. If photovoltaics is going to become a major energy contender, the countries where the support has been lacking will remain technologically inferior with respect to those, where the support has

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 746

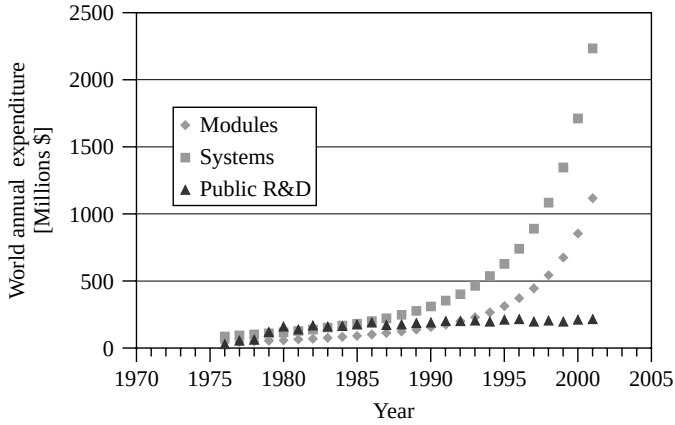


Figure 1.4 Public funding for R&D (triangles) compared to module (diamonds) and system (squares) sales. (This curve is drawn from the data of Eckhart *et al*. in Chapter 24, "Financing PV Growth", in this book)

been stronger. This should be taken into account while making decisions about energy policy and public or private financing.

The critical question then is: Should the support be focused in R&D, or is PV technology already mature enough (as many claim) to focus on the cost reduction via the economy of scale permitted by the larger volume of production required by a subsidized market? This point will be discussed latter in this chapter.

5. *Photovoltaics is polluting just like all high-technology or high-energy industries only with different toxic emissions*:

   One of the most valuable characteristics of photovoltaics is its well-deserved image as an environmentally clean and "green" technology. This healthy image obviously results from the cleaner operation of a PV electricity generator compared to a fossil-fuel fired generator, but this must also extend to the manufacturing process itself as well as the recycling of discarded modules. Manufacturing of PV modules on a large scale requires the handling of large quantities of hazardous or potentially hazardous materials (e.g. heavy metals, reactive chemical solutions, toxic gases). Let it be stated at the beginning that the present Si-based PV technology which dominates the market has few environmental concerns and is considered totally safe to the public.

   The PV industry is very aware of the value of its clean "green" image and has worked hard over the years to establish and maintain high standards of environmental responsibility [6, 7]. Conferences on PV Safety and Environmental Issues have been held since the late 1980s and their proceedings have been published [8, 9]; the PV Environmental Health Safety Assistance Center at Brookhaven National Laboratory in New York, USA provides worldwide leadership in risk analysis and safety recommendations for the PV industry [10].

   Safe handling procedures for some of the materials and processes were already well established from the integrated circuit or glass coating industries. But in the case of unique materials and processes, safety procedures had to be developed by the PV industry. This is especially true of the thin-film technologies [11]. The PV industry recognized early that being proactive and designing safety into the process, from the

beginning, was the responsible thing to do and would ultimately result in reduced costs. The international nature of the PV industry introduces some variability in the standards which must be met.

Hazards can be classified by whether they affect workers at a PV manufacturing plant, customers with photovoltaics on or near their homes, or the public who consumes air and water near the PV plant. The population with greatest potential health risks are employees in PV manufacturing. Very little risk is associated with the public or the PV owner or installer. Among the most heavily studied issues unique to the PV industry is the potential toxicity of semiconductor CdTe and the safe usage of hydride gases AsH_3, SiH_4, GeH_4, PH_3, B_2H_6, and H_2Se, which are used in the growth of GaAs, a-Si, a-SiGe, and $Cu(InGa)Se_2$ layers. There has been considerable research and risk analysis of CdTe as a PV material [12–14]. The general conclusion is that CdTe in modules does not pose a risk to the public. Similarly, procedures and hardware ensuring safe usage of the hydride gases listed above have been well established in both the electronics and PV industries [15].

Environmental monitoring of the workplace for hazardous levels in the air or on surfaces, and biological monitoring of the employee for evidence of exposure should be routine. Once the module is manufactured, the only way for the public to be exposed to hazardous materials existing in some kind of modules is by absorbing them via ingestion or inhalation. Accordingly, accidental human absorption is not at all likely. Even in event of a house fire, studies have shown that PV modules do not release any potentially hazardous materials [16].

A related issue is what to do with thin film PV modules at the end of their projected 25- to 30-year life. An excellent strategy is to recycle the modules. This solves two problems at once, namely, keeping potentially hazardous materials out of the environment and reducing the need for additional mining and/or refining of new materials. Semiconductor vendors have indicated a willingness to accept used modules, and to extract and purify the CdTe, CdS, or $Cu(InGa)Se_2$ for resale and reuse [17, 18].

Thus, we can say with confidence that photovoltaics is nearly the cleanest and safest technology with which to generate electricity. It is especially true of the present Si technology.

6. *PV modules never recover all of the energy required in making them, thus they represent a net energy loss*:

The focus of photovoltaics is on generating energy (specifically electrical energy) with many beneficial characteristics as noted in Table 1.1. Among those who envision photovoltaics having an increasingly larger role in producing the world's electricity, there is awareness that photovoltaics must produce much more energy than was required to produce the PV system. Otherwise, it would be a net energy loss not a net energy source. The "energy payback" has been widely studied. It is described in terms of how many years the PV system must operate to produce the energy required for its manufacture. After the payback time, all of the energy produced is truly new energy.

This topic is discussed in Chapter 21. An excellent review has been given by Alsema [19]. In general, results of several studies have arrived at some general conclusions. Specific payback times have ranged from 3 to 5 years for crystalline Si and 1 to 4 years for thin films. For crystalline Si, forming the crystalline Si wafers is

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 748

the major energy requirement. For thin films, the semiconductor layers are 100 times thinner, and deposited at ~1000°C lower temperature, so their energy requirement is negligible, in comparison. Instead, it is the energy embodied in the glass or stainless steel substrate, which is the major energy sink. Also, a seemingly insignificant component, the cosmetic Al frame around the module, is responsible for a surprisingly large fraction of energy. In fact, this can be the dominant energy sink for thin-film a-Si or Cu(InGa)Se_2 modules [20, 21]. Although thin-film modules have a shorter energy payback, they also have lower efficiency, which means a larger BOS is needed to support the larger number of modules. Thus, a larger amount of energy is embodied in the BOS for thin-film photovoltaics compared to crystalline Si photovoltaics.

The case of concentrators is less studied, but again the use of semiconductor is reduced and the BOS becomes more important than even for the thin films because the concentrating structures are very massive. However, their efficiency is higher. In summary, we can guess that in this case the situation will be similar to the case of thin films.

1.4 HISTORY OF PHOTOVOLTAICS

The history of photovoltaics goes back to the nineteenth century, as shown in Table 1.2. The first functional, intentionally made PV device was by Fritts [22] in 1883. He melted Se into a thin sheet on a metal substrate and pressed a Au-leaf film as the top contact. It was nearly 30 cm^2 in area. He noted, "the current, if not wanted immediately, can be either stored where produced, in storage batteries, ... or transmitted a distance and there used." This man foresaw today's PV technology and applications over a hundred years ago. The modern era of photovoltaics started in 1954 when researchers at Bell Labs in the USA accidentally discovered that *pn* junction diodes generated a voltage when the room lights were on. Within a year, they had produced a 6% efficient Si *pn* junction solar cell [23]. In the same year, the group at Wright Patterson Air Force Base in the US published results of a thin-film heterojunction solar cell based on Cu_2S/CdS also having 6% efficiency [24]. A year later, a 6% GaAs *pn* junction solar cell was reported by RCA Lab in the US [25]. By 1960, several key papers by Prince [26], Loferski [27], Rappaport and Wysoski [28], Shockley (a Nobel laureate) and Queisser [29], developed the fundamentals of *pn* junction solar cell operation including the theoretical relation between band gap, incident spectrum, temperature, thermodynamics, and efficiency. Thin films of CdTe were also producing cells with 6% efficiency [30]. By this time, the US space program was utilizing Si PV cells for powering satellites. Since space was still the primary application for photovoltaics, studies of radiation effects and more radiation-tolerant devices were made using Li-doped Si [31]. In 1970, a group at the Ioffe Institute led by Alferov (a Nobel laureate), in the USSR, developed a heteroface GaAlAs/GaAs [32] solar cell which solved one of the main problems that affected GaAs devices and pointed the way to new device structures. GaAs cells were of interest due to their high efficiency and their resistance to the ionizing radiation in outer space. The year 1973 was pivotal for photovoltaics, in both technical and nontechnical areas. A significant improvement in performance occurring in 1973 was the "violet cell" having an improved short wavelength response leading to a 30% relative increase in efficiency over state-of-the-art Si cells [33]. GaAs heterostructure cells were also developed at IBM in

Table 1.2 Notable events in the history of photovoltaics

- 1839 Becquerel (FR) discovered photogalvanic effect in liquid electrolytes
- 1873 Smith (UK) discovered photoconductivity of solid Se
- 1877 Adams and Day (UK) discover photogeneration of current in Se tubes; the first observation of PV effect in solids
- 1883 Fritts (US) makes first large area solar cell using Se film
- 1954 First 6% efficient solar cells reported: Si (Bell Lab, USA) and Cu_2S/CdS (Air Force, USA)
- 1955 Hoffman Electronics (USA) offers 2% efficient Si PV cells at $1500/W
- 1958 NASA Vanguard satellite with Si backup solar array
- 1959 Hoffman Electronics (USA) offers 10% efficient Si PV cells
- 1963 Sharp Corp (JP) produces first commercial Si modules
- 1966 NASA Orbiting Astronomical Observatory launched with 1 kW array
- 1970 First GaAs heterostructure solar cells by Alferov, Andreev *et al.* in the USSR
- 1972 First PV conference to include a session on terrestrial applications (IEEE)
- 1973 A big year in photovoltaics: Worldwide oil crisis spurs many nations to consider renewable energy including photovoltaics; Cherry Hill Conference in USA (established photovoltaics' potential and legitimacy for government research funding); World's first solar powered residence (University of Delaware, USA) built with Cu_2S (not c-Si!) solar modules
- 1974 Project Sunshine initiated in Japan to foster growth of PV industry and applications; Tyco (USA) grows 2.5 cm wide Si ribbon for photovoltaics, first alternative to Si wafers
- 1975 First book dedicated to PV science and technology by Hovel (USA)
- 1980 First thin-film solar cell >10% using Cu_2S/CdS (USA)
- 1981 350 kW Concentrator array installed in Saudi Arabia
- 1982 First 1 MW utility scale PV power plant (CA, USA) with Arco Si modules on 2-axis trackers
- 1984 6 MW array installed in Carrisa Plains CA, USA [35]
- 1985 A big year for high-efficiency Si solar cells: Si solar cell >20% under standard sunlight (UNSW, Australia) [36] and >25% under 200X concentration (Stanford Univ. USA) [37]
- 1986 First commercial thin-film power module, the a-Si G4000 from Arco Solar (USA)
- 1987 Fourteen solar powered cars complete the 3200 km World Solar Challenge race (Australia) with the winner averaging 70 kph
- 1994 GaInP/GaAs 2-terminal concentrator multijunction >30% (NREL, USA) [38]
- 1995 "1000 roofs" German demonstration project to install photovoltaics on houses, which triggered the present favorable PV legislation in Germany, Japan and other countries
- 1996 Photoelectrochemical "dye-sensitized" solid/liquid cell achieves 11% (EPFL, Switzerland) [39]
- 1997 Worldwide PV production reaches 100 MW per year
- 1998 Cu(InGa)Se_2 thin-film solar cell reaches 19% efficiency (NREL, US) [40] comparable with multicrystalline Si. First concentrating array for space launched on Deep Space 1 by US (5 kW using high efficiency GaInP/GaAs/Ge triple junction cells)
- 1999 Cumulative worldwide installed photovoltaics reaches 1000 MW
- 2000 Olympics in Australia highlight wide range of PV applications, and the awarding of the first Bachelor of Engineering degrees in Photovoltaics and Solar Engineering (UNSW, Australia)
- 2002 Cumulative worldwide installed photovoltaics reaches 2000 MW. It took 25 years to reach the first 1000 MW and only 3 years to double it; production of crystalline Si cells exceeds 100 MW per year at Sharp Corp. (Japan). BP Solar ceases R&D and production of a-Si and CdTe thin-film modules in USA ending >20 years of effort

the USA having 13% efficiency [34]. Also in 1973, a crucial nontechnical event occurred called the *Cherry Hill Conference*, named after the town in New Jersey, USA, where a group of PV researchers and heads of US government scientific organizations met to evaluate the scientific merit and potential of photovoltaics. The outcome was the decision that photovoltaics was worthy of government support, resulting in the formation of the US Energy Research and Development Agency, the world's first government group whose mission included fostering research on renewable energy, which ultimately became the US Dept. of Energy. Finally, in October 1973, the first World Oil Embargo was instituted by the Persian Gulf oil producers. This sent shock waves through the industrialized world, and most governments began programs to encourage renewable energy especially solar energy. Some would say this ushered in the modern age of photovoltaics and gave a new sense of urgency to research and application of photovoltaics in terrestrial applications.

In the 1980s, the industry began to mature, as emphasis on manufacturing and costs grew. Manufacturing facilities for producing PV modules from Si wafer *pn* junction solar cells were built in the USA, Japan, and Europe. New technologies began to move out of the government, university and industrial laboratories, and into precommercialization or "pilot" line production. Companies attempted to scale up the thin-film PV technologies like a-Si and $CuInSe_2$, which had achieved $>10\%$ efficiency for small area (1 cm^2) devices made with carefully controlled laboratory scale equipment. Much to their disappointment, they found that this was far more complicated than merely scaling the size of the equipment. Most large US semiconductor companies, gave up their R/D efforts (IBM, General Electric, Motorola) lacking large infusions of private or government support to continue. One common result was the purchase of American companies and their technologies by foreign companies. In 1990, the world's largest solar manufacturer was Arco Solar (CA, USA), owned by oil company Atlantic Richfield, which had c-Si and thin-film a-Si in production and thin-film $CuInSe_2$ in precommercialization. They were sold to the German firm Siemens and renamed Siemens Solar (in 2001, the Dutch company Shell Solar would buy Siemens, becoming another large internationally based company with multiple PV technologies in production). Also in 1990, Energy Conversion Devices (MI, USA) formed a joint venture called United Solar Systems Corp. with the Japanese manufacturer Canon to commercialize ECD's roll-to-roll triple-junction a-Si technology. In 1994, Mobil Solar Energy (MA, USA), which had developed a process for growing solar cells on Si ribbon (called the Edge defined film growth or EFG process) instead of more costly wafers, was sold to the German company ASE and renamed ASE Americas. The British solar company BP Solar acquired patents to electrodeposition of thin-film CdTe solar cells in 1989, when it's parent company purchased the American oil giant Standard Oil of Ohio. At the same time, it acquired the patents of the University of New South Wales (Australia) to fabricate the Laser-Grooved Buried-Grid (LGBG) cells, which became the most efficient silicon cells in fabrication. In 1996, it signed a license agreement with the Polytechnic University of Madrid (Spain) to exploit the Euclides concentration technology that used their LGBG cells as concentrator cells. In 1999, BP Solar acquired Solarex from Enron (another huge fossil-fuel energy company) that had crystalline and amorphous Si solar cell technology. Thus, BP Solar established themselves with manufacturing interests in all three technology options (standard Si wafers, thin films

and concentrators).[1] Meanwhile, the Japanese PV industry began to take off. Production of c-Si modules and intensive research on thin-film technology in Japan led to many innovative device designs, improved materials processing, and growing dominance in the world PV market.

Along with the maturing of the solar cell technology, the BOS needed to grow. Many products like inverters, which convert the DC power into AC power, and sun trackers had only limited application outside of a PV power system, so once again there was only limited technical and financial resources for development. In many system evaluations, the inverter was repeatedly identified as the weak link in terms of reliability and AC power quality [41]. Their costs have not fallen nearly as fast as those for the PV modules. While much effort and resources had been focused on the solar cell cost and performance, little attention had been paid to installation and maintenance costs. It was quickly discovered that there was room for much improvement.

An early development that helped many companies was to sell PV cells for consumer-sized, small-scale power applications. The solar-powered calculator, pioneered by Japanese electronics companies as a replacement for battery-powered calculators in the early 1980s, is the best-known example. This led to the early use of thin-film a-Si PV cells for various applications. Another example was solar-powered outdoor lighting. These novel consumer applications, where portability and convenience were more valued than low price, allowed the PV companies to maintain some small income while continuing to develop power modules.

Another application was the rural electrification of remote villages in an attempt to help roughly one-third of the world's citizens to gain access to a modest amount of modern communication and lighting. Most of these PV installations were very small, on the order of 10 to 40 W per household (100 times smaller than the "needs" of a modern home in the developed world.) Most of these installations were funded by some international aid agency. Reviews and follow-up studies of these programs have indicated very large failure rates, primarily due to lack of technical infrastructure [42], training, cultural misunderstandings, design of the payment structure, and other nontechnical reasons [43]. Rarely have the PV modules failed. Even with subsidies from the international agencies, the high initial cost of ownership ($100–1000) was still a major barrier in much of the world where this represents a year's income for a family [44].

On the opposite end of the size scale were the MW-size PV plants installed by utilities in developed countries in the 1980s to evaluate their potential in two applications: as a peak-load-reduction technology, where the photovoltaics provides additional power primarily to meet the peak demand during the afternoon [45]; or as distributed generators, to reduce transmission and distribution losses [46]. Several American utilities investigated these applications, to assess the technical as well as financial benefits for photovoltaics in utility scale applications. Other novel configurations of grid-tied PV

[1] While this book was going to press in November 2002, BP Solar suddenly announced the closure of its two thin-film manufacturing efforts in the United States (a-Si in Virginia and CdTe in California) in order to focus more resources on its multicrystalline Si wafer PV production. This was a great disappointment to all those who worked so hard to establish these thin-film technologies and facilities, which were among the most advanced thin-film PV products in the world.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 752

systems were evaluated as so-called "demand side management" options where the on-site distributed photovoltaics is used to reduce demand rather than increase supply [47]. Although American utilities lost interest in PV in the late 90s due to deregulation, grid-connected applications in Europe and Japan began to grow rapidly, primarily owing to strong government support. Both small- and large-scale grid connected PV installations are blossoming in these countries [48, 49].

Yet another important development in the application of PV in the late 1990s, was building integrated PV (BIPV [50]), where PV cells are incorporated into a standard building product, such as a window or a roof shingle, or into an architectural feature like an exterior sun awning or semitransparent skylight. In this way, the cost of the PV is partially offset by the cost of the building materials, which would have been required anyway, so the incremental cost of the photovoltaics is much lower. BIPV is discussed in Chapter 22. The success of grid-connected residential or BIPV commercial applications has been possible because several countries led by Germany have established high rates to pay for the PV electricity produced by solar installations in private houses. In this scheme, the installation owner receives \$0.5/kWh for the electricity they feed into the public electric grid (as of 2001). But the owner buys the electricity consumed in their own house at the normal cost of ~\$0.1/kWh from the grid. Additionally, German banks provided generous loans for purchasing the installation. Similar concepts are used in Spain, the Netherlands, and other countries in Europe. But, the success has been still bigger in Japan where homebuilders receive a rebate from the government for about 30% of the PV system cost. Then, their electric bill is determined by the utility using the "net metering" where the customer pays only the net difference between what they used and what they generated. Rebates and net metering are available in some, but not all, states in the USA as of 2002. Interestingly, government support of photovoltaics in Japan has been decreasing while the market for PV homes has continued showing an impressive growth rate.

1.5 PV COSTS, MARKETS AND FORECASTS

In the first 20 years of PV research, from the mid 1960s to the mid 1980s, the main focus was to make the product more efficient so it produced more power. Impressive gains in cell and module efficiency were made. Costs also fell dramatically as solar cells moved from pilot scale to semiautomated production.

Although the important figure of merit for cost is \$/kWh, typically \$/W_P is often used. Modules are rated in Watts of peak power (W_P). This is the power the module would deliver to a perfectly matched load when the module is illuminated with 1 kW/m^2 of luminous power of a certain standard spectrum while the cell temperature is fixed at 25°C. (By the way, these "standard test conditions" or STC rarely occur in real outdoor applications! See Chapter 16 for a complete discussion of testing conditions and Chapter 20 for real outdoor conditions.).

Figure 1.5 shows costs (\$/W_P) and production measured in MW_P over the commercial history of photovoltaics. Up until about year 2000, these values represent mostly c-Si solar cell technology. These two curves are typical of most new technologies. Initially, prices are high since volume production is low, so development and start-up

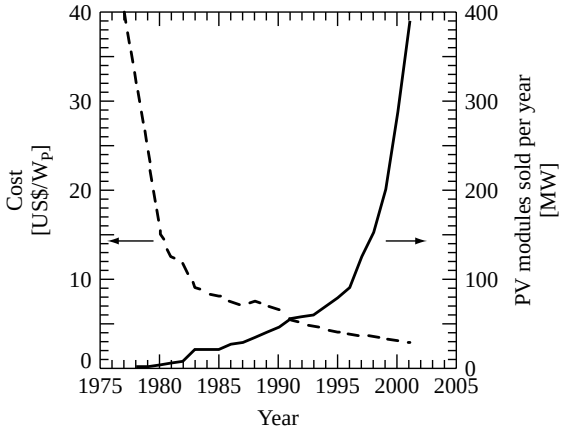


Figure 1.5 Historical trends of cost per Watt for solar cells and volume of production. Data from various sources. Beware: these costs are for PV modules not completed systems, which typically increase by a factor of 2 to 3

costs are spread over the relatively few units sold. The high price excludes most buyers except unique niche applications (i.e. remote telecommunications transmitters, where the unique properties of photovoltaics makes it the most appropriate source of electricity) government-sponsored programs (i.e. satellites, weather monitoring stations, military outposts and also human development programs in remote areas including water pumping), and curious wealthy pioneers (i.e. private homes in the mountains for environmentally concerned millionaires). As volume production increases, costs fall as economies of scale take over. The technology is now within economic reach of wider markets and demand grows rapidly as people with moderate incomes can afford the product. Eventually, the decrease in price slows, and it becomes harder to improve the cost and performance of a given product. But each small decrease in cost opens up larger markets and applications. Once a certain price is reached, a massive new market will open up with ample opportunity for investors to finance new manufacturing capacity.

This relation between cumulative production of PV modules in MW_P (M) and price in $/$W_P$ (p) can be described by an experience curve, which is characterized by a parameter E called the experience exponent [51, 52] or

$$\frac{p(t)}{p_0} = \left[\frac{M(t)}{M_0}\right]^{-E} \tag{1.1}$$

where M_0 and p_0 are the cumulative market and the price at an arbitrary initial time $t = 0$ (that we can take at the beginning of the early commercialization). The experience curve for photovoltaics is shown in Figure 1.6 where lowest price per W_P for a given year is plotted against the cumulative module production up to that year. When graphed as a log–log plot, it is the slope that is of significance since it defines the experience factor given as $1-2^{-E}$. This quantity indicates how much costs are reduced for every doubling of cumulative production. Figure 1.6 presents an exponent $E = 0.30$ which gives an experience factor of 0.19. Thus, prices have fallen 19% for every doubling in

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 754

cumulative production over the past 30 years. If the trend continues, the price of $\$1/W_P$ will be reached when the cumulative production reaches 10^5 MW_P.

It should be said that while the annual growth in sales of photovoltaics is quite spectacular, averaging 33% per year from 1995 to 2000, the experience factor of 0.19 is rather mediocre. For example, for semiconductor memories it is about 0.32, although for wind power it is only 0.15. PV technology has not reduced prices very effectively. This supports the idea that R&D must be supported to look for innovative options able to reduce prices beyond the safe path of the experience curve (additional argument for myth 4).

When this cumulative market will be reached can be determined, if we know the demand elasticity *S*. Technically, this is the logarithmic derivative of the annual market with respect to the price (changed of sign) and shows that one percent of price decrement will produce *S* percent of market increase. This parameter allows us to determine when in the future a certain level of price is reached [52].

Based on this, the installed PV power is given in Figure 1.7 for reasonable values of the demand elasticity. We can observe a fast initial growth, followed by a period of moderate growth. This second period is determined by the investment that society is willing to invest in this expensive energy technology. Extreme curves of the shaded area correspond to the expenditure for photovoltaics of the case that 0.05% (lower curve) and the 0.2% (upper curve) of the GDP of the industrialized countries is invested every year in PV electricity. As a reference, about 0.3% of the industrialized countries GNP is being given today as aid for development.

In 1992, a study (the RIGES scenario [54]) was presented to the Rio Summit analyzing the possibility of reducing the CO_2 emissions while maintaining a high economic growth rate in the developing countries. The dots represent the amount of photovoltaics to attain the RIGES requirements. In other words, the dots are the amount of installed photovoltaics required to be environmentally relevant and socially advanced. Note that

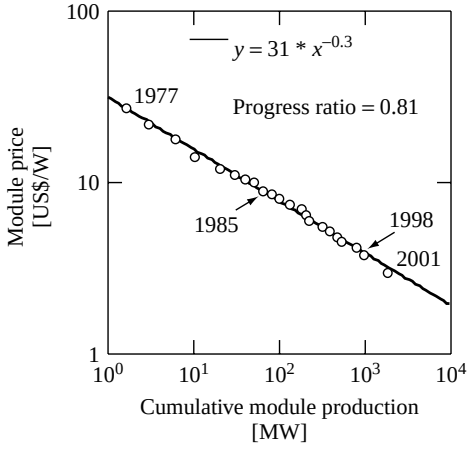


Figure 1.6 Experience curve for photovoltaics from 1976 until 1998 [53]. Straight line is fit indicating an experience factor of $1 - 2^{-0.3} = 0.19$

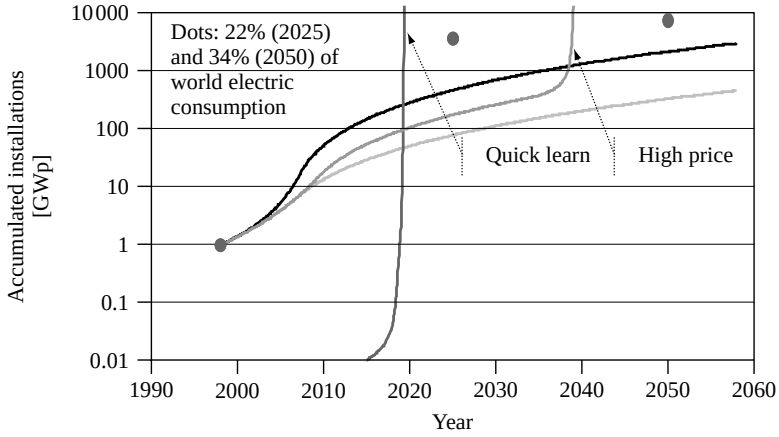


Figure 1.7 Long-term forecast of cumulative installed peak PV power. The dashed area represents the expected cumulative installed capacity for a range of assumptions associated with the demand elasticity. PV prices (not shown) will not be competitive in the period shown unless the conventional electricity doubles in price (in constant dollars). However, photovoltaics will be competitive if some innovation develops and is commercialized with a "quick learn" experience factor like that of the microelectronic memories (0.32). Dots represent the level of electricity penetration to reach environmental goals (see text)

the goal for 2050 is to produce the 34% of the total world electricity production; by no means a niche market to be produced by a cottage industry!

It is observed that with the present low experience factor, the environmental goals are not achieved. Nevertheless, this forecast predicts that the PV industry will be very large by the middle of the century.

The curve labeled high price refers to the case where conventional electricity prices are doubled. In this case, photovoltaics will reach price competitiveness with the existing electricity before the middle of the century (the almost-vertical line means photovoltaics is cheaper than conventional electricity but realistic growth will occur more slowly). Perhaps a hidden but practical conclusion of this analysis is that support to the PV industry will result in an additional element of security in the supply of electricity. Energy security is of increasing interest for the public officials as well as for citizens.

The preceding cases correspond to the situation governed by a constant experience factor. No technological breakthrough is considered. What would be the situation if a breakthrough were produced? A breakthrough technology would be characterized by a higher experience factor and should be able to reduce costs by experience faster than present PV technology. The curve labeled quick learning in Figure 1.7 shows the case of a technology with the experience factor of the microelectronic memories. Note that in very few years (after real commercialization starts) it would be able to reach competition with conventional electricity.

However, note that this success is based on reaching a certain cumulative market (in this exercise 10 000 MW) despite the higher price of this technology at its early stage as compared to the competing PV technology. This is a commonly occurring situation and

must be taken into account by investors. Comparison must not be made between prices of different alternatives at the same moment in time but for equal cumulated market. In other words, we must compare the price of an option today with that of its more mature competitor in some moment of the past and we must also consider the experience factor. In other words, sometimes a product that at a given moment is more expensive than its competition can become cheaper sooner because it bears internally the seed for lower cost and comparable performance. But, of course, not every more-expensive novel product bears this seed. This is the risky nature of entrepreneurship.

We conclude, and this is our position concerning the controversy pointed out in myth 4, that basic R&D in photovoltaics must be looking for breakthroughs if we want photovoltaics to fulfill the goals that society requests. At the same time, support to the market will help create a sizeable industry that will be able to commercialize breakthroughs worldwide. In any case we are not talking of a cottage industry. The PV industry has the potential of becoming a major electricity supplier in the twenty-first century and to constitute a powerful industry able to abate environmental stresses, to facilitate the human development of the poorest, and to constitute an element of safety in our electric supply.

1.6 WHAT ARE THE GOALS OF TODAY'S PV RESEARCH AND MANUFACTURING?

Chapters 5 to 15 in this book focus on individual technologies and will give specific examples of where the research and manufacturing effort is concentrating to reduce costs and improve performance. But several trends are common to all. These are listed in Table 1.3.

Since the overall goal is to produce a low cost PV *system*, we need more than low-cost-efficient solar cells, we need a low cost efficient system including mounting hardware, power conditioning electronics, fuses, cables, storage, tracking, and so on. Less research and development has gone into these areas than into PV solar cells and modules.

Table 1.3 Goals of current solar cell research and manufacturing

- Use less semiconductor material by making thinner cells or
- Use less expensive semiconductor materials. These tend to be less pure and less perfect.
- Improve solar cell performance with less expensive, less perfect semiconductors
- Even with this poorer material keep a high production yield, that is, reduce the number of cells or modules rejected by the quality control.
- Increase material utilization by reducing waste in semiconductor and cell fabrication
- Increase solar cell flux on the solar cells by using concentrators without increasing cost or optical losses too much. In this way, less semiconductor material is used.
- Increase solar radiation utilization by absorbing more of the spectrum efficiently
- Increase speed and throughput of manufacturing processes
- Simplify processing steps (this reduces fabrication costs and increases the yield) and reduce equipment costs
- Reduce costs and improve reliability of BOS (auxiliary elements).

1.7 GLOBAL TRENDS IN PERFORMANCE AND APPLICATIONS

Figure 1.8 shows the trends in efficiency achieved over the past 20 years for all the major PV technologies. These results are for small area "champion cells", the one-of-a-kind result that establishes the potential of a given material system and device technology. The highest efficiency is for the most expensive and complex devices, based on III–V technologies like GaAs and GaInP, consisting of multiple devices with different optical and electrical properties, grown on top of each other. They are discussed in Chapter 9. These multijunction (MJ) cells outperform other cell technologies for three reasons. They are made from very perfect and high-purity crystalline materials, they can capture either a wider range of the solar spectrum or the same range more efficiently than other devices, and they are operated with high concentration factors, using lenses, which increases the efficiency for reasons discussed in Chapter 3. Not surprisingly, they are also extremely expensive.

The workhorse of the PV industry is still Si as discussed in Chapters 5, 6, and 7. Si wafers in the form of either single crystal Si or multicrystalline (also called polycrystalline) Si accounts for 90% of the PV market. Although champion single crystal (c-Si) and multicrystalline Si (multi-Si) cells have been recorded with 25% and 20% efficiency, respectively (both made at UNSW in Australia), the difference in module performance between the two Si-wafer technologies is much smaller. They have typically 14% and 12% efficiency in commercial modules, respectively. Despite three decades of research and manufacturing, clever scientists and engineers are still finding ways to improve the performance of Si-wafer photovoltaics. They are also finding ways of reducing the cost.

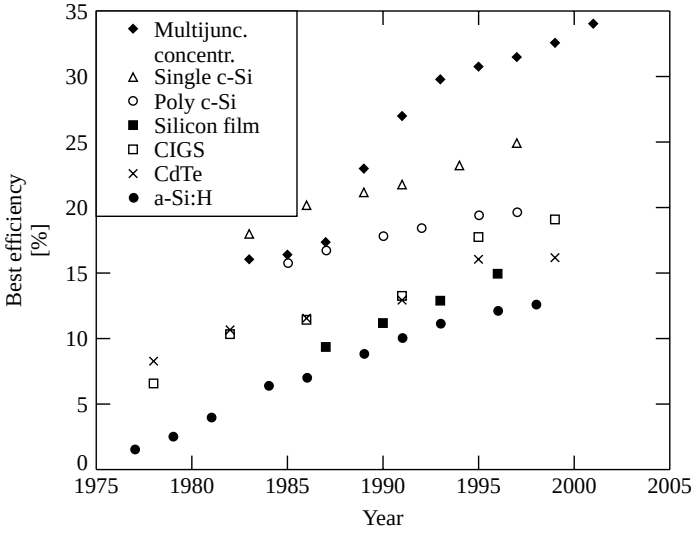


Figure 1.8 Best small area (0.5–5 cm^2) efficiency for various cell technologies measured under standard laboratory test conditions. MJ concentrators are double junctions before 1995, and triple junctions after. a-Si represents stabilized efficiency after extended light soaking and are MJ after 1990 (see Chapter 12)

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 758

Among these attempts are the developments of silicon sheets, which avoid the wafering of bulky Si ingots, a very expensive and wasteful process. Ribbon sheets have already reached a noticeable fraction of the market with two companies: RWE in Germany/USA (leading) and Evergreen Solar in the USA.

Continuing down the efficiency axis we come to the three leading thin-film PV contenders, $Cu(InGa)Se_2$ (CIGS) (Chapter 13), CdTe (Chapter 14), and a-Si (Chapter 12), in that order. The main motivation for interest in thin-film photovoltaics has always been the possibility of lower cost, not higher performance. Their champion cell performance has always been a factor of about 2 lower than Si-wafer technologies until ~2000, when $Cu(InGa)Se_2$ cells with efficiencies of 19% were reported, putting them in potentially close competition with multi-Si, although there are vast differences in the manufacturing experience base between $Cu(InGa)Se_2$ (no base) and multi-Si (large base). The benefits and challenges of thin-film PV technologies are discussed in detail in Section 1.9. Finally, we note that the category of "Si film" solar cells in Figure 1.8 is a special case of solar cells developed entirely by one company (Astropower in the USA), in which they are attempting to achieve the high performance of polycrystalline Si with the low-cost approach of thin films. More information about thin-film Si technology is given in Chapter 8.

The results in Figure 1.8 clearly show that there are many promising technologies in terms of their possibility of achieving rather high efficiency. But the reality is that, as seen in Figure 1.9, almost 90% of the world's PV power modules are either single c-Si or multi c-Si. The evolution shows a trend away from c-Si towards the multicrystalline-Si technology with the market share of Si sheet also increasing. About 10% of worldwide PV sales are a-Si/a-Si or a-Si/a-SiGe multijunctions and the remaining <1% is $Cu(InGa)Se_2$, CdTe, and concentrators. The multijunction concentrators based on GaInP/GaAs cells have yet to find commercial application on Earth but nonconcentrating GaInP/GaAs cells are commonly used for space missions where their high efficiency is more important than their high cost. High performance GaInP-based technology is discussed in Chapter 9 and PV space power is described in Chapter 10.

Who is making all the PV modules? Figure 1.10 shows the breakdown by the three major geographic regions of Europe, Japan, and the United States. Note that this is a logarithmic scale, indicating very steady growth for the past decade of 20 to 30%. Production in 2001 was actually 36% higher than in 2000. The USA has had steady growth and led the world in photovoltaics between 1992 and 1998 when Europe and Japan

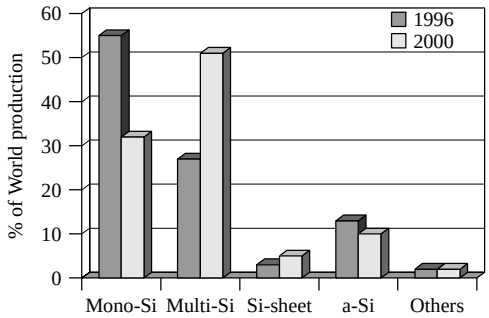


Figure 1.9 Distribution of the PV market by technologies

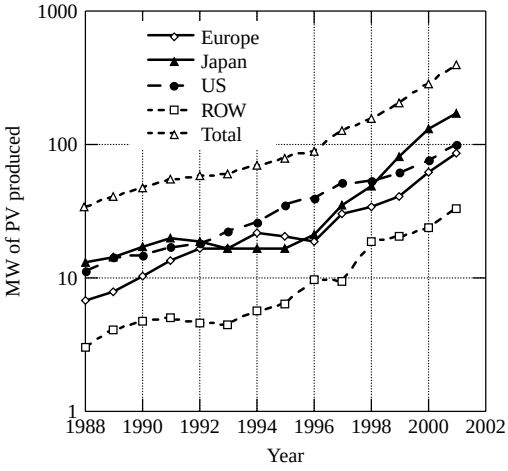


Figure 1.10 World production of PV modules

had static manufacturing growth. Then in 1998, progressive and supportive government policies in many European countries and in Japan resulted in a substantial increase in production. These policies were driven partly by a strong commitment to CO_2 reduction as proscribed by the Kyoto Protocol. European production is dominated by Spain closely followed by Germany, and to a lesser extent France, Italy, and The Netherlands. Note that these figures do not indicate the final residence of the PV module, only its place of birth. For example, a large fraction of PV modules made in the US (∼70%) and Spain (∼80%) are in fact exported.

We present in Figure 1.11 the top ten PV manufacturers in 2001 [55]. The list is headed by the Japanese electronics company, Sharp, followed in the third position

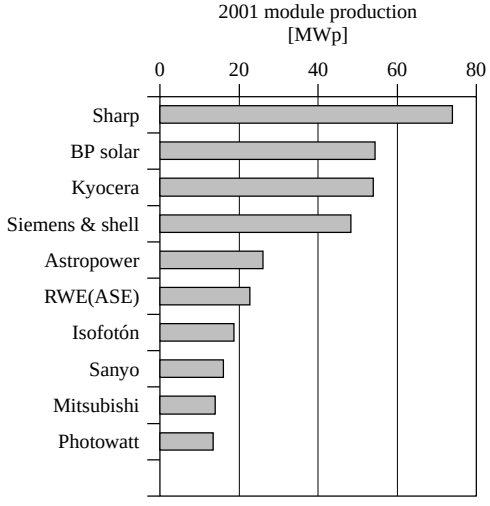


Figure 1.11 The top ten PV cell manufacturers in 2001

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 760

by Kyocera, another Japanese company. The second place is held by the UK-based oil multinational BP Solar with c-Si and multi-Si plants in United States, Spain, Australia, and India. BP Solar also had two thin-film technologies in commercialization in the United States, namely a-Si and CdTe, (see footnote in Section 1.4). The biggest PV manufacturer in the USA today is Shell Solar (formerly Siemens Solar) but is held by European capital. As of this writing, of all the companies in Figure 1.11, only Astropower (US), Isofoton (Spanish), and Photowatt (French but held by Canadian investors) focus their business exclusively on photovoltaics. The others are either divisions or subsidiaries of large companies with diverse manufacturing interests.

Figure 1.3 showed the worldwide trend for various applications. Growth has been driven by distributed, grid-connected PV applications since 1996, mostly in Europe and Japan, as discussed above. There is steady growth in use of photovoltaics for diesel hybrid and communication. These are off-grid applications, typically in remote locations. Previously, operating a diesel generator 24 h a day or replacing large battery packs was the only alternative. PV diesel hybrids can be cost-effective in these cases [56].

Note that Figure 1.3 shows that large-scale, centralized solar power plants are almost nonexistent. These huge "solar electric farms" were envisioned in the early days of photovoltaics to be built in sunny arid deserts, where land was essentially worthless for other uses (note, that photovoltaics operates without water in contrast to conventional thermal power plants). These huge facilities would replace conventional power plants, at least for daytime power. The world's largest centralized PV power plant to date was installed in central California between 1984 and 1985. The operation and performance over several years was reported [57], including operation and maintenance costs [58]. After several years of operation, the installation was disassembled and the modules were sold individually on the market. Presently, the largest centralized PV power plant is 3 MW in Serre, Italy [59]. There have been several other large-scale PV installations, and their installation techniques, labor, and operation costs have been well documented. In particular, the 0.4 MW a-Si plant in California, USA [60] and the 0.48 MW concentration plant [61] in Tenerife, Spain are notable because they do not use the ubiquitous unconcentrated c-Si technology.

1.8 CRYSTALLINE SILICON PROGRESS AND CHALLENGES

Figure 1.9 showed that c-Si, as either single or multicrystalline wafers or ribbons, was responsible for almost 90% of worldwide PV production in 2001. How did its dominance occur?

First, there was a tremendous worldwide scientific and technical infrastructure for Si starting in the 1960s. Huge government and industrial investments were made in programs for understanding the chemical and electronic properties of Si, how to grow it with the required purity and crystalline structure, and to create the equipment needed to perform all the processing steps. The motivation was not just idle scientific curiosity, but rather the competitive drive to manufacture increasingly complex integrated circuit chips, which created first the analog then the digital electronic revolutions leading to our current information, entertainment, and telecommunications industries. The promise of wealth and market dominance led public and private organizations to unlock many

secrets of Si technology. The silicon band gap, of 1.1 eV, is almost optimum to make a good solar converter, as explained in Chapters 3 and 4. The PV industry could utilize the preceding gains for their own application without having to re-create this scientific and technological infrastructure. In addition, Si is one of the most abundant minerals in the Earth's crust. Thus, there was no physical limitation to providing a huge fraction of the Earth's electricity needs with the known Si reserves.

However, for mechanical reasons (it is brittle), silicon requires relatively thick cells, with a typical wafer thickness of about 300 μm. Therefore, some of the electrons pumped by the photons to the conduction band have to travel large distances, on the order of the thickness, to be extracted by the front face through the selective contact to this band (the *pn* junction). Consequently, a good material with high chemical purity and structural perfection is required to fight the natural tendency of the conduction-band electrons to return to the valence band. This loss process is called *recombination*. To avoid this loss, the electrons must be highly mobile, as they are in perfect silicon. Impurities and imperfections must be avoided as they can absorb the extra energy of the conduction-band electrons and convert it into heat, thus eliminating the free electron from traveling through the circuit by immediately restoring it to the valence band energy. Producing heat, which is desirable in solar thermal panels, where this heat is transferred to a fluid, is undesirable in PV modules, where we try to recover the solar energy as electricity, of much higher value.

Metallurgical Grade (MG) Silicon is obtained by reduction of quartz with coke in an arc furnace. Then it is strongly purified by a method developed by and named after the Siemens Company consisting of the fractional distillation of chlorosilanes, which are obtained from the reaction of HCl with Si. Finally, silanes are reduced with hydrogen at high temperatures to produce hyperpure silicon, usually called *Semiconductor Grade* (SG) *Silicon* or just *polysilicon* (it is called *polysilicon* because it has many grains of crystalline Si, typically of about 1 mm).

The polysilicon now has the desired chemical purity (unwanted impurities below the parts per billion (ppb) level, or less than one impurity atom for every 10^{12} Si atoms, for some impurity atoms), but its structural quality is deficient. The structural quality is improved by melting the polysilicon ($>1400^{\circ}$C) and "freezing" or allowing it to solidify very slowly around a rotating crystalline seed, usually by the Czochralski (Cz) method. In this way, a cylindrical single crystalline ingot is obtained of about 25 cm diameter and of 100 cm length. In this step, a very small number of atoms of boron are introduced in the melt to allow the appropriate metallic contacts deposited later to be selective to the valence-band electrons. This forms the *p*-type side of the *pn* junction.

The ingot is now cut in wafers with a saw. For this a very long wire (up to 500 Km) is wound many times on rotating drums cuts with slurry the silicon ingot into wafers. However, the process is slow and about half of the silicon is lost in the sawdust. The challenge here is to cut the wafers thinner so as to make more profit from the silicon. Wafers of 150 μm instead of the standard 300 μm are used in some companies. The techniques and challenges related to crystal growth and sawing are described in detail in Chapter 6.

The wafer is now etched slightly to remove the saw damage and to condition (texture) the surface for better light absorption. Then the conduction-band selective contacts

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 762

are made by introducing atoms of phosphorus to one surface of the wafer, making it the *n*-type partner in the *pn* junction. This is done by locating the wafers in a phosphorus rich atmosphere at high temperature so that these atoms penetrate slightly ($\sim$0.2 μm) into the silicon wafer.

Then metallic grids are printed on the boron and phosphorus-doped zones (some tricks, sometimes proprietary, are used to separate boron- and phosphorus-doped regions properly) and the solar cell is thus finished. The grids make it easier to collect the electricity without resistance losses and are commonly applied with low cost screen-printing methods. But a solar cell is brittle and produces low voltage (about 0.5 V) so that some 36 cells (or multiples of this number) are interconnected with tinned copper ribbons and encapsulated in a sandwich formed of a sheet of tempered glass, an embedding polymer that surrounds the solar cells, and a back sealing plastic layer. The reason for multiples of 36 cells is so that their output voltage $\sim$15 V will be compatible with most DC battery charging applications.

The lowest (publicly offered) module selling prices in 2002 were about \$3/W_p. The breakdown of costs, as given in Chapter 6 are presented in Figure 1.12. The wafer itself represents about 65% of the module cost, approximately equally divided between purification, crystallization, and sawing. This hyperpure silicon is found today as a scrap or waste product from the microelectronics industry at a price of $\sim$\$50/kg. The increase of the PV market in the 1990s has nearly exhausted this market. Additional supplies are coming from the former Soviet Union, whose microelectronic industry has disappeared due to western competition. The Siemens method is considered ultimately too expensive for photovoltaics, but the purity it provides seems necessary for the fabrication of solar cells. Attempts in the 1980s to fabricate a low-cost solar silicon (SolSil) did not succeed due in part to the scarce interest generated by the small markets of the time. Today, new attempts are being made. These might include purification in the MG silicon production steps and in the crystal growth steps avoiding the expensive chlorination procedure and using the molten step of the crystal growth for further purification, or alternately, reducing the chlorosilanes in the molten phase prior to the crystal growth. It is not clear whether they can achieve the needed purity level. Even if feasible it is uncertain whether such wafers would be less expensive than the standard Siemens process or some simplified versions of it. There are fears that a shortage of hyperpure polycrystalline silicon availability might seriously hinder the growth in cell production demanded by the market. Chapter 5

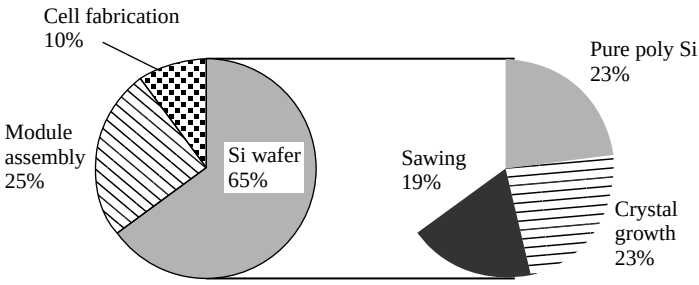


Figure 1.12 Breakdown of costs in the fabrication of a Si-wafer-based PV module. The right side presents the wafer costs

describes in detail the problems and challenges associated with silicon production and purification.

An important advance in solar cell fabrication was the demonstration that solar cells with high efficiency can be fabricated from wafers containing hundreds of large-grain (1–10 mm) multicrystals, called multicrystalline (multi-Si) or polycrystalline (poly-Si), although this later term is less favored because it may cause confusion with the feedstock (polysilicon). The multi-Si growth procedure is much faster and the wafer is cheaper. The loss of efficiency of a few percent (absolute) caused by the random orientation of crystalline grains in a multi-Si wafer compared to a single c-Si wafer is balanced by the lower cost so that the price per watt peak is the same on a module basis. But the simplicity of the multi-Si wafer-growing equipment and process is producing a clear trend towards the use of the multicrystalline option as seen in Figure 1.9.

An interesting option in Si solar cell manufacturing is the growth of ribbons [62, 63]. Ribbons do not require the expensive sawing process. However, the growth of the ribbon crystal is slower because they usually grow in the plane perpendicular to the ribbon surface, with very small area (the ribbon width times the thickness). In contrast, wafers grow in the plane of the wafer surface whose area is the wafer area. The standard ingot solidification process is a very effective purification process due to the preferential segregation of impurities to the molten silicon. However, in ribbon Si the plane of solidification moves faster (although with very small area), so the segregation is less effective. In summary, the ribbon cells are almost as good as the multicrystalline bulk-grown cells and possibly cheaper. Challenges lie in increasing the growth speed and the resulting cell efficiency.

Although the solar cell manufacturing process represents a relatively small fraction of the total cell cost, it strongly affects the overall cost in $\$/W_P$ because it determines the cell and module efficiency. This efficiency depends on the quality of the wafer or ribbon utilized but it also greatly depends on the cell process itself.

As a matter of fact, an efficiency of 25% has been achieved for laboratory cells in a long complex process where every possible efficiency-improving detail has been implemented to produce a complicated but nearly ideal device structure. This is explained in Chapter 7. However, most factories use some variation of the wafer and cell fabrication process described above, including the screen-printing process, that leads to 15% single crystal cells or 13% multicrystal cells. In modules, these efficiencies are reduced to 14% or 12%, mainly due to the redefinition of area that now includes the module frame. This process is considered the best compromise between costs and performance.

The existence of the large efficiency gap between laboratory and commercial cells, together with the increasing markets, suggest that novel, high efficiency commercial cell processes will appear in the next years. Some companies (BP Solar or Sanyo, for instance) are already on this path and have different processes leading to 17 to 18% cells in production. It is worth noting, that the ribbon technology is incompatible (see Chapter 7) with the ubiquitous screen-printing processing, so that a new processing, which may not be so cheap, is required to fabricate cells with this material.

An additional factor very seriously affecting the cell cost is the production yield. In the fabrication of any semiconductor device, not every unit introduced into the production

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 764

line is successfully completed. In good single crystal Si cells, the manufacturing yield is 95%. Many supposedly cheap technologies find their Achilles' heel in the low yield.

Finally, the module fabrication requires interconnecting and encapsulating the cells. These steps also have room for some cost reduction. The use of cheaper materials may help somewhat, as well as better automation, better module interconnection, and integration designs.

1.9 THIN FILM PROGRESS AND CHALLENGES

One might ask "why develop a totally different semiconductor technology for photovoltaics when Si is so well established?". The simplest answer is "to achieve lower cost and improved manufacturability at larger scales than could be envisioned for Si wafer-based modules." In fact, we have already defended our belief that Si technology, very important in the next decades, will not be able to reach the ultimate goals required for mass worldwide penetration of photovoltaics (Section 1.5). What were the disadvantages of c-Si that led to the early investigation and eventual commercialization of alternatives? It was recognized early in the development of photovoltaics that Si crystals were expensive and slow to grow. It was also recognized that of all the viable semiconductors, Si would require the greatest thickness to absorb sunlight, due to its unique optical properties. Si is the most weakly absorbing semiconductor used for solar cells because it has an indirect band gap while most of the other semiconductors have a direct band gap (see Chapter 3 for a more complete explanation of direct and indirect band gaps). Therefore, at least ten times more crystalline Si is needed to absorb a given fraction of sunlight compared to other semiconductors like GaAs, CdTe, $Cu(InGa)Se_2$, and even other forms of Si such as a-Si. Thicker semiconductor material means higher material volume but also a higher quality material because of the longer paths that the high-energy electrons excited by the photons must travel before they are delivered to the external circuit to produce useful work. All this leads, as seen before, to high material cost. In addition, we mention that, presently, much of the Si-PV industry relies on buying scrap material from the electronics industry. As photovoltaics' demand grows, the supply of scrap material might become insufficient (see Chapter 5).

It was recognized almost as early as c-Si PV cells were developed in the 1950s that other semiconductors could make good solar cells. Most of them exist in a form called thin films. When they are fabricated into useful devices, they are so thin that they must be deposited on a foreign material called a substrate for mechanical support like a layer of paint on a piece of wood or the reflective metal coating on glass to form a mirror. A framework for analyzing the material properties, device structures, and manufacturing issues unique to thin-film solar cells (TFSC) was developed [64] since they differ considerably from Si wafers. Throughout the 1970s, progress in Cu_2S/CdS solar cells led to the development of new theories to explain the device performance, new methods of materials processing, and new concepts in semiconductor device manufacturing [65, 66]. Between 1981 and 82, four thin-film technologies demonstrated the ability to cross the magical 10% efficiency barrier, thus becoming candidates for serious consideration: Cu_2S/CdS [67], a-Si [68], $CuInSe_2$/CdS [69], and CdTe/CdS [70]. (It is an inexplicable fact in this business that 10% efficiency seems to suddenly confer respectability and status to any PV technology.) Of these four TFSC technologies, Cu_2S/CdS would soon be

rejected for commercialization due to fundamental and fatal stability problems related to electrochemical decomposition [71]. In contrast, a-Si has a minor stability problem that is not catastrophic and has not been a major barrier to further development and production as discussed in Chapter 12. No fundamental stability problem has been found with $Cu(InGa)Se_2$ and CdTe modules. Consequently, significant industrial and government-sponsored research and development resources have been directed worldwide at TFSC technology. This has led to steady progress in champion cell efficiencies as seen in Figure 1.8.

The main advantage of TFSC is that they will eventually have lower costs than c-Si-wafer PV technology when they are produced in sufficiently large volumes to offset the initial capital investment. The lower costs of TFSC derive from the following characteristics: they are typically 100 times thinner than Si wafers ($\sim$1–3 μm for all the semiconductor layers) deposited onto relatively low-cost substrates such as glass, metal foils, and plastics; they are deposited continuously over large areas at much lower temperature (200 to 500°C vs $\sim$1400°C for c-Si); they can tolerate higher impurities (thus needing less expensive purification of raw materials); and they are easily integrated into a monolithic interconnected module. For a reference, the semiconductors in typical TFSC are 10 times thinner than a human hair. TFSC are either polycrystalline with small $\sim$1 μm sized grains such as $Cu(InGa)Se_2$ or CdTe, or else amorphous like a-Si. This is a consequence of being deposited at temperatures too low to allow perfect crystalline bond formation. TFSCs typically consist of 5 to 10 different layers whose functions include reducing resistance, forming the *pn* junction, reducing reflection losses, and providing a robust layer for contacting and interconnection between cells. Some of the layers are only $\sim$20 atoms thick (10 nm), yet they may be a meter wide! This requires excellent process control. The manufacturing process is designed such that they are deposited sequentially on moving substrates as in a continuous process line. This minimizes handling and facilitates automation, including laser scribing, to isolate and interconnect individual cells on the module, called monolithic integration.

With all the advantages of TFSCs, why does c-Si or multi-Si still dominate 90% of the world market? This brings us to the disadvantages of TFSC: they have lower efficiencies (so far), and they have a much less-developed knowledge and technology base compared to c-Si. Consequently, under-capitalized companies have had to struggle to develop not only an understanding of the materials and devices but also the equipment and processing to manufacture them. The thin-film PV industry has had to develop the technologies all by itself with considerably less financial resources than the Si industry had. They were not able to adopt a mature technology like the Si PV community did from the Si electronics industry.

What are the strengths and remaining challenges for the TFSC industry? We will review the salient characteristics of the three leaders: a-Si, $Cu(InGa)Se_2$/CdS, and CdTe/CdS.

Amorphous Si (Chapter 12) is deposited from hydride gases such as SiH_4 using plasma to decompose the gas. This is called *plasma-enhanced CVD* (PECVD) and allows for large areas to be coated rather uniformly and with excellent control. However, the utilization of gases is only around 10 to 30%, meaning much of the source material is

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 766

wasted. The material has 1 to 10% hydrogen bonded to the Si, and is often designated as *a-Si:H*. The H atoms passivate a large number of the defects resulting from the incomplete bonding of the Si atoms. The atomic structure has no long-range order like all other crystalline or polycrystalline materials. This can be an advantage. Films are typically deposited between 150 to 300°C, the lowest temperature of any of the TFSC materials, allowing the use of lower-cost, low-temperature substrates. a-Si solar cells are deposited on glass, stainless steel foil, or plastic. The last two substrates are flexible allowing for "roll-to-roll" manufacturing where all the layers are deposited as the roll moves through their process zone. The *pn* junction is formed by doping the thin contact layers as they grow with dopant gases containing the boron or phosphorous atoms. All practical a-Si modules contain multiple junction devices where two or three junctions are grown on top of each other. This allows for more efficient utilization of the sunlight. The total thickness, including multiple junctions and all the contact layers, is less than 1 μm excluding the substrate. The highest reported efficiency was 15% for a triple junction, which degraded to about 13% before stabilizing. While a-Si TFSCs cells may have slightly poorer performance compared to other TFSCs when tested under laboratory conditions (Figure 1.8), they have a unique feature that improves their relative performance outside in real conditions; namely, their efficiency is temperature-independent while for all other PV technologies, c-Si or thin-film, the efficiency decreases as the module heats up as in real outdoor conditions. This can result in those other modules losing 2 to 4% (absolute) of their rated output in the summer time and helps a-Si look more favorable. The three major challenges for a-Si technology are: 1) to improve the efficiency from today's 6 to 8% up to 10 to 12%; 2) minimize or eliminate the self-limited degradation which reduces efficiency by 2 to 3% (absolute); and 3) to increase the deposition rate of the layers and the utilization of the gases to allow for faster, lower-cost manufacturing.

Polycrystalline layers of $Cu(InGa)Se_2$ (Chapter 13) alloys have produced the highest efficiency TFSC devices and modules. TFSCs based on $CuInSe_2$ (no Ga) achieved 12 to 15% efficiency but were limited by the low band gap. Alloying with Ga and/or S increases the band gap and increases the efficiency of delivering the electrons to the circuit (as discussed in Section 1.2). While many deposition methods have been explored in the laboratory, there are two different processes under commercial development. Co-evaporation forms the alloy by simultaneous evaporation of the Cu, In, Ga, and Se from sources onto a heated substrate. The other process is called *selenization*, because layers of Cu, In, and Ga are deposited by various means onto a substrate, then heated in the presence of Se from a gas such as H_2Se or a Se vapor, thus contributing the fourth constituent of the alloy. Substrate temperatures typically reach 500 to 600°C during some stage of the growth. Substrates of Mo-coated glass are typically used although metal foils or plastic are being investigated. The $Cu(InGa)Se_2$-films are *p*-type, typically 1 to 3 μm thick and have crystallites or grains on the order of 1 μm. The *pn* junction is formed by depositing an *n*-type layer of CdS, ZnO, or other new materials under development to replace the CdS. The highest reported cell efficiency is presently 19% and several companies have reported modules with >10% efficiency. However, progress has been largely empirical since little fundamental understanding of the materials or devices is available. A very active area of research is developing methods to incorporate other alloys to increase the band gap even further. The three major challenges for $Cu(InGa)Se_2$-related technology are: 1) to control the composition (Ga, S, Se, or Na) of the alloy through the film in a

manufacturing environment on a moving substrate; 2) to find alternative junction partners to replace CdS; and 3) to find new alloys or new deposition methods to give high performance devices with higher band gap alloys.

Polycrystalline layers of CdTe (Chapter 14) have been investigated for photovoltaics since the 1970s. In contrast to limited process options for a-Si or $Cu(InGa)Se_2$, there are over 10 methods to deposit the CdTe films that have produced CdTe solar cells exceeding 10% efficiency. Four have reached precommercialization: spray pyrolysis (SP), electrodeposition (ED), vapor deposition (VD) and close spaced sublimation (CSS). Some take place in liquid baths that are barely warm $\sim 50^{\circ}C$ with CdTe deposition rates of μm/hr (ED) while others take place in vacuum systems at temperatures high enough to soften glass $\sim 600^{\circ}C$ with CdTe deposition rates of μm/min (CSS). There seem to be three critical steps, however, that all efficient CdTe solar cells require. First, they need a post-deposition anneal in the presence of Cl and O_2 at around $400^{\circ}C$. This chemical/thermal treatment enlarges the grains, passivates the grain boundaries, and improves the electronic quality of the CdTe. Second, all CdTe layers need a surface treatment before applying a contact. This treatment can be a wet or dry process and prepares the CdTe surface by etching away unwanted oxides and leaving a Te-rich layer needed to make a low-resistance contact. Third, all high efficiency devices have a Cu-containing material somewhere in their CdTe contact process but again, there are many ways this can be achieved. Whichever process is used to deposit the CdTe, it has been found that the entire device process is highly coupled since processing steps strongly influence previous layers. This is partially due to the CdTe grain boundaries which act like paths for interdiffusion.

The *pn* junction is formed by first depositing an *n*-type layer of CdS on a transparent conductive oxide substrate followed by the CdTe layer and appropriate chemical annealing. Once the solar cell is made, the CdTe films are slightly *p*-type, typically, 2 to 8 μm thick and have crystallites or grains on the order of 1 μm. The highest reported efficiency for a CdTe/CdS device is presently 16%. Some CdTe modules have been in outdoor field-testing for over five years with negligible degradation, yet other CdTe devices degrade during accelerated life testing indoors. Of the three leading TFSC technologies, CdTe may have the most challenges. The dual role of Cu must be resolved; it seems to be required to produce a high-efficiency device but it is also implicated in long-term stability problems. The various optimizing treatments need to be better understood so they can be simplified and transferred into production. CdTe modules may be more sensitive to atmospheric interaction (O_2, H_2O) requiring better encapsulation methods. Finally, safe and cost-effective Cd usage in the workplace followed by recycling at the end of the module's life need to be determined.

In fact, technically astute investors know that other factors can be more important than efficiency in selecting a technology for development. This point is made loud and clear by examining the relative performance of the three major TFSC technologies – $Cu(InGa)Se_2$, CdTe, and a-Si – in Figure 1.8. Note that a-Si has always had the lowest efficiency. Yet, of the three, it was a-Si that has been commercialized much earlier and more widely. It enjoys by far a much greater manufacturing capacity. As of 2001 [72], a-Si accounted for almost 9% (34 MW) of the world PV power module production, while CdTe and $Cu(InGa)Se_2$ power modules together accounted for less than 0.3% (1 MW), despite a-Si champion-cell-stabilized efficiency lagging the others by several percent. In

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 768

reality, typical production modules of a-Si and CdTe are in the 7 to 8% range while Cu(InGa)Se_2 is 9 to 10%.

Why has so much capital been invested to develop a-Si technology over CdTe and Cu(InGa)Se_2? A major factor was that a-Si had a stronger worldwide scientific research base, which ensured that the relation between deposition conditions and fundamental material and device properties were well characterized. In contrast, CdTe and Cu(InGa)Se_2 are "orphans" because they have no real application outside of photovoltaics. Therefore, the entire knowledge base and technical infrastructure had to be developed mostly by underfunded industrial groups and a small number of university and government research labs. This shows that translating research-grade champion cell performance into production modules coming off-the-line day after day is a very challenging task. Figure 13.23 in Chapter 13 shows the time delay in translating efficiency achieved in the lab for a small area cell to a prototype large area module for the Cu(InGa)Se_2 technology. Delays of six to eight years are typical. To conclude, the highest efficiency technology is not always going to be the best choice for a low-cost, high-yield process, at least not until much of the technical background is in place.

Recognizing that the ideal PV technology would have some of the merits of c-Si but be deposited as a thin film a few microns thick, several groups have elected to try to achieve the "best of both worlds" by developing thin films of multi-Si deposited on a substrate. This is the subject of Chapter 8. At present, the best thin-film multi-Si solar cells have the same efficiency as their $CuInGaSe_2$, CdTe, or a-Si based predecessors. This is partly because multi-Si thin-film photovoltaics also inherits some of the problems of both c-Si and thin films. In particular, passivation of grain boundaries and surfaces seems to be a major problem, yet many of the well-established passivation methods from c-Si are not applicable to multi-Si thin films.

But there are new thin-film technologies such as the dye-sensitized solar cell that operate on a very different principle than an all-solid-state solar cell, almost more like photosynthesis than photovoltaics. This fascinating new solid–liquid technology, not free of challenges either, is described in Chapter 15.

1.10 CONCENTRATION PV SYSTEMS

If solar cells are expensive, focusing concentrated sunlight onto fewer solar cells was considered from the earliest times as a way of reducing costs. For example, instead of a typical 100 cm^2 solar cell absorbing 100 cm^2 worth of sunlight, one could focus the sunlight from 100 cm^2 onto a 1 cm^2 of solar cell, thus reducing the solar cell cost by 99% while still utilizing the same amount of sunlight. This neglects the cost of the optical focusing elements and other special equipment needed for concentrator technology. Of course, two conditions were required. One is that the optical surface collecting the light and redirecting it to the cells had to be cheaper than the cell area it replaced and the second is that the efficiency of the cells under concentrated sunlight should not decrease substantially. The first condition is generally fulfilled if we consider only the optical surfaces. However, keeping the cells constantly in focus requires a moving structure to keep the optics pointing at the sun as it moves across the sky, which adds cost to the system. As for the cell efficiency, there are fundamental reasons (see Chapter 3) why

it should increase with increasing luminous flux. However, in practice, ohmic resistance losses caused by the handling of large currents limits the efficiency increase. Thus, cells for concentrator applications must be carefully designed to minimize such losses and therefore they become more expensive. Yet, the small area of cells used in a concentration system, or alternatively, the large amount of electric power produced by each cell, allows for paying higher costs for the cells, and therefore allows them to incorporate many refinements in order to make such cells very efficient.

One factor that reduces the system efficiency is the loss associated with the optical hardware used to concentrate the light. Additionally, only the direct sunbeam is collected since scattered (diffuse) light is not focused. This again reduces the electric output by at least 15%. However, this last reduction is compensated for in many sunny locations by the fact that the tracking system always aims the cell directly at the sun (at least for two-axis tracking systems). In contrast, with the more typical stationary modules, the output power varies like the cosine of the angle of the sun, so this is very low in morning and late afternoon, when the sun is at an oblique shallow angle relative to the module (equations describing the sun's motion and the relative illumination on a module versus time of day or time of year is given in Chapter 20). Accounting for these gains and losses, it is generally found that concentrator efficiency today tends to be somewhat higher than flat module efficiency and this tendency will increase in the future with the adoption of higher efficiency cells. It is also believed that concentrators should ultimately be cheaper than flat module silicon solar cells. However, this statement has not been confirmed in practice due to the lack of a real market, apart from a few purchases for demonstration purposes.

Actually, concentrators are appropriate for relatively large installations while the PV market has evolved so far in smaller installations such as grid-connected houses, remote homes, or telecommunication applications whose size is seldom bigger than 5 kW. Therefore, only very few companies fabricate concentrator cells today, and their prices are very high because they have large general costs for very small production volumes. However, the situation may change. The general increase of the PV markets will probably stimulate the appearance of niches better adapted to concentrators.

The difficulty of developing concentrators must not be underestimated. Combining the requirement for high performance with the low cost is a formidable challenge. In particular, the optics must be low cost, yet permit highly accurate focussing, high optical efficiency, and equal illumination in all cells. The tracking structure must be cheap and accurate, cells must be efficient and not too expensive, and finally the cooling and current extraction must be effective and cheap. Chapter 11 deals in detail with concentrator issues.

1.11 BALANCE OF SYSTEMS

A photovoltaic system consists of more than PV modules composed of solar cells. In addition, it requires elements that are generically known together as "*Balance Of System*" (BOS). The BOS is, typically, composed of the battery, the control unit and the inverter, the mechanical support structure, the electric cabling, and protection devices such as fuses, grounding rods, and disconnect switches. We present in Figure 1.13 the cost of a PV stand-alone installation with storage, as presented in Chapter 2.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 770

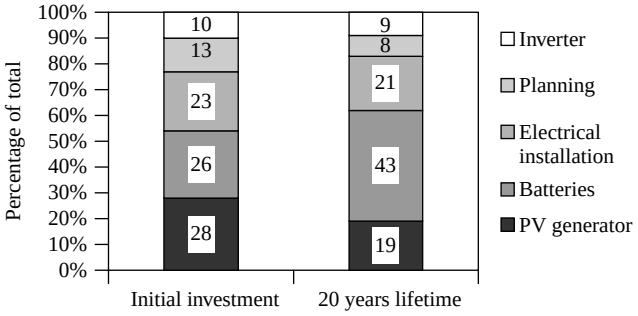


Figure 1.13 Breakdown of costs of a stand-alone PV installation with an optimum size storage. Differences in the initial investment and the life cycle investment are due to replacement of batteries during the 20 year life

We can see that in this case the PV modules represent approximately a quarter of the cost, while the cost of the batteries exceed that of the PV modules, especially when their periodic replacement over 20 years is included. This makes a very important and little appreciated point: even if we could make PV modules for free, the life cycle cost of the stand-alone system in Figure 1.13 would only be reduced by 25%! Clearly, batteries must be examined more closely. It is fortunate that many important applications do not require battery storage and therefore are free from this major cost burden.

Batteries are in most cases the lead-acid type. While automobile batteries are optimized for providing strong current for a short period to start the car, the ideal batteries for PV systems are so-called “deep cycle” batteries that can yield a large fraction of their charge (deep discharge cycle) and must operate with high efficiency and long duration. Yet, many PV applications use standard “shallow discharge” auto batteries due to their ubiquitous availability and lower initial cost (due to massive worldwide markets and applications) to the detriment of the long-term PV system cost. Some of the modern batteries such as those based on lithium ion or lithium polymer used in laptop computers or mobile phones could be used in solar applications but they are too expensive and are not significantly better than properly managed lead-acid batteries. This is studied in detail in Chapter 18. As said before, the relatively high cost of the batteries is further increased when we consider the costs over the PV system’s life cycle since the batteries have to be replaced every four to eight years due to their relatively short lifetime. Therefore, good maintenance procedures to increase the battery lifetime are important but not always applied. Alternative methods of storage exist but they will not replace lead-acid battery, at least, not in the next ten years.

The battery charge controller is essential for the long life of the battery. It is an electronic device that prevents overcharging and excessive discharging, both of which can dramatically shorten the battery’s life. In large systems, equilibration of the battery charging (so that all battery cells charge equally) should also be incorporated. In hybrid systems, which combine photovoltaics and a diesel or wind generator, the control unit must connect and disconnect the different generators according to a plan. Also, loads can be categorized, so that in case of low battery charge and low PV output, some loads can be disconnected while some essential ones are maintained active.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 771

In many cases, battery storage is not necessary. For PV-powered water pumping, the water is pumped and stored while the sun is shining. For the grid-connected homes (one the fastest growing applications) and office buildings, the photovoltaics produces energy during the day and the grid supplies the energy during the night or on cloudy days, thus eliminating the need for batteries, simplifying system design, and reducing BOS costs. In PV-powered generating plants, the electricity is produced while the sun shines. The success of grid-connected PV applications is very sensitive to price competition with conventional electricity (even if the PV electricity is subsidized, as it occurs in Europe and Japan). Luckily the expensive batteries do not hamper this competition.

PV modules produce direct current (DC) which is suitable for directly charging batteries or powering a small number of special products. However, most appliances run on alternating current (AC). Consequently, an inverter must be used to convert the DC into AC. Inverters are widely used for many industrial applications. They are fabricated in large quantities as uninterruptible power supplies (UPS), to convert the DC electricity stored in batteries into AC electricity in the case of grid failure. They are used in hospitals and other installations where electricity failure is not tolerable. Small UPSs are often used in computers that must operate continuously. The PV inverter has an additional and important role: to vary the electrical operating point of the PV array to maintain its output at the maximum value, that is, the variable bias point at which the PV array produces highest power extraction. Changes of temperature and insolation change the voltage where maximum power extraction occurs. The electronics of the inverter typically include maximum power tracking. The inverters used in photovoltaics and the rest of the power conditioning electronics are explained in Chapter 19. Inverters have often been the source of poor reliability in early systems. Feedback to manufacturers and more robust components has greatly reduced these problems.

Because of the cost, the electronics for power conditioning is sometimes considered as a serious hindrance in the development of photovoltaics. However, we think the appearance of a significant market will reduce costs to reasonable limits. One fresh approach is to have a small inverter on the back of every module instead of a centralized inverter for the entire system. This modular approach has many advantages and is being put into production in USA [73] and Germany [74], but at present it is more expensive and will be so unless mass production reduces costs with respect to the theoretically cheaper bigger inverters.

The system-mounting structure is also important, in particular, in concentrating systems. In fact, this is the second most important cost element in concentrating PV systems, after the modules. In contrast, the power conditioning cost is comparable to many other small costs associated with the plant construction. This can be seen in Figure 1.14, where we present the breakdown of costs for a TFSC ($CuInSe_2$, an ancestor of today's $Cu(InGa)Se_2$) flat module central plant and a concentrating central plant. Both sets of calculations are based on the same criteria for the two technologies involved (but based on realistic but futuristic assumptions) and presented in Chapter 21.

The plant cost is about twice the module cost: more for concentrators, less for "flat-plate" modules. Notice, that the cost of land is absolutely negligible land (provided the plant is built where land prices are low). It has to be stressed again that modules at the costs in the figure are not yet achievable.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 772

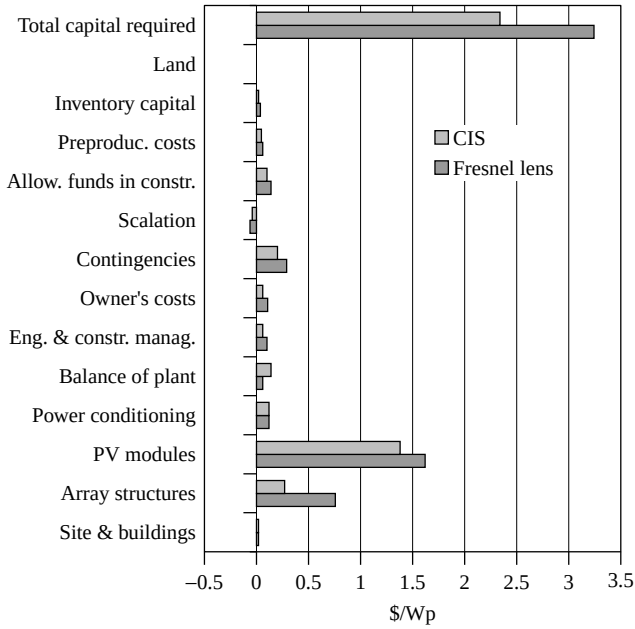


Figure 1.14 Calculated breakdown of costs of the construction of a PV central plant based on two anticipated technologies. Above, CIS flat module cells; below, Fresnel lens concentrator with Si cells operating at 500 suns. (Data from Whisnant *et al*. in Chapter 21)

We present in Figure 1.15 a calculation of costs of different types of modules in an advanced manufacturing process calculated with the same procedure. Again, the origin of this is Chapter 21, devoted to cost calculations. Although cost calculations for the $Cu(InGa)Se_2$ technology is not available, their costs are not expected to be too different to those of a-Si. Indeed, the cost model in Chapter 21 includes many more parameters and assumptions and the results depend on such choices. Nevertheless, according to the model, the common c-Si Cz module (with efficiency of about 14 to 15%) is clearly more expensive than any one of the other options (provided they have a minimal efficiency of 5 to 7%). Dendritic web technology refers to one form of Si ribbon technology that is starting to find a place in the market.

Using again the cost model in Chapter 21 we can calculate the cost of the electricity generated. Table 1.4 shows the difference of electric output of the two equally rated plants.[2] Obviously, the yearly efficiency (referred to as the *module aperture area*) is different. Furthermore, they both present different capacity factors as defined by the ratio of the produced energy (in kWh or MWh) to the rated power (in kWp or MWp) times the total number of hours in a year (365×24). To calculate the electricity cost (in \$/kWh), one term is the financial cost of all the expenditures and another much smaller one is the operation and maintenance costs (very small as stated at the beginning). The price per kWh must be compared to the typical average price of \$0.05/kWh for conventional

[2] Figure 1.15 and Table 1.4 present cost data scaled to 1985 or 1990 equivalent dollars, as does discussion and results in Chapter 21. Although quantitatively "out of date", they are still quite useful for relative comparison.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 773

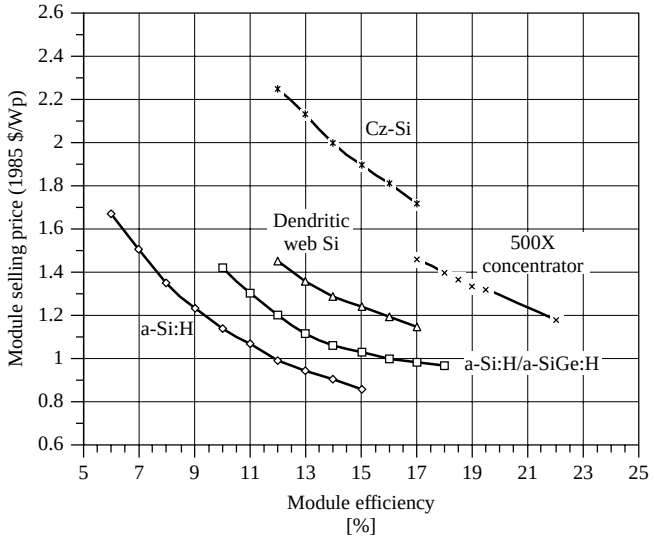


Figure 1.15 Effect of cell technology and efficiency on module price. (Source: Whisnant *et al*. in Chapter 21)

Table 1.4 Annual performance and energy cost summary for central station plants. Calculations for constant 1990$, for a 50 Mw_p plant in Central California, USA. Source Whisnant *et al*., Chapter 21

	Fresnel lens concentrator	CIS Flat-plate
Energy output (MWh)	140 100	112 000
Capacity factor	32.0%	25.8%
Annual energy efficiency	18.8%	9.9%
Annual expenses (10^6)		
Capital charge	16.69	11.95
Operation & maintenance expense	0.61	0.18
Total	17.30	12.13
30-YR Levelized energy cost ($/kWh)		
Capital charge	0.119	0.106
Operation & maintenance costs	0.004	0.002
Total ($/kWh)	0.123	0.106

electricity. Photovoltaics is more than twice as expensive. However, this cost can be attractive for a part of the electricity generated by a utility, such as when it meets peak power demand.

For the end-user, the cost of the conventional electricity can be very similar to the one in the table. The cost of installing a PV grid-connected system in a house or building is not much more expensive than the centralized power plant presented in the table but the marketing cost of this distributed residential product will increase the final installation cost. It is to be stressed, that the costs used here are merely an indication

of the relative significance of various factors. These assumptions and model parameters may not apply to some immature or as-yet undeveloped technologies. Yet, these exercises can be useful to decision makers, especially if complemented by sensitivity studies (as in Chapter 21).

As a final comment, the costs presented in this section are final costs calculated by a detailed account of all the elements utilized in a plant or module and an estimation of the performances achieved. In contrast, the costs in Section 1.5 are based on fundamental economic laws that are found empirically to occur regardless of technological details, yielding costs as a function of the time. Thus, the costs, as calculated in the present section, may be considered as a point of that curve, to occur somewhere in the future.

Of course, all components in the system have some parasitic power loss. Yearly average AC efficiencies of 10% are common in a well-managed grid-connected system with modules starting at 14%. Much of this loss can be attributed to temperature: cells operate at 20 to 30°C over the ambient which reduces their efficiency and output since efficiency decreases with temperature (except for a-Si). Additionally, 8 to 10% relative losses are expected for the DC–AC conversion in the inverter. Round trip battery losses (charge–discharge) in stand-alone applications can be an additional relative 10 to 20%.

1.12 FUTURE OF EMERGING PV TECHNOLOGIES

The solar resource is huge although its energy density is rather low. However, it is not so low as to lose any hope of massive utilization but it is not high enough to make it easy.

Obviously, the proper strategy for recovering a dispersed resource is to do it with high efficiency at a low cost per area. But the standard PV-effect, as described in this chapter, only delivers to the external circuit with high efficiency those charge carriers generated by the few photons with energy close to the band gap. The excess energy of photons whose energy is greater than the band gap is typically wasted as heat. Even worse, all of the energy of the photons whose energy is below the band gap is wasted since they are not absorbed and therefore generate no charge carriers.

Thus, as described in detail in Chapter 4, the maximum efficiency that can be obtained under the best conditions from a single junction solar cell is in the range of 40%. The best efficiency so far obtained for single-junction solar cells is 27.6%, with GaAs research-type cells [75] under concentrated sunlight of 255 suns, that is, of 255 times the unconcentrated standard power density (i.e. at 255 kW/m^2). Typical commercial silicon cell efficiency is ~15% measured at standard conditions (input optical power density of 1 kW/m^2, 25°C and standard terrestrial solar spectrum).

One way of extracting more power from the sun is to use stacks of cells of semiconductors having different band gaps. Higher band gap semiconductors are located on top of the stack allowing photons of energy less than their band gap to pass through, where they can be absorbed by inner cells of lower band gap. The limit efficiency of these stacks, as presented in Chapter 4, with infinite number of cells of different band gap is 86%, as compared with the 40% of the single band gap cells. Efficiencies up to 32% (under standard unconcentrated terrestrial solar spectrum) have been achieved for a monolithic three band gap stacked cells of GaInP/GaAs/Ge [76].

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 775

The interest in multijunction cells has been reawakened by the requirements of space cells, where price is less relevant than the performance in many cases (Chapter 10). However, they can be used in terrestrial applications provided they are operating at very high concentration. There is a trend to develop cells operating at 1000 suns. Efficiencies up to 26% with a single band gap GaAs solar cell [77] and of over 29% with a double band gap GaInP/GaAs cells [78] have been achieved (Chapter 9). Also, the development of low-cost concentrators able to operate at 1000 suns is a subject of current research [79]. The prospects are very promising because such technologies predict in the long-term to produce electricity competitive with conventional sources. A cost estimate is presented in Table 1.5. In the 1-J no learning case, the costs are similar to those in Table 1.4. However, in the learning case the costs are, in extremely good locations (EGL), very competitive with conventional electricity, provided that we achieve very high efficiencies. In this calculation, the experience factor for the cells is the same as in microelectronics; for the rest of the elements the learning curve is same as the present one for modules [80].

The role of the experience factor has been stressed when describing Figures 1.6 and 1.7. Conventional cells have a relatively low experience factor, we think, because they are limited in the maximum efficiency they can reach. Multijunction cells, in contrast, have a much higher efficiency limit and therefore they can progress in efficiency for a longer time. This is one reason to attribute to them a faster experience factor.

Multijunction cells are also crucial to the success of the thin-film photovoltaics. In the a-Si thin-film PV technology, the highest cell and module efficiencies being reported for the past decade are for triple junctions as described in Chapter 12. The use of multijunction cells in thin films might lead to a faster learning curve and hence reduced costs. The band gap of various polycrystalline alloys of $Cu(InGa)Se_2$, $Cu(InAl)Se_2$, Cu(InGa)(SeS), or CdZnTe can be varied with alloying. The theoretical efficiency of a tandem device with a 1.6 to 1.8 eV band gap top cell and a 1.0 to 1.2 eV band gap bottom cell exceeds 30%. In all cases, the top cell must provide the majority of the power.

Table 1.5 Costs for very high-efficiency 1000 suns-concentrating systems [80] for one junction (1-J) and four junctions (4-J) cells. NDI stands for Normal Direct Irradiation. EGL stands for extremely good location with NDI = 2700 $W \cdot m^{-2} \cdot year^{-1}$. "No learning" means with present costs (2002) while "learning" means they are reduced by a learning curve with experience factor of 0.32

Cost element	1-J, no learning	4-J, learning
Cells ($ per cm^2 cell area)	13.4	4.43
Module ($ per aperture area)	265	113
Cell efficiency (%)	23.1	45
Module efficiency (%)	19.0	37.1
Plant price ($ per m^2 aperture area)	526	271
Madrid NDI ($W \cdot m^{-2} \cdot year^{-1}$)	1826	1826
Performance ratio	0.606	0.606
Electricity costs in Madrid ($ per kWh)	0.186	0.050
Electricity costs in EGL ($ per kWh)	0.131	0.035

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 776

Yet, in all these alloy systems, material quality and device performance degrades substantially for band gaps exceeding 1.4 to 1.6 eV, depending on the materials (see Chapter 13). A very productive area for research will be to either improve these materials when they are produced with high band gaps or to develop new alloys. Any multijunction process must be sequentially compatible from beginning to end.

Besides the multijunction effort, other concepts are in place for attempting a better utilization of the solar spectrum, such as the Intermediate Band (IB) concept [81], which is described further in Chapter 4. A band, with electron states in the center of the band gap, would permit the passage of electrons from the valence band to the conduction band by means of two low-energy photons, one pumping electrons from the valence to the intermediate band and the other pumping the electron from this band to the conduction band. This concept presents a potential behavior which is somewhat better than the multijunction cell stack with two semiconductors. Nanotechnology is a means of producing this intermediate band [82], and the basic effect sketched above seems to have been proved already by using quantum dots [83]. Quantum dots are droplets of one material in a host of another material of higher band gap. The droplets are very small and exhibit quantum effects among which is the appearance of intermediate bands or levels. Alloys might also be found with intermediate band [84], but no practical realization of this concept has been shown so far. Other new type devices, although not easy to realize today, are also envisioned. The theoretical basis of all these new devices can be found in Chapter 4.

Finally, solar cells have been constructed which do not operate on the photovoltaic effect but on charge transfer between molecular orbitals, as in photosynthesis. Their potential for efficient absorption of the spectrum depends on "tuning" the chemistry of organic dyes, as described in Chapter 15. Presently the subject of much research, dye-sensitized solar cells have achieved $>10\%$ efficiency but have many challenges regarding stability and manufacturability.

1.13 CONCLUSIONS

Photovoltaics constitutes a new form of producing electric energy that is environmentally clean and very modular. In stand-alone installations, it must use storage or another type of generator to provide electricity when the sun is not shining. In grid-connected installations storage is not necessary: in the absence of sunlight, electricity is provided from other (conventional) sources.

PV electricity is highly appreciated by the public. It is unique for many applications of high social value such as providing electricity to people who lack it in remote areas. Often, international donor agencies are providing the funding, as many of the users are very poor. Photovoltaics is very suitable as the power supply for remote communication equipment. Its use is increasing rapidly to produce electricity in grid-connected houses and buildings in industrialized countries, despite a 5 to 10 times higher cost than conventional electricity. Often, publicly funded programs are required to enable photovoltaics to compete by partially offsetting its higher costs.

Largely, because of grid-connected PV applications such as homes and businesses, the expansion of the PV market has been very rapid in the last years of the twentieth

century and it is expected to continue during the next few years of the twenty-first century. Then, the growth will probably continue at a slower pace unless new technological advances are developed and commercialized. In that case, growth could accelerate.

Photovoltaics is poised to become a large global high-tech industry, manufacturing and selling modules in nearly every country. Governments and entrepreneurs should be aware of this. Public R&D support has always been generous. It must continue to be so for those countries that want to maintain leadership in this technology. Partial subsidization of PV installations is permitting an unprecedented development of the PV industry and will also help the industry of the countries involved in this endeavor to take the lead.

Crystalline Si technology, both monocrystalline and multicrystalline is today clearly dominant, with about 90% of the market. It will remain dominant at least for the next ten years. The trend is towards the multicrystalline option. Si is one of the most abundant elements in the Earth's crust but the purified Si used in today's solar cells is obtained primarily as off-grade poly-Si and scrap wafers from the microelectronic industry. Soon it will not be enough for the growing PV industry. Thus, some concerns exist regarding a shortage of the purified silicon for the PV industry. However, it is doubtful that Si technology will be able to reach competition with conventional electricity. Consequently, low-cost alternatives and high-efficiency novel concepts, many already in development, are needed.

Thin-film technology is one of the candidates to take over from Si technology in the long-term. There are many technological options regarding thin-film materials and methods of deposition but their primary claim to the throne currently occupied by Si is that they can be ultimately produced at much lower cost.

Concentration of sunlight is another candidate for mass penetration of photovoltaics, although it will not be easily accepted for the grid-connected houses, one of the most promising applications today. Concentrators will probably find incipient niche markets in big stand-alone applications or as small, central-power plants during the present decade.

Finally, new materials and device designs based on III-V semiconductor alloys, such as GaInP allowing more efficient use of the solar spectrum are now being developed for space applications. With the use of concentrators, they may be of interest for terrestrial applications, with the potential of reaching competitive costs with conventional electricity. Other options, such as quantum dots and dye-sensitized solar cells, are still in the initial research phase. They must compete not only with the ubiquitous Si but also with the other options, mentioned above for funding for further development.

Thus, photovoltaics possesses a panoply of novel technologies that almost ensures that alternatives will be available when the current Si wafer technology cannot reach prices low enough to compete with conventional electricity. If this happens, a strong industry and infrastructure – first developed for the Si technology – will be able to seamlessly take this new PV technology and apply it worldwide. In any case, the present "subsidies" to research or application must be considered as public investment in a policy with strong public support and long-term human benefits.

The widespread contribution of PV electricity in electric grids will require a new type of grid management that can accept small generators as well as small consumers. On

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 778

the other hand, hybrid forms of electricity generation, including photovoltaics, will play an important role in the future development of stand-alone applications and minigrids. The general cost reduction will make photovoltaics available to more and more people in developing countries helping their development with little degradation of air quality associated with fossil-fuel generators.

In summary, it is very likely that photovoltaics will become in the next half century an important source of world electricity. Public support and global environmental concerns will keep photovoltaics viable, visible, and vigorous both in new technical developments and user applications. Nations which encourage photovoltaics will be leaders in this shining new technology, leading the way to a cleaner, more equitable twenty-first century, while those that ignore or suppress photovoltaics will be left behind in the green, economic energy revolution.

REFERENCES

1. Benka S, *Phys. Today* 38, 39 (2002); adapted from Pasternak A, Lawrence Livermore Natl. Lab report UCRL_ID_140773 (October 2000).
2. We acknowledge Dr. Larry Kazmerski of NREL, and coauthor of Chap. 24, for developing the "myths of PV" concept as a way of dispelling common misunderstandings and issues in PV. See his paper Renewable Energy World, August 2002, 175–183.
3. "*Energy System Emissions and Material Requirements*", Meridian Corporation (Alexandria VA) report prepared for the Deputy Assistant Secretary for Renewable Energy of the USA (1989).
4. Kelly H, Weinberg C, in Johansson T *et al*. (Eds), *Renewable Energy*, 1011–1069, Island Press, Washington, DC (1993).
5. Maycock P, *Renewable Energy World* **3**, 59–74 (2000).
6. Fthenakis V, Moskowitz P, *Prog. Photovolt.* **8**, 27–38 (2000).
7. Tsou Y *et al*., *Proc. 2nd World Conf. Photovoltaic Energy Conversion*, 1199–1204 (1998).
8. See Papers contained in Proc. of Photovoltaic Safety Conf., *Sol. Cells* **19**, 189–397 (1987).
9. See Papers contained in Proc. of Workshop on Environmental Aspects of PV Systems, *Prog. Photovolt.* **6**, 87–146 (1998).
10. *www.pv.bnl.gov*.
11. Fthenakis V, Moskowitz P, *Prog. Photovolt.* **3**, 295–306 (1995).
12. Patterson M, Turner A, Sadeghi M, Marshall R, *Proc. 12th Euro Photovoltaic Solar Energy Conf.*, 951–953 (1994).
13. Fthenakis V *et al*., *Prog. Photovolt.* **7**, 489–497 (1999).
14. Bohland J, Smigielski K, *Proc. 28th IEEE Photovoltaic Specialist Conf.*, 575–578 (2000).
15. Moskowitz P, Fthenakis V, *Sol. Cells* **31**, 513–525 (1993).
16. Moscowitz P, Fthenakis V, *Sol. Cells* **29**, 63–71 (1990).
17. Fthanakis V, Eberspacher C, Moskowitz P, *Prog. Photovolt.* **4**, 447–456 (1996).
18. Bohland J, Dapkus T, Kamm K, *Proc. 2nd IEEE World Photovoltaic Specialists Conf.*, 716–719 (1998).
19. Alsema E, *Prog. Photovolt.* **8**, 17–25 (2000).
20. Keoleian G, Lewis G, *Prog. Photovolt.* **5**, 287–300 (1997).
21. Knapp K, Jester T, *16th Euro Photovoltaic Solar Energy Conf.*, 2053–2056 (2000).
22. Fritts C, *Proc. Am. Assoc. Adv. Sci.* **33**, 97 (1883).
23. Chapin D, Fuller C, Pearson G, *J. Appl. Phys.* **25**, 676, 677 (1954).
24. Reynolds D, Leies G, Antes L, Marburger R, *Phys. Rev.* **96**, 533, 534 (1954).
25. Jenny D, Loferski J, Rappaport P, *Phys. Rev.* **101**, 1208, 1209 (1956).
26. Prince M, *J. Appl. Phys.* **26**, 534–540 (1955).

27. Loferski J, *J. Appl. Phys.* **27**, 777–784 (1956).
28. Wysocki J, Rappaport P, *J. Appl. Phys.* **31**, 571–578 (1960).
29. Shockley W, Queisser H, *J. Appl. Phys.* **32**, 510–519 (1961).
30. Cusano D, *Solid State Electron.* **6**, 217–232 (1963).
31. Wysocki J *et al*., *Appl. Phys. Lett.* **9**, 44–46 (1966).
32. Alferov ZhI, *Fiz. Tekh. Poluprovodn.* **4**, 2378 (1970).
33. Lindmayer J, Allsion J, *COMSAT Tech. Rev.* **3**, 1–22 (1973).
34. Hovel H, Woodall J, *Proc. 10th IEEE Photovoltaic Specialist Conf.*, 25–30 (1973).
35. Sumner D, Whitaker C, Schlueter L, *Proc. 20th IEEE Photovoltaic Specialist Conf.*, 1289–1292 (1988).
36. Green M *et al*., *Proc. 18th IEEE Photovoltaic Specialist Conf.*, 39–42 (1985).
37. Sinton R, Kwark Y, Gruenbaum P, Swanson R, *Proc. 18th IEEE Photovoltaic Specialist Conf.*, 61–65 (1985).
38. Friedman D *et al*., *Prog. Photovolt.* **3**, 47 (1995).
39. Gratzel M, *Prog. Photovolt.* **8**, 171–186 (2000).
40. Contreras M *et al*., *Prog. Photovolt.* **7**, 311–316 (1999).
41. Bower W, *Prog. Photovolt.* **8**, 113–126 (2000).
42. Fitzgerald M, Mrohs M, *Proc. 26th IEEE Photovoltaic Specialist Conf.*, 1225–1229 (1997).
43. Nieuwenhout F *et al*., *Prog. Photovolt.* **9**, 455–474 (2001).
44. Cabraal A, Cosgrove-Davies M, Schaeffer L, *Proc. 25th IEEE Photovoltaic Specialist Conf.*, 1357–1362 (1996).
45. Hoff T, Jennings C, *Proc. 18th IEEE Photovoltaic Specialist Conf.*, 235–240 (1985).
46. Shugar D, Hoff T, *Prog. Photovolt.* **1**, 233–250 (1993).
47. Byrne J, Hegedus S, Wang Y, *Prog. Photovolt.* **2**, 235–248 (1994).
48. Yordi B, Gillett W, *Prog. Photovolt.* **5**, 175–185 (1997).
49. Iuchi M, *2nd World Conference and Exhibition on Photovoltaic Solar Energy Conversion*, 3304–3307 (1998).
50. Strong S, *Proc. 25th IEEE Photovoltaic Specialist Conf.*, 1197–1202 (1996).
51. Cody G, Tiedje T, *Proc. 25th IEEE Photovoltaic Specialist Conf.*, 1521–1524 (1996).
52. Luque A, *Prog. Photovolt.* **9**, 303–312 (2001).
53. Parente V, Goldemberg J, Zilles R, *Prog. Photovolt.* **10**, 571–574 (2002).
54. Johansson T, Kelly H, Reddy A, Williams R, *Renewable Energy*, Chap. 1, Island Press, Washington, DC, 1–71 (1993).
55. *Photon Int.* 38–43 (March 2002).
56. Aulich H, Raptis F, Schmid J, *Prog. Photovolt.* **6**, 325–340 (1998).
57. Sumner D, Whitaker C, Schlueter L, *Proc. 20th IEEE Photovoltaic Specialist Conf.*, 1289–1292 (1988).
58. Wenger H, Jennings C, Ianucci J, *Proc. 21st IEEE Photovoltaic Specialist Conf.*, 844–849 (1990).
59. Iliceto A, Belli G, Lambri L, Toninelli F, *Proc. 14th Euro. Photovoltaic Solar Energy Conf.*, 2670–2674 (1997).
60. Matlin R, Shugar D, *Proc. 12th Euro Photovoltaic Solar Energy Conf.*, 842–845 (1994).
61. Sala G *et al*., *2nd World Conference and Exhibition on Photovoltaic Solar Energy Conversion*, 1963–1968 (1998).
62. Kardauskas M, Rosenblum M, Mackintosh B, Kalejs J, *Proc. 25th IEEE Photovoltaic Specialist Conf.*, 383–388 (1996).
63. Schmidt W, Woesten B, Kalejs J, *Prog. Photovolt.* **10**, 129–140 (2002).
64. Barnett A, Rothwarf A, *IEEE Trans. Electron Devices* **27**, 615–630 (1980).
65. Hewig G *et al*., *Proc. 16th IEEE Photovoltaic Specialist Conf.*, 713–718 (1982).
66. Bragagnolo J *et al*., *IEEE Trans. Electron Devices* **27**, 645–650 (1980).
67. Hall R, Birkmire R, Phillips J, Meakin J, *Appl. Phys. Lett.* **38**, 925, 926 (1981).
68. Catalano A *et al*., *Proc. 16th IEEE Photovoltaic Specialist Conf.*, 1421, 1422 (1982).

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 780

69. Mickelson R, Chen W, *Proc. 16th IEEE Photovoltaic Specialist Conf.*, 781–785 (1982).
70. Tyan Y, Perez-Albuerne E, *Proc. 16th IEEE Photovoltaic Specialist Conf.*, 794–799 (1982).
71. Phillips J, Birkmire R, Lasswell P, *Proc. 16th IEEE Photovoltaic Specialist Conf.*, 719–722 (1982).
72. Maycock P, *PV NEWS* **21**(4), 1 (April 2002).
73. Wills R, Hall F, Strong S, Wohlgemuth J, *Proc. 25th IEEE Photovoltaic Specialist Conf.*, 1231–1234 (1996).
74. Knaupp W *et al*., *Proc. 25th IEEE Photovoltaic Specialist Conf.*, 1235–1238 (1996).
75. Vernon S *et al*., *Proc. 22nd IEEE Photovoltaic Specialist Conf.*, 353–357 (1991).
76. Cotal H *et al*., *Proc. 28th IEEE Photovoltaic Specialist Conf.*, 955 (2000).
77. Algora C *et al*., *IEEE Trans. Electron Devices* **48**, 840–844 (2001).
78. Bett A *et al*., *Proc. 28th IEEE Photovoltaic Specialist Conf.*, 961 (2000).
79. Miñano J, González J, Benítez P, *Appl. Opt.* **34**, 7850 (1995).
80. Yamaguchi M, Luque A, *IEEE Trans. Electron Devices* **46**(10), 2139–2144 (1999).
81. Luque A, Martí A, *Phys. Rev. Lett.* **78**(26), 5014–5017 (1997).
82. Martí A, Cuadra L, Luque A, *Proc. 28th IEEE Photovoltaic Specialist Conf.*, 940–943 (2000).
83. Cassaboisa G *et al*., *Physica* **E13**, 105–108 (2002).
84. Wahnon P, Tablero C, *Phys. Rev. B* **65**, 165115-1–165115-10 (2002).

2

Motivation for Photovoltaic Application and Development

Joachim Luther

Fraunhofer Institut für Solare Energiesysteme, Heidenhofstr 2 D-79110 Freiburg, Germany

Photovoltaics is the technological symbol for a future sustainable energy supply system in many countries. A considerable amount of money is invested in research, development and demonstration; several governments set up substantial market introduction programs and industry invests in larger production facilities. No other renewable energy technology receives such a strong appreciation by the public and to an increasing extent also by the politicians and the industrial and financial sectors. This is a remarkable situation since at the same time photovoltaic (PV) electricity is regarded as much too expensive compared to conventional grid electricity. The high and justified recognition of photovoltaics may be understood on the basis of a description of the main positive features of this kind of solar electricity conversion.

2.1 CHARACTERISTICS OF PHOTOVOLTAIC ENERGY CONVERSION

Photovoltaics aims at two areas of application. One is the power supply for off-grid professional devices and supply systems (e.g. telecommunication equipment, solar home systems) and the other is large-scale electricity generation as a substitute for and a complement to today's non-sustainable energy processes. With respect to the latter, the global potential of PV electricity is of key importance. Figure 2.1 shows the technical and the theoretical potential of several renewable energy sources. The theoretical potential does not take into account land use restrictions, conversion efficiencies, storage requirements and so on. The technical potential on the other hand must not be confused with short-term economic potentials, since price situations and capital requirements for activating these energy sources on a large scale are not considered.

Handbook of Photovoltaic Science and Engineering. Edited by A. Luque and S. Hegedus
© 2003 John Wiley & Sons, Ltd ISBN: 0-471-49196-9

Resource	Current use	Technical potential	Theoretical potential
Hydropower	9	50	147
Biomass energy	50	>276	2900
Solar energy	0.1	>1575	3900000
Wind energy	0.12	640	6000

Units: exajoule per year

Figure 2.1 Current use and current potentials of selected renewable energy sources [1]. For comparison: global primary energy demand 402 exajoule/annum (1998). The electricity part of "current use" has been converted to primary energy utilising an efficiency factor of 0.385

Nevertheless, two important conclusions may be drawn from Figure 2.1: (1) even under strong area restrictions (e.g. utilisation of a small percentage of the land area) and guarded assumptions of overall technical efficiencies, solar energy conversion alone could in principle produce considerably more technical energy than is consumed today and (2) compared to other renewable sources, solar radiation is by far the largest. A sustainable global energy system that is strongly based on renewable sources will in the long run mainly be a solar energy system. With respect to the technology of solar electricity production, this means solar thermal power plants and photovoltaics. The application of thermal power plants is restricted to areas with high and direct insulation; flat-plate (standard) PV modules may be applied practically everywhere in the world since they convert diffuse and direct (beam) radiation with approximately the same efficiency.

Photovoltaic energy conversion meets the important requirements of a sustainable energy production in an obvious way. During operation there is no harmful emission or transformation of matter (generation of pollutants), nor any production of noise or other by-products. PV energy conversion is a technologically elegant one-step process avoiding conventional thermodynamic or mechanical intermediate steps. On the other hand, production of PV modules and system components will – as any industrial device production – include material transformation and the production of wastes. Thus, it is extremely important to realise PV conversion technologies that comply with the requirements of environmentally benign production schemes.

Though characterised by the high global potential given in Figure 2.1, the area-specific power density of solar radiation is relatively low, that is, approximately 100 W/m^2 on the average. This means that the global harvesting of solar resources necessarily requires a large-area production of energy converters. Appropriate recycling strategies will thus be essential for the energy-relevant application of photovoltaics. Already today standard silicon-wafer PV-technology meets in principle the requirements with respect to recycling and sustainable production.

Photovoltaic energy conversion is highly modular. Installations may vary between milliwatts for consumer products (watches) to megawatts in the case of grid-connected power plants. From a market point of view, this gives rise to a broad variety of PV applications. For the professional energy supply business, modularity is especially important with

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 783

respect to the development of electricity supply systems in many rural and remote areas where grid extension is economically not feasible. Starting from low power installations, PV modules may be gradually added to suitable systems in order to cope with the growing energy demands. In this way electricity supplies can be realised avoiding too high initial investments. In the context of rural electrification and also for the professional energy supply of off-grid electronic devices, the low maintenance characteristic of photovoltaics is regarded as a considerable advantage. The absence of moving parts, the robustness against harsh environment and the lack of fuel supply requirements (exception: hybrid systems) make photovoltaics a well-suited energy supply technique for a vast area of stand-alone energy supplies.

Another positive characteristic of PV is that it may be seen as part of the rapidly growing semiconductor industry/market. This relation facilitates a strong technology transfer from a mature industry to the emerging PV industry. Parallel to this the affinity opens the prospect of new big industries creating considerable business and employment opportunities.

The strong market growth of photovoltaics over the last two decades (Figure 2.2) and the high recognition of this technology could not be understood if there had not been a continuous and strong reduction in PV system and module prices.

Figure 2.3 shows the price-experience curve (learning curve) until 2000. Each doubling of the module shipment resulted on the average in a market price reduction of 20%. It is generally accepted that it is possible to cut down the prices by another factor of two utilising mass production lines based on today's crystalline silicon-wafer technology. Considerable further cost reduction seems feasible since new and most probably least costly PV technologies are already in the state of pilot production (thin-film cells) or under investigation in laboratories: cells based on III/V materials and optical concentration, cells using dyes or organic compounds for energy conversion and so on. Thus, there is a good

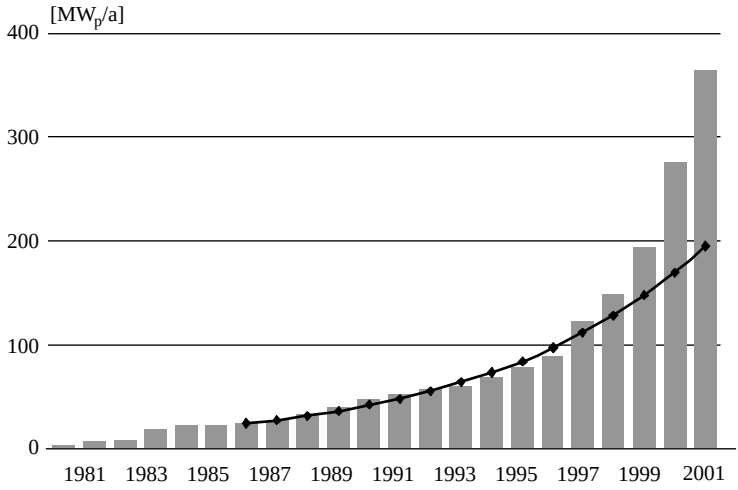


Figure 2.2 Evolution of the global photovoltaic market [2]. Until 1996 the growth rate was approximately 15%/annum. Today's growth rates are in the order of 30%/annum

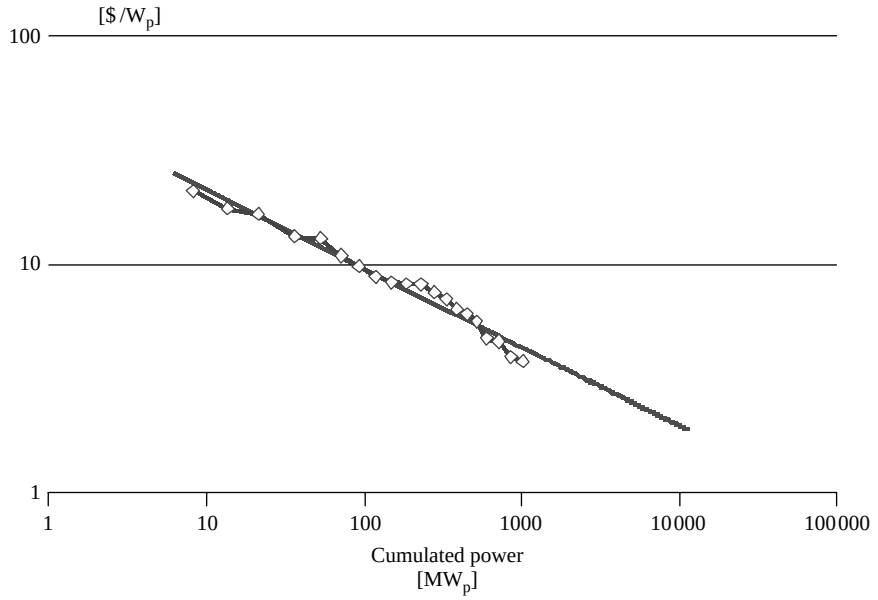


Figure 2.3 Price-experience curve of the photovoltaic market until 2000 [2]. Scale: double logarithmic. The slope of the curve is characterised by a factor $f = 0.8$. This means that a doubling of the photovoltaic shipment resulted in a price reduction down to 80%

chance of at least a break even between PV electricity and conventional electricity prices, given that the external (social) costs of non-sustainable electricity are properly taken into account.

2.2 A LONG-TERM SUBSTITUTE FOR TODAY'S CONVENTIONAL ELECTRICITY PRODUCTION – THE ECOLOGICAL DIMENSION OF PHOTOVOLTAICS

A sustainable development has to respect certain guard rails. With respect to the energy supply system, these are mainly determined by the global warming/CO_2 problem (Figure 2.4). A global warming of more than 2°C (with respect to the pre-industrialised period) and a warming rate of 0.2°C per decade seems to be the upper limits for a development that may be kept under control. These settings are in accordance with the results of recent studies of the International Panel on Climate Change [3].

Since preservation and expansion and especially the transformation of energy systems require important investments, a strategy of large stopping-distances seems reasonable. For example, the German Advisory Council on Global Change proposes a continuous CO_2 reduction rate of 1% per year over the next 150 years [3]. If one strives for a fair and rational distribution of the worldwide reduction obligations, the highly industrialised nations have to reduce their CO_2 emissions until 2050 by almost 80% relative to the 1990 level [3]. This reduction mainly affects the energy sector including fuels for transportation.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 785

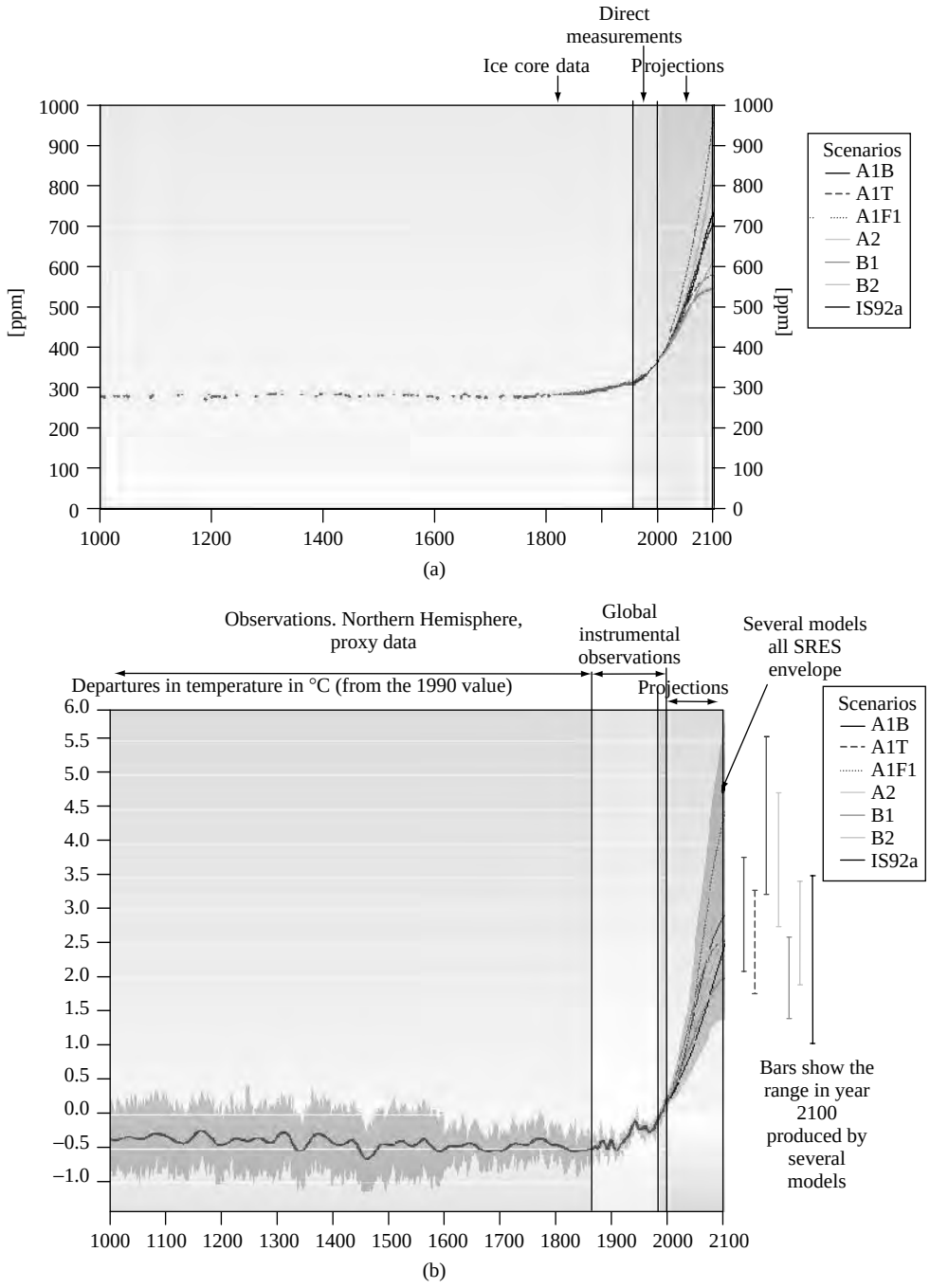


Figure 2.4 Past and future CO_2 atmospheric concentrations (a) and variations of the Earth's surface temperature (b). Time span: year 1000 to year 2100. The projections on the right sides are based on scenarios of the Intergovernmental Panel on Climate Change (IPCC) [4]. Reproduced by permission of Cambridge University Press

As far as can be seen today, such reduction goals may – in a socially acceptable way – only be realised if renewable energy technologies are applied worldwide on a really large scale. For electricity production, this means above all the use of photovoltaics and to a lesser (but important) extent wind energy, biomass conversion and solar thermal power plants. Thus, the principal ecological dimension of photovoltaics is extremely high. The question arises of course whether we can activate this potential in due time. With respect to this, at least two sub-problems have to be addressed: (1) how fast can the industry produce the required large-area PV energy conversion systems and (2) is it possible to establish a strong long-lasting market-driving mechanism that conforms with today's ideas of a competitive market.

During the last few years, annual market growth rates and thus the increase in industrial PV production have been in the order of 15 to 30% (Figure 2.2). The main drivers behind this impressive growth are several governments and consumers who see an urgent need for a transformation of today's energy systems towards sustainability [5–7]. Financial-support mechanisms have been developed that include clear components of competition inside well-defined areas of energy supply technologies. The financial support is justified by the high potential of PV electricity and its ecologically benign character. On the other hand, such schemes have to be limited in time, which is especially important if the promoted markets evolve exponentially. Thus, the question arises whether grid-coupled PV electricity will become cost-effective within reasonable time frames. Figure 2.5 shows a cost projection of base-load, peak-load and photovoltaic electricity.

The assessment is based on an extrapolation of the prize experience curve given in Figure 2.3, a PV-market growth rate of 20%/annum and an increase of conventional electricity costs by 2% per year. Under these assumptions, a break even between peak load and

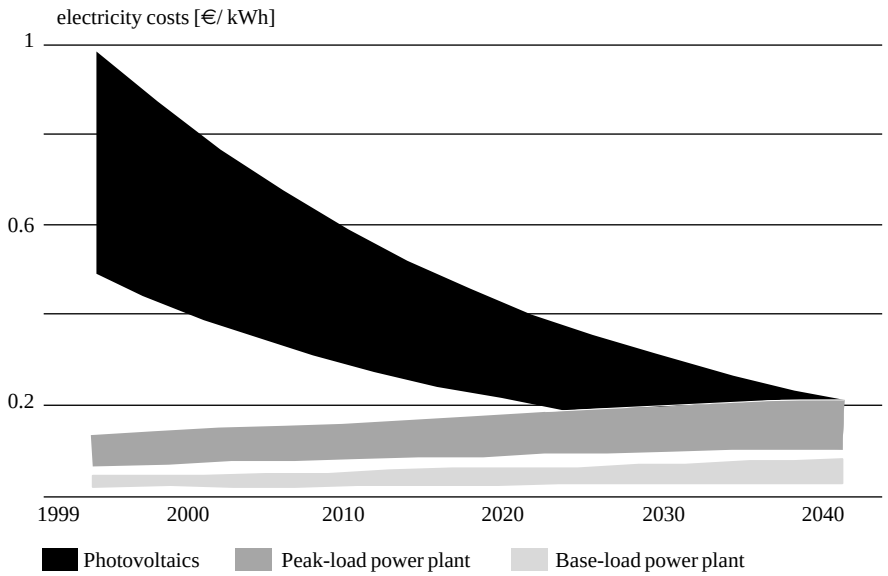


Figure 2.5 Cost projections for photovoltaic electricity. The upper boundary of photovoltaic costs reflects the meteorological situation of Germany, the lower boundary that of Southern Europe [8]

solar electricity is found around 2025; a break even with costs from conventional base-load power plants is not reached in the first half of this century.

The extrapolation of a prize experience curve over several decades bears, of course, a considerable risk. On the other hand, there are good technological arguments that the price-experience curve will continue to decrease in a similar way as in the past for a considerable time span; a strong reduction in the slope of the curve most probably will not be encountered before a further price reduction by a factor of 3 to 5 is reached. It is important (and essential) to mention that most probably this statement is also valid for the balance of system components such as power electronics and so on.

Many conclusions may be drawn from a graph such as that given in Figure 2.5 (based partly on an arbitrary set of parameters). If one aims, for example, at an efficient transformation of today's electricity supply system, one may propose an energy tax on environmentally and socially non-benign energy sources in order to establish competitive market situations at a significantly earlier time. This tax could be based on the internalisation of external costs of the different competing energy supply technologies. In order to allow for a smooth evolution towards sustainability, such a tax should be introduced gradually. If one assumes, for example, a continuous additional tax-induced cost increase of electricity from conventional peak- and base-load power plants of 2% per year, the break even in Figure 2.5 for peak load would occur around 2015 and that for base load around 2030. The public income from such an energy tax could be invested in research and development in the field of renewable energies and in energy efficiency measures. This would support, for example, the continuity of cost reduction of photovoltaics (Figure 2.3) as well as the reduction of the gross energy demand and thus of the energy-related financial burdens.

From an industrial point of view, a strong evolution of PV electricity generation could produce an interesting global market. A long-term price-experience factor $f = 0.82$ (see caption to Figure 2.3) and a market growth rate of 20% per year would generate, for example, the following global markets for PV-modules: 2010 2×10^9 €, 2020 9×10^9 € and 2030 30×10^9 € [8]. The additional market for system components and integration will probably be of the same order. The prospects of such substantial markets are the motivation of today's considerable industrial investments in PV production capacities.

It has been shown so far that if it is possible to maintain a price-experience evolution as shown in Figure 2.3 and if governments adopt a serious energy tax strategy, electricity from photovoltaics may become cost-effective between approximately 2015 (peak power) and 2030 (bulk power). Attractive high-technology markets could develop in the near future. One question still to be answered is 'when will photovoltaics contribute considerably to the global electricity generation?'

In a speculative way, this question may be addressed by stipulating conceivable long-term growth rates for PV electricity generation (Figure 2.6). Rates of 26%/annum results in a growth factor of 10 over one decade. Starting at present-days low level it will – even under these conditions – take at least three to four decades for photovoltaics to contribute in a substantial way to the global electricity demand (for rural electrification see Section 2.3). After this starting period, photovoltaics can become one main supplier of electric energy in the future energy mix. Growth rates of 15% per year on the other hand produce just a growth factor of four over one decade; under such conditions, photovoltaics

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 788

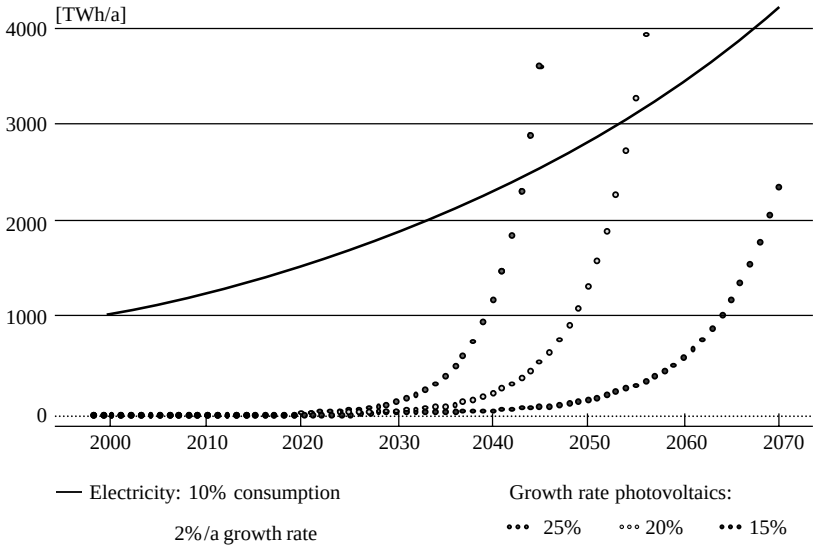


Figure 2.6 Scenarios of global photovoltaic electricity production under the assumption of constant growth rates of 25%/annum, 20%/annum and 15%/annum. The solid curve represents 10% of the global electricity demand assuming a constant growth rate of 2%/annum. A linear scale is used for the energy axis. Thus, although growing exponentially, the energy production by means of photovoltaics is not visible on the scale used until the year 2020

may become a major player only in the second half of our century. Since the man-made climatic problems call for a fast and forceful solution that is both environmentally benign and socially acceptable, high and continuously high growth rates for photovoltaics are extremely advisable.

The liberalisation of the energy market has led to new and innovative concepts, especially in the electricity sector. One of these is the distributed generation scheme – kilowatt to several megawatt installations instead of central units in the gigawatt range. Its main advantages are dispersed on-site heat and power cogeneration, electrical stabilisation of grids especially at the low voltage level, distributed reserves in the case of supply problems and partly favourable investment situations. PV electricity generation is well adapted to the distributed generation scheme (Figure 2.7). This is especially the case if peak-load conditions in the grids are considered that are strongly correlated with solar insolation: cooling and climatisation units. Also, the extreme modularity of photovoltaics turns out to be advantageous for distributed generation. Distributed peak power production will be most probably the first energy market application for optical concentrator PV power stations.

If performed on a continental or even intercontinental level, the highly dispersed application of photovoltaics will also lead to a strong levelling out of its stochastic local power generation characteristics. The remaining slowly varying time pattern of the lumped electricity production (diurnal, seasonal variations) can be compensated (of course, at extra costs) by the control of complementary electricity sources or in the long run by means of storage technologies.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 789



Figure 2.7 Distributed power generation by means of photovoltaics. The roof-integrated photovoltaic system has a peak power of 5 kW

Another area that should be highlighted separately is façade- and roof-integrated grid-coupled photovoltaics. If done in an aesthetically promising way, PV claddings may substitute prestigious building elements. These applications may become the first mass market for homogeneous-looking thin-film PV modules. Even if the area-related investment costs of a PV building element is higher than the cost of conventional high-end elements, the electricity generated and the benefit of architectural double uses (Figure 2.8) may compensate for this difference within a reasonable time.



Figure 2.8 Photovoltaic integration into the glazing of a shed roof. In this case, photovoltaic cells serve also as optical shading elements for the space below, preventing overheating under summer conditions

2.2.1 In Summary

The technical potential of PV electricity is high enough to contribute considerably to the abatement of the man-made global CO_2 problem. With the help of medium-term financial supports and long-term energy taxes (motivated by external costs), the technical potential may be exploited economically. Under grid-coupled operation, photovoltaics will most probably become cost-effective for distributed peak power production and for applications in the building sector (as cladding element). Solar electricity generation has to grow by three orders of magnitude before it encounters a (extrapolated) 10% level of the global electricity demand. Assuming ambitious growth rates, this process of transformation will take three to four decades.

2.3 A TECHNOLOGICAL BASIS FOR OFF-GRID ELECTRICITY SUPPLY – THE DEVELOPMENT DIMENSION OF PHOTOVOLTAICS

Two billion people worldwide have no access to commercial electricity. When possible, they cover their needs for electricity-based services by means of primary batteries, rechargeable batteries (car batteries) or small fossil fuel–driven generator sets. In many (if not in most) cases these supply schemes are very costly, uncomfortable, unreliable and ecologically questionable. This results in the fact that for a large fraction of the global population, services such as electric light, radio, television, telecommunication, health services, clean water, cooling, electromechanical energy and so on are not at all available or are available only at a very low level. It must be stressed that these electricity-based benefits are amongst others crucial for education, business and small commercial activities (handicrafts, agriculture, food processing etc.). Thus the lack of a stable and affordable electricity supply hinders the development of many rural and remote areas. This is ethically not acceptable; it may lead to a destabilisation of regions and foster the growth of megacities, just to name three important reasons why this situation cannot be regarded as socially sustainable.

Beyond the low power level of primary batteries, there are mainly three worldwide applicable technology groups that are able to contribute in a professional way to the electricity supply of remote areas: grid extension, diesel generator sets and renewable energies. The most prominent amongst the renewables are small hydro, wind and photovoltaics. All technologies mentioned will eventually contribute to the electricity supply of rural and remote areas.

At first sight, grid extensions seem to be the most natural and technologically elegant way to overcome shortages in the electricity supply. It turns out, however, that the initial investments for such a strategy are prohibitively high if the consumers are scattered over large areas and if the average electricity demand, that is, the energy transported over the lines, is relatively small. Under these situations it is mostly impossible to recover the investments by selling electricity to the dispersed consumers or to raise it through subsidies from governmental organisations.

The alternatives to grid extension are renewables and diesel electric systems. In the following text, the benefits and characteristics of photovoltaics will be discussed.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 791

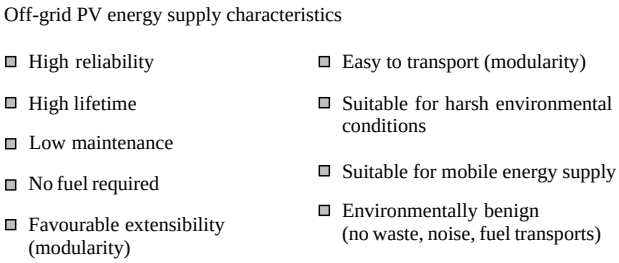


Figure 2.9 Characteristics of off-grid photovoltaic energy supply systems

Figure 2.9 summarises some of the salient features of photovoltaics for off-grid electricity supply. The main disadvantage of PV systems is the comparatively high initial investments. Diesel systems on the other hand produce relatively low initial investments but high running costs especially due to fuel requirements.

Off-grid electricity supply systems generally need energy storages. In the case of diesels these are the fuel tanks, and in PV systems electrochemical batteries are applied. Especially because of this fact, solar electric systems are not maintenance-free (though their behaviour with respect to maintenance and wear is generally significantly more favourable than it is for diesel systems). For reliable operation, the storage battery has to be exchanged at regular intervals. Figure 2.10 indicates clearly that at present it is not the PV module but the system electronics (including installation) and the storage devices that constitute the main lifetime cost factors. Thus research and development in the fields of system technology and solar batteries would be highly supportive for an accelerated introduction of PV off-grid systems worldwide.

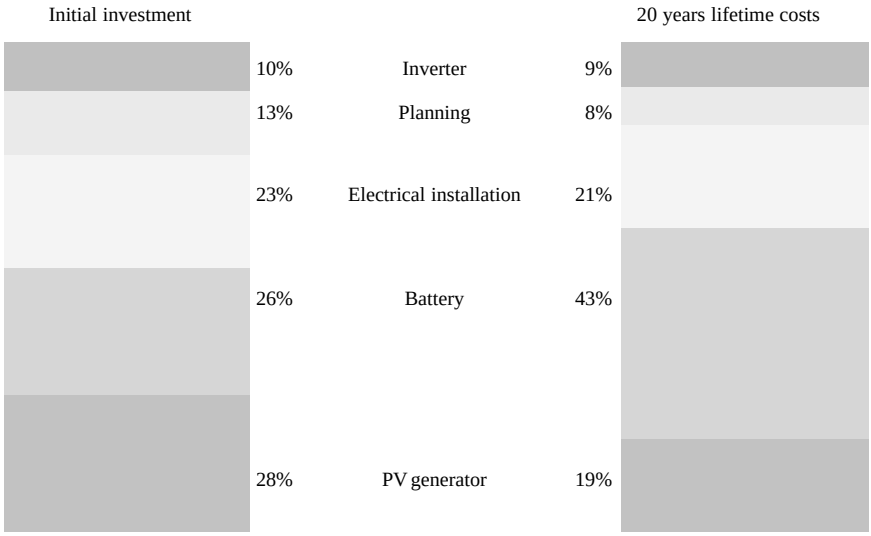


Figure 2.10 Initial versus lifetime costs of small professional autonomous photovoltaic systems – a typical example. System data: PV 1.3 kW_p, battery 35 kWh, battery lifetime 5.4 a, loss of load probability 0.1%, system site Mexico

At present, the main applications of photovoltaics in rural and remote areas are (1) photovoltaic (car-) battery-charging stations where the charged battery is transported to the electricity consumer, (2) photovoltaic solar home systems for the energy supply of electric lights, radios and simple television sets and (3) photovoltaic powered pumping stations for water supply for human requirements and agricultural irrigation.

In particular, the solar home systems have to be highlighted. As might have become clear from the above discussion, they present today in many cases the technical and economic optimal solution for a basic electrification in the areas under discussion. The market volume is in the order of 400 000 systems a year. System prices show increasingly favourable values (Figure 2.11).

The main barriers to a further accelerated introduction of these technologies into practice are non-technological in character. Appropriate financing schemes, social integration, training of engineers and trade and industry structures have to be developed and efficiently integrated into the societies. Photovoltaics may then become one of the main pillars of rural electrification. A market volume of 15 to 30 GW per year is expected in 10 to 20 years [8].

It has already been mentioned that modularity is one of the important characteristics of photovoltaics. In principle, solar home systems may be upgraded step by step. In suitable cases it will be possible to interconnect the individual systems to form PV village power supplies. Besides the social management of such structures, a prerequisite for their widespread applications is the development and industrial production of appropriate power electronics, load management devices and information schemes. For larger installations, hybrid systems (e.g. diesel/photovoltaic) may dominate the market in the near future. It is also conceivable that village power systems may be interconnected by transmission lines or even connected to the main (possibly weak) grids. All these measures increase the complexity and most probably also the cost of such systems. On the other hand, because of technical redundancies and load levelling effects, the performance of such advanced PV-based electricity supply systems could increase considerably.

	Price (US $)	Price (%)	Lifetime (years)
PV-module (53 W) and support	200	47	>20
Battery (70 Ah)	40	9	4
Battery charge controller	35	8	10
Lamps, wiring, switches	35	8	5
Delivery, installation, retail margins	75	18	-
Duties and taxes	40	10	-
Total	425	100	

Figure 2.11 Cost break down of a typical solar home system

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 793

Off-grid PV electricity supply is by no means restricted to the rural areas of developing countries. Mountain lodges, remotely situated farms, holiday houses and so on represent a considerable market for PV energy supply also in industrialised countries. Such installations may be connected to the central grid as well. In these cases photovoltaics serves as a back-up electricity supply in case of line failure or unreliable electricity supply from the utility side.

2.3.1 In Summary

Photovoltaics is an excellently suitable solution for low power electricity supply in rural and remote areas in developing countries. Photovoltaics have a similar but smaller market in industrialised countries as well. In total, one-third of today's world population could benefit from off-grid PV installations. System electronics and storage batteries have to be further developed in order to address the market in an optimal way. Financial engineering, social integration of the new technology, training and the installation of an appropriate trade and industrial infrastructure are the most important prerequisites for a successful worldwide implementation of off-grid PV electricity supplies.

2.4 POWER SUPPLY FOR INDUSTRIAL SYSTEMS AND PRODUCTS – THE PROFESSIONAL LOW POWER DIMENSION

This segment of PV application comprises the electricity supply of consumer products, professional industrial systems and remote low power systems. In this area, photovoltaics substitutes or complements the traditional power supply via batteries or small fossil fuel motor generator sets. The average power demand, especially in the area of consumer products and professional systems, lies generally in the power range of several milliwatts to 100 W. On the other hand, a mass market of billions of units may be envisaged gaining market volumes in the 10-GW range in several decades [8]. The drivers of these markets are strong technological innovations in the field of telecommunication, dispersed electronic intelligence and remote sensors/actuators and so on (Figures 2.12. and 2.13).

Such equipments need small amounts of electric energy for operation and standby. Even if the electric grid is ready-to-hand, wireless and maintenance-free systems are generally preferred by the users. Maintenance-free here means that costly primary batteries with limited energy content can be avoided. In many cases photovoltaics may enter this field of energy supply (1) if the devices are at least temporarily under illumination and (2) if suitable storage devices are applicable. Thus the availability of small and efficient storages (rechargeable batteries, capacitors) is a prerequisite for a widespread application of PV cells in this area.

In this market segment there is a wide range of specific requirements for PV cells. Communication equipment like mobile phones being characterised by low surface area and relatively high energy demand call for high efficiency (e.g. 20%) solar cells; wrist watches on the other hand can be powered by PV cells having only some percentage of efficiency. PV car roofs need aesthetically optimised precision products, while powering a last mile repeater for telecommunication links may be done with small standard PV

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 794

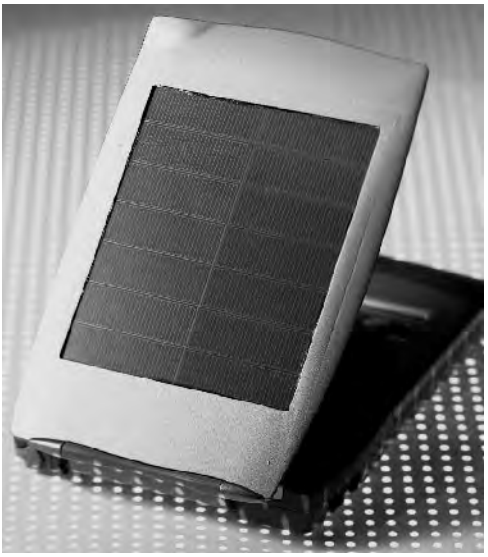


Figure 2.12 Personal digital assistant powered by high efficient silicon solar cells. Module efficiency at one sun >20%; efficiency at 1/100 sun, 17%



Figure 2.13 Autonomous emergency phone and information system using photovoltaic power supply

modules. The optimal integration of PV cells into the surface of electronic devices may ask for flexible, geometrically tailored or even coloured components.

Especially for indoor applications it is essential to have light-energy converters with high conversion efficiencies at low illumination levels.

These diverse and partly conflicting requirements will result in the development of different types of PV converters, well suited to the individual applications. Dye or organic solar cells may find their first application in such market segments.

2.5 POWER FOR SPACECRAFT AND SATELLITES – THE EXTRATERRESTRIAL DIMENSION OF PHOTOVOLTAICS

The power supply of satellites has been the first professional application of photovoltaics (Vanguard I, 1958). While at the beginning military activities have been in the foreground of space power supply, today commercial activities play an important role as well: powering satellites for telecommunication, remote sensing, navigation and so on.

In these obviously off-grid applications, PV power supplies are more or less without technological competitors. Thus, solar electric systems of this kind would have been

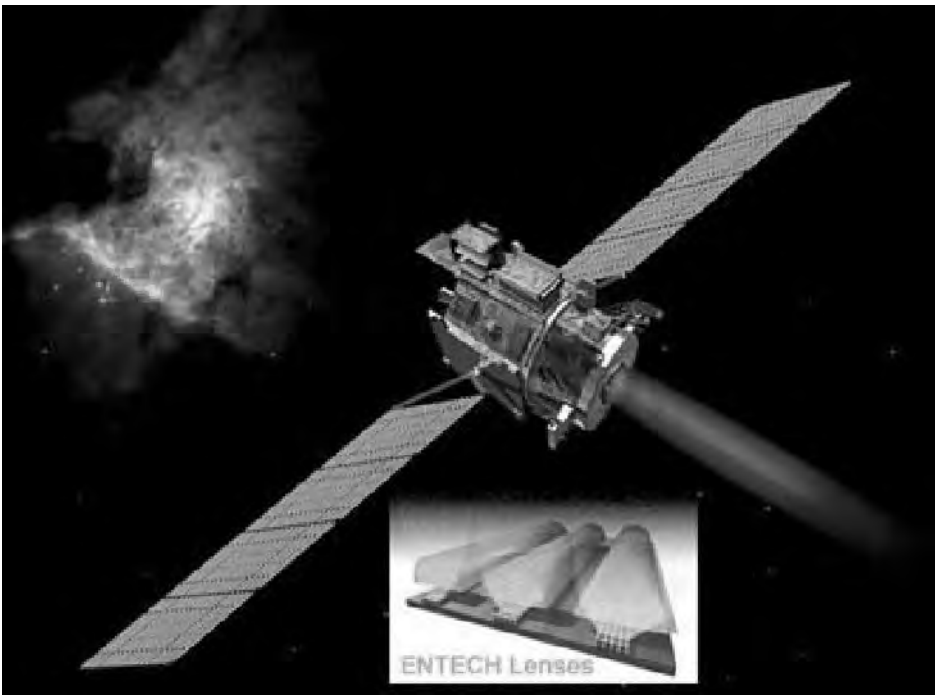


Figure 2.14 Artists view of the "Scarlet Deep Space One Satellite". The satellite is equipped with triple-junction GaInP/GaAs/Ge photovoltaic cells. For the first time, optical concentration was applied with III/V space cells. Courtesy of NASA/JPL/Caltech

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 796

successfully developed even without the terrestrial market addressed above. On the other hand, the parallel (but partly deferred) evolution of space and terrestrial photovoltaics has led to many fruitful and essential cross-stimulations.

Solar modules for space applications (Figure 2.14) have to meet other and mostly more stringent requirements than terrestrial devices such as lowest weight and highest efficiency, extremely high reliability, high resistivity against extraterrestrial particle radiation (high lifetime) and spectral sensitivity that is well matched to the extraterrestrial solar spectrum.

Photovoltaic modules meeting corresponding standards have been realised in a relatively short time span. This success was based on many ingenious inventions and on the fact that cost reduction has not been the most important issue in the first decade of satellite technology. Thus, space photovoltaics became, at least initially, the technological breeder for today's silicon-wafer-based terrestrial solar cell technology.

A similar technology fertilisation might occur again: About one-half of today's space solar cells are already made from GaAs and related compound semiconductors (Figure 2.14). Such solar cells exhibit considerable higher efficiencies and lifetimes than Silicon cells. There is a realistic chance that these space-proven technologies become the basis for a new class of solar cells having light conversion efficiencies in the range of 40 or even 50%. Besides advanced flat-plate solar modules such solar cells, in conjunction with optical concentrators, constitute a strong technological option for future terrestrial large-scale electricity generation.

REFERENCES

1. United Nations Development Programme (UNDP) World Energy Assessment, ISBN 92-1-126126-0, New York (2000).
2. Data Source: Projektgesellschaft Solare Energiesysteme (PSE GmbH), Freiburg, Germany (2002).
3. Scenario for the derivation of CO_2-reduction targets and implementation strategies. Statement of the Germany Advisory Council on Global Change (WBGU) on the occasion of the COP-1 of the UNFCCC in Berlin 1995. Internet: *http://www.wbgu.de/wbgu_sn1995_engl.html* and IPCC – Intergovernmental Panel on Climate Change, Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the IPCC, Cambridge University Press, Cambridge, New York (2001).
4. IPCC – Intergovernmental Panel on Climate Change, Climate Change 2001–Synthesis Report. Contribution of Working Groups I, II and III to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, New York (2001).
5. Saitoh T, "Restructuring of photovoltaic R&D programs in Japan", in Sopori B, Ed, *Proc. 11th Workshop on Crystalline Silicon Solar Cell Materials and Processes*, NREL, Estes Park, CO (2001).
6. German Government, *Sol. Energy* **70**, 489–504 (2001).
7. Photovoltaics – Energy for the New Millennium, The National 2000–2004 Photovoltaic Program Plan, DOE/GO-10099-940.
8. Hoffmann W, "PV solar electricity: one among the new millennium industries", *Proceedings of the 17th European Photovoltaic Solar Energy Conference*, Munich, Germany (2001).

3

The Physics of the Solar Cell

Jeffery L. Gray

Purdue University, West Lafayette, Indiana, USA

3.1 INTRODUCTION

Semiconductor solar cells are fundamentally quite simple devices. Semiconductors have the capacity to absorb light and to deliver a portion of the energy of the absorbed photons to carriers of electrical current – electrons and holes. A semiconductor diode separates and collects the carriers and conducts the generated electrical current preferentially in a specific direction. Thus, a solar cell is simply a semiconductor diode that has been carefully designed and constructed to efficiently absorb and convert light energy from the sun into electrical energy. A simple conventional solar cell structure is depicted in Figure 3.1. Sunlight is incident from the top on the front of the solar cell. A metallic grid forms one of the electrical contacts of the diode and allows light to fall on the semiconductor between the grid lines and thus be absorbed and converted into electrical energy. An antireflective layer between the grid lines increases the amount of light transmitted to the semiconductor. The semiconductor diode is fashioned when an n-type semiconductor and a p-type semiconductor are brought together to form a metallurgical junction. This is typically achieved through diffusion or implantation of specific impurities (dopants) or via a deposition process. The diode's other electrical contact is formed by a metallic layer on the back of the solar cell.

All electromagnetic radiation, including sunlight, is composed of particles called photons, which carry specific amounts of energy determined by the spectral properties of their source. Photons also exhibit a wavelike character with the wavelength, λ, being related to the photon energy, E_λ, by

$$E_\lambda = \frac{hc}{\lambda} \tag{3.1}$$

where h is Plank's constant and c is the speed of light. Only photons with sufficient energy to create an electron–hole pair, that is, those with energy greater than the semiconductor

Handbook of Photovoltaic Science and Engineering. Edited by A. Luque and S. Hegedus
© 2003 John Wiley & Sons, Ltd ISBN: 0-471-49196-9

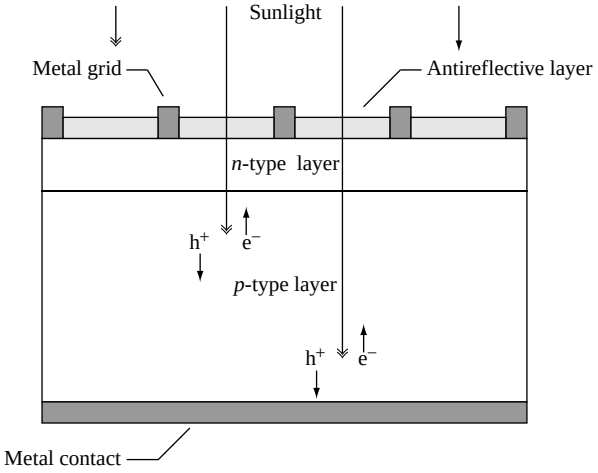


Figure 3.1 A schematic of a simple conventional solar cell. Creation of electron–hole pairs, e^- and h^+, respectively, is depicted

band gap (E_G), will contribute to the energy conversion process. Thus, the spectral nature of sunlight is an important consideration in the design of efficient solar cells.

The sun has a surface temperature of 5762 K and its radiation spectrum can be approximated by a black-body radiator at that temperature. Emission of radiation from the sun, as with all black-body radiators, is isotropic. However, the Earth's great distance from the sun (approximately 93 million miles) means that only those photons emitted directly in the direction of the Earth contribute to the solar spectrum as observed from Earth. Therefore, for practical purposes, the light falling on the Earth can be thought of as parallel streams of photons. Just above the Earth's atmosphere, the radiation intensity, or Solar Constant, is about 1.353 kW/m^2 [1] and the spectral distribution is referred to as an *air mass zero* (AM0) radiation spectrum. The Air Mass is a measure of how absorption in the atmosphere affects the spectral content and intensity of the solar radiation reaching the Earth's surface. The Air Mass number is given by

$$\text{Air Mass} = \frac{1}{\cos\theta} \tag{3.2}$$

where θ is the angle of incidence ($\theta = 0$ when the sun is directly overhead). The Air Mass number is always greater than or equal to one at the Earth's surface. An easy way to estimate the Air Mass has been given by Green [1] as

$$\text{Air Mass} = \sqrt{1 + (S/H)^2} \tag{3.3}$$

where S is the length of a shadow cast by an object of height H. A widely used standard for comparing solar cell performance is the AM1.5 spectrum normalized to a total power density of 1 kW/m^2. The spectral content of sunlight at the Earth's surface also has a diffuse (indirect) component owing to scattering and reflection in the atmosphere and surrounding landscape and can account for up to 20% of the light incident on a solar

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 799

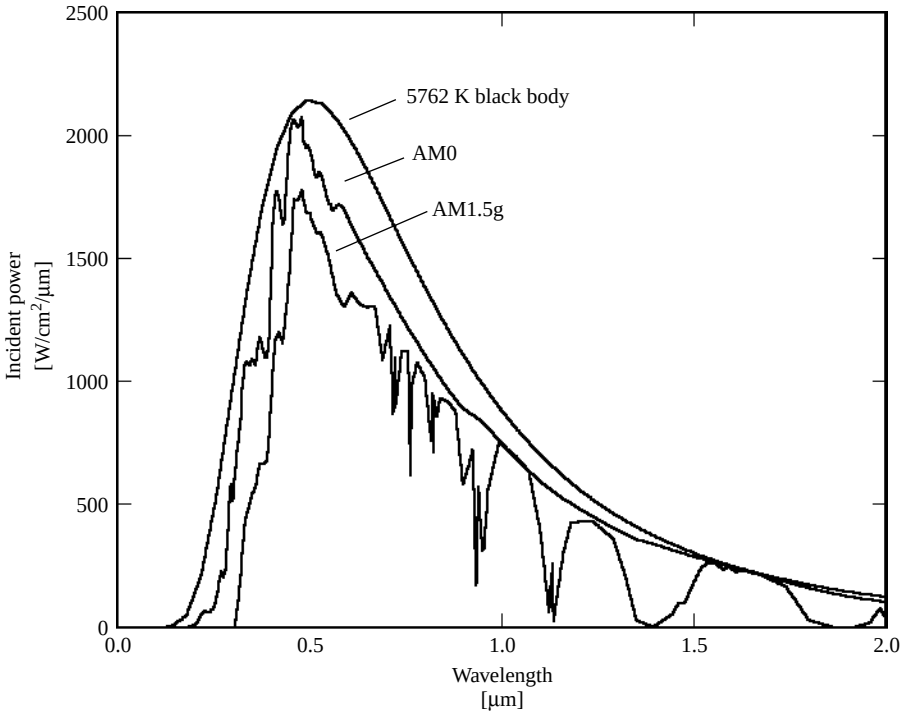


Figure 3.2 The radiation spectrum for a black body at 5762 K, an AM0 spectrum, and an AM1.5 global spectrum

cell. The Air Mass number is therefore further defined by whether or not the measured spectrum includes the diffuse component. An AM1.5g (global) spectrum includes the diffuse component, while an AM1.5d (direct) does not. Black body ($T = 5762$ K), AM0, and AM1.5g radiation spectrums are shown in Figure 3.2. The Air Mass and solar radiation are described in more detail in Chapters 16 and 20.

The physical principles underlying the operation of solar cells are the subject of this chapter. First, a brief review of the fundamental properties of semiconductors is given that includes an overview of semiconductor band structure and carrier generation, recombination, and transport. Next, the electrostatic properties of the *pn*-junction diode are reviewed, followed by a description of the basic operating characteristics of the solar cell, including the derivation (based on the solution of the minority-carrier diffusion equation) of an expression for the current–voltage characteristic of an idealized solar cell. This is used to define the basic solar cell figures of merit, namely, the open-circuit voltage, V_{OC}; the short-circuit current, I_{SC}; the fill factor, FF; the conversion efficiency, η; and the collection efficiency, η_C. Much of the discussion here centers on how carrier recombination is the primary factor controlling solar cell performance. Finally, some additional topics relevant to solar cell operation, design, and analysis are presented. These include the relationship between band gap and efficiency, the solar cell spectral response, parasitic resistive effects, temperature effects, a brief introduction to some modern cell design concepts, and a short overview of detailed numerical modeling of solar cells.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 800

3.2 FUNDAMENTAL PROPERTIES OF SEMICONDUCTORS

An understanding of the operation of semiconductor solar cells requires familiarity with some basic concepts of solid-state physics. Here, an introduction is provided to the essential concepts needed to examine the physics of solar cells. More complete and rigorous treatments are available from a number of sources [2–6].

Solar cells can be fabricated from a number of semiconductor materials, most commonly silicon (Si) – crystalline, polycrystalline, and amorphous. Solar cells are also fabricated from GaAs, GaInP, Cu(InGa)Se_2, and CdTe, to name but a few. Solar cell materials are chosen largely on the basis of how well their absorption characteristics match the solar spectrum and their cost of fabrication. Silicon has been a common choice due to the fact that its absorption characteristics are a fairly good match to the solar spectrum, and silicon fabrication technology is well developed as a result of its pervasiveness in the semiconductor electronics industry.

3.2.1 Crystal Structure

Electronic grade semiconductors are very pure crystalline materials. Their crystalline nature means that their atoms are aligned in a regular periodic array. This periodicity, coupled with the atomic properties of the component elements, is what gives semiconductors their very useful electronic properties. An abbreviated periodic table of the elements is given in Table 3.1.

Note that silicon is in column IV, meaning that it has four valence electrons, that is, four electrons that can be shared with neighboring atoms to form covalent bonds with those neighbors. In crystalline silicon, the atoms are arranged in a *diamond lattice* (carbon is also a column IV element) with tetrahedral bonding – four bonds from each atom where the angle between any two bonds is 109.5°. Perhaps surprisingly, this arrangement can be represented by two interpenetrating face-centered cubic (*fcc*) unit cells where the second *fcc* unit cell is shifted one-fourth of the distance along the body diagonal of the first *fcc* unit cell. The lattice constant, ℓ, is the length of the edges of the cubic unit cell. The entire lattice can be constructed by stacking these unit cells. A similar arrangement, the *zincblende* lattice, occurs in many binary III–V and II–VI semiconductors such as GaAs (a III–V compound) and CdTe (a II–VI compound). For example, in GaAs, one interpenetrating *fcc* unit cell is composed entirely of Ga atoms and the other entirely of As atoms. Note that the average valency is four for each compound, so that there are four bonds to and from each atom with each covalent bond involving two valence electrons. Some properties of semiconductors are dependent on the orientation of the crystal

Table 3.1 Abbreviated periodic table of the elements

I	II	III	IV	V	VI
		B	C	N	O
		Al	Si	P	S
Cu	Zn	Ga	Ge	As	Se
Ag	Cd	In	Sn	Sb	Te

lattice, and casting the crystal structure in terms of a cubic unit cell makes identifying the orientation easier using Miller indices.

3.2.2 Energy Band Structure

Of more consequence to the physics of solar cells, however, is how the periodic crystalline structure of the semiconductor establishes its electronic properties. An electron moving in a semiconductor material is analogous to a particle confined to a three-dimensional box that has a complex interior structure due primarily to the potential fields surrounding the component atom's nucleus and tightly bound core electrons. The dynamic behavior of the electron can be established from the electron wave function, ψ, which is obtained by solving the time-independent Schrödinger equation

$$\nabla^2\psi + \frac{2m}{\hbar^2}[E - U(\vec{r})]\psi = 0 \tag{3.4}$$

where m is electron mass, h is the reduced Planck constant, E is the energy of the electron, and $U(\vec{r})$ is the periodic potential energy inside the semiconductor. Solving this quantum-mechanical equation is beyond the scope of this work, but suffice it to say that the solution defines the band structure (the allowed electron energies and the relationship between the electron's energy and momentum) of the semiconductor and, amazingly, tells us that the quantum mechanically computed motion of the electron in the crystal is, to a good approximation, like that of an electron in free space if its mass, m, is replaced by an effective mass, m^*, in Newton's law of motion from classical mechanics

$$F = m^* a \tag{3.5}$$

where F is the applied force and a is the acceleration of the electron. A simplified energy band structure is illustrated in Figure 3.3. The allowed electron energies are plotted against the crystal momentum, $p = \hbar k$, where k is the wave vector (represented here as a scalar for simplicity) corresponding to the wave function solutions of the Schrödinger equation. Only the energy bands of immediate interest are shown – energy bands below the valence band are presumed to be fully occupied by electrons and those above the conduction band are presumed to be empty. The electron effective mass is thus defined as

$$m^* \equiv \left[\frac{\mathrm{d}^2E}{\mathrm{d}p^2}\right]^{-1} = \left[\frac{1}{\hbar^2}\frac{\mathrm{d}^2E}{\mathrm{d}k^2}\right]^{-1} \tag{3.6}$$

Notice that the effective mass is not constant within each band. In addition, near the top of the valence band, the effective mass is actually negative. Electrons (*) fill the states from bottom to top and the states near the top of the valence band are empty (o) due to some electrons being thermally excited into the conduction band. These empty states can conveniently be regarded as positively charged carriers of current called *holes* with a positive effective mass. It is conceptually much easier to deal with a relatively few number of holes that have a positive effective mass since they will behave like classical positively charged particles. The top of the valence band and the bottom of the conduction band are approximately parabolic in shape and therefore the electron effective mass (m_n^*)

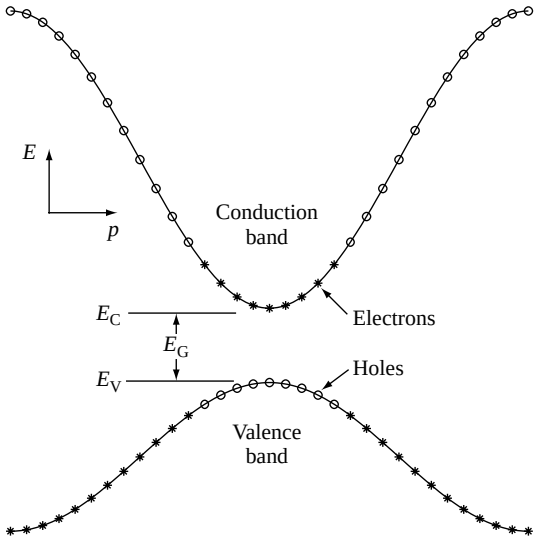


Figure 3.3 A simplified energy band diagram at $T > 0$ K for a direct band gap (E_G) semiconductor. Electrons near the maxima in valence band have been thermally excited to the empty states near the conduction-band minima, leaving behind holes. The excited electrons and remaining holes are the negative and positive mobile charges that give semiconductors their unique transport properties

near the bottom of the conduction band is a constant, as is the hole effective mass (m_p^*) near the top of the valence band. This is a very practical assumption that greatly simplifies the analysis of semiconductors.

When the minimum of the conduction band occurs at the same value of the crystal momentum as the maximum of the valence band, as it does in Figure 3.3, the semiconductor is a *direct band gap* semiconductor. When they do not align, the semiconductor is said to be an *indirect band gap* semiconductor. This is especially important when the absorption of light by a semiconductor is considered later in this chapter.

Even amorphous materials exhibit a similar band structure. Over short distances, the atoms are arranged in a periodic manner and an electron wave function can be defined. The wave functions from these small regions overlap in such a way that a *mobility gap* can be defined with electrons above the mobility gap defining the conduction band and holes below the gap defining the valence band. Unlike crystalline materials, however, there are a large number of localized energy states within the mobility gap (band tails and dangling bonds) that complicate the analysis of devices fabricated from these materials. Amorphous silicon (a-Si) solar cells are discussed in Chapter 12.

3.2.3 Conduction-band and Valence-band Densities of State

Now that the dynamics of the electron motion in a semiconductor has been approximated by a negatively charged particle with mass m_n^* in the conduction band and by a positively charged particle with mass m_p^* in the valence band, it is possible to calculate the density

of states in each band. This again involves solving the time-independent Schrödinger equation for the wave function of a particle in a box, but in this case the box is empty. All the complexities of the periodic potentials of the component atoms have been incorporated into the effective mass. The density of states in the conduction band is given by

$$g_{\mathrm{C}}(E) = \frac{m_n^* \sqrt{2m_n^*(E - E_{\mathrm{C}})}}{\pi^2 \hbar^3} \mathrm{cm}^{-3}\,\mathrm{eV}^{-1} \tag{3.7}$$

while the density of states in the valence band is given by

$$g_{\mathrm{V}}(E) = \frac{m_p^* \sqrt{2m_p^*(E_{\mathrm{V}} - E)}}{\pi^2 \hbar^3} \mathrm{cm}^{-3}\,\mathrm{eV}^{-1} \tag{3.8}$$

3.2.4 Equilibrium Carrier Concentrations

When the semiconductor is in thermal equilibrium (i.e. at a constant temperature with no external injection or generation of carriers), the Fermi function determines the ratio of filled states to available states at each energy and is given by

$$f(E) = \frac{1}{1 + \mathrm{e}^{(E - E_{\mathrm{F}})/kT}} \tag{3.9}$$

where E_{F} is the Fermi energy, k is Boltzmann's constant, and T is the Kelvin temperature. As seen in Figure 3.4, the Fermi function is a strong function of temperature. At absolute zero, it is a step function and all the states below E_{F} are filled with electrons and all those above E_{F} are completely empty. As the temperature increases, thermal excitation will leave some states below E_{F} empty, and the corresponding number of states above E_{F} will be filled with the excited electrons.

The equilibrium electron and hole concentrations (#/cm^3) are therefore

$$n_o = \int_{E_{\mathrm{C}}}^{\infty} g_{\mathrm{C}}(E) f(E)\,\mathrm{d}E = \frac{2N_{\mathrm{C}}}{\sqrt{\pi}} F_{1/2}((E_{\mathrm{F}} - E_{\mathrm{C}})/kT) \tag{3.10}$$

$$p_o = \int_{-\infty}^{E_{\mathrm{V}}} g_{\mathrm{V}}(E)[1 - f(E)]\,\mathrm{d}E = \frac{2N_{\mathrm{V}}}{\sqrt{\pi}} F_{1/2}((E_{\mathrm{V}} - E_{\mathrm{F}})/kT) \tag{3.11}$$

where $F_{1/2}(\xi)$ is the Fermi–Dirac integral of order 1/2,

$$F_{1/2}(\xi) = \int_0^{\infty} \frac{\sqrt{\xi'}\,\mathrm{d}\xi'}{1 + \mathrm{e}^{\xi' - \xi}} \tag{3.12}$$

The conduction-band and valence-band effective densities of state (#/cm^3), N_{C} and N_{V}, respectively, are given by

$$N_{\mathrm{C}} = 2\left(\frac{2\pi m_n^* kT}{h^2}\right)^{3/2} \tag{3.13}$$

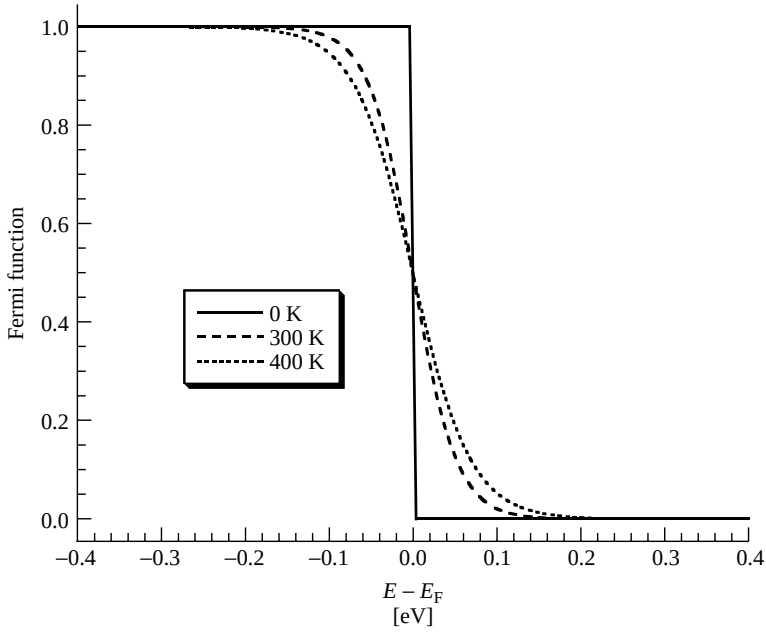


Figure 3.4 The Fermi function at various temperatures

and

$$N_{\mathrm{V}} = 2\left(\frac{2\pi m_p^* kT}{h^2}\right)^{3/2}. \tag{3.14}$$

When the Fermi energy, E_{F}, is sufficiently far ($>3\ kT$) from either bandedge, the carrier concentrations can be approximated (to within 2%) as [7]

$$n_o = N_{\mathrm{C}} \mathrm{e}^{(E_{\mathrm{F}} - E_{\mathrm{C}})/kT} \tag{3.15}$$

and

$$p_o = N_{\mathrm{V}} \mathrm{e}^{(E_{\mathrm{V}} - E_{\mathrm{F}})/kT} \tag{3.16}$$

and the semiconductor is said to be *nondegenerate*. In nondegenerate semiconductors, the product of the equilibrium electron and hole concentrations is independent of the location of the Fermi energy and is just

$$p_o n_o = n_{\mathrm{i}}^2 = N_{\mathrm{C}} N_{\mathrm{V}} \mathrm{e}^{(E_{\mathrm{V}} - E_{\mathrm{C}})/kT} = N_{\mathrm{C}} N_{\mathrm{V}} \mathrm{e}^{-E_{\mathrm{G}}/kT}. \tag{3.17}$$

In an undoped (intrinsic) semiconductor in thermal equilibrium, the number of electrons in the conduction band and the number of holes in the valence band are equal; $n_o = p_o = n_{\mathrm{i}}$, where n_{i} is the intrinsic carrier concentration. The intrinsic carrier concentration can be computed from (3.17), giving

$$n_{\mathrm{i}} = \sqrt{N_{\mathrm{C}} N_{\mathrm{V}}} \mathrm{e}^{(E_{\mathrm{V}} - E_{\mathrm{C}})/2kT} = \sqrt{N_{\mathrm{C}} N_{\mathrm{V}}} \mathrm{e}^{-E_{\mathrm{G}}/2kT}. \tag{3.18}$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 805

The Fermi energy in an intrinsic semiconductor, $E_i = E_F$, is given by

$$E_i = \frac{E_V + E_C}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right) \tag{3.19}$$

which is typically very close to the middle of the band gap. The intrinsic carrier concentration is typically very small compared to the densities of states and typical doping densities ($n_i \approx 10^{10}$ cm^{-3} in Si) and intrinsic semiconductors behave very much like insulators; that is, they are not very useful as conductors of electricity.

The number of electrons and holes in their respective bands, and hence the conductivity of the semiconductor, can be controlled through the introduction of specific impurities, or dopants, called *donors* and *acceptors*. For example, when semiconductor silicon is doped with phosphorous, one electron is donated to the conduction band for each atom of phosphorous introduced. From Table 3.1, it can be seen that phosphorous is in column V of the periodic table of elements and thus has five valence electrons. Four of these are used to satisfy the four covalent bonds of the silicon lattice and the fifth is available to fill an empty state in the conduction band. If silicon is doped with boron (valency of three, since it is in column III), each boron atom accepts an electron from the valence band, leaving behind a hole. All impurities introduce additional localized electronic states into the band structure, often within the forbidden band between E_C and E_V, as illustrated in Figure 3.5. If the energy of the state, E_D, introduced by a donor atom is sufficiently close to the conduction bandedge (within a few kT), there will be sufficient thermal energy to allow the extra electron to occupy a state in the conduction band. The donor state will then be positively charged (ionized) and must be considered when analyzing the electrostatics of the situation. Similarly, an acceptor atom will introduce a negatively charged (ionized) state at energy E_A. The controlled introduction of donor and acceptor impurities into a semiconductor allows the creation of the n-type (electrons are the primary source of electrical conduction) and p-type (holes are the primary source of electrical conduction) semiconductors, respectively. This is the basis for the construction of all semiconductor devices, including solar cells. The number of ionized donors and acceptors are given by [7]

$$N_D^+ = \frac{N_D}{1 + g_D e^{(E_F - E_D)/kT}} = \frac{N_D}{1 + e^{(E_F - E_D')/kT}} \tag{3.20}$$

and

$$N_A^- = \frac{N_A}{1 + g_A e^{(E_A - E_F)/kT}} = \frac{N_A}{1 + e^{(E_A' - E_F)/kT}} \tag{3.21}$$

where g_D and g_A are the donor and acceptor site degeneracy factors. Typically, $g_D = 2$ and $g_A = 4$. These factors are normally combined into the donor and the acceptor energies so that $E_D' = E_D - kT \ln g_D$ and $E_A' = E_A + kT \ln g_A$. Often, the donors and acceptors are assumed to be completely ionized so that $n_o \simeq N_D$ in n-type material and $p_o = N_A$ in p-type material. The Fermi energy can then be written as

$$E_F = E_i + kT \ln \frac{N_D}{n_i} \tag{3.22}$$

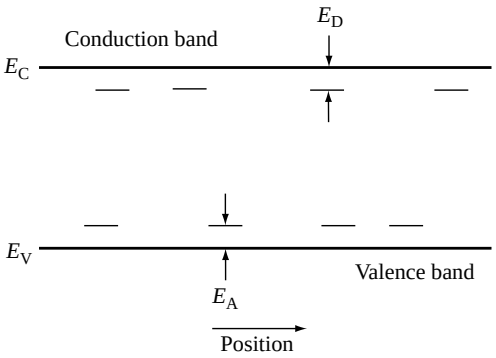


Figure 3.5 Donor and acceptor levels in a semiconductor. The nonuniform spatial distribution of these states reinforces the concept that these are localized states

in n-type material and as

$$E_F = E_i - kT \ln \frac{N_A}{n_i} \tag{3.23}$$

in p-type material.

When a very large concentration of dopants is introduced into the semiconductor, the dopants can no longer be thought of as a minor perturbation to the system. Their effect on the band structure must be considered. Typically, this so-called heavy doping effect manifests itself as a reduction in the band gap, E_G, and thus an increase in the intrinsic carrier concentration, as can be seen from equation (3.18). This band gap narrowing (BGN) [8] is detrimental to solar cell performance and solar cells are typically designed to avoid this effect, though it may be a factor in the heavily doped regions near the solar cell contacts.

3.2.5 Light Absorption

The creation of electron–hole pairs via the absorption of sunlight is fundamental to the operation of solar cells. The excitation of an electron directly from the valence band (which leaves a hole behind) to the conduction band is called *fundamental absorption*. Both the total energy and momentum of all particles involved in the absorption process must be conserved. Since the photon momentum, $p_\lambda = h/\lambda$, is very small compared to the range of the crystal momentum, $p = h/\ell$, the photon absorption process must, for practical purposes, conserve the momentum of the electron.[1] The absorption coefficient for a given photon energy, $h\nu$, is proportional to the probability, P_{12}, of the transition of an electron from the initial state E_1 to the final state E_2, the density of electrons in the initial state, $g_V(E_1)$, and the density of available final states, and is then summed over

[1] The wavelength of sunlight, λ, is on the order of a micron (10^{-4} cm), while the lattice constant is a few angstroms (10^{-8} cm). Thus, the crystal momentum is several orders of magnitude larger than the photon momentum.

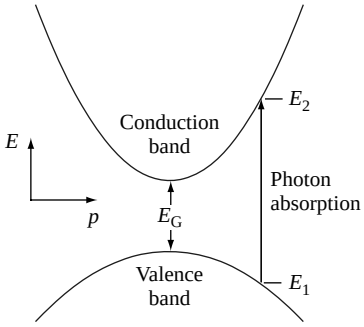


Figure 3.6 Photon absorption in a direct band gap semiconductor for an incident photon with energy $h\nu = E_2 - E_1 > E_G$

all possible transitions between states where $E_2 - E_1 = h\nu$ [9],

$$\alpha(h\nu) \propto \sum P_{12} g_V(E_1) g_C(E_2), \tag{3.24}$$

assuming that all the valence-band states are full and all the conduction-band states are empty. Absorption results in creation of an electron-hole pair since a free electron is excited to the conduction band leaving a free hole in the valence band.

In direct band gap semiconductors, such as GaAs, GaInP, CdTe, and Cu(InGa)Se_2, the basic photon absorption process is illustrated in Figure 3.6. Both energy and momentum must be conserved in the transition. Every initial electron state with energy E_1 and crystal momentum p_1 in the valence band is associated with a final state in the conduction band at energy E_2 and crystal momentum p_2. Since the electron momentum is conserved, the crystal momentum of the final state is the same as the initial state, $p_1 \approx p_2 = p$.

Conservation of energy dictates that the energy of the absorbed photon is

$$h\nu = E_2 - E_1 \tag{3.25}$$

Since we have assumed parabolic bands,

$$E_V - E_1 = \frac{p^2}{2m_p^*} \tag{3.26}$$

and

$$E_2 - E_C = \frac{p^2}{2m_n^*} \tag{3.27}$$

Combining equations (3.25), (3.26), and (3.27) yields

$$h\nu - E_G = \frac{p^2}{2}\left(\frac{1}{m_n^*} + \frac{1}{m_p^*}\right) \tag{3.28}$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 808

and the absorption coefficient for direct transitions is [9]

$$\alpha(h\nu) \approx A^*(h\nu - E_G)^{1/2}, \tag{3.29}$$

where A^* is a constant. In some semiconductor materials, quantum selection rules do not allow transitions at $p = 0$ but allow them for $p \neq 0$. In such cases [9]

$$\alpha(h\nu) \approx \frac{B^*}{h\nu}(h\nu - E_G)^{3/2}, \tag{3.30}$$

where B^* is a constant.

In indirect band gap semiconductors like Si and Ge, where the valence-band maximum occurs at a different crystal momentum than the conduction-band minimum, conservation of electron momentum necessitates that the photon absorption process involve an additional particle. Phonons, the particle representation of lattice vibrations in the semiconductor, are suited to this process because they are low-energy particles with relatively high momentum. This is illustrated in Figure 3.7. Notice that light absorption is facilitated by either phonon absorption or phonon emission. The absorption coefficient, when there is phonon absorption, is given by

$$\alpha_a(h\nu) = \frac{A(h\nu - E_G + E_{ph})^2}{e^{E_{ph}/kT} - 1} \tag{3.31}$$

and by

$$\alpha_e(h\nu) = \frac{A(h\nu - E_G - E_{ph})^2}{1 - e^{-E_{ph}/kT}} \tag{3.32}$$

when a phonon is emitted [9]. Because both processes are possible,

$$\alpha(h\nu) = \alpha_a(h\nu) + \alpha_e(h\nu). \tag{3.33}$$

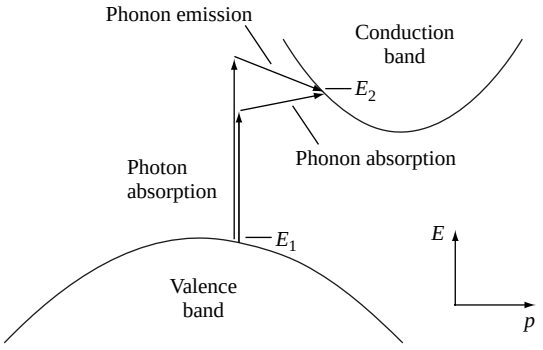


Figure 3.7 Photon absorption in an indirect band gap semiconductor for a photon with energy $h\nu < E_2 - E_1$ and a photon with energy $h\nu > E_2 - E_1$. Energy and momentum in each case are conserved by the absorption and emission of a phonon, respectively

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 809

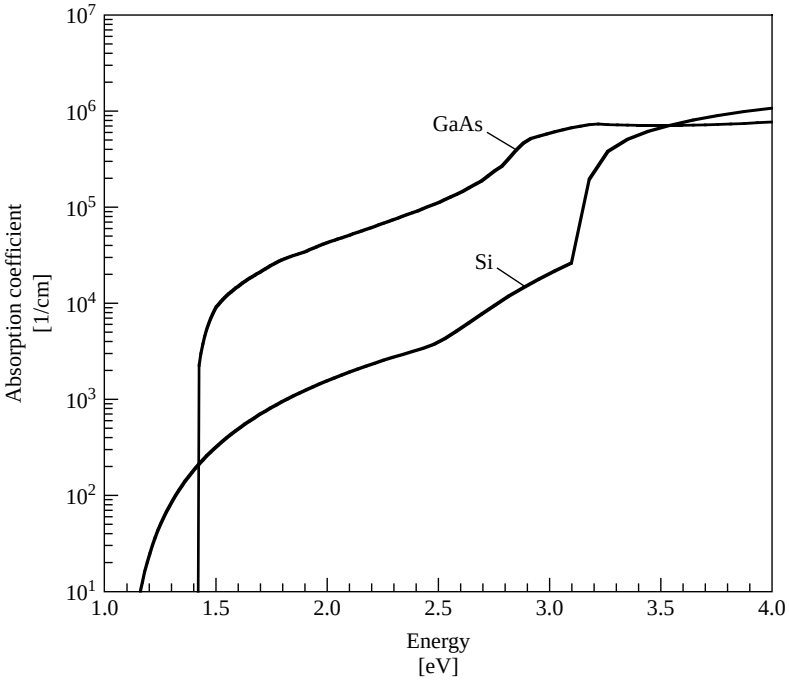


Figure 3.8 Absorption coefficient as a function of photon energy for Si (indirect band gap) and GaAs (direct band gap) at 300 K. Their band gaps are 1.12 and 1.4 eV, respectively

Since both a phonon and an electron are needed to make the indirect gap absorption process possible, the absorption coefficient depends not only on the density of full initial electron states and empty final electron states but also on the availability of phonons (both emitted and absorbed) with the required momentum. Thus, compared to direct transitions, the absorption coefficient for indirect transitions is relatively small. As a result, light penetrates more deeply into indirect band gap semiconductors than direct band gap semiconductors. This is illustrated in Figure 3.8 for Si, an indirect band gap semiconductor, and GaAs, a direct band gap semiconductor. Similar spectra are shown for other semiconductors elsewhere in this handbook.

In both direct band gap and indirect band gap materials, a number of photon absorption processes are involved, though the mechanisms described above are the dominant ones. A direct transition, without phonon assistance, is possible in indirect band gap materials if the photon energy is high enough (as seen in Figure 3.8 for Si at about 3.3 eV). Conversely, in direct band gap materials, phonon-assisted absorption is also a possibility. Other mechanisms may also play a role in defining the absorption process in semiconductors. These include absorption in the presence of an electric field (the Franz–Keldysh effect), absorption aided by localized states in the forbidden gap, and degeneracy effects when a significant number of states in the conduction band are not empty and/or when a significant number of state in the valence band are not full, as can happen in heavily doped materials (BGN) and under high-level injection (the Burstein–Moss shift). The net absorption coefficient is then the sum of the absorption coefficients due to all absorption

processes or

$$\alpha(h\nu) = \sum_{\mathrm{i}} \alpha_{\mathrm{i}}(h\nu). \tag{3.34}$$

In practice, measured absorption coefficients or empirical expressions for the absorption coefficient are used in analysis and modeling. The rate of creation of electron–hole pairs (# of electron–hole pairs per cm^3 per second) as a function of position within a solar cell is

$$G(x) = (1-s)\int_{\lambda} (1-r(\lambda))f(\lambda)\alpha(\lambda)\mathrm{e}^{-\alpha x}\,\mathrm{d}\lambda \tag{3.35}$$

where s is the grid-shadowing factor, $r(\lambda)$ is the reflectance, $\alpha(\lambda)$ is the absorption coefficient, $f(\lambda)$ is the incident photon flux (number of photons incident per unit area per second per wavelength), and the sunlight is assumed to be incident at $x = 0$. Here, the absorption coefficient has been cast in terms of the light's wavelength through the relationship $h\nu = hc/\lambda$. The photon flux, $f(\lambda)$, is obtained by dividing the incident power density at each wavelength by the photon energy.

Free-carrier absorption, in which electrons in the conduction band absorb the energy of a photon and move to an empty state higher in the conduction band (correspondingly for holes in the valence band), is typically only significant for photons with $E < E_{\mathrm{G}}$ since the free-carrier absorption coefficient increases with increasing wavelength,

$$\alpha_{fc} \propto \lambda^{\gamma} \tag{3.36}$$

where $1.5 < \gamma < 3.5$ [9]. Thus, in single-junction solar cells, it does not affect the creation of electron–hole pairs and can be ignored (although free-carrier absorption can be exploited to probe the excess carrier concentrations in solar cells for the purpose of determining recombination parameters [10]). However, free-carrier absorption is a consideration in tandem solar cell systems in which a wide band gap (E_{G1}) solar cell is stacked on top of a solar cell of smaller band gap ($E_{\mathrm{G2}} < E_{\mathrm{G1}}$). Photons with energy too low to be absorbed in the top cell ($h\nu < E_{\mathrm{G1}}$) will be transmitted to the bottom cell and be absorbed there (if $h\nu > E_{\mathrm{G2}}$). Of course, more solar cells can be stacked as long as $E_{\mathrm{G1}} > E_{\mathrm{G2}} > E_{\mathrm{G3}} \ldots$, and so on. The number of photons transmitted to the next cell in the stack will be reduced by whatever amount of free-carrier absorption occurs. Tandem cells are discussed more completely in Chapters 9 and 12.

3.2.6 Recombination

When a semiconductor is taken out of thermal equilibrium, for instance by illumination and/or injection of current, the concentrations of electrons (n) and holes (p) tend to relax back toward their equilibrium values through a process called *recombination* in which an electron falls from the conduction band to the valence band, thereby eliminating a valence-band hole. There are several recombination mechanisms important to the operation of solar cells – recombination through traps (defects) in the forbidden gap, radiative (band-to-band) recombination, and Auger recombination – that will be discussed here. These three processes are illustrated in Figure 3.9.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 811

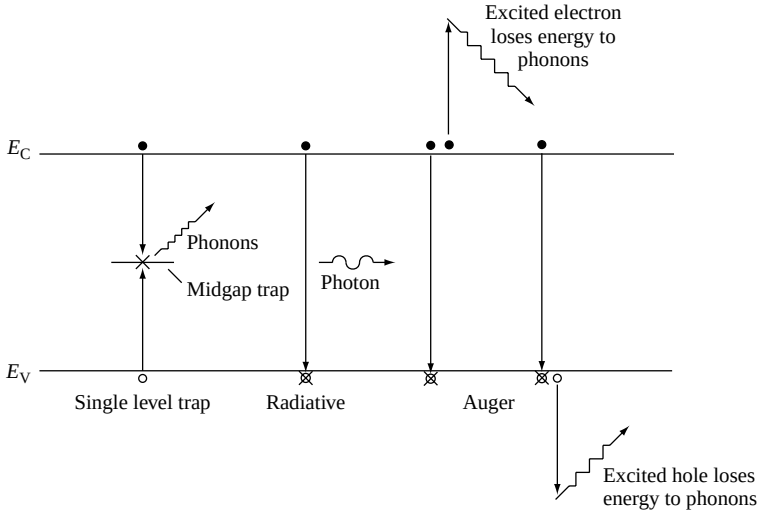


Figure 3.9 Recombination processes in semiconductors

The net recombination rate per unit volume per second through a single level trap (SLT) located at energy $E = E_T$ within the forbidden gap, also commonly referred to as *Shockley–Read–Hall recombination,* is given by [11]

$$R_{SLT} = \frac{pn - n_i^2}{\tau_{SLT,n}(p + n_i e^{(E_i - E_T)/kT}) + \tau_{SLT,p}(n + n_i e^{(E_T - E_i)/kT})} \tag{3.37}$$

where the carrier lifetimes are given by

$$\tau_{SLT} = \frac{1}{\sigma v_{th} N_T} \tag{3.38}$$

where σ is the capture cross section, v_{th} is the thermal velocity of the carriers, and N_T is the concentration of traps. The capture cross section can be thought of as the size of the target present to a carrier traveling through the semiconductor at velocity v_{th}. Small lifetimes correspond to high rates of recombination. If a trap presents a large target to the carrier, the recombination rate will be high (low carrier lifetime). When the velocity of the carrier is high, it has more opportunity within a given time period to encounter a trap and the carrier lifetime is low. Finally, the probability of interaction with a trap increases as the concentration of traps increases and the carrier lifetime is therefore inversely proportional to the trap concentration.

Some reasonable assumptions allow equation (3.37) to be simplified. If the material is p-type ($p \approx p_o \gg n_o$), in low injection ($n_o \leq n \ll p_o$), and the trap energy is near the middle of the forbidden gap ($E_T \approx E_i$), the recombination rate can be written as

$$R_{SLT} \approx \frac{n - n_o}{\tau_{SLT,n}}. \tag{3.39}$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 812

Notice that the recombination rate is solely dependent on the minority carrier (also called the limiting carrier). This is reasonable since there are far fewer minority carriers than majority carriers and one of each is necessary for there to be recombination.

If high-injection conditions prevail ($p \approx n \gg p_o, n_o$),

$$R_{\text{SLT}} \approx \frac{n}{\tau_{\text{SLT},p} + \tau_{\text{SLT},n}} \approx \frac{p}{\tau_{\text{SLT},p} + \tau_{\text{SLT},n}}. \tag{3.40}$$

In this case, the effective recombination lifetime is the sum of the two carrier lifetimes. While the recombination rate is high due to the large number of excess holes and electrons, the carrier lifetime is actually longer than in the case of low injection. This can be of significance in the base region of solar cells, especially concentrator cells (solar cells illuminated with concentrated sunlight), since the base is the least doped layer.

Radiative (band-to-band) recombination is simply the inverse of the optical generation process and is much more efficient in direct band gap semiconductors than in indirect band gap semiconductors. When radiative recombination occurs, the energy of the electron is given to an emitted photon – this is how semiconductor lasers and light emitting diodes (LEDs) operate. In an indirect band gap material, some of that energy is shared with a phonon. The net recombination rate due to radiative processes is given as

$$R_\lambda = \text{B}(pn - n_\text{i}^2). \tag{3.41}$$

If we have an n-type ($n \approx n_o \gg p_o$) semiconductor in low injection ($p_o \leq p \ll n_o$), the net radiative recombination rate can be written in terms of an effective lifetime, $\tau_{\lambda,p}$,

$$R_\lambda \approx \frac{p - p_o}{\tau_{\lambda,p}} \tag{3.42}$$

where

$$\tau_{\lambda,p} = \frac{1}{n_o \text{B}}. \tag{3.43}$$

A similar expression can be written for p-type semiconductors.

Auger recombination is somewhat similar to radiative recombination, except that the energy of transition is given to another carrier (in either the conduction band or the valence band), as shown in Figure 3.9. This electron (or hole) then relaxes thermally (releasing its excess energy and momentum to phonons). Just as radiative recombination is the inverse process to optical absorption, Auger recombination is the inverse process to *impact ionization*, where an energetic electron collides with a crystal atom, breaking the bond and creating an electron–hole pair. The net recombination rate due to Auger processes is

$$R_{\text{Auger}} = (\Lambda_n n + \Lambda_p p)(pn - n_\text{i}^2) \tag{3.44}$$

In an n-type material in low injection (and assuming Λ_n and Λ_p are of comparable magnitudes), the net Auger recombination rate becomes

$$R_{\text{Auger}} \approx \frac{p - p_o}{\tau_{\text{Auger},p}} \tag{3.45}$$

with

$$\tau_{\text{Auger},p} = \frac{1}{\Lambda_n n_o^2}. \tag{3.46}$$

A similar expression can be derived for minority electron lifetime in p-type material.

Each of these recombination processes occurs in parallel and there can be multiple and/or distributed traps[2] in the forbidden gap; thus the total recombination rate is the sum of rates due to each process

$$R = \left[\sum_{\text{traps } i} R_{\text{SLT},i}\right] + R_\lambda + R_{\text{Auger}}. \tag{3.47}$$

An effective minority-carrier lifetime for a doped material in low-level injection is given as

$$\frac{1}{\tau} = \left[\sum_{\text{traps } i} \frac{1}{\tau_{\text{SLT},i}}\right] + \frac{1}{\tau_\lambda} + \frac{1}{\tau_{\text{Auger}}}. \tag{3.48}$$

The distribution of traps in the energy gap for specific materials is given in other chapters.

Interfaces between two dissimilar materials, such as, those that occur at the front surface of a solar cell, have a high concentration defect due to the abrupt termination of the crystal lattice. These manifest themselves as a continuum of traps within the forbidden gap at the surface; electrons and holes can recombine through them just as with bulk traps. This is illustrated in Figure 3.10. Rather than giving a recombination rate per unit volume per second, surface traps give a recombination rate per unit area per second. A general expression for surface recombination is [11]

$$R_{\text{S}} = \int_{E_{\text{V}}}^{E_{\text{C}}} \frac{pn - n_{\text{i}}^2}{(p + n_{\text{i}} e^{(E_{\text{i}} - E_{\text{t}})/kT})/s_n + (n + n_{\text{i}} e^{(E_{\text{t}} - E_{\text{i}})/kT})/s_p} D_\Pi(E_{\text{t}})\, \text{d}E_{\text{t}} \tag{3.49}$$

where E_{t} is the trap energy, $D_\Pi(E_{\text{t}})$ is the surface state concentration (the concentration of traps is probably dependent on the trap energy), and s_n and s_p are surface recombination velocities, analogous to the carrier lifetimes for bulk traps. The surface recombination rate is generally written, for simplicity, as [11]

$$R_{\text{S}} = S_p(p - p_o) \tag{3.50}$$

in n-type material and as

$$R_{\text{S}} = S_n(n - n_o) \tag{3.51}$$

in p-type material. S_p and S_n are effective surface recombination velocities. It should be mentioned that these effective recombination velocities are not necessarily constants, though they are usually treated as such.

[2] It is unlikely that more than one trap will be involved in a single recombination event since the traps are spatially separated.

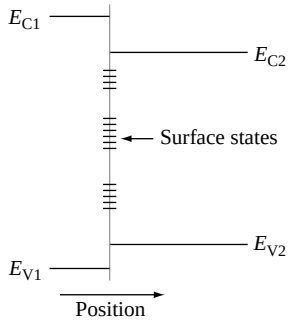


Figure 3.10 Illustration of surface states at a semiconductor surface or interface between dissimilar materials such as two different semiconductors (heterojunction) or a metal and a semiconductor (Schottky contact)

3.2.7 Carrier Transport

As has already been established, electrons and holes in a semiconductor behave much like a free particle of the same electronic charge with effective masses of m_n^* and m_p^*, respectively. Thus, they are subject to the classical processes of drift and diffusion. Drift is a charged particle's response to an applied electric field. When an electric field is applied across a uniformly doped semiconductor, the bands bend upward in the direction of the applied electric field. Electrons in the conduction band, being negatively charged, move in the opposite direction of the applied field and holes in the valence band, being positively charged, move in the same direction of the applied field (Figure 3.11) – in other words, electrons *sink* and holes *float*. This is a useful conceptual tool for analyzing the motion of holes and electrons in semiconductor devices. With nothing to impede their motion, the holes and electrons would continue to accelerate without bound. However, the semiconductor crystal is full of objects with which the carriers collide and are scattered. These objects include the component atoms of the crystal, dopant ions, crystal defects, and even other electrons and holes. On a microscopic scale, their motion is much like that of a ball in pinball machine, the carriers are constantly bouncing (scattering) off objects in the crystal, but generally moving in the direction prescribed by the applied electric field, $\vec{E} = -\nabla\phi$, where ϕ is the electrostatic potential. The net effect is that the carriers appear to move, on a macroscopic scale, at a constant velocity, v_{d}, the drift velocity. The drift velocity is directly proportional to the electric field

$$|\vec{v}_d| = |\mu\vec{E}| = |\mu\nabla\phi| \tag{3.52}$$

where μ is the carrier mobility. The carrier mobility is generally independent of the electric field strength unless the field is very strong, a situation not typically encountered in solar cells. The drift current densities for holes and electrons can be written as

$$\vec{J}_p^{\,\mathrm{drift}} = qp\vec{v}_{\mathrm{d},p} = q\mu_p p\vec{E} = -q\mu_p p\nabla\phi \tag{3.53}$$

and

$$\vec{J}_n^{\,\mathrm{drift}} = -qn\vec{v}_{\mathrm{d},n} = q\mu_n n\vec{E} = -q\mu_n n\nabla\phi. \tag{3.54}$$

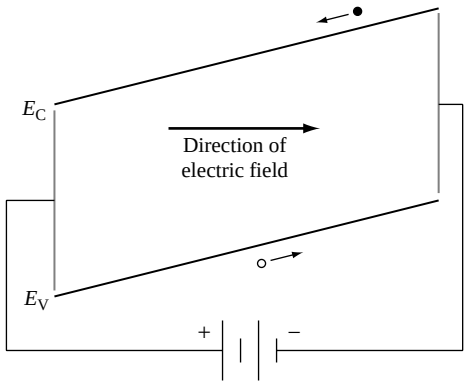


Figure 3.11 Illustration of the concept of drift in a semiconductor. Note that electrons and holes move in opposite directions. The electric field can be created by the internal built-in potential of the junction or by an externally applied bias voltage

The most significant scattering mechanisms in solar cells are lattice (phonon) and ionized impurity scattering. These component mobilities can be written as

$$\mu_L = C_L T^{-3/2} \tag{3.55}$$

for lattice scattering and as

$$\mu_I = \frac{C_I T^{3/2}}{N_D^+ + N_A^-} \tag{3.56}$$

for ionized impurity scattering. These can then be combined using Mathiessen's rule to give the carrier mobility [12]

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I}. \tag{3.57}$$

This is a first-order approximation that neglects the velocity dependencies of the scattering mechanisms. These two types of mobility can be distinguished experimentally by their different dependencies on temperature and doping. A better approximation is [12]

$$\mu = \mu_L \left[1 + \left(\frac{6\mu_L}{\mu_I} \right) \left(\mathrm{Ci}\left(\frac{6\mu_L}{\mu_I} \right) \cos\left(\frac{6\mu_L}{\mu_I} \right) + \left[\mathrm{Si}\left(\frac{6\mu_L}{\mu_I} \right) - \frac{\pi}{2} \right] \sin\left(\frac{6\mu_L}{\mu_I} \right) \right) \right], \tag{3.58}$$

where Ci and Si (not to be confused with the abbreviation for silicon) are the cosine and sine integrals, respectively.

When modeling solar cells, it is more convenient to use measured data or empirical formulas. Carrier mobilities in Si at 300 K are well approximated by [12]

$$\mu_n = 92 + \frac{1268}{1 + \left(\dfrac{N_D^+ + N_A^-}{1.3 \times 10^{17}} \right)^{0.91}} \mathrm{cm}^2/\mathrm{V\text{-}s} \tag{3.59}$$

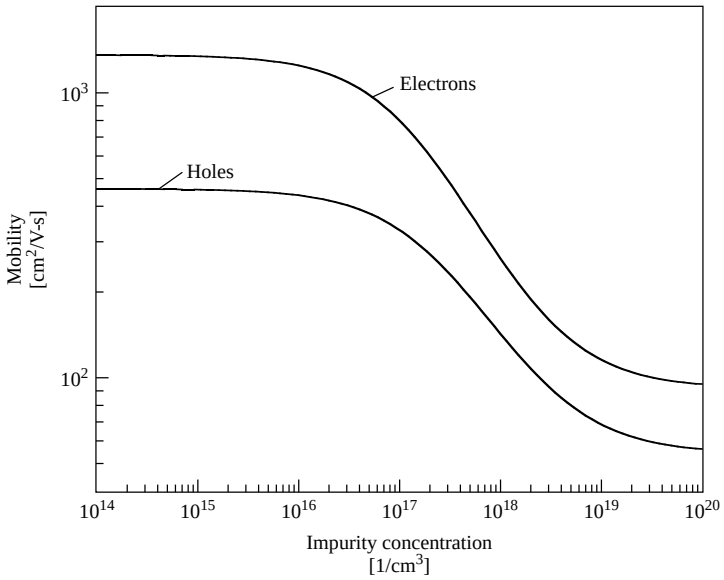


Figure 3.12 Electron and hole mobilities in silicon for $T = 300$ K

$$\mu_p = 54.3 + \frac{406.9}{1 + \left(\dfrac{N_{\rm D}^+ + N_{\rm A}^-}{2.35 \times 10^{17}}\right)^{0.88}} \text{cm}^2/\text{V-s} \tag{3.60}$$

and are plotted in Figure 3.12. At low impurity levels, the mobility is governed by intrinsic lattice scattering, while at high levels the mobility is governed by ionized impurity scattering.

Electrons and holes in semiconductors tend, as a result of their random thermal motion, to move (diffuse) from regions of high concentration to regions of low concentration. Much like how the air in a balloon is distributed evenly within the volume of the balloon, carriers, in the absence of any external forces, will also tend to distribute themselves evenly. This process is called *diffusion* and the diffusion current densities are given by

$$\vec{J}_p^{\,\rm diff} = -q D_p \nabla p \tag{3.61}$$

$$\vec{J}_n^{\,\rm diff} = q D_n \nabla n \tag{3.62}$$

where D_p and D_n are the hole and electron diffusion coefficients, respectively. Note that they are driven by the gradient of the carrier densities.

In thermal equilibrium, there can be no net hole current and no net electron current – in other words, the drift and diffusion currents must exactly balance. In nondegenerate materials, this leads to the Einstein relationship

$$\frac{D}{\mu} = \frac{kT}{q} \tag{3.63}$$

and allows the diffusion coefficient to be directly computed from the mobility. Generalized forms of the Einstein relationship, valid for degenerate materials, are

$$\frac{D_n}{\mu_n} = \frac{1}{q} n \left[\frac{\mathrm{d}n}{\mathrm{d}E_\mathrm{F}} \right]^{-1} \tag{3.64}$$

and

$$\frac{D_p}{\mu_p} = \frac{-1}{q} p \left[\frac{\mathrm{d}p}{\mathrm{d}E_\mathrm{F}} \right]^{-1}. \tag{3.65}$$

The diffusion coefficient actually increases when degeneracy effects come into play.

The total hole and electron currents (vector quantities) are the sum of their drift and diffusion components

$$\vec{J}_p = \vec{J}_p^{\,\mathrm{drift}} + \vec{J}_p^{\,\mathrm{diff}} = q\mu_p p\vec{E} - qD_p \nabla p = -q\mu_p p \nabla\phi - qD_p \nabla p \tag{3.66}$$

$$\vec{J}_n = \vec{J}_n^{\,\mathrm{drift}} + \vec{J}_n^{\,\mathrm{diff}} = q\mu_n n\vec{E} + qD_n \nabla n = -q\mu_n n \nabla\phi + qD_n \nabla n \tag{3.67}$$

The total current is then

$$\vec{J} = \vec{J}_p + \vec{J}_n + \vec{J}_\mathrm{disp} \tag{3.68}$$

where $\vec{J}_\mathrm{disp}$ is the *displacement current* given by

$$\vec{J}_\mathrm{disp} = \frac{\partial \vec{D}}{\partial t}. \tag{3.69}$$

$\vec{D} = \varepsilon\vec{E}$ is the dielectric displacement field, where ε is the electric permittivity of the semiconductor. The displacement current is typically neglected in solar cells since they are static (dc) devices.

3.2.8 Semiconductor Equations

The operation of most semiconductor devices, including solar cells, can be described by the so-called semiconductor device equations, first derived by Van Roosbroeck in 1950 [13]. A generalized form of these equations is given here.

$$\nabla \cdot \varepsilon\vec{E} = q(p - n + N) \tag{3.70}$$

This is a form of Poisson's equation, where N is the net charge due to dopants and other trapped charges. The hole and electron continuity equations are

$$\nabla \cdot \vec{J}_p = q\left(G - R_p - \frac{\partial p}{\partial t}\right) \tag{3.71}$$

$$\nabla \cdot \vec{J}_n = q\left(R_n - G + \frac{\partial n}{\partial t}\right) \tag{3.72}$$

where G is the optical generation rate of electron–hole pairs. Thermal generation is included in R_p and R_n. The hole and electron current densities are given by

$$\vec{J}_p = -q\mu_p p\nabla(\phi - \phi_p) - kT\mu_p \nabla p \tag{3.73}$$

and

$$\vec{J}_n = -q\mu_n n\nabla(\phi + \phi_n) + kT\mu_n \nabla n. \tag{3.74}$$

Two new terms, ϕ_p and ϕ_n, have been introduced here. These are the so-called band parameters that account for degeneracy and a spatially varying band gap and electron affinity [14]. These terms were ignored in the preceding discussion and can usually be ignored in nondegenerate homostructure solar cells.

The intent here is to derive an analytic expression for the current–voltage characteristic of a simple solar cell, and so some simplifications are in order. It should be noted, however, that a complete description of the operation of solar cells can be obtained by solving the complete set of coupled partial differential equations, equations (3.70) through (3.74). The numerical solution of these equations is addressed later in this chapter.

3.2.9 Minority-carrier Diffusion Equation

In a uniformly doped semiconductor, the band gap and electric permittivity are independent of position. Since the doping is uniform, the carrier mobilities and diffusion coefficients are also independent of position. As we are mainly interested in the steady state operation of the solar cell, the semiconductor equations reduce to

$$\frac{\mathrm{d}\vec{E}}{\mathrm{d}x} = \frac{q}{\varepsilon}(p - n + N_\mathrm{D} - N_\mathrm{A}) \tag{3.75}$$

$$q\mu_p \frac{\mathrm{d}}{\mathrm{d}x}(p\vec{E}) - qD_p \frac{\mathrm{d}^2 p}{\mathrm{d}x^2} = q(G - R) \tag{3.76}$$

and

$$q\mu_n \frac{\mathrm{d}}{\mathrm{d}x}(n\vec{E}) + qD_n \frac{\mathrm{d}^2 n}{\mathrm{d}x^2} = q(R - G) \tag{3.77}$$

In regions sufficiently far from the pn-junction of the solar cell (quasi-neutral regions), the electric field is very small. When considering the minority carrier (holes in the n-type material and electrons in the p-type material) and low-level injection ($\Delta p = \Delta n \ll N_\mathrm{D}, N_\mathrm{A}$), the drift current can be neglected with respect to the diffusion current. Under low-level injection, R simplifies to

$$R = \frac{n_P - n_{Po}}{\tau_n} = \frac{\Delta n_P}{\tau_n} \tag{3.78}$$

in the p-type region and to

$$R = \frac{p_N - p_{No}}{\tau_p} = \frac{\Delta p_N}{\tau_p} \tag{3.79}$$

in the n-type region. Δp_N and Δn_P are the excess minority-carrier concentrations. The minority-carrier lifetimes, τ_n and τ_p, are given by equation (3.48). For clarity, the capitalized subscripts, "P" and "N", are used to indicate quantities in p-type and n-type regions, respectively, when it may not be otherwise apparent. Lower-case subscripts, "p" and "n", refer to quantities associated with minority holes and electrons, respectively. Equations (3.76) and (3.77) thus each reduce to what is commonly referred to as the *minority-carrier diffusion equation*. It can be written as

$$D_p \frac{\mathrm{d}^2 \Delta p_N}{\mathrm{d}x^2} - \frac{\Delta p_N}{\tau_p} = -G(x) \tag{3.80}$$

in n-type material and as

$$D_n \frac{\mathrm{d}^2 \Delta n_P}{\mathrm{d}x^2} - \frac{\Delta n_P}{\tau_n} = -G(x) \tag{3.81}$$

in p-type material. For example, Δn_P is the minority electron concentration in the p-type material. The minority-carrier diffusion equation is often used to analyze the operation of semiconductor devices, including solar cells.

3.3 *PN*-JUNCTION DIODE ELECTROSTATICS

Where an n-type semiconductor comes into contact with a p-type semiconductor, a *pn*-junction is formed. In thermal equilibrium there is no net current flow and by definition the Fermi energy must be independent of position. Since there is a concentration difference of holes and electrons between the two types of semiconductors, holes diffuse from the p-type region into the n-type region and, similarly, electrons from the n-type material diffuse into the p-type region. As the carriers diffuse, the charged impurities (ionized acceptors in the p-type material and ionized donors in the n-type material) are uncovered – that is, no longer screened by the majority carrier. As these impurity charges are uncovered, an electric field (or electrostatic potential difference) is produced, which limits the diffusion of the holes and electrons. In thermal equilibrium, the diffusion and drift currents for each carrier type exactly balance, so there is no net current flow. The transition region between the n-type and the p-type semiconductors is called the *space-charge region*. It is also often called the *depletion region*, since it is effectively depleted of both holes and electrons. Assuming that the p-type and the n-type regions are sufficiently thick, the regions on either side of the depletion region are essentially charge-neutral (often termed *quasi-neutral*). The electrostatic potential difference resulting from the junction formation is called the *built-in voltage*, V_{bi}. It arises from the electric field created by the exposure of the positive and the negative space charge in the depletion region.

The electrostatics of this situation (assuming a single acceptor and a single donor level) are governed by Poisson's equation

$$\nabla^2 \phi = \frac{q}{\varepsilon}(n_o - p_o + N_{\mathrm{A}}^- - N_{\mathrm{D}}^+) \tag{3.82}$$

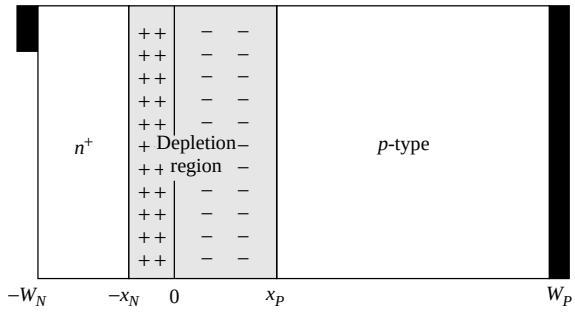


Figure 3.13 Simple solar cell structure used to analyze the operation of a solar cell. Free carriers have diffused across the junction ($x = 0$) leaving a space-charge or depletion region practically devoid of any free or mobile charges. The fixed charges in the depletion region are due to ionized donors on the n-side and ionized acceptors on the p-side

where ϕ is the electrostatic potential, q is magnitude of the electron charge, ε is the electric permittivity of the semiconductor, p_o is the equilibrium hole concentration, n_o is the equilibrium electron concentration, N_A^- is the ionized acceptor concentration, and N_D^+ is the ionized donor concentration. Equation 3.82 is a restatement of equation 3.70 for the given conditions.

This equation is easily solved numerically; however, an approximate analytic solution for an abrupt *pn*-junction can be obtained that lends physical insight into the formation of the space-charge region. Figure 3.13 depicts a simple one-dimensional (1D) *pn*-junction solar cell (diode), with the metallurgical junction at $x = 0$, which is uniformly doped N_D on the n-type side and N_A on the p-type side. For simplicity, it is assumed that the each side is nondegenerately doped and that the dopants are fully ionized.

Within the depletion region, defined by $-x_N < x < x_P$, it can be assumed that p_o and n_o are both negligible compared to $|N_A - N_D|$ so that equation (3.82) can be simplified to

$$\begin{aligned} \nabla^2\phi &= -\frac{q}{\varepsilon}N_D, \quad \text{for} - x_N < x < 0 \quad \text{and} \\ \nabla^2\phi &= \frac{q}{\varepsilon}N_A, \quad \text{for} \quad 0 < x < x_P \end{aligned} \tag{3.83}$$

Outside the depletion region, charge neutrality is assumed and

$$\nabla^2\phi = 0 \quad \text{for} \quad x \le -x_N \quad \text{and} \quad x \ge x_P. \tag{3.84}$$

This is commonly referred to as the *depletion approximation*. The regions on either side of the depletion regions are the quasi-neutral regions.

The electrostatic potential difference across the junction is the built-in voltage, V_{bi}, and can be obtained by integrating the electric field, $\vec{E} = -\nabla\phi$.

$$\int_{-x_N}^{x_P} \vec{E}\,dx = -\int_{-x_N}^{x_P} \frac{d\phi}{dx}dx = -\int_{V(-x_N)}^{V(x_P)} d\phi = \phi(-x_N) - \phi(x_P) = V_{bi} \tag{3.85}$$

Solving equations (3.83) and (3.84) and defining $\phi(x_P) = 0$, gives

$$\phi(x) = \begin{cases} V_{\text{bi}}, x \leq -x_N \\ V_{\text{bi}} - \dfrac{qN_{\text{D}}}{2\varepsilon}(x + x_N)^2, -x_N < x \leq 0 \\ \dfrac{qN_{\text{A}}}{2\varepsilon}(x - x_P)^2, 0 \leq x < x_P \\ 0, x \geq x_P \end{cases} \tag{3.86}$$

The electrostatic potential must be continuous at $x = 0$. Therefore, from equation (3.86),

$$V_{\text{bi}} - \frac{qN_{\text{D}}}{2\varepsilon}x_N^2 = \frac{qN_{\text{A}}}{2\varepsilon}x_P^2 \tag{3.87}$$

In the absence of any interface charge at the metallurgical junction, the electric field is also continuous at this point (really, it is the displacement field, $\vec{D} = \varepsilon\vec{E}$, but in this example, ε is independent of position), and

$$x_N N_{\text{D}} = x_P N_{\text{A}} \tag{3.88}$$

This is simply a statement that the total charge in either side of the depletion region exactly balance each other and therefore the depletion region extends furthest into the more lightly doped side.

Solving equations (3.87) and (3.88) for the depletion width, W_{D}, gives[3]

$$W_{\text{D}} = x_N + x_P = \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_{\text{A}} + N_{\text{D}}}{N_{\text{A}}N_{\text{D}}}\right)V_{\text{bi}}}. \tag{3.89}$$

Under nonequilibrium conditions, the electrostatic potential difference across the junction is modified by the applied voltage, V, which is zero in thermal equilibrium. As a consequence, the depletion width is dependent on the applied voltage,

$$W_{\text{D}}(V) = x_N + x_P = \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_{\text{A}} + N_{\text{D}}}{N_{\text{A}}N_{\text{D}}}\right)(V_{\text{bi}} - V)}. \tag{3.90}$$

As previously stated, the built-in voltage, V_{bi}, can be calculated by noting that under thermal equilibrium the net hole and electron currents are zero. The hole current density is

$$\vec{J}_p = q\mu_p p_o \vec{E} - qD_p \nabla p = 0. \tag{3.91}$$

[3] A somewhat more rigorous treatment of equation 3.89 would yield a factor of $2kT/q$ which is ~50 mV at 300 K, or

$$W_{\text{D}} = \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_{\text{A}} + N_{\text{D}}}{N_{\text{A}}N_{\text{D}}}\right)(V_{\text{bi}} - 2kT/q)} \quad [3].$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 822

Thus, in 1D, utilizing the Einstein relationship, the electric field can be written as

$$\vec{E} = \frac{kT}{q}\frac{1}{p_o}\frac{\mathrm{d}p_o}{\mathrm{d}x} \tag{3.92}$$

Rewriting equation (3.85) and substituting equation (3.92) yields

$$V_{\mathrm{bi}} = \int_{-x_N}^{x_P} E\vec{\mathrm{d}}x = \int_{-x_N}^{x_P} \frac{kT}{q}\frac{1}{p_o}\frac{\mathrm{d}p_o}{\mathrm{d}x}\mathrm{d}x = \frac{kT}{q}\int_{p_o(-x_N)}^{p_o(x_P)} \frac{\mathrm{d}p_o}{p_o} = \frac{kT}{q}\ln\left[\frac{p_o(x_P)}{p_o(-x_N)}\right] \tag{3.93}$$

Since we have assumed nondegeneracy, $p_o(x_P) = N_{\mathrm{A}}$ and $p_o(-x_N) = n_{\mathrm{i}}^2/N_{\mathrm{D}}$. Therefore,

$$V_{\mathrm{bi}} = \frac{kT}{q}\ln\left[\frac{N_{\mathrm{D}}N_{\mathrm{A}}}{n_{\mathrm{i}}^2}\right]. \tag{3.94}$$

Figure 3.14 shows the equilibrium energy band diagram, electric field, and charge density for a simple abrupt *pn*-junction silicon diode in the vicinity of the depletion region. The conduction bandedge is given by $E_{\mathrm{C}}(x) = E_0 - q\phi(x) - \chi$, the valence bandedge

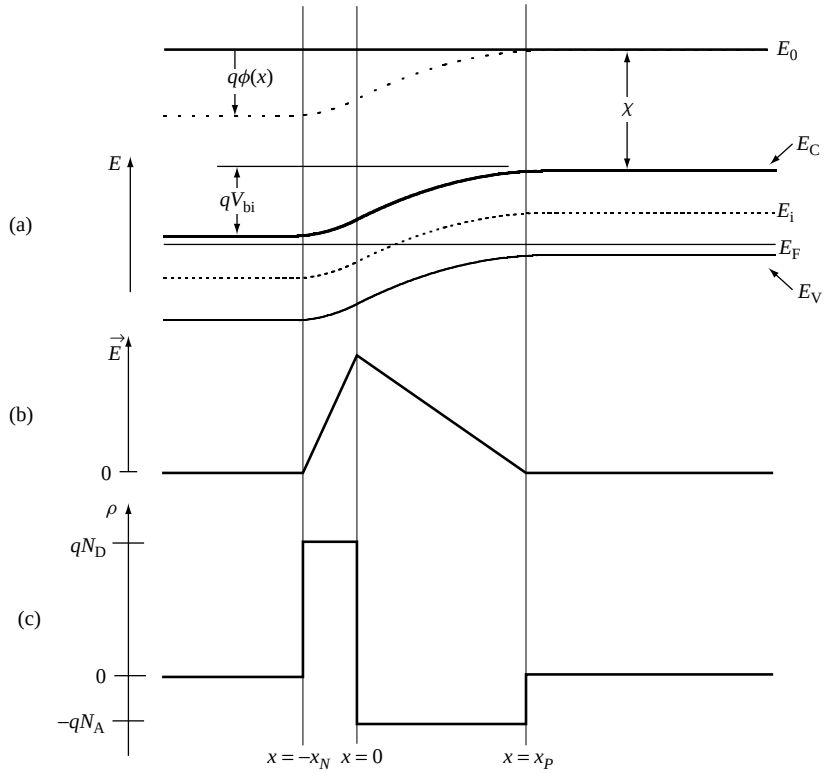


Figure 3.14 Equilibrium conditions in a solar cell: (a) energy bands; (b) electric field; and (c) charge density

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 823

by $E_V(x) = E_C(x) - E_G$, and the intrinsic energy by equation (3.19). E_0, defined as the vacuum energy, serves as a convenient reference point and is universally constant with position. An electron at the vacuum energy is, by definition, completely free of influence from all external forces. The electron affinity, χ, is the minimum energy needed to free an electron from the bottom of the conduction band and take it to the vacuum level. The electric field is a result of the uncovered ionized donors and acceptors and opposes the diffusion of electrons and holes in the quasi-neutral regions. The charge density plot illustrates the balance of charge between the two sides of the depletion region. In heterojunctions, both the band gap and the electron affinity are position-dependent – making the calculation of the junction electrostatics and energy band diagram more involved.

The basic solar cell structure has now been established (Figure 3.13). It is simply a *pn*-junction diode consisting of two quasi-neutral regions on either side of a depletion region with an electrical contact made to each quasi-neutral region. Typically, the more heavily doped quasi-neutral region is called the *emitter* (the *n*-type region in Figure 3.13) and the more lightly doped region is called the *base* (the *p*-type region in Figure 3.13). The base region is also often referred to as the *absorber region* since the emitter region is usually very thin and most of the light absorption occurs in the base. This basic structure will serve as the basis for deriving the fundamental operating characteristics of the solar cell.

3.4 SOLAR CELL FUNDAMENTALS

The basic current–voltage characteristic of the solar cell can be derived by solving the minority-carrier diffusion equation with appropriate boundary conditions.

3.4.1 Solar Cell Boundary Conditions

At $x = -W_N$, the usual assumption is that the front contact can be treated as an ideal ohmic contact. Hence,

$$\Delta p(-W_N) = 0. \tag{3.95}$$

However, since the front contact is usually a grid with metal contacting the semiconductor on only a small percentage of the front surface, modeling the front surface with an effective surface recombination velocity is more realistic. This effective recombination velocity models the combined effects of the ohmic contact and the antireflective passivation layer (SiO_2 in silicon solar cells). In this case, the boundary condition at $x = -W_N$ is

$$\frac{d\Delta p}{dx} = \frac{S_{F,\text{eff}}}{D_p}\Delta p(-W_N) \tag{3.96}$$

where $S_{F,\text{eff}}$ is the effective front surface recombination velocity. As $S_{F,\text{eff}} \to \infty$, $\Delta p \to 0$, and the boundary condition given by equation (3.96) reduces to that of an ideal ohmic contact (equation 3.95). In reality, $S_{F,\text{eff}}$ depends upon a number of parameters and is bias-dependent. This will be discussed in more detail later.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 824

The back contact could also be treated as an ideal ohmic contact, so that

$$\Delta n(W_P) = 0. \tag{3.97}$$

However, solar cells are often fabricated with a *back-surface field* (BSF), a thin more heavily doped region at the back of the base region. The BSF keeps minority carriers away from the back ohmic contact and increases their chances of being collected and it can be modeled by an effective, and relatively low, surface recombination velocity. This boundary condition is then

$$\left.\frac{\mathrm{d}\Delta n}{\mathrm{d}x}\right|_{x=W_P} = -\frac{S_{\mathrm{BSF}}}{D_n}\Delta n(W_P), \tag{3.98}$$

where S_{BSF} is the effective surface recombination velocity at the BSF.

All that remains now is to determine suitable boundary conditions at $x = -x_N$ and $x = x_P$. These boundary conditions are commonly referred to as the *law of the junction*.

Under equilibrium conditions, zero applied voltage and no illumination, the Fermi energy, E_{F}, is constant with position. When a bias voltage is applied, it is convenient to introduce the concept of quasi-Fermi energies. It was shown earlier that the equilibrium carrier concentrations could be related to the Fermi energy (equations 3.15 and 3.16). Under nonequilibrium conditions, similar relationships hold. Assuming the semiconductor is nondegenerate,

$$p = n_{\mathrm{i}}\mathrm{e}^{(E_{\mathrm{i}}-F_P)/kT} \tag{3.99}$$

and

$$n = n_{\mathrm{i}}\mathrm{e}^{(F_N-E_{\mathrm{i}})/kT} \tag{3.100}$$

It is evident that under equilibrium conditions $F_P = F_N = E_{\mathrm{F}}$. Under nonequilibrium conditions, assuming that the majority carrier concentrations at the contacts retain their equilibrium values, the applied voltage can be written as

$$qV = F_N(-W_N) - F_P(W_P) \tag{3.101}$$

Since, in low-level injection, the majority carrier concentrations are constant throughout the quasi-neutral regions, that is, $p_P(x_P \le x \le W_P) = N_{\mathrm{A}}$ and $n_N(-W_N \le x \le -x_N) = N_{\mathrm{D}}$, $F_N(-W_N) = F_N(-x_N)$ and $F_P(W_P) = F_P(x_P)$. Then, assuming that both the quasi-Fermi energies remain constant inside the depletion region,

$$qV = F_N(x) - F_P(x) \tag{3.102}$$

for $-x_N \le x \le x_P$, that is, everywhere inside the depletion region. Using equations (3.99) and (3.100), this leads directly to the *law of the junction*, the boundary conditions used at the edges of the depletion region,

$$p_N(-x_N) = \frac{n_{\mathrm{i}}^2}{N_{\mathrm{D}}}\mathrm{e}^{qV/kT} \tag{3.103}$$

and

$$n_P(x_P) = \frac{n_i^2}{N_A} e^{qV/kT}. \tag{3.104}$$

3.4.2 Generation Rate

For light incident at the front of the solar cell, $x = -W_N$, the optical generation rate takes the form (see equation 3.35)

$$G(x) = (1 - s) \int_\lambda (1 - r(\lambda)) f(\lambda) \alpha(\lambda) e^{-\alpha(x + W_N)} \, d\lambda. \tag{3.105}$$

Only photons with $\lambda \leq hc/E_G$ contribute to the generation rate.

3.4.3 Solution of the Minority-carrier Diffusion Equation

Using the boundary conditions defined by equations (3.96), (3.98), (3.103), and (3.104) and the generation rate given by equation (3.105), the solution to the minority-carrier diffusion equation, equation (3.80), is easily shown to be

$$\Delta p_N(x) = A_N \sinh[(x + x_N)/L_p] + B_N \cosh[(x + x_N)/L_p] + \Delta p'_N(x) \tag{3.106}$$

in the n-type region and

$$\Delta n_P(x) = A_P \sinh[(x - x_P)/L_n] + B_P \cosh[(x - x_P)/L_n] + \Delta n'_P(x) \tag{3.107}$$

in the p-type region. The particular solutions due to $G(x)$, $\Delta p'_N(x)$, and $\Delta n'_P(x)$ are given by

$$\Delta p'_N(x) = -(1 - s) \int_\lambda \frac{\tau_p}{(L_p^2 \alpha^2 - 1)} [1 - r(\lambda)] f(\lambda) \alpha(\lambda) e^{-\alpha(x + W_N)} \, d\lambda \tag{3.108}$$

and

$$\Delta n'_P(x) = -(1 - s) \int_\lambda \frac{\tau_n}{(L_n^2 \alpha^2 - 1)} [1 - r(\lambda)] f(\lambda) \alpha(\lambda) e^{-\alpha(x + W_N)} \, d\lambda. \tag{3.109}$$

Using the boundary conditions set above, A_N, B_N, A_P, and B_P are easily obtained.

3.4.4 Terminal Characteristics

The minority-carrier current densities in the quasi-neutral regions are just the diffusion currents, since the electric field is negligible. Using the active sign convention for the current (since a solar cell is typically thought of as a battery) gives

$$\vec{J}_{p,N}(x) = -q D_p \frac{d\Delta p_N}{dx} \tag{3.110}$$

and

$$\vec{J}_{n,P}(x) = q D_n \frac{\mathrm{d}\Delta n_P}{\mathrm{d}x} \tag{3.111}$$

The total current is given by

$$I = A[J_p(x) + J_n(x)] \tag{3.112}$$

and is true everywhere within the solar cell (A is the area of the solar cell). Equations (3.110) and (3.111) give only the hole current in the n-type region and the electron current in the p-type region, not both at the same point. However, integrating equation (3.72), the electron continuity equation, over the depletion region, gives

$$\int_{-x_N}^{x_P} \frac{\mathrm{d}\vec{J}_n \mathrm{d}x}{\mathrm{d}x} \mathrm{d}x = \vec{J}_n(x_P) - \vec{J}_n(-x_N) = q \int_{-x_N}^{x_P} [R(x) - G(x)]\,\mathrm{d}x \tag{3.113}$$

$G(x)$ is easily integrated and the integral of the recombination rate can be approximated by assuming that the recombination rate is constant within the depletion region and is $R(x_m)$ where x_m is the point at which $p_\mathrm{D}(x_m) = n_\mathrm{D}(x_m)$ and corresponds to the maximum recombination rate in the depletion region. If recombination via a midgap single level trap is assumed, then, from equations (3.37), (3.99), (3.100), and (3.102), the recombination rate in the depletion region is

$$R_\mathrm{D} = \frac{p_\mathrm{D} n_\mathrm{D} - n_\mathrm{i}^2}{\tau_n(p_\mathrm{D} + n_\mathrm{i}) + \tau_p(n_\mathrm{D} + n_\mathrm{i})} = \frac{n_\mathrm{D}^2 - n_\mathrm{i}^2}{(\tau_n + \tau_p)(n_\mathrm{D} + n_\mathrm{i})} = \frac{n_\mathrm{D} - n_\mathrm{i}}{(\tau_n + \tau_p)} = \frac{n_\mathrm{i}(\mathrm{e}^{qV/2kT} - 1)}{\tau_\mathrm{D}} \tag{3.114}$$

where τ_D is the effective lifetime in the depletion region. From equation (3.113), $\vec{J}_n(-x_N)$, the majority carrier current at $x = -x_N$, can now be written as

$$\begin{aligned}\vec{J}_n(-x_N) &= \vec{J}_n(x_P) + q \int_{-x_N}^{x_P} G(x)\,\mathrm{d}x - q \int_{-x_N}^{x_P} R_\mathrm{D} \mathrm{d}x \\ &= \vec{J}_n(x_P) + q(1-s) \int_\lambda [1 - r(\lambda)] f(\lambda) [\mathrm{e}^{-\alpha(W_N - x_N)} - \mathrm{e}^{-\alpha(W_N + x_P)}]\,\mathrm{d}\lambda \\ &\quad - q \frac{W_\mathrm{D} n_\mathrm{i}}{\tau_\mathrm{D}} (\mathrm{e}^{qV/2kT} - 1)\end{aligned} \tag{3.115}$$

where $W_\mathrm{D} = x_P + x_N$. Substituting into equation (3.112), the total current is now

$$I = A \left[J_p(-x_N) + J_n(x_P) + J_\mathrm{D} - q \frac{W_\mathrm{D} n_\mathrm{i}}{\tau_\mathrm{D}} (\mathrm{e}^{qV/2kT} - 1) \right] \tag{3.116}$$

where

$$J_\mathrm{D} = q(1-s) \int_\lambda [1 - r(\lambda)] f(\lambda) (\mathrm{e}^{-\alpha(W_N - x_N)} - \mathrm{e}^{-\alpha(W_N + x_P)})\,\mathrm{d}\lambda \tag{3.117}$$

is the generation current from the depletion region and A is the area of the solar cell. The last term of equation (3.116) represents recombination in the space-charge region.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 827

The solutions to the minority-carrier diffusion equation, equations (3.106) and (3.107), can be used to evaluate the minority-carrier current densities, equations (3.110) and (3.111). These can then be substituted into equation (3.116), which, with some algebraic manipulation, yields

$$I = I_{\mathrm{SC}} - I_{o1}(\mathrm{e}^{qV/kT} - 1) - I_{o2}(\mathrm{e}^{qV/2kT} - 1) \tag{3.118}$$

where I_{SC} is the short-circuit current and is the sum of the contributions from each of the three regions: the n-type region ($I_{\mathrm{SC}N}$), the depletion region ($I_{\mathrm{SCD}} = AJ_{\mathrm{D}}$), and the p-type region ($I_{\mathrm{SC}P}$)

$$I_{\mathrm{SC}} = I_{\mathrm{SC}N} + I_{\mathrm{SCD}} + I_{\mathrm{SC}P} \tag{3.119}$$

where

$$I_{\mathrm{SC}N} = qAD_p \left[\frac{\Delta p'(-x_N)T_{p1} - S_{\mathrm{F,eff}}\Delta p'(-W_N) + D_p \left.\dfrac{\mathrm{d}\Delta p'}{\mathrm{d}x}\right|_{x=-W_N}}{L_p T_{p2}} - \left.\frac{\mathrm{d}\Delta p'}{\mathrm{d}x}\right|_{x=-x_N} \right] \tag{3.120}$$

with

$$T_{p1} = D_p/L_p \sinh[(W_N - x_N/L_p] + S_{\mathrm{F,eff}} \cosh[(W_N - x_N/L_p] \tag{3.121}$$

$$T_{p2} = D_p/L_p \cosh[(W_N - x_N)/L_p] + S_{\mathrm{F,eff}} \sinh[(W_N - x_N)/L_p] \tag{3.122}$$

and

$$I_{\mathrm{SC}P} = qAD_n \left[\frac{\Delta n'(x_P)T_{n1} - S_{\mathrm{BSF}}\Delta n'(W_P) + D_n \left.\dfrac{\mathrm{d}\Delta n'}{\mathrm{d}x}\right|_{x=W_P}}{L_n T_{n2}} + \left.\frac{\mathrm{d}\Delta n'}{\mathrm{d}x}\right|_{x=x_P} \right] \tag{3.123}$$

with

$$T_{n1} = D_n/L_n \sinh[(W_P - x_P)/L_n] + S_{\mathrm{BSF}} \cosh[(W_P - x_P)/L_n] \tag{3.124}$$

$$T_{n2} = D_n/L_n \cosh[(W_P - x_P)/L_n] + S_{\mathrm{BSF}} \sinh[(W_P - x_P)/L_n] \tag{3.125}$$

I_{o1} is the dark saturation current due to recombination in the quasi-neutral regions,

$$I_{o1} = I_{o1,p} + I_{o1,n} \tag{3.126}$$

with

$$I_{o1,p} = qA\frac{n_{\mathrm{i}}^2}{N_{\mathrm{D}}}\frac{D_p}{L_p}\left\{ \frac{D_p/L_p \sinh[(W_N - x_N)/L_p] + S_{\mathrm{F,eff}} \cosh[(W_N - x_N/L_p]}{D_p/L_p \cosh[(W_N - x_N)/L_p] + S_{\mathrm{F,eff}} \sinh[(W_N - x_N)/L_p]} \right\} \tag{3.127}$$

and

$$I_{o1,n} = qA\frac{n_i^2}{N_A}\frac{D_n}{L_n} \times \left\{\frac{D_n/L_n \sinh[(W_P - x_P)/L_n] + S_{BSF}\cosh[(W_P - x_P)/L_n]}{D_n/L_n \cosh[(W_P - x_P)/L_n] + S_{BSF}\sinh[(W_P - x_P)/L_n]}\right\} \quad (3.128)$$

These are very general expressions for the dark saturation current and reduce to more familiar forms when appropriate assumptions are made, as will be seen later.

I_{o2} is the dark saturation current due to recombination in the space-charge region,

$$I_{o2} = qA\frac{W_D n_i}{\tau_D} \quad (3.129)$$

and is bias-dependent since the depletion width, W_D, is a function of the applied voltage (equation 3.89).

3.4.5 Solar Cell $I-V$ Characteristics

Equation (3.118), repeated here, is a general expression for the current produced by a solar cell.

$$I = I_{SC} - I_{o1}(e^{qV/kT} - 1) - I_{o2}(e^{qV/2kT} - 1) \quad (3.130)$$

The short-circuit current and dark saturation currents are given by rather complex expressions (equations 3.119, 3.128, and 3.129) that depend on the solar cell structure, material properties, and the operating conditions. A full understanding of solar cell operation requires detailed examination of these terms. However, much can be learned about solar cell operation by examining the basic form of equation (3.130). From a circuit perspective, it is apparent that a solar cell can be modeled by an ideal current source (I_{SC}) in parallel with two diodes – one with an ideality factor of "1" and the other with an ideality factor of "2", as shown in Figure 3.15. Note that the direction of the current source is opposed to the current flow of the diodes – that is, it serves to forward-bias the diodes.

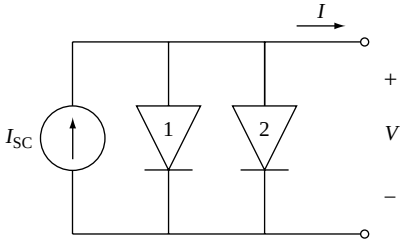


Figure 3.15 Simple solar cell circuit model. Diode 1 represents the recombination current in the quasi-neutral regions ($\propto e^{qV/kT}$), while diode 2 represents recombination in the depletion region ($\propto e^{qV/2kT}$)

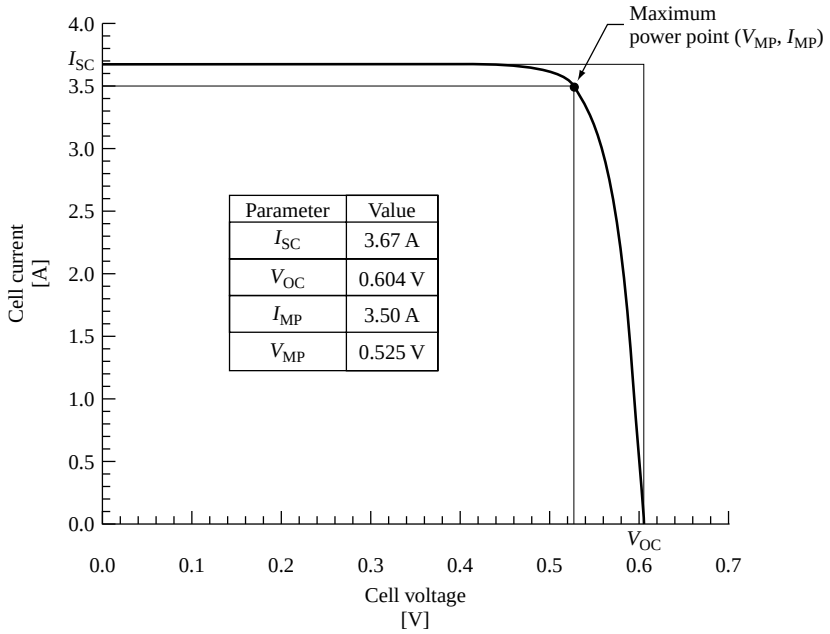


Figure 3.16 Current–voltage characteristic of the silicon solar cell defined by Table 3.2

The current–voltage ($I-V$) characteristic of a typical silicon solar cell is plotted in Figure 3.16 for the parameter values given in Table 3.2. For simplicity, the dark current due to the depletion region (diode 2) has been ignored (a reasonable and common assumption for a good silicon solar cell, especially at larger forward biases). It illustrates several important figures of merit for solar cells – the short-circuit current, the open-circuit voltage, and the fill factor. At small applied voltages, the diode current is negligible and the current is just the short-circuit current, I_{SC}, as can be seen when V is set to zero in equation (3.130). When the applied voltage is high enough so that the diode current (recombination current) becomes significant, the solar cell current drops quickly.

Table 3.2 shows the huge asymmetry between the n-emitter and the p-base in a typical solar cell. The emitter is $\sim$1000 times thinner, 10 000 times more heavily doped, and its diffusion length is $\sim$100 times shorter than the corresponding quantities in the base.

At open circuit ($I = 0$), all the light-generated current, I_{SC}, is flowing through diode 1, so the open-circuit voltage can be written as

$$V_{OC} = \frac{kT}{q} \ln \frac{I_{SC} + I_{o1}}{I_{o1}} \approx \frac{kT}{q} \ln \frac{I_{SC}}{I_{o1}}, \tag{3.131}$$

where $I_{SC} \gg I_{o1}$.

Of particular interest is the point on the $I-V$ curve where the power produced is at a maximum. This is referred to as the *maximum power point* with $V = V_{MP}$ and $I = I_{MP}$. As seen in Figure 3.18, this point defines a rectangle whose area, given by $P_{MP} = V_{MP} I_{MP}$, is the largest rectangle for any point on the $I-V$ curve. The maximum

Table 3.2 Si solar cell model parameters

Parameter	Value
A	100 cm^2
W_N	0.35 μm
$N_{\rm D}$	1×10^{20} cm^{-3}
D_p	1.5 cm^2/V-s
$S_{\rm F,eff}$	3×10^4 cm/s
τ_p	1 μs
L_p	12 μm
W_P	300 μm
$N_{\rm A}$	1×10^{15} cm^{-3}
D_n	35 cm^2/V-s
$S_{\rm BSF}$	100 cm/s
τ_n	350 μs
L_n	1100 μm

power point is found by solving

$$\left.\frac{\partial P}{\partial V}\right|_{V=V_{\rm MP}} = \left.\frac{\partial (IV)}{\partial V}\right|_{V=V_{\rm MP}} = \left.\left[I + V\frac{\partial I}{\partial V}\right]\right|_{V=V_{\rm MP}} = 0 \tag{3.132}$$

for $V = V_{\rm MP}$. The current at the maximum power point, $I_{\rm MP}$, is then found by evaluating equation (3.130) at $V = V_{\rm MP}$.

The rectangle-defined $V_{\rm OC}$ and $I_{\rm SC}$ provides a convenient means for characterizing the maximum power point. The fill factor, FF, is a measure of the *squareness* of the I–V characteristic and is always less than one. It is the ratio of the areas of the two rectangles shown in Figure 3.16 or

$$FF = \frac{P_{\rm MP}}{V_{\rm OC} I_{\rm SC}} = \frac{V_{\rm MP} I_{\rm MP}}{V_{\rm OC} I_{\rm SC}} \tag{3.133}$$

An empirical expression for the fill factor is [15]

$$FF = \frac{V_{\rm OC} - \dfrac{kT}{q}\ln[qV_{\rm OC}/kT + 0.72]}{V_{\rm OC} + kT/q}. \tag{3.134}$$

Arguably, the most important figure of merit for a solar cell is its power conversion efficiency, η, which is defined as

$$\eta = \frac{P_{\rm MP}}{P_{\rm in}} = \frac{FFV_{\rm OC} I_{\rm SC}}{P_{\rm in}} \tag{3.135}$$

The incident power, $P_{\rm in}$, is determined by the properties of the light spectrum incident upon the solar cell. Further information regarding experimental determination of these parameters appears in Chapter 16.

Another important figure of merit is the collection efficiency, which can be defined relative to both optical and recombination losses as an *external* collection efficiency

$$\eta_{\mathrm{C}}^{\mathrm{ext}} = \frac{I_{\mathrm{SC}}}{I_{ph}} \tag{3.136}$$

where

$$I_{ph} = qA \int_{\lambda<\lambda_{\mathrm{G}}} f(\lambda)\,\mathrm{d}\lambda \tag{3.137}$$

is the maximum possible photocurrent that would result if all photons with $E > E_{\mathrm{G}}$ ($\lambda < \lambda_{\mathrm{G}} = hc/E_{\mathrm{G}}$) created electron–hole pairs that were collected. The collection efficiency can also be defined with respect to recombination losses as the internal collection efficiency

$$\eta_{\mathrm{C}}^{\mathrm{int}} = \frac{I_{\mathrm{SC}}}{I_{\mathrm{gen}}} \tag{3.138}$$

where

$$I_{\mathrm{gen}} = qA(1-s) \int_{\lambda<\lambda_{\mathrm{G}}} [1-r(\lambda)] f(\lambda)(1-\mathrm{e}^{-\alpha(W_N+W_P)})\,\mathrm{d}\lambda \tag{3.139}$$

is the light-generated current. This represents what the short-circuit current would be if every photon that is absorbed is collected and contributes to the short-circuit current. $I_{\mathrm{gen}} = I_{\mathrm{inc}}$ when there is no grid shadowing, no reflective losses, and the solar cell has infinite optical thickness.

3.4.6 Properties of Efficient Solar Cells

Using these figures of merit, the properties of a good (efficient) solar cell can be ascertained. From equation (3.135), it is clear that an efficient solar cell will have a high short-circuit current, I_{SC}, a high open-circuit voltage, V_{OC}, and a fill factor, FF, as close as possible to 1.

$I_{\mathrm{SC}} = \eta_{\mathrm{C}}^{\mathrm{int}} I_{\mathrm{gen}}$ is directly proportional to both the internal collection efficiency and the light-generated current, I_{gen}. The internal collection efficiency is solely dependent on the recombination in the solar cell and will approach 1 as $\tau \to \infty$ and $S \to 0$. To maximize I_{gen} (i.e. $I_{\mathrm{gen}} \to I_{\mathrm{inc}}$), the solar cell should be designed with a minimum amount of grid shadowing (s), minimum reflectance ($r(\lambda)$), and be optically thick enough such that nearly all the photons with $E > E_{\mathrm{G}}$ are absorbed.

The open-circuit voltage

$$V_{\mathrm{OC}} \approx \frac{kT}{q} \ln \frac{I_{\mathrm{SC}}}{I_{o1}} \tag{3.140}$$

is logarithmically proportional to the short-circuit current and to the reciprocal of the reverse saturation current, I_{o1} (the same is true for I_{o2}). Therefore, reducing the saturation

current will increase the open-circuit voltage. From equations (3.127) and (3.128), it is obvious that $I_{o1} \to 0$ as $\tau \to \infty$ and $S \to 0$.

From equation (3.134) it is clear that increasing $V_{\rm OC}$ will increase the fill factor, FF. Thus, the design and the operation of an efficient solar cell has two basic goals:

1) Minimization of recombination rates throughout the device.
2) Maximization of the absorption of photons with $E > E_{\rm G}$.

It is evident that, despite the apparent complexity of the expressions describing the operation of solar cells, the basic operating principles are easy to understand. Electron–hole pairs are created inside the solar cell as a result of absorption of the photons incident on the solar cell from the sun. The objective is to collect the minority carriers before they are lost to recombination.

3.4.7 Lifetime and Surface Recombination Effects

The solar cell characteristics previously derived (equations 3.118 through 3.128) allow examination of the dependence of the solar cell performance on specific sources of recombination. Figure 3.17 shows how the base minority-carrier lifetime affects $V_{\rm OC}$, $I_{\rm SC}$, and the FF. Unless otherwise stated, the parameters of Table 3.2 are used to compute the solar cell performance. Short lifetimes mean that the diffusion length in the base is much less than the base thickness and carriers created deeper than about one diffusion length in the base are unlikely to be collected. When this is true ($L_n \ll W_P$), the contribution to the dark saturation current in the base (equation 3.128) becomes

$$I_{o1,n} = qA\frac{n_{\rm i}^2}{N_{\rm A}}\frac{D_n}{L_n} \tag{3.141}$$

and is commonly referred to as the *long-base approximation*. In this case, the BSF has no effect on the dark saturation current. On the other hand, when the base minority-carrier lifetime is long ($L_n \gg W_P$), the carriers readily come in contact with the BSF and the dark saturation current is a strong function of $S_{\rm BSF}$

$$I_{o1,n} = qA\frac{n_{\rm i}^2}{N_{\rm A}}\frac{D_n}{(W_P - x_P)}\frac{S_{\rm BSF}}{S_{\rm BSF} + D_n/(W_P - x_P)} \tag{3.142}$$

When $S_{\rm BSF}$ is very large (i.e. no BSF), this reduces to the more familiar short-base approximation

$$I_{o1,n} = qA\frac{n_{\rm i}^2}{N_{\rm A}}\frac{D_n}{(W_P - x_P)}. \tag{3.143}$$

Figure 3.18 shows how $S_{\rm BSF}$ affects $V_{\rm OC}$, $I_{\rm SC}$, and the FF. Notice that the break point in the curves occurs when $S_{\rm BSF} \approx D_n/W_P = 1000$ cm/s, as can be inferred from equation (3.142).

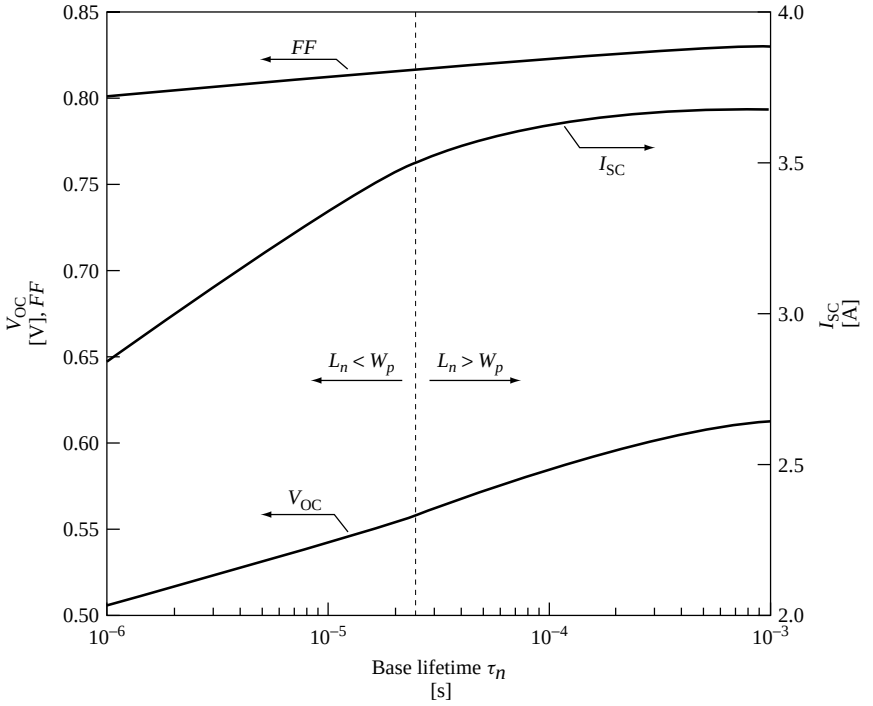


Figure 3.17 Effect of base lifetime on solar cell performance for the solar cell parameters in Table 3.2. The minority-carrier diffusion length ($L_n = \sqrt{D_n \tau_n}$) is equal to the base thickness (W_P) when $\tau_n = 25.7\ \mu s$

Front surface recombination for solar cells with contact grids on the front of the device is really an average over the front surface area of the relatively low surface recombination velocity between the grid lines and the very high surface recombination velocity of the ohmic contact. An expression for the effective front surface recombination velocity is given by [16]

$$S_{\mathrm{F,eff}} = \frac{(1-s)S_{\mathrm{F}}\overline{G}_N\tau_p\left(\cosh\dfrac{W_N}{L_p}-1\right)+p_o(\mathrm{e}^{qV/A_okT}-1)\left[s\dfrac{D_p}{L_p}\dfrac{\cosh\dfrac{W_N}{L_p}}{\sinh\dfrac{W_N}{L_p}}+S_{\mathrm{F}}\right]}{(1-s)\left[p_o(\mathrm{e}^{qV/A_okT}-1)+\overline{G}_N\tau_p\left(\cosh\dfrac{W_N}{L_p}-1\right)\right]} \tag{3.144}$$

where S_{F} is the surface recombination velocity between the grid lines and $\overline{G}$ is the average generation rate in the emitter region. It is obvious that $S_{\mathrm{F,eff}}$ is dependent upon the solar cell operation point. This is better seen in Table 3.3 where some special cases are illustrated (assuming $L_p \gg W_N$).

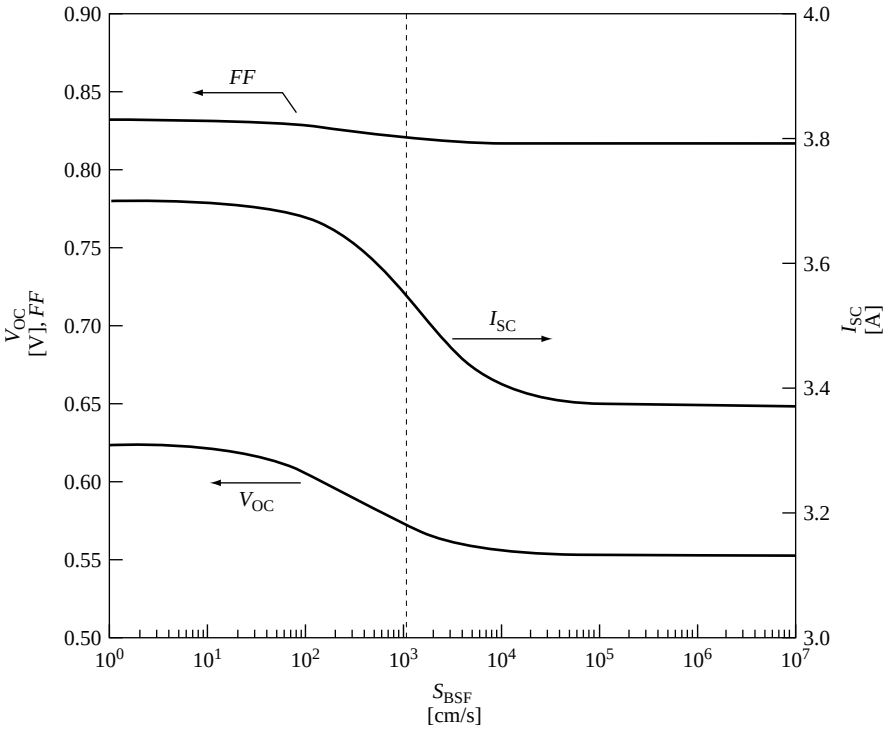


Figure 3.18 Effect of the back-surface field recombination velocity on solar cell performance. All other parameters are from Table 3.2

Table 3.3 Special cases of $S_{F,eff}$

No grid ($s = 0$)	$S_{F,eff} = S_F$
Full grid ($s = 1$)	$S_{F,eff} \rightarrow \infty$
Dark ($\overline{G} = 0$)	$S_{F,eff} = \dfrac{S_F + sD_p/W_N}{1 - s}$
Short circuit (V = 0)	$S_{F,eff} = S_F$
V large ($\approx V_{OC}$)	$S_{F,eff} = \dfrac{S_F + sD_p/W_N}{1 - s}$

3.4.8 An Analogy for Understanding Solar Cell Operation: A Partial Summary

The following analogy illustrates the importance of minimizing all sources of recombination in the solar cell.[4] Imagine a funnel that has a variety of holes of different shapes and sizes in it. It also has a stopcock at the bottom for controlling the flow of liquid

[4] This analogy was developed on the basis of discussions with Professor Richard J. Schwartz of Purdue University, West Lafayette, Indiana.

through the bottom of the funnel. Water, representing the incident sunlight, is poured into the top of the funnel. Water flowing out of the funnel through the stopcock represents the current delivered by the solar cell. Since the funnel is full of holes, some of the water leaks out instead of flowing through the stopcock. This leakage represents the recombination of minority carriers in the solar cell. The different-shaped holes represent different sources of recombination. For instance, square holes might represent recombination in the base region, round holes might be recombination in the space-charge region, triangular holes might be surface recombination at the back contact, and so on. The rate at which water pours in is proportional to the light intensity. At steady state, the water will find a height such that the flow of water in, I_{gen}, is equal to the flow through the stopcock (I) plus the water that leaks through the holes (I_{recomb}). This height represents the solar cell voltage (V).

When the stopcock is fully open, the water flows out through the stopcock at its maximum rate (I_{SC}), although some water will leak out through the holes so that $I_{\text{SC}} < I_{\text{gen}}$. This is analogous to the collection efficiency, η_{C}, of a solar cell – the objective being to minimize the amount of leakage (recombination) so that η_{C} is as close to unity as possible. Smaller holes means less recombination and $I_{\text{SC}} \rightarrow I_{\text{gen}}$.

As the stopcock is slowly closed, the level of liquid in the funnel rises – just as the solar cell voltage increases as the current decreases. When the stopcock is completely closed, the height of the water is representative of the open-circuit voltage (V_{OC}). At open circuit all the minority carriers must recombine just as all the water must leak out of the funnel in this analogy. If the holes are all big, the height of the water will be low. This is equivalent to short minority-carrier lifetimes and large surface recombination velocities that result in a low V_{OC}. By reducing the size of the holes (i.e. increasing the minority-carrier lifetimes and reducing the surface recombination velocities), the height of the water in the funnel (i.e. V_{OC}) is increased. Reducing the size of only the square holes (by increasing the minority-carrier lifetime in the base) will not increase the height of the water as much as might be expected since the round holes (recombination in the depletion region) are still large. All leaks (recombination sources) must be plugged (recombination rates minimized) before the height of the water (V_{OC}) increases substantially.

3.5 ADDITIONAL TOPICS

3.5.1 Efficiency and Band gap

Since only photons with $h\nu > E_{\text{G}}$ can create electron – hole pairs and contribute to the output of the solar cell, it is obvious that the band gap determines how well the solar cell is coupled to the solar spectrum. A simple analysis can be performed to predict the maximum solar cell efficiency. More complete analyses of the theoretical limits of solar cells are given elsewhere [17–19] and are also discussed in Chapter 4 of this handbook. Assuming the maximum energy that can be extracted from an absorbed photon is E_{G}, the maximum efficiency can be expressed as

$$\eta_{\max}(E_{\text{G}}) = \frac{\frac{1}{q} E_{\text{G}} I_{\text{inc}}}{P_{\text{in}}} = \frac{E_{\text{G}}}{(P_{\text{in}}/A)} \int_{\lambda<\lambda_{\text{G}}} f(\lambda)\, \mathrm{d}\lambda. \tag{3.145}$$

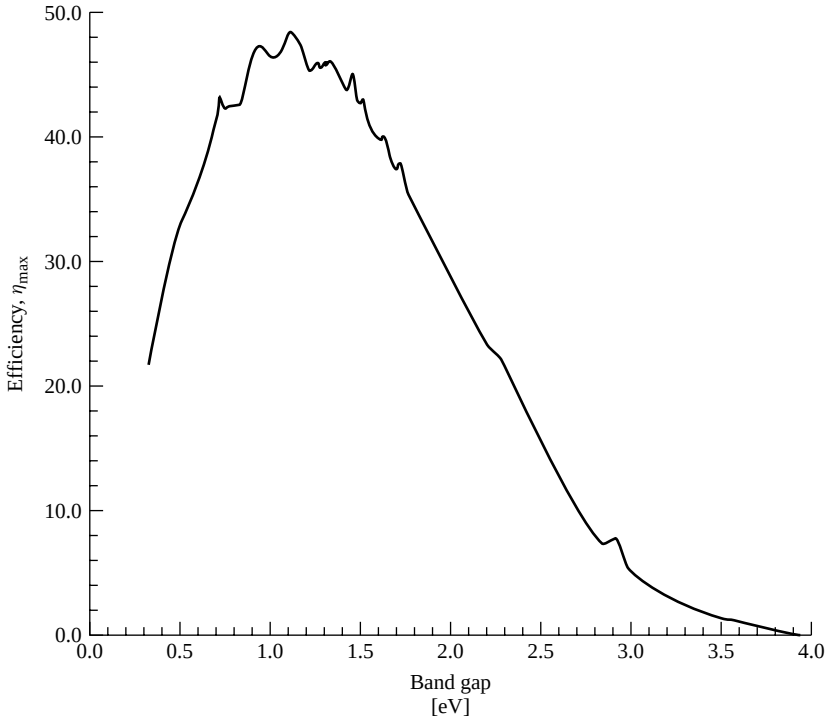


Figure 3.19 Theoretical maximum efficiency as a function of semiconductor band gap for an AM1.5 global spectrum

This is plotted in Figure 3.19 for an AM1.5 global spectrum and shows a maximum efficiency of 48% at about $E_G = 1.1$ eV, close to the band gap of silicon. Of course, this is only a simple estimate and assumes that $V_{OC} = 1/qE_G$ and $FF = 1$, which are obvious exaggerations. Perfect light trapping was also assumed so that $I_{SC} = I_{inc}$, but that is a more realistic prospect. Under nonconcentrated solar illumination, the actual maximum theoretical efficiency for a silicon solar cell is approximately 30%. However, this simple approach does serve to demonstrate the important role the semiconductor band gap plays in determining solar cell performance and shows that band gaps between 1.0 and 1.6 eV have nearly equivalent maximum theoretical efficiencies.

3.5.2 Spectral Response

The spectral response, $SR(\lambda)$, of a solar cell permits an examination of how photons of different wavelengths (energy) contribute to the short-circuit current. Just as the collection efficiency can be measured as either an external or an internal collection efficiency, so can the spectral response. The spectral response is defined as the *short-circuit current*, $I_{SC}(\lambda)$, resulting from a single wavelength of light normalized by the maximum possible current. The external spectral response is defined as

$$SR_{ext} = \frac{I_{SC}(\lambda)}{qAf(\lambda)} \tag{3.146}$$

and the internal spectral response as

$$SR_{\text{int}} = \frac{I_{\text{SC}}(\lambda)}{qA(1-s)(1-r(\lambda))f(\lambda)(\mathrm{e}^{-\alpha(\lambda)W_{\text{opt}}}-1)}, \tag{3.147}$$

where W_{opt} is the optical thickness of the solar cell (technically, also a function of wavelength). Experimentally, the external spectral response is measured. The internal spectral response is determined from it along with the knowledge of the grid shadowing, reflectance, and optical thickness. W_{opt} can be greater than the cell thickness if light-trapping methods are used. Such methods include textured surfaces [20] and back-surface reflectors [21] and are discussed in Chapter 8. The short-circuit current can be written in terms of the external spectral response as

$$I_{\text{SC}} = \int_{\lambda} SR_{\text{ext}}(\lambda) f(\lambda)\, \mathrm{d}\lambda. \tag{3.148}$$

The internal spectral response gives an indication of which sources of recombination are affecting the cell performance. This is demonstrated in Figure 3.20 where the internal spectral response of the silicon solar cell described by the parameters of Table 3.2 is shown. Also shown is the spectral response when $S_{\text{F,eff}} = 100$ cm/s (a passivated front surface) and the spectral response when $S_{\text{BSF}} = 1 \times 10^7$ cm/s (in effect, no BSF). The short wavelength response improves dramatically when the front surface is passivated

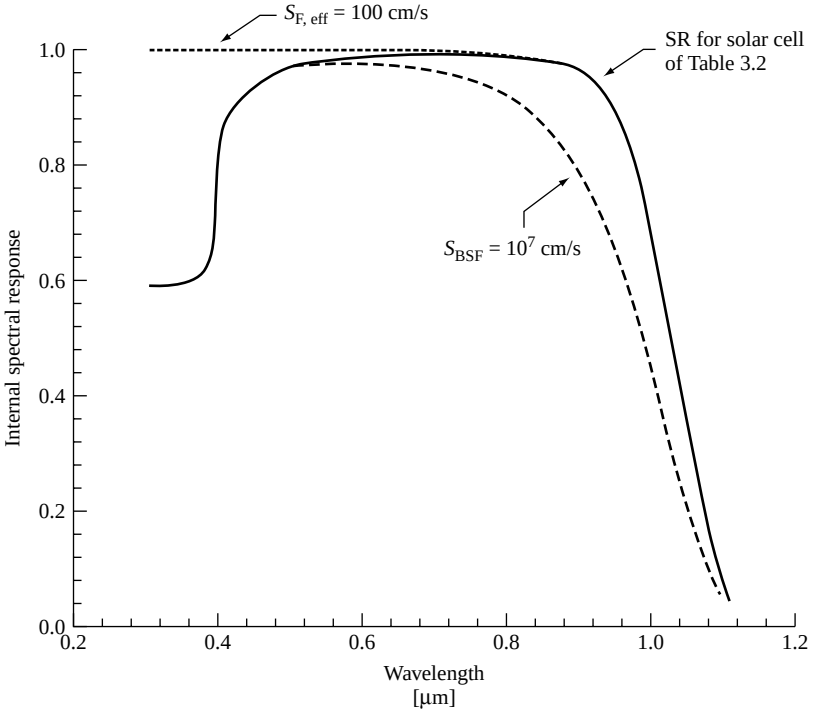


Figure 3.20 Internal spectral response of the silicon solar cell defined in Table 3.2

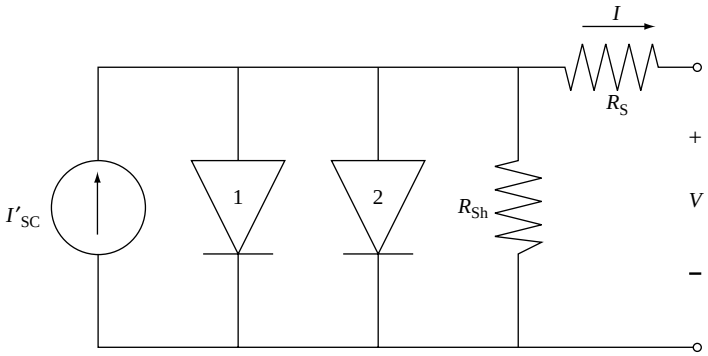


Figure 3.21 Solar cell circuit model including the parasitic series and shunt resistances

since the absorption coefficient is highest for short wavelength (high energy) photons. Conversely, removing the BSF makes it more likely that electrons created deep within the base region of the solar cell (those created by the long wavelength, low-energy photons) will recombine at the back contact and therefore, the long wavelength response is dramatically reduced.

3.5.3 Parasitic Resistance Effects

Equation (3.130) neglects the parasitic series and shunt resistances typically associated with real solar cells. Incorporating these resistances into the circuit model of Figure 3.15, as shown in Figure 3.21, yields

$$I = I'_{\mathrm{SC}} - I_{o1}(\mathrm{e}^{q(V+IR_{\mathrm{S}})/kT} - 1) - I_{o2}(\mathrm{e}^{q(V+IR_{\mathrm{S}})/2kT} - 1) - \frac{(V + IR_{\mathrm{S}})}{R_{\mathrm{Sh}}} \qquad (3.149)$$

where I'_{SC} is the short-circuit current when there are no parasitic resistances. The effect of these parasitic resistances on the $I-V$ characteristic is shown in Figures 3.22 and 3.23. As can also be seen in equation (3.149), the shunt resistance, R_{Sh}, has no effect on the short-circuit current, but reduces the open-circuit voltage. Conversely, the series resistance, R_{S}, has no effect on the open-circuit voltage, but reduces the short-circuit current. Sources of series resistance include the metal contacts, particularly the front grid, and the transverse flow of current in the solar cell emitter to the front grid.

It is often more convenient to rewrite equation (3.149) as

$$I = I'_{\mathrm{SC}} - I_o(\mathrm{e}^{q(V+IR_{\mathrm{S}})/A_o kT} - 1) - \frac{(V + IR_{\mathrm{S}})}{R_{\mathrm{Sh}}} \qquad (3.150)$$

where A_o is the diode ideality (quality) factor and typically has a value between 1 and 2, with $A_o \approx 1$ for diode dominated by recombination in the quasi-neutral regions and $A_o \to 2$ when recombination in the depletion region dominates. In solar cells where the

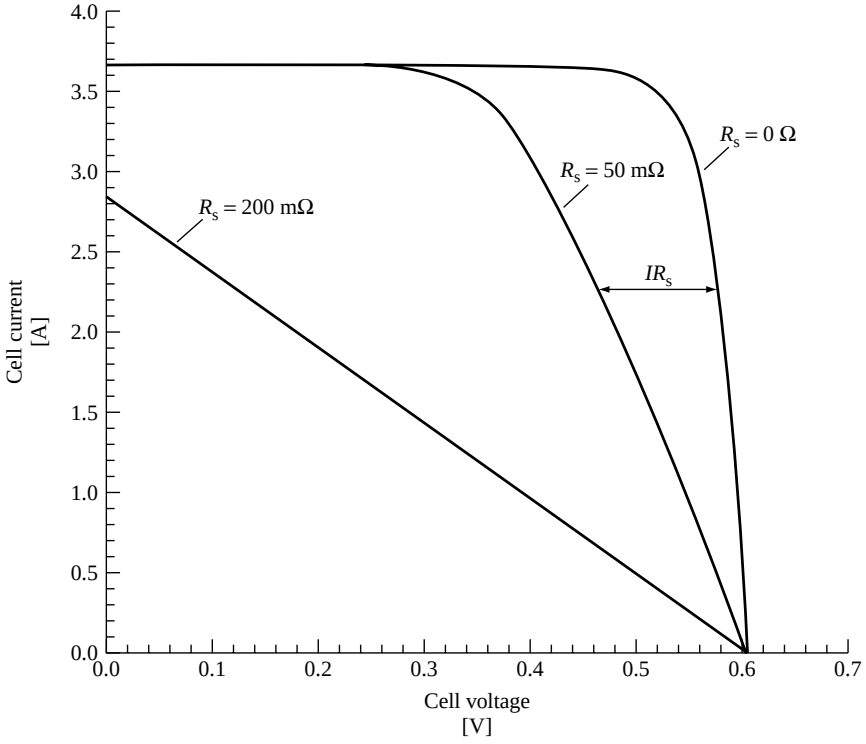


Figure 3.22 Effect of series resistance on the current–voltage characteristic of a solar cell ($R_{\mathrm{Sh}} \to \infty$)

recombination in each region is comparable, A_o is somewhere in-between. At short circuit, equation (3.150) becomes

$$I_{\mathrm{SC}} = I'_{\mathrm{SC}} - I_o(\mathrm{e}^{qI_{\mathrm{SC}}R_{\mathrm{S}}/A_okT} - 1) - I_{\mathrm{SC}}R_{\mathrm{S}}/R_{\mathrm{Sh}} \tag{3.151}$$

and at open circuit, it becomes

$$0 = I'_{\mathrm{SC}} - I_o(\mathrm{e}^{V_{\mathrm{OC}}/A_okT} - 1) - V_{\mathrm{OC}}/R_{\mathrm{Sh}}. \tag{3.152}$$

When $\log(I_{\mathrm{SC}})$ is plotted versus V_{OC} (where I_{SC} and V_{OC} are obtained over a range of illumination intensities), there is typically a regime where neither the series nor shunt resistances are important, as illustrated in Figure 3.24. The slope of this line will yield the diode ideality factor A_o, while the y-intercept will give I_o. In the regime where only series resistance is important, equations (3.151) and (3.152) can be combined to give

$$I_{\mathrm{SC}}R_{\mathrm{S}} = \frac{A_okT}{q} \ln\left[\frac{I_o\mathrm{e}^{qV_{\mathrm{OC}}/A_okT} - I_{\mathrm{SC}}}{I_o}\right] \tag{3.153}$$

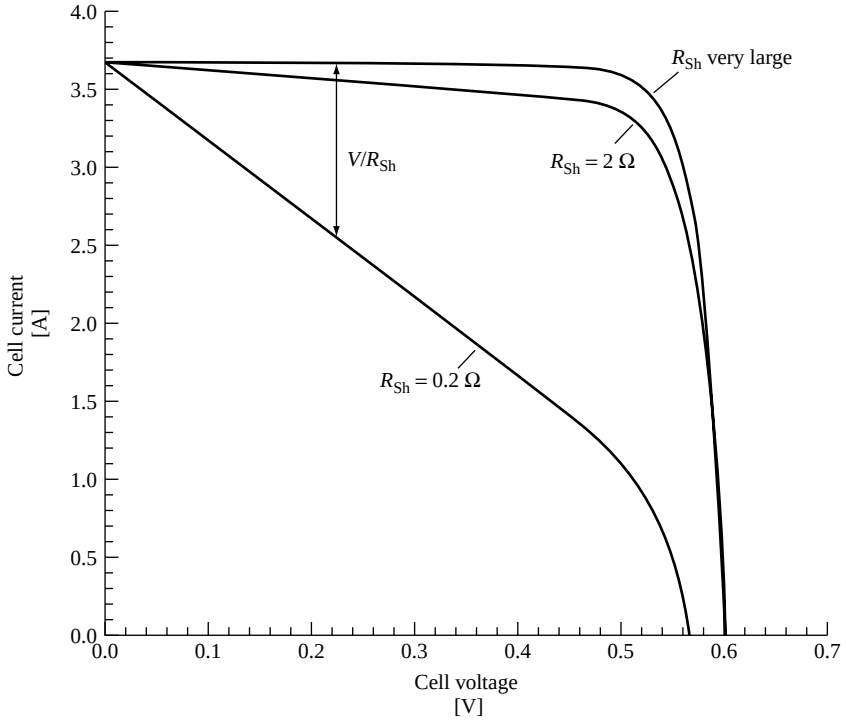


Figure 3.23 Effect of shunt resistance on the current–voltage characteristic of a solar cell ($R_S = 0$)

and a plot of I_{SC} versus $\log[I_o e^{qV_{OC}/A_o kT} - I_{SC}]$ will then permit R_S to be extracted from the slope of this line. Similarly, in the regime where only R_{Sh} is important, equations (3.151) and (3.152) can be combined to give

$$\frac{V_{OC}}{R_{Sh}} = I_{SC} - I_o e^{qV_{OC}/A_o kT} \tag{3.154}$$

and R_{Sh} can be determined from the slope of the line given by plotting V_{OC} versus $[I_{SC} - I_o e^{qV_{OC}/A_o kT}]$. If the series and shunt resistances are such that there is no regime where they can be neglected, the parameters can, with patience, be extracted through the process of trial and error.

3.5.4 Temperature Effects

From equations (3.127) and (3.128), it is apparent that

$$I_{o1,n},\ I_{o1,p} \propto n_i^2 \tag{3.155}$$

and from equation (3.129) that

$$I_{o2} \propto n_i. \tag{3.156}$$

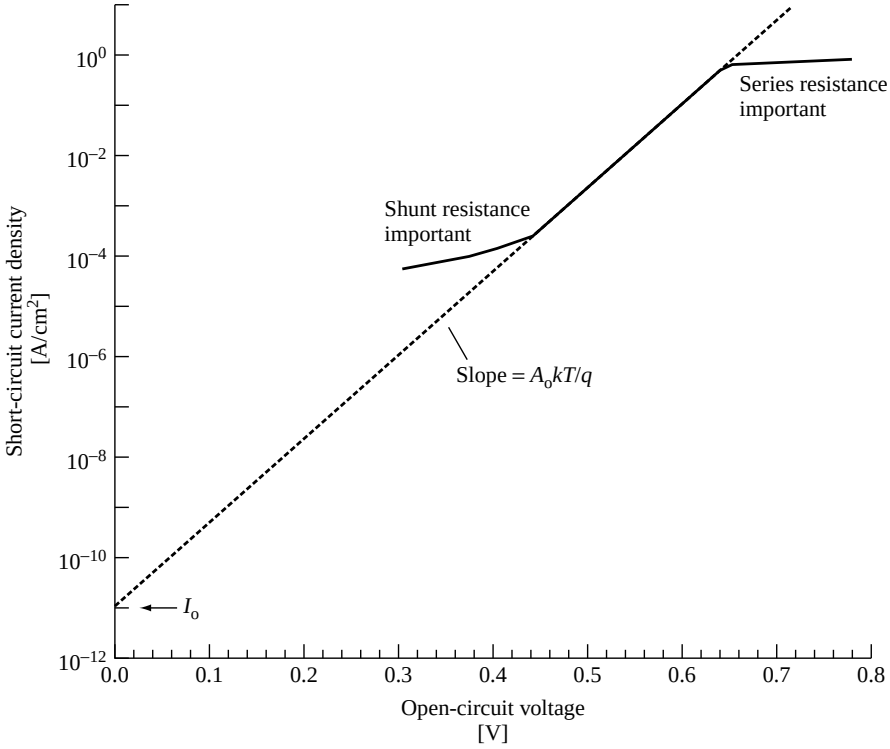


Figure 3.24 Short-circuit current versus open-circuit voltage plot illustrating parameter extraction

An increase in the intrinsic carrier concentration increases the dark saturation (recombination) current and results in a decrease in the open-circuit voltage, as can be seen from equation (3.140). The dark saturation current contains other temperature-dependent terms (D, τ, and S), but the temperature dependence of the intrinsic carrier concentration dominates. The intrinsic carrier concentration is given by equation (3.18), which when combined with equations (3.13) and (3.14) yields

$$n_{\mathrm{i}} = 2(m_n^* m_p^*)^{3/4} \left(\frac{2\pi kT}{h^2}\right)^{3/2} \mathrm{e}^{-E_{\mathrm{G}}/2kT}. \tag{3.157}$$

The effective masses are generally taken to be weak functions of temperature. The band gap decreases with temperature and its temperature dependence is well modeled by

$$E_{\mathrm{G}}(T) = E_{\mathrm{G}}(0) - \frac{aT^2}{T+\beta}. \tag{3.158}$$

where α and β are constants specific to each semiconductor. It is clear that as the temperature increases, n_{i} increases, and thus recombination increases, and cell performance is impaired. Band gap narrowing, referred to earlier, is a reduction in band gap due to high doping and also serves to increase n_{i} and impair solar cell performance.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 842

The open-circuit current expression, equation (3.140), can be rearranged and the temperature dependence explicitly included to give

$$I_{SC} \approx I_{o1} e^{qV_{OC}/kT} \approx BT^{\zeta} e^{-E_G(0)/kT} e^{qV_{OC}/kT} \tag{3.159}$$

where B is a temperature-independent constant and $T^{\zeta} e^{-E_G(0)/kT}$ accounts for the temperature dependence of the saturation current. The short-circuit current is relatively unaffected by temperature under typical operating conditions, so by differentiating with respect to T, the temperature dependence of the open-circuit voltage can be expressed as [15]

$$\frac{dV_{OC}}{dT} = -\frac{\frac{1}{q}E_G(0) - V_{OC} + \zeta\frac{kT}{q}}{T} \tag{3.160}$$

which for silicon at 300 K corresponds to about -2.3 mV/°C. Equation (3.159) can be rearranged as follows:

$$V_{OC}(T) = \frac{1}{q}E_G(0) - \frac{kT}{q}\ln\left(\frac{BT^{\zeta}}{I_{SC}}\right). \tag{3.161}$$

V_{OC} varies roughly linearly and inversely with temperature and an extrapolation of V_{OC} to $T = 0$ is approximately the band gap since $\lim_{T\to 0}[T \ln T] = 0$.

3.5.5 Concentrator Solar Cells

Operating solar cells under concentrated illumination offers two main advantages. The first is that since fewer solar cells are required to collect the sunlight falling on a given area, their cost of manufacture can be higher than that for cells designed for nonconcentrated illumination, and they are therefore presumably of higher quality (efficiency). The second is that operation under concentrated illumination offers an advantage in the solar cell efficiency. If sunlight is concentrated by a factor of X (X sun illumination), the short circuit at that concentration is

$$I_{SC}^{Xsuns} = XI_{SC}^{1sun}. \tag{3.162}$$

This is assuming that the semiconductor parameters are unaffected by the illumination level and that the cell temperature is the same at both levels of illumination – not necessarily valid assumptions especially at very large X, that is, X > 100. However, these assumptions will allow the demonstration of the potential efficiency of concentrator solar cells. Substituting equation (3.162) into equation (3.135) gives

$$\eta = \frac{FF^{Xsuns}V_{OC}^{Xsuns}I_{SC}^{Xsuns}}{P_{in}^{Xsuns}} = \frac{FF^{Xsuns}V_{OC}^{Xsuns}XI_{SC}^{1sun}}{XP_{in}^{1sun}} = \frac{FF^{Xsuns}V_{OC}^{Xsuns}I_{SC}^{1sun}}{P_{in}^{1sun}} \tag{3.163}$$

From equation (3.131),

$$V_{OC}^{Xsuns} = V_{OC}^{1sun} + \frac{kT}{q}\ln X. \tag{3.164}$$

The FF is a function of V_{OC} (equation 3.134), so

$$\eta^{\text{Xsuns}} = \eta^{\text{1sun}} \left(\frac{FF^{\text{Xsuns}}}{FF^{\text{1sun}}} \right) \left(1 + \frac{\dfrac{kT}{q} \ln \text{X}}{V_{\text{OC}}^{\text{1sun}}} \right). \tag{3.165}$$

Both factors multiplying the 1 sun efficiency increase as the illumination concentration increases. Therefore, the efficiency of concentrator cells increases as the illumination concentration increases, as shown in Figure 3.25.

Of course, there are many obstacles to achieving this. Concentrator cells must be cooled, since an increase in operating temperature reduces V_{OC}, and hence the cell efficiency. The FF^{Xsuns} eventually decreases with increasing X and current due to the parasitic series resistance. Concentrator solar cells are discussed in more detail in Chapter 11.

3.5.6 High-level Injection

In high-level injection, the excess carrier concentrations greatly exceed the doping in the base region, so $\Delta p \approx \Delta n \approx n \approx p$ if the carriers are moving generally in the same direction. This occurs with back-contact solar cells, such as the silicon point-contact solar

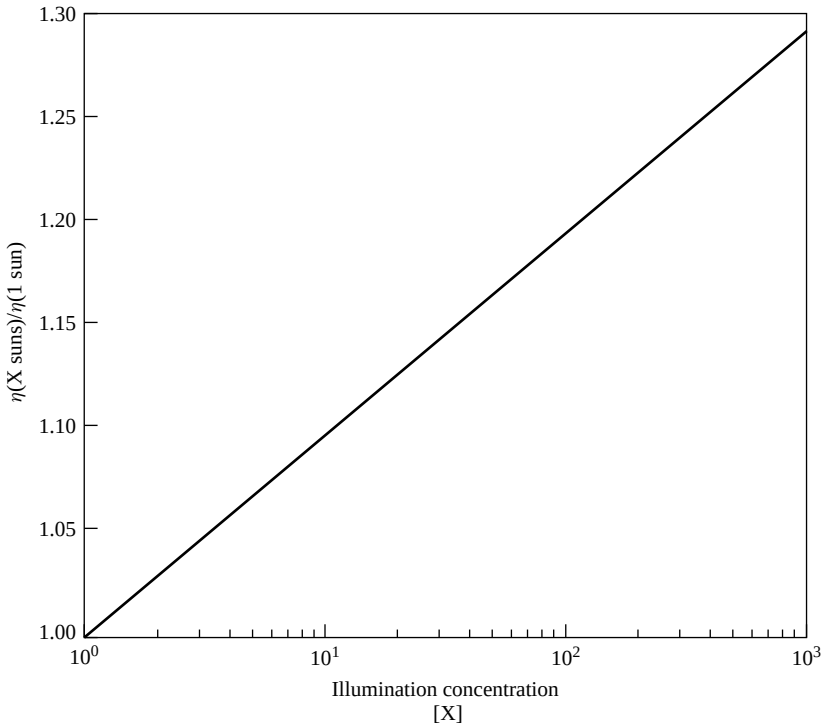


Figure 3.25 Relative efficiency as a function of illumination concentration

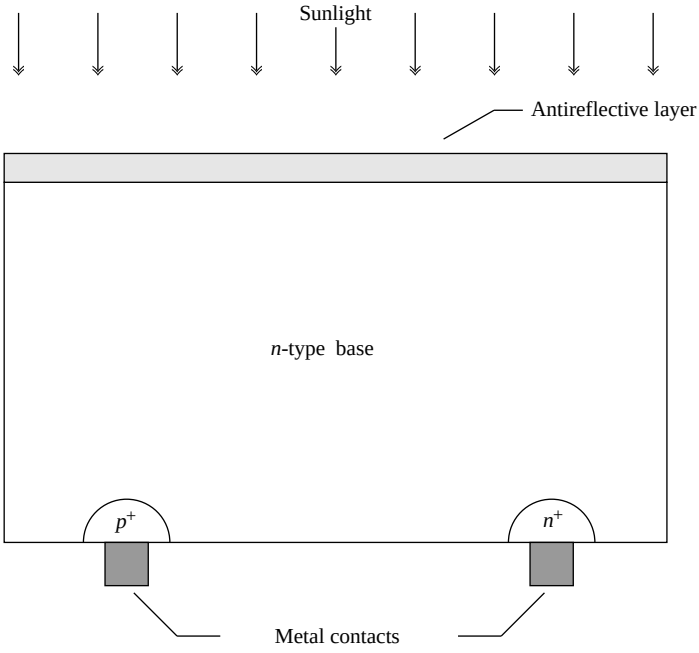


Figure 3.26 Schematic of a back-contact solar cell

cell [22], as illustrated in Figure 3.26. Since both electrical contacts are on the back, there is no grid shadowing. These cells are typically used in concentrator application and high-level injection conditions pervade. Assuming high-level injection, a simple analysis is possible.

Returning to equations (3.76) and (3.77), it can be seen that in high-level injection, the electric field can be eliminated (it is not necessarily zero), resulting in the ambipolar diffusion equation

$$D_{\mathrm{a}}\frac{\mathrm{d}^2 p}{\mathrm{d}x^2} - \frac{p}{\tau_n + \tau_p} = -G(x), \tag{3.166}$$

where the ambipolar diffusion coefficient is given by

$$D_{\mathrm{a}} = \frac{2D_n D_p}{D_n + D_p}. \tag{3.167}$$

The ambipolar diffusion coefficient is always less than the larger diffusion coefficient and greater than the smaller one.

In silicon, where $D_n/D_p \approx 3$ over a wide range of doping, the ambipolar diffusion coefficient is $D_{\mathrm{a}} \approx 3/2D_p \approx 1/2D_n$ and, if we also assume $\tau_p \approx \tau_n$, the ambipolar diffusion length is

$$L_{\mathrm{a}} \approx \sqrt{3}L_p \approx L_n. \tag{3.168}$$

Thus, the increased high-injection lifetime (see equation 3.40) offsets the reduced ambipolar diffusion coefficient.

It is crucial that the front surface of a back-contacted cell be well passivated, so we will assume that $S_F = 0$. We will further assume that optical generation is uniform throughout the base region. At open circuit, with these assumptions, $d^2p/dx^2 = 0$ and therefore

$$V_{OC} = \frac{2kT}{q} \ln\left[\frac{G(\tau_n + \tau_p)}{n_i}\right]. \tag{3.169}$$

The short-circuit current (with $p \simeq 0$ at the back of the cell) is

$$I_{SC} = qAL_aG \sinh(W_B/L_a) \tag{3.170}$$

which, when $L_a \gg W_B$, becomes

$$I_{SC} = qAW_BG. \tag{3.171}$$

3.5.7 *p-i-n* Solar Cells

The *p-i-n* solar cell takes advantage of the fact that in many semiconductor materials, especially direct band gap semiconductors (i.e. large absorption coefficient), most of the electron–hole pairs are created very near the surface. If an intrinsic (undoped) layer is placed between the (very thin) *n* and *p* regions, the depletion region thickness is the most significant fraction of the total solar cell thickness, as illustrated in Figure 3.27. Carrier collection is now aided by the electric field in the depletion region, which helps offset the low lifetimes in some materials, such as amorphous silicon (see Chapter 12). The $I-V$ characteristic of a *p-i-n* solar cell can be described with minor modifications to the previously derived expressions. The most significant modification is to equation (3.129) where the depletion width is now written as

$$W_D = \chi_N + W_I + \chi_P \tag{3.172}$$

where W_I is the thickness of the intrinsic layer. Since χ_N and χ_P are very thin, short-base approximations are in order for equations (3.127) and (3.128). Also, there is no BSF ($S_{BSF} \to \infty$).

3.5.8 Detailed Numerical Modeling

While analytic solutions such as those discussed thus far in this chapter provide an intuitive understanding of solar cell performance and are therefore very important, they are limited in their accuracy due to the many simplifying assumptions that must be made in order to obtain them. It is rather straightforward to solve the semiconductor equations numerically without the need to make so many simplifications. Several computer codes have been written that solve the semiconductor equations for the explicit purpose of modeling solar cells – PC-1D [23], AMPS [24], ADEPT [25], and its predecessors [16, 26, 27], for example. The basic design of these computer programs is very similar. The semiconductor

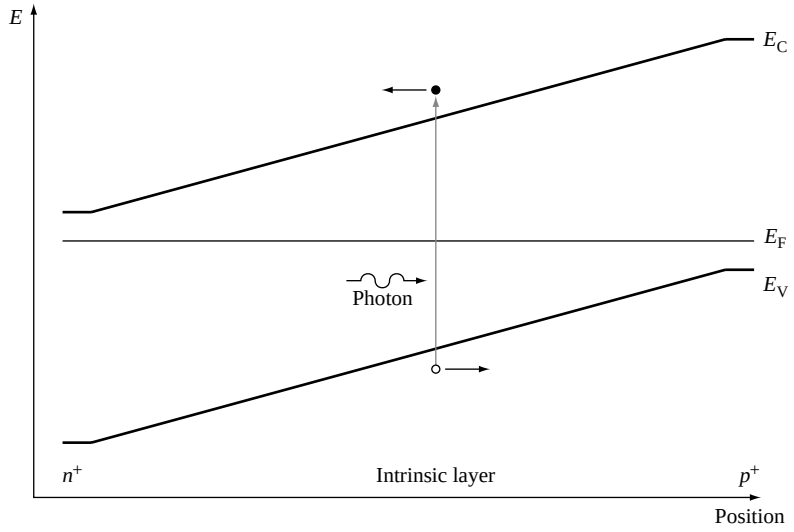


Figure 3.27 Band diagram of a *p-i-n* solar cell illustrating field-enhanced collection

equations (three coupled nonlinear partial differential equations) are cast in a normalized form [28] to simplify the calculations. Finite difference or finite element methods are then used to discretize the equations on a mesh (grid), resulting in a set of three coupled nonlinear difference equations. Using appropriately discretized boundary conditions, these equations are solved iteratively using a generalized Newton method to obtain the carrier concentrations and electric potential at each mesh point. Each Newton iteration involves the solution of a very large matrix equation or order $3N$, where N is the number of mesh points. One-dimensional simulations typically utilize on the order of 1000 mesh points, so the matrix is 3000×3000. In 2D, the minimum mesh is typically at least 100×100, so $N = 10^4$ and the matrix is of order 3×10^4 and contains 9×10^8 elements. Fortunately, the matrices are sparse and can be solved using considerably less computer memory than one would expect at first glance.

Numerical simulation allows analysis of solar cell designs and operating conditions for which simple analytic expressions are inadequate. The necessity of ignoring the spatial variation of parameters is eliminated and more accurate representations of the solar cell are possible. In particular, nonuniform doping, heterojunction solar cells (band gap varies spatially), amorphous silicon solar cells (complex trapping/recombination mechanisms, field-assisted collection), and concentrator solar cells (high-level injection, 2D/3D effects) can all be modeled with more precision.

3.6 SUMMARY

It has been the objective of this chapter to give the reader a basic understanding of the physical principles that underlie the operation of solar cells. Toward that end, the fundamental physical characteristics of solar cell materials that permit the conversion of light into electricity have been reviewed. These characteristics include the ability of

semiconductors to absorb photons by conferring that energy to carriers of electrical current and the ability of semiconductor materials to conduct electricity.

The basic operating principles of the solar cell (a carefully designed *pn*-junction diode) were derived from the (simplified) equations describing the dynamics of holes and electrons in semiconductors. This led to the definition of the solar cell figures of merit – the open-circuit voltage (V_{OC}), the short-circuit current (I_{SC}), the fill factor (FF), and the cell efficiency (η). The two key factors determining solar cell efficiency – electron–hole pair generation and recombination – were identified and discussed. In particular, the need to minimize all sources of recombination in the solar cells was demonstrated through examples and by way of a simple analogy.

The importance of matching the band gap of the solar cell material to the solar spectrum was also discussed and it was shown that silicon, with a band gap of 1.12 eV, is an excellent match to the solar spectrum. The effects of parasitic resistances and temperature on solar cell performance were examined and, finally, some advanced cell concepts were briefly introduced. Many of these topics will be expanded upon in the following chapters of this handbook.

REFERENCES

1. Green M, *Solar Cells: Operating Principles, Technology, and System Applications*, Chap. 1, Prentice Hall, Englewood Cliffs, NJ, 1–12 (1982).
2. Pierret R, in Pierret R, Neudeck G (Eds), *Modular Series on Solid State Devices, Volume VI: Advanced Semiconductor Fundamentals*, Addison-Wesley, Reading, MA (1987).
3. Sze S, *Physics of Semiconductor Devices*, 2nd Edition, John Wiley & Sons, New York, NY (1981).
4. Böer K, *Survey of Semiconductor Physics: Electrons and Other Particles in Bulk Semiconductors*, Van Nostrand Reinhold, New York, NY (1990).
5. Shur M, *Physics of Semiconductor Devices*, Prentice Hall, Englewood Cliffs, NJ (1990).
6. Singh J, *Physics of Semiconductors and Their Heterostructures*, McGraw-Hill, New York, NY (1993).
7. Pierret R, *Semiconductor Device Fundamentals*, Chap. 2, Addison-Wesley, Reading, MA, 23–74 (1996).
8. Slotboom J, De Graff H, *Solid-State Electron.* **19**, 857–862 (1976).
9. Pankove J, *Optical Processes in Semiconductors*, Chap. 3, Dover Publications, New York, NY, 34–81 (1971).
10. Sanii F, Giles F, Schwartz R, Gray J, *Solid-State Electron.* **35**, 311–317 (1992).
11. Pierret R, in Pierret R, Neudeck G (Eds), *Modular Series on Solid State Devices, Volume VI: Advanced Semiconductor Fundamentals*, Chap. 5, Addison-Wesley, Reading, MA, 139–179 (1987).
12. Pierret R, in Pierret R, Neudeck G (Eds), *Modular Series on Solid State Devices, Volume VI: Advanced Semiconductor Fundamentals*, Chap. 6, Addison-Wesley, Reading, MA (1987).
13. Van Roosbroeck W, *Bell Syst. Tech. J.* **29**, 560–607 (1950).
14. Lundstrom M, Schulke R, *IEEE Trans. Electron Devices* **30**, 1151–1159 (1983).
15. Green M, *Solar Cells: Operating Principles, Technology, and System Applications*, Chap. 5, Prentice Hall, Englewood Cliffs, NJ, 85–102 (1982).
16. Gray J, *Two-Dimensional Modeling of Silicon Solar Cells*, Ph.D. thesis, Purdue University, West Lafayette, IN (1982).
17. Gray J, Schwartz R, *Proc. 18th IEEE Photovoltaic Specialist Conf.*, 568–572 (1985).

18. Mathers C, *J. Appl. Phys.* **48**, 3181, 3182 (1977).
19. Shockley W, Queisser H, *J. Appl. Phys.* **32**, 510 (1961).
20. Baraona C, Brandhorst Jr. H, *Proc. 11th IEEE Photovoltaic Specialist Conf.*, 44–48 (1975).
21. Chai A, *Proc. 14th IEEE Photovoltaic Specialist Conf.*, 156–160 (1980).
22. Sinton R, Kwark Y, Swanson R, *Proc. 18th IEEE Photovoltaic Specialist Conf.*, 61–64 (1984).
23. Rover D, Basore P, Thorson G, *Proc. 18th IEEE Photovoltaic Specialist Conf.*, 703–709 (1984).
24. Rubinelli F *et al.*, *Proc. 22nd IEEE Photovoltaic Specialist Conf.*, 1405–1408 (1991).
25. Gray J, *Proc. 22nd IEEE Photovoltaic Specialist Conf.*, 436–438 (1991).
26. Lundstrom M, *Numerical Analysis of Silicon Solar Cells*, Ph.D. thesis, Purdue University, West Lafayette, IN (1980).
27. Gray J, *IEEE Trans. Electron Devices* **36**, 906–912 (1989).
28. Snowden C, *Introduction to Semiconductor Device Modeling*, Chap. 2, World Scientific, Singapore, 14–36 (1986).

4

Theoretical Limits of Photovoltaic Conversion

Antonio Luque and Antonio Martí

Instituto de Energía Solar, UPM, Spain

4.1 INTRODUCTION

Efficiency is an important matter in the photovoltaic (PV) conversion of solar energy because the sun is a source of power whose density is not very low, so it gives some expectations on the feasibility of its generalised cost-effective use in electric power production. However, this density is not so high as to render this task easy. After a quarter of a century of attempting it, cost still does not allow a generalised use of this conversion technology.

Efficiency forecasts have been carried out from the very beginning of PV conversion to guide the research activity. In solar cells the efficiency is strongly related to the generation of electron–hole pairs caused by the light, and their recombination before being delivered to the external circuit at a certain voltage. This recombination is due to a large variety of mechanisms and cannot be easily linked to the material used to make the cell. Nevertheless, already in 1975 Lofersky [1] had established an empirical link that allowed him to predict which materials were most promising for solar cell fabrication.

In 1960, Shockley and Queisser [2] pointed out that the ultimate recombination mechanisms – impossible to avoid – was just the detailed balance counterpart of the generation mechanisms. This allowed them to determine the maximum efficiency to be expected from a solar cell. This efficiency limit (40.7% for the photon spectrum approximated by a black body at 6000 K) is not too high because solar cells make rather ineffective use of the sun's photons. Many of them are not absorbed, and the energy of many of the absorbed ones is only poorly exploited.

Revisiting the topic of efficiency limits is pertinent because today there are renewed attempts to invent and develop novel concepts in solar cells – sometimes known as

Handbook of Photovoltaic Science and Engineering. Edited by A. Luque and S. Hegedus
© 2003 John Wiley & Sons, Ltd ISBN: 0-471-49196-9

third-generation cells [3] – aimed towards obtaining a higher efficiency. The oldest of them, and well established today, is the use of multijunction solar cells, to be studied in detail in Chapter 9 of this book. However, novel attempts are favoured by the general advancement of science and technology that puts new tools in the inventor's hands, such as nanotechnology, not available just a decade ago.

Conventional solar cells are semiconductor devices in which an interaction between electrons and holes is produced by absorption of light photons in order to produce electric work. Unlike solar thermal converters, they can operate at room temperature.[1] In this chapter we shall start by presenting some of the irreversible thermodynamic background regarding the photon–electron interaction. Special attention is paid to the conditions for complying with the second law of the thermodynamics as a guide for new device inventors.

This thermodynamic approach will give us the efficiency that a certain solar converter can obtain under certain ideal conditions. This will be applied not only to the present solar cells but also to a number of proposed new converters, some of them actually tried experimentally and some not.

The solar converter theory developed in this chapter looks very different from the one presented in Chapter 3 and is complementary to it. There the solar cell will be analysed in a solid-state physics context, on the basis of electric carrier transport and recombination, taking into account the subtleties that a given material may impose. In this chapter all materials will be considered ideal, entropy-producing mechanisms will be reduced to those inherent to the concept being studied and all other mechanisms will be ignored. In this way, the efficiencies given in this chapter are to be considered as upper bounds of the solar cells studied both in Chapter 3 and the remaining chapters of this book. The lower values found there are not necessarily due to poor technology. They may be fundamental when linked to the actual materials and processes used to manufacture real solar cells. However, they are not fundamental in the sense that other materials and processes could, in principle, be found where different materials or process limitations will apply, which perhaps might be less restrictive.

4.2 THERMODYNAMIC BACKGROUND

4.2.1 Basic Relationships

Thermodynamics defines a state function for a system in equilibrium. This function can be the entropy, the internal energy, the canonical grand potential, the enthalpy and so on, all of them mutually equivalent, containing the same information of the system and related to each other through the Legendre's transformation [4]. The variables used to describe the macroscopic state of a system are divided into *extensive* (volume U, number of particles N, entropy S, internal energy E, etc.) and *intensive* (pressure P, electrochemical potential μ, temperature T, etc.) variables. A number of relationships may be written with them. For instance,

$$\mathrm{d}E = T\,\mathrm{d}S - P\,\mathrm{d}U + \mu\,\mathrm{d}N \tag{4.1}$$

[1] Solar cells usually operate at some 40 to 60°C, but there is nothing theoretical against improving cooling to reach a cell temperature as close to the ambient one as desired.

For our purposes it is convenient to choose the grand potential Ω as our preferred state function. It is the state function that describes the state of a system when the electrochemical potential, the volume and the temperature are chosen as independent variables. It is also the Legendre transform of the energy with respect to the temperature and the electrochemical potential:

$$\Omega \doteq E - TS - \mu N = -PU \tag{4.2}$$

Its definition is given by the equal-by-definition sign ($\doteq$). The right-hand equality represents a property proven with generality [5].

Other thermodynamic variables that characterise the system in equilibrium can be obtained from the grand potential as derivatives. Hence, the number of particles, entropy and pressure of the system under consideration can be obtained as

$$N = -\left.\frac{\partial \Omega}{\partial \mu}\right|_{U,T}; \quad S = -\left.\frac{\partial \Omega}{\partial T}\right|_{U,\mu}; \quad P = -\left.\frac{\partial \Omega}{\partial U}\right|_{\mu,T} \tag{4.3}$$

For describing systems in non-equilibrium, the system under study is assumed to be subdivided into small subsystems each one comprising an elementary volume in the space-of-phases[2] (x, y, z, v_x, v_y, v_z) and having a size sufficient to allow the definition of thermodynamic magnitudes in it. Within such volumes, the subsystems are assumed to be in equilibrium. Thus, thermodynamic magnitudes are dependent on the position $\boldsymbol{r} \doteq (x, y, z)$ of the elementary volume and the velocity of motion $\boldsymbol{v} \doteq (v_x, v_y, v_z)$ of the elementary bodies or particles in it.

For describing the system in non-equilibrium, it is necessary to introduce the concept of *thermodynamic current densities* [6], $\boldsymbol{j}_x$. They are related to the extensive variables X and are defined for those elementary bodies with velocity $\boldsymbol{v}$ at the point $\boldsymbol{r}$ as follows:

$$\boldsymbol{j}_x(\boldsymbol{r}, \boldsymbol{v}) = x(\boldsymbol{r}, \boldsymbol{v})\, \boldsymbol{v} \tag{4.4}$$

where $x = X/U$ is the contribution to the extensive variable X, per unit of volume U at the point $\boldsymbol{r}$ of the elementary bodies with velocity $\boldsymbol{v}$.

Equation (4.2) can be applied to the thermodynamic current densities allowing us to write

$$\boldsymbol{j}_\omega = \boldsymbol{j}_e - T(\boldsymbol{r}, \boldsymbol{v})\boldsymbol{j}_s - \mu(\boldsymbol{r}, \boldsymbol{v})\boldsymbol{j}_n = -P(\boldsymbol{r}, \boldsymbol{v})\, \boldsymbol{v} \tag{4.5}$$

For the thermodynamic current densities, we can write the following continuity equations [7]:

$$g \doteq \frac{\partial n}{\partial t} + \nabla \cdot \boldsymbol{j}_n \tag{4.6}$$

$$\upsilon \doteq \frac{\partial e}{\partial t} + \nabla \cdot \boldsymbol{j}_e \tag{4.7}$$

$$\sigma \doteq \frac{\partial s}{\partial t} + \nabla \cdot \boldsymbol{j}_s \tag{4.8}$$

[2] x, y and z are the spatial co-ordinates giving the position of the particle and v_x, v_y and v_z are the co-ordinates of its velocity.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 852

where g, υ and σ are formally defined as the number of particles, energy and entropy generation rates per unity of volume. The symbol "$\nabla\cdot$" is the divergence operator.[3]

4.2.2 The Two Laws of Thermodynamics

Equations (4.7) and (4.8) have close links with the laws of thermodynamics. A certain elementary subsystem or body can draw energy from or release energy to another body close to it, but the first law of thermodynamics states that the sum of the energies generated at all the i elementary bodies at a given position $\boldsymbol{r}$ must be zero, that is,

$$\sum_i \upsilon(\boldsymbol{r}, \boldsymbol{v}_i) = 0 \tag{4.9}$$

In the same way, the entropy generated by an elementary body can possibly be negative, but the second law of thermodynamics, as stated by Prigogine [7], determines that the sum of all the entropy generated by all the bodies, σ_{irr}, must be non-negative everywhere.

$$\sum_i \sigma(\boldsymbol{r}, \boldsymbol{v}_i) = \sigma_{\text{irr}}(\boldsymbol{r}) \geq 0 \tag{4.10}$$

4.2.3 Local Entropy Production

It is illustrative to have a look at the sources for entropy production. Using equation (4.2) in the basic thermodynamic relationship of equation (4.1), we obtain an interesting relationship between thermodynamic variables per unity of volume:

$$\mathrm{d}s = \frac{1}{T}\,\mathrm{d}e - \frac{\mu}{T}\,\mathrm{d}n \tag{4.11}$$

If this relationship and equation (4.5) are substituted in equation (4.8), we find that

$$\sigma = \frac{1}{T}\frac{\partial e}{\partial t} - \frac{\mu}{T}\frac{\partial n}{\partial t} + \nabla\cdot\left(\frac{1}{T}\boldsymbol{j}_e - \frac{1}{T}\boldsymbol{j}_w - \frac{\mu}{T}\boldsymbol{j}_n\right) \tag{4.12}$$

Introducing equations (4.6) and (4.7) in (4.12) and after some mathematical handling we obtain that [8]

$$\sigma = \frac{1}{T}\upsilon + \boldsymbol{j}_e\nabla\frac{1}{T} - \frac{\mu}{T}g - \boldsymbol{j}_n\nabla\frac{\mu}{T} + \nabla\cdot\left(-\frac{1}{T}\boldsymbol{j}_\omega\right) \tag{4.13}$$

where "∇" is the gradient operator.[4] This is an important equation allowing us to identify the possible sources of entropy generation in a given subsystem. It contains terms involving energy generation (from another subsystem: υ) and transfer (from the surroundings:

[3] The linear unbounded operator "$\nabla\cdot$" applied to the vector $\boldsymbol{A} \doteq (A_x, A_y, A_z)$ is defined as $\nabla\cdot\boldsymbol{A} = \dfrac{\partial A_x}{\partial x} + \dfrac{\partial A_y}{\partial y} + \dfrac{\partial A_z}{\partial z}$.

[4] The gradient operator, "∇," (without being followed by a dot) is a linear unbounded operator whose actuation on a scalar field $f(x, y, z)$ is defined as $\nabla f = \dfrac{\partial f}{\partial x}\boldsymbol{e_1} + \dfrac{\partial f}{\partial y}\boldsymbol{e_2} + \dfrac{\partial f}{\partial z}\boldsymbol{e_3}$ where $\boldsymbol{e_1}$, $\boldsymbol{e_2}$, $\boldsymbol{e_3}$ are the orthonormal basis vectors of the Cartesian framework being used as reference.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 853

$\boldsymbol{j}_e \nabla 1/T$), free energy ($\mu g$) generation, Joule effect ($\boldsymbol{j}_n \nabla \mu$) and expansion of the volume that contains the particles ($\nabla \boldsymbol{j}_\omega / T$). This equation is very important and will be used to prove the thermodynamic consistence of solar cells.

4.2.4 An Integral View

Fluxes, $\dot{X}$, of the thermodynamic currents, $\boldsymbol{j}_x$, will be frequently used in this paper. In this text, they will be also called *thermodynamic variable rates*. By definition, the following relationship exists:

$$\dot{X} \doteq \int_A \sum_i \boldsymbol{j}_x \, \mathrm{d}A \tag{4.14}$$

where the sum refers to the different subsystems with different velocities to be found at a given position. A is the surface through which the flux is calculated. Actually, $\boldsymbol{j}_x \, \mathrm{d}A$ represents the scalar product of the current density vector $\boldsymbol{j}_x$ and the oriented surface element $\mathrm{d}A$ (orientation is arbitrary and if a relevant volume exists the orientation selected leads to the definition of *escaping* or *entering* rates).

4.2.5 Thermodynamic Functions of Radiation

The number of photons in a given mode of radiation is given [9] by the well-known Bose–Einstein factor $f_{\mathrm{BE}} = \{\exp[(\varepsilon - \mu)/kT] - 1\}^{-1}$, which through equation (4.3) is related to the grand canonical potential $\Omega = kT \ln\{\exp[(\mu - \varepsilon)/kT] - 1\}$. In these equations, most of the symbols have been defined earlier: ε is the photon energy in the mode and k is the Boltzman constant. The corresponding thermodynamic current densities for these photons are

$$\boldsymbol{j}_n = f_{\mathrm{BE}} \mathbf{c}/(U n_r); \quad \boldsymbol{j}_e = \varepsilon \boldsymbol{j}_n f_{\mathrm{BE}} \mathbf{c}/(U n_r); \quad \boldsymbol{j}_\omega = \Omega \mathbf{c}/(U n_r) \tag{4.15}$$

where $\mathbf{c}$ is the light velocity (a vector since it includes its direction) in the vacuum and n_r is the index of refraction of the medium in which the photons propagate, which is assumed to be independent of the direction of propagation. Thus, $\mathbf{c}/n_r$ is the velocity of the photons in the medium.

The number of photon modes with energy between ε and $\varepsilon + \mathrm{d}\varepsilon$ is $8\pi U n_r^3 \varepsilon^2/(h^3 \mathbf{c}^2) \, \mathrm{d}\varepsilon$. When the modes with energies $\varepsilon_m < \varepsilon < \varepsilon_M$ are taken into account, the total grand canonical potential of the photons, Ω_{ph}, associated with these modes is the sum of the contributions from each mode and can be written as

$$\Omega_{\mathrm{ph}}(U, T, \mu) = \frac{8\pi U n_r^3}{h^3 \mathbf{c}^3} \int_{\varepsilon_m}^{\varepsilon_M} \varepsilon^2 kT \ln(1 - \mathrm{e}^{(\mu - \varepsilon)/kT}) \, \mathrm{d}\varepsilon \tag{4.16}$$

where h is the Planck's constant.

Photons do not interact among themselves, such that temperatures and chemical potentials can be different for each mode. This means that they can be a function of the energy and of the direction of propagation. In the non-equilibrium case they can also be a

function of the position. If we only take into account the photons' propagation in a small solid angle $\mathrm{d}\varpi$, the grand potential in equation (4.16) must be multiplied by $\mathrm{d}\varpi/4\pi$. The same coefficient affects other thermodynamic variables of the radiation.

The flux of a thermodynamic variable X of the radiation through a surface A with a solid angle ϖ is then given by

$$\dot{X} = \int_A \sum_i \boldsymbol{j}_x \,\mathrm{d}A = \int_{A,\varpi} \frac{1}{4\pi}\frac{X}{U}\frac{\mathbf{c}}{n_r}\cos\theta\,\mathrm{d}\varpi\,\mathrm{d}A = \int_H \frac{1}{4\pi}\frac{X}{U}\frac{\mathbf{c}}{n_r^3}\,\mathrm{d}H \tag{4.17}$$

where the angle θ is defined in Figure 4.1. In this case, the sum of equation (4.14) has been substituted by the integration on solid angles. In many cases the integration will be extended to a restricted domain of solid angles. It is, in particular, the case of the photons when they come from a remote source such as the sun.

The differential variable $\mathrm{d}H = n_r^2\cos\theta\,\mathrm{d}\varpi\,\mathrm{d}A$, or its integral on a certain domain (at each position of A it must include the solid angle ϖ containing photons), is the so-called multilinear Lagrange invariant [10]. It is invariant for any optical system [11]. For instance, at the entry aperture of a solar concentrator (think of a simple lens), the bundle of rays has a narrow angular dispersion at its entry since all the rays come from the sun within a narrow cone. Then, they are collected across the whole entry aperture. The invariance for H indicates that it must take the same value at the entry aperture and at the receiver, or even at any intermediate surface that the bundle may cross. If no ray is turned back, all the rays will be present at the receiver. However, if this receiver is smaller than the entry aperture, the angular spread with which the rays illuminate the receiver has to be bigger than the angular spread that they have at the entry. In this way, H becomes a sort of measure of a bundle of rays, similar to its four-dimensional area with two spatial dimensions (in $\mathrm{d}A$) and two angular dimensions (in $\mathrm{d}\varpi$). Thus, we may talk of the H_{sr} of a certain bundle of rays linking the sun with a certain receiver.

Besides *Lagrange invariant*, this invariant receives other names. In treatises of thermal transfer it is called *vision* or *view factor*, but Welford and Winston [12] have recovered for this invariant the old name given by Poincaré that in our opinion accurately reflects its properties. He refers to it as *étendue* (extension) of a bundle of rays. We shall adopt in this chapter this name as a shortened denomination for this *multilinear Lagrange invariant*.

When the solid angle of illumination consists of the total hemisphere, $H = n_r^2\pi A$, where A is the area of the surface traversed by the photons. However, in the absence of optical elements, the photons from the sun reach the converter located on the Earth within

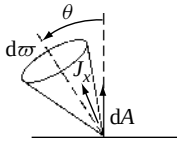


Figure 4.1 Drawing used to show the flux of a thermodynamic variable across a surface element $\mathrm{d}A$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 855

Table 4.1 Several thermodynamic fluxes for photons with energies between ε_m and ε_M distributed according to a black body law determined by a temperature T and chemical potential μ. The last line involves the calculations for a full-energy spectrum ($\varepsilon_m = 0$ and $\varepsilon_M = \infty$). One of the results (the one containing T^4) constitutes the Stefan–Boltzman law

$$\dot{\Omega}(T, \mu, \varepsilon_m, \varepsilon_M, H) = kT\frac{2H}{h^3\mathbf{c}^2}\int_{\varepsilon_m}^{\varepsilon_M} \ln(1 - \mathrm{e}^{(\mu-\varepsilon)/kT})\varepsilon^2\,\mathrm{d}\varepsilon = \int_{\varepsilon_m}^{\varepsilon_M} \dot{\omega}(T, \mu, \varepsilon, H)\,\mathrm{d}\varepsilon \qquad \text{(I-1)}$$

$$\dot{N}(T, \mu, \varepsilon_m, \varepsilon_M, H) = \frac{2H}{h^3\mathbf{c}^2}\int_{\varepsilon_m}^{\varepsilon_M} \frac{\varepsilon^2\,\mathrm{d}\varepsilon}{\mathrm{e}^{(\varepsilon-\mu)/kT}} - 1 = \int_{\varepsilon_m}^{\varepsilon_M} \dot{n}(T, \mu, \varepsilon, H)\,\mathrm{d}\varepsilon \qquad \text{(I-2)}$$

$$\dot{E}(T, \mu, \varepsilon_m, \varepsilon_M, H) = \frac{2H}{h^3\mathbf{c}^2}\int_{\varepsilon_m}^{\varepsilon_M} \frac{\varepsilon^3\,\mathrm{d}\varepsilon}{\mathrm{e}^{(\varepsilon-\mu)/kT}} - 1 = \int_{\varepsilon_m}^{\varepsilon_M} \dot{e}(T, \mu, \varepsilon, H)\,\mathrm{d}\varepsilon \qquad \text{(I-3)}$$

$$\dot{S} = \frac{\dot{E} - \mu\dot{N} - \dot{\Omega}}{T}; \quad \dot{F} \doteq \dot{E} - T\dot{S} = \mu\dot{N} + \dot{\Omega} \qquad \text{(I-4)}$$

$$\dot{E}(T, 0, 0, \infty, H) = (H/\pi)\sigma_{\mathrm{SB}}T^4; \dot{S}(T, 0, 0, \infty, H) = (4H/3\pi)\sigma_{\mathrm{SB}}T^3;$$
$$\sigma_{\mathrm{SB}} = 5.67 \times 10^{-8}\ \mathrm{Wm^{-2}\,K^{-4}} \qquad \text{(I-5)}$$

a narrow cone of rays. In this case, $H = \pi A \sin^2\theta_s$, where θ_s is the sun's semi-angle of vision (taking into account the sun's radius and its distance to Earth), which is equal [13] to 0.267°. No index of refraction is used in this case because the photons come from the vacuum.

Fluxes of some thermodynamic variables and some formulas of interest, which can be obtained from the above principles, are collected in Table 4.1.

4.2.6 Thermodynamic Functions of Electrons

For electrons, the number of particles in a quantum state with energy ε is given by the Fermi–Dirac function $f_{\mathrm{FD}} = \{\exp[(\varepsilon - \varepsilon_F)/kT] + 1\}^{-1}$ where the electrochemical potential (the chemical potential including the potential energy due to electric fields) of the electrons is usually called *Fermi level*, ε_F. Unlike photons, electrons are in continuous interaction among themselves, so that finding different temperatures for electrons is rather difficult. In fact, once a monochromatic light pulse is shined on the semiconductor, the electrons or holes thermalise to an internal electron temperature in less than 100 fs. Cooling to the network temperature will take approximately 10 to 20 ps [14]. However, electrons in semiconductors are found in two bands separated by a large energy gap in which electron states cannot be found. The consequence of this is that, in non-equilibrium, different electrochemical potentials, ε_{Fc} and ε_{Fv} (also called *quasi-Fermi levels*), can exist for the electrons in the conduction band (CB) and the valence band (VB), respectively. Sometimes we prefer to refer to the electrochemical potential of the holes (empty states at the VB), which is then equal to $-\varepsilon_{Fv}$. Once the excitation is suppressed, it can even take milliseconds before the two quasi-Fermi levels become a single one, as it is in the case of silicon.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 856

4.3 PHOTOVOLTAIC CONVERTERS

4.3.1 The Balance Equation of a PV Converter

In 1960, Shockley and Queisser (SQ) published an important paper [2] in which the efficiency upper limit of a solar cell was presented. In this article it was pointed out, for the first time in solar cells, that the generation due to light absorption has a detailed balance counterpart, which is the radiative recombination. This SQ efficiency limit, detailed below, occurs in ideal solar cells that are the archetype of the current solar cells. Such cells (Figure 4.2) are made up of a semiconductor with a VB and a CB, more energetic and separated from the VB by the band gap, ε_g. Each band is able to develop separate quasi-Fermi levels, ε_{Fc} for the CB and ε_{Fv} for the VB, to describe the carrier concentration in the respective bands. In the ideal SQ cell the mobility of the carriers is infinite, and since the electron and hole currents are proportional to the quasi-Fermi level gradients, it follows that the quasi-Fermi levels are constant. Contact to the CB is made by depositing a metal on an n^+-doped semiconductor. The carriers going through this contact are mainly electrons due to the small hole density. The few holes passing through this contact are accounted for as surface recombination, assumed zero in the ideal case. Similarly, contact with the VB is made with a metal deposited on a p-doped semiconductor. The metal Fermi levels ε_{F+} and ε_{F-} are levelled, respectively, to the hole and the electron quasi-Fermi levels ε_{Fv} and ε_{Fc} at each interface. In equilibrium, the two quasi-Fermi levels become just one.

The voltage V appearing between the two electrodes is given by the splitting of the quasi-Fermi levels, or, more precisely, by the difference in the quasi-Fermi levels of majority carriers at the ohmic contact interfaces. With constant quasi-Fermi levels and ideal contacts, the split is simply

$$qV = \varepsilon_{Fc} - \varepsilon_{Fv} \tag{4.18}$$

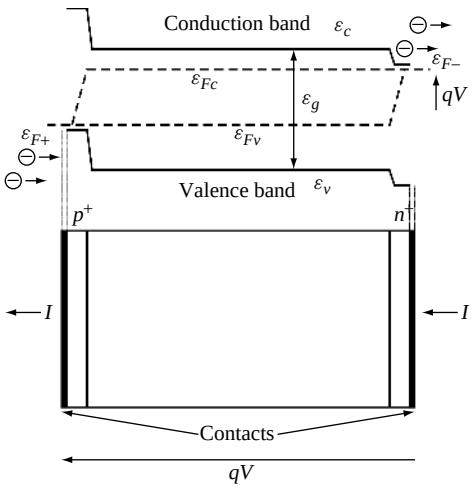


Figure 4.2 Band diagram of a solar cell with its contacts

Photons are absorbed by pumping electrons from the VB to the CB through the process known as electron–hole pair generation. However, as required by the detailed balance, the opposite mechanism is also produced so that a CB electron can decay to the VB and emit a photon, leading to what is called a *radiative recombination process*, responsible for luminescent light emission. In fact, many of such luminescent photons whose energy is slightly above the band gap are reabsorbed, leading to new electron–hole pair generations and balancing out the recombinations. Only the recombination processes leading to the effective emission of a photon out of the semiconductor produce a net recombination. Taking into account that the luminescent photons are emitted isotropically, only those photons emitted near the cell faces, at distances in the range of or smaller than the inverse of the absorption coefficient, and directed towards the cell faces with small angles (those reaching the surface with angles higher that the limit angle will be reflected back) have chances to actually leave the semiconductor, and thus to contribute to the net radiative recombination. The rare device analyses that account for this re-absorption – not yet very common as they are rather involved and the radiative recombination is small in silicon and in many thin-film cells – are said to include photon recycling [15].

In the ideal SQ cell any non-radiative recombination mechanism, which is an entropy-producing mechanism, is assumed to be absent.

The difference between the electrons pumped to the CB by external photon absorption and those falling again into the VB and effectively emitting a luminescent photon equals the current extracted from the cell. This can be presented in an equation form as

$$I/q = \dot{N}_s - \dot{N}_r = \int_{\varepsilon_g}^{\infty} (\dot{n}_s - \dot{n}_r)\,\mathrm{d}\varepsilon \tag{4.19}$$

where $\varepsilon_g = \varepsilon_c - \varepsilon_v$ is the semiconductor band gap and $\dot{N}_s$ and $\dot{N}_r$ are the photon fluxes entering or leaving the solar cell, respectively, through any surface. When the cell is properly contacted, this current is constituted by the electrons that leave the CB through the highly doped n-contact. In a similar balance, in the VB, I/q are also the electrons that enter the VB through the highly doped p-contact. Note that the sign of the current is the opposite to that of the flow of the electrons.

Using the nomenclature in Table 4.1, the terms in equation (4.19) for unit-area cells are $\dot{N}_s = a\dot{N}(T_s, 0, \varepsilon_g, \infty, \pi \sin^2\theta_s)$ for the cell facing the sun directly or $\dot{N}_s = a\dot{N}(T_s, 0, \varepsilon_g, \infty, \pi)$ for the cell under full concentration and $\dot{N}_r = \xi\dot{N}(T_a, qV, \varepsilon_g, \infty, \pi)$ where a and ξ are the absorptivity and emissivity of the cell. T_s is the sun temperature and T_a is the room temperature. Full concentration means using a concentrator without losses that is able to provide isotropic illumination; this is the highest illuminating power flux from a given source. The conservation of the étendue requires this concentrator to have a concentration C fulfilling the equation $C\pi \sin^2\theta_s = \pi n_r^2$, that is, $C = 46050 n_r^2$. This concentration is indeed unrealistic, but it does lead to the highest efficiency. Furthermore, it can be proven [16] that when the quasi-Fermi level split is uniform in the semiconductor bulk, then $a = \xi$. We shall assume – from now on in this chapter – that the solar cell is thick enough and perfectly coated with antireflection layers so as to fully absorb any photon with energy above the band gap energy so that $a = \xi = 1$ for these photons.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 858

The assumption $\dot{N}_r = \dot{N}(T_a, qV, \varepsilon_g, \infty, \pi)$ states that the temperature associated with the emitted photons is the room temperature T_a. This is natural because the cell is at this temperature. However, it also states that the chemical potential of the radiation emitted, μ_{ph}, is not zero but

$$\mu_{\text{ph}} = \varepsilon_{Fc} - \varepsilon_{Fv} = qV \tag{4.20}$$

This is so because the radiation is due to the recombination of electron–hole pairs, each one with a different electrochemical potential or quasi-Fermi level. A plausibility argument for admitting $\mu_{\text{ph}} = \varepsilon_{Fc} - \varepsilon_{Fv}$ is to consider that photons and electron–hole pairs are produced through the reversible (i.e. not producing entropy) equation electron + hole $\leftrightarrow$ photon. Equation (4.20) would then result as a consequence of equalling the chemical potentials before and after the reaction. Equation (4.20) can be also proven by solving the continuity equation for photons within the cell bulk [16, 17].

When the exponential of the Bose–Einstein function is much higher than one, the recombination term in equation (4.19) for full concentration can be written as

$$\begin{aligned}\dot{N}_r &= \frac{2\pi}{h^3\mathbf{c}^2}\int_{\varepsilon_g}^{\infty} \varepsilon^2 \exp\left(\frac{\varepsilon - qV}{kT_a}\right) \mathrm{d}\varepsilon \\ &= \frac{2\pi kT}{h^3\mathbf{c}^2}[4(kT)^2 + 2\varepsilon_g kT + \varepsilon_g^2] \exp\left(\frac{qV - \varepsilon_g}{kT_a}\right)\end{aligned} \tag{4.21}$$

This equation is therefore valid for $\varepsilon_g - qV \gg kT_a$. Within this approximation, the current–voltage characteristic of the solar cell takes its conventional *single* exponential appearance. In fact, this equation, with the appropriate factor $\sin^2\theta_s$, is accurate in all the ranges of interest of the current–voltage characteristic of ideal cells under unconcentrated sunlight.

The SQ solar cell can reach an efficiency limit given by

$$\eta = \frac{\{qV[\dot{N}_s - \dot{N}_r(qV)]\}_{\max}}{\sigma_{\text{SB}} T_s^4} \tag{4.22}$$

where the maximum is calculated by optimising V. This efficiency limit was first obtained by Shockley and Queisser [2] (for unconcentrated light) and is plotted in Figure 4.3 for several illumination spectra as a function of the band gap.

Outside the atmosphere the sun is seen quite accurately as a black body whose spectrum corresponds to a temperature of 5758 K [19]. To stress the idealistic approach of this chapter, we do not take this value in most of our calculations but rather 6000 K for the sun temperature and 300 K for the room temperature.

It must be pointed out that the limiting efficiency obtained for full concentration can be obtained also at lower concentrations if the étendue of the escaping photons is made equal to that of the incoming photons [16]. This can be achieved by locating the cell in a cavity [20] that limits the angle of the escaping photons.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 859

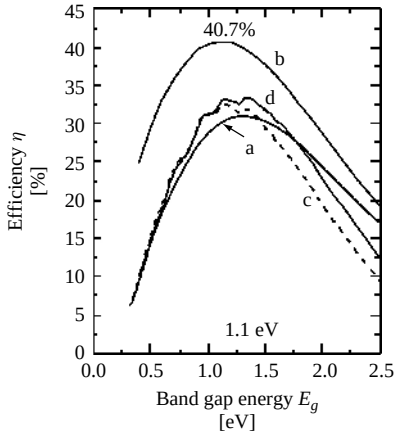


Figure 4.3 SQ efficiency limit for an ideal solar cell versus band gap energy for unconcentrated black body illumination, for full concentrated illumination and for illumination under the terrestrial sun spectrum: (a) unconcentrated 6000 K black body radiation (1595.9 Wm^{-2}); (b) full concentrated 6000 K black body radiation (7349.0×10^4 Wm^{-2}); (c) unconcentrated AM1.5-Direct [18] (767.2 Wm^{-2}) and (d) AM1.5 Global [18] (962.5 Wm^{-2})

It is interesting to note that the SQ analysis limit does not make any reference to semiconductor *pn* junctions. William Shockley, who first devised the *pn*-junction operation [21], was also the first in implicitly recognising [2] its secondary role in solar cells. In fact, a *pn* junction is not a fundamental constituent of a solar cell. What seems to be fundamental in a solar cell is the existence of two gases of electrons with different quasi-Fermi levels (electrochemical potentials) and the existence of selective contacts [22] that are traversed by each one of these two gases. The importance of the role of the existence of these selective contacts has not been sufficiently recognised. This is achieved today with *n*- and *p*-doped semiconductor regions, not necessarily forming layers, as in the point contact solar cell [23], but in the future it might be achieved otherwise, maybe leading to substantial advancements in PV technology. The role of the semiconductor, of which the cell is made, is to provide the two gases of electrons that may have different quasi-Fermi levels owing to the gap energy separation that makes the recombination difficult.

So far, for non-concentrated light, the most efficient single-junction solar cell, made of GaAs, has achieved an efficiency of 25.1% [24] of AM1.5G spectrum. This is only 23% below the highest theoretical efficiency of 32.8% for the GaAs band gap, of 1.42 eV, for this spectrum [25]. The theoretical maximum almost corresponds to the GaAs band gap. However, most cells are manufactured so that the radiation is also emitted towards the cell substrate located in the rear face of the cell, and little radiation, if any, turns back to the active cell. The consequence of this is that the étendue of the emitted radiation, which is π for a single face radiating to the air, is enlarged. The étendue is then $\pi + \pi n_r^2$. The term πn_r^2 proceeds from the emission of photons towards the substrate of the cell, which has a refraction index n_r. This reduces the limiting efficiency of the GaAs solar cell from 32.8 to 30.7%. Taking this into account, the efficiency of this best experimental cell is only 18.2% below the achievable efficiency of this cell, only on the basis of radiative recombination. Some substantial increase in efficiency might then be achieved by putting a reflector at the rear side of the active layers

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 860

of the cells [26] (not behind the substrate!). This would require the use of thin GaAs solar cells [27] or the fabrication of Bragg reflectors [28] (a stack of thin semiconductor layers of alternating refraction indices) underneath the active layers. Bragg reflectors have been investigated for enhancing the absorption of the incoming light in very thin cells, but the reduction of the luminescent emission towards the substrate might be an additional motivation.

4.3.2 The Monochromatic Cell

It is very instructive to consider an ideal cell under monochromatic illumination. When speaking of monochromatic illumination, we mean that, in fact, the cell is illuminated by photons within a narrow interval of energy $\Delta\varepsilon$ around the central energy ε. The monochromatic cell must also prevent the luminescent radiation of energy outside the range $\Delta\varepsilon$ from escaping from the converter.

For building this device an ideal concentrator [29] can be used that collects the rays from the solar disc, with an angular acceptance of just θ_s, with a filter on the entry aperture, letting the aforementioned monochromatic illumination to pass through. This concentrator is able to produce isotropic illumination at the receiver. By a reversal of the ray directions, the rays issuing from the cell in any direction are to be found also at the entry aperture with directions within the cone of semi-angle θ_s. Those with the proper energy will escape and be emitted towards the sun. The rest will be reflected back into the cell where they will be recycled. Thus, under ideal conditions no photon will escape with energy outside the filter energy and, furthermore, the photons escaping will be sent directly back to the sun with the same étendue of the incoming bundle H_{sr}.

The current in the monochromatic cell, ΔI, will then be given by

$$\begin{aligned}\Delta I/q &\equiv i(\varepsilon, V)\Delta\varepsilon/q = (\dot{n}_s - \dot{n}_r)\Delta\varepsilon \\ &= \frac{2H_{sr}}{h^3\mathbf{c}^2}\left[\frac{\varepsilon^2\Delta\varepsilon}{\exp\left(\dfrac{\varepsilon}{kT_s}\right)-1} - \frac{\varepsilon^2\Delta\varepsilon}{\exp\left(\dfrac{\varepsilon - qV}{kT_a}\right)-1}\right]\end{aligned} \tag{4.23}$$

This equation allows for defining an equivalent cell temperature T_r,

$$\frac{\varepsilon}{kT_r} = \frac{\varepsilon - qV}{kT_a} \Rightarrow qV = \varepsilon\left(1 - \frac{T_a}{T_r}\right) \tag{4.24}$$

so that the power produced by this cell, $\Delta\dot{W}$, is

$$\begin{aligned}\Delta\dot{W} &= \frac{2H_{sr}}{h^3\mathbf{c}^2}\left[\frac{\varepsilon^3\Delta\varepsilon}{\exp\left(\dfrac{\varepsilon}{kT_s}\right)-1} - \frac{\varepsilon^3\Delta\varepsilon}{\exp\left(\dfrac{\varepsilon}{kT_r}\right)-1}\right]\left(1 - \frac{T_a}{T_r}\right) \\ &= (\dot{e}_s - \dot{e}_r)\Delta\varepsilon\left(1 - \frac{T_a}{T_r}\right)\end{aligned} \tag{4.25}$$

We realise that the work extracted from the monochromatic cell is the same as that extracted from a Carnot engine fed with a heat rate from the hot reservoir $\Delta\dot{q} = (\dot{e}_s - \dot{e}_r)\Delta\varepsilon$. However, this similarity does not hold under a non-monochromatic illumination because, for a given voltage, the cell equivalent temperature would depend on the photon energy ε being unable to define a single equivalent temperature for the whole spectrum. Note that the equivalent cell temperatures corresponding to short-circuit and open-circuit conditions are T_a and T_s, respectively.

To calculate the efficiency, $\Delta\dot{W}$ in equation (4.25) must be divided by the appropriate denominator. We could divide by the black body incident energy $\sigma_{\mathrm{SB}}T_s^4$, but this would be unfair because the unused energy that is reflected by the entry aperture could be deflected with an optical device and used in other solar converters. We can divide by $\dot{e}_s\Delta\varepsilon$, the rate of power received at the cell, thus obtaining the *monochromatic* efficiency, η_{mc}, given by

$$\eta_{\mathrm{mc}} = \left.\frac{q(\dot{n}_s - \dot{n}_r)V}{\dot{e}_s}\right|_{\max} = \left.\left(1 - \frac{\dot{e}_r}{\dot{e}_s}\right)\left(1 - \frac{T_a}{T_r}\right)\right|_{\max} \tag{4.26}$$

that is represented in Figure 4.4 as a function of the energy ε.

Alternatively, we could have used the standard definition of efficiency used in thermodynamics [30, 31] to compute the efficiency of the monochromatic cell. In this context, we put in the denominator the energy really wasted in the conversion process, that is, $(\dot{e}_s - \dot{e}_r)\Delta\varepsilon$. Actually, the energy $\dot{e}_r\Delta\varepsilon$ is returned to the sun, perhaps for later use (slowing down, for example, the sun's energy loss process!). This leads to the *thermodynamic* efficiency:

$$\eta_{\mathrm{th}} = \left(1 - \frac{T_a}{T_r}\right) \tag{4.27}$$

This efficiency is the same as the Carnot efficiency obtained by a reversible engine operating between an absorber at temperature T_r and the ambient temperature and suggests that an ideal solar cell may work reversibly, without entropy generation. Its maximum,

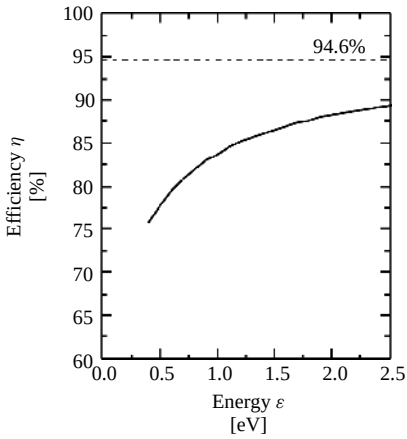


Figure 4.4 Monochromatic cell efficiency versus photon energy

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 862

for $\Delta\dot{w} \geq 0$, is obtained for $T_r = T_s$, although unfortunately it occurs when $\Delta\dot{w} = 0$, that is, when negligible (actually, none) work is done by the cell. However, this is a general characteristic of reversible engines, which yield the Carnot efficiency only at the expense of producing negligible power.

4.3.3 Thermodynamic Consistence of the Shockley–Queisser Photovoltaic Cell

Electrons and photons are the main particles interacting in a solar cell [8]. However, other interactions occur as well. In general, looking at equation (4.13), the generation of entropy for each kind of particle (electrons, photons and others) is written as

$$\sigma = \sum_i \left[\frac{1}{T}\upsilon + \boldsymbol{j}_e \nabla \frac{1}{T} - \frac{\mu}{T} g - \boldsymbol{j}_n \nabla \frac{\mu}{T} + \nabla \cdot \left(-\frac{1}{T} \boldsymbol{j}_\omega \right) \right] \tag{4.28}$$

where the summation extends to the different states of the particle.

First we analyse the generation of entropy by electrons, σ_{ele}. Except for the case of ballistic electrons, the pressure – appearing inside $\boldsymbol{j}_\omega$ as shown in equation (4.5) – is very quickly equilibrated; it is the same for $+\boldsymbol{v}$ and for $-\boldsymbol{v}$ as a result of the frequent elastic collisions. Thus, the term in $\boldsymbol{j}_\omega$ for electrons disappears when the sum is extended to all the states in each energy. These collisions also cause the temperature and the free energy to become the same for all the directions, at least for a given energy. Furthermore, in conventional solar cells, the temperature of the electrons is the same for any energy and is equal to the lattice temperature T_a. Also, the electrochemical potential is the same for all the electrons within the same band at a given $\boldsymbol{r}$.

In a real cell, the energy flow $\boldsymbol{j}_e$ goes from the high-temperature regions towards the low-temperature ones $[\nabla(1/T) \geq 0]$ so that the term in (4.28) involving $\boldsymbol{j}_e$ produces positive entropy. The electron flow, $\boldsymbol{j}_n$, opposes the gradient of electrochemical potential, thus also producing positive entropy for constant temperature. However, in the ideal cell, the lattice temperature, which is also that of the electrons, is constant and the term involving $\nabla(1/T)$ disappears. Furthermore, in the SQ [2] ideal cell, mobility is infinite and, therefore, its conduction and valence band electrochemical potentials or quasi-Fermi levels (ε_{Fc}, ε_{Fv}) are constant throughout the whole solar cell and their gradients are also zero. Therefore, all the gradients in equation (4.28) disappear and the electron contribution to entropy generation, σ_{ele}, is given by

$$\sigma_{\text{ele}} = \sum_{i-\text{ele}} \left[\frac{1}{T_a} \upsilon_{i-\text{ele}} - \frac{\varepsilon_{Fc(\upsilon)}}{T_a} g_{i-\text{ele}} \right] \tag{4.29}$$

The quasi-Fermi level to be used in this case is ε_{Fc} or ε_{Fv} depending on the band to which the electronic state i-ele belongs. This is represented by $\varepsilon_{Fc(\upsilon)}$.

Other interactions may occur in the cell involving other particles besides the electrons and the photons. We shall assume that in these interactions the bodies involved (labelled as *others*) also have a direction-independent pressure and that they are also at

the lattice temperature. Furthermore, if these particles exist, they are assumed to have zero chemical potential (as it is the case, for example, of the phonons). Using these assumptions, the contribution to the irreversible entropy generation rate from these *other* particles becomes

$$\sigma_{\text{others}} = \sum_{i-\text{others}} \left[\frac{1}{T_a} \upsilon_{i-\text{others}} \right] \tag{4.30}$$

For the case of the photons, the situation is rather different. As said before, photons do not interact with each other and, therefore, they are essentially ballistic. Their thermodynamic intensive variables may experience variations with the photon energy and also with their direction of propagation. In fact, they come from the sun in a few directions only, and only in these directions do they exert a pressure. The direct consequence is that a non-vanishing current of grand potential exists for the photons. (It vanishes in gases of photons that are confined and in thermal equilibrium with the confining walls. Using this condition for photon beams from the sun is not correct.)

N_{ph} being the number of photons in a mode corresponding to a certain ray moving inside the semiconductor, their evolution along a given ray path corresponding to a radiation mode is given by [17]

$$N_{\text{ph}}(\varsigma) = f_{\text{BE}}(T, qV)\lfloor 1 - \mathrm{e}^{-\alpha\varsigma} \rfloor + N_{\text{ph}}(0)\mathrm{e}^{-\alpha\varsigma} \tag{4.31}$$

where ς is the length of the ray, f_{BE} is the Bose–Einstein factor for luminescent photons whose chemical potential is the separation between the conduction and the valence band electron quasi-Fermi levels – in this case equalling the cell voltage V (times q) – and α is the absorption coefficient. Equation (4.31) shows a non-homogeneous profile for N_{ph} contributed to by luminescent photons that increase with ς (first term on the right-hand side) and externally fed photons that decrease when the ray proceeds across the semiconductor (second term), and these photons are absorbed. $N_{\text{ph}}(0) = f_{\text{BE}}(T_s, 0)$ is usually taken in solar cells that correspond to illumination by free (i.e. with zero chemical potential) radiation at the sun temperature T_s.

In general, the photons in a mode of energy ε are considered as a macroscopic body [9] for which temperature and chemical potential can be defined. However, thermodynamically, they can be arbitrarily characterised by a chemical potential μ and a temperature T as long as $(\varepsilon - \mu)/T$ takes the same value. For example, the incident solar photons may be considered at the solar temperature T_s with zero chemical potential or, alternatively, at room temperature T_a with an energy variable chemical potential $\mu_s = \varepsilon(1 - T_a/T_s)$. This property has already been used in the study of the monochromatic cell.

Indeed, this arbitrary choice of T and μ does not affect the entropy production. This becomes evident if we rewrite equation (4.28) in the case of photons as

$$\sigma_{\text{ph}} = \sum_{i-\text{ph}} \left[\frac{\varepsilon - \mu}{T} g + \boldsymbol{j}_n \nabla \left(\frac{\varepsilon - \mu}{T} \right) + \nabla \cdot \left(-\frac{1}{T} \boldsymbol{j}_\omega \right) \right] \tag{4.32}$$

where we have made use of equation (4.15) and of the fact that $\upsilon = \varepsilon g$. This equation depends explicitly only on $(\varepsilon - \mu)/T$. This is less evident in the term in $\boldsymbol{j}_\omega/T$, but as

discussed in the context of equation (4.15), $\boldsymbol{j}_\omega$ is proportional to T and thus $\boldsymbol{j}_\omega/T$ only depends on $(\varepsilon-\mu)/T$. Note, however, that $\boldsymbol{j}_\omega$ is affected by the specific choice of T.

For simplicity, we shall consider the photons at room temperature and we shall calculate their chemical potential, $\mu_{\rm ph}$, from setting the equality

$$N_{\rm ph}(\varsigma,\varepsilon)=\frac{\varepsilon^2}{\exp\left[\dfrac{\varepsilon-\mu_{\rm ph}(\varsigma,\varepsilon)}{kT_a}\right]-1} \tag{4.33}$$

However, we might have chosen to leave $\mu_{\rm ph}=0$, and then the effect of the absorption of light when passing through the semiconductor could have been described as a cooling down of the photons (if $N_{\rm ph}$ actually decreases with ς).

With the room-temperature luminescent photon description, the production of entropy by photons is given by

$$\sigma_{\rm ph}=\sum_{i-{\rm ph}}\left[\frac{\upsilon_{i-{\rm ph}}}{T_a}-\frac{\mu_{i-{\rm ph}}g_{i-{\rm ph}}}{T_a}-\frac{\boldsymbol{j}_{n,i-{\rm ph}}\nabla\mu_{i-{\rm ph}}}{T_a}-\frac{\nabla\boldsymbol{j}_{\omega,i-{\rm ph}}}{T_a}\right] \tag{4.34}$$

However, using equations (4.15) and (4.3),

$$\nabla\boldsymbol{j}_\omega=\frac{c}{Un_r}\frac{{\rm d}\Omega}{{\rm d}\mu}\nabla\mu=-\frac{c}{Un_r}f_{\rm BE}\nabla\mu=-\boldsymbol{j}_n\nabla\mu \tag{4.35}$$

and

$$\sigma_{\rm ph}=\sum_{i-{\rm ph}}\left[-\frac{\upsilon_{i-{\rm ph}}}{T_a}+\frac{\mu_{i-{\rm ph}}g_{i-{\rm ph}}}{T_a}\right] \tag{4.36}$$

where

$$g=(c/Un_r)\alpha f_{\rm BE}(T_a,qV){\rm e}^{-\alpha\varsigma}-(c/Un_r)\alpha f_{\rm BE}(T_s,0){\rm e}^{-\alpha\varsigma};\quad \upsilon=\varepsilon g \tag{4.37}$$

The total irreversible entropy production is obtained by adding equations (4.29), (4.30) and (4.36), taking into account equation (4.37). Now, the terms in energy generation, all at the same temperature, must balance out by the first principle of thermodynamics. The net absorption of photons corresponds to an electron transfer (positive and negative generations) between states gaining an electrochemical potential qV, so the terms $\mu g/T$ corresponding to electrons and photons subtract each other. No additional generations are assumed to take place. Thus the total irreversible entropy generation rate can be written as

$$\sigma_{\rm irr}=\frac{c\alpha}{Un_r}\sum_{i-{\rm ph}}\frac{(\mu_{i-{\rm ph}}-qV)[f_{\rm BE}(T_s,0){\rm e}^{-\alpha\varsigma}-f_{\rm BE}(T_a,qV){\rm e}^{-\alpha\varsigma}]}{T_a} \tag{4.38}$$

For a given mode the second factor balances out when $f_{\rm BE}(T_s,0)=f_{\rm BE}(T_a,qV)$, and so does the irreversible entropy generation. In this case, $N_{\rm ph}=f_{\rm BE}(T_a,qV)$ is constant along the ray and $\mu_{i-{\rm ph}}=qV$ is constant at all points. The irreversible entropy generation rate is

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 865

then zero everywhere. If $f(T_s, 0) > f(T_a, qV)$, then $N_{\text{ph}} > f(T_a, qV)$ and $\mu_{i-\text{ph}} > qV$, so that both factors are positive and so is the product. If $f(T_s, 0) < f(T_a, qV)$, then $N_{\text{ph}} < f(T_a, qV)$ and $\mu_{i-\text{ph}} < qV$. In this case both factors are negative and the product is positive.

Thus, we have proven that every mode contributes non-negatively to the entropy. We can then state that the SQ cell produces non-negative entropy and, in this sense, it complies with the second law of thermodynamics.

For less idealised cases, as we have mentioned above, the existence of quasi-Fermi levels or temperature gradients generally produces additional positive entropy. Non-radiative net recombination of electrons from the conduction to the valence band also produces positive entropy. However, net generation would contribute to the production of negative entropy and, therefore, it may incur in violation of the second law of thermodynamics, if no other mechanism contributing to the creation of positive entropy exists. So the inclusion of imaginative carrier generation rates in novel device proposals, without counterparts, must be considered with some caution.

4.3.4 Entropy Production in the Whole Shockley–Queisser Solar Cell

The preceding approach is applicable to regions in which the physical properties of the system are continuous and differentiable, but not to abrupt interfaces. For testing the compliance with the second law, the continuity equations must be integrated in such cases by choosing volumes surrounding the suspected interfaces. This integral approach can also be extended to the whole converter to check for any violation of the second law and also to calculate the whole entropy production of a device. However, the integral approach, if not complemented with the local approach, is valid for proving thermodynamic inconsistency, but not for proving consistency that has to be proven at every point. An example of the use of such an integral approach is given here.

In the integral analysis we follow steps similar to those used in the local analysis. In particular, the first law of thermodynamics is applied by integrating equation (4.7) and using the first law expressed by equation (4.9). Then, we obtain, for the stationary case

$$0 = \int_A \sum_i \boldsymbol{j}_{e,i}\, \mathrm{d}A = +\dot{E}_r - \dot{E}_s + \dot{E}_{mo} - \dot{E}_{mi} + \dot{E}_{\text{others}} \tag{4.39}$$

where $\dot{E}_s$ and $\dot{E}_r$ are the radiation energies entering or escaping from the converter, $\dot{E}_{mi}$ and $\dot{E}_{mo}$ are the energies of the electrons entering the VB and leaving the CB, respectively, and $\dot{E}_{\text{others}}$ is the net flow of energy leaving the semiconductor as a result of other mechanisms.

Taking equation (4.5) into account, the fact that no chemical potential is associated with *other* particles and processes, and the annihilation of the grand canonical potential current density, the term corresponding to the other elements is more conventionally written as

$$\dot{E}_{\text{others}} = T_a \dot{S}_{\text{others}} \doteq \dot{Q} \tag{4.40}$$

where $\dot{Q}$ is defined as the rate of heat leaving the converter.

The second law of thermodynamics, expressed by equations (4.8) and (4.10), integrated into the whole volume of the converter can be written, for the stationary case, as

$$\dot{S}_{\text{irr}} = \int_U \sigma_{\text{irr}}\, \mathrm{d}U = \int_A \sum_i \boldsymbol{j}_{s,i}\, \mathrm{d}A = \dot{S}_r - \dot{S}_s + \dot{S}_{mo} - \dot{S}_{mi} + \dot{S}_{\text{others}} \tag{4.41}$$

The irreversible rate of entropy production is obtained by the elimination of the terms with subscript *others* by multiplying equation (4.41) by T_a, subtracting equation (4.39) from the result and taking into account (4.40). In this way we obtain

$$T_a \dot{S}_{\text{irr}} = (\dot{E}_s - T_a \dot{S}_s) - (\dot{E}_r - T_a \dot{S}_r) + (\dot{E}_{mi} - T_a \dot{S}_{mi}) - (\dot{E}_{mo} - T_a \dot{S}_{mo}) \tag{4.42}$$

From equation I-4 (Table 4.1) and considering the annihilation of the grand canonical potential flow for electrons, $(\dot{E}_{mi} - T_a \dot{S}_{mi}) = \varepsilon_{Fv} \dot{N}_{mi}$ and $(\dot{E}_{mo} - T_a \dot{S}_{mo}) = \varepsilon_{Fc} \dot{N}_{mo}$, so that we can write

$$T_a \dot{S}_{\text{irr}} = \varepsilon_{Fv} \dot{N}_{mi} - \varepsilon_{Fc} \dot{N}_{mo} + (\dot{E}_s - T_a \dot{S}_s) - (\dot{E}_r - T_a \dot{S}_r) \tag{4.43}$$

Taking into account that $\dot{N}_{mi} = \dot{N}_{mo} = I/q$ and $\varepsilon_{Fc} - \varepsilon_{Fv} = qV$, equation (4.43) is now rewritten as

$$T_a \dot{S}_{\text{irr}} = -\dot{W} + (\dot{E}_s - T_a \dot{S}_s) - (\dot{E}_r - T_a \dot{S}_r) \tag{4.44}$$

Here, this equation has been derived from the *local* model. However, it can also be obtained with more generality from a classical formulation of the second law of thermodynamics [32]. It is valid for ideal as well as for non-ideal devices. The values of the thermodynamic variables to be used in equation (4.44) are given in Table 4.1. Power – which in other cases will be that of the converter under study – corresponds in this case to the power of an SQ ideal solar cell and is given by the product of equations (4.18) and (4.19).

We have already discussed the basic ambiguity for the thermodynamic description of any radiation concerning the choice of the temperature and the chemical potential. A useful corollary is derived from this fact [32]. If the power rate produced by a radiation converter depends on the radiation only through its rate of incident energy or number of photons, then any radiation received or emitted by the converter can be changed into a luminescent radiation at room temperature, T_a, and with chemical potential, μ_x, without affecting the rate of power and of irreversible entropy produced. The chemical potential μ_x of this equivalent luminescent radiation is linked to the thermodynamic parameters of the original radiation, T_{rad} and μ_{rad}, through the equation

$$\frac{\varepsilon - \mu_{\text{rad}}}{kT_{\text{rad}}} = \frac{\varepsilon - \mu_x}{kT_a} \Rightarrow \mu_x = \varepsilon \left(1 - \frac{T_a}{T_{\text{rad}}}\right) + \mu_{\text{rad}} \frac{T_a}{T_{\text{rad}}} \tag{4.45}$$

Note that, in general, μ_x is a function of the photon energy, ε, as it may also be T_{rad} and μ_{rad}.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 867

The proof of the theorem is rather simple and is brought in here because it uses relationships of instrumental value. It is straightforward to see that

$$\dot{n}_x = \dot{n}_{\text{rad}}; \quad \dot{e}_x = \varepsilon \dot{n}_x = \dot{e}_{\text{rad}}; \quad \dot{\omega}_x / T_a = \dot{\omega}_{\text{rad}} / T_{\text{rad}}; \quad \dot{e}_{\text{rad}} - T\dot{s}_{\text{rad}} = \mu_x \dot{n}_x + \dot{\omega}_x \tag{4.46}$$

where, again, the suffix *rad* labels the thermodynamic variables of the original radiation.

The equality of the energy and the number of photons of the original and equivalent radiation proves that the power production is unchanged. Furthermore, the last relationship can be introduced in equation (4.44) to prove that the calculation of the entropy production rate also remains unchanged.

Taking all this into account, the application of equation (4.44) to the SQ solar cell is rather simple. Using the SQ cell model for the power and using the room-temperature equivalent radiation (of chemical potential μ_s) of the solar radiation, we obtain

$$\begin{aligned} T_a \dot{S}_{\text{irr}} &= -\int_{\varepsilon_g}^{\infty} qV[\dot{n}(T_a, \mu_s) - \dot{n}(T_a, qV)]\,\mathrm{d}\varepsilon + \int_{\varepsilon_g}^{\infty} [\mu_s \dot{n}(T_a, \mu_s) + \dot{\omega}(T_a, \mu_s)]\,\mathrm{d}\varepsilon \\ &\quad - \int_{\varepsilon_g}^{\infty} [qV\dot{n}(T_a, qV) + \dot{\omega}(T_a, qV)]\,\mathrm{d}\varepsilon \\ &= \int_{\varepsilon_g}^{\infty} [\dot{\omega}(T_a, \mu_s) - \dot{\omega}(T_a, qV)]\,\mathrm{d}\varepsilon + \int_{\varepsilon_g}^{\infty} [(\mu_s - qV)\dot{n}(T_a, \mu_s)]\,\mathrm{d}\varepsilon \end{aligned} \tag{4.47}$$

The integrand is zero for $qV = \mu_s$, but as μ_s varies with ε it is not zero simultaneously for all ε. To prove that it is positive, we differentiate with respect to qV. Using $\mathrm{d}\dot{\omega}/\mathrm{d}\mu = -\dot{n}$,

$$\frac{\mathrm{d}(T_a \dot{S}_{\text{irr}})}{\mathrm{d}qV} = -\int_{\varepsilon_g}^{\infty} [\dot{n}(T_a, \mu_s) - \dot{n}(T_a, qV)] = I/q \tag{4.48}$$

Thus, for each energy, the minimum of the integrand also appears for $qV = \mu_s(\varepsilon)$. Since this minimum is zero, the integrand is non-negative for any ε and so is the integral, proving also the thermodynamic consistence of this cell from the integral perspective. Furthermore, the minimum of the entropy, which for the non-monochromatic cell is not zero, occurs for open-circuit conditions. However, the ideal monochromatic cell reaches zero entropy production, and therefore reversible operation, under open-circuit conditions, and this is why this cell may reach the Carnot efficiency, as discussed in a preceding section.

4.4 THE TECHNICAL EFFICIENCY LIMIT FOR SOLAR CONVERTERS

We have seen that with the technical definition of efficiency – in whose denominator the radiation returned to the sun is not considered – the Carnot efficiency cannot be reached with a PV converter. What then is the technical efficiency upper limit for solar converters?

We can consider that this limit would be achieved if we could build a converter producing zero entropy [33]. In this case, the power that this converter can produce,

$\dot{W}_{\lim}$, can be obtained from equation (4.44) by setting the irreversible entropy-generation term to zero. Substituting the terms involving emitted radiation by their room-temperature luminescent equivalent, equation (4.44) becomes

$$\dot{W}_{\lim} = \int_{\varepsilon_g}^{\infty} \{[\dot{e}_s - T_a \dot{s}_s] - [\mu_x(\varepsilon)\dot{n}_x + \dot{\omega}_x]\}\, \mathrm{d}\varepsilon = \int_{\varepsilon_g}^{\infty} \dot{w}_{\lim}(\varepsilon, \mu_x)\, \mathrm{d}\varepsilon \tag{4.49}$$

The integrand should now be maximised [32] with respect to μ_x. For this, we calculate the derivative

$$\frac{\mathrm{d}\dot{w}_{\lim}}{\mathrm{d}\mu_x} = -\dot{n}_x - \frac{\mathrm{d}\dot{\omega}_x}{\mathrm{d}\mu_x} - \mu_x \frac{\mathrm{d}\dot{n}_x}{\mathrm{d}\mu_x} = -\mu_x \frac{\mathrm{d}\dot{n}_x}{\mathrm{d}\mu_x} \tag{4.50}$$

where we have used that the derivative of the grand potential with respect to the chemical potential is the number of particles with a change of sign. This equation shows that the maximum is achieved if $\mu_x = 0$ for any ε, or, in other words, if the radiation emitted is a room-temperature thermal radiation. This is the radiation emitted by all the bodies in thermal equilibrium with the ambient. However, the same result will be also achieved if the emitted radiation is any radiation whose room-temperature luminescent equivalent is a room-temperature thermal radiation.

Now, we can determine this efficiency, according to Landsberg [33], as

$$\begin{aligned} \eta &= \frac{\left(\frac{H_{sr}}{\pi}\right)\left[\left(\sigma T_s^4 - \frac{4}{3}\sigma T_a T_s^3\right) - \left(\sigma T_a^4 - \frac{4}{3}\sigma T_a^4\right)\right]}{(H_{sr}/\pi)\sigma T_s^4} \\ &= 1 - \frac{4}{3}\left(\frac{T_a}{T_s}\right) + \frac{1}{3}\left(\frac{T_a}{T_s}\right)^4 \end{aligned} \tag{4.51}$$

which for $T_s = 6000$ K and $T = 300$ K gives 93.33% instead of 95% of the Carnot efficiency.

No ideal device is known that is able to reach this efficiency. Ideal solar thermal converters, not considered in this chapter, have a limiting efficiency of 85.4% [34, 35] and therefore do not reach this limit. Other high-efficiency ideal devices considered in this chapter do not reach it either. We do not know whether the Landsberg efficiency is out of reach. At least it is certainly an upper limit of the technical efficiency of any solar converter.

4.5 VERY HIGH EFFICIENCY CONCEPTS

4.5.1 Multijunction Solar Cells

A conceptually straightforward way of overcoming the fundamental limitation of the SQ cell, already pointed out by SQ, is the use of several solar cells of different band gaps to convert photons of different energies. A simple configuration to achieve it is by just stacking the cells so that the upper cell has the highest band gap and lets the photons pass through towards the inner cells (Figure 4.5). The last cell in the stack is the one with the narrowest band gap. Between cells we put low-energy pass filters so that the

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 869

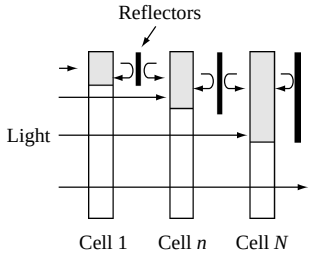


Figure 4.5 Stack of solar cells ordered from left to right in decreasing band gap ($E_{g1} > E_{g2} > E_{g3}$). (Reprinted from Solar Energy Materials and Solar Cells V. 43, N. 2, Martí A. and Araújo G. L, *Limiting Efficiencies for Photovoltaic Energy Conversion in Multigap Systems*, 203–222, © 1996 with permission from Elsevier Science)

reflection threshold of each filter is the band gap of the cell situated above. This prevents the luminescent photons from being emitted for energies different to those with which the photons from the sun are received in each cell. In this configuration, every cell has its own load circuit, and therefore, is biased at a different voltage. It has been shown [31] that a configuration without back reflectors leads to a lower efficiency if the number of cells is finite. For the case of a stack with an infinite number of cells, the limiting efficiency is found to be independent of whether reflectors are placed or not.

For example, for the case of a stack with two cells, the power generated is

$$W = qV_l\lfloor \dot{N}(T_s, 0, \varepsilon_{gl}, \varepsilon_{gh}, H_s) - \dot{N}(T_a, qV_l, \varepsilon_{gl}, \varepsilon_{gh}, H_r)\rfloor + qV_h[\dot{N}(T_s, 0, \varepsilon_{gh}, \infty, H_s) - \dot{N}(T_a, qV_h, \varepsilon_{gh}, \infty, H_r)] \tag{4.52}$$

where the suffixes l and h (low and high) refer to the band gap and the voltage of the two cells involved. The maximum power is obtained by optimising this function with respect to the variables V_l, V_h, ε_{gl} and ε_{gh}. We present in Figure 4.6 the efficiency of

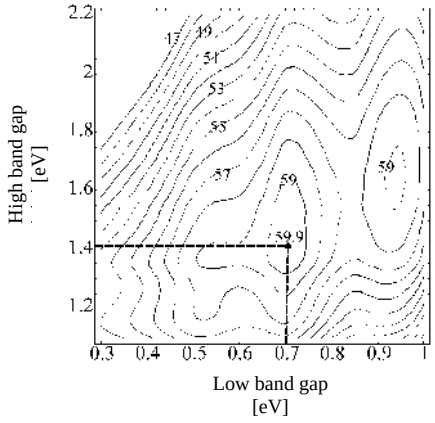


Figure 4.6 Efficiency of a tandem of two ideal cells under AM1.5D illumination as a function of the two cells' band gap ε_l and ε_h

two ideal cells as a function of the two cells' band gaps, which is optimised only with respect to V_l and V_h. The generation term is in this case not the one in equation (4.52) but the one corresponding to a standard spectrum AM1.5D [18]. Power is converted into efficiency by dividing by 767.2 Wm^{-2}, which is the power flux carried by the photons in this spectrum. In the figure, we can observe that the efficiency maximum is very broad, allowing for a wide combination of materials.

A lot of experimental work has been done on this subject. To our knowledge, the highest efficiency so far achieved, 34% (AM1.5 Global), has been obtained by Spectrolab [24, 36], in 2001, using a monolithic (made on the same chip) two-terminal tandem InGaP/GaAs stuck on a Ge cell and operating at a concentration factor of 210, that is, at 21 Wcm^{-2}.

The maximum efficiency is obtained with an infinite number of solar cells, each one biased at its own voltage $V(\varepsilon)$ and illuminated with monochromatic radiation. The efficiency of this cell is given by

$$\eta = \frac{\int_0^\infty \eta_{\mathrm{mc}}(\varepsilon)\dot{e}_s \,\mathrm{d}\varepsilon}{\int_0^\infty \dot{e}_s \,\mathrm{d}\varepsilon} = \frac{1}{\sigma_{\mathrm{SB}} T_s^4}\int_0^\infty \eta_{\mathrm{mc}}(\varepsilon)\dot{e}_s \,\mathrm{d}\varepsilon = \frac{1}{\sigma_{\mathrm{SB}} T_s^4}\int_0^\infty i(\varepsilon, V)V|_{\max}\,\mathrm{d}\varepsilon \quad (4.53)$$

where $\eta_{\mathrm{mc}}(\varepsilon)$ is the monochromatic cell efficiency given by equation (4.26) and $i(\varepsilon, V)$ was defined in equation (4.23). For $T_s = 6000$ K and $T_a = 300$ K, the sun and ambient temperature, respectively, this efficiency is [34] 86.8%. This is the highest efficiency limit of known ideal converter.

Tandem cells emit room-temperature luminescent radiation. This radiation presents, however, a variable chemical potential $\mu(\varepsilon) = qV(\varepsilon)$ and therefore it is not a radiation with zero chemical potential (free radiation). In addition, the entropy produced by this array is positive since the entropy produced by each one of the monochromatic cells forming the stack is positive. None of the conditions for reaching the Landsberg efficiency (zero entropy generation rate and emission of free radiation at room temperature) is then fulfilled and, therefore, tandem cells do not reach this upper bound.

It is highly desirable to obtain monolithic stacks of solar cells, that is, on the same chip. In this case, the series connection of all the cells in the stack is the most compact solution. Chapter 9 will deal with this case extensively. If the cells are series-connected, a limitation appears that the same current must go through all the cells. For the case of two cells studied above, this limitation is expressed by the equation

$$\begin{aligned} I &= q\lfloor \dot{N}(T_s, 0, \varepsilon_{gl}, \varepsilon_{gh}, H_s) - \dot{N}(T_a, qV_l, \varepsilon_{gl}, \varepsilon_{gh}, H_r)\rfloor \\ &= q[\dot{N}(T_s, 0, \varepsilon_{gh}, \infty, H_s) - \dot{N}(T_a, qV_h, \varepsilon_{gh}, \infty, H_r)] \end{aligned} \quad (4.54)$$

This equation establishes a link between V_l and V_h (for each couple ε_l, ε_h), which reduces the value of the maximum achievable efficiency. The total voltage obtained from the stack is $V = V_l + V_h$.

Our interest now is to determine the top efficiency achievable in this case when the number of cells is increased to infinity. Surprisingly enough, it is found [37, 38] that

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 871

the solution is also given by equation (4.53) and therefore, the limiting efficiency of a set of cells that is series-connected is also 86.8%.

4.5.2 Thermophotovoltaic Converters

Thermophotovoltaic (TPV) converters are devices in which a solar cell converts the radiation emitted by a heated body into electricity. This emitter may be heated, for example, by the ignition of a fuel. However, in our context, we are more interested in solar TPV converters in which the sun is the source of energy that heats an absorber at temperature T_r, which then emits radiation towards the PV converter. Figure 4.7 draws the ideal converter for this situation.

An interesting feature of the TPV converter is that the radiation emitted by the solar cells is sent back to the absorber assisting in keeping it hot. To make the cell area different from the radiator area (so far cells in this chapter have all been considered to be of area unity), the absorber in Figure 4.7 radiates into a reflecting cavity containing the cell. The reflective surfaces in the cavity and the mirrors in the concentrator are assumed to be free of light absorption.

In the ideal case [39], the radiator emits two bundles of rays, one of étendue H_{rc} (emitted by the right side of the absorber in the Figure 4.7a), which is sent to the cell after reflections in the cavity walls, with energy E_r' without losses, and another (unavoidable) one H_{rs}, with energy E_r, which is sent back to the sun (by the left face of the absorber) after suffering reflections in the left-side concentrator. Ideally, the radiator illuminates no other absorbing element or dark part of the sky: all the light emitted by the radiator's left side is sent back to the sun by the concentrator. Rays emitted by the radiator into the cavity may return to the radiator again without touching the cell, but since no energy is transferred, such rays are not taken into account.

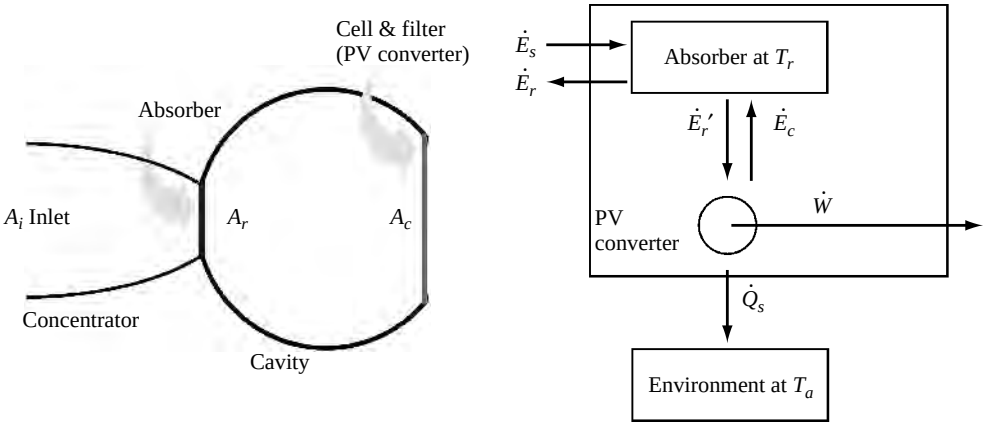


Figure 4.7 (a) Schematic of an ideal TPV converter with the elements inserted in loss-free reflecting cavities and (b) illustration of the thermodynamic fluxes involved. In the monochromatic case, $\dot{E}_s \equiv \dot{E}(T_s, 0, 0, \infty, H_{rs})$, $\dot{E}_r \equiv \dot{E}(T_r, 0, 0, \infty, H_{rs})$, $\dot{E}_r' \equiv \dot{e}(\varepsilon, T_r, 0, H_{rc})\Delta\varepsilon$, $\dot{E}_c \equiv \dot{e}(\varepsilon, T_a, qV, H_{rc})\Delta\varepsilon$

On the other hand, the radiator is illuminated by a bundle of rays, coming from the sun, of étendue $H_{sr} = H_{rs}$ and by the radiation emitted by the cell itself, of étendue $H_{cr} = H_{rc}$. Also, the cell may emit some radiation into the cavity, which returns to the cell again. This radiation is therefore not accounted for as an energy loss in the cell. In addition, we shall assume that the cell is coated with an ideal filter that allows only photons with energy ε and within a bandwidth $\Delta\varepsilon$ to pass through, while the others are totally reflected. In this situation the energy balance in the radiator becomes

$$\dot{E}(T_s, 0, 0, \infty, H_{rs}) + \dot{e}(\varepsilon, T_a, qV, H_{rc})\Delta\varepsilon = \dot{e}(\varepsilon, T_r, 0, H_{rc})\Delta\varepsilon + \dot{E}(T_r, 0, 0, \infty, H_{rs}) \tag{4.55}$$

where the first equation member is the net rate of energy received by the radiator and the second member is the energy emitted.

Using $\dot{e}(\varepsilon, T_r, 0, H_{rc})\Delta\varepsilon - \dot{e}(\varepsilon, T_a, qV, H_{rc})\Delta\varepsilon = \varepsilon\Delta i/q = \varepsilon\Delta\dot{w}/(qV)$, where Δi is the current extracted from the monochromatic cell and $\Delta\dot{w}$ is the electric power delivered, it is obtained that

$$\frac{\varepsilon\Delta\dot{w}}{qV} = \frac{H_{rs}\sigma_{SB}}{\pi}(T_s^4 - T_r^4) \Leftrightarrow \frac{\varepsilon^2\Delta i}{qH_{rc}\Delta\varepsilon} = \frac{H_{rs}\varepsilon}{H_{rc}\Delta\varepsilon}\frac{\sigma_{SB}}{\pi}(T_s^4 - T_r^4) \tag{4.56}$$

This equation can be used to determine the operation temperature of the radiator, T_r, as a function of the voltage V, the sun temperature T_s, the energy ε and the dimensionless parameter $H_{rc}\Delta\varepsilon/H_{rs}\varepsilon$. Notice that the left-hand side of equation (4.56) is independent of the cell étendue and of the filter bandwidth (notice that $\Delta i \propto H_{rc}\Delta\varepsilon$).

Dividing by $H_{rs}\sigma_{SB}T_s^4/\pi$, the solar input power, allows expressing the efficiency of the TPV converter as

$$\eta = \left(1 - \frac{T_r^4}{T_s^4}\right)\left(\frac{qV}{\varepsilon}\right) = \left(1 - \frac{T_r^4}{T_s^4}\right)\left(1 - \frac{T_a}{T_c}\right) \tag{4.57}$$

where T_c is the equivalent cell temperature as defined by equation (4.24). As presented in Figure 4.8, this efficiency is a monotonically increasing function of $H_{rc}\Delta\varepsilon/H_{rs}\varepsilon$. For $(H_{rc}\Delta\varepsilon/H_{rs}\varepsilon) \to \infty$, $\Delta i \to 0$ and $T_r \to T_c$. In this case an optimal efficiency [40] for $T_s = 6000$ K and $T_a = 300$ K is found to be 85.4% and is obtained for a temperature $T_c = T_r = 2544$ K. This is exactly the same as the optimum temperature of an ideal solar thermal converter feeding a Carnot engine. In reality, the ideal monochromatic solar cell is a way of constructing the Carnot engine. This efficiency is below the Landsberg efficiency (93.33%) and slightly below the one of an infinite stack of solar cells (86.8%).

It is worth noting that the condition $(H_{rc}\Delta\varepsilon/H_{rs}\varepsilon) \to \infty$ requires that $H_{rc} \gg H_{rs}$, and for this condition to be achieved, the cell area must be very large compared to the radiator area. This compensates the narrow energy range in which the cell absorbs. This is why a mirrored cavity must be used in this case.

4.5.3 Thermophotonic Converters

A recent concept for solar conversion has been proposed [3] with the name of thermophotonic (TPH) converter. In this concept, a solar cell converts the luminescent radiation

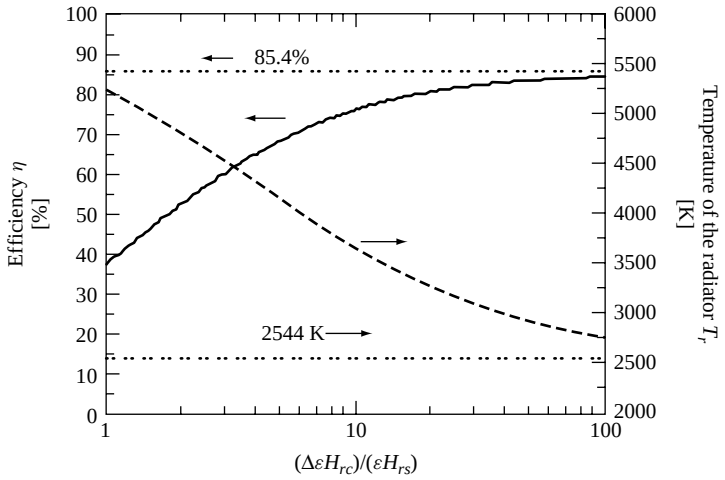


Figure 4.8 TPV ideal converter efficiency versus $H_{rc}\Delta\varepsilon/H_{rs}\varepsilon$. The energy ε and the cell voltage V are optimised

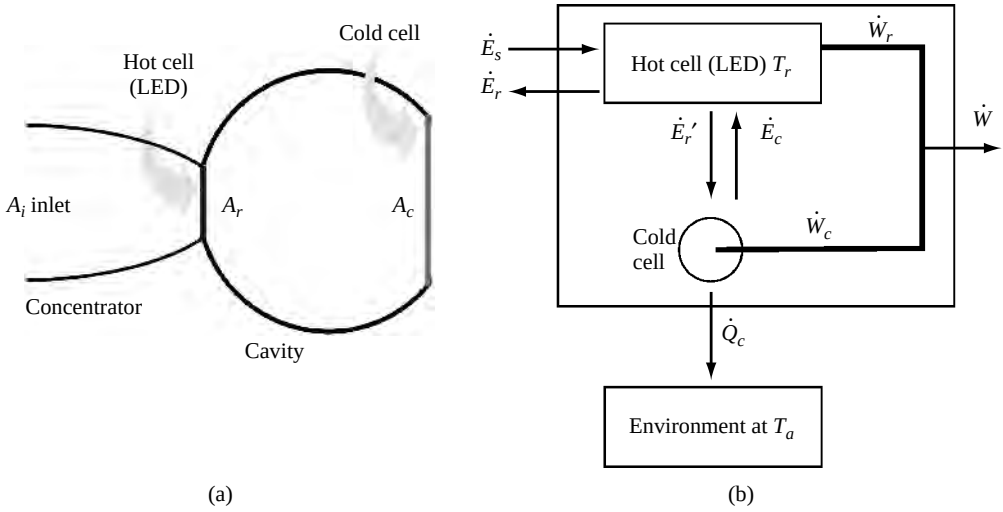


Figure 4.9 Diagram of the TPH converter with cavity. In (b), the energy fluxes are given by $\dot{E}_s \equiv E(T_s, 0, 0, \infty, H_{rs})$, $\dot{E}_r \equiv E(T_r, qV_r, 0, \infty, H_{rs})$, $\dot{Q}_r = \dot{E}_s - \dot{E}_r$, $\dot{E}_r' \equiv E(T_r, qV_r, 0, \infty, H_{rc})$, $\dot{E}_c' \equiv E(T_a, qV_c, \varepsilon_g, \infty, H_{rc})$

emitted by a heated light emitting diode (LED) into electricity. A diagram of this device is found in Figure 4.9. As in the TPV device, the LED can be heated with a fuel, but in our context it is heated as well with radiation absorbed from the sun. To emit luminescent radiation, the LED absorbs electric power, $\dot{W}_r$, in addition to the power delivered from the photons that illuminate the absorber. This power is to be subtracted from the electric power converted by the solar cell $\dot{W}_c$.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 874

As a mater of fact, this converter is a generalisation of the TPV converter. If the LED is in short circuit, the TPH converter is exactly the same as the TPV converter. The application of a voltage to the LED changes the converter properties.

Ideal LEDs are like solar cells. Thus, to describe the ideal TPH converter, we shall refer to the LED as the hot cell, as opposed to the electric power generator, which is the cold cell. When they are considered as ideal devices, the current–voltage characteristics of both, the LED and the solar cell, are the same. This causes the current passing by the cold cell to be faithfully replicated (with the sign changed) by the hot cell, owing to the reciprocal illumination of both devices, as expressed by the two following equations:

$$I_c/q = \dot{N}(T_r, qV_r, \varepsilon_g, \infty, H_{rc}) - \dot{N}(T_a, qV_c, \varepsilon_g, \infty, H_{rc}) \tag{4.58}$$

$$I_r/q = \dot{N}(T_a, qV_c, \varepsilon_g, \infty, H_{rc}) - \dot{N}(T_r, qV_r, \varepsilon_g, \infty, H_{rc}) \tag{4.59}$$

We shall use the subindex c for the cold solar cell current and voltage and r for the hot solar cell. The power generated by each cell is $\dot{W}_c = I_c V_c$ and $\dot{W}_r = I_r V_r$. In a normal operation, I_r is negative and the power generated by the hot cell is negative.

The application of the first law of thermodynamics to the cold and the hot cell, respectively, leads to

$$\dot{E}(T_r, qV_r, \varepsilon_g, \infty, H_{rc}) = \dot{E}(T_c, qV_c, \varepsilon_g, \infty, H_{rc}) + \dot{W}_c + \dot{Q}_c \tag{4.60}$$

$$\dot{Q}_r + \dot{E}(T_c, qV_c, \varepsilon_g, \infty, H_{rc}) = \dot{E}(T_r, qV_r, \varepsilon_g, \infty, H_{rc}) + \dot{W}_r \tag{4.61}$$

where $\dot{Q}_c$ is the heat rate delivered by the cold cell in the heat sink, while

$$\dot{Q}_r = \dot{E}(T_s, 0, 0, \infty, H_{rs}) - \dot{E}(T_r, 0, 0, \infty, H_{rs}) = \frac{H_{sr}\sigma_{\mathrm{SB}}}{\pi}(T_s^4 - T_r^4) \tag{4.62}$$

is the heat injected by the sun into the hot cell. The sum of equations (4.60) and (4.61) leads to

$$\dot{Q}_r - \dot{Q}_c = \dot{W}_r + \dot{W}_c = I_r(V_r - V_c) = I_c(V_c - V_r) = \dot{W} \tag{4.63}$$

where $\dot{W}$ is the algebraic sum of the powers produced by both cells (usually the hot cell will be absorbing, not producing, power).

In the case of monochromatic cell and LED operation, a proper filter of bandwidth $\Delta\varepsilon$ and centred at the energy ε is to be located somewhere in the optical system to allow for an interchange of photons only within the filter bandwidth. The preceding equations can be easily particularised in this case. An interesting relationship holding in this case, $\dot{e}\Delta\varepsilon = \varepsilon\dot{n}\Delta\varepsilon \equiv \dot{E}$, allows for a simple expression of the heat rate in both cells,

$$\dot{Q}_r = (V_r - \varepsilon/q)I_r \tag{4.64}$$

$$\dot{Q}_c = (\varepsilon/q - V_c)I_c \tag{4.65}$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 875

and this allows for writing the efficiency, using the equations (4.63) to (4.65) as

$$\eta = \frac{\dot{Q}_r}{H_{sr}\sigma_{\mathrm{SB}}T_s^4/\pi}\,\frac{\dot{W}}{\dot{Q}_r} = \left(1-\frac{T_r^4}{T_s^4}\right)\left(1-\frac{\dot{Q}_c}{\dot{Q}_r}\right) = \left(1-\frac{T_r^4}{T_s^4}\right)\left(\frac{qV_c/\varepsilon - qV_r/\varepsilon}{1-qV_r/\varepsilon}\right) \tag{4.66}$$

As said before, the TPH converter has an extra degree of freedom in the design, as compared to the TPV. It is the hot cell voltage V_r. The hot cell temperature in the monochromatic case depends on the same parameter $H_{rc}\Delta\varepsilon/H_{rs}\varepsilon$ as in the TPV case. When this parameter tends to infinity, the cold cell tends to be in open circuit, V_{coc}, given by

$$\frac{(\varepsilon - qV_{\mathrm{coc}})}{kT_a} = \frac{(\varepsilon - qV_r)}{kT_r} \Rightarrow qV_{\mathrm{coc}} = \varepsilon\left(1-\frac{T_a}{T_r}\right) + qV_r\frac{T_a}{T_r} \tag{4.67}$$

Using this equation in equation (4.66), we obtain the same limit equation for the efficiency as the one for the TPV converter (equation 4.57):

$$\eta = \left(1-\frac{T_r^4}{T_s^4}\right)\left(1-\frac{T_a}{T_r}\right) \tag{4.68}$$

An interesting feature of this formula is that it holds for any value of V_r. Thus the TPH converter has the same upper limit as the TPV one. Furthermore, T_r is the same as in the TPV.

To our knowledge, this novel concept has not yet been fully explored. However, we present in Figure 4.10 the efficiency versus the cold cell voltage V_c (usually, in other contexts, we call efficiency to the maximum of this curve) for V_r in forward bias, zero bias (TPV mode) and reverse bias. In Figure 4.11, we present the hot cell temperature. We can

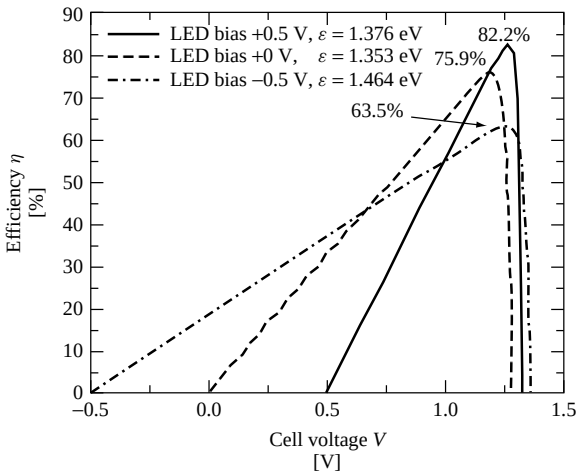


Figure 4.10 TPH efficiency versus V_c for three values of V_r for $H_{rc}\Delta\varepsilon/H_{rs}\varepsilon = 10/\pi$. The energy ε has been optimised

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 876

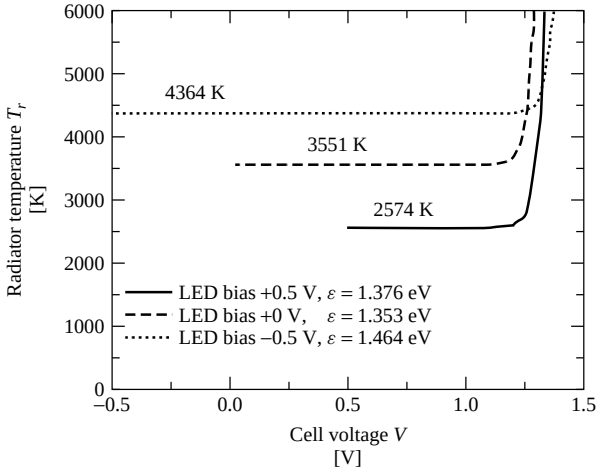


Figure 4.11 Hot cell temperature as a function of the cold cell voltage for several biasing voltages of the hot cell (LED) and for the energy that leads to maximum efficiency ($H_{rc}\Delta\varepsilon/H_{rs}\varepsilon = 10/\pi$). The energy ε has been optimised

see that higher efficiencies at lower temperatures can be obtained using the TPH concept under direct bias. This is not in contradiction to the fact that the limiting efficiency and the temperature for it is the same in both cases.

4.5.4 Higher-than-one Quantum Efficiency Solar Cells

One of the drawbacks that limit the efficiency of single-junction solar cells is the energy wasted from each photon that is absorbed because it is not converted into electrical power. Werner, Kolodinski, Brendel and Queisser [41, 42] have proposed a cell in which each photon may generate more than one electron–hole pair, thus leading to higher-than-one quantum efficiency solar cells. Without discussing which physical mechanisms may allow for this behaviour, let us examine its implications. Admitting that every photon may create $m(\varepsilon)$ electron–hole pairs, the current extracted from the device would be given by

$$I/q = \int_{\varepsilon_g}^{\infty} [m(\varepsilon)\dot{n}_s - m(\varepsilon)\dot{n}_r(T, \mu)]\,\mathrm{d}\varepsilon \tag{4.69}$$

In this equation, ε_g is the energy threshold for photon absorption and the factor m in the generation term is our initial assumption. The same term must appear in the recombination term to reach the detailed balance: if the sun temperature is brought to the ambient temperature, the current will be zero when $\mu = 0$, only if the factor m also appears in the recombination term. For the moment we are saying nothing about the chemical potential μ of the photons emitted.

The power delivered, $\dot{W}$, is given by

$$\dot{W} = \int_{\varepsilon_g}^{\infty} qV[m\dot{n}_s - m\dot{n}_r(T, \mu)]\,\mathrm{d}\varepsilon \tag{4.70}$$

Let us consider a monochromatic cell and calculate the irreversible entropy generation rate $\dot{S}_{\rm irr}$ in the whole device. With the aid of the general equation (4.44) and equation I-4 in Table 4.1, it is given by

$$T_a \dot{S}_{\rm irr}/\Delta\varepsilon = (\mu_x \dot{n}_x + \dot{\omega}_x) - (\mu \dot{n}_r + \dot{\omega}_r) - qV(m\dot{n}_x - m\dot{n}_r) \tag{4.71}$$

where the source of photons has been substituted by its equivalent room-temperature luminescent radiation characterised by the chemical potential μ_x and the ambient temperature T_a. The open-circuit conditions are achieved when $\mu_{\rm OC} = \mu_x$. For this value the entropy rate is zero since then $\dot{n}_x = \dot{n}_r$ and $\dot{\omega}_x = \dot{\omega}_r$.

Let us calculate the derivative of the irreversible entropy generation rate (equation 4.71) with respect to μ and particularise it for the open-circuit value of μ. Considering what follows V as only an unknown function of μ and independent of the way of obtaining the excitation (which is the case for infinite mobility) and using the fundamental relationship $\partial\dot{\omega}_r/\partial\mu = -\dot{n}_r$, the result is

$$\left[\frac{{\rm d}(T_a \dot{S}_{\rm irr}/\Delta\varepsilon)}{{\rm d}\mu}\right]_{\mu_{\rm OC}} = (qmV_{\rm OC} - \mu_{\rm OC})\left[\frac{{\rm d}\dot{n}_r}{{\rm d}\mu}\right]_{\mu_{\rm OC}} \tag{4.72}$$

This derivative is only zero if $qmV_{\rm OC} = \mu_{\rm OC}$. Since $\mu_{\rm OC} = \mu_x$ can take any value by changing the source adequately, we obtain the result $qmV = \mu$. Any other value would produce a negative rate of entropy generation in the vicinity of the open circuit, against the second law of thermodynamics. This is a demonstration, based on the second law of thermodynamics, of the relationship between the chemical potential of the photons and the voltage (or electron and hole quasi-Fermi level split).

If we could choose m freely, the maximum power is achieved if we can maximise the integrand of equation (4.70) for each value of the energy [43, 44]. Once this is done, the reduction in ε_g so that it tends towards zero increases the power output. For the limit of $\varepsilon_g \to 0$, the maximum efficiency is the same as in equation (4.53), where a stack of an infinite number of cells was studied. Here $qVm(\varepsilon)$ is the variable that plays the same role as $qV(\varepsilon)$ earlier, although here V is the same for all the terms. In consequence, the upper efficiency is the same as for the tandem cell stack, 86.8%.

The higher-than-one quantum efficiency behaviour has been actually found [45, 46], although very close to one, for visible photons of high-energy and UV photons. The effect is attributed to impact ionisation, a mechanism in which the electron or the hole created by the high-energy photon, instead of thermalising by scattering with phonons, by means of impact processes transfers its high energy to a valence-band electron that gets pumped into the conduction band. This mechanism has a detailed balance counterpart that is the Auger recombination, in which the energy recovered in the recombination is transferred to an electron or a hole, which thus acquires a high kinetic energy.

4.5.5 Hot Electron Solar Cells

Wurfel [47] has studied the impact ionisation cells from another perspective and has shown that they become identical to the hot carrier solar cells proposed by Ross and Nozik [48]

pointing out also in this respect the necessity of providing the cell with special contacts as will be explained below.

In the hot carrier solar cell, the electron gas (semiconductor properties exhibiting a band gap does not seem to be essential for its operation) is in equilibrium at the lattice temperature T_a despite the absence of phonon interaction. This equilibrium is reached through other interactions as, for example, the interaction with thermal photons. Out of the equilibrium, electrons are thermalised by elastic scattering between them that allows a set of energy-dependent chemical potentials, one for the electrons of each energy. Furthermore, transference of energy between electrons exists through mechanisms that include impact ionisation and Auger recombination (this terminology is used in the context of semiconductors, which, as mentioned, we think is not necessary for this cell) according to reactions like $e_1^- + e_2^- \leftrightarrow e_3^- + e_4^-$. Actually in this equation transference of energy between electrons is produced by couples so that $\hat{\varepsilon}_1 + \hat{\varepsilon}_2 = \hat{\varepsilon}_3 + \hat{\varepsilon}_4$ where $\hat{\varepsilon}$ is electron (not photon) energy. However, the equilibrium of the reaction also establishes that the electrochemical potentials of the electrons (that we still call ε_F) are related by $\varepsilon_F(\hat{\varepsilon}_1) + \varepsilon_F(\hat{\varepsilon}_2) = \varepsilon_F(\hat{\varepsilon}_3) + \varepsilon_F(\hat{\varepsilon}_4)$. Therefore, the electrochemical potential of the electrons is a linear function of the energy in the form

$$\varepsilon_F(\hat{\varepsilon}) = \beta\hat{\varepsilon} + \varepsilon_{F0} \tag{4.73}$$

where the Fermi energy reference ε_{F0} becomes the same for both electrons and holes.

Since the Fermi function for electrons is written as

$$\frac{1}{\exp\left[\dfrac{\hat{\varepsilon} - \beta\hat{\varepsilon} - \varepsilon_{F0}}{kT_a}\right] + 1} = \frac{1}{\exp\left[\dfrac{\hat{\varepsilon} - \varepsilon_{F0}/(1-\beta)}{kT_a/(1-\beta)}\right] + 1} \tag{4.74}$$

the electron distribution can be regarded equivalently as a distribution at the lattice temperature T_a but characterised by a varying electrochemical potential given by equation (4.73) or as a hot carrier distribution with a constant electrochemical potential given by $\mu_{hc} = \varepsilon_{F0}/(1-\beta)$ and a *hot carrier* temperature $T_{hc} = T_a/(1-\beta)$.

If the interaction with phonons is introduced, through the reaction $e_1^- + \text{phonon} \leftrightarrow e_2^-$, where $e_{1(2)}^-$ represents an electron with energy $\hat{\varepsilon}_{1(2)}$, the fact that the phonon chemical potential is zero leads to $\varepsilon_F(\hat{\varepsilon}_1) = \varepsilon_F(\hat{\varepsilon}_2)$. However, $\hat{\varepsilon}_1 \neq \hat{\varepsilon}_2$ but $\hat{\varepsilon}_2 = \hat{\varepsilon}_1 + \varepsilon$, where ε is the phonon energy and, therefore, $\varepsilon_F(\hat{\varepsilon}_1) = \varepsilon_F(\hat{\varepsilon}_2)$ is only fulfilled if $\beta = 0$ and consequently all the electrons have the same electrochemical potential ε_{F0} and are at the lattice temperature. Phonon interaction is not considered to occur in what follows.

If the interaction with photons is considered now, $e_1^- + \text{photon} \leftrightarrow e_2^-$, the equilibrium of the reaction is represented by $\varepsilon_F(\hat{\varepsilon}_1) + \mu_{\text{ph}} = \varepsilon_F(\hat{\varepsilon}_2)$ where μ_{ph} is the photon chemical potential. Taking into account equation (4.73), it is obtained that $\mu_{\text{ph}} = \beta(\hat{\varepsilon}_1 - \hat{\varepsilon}_2) = \beta\varepsilon$ where ε is the energy of the photon involved. The result is that we have an energy-dependent photon chemical potential μ_{ph}. The Bose function that describes the occupation probability of the photon energy level ε becomes

$$\frac{1}{\exp\left[\dfrac{\varepsilon - \beta\varepsilon}{kT_a}\right] - 1} = \frac{1}{\exp\left[\dfrac{\varepsilon}{kT_a/(1-\beta)}\right] - 1} \tag{4.75}$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 879

which can be seen as free radiation (zero chemical potential) at the hot carrier temperature.

Würfel also points out that, provided we can find a material in which a weak phonon coupling is produced in order to manufacture a solar cell, we still should be careful with the contacts because they do not allow for the maintenance of hot carriers inside the absorbing material. These special contacts must be able to cool the electrons from the hot temperature T_{hc} to the contact temperature T_a reversibly by changing their electrochemical potential from ε_{F0} to the electron electrochemical potential at the contacts (Fermi level at the metals) ε_{F+} and ε_{F-}. In Wurfel's Reference [47], these special contacts are devised as selective membranes (Figure 4.12) that only allow electrons with energy centred around ε_e (left contact) and ε_h (right contact) to pass through. The reversible change in temperature and electrochemical potential of the electrons at the membranes is obtained by setting

$$\begin{aligned}\frac{\hat{\varepsilon}_e - \mu_{hc}}{kT_{hc}} &= \frac{\hat{\varepsilon}_e - \varepsilon_{F-}}{kT_a} \Leftrightarrow \varepsilon_{F-} = \hat{\varepsilon}_e\left(1 - \frac{T_a}{T_{hc}}\right) + \mu_{hc}\frac{T_a}{T_{hc}} \\ \frac{\hat{\varepsilon}_h - \mu_{hc}}{kT_{hc}} &= \frac{\hat{\varepsilon}_h - \varepsilon_{F+}}{kT_a} \Leftrightarrow \varepsilon_{F+} = \hat{\varepsilon}_h\left(1 - \frac{T_a}{T_{hc}}\right) + \mu_{hc}\frac{T_a}{T_{hc}}\end{aligned} \tag{4.76}$$

where $T_{hc} = T_a/(1-\beta)$. The cell voltage will therefore be given by

$$qV = \varepsilon_{F-} - \varepsilon_{F+} = (\hat{\varepsilon}_e - \hat{\varepsilon}_h)\left(1 - \frac{T_a}{T_{hc}}\right) \tag{4.77}$$

The current extracted from the cell is determined by the rate at which electron–hole pairs of energy $\hat{\varepsilon}_e - \hat{\varepsilon}_h$ can be withdrawn from the cell. Since no energy is lost as heat, the energy balance equation (first principle) leads to

$$I(\hat{\varepsilon}_e - \hat{\varepsilon}_h)/q = \dot{E}(T_s, 0, \varepsilon_g, \infty, H_s) - \dot{E}(T_{hc}, 0, \varepsilon_g, \infty, H_r) \tag{4.78}$$

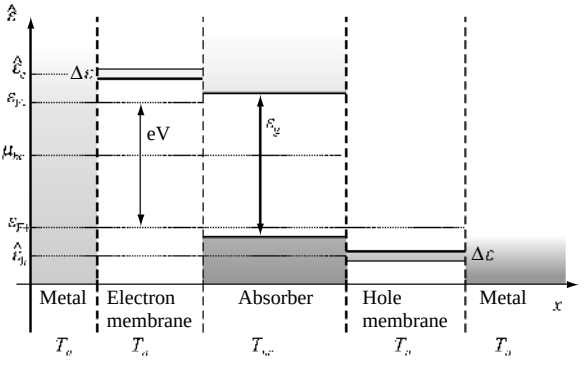


Figure 4.12 Band structure of a hot electron solar cell showing contacting scheme by means of selective membranes. (Reprinted from Solar Energy Materials and Solar Cells V. 46, N. 1, Würfel P., *Solar Energy Conversion with Hot Electrons from Impact Ionisation,* 43–52, © 1997 with permission from Elsevier Science)

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 880

and the power extracted from the cell can be finally computed as

$$\dot{W} = IV = [\dot{E}(T_s, 0, \varepsilon_g, \infty, H_s) - \dot{E}(T_{hc}, 0, \varepsilon_g, \infty, H_r)]\left(1 - \frac{T_a}{T_{hc}}\right) \tag{4.79}$$

which is independent of the carrier-extracting energies of the contacts. In other words, a large separation of the extracting energies leads to high voltage and low current and *vice versa*. Note that T_{hc} is a parameter for equation (4.79) and for equation (4.77). By elimination, we obtain the $W(V)$ and from it the derived IV curve. The cell efficiency is obtained from the maximum of $W(V)$ or $W(T_{hc})$.

For $\varepsilon_g \to 0$ the limit efficiency becomes

$$\eta = \left(1 - \frac{T_{hc}^4}{T_s^4}\right)\left(1 - \frac{T_a}{T_{hc}}\right) \tag{4.80}$$

just as in the TPV converters, leading to a limiting efficiency of 85.4%.

The monoenergetic membrane for electron and hole transfer to the contacting metals might perhaps be an insulator with an impurity band, but the nature of the phonon-insulated absorber is totally unknown. Ross and Nozik [48, 49] who have developed concepts close to those expressed here are investigating the poorer coupling of the electrons with the photons in materials with quantum dots with the idea of favouring this hot electron type of cell or the impact ionisation one that we have described in the preceding section.

4.5.6 Intermediate Band Solar Cell

One of the causes of efficiency reduction in single-junction solar cells is the transparency of the semiconductor to sub−band gap photons [50]. The inclusion of an intermediate band (IB) may greatly increase the efficiency. We show in Figure 4.13 a band diagram of the photon absorption and emission in this intermediate band material. Photons are absorbed not only by pumping electrons from the VB to the CB as in a traditional solar cell (photons with energy $h\nu_3$) but also by transitions from the VB to the IB (photons with energy $h\nu_2$) and from the IB to the CB (photons with energy $h\nu_1$). In total, two low-energy photons are used to pump an electron from the valence band to the conduction band, passing through the intermediate band. This certainly increases the cell current.

The three absorption mechanisms detailed above are effective if the IB is a band partially filled with electrons. In this way, there are empty states in it to accommodate the electrons from the VB and there are electrons to sustain a strong pumping to the CB. The detailed balance imposes photon emissions that are opposite to each one of the three absorption mechanisms.

The cell is contacted as shown in Figure 4.14. The electrons are extracted from the VB and returned to the CB using two layers of n and p ordinary semiconductors. A high voltage is produced as the sum of the two semiconductor junction voltages occurring at both sides of the IB material.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 881

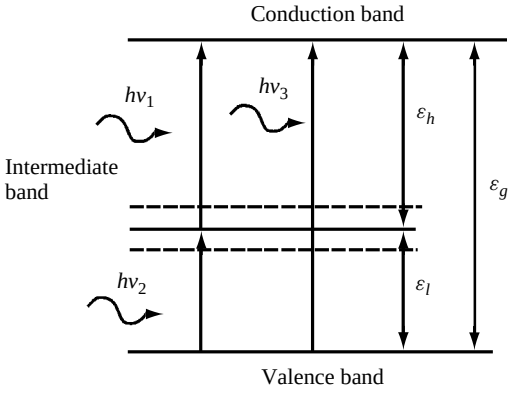


Figure 4.13 The band diagram of an intermediate band solar cell

In this IB material we admit that there are three quasi-Fermi levels, one for each band. The cell voltage is q times the splitting of the CB and VB quasi-Fermi levels.

In general, there is an energy threshold for each one of the absorption mechanisms described earlier. However, the ideal structure is the one in which the upper energy (lintel) of a photon that can be absorbed in certain mechanisms – involving, for example, transitions from the VB to the IB – is the threshold of the next one – for example, transitions from the IB to the CB. More specifically, calling ε_g the energy interval between the conduction and the valence band, ε_l the interval between the Fermi level in the intermediate band and the valence band and $\varepsilon_h = \varepsilon_g - \varepsilon_l$, and assuming that $\varepsilon_l < \varepsilon_h$, we consider that all the photons in the interval $(\varepsilon_l, \varepsilon_h)$ are absorbed by transitions from the VB to the IB, all the photons in the interval $(\varepsilon_h, \varepsilon_g)$ are absorbed by transitions from the IB to the CB and all the photons in the interval (ε_g, ∞) are absorbed by transitions from the VB to the CB.

Under such conditions the equations that rule the current–voltage characteristic of the cell are

$$I/q = \lfloor \dot{N}(T_s, 0, \varepsilon_g, \infty, \pi) - \dot{N}(T, \mu_{CV}, \varepsilon_g, \infty, \pi) \rfloor + [\dot{N}(T_s, 0, \varepsilon_h, \varepsilon_g, \pi) - \dot{N}(T, \mu_{CI}, \varepsilon_h, \varepsilon_g, \pi)] \tag{4.81}$$

$$\dot{N}(T_s, 0, \varepsilon_l, \varepsilon_h, \pi) - \dot{N}(T, \mu_{IV}, \varepsilon_l, \varepsilon_h, \pi) = \dot{N}(T_s, 0, \varepsilon_h, \varepsilon_g, \pi) - \dot{N}(T, \mu_{CI}, \varepsilon_h, \varepsilon_g, \pi) \tag{4.82}$$

with μ_{XY} being the photon chemical potentials equal to the separation of quasi-Fermi levels between band X and band Y. Equation (4.81) states the balance of electrons in the CB (addition of those that are pumped from the IB and the VB by means of the absorption of the corresponding photon less those that radiatively recombine). Equation (4.82) states a similar balance equation for the electrons in the IB and takes into account that no current is extracted from the IB. In addition,

$$qV = \mu_{CV} = \mu_{CI} + \mu_{IV} \tag{4.83}$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 882

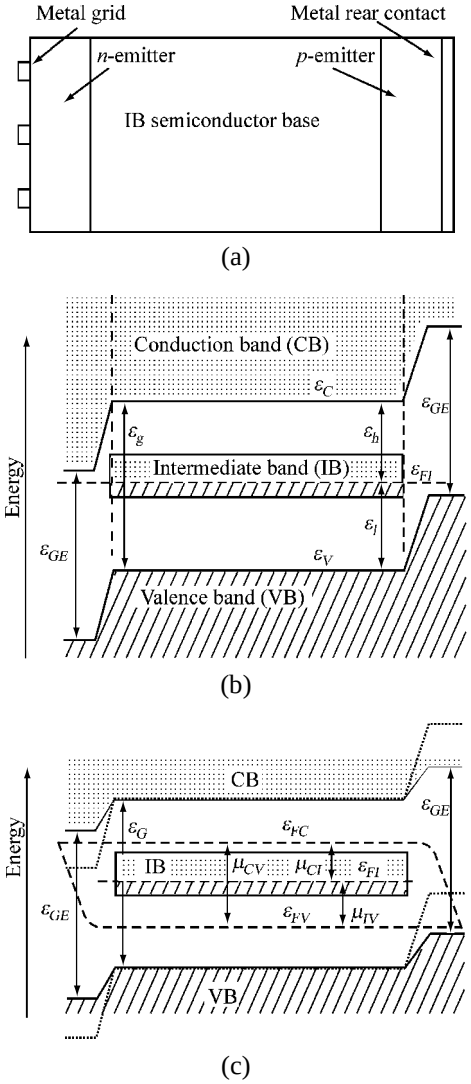


Figure 4.14 (a) Structure of the IB solar cell; (b) band diagram in equilibrium and (c) band diagram of forward bias conditions. (Reproduced from *A Metallic Intermediate Band High Efficiency Solar Cell*, Luque A. and Martí A., by permission of John Wiley & Sons Limited © 2001)

By eliminating μ_{CI} and μ_{IV} from the last three equations, we obtain the current–voltage characteristic of the cell. Efficiencies for different values of ε_l and ε_g are plotted in Figure 4.15.

The maximum efficiency of 63.2% is achieved for a cell of gap 1.95 eV with the IB Fermi level located at 0.71 eV from one of the bands. This efficiency is higher than the one corresponding to two series-connected ideal cells in tandem, of 54.5% (for band gaps of 0.8 and 1.54 eV). A detailed analysis of this cell operation can be found in the References [50–52]. The generalisation of the concept to more than two intermediate

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 883

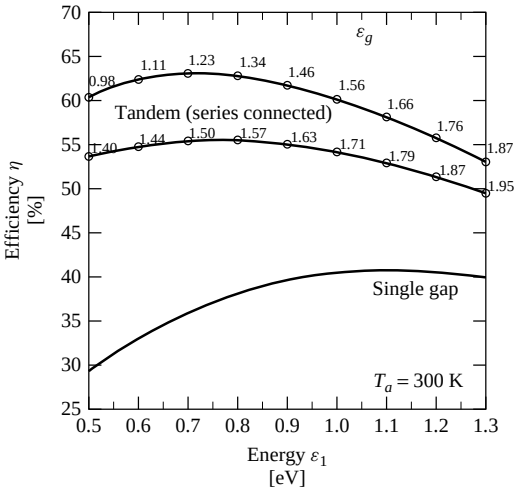


Figure 4.15 Limiting efficiency for the IB solar cell under maximum concentrated sunlight. For comparison, the limiting efficiency of single-gap solar cells (in this case, $\varepsilon_l \equiv \varepsilon_g$) and a tandem of two cells that are series-connected are also shown (in this case, ε_l labels the lowest band gap of the cells and the figures on the curve correspond to the highest of the band gaps). (Reproduced from *Increasing the Efficiency of Ideal Solar Cells by Photon Induced Transitions at Intermediate Levels*, Luque A. and Martí A., Physical Review Letters V. 78 N. 34, 5014–5017, by permission of American Physical Society © 1997)

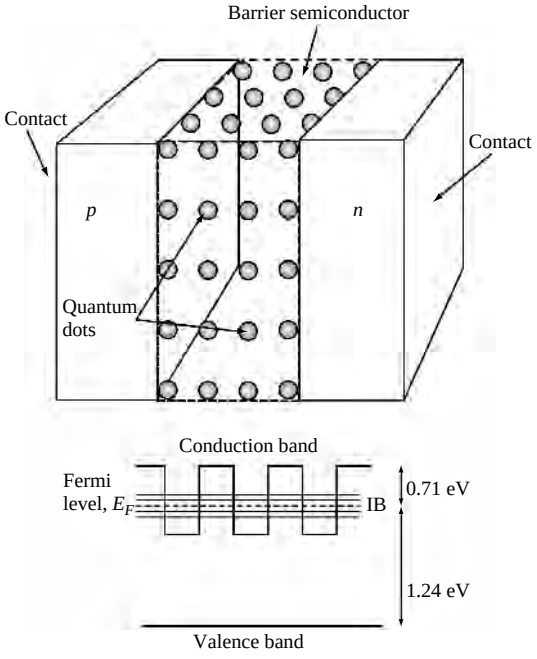


Figure 4.16 IB formation in QD arrays

band gaps (multiband solar cells) has been studied in [53, 54]. Some related concepts are that of an impurity band, in which the IB is made up of impurities [55], and the use of intermediate band materials as up- and down-converters [56, 57].

Suitable materials for a metallic IB are not very common. Accurate theoretical calculations establish that some alloys of the type III_3-V_4-IVb (with four electrons in the outer shell) present the required band [58]. No experiment has been carried out so far for implementing these materials, and therefore it is not certain whether the required metallic IB is actually formed or if the corresponding alloys are thermodynamically stable or separate into phases.

Another proposal [59] for forming the required IB is the use of nanotechnology. In particular, closely spaced quantum dot arrays might produce the desired band Figure 4.16. $In_{0.58}Ga_{0.42}As$ dots (of band gap 0.87 eV) of a radius of 39 nm in a barrier material of $Al_{0.4}Ga_{0.6}As$ of band gap 1.95 eV could produce an IB centred at 0.71 eV from the CB (if strain effects could be avoided and the offset in the VB could be suppressed [60]). The half filling of the IB can be achieved by barrier material doping [61, 62]. Experiments to prove the principle of operation of this type of cell have not yet been carried out.

4.6 CONCLUSIONS

In this chapter we have provided a thermodynamic basis that allows evaluating the thermodynamic consistency of classical and newly proposed solar cells. Also, we have assessed the efficiency limit of several PV concepts.

As deduced by Shockley and Queisser, the upper limit reachable with a single-junction solar cell is 40.7% for the cell illuminated by a black body source of photons at 6000 K and assuming the cell temperature at 300 K. This value is rather low if we take into account that the Carnot limiting efficiency for a reversible engine operating between hot and cold heat reservoirs at 6000 K and 300 K, respectively, is 95%. High-efficiency devices that can ideally surpass the Shockley–Queisser efficiency have been called *third generation PV converters*. Thus, the following question arises: could we invent a solar converter that exhibits this Carnot efficiency?

The answer is negative and the reason for it lies in the definition of efficiency. In the definition of the Carnot efficiency, the term entering in the denominator is the power *consumed*, that is, the power *arriving* at the converter less the power *leaving* the converter owing to the radiation that is emitted. In the conventional definition of the efficiency for solar converters, the term entering in the denominator is the power *arriving* at the converter and because it is higher than the *consumed* power this leads to a lower efficiency. With this definition, the higher achievable efficiency is the Landsberg efficiency of 93.33%. However, this efficiency, if reachable, cannot be reached with any known ideal solar converter.

A very high efficiency of 85.4% can ideally be reached with several devices such as the TPV converter, constituted by one ideal solar cell and one absorber and the TPH converter, conceived with one ideal solar cell and one LED that also plays the role of an absorber, or even a hot carrier solar cell. To reach this efficiency they all must emit radiation with zero chemical potential (free radiation) at 2544 K. This efficiency is also the limit for solar thermal devices.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 885

However, this is not the highest efficiency that can be reached in solar converters. Efficiencies of up to 86.8% can be achieved using an array of solar cells of different band gap, either series-connected or independently connected.

Except for the TPH case, all solar cells operate at ambient temperature. This is a highly desirable feature. The TPH concept, which allows for higher efficiencies at lower temperatures than the TPV concept, may bring some advances. However, such devices to be practical require an almost ideal external quantum efficiency of the LED and the ability of working at high temperatures, both requirements being very difficult to acquire.

Experimental work on IB solar cells – whose upper limit efficiency is 63.2% – is now starting. It is conceivable that IBs may substitute two junction materials with perhaps less complexity. They may also be combined in tandem with more ordinary or IB cells.

Finally, it seems unlikely to us that a semiconductor can be found where the electrons are uncoupled enough from the phonons as to allow effective impact ionisation cells. However, a multilevel organic dye may be found or an array of quantum dots may be engineered where this coupling is reduced. In any case, the operation principles of this cell, and not only that of the usual solar cells, should be taken into account by those trying to bring other technologies into the manufacture of PV converters.

From the preceding statements it is clear that very high efficiencies are possible in a device operating in the PV mode. But what is possible in practice? This is a very difficult question. Certainly, stimulated by space research, there is a trend towards the development of multijunction cells. We stress again here that efficiencies of 34% (Global AM1.5) with a monolithic tandem of InGaP/GaAs, stuck on a Ge cell operating at 212 suns have been achieved. It can be asked whether such solutions are not too expensive for terrestrial use. However, we disagree. The use of very high concentration elements, in the range of 1000 or more, may make very expensive cells [63] usable. Wide acceptance angle concentrators with this concentration factor have already been developed that seem very suitable for mounting in low-cost tracking structures [64].

REFERENCES

1. Lofersky J, *Postepy-Fizkyi* **26**, 535–560 (1975).
2. Shockley W, Queisser H, *J. Appl. Phys.* **32**, 510–519 (1961).
3. Green M, *Prog. Photovolt.* **9**, 123–135 (2001).
4. Callen H, *Thermodynamics*, John Wiley & Sons, New York (1981).
5. Landau L, Lifchitz E, *Physique Statistique*, Chap. I §24, Mir, Moscou (1967).
6. Badescu R, *Equilibrium and Nonequilibrium Statistical Mechanics*, Wiley, New York (1975).
7. Kondepudi D, Prigogine I, *Modern Thermodynamics*, Wiley, Chichester (1999).
8. Luque A, Martí A, Cuadra L, *IEEE Trans. Elec. Dev.* **48**, N. 9, 2118–2124 (2001).
9. Landau L, Lifchitz E, *Physique Statistique*, Chap. V §52, Mir, Moscou (1967).
10. Welford W, Winston R, *The Optics of Nonimaging Concentrators*, Appendix I, Academic Press, New York, NY (1978).
11. Landau L, y Lifchitz E, *Mécanique*, Chap. VII §46 La Paix, Moscou, (year not shown: prior to 1965).
12. Welford W, Winston R, *The Optics of Non-imaging Concentrators*, Chapter 2, §2.7. Academic Press, New York (1978).
13. Campbell P, Green A, *IEEE Trans. Elec. Dev.* **33**, 234–239 (1986).

14. Nozik A, *Annu. Rev. Phys. Chem.* **52**, 193–231 (2001).
15. Martí A, Balenzategui J, Reyna R, *J. Appl. Phys.* **82**, 4067–4075 (1997).
16. Araújo G, Martí A, *Sol. Energy Mater. Sol. Cells*, **31**, 213–240 (1994).
17. Luque A, Martí A, *Phys. Rev. Lett.* **78**, 5014–5017 (1997).
18. Hulstrom R, Bird R, Riordan C, *Sol. Cells* **15**, 365–391 (1985).
19. De Vos A, *Endoreversible Thermodynamics of Solar Energy Conversion*, Chap. 2 §2.1, Oxford University, Oxford (1992).
20. Miñano J, "*Optical Confinement in Photovoltaics*", in Luque A, Araújo G, Eds, *Physical Limitations to Photovoltaic Energy Conversion*, 50–83, Adam Hilger, Bristol (1990).
21. Shockley W, *Bell Syst. Tech.* **28**, 435–489 (1949).
22. Würfel P, *Physica E* **14**, 18–26 (2002).
23. Sinton R, Kwark Y, Gan J, Swanson R, *IEEE Electron. Dev. Lett.* **EDL7**, 567–569 (1986).
24. Green M, *Prog. Photovolt.* **9**, 137–144 (2001).
25. Araújo G, "*Limits to Efficiency of Single and Multiple Bandgap Solar Cells*", in Luque A, Araújo G, Eds, *Physical Limitations to Photovoltaic Energy Conversion*, 119–133, Adam Hilger, Bristol (1990).
26. Araújo G, Martí A, *IEEE Trans. Elec. Dev.* **37**, 1402–1405 (1998).
27. Gale R, King B, Fan J, *Proc. 19th IEEE PSC*, 293–295, IEEE, New York (1987).
28. Tobin S, Vernon S, Sanfacon M, Mastrovito A, *Proc. 22th IEEE PSC*, 147–152, IEEE, New York (1991).
29. Miñano J, *J. Opt. Soc. Am. A* **3**, 1345–1353 (1986).
30. Parrot J, in Luque A, Araújo G, Eds, *Physical Limitations to Photovoltaic Energy Conversion*, Adam Hilger, Bristol (1990).
31. Martí A, Araújo G, *Sol. Energy Mater. Sol. Cells* **43**, 203–222 (1996).
32. Luque A, Martí A, *Phys. Rev. B* **55**, 6994–6999 (1997).
33. Landsberg P, Tonge G, *J. Appl. Phys.* **51**, R1–20 (1980).
34. De Vos A, Pauwels H, *Appl. Phys.* **25**, 119–125 (1981).
35. Luque A, Martí A, *Sol. Energy Mater. Sol. Cells* **58**, 147–165 (1999).
36. Karam N *et al*., *Sol. Energy Mater. Sol. Cells* **66**, 453–466 (2001).
37. Brown A, Green M, *Prog. in Photovolt.: Res. Appl.* **10**, 299–307 (2002).
38. Tobías I, Luque A, *Prog. in Photovolt.: Res. Appl.* **10**, 323–329 (2002).
39. Luque A, "*Coupling Light to Solar Cells*", in Prince M, Ed, *Advances in Solar Energy*, Vol. 8, ASES, Boulder, CO (1993).
40. Castañs M, *Revista Geofísica* **35**, 227–239 (1976).
41. Werner J, Kolodinski S, Queisser H, *Phys. Rev. Lett.* **72**, 3851–3854 (1994).
42. Werner J, Brendel R, Queisser H, *Appl. Phys. Lett.* **67**, 1028–3014 (1995).
43. De Vos A, Desoete B, *Sol. Energy Mater. Sol. Cells* **51**, 413–424 (1998).
44. Spirkl W, Ries H, *Phys. Rev. B* V. 52 N. 15, 319–325 (1995).
45. Kolodinski S, Werner J, Queisser H, *Appl. Phys. Lett.* **63**, 2405–2407 (1993).
46. Kolodinski S, Werner J, Queisser H, *Sol. Energy Mater. Sol. Cells* **33**, 275–285 (1994).
47. Wurfel P, *Sol. Energy Mater. Sol. Cells* **46**, 43–52 (1997).
48. Ross R, Nozik A, *J. Appl. Phys.* **53**, 3813–3818 (1982).
49. Nozik A, *Photovoltaics for the 21st Century, Proceedings of the International Symposium* in McConnel R, Kapur V, Eds, 61–68, The Electrochemical Society, Pennington, NJ (2001).
50. Luque A, Martí A, *Phys. Rev. Lett.* **78**, 5014–5017 (1997).
51. Luque A, Martí A, *Prog. Photovolt.: Res. Appl.* **9**, 73–86 (2001).
52. Luque A, Martí A, Cuadra L, *Proc. of the 16th European Photovoltaic Solar Energy Conference*, 59–61, James & James Ltd, London (2000).
53. Green M, *Prog. Photovolt.: Res. Appl.* **9**, 137–144 (2001).
54. Brown A, Green M, "*Limiting Efficiency for a Multi-Solar Cell Containing Three and Four Bands*" International Workshop on Photovoltaics in nanostructures, Dresden, Germany, Proc. to be published in *Physica E* (Private Communication) (2001).

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 887

55. Beaucarne G, Brown A, Keevers M, Corkish R, Green M, *Prog. in Photovolt.: Res. Appl.* **10**, 345–353 (2002).
56. Trupke T, Green M, Würfel P, *J. of Appl. Phys.* **52**, 1668–1674 (2002).
57. Trupke T, Green M, Würfel P, *J. of Appl. Phys.* **92**, 4117–4122 (2002).
58. Wahnón P, Tablero C, *Physical Review B* **65**, 165115, 1–10 (2002).
59. Martí A, Cuadra L, Luque A, *Proc. of 28th IEEE Photovoltaic Specialist Conf.*, 940–943, IEEE, New York (2000).
60. Cuadra L *et al*., *Proc. of the 17th European Photovoltaic Solar Energy Conference*, 98–101, James&James (Munich, Germany, 2001).
61. Martí A *et al*., *Photovoltaics for the 21st Century*, *Proc. de la 199th Electrochemical Society Meeting*, 46–60, The Electrochemical Society, Pennington, NJ (2001).
62. Martí A, Cuadra L, Luque A, *IEEE Trans. Elec. Dev.* **48**, 2394–2399 (2001).
63. Yamaguchi M, Luque A, *IEEE Trans. Electron. Dev.* **46**, 2139–2144 (1999).
64. Miñano J, González J, Benítez P, *Appl. Opt.* **34**, 7850–7856 (1995).

5

Solar Grade Silicon Feedstock

Bruno Ceccaroli[1] and Otto Lohne[2]

[1]*Silicon Technologies AS, Kristiansand, Norway,* [2]*Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

5.1 INTRODUCTION

The Photovoltaic (PV) industry is still in its infancy and at the moment it is very difficult to predict which technical, economical and social patterns its deployment will follow before reaching maturity. However, if photovoltaics are to become a major energy source in the future, it is appropriate to question which materials and which natural elements are critical to secure the long-term sustainability of this energy source. This is particularly valid for the semiconductor materials whose band gap has to perform the efficient conversion of sunlight to electricity. The recent history of photovoltaics (from the 1950s) reveals an intense activity of research and development embracing a broad range of disciplines and leading to a healthy multitude of innovations. Organic versus inorganic semiconductors, intrinsic versus extrinsic semiconductors, homojunctions versus heterojunctions and amorphous versus crystalline structures are a few dilemma that new research achievements steadily bring to the scientific and industrial community. It will take years, perhaps decades, before mankind is able to solve the challenge and answer the questions addressed above. (These aspects are also described in Chapters 6 to 16 of this handbook.)

Up to now, the dominant semiconductor material used in photovoltaics is silicon, particularly crystalline silicon. The most recent market surveys issued by commercial consultants, (trans)governmental organisations and agencies clearly state this fact.[1] Analysing annual growth by means of technology and material from these quoted sources, multicrystalline silicon takes the lion's share of the growth. In the first half of the 1990s,

[1] For PV News edited by Maycock (2001), commercial shipments to terrestrial and indoor applications of cells or modules based on non-silicon technologies accounted in 2000 for just 1.2 MW of a total 288 MW, that is, 4.2 per thousand. Amorphous silicon was credited in the same period 27 MW shipment or 9.4% and crystalline

Handbook of Photovoltaic Science and Engineering. Edited by A. Luque and S. Hegedus
© 2003 John Wiley & Sons, Ltd ISBN: 0-471-49196-9

multicrystalline sawn wafers accounted for just half of the single-crystal shipments. In 1998 both technologies were equivalent in size. Two years later, multicrystalline technology surpassed single-crystal technology by 55%.

The consensus is that crystalline silicon is and will remain for at least a long decade the workhorse of this growing market. Long-term visionary forecasts predict that by 2050, 30 000 TWh PV electricity will be generated annually worldwide. This will require an installed PV output capacity totalling approximately 15 million metric tons (MT) of solar grade silicon feedstock, assuming that silicon remains dominant and that cell efficiency and material yields have steadily improved. To build up such a capacity over fifty years will represent an annual production of 300 000 MT solar grade silicon feedstock. The annual present consumption of pure silicon for photovoltaics is approximately a hundredth part of that (i.e. 4000 MT), whereas production of metallurgical grade silicon for all purposes in 2000 was approximately one million metric tons. This clearly poses the question of material feasibility and availability. The present chapter is therefore dedicated to silicon, its extraction, purification and availability by current and future practice.

5.2 SILICON

Silicon (Si) is the second member in the Group IVA in the periodic system of elements. It never occurs free in nature, but in combination with oxygen forming oxides and silicates. Most of the Earth's crust is made up of silica and miscellaneous silicates associated with aluminium, magnesium and other elements. Silicon constitutes about 26% of the Earth's crust and is the second most abundant element in weight, oxygen being the largest.

5.2.1 Physical Properties of Silicon Relevant to Photovoltaics

Silicon is a semiconductor with a band gap E_g of 1.12 eV at 25°C. At atmospheric pressure, silicon crystallises into a diamond cubic structure, which converts into a body-centred lattice when subjected to ca 15 GPa. Under some circumstances, slow-growing

silicon the balance, that is, close to 90% of which 49% (141 MW) are cells made of multicrystalline sawn wafers, 31% (90 MW) of single-crystal wafers and 10% (28 MW) are cells based on various crystalline silicon technologies, for example, ribbon (15 MW or 5.1%), amorphous silicon on single-crystal silicon slices (12 MW or 4.2%) and others (approx. 1 MW or 3.5 per thousand).

Solar Modules Shipment by Technology (Source: PV News 2001)

Technologies	Shipped output in 2000
Amorphous silicon	27 MW (9%)
Single crystal silicon wafers	90 MW (31%)
Multicrystalline sawn silicon wafers	141 MW (49%)
Various crystalline silicon technologies (i.e. ribbon, films, amorphous on single crystal)	28 MW (10%)

Other sources available in May–November 2001 at the time of writing this chapter, such as the German magazine Photon International March 2001 published congruent data for the same period. Statistics from non-commercial entities as those edited by the International Energy Agency (IEA-PVPS programme) covering 21 industrialised countries, members of the OECD-organisation, show over years the same trends [1–4].

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 890

faces of silicon are (111) but in epitaxial films and polysilicon deposition (111) is the fastest growth direction. Vapour deposition below 500°C results in amorphous silicon. If reheated above this temperature, crystallisation will occur.

Unlike most of the compounds and elements, silicon contracts when melting or expands when solidifying.

Impurities incorporated in the silicon lattice during the crystal growth or during the post-treatment (diffusion, implantation etc.) ionise at low temperatures, thus providing either free electrons or holes. Impurities from the Group IIIA replace a Si atom in the atomic lattice to supply electrons and are called *n*-dopants or donors, whereas elements from the Group VA substitute for a Si atom to supply holes and are called *p*-dopants or acceptors (see Section 5.6.3). Phosphorus and boron represent these groups and are used in PV processing to control the semiconductor properties (doping levels) of silicon. Impurity concentrations are expressed in atoms of impurity per cubic centimeter of the host material (silicon). In silicon semiconductor devices, these vary from 10^{14} to 10^{20} atoms per cm^3 and can be directly measured by analytical instruments. An indirect measure of impurity concentration is the minority-carrier lifetime. This is the time that elapses before a free electron in the lattice recombines with a hole. The transition metals, Fe, Cr, Ni, degrade the minority-carrier lifetime and the solar cell performance. High-purity silicon crystals with metal content less than 10 ppb(w) have minority-carrier lifetime values as high as 10 000 μs. Semiconductor wafers with phosphorus and boron dopants have values from 50 to 300 μs. Solar cells require minority-carrier lifetime value of at least 25 μs.

The relatively high refractive index limits the optical applications of silicon. The absorption/transmission properties in the 0.4 to 1.5 μm wavelength spectra are important in the performance of PV cells and photoconductive devices. In PV applications antireflective layers applied to silicon are commonly used.

Silicon even when alloyed with small quantities of impurities is brittle. Shaping silicon for PV applications requires sawing and grinding. Microelectronic applications require polishing. These mechanical operations are very similar to those applied to glasses.

Various thermal and mechanical properties are reported in Table 5.1.

For more details the reader is invited to consult the References [5–8].

Table 5.1 Thermal and mechanical properties of silicon

Property	Value
Atomic weight	28.085
Atomic density (atoms/cm^3)	5.0×10^{22}
Melting point (°C)	1410
Boiling point (°C)	2355
Density (g/cm^3 at 25°C)	2.329
Heat of fusion (kJ/g)	1.8
Heat of vaporisation at MP (kJ/g)	16
Volume of contraction on melting (%)	9.5

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 891

5.2.2 Chemical Properties Relevant to Photovoltaics

Silicon is stable in the tetravalent state and has a strong affinity for oxygen, forming stable oxides and silicates, the only natural occurrences known for silicon. Artificially isolated elemental silicon immediately oxidises, forming a thin protective film of silica of less than 100 Å, which prevents further oxidation. Oxygen plays an important role in silicon-semiconductor devices, for instance, in manufacturing metal oxide semiconductor (MOS) transistors.

Silicon and carbon (Group IVA) form a strong Si−C bond and stable products. Silicon carbide is artificially synthesised in several allomorphic structures, finding various applications in photovoltaics and electronics. Primary uses are the abrasive properties of SiC for wafering silicon crystals and the emerging applications of SiC semiconductors. The strong Si−C bond is also the origin of the rich organosilicon chemistry encompassing numerous polysiloxanes (commonly named *silicones*) and organosilanes in which organic radicals are attached to silicon atoms through a covalent Si−C bond.

The tetravalence and the similarity of silicon and carbon are illustrated in the ability of silicon to form bonds with itself, Si−Si, and to form polymers, for example, $-(SiH_2)_p-$, $-(SiF_2)_p-$, comparable to hydrocarbons and fluorocarbons, although the length of the chains remains modest in the case of the silanes.

Silicon forms hydrides; Monosilane (SiH_4) is a key chemical compound for the production of amorphous silicon and the purification of silicon to semiconductor grade (see later in this chapter).

The chemical reactivity of silicon with chlorine is also extremely important. Alkyl- and arylchlorosilanes are the necessary intermediates to build the polysiloxane chains (silicones). Trichlorosilane and tetrachlorosilane, because they are volatile at low temperature and can be decomposed to elemental silicon at high temperature, are both the intermediates and the by-products of the purification processes upgrading metallurgical grade silicon to semiconductor purity (see later in this chapter). Other chlorosilanes or halogenosilanes are also used in chemical vapour deposition applications. The halogen atom is easily substituted by a hydroxyl group, −OH, through hydrolysis. Such a hydroxyl group tends to react with other functional groups by exchanging the hydrogen atom. This is the basis of a rich surface chemistry.

Silicon and germanium (Group IVA) are isomorphous and mutually soluble in all proportions.

Tin and lead, also elements of Group IVA, do not react with silicon and are not miscible in silicon, which is mentioned as a remarkable curiosity.

For more details on the chemical properties the reader is invited to consult the References [8−11].

5.2.3 Health Factors

The surface of elemental silicon is oxidised and is relatively inert and is considered as non-toxic. Hazard risks with elemental silicon are high when silicon occurs as a fine

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 892

powder in the presence of an ignition source. Damaging and fatal explosions have been reported by the silicon industry. The raw material, from which silicon is made, quartz or quartzite, is one of the sources of silicosis. Most of the hazards are related when quartz/quartzite is quarried, exposure taking place during drilling, crushing, loading and bulk handling. Several protection methods must be applied to the quarries and to the metallurgical plants to prevent silicon dust explosion and silicosis. Volatile silanes such as monosilane and chlorosilanes are extremely reactive in the presence of oxygen, water or moisture. They are also classified as hazardous chemical substances whose handling requires special care. Saturated long chain silanes, polysiloxanes as well as amorphous silica are known to be chemically inert and not toxic. Because of that they are widely used in pharmacy, food industry and cosmetics.

The production of metallurgical silicon and electronic grade silicon has an environmental impact through energy consumption, associated with climatic and polluting gases principally CO_2, NO_x and SO_2. However, it must be noticed that the corresponding nuisances and energy consumption involved in manufacturing and installing PV systems are "paid back" by the same system in the form of emission-free "green" electricity only after about four to five years of an average existence of more than 25 years. More quantified examples about environmental and energy "payback" are reported from Europe, Japan and Australia and commented in [12].

5.2.4 History and Applications of Silicon

Since antiquity, silicon has been of great importance to humanity. However, the first applications were based on naturally occurring forms of silicon, for instance, flint (silex-silicis in Latin), a variety of quartz used from the Stone Age to the Neolithic Era to make tools, weapons and later potteries. Glass made of silicate dated back to 12 000 B.C. Elemental silicon was prepared for the first time in 1824 by Berzelius, passing silicon tetrachloride over heated potassium. Silicon tetrachloride could be prepared by chlorinating silicate/silica. The first crystalline silicon was made accidentally in 1854 by Sainte-Claire Deville working on aluminium electrolysis. The first preparation of silicon/silicon rich alloys in an electric arc furnace was performed by Moisan in 1895 and the industrial production was by Bozel and Rathenau independently from 1897 to 1898. Acheson also discovered accidentally in this period silicon carbide while trying to make artificial diamond. Silicon alloys, particularly ferroalloys, have from the end of the nineteenth (XIX) century played an important role in the production of steel. Silicon metal (silicon content higher than 96% according to definitions outlined by trade organisations) was not current until the Second World War. Three major applications have since greatly stimulated the production and purification of silicon, that is, aluminium, silicones and solid-state electronics. Silicon carbide has also found a broad range of applications taking advantage of its hardness and chemical noble character. More recently SiC has found applications in electronics because of its excellent semiconductor properties and tends to become a strategic material for cutting silicon and boules into thin wafers.

At the beginning of the new 2000 millennium, approximately one million metric tons of metallurgical grade silicon, also called *silicon metal* in the industry because of its appearance and not because of its physical properties, are produced and sold in the world

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 893

Table 5.2 Chemical characteristics of commercial metallurgical grade silicon

Element	O	Fe	Al	Ca	C	Mg	Ti	Mn	V	B	P
Low (ppm)	100	300	300	20	50	5	100	10	1	5	5
High (ppm)	5000	25 000	5000	2000	1500	200	1000	300	300	70	100
Element	Cu	Cr	Ni	Zr	Mo						
Low (ppm)	5	5	10	5	1						
High (ppm)	100	150	100	300	10						

market. This is a relatively small amount compared to the multimillion metric ton markets of crude iron, steel, aluminium or ferroalloys. Industrial location to produce silicon metal has been guided by the vicinity of rich and pure quartz deposits and/or the availability of abundant electrical power. Leading producing countries are PR China, the United States, Brazil, Norway and France. In spite of some recent mergers, the industry remains fragmented. One may find three dozens of companies producing and marketing silicon metal, most of the plants having an annual output of 20 000 to 60 000 MT (see production of silicon metal later in this chapter). The chemical characteristics of commercial metallurgical grade silicon are indicated in Table 5.2.

The silicon metal market is traditionally divided into two main subgroups, that is, the aluminium and the chemical segments each consuming approximately half of the worldwide output. There are, however, some differences in the characteristics requested by each.

5.2.4.1 Applications in aluminium

In the aluminium industry, silicon is added to molten aluminium in which it is dissolved. A simple eutectic composition occurs at 12.6% silicon in aluminium. This has important consequences for industrial applications in the aluminium industry. Silicon is used in order to improve the viscosity, the fluidity of liquid aluminium and the mechanical properties of commercial alloys. The iron, calcium and phosphorus content in silicon are particularly critical for such applications.

There are two important groups of aluminium alloys in which silicon is one of the main alloying elements.

5.2.4.1.1 Casting alloys

By adding silicon to the melt, the fluidity is improved. Aluminium alloys near the eutectic composition are therefore used in thin-walled castings. Typical concentrations are 7 to 12%. If a few tenths of a percent of magnesium is added, the alloys may be age-hardened and thereby nearly double their yield strength.

To counteract the formation of large needle-shaped particles, the alloys are normally modified with sodium, strontium or phosphorus.

The alloys present good corrosion properties.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 894

5.2.4.1.2 Wrought alloys

AlMgSi alloys (6xxx series) are widely used as medium-strength structural alloys. Typical silicon content is 0.5 to 1.0%.

The alloys of the 6xxx series have good hot-working properties and are age-hardening. These alloys are therefore well suited for extrusion of profiles, which by heating at 150 to 200°C are given their final strength.

The alloys present good corrosion and weldable properties. Typical markets are building and transport industries.

5.2.4.2 Applications in chemistry

5.2.4.2.1 Silicones

Since the discovery of the direct synthesis (reaction 5.1) of dimethyldichlorosilane, during the Second World War independently by Rochow and Müller, the silicones industry has developed to become a strong and growing chemical business consuming (year 2000) about 400 000 MT of silicon [8–11].

$$Si(s) + 2CH_3Cl(g) \rightarrow (CH_3)_2SiCl_2 \text{ (Cu catalyst, 260 to 370°C)} \quad (5.1)$$

$$(CH_3)_2SiCl_2 + 2H_2O \rightarrow (CH_3)_2Si(OH)_2 + 2HCl \quad (5.2)$$

$$n(CH_3)Si(OH)_2 \rightarrow [(CH_3)_2SiO]_n + nH_2O \quad (5.3)$$

The direct synthesis (5.1) is industrially performed in a fluidised bed reactor requesting small particles or powder of silicon (20–300 μm). The reaction is exothermic and needs to be activated with copper catalysts as well as promoters Zn, Sn, P and others. Whereas Fe does not seem to play an important role, Ca and Al have shown to take an active part in the overall reaction.

5.2.4.2.2 Synthetic silica

Varieties of synthetic silica such as pyrogenic silica (also called *fumed silica*) or silica ingots as feedstock to optical fibres are industrially prepared by burning silicon tetrachloride:

$$SiCl_4(g) + 2H_2(g) + O_2(g) \rightarrow SiO_2(s) + 4HCl(g) \quad (5.4)$$

Silicon tetrachloride may be prepared by chlorination of natural silica. However, industrially, silicon tetrachloride is produced by reacting chlorine with metallurgical grade silicon in a direct synthesis performed either in a fluidised bed or a fixed bed reactor:

$$Si(s) + 2Cl_2(g) \rightarrow SiCl_4(g) \quad (5.5)$$

$$Si(s) + 4HCl(g) \rightarrow SiCl_4(g) + 2H_2(g) \quad (5.6)$$

The present fumed silica market is about 100 000 to 120 000 MT, which in turn consumes around 50 000 to 60 000 MT of metallurgical silicon. It is noted that the main market for

fumed silica is as additives in silicone rubbers used to increase the mechanical strength and the elasticity of these elastomers.

5.2.4.2.3 Functional silanes

This generic term covers a broad range of products built on silane molecules, in which an atom of hydrogen or of chlorine is substituted with an organic radical bearing a functional group, for example, amine, acid, ester, alcohol and so on.

$$X_{4-(n+p+q)}SiR_nR'_pR''_q$$

represents a general formula of functional silanes, in which R, R′, R″ are organic radicals and X is Cl or OH. There exist a multitude of functional silanes. One of their major applications is as coupling agent between inorganic and organic compounds, for example, inorganic fillers (glass, silica, clays etc.) in organic matrices (epoxy, polyester etc.)

Several routes may be imagined to generate such molecules. The current industrial praxis dictated by economical efficiency is the functionalisation of trichlorosilane according to

$$SiHCl_3(g) + RX(l) \rightarrow SiRCl_3(l) + HX(g)\ (X = Cl, OH, \ldots) \tag{5.7}$$

whereas trichlorosilane is produced by a direct synthesis in a fluidised or fixed bed reactor:

$$Si(s) + 3HCl(g) \rightarrow SiHCl_3(g) + H_2(g) \tag{5.8}$$

$$Si(s) + 3SiCl_4(g) + 6H_2 \rightarrow 4SiHCl_3(g) + 4H_2 \tag{5.9}$$

$$SiCl_4(g) + H_2(g) \rightarrow SiHCl_3(g) + HCl(g) \tag{5.10}$$

A variant recently developed for some classes of functional silanes is the direct synthesis involving silicon and methanol:

$$Si(s) + 3CH_3OH(l) \rightarrow SiH(OCH_3)_3 + H_2(g)\ (Cu\ catalyst) \tag{5.11}$$

$$SiH(OCH_3)_3 + RX \rightarrow SiR(OCH_3)_3 + HX \tag{5.12}$$

Orthoethylsilicate or tetraethoxysilane, $Si(OC_2H_5)_4$, is an important chemical molecule for glass, ceramic, foundry and painting. It is produced industrially by substitution of Cl in $SiCl_4$ or by direct reaction of ethanol with silicon in the presence of a catalyst following the same pattern as (5.7) and (5.12).

The total consumption of silicon metal to functional silanes and orthosilicates may be estimated at 10 000 to 20 000 MT per year.

5.2.4.3 Semiconductor silicon

Silicon is by far the most important and popular semiconductor material since the emergence of solid-state electronics in the late fifties and the early sixties. Ultra-pure

silicon (commercially called *polysilicon*) with the adequate semiconductor properties is industrially prepared through the distillation and the thermal decomposition of volatile silicon compounds, for example, trichlorosilane, $SiHCl_3$, and monosilane, SiH_4. These operations are performed in large chemical plants, which for synergy reasons are sometimes incorporated in plants producing other silicon-based compounds as those described above. Although the ultimate application in the case of polysilicon is in the semiconductor industry, this particular process is, from a silicon raw material perspective, counted among the chemical applications of silicon.

The current production of polysilicon in 2000 was approximately 20 000 MT, whereas the installed worldwide capacity was estimated around 25 000 MT. Compared to the other applications of silicon, the use of semiconductor silicon in terms of volume is very modest. However, it is a high-value product. For example, the silicon value is multiplied by a factor of 30 to 50 through upgrading metallurgical grade silicon to polysilicon. This is also the fastest-growing application of silicon with an annual growth rate of approximately 10%, whereas silicones grows at a 5 to 7% rate, aluminium at 2 to 3% and silicon metal overall at 4%.

The present raw material for all silicon solar cells originates from this route. Therefore, a more detailed description of these processes will be given later in the present chapter.

5.2.4.4 Other applications

There are a few other applications of silicon in various fields such as explosives (silicon powder), refractories and advanced ceramics (silicon nitride and carbide). These applications presently do not account for more than 1% of the worldwide silicon metal output.

Because of their anticipated excellent mechanical and chemical resistant properties, alloys rich in silicon have a bright future. They may be prepared by powder metallurgy, mixing and sintering silicon powder with metallic powders (e.g. Cu, Al, Ti, Co, V etc.) [7].

The present chapter does not review the applications of silicon, such as glasses, ferroalloys and silicon carbide, in which silicon is usually added to the production process as natural silicate, quartz, quartzite or other silicon alloys.

5.3 PRODUCTION OF METALLURGICAL GRADE SILICON

5.3.1 The Carbothermic Reduction of Silica

Metallurgical grade silicon, also called *silicon metal*, with a typical purity of 98.5% Si is produced in submerged electric arc furnaces. In principle, this process is much the same as it was at the beginning of the twentieth (XX) century when it was first developed for ferrosilicon and other alloys. However, practical execution has greatly improved with larger furnaces, more efficient material handling and improved control of the operations. This has led to a continuous decrease of the specific energy consumption concomitant to higher degrees of raw material utilisation.

The furnace consists essentially of a crucible filled with quartz and carbon materials. Silicon is freed by the carbothermic reduction of silica according to the overall reaction:

$$SiO_2(s) + 2C(s) = Si(l) + 2CO(g) \tag{5.13}$$

Contrary to what is often claimed in popular articles or reviews, silica sand is currently not used for this purpose. Lumpy quartz (e.g. 10–100 mm) with appropriate purity and thermal resistance is preferred. Carbon raw material generally consists of metallurgical grade coal as well as woodchips and/or charcoal and coke. The metallurgical coal is co-produced with coal used for crude steel production. As a rule this coal needs to be washed in order to remove most of the ash containing unsuitable impurities. Raw materials, both quartz and carbon, are selected in order to achieve high product quality (silicon and silica fumes), to maximise furnace performances and to minimize the environmental damages (i.e. SO_2 and NO_X emissions). The raw material reactivity and the consistency of the mix of raw materials in the charge, for instance its porosity, are extremely important factors in achieving good furnace performance in terms of high material yield, lower power consumption and good product quality.

The raw material mix or charge is heated by means of an intense electric arc sustained between the tip of three submerged electrodes and the electrical ground of the furnace. Although important exceptions exist, the current practice is to run this process in a three-phase current, open and rotating furnace at a working electrical load normally between 10 and 30 MW, depending on the size of the furnace. The tendency is to increase the furnace size and the electrical load in order to achieve higher output and productivity.

Electrodes are also made of carbon. Originally, expensive graphite electrodes were used. They were displaced by pre-baked electrodes, which in turn tend to be replaced by more cost-efficient self-baking electrodes. The electrode technology is an important aspect to the present development of this industry: half a dozen electrode types ranging from pre-baked to self-baking electrodes of Søderberg type are currently used or are in the process of development.

Liquid silicon metal is tapped from the bottom of the furnace, and the thoroughly mixed raw materials are charged on the top. The reaction co-product, carbon monoxide CO(g), is further oxidised to carbon dioxide CO_2(g) in open furnaces and released into the atmosphere. In open furnaces, side-reactions leading to the formation of silica fumes play an important role for the overall economics of the process:

$$Si(l) + \tfrac{1}{2}O_2(g) = SiO(g) \tag{5.14}$$

$$SiO(g) + \tfrac{1}{2}O_2 = SiO_2(s) \tag{5.15}$$

The silica fumes, which consist mainly of very fine particles of amorphous silica less than 1 μm, are passed through filter cloths installed in large bag-house systems adjacent to the furnaces. The collected amorphous finely divided silica finds valuable applications as additives in concrete and refractory. Depending on the quality of the raw materials used and the operational strategy and skills, the silicon yield as metallurgical silicon ranges from 80 to 90%, the balance resulting in silica fume.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 898

Reactions (5.13) to (5.15) are a simplification of the complex system. Several main principles can be understood from a more detailed description of the chemistry. There are two important intermediate compounds: the gaseous silicon monoxide SiO(g) as already mentioned in reaction (5.14) and the solid silicon carbide SiC(s). To interpret the chemistry occurring in the furnace, it is convenient to conceptually split the furnace reaction inner space into an inner hot zone and an outer cooler zone. Liquid silicon is produced in the inner zone, where the dominant chemistry is described by the reactions

$$2SiO_2(l) + SiC(s) = 3SiO(g) + CO(g) \quad (5.16)$$

$$SiO(g) + SiC(s) = 2Si(l) + CO(g) \quad (5.17)$$

The temperature in the inner zone is in the range of 1900 to 2100°C, allowing a high proportion of SiO(g) in this zone, which is absolutely indispensable for further reduction according to reaction (5.17).

In the outer zone, where temperature is below 1900°C, SiO(g) and CO(g) convected away from the inner zone meet and react with free carbon. Consequently, silicon carbide SiC(s) and condensation products of Si(l) in a matrix of SiO_2(s,l) are formed as the partial pressure of SiO(g) drops:

$$SiO(g) + 2C(s) = SiC(s) + CO(g) \quad (5.18)$$

$$2SiO(g) = Si(l) + SiO_2(s) \quad (5.19)$$

A schematic description of the furnace is given in Figure 5.1.

The high-temperature nature of this process implies operation as continuous as possible. Raw materials are therefore fed in small batches with frequent intervals and are judiciously distributed on the top of the charge. Liquid silicon is continuously, or at frequent intervals, drained out from the bottom of the furnace, whereas gas exhaust and fumes are constantly passing through the filter to clean the fumes and recapture the silica.

Liquid crude silicon contains 1 to 3% impurities depending on the raw materials and the type of electrodes. The main impurities are

Fe: 0.2–1%
Al: 0.4–0.7%
Ca: 0.2–0.6%
Ti: 0.1–0.02%
C: 0.1–0.15%

5.3.2 Refining

Most of the applications of silicon as described above request further refining. The crude silicon is therefore tapped as liquid in large ladles (containing up to 10 MT of silicon) and treated when still liquid with oxidative gas and slag-forming additives, mainly silica sand (SiO_2) and lime/limestone ($CaO/CaCO_3$). Other chemicals such as dolomite (CaO–MgO), calcium fluoride (CaF_2) and others are used depending on plant practice and customer requirements. Elements less noble than silicon such as Al, Ca and Mg are oxidised and

the degree of refining is determined by distribution equilibria (5.20) to (5.23), where the (parentheses) refer to components dissolved in a slag phase and the underscored symbols refer to dissolved elements in liquid silicon:

$$4\underline{Al} + 3(SiO_2) = 3Si(l) + 2(Al_2O_3) \quad (5.20)$$

$$2\underline{Ca} + SiO_2 = Si(l) + 2(CaO) \quad (5.21)$$

$$2\underline{Mg} + SiO_2 = Si(l) + 2(MgO) \quad (5.22)$$

$$S(l) + O_2 = (SiO_2) \quad (5.23)$$

Theoretically it is possible to remove Al and Ca to very low levels, but in practice this is prevented by the large heat losses occurring during this operation. Temperature drops to 1700 to 1500°C, and to avoid freezing of the melt, some of the silica needed for slag

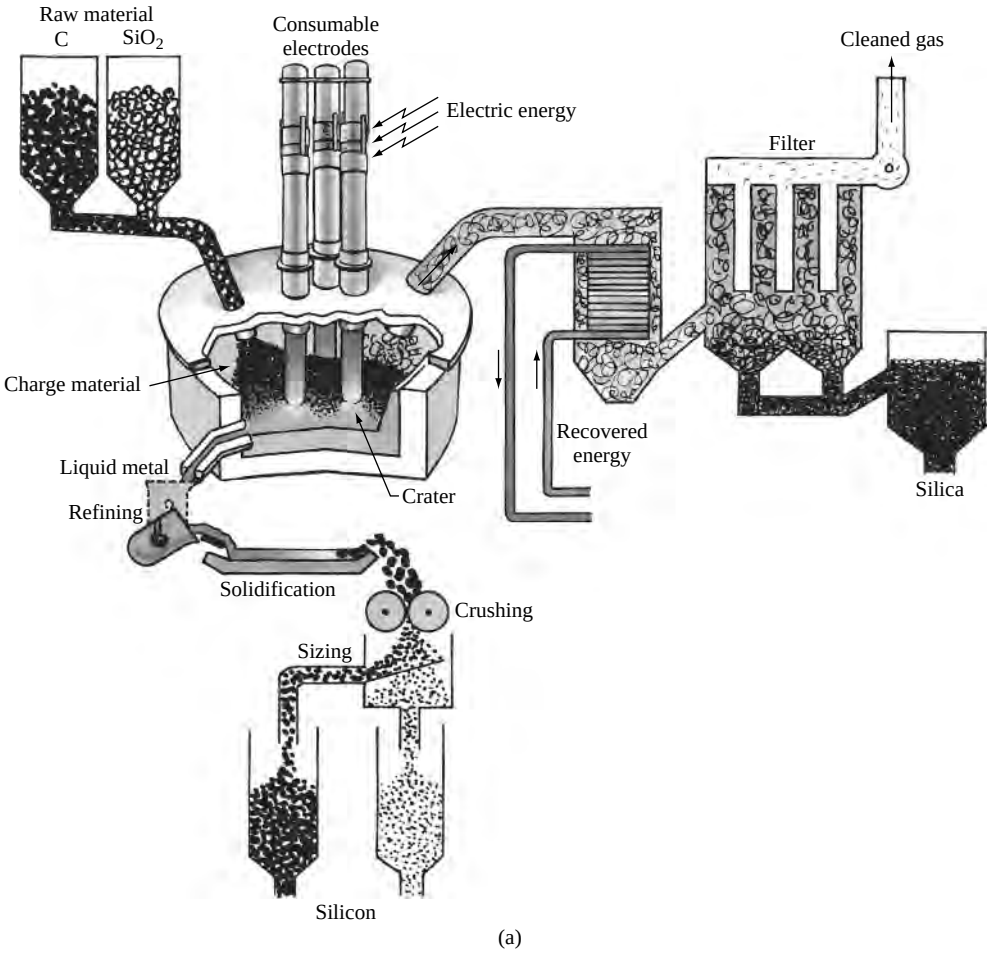
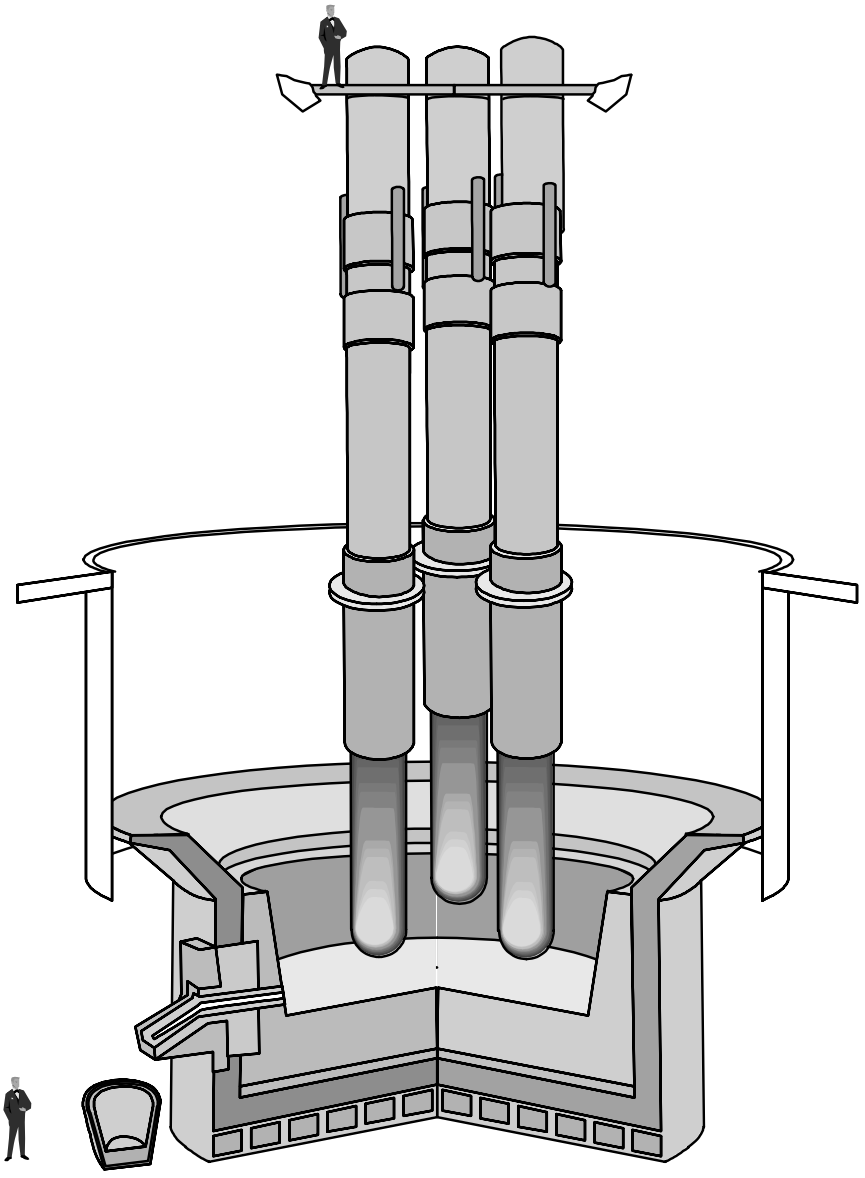


Figure 5.1 (a) and (b) Schematic representation of a furnace for production of metallurgical grade silicon. Reproduced from Schei A, Tuset J, Tveit H, *Production of High Silicon Alloys*, Tapir forlag, Trondheim (1998) with permission by Haluard Tueit

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 900



(b)

Figure 5.1 (*continued*)

formation is provided by direct oxidation of Si(l) with added oxygen (oxygen is added to heat silicon to keep it liquid). A disadvantage of this operation is also a partial oxidation of silicon resulting in expensive material losses.

After completion of oxidative refining in the ladle, the slag, which contains part of the impurities, is removed mechanically or by gravity and liquid silicon is poured into a casting mold. The slag-forming additives influence the slag density and viscosity, hence

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 901

the practical separation of slag and the ultimate purity of the poured silicon. For instance, a high CaO content will lead to a low viscosity slag, which will sink to the bottom of the ladle, while CaF will increase the viscosity. Sufficiently different properties in density and viscosity of both the slag and the molten silicon are required to achieve a good separation. Many studies and practical on-site developments have been devoted to this step of the process [13].

Carbon is present in crude liquid silicon mainly as dissolved C and suspended SiC particles. The fraction of SiC increases as the temperature is lowered; SiC particles are then efficiently captured by the slag phase and thus are removed from liquid silicon during the ladle treatment and the subsequent pouring. SiC is removed simply by mechanical separation, precipitated particles sticking to the walls of the ladle and the other devices containing the liquid silicon [14]. Dissolved carbon in the range of 80 to 100 ppm(w) in best cases will finally remain in the purified alloy of metallurgical silicon.

The use of these refining principles to prepare solar grade silicon will be further discussed later in the present chapter.

5.3.3 Casting and Crushing

The refined melt is poured from the ladle into a cast iron mold or onto a bed of silicon fines. The casting should preferably be removed from the mold while bleeding, that is, not fully solidified. After solidification in standard industrial conditions, metallurgical grade silicon is multicrystalline. The individual Si grains vary in size typically from 1 mm close to the iron mold wall to up to more than 100 mm in the centre section if cast on a bed of silicon fines [10]. The impurities are generally located at the Si grain boundaries as silicides and intermetallic compounds, but may also be incorporated in the Si grains if solidification has been sufficiently rapid [15]. Oxides and carbides are found as inclusions located at the grain boundaries and to a lesser degree inside the Si grains.

To be used in customers' processes, solidified silicon needs to be further crushed down to small lumps up to 100 mm. This is performed in jaw crushers and roll crushers, since at room temperature metallurgical grade silicon is hard and brittle. This operation generally generates significant amount of fines, which are undesirable because they may be contaminated by impurities and are difficult to handle during further transport and handling. Therefore, fines are removed after the primary crushing. The dominant fracture mode was found by Forwald *et al*. [7, 16] to be transgranular. For chemical applications, silicon lumps need to be further reduced to small powder particles of a few tens to a few hundreds of micrometers. This is carried out in industrial equipment such as ball mills.

Alternative methods based on rapid cooling have recently been developed to increase the homogeneity of the solidified structure of silicon through an even distribution of the impurities and intermetallic phases. Granulation in water, resulting in small granules of a few millimeters and thus avoiding casting and subsequent coarse crushing, has become a standard practice for several producers [17–19]. In an earlier attempt to avoid casting, crushing and milling, gas atomisation was tested by producers and users, but was not further industrialised for economical reasons [20, 21].

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 902

5.3.4 Economics

The carbothermic reduction of quartz in the submerged arc furnace consumes large amounts of energy and material. Best industrial performances are 10 to 11 MWh per metric ton (MT) of silicon metal and 90% silicon yield. Both availability and price of electrical power and raw material such as quartz and coal are therefore extremely sensitive to the silicon metal economics. Boardwine *et al*. [22] have presented the average structure for direct cost for Western producers as follows:

Reduction materials (as coal)	20%
Quartz	9%
Electrodes	12%
Electric power	21%
Supplies and equipment	16%
Labour	17%
Transport (to customers)	5%

In spite of dramatic changes and improvement for the individual producers, particularly with the development of new electrode technologies, the above figures are probably still valid for a large group of plants.

Most of the silicon plants are 20 years or more old. New capacity has been added during the past 10 years mainly by converting ferrosilicon furnaces to silicon. High capital expenditure is a barrier to new expansion. Investment of one million US dollars per one thousand MT of silicon is proposed as an indicating information for new furnaces.

A general and detailed source of information on production of metallurgical grade silicon is found in the recent book *Production of High Silicon Alloys* by Schei *et al*. [23].

5.4 PRODUCTION OF SEMICONDUCTOR GRADE SILICON (POLYSILICON)

Impurities in the ppb(a)–ppt(a) range are required for polysilicon supplied to the semiconductor industry. The ultra-high purity is needed to ensure exacting semiconductor properties in the grown silicon crystals. This is achieved first by the preparation of a volatile silicon hydride and its purification generally using fractional distillation. This is followed by the decomposition of this hydride to hyperpure elemental silicon by reductive pyrolysis or chemical vapour deposition. The preparation of the volatile Si compound involves external reactants and its decomposition generates by-products, which need to be recycled. The various polysilicon routes therefore must control four successive steps. All have a strong impact on the overall feasibility and economics of the suitable polysilicon products:

1. preparation/synthesis of the volatile silicon hydride
2. purification

3. decomposition to elemental silicon
4. recycling of by-products.

Many processes to produce polysilicon have been tested, patented and a few operated for many years. Only three large commercial processes are currently active:

1. The most popular process is based on the thermal decomposition of trichlorosilane at 1100°C on a heated silicon rod placed inside a deposition chamber. This process, which was developed in the late fifties, is commonly referred to as the Siemens process with reference to the company that carried out its early development.

$$2SiHCl_3 = SiCl_4 + Si + 2HCl \tag{5.24}$$

In 2001 this process still accounted for at least 60% of the worldwide production of polysilicon.

2. In a more recent process developed by Union Carbide Chemicals in the United States of America, the trichlorosilane has been replaced by monosilane SiH_4, but the principle of decomposition on a heated silicon rod inside a closed deposition chamber is maintained.

$$SiH_4 = Si + 2H_2 \tag{5.25}$$

This process, presently run by the company *Advanced Silicon Materials, LLC.* has gained during the past 15 years a significant market acceptance.

3. Finally, in the third process, also making use of monosilane SiH_4, the heated silicon rod in the closed reaction chamber has been replaced by a fluidised bed of heated silicon particles. The particles act as seeds on which SiH_4 is continuously decomposed to larger granules of hyper-pure silicon. Unlike (1) and (2) this process is a continuous one. This process is known as the Ethyl Corporation process, after the name of the US chemical company that developed it. This process is presently run by the US corporation MEMC in Pasadena, Texas.

The respective features, advantages and disadvantages of these different routes are described in the following sections.

5.4.1 The Siemens Process

A schematic overview of the process is given in Figure 5.2.

Trichlorosilane $HSiCl_3$ is prepared by hydrochlorination of metallurgical grade silicon in a fluidised bed reactor:

$$Si(s) + 3HCl = HSiCl_3 + H_2 \tag{5.26}$$

This reaction occurs at 350°C normally without a catalyst. A competing reaction is

$$Si(s) + 4HCl = SiCl_4 + 2H_2 \tag{5.27}$$

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 904

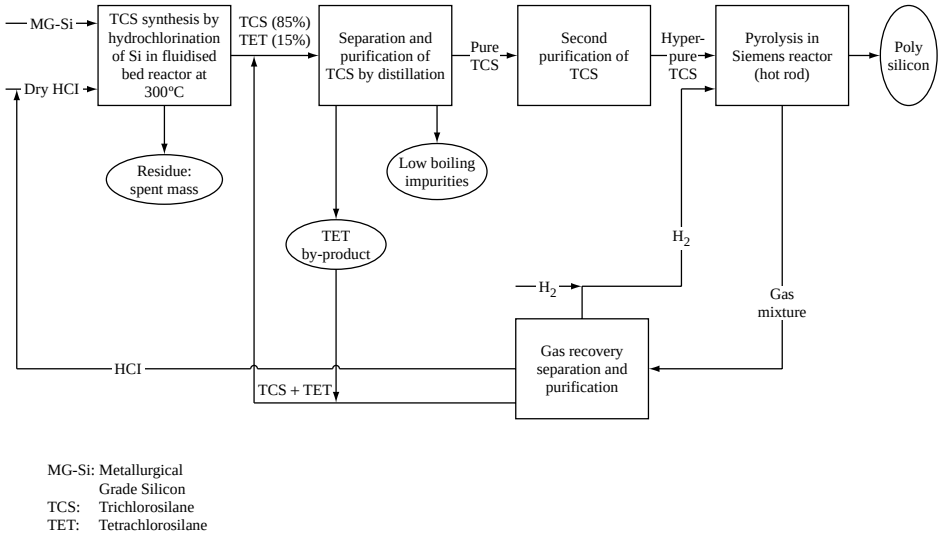


Figure 5.2 Schematic representation of the Siemens process

contributing to the formation of unsuitable tetrachlorosilane in molar proportion of 10 to 20%.

Trichlorosilane is chosen because of its high deposition rate, its low boiling point (31.8°C) and its comparatively high volatility and hence the ease of purification with respect to boron and phosphorus down to the ppb level. The boiling point of other silanes frequently found with trichlorosilane are as follows: SiH_4 (−112°C), SiH_2Cl_2 (8.6°C) and $SiCl_4$ (57.6°C). The suitable trichlorosilane undergoes a double purification through fractional distillation, the first step removing the heaviest components resulting from the direct synthesis and the second step eliminating the components lighter than trichlorosilane, also called *volatiles*.

High-purity $SiHCl_3$ is then vaporised, diluted with high-purity hydrogen and introduced into the deposition reactors. The gas is decomposed onto the surface of heated silicon seed rods, electrically heated to about 1100°C, growing large rods of hyperpure silicon.

The main reactions are:

$$2SiHCl_3 = SiH_2Cl_2 + SiCl_4 \tag{5.28}$$

$$SiH_2Cl_2 = Si + 2HCl \tag{5.29}$$

$$H_2 + HSiCl_3 = Si + 3HCl \tag{5.30}$$

$$HCl + HSiCl_3 = SiCl_4 + H_2 \tag{5.31}$$

The stream of reaction by-products, which leaves the reactor, contains H_2, HCl, $HSiCl_3$, $SiCl_4$ and H_2SiCl_2.

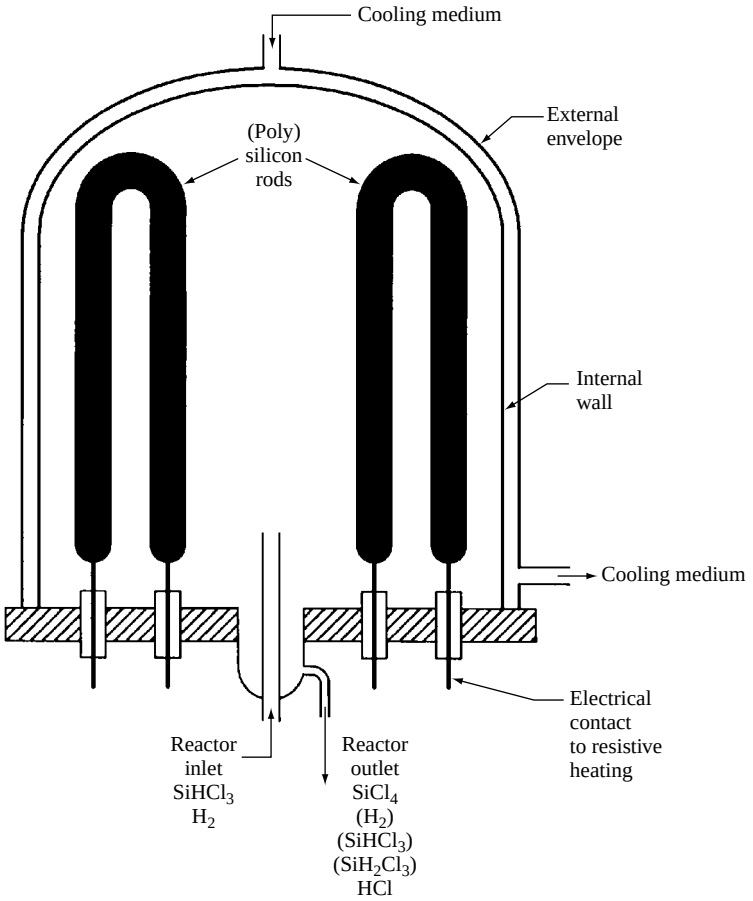


Figure 5.3 Schematic representation of the traditional Siemens reactor

A schematic representation of the Siemens reactor is given in Figure 5.3.

The Siemens process is highly energy consuming, a major part of the energy being dispersed and lost. To avoid deposition on the inner surfaces of the reaction chamber, this has to be cooled. Originally, the decomposition chamber consisted of a quartz bell jar containing one single inverted U-shaped silicon seed rod. A major advancement in polysilicon production was the utilisation of metal bell jars in place of the quartz bell jars. Quartz bell jars could not be produced in large diameters and were susceptible to breakage. The development of steel bell jars made it possible to accommodate 30 or more inverted U-rods in each reactor. This dramatically increased the productivity while decreasing the energy consumption per kilogram of produced polysilicon.

As reactions and equilibria (5.28) to (5.31) show, the deposition process generates by-products. Unfortunately, for each mole of Si converted to polysilicon, 3 to 4 moles are converted to $SiCl_4$, binding large amounts of chlorine and valuable silicon. The main industrial application of tetrachlorosilane is as a source material to produce pyrogenic (also called *fumed*) silica as described above in this chapter. The present market of fumed

silica is about 60 000 MT measured in terms of silicon unit. This presently corresponds to three times the output of polysilicon in 2000. Moreover, a significant portion of fumed silica is also produced by burning derivated by-products from the silicones industry. In the early stages of the polysilicon industry, the fumed silica business could absorb the excess of silicon tetrachloride generated by the Siemens process. This explains the arrangements between polysilicon and fumed silica producers all around the world. With polysilicon production growing much faster than silicones and fumed silica, the question was whether to eliminate or recycle the tetrachlorosilane. This became an environmental and economical necessity. The concept of recycling on site the by-product back to the valuable starting material to form a closed-loop production process is generally an ideal preferred solution today. There are two basic chemical processes applicable to reconvert $SiCl_4$ to $SiHCl_3$:

1. The high temperature reduction of silicon tetrachloride with hydrogen.

$$SiCl_4 + H_2 = SiHCl_3 + HCl \tag{5.32}$$

At about 1000°C, a 1:1 molar mixture of $SiCl_4$ and H_2 produces approximately 20 to 25% molar $SiHCl_3$ in the gaseous mixture. This process requires a fair amount of electrical energy but has a distinct advantage that the trichlorosilane produced is of very high quality because both reactants, silicon tetrachloride and hydrogen, are basically electronic grade when produced by equations (5.28) and (5.31).

2. The hydrogenation of silicon tetrachloride in a mass bed of metallurgical silicon.

$$3SiCl_4 + 2H_2 + Si = 4SiHCl_3 \tag{5.33}$$

This hydrogenation reaction produces approximately 20% trichlorosilane at 500°C, 35 atm with a 1:1 ratio of $SiCl_4$ to H_2 in one pass through a mass bed of metallurgical grade silicon in a fluidised bed reactor.

In spite of its widespread and dominant position in the industry, the Siemens process as described above suffers from the following disadvantages:

- High energy consumption, over 90% of the input power is lost to the cold walls of the reactor.
- Two power supplies and preheating of the seed rods are normally required because the high-resistivity (~230 000 ohm cm) seed rods require very high power supplies and high initial power rates to heat the rods. Therefore, a separate power supply for quartz lamps or graphite rod induction heating is used to preheat seed rods to about 400°C (~0.1 ohm cm). Lower power electrical supplies can then be used to provide continued heating and control.
- Electrical contacts to seed rods are made of graphite, which is a source of contamination.
- Power failure (especially when starting the process) causes run abortion.
- Hot spot formation and filament burn out may occur.
- Problems arise owing to gas inclusions and to non-uniform deposition at the joints.
- Gas flows and electrical power have to be adjusted during the process to obtain optimal deposition rate.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 907

- The process is operated batchwise.
- Large amount of by-products need to be handled or recycled.

More recent developed processes have attempted to overcome some of these disadvantages.

5.4.2 The Union Carbide Process

Research on this process was initiated in 1976 after the international oil crisis. The US Government funded several projects with the objective of finding a route to inexpensive solar grade polysilicon. The Union Carbide process for silane production, with fluidised bed production of polysilicon, was selected for further funding. When the funding was not provided for political reasons, Union Carbide decided to use the silane technology for the production of semiconductor grade polysilicon. Silane deposition technology for polysilicon rods was licensed from Komatsu Electronic Metals, Japan. In 1990, the business was sold to Komatsu, which became Advanced Silicon Materials (ASiMI). By 1998, two large industrial plants were constructed in the United States with the nominal capacity of 5500 MT polysilicon, thus ranking ASiMI as number two or three worldwide among the polysilicon producers. A schematic overview of the process is given in Figure 5.4.

The main process steps are as follows:

The hydrogenation of tetrachlorosilane through a mass bed of silicon metal is carried out in a fluidised bed reactor as already described by equation (5.33).

The trichlorosilane is separated by distillation while the unreacted tetrachlorosilane is recycled back to the hydrogenation reactor.

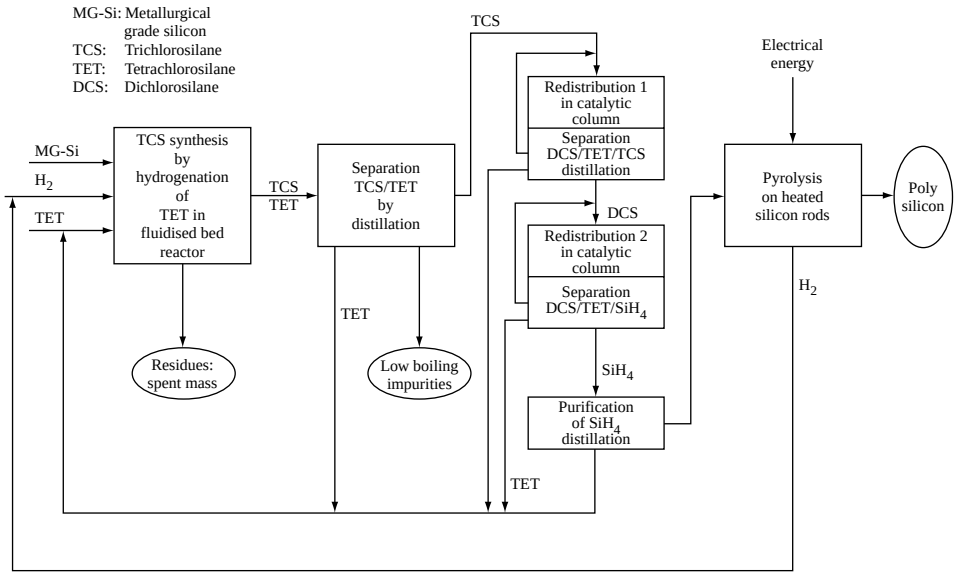


Figure 5.4 A schematic representation of the Union Carbide Polysilicon process

The purified trichlorosilane is then redistributed in two separate steps through fixed bed columns filled with quarternary ammonium ion exchange resins acting as catalyst to both redistribution equations (5.34) and (5.35):

$$2HSiCl_3 = H_2SiCl_2 + SiCl_4 \tag{5.34}$$

$$3H_2SiCl_2 = SiH_4 + 2HSiCl_3 \tag{5.35}$$

Products of (5.34) and (5.35) are separated by distillation. Tetrachlorosilane and trichlorosilane are recycled to the hydrogenation (5.33) reactor and the first redistribution step (5.34), respectively. Silane is further purified by distillation and then pyrolysed to produce polysilicon onto heated silicon seed rods mounted in a metal bell-jar reactor:

$$SiH_4 = 2H_2 + Si \tag{5.36}$$

With hydrogen and chlorine recycled, the only raw material requirement is of metallurgical grade silicon in granular form designed for fluidisation. Because equations (5.33) to (5.35) yield low portions of suitable products and because distillation has to take place after each of them, the intermediates tri- and tetrachlorosilane are recycled and purified many times before conversion to silane. This results in extremely high purity for the silane and the subsequent polysilicon. This is operated as a closed-loop process, not as a batch process.

Other advantages of using SiH_4 are that the pyrolysis may be operated at significantly lower temperature, the decomposition is complete, conversion efficiency is higher and no corrosive compounds are formed. Uniform, large-diameter, long, dense, void-free cylindrical rods of polysilicon produced this way are particularly suitable for single-crystal manufacturing by the floating zone (FZ) method.

The disadvantage of the monosilane-based process is the higher cost of the volatile molecule, since additional steps are requested to convert trichlorosilane to monosilane. Moreover, the recycling of unsuitable chlorosilanes is compulsory when choosing this option because the redistribution equations yield only a small percentage of the suitable silane.

5.4.3 The Ethyl Corporation Process

This process was developed by the US company Ethyl Corporation at the same time, in similar conditions and political context as the above-described Union Carbide process. Although managed differently, the outcome of both projects was similar in the sense that both shifted their focus from solar grade polysilicon and ended up as new commercial polysilicon processes serving the electronic industry. The Ethyl Corporation process is, by comparison with the Siemens and the Union Carbide processes, revolutionary in all aspects except the concept of purifying and decomposing a volatile silicon compound by pyrolysis.

The first radical change was the choice not to use metallurgical grade silicon as the primary raw material for silane. The idea was to make use of silicon fluoride, which is a waste by-product of the huge fertiliser industry. Tens of thousands of tonnes of silicon fluoride every year are available. This is potentially a very low-cost starting

material. Silicon fluoride is hydrogenated to monosilane by metal hydrides such as lithium aluminium hydride or sodium aluminium hydride.

$$2H_2 + M + Al = AlMH_4, \text{ M being Na or Li} \quad (5.37)$$

$$SiF_4 + AlMH_4 = SiH_4 + AlMF_4 \quad (5.38)$$

$AlMF_4$ is believed to find application in the aluminium industry, making it a valuable saleable product.

After distillation, monosilane SiH_4 is thermally decomposed to polysilicon as described by (5.36). However, to realise this process, Ethyl Corporation introduced a second radical change, not using static silicon seed rods in a bell-jar reactor but dynamic silicon seed spheres in a fluidised bed sustained by a gas stream of silane and hydrogen. A schematic representation of a fluidised bed reactor is given in Figure 5.5.

The fluidised bed reactor offers some significant advantages compared to the bell-jar reactor. Most of the shortcomings identified for the Siemens process are then eliminated. The energy losses and hence the energy consumption are considerably reduced because the decomposition operates at a lower temperature and because the same requirement to cool the bell jar is not there. Another advantage is that large reactors may be constructed and operated continuously, reducing further the capital and operating costs.

The end products are small granules of polysilicon that may present some advantages (e.g. when continuous feeding in customer process is requested) or disadvantages (e.g. not usable for direct float zone crystallisation).

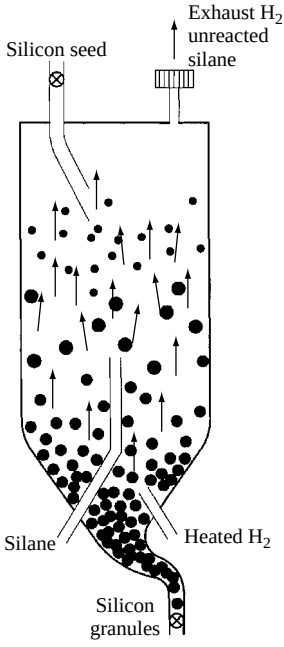


Figure 5.5 A schematic representation of a fluidised bed reactor for polysilicon production

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 910

A disadvantage of the process is the generation of powder due to homogeneous decomposition of SiH_4 in the free reactor space and the hydrogen absorption into the polysilicon deposition layer.

5.4.4 Economics and Business

The polysilicon business is completely dedicated to the semiconductor industry and therefore is a high-technology, fast-growing (annual growth approximately 10% on average) and cyclical business. With an average price of 50 US$ per kg and a sale of 20 000 MT, it is a global business of one billion US$ supporting a multibillion US$ semiconductor market. Physical volumes are small but added value is high. This may be judged by taking into consideration that metallurgical grade silicon at 1 US$ per kg is the only raw material necessary. Small amounts of hydrogen and chlorine or tetrachlorosilane are purchased as make-up to compensate losses occurring in the closed-loop process.

The industry is extremely capital-intensive. Plants have capacities ranging from 1000 to 5000 MT. One unit of 1000 MT may require a capital expenditure of 100 million US$.

The second characteristic is that the decomposition process requires large amounts of energy. In the early stages of the industry, consumption of 350 kWh per kg polysilicon was not unusual. Efficiencies improved to 150 to 160 kWh per kg, and advanced processes are now about 100 kWh per kg. The fluidised bed process is even more efficient.

The industry is concentrated in the United States (five plants, four producers), the European Union (two plants, two producers) and Japan (three plants, two producers). Small plants are also operating in PR China and the Commonwealth of Independent States (CIS).

Polysilicon is then melted and solidified into a single crystal either by the Czochralski (CZ) or by the floating zone method (FZ). Large single crystal ingots or boules are then sliced into wafers onto which semiconductor devices are built. About 15% of polysilicon production is used to make single- and multicrystalline silicon wafers for solar cell applications.

As it is an important aspect of the PV industry and sciences, crystallisation technologies are covered in Chapter 6 of this handbook and some basic principles are given in the next section of the present chapter.

5.5 CURRENT SILICON FEEDSTOCK TO SOLAR CELLS

Elemental silicon is used in photovoltaics as the main semiconductor material converting light to electricity. Two main classes of silicon must be distinguished: amorphous and crystalline. Crystalline cells are either single- or multicrystalline. Within each group of technology several variants may be distinguished. The elaboration of the cells by different silicon-based technologies and their characteristics are described in Chapter 7. According to recent market surveys [1–3], crystalline silicon is strengthening its dominant position, particularly multicrystalline silicon (including ribbon) with new capacities steadily coming on stream. (See Table 5.3 and References [1–3].)

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 911

Table 5.3 Solar modules shipment by technology Source: PV News 2001, Reference [1]

Technology	1998 [MW]	1998 [%]	2000 [MW]	2000 [%]
Single-crystalline silicon wafers	61.5	39.7	89.7	31.2
Multicrystalline silicon wafers	67.0	43.3	140.6	48.9
Ribbon/multicrystalline silicon film	6.0	3.8	16.7	5.8
Amorphous silicon/single-crystalline silicon	–	–	12.0	4.2
Amorphous silicon	19.0	12.3	27.0	9.4
Others	1.4	0.9	1.7	0.6
Total terrestrial PV shipments	154.9	100	287.7	100

(i) Amorphous silicon. Monosilane (SiH_4) is the source of silicon required for the deposition of silicon amorphous thin film in a glow discharge or low temperature plasma. Silane is mass-produced by the Union Carbide (ASiMI) and Ethyl Corporation (MEMC) methods, which are described above. The global annual output capacity is approximately 7000 MT including minor volumes produced in Japan. The major part of it is used to produce polysilicon on site (United States), the balance being sold through distributors, for example industrial gas companies, to a vast group of customers. The worldwide silane market for silicon films in the semiconductor, PV, glass and ceramic industries is about 500 MT. Applications include passivating and semiconducting layers for integrated circuits, epitaxial films, architectural glass coatings, special ceramics, surface treatment and amorphous silicon. Silane is available in quantities and purity exceeding the need of the PV market. Quantity and cost of silicon is of less importance for amorphous silicon since the specific consumption per watt output is 50 to 100 times less than that for crystalline silicon cells, that is, 100 to 400 mg/W versus 10 to 20 g/W.

(ii) Crystalline silicon. Regardless of the technology, significant quantities of silicon are required to produce multicrystalline silicon solar cells. In the present stage of the technology, 15 to 17 g/W (estimated average value) are consumed. In 2000 this had represented a quantity of approximately 4000 MT. It is expected that this specific consumption per watt will decrease with technology improvement. Firstly, up to 60% of the purchased silicon feedstock is presently wasted during the manufacturing process. Secondly, the thickness of the silicon wafer is unnecessarily high with respect to light conversion. It is presently a mechanical requirement for the further handling of the wafers and the cells. Assuming continuous growth at the same level as during the last decade, several market forecasts predict that the requirement for silicon feedstock will in 2010 reach at least 12 000 MT. If the growth is accelerated as the more recent trends seem to show, the volume demand may be as high as 30 000 MT.

Commercial silicon bulk material is of two types:

- The metallurgical grade, global annual output: 1 million MT, price 0.8 to 1.5 US$/kg.
- The semiconductor grade, global annual output: 20 000 to 25 000 MT, price 35 to 55 US$/kg.

Chemical purity of typical metallurgical grade is given in Table 5.2. For some price premium, silicon producers could upgrade it to higher purity by metallurgical treatment

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 912

Table 5.4 Metallurgical silicon expected characteristics if upgraded with best currently known metallurgical technology

Impurity	Target concentration [ppm(w)]	Target price [US$/kg]
Fe	50	5–7
Al	50	
Ca	50	
Ti	5	
C	50	
B	7	
P	7	
O	200	
Other impurities	Less than 5	

Table 5.5 Specification of chemical impurities in lowest-grade silicon currently purchased to produce multicrystalline silicon wafers

Impurity	Specification
Fe, Al, Ca, Ti, metallic impurities	Less than 0.1 ppm(w) each
C	Less than 4 ppm(w)
O	Less than 5 ppm(w)
B	Less than 0.3 ppm(w)
P	Less than 0.1 ppm(w)

Table 5.6 Electrical specification of lowest-grade silicon currently purchased to produce multicrystalline silicon wafers

Property	Specification
Resistivity	Higher than 1 ohm cm, *p*-type
Minority-carrier lifetime	Higher than 25 μs

as tentatively indicated in Table 5.4. However, impurity contents, particularly boron and phosphorus, prevent the use of such material in solar cells. The current minimum purity requested for growing multicrystalline silicon ingots for wafering is given in Tables 5.5 and 5.6.

For semiconductor grade silicon, impurity levels are at least at the ppb level and resistivity within 1000 to 30 000 ohm cm. Although purity requirements for solar cells are not as stringent as for semiconductors, the industry has up to now been forced to select its silicon raw material from the semiconductor silicon source. Silicon is one of the main, if not the largest, cost drivers of the installed PV system. It accounts for approximately 25% of the wafer cost and 15% of the cell cost. Taking into account the dominant share of silicon crystalline technology, the industry is very sensitive to the silicon

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 913

price. To reduce price, the industry has therefore selected its silicon raw material from various second-grade classifications of semiconductor silicon, being less expensive than prime-grade polysilicon. We may assume that this feedstock has cost the PV industry, on average, one-third of what prime-grade polysilicon would have cost. The present silicon feedstock for crystalline cells consists of the following factors.

1. *Rejects from crystal growth*: Crystal growth, particularly by the CZ method, generates significant amounts of unusable silicon as follows:
   - The head (or top) as well as the bottom (or tail) of the single crystal (also called *ingot* or *boule*). These sections are rejected because of size, defects in the crystalline structure and higher impurity levels.
   - Ingots from aborted runs, for example, because of electrical power failure during crystal pulling.
   - Ingots or ingot sections that fail to meet the specifications of crystal structure, electrical characteristics or carbon (C) and oxygen (O) content.
   - Crucible leavings or pot scrap.

As late as ten years ago, almost half of the polysilicon purchased for crystal growth was rejected for the above reasons. Crystal growth yields have improved and processing techniques have been developed to recycle the higher purity reject material into test or so-called *dummy* wafers. These improvements have reduced the quantity of feedstock available to the PV industry. Rejects from crystal growth, except for the pot scrap, *n*-type, and heavily doped *p*-type material, have been the ideal silicon feedstock for crystalline solar cells. Acceptable quality was available at reasonable prices. We estimate this source at 5 to 6% of all polysilicon shipped, that is, in 2000 approximately 1000 MT. Pot scrap is the remainder of the silicon melt in the quartz crucible after the CZ crystal growth is completed. After solidification, the crucible cracks and pieces of it adhere to the solidified silicon residue. Driven by the need to find new cost-effective sources, the PV industry has found ways to clean mechanically (sand blasting) and chemically (treatment by hydrofluoric acid) selected crucible residues. This has resulted in added quantities to the PV silicon feedstock. We estimated this source at 500 to 700 MT in 2000.

2. *Rejects from polysilicon*: Four to five percent of polysilicon produced is also rejected for quality reasons. These are
   - chunks and rods from aborted runs (failure under operation);
   - broken or imperfect seed rods;
   - "carbon ends," parts of the rod close to the electrical graphite contacts;
   - fines and small chips generated during processing of rods into chunk form;
   - "popcorn" or excessive dendritic growth with unacceptable surface texture that may occur under certain operative conditions;
   - small granules from the fluidised bed process;
   - rod sections that have been sawn or fabricated, so have unacceptable purity.

At 5% of polysilicon produced this amounts to another 1000 MT in 2000.

Rejects from crystal growth and from polysilicon plants in 2000 account in total for 2500 to 3000 MT at most, whereas the consumption has been at least 3500 MT and most probably 4000 MT. The balance has necessarily been covered by standard semiconductor grade polysilicon, either prime grade or a grade produced on purpose for PV customers by some polysilicon producers. As pointed out the present polysilicon capacity exceeds the actual demand by some 5000 MT. It may therefore be attractive for some producers to offer a solar grade in order to optimise their overall costs by operating their plants at higher capacity. This option will be valid for a few years (2–4) as long as installed polysilicon capacity exceeds the semiconductor demand.

Driven by the semiconductor business, polysilicon capacity will not grow as fast as the solar cell market demand. As the semiconductor market matures, the historical growth rates of 15 to 20% appear difficult to sustain; more conservative annual growth rates of 7 to 10% have been projected. Having gone through several high growth/negative growth cycles, the polysilicon producers are now adding capacity only on demand from the semiconductor industry, with guaranteed supply contracts. The producers attempt to avoid large oversupply and inventories, thereby reducing the quantity of polysilicon feedstock available to the PV industry. An annual growth of 20% is predicted as an average or "business as usual" growth rate for the solar market. The polysilicon feedstock shortage for PV can therefore only become worse. The challenge is to develop an alternative source of silicon feedstock competitive in quality and price. The fast-growing solar cell market needs to have a feedstock supply independent of the semiconductor market cycles and growth limitations.

This source must emerge as soon as possible and no later than 2004 to 2005, when polysilicon capacity has been planned to be fully booked again by semiconductor demand. This source should be readily scalable to provide more than 10 000 MT by 2010 at a price less than half the price of semiconductor grade polysilicon.

5.6 REQUIREMENTS OF SILICON FOR CRYSTALLINE SOLAR CELLS

Before exploring the potential routes to solar grade silicon, it is worth looking at the structural and chemical limitations imposed on silicon to reach optimal performances for crystalline and particularly for multicrystalline solar cells. After a short description of the valid principles of crystallisation, we will examine the effect of various impurities and the effect of structural imperfections.

5.6.1 Solidification

When silicon solidifies, a homogeneous melt will normally not result in a homogeneous solid. Concentration gradients will appear as shown in the Figures 5.6 and 5.7.

A schematic phase diagram of the silicon corner is shown in Figure 5.6. When cooling a melt of composition X_0, it starts to solidify at a temperature T_0. At a lower temperature, T_1, a fraction has solidified. At equilibrium the fractional compositions of

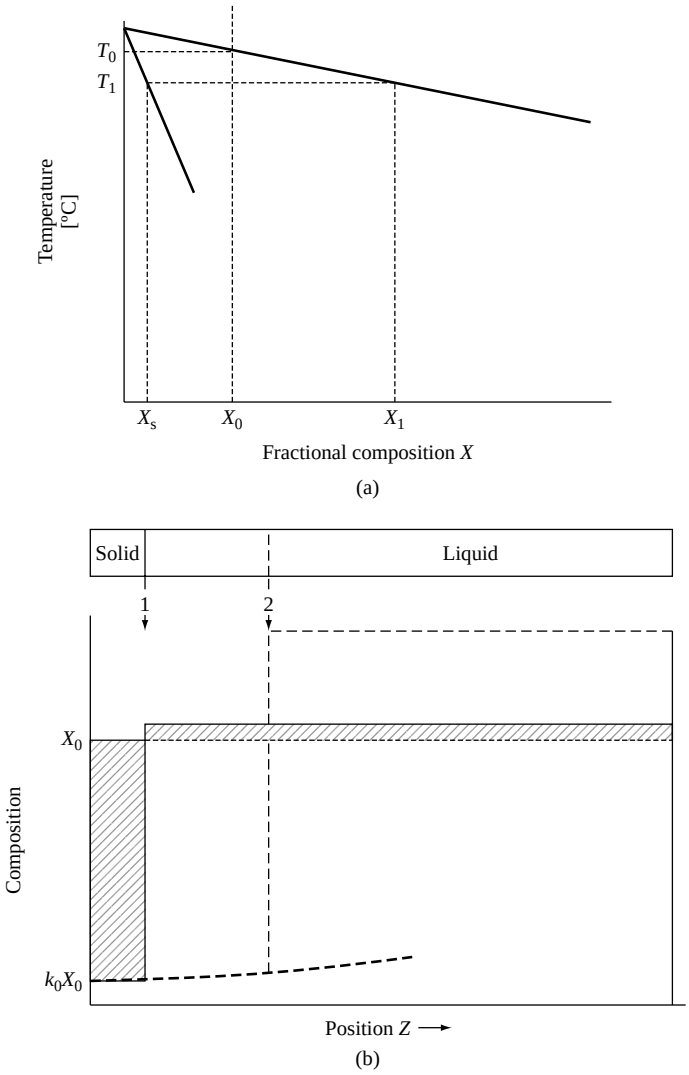


Figure 5.6 (a) Left side of a phase diagram with $k_0 < 1$ and (b) composition development in solid and liquid along a rod solidified from the left end under "normal freezing"

the solid and liquid are X_s and X_l, respectively. The equilibrium distribution coefficient k_0 at T_1 is defined as the ratio of the equilibrium solid and liquid compositions:

$$k_0 = X_s / X_l$$

(at *equilibrium* both the solid and the liquid have homogeneous compositions X_s and X_l, respectively).

During normal freezing the solid and the liquid are not in equilibrium. This principle can be used to purify the material. If (1) the solid–liquid interface is flat, (2) the

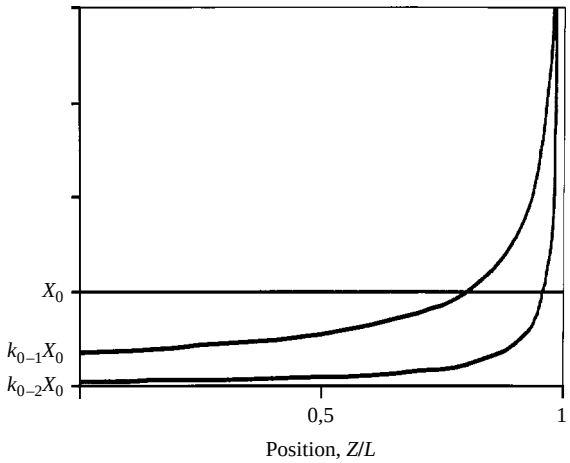


Figure 5.7 Composition profile after the entire rod of Figure 5.6 has solidified, illustrating normal solidification at two different values of k_0, $k_{0-1} = 0.35$ (upper curve) and $k_{0-2} = 0.05$

melt has a uniform composition and (3) negligible diffusion takes place in the solid, a composition profile like the one in Figure 5.6 is obtained when equilibrium prevails at the interface.

A horizontal cylinder of liquid alloy of initially uniform composition X_0 is cooled at the left end (see Figure 5.6b). Let a small amount of solid form so that the solid–liquid interface is located at position 1. When a small volume has solidified, the composition in that volume has dropped from X_0 to k_0X_0. A mass of solute proportional to the cross-hatched area to the left of position 1 has been removed from the solid and rejected into the remaining liquid. This will increase the liquid composition to a level above X_0. If we consider that local equilibrium prevails at the solid–liquid interface during solidification, the liquid and solid compositions at the interface are tied together by the equation $X_s = k_0X_l$. So when the composition of the liquid is raised, the solid composition must also rise as solidification proceeds. When the solid–liquid interface has moved to position 2, the solid composition will have gradually increased, as shown in Figure 5.6(b).

It is shown in Figure 5.7 that k_0 will have a marked effect on the distribution of impurities in solid silicon. Values of k_0 for different elements are given in Table 5.7 [24, 25]. For impurities having a low value of k_0, for example, $k_{0,\,Fe} = 8 \times 10^{-6}$, the solidification process has a large purifying effect: only one Fe atom out of about 100 000 in the melt will enter the solid when solidification starts. For elements in which k_0 is near to 1, no marked change in the concentration of impurities will appear between the melt and the solid silicon.

During the solidification process, the impurity concentration at the interface varies with the fraction of melt solidified and k_0. If the solute rejected into the boundary region is not transported immediately into the melt in front of the interface, solute will build up at the boundary. As the solute builds up, however, its concentration gradient across the boundary layer becomes steeper and the rate of transport through the boundary layer by diffusion increases until a balance is obtained between the solute being rejected into and

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 917

Table 5.7 Distribution coefficients, k_0, for some elements in silicon at the melting point [24, 25]

Element	k_0	Element	k_0
B	0.8	Ti	3.6×10^{-4}
Al	0.002	Cr	1.1×10^{-5}
Ga	0.008	Mn	1×10^{-5}
N	7×10^{-4}	Fe	8×10^{-6}
P	0.35	Co	8×10^{-6}
As	0.3	Ni	8×10^{-6}
C	0.07	Cu	4×10^{-4}
O	0.25–1.25	Zn	1×10^{-5}

out of the boundary layer. At this point the ratio $(X_l)_{\text{interface}}/(X_l)_{\text{bulk}}$ becomes constant. The effective distribution function k_{eff} is defined as

$$k_{\text{eff}} = X_{\text{s, interface}} / X_{\text{l, bulk}}$$

and tells a great deal about liquid mixing in front of the interface. Poor mixing in the melt results in $k_{\text{eff}} = 1$. Good mixing (no build-up) means $k_{\text{eff}} = k_0$.

Burton *et al*. [26] have derived an expression for k_{eff} when a rotating crystal is pulled from the melt at a velocity v:

$$k_{\text{eff}} = k_0/[k_0 + (1 - k_0]\exp^{-\Delta})$$

where

$$\Delta = v\delta/D_l \text{ and } \delta = 1.6\ D_l^{1/3} \cdot \nu^{1/6}\ \omega^{-1/2}$$

where δ is the distance from the growing interface to where the concentration in the melt is uniformly equal to $X_{l,\text{bulk}}$, D_l is the diffusion coefficient of the solute, ν is the viscosity and ω the crystal rotation rate.

Kodera [27] has applied this to obtain δ/D_l and D_l values for melts doped with various impurities at different rotating rates, as shown in Table 5.8. For CZ-grown silicon where v is around 1 mm/min, Δ is small and k_{eff} is close to k_0 [24].

The ability to purify molten feedstock by solidification will therefore vary for the different impurities and with solidification velocity. The validity of the assumption of no diffusion in the solid will not be met at a slow solidification rate and some "back" diffusion will take place. In addition, impurities diffuse from the crucible into the solid, the rate depending on temperature and time. Selection of the optimal casting process parameters is complicated and improvements will depend on the ability to model the total solidification process.

5.6.2 Effect of Crystal Imperfections

Silicon has a diamond lattice atomic structure being cubic with atoms joined to four nearest neighbours by covalent bonds at tetrahedral angles. If the atoms are represented

Table 5.8 Values of δ/D_l and D_l for some elements [27]

Impurity element	Rotation rate [rpm]	$\frac{\delta}{D_l}$ [s/cm]	Diffusion coefficient, D_l [cm^2/s]
B	10	170 ± 19	$(2.4 \pm 0.7) \times 10^{-4}$
	60	84 ± 37	$(2.4 \pm 0.7) \times 10^{-4}$
Al	10	86 ± 34	$(7.0 \pm 3.1) \times 10^{-4}$
	60	40 ± 17	$(7.0 \pm 3.1) \times 10^{-4}$
P	5	127 ± 36	$(5.1 \pm 1.7) \times 10^{-4}$
	55	60 ± 19	$(5.1 \pm 1.7) \times 10^{-4}$
As	5	190 ± 53	$(3.3 \pm 0.9) \times 10^{-4}$
	55	79 ± 16	$(3.3 \pm 0.9) \times 10^{-4}$

as hard spheres, the diamond structure fills 0.34 of the available space, compared to 0.74 for a close-packed structure such as aluminium. The formation of defects in the silicon lattice creates local electronic disturbances because of the nature of the atomic bonding.

The following description follows standard textbook presentations in physical metallurgy and concentrates on some lattice defects appearing in silicon, which are important when discussing the effect of impurities on the electronic properties.

It is common practice to divide the lattice defects into the following categories:

Plane defects
- Stacking faults
- Twin boundaries
- Grain boundaries

Line defects
- Dislocations

Point defects
- Vacancies
- Self-interstitials
- Substitutional impurities
- Interstitial impurities

Volume defects
- Precipitates

Single-crystal material consists of one grain of a continuous crystal structure. Multicrystalline (mc-)/polycrystalline (pc-) material consists of many separate grains with inclined crystal planes that meet at *grain boundaries*. Depending upon the misorientation between adjoining grains, the interfaces are termed *subgrain* boundaries ($0-5^\circ$) and *grain* boundaries. *Twin* boundaries arise where one crystal grain is a mirror image of its neighbour. Grain boundaries may have a more open and less well organised atomic structure. Atomic diffusion is therefore easier along grain boundaries than in the bulk crystal. Because of the lack of crystal order, some of the covalent bonds are broken and the grain boundaries may therefore carry electric charge. The electrical properties of the grain boundaries vary with the amount of atomic disregister.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 919

The concept of a linear lattice imperfection called a *dislocation* arose primarily from the study of plastic deformation of crystalline materials. The dislocations are borderlines where the atomic planes are out of register. At dislocations some of the covalent bonds in silicon are broken. The dislocations may therefore carry electric charge.

Because of the disregister, the dislocations are surrounded by stress fields and may attract impurity atoms.

In *n*-type silicon, donor impurities can provide electrons to fill the missing charge from a missing bond at the dislocation. As a result of the acquisition of electrons, the dislocation becomes negatively charged, which again may attract the positively charged donor impurities. This situation may lead to a space charge region in the form of a cylinder in which positively charged donor ions surround a negatively charged dislocation.

In *p*-type material, the dopant atoms may accept electrons from a dislocation. In this case the cylinder-shaped space charge region has a positively charged dislocation surrounded by negatively charged acceptor ions.

The category of *point defects* include vacancies, interstitials and impurities either present as intentionally dopant atoms added to control the properties of silicon or unintentionally incorporated as contaminants from the raw materials, processing or crystal growth.

A *vacancy* is a missing atom at a silicon site. The result of the removal of a silicon atom is the formation of four dangling atomic bonds with unpaired electrons and some lattice relaxation. Vacancies thus tend to exhibit acceptor-like behaviour. The concentration of vacancies, n/N_0, in silicon at a given temperature can be determined from the following expression:

$$n = N_0 \exp(-E_v/kT)$$

where n = number of vacancies/volume, N_0 = number of atoms/volume, k = Boltzmann's constant and T = temperature in K. E_v is the formation energy, which is the energy required to take an atom from a lattice site inside the crystal to a lattice site on the surface. In silicon, $E_v \approx 2.3$ eV.

A *self-interstitial* can be formed by inserting a silicon atom into one of the holes in the structure. The energy of formation of interstitials, E_i, in the loosely packed diamond structure is lower than the formation energy of vacancies, $E_i \approx 1.1$. eV.

An interstitial has four valence electrons that are not involved in covalent bonding with the adjoining atoms. These electrons may be lost to the conduction band and the interstitial may behave as a donor.

At low concentrations the impurities exist as single atoms in the matrix – so-called *solid solutions*. Atoms in the solid solution are incorporated in the matrix in two ways. They can substitute for an atom of the host crystal and maintain the regular atomic structure of the crystal. In this case they are known as *substitutional impurities.* Alternatively, the impurity atoms can occupy positions squeezed in between the atoms of the host crystal. Then they are known as *interstitial impurities*.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 920

The energy levels associated with point defects in silicon are fairly deep. They serve as centres for minority-carrier recombination and therefore reduce the carrier lifetime. The lifetime is inversely proportional to the concentration of point defects.

The diffusion of atoms in a crystal proceeds by thermally activated jumps from sites to sites. The diffusivity is expressed by the diffusion coefficient, D, with units cm^2/s:

$$D = D_0 \exp(-E_D/kT)$$

where D_0 is a constant and E_D is the activation energy for the jumping process.

At higher concentrations the impurities may agglomerate. The ability to be in solid solution increases with temperature (see Section 5.6.3). The impurities in a material at a certain purity level may thus be in solid solution at elevated temperatures and form *pairs* with other atoms or *precipitate* at dislocations, at grain boundaries or with other impurities at lower temperatures. The sequence of precipitation is characterised by *nucleation, growth* in which excess atoms diffuse to precipitates and Ostwald *ripening* in which a growth competition exists where large precipitates grow at the expense of small ones.

Minority carriers (being electrons on the p-side and holes on the n-side of the solar cells) flow more readily within a grain than across grain boundaries. Minority carriers may recombine with majority carriers at recombination centres like impurities, precipitates, dislocations and grain boundaries. The average distance that a minority carrier will travel within silicon depends upon the density of recombination centres and their recombination ability. The longer the distance between recombination centres, the better the efficiency of the cell. Precipitation of impurities reduces the number of atoms in solid solution and may therefore change the minority-carrier diffusion length. The density of precipitates is governed by the cooling rate, the ramping sequences and the diffusivity of the impurity elements.

At the surface there are many recombination sites. In the final cell production sequence, gettering and passivation will change the conditions near/at the surface.

Measurements of the local electronic properties in individual grains and near grain boundaries show that the properties may vary from grain to grain. It is hoped that the results of such measurements may stimulate further theoretical and experimental work to clarify the relations between grain structure and the electronic properties in multicrystalline silicon.

The density of dislocations influences the lifetime of the minority carriers. Experiments show good correlation between areas having a high dislocation density and a short effective lifetime. To control the dislocation density has therefore become an important task. The dislocation movements and multiplication and their interaction with impurity atoms being in solid solution or existing as complexes during the production processing steps challenges our ability to model complex relationships. However, it is being claimed that numerical simulation is used as a valuable tool to optimise the crystallisation and cooling processes at plants. This is the kind of work that certainly will develop during the coming years.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 921

5.6.3 Effect of Various Impurities

It is well known that impurity atoms have a strong effect on the efficiency of silicon as a photovoltaic material. It is also known that the effect of impurities can be changed by heat treatments and by exposing the material to gettering atmospheres in which selected elements diffuse into silicon and combine with the impurities (a semantic parallel to getter pumps in which oxygen is adsorbed by certain elements).

The impurity atoms may appear as solid solutions, as pairs with other elements, for example FeB, or as larger aggregates/precipitates with silicon and/or other elements, for example Fe_2Si. This depends upon the temperature, the concentration and the density of the imperfections (dislocations, grain boundaries). If the temperature or the (chemical) surroundings are altered, it will take some time before a new equilibrium is established. The time to reach equilibrium may depend on parameters such as temperature, cooling/heating rate, and chemical composition, grain size, dislocation density and others. When comparing results from literature values in which the specifications of relevant parameters are not defined, it is likely that differences may appear.

Most of the impurities in silicon used for PV cells exist at very low concentrations. Since measurements of trace quantities are difficult, much of the progress has occurred when new and better instrumentation has become available. Over the years many review articles and books dealing with the effect of impurities in crystalline silicon have been published. The interested reader is therefore encouraged to look up some of the references for a wider treatment of the subject [25, 28–33]. The following will therefore be an attempt to briefly summarise and update the general knowledge of the subject.

The maximum solid solubilities, $X_s(\max)$, of impurities in silicon are related to the distribution coefficients at the melting point according to an empirical relation found by Fischler [34]:

$$X_s(\max) = 0.1k_0$$

or in units of atoms/cm^3:

$$C(\max) = 5.2 \times 10^{21} k_0$$

Although deviations have been found for nitrogen, carbon and oxygen [35], the relation may be useful.

5.6.3.1 Atoms from Groups IIIA (B, Al, Ga. . .) or VA (N, P, As, Sb. . .)

These atoms act as substitutional impurities in silicon. At a site where a Group VA impurity (e.g. phosphorus) has replaced a silicon atom, four *d*-electrons are bound to the Si neighbours, while the fifth electron is weakly tied to the Group VA atom. The fifth electron is not completely free to move, but it is easily activated to the conduction band. Group VA atoms are therefore *donor* atoms.

In an analogous way, Group IIIA atoms (e.g. boron) do not have enough valence electrons to satisfy the four covalent neighbour bonds. This gives rise to holes weakly tied to the Group IIIA atoms. These impurities therefore create energy levels for electrons in the forbidden gap just above the valence band edge. Group IIIA atoms are called *acceptors*.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 922

To be able to control the level of doping, the concentration of unwanted elements from the Groups IIIA and VA must be well below the concentration of the doping element.

Our theoretical understanding of the properties of impurities in silicon other than those from the Groups IIIA and VA is less well developed. However, as new instrumentation has become available, a lot of experimental results of high scientific standard have been published.

5.6.3.2 Carbon

Carbon is a common substitutional impurity. Like silicon it has four valence electrons and is therefore electrically neutral. The carbon atom is smaller than the silicon atom and may therefore be involved in precipitation of species expanding the lattice like silicon oxide.

The solid solubility limit is $C_S = 3.5 \times 10^{17}$ atom/cm^3 at the melting point or $C_S(T) = 4 \times 10^{24} \exp(-2.3 \text{ eV}/kT)$ atoms/cm^3, where k is the Boltzmann's constant [89, 91].

In metallurgical grade (MG) silicon, carbon is present at levels above the solid solubility limit and SiC precipitates are commonly present. In electronic grade (EG) the concentration of carbon is low. But in contact with the crucible and the carbon-rich atmosphere (graphite), it is difficult to avoid contamination of the melt.

As a substitutional element, carbon diffuses rather fast [$D = 1.9 \exp(-3 \text{ eV}/kT)$ cm^2/s], but much slower than interstitial impurities.

5.6.3.3 Oxygen

The subject of oxygen atoms in silicon has been studied for many years. Since results from experiments performed under different thermal histories have varied, the discussion goes on.

Oxygen atoms in solid solution are electrically inactive and predominantly enter interstitial sites. The equilibrium value of the solid solubility at the melting point is generally accepted to be 1×10^{18} atom/cm^3. However, the results from measurements of solid solubility as a function of temperature differ markedly. The value of k_0 will therefore vary (see Table 5.7).

In a work by Itoh and Nozaki [32] an important point was made that the concentration of oxygen at equilibrium at a given temperature was determined by the type of oxide that was stable at that temperature. So if the type of oxide varied, it would be expected that bends in solid solubility curves at temperatures at which the silicon oxides change would be observed. However, up to 1200°C, SiO_2 is the stable chemical form and the solubility $L = \exp(\Delta S/R) \exp(-\Delta H/kT)$ will have constant values of ΔS and ΔH.

On cooling, super saturation is easily obtained and oxygen precipitates at a rate that depends on oxygen content, temperature, and time at temperature and nucleation sites. Carbon is also found to influence the precipitation of silicon oxides, possibly because the carbon atoms reduce the expansion of the lattice when silicon oxide is growing.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 923

Oxygen has a high diffusion coefficient, $D = 0.13 \exp(-2.53/kT)$ [36]. By heat treatments, the distribution of oxide particles may change since particles may get dissolved when heated and grow when cooled. Oxygen atoms are found to change the effect of other impurities, a process known as internal gettering.

One important oxygen source is the crucible. Silicon is normally melted in high-purity-fused quartz (SiO_2) crucibles (single crystals) or quartz crucibles coated with high-purity Si_3N_4 (multicrystalline ingots). If holes appear in the coating, quartz will easily dissolve in the melt and raise the level of oxygen in the solid.

The variety of effects observed with oxygen in silicon will truly stimulate further research for a long time.

5.6.3.4 Transition metals

Properties of transition-metal impurities in silicon have been reported in a large number of articles, reviews and books [28–30, 33]. The transition metals are presented by the symbols 3d, 4d and 5d, which specify the outer electron configuration of a neutral atom. Most of the metals forming deep energy levels ("midway" between the conduction band and the valence band) in silicon belong to this group and have therefore a large influence on the electronic properties of silicon.

The main impurities found in silicon belong to the 3d transition metals (Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu). In general, these are present as interstitial impurities.

The diffusivity increases with increasing atomic number in the 3d row (see Figure 5.8), with Ni and Cu having the largest diffusion coefficients known in silicon.

The concentration of vacancies is very low in silicon. The high diffusion coefficients of 3d elements can therefore only be explained by interstitial diffusion mechanisms independent of vacancies.

The 3d transition metals have a steep temperature-dependent solid solution limit that make them easily supersaturated during cooling (see Figures 5.9 and 5.10). They therefore frequently form complexes/precipitates at dislocations, grain boundaries or other lattice defects.

The solubility at room temperature is very low, of the order of 1 atom/cm^3. However, some 3d impurities are mobile even at room temperature. So the atoms with the highest mobility (Co, Ni, Cu) will be out of the solid solution during or just after the cooling. The 3d metal atoms with a low diffusivity may remain in the solid solution at interstitial sites for a much longer time after cooling. This may depend upon crystal perfection, which determines the diffusion length to reach sinks such as dislocations, grain boundaries and precipitates.

The solid solubility of the 3d transition metals in silicon is shown in Figure 5.10.

A severe deterioration of the electrical properties is expected from solid solution impurities capturing the minority carriers (being electrons on the *p*-side of a *p-n*

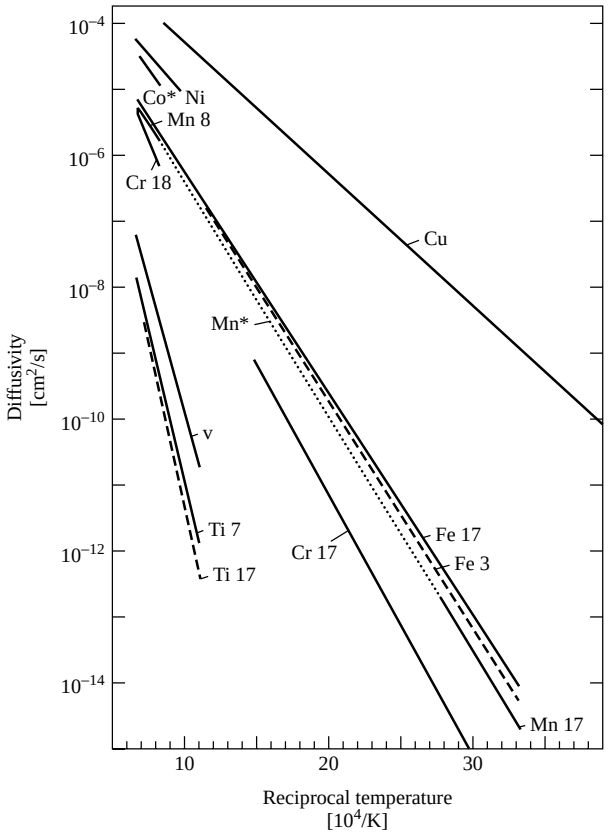


Figure 5.8 Diffusivities of 3d transition elements. Reproduced from *Metal Impurities in Silicon-Device Fabrication*, Graff K, 29, 2000, © Springer-Verlag GmbH & Co. KG

junction). The low-injection-level minority-carriers lifetime, τ_0, is inversely proportional to the impurity concentration, $N(1/\text{cm}^3)$ [28]:

$$\tau_0 = (\sigma v N)^{-1}$$

where v is the thermal velocity, which is the average speed of the electrons as they randomly collide with atoms, impurities or other defects, and σ, having the units of cm^2, represents the impurity atoms effective cross-section for the capture of a minority carrier.

Here the carrier capture cross-section for electrons, σ_e (cm^2), must be inserted for p-type silicon and σ_h for holes in n-type silicon. The thermal diffusion velocity v of electrons at room temperature is 2×10^7 cm/s.

The capture cross-sections for different transition metals can differ by several orders of magnitude. As a consequence, the carrier lifetime of a silicon sample can even be determined by an impurity of minor concentration if this is a “lifetime killer” with a high

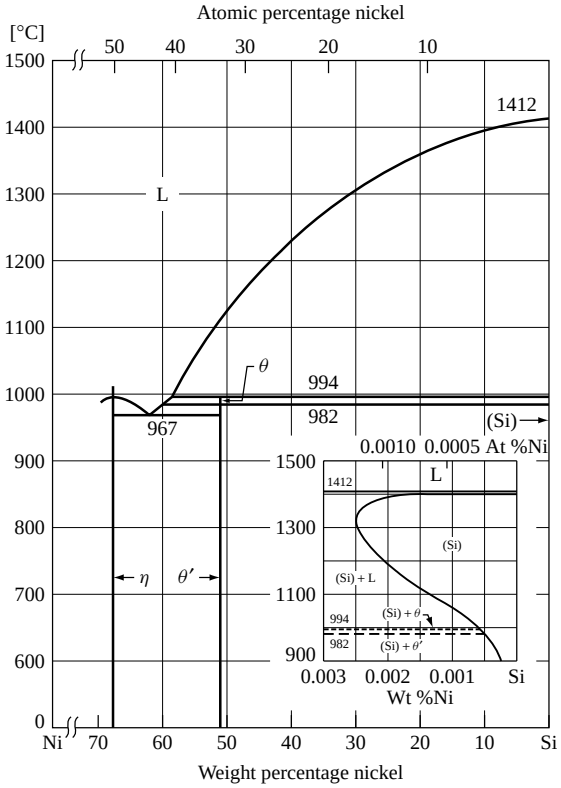


Figure 5.9 The Ni-Si phase diagram. The shape of the solidus line on the Si side is shown in the inserted window. It is seen that the solid solubility of Ni in Si is low, that it increases at temperatures above the peritectic point (at 994°C) (retrograde solubility) and reaches a maximum at about 1300°C. Redrawn from ASM Handbook, Vol. 8, *Metallography, Structures and Phase Diagrams* 8th Edition, ASM International, Materials Park, Ni-Si Phase Diagram, p. 325

minority-carrier capture cross-section. Therefore, the tolerable impurity concentration for acceptable lifetime values depends upon the chemical nature of the respective impurity, and its carrier capture cross-section for electrons in *p*-type silicon and for holes in *n*-type silicon. Both parameters can differ by orders of magnitude and consequently the acceptable concentration for a defined impurity can be quite different in *p*- and *n*-type silicon (Figures 5.11 and 5.12; [30]).

Copper and nickel have high diffusivities and low capture cross-sections. These elements will rapidly enter a low solid solution level after cooling to room temperature and may therefore be expected to have less effect on lifetime than elements with lower diffusivities (Fe, Ti) and high capture cross-sections.

5.6.3.5 Precipitates

Except for copper, which forms Cu_3Si particles when cooled, the other 3d metals form $MeSi_2$ precipitates. Investigations of crystallographic structure, morphology and

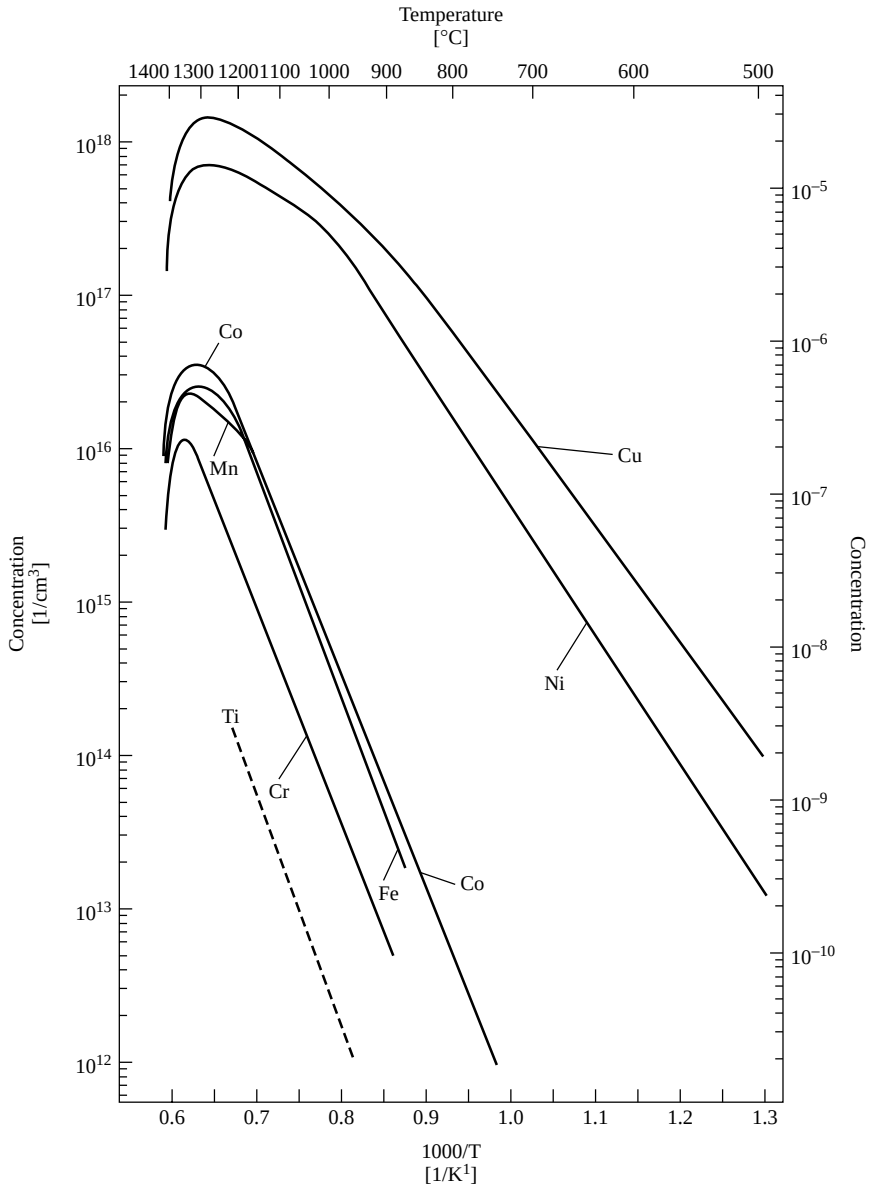


Figure 5.10 Solid solubility of 3d transition elements in silicon. Sc is still missing, but is assumed to be a little below the values of Ti. Adapted from *J. Appl. Phys.* Weber E, **A30**, 1–22 1983, © Springer-Verlag GmbH & Co. KG. The values for Ti are taken from [37]

composition of several transition-metal precipitates in silicon identify crystalline silicides, $FeSi_2$, $CoSi_2$, $NiSi_2$, as compounds that influence the minority-carrier diffusion length. The morphology and density depend upon temperature and time at temperature. Elements having a high diffusivity like Ni and Cu precipitate easily. This reduces the number of atoms in solid solution and may therefore change the electric properties. Kittler and

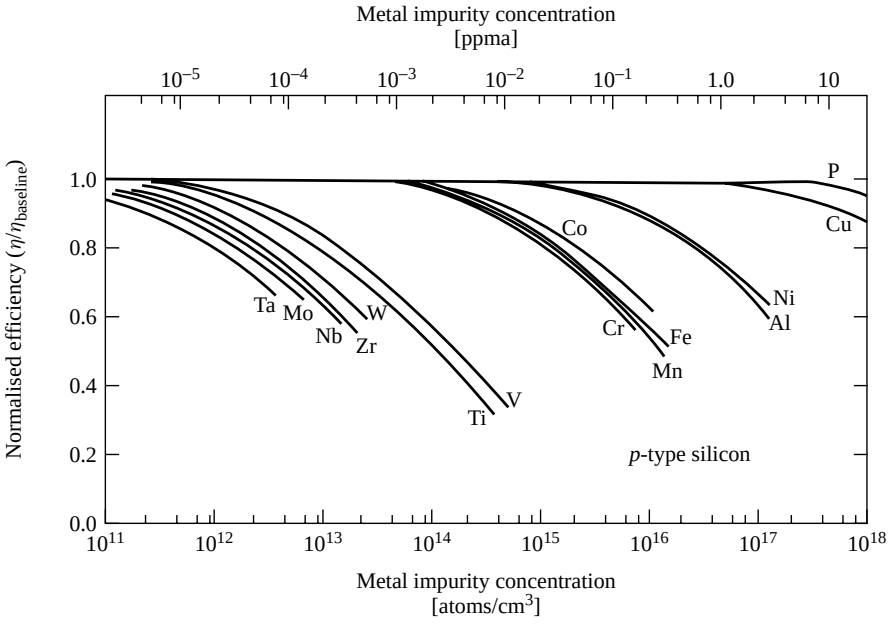


Figure 5.11 Solar cell efficiency versus impurity concentration for 4 ohm cm p-base devices [30]. Reproduced from Davis Jr. J *et al.*, *IEEE Trans. Electron Devices* © 1980 IEEE

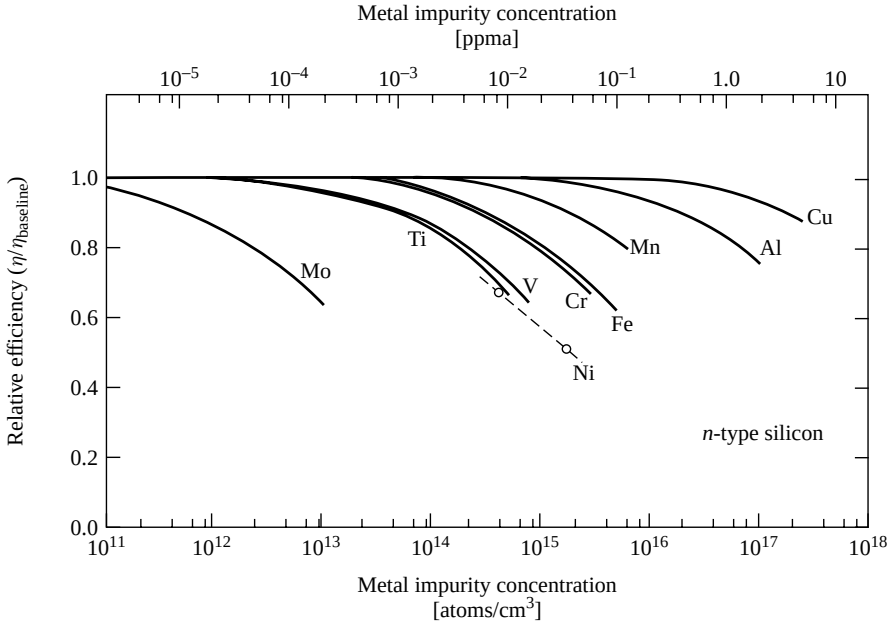


Figure 5.12 Solar cell efficiency versus impurity concentration for 1.5 ohm cm n-base devices [30]. Reproduced from Davis Jr. J *et al.*, *IEEE Trans. Electron Devices* © 1980 IEEE

co-workers [31] studied the precipitation and coarsening of $NiSi_2$ and the effect of precipitates upon the minority-carrier diffusion length in *n*-Si single crystals (FZ material). Scanning electron microscopy (SEM) using the method of electron beam–induced current (EBIC) revealed that $NiSi_2$ precipitates were efficient recombination centres and that the minority-carrier diffusion length, L_D, was related to the precipitate density, N_P, by

$$L_D = 0.7 \cdot N_P^{-1/3}$$

This relationship revealed that the diffusion length depends only on the density of precipitates and not on the concentration of impurities. Therefore, a suitable temperature process can increase the diffusion length, L_D. This occurs if during this process large precipitates grow at the expense of the smaller ones, increasing the free distance between the precipitates. This ripening process of precipitates is observed repeatedly during heat treatments.

5.7 ROUTES TO SOLAR GRADE SILICON

To design a specific route to solar grade silicon between the two existing commercial grades, that is, metallurgical and semiconductor (polysilicon), is not a new challenge. It was first seriously and massively addressed in the mid-seventies after the first international oil crisis. Considerable R&D efforts were made in the years 1975 to 1985 particularly in the United States under the guidance of the Department of Energy (DoE). During and at the end of this productive period, numerous specific articles, exhaustive review-studies and symposium proceedings devoted to these topics were published [38–40]. The following gives a summary and an update of these efforts. All approaches investigated are classified within four subgroups.

1. Crystallisation is generally an effective purification technique. Crystallisation is also a necessary step to produce the silicon device. The question is whether the crystallisation steps in the manufacturing chain to the solar cell can advantageously be combined with the purification.
2. Upgrading the purity of metallurgical silicon is a technique with the potential to produce millions of tonnes of silicon as we have demonstrated earlier in this chapter. The goal of and the challenge for this alternative are to achieve adequate purity at acceptable cost. A guiding number for this option has been and should remain less than 10 US$/kg.
3. The polysilicon process involving the synthesis, the purification and the decomposition to elemental silicon of volatile silicon hydrides is proven technology. This is presently the sole reliable source of silicon for solar cells. Simplifying it, exploring or revisiting related routes seems therefore an attractive alternative. The challenge in this particular case is to achieve sufficient output at acceptable cost.
4. Electrolysis, electrolytic methods, metallic reduction of silicon and so on are summarised in a fourth category of methods.

5.7.1 Crystallisation

The efficiency of the crystallisation processes may be predicted by the segregation or the distribution coefficient of each impurity element (see Section 5.6.1) between the solid and

the liquid silicon phases. Published data (see Table 5.7) clearly show that the elements belonging to the Group IIIA (B, Al, Ga) and the Group VA (P, As, Sb) have a coefficient close to 1, making it difficult to separate them from the silicon. These elements are the electronic dopants for silicon and their concentration must be closely controlled during all steps of the manufacturing process. Their chemical and physical behaviour in interaction with silicon are two aspects of their close neighbourhood to this element. Crystallisation as such cannot be satisfactory without prior treatments, specifically to remove these elements, particularly P and B.

- *Crystallisation from a silicon-melt*: Various methods have been claimed to be useful for refining silicon by directional solidification resulting in large oriented crystals. All these various crystallisation methods are also used in manufacturing the silicon photovoltaic devices, and they are described in Chapter 6 of this handbook. Other valuable references are to be found in [39, 40].
- *Crystallisation from an aluminium melt*: Silicon forms a unique eutectic phase at the concentration of 12.6%(w) in aluminium. Several companies (Union Carbide, Alcoa and Wacker) have tried to take advantage of this property in order to purify silicon. By cooling a hypereutectic composition, pure silicon phases crystallise and precipitate from the aluminium melt. After solidification of the hypereutectic alloy, the aluminium matrix can be dissolved by leaching, releasing the pure silicon crystals [39, 41–44]. Published data indicated that the aluminium content in the resulting silicon was as high as 300 ppm(w). The segregation coefficient of aluminium in silicon limits further separation. A further reduction of aluminium, by directional solidification of the silicon thus obtained, would add unacceptable cost just for this element. Other metallurgical routes to eliminate Al at 1 ppm(w) level or less need to go through silicon melting and subsequent extraction with the double risk of recontamination and heavy material losses.

5.7.2 Upgrading Purity of the Metallurgical Silicon Route

The processes used in the elaboration of metallurgical grade silicon for the aluminium and the chemical industries proves that most of the metallic impurities could by repetition and combination of methods be decreased to a rather low level, for example 100 ppm(w) (see Table 5.9). Since metals could be further decreased by a post-directional solidification, it was rational to explore the full potential of this expected low-cost route.

5.7.2.1 Use of pure raw materials and pure lining

The use of pure linings in the furnaces and the intermediate vessels containing the liquid silicon was a first approach. It was investigated by several companies and research

Table 5.9 Typical best results achieved with the carbothermic reduction using high-purity raw materials

Impurity	B	P	Al	Fe	Ti
Minimum content ppm(w)	2	1	100	100	10
Maximum content ppm(w)	4	3	300	200	20

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 930

consortiums such as Dow-Corning, Elkem/Exxon, Siemens, NEDO, to mention the most accomplished [45–52]. Quartz or silica sand selected for their original purity were chemically treated by leaching or precipitated from water glass solutions. Purified carbon black was used as the reduction material in most of these approaches. The morphology of sand or precipitated silica and carbon black are quite different from, respectively, lumpy crystalline quartz and bulk carbon, charcoal or woodchips currently used in the carbothermic reduction. New technologies, in particular, to agglomerate the powder-like raw materials had to be developed. Pellets consolidated by means of organic non-contaminating binders (e.g. sucrose) or rice hulls perceived as potential high-volume, low-cost raw materials since they contained both silica and carbon are worth mentioning. Extruded pellets of cocked rice hulls, with sucrose, as a binder appeared to be reactive in the submerged arc furnace technology. Using these special raw materials, boron and phosphorus were found in the range of 1 to 4 ppm(w). In standard metallurgical grade, B and P range between 7 and 50 ppm(w), typical average values being around 25 ppm. The boron level achieved with rice hulls could be as low as 1 ppm(w). However, phosphorus was as high as 40 ppm(w) and would request an additional and specific treatment to make use of this source of silicon in solar cells.

5.7.2.2 Post-treatment by chemical leaching

Metallic impurities segregate at the grain boundaries, forming intermetallic phases with silicon, basically made of silicides and silicates. Hydrometallurgical upgrading of silicon by acid treatment and leaching has a long history. The *Silgrain* © process, a trademark of Elkem ASA, is currently producing 30 000 MT metallurgical grade silicon by refining ferrosilicon (90% Si) in an acid liquor of ferric chloride. The process is particularly efficient in removing iron and transition elements because these have very low segregation coefficients in silicon. The dissolution of iron and the transition elements in the leaching liquor is eased and enhanced by the formation of stable intermetallic phases with Al, Ca or isostructural elements, for example, Sr, Ba, Ga or a lanthanide. The process has been further developed during the past 15 to 20 years with solar grade silicon as an objective [53–56].

Surface treatment or leaching by hydrofluoric acid or by HCl/HF mixtures is also well known to wash out residual impurities present as oxides and silicates. Grinding silicon prior to leaching to increase the surface exposure is a way to enhance the purification [57, 58]. Acid leaching has therefore been envisaged by several groups and companies, for example Wacker and Elkem, as a post-treatment purification after the carbothermic reduction [53–58].

However, it must be emphasised that this method cannot in practice refine silicon below the dissolution limit characteristic of each element at the solidification composition. For the major impurities Fe, Ca and Al present in metallurgical grade silicon the limit of dissolution as obtained from classical refining in a ladle exceeds by far the ppm level. Effective post-treatment leaching of metallurgical grade silicon has made it possible to achieve less than 10 ppm(w) for these major elements and less than 1 ppm(w) for minor transition elements. One pass of directional solidification was then envisaged as sufficient to achieve the requested purity for solar cells.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 931

The acid treatment or leaching methods is not effective in removing interstitial and substitution impurities such as boron, carbon and oxygen. Adding Ca to the silicon alloy prior to acid treatment, however, proves that P could be reduced by a factor of 5 down to a concentration less than 5 ppm(w), probably because P is dissolved in calciumsilicide [55, 56]. Adding barium also proves some effect on boron dissolution.

5.7.2.3 Post-treatment by extraction metallurgy in ladle

Post-treatments of liquid silicon are common practices in refining metallurgical grade silicon for applications in aluminium and chemical industry. The objective is then to adjust Al, Ca and possibly C to a suitable concentration of some hundreds or thousands of ppm(w). To that respect the reader may with great benefit consult Schei's *et al*. comprehensive handbook [13, 23]. As described above, crystallisation and leaching are efficient means of removing chemical elements with high ability to segregate from liquid to solid silicon, that is, Fe, and most of the metallic transition elements. Extraction metallurgy from a liquid phase of silicon, either liquid–liquid, liquid–solid or liquid–gas, has received considerable attention in order to remove the critical elements P, B and C. When silicon is kept liquid, it is possible to displace the equilibrium between both phases present and gradually remove the unsuitable impurity, by continuously extracting it.

Impurity(liquid silicon) = Impurity(liquid slag)	*Ksi/ls*
Impurity(liquid silicon) = Impurity(solid slag)	*Ksi/ss*
Impurity(liquid silicon) = Impurity(gas)	*Ksi/g*

Particular attention has been paid to boron and phosphorus because these elements are the major *p*- and *n*-type dopants of silicon and because they coexist in metallurgical grade silicon in concentration of one or two orders of magnitude too high for solar cell applications.

Boron has nearly the same affinity towards oxygen as silicon. Boron forms gaseous suboxides BO being analogous to SiO, and its stable oxide B_2O_3 behaves similar to SiO_2 in the presence of alkaline earths at slag-forming temperatures. Therefore, there are good reasons to expect the removal of boron as an oxide constituent of the extracting slag or as a gaseous suboxide at elevated temperature. Both theoretical possibilities have been experimentally verified.

Since the work published by Theuerer in 1956 [59], it has been known that liquid silicon becomes purified with respect to boron when brought in contact with a gas mixture of $Ar–H_2–H_2O$. The sole role of H_2 and H_2O assisting the extraction has been emphasised by several authors such as Khattak *et al*. [60–63], whereas Amouroux, Morvan *et al*. of the University of Paris [64–67] and the Japanese group of Kawasaki Steel/NEDO [68–72] have underlined the benefit of using an oxidative plasma in the presence of moisture and hydrogen. Amouroux, Morvan *et al*. have also shown that boron elimination was enhanced when fluoride, for instance in form of CaF_2, was injected into the plasma gas.

Experiments in removing B by slag extraction have been done by several companies and groups including Kema Nord, Wacker, Elkem and NEDO/Kawasaki. Schei [73] has

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 932

described a counter-flow solid–liquid reactor to remove boron by fractional extraction in a semi-continuous process in a patent assigned to Elkem ASA of Norway.

It has been demonstrated that phosphorus could be evaporated from a silicon melt under vacuum conditions [74, 75]. Miki *et al*. [76] have explained the thermodynamics of this process by reactions involving mono- and diatomic phosphorus in the gas phase:

$$P_2(g) = 2\underline{P}(1\%\ \text{inSi (l)}) \tag{5.39}$$

$$P(g) = \underline{P}(1\%\ \text{inSi (l)}) \tag{5.40}$$

Underscored symbols refer to dissolved element in liquid silicon as already defined in Section 5.3.2.

Silicon produced by carbothermic reduction is so to speak supersaturated in SiC when tapped from the furnace and may contain as much as 1000 to 1500 ppm(w) C. As this silicon is cooled down to the solidification temperature, the majority of this carbon precipitates out as SiC particles leaving around 50 to 60 ppm(w) in liquid silicon. Carbon removal from liquid silicon is therefore a two-step operation:

1. the removal of precipitated SiC as close to the solidification temperature as possible
2. the removal of dissolved $\underline{C}$ by oxidation to CO(g).

As already mentioned, SiC particles become effectively captured by the slag phase during oxidative refining in which the main objective may be to remove Al and Ca as industrially practised today or in a similar operation with the purpose of removing B or P. Depending on the temperature and the degree of slag/molten silicon intermixing, this treatment may give a product with 80 to 100 ppm(w) C. Other methods, which have been applied and proved effective at a temperature closer to the solidification temperature, are filtration, centrifugation or settling combined with slow cooling. Several studies have provided valuable contributions suggesting several methods, for example, settling in combination with directional solidification [77], filtration combined with oxidation [78], oxidative plasma [70–72] and decarburisation by inert gas purging or under vacuum [14, 78]. Klevan [14] has developed a mathematical model, which describes the kinetics of decarburisation when inert purging is applied. Mechanical removal is, however, not efficient enough to affect the substitution carbon. Stronger methods able to displace the equilibria, such as oxidative plasma and vacuum vaporisation, are believed to be more powerful techniques.

It is worth noting that these types of operation can all be carried out in carbon-lined ladles. The several steps dealing with the removal of dissolved carbon $\underline{C}$ from liquid silicon Si(l), however, has to be carried out in the absence of C and at a highest possible temperature in order to optimise the equilibrium and the kinetic conditions for the reaction:

$$\underline{C} + \tfrac{1}{2}O_2 = CO(g) \tag{5.41}$$

A parallel reaction, which affects the silicon yield, unfortunately, also takes place:

$$Si(l) + \tfrac{1}{2}O_2 = SiO(g) \qquad (5.42) = (5.14)$$

The value of solid solubility of carbon in silicon is approximately 10 ppm(a), corresponding to the homogeneous distribution of substituted carbon atoms on silicon lattice sites. This type of carbon impurity is detected by infrared spectroscopy. Higher concentrations of carbon result in SiC precipitates of different size and morphology. This type of carbon is detected and analysed by combustion methods or secondary ion mass spectroscopy (SIMS).

5.7.2.4 Challenges and achievements

The improved carbothermic reduction of silica, followed by leaching and directional solidification, has the capability to produce silicon with sub-ppm concentration of all metallic impurities. Controlling the non-metallic elements, particularly B, P, C and O, at the level required for solar cell applications remains the major challenge for the metallurgical route. Unfortunately, there is no universal method for reducing simultaneously these critical elements. As a consequence, several refining steps are necessary, with the risk of reducing the silicon yield. Another risk is the recontamination by impurities from the reactants and the lining during handling and treatment of liquid silicon. Extensive studies were done on this route in the early eighties, but efforts since then have been considerably reduced. The Japanese program headed by NEDO is to our knowledge the most accomplished project representing this route (Table 5.10; [71]). On the basis of this project's results, Kawasaki has built a 60-MT pilot plant, but the economical feasibility of the route remains uncertain. In the NEDO process, metallurgical grade silicon is purified through

1. melting of silicon by electron beam and evaporation under vacuum
2. first directional solidification
3. remelting of silicon and gas treatment ($O_2 + H_2O$) assisted by a plasma torch
4. second directional solidification.

Table 5.10 Solar grade as obtained through upgrading metallurgical grade silicon by the NEDO method according to [71]

B [ppm(w)]	P [ppm(w)]	Al [ppm(w)]	Fe [ppm(w)]	Ti [ppm(w)]	O [ppm(w)]	C [ppm(w)]	Resistivity [ohm cm]	Lifetime [μs]
0.04–0.10	0.03–0.14	<0.01	<0.05	<0.01	<6	<5	0.8–1.2	>7.7

5.7.3 Simplification of the Polysilicon Process

The Union Carbide and Ethyl Corporation polysilicon processes (Sections 5.4.2 and 5.4.3) resulted from attempts to make the Siemens process more economical. Project goals were to meet a price target of 10 US$/kg set in 1975 by the US Department of Energy (DoE). The Siemens process was fully developed around 1960, but polysilicon for electronic devices and other purposes had been produced since the forties. The Siemens process was the first design of a rational industrial process, which gained international recognition through rapid and broad licensing. At least a dozen processes had been developed prior to the Siemens process and coexisted with the Siemens process until the beginning of the seventies. Purity and high production rates based on semiconductor industry

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 934

demand were the primary criteria for the process design. Capital expenditure and energy consumption were considered secondary criteria. While large production quantities of ultra-pure polysilicon were available at acceptable prices for semiconductors, this price was too expensive for the development of low-cost solar systems. For solar cells, a process with reduced costs, reduced energy consumption and increased production rates, with purity levels not as important, is needed. Since several processes had been developed in the past, a review and re-evaluation of the silicon chemistry of these processes was suggeste (see Tables 5.11 and 5.12).

Table 5.11 Historical processes to manufacture polysilicon (according to Strategies Unlimited [79])

Companies	Volatile silicon source	Reduction agent	Reactor type
Dupont	$SiCl_4$	Zinc	Inside-quartz tube
Bell lab.	$SiCl_4$	Hydrogen	Ta-filament
Union Carbide	$SiHCl_3$	Hydrogen	Inside-quartz tube
Int'l Telephone	$SiCl_4$	Sodium hydride NaH	Ta-filament
Mallinckrodt	SiI_4	Hydrogen	Inside-quartz tube
Transitron	SiH_4	Decomposition	Inside-quartz tube
Texas Instruments	$SiCl_4$	Hydrogen	Outside-quartz tube
Foot Mineral	SiI_4	Decomposition	Si-filament
Chisso	$SiCl_4$	Hydrogen	Inside-quartz tube
Siemens	$SiHCl_3$	Hydrogen	Si-filament
Komatsu	SiH_4SiHCl_3	Decomposition	Si-filament
Motorola	$SiHCl_3$	Hydrogen	W-filament
Phoenix Materials	$SiCl_4$	Hydrogen	Si-filament
Texas Instruments	$SiHCl_3$	Hydrogen	Fluidised bed spheres

Table 5.12 Polysilicon research projects stimulated by the objectives of low-cost solar cells programs after 1975

Companies/groups	Volatile silicon source	Reduction agent	Reactor type
Aerochem Res. Lab.	$SiCl_4$	Sodium	Free space
Eagle Picher/General Atomic/Allied	SiH_4	Decomposition	Fluidised bed
Battelle	$SiCl_4$	Zink	Fluidised bed
Hemlock	SiH_2Cl_2	Hydrogen	Si-filament/rod
Union Carbide	SiH_4	Decomposition	Free space
Union Carbide	SiH_4	Decomposition	Fluidised bed
Ethyl Corp	SiH_4	Decomposition	Fluidised bed
Motorola	Si_nF_{2n+2}	Decomposition	
NEDO	$SiHCl_3$	Hydrogen	Fluidised bed
Rhône–Poulenc	SiH_4	Decomposition	
Schumacher	$SiHBr_3$	Hydrogen	Fluidised bed
SRI International	SiF_4	Sodium	Solid separation
Westinghouse	$SiCl_4$	Sodium	Free space
Bayer	$SiCl_4$	Aluminium	Melt
Bayer	SiH_4	Decomposition	Fluidised bed
Wacker	$SiHCl_3$	Hydrogen	Fluidised bed

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 935

To make the process cost effective, the economics of both the production of the volatile silicon compound and its decomposition or reduction to elemental silicon had to be reviewed. Compared to the historical processes, the research projects did not discover radically new silicon compounds. Silicon halides $SiX_{4-n}H_n$ could apparently not be avoided. However, innovative methods were discovered to produce these silicon compounds. Metallurgical grade silicon remained in most cases the starting point, but silicon fluoride, obtained as co-products from phosphates leaching (fertilisers) and direct chlorination of natural silica to produce tetrachlorosilane, were envisaged as serious challenges. The research abandoned completely the concept of heated filament or seed rod that was clearly perceived as too expensive. Texas Instruments had demonstrated the benefit of the fluidised bed technology, that is,

- larger throughput
- lower energy consumption
- continuous operation
- lower capital expenditure.

Although the quality was not acceptable for making microelectronics devices, this was not a drawback for the development of the solar cell market.

In addition to the fluidised bed, free space reactor implying spontaneous seedless formation of solid silicon particles through homogeneous decomposition of silane has also been envisaged among others by Union Carbide [38].

None of these projects have resulted yet in establishing a new polysilicon route devoted to solar cells and decoupled from the electronic feedstock. The research has to a large extent developed and strengthened the classical polysilicon business by developing new products and processes to produce semiconductor grade silicon (Union Carbide, Ethyl Corporation). However, it is believed that for quality reasons, a significant portion of the Ethyl Corporation process production is still to find outlet in the solar market, making this product to some extent a prime solar grade polysilicon.

The other polysilicon producers have declared that by relaxing the production procedures and quality control they could, when excess capacity is available, produce specifically a solar grade polysilicon at reduced cost. It seems reasonable to expect a price reduction of approximately 35%. The same producers have underlined that specific investments to produce solar grade silicon by this route should not be expected because of too low profitability [80, 81].

At the time of writing, three projects aimed at developing a solar polysilicon route were in the process of being developed:

1. The German chemical company Bayer AG has for several years re-explored the decomposition of silane in fluidised bed reactor. The method combines the cost-effective synthesis of silane as in the Union Carbide process and the fluidised bed reactor as in the Ethyl Corporation process. The company has published the results of its basic assumption and made positive announcements [82–84]. Meantime, Bayer AG has withdrawn from the solar wafer business, which was the main driver for this feedstock development. To date it is therefore uncertain what the future of this project will be.

2. Another German company Wacker Chemie, which is one of the leading suppliers of electronic grade polysilicon and a great contributor to the development of materials to photovoltaics in the past, has decided to build a pilot plant to develop solar grade polysilicon by the reduction of trichlorosilane in a fluidised bed reactor. Under industrial conditions, each fluidised bed reactor should produce more than 500 MT/year and should be economical enough to meet the cost target of the solar cell industry.
3. The Japanese company Tokuyama Corporation, also a leading polysilicon manufacturer, has announced at a recent symposium its intention to develop a new deposition process using trichlorosilane in order to specifically produce solar grade silicon [85]. More technical details had not been presented at the time of writing this chapter. The reader is invited to consult later publications.

5.7.4 Other Methods

Besides the large avenues described above, one finds some other routes. The following processes are worth noting.

1. The aluminothermic reduction of silica

$$3SiO_2 + 4Al = 3Si + 2Al_2O_3 \tag{5.43}$$

explored by Wacker [39].

2. The electrolytic transfer of silicon from an anode made partly of metallurgical grade silicon alloyed with Cu in (Cu_3Si-Si) to a graphite cathode through a liquid electrolyte made of KF:LiF:SiK_2F_6 as described by Olson *et al.* [86–88].

$$\text{Si(metallurgical grade)} \rightarrow \text{Si}(4+) + 4\text{e}(-) \rightarrow \text{Si(refined)} \tag{5.44}$$

Although revisited by several R&D groups, these methods do not seem to have been the subject of extensive new programmes since the eighties.

5.8 CONCLUSIONS

At the time this handbook was written, silicon is the dominant PV material with more than 98% of cells and modules based on it. Crystalline technologies account for nearly 90%. We do not see new materials displacing silicon in the foreseeable future. The crystalline technologies require significant consumption of pure silicon. Purity requirements have not been exactly defined with the industry selecting silicon raw materials among the residuals from single-crystal pulling and second-grade polysilicon. At present, these sources are not sufficient to cover the rapidly growing demand from the solar industry. Polysilicon produced specifically for solar applications and prime-grade polysilicon are extensively used. This has the negative consequence of increasing the cost of solar cells and hence of slowing the market deployment of solar electrical energy. Research on a new dedicated route to solar grade silicon decoupled from the semiconductor business cycles has been going on for almost 30 years. Two directions still retain most of the attention of scientists and technologists: the further development of pyrometallurgical treatments of liquid silicon

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 937

and the simplification of the polysilicon processes. The crystallisation processes applied to silicon to produce the PV devices play a role in purifying and shaping silicon and should affect to a larger extent the development and the final definition of solar grade silicon. Quality and cost requirements have imposed strict limits to solar grade silicon. Capital cost is perceived as a significant part of the overall cost and adds to the barriers that need to be overcome.

REFERENCES

1. Maycock P (Ed), PV Energy Systems, *PV NEWS* **20**(3), 2 (2001).
2. Ikki O (Ed), Resources Total System Co, *PV Activities Japan* **7**(3), 1–6 (2001).
3. Schmela M, *PHOTON Int.* **3**, 32–35 (2001).
4. Knight J, *Trends in Photovoltaic Applications in Selected IEA Countries Between 1992 and 1999*, Report IEA-PVPS T1-08 (2000).
5. Murphy G, Brown R, *Silicon in Bulletin 675 Mineral Facts and Problems*, Bureau of Mines, US Department of Interior, Washington DC (1985).
6. Kerkhove D, *Silicon Production Technology*, Technische Universiteit Delft, Delft, Netherlands (1994).
7. Forwald K, *Dissertation NTNU*, Norway, MI-47 (1997).
8. Dosaj V, *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th Edition, Vol. 21, 1104–1122, Wiley, New York (1997).
9. Moretto H, Schulze M, Wagner G, *Ullmann's Encyclopaedia Ind. Chem.* **A24**, 57–93 (1993).
10. Rong H, *Dissertation NTNU*, Norway, IUK-67 (1992).
11. Sørheim H, *Dissertation NTNU*, Norway, IUK-74 (1994).
12. Watt M, *Added Values of Photovoltaic Power Systems*, Report IEA-PVPS T1-09 (2001).
13. Schei A, Tuset J, Tveit H, *Production of High Silicon Alloys*, Chap. 12, Tapir forlag, Trondheim (1998).
14. Klevan O, *Dissertation NTNU*, Norway, MI-167 (1997).
15. Anglezio J, Servant C, *J. Mater. Res.* **5**, 1894–1899 (1990).
16. Forwald K, Schüssler G, *Proc. Silicon for the Chemical Industry I*, 39–46 (Norway, 1992).
17. Nygaard L, Brekken H, *Proc. Silicon for the Chemical Industry II*, 61–67 (Norway, 1994).
18. Brekken H, Nygaard L, Andresen B, *Proc. Silicon for Chemical Industry III*, 33–45 (Norway, 1996).
19. Pachaly B, *Proc. Silicon for Chemical Industry II*, 55–60 (Norway, 1994).
20. Forwald K, Soerli Oe, Schüssler G, EP 0 372 918 B1 (1989).
21. Schulze M, Licht E, *Proc. Silicon for the Chemical Industry I*, 131 (Norway, 1992).
22. Boardwine C *et al*., Progress in Organic Chemistry, *Proc. of Int. Symp. on Organosilicon Chemistry X (1993)*, 555–569, Gordon & Breach Science Publishers, Amsterdam (1995).
23. Schei A, Tuset J, Tveit H, *Production of High Silicon Alloys*, Tapir forlag, Trondheim (1998).
24. Kobayashi S, in Hull R (Ed), *Properties of Crystalline Silicon*, 6–22, University of Virginia, USA (1999).
25. Trumbore F, *Bell Syst. Tech. J. (USA)* **39**, 205 (1960).
26. Burton J, Kolb E, Slichter W, Struthers J, *J. Chem. Phys.* **21**, 1991–1996 (1953).
27. Kodera H, *Japan J. Appl. Phys.* **2**, 212–219 (1963)
28. Graff K, *Metal Impurities in Silicon-Device Fabrication*, 29, Springer, Berlin (2000).
29. Istratov A, Hieslmair H, Weber E, *Appl. Phys.* **A70**, 489–534 (2000).
30. Davis Jr. J *et al*., *IEEE Trans. Electron Devices* **Ed-27**, 677–687 (1980).
31. Kittler M, Lärz J, Seifert W, *Appl. Phys. Lett.* **58**, 911–913 (1991).
32. Itoh Y, Nozaki T, *Japan J. Appl. Phys.* **24**, 279–284 (1985).
33. Weber E, *J. Appl. Phys.* **A30**, 1–22 (1983).

34. Fischler S, *J. Appl. Phys.* **33**, 1615 (1962).
35. Jaccodine R, Pearce C, in Bullis W, Kimerling L (Eds), *Defects in Silicon*, 115–119, The Electrochemical Society, Pennington, NJ (1983).
36. Mikkelsen Jr. J, *Mater. Res. Soc. Symp. Proc. (USA)* **59**, 19 (1998).
37. Hocine S, Mathiot D, *Mater. Sci. Forum* **38-41**, 725 (1989).
38. Lutwack G, *Proc. of the Flat-Plate Solar Array Project Workshop on Low-Cost Polysilicon for Terrestrial Photovoltaic Solar-Cell Applications*, Contract DOE/JPL-1012-122, JPL Publication 86-11 (1986).
39. Dietl J, *Silicon for Photovoltaics*, Vol. 2, 285–352, North Holland, Amsterdam (1987).
40. Lanier F, Ang T, *Photovoltaic Engineering Handbook*, 3–17, Adam Hilger, Bristol (1990).
41. Dawless R, US Patent 4,246,249 (1981).
42. Smith F, Dawless R, *Electrochem. Soc.* **81-2**, 1147 (1981).
43. Kotval P, Strock H, US Patent 4,195,067 (1980).
44. Hanoka J, Strock H, Kotval P, *J. Appl. Phys.* **52**, 5829–5832 (1981).
45. Hunt L, Dosaj V, Final Report, Contract DOE/JPL 954 559-78/7 (1979).
46. Hunt L, Dismukes J, Amick J, *Proc. Symp. Materials and Processing Technologies for Photovoltaics*, 106, The Electrochemical Society, Pennington, NJ (1983).
47. Amick J *et al*., *Proc. Symp. Materials and Processing Technologies for Photovoltaics*, 67, The Electrochemical Society, Pennington, NJ (1983).
48. Amick J *et al*., *Proc. 5th EC Photovoltaic Solar Energy Conf.*, 336, Kovouri, Athens (1983).
49. Aulich H, Eisenrith K, Urbach H, Grabmaier J, *Proc. 3rd Symp. Materials and Processing Technologies for Photovoltaics*, 177 The Electrochemical Society. Pennington, NJ (1982).
50. Aulich H *et al*., *Proc. 4th EC Photovoltaic Solar Energy Conf.*, 868 (Stresa, 1982).
51. Aulich H, *Proc. 5th EC Photovoltaic Solar Energy Conf.*, 936 Kovouri, Athens (1983).
52. Yoshiyagawa M *et al*., "Production of SOG-Si by Carbothermic Reduction of High Purity Silica", Presented at *Silicon for Solar Cells Workshop* (Schliersee, 1981).
53. Aas H, Kolflaath J, US Patent 3,809,548 (1974).
54. Aas H, *The Met. Soc. AIME, TMS Paper Selection* **A71-47**, 651–667 (1971).
55. Halvorsen G, US Patent 4 539 194 (1985).
56. Ceccaroli B, Friestad K, Norwegian Patent Application WO 01/42 136 (2000).
57. Pizzini S, *Sol. Energy Mater.* **6**, 253 (1982).
58. Dietl J, *Sol. Cells* **10**, 145 (1983).
59. Theuerer H, *J. Met.* **8**, 1316 (1956).
60. Khattak C, Schmid F, Hunt L, *Proc. of the Symp. on Electronic Properties of Polycrystalline or Impure Semiconductors and Novel Silicon Growth Methods*, 223–232, Saint Louis, MO, The Electrochemical Society, Pennington, NJ (1980).
61. Khattak C, Schmid F, *Proc. Symp. Materials and Processing Technologies for Photovoltaics*, 478–489, The Electrochemical Society. Pennington, NJ (1983).
62. Khattak C, Schmid F, *Silicon Processing for Photovoltaics II*, 153–183, Elsevier Science Publishers B.V., Amsterdam (1987).
63. Schmid F, Khattak C, US Patent 5,972,107 (1999).
64. Amouroux J, Morvan D, *High Temp. Chem. Processes* **1**, 537–560 (1992).
65. Cazard-Juvernat I, Bartagnon O, Erin J, *High Temp. Chem. Processes* **3**, 459–466 (1994).
66. Combes R, Morvan D, Picard G, Amouroux J, *J. Phys. III France* **3**, 921–943 (1993).
67. Erin J, Morvan D, Amouroux J, *J. Phys. III France* **5**, 585–604 (1995).
68. Suzuki K, Sakaguchi K, Takano K, Sano N, *J. Jpn. Met.* **54**, 168–172 (1990).
69. Suzuki K, Kumagai T, Sano N, *ISJN Int.* **32**, 630–634 (1992).
70. Baba H *et al*., *Proc. 13th Euro. Conf. Photovoltaic Solar Energy Conversion*, 390–394 (Nice, 1995)
71. Nakamura N *et al*., *Proc. 2nd World Conf. on Photovoltaic Solar Energy Conversion* (Vienna, 1998).
72. Nakamura N *et al*., EP Patent 0 855 367 A1 (1998).

73. Schei A, US Patent 5,788,945 (1998).
74. Suzuki K, Sakaguchi K, Nakagiri T, Sano N, *J. Jpn. Inst. Met.* **54**, 61 (1990).
75. Ikeda T, Maeda M, *ISIJ Int.* **32**, 635–642 (1992).
76. Miki T, Morita K, Sano N, *Met. Mater. Trans.* **27B** 937–941 (1996).
77. Aulich H, Schulze F, Urbach H, Lerchenberger A, *Proc. of the Flat-Plate Solar Array Project Workshop on Low-Cost Polysilicon for Terrestrial Photovoltaic Solar-Cell Applications*, Contract DOE/JPL-1012-122, JPL Publication 86-11, 267–278 (1986).
78. Sakaguchi K, Maeda M, *Met. Trans. B* **23B**, 423–427.
79. *Silicon Industry Vol. 2 – Technology Assessment*, Strategies Unlimited, Mountain View, CA 94040, USA (1983)
80. Ossenbrink H, Bishop J, Aulich H (Eds), *Dependable and Economic Silicon Materials Supply for Solar Cell Production*, Report on Workshop, JRC Ispra (VA) Italy, 7th December 1998 Special Publication No. I.99.56 (1999).
81. Maurits J, *Proc. of the 8th Workshop on Crystalline-Silicon Solar Cell Materials and Processes*, 10–17 (1998).
82. Block H, Wagner G, *Proc. of the 16th EPSEC*, 1059–1063 (Glasgow, 2000).
83. Block H, Wagner G, *Proc. of Silicon for the Chemical Industry V*, 271–280 (Norway, 2000).
84. Tilg A, Mleczko L, *Proc. of Silicon for the Chemical Industry V*, 281–295 (Norway, 2000).
85. EPIA Round Table on Solar Grade Silicon Feedstock: The Global Perspective, *17*th *European Solar Energy Conf. and Exhib.* (Munich, Germany, 2001).
86. Olson J, Carleton K, Kibbler A, *Proc. 16th IEEE Photovoltaic Specialists Conf.*, 123–127 (1982).
87. Olson J, Carleton K, *J. Electrochem. Soc.* **128**, 2698 (1981).
88. Carleton K, Olson J, Kibbler A, *J. Electrochem. Soc.* **130**, 782–786 (1983).
89. Nozaki T, Yatsurgui Y, Akiyama N, Endo Y, Makida Y, Behaviour of Light Impurity Elements in the Production of Semiconductor Silicon, *J. Radioanal. Chem.* **19**, 109–128 (1974).
90. ASM Handbook, Vol. 8, *Metallography, Structures and Phase Diagrams* 8th Edition, ASM International, Metals Park, Ni-Si Phase Diagram, p. 325.
91. Bean A, Newmon R, *J. Phys. Chem. Solids* **32**, 1211–1219 (1971).

6

Bulk Crystal Growth and Wafering for PV

W. Koch[1], A. L. Endrös[2], D. Franke[3], C. Häßler[1], J. P. Kalejs[4] and H. J. Möller[5]

[1]*Bayer AG, Krefeld, Germany,* [2]*Siemens and Shell Solar GmbH, Munich, Germany,* [3]*Access e.V., Aachen, Germany,* [4]*RWE Schott Solar, Massachusetts, USA,* [5]*TU Bergakademie Freiberg, Freiberg, Germany*

6.1 INTRODUCTION

The workhorse of the photovoltaics industry is silicon. More than four out of five amperes of solar-module produced current come from crystalline silicon modules. Of this amount, over 50% is produced by a steadily increasing share from multicrystalline silicon material.

Silicon solar cells first were made about 50 years ago from Czochralski (Cz)-pulled monocrystals with technology adapted from the microelectronics industry. Subsequently, world-record cell efficiencies have been achieved at a very high cost on a laboratory scale with Float Zone (FZ) monocrystals.

Cost pressures have forced the development of multicrystalline material solidification processes for production of very large silicon ingots (blocks), which have reached typical sizes of 250 kg in 2000. A good theoretical understanding of growth processes, together with numerical simulations of the entire process down to a microstructural defect level in the crystal today has resulted in the economical production of high-quality material. Sawing of silicon crystal into the thin wafers required for the best performance in solar cells wastes about 50% of expensive, pure silicon feedstock and is very costly. This has led to the development of several crystalline silicon foil, that is, ribbon production processes, and these are now in various stages of R&D and commercialization.

Handbook of Photovoltaic Science and Engineering. Edited by A. Luque and S. Hegedus
© 2003 John Wiley & Sons, Ltd ISBN: 0-471-49196-9

6.2 BULK MONOCRYSTALLINE MATERIAL

The dominant issue of the photovoltaic industry is to fabricate solar cells in large volumes that are both highly efficient and cost-effective. An overall industrial goal is to significantly lower the costs per watt. The dominant absorber material used today for the majority of the commercially produced solar cells is the Czochralski-grown crystalline silicon (c-Si) in monocrystalline and block-cast material in multicrystalline form (mc-Si). Up to now, a lot of effort has been undertaken to increase the electrical efficiency of solar cells reproducibly towards and even above 20% [1], whereas much higher efficiencies were indeed claimed [2] but were most probably never reached [3]. Unfortunately, efficiency improvements are often reached only with the help of cost-intensive process steps so that most steps cannot be directly implemented into industrial products but have to be reengineered for low-enough cost. Hence, there still remains a significant efficiency gap between monocrystalline laboratory cells with efficiencies up to 24% and cost-effective, commercial Cz solar cells that are presently produced and sold in high volume at approximately 14 to 17% efficiency.

While some years ago the cost of a module was driven almost equally by the cost of the wafer (33%), the cell process (33%) and the module making (33%), this well-known ratio has changed significantly – both for C-Si and mc-Si. Today, in most products the wafer attributes to sometimes more than 50% (!) of the module cost, whereas the cell process and the module process attribute to the rest with similar portions of ∼25%. The main reason for this is on one hand a steady cost reduction in the cell and module processes and on the other hand a significant increase in feedstock price together with an almost unchanged wafer thickness of 250 to 350 μm in production.

One way to meet today's demand of lower wafer cost is to (1) reduce the cost of crystal growth by improvement of productivity and material consumption at constant wafer quality, (2) reduce the cost of the wire-sawing process and (3) cut thinner wafers. While commercial Si solar cells still have a present wafer thickness of 250 to 350 μm for reasons of mechanical stability, a thickness of only 60 to 100 μm has been calculated to be the physical optimum thickness for silicon solar cells [4]. In this thickness regime the maximum theoretical efficiency for c-Si solar cells can be reached. In this optimum thickness regime, however, monocrystalline silicon becomes very fragile so that not only the electronic but also the mechanical properties of the wafer as well as the handling and processing techniques become of utmost importance for a good overall fabrication yield. Besides the mere mechanical wafer stability, manufacturing processes have to be adapted, redesigned or newly developed to avoid bending and breaking of ultra-thin wafers. With reduced wafer thickness, the necessity for surface passivation has to be taken into account. Since this cannot be done without adding to the cost, any added process step has to add adequate efficiency to remain cost-effective. Other important issues to get as much efficiency as possible out of the "valuable" wafer are improvements in anti-reflection (AR) coating, grid shadowing, "blue" response of the emitter and both surface and volume passivation.

The demand for high-quality polysilicon feedstock in the world market grew quickly not only in the microelectronic but also in the photovoltaic industry. In 1980 the worldwide production of single-crystal silicon amounted to approximately 2000 metric tons per year. This number was equivalent to ∼100 000 silicon crystals every year

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 942

by either the Czochralski technique (80%) or the floating-zone technique (20%). The PV industry used both the high-quality tops and tails of the microelectronic crystals for less than 5$/kg to fit the feedstock demand and the depreciated Cz pullers of the "big brother" microelectronic industry. Since the microelectronic industry was and is still driven by continuously increasing ingot diameters, the "small" Cz machines became available for the PV-industry at interesting prices. During the last expansion phase of the microelectronic industry (1993–99), the PV industry had to struggle with a severe shortage of affordable feedstock. To reach the production volume, even pot scrap Si material had to be used. New and demanding techniques were developed in a hurry to separate the Si from the quartz crucible parts and to pre-select and pre-clean this material. Also, fine grain material had to be made usable. However, the annual world requirement for solar quality silicon in 2010 is estimated at 8000 to 10 000 metric tons for PV use, that is, the silicon demand will be roughly as high in volume as today's world production for microelectronics. A dedicated Si feedstock supply only for PV is therefore a necessity.

It can no longer be denied that the growing of silicon crystals has matured from an art into a scientific business. In today's PV-business, some of the bigger companies convert more than ~1 to 2 tons of silicon per day into solar grade Cz crystals and solar cells. Since PV has different main requirements for crystal growing from the microelectronic industry, the focus of machine and process development differ.

6.2.1 Cz Growth of Single-crystal Silicon

Solar cells made out of Czochralski (Cz)-grown crystals and wafers play – together with multicrystalline cells – a dominant part in today's PV industry. This is due to the following advantages.

Cz crystals can be grown from a wide variety of differently shaped and doped feedstock material. This enables the PV industry to buy cost-effective feedstock Si with sufficient quality even on spot markets. Since the feedstock is molten in a crucible, the shape, the grain size and the resistivity of the different feedstock materials can be mixed for the required specifications, although a given feedstock alone would fail. However, special care must be taken to avoid any macroscopic particles (SiO_2, SiC) that would not be dissolved in the melt especially when pot scrap material is used.

The Cz process acts as a purification step with respect to lifetime-limiting elements. The effective distribution coefficients of the most dominant lifetime-limiting metals (Fe, Ni, Au, Ti, Pt, Cr) are in the range of 10^{-5} or below. Together with appropriate gettering steps during cell processing, highly efficient commercial solar cells can even be made out of ingots grown from low-grade pot scrap material. The targeted iron equivalent concentration in the finished cell must be $<10^{12}$ atoms/cm^3 to achieve a minority carrier diffusion length well above ~150 μm.

The Cz process itself acts as a quality control step since proper crystallisation, that is, dislocation-free growth of an ingot, can only take place in a well-defined process window. The homogeneity of a well-grown solar grade Cz ingot for PV application is excellent with respect to the bandwidth of electronic and structural properties, whereas mc-Si block casting produces specifications with higher variances in most parameters. Cell

processes with Cz-Si can therefore use high-efficiency processes with smaller process windows that require well-defined starting material.

Cz technology is mature and cost-effective. Equipment and processes for semi-automated growing of crystals are commercially available so that several Cz pullers can be run by a single operator. Owing to the robust making of the machines, many Cz growers more than 20 years age are still in production.

The ingot can be pulled in a defined <100> orientation. This is a big economic advantage since the solar cell process can use this crystallographic property to homogeneously texture solar cells with a very cost-effective wet chemical etching step. By anisotropic etching, a surface structure with random pyramids is built that couples the incoming light very effectively into the solar cell. This effect together with the usually higher diffusion length of Cz crystals gives rise to the increased efficiency of Cz-Si solar cells compared to similarly processed mc-Si cells.

There exists a high potential for increasing the net pulling speed, that is, the productivity of a puller by a clever design of the hot zone, by sophisticated recharging concepts of Si in the hot crucible and by tuning the growth recipe to the optimum pull speed. Here the PV industry is in the novel position that it can neglect most specifications that are required in the microelectronic industry.

This is possible since the PV specifications are strongly reduced in the number of required parameters in contrast to microelectronic material. A PV specification "simply" focuses on the maximum productivity, a minority-carrier diffusion length of the material that should be exceeding the cell thickness and a shallow p-type doping that leads to a specific resistivity between 0.3 and 10 Ωcm, depending on the fabricated solar cell type.

One of the main disadvantages of Cz crystallisation of silicon is the fact that square cells are best suited to built a highly efficient solar module, whereas Cz ingots have a round cross section. In order to use both the crystal and the module area in the best manner, the ingots are usually cut into a pseudosquare cross section before they are cut into wafers. Additionally, the tops and tails of the ingots cannot be used for wafer production. The cropped and slapped materials, that is, tops and tails and so on, are then fed back into the growth process again.

The equipment and the basic principle for Cz pulling is shown in Figure 6.1. The Cz equipment consists of a vacuum chamber in which feedstock material, that is, polycrystalline silicon pieces or residues from single crystals, is melted in a crucible and a seed crystal is first dipped into the melt. Then the seed is slowly withdrawn vertically to the melt surface whereby the liquid crystallises at the seed. High vacuum conditions can be used as long as the melt weight is small (<1–2 kg), but with larger melts (often more than 30 kg) only pulling under argon inert gas stream is practicable. Owing to the reduced argon consumption, the argon pressure is set in the 5 to 50 mbar regime in the PV industry, whereas in the microelectronic industry, atmospheric pressure is also used.

After the silicon is completely molten, the temperature of the melt is stabilised to achieve the required temperature to lower the seed into the melt. The temperature must be chosen so that the seed is not growing in diameter (melt too cold) or decreasing in diameter (melt too hot). In PV the seed is usually <100>-oriented, is monocrystalline and

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 944



(a) (b)

Figure 6.1 Cz pullers in a PV-production environment (a) and growing Cz crystal (b) in a quartz crucible

is pulled upwards to grow a "crystal neck". Since dislocations propagate on (111) planes that are oblique in an <100>-oriented crystal, the dislocations grow out of the crystal neck after a couple of centimetres so that the rest of the crystal grows dislocation-free even if the growth was started from a dislocated seed. The dislocation-free state of the grown crystal manifests itself in the development of "ridges" on the crystal surface. If this state is achieved, the diameter of the crystal can be enlarged by slower pulling until it reaches the desired value. The transition region from the seed node to the cylindrical part of the crystal has more or less the shape of a cone and is therefore called the "seed cone". This cone can be pulled differently, either flat or steep.

Shortly before the desired diameter is reached, the pulling velocity is raised to the specific value at which the crystal grows with the required diameter. Owing to the seed rotation, the crystal cross section is mostly circular. In general, the pulling velocity during the growth of the cylindrical part is not kept constant, but is reduced towards the bottom end of the crystal. This is mainly caused by the increasing heat radiation from the crucible wall as the melt level sinks. The heat removal of the crystallisation thus becomes more difficult and more time is needed to grow a certain length of the crystal. Standard pull speeds in the body range from 0.5 to 1.2 mm/min. The diameter of the crystal in PV is often chosen between 100 and 150 mm. This is due to the short-circuit current of big solar cells where values of 6 A per cell are exceeded. It is difficult to provide a proper contacting scheme in screen print technology that can handle such high currents in the front contacts without high series-resistance losses. With even larger cell sizes, this effect becomes more problematic.

To complete the crystal growth free of dislocations, the crystal diameter has to be reduced gradually to a small size, whereby an end cone develops. For this purpose, the pulling speed is raised and the crystal diameter is decreased. If the diameter is small enough, the crystal can be separated from the melt without a dislocation forming in the cylindrical part of the crystal. The withdrawal of the crystal from the residual melt can be done with a rather high velocity, but not too fast, because thermal shock would cause plastic deformation called "slip" in the lower part of the crystal. The final crystal length is dependent on the crucible charge and varies between 40 and 150 cm.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 945

Nowadays, the seed crystals used for Cz crystal growth are usually dislocation-free. However, each time the seed crystal is dipped into the melt, dislocations are generated by the temperature shock and by surface tension effects between the melt and the crystal. Normally these dislocations are propagated, or move into the growing crystal, particularly if the crystals have large diameters. The movement of dislocations is affected by cooling strain and faulty crystal growth.

The strain that occurs as a result of different cooling rates between the inner and the outer parts of the crystal is probably the main reason for the dislocation movement in the case of large crystals. At the usual <100>-crystal orientation, no (111) lattice plane, that is, no main glide plane, extends parallel to the crystal axis. All (111) glide planes are oblique to the crystal axis and as a result all dislocations that move only on one glide plane are conducted out of the crystal at some time. For movement in the pulling direction, the dislocations have to move downwards in a zigzag motion using at least two of the four different (111) glide planes. Dislocation-free crystal growth is relatively stable even for large crystal diameters, in spite of the higher cooling strain. This is so because it is difficult for a perfect crystal to generate a first dislocation. However, if a first dislocation has been formed, it can multiply and move into the crystal. In this way, numerous dislocations are generated and spread out into the crystal until the strain becomes too low for further movement of dislocations. Therefore, if a dislocation-free growing crystal is disturbed at one point, the whole cross section and a considerable part of the already-grown good crystal are inundated with backward-moving slip dislocations. The length of the slip-dislocated area is approximately equal to the diameter. Wafers and cells that show these types of dislocations can easily be hydrogen-passivated.

After losing the dislocation-free state, the crystal continues to grow with a high density of dislocations that usually are irregularly arranged. They are partly grown-in into the crystal and partly generated later by strain-induced processes. At high temperatures, climb processes take place that distort the dislocation array even more. This further increases the irregular shape and the distribution of the dislocations. In contrast to simple "slip" dislocations, these "grown-in" dislocations cannot be passivated well by a hydrogen-passivation step later in the solar cell processing. With crystal diameters above 30 mm, the monocrystalline but dislocated growth is not stable and in most cases changes to polycrystalline growth because of the tendency of a Si crystal to form twins in the presence of strain and dislocations. These twins also multiply and form higher-order twins and thus rapidly form a polycrystal. This fine-grained poly-Si material is not usable for solar cell production. Known causes for the generation of the critical first dislocation are either solid particles in the melt that move to the solidification front, gas bubbles that are trapped at the solidification front, impurities that exceed the solubility limit in the melt, vibrations of crystals and melt, thermal shocks or a too high cooling strain.

The growth of the seed cone is the most critical stage in the pulling of the Cz crystal. For productivity reasons very flat seed cones are preferable since the time for the making of the cone is not productive. However, the probability of introducing dislocations in the seed cone is lowest for tapered cones, although this means an increase in the pulling time by 15 to 25% for the same body length and additional loss in usable material. The loss of time and material gets worse for larger ingot diameters.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 946

Owing to the reaction between the liquid Si and the quartz crucible, the crucible is of considerable importance to the growth. The silica of the crucible supplies considerable amounts of oxygen to the melt and, owing to the high purity of the silica, only small amounts of other impurities. However, the crucible tends to dissolve after a long-standing time so that the risk for particles in the melt from the crucible is increased with increased pulling time. The oxygen of the melt adds up to 10^{18} oxygen atoms/cm^3 to the growing crystal, whereas carbon is usually $<10^{17}$/cm^3 and has only little impact on the solar cell performance. Oxygen effects like thermal donors and precipitates can be well controlled in Cz cell processing.

6.2.2 Tri-crystalline Silicon

The mechanical properties of tri-crystalline silicon (tri-Si) allow slicing of ultra-thin wafers with higher mechanical yield than monocrystalline silicon (see [5, 6]). Tri-Si is a crystal compound consisting of three mutually tilted monocrystalline silicon grains [5, 6]. The crystal compound has a (110) surface orientation in all grains in contrast to the standard (100) orientation of wafers for today's solar cell production. While two of the grain boundaries in a tri-crystal are $\sum 3$ classified, that is, first-order twins, the third grain boundary is a $\sum 9$ structure, that is, a second-order twin. All boundaries are perpendicular to the (110) wafer plane and meet at the ingot centre forming a characteristic tri-star. Tri-Si growth is fully compatible with standard Cz monocrystalline growth, but it is faster. Using a tri-crystal seed that contains the complete generic information, tri-Si ingots of 100 to 150 mm diameter and up to 700 mm length are grown in standard commercial crystal pullers using optimised growth parameters for neck, crown and body (see Figure 6.2).

As the pulling axis is parallel to the common <110> orientation of all three grains, the dislocations often cannot be completely eliminated during necking, resulting in ingots that are not dislocation-free. The maximum local dislocation densities are about 105/cm^2 in the ingot top and about 107/cm^2 in the ingot bottom. The dislocations are often arranged



Figure 6.2 Tri-crystal seed (front), tri-crystal with cropped crown (left) and tri-crystal crown (right) with typical facets

in a streetlike geometry, while the dislocation density is significantly reduced in between the streets. However, despite the presence of dislocation from the very beginning (!), ingots of up to 700-mm length can be Cz-grown with multiple recharges at increased pull speed (see Figure 6.3). This is in fundamental contrast to the traditional c-Si growth with (100) orientation where dislocations must be completely avoided because otherwise the crystalline structure is totally lost after only a few centimetres of dislocated growth due to rapid dislocation multiplication. This peculiar structural stability of tri-Si, studied in detail through a geometrical stress model, is attributed to a reduction of cross slip in the tri-Si structure [7, 8].

Regarding electrical properties, ELYMAT mappings [9] show that there is a good spatial correlation between dislocation density and minority-carrier diffusion length: areas of high dislocation density show low diffusion length values. For dislocation lean ingots with 4- to 10-Ωcm resistivity, diffusion length values of more than 1000 μm (see Figure 6.4) can be obtained before light-induced degradation (LID).

The LID of the diffusion length of Cz-Si has recently been investigated extensively by several groups [1, 10, 11]. Although the identification of the defect structure is yet

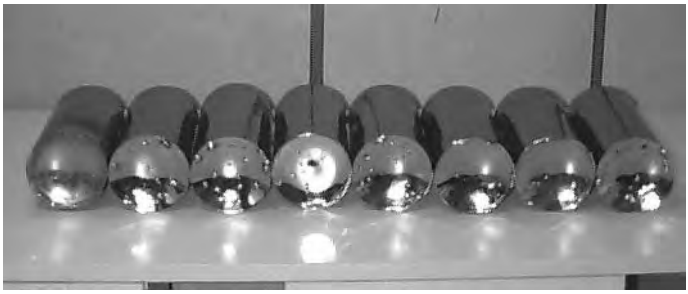


Figure 6.3 Tri-Si ingots (Ø140 mm) from quasi-continuous pulling

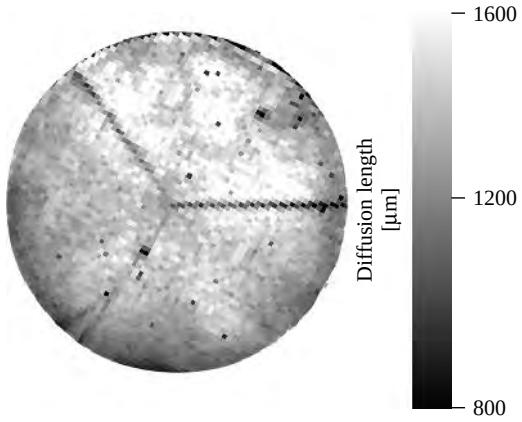


Figure 6.4 Diffusion length map of a 10-Ωcm as-grown wafer cut from the top of a dislocation lean tri-Si ingot. Lateral average diffusion length is 1300 μm

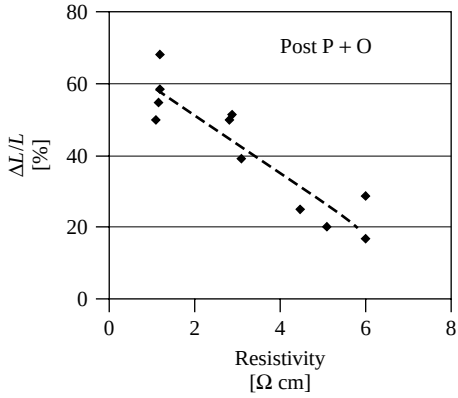


Figure 6.5 Relative degradation of diffusion length versus base resistivity for tri-crystalline silicon, measured after a standard P-diffusion and oxidation process

to be accomplished, a strong correlation of the effect with boron and oxygen is well established. Because tri-Si is grown with boron doping using standard Cz technology, an impact of LID on electrical quality is also present. Figure 6.5 shows a plot of the relative degradation as a function of base resistivity after a standard P-diffusion and oxidation process. Relative degradation values of less than 30% can be reached with 4- to 6-Ωcm material.

Tests on Si crystals fabricated from virgin poly-feedstock material with gallium, indium or aluminium as doping material have been performed to study LID. Best results have been obtained for Indium-doped material, for which the LID effect is reduced to almost below 3% of the diffusion length, which is within the accuracy limit of the ELYMAT measurements. With this material, light stable diffusion length values above 1 mm (!) can be achieved. This material is therefore best suited for high-efficiency cell processes.

In order to test the tri-Si material with standard solar cell processes, test cells are manufactured with a modified "Siemens Solar Boron Back Surface Field" (BSF) process with screen-printed contacts, as described in [12]. The final cell thickness was between 120 and 250 μm. A boron-BSF process was chosen since the BSF is crucial for high efficiency at reduced wafer thickness and beneficial for material with high diffusion length. SiN deposited with a commercial low pressure chemical vapour deposition (LP-CVD) method was applied as an AR coating, that is, no volume passivation can be expected since no molecular hydrogen was present during SiN deposition. In this case no complicated passivation or activation effects of hydrogen in Si must be taken into account [13–16]. In order to eliminate the effect of the different surface orientations and in order to focus on the pure material response, the wafers were *not* textured.

Table 6.1 shows that the efficiencies reach 15.5% in a lab-scale average. This efficiency compares well and slightly exceeds the results of corresponding solar cells from *solar-grade* <100> mono-material that was grown in standard production. The champion tri-Si cell efficiency without surface texturing was 15.9% using this cost-effective process.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 949

Table 6.1 AM1.5 cell efficiencies for non-textured BSF cells on tri-Si and mono-Si substrates. Cell area is 103×103 mm^2 pseudo square

Material (non-textured)	Thickness average [μm]	Average efficiency [%]	V_{OC} [mV]	I_{SC} [mA/cm^2]	FF [%]	Rho [ohm cm]
Tri-Si	140	15.5	615	33.4	75.5	4
Mono-Si	200	15.2	612	32.7	76	1

6.3 BULK MULTICRYSTALLINE SILICON

Multicrystalline silicon besides monocrystalline silicon represents the basis of today's photovoltaic technology. Multicrystalline silicon offers advantages over monocrystalline silicon with respect to manufacturing costs and feedstock tolerance at, however, slightly reduced efficiencies. Another inherent advantage of multicrystalline silicon is the rectangular or square wafer shape yielding a better utilisation of the module area in comparison to the mostly round or pseudosquare monocrystalline wafers. The efficiencies of multicrystalline silicon solar cells are affected by recombination-active impurity atoms and extended defects such as grain boundaries and dislocations. A key issue in achieving high solar cell efficiencies is a perfect temperature profile of both ingot fabrication and solar cell processing in order to control the number and the electrical activity of extended defects. Moreover, the implementation of hydrogen-passivation steps in solar cell processing turned out to be of particular importance for multicrystalline silicon. With the introduction of modern hydrogen-passivation steps by Si_3N_4 layer deposition, the efficiencies of industrial multicrystalline silicon solar cells were boosted to the 14 to 15% efficiency range and consequently market shares were continuously shifted towards multicrystalline silicon as the standard material of photovoltaics.

6.3.1 Ingot Fabrication

Two different fabrication technologies for multicrystalline silicon, the Bridgman and the block-casting process (see illustrations in Figures 6.6 and 6.7) are employed. In both processes the solidification of high-quality multicrystalline silicon ingots with weights of 250 to 300 kg, dimensions of up to 70×70 cm^2 and heights of more than 30 cm have been successfully realised. While the Bridgman technology is a quite commonly used technique, the only two companies mainly employing the casting technology are Kyocera (Japan) and Deutsche Solar GmbH (Germany) [17, 18].

The main difference between both the techniques is that for the melting and crystallisation process only one crucible (Bridgman) is used, whereas for the crystallisation process a second crucible (block casting) is used.

In the case of the Bridgman process, a silicon nitride (Si_3N_4)-coated quartz crucible is usually employed for melting of the silicon raw material and subsequent solidification of the multicrystalline ingot. The Si_3N_4 coating thereby serves as an anti-sticking layer preventing the adhesion of the silicon ingot to the quartz crucible walls that owing to the volume expansion during crystallisation of the silicon material would inevitably lead to a destruction of both the silicon ingot and the crucible. Concerning the block-casting

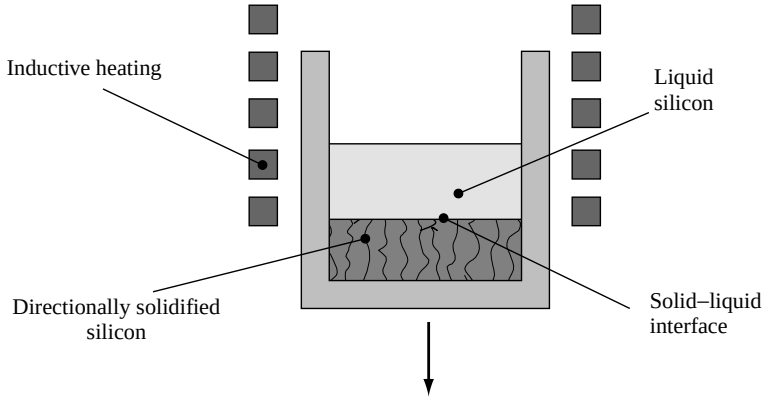


Figure 6.6 Conventional Bridgman technique that still is mainly used for the fabrication of multicrystalline ingots. Both melting and crystallisation of the silicon is performed in a Si_3N_4-coated quartz crucible. Crystallisation is realised by slowly moving downward the liquid silicon-containing crucible out of the inductively heated hot zone of the process chamber

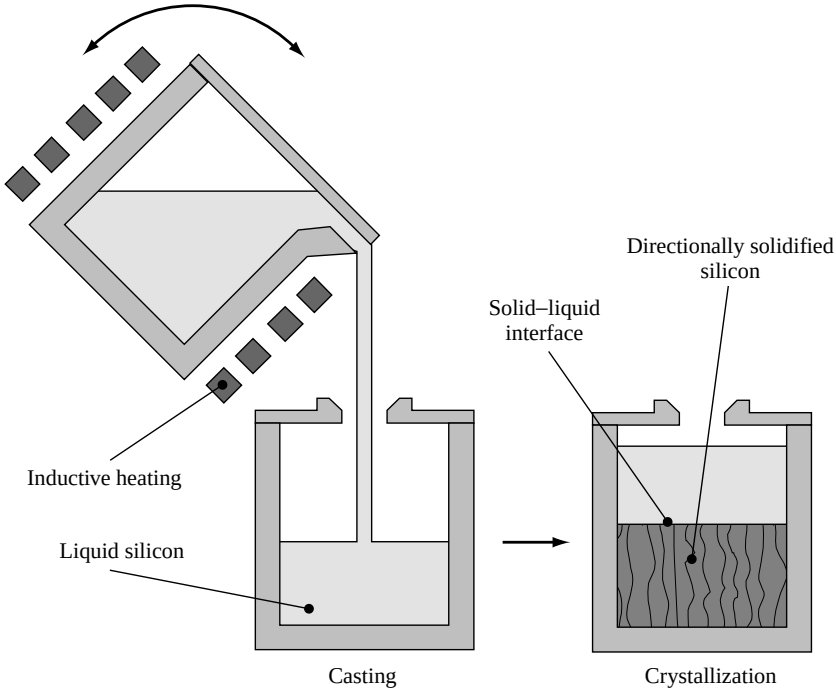


Figure 6.7 Block-casting process for the fabrication of multicrystalline silicon. After melting the silicon in a quartz pot, the silicon is poured into a second quartz crucible with a Si_3N_4 coating. The heating elements of the crystallisation crucible are not shown in the figure. In comparison with the Bridgman technique (see Figure 6.6), shorter crystallisation and cooling times can be realised by employing a more variable heater system

process, the melting is performed in a quartz crucible without a coating, whereas – after pouring the molten silicon into a second crucible – for the crystallisation also a Si_3N_4-coated one is used.

Usually, in both production technologies, crystallisation starts at the bottom of the crucible by lowering the temperature below the melting temperature (1410°C) of silicon. Within the Bridgman process the temperature reduction is achieved by simply descending the liquid silicon-containing crucible out of the hot area of the crystallisation furnace. Contrarily, the temperature control during the block-casting process is achieved by a corresponding adjustment of the heaters, whereas the crucible itself is not moved during solidification.

After solidification starts in the bottom region, the crystallisation front, that is, the liquid–solid interphase, moves in a vertical direction upwards through the crystallisation crucible. This so-called directional solidification results in a columnar crystal growth and consequently adjacent wafers fabricated out of the ingots show nearly identical defect structures (grain boundaries and dislocations).

Common crystallisation speeds used for the Bridgman technology are in a range of about 1 cm/h (corresponding to a weight of approximately 10 kg/h for large ingots). With regard to the increase in crystallisation speed, that is, productivity, mainly cooling of the already crystallised fraction of the ingot has to be taken into account. Too high process speeds cause large thermal gradients within the solidified silicon that may result in cracks or even destruction of the ingot. For the block-casting technology, however, owing to the more versatile and sophisticated heater system, considerably higher crystallisation speeds can be achieved [18].

6.3.2 Doping

Standard multicrystalline silicon is a boron-doped p-type material with a specific electrical resistivity of about 1 Ωcm, which corresponds to a boron concentration of about $2 \times 10^{16}/cm^3$. The specific resistivity is adjusted in a way such that optimal solar cell performance is guaranteed. Naturally, the boron concentration can be varied according to the requirements of specific solar cell processes. Specific resistivities in a range of 0.1 to 5 Ωcm have been used for solar cell fabrication so far. The boron concentration is normally adjusted by adding the equivalent amount of B_2O_3 to the silicon raw material prior to melting of the silicon. Considering alternative doping elements like gallium (*p*-type) or phosphorous (*n*-type) first, the segregation coefficient governing the resistivity decrease with increasing block height has to be considered. With a segregation coefficient of 0.8, boron nearly always is the optimal doping element giving only a small resistivity change over the silicon ingot (see Figure 6.8), whereas gallium and phosphorous with segregation coefficients of 0.008 and 0.35, respectively, are less favourable.

For phosphorous as an *n*-type dopant, additionally the disadvantage of a lower minority charge carrier (i.e. holes) mobility and a more complicated solar cell process using, for example, higher process temperatures for boron instead of phosphorous diffusion is encountered. However, recent results indicate that the activity of extended defects in *n*-type multicrystalline silicon is unexpectedly low, which could render *n*-type material nevertheless an attractive new feedstock source for photovoltaics.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 952

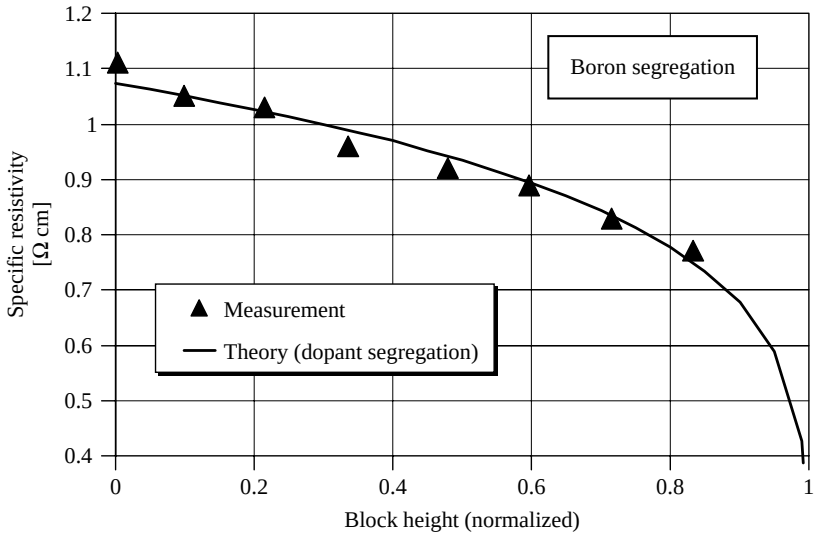


Figure 6.8 Decrease in the specific resistivity of *p*-type multicrystalline silicon due to segregation of the doping element boron

6.3.3 Crystal Defects

The main crystal defects in multicrystalline silicon are grain boundaries and dislocations. Concerning the attainable efficiencies of solar cells, not only the concentration of these defects but also their electrical activity is considered as crucial.

With respect to the grain boundaries and the grain size, basically smaller grains are observed at the beginning of the crystallisation process in the ingot bottom part. With increasing block height, individual grains prevail at the expense of surrounding grains and thus give rise to an increase in the mean grain size. This increase of grain size, however, depends on the crystallisation speed (see Figure 6.9). A higher crystallisation speed also means higher temperature gradients and thus an increased probability for the formation of crystal seeds in the melt that in turn lead to a limitation of the grain size. This is also the reason for faster crystallised block-cast material usually exhibiting smaller grains than conventional Bridgman-type multicrystalline silicon.

On the other hand, the grain sizes achieved with modern block-cast material are still large enough to not degrade solar cell efficiencies provided that the electrical activity of the grain boundaries is low enough. Grain boundaries and dislocations, if electrically charged, effectively attract minority charge carriers and consequently represent highly active recombination centres for photo-generated charge carriers. The electrical activity of grain boundaries and dislocations is determined by their impurity decoration (specifically by transition metals) and strongly increases with higher impurity concentrations.

Moreover, it was discovered that the shape of the crystallisation front during solidification also has considerable influence on the grain boundary activity [19]. Preserving a strictly planar solidification front clearly leads to a reduced grain boundary activity.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 953

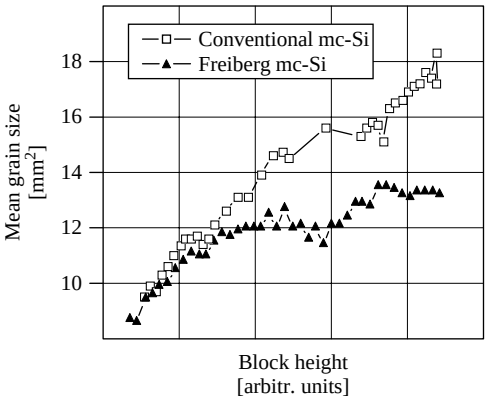


Figure 6.9 Mean grain size as a function of the block height for conventional Bridgman-type multicrystalline silicon (mc-Si) and the faster crystallised block-cast material from the Freiberg production facility of Deutsche Solar GmbH. Reduced crystallisation times lead to slightly lower grain sizes of the Freiberg mc-Si

Because also in modern high-throughput production processes a nearly perfectly planar solidification front is kept throughout the crystallisation process, grain boundaries show only weak electrical activities and therefore are generally considered as less important for solar cell efficiencies.

Crystal dislocations, however, turned out to be the most efficiency-relevant defects in multicrystalline silicon for solar cells. The dislocation density that is experimentally accessible by counting micrometer-small etch pits after appropriate chemical etching steps nearly shows a perfect correlation to the wafer lifetime and diffusion length (Figure 6.10) that are closely linked to solar cell performance. Dislocations are induced and multiplied

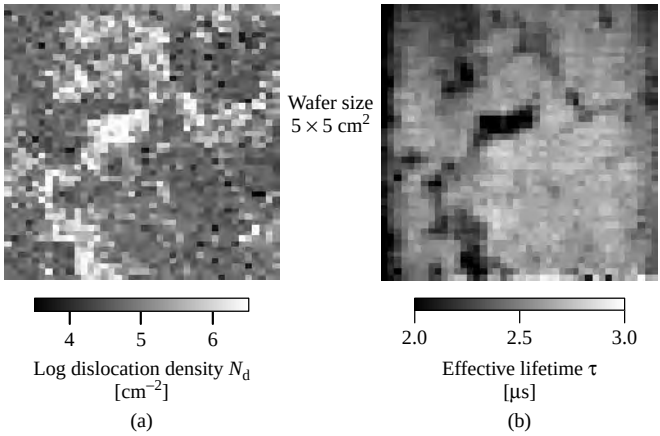


Figure 6.10 Topographies of the dislocation density N_d (a) and the effective lifetime τ_{eff} (b) of a typical multicrystalline silicon wafer showing the excellent correlation of both parameters. The effective lifetime measurements were conducted out without any surface passivation and thus are limited to less than approximately 3 μs by surface recombination processes

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 954

by thermal stress that is originated from temperature non-homogeneities during crystallisation and cooling of the ingot. The reduction of these temperature variations while keeping high process speed is therefore considered one of the most important issues for the further improvement of multicrystalline silicon.

An optimal process scenario for the production of multicrystalline silicon from both the crystal defect and the productivity point of view starts with a small crystallisation speed and minimal temperature gradients in order to secure a low-defect density bottom region of the ingot. After that, crystallisation speed should be largely increased for productivity reasons while keeping the solidification front planar and thermal gradients within the solidified silicon low.

6.3.4 Impurities

Despite boron being the standard dopant and thus an intentionally introduced impurity, even higher impurity concentrations in multicrystalline silicon are observed for both oxygen and carbon.

The interstitial oxygen concentration in multicrystalline silicon is affected by two processes, which are oxygen incorporation via the quartz crucible during melting and oxygen loss through evaporation of SiO, that is, the evaporating gaseous silicon monoxide that is stable at high temperatures only.

Because the segregation coefficient >1, the oxygen content decreases with increasing block height. Typical concentrations of the interstitial oxygen content of Bridgman and block-cast material are given in Table 6.2. Obviously, although the silicon melt never stays in direct contact with the quartz crucible, lower oxygen concentrations with Bridgman-type material compared to silicon from the block-casting process are not feasible. We therefore conclude that there also has to exist an oxygen release from the Si_3N_4 coating (containing some percentage of oxygen) into the silicon melt during the Bridgman process. In addition, we observe a much more rapid decrease of the oxygen concentration with increasing block height for block-cast material, which is attributed to the normally employed lower ambient pressure and enhanced gas exchange during this process.

Table 6.2 Typical concentrations of interstitial oxygen $[O_i]$ for block-cast material from the Freiberg production plant of Deutsche Solar GmbH and for typical material coming from a Bridgman process. For the determination of the oxygen concentration by Fourier Transform Infrared Spectroscopy (FTIR), a conversion factor of $2.45 \times 10^{17}/cm^2$ was used

Ingot position	Interstitial oxygen concentration $[O_i]$ $[10^{17}/cm^3]$	
	Block-casting process	Bridgman process
Bottom	6.5	6
Middle	0.9	3.5
Top	0.5	2

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 955

Although oxygen residing on interstitial lattice sites is not electrically active, recombination-active oxygen complexes such as thermal donors [20–22], new donors [23, 24] and oxygen precipitates may be formed after annealing steps, specifically in the high oxygen concentration bottom part of the ingot (see an example of the donor activity in Figure 6.11).

Specifically, thermal donors turned out to be mainly responsible for a broad low-lifetime region with a width of 4 to 5 cm in the bottom part of Bridgman-type ingots [25]. Owing to the instability of the thermal donors in high-temperature steps during solar cell processing, these low lifetimes, however, does not lead to low efficiencies. It is worth noting that the width of this low-lifetime region in the bottom part of the ingots is largely reduced for material from the block-casting process. The most likely explanation for this is the shorter process times that consequently give less time for the formation of oxygen complexes out of interstitial oxygen atoms.

Similar to metals, oxygen segregation at grain boundaries and dislocations enhances the recombination strength of these extended defects. Oxygen precipitates may also getter metal impurities during crystallisation, which are released afterwards during solar cell processing as highly recombination-active point defects.

Generally, the manifold involvement of oxygen in efficiency-relevant microscopic processes makes the reduction of the oxygen incorporation into multicrystalline silicon one of the most important targets of material improvement efforts.

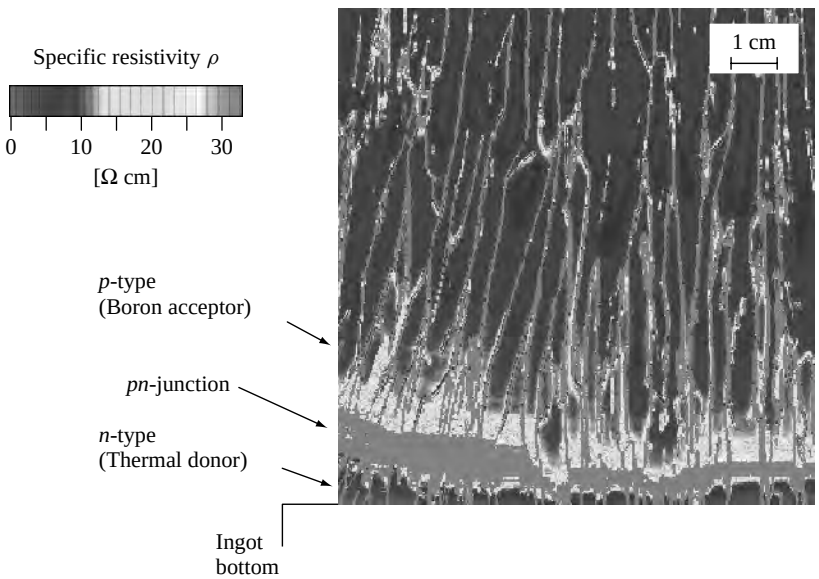


Figure 6.11 High resolution map of the specific resistivity (van-der-Pouw measurement technique) of a vertically cut wafer from a special high resistivity *p*-type multicrystalline silicon test ingot. Owing to the increased oxygen content, the formation of thermal oxygen donors changes the conductivity to *n*-type in the bottom part. The *pn*-junction can be identified by a remarkable increase in the specific resistivity

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 956

Like oxygen, carbon in multicrystalline silicon appears in concentrations considerably higher than those of the boron dopant concentrations. Typical concentrations of substitutional carbon are in the range of $2-6 \times 10^{17}/cm^3$ generally increasing with increasing block height. The incorporation of carbon into the silicon melt takes place via gaseous CO formation inside the crystallisation chamber by SiO chemically reacting with the graphite heaters. The main problem that is caused by an increased carbon concentration is the formation of needle-shaped SiC crystals (often associated with oxygen and nitrogen, see Figure 6.12) within the silicon material. SiC representing an electrically conductive semiconductor material effectively shorts the solar cell *pn*-junction, thereby leading to drastically reduced efficiencies. The problem of SiC formation, however, usually occurs only in the uppermost region of the ingot that is anyway rejected because of segregation of metallic impurities.

Despite oxygen and carbon being present in much higher concentrations, transition metals like iron or titanium are considered as much more important with regard to solar cell efficiencies with the exception of the outer edges (width 5–10 mm) of an ingot where in-diffusion from the Si_3N_4 coating may occur and the top segregation region metal point defects in high-quality multicrystalline silicon are present in concentration levels below the detection limit of Deep Level Transient Spectroscopy (DLTS) measurements,

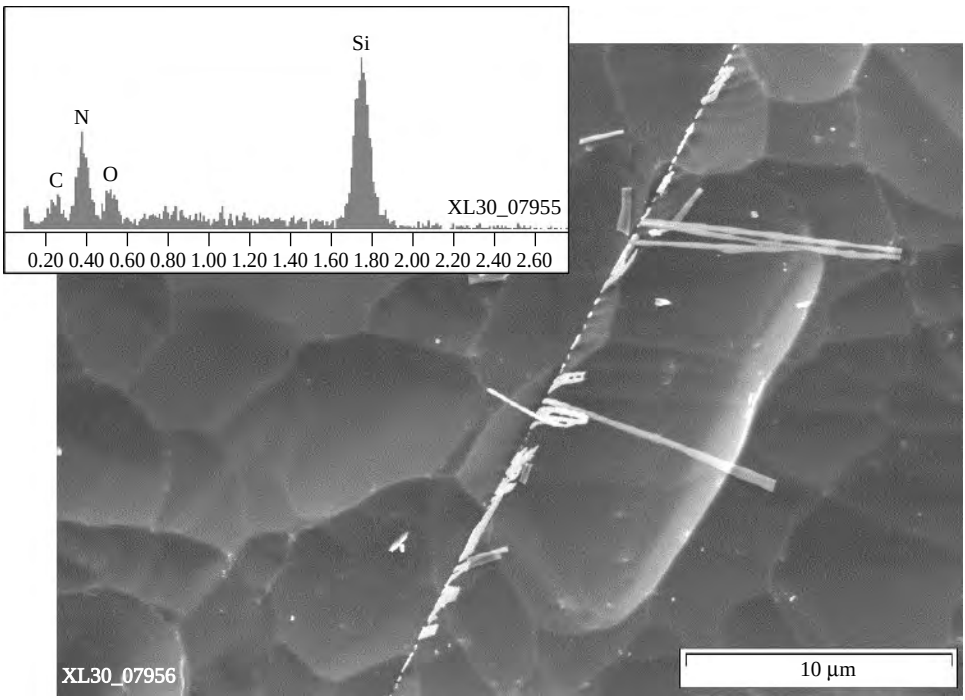


Figure 6.12 SEM (Scanning Electron Microscope) image of a heavily shunted solar cell region. The microscopic investigations reveal needle-shaped structures containing silicon, nitrogen, oxygen and carbon. The shunting mechanism is assumed to be due to electrically conductive SiC that short-circuits the solar cell *pn*-junction

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 957

that is, below approximately $10^{12}/cm^3$. The importance of metal impurities for multicrystalline silicon solar cells is, however, based on metal impurities controlling the activity of extended defects, specifically that of crystal dislocations.

We anticipate that metallic impurities are, for example, responsible for the observed systematic changes of the lifetime of multicrystalline silicon wafers after high-temperature steps in the range of 800 to 1000°C (e.g. the phosphorous diffusion step for fabrication of the solar cell *pn*-junction). The wafer lifetime quite commonly decreases in annealing steps above 900°C, where this decrease is even more significant at enhanced cooling speeds after the anneal. The proposed mechanism behind this lifetime degradation is a release of metal atoms from extended defects like dislocations into the wafer bulk material and a subsequent quenching of the metal atoms as highly recombinative point defects.

Another hint of an extensive interaction between extended defects and metal impurities is given in Figure 6.13 that depicts the theoretical segregation profile of iron in an intentionally contaminated multicrystalline ingot (mean iron concentration: $7.9 \times 10^{17}/cm^3$) as compared to the experimental one. We clearly can state a reduced experimentally determined segregation effectiveness most probably caused by iron segregation into extended defects competing with the segregation process in the liquid silicon phase during crystallisation.

In order to prevent such defect–metal interaction processes leading to enhanced recombination activity, a very effective segregation of metallic impurities into the ingot top region has to be assured. This segregation effectiveness, however, decreases with both increasing crystallisation speed and increasing concentration of extended defects.

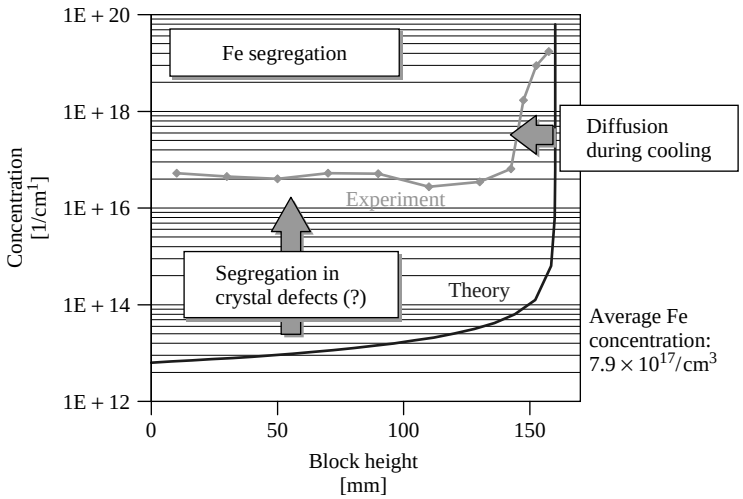


Figure 6.13 Experimentally determined iron concentration of an intentionally contaminated multicrystalline silicon test ingot (mean Fe concentration: $7.9 \times 10^{17}/cm^3$). The experimental data is given as a function of the block height and in comparison to the theoretically expected profile. The much higher than theoretically predicted concentration in the bottom and middle part of the ingot is attributed to a segregation not only into the silicon melt but also into extended crystal defects during solidification

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 958

This in turn verifies the importance of a properly adjusted and controlled crystallisation speed. In order to assure an effective impurity segregation for high-quality multicrystalline silicon, specifically in regions with increased defect densities (e.g. ingot bottom part), solidification at a low crystallisation speed is essential.

6.4 WAFERING

More than 80% of the current solar cell production requires the cutting of large silicon crystals. Multicrystalline ingots grown by the Bridgman or gradient freeze technique now reach cross sections of more than 50×50 cm^2 and weigh over 250 kg; monocrystalline Cz crystals have diameters of up to 20 cm today. While in the last few years the cost of solar cell processing and module fabrication could be reduced considerably, the sawing costs remain high.

Figure 6.14 shows that the sawing costs are a substantial part (29%) of the wafer production cost and thus contribute considerably to the total module cost. Since the sawing of the crystals is connected with high material losses (about 50%), ribbon growth techniques or the thin film technology, which avoid the sawing step, have a high potential for developing cheaper solar cells. However, both technologies still have to overcome serious difficulties and their development will probably take another 5 to 10 years. The present task is therefore to optimise the sawing technique for further cost reduction in mass production.

At the beginning of the PV industry, the available sawing technology of the microelectronic industry was used. The ingots were mainly cut by inner diameter (ID) saws. This technology is, however, relatively slow and not economical for mass production [26]. It was therefore gradually replaced by the multi-wire slicing technology [27]. The advantages are the higher throughput of about 500 to 700 wafers per day and per machine, a smaller kerf loss of about 180 μm and almost no restrictions on the size of the ingots. Currently, wafers between 250 and 350 μm are usually cut, but a wafer thickness down to about 100 μm can be achieved by the technique. Since the technology is relatively new and still under development, most wafer manufacturers have to optimise the sawing process by their own experience. The sawing process depends on several variable parameters as will be described next, which makes it difficult to optimise the process in view of throughput, material losses, reduction of supply materials and wafer quality.

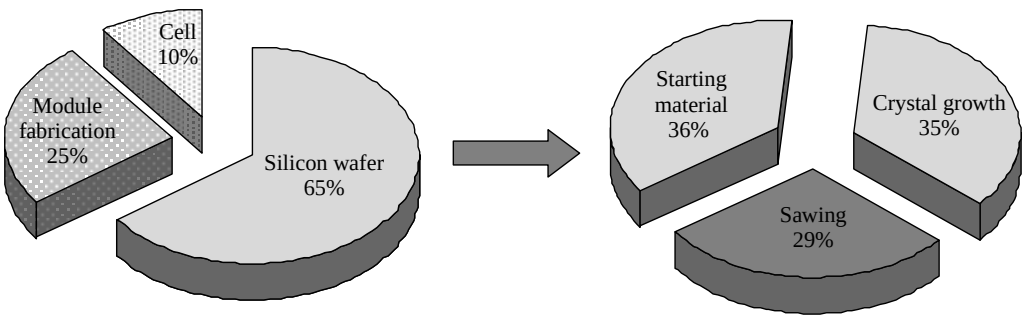


Figure 6.14 Cost distribution for modules and silicon wafers

Basic knowledge about the microscopic details of the sawing process is required in order to slice crystals in a controlled way. In the following section, the principles of the sawing process will be described as far as they are understood today.

6.4.1 Multi-wire Wafering Technique

After crystal growth the silicon ingots are cut in a first step by band saws into columns with a cross section that is determined by the final wafer size. Standard sizes are about 10×10 cm^2, but larger wafers sizes up to 15×15 cm^2 are increasingly used in the solar cell technology. The columns are glued to a substrate holder and placed in a multi-wire saw that slices them into the final wafers. The principle of the multi-wire technology is depicted in Figure 6.15. A single wire is fed from a supply spool through a pulley and tension control unit to the four wire guides that are grooved with a constant pitch. Multiple strands of a wire net (known as a web) are formed by winding the wire on the wire guides through the 500 to 700 parallel grooves. A take-up spool collects the used wire. The wire is pulled by the torque exerted by the main drive and slave as shown in the figure. The tension on the wire is maintained by the feedback control unit at a prescribed value. The silicon column on the holder is pushed against the moving wire web and sliced into hundreds of wafers at the same time. The wire either moves in one direction or oscillates back and forth. Solar cell wafers are mainly cut by a wire that is moving in one direction, whereas wafers for the microelectronic industry are cut by oscillating wires. Cutting in one direction allows higher wire speeds between 5 and 20 m/s, but yields less planar surfaces. Smoother and more even surfaces are obtained by oscillating sawing. Depending on the pulling speed, the wires have a length between 150 and 500 km in order to cut a single column in one run. The wire material is usually stainless steel.

Cutting is achieved by an abrasive slurry, which is supplied through nozzles over the wire web and carried by the wire into the sawing channel. The slurry consists of a

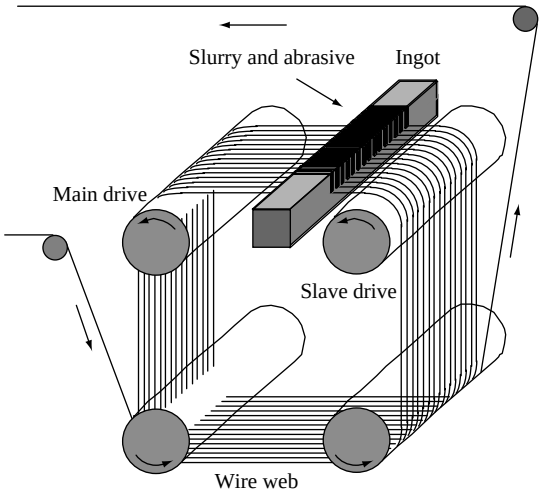


Figure 6.15 Schematic diagram depicting the principle of the multi-wire sawing technique

suspension of hard grinding particles. Today, SiC and diamond are the most commonly used abrasives. Both materials are very expensive and account for 25 to 35% of the total slicing cost. The volume fraction of solid SiC particles can vary between 20 and 60% and the mean grain size between 5 and 30 μm. For polishing smaller grain sizes below 1 μm are used. The main purpose of the slurry is to transport the abrasive particles to the sawing channels and to the crystal surface. It also has to keep the particles apart and must prevent their agglomeration. The entry of the slurry is a result of the interaction between the wire and the highly viscous slurry. Normally, only a small amount of slurry enters the cutting zone. The factors that are important here are the viscosity and the wire speed, but to understand the fluid mechanical problems that are involved a complex physical modelling is required. First attempts of a description have been reported recently [28–31].

Most of the commercial slurries are based on oil, but water-based or water-washable slurries based on ethylene glycol have been tested as well. Oil slurries have several drawbacks. The wafers can stick together and are difficult to separate after sawing. This problem will become even more severe when the wafer thickness will be reduced in the future. The removal of oil from the wafer surfaces requires comprehensive cleaning procedures. Since large quantities of slurry and SiC are used during sawing and recycling is not possible at present, the disposal of these materials has to be considered as well. The disposal causes, however, environmental hazards and is therefore complicated and expensive. On the contrary, water-based slurries or slurries that are very hygroscopic have the problem that hydrogen gas is generated from the interaction of water and silicon, which can cause the hazard of explosion. From the environmental point of view, water-washable slurries may be the choice of the future.

Material is continuously removed through the interaction of the SiC particles below the moving wire and the silicon surface. The abrasive action of the SiC depends on many factors such as wire speed, force between wire and crystal, the solid fraction of SiC in the suspension, the viscosity of the suspension, the size distribution and the shape of the SiC particles. The viscosity of the slurry depends on the temperature and the solid fraction of particles. Since the temperature rises as a result of the cutting process, the suspension has to be cooled and the temperature controlled during sawing. The viscosity of the slurry also changes because of the continuous abrasion of silicon and iron from the wire. This gradually deteriorates the abrasive action and the slurry has to be replaced or mixed with new slurry after some time.

The kerf loss and surface quality are determined by the diameter of the wire, the size distribution of the SiC particles and the transverse vibrations of the wire. The amplitude of vibration is mainly sensitive to the tension of the wire, but it also depends on the damping effect of the slurry. Increasing the tension will reduce the amplitude of vibration, hence the kerf loss [32]. Typical wire diameters are around 180 μm and the mean size of active particles can vary between 5 and 30 μm. This yields kerf losses around 200 to 250 μm per wafer.

The objective of efficient sawing is to slice with a high throughput, with a minimum loss of slurry and silicon and with a high quality of the resulting wafers. Since many parameters can be changed, the optimisation of sawing becomes a difficult task and today it is mainly done by the wafer manufacturers. They are mostly guided by experience. In

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 961

the following section, the main results of investigations are summarised, which describe the current understanding of the microscopic details of the wire sawing and yield some guidelines to optimise the process.

6.4.2 Microscopic Process of Wafering

Figure 6.16 shows schematically a cross section of the wire in the cutting zone. The space between the wire and the crystal surface is filled with slurry and SiC particles. The pressure of the wire on the particles varies along the contact area. The forces are maximal directly below the wire and decrease towards the side faces. Because of the transverse vibrations, the wire also exerts forces sideward, which determines the surface quality of the sliced wafers. The interaction between the abrasive SiC particles and the silicon crystal yields a distinct damage pattern on the surface that can be analysed by microscopic techniques. A typical surface structure as seen under an optical microscope is shown in Figure 6.17. Similar structures are obtained along the entire contact zone, which shows that the abrasive process is the same in all directions.

The surface structure consists of local indentations with a mean diameter of a few micrometers. Such a uniform structure can be explained by the interaction of loose, rolling particles that are randomly indented into the crystal surface until small silicon pieces are chipped away. Since SiC particles are facetted and contain sharp edges and

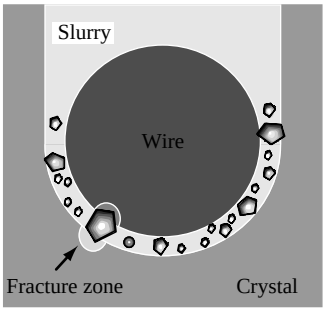
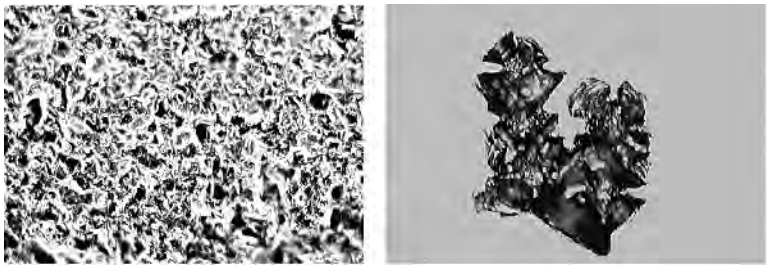


Figure 6.16 Cross section of wire, slurry with abrasive and crystal in the cutting zone



(a) (b)

Figure 6.17 Optical micrograph of the surface of an as-cut silicon wafer (a) and several micro-indentations on a polished silicon surface (b)

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 962

tips, they can exert very high local pressures on the surface. This "rolling grain" model forms the physical basis of the wire sawing process. Similar surface structures also form after lapping semiconductor surfaces with loose abrasive particles.

The individual process of the interaction of a single particle with sharp edges and the surface of a brittle material can be studied by micro-indentation experiments. This is shown in Figure 6.17(b) for a silicon surface. The damage structure of several overlapping micro-indentations with a Vickers diamond indenter resembles the structure of an as-cut wafer. Numerous micro-indentation experiments on monocrystalline silicon have been carried out in the past to investigate the damage structure quantitatively (e.g. [33–37]). The main results are summarised schematically in Figure 6.18 for a "sharp" Vickers indenter with pyramid geometry. Loading by sharp indenters first leads to the generation of a remnant plastic impression in the surface known as the elastic–plastic zone. Recent Raman investigations of this region have shown that under high pressures the silicon lattice transforms into other crystal structures. Several phase changes have been observed directly under the indenter, in particular a metallic high-pressure phase [38, 39]. Under loading at 11.8 GPa an endothermic transformation to metallic silicon (Si II) occurs ($\Delta G = 38$ kJ/mol), which partly transforms back to another high-pressure phase (Si III at 9 GPa, $\Delta G = -8.3$ kJ/mol). In the metallic state the silicon can plastically deform and

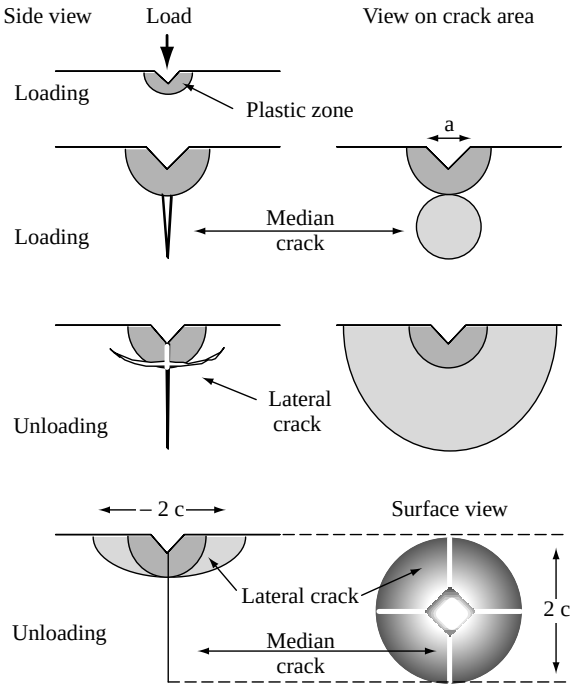


Figure 6.18 Schematic diagram of the development of the crack system below a sharp indenter upon loading and unloading. Dark gray areas indicate the plastic zone below the indenter. The dotted areas are the crack planes of the halfpenny-shaped median crack system. They are viewed from end-on (left side) or perpendicular to the plane (right side). In case radial cracks also occur, they may coalesce with the median crack and form a similar crack pattern

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 963

the material can be removed by processes known for ductile metals. This is, however, a slow but moderate process.

With increasing pressure the material begins to break and cracks are generated parallel to the load axis emanating from the plastic zone. Median cracks are generated beneath the plastic zone, where the tensile stresses are maximal, in the form of full or truncated circles. At a critical size they become unstable and extend towards the surface. In addition, shallow radial cracks may be generated at the edges of the plastic zone. Both radial and median cracks may coalesce to form halfpenny-shaped cracks that are visible at the surface (as shown in Figure 6.19). Upon unloading, residual stresses from the elastic–plastic zone can lead to lateral cracks parallel to the surface. When these lateral cracks reach the surface, material is chipped away. This is the main process for material removal during sawing. Chipping requires a certain minimum load to occur (chipping threshold). Above the limit when material is removed only the median and radial cracks remain. They are finally part of the saw damage.

A quantitative model based on the rolling grain interaction described above has been developed. Results have been compared with experimental investigations of the sawing process on commercial multi-wire saws, allowing for the extraction of useful conclusions. Details can be found in Reference [40].

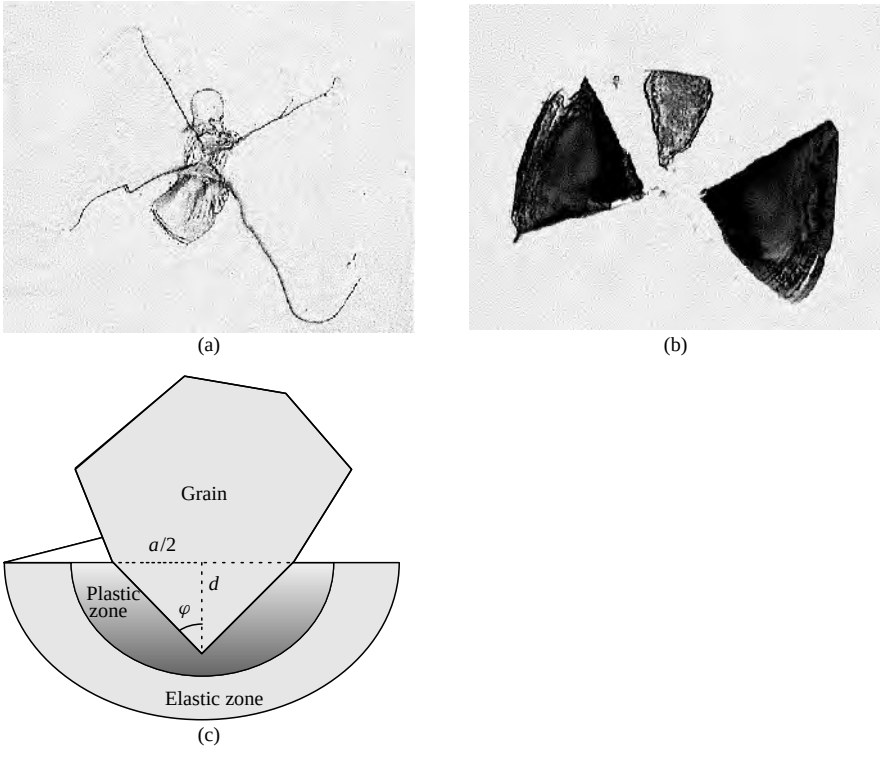


Figure 6.19 (a) Optical micrograph of median and (b) lateral cracks (below the surface) at a Vickers indentation. (c) Schematic representation of the impression of a sharp grain into a surface

6.4.3 Wafer Quality and Saw Damage

Several factors are currently considered to determine the quality of the wafers: fracture behaviour, crack density, thickness variations, surface roughness and cleanness. After sawing the surface of the wafer is damaged from the fracture processes and contaminated with organic and inorganic remnants from the slurry. Therefore, the wafers have to be cleaned and the saw damage removed by etching before a solar cell can be fabricated. In addition, the thickness and surface roughness of the wafer may vary, which may be detrimental for some of the further processing steps. All factors are related to the sawing process. Figure 6.20 shows an example of the topology of an as-cut surface. It consists of thickness variations on different length scales. On the scale of millimetres, one can observe grooves parallel to the direction of the wire. They occur particularly under higher loads and can be caused by a deficit of slurry, mechanical vibrations or inhomogeneities of the material. Mostly a large number of parallel wafers are then affected. Grooves on wafers cannot be removed by etching and thus reduce the quality of a wafer.

On a length scale of about 100 μm, the surface may have a wavy topology that is not detrimental unless sharp steps occur. On the micrometer-length scale the surface shows a certain roughness, which is directly related to the microscopic sawing process as described before. The extent of saw damage, which has to be etched away before solar cell processing, lies typically in the range of 5 to 10 μm.

Saw damage also occurs in the abrasive grains and the wire itself. Although the fracture strength of the SiC particles is higher than that of silicon, the grains eventually lose their sharpness owing to breakage that reduces their sawing performance. To reduce the abrasion of the grains, sawing should be done in a stress range where the load on the individual grains lies above the fracture strength of silicon but below that of SiC.

Typically, the wires have diameters between 150 and 180 μm and a length of about 150 to 500 km. They are made of stainless steel and coated with a brass layer.

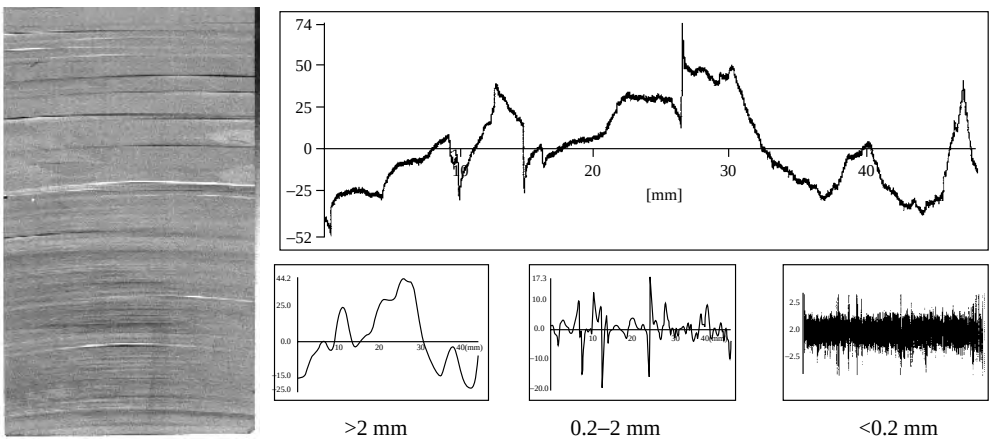


Figure 6.20 Surface structure of a wafer with grooves resulting from uneven cutting. It also shows the bowing of the wire under load during sawing. The surface profile measured by a laser scanner profiler is depicted on the right. Different wavelengths filtered from the profile are shown below

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 965

It is important that the thickness is very uniform over the entire length, because sudden changes in the diameter can lead to fracture of the wire or damage to the wafer surface. The abrasion of the steel wires is also due to the interactions with the grains. Excessive wear can lead to breakage, which is undesirable during sawing because it is very time consuming to build up the wire web inside the machine. Some manufacturers are beginning to develop *in situ* detection systems to control the sawing process and thus prevent the wire breakage.

6.4.4 Cost and Size Considerations

The investigations of the microscopic processes of wire sawing have laid the basis for the selection of the best range of parameters and for further modifications. It allows one to increase the sawing performance, to reduce the consumption of slurry, SiC powder, wire material and etchant, and hence directly the costs of slicing. Furthermore, the quality of the wafers such as roughness, flatness and saw damage of the surfaces can be improved. This is important in view of the development of thinner wafers for solar cells, which will reduce the consumption of expensive silicon. The current sawing technique in production allows the sawing of wafers with thickness down to about 200 μm. The goal is to further reduce the thickness to about 100 to 150 μm in production. Sawing of thinner wafers is possible but at present still at the expense of more breakage. The problem becomes even more severe when the wafer size increases at the same time to 15×15 cm^2 or more. In mass production such a development will only become possible by a careful selection of the parameter range and an *in situ* control of all the factors that determine the slicing process.

6.5 SILICON RIBBON AND FOIL PRODUCTION

Research and development on crystal growth technologies for production of crystalline silicon ribbon have been under way for three decades. Interest in methods of crystalline silicon wafer production was initiated during the oil crises of the mid-1970s. Out of this period arose the first large-scale efforts in R&D to develop low-cost methods of producing substrates for solar cell manufacture. A seminal program was conducted in the US, which was led between 1975 and 1985 by the Jet Propulsion Laboratory (JPL) Flat Plate Array Project [41]. It was the activity in this project in this time period, combined with larger investments from the private sector both in the US and internationally, that developed the seeds of the technology of crystalline silicon ribbon and foil production methods being commercialised today.

The past decade of R&D on crystalline silicon materials has culminated in the expansion of wafer manufacturing at an unprecedented pace. While established methods of production based on Cz growth, directional solidification and ingot casting have flourished, a new generation of ribbon technologies has moved past the R&D stage into large-scale manufacturing and is in competition with these conventional approaches. Ribbon technologies, some of which had entered R&D already in the early 1970s, and have now reached maturity with the start up of manufacturing on a megawatt (MW) scale, include Edge-defined Film-fed Growth (EFG), String Ribbon (STR) and Silicon Film^{TM} (SF).

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 966

Dendritic Web (WEB) production and the Ribbon Growth on Substrate (RGS) technique are moving to pilot demonstration phases. A summary of the changes in the status of leading ribbon/foil technologies over the past decade and projections for manufacturing capacities are given in Table 6.3. It is anticipated that the ribbon production will contribute in excess of 30 MW of wafers to world solar energy markets by the end of 2001.

Development has not been continuous for most of the methods listed above. The R&D has been interrupted and then restarted in several cases when the technological status changed to generate new opportunities for cost-effective production. EFG development has the longest continuous history. After initial technology development on EFG started at Tyco Laboratories in 1971, it was subsequently augmented with funding from Mobil Oil, starting in 1974. In the time span from 1971 to the present, pilot lines using five different variations of the EFG process have been evaluated, starting with single ribbons in 1971 to the octagonal crystal tube now being commercialised. Ownership transferred to ASE Americas in 1994, at which time the transition to manufacturing was initiated. After periods of decreased activity, WEB, STR and RGS have all been strengthened with R&D in the past several years subsequent to being revitalised by new owners. WEB development was initiated with funding from Westinghouse in the 1970s, but now is being carried out by EBARA Solar. STR technology underwent an R&D phase in the early 1980s under the name of Edge-Stabilised Ribbon (ESR) and Edge-Supported Pulling (ESP) at the National Renewable Energy Laboratory and at Arthur D. Little, respectively, before being taken up in 1994 by Evergreen Solar. RGS development was initiated at Bayer, but is currently continuing with ECN of the Netherlands. If successful, a future commercialisation is anticipated by Deutsche Solar in Germany and Sunergy in the Netherlands.

Ribbon and foil technologies must meet the challenges of the photovoltaic marketplace and overcome a number of existing technical barriers if they are to continue to expand manufacturing and to position themselves to remain competitive in the next decade. Challenges to be met are productivity increases on a per furnace basis to drive down labour and overhead (capital) costs, improved mechanical and electronic quality of ribbon wafers together with the development of low-cost solar cell designs that will raise efficiencies to 18 to 20% and reduction of wafer thickness while maintaining high yields in order to reduce demand on silicon feedstock. Achievement of these goals in the next decade can lead to cost decreases, which will drive additional volume expansion for

Table 6.3 Historic record on R&D and manufacturing status of leading ribbon/foil technologies of the past decade

Wafer process/ year started	1990 status level		2000	2001	Schematic
WEB/1967	R&D	<0.1 MW	R&D <0.2 MW	Pilot −0–1 MW	Figure 6.20
EFG/1971 (Ribbon); 1988 (Octagon)	Pilot	1.5 MW	Production −12 MW	Production −20 MW	Figure 6.21
ESP (STR)/1980	R&D	–	Pilot <0.5 MW	Production <5 MW	Figure 6.23
SF/1983	–	–	Pilot −1–2 MW	Production >5 MW	–
RGS/1983	R&D	–	R&D	Pilot <1 MW	Figure 6.24

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 967

ribbon and foils and allow these new-generation technologies to become market leaders in silicon wafer production. Technology description, the status of each of the growth approaches and the barriers for each of the technologies to overcome in order to remain competitive are the topics of the following sections.

6.5.1 Process Description

The ribbon technologies that have been proposed over the past three decades and that have survived till the commercial manufacturing and R&D phases (Table 6.3) may be grouped into two basic approaches: "vertical" and "horizontal" growth (pulling) methods. The latter category is further subdivided into methods in which either a substrate is used to assist in the formation of the crystal or foil or those that do not use any substrate material. The horizontal methods refer not so much to the geometrical aspects related to the ribbon-pulling direction as to the disposition of the temperature gradients, which act at the interface and influence growth characteristics. The EFG, WEB and STR methods are examples of the vertical method category, while both the RGS foil and the SF methods grow crystals in a horizontal-pulling configuration with the aid of a substrate. The term "foil" is used interchangeably with wafer, but here we use it to refer more specifically to the RGS wafer to distinguish a unique aspect: the wafer is crystallised upon contact with a substrate, and is then detached and the substrate material recycled.

Ribbon/foil growth techniques have historically been evaluated in a number of variants and modifications of the techniques listed in Table 6.3. Successes and failures in many of these variants often spawned new processes or led to evolution and modifications in old variants. A bibliography and descriptions of the techniques and a detailed historical perspective of the many variant ribbon technologies that have been pursued can be found in the endnote [42] and in References [43, 44].

Fundamental differences exist in the heat transfer and the interface temperature gradients during growth for these two general categories of ribbon and foil production methods. These lead to very different process limits in several important areas: the capacity, or throughput potential in a single furnace configuration, crystallite or grain nucleation characteristics and the pulling speed. The speed is constrained as a consequence of the thermoelastic stress acting on the crystal during growth. The pull speed and stress affect the defect density and electronic quality. For example, for the vertical techniques – WEB, EFG and STR – the crystal growth direction and dominant heat transfer of latent heat from the interface are both parallel to the pulling axis of the ribbon and essentially perpendicular to the growth interface. The latent heat conducted along the ribbon is radiated to the environment. The pulling speed and the interface growth velocity are the same. For RGS and SF, crystals nucleate on the substrate and grow nearly perpendicular to the substrate-pulling direction, while the growth interface tends to be angled towards the pulling axis of the substrate. Thermal conduction of latent heat from the growth interface is augmented in the direction perpendicular to the pulling axis, that is, through the thickness of the ribbon, because of conduction into the substrate. This augmented heat removal allows very high ribbon production rates, whereby low interface growth rates, v_I, are realised with high pull rates, v_P, that is,

$$v_P = v_I / \cos(\theta)$$

where θ is the angle between the normal to the interface and the pull direction and is close to 90°. The low interface growth rate, in turn, reduces the need for the high interface temperature gradients required to maintain growth stability in vertical ribbon growth. The gradients in the vertical methods are the cause of high thermoelastic stresses and set practical productivity limits when low defect densities and flat ribbon are required. Details on the process limits affecting the horizontal growth techniques may be found in other publications [45, 46]. We next give a description of each of the techniques listed in Table 6.3.

WEB. WEB is grown directly from melted silicon in a crucible with no shaping device (Figure 6.21) [47]. A dendritic seed or button is lowered into a supercooled melt. The seed spreads laterally to form a button. When the seed is withdrawn, two secondary dendrites propagate from the ends of the button into the melt, forming a frame to support the freezing ribbon. The dendrites grow into the melt that has been supercooled by several degrees. Very accurate melt temperature control is required in order to maintain the supercooled interface condition and prevent "pull-out", whereby the growth terminates by voiding of the meniscus. The width of the ribbon is controlled by the position of the two dendrites that support the liquid film. The growth velocity is determined by the rate of removal of the latent heat into the ribbon and of the heat conducted through the melt through the meniscus. Typical growth rates are from 1 to 3 cm/min.

In vertical ribbon growth, the meniscus contains the suspended melt volume that connects the bulk melt to the growth interface and crystal. Its shape and the heat conduction taking place within it critically affect impurity segregation and the crystallisation

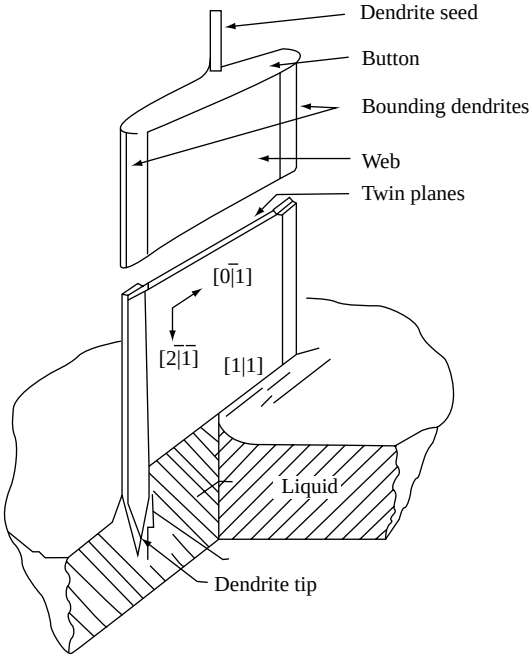


Figure 6.21 Schematic of Dendritic Web (WEB) growth process for ribbon

conditions and shape of the crystal. The meniscus height, h, away from the influence of the dendrites, is fixed by the surface tension, liquid-silicon contact angle and liquid density. This can be calculated from the solution of the Young–Laplace equation requiring a contact angle of $\Theta = 11^\circ$ at the liquid–solid interface that gives

$$h = a[1 - \sin(\Theta)] \quad \text{where} \quad a = (2\sigma/\rho g)^{1/2}$$

where $\sigma = 720$ dyne/cm is the surface tension, $\rho = 2.53$ g/cm^3 is the density of liquid silicon and g is the gravitational acceleration. Because thermal radiation dominates the heat flow from the ribbon, the geometry of the heat shields and the susceptor lid controls the isotherms in the melt and the ribbon.

Accurate temperature control, within a few tenths of one degree, is necessary to ensure a uniform ribbon width and thickness. The temperature of the melt surface must be constant over the width of the growing web to prevent the dendrites from growing in or out. The dominant impurity in the WEB in production today is oxygen since a quartz crucible is used. Typical WEB thicknesses range from 100 to 150 μm and widths up to 8 cm have been grown. Growth lengths between seedings in pilot production extend to many tens of meters.

EFG. In this technique, the geometry of the ribbon is controlled by a slotted graphite die through which silicon is fed via capillary action (Figure 6.22) [48]. A seed crystal is

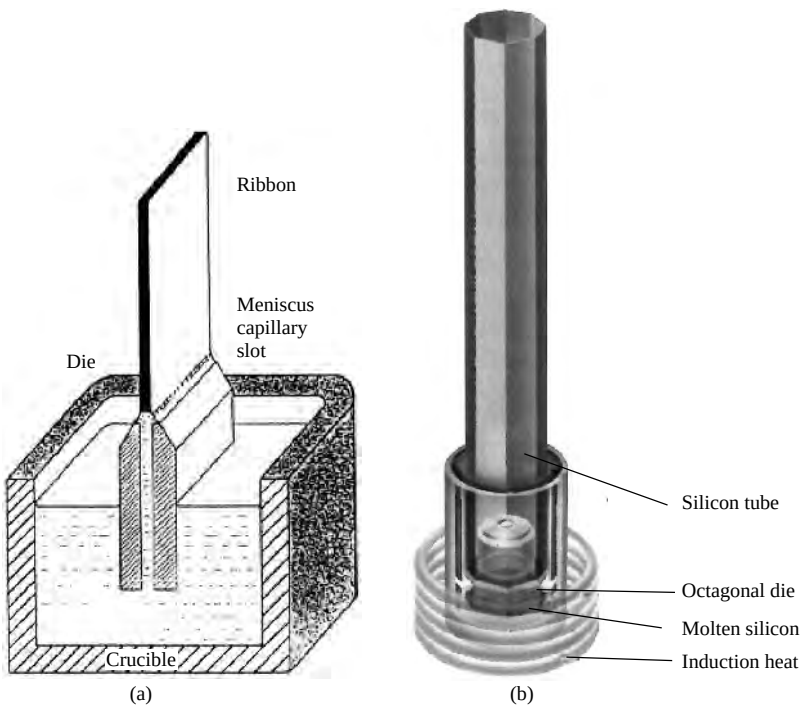


Figure 6.22 Schematics of Edge-defined Film-fed Growth (EFG) growth process: (a) ribbon die and crucible configuration and (b) octagon configuration

lowered until it contacts the liquid in the capillary. The liquid spreads out over the top of the die to the edges where it is pinned by surface tension. The seed is withdrawn, pulling the liquid up while more liquid flows upward through the capillary. As the ribbon is withdrawn, the liquid freezes on the solid crystal. The die and the crucible are integral, that is, made of the same piece of graphite. The thickness of the sheet material is fixed by the width of the die top, distance between the die tip and melt level, meniscus shape, heat loss from the sheet and the pull rate. The shape of the liquid–gas interface, or meniscus, which connects the die to the solidifying ribbon, is described by the Young–Laplace or the capillary equation. As with WEB, the growth rate is controlled by how fast heat can be conducted away from the interface and lost by radiation or convection from the solid crystal. Growth is self-stabilising because the meniscus height increases with an increase in pull rate. The curvature of the meniscus causes the thickness of the crystal to decrease. This increases the rate of heat removal per unit area of the interface, thus increasing the growth rate until it is again equal to the pull rate.

The dominant impurity in EFG ribbon is carbon, which is in supersaturation. Temperature control of a few degrees along the interface is sufficient to prevent ribbon pull-out or freezing of a growing ribbon to the die top. Over time, the die becomes eroded affecting ribbon properties and leading to a non-uniform ribbon thickness and growth difficulties.

Ribbons with thicknesses from 400 μm to as little as 100 μm have been grown. Rather than a single flat ribbon, hollow EFG polygons are grown to enhance the rate of throughput. The favoured geometries for commercial development today are octagons with 10-cm- or 12.5-cm-wide faces, equivalent to growth of up to a 100-cm-wide ribbon from a single furnace. Various closed geometries, including nonagons with 5-cm faces, and large-diameter cylinders have been grown. Growth velocities for the EFG octagon are 1.7 cm/min. The relationships between the EFG process parameters and the silicon ribbon characteristics, including thermal stress and the influence of impurities and defects on the quality of the material, have been extensively examined and are reviewed in Reference [49].

An extension of the EFG process to growth of 50-cm-diameter cylinders has recently been demonstrated [50]. An example of such a cylinder 1.2 m in length is shown in Figure 6.23. The cylindrical geometry offers some relief from the large thermoelastic stresses generated in plane ribbon. This allows consideration of higher productivity furnaces from a combination of larger perimeters and potentially higher growth speeds. Growth of EFG cylinders with average wall thickness down to 100 μ has been demonstrated, and solar cells have been made on this material [51].

STR. In this technique, ribbon growth takes place directly from a pool of melted silicon without a die (Figure 6.24) in a process mirroring the WEB geometrically. Rather than dendrites, as with WEB, the position of the ribbon edges in STR is maintained by two strings fed through holes in the bottom of the crucible. The strings are drawn upward out of the melt to support the meniscus and the ribbon, and their pull rate determines the growth speed of the ribbon. The thickness of the ribbon is controlled by surface tension, heat loss from the sheet and pull rate. An important difference of the STR process from WEB growth is that the constraints of maintaining propagating dendrites and a supercooled melt are eliminated, and this relaxes the high degree of temperature control required in the WEB furnace. The high meniscus, 7 mm (see Equation 2), allows simple control of



Figure 6.23 Experimental 50-cm-diameter EFG cylinder exiting from furnace

the growth process and maximises its stability to mechanical and thermal fluctuations. Depending upon the wetting qualities of the strings and their diameter, the meniscus height at the strings near the edges differs from that at the centre, and it is usually much lower at the edges [52, 53].

For comparable thicknesses, the growth velocities of STR are similar to EFG and WEB. Careful adjustment of the growth parameters can allow very thin ribbon, down to 5 μm, to be grown [54]. Generally passive after heaters are used, but some work on an active after heater has been carried out to allow low-stress, 100-μm-thick ribbon to be grown. This material was sufficiently flat to be made into solar cells. Because of the concave downward meniscus curvature at the string, any grains nucleated at the strings can propagate into the ribbon.

RGS. In this growth technique, the silicon melt reservoir and die are placed in close proximity to the top surface of a substrate, on which the ribbon/foil grows. The substrate may be graphite or ceramic (Figure 6.25) [55]. The principle is to have a large wedge-shaped crystallisation front. The die contains the melt and acts to fix the width

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 972

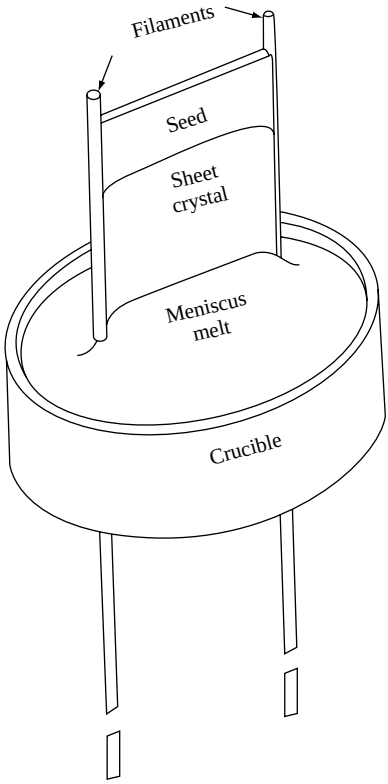


Figure 6.24 Schematic of String Ribbon (STR) growth system

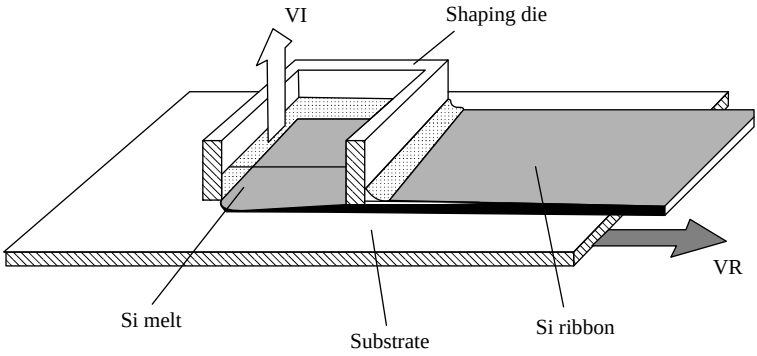


Figure 6.25 Schematic of Ribbon Growth on a Substrate (RGS) configuration

of the foil. The thickness of the foil is controlled by the heat-removal capacity of the substrate, pull rate and surface tension. The direction of crystallisation and growth are nearly perpendicular. The area of the growth interface now can be very large compared to the foil thickness. The latent heat is extracted by conduction into the substrate. The thermal gradients near the interface are small, thus reducing thermally induced stress in

the wafer. Growth rates from 4 to 9 m/min have been demonstrated. One example was an 8.6-cm-wide foil, 300-μm thick, grown at 6.5 m/min.

An important goal in the R&D phase of RGS has been to make a substrate that can be reused. After cooling, the silicon foil may be separated from the substrate by stresses arising from differences in thermal expansion between substrate and silicon. Experimentation with coated foils is in progress and offers the most promise in providing a cost-effective reusable substrate. Thicknesses between 100 and 500 μm have been grown. By working with the lower thermal gradients in the foil thickness direction, but still large enough for rapid growth, fluctuations in the pulling speed and gradient only affect the foil thickness slightly [55].

SF. The details of the SF process are proprietary. The silicon crystal is grown in a thin layer directly upon either an insulating or a conducting substrate, with a barrier layer that promotes nucleation [56]. In the case of an insulating substrate, the barrier layer must also act as a conductor to collect the current generated in the cell. In the case of a conducting substrate, the substrate can also act as an electrical conductor if *vias*, or holes, are provided to connect the thin silicon crystal layer and the substrate. The SF thin film and barrier layer do not separate from the substrate on cooling as in RGS, but become the active part of the solar cell. The grown polycrystalline silicon layer is made very thin ($\ll$100 μm), thus reducing the amount of silicon required. Currently, layers of 20-μm thickness are under development. A variety of substrate materials have been used including steel, ceramics and graphite cloth [56, 57]. It is necessary that the barrier layer prevent the transport of impurities from the substrate into the silicon. The barrier layer allows wetting and nucleation during growth. It should also act to electrically passivate the back surface and have a high optical reflectivity.

An insulating barrier layer has been reported that promotes growth of large columnar grains (greater than 1 mm) through the thickness of the grown SF silicon film [56]. As-grown films on coated ceramic substrates exhibit very low diffusion lengths of less than 10 μm. The new barrier layer and the substrate result in longer diffusion lengths, 20 to 40 μm, and the silicon has an improved response to phosphorous gettering.

6.5.2 Productivity Comparisons

Ribbon crystal growth technology for production of silicon wafers has been historically faced with the evaluation of the trade-off between bulk electronic quality and throughput (productivity per furnace). The choice of crystal growth conditions is made on the basis of wafer cost parameters and the premium imposed by the marketplace on solar cell efficiency. Material quality that can translate into high solar cell efficiencies has always been a primary market driver guiding ribbon growth process development. Ribbon wafers have inherently lower wafer production costs than those obtained from directionally solidified and cast ingots, or Cz boules, because ribbon growth avoids the large material losses due to sawing, which exceed 50% of the starting feedstock. The ribbon geometry has an additional cost advantage in that high levels of radiative cooling allow very rapid pulling rates. On the other hand, the higher wafer cost for the cast ingot or Cz boule production methods demand that these products maintain an advantage in bulk electronic quality, and higher solar cell and module efficiencies, in order to stay competitive with respect

Table 6.4 Single-furnace performance metrics for ribbon technologies under development and in commercialisation

Method/ parameter	Pull speed [cm/min]	Width [cm]	Throughput [cm^2/min]	Furnaces per 100 MW[a]
WEB	1–2	5–8	5–16	2000
EFG Octagon	1.65	8 × 12.5	165	100
STR	1–2	5–8	5–16	1175
SF	[b]	15–30	[b]	[b]
RGS	600–1000	12.5	7500–12 500	2–3

[a] Furnace data are taken from Reference [58], where throughput is normalised for comparison purposes to an overall yield of 90% and cell efficiency of 15% for all processes
[b] Pulling rate parameters for the SF process are not available

to ribbon technology. Superior electronic quality is generally achieved at the expense of throughput in bulk crystal growth from the melt.

A summary of performance metrics for ribbon technologies currently under development is given in Table 6.4. A critical driving parameter in technology development of ribbon methods for large-scale manufacture and commercialisation is the productivity per furnace. Productivity governs the capital cost of installed capacity and direct labour costs, which constitute significant barriers to ongoing commercialisation on a large scale for all ribbon technologies.

6.5.3 Manufacturing Technology

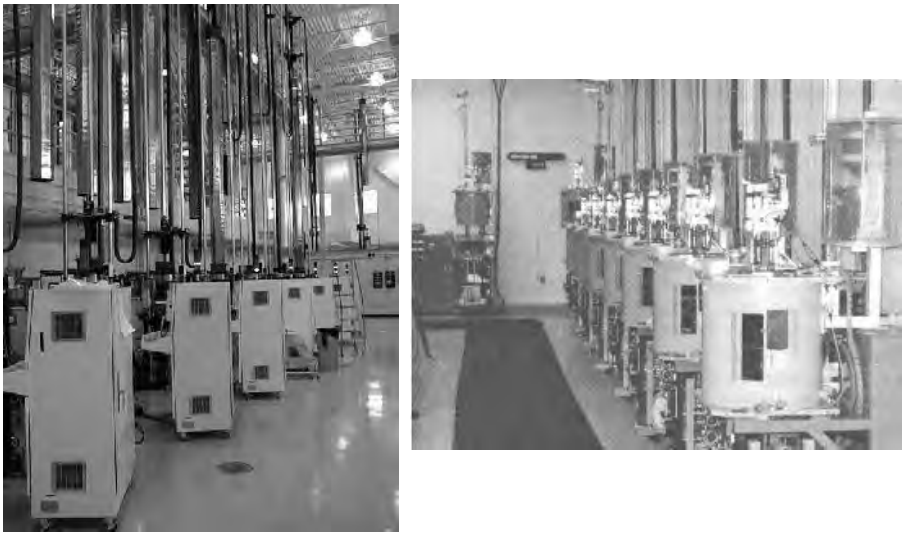
Table 6.4 shows that development of ribbon technologies is proceeding along two distinct paths. WEB and STR rely on low furnace cost to remain competitive in wafer costs. Development of SF and RGS technologies is focused on achieving superior throughput per furnace.

Scaling of ribbon factories significantly beyond the manufacturing levels practised now, for example, 100 MW, poses different challenges for technology development in these two cases. The low throughput ribbon furnace (WEB, STR) requires a simple and low-cost furnace design, a high level of automation and low infrastructure costs. In contrast, high reliability and uptime of furnaces and the growth process are most critical for the high throughput technologies SF and RGS. EFG technology development is moving in a direction that is trending towards the middle of these two extremes.

Examples of commercial installations and equipment now in manufacturing for EFG and STR technologies are shown in Figure 6.26. Figure 6.26(a) pictures a group of furnaces in the EFG octagon manufacturing line, with a high bay area to accommodate the 5.4-m octagon growth lengths. EFG wafers of 10-cm width and of 10- or 15-cm lengths are standard products and are cut from the octagon tubes using high-speed lasers (not shown). More detail on this technology is given in Reference [59].

Single ribbon furnaces for the growth of 8-cm-wide ribbon of the STR manufacturing line are shown in Figure 6.26(b). Ribbon sections up to a meter long are scribed

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 975



(a) (b)

Figure 6.26 Manufacturing crystal growth equipment in commercialisation for (a) EFG and (b) STR technology

from the ribbon while it is growing, and then further cut into wafers of 10-cm length prior to processing.

WEB, SF and RGS ribbon technologies all are in various stages of R&D and pilot demonstrations leading to commercialisation. SF is perhaps the closest to successfully scaling up the technology, as a wider (20 cm) and higher throughput furnace is reaching the final demonstration phase. WEB is basing its expansion to a 1- to 2-MW pilot operation on a single ribbon furnace for producing 5-cm-wide wafers. RGS is moving towards the use of 12.5-cm-wide ribbon in which high sustained throughput technology and consistent high material quality needs to be demonstrated.

6.5.4 Ribbon Material Properties and Solar Cells

Except for WEB, all the growth processes produce multicrystalline ribbon. WEB ribbon grows with (111) crystallographic faces (see Figure 6.21). It typically does not have any grain boundaries, but has a single multiple-twin boundary located about mid-way through the ribbon thickness. Each (111) surface is made up of a single grain, and the dislocation density is the lowest of any ribbon, 10^3 to $10^4/cm^2$.

For the other two vertical ribbon growth cases, EFG and STR, extraneous crystals are generated most often at the sides of the ribbon (i.e. octagon tube corners) and propagate along the growth axis of the ribbon. These crystals form elongated grains often many centimetres in length along the growth axis, and which extend through the ribbon thickness. In EFG, the grains are interspersed with numerous twin boundary arrays. The grains at the ribbon edge in STR are generally smaller than in the centre. Because the meniscus near each string is concave downward, the grains nucleated at the string can

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 976

propagate into the ribbon centre. In both EFG and STR, the grain dimensions typically are large compared to the ribbon thickness and the as-grown diffusion length, and the charge collection and solar cell efficiency are minimally influenced by grain boundary recombination.

For SF and RGS, the substrates provide the dominant nucleation sites for crystals. The grains usually are columnar and extend through the ribbon thickness with dimensions that can be made large compared to the diffusion length with proper adjustments of the pulling speed and interface inclination.

Fast movement of the solid–liquid interphase reduces the ability of the freezing process to segregate impurities to the melt. This is represented in Table 6.5 together with the crystalline aspect and the dislocation density of the different technologies.

Stresses produced by thermal gradients during growth generate most of the intragranular dislocations in ribbon material. In the best quality EFG material, it has been shown that the loss in background current is highly correlated with the dislocation density and not grain boundaries [60]. It is not clear if the intrinsic qualities of the dislocations act as recombination centres or if the associated impurity cloud is responsible. Dislocations decorated with SiO_x precipitates have been reported to limit the lifetime in WEB [61] and RGS material [62]. Recent photoluminescence studies suggest similar causes for dislocation recombination activity in EFG material [63].

A critical parameter for solar cell efficiency is the ribbon thickness. As shown by Bowler and Wolf [64], an optimum thickness for peak efficiency occurs. This thickness is dependent on the fabrication technique and material properties, including front and back surface recombination velocities, minority-carrier lifetime and base resistivity among other parameters. For a "typical" $n^+ pp^+$ structure with a 200-μm diffusion length, L_d, a back surface field and a single layer anti-reflection coating, the optimum thickness region consists of a broad peak near 80 μm (Figure 6.27) as calculated using PC-1D [65]. For a 100-μm L_d the optimum thickness peaks at about 50 μm and for a 400-μm L_d it is near 120 μm. A front surface recombination velocity of 10^5 cm/s is assumed.

Table 6.5 Comparison of silicon ribbon material characteristics. In all cases the columnar grains extend through the thickness of the ribbon. An equilibrium segregation coefficient of $k_0 \sim 10^{-5}$ is typical of the most detrimental impurities for ribbon bulk lifetime

Material	Crystallinity	Dislocation density [1/cm^2]	Effective segregation	Thickness [μm]
EFG	Columnar grains in growth direction	10^5–10^6	$k_0 < k_{eff} < 10^{-3}$	250–350
WEB	Single (111) face central twin planes	10^4–10^5	$k_0 < k_{eff} < 10^{-3}$	75–150
STR	Columnar grains in growth direction	5×10^5	$k_0 < k_{eff} < 10^{-3}$	100–300
SF	Columnar grains through thickness	10^4 to 10^5	$k_{eff} < 1$	50–100
RGS	Columnar grains through thickness	10^5–10^7	$k_{eff} < 1$	300–400

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 977

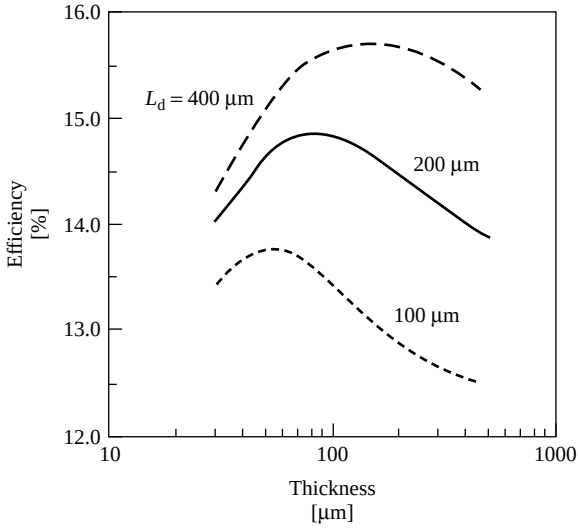


Figure 6.27 Solar cell efficiency versus thickness. (See the text for a description of the solar cell parameters)

The thickness of WEB and SF is closer to the optimum than the other three ribbon growth techniques. Light trapping will move the optimum to a thinner base, but such optical structures on multicrystalline ribbon are not yet practical. For growth on a substrate, it is possible to texture the substrate to trap light. This has been shown on SF wafers [57].

Solar cell fabrication processes used for conventional Cz and cast material wafers are commonly applied to silicon ribbon. If the ribbon is doped *p*-type with boron, the *n*-type emitter typically is formed by phosphorus diffusion either from $POCl_3$, PH_3 or a spin-on source. Front contacts are usually screen-printed or evaporated with diffused or alloyed aluminium as back contact to produce a back surface field. A double or single layer anti-reflection coating may be used. For SF grown on an insulating substrate, contact with the back surface requires etching holes to allow contact with the conducing barrier layer. Table 6.6 summarises material characteristics and solar cell performance potential of the various forms of ribbon material. Boron dopes the ribbon *p*-type. Antimony can be used to produce *n*-type material as reported for WEB.

Table 6.6 Some "best" solar cell efficiency levels for various ribbon technologies

Material	Resistivity [Ω cm]	Carbon [1/cm^3]	Oxygen [1/cm^3]	Efficiency [%]
EFG (Reference [66])	2–4, *p*-type	10^{18}	$<5 \times 10^{16}$	15–16
WEB	5–30, *n*-type	Not detected	10^{18}	17.3
STR	1–3, *p*-type	4×10^{17}	$<5 \times 10^{16}$	15–16
SF (Reference [67])	1–3, *p*-type	5×10^{17}	5×10^{17}	16.6
RGS (Reference [68])	2, *p*-type	10^{18}	2×10^{18}	12.0

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 978

Ribbon cell processing, nevertheless, needs to recognise the unique growth constraints of all of the techniques. As noted above, a common material characteristic for ribbon materials historically has been the compromised as-grown bulk electronic quality, dictated by the development of desired high-throughput growth configurations. This strategy has its base in a commercial demand for very low cost wafers that must compete at relatively low volumes with already established dominant products based on single-crystal Cz wafers or directionally solidified and cast ingots. As-grown ribbon diffusion lengths most often are less than 100 μm. To obtain the maximum cell efficiency on this ribbon, with higher dislocation densities and contaminating impurities than in competitor wafers, a solar cell processing strategy was devised early in the history of ribbon technology development that incorporates special bulk lifetime upgrading steps. For example, bulk lifetime upgrading via aluminium alloying and hydrogen is particularly effective for EFG material [69, 70]. Another approach is to use plasma-enhanced chemical vapour deposition (PECVD) of silicon nitride to generate hydrogen for passivating the silicon bulk [71].

6.5.5 Ribbon/Foil Technology – Future Directions

Ribbon/foil wafer production is poised to move on to face a new round of challenges in the construction of large (50–100 MW) manufacturing facilities for crystalline silicon ribbon wafers. The RGS foil technology, with the greatest potential of all ribbon methods for cost reduction on the basis of a high throughput per furnace, is entering a pilot demonstration phase. This process faces challenges in process and equipment development before it can enter high-volume manufacturing of wafers. In the next pilot phase, we may expect that RGS will demonstrate a consistent material quality sufficient for improving cell efficiency to greater than 12% from the current 10 to 11%; process control capable of reproducibly producing a low stress, regular structure, with a shaped 12.5-cm-wide wafer suitable for high-yield cell processing at about a 300-μ thickness and a reliable prototype furnace with melt replenishment to enable continuous production, which will gain full benefit from the high throughput growth concept.

The future focus of WEB and STR ribbon development is on process automation and capital cost reduction for furnaces and infrastructure based on a concept of a low throughput (per furnace) process. Production will probably grow to between 1 and 10 MW for each of these approaches over the next several years. Process parameters that will be practised in this next round of manufacturing equipment expansion for both of them are narrowed down to a single ribbon per furnace concept with similar throughput parameters – a ribbon width of about 8 cm and a pull speed in the 1 to 2 cm/min range. The WEB ribbon technology is embarking on its pilot expansion with a unique process for the growth and manufacturing of solar cells at a 100-μm wafer thickness. STR is expanding its manufacturing with a 300-μm-thick wafer. Although wafer bulk quality is demonstrated on an R&D level to be capable of achieving 15 to 16% cell efficiencies for STR, and over 17% for WEB, quality and cell efficiency levels on a multi-megawatt scale are yet to be established. Both approaches will attempt to demonstrate the cost-effectiveness of operating on a multi-MW level in the next few years. R&D directions, which would appear to have the most potential for the reduction of wafer material costs for these techniques, are growth of wider ribbons and more ribbons per furnace.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 979

EFG and SF ribbon technology have successfully completed their initial scale up of wafer production to the multi-megawatt level. Process control and equipment reliability improvements, which can drive throughput and yield higher, become increasingly more dominant in determining the manufacturing cost. As throughput per furnace increases, capital cost impact on variable manufacturing costs from the growth furnace decreases. There is pressure on all ribbon technologies to concentrate on reducing the capital cost of the wafer production equipment if the transition to large-scale wafer manufacturing of 50- to 100-MW annual capacity factories is to be sustained.

Process variable ranges are firmly established for the EFG process. Octagon tube length, throughput and wafer thickness parameters will remain within the ranges given in Table 6.4 for the next generation of equipment, while octagon face width, and hence the EFG wafer dimension, will increase from 10 to 12.5 cm. The major thrust in R&D on the EFG process in this phase will be on process control and process and equipment automation.

The benefits of the savings in silicon feedstock and potential gains in cell efficiency with a reduction of the ribbon/foil thickness are well understood for all these technologies. However, the pressure to carry out R&D in this direction for the case of wafers made from ribbons is not as acute as for conventional crystalline silicon wafer manufacturing methods because of the large benefit in feedstock savings already realised for ribbons on account of their favourable geometry. The R&D for the next generation of vertical ribbon technology beyond about five years will target the demonstration of production methods for very thin wafers. A strong motivator driving thickness reduction will be the pressure to increase the cell efficiency, which is seen to be capped in the 16 to 17% range (see Table 6.6) for current cell designs and wafer bulk quality. Low-cost cell designs, which can break this barrier and achieve desired targets of 18 to 20% for ribbon, are most easily found for thinner wafers, but this also requires improvements in bulk electronic quality to be achieved concurrently.

The major problem in this development for all vertical growth techniques will be to find methods to reduce the effects of thermal stress. At present, the only means by which this can be done is to reduce the pull speed. The cylinder geometry has the potential to offer some relief to the EFG process at the expense of having to work with thin curved wafers in cell and module processing. Although thermal stress is not a problem with substrate-assisted growth techniques, there probably will be a trade-off between good bulk quality with large grains and throughput.

6.6 NUMERICAL SIMULATIONS OF CRYSTAL GROWTH TECHNIQUES

Commercial finite element simulation tools for structural analysis in computer-aided engineering started to develop at the beginning of the seventies. Today, simulation tools are an essential part in various industry productions; see crash test simulation for automobile development or airflow simulations in the aerospace industry. As an advanced application the descriptions of whole production processes are the goal of the strategies for simulations. If these strategies are successful, computer modelling opens the opportunity to shorten development time for production facilities, to reduce the costs for the

engineering and to speed up process optimisation. In this chapter we will report on those simulation tools, various thermal models and examples of numerical simulations of silicon crystallisation processes.

6.6.1 Simulation Tools

Numerical simulation tools can be distinguished in universal and special-purpose programs. Examples of commercial universal-purpose programs are ABAQUS [72], ANSYS [73] and MARC [74] with a wide range of applications in structural analysis, thermal and fluid-flow simulations or electromagnetic field simulations. The number of special-purpose programs cannot be estimated seriously. Many universities and companies are working with specially developed software tools to obtain solutions of their specific problems. Recently, the large commercial programs both compromise and enable the user to add their specific subroutines to a program run.

The main structure of most of the simulation tools is similar: A pre-processing is designed to define the initial and boundary conditions of a simulation run and includes the generation of the simulation domain (finite element mesh) as well as a set of physical data that describes the material properties. The main-processing is normally not interactive and contains the solver of the mathematical formulations. The post-processing visualises the simulation results.

The demand for simulation tools depends on the complexity of the physical problem or on the technical process that the user wants to simulate. In general, the description of all physical relationships is reachable only in relatively simple and well-known problems. The full description of an industrial production facility by numerical simulations is not possible today, and neither is it reachable in the near future because too many details are too complex to be described by the numerical models. Therefore, the development of useful simplifications is one of the important keys to a successful simulation. This demand requires an integrated teamwork between the user of a simulation tool and the operators at the production facility and other process specialists.

Not the another important requirement is the validation of simulation results by experimental data. At least two experiments are necessary to validate simulation results concerning the process behaviour of a production facility. This means that the simulation model should be validated by measurements during a standard process and in a worst-case scenario to ensure the correctness of the results in an enlarged area of validity. Normally, these experiments are expensive and difficult to realise during a running production, but otherwise, running a non-optimised production would be quite more expensive. Anyhow, the validation of simulation results is necessary to ensure the success of the simulation method.

6.6.2 Thermal Modelling of Silicon Crystallisation Techniques

The wafer material for crystalline silicon solar cells can be divided into those from ribbon and bulk crystals. For most of wafer production processes, numerical simulations are in use to describe the thermal conditions during the crystallisation. In the case of ribbon crystals, only the EFG [75, 76] and the STR process [77] have reached a market

production. The RGS process [78] is now in preparation for a commercial production. The Cz crystal-pulling technique is the standard process for microelectronic single-crystal wafers and covers an essential part of the PV market share [79, 80]. The TriSi crystal is a new variation of this process, especially for photovoltaic applications [81]. The characteristics of ingot crystallisation can be explained by the shape of their liquid–solid interface. Anyhow, today's ingot crystallisation goes more and more towards a mostly planar solidification. For the use of numerical simulations of the Cold Wall process, see [82, 83]; for the Heat Exchange Method (HEM), see [84] and for the Solidification by Planar Interface (SOPLIN) processes, see [85, 86].

To simulate the temperature history during crystallisation, various thermal effects must be taken into account. In Figure 6.28 the scheme of thermal conditions for the ribbon growth and ingot crystallisation is presented. The biggest difference between the two is the strong variation of the cooled surface to volume relation (SV) during crystal growth. This relationship can be used to qualify the cooling behaviour of the different crystals in an equivalent surrounding. For ribbon growth SV is given as 2/ribbon thickness and for ingots as 1/ingot height. The high number for ribbon growth (e.g. SV = 66/cm) means that the surface affects the crystallisation, while the low number for an ingot geometry (e.g. SV = 0.033/cm) shows that volume effects are more important for crystallisation. By this, the SV parameter characterises the requirements for the modelling of different crystallisation techniques. In the case of bulk crystallisation, the latent heat at the liquid–solid interface must be lead away by a heat sink at the bottom of the ingot. By this, the crystallisation is propagated by a conductive heat flow through the solid ingot volume and the temperature gradients inside the volume have to be simulated with high attention. In the case of ribbon growth, heat flow by convection and radiation at the silicon surface is the dominant heat-transport mechanism to lead away latent heat and propagate the solidification. Therefore, simulation results are very sensitive to heat-transition coefficients and the emission behaviour at the ribbon surface.

Furthermore, both techniques can be distinguished into quasi-steady state and moving boundary processes. Assuming a constant pulling speed, the ribbon growth is

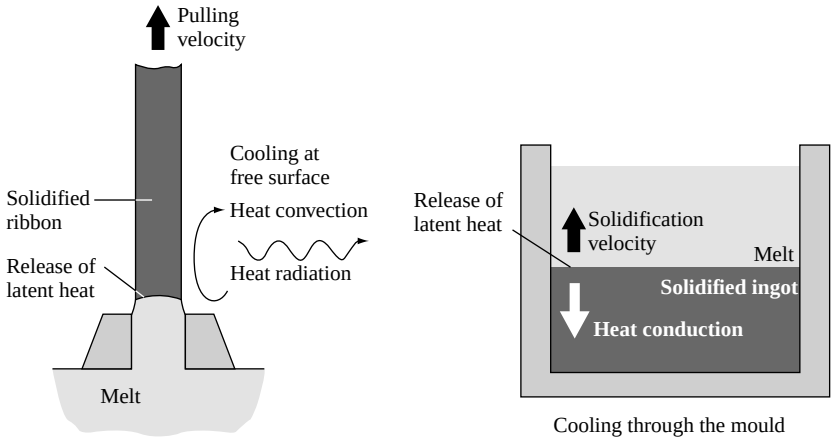


Figure 6.28 Scheme of thermal effects during ribbon growth and ingot crystallisation

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 982

characterised by a steady-state temperature field and the liquid–solid phase boundary can be modelled by a fixed-temperature boundary condition, as the silicon melting temperature of 1410°C. In the case of the ingot casting, the phase boundary moves through the crystal and the release of latent heat can be modelled by an enthalpy formulation. By this, the release of the latent heat of finite elements can be taken into account directly after their total solidification, or the fraction of latent heat must be considered for partly solidified elements [87, 88]. The importance of an accurate modelling of the release of latent heat may become more clear by estimating the crystallised volume rate in typical ingot processes to be around 9000 cm^3 per hour, which means a latent heat source of more than 8 kW at the location of the phase boundary. Pulling one 10-cm-wide ribbon, the crystallised volume rate is about 30 cm^3 per hour with a latent heat release of 0.03 kW.

The Czochralski and TriSi crystal-pulling techniques can be classified between ribbon and ingot crystallisation. The temperature profile can be assumed to be stationary and the SV parameter, given as 1/crystal diameter, lies, for example, in the range of 0.066/cm.

In general, the heat flow in silicon during crystallisation can be described by the heat-transport equation [89–91]:

$$\rho c_p \frac{\partial T}{\partial t} = \lambda \nabla^2 T + L \frac{\partial f_c}{\partial t}$$

with the silicon data:

density of solid silicon	$\rho_{(1410°C)}$	=	2.30 [g/cm^3]
density of liquid silicon	$\rho_{(1411°C)}$	=	2.53 [g/cm^3]
heat capacity	$c_{p(20°C)}$	=	0.83 [J/g K]
	$c_{p(1410°C)}$	=	1.03 [J/g K]
heat conductivity	$\lambda_{(20°C)}$	=	1.68 [W/cm K]
	$\lambda_{(1410°C)}$	=	0.31 [W/cm K]
latent heat of phase change	L	=	3300 [J/cm^3]

The time t and the temperature T are variable and result from the simulation. For the moving boundary case the solid fraction f_c becomes important. This parameter depends on the finite element temperature and varies between zero for a completely liquid finite element and one for a solid element.

Additionally to these material properties and the heat flow mechanisms in the silicon material, the description of the internal furnace construction must be taken into account to perform simulations of crystallisation facilities. This includes the geometrical description and material properties of the internal set-up as well as the radiative heat exchange with heaters and cooling facilities.

6.6.3 Simulation of Bulk Silicon Crystallisation

As an example of the temperature simulation of silicon ingot crystallisation, the SOPLIN casting technique is selected. To simulate this process, a finite element mesh of about 230 000 elements was built to describe the furnace geometry. This mesh includes the silicon ingot, the mould, all insulation materials and active heating and cooling facilities,

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 983

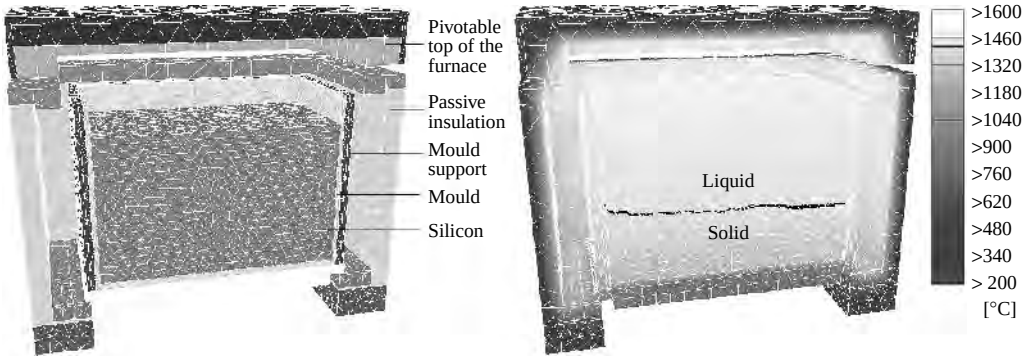


Figure 6.29 Finite element geometry of an ingot casting furnace and simulated temperature distribution during a reference process. The liquid–solid interface is marked by the black line

as shown in Figure 6.29. Because of confidentiality agreements with the industry, the heating and cooling systems are not shown in detail in this figure. All heat conductance and capacity effects as well as the non-stationary release of latent heat are taken into account. All material contact regions between silicon, mould or insulation materials are modelled by heat flow–resistance parameters. To describe the heat flux by radiation inside the furnace, a view-factor model is included in the software. All material data are treated in their temperature dependency and all the internal control systems of the furnace are added to the simulation software.

To start one simulation run, only the cooling water temperature and the time-dependent process control information are necessary as input data, as they are entered in the crystallisation furnace. Output from one calculation is the three-dimensional temperature history in the furnace, beginning after pouring the melt and ending with a homogeneous temperature of about 300°C inside the ingot. This calculation needs less than 6 h on a common one-processor workstation.

In Figure 6.29, an example of the temperature distribution during a reference process is shown in the middle, cut through the furnace. The liquid–solid interface is marked by the melting temperature isotherm. The solidification front is mostly flat, and a slight non-symmetry is caused by the specific construction of the heating system. These simulation results are verified in an experimental crystallisation furnace with good agreement to the measurement in the ingot volume during crystallisation.

In general, the shape of the solidification front is controlled by the lateral heat flux, while the vertical heating and cooling conditions control the solidification velocity. To investigate these general reflections for the described furnace, variations of the process control were simulated. In Figure 6.30, two variations are presented. By a 30% raise of heating power at the side walls of the ingot, the shape of the solidification front becomes more convex. Otherwise, a reduction of heating power by about 20% turns the solidification front to a more concave shape. Additionally, to this more or less predictable effect, simulation results show an increase in the solidification time for the convex crystallisation of 44% and a 30% reduced processing time for the concave solidification. Both effects are due to the total varying power input in the furnace. These simulation results enable

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 984

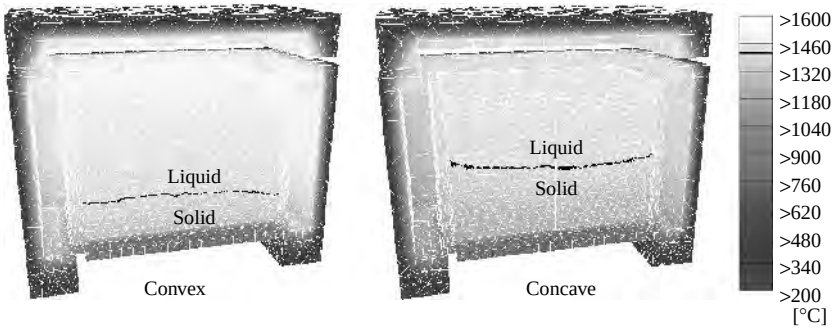


Figure 6.30 Examples of convex and concave liquid–solid interfaces due to the variation of side wall heating power. Both pictures are taken at the same process time

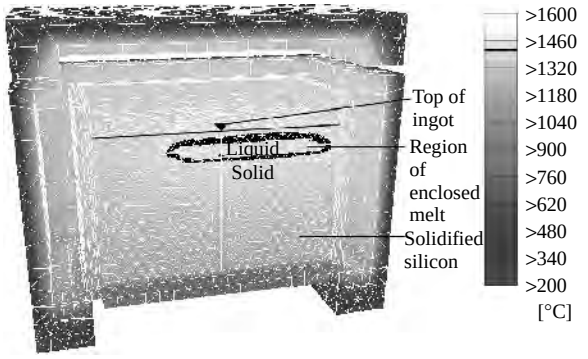


Figure 6.31 Simulation result representing a remarkable decrease of heating power at the top region of the ingot. By this, the solidification time is reduced to half, but solidification ends with an inclusion of silicon melt

the furnace operator to find the balance between material quality, which is known to be high in the case of planar solidification, and process economy.

In Figure 6.31, simulation results are shown, representing a noticeable reduction of heating power at the ingot top. By this, the solidification velocity can be speeded up and the time for solidification is reduced to half. However, in the simulated case study, the solidification ends with an encapsulation of the melt by solidified silicon. Because of the 10% higher density of liquid silicon with respect to the solid phase, this process scenario causes a burst out of melt from ingots volume. This can lead to a crack in the mould and to a damage of the furnace. These case-study simulations can find worst-case process conditions that must be avoided in production. For more simulation results of the SOPLIN process, see [92–94].

6.6.4 Simulation of Silicon Ribbon Growth

As a second example of silicon crystallisation processes, the RGS is taken. The basic idea of this process is the de-coupling of the pulling direction of substrates on which silicon

solidifies and the solidification direction of the silicon itself, as shown in Figure 6.32. By this, a very high production rate of one wafer per second is realisable. For this process numerical simulations were performed describing the crystallisation of the silicon ribbon in further detail. This simulation is realised by a phase-field approach [95]. In Figure 6.33 two nucleation states of the silicon on the substrate are compared by their temperature field in two-dimensional simulations. In the first case a supercooled region in front of the tip of the growing crystal leads to a more dendritic growth mode of the crystal, because the liquid–solid interface is morphologically non-stable. In the second crystallisation mode,

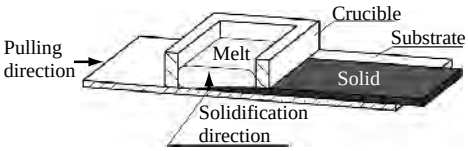


Figure 6.32 Scheme of the RGS process. The latent heat is removed through the substrate. By this, solidification is propagated vertical to the substrate and is de-coupled from substrates' pulling direction

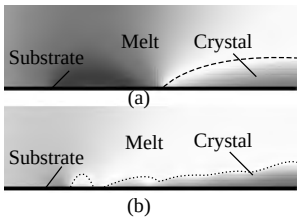


Figure 6.33 Simulated temperature profile on the RGS substrate. Light grey scales indicate high temperatures. The liquid–solid interface is marked with the dotted line. Unstable crystal growth into a supercooled melt at the tip of the ribbon. Nucleation of new grains limits the supercooling and leads to a stable columnar grain growth

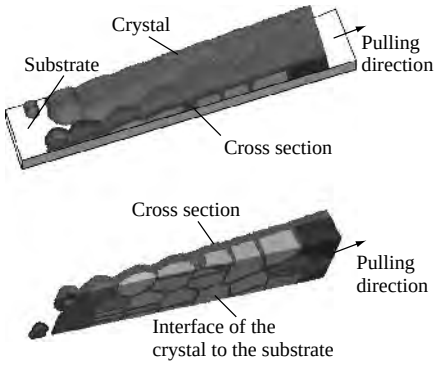


Figure 6.34 Simulation results of the growth of silicon grains on the RGS substrate during solidification. Different grains are marked by different grey scales. The liquid–solid interface is marked by a darker grey scale and envelops the grains. For a better visibility, the melt is not shown in this figure

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 986

the supercooling decreases owing to a nucleation of new grains on the substrate surface, which leads to a more columnar growth of the silicon sheet. Both crystallisation modes depend on the surface of the substrate and on the heating power controlled temperatures of the substrate and the melt. Numerical simulations allow investigating this crystal-growth tendency as well as the temperature field at the contact region of the ribbon on the substrate. In Figure 6.34 an example of the simulation of single silicon grain growth on the substrate is given [96]. In this way, numerical simulations can study the grain growth under various temperature conditions and grain-selection mechanisms of silicon crystals.

6.7 CONCLUSIONS

At the present time, the PV industry relies on solar cells made on crystalline silicon wafers, which provide around 90% of the total PV power installed. It is expected that monocrystalline and multicrystalline solar cells will continue dominating the industry for the next 10 years.

With rejects from the microelectronic industry as feedstock, the PV industry grows monocrystalline ingots by the Czochralski technique. Owing to more relaxed specifications than in microelectronics, the throughput of PV Cz pullers can be increased and still produce high-quality silicon, allowing the achievement of 15 to 17% efficient solar cells.

Tri-crystalline silicon (tri-Si) can be grown in standard production Cz growers using a quasi-continuous pulling with multiple recharging with high productivity and feedstock usage. Tri-Si allows slicing of ultra-thin wafers with higher mechanical yield than monocrystalline Si and obtaining solar cells of similar performance.

Multicrystalline Si can be manufactured at a lower cost than monocrystalline Si, but produces less efficient cells, mainly owing to the presence of dislocations and other crystal defects. With the introduction of new technologies, the gap between multicrystalline and monocrystalline solar cells is reducing.

Si ingots are sliced into thin wafers with multi-wire saws, with throughputs of about 500 to 700 wafers per day and per machine. Sawing is responsible for high material losses, and amounts to a substantial part of the wafer production cost. Understanding the microscopic processes of wire sawing allows optimising the technique and improving the sawing performance.

Silicon ribbon wafer/foil technologies have matured in the past decade to where they are serious contenders for competing on a scale equal to that of conventional silicon wafers. Several ribbon technologies have already demonstrated robust and reproducible processes, which have been scaled up to megawatt levels. The next round of expansion will probably reach the 50 to 100 MW factory size for individual technologies within 5 years. Factories of this scale will require considerable development of automated equipment and infrastructure in addition to continuous improvement on the basic process control and material quality. EFG and SF ribbon technologies are most notably in a position today to manufacture wafers on this scale and lead the anticipated growth of the photovoltaic industry.

During the last decade, computer power has been increasing strongly and there is a high progress in the modelling of physical phenomena and in the development of

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 987

simulation tools. By this, computer simulations become a powerful tool in science and industrial applications. The simulation results of industrial crystallisation processes that are shown and the detailed study of the crystallisation mode by numerical simulations are some examples of the possibilities today. These numerical simulations offer a wide range of possibilities to increase the knowledge about the basic physics of crystallisation and technical processes. One insistent demand on computer simulations is to close the gap between science and engineering to get a closer picture of reality.

6.8 ACKNOWLEDGEMENTS

This chapter has been coordinated by W. Koch. Section 6.2 has been prepared by A.L. Endrös, Section 6.3 by C. Häßler, Section 6.4 has been prepared by H. J. Möller, Section 6.5 by J. P. Kalejs and Section 6.6 by D. Franke. The editor also wants to recognise the work of C. del Cañizo, senior researcher at his institute (IES-UPM), who greatly reduced the original text to keep it under the budgeted extension without losing readability and homogeneity.

REFERENCES

1. Knobloch J *et al*., *Proc. 13th Euro. Conf. Photovoltaic Solar Energy Conversion*, 9 (1995).
2. Li J *et al*., *Appl. Phys. Lett.* **60**, 2240 (1992).
3. Zundel M, Csaszar W, Endrös A, *Appl. Phys. Lett.* **67**, 3945 (1995).
4. Güttler G, Queisser H, *Energy Conversion* **10**, 51 (1970).
5. Martinelli G, Kibizov R, *Appl. Phys. Lett.* **62**, 3262 (1993).
6. Endrös A, Einzinger R, Martinelli G, *Proc. 14th Euro. Conf. Photovoltaic Solar Energy Conversion*, 112 (1997).
7. Palm J *et al*., *Proc. 16th Euro. Conf. Photovoltaic Solar Energy Conversion*, 1222–1225 (2000).
8. Palm J *et al*., *Proc. 28th IEEE Photovoltaic Specialist Conf.*, 40–45 (2000).
9. Carstensen J, Lippik W, Föll H, "Mapping of Defect Related Silicon Bulk and Surface Properties with the ELYMAT Technique", in Huff H, Bergholz W, Sumino K, Eds, *Semiconductor Silicon 94*, 1105–1116, The Electrochemical Society, Pennington, NJ (1994).
10. Schmidt J, Aberle A, Hezel R, *Proc. 26th IEEE Photovoltaic Specialist Conf.*, 13 (1997).
11. Glunz S *et al*., *Proc. 2nd World Conf. Photovoltaic Solar Energy Conversion*, 1343 (1998).
12. Münzer A *et al*., to be published.
13. Endrös A, *Phys. Rev. Lett.* **63**, 70 (1989).
14. Endrös A, *Solid State Phenomena* **32–33**, 143 (1993).
15. Csaszar W, Endrös A, *Phys. Rev. Lett.* **73**, 2 (1994).
16. Csaszar W, Endrös A, *Mater. Sci. Eng.* **B36**, 112 (1996).
17. Koch W *et al*., *Solid State Phenomena* **401**, 57–58 (1997).
18. Häßler C *et al*., *Proc. 2nd World Conf. Photovoltaic Solar Energy Conversion*, 1886 (1998).
19. Koch W, Krumbe W, Schwirtlich I, *Proc. 11th EU Photovoltaic Specialist Conf.*, 518 (1992).
20. Fuller C, Logan R, *J. Appl. Phys.* **28**, 1427 (1957).
21. Kaiser W, Frisch H, Reiss H, *Phys. Rev.* **112**, 1546 (1958).
22. Wagner P, Hage J, *Appl. Phys. A: Solids Surf.* **49**, 123 (1989).
23. Cazcarra V, Zunino P, *J. Appl. Phys.* **51**, 4206 (1980).
24. Pensl G *et al*., *J. Appl. Phys. A* **48**, 49 (1989).
25. Häßler C *et al*., *Proc. 14th Euro. Conf. Photovoltaic Solar Energy Conversion*, 720 (1997).
26. Chonan S, Jiang Z, Yuki Y, *J. Vib. Acoustics* **115**, 529 (1993).
27. Wells R, *Solid State Technol.* **30**, 63 (1987).

28. Sahoo R *et al*., in Subramania K, Ed, *ASME – IMECE Manufacturing Science and Engineering*, 131, ASME press, New York (1996).
29. Li J, Kao I, Prasad V, *ASME – IMECE Manufacturing Science and Engineering*, 439, ASME press, New York (1997).
30. Yang F, Kao J, *J. Electron. Packaging* **121**, 191 (1999).
31. Bhagavat M, Kao I, Prasad V, *ASME J. Tribology* (2001) in press.
32. Kao I, Wie S, Chiang P, *Proc. of NSF Design & Manufacturing Grantees Conf.* 239 (1997).
33. Chen C, Leipold M, *J. Am. Ceram. Soc.* **59**, 469 (1980).
34. Lawn B, Marshall D, *J. Am. Ceram. Soc.* **62**, 347 (1979).
35. Anstis G, Chantikul P, Lawn B, Marshall D, *J. Am. Ceram. Soc.* **64**, 533 (1981).
36. Evans A, Charles E, *J. Am. Ceram. Soc.* **59**, 371 (1976).
37. Lawn B, Evans A, *J. Mater. Sci.* **12**, 2195 (1977).
38. Gogots Y, Baek C, Kirscht F, *Semicond. Sci. Technol.* **14**, 936 (1999).
39. Weppelmann E, Field J, Swain M, *J. Mater. Res.* **8**, 246 (1993).
40. Borst C, Möller H, German BMBF VEDRAS Report, 23 (1999).
41. "Electricity from Solar Cells" *Flat Plate Array Project*, 10 Years of Progress, JPL Publication 400−279 10/85 (Oct. 1985).
42. For comprehensive reviews see bibliographies in *J. Cryst. Growth*, **50** (1980); *J. Cryst. Growth*, **82** (1987); *J. Cryst. Growth*, **104** (1990).
43. Ciszek T, *J. Cryst. Growth* **66**, 655 (1984).
44. Bell R, Kalejs J, *J. Mat. Res.* **13**, 2732 (1998).
45. Chalmers B, *J. Cryst. Growth* **70**, 3 (1984).
46. Thomas P, Brown R, *J. Cryst. Growth* **82**, 1 (1987).
47. Hopkins R *et al*., *J. Cryst. Growth* **82**, 142 (1987).
48. Ravi K, *J. Cryst. Growth* **39**, 1 (1977).
49. Kalejs J, in Khattak C, Ravi K, Eds, *Silicon Processing for Photovoltaics II*, 185−254, North Holland, Amsterdam (1987).
50. Garcia D *et al*., *J. Cryst. Growth* **225**, 566 (2001).
51. Mackintosh B *et al*., *Proc. 28th IEEE Photovoltaic Specialist Conf.*, 46 (2000).
52. Sachs E, in *Proc. of the Flat-Plate Solar Array Project Research Forum on the High-Speed Growth and Characterisation of Crystals for Solar Cells*, JPL, 279 (1984).
53. Sachs E, Ely D, Serdy J, *J. Cryst. Growth* **82**, 117 (1987).
54. Wallace R *et al*., Sixth Workshop on the Role of Impurities and Defects in Silicon Device Processing – Extended Abstracts and Papers, NREL/SP-413-21550, 203 (1996).
55. Lange H, Schwirtlich I, *J. Cryst. Growth* **104**, 108 (1990).
56. Cotter J *et al*., *Proc. 13th Euro. Conf. Photovoltaic Solar Energy Conversion*, 1732 (1995).
57. Barnett A *et al*., *Proc. 18th IEEE Photovoltaic Specialist Conf.*, 1094 (1985).
58. Ruby D, Ciszek T, Sopori B, *NCPV Program Review Meeting*, April, 16−19, Denver, CO, unpublished.
59. Schmidt W, Woesten B, Kalejs J, *Proc. 16th EPVSEC*, 1082−1086 (2000).
60. Sawyer W, Bell R, Schoenecker A, *Solid State Phenomena* **37−38**, 3 (1994).
61. Meier D, Hopkins R, Campbell R, *J. Propulsion Power* **4**, 586 (1988).
62. Koch W *et al*., *Proc. of 2nd World Conf.* 1254 (1998).
63. Koshka Y *et al*., *Appl. Phys. Lett.* **74**, 1555 (1999).
64. Bowler D, Wolf M, *IEEE Trans. Components, Hybrids Manufacturing Technology* **3**, 464 (1980).
65. Basore P, Clugston D, PC1D Version 4.2 for Windows, Copyright by University of New South Wales.
66. Bathey B *et al*., *Proc. 28th IEEE Photovoltaic Specialist Conf.*, 194 (2000).
67. Sims P *et al*., *Annual Report on DOE/NREL Subcontract No. DE-AC36-98-GO20337*, Publication No. NREL/SR-520-28547 (2000).
68. Hahn G *et al*., *Proc. 16th EPVSEC*, **1095** (2000).

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 989

69. Sana P, Rohatgi A, Kalejs J, Bell R, *Appl. Phys. Lett.* **64**, 97 (1994).
70. Bailey J, Kalejs J, Keaveny C, *Proc. 24th IEEE Photovoltaic Specialist Conf.*, **1356** (1994).
71. The earliest work is reported in the 3rd *Jpn. Photovoltaic Solar Energy Conversion* by Morita H *et al*., *Jpn. J. Appl. Phys.* **21** 47 (1982); U.S. patent No. 4,640,001 (1987) and Australian patent No. 609424 (1991); a comprehensive review of recent developments is given by Szlufcik J *et al*., *Proc. E-MRS Spring Meeting*, E-VI.1 (2001).
72. See: *www.abaqus.com*.
73. See: *www.ansys.com*.
74. See: *www.marc.com*.
75. Lambropoulos J *et al*., *J. Cryst. Growth* **65**, 324–330 (1983).
76. Kalejs J, Schmidt W, *Proc. 2nd World Conf. Photovoltaic Solar Energy Conversion*, 1822–1825 (1998).
77. Wallace R, Janoch R, Hanoka J, *Proc. 2nd World Conf. Photovoltaic Solar Energy Conversion*, 1818–1821 (1998).
78. Koch W, Höfs H, Häßler C, Thurm S, *Proc. 2nd World Conf. Photovoltaic Solar Energy Conversion*, 1254–1259 (1998).
79. Chang C, Brown R, *J. Cryst. Growth* **63**, 343–352 (1983).
80. Zulehner W, Huber D, *Crystals*, Vol. 8, 3–143, Springer Verlag, Berlin, Heidelberg, New York (1982).
81. Palm J *et al*., *Proc. 16th Euro. Conf. Photovoltaic Solar Energy Conversion*, 1222–1225 (2000).
82. Helmreich D, "*The Wacker Ingot Casting Process*", in Khattak C, Ravi K, Eds, *Silicon Processing for Photovoltaics*, Vol. II, 97–115, North-Holland, Amsterdam (1987).
83. Schätzle P *et al*., *Proc. 11th Euro. Conf. Photovoltaic Solar Energy Conversion*, 465–468 (1992).
84. Khattak C, Schmid F, *Proc. 25th IEEE Photovoltaic Solar Energy Conversion*, 597–600 (1996).
85. Koch W *et al*., *Proc. 12th Euro. Conf. Photovoltaic Solar Energy Conversion*, 797–798 (1994).
86. Häßler C *et al*., *Proc. 2nd World Conf. Photovoltaic Solar Energy Conversion*, 1886–1889 (1998).
87. Zabaras N, Ruan Y, Richmond O, *Computer Methods in Applied Mechanics and Engineering*, Vol. 8, 333–364, Elsevier Science Publisher B.V., Amsterdam (1990).
88. Diemer M, Franke D, "*Modelling of Thermal Stress Near the Liquid-Solid Phase Boundary*", in Thomas B, Beckermann C, Eds, *Modelling of Casting, Welding and Advanced Solidification*, Vol. 8, 907–914, TMS (1998).
89. Zienkiewicz O, Taylor R, *The Finite Element Method*, McGraw-Hill, London (1989).
90. Bathe K, *Finite Element Procedures in Engineering Analysis*, München Verlag, Wien (1984).
91. Kurz W, Fischer D, *Fundamentals of Solidification*, Trans Tech Publications, Zürich (1998).
92. Franke D *et al*., *Proc. 25th IEEE Photovoltaic Specialist Conf.*, 545–548 (1996).
93. Häßler C *et al*., *Solid State Phenomena* **67–68**, 447–452 (1999).
94. Franke D, Apel M, Häßler C, Koch W, *Proc. 16th Euro. Conf. Photovoltaic Solar Energy Conversion*, 1317–1320 (2000).
95. Steinbach I *et al*., *Physica D* **94**, 135–147 (1996).
96. Steinbach I, Höfs H, *Proc. 26th IEEE Photovoltaic Solar Energy Conversion*, 91–93 (1997).

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 990

7

Crystalline Silicon Solar Cells and Modules

Ignacio Tobías[1], Carlos del Cañizo[1] and Jesús Alonso[2]

[1]*Universidad Politécnica de Madrid, Madrid, Spain,* [2]*Isofotón S.A., Málaga, Spain*

7.1 INTRODUCTION

Crystalline silicon solar cells and modules have dominated photovoltaic (PV) technology from the beginning. They constitute more than 85% of the PV market today, and although their decline in favor of other technologies has been announced a number of times, they presumably will retain their leading role for a time, at least for the next decade.

One of the reasons for crystalline silicon to be dominant in photovoltaics is the fact that microelectronics has developed silicon technology greatly. On the one hand, not only has the PV community benefited from the accumulated knowledge but also silicon feedstock and second-hand equipment have been acquired at reasonable prices. On the other hand, Microelectronics has taken advantage of some innovations and developments proposed in Photovoltaics.

For several decades, the terrestrial PV market has been dominated by *p*-type Czochralski silicon substrates. Continuous improvements in performance, yields and reliability have allowed an important cost reduction and the subsequent expansion of the PV market. Because of the lower cost of mc-Si wafers, multicrystalline (MC) silicon cells emerged in the 1980s as an alternative to single-crystal ones. However, their lower quality precluded the achievement of similar efficiencies to those of Cz, so that the figure of merit \$ W^{-1} has been quite similar for both technologies over a long time (see Table 7.1).

Deeper understanding of the physics and optics of the mc-Si material led to improved device design, which allowed a wider spread of the technology. A combination

Handbook of Photovoltaic Science and Engineering. Edited by A. Luque and S. Hegedus
© 2003 John Wiley & Sons, Ltd ISBN: 0-471-49196-9

Table 7.1 Breakdown of costs of fabrication of single-crystalline (SX) and multicrystalline (MC) solar cells (corresponding to year 1990) [1]

Item	SX	MC
Pure Si	38	38
Ingot formation	115	35
Sawing	77	77
Wafer cost	**230**	**150**
Cell fabrication	80	80
Total components	310	230
Yield	0.95	0.9
Cell cost	**326**	**256**
Module assembling	75	75
Lamination	75	75
Module cost (Euro m^{-2})	**476**	**406**
Efficiency	0.14	0.12
Module cost (Euro Wp^{-1})	**3.40**	**3.38**

Table 7.2 Market share of monocrystalline and multicrystalline solar cells [2]

Year	Cz-Si solar cells		MC-Si solar cells	
	Output [MW]	Market share [%]	Output [MW]	Market share [%]
1996	48.7	55	28.4	32
2000	92.0	32	146.7	51

of improved material quality and material processing has allowed higher efficiencies at a still lower cost, increasing the share of MC in the PV market, well ahead of monocrystalline. Recent evolution of the market can be seen in Table 7.2 [2].

This chapter offers an overview of silicon solar cell and module technology. First, Si properties justifying its use as photovoltaic material are presented. Then, design of Si solar cells is reviewed, highlighting the benefits and limits of different approaches. Manufacturing processes are described, paying special attention to technologies that are currently implemented at the industrial level, mostly based on screen-printing metallization technology. Considerations of ways of improving solar cell technology are also specified. Peculiarities of multicrystalline material are explained in Section 7.6, while other approaches that are already in industrial production are also described briefly. Next, crystalline Si modules are reviewed, pertaining to electrical performance, fabrication sequence and reliability concerns.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 992

7.2 CRYSTALLINE SILICON AS A PHOTOVOLTAIC MATERIAL

7.2.1 Bulk Properties

Crystalline silicon has a fundamental indirect band gap $E_G = 1.17$ eV and a direct gap above 3 eV [3] at ambient temperature. These characteristics determine the variation of optical properties of Si with wavelength, including the low absorption coefficient for carrier generation for near band gap photons [4]. At short ultraviolet (UV) wavelengths in the solar spectrum, the generation of two electron–hole pairs by one photon seems possible, though quantitatively this is a small effect [5]; at the other extreme of the spectrum parasitic free-carrier absorption competes with band-to-band generation [6]. The intrinsic concentration is another important parameter related to the band structure; it links carrier disequilibrium with voltage [7].

At high carrier densities, doping- or excitation-induced, the band structure is altered leading to an increase in the effective intrinsic concentration: this is one of the so-called heavy doping effects that degrade the PV quality of highly doped regions [8].

Recombination in Si is usually dominated by recombination at defects, described with Shockley–Read–Hall (SRH) lifetimes. The associated lifetime τ (which can also be described in terms of diffusion length L) increases for good quality materials. Auger recombination, on the contrary, is a fundamental process that becomes important at high-carrier concentration [9]. The Auger coefficients are reported to be higher at moderate carrier densities due to excitonic effects [10]. Band-to-band direct recombination is also a fundamental process but quantitatively negligible (it is instructive, however, to notice that record-efficiency solar cells have such extraordinarily low SRH recombination levels that they perform as 1%-efficient light-emitting diodes, that is, radiative recombination is significant [11]).

At low and moderate doping, electrons present mobilities about three times higher than holes, both limited by phonon scattering. Impurity scattering dominates for higher doping densities [12, 13]. Carrier–carrier scattering affects transport properties in highly injected material [14].

7.2.2 Surfaces

7.2.2.1 Contacts

Contacts are structures built on a semiconductor surface that allow charge carriers to flow between the semiconductor and the external circuit. In solar cells, contacts are required to extract the photogenerated carriers from the absorbing semiconductor substrate. They should be selective, that is, should allow one type of carrier to flow from Si to metal without energy loss while blocking the transport of carriers of the opposite type.

Direct Si-metal contacts, in general, do not behave this way. As an exception, good hole contacts to highly doped p-Si substrates with aluminum are possible. But the

most used approach is to create heavily doped regions under the metal, *p*-type for hole extraction and *n*-type for electron extraction. Majority carriers in this region can flow through the contact with low voltage loss. The transport of minority carriers is described by a surface recombination velocity (SRV), S. Although the SRV is high, limited only by thermal diffusion so that $S \cong 10^6$ cm·s^{-1} [15], the concentration of minority carriers, for a given *pn* product, is suppressed by the high doping and the flow is reduced.

As will be seen later on, the contact for the minority carriers is usually placed at the front (illuminated) face of the substrate, and the corresponding heavily doped layer is usually called *emitter*. The doped region under the majority carrier contact at the back is called a *Back Surface Field* (BSF).

Recombination at these heavily doped regions is described by the saturation current density J_0 that includes volume and true contact recombination. Their thickness w should be much higher than the minority-carrier diffusion length L so that few excess carriers reach the contact, and the doping level must be very high to decrease contact resistance and the minority-carrier concentration, although heavy doping effects may limit the doping level advisable for such regions. The recombination activity of BSF layers is often described in terms of an effective SRV instead of the saturation current density.

Typical 10^{-13} to 10^{-12} A·cm^{-2} J_0 values are achieved [16, 17]. Diffused phosphorus is used for *n*-contacts. Aluminum alloying has the advantages over boron for *p*-contacts that very thick p^+ layers can be formed in a short time at moderate temperatures and that gettering action is achieved [18]. As a shortcoming, the p^+ layer is nonhomogeneous and can even be locally absent; the obtained J_0 is larger than expected for a uniform layer. In comparison to aluminum, boron offers higher doping levels because of a larger solubility [19] and transparency to light so that it can also be used at illuminated surfaces.

Other structures have been tested to obtain selective contacts: metal-insulator-semiconductor (MIS) contacts [20], polysilicon contacts [21] and heterojunction to a-Si or other wide band gap material [22].

7.2.2.2 Noncontacted surfaces

Because of the severe alteration of the bonding of Si atoms, a large number of band gap states exist at a bare Si surface which, acting as SRH recombination centers, make the SRV very large, around 10^5 cm·s^{-1} [23, 24]. In order to reduce surface recombination, two main approaches are followed [25].

In the first one, the density of electron surface states in the gap is decreased. This is accomplished by depositing or growing a layer of an appropriate material that partially restores the bonding environment of surface Si atoms. This material must be an insulator.

Thermal SiO_x is grown in an oxygen-rich atmosphere at the expense of substrate Si atoms at high temperatures around 1000°C. SiN_x is deposited by plasma-enhanced chemical vapor deposition (PECVD) [26] at low temperatures between 300 and 400°C range. The quality of both techniques is very sensitive to subsequent treatments, with hydrogen playing a major role in obtaining low SRV values below 100 cm·s^{-1}.

As a general rule, S increases with the doping of the substrate [23, 24]. It also depends on the injection level and doping type, because the interfaces contain positive

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 994

charges that affect the number of carriers at the surface, and because the capture probability for electrons and holes is different [27]. N-type or intrinsic surfaces are usually better than p-type ones. Stability under UV exposure is another fundamental concern.

In the second approach, the excess carrier density at the interface is decreased relative to the bulk. This effect has already been commented upon with respect to contacts and results in a reduced effective SRV at the bulk edge of the corresponding space charge region. It can be produced by charges in the surface layer, by the electrostatics associated to a MOS structure [28] or by doping.

The surface layer can be accumulated or inverted, or, correspondingly, doped with the same or the opposite type of the substrate. Its recombination activity is better described by a constant saturation current density, J_0, whose minimization follows the same rules as mentioned for the contacts if S at the surface is high. On the contrary, if S is low when compared to D/L for minority carriers (with D the diffusion constant and L the diffusion length in the substrate) in the surface layer, it is better for it to be thin or "transparent" to minority-carrier flow ($w < L$). The optimum doping level is a compromise between reduction of the excess carriers, and heavy doping effects and the increase of the SRV with doping. Moderate doping levels are favored in this case. Under passivated surfaces, J_0 values around 10^{-14} A·cm^{-2} are achievable, with phosphorus-doped substrates giving better results than boron-doped ones [16, 17].

In conclusion, recombination at a noncontacted surface can be made much smaller than at a metallized one, and this has a deep influence in the evolution of Si solar cell design.

7.3 CRYSTALLINE SILICON SOLAR CELLS

7.3.1 Cell Structure

Several studies have been carried out to find the limiting efficiency and the optimum structure of a-Si solar cell [29–31]. All avoidable losses are assumed to be suppressed:

1. no reflection losses and maximum absorption as achieved by ideal light-trapping techniques,
2. minimum recombination: SRH and surface recombination are assumed avoidable and only Auger recombination remains,
3. the contacts are ideal: neither shading nor series resistance losses,
4. no transport losses in the substrate: the carrier profiles in the substrate are flat so that recombination is the minimum possible for a given voltage.

The optimum cell should use intrinsic material, to minimize Auger recombination and free-carrier absorption, and should be around 80-μm thick, the result of the trade-off between absorption and recombination. It could attain nearly 29% efficiency at one sun AM1.5 Global, 25°C [30].

This ideal case does not tell us where to put the contacts. To realize Condition 4 mentioned above, the contacts should be located at the illuminated or the front face, closest to photogeneration (Figure 7.1a). Because of metal shading losses this threatens

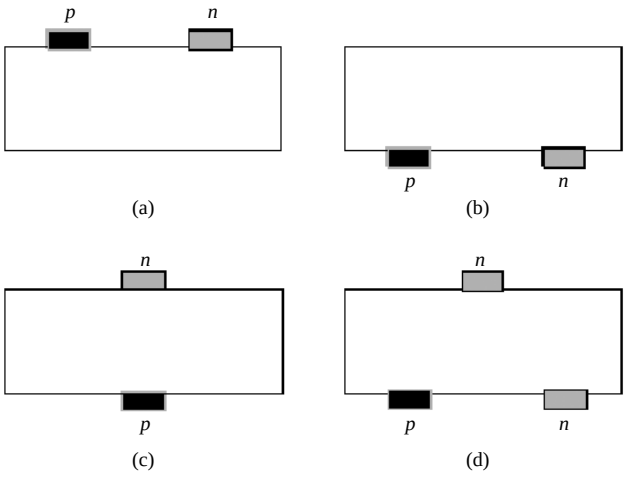


Figure 7.1 Contacting structures: (a) both contacts at the front; and (b) at the back; (c) both faces contacted; and (d) one carrier extracted at both faces. The structures with interchanged *n*- and *p*-types are also possible

Condition 3 mentioned above. This solution is being considered for concentration [32]. Putting both contacts at the back works the other way round (Figure 7.1b). Back-contacted cells exhibit record efficiencies under concentration and demonstrate very high values near 23% at one sun [33].

In most cells, each contact is placed on a different face, which is technologically simpler (Figure 7.1c). Minority carriers in the substrate are usually collected at the front since their extraction is more problematic because of their low density. The diffusion length describes the maximum distance from where they can be collected. Majority carriers can drift to the back contact with low loss. Several designs extract minority carriers both at the front and at the back (Figure 7.1d) [34], thus increasing the volume of profitable photogeneration.

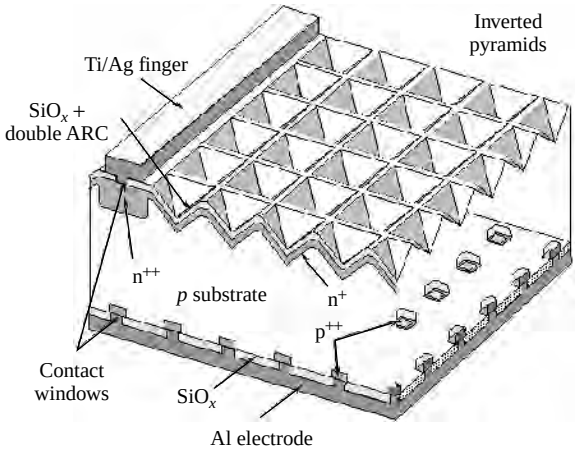
Bifacial cells are designed to collect light incident at both faces, which allows a boost in output power if there is a significant albedo component. They can be implemented with any structure in Figure 7.1 on the condition that both surfaces allow light through [35].

Both 24.7% record-efficiency laboratory solar cells [36] (Figure 7.2a) and mass-produced industrial devices (Figure 7.2b), typically 15% efficient, display the contacting structure in Figure 7.1(c). They will be described in the paragraphs that follow with the aim of illustrating the amplitude and the reasons for the performance gaps between the ideal and the best Si cell, and between this cell and industrial cells.

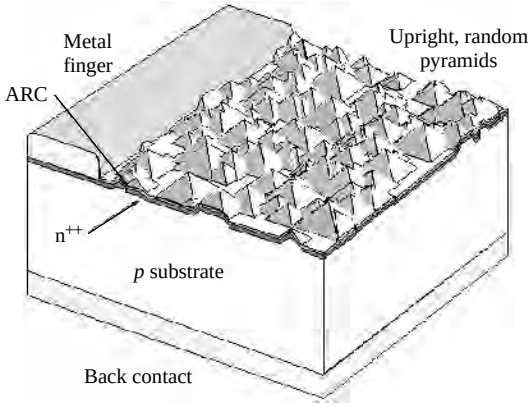
7.3.2 Substrate

7.3.2.1 Materials and processing

Highest efficiencies are achieved with monocrystalline float zone (FZ-Si) material, which in addition to extreme crystalline perfection shows the lowest contamination levels of both



(a)



(b)

Figure 7.2 (a) Passivated emitter and rear locally diffused (PERL) cell; and (b) industrial cell with screen-printed contacts. (Not to scale)

metallic and light (O, C, N) impurities. This translates into the longest post-processing SRH lifetimes in the millisecond range, but still shorter than the Auger limit. Magnetic Czochralski (MCz) material contains much less oxygen than conventional Czochralski and also allows very high efficiencies to be obtained [36].

Industrial cells use Czochralski (Cz-Si) wafers because of their availability. Cz wafers are also perfect crystals but they contain a high concentration of oxygen that affects lifetime in several ways [37]. Some commercial devices are made on multicrystalline (mc-Si) substrates grown in blocks or ribbons with procedures specially developed for photovoltaics. In addition to crystal defects, such as grain boundaries and dislocations, the potential content of metallics is higher because of lower segregation to the melt during the faster solidification process. As a result, the lifetime of mc-Si is lower.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 997

But lifetime is important at the end of solar cell fabrication during which it can undergo strong variations. This issue is handled in different ways in a laboratory and a factory environment. In the laboratory, measures are taken to maintain long initial lifetimes by avoiding contamination during high-temperature steps: furnace cleaning, ultra pure chemicals and so on. In a rough, industrial environment and with defect-containing (Cz- and mc-Si) materials, the problem is more complex: in addition to contamination from the surroundings, impurities and defects in the substrate move, interact and transform at high temperature. The solution is to integrate gettering steps [38] in the fabrication flow that reduce the impact of contamination, and to tailor the thermal treatments to the peculiarities of the material. Final substrate lifetimes of industrial cells range from 1 to 10 μs.

Gettering techniques eliminate or reduce contaminant impurities in a wafer, and so neutralize the effect of lifetime reduction. Although gettering processes are not always well understood, it is admitted, in general, that in a gettering process a sink region is formed, which is able to accommodate the lifetime-killing impurities in such a way that they are not harmful to the device being manufactured, or at least they are where they can be easily removed.

In solar cell fabrication, we take advantage of the fact that phosphorus and aluminum diffusions, appropriate candidates for emitter and BSF layers, respectively, produce gettering in certain conditions [39]. Other techniques have been explored [40, 41], but their integration in a solar cell process is not so straightforward.

P gettering effect has been proved for a wide variety of P diffusion techniques (spin-on, $POCl_3$, PH_3 etc), provided diffusion is done in supersaturation conditions (i.e. over its solid solubility in silicon). Unfortunately, this leaves a “dead layer” of electrically inactive phosphorus near the surface, which reduces UV response of the cells in case it is not etched away [42]. Another phenomenon related to this supersaturated P is the injection of silicon self-interstitials to the bulk of the material, which is responsible for an enhancement of the gettering effect [43].

When Al is deposited on Silicon (by different techniques such as sputtering, vacuum evaporation or screen printing) and annealed over the eutectic temperature (577°C), a liquid Al-Si layer is formed, where impurities tend to segregate because of their enhanced solubility [44]. They will remain in this gettering layer while cooling, so that bulk lifetime will improve after the process.

Another approach to improve material quality is “bulk passivation”, a treatment with hydrogen, for example, during SiN_x deposition, to which some defected materials respond very well [45].

7.3.2.2 Doping level and type

Both laboratory record-efficiency and industrial cells use boron-doped substrates. Rather than fundamental advantages [46], there are practical (properties of P diffusion, easiness of Al alloying) and historical reasons for this preference [47]. However, the situation may change once the role of boron in the degradation of Cz-Si cells under illumination has been established [48].

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 998

The optimum substrate doping depends on the cell structure and dominant recombination mechanism. Though intrinsic substrates present the advantage of highest Auger-limiting lifetimes, higher doping is favored when SRH recombination is present, since recombination is proportional to the excess density that decreases, for a given voltage, as doping increases [29]. This is balanced with a reduction of the lifetime itself.

A high doping also helps in minimizing the series resistance losses associated to the transport of carriers to the back face in thick cells with the majority carrier contact at the back.

Doping levels in the 10^{16} cm^{-3} range are found in the substrate of industrial cells. Very high efficiencies have been obtained with both low (1 Ω·cm for PERL cells) and high substrate resistivities, as in the point-contact cells [33].

7.3.2.3 Thickness

From the point of view of electrical performance, the choice of the optimum substrate thickness also depends on the structure and the quality of the materials and involves several considerations. In cells with diffusion lengths longer than the thickness, the most important issue is surface recombination: if S at the back is higher than D/L for the minority carriers in the substrate (around 250 $cm \cdot s^{-1}$ for the best cells), thinning the cell increases recombination at a given voltage, and vice versa. Thinner cells always absorb less light as well, which is attenuated by light-trapping techniques. Passivated emitter and rear locally diffused (PERL) cells were reported to improve when going from 280- to 400-μm thickness because of a (relatively) high rear surface recombination and nonideal light-trapping [49].

The losses associated with the transport of carriers extracted at the nonilluminated face decrease with thinning: in conventional structure cells; this leads to decreased series resistance. In back-contacted cells, both types of carriers benefit from thinning and the trade-off in absorption leads to lower w values, around 150 to 200 μm.

The diffusion length of industrial cells (around 100 μm) is generally lower than thickness. These cells are rather insensitive to thinning because they collect only the generation near the contact and are not affected by rear surface recombination. The driving criteria are cost and fabricability. A thickness in the 200- to 300-μm range is usually employed but there is a clear trend toward thinner wafers for saving expensive silicon material [47], and advanced wafering techniques and procedures to process very thin, large-area substrates without breaking are being developed: light-trapping and back recombination will become increasingly important.

7.3.3 The Front Surface

7.3.3.1 Metallization techniques

Metal grids are used at the front face to collect the distributedly photogenerated carriers. The compromise between transparency and series resistance requires metallization technologies able to produce very narrow but thick and highly conductive metal lines with

a low contact resistance to Si. Laboratory cells use photolithography and evaporation to form 10- to 15-μm metal fingers. Ti/Pd/Ag structures combine low contact resistance to *n*-Si and high bulk conductivity. These processes are not well suited to mass production that relies on thick-film technologies. Ag pastes are screen printed, resulting in over 100-μm wide lines with higher bulk and contact resistance. Laser-grooved buried-grid (LGBG) industrial cells implement a finer metallization technology [50], which will be presented in Section 7.7.3.

Coarser metallization techniques imply higher shading and resistance losses, and restrict the efficiency enhancement that could be achieved by internal cell design.

7.3.3.2 Homogeneous emitters

Under the metal lines, the substrate must be heavily doped to make the contact selective. Usually the doped region, the emitter, extends all across the front surface, acting as a "transparent electrode" by offering minority carriers in the substrate a low resistance path to the metal lines.

When the exposed surface is not passivated (Figure 7.3a), the emitter should be as thin as possible, because the high SRV makes the light absorbed in this region poorly collected, and also highly doped to decrease recombination. On the other hand, a sufficient low-sheet resistance is to be achieved. The solution is to make very thin and highly doped emitters.

If the surface is passivated (Figure 7.3b), the collection efficiency of the emitter can be raised by lowering the doping level thus avoiding heavy doping and other detrimental

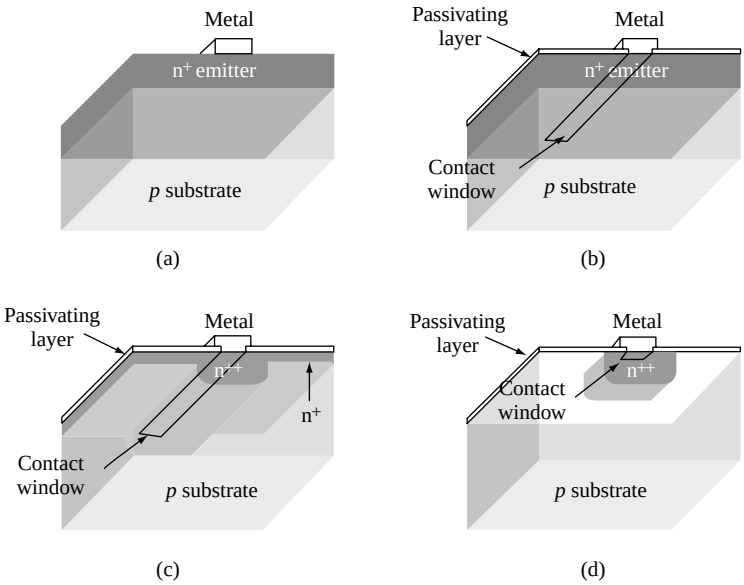


Figure 7.3 Different emitter structures: (a) homogenous emitter without surface passivation; (b) homogenous emitter with surface passivation; (c) selective emitter; and (d) localized emitter

effects. This must be balanced with contact resistance. Etching off of the passivating layer before metallization is usually needed (not in screen-printed cells). To maintain low sheet resistance and diminish recombination at the metallized fraction, the emitter is deep (around 1 μm). Note that the collection of carriers near the surface implies that the emitter is thin in terms of the minority-carrier diffusion length ($w < L$) and so it is very sensitive to surface recombination. Recombination is further reduced, by making the contact window narrower than the finger width, as illustrated in Figure 7.3 [51].

Control of both the surface concentration and the depth of the emitter, is achieved by depositing, in a thermal step, the desired amount of phosphorus or boron (predeposition) and then diffusing it into the substrate (drive-in) during subsequent furnace steps. The MIS solar cell, on the contrary, uses no diffusion for the *n* region, which is electrostatically induced by charges on top of the surface [20].

The J_0 of the emitter is the average, weighted by the contacted area, of the J_0 of contacted and noncontacted portions.

7.3.3.3 Selective and point emitters

A further improvement involves making separate diffusions for the different regions since the requirements are so different (Figure 7.3c) [52]: a heavily doped and thick region under the contacts, a thin and lowly doped region under the passivating layer. These structures, known as "selective emitters", come at the expense of more complicated processing usually involving photolithographic delineation and alignment of the diffusions (not in LGBG cells to be seen later).

If a very low SRV is possible, it would be best to have no emitter at all since doping always degrades bulk lifetime (Figure 7.3d). Examples are the back point-contact solar cell and the point-emitter design with bifacial contact [53], originally designed for concentration but capable of very high one-sun efficiency as well.

With localized contacts, surface recombination decreases with the penalty of an increase in transport losses in the substrate: deeper gradients for minority carriers, or increased series resistance for majority carriers, because of current crowding near the contacts. The trading is more favorably solved as the contact size shrinks [54]. Light and/or localized diffusions also have the drawback of decreased gettering action.

7.3.3.4 Industrial cells

Screen printing drastically affects the design of the emitter: it must be very highly doped to decrease the high-contact resistance and not very shallow so that it is not perforated during paste firing, which would short-circuit the junction [55]. Besides, the wide metal lines must be placed well apart; and in order to keep shading losses moderate, the emitter lateral conductance must be high, which also advises deep and highly doped regions. These characteristics are good to decrease recombination at the contacts, but are far from optimum at the exposed surface.

Industrial phosphorus emitters, typically, feature surface concentrations over 10^{20} cm^{-3} and 0.4-μm depth, and result in a sheet resistance around 40 Ω. As already

mentioned, the very highly doped region exhibits almost no photovoltaic activity due to the presence of precipitates ("dead layer"). As a result, the collection of short-wavelength light is very poor and the J_0 large, irrespective of a hypothetical surface passivation is therefore not implemented. The advantage is that heavy phosphorus diffusions produce very effective gettering. Some ways of incorporating selective emitters to screen-printed cells are being considered, SiN_x appearing very well suited for surface passivation [55]. This, however, must be accompanied by a decrease in the finger width so that lower sheet resistances are tolerated.

7.3.4 The Back Surface

The majority carrier, back *p*-contact of industrial solar cells is usually made by printing, and subsequently, firing an aluminum-containing Ag-conducting paste.

The p^+ layer would be useful in decreasing contact recombination, as explained, but this is immaterial to the electrically thick ($w > L$) present industrial cells and it is not optimized for this purpose.

For high-efficiency cells, contact passivation is essential. A BSF is a first step (Figure 7.4a). Localized contacts, as shown in Figure 7.4(b), further reduce recombination. This structure is presented by some of the bifacial cells [56].

If the surface passivation is good, the BSF is restricted to point contacts, some microns in size, as in PERL and similar cells (Figure 7.4c) [57]. The back face of bifacial cells passivated with SiN_x is shown in Figure 7.4(d).

A shallow and light diffusion helps in decreasing surface recombination (Figure 7.4e). The diffusion can be the same type of the substrate, or the opposite one: so-called PERT (Passivated emitter rear totally diffused) [36] and PERF (Passivated emitted rear floating junction) [58] cells demonstrate these concepts. The latter structure benefits from the lower J_0 values of n^+ layers, and it is essential that no electron flow is injected from the *n* region to the *p* contact: the junction must be in open-circuit (a "floating" junction). Note that this structure has nothing to do with structures sometimes found at the back of industrial cells (Figure 7.4f) where a coarse metal mesh is fired through the parasitic n^+ layer formed during front diffusion. A mesh is preferred to a continuous layer for mechanical reasons. The junction is shorted and presents to the substrate a high SRV.

7.3.5 Size Effects

Substrate edges are highly recombining surfaces that adversely affect cell performance, especially for small size, large diffusion length devices. For laboratory cells, efficiency is defined on the basis of a design area. The emitter is limited to it by planar masking or mesa etching. The true edge is thus placed far away from the cell limit, and then recombination is reduced. For real applications, on the contrary, only the substrate area counts, and edge optimization is more complex. Advanced passivation schemes such as edge diffusions are being considered [59, 60]. In large industrial cells, this recombination is much less important.

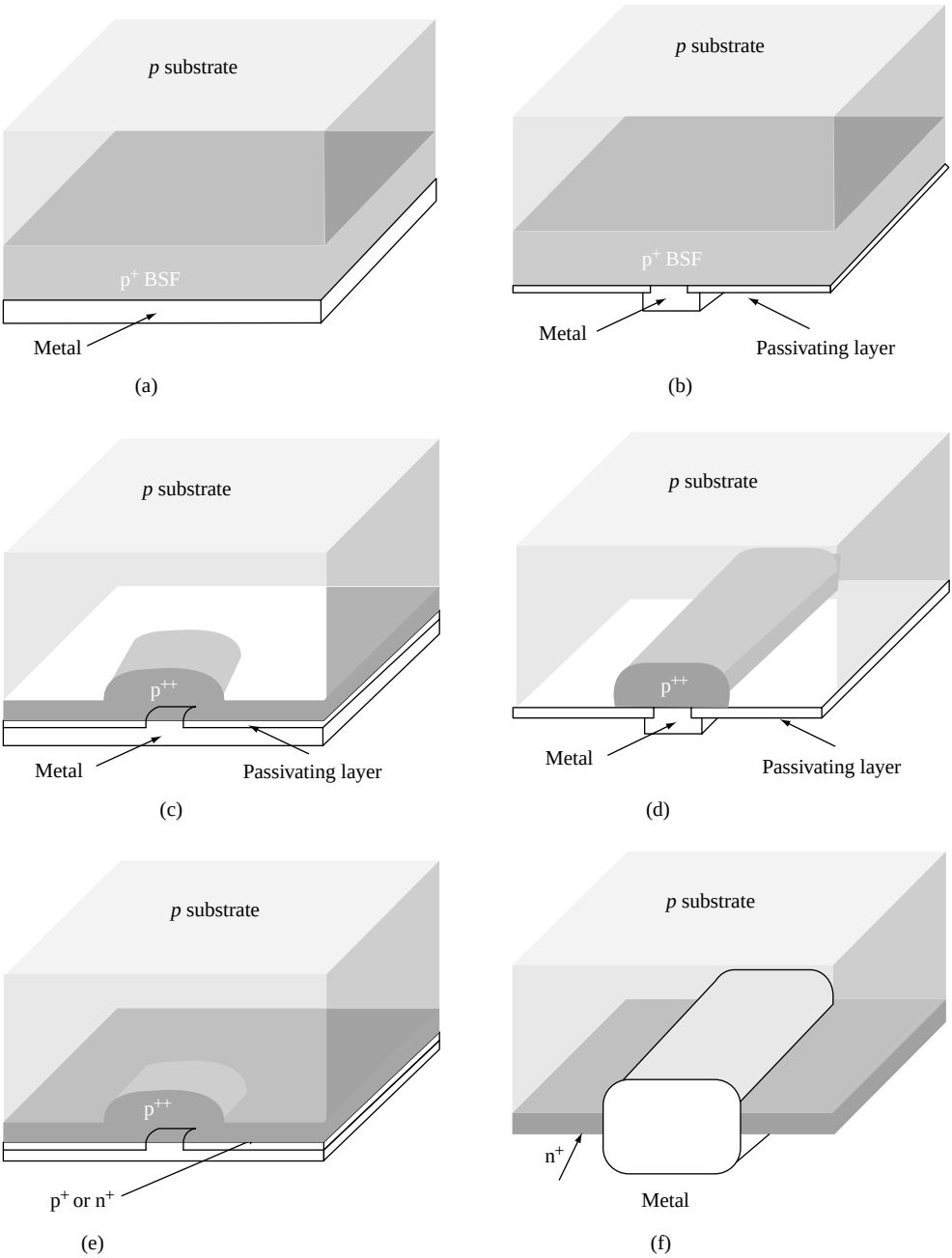


Figure 7.4 Rear contact structures: (a) continuous BSF; (b) bifacial cell; (c) local BSF; (d) local BSF, bifacial cell; (e) selective emitter or floating junction passivation; and (f) shorted junction at the back face of industrial cells

Large cell sizes are preferred by the industry, 10×10 cm^2 or 12.5×12.5 cm^2 being standard. Apart from fabricability concerns, a bigger cell means that more current must be collected at the terminals, making Joule losses grow: the longitudinal resistance of the metal lines increases quadratically with their length. This problem, more severe for coarser metallization techniques, decreases efficiency with increasing size. To alleviate series resistance at the price of increased shading, terminals are soldered to metal bus bars inside the cell's active area, thus decreasing the distance from where current must be collected along the fingers.

7.3.6 Cell Optics

Flat plate solar cells in operation are illuminated from a large portion of the sky, not only because of the isotropic components of radiation, but also because of the sun's apparent motion over the day and the year. So, with regard to angular distribution, these cells must accept light from the whole hemisphere. The spectral distribution also varies with time, weather conditions, and so on. For calibration purposes, a standard spectral distribution AM1.5 Global is adopted as a representative condition, usually specified at 0.1 W·cm^{-2}.

A solar cell should absorb all useful light. For nonencapsulated cells, the first optical loss is the shading by the metal grid at the illuminated face, if any. This loss can amount to more than 10% for industrial cells while for laboratory cells using fine metallization it is much lower. Though several techniques have been proposed to decrease the effective shading, such as shaped fingers, prismatic covers, or cavities [61], their efficacy depends upon the direction of light and so they are not suited to isotropic illumination.

7.3.6.1 Antireflection coatings

Next, loss comes from the reflectance at the Si interface, more than 30% for bare Si in air due to its high refraction index. A layer of nonabsorbing material with a lower refraction index (n_{ARC}) on top of the Si substrate decreases reflectance: this is a step toward the zero-reflection case of a smoothly varying refraction index. If the layer is thick in terms of the coherence length of the illumination, around 1 μm for sunlight, there are no interference effects inside it. The encapsulation (glass plus lamination) belongs to this category.

Antireflection coating (ARC) means an optically thin dielectric layer designed to suppress reflection by interference effects. Reflection is at a minimum when the layer thickness is (an odd multiple of) $n_{\mathrm{ARC}}\lambda_0/4$, with λ_0 the free space wavelength, since in this case reflected components interfere destructively. At other wavelengths reflection increases, but is always below the value with no ARC or, at most, equal [62]. The ARC is usually designed to present the minimum at around 600 nm, where the flux of photons is a maximum in the solar spectrum. For reflection to become zero at the minimum, the coating index should be the geometric average of those of air and silicon, that is, 2.4 at 600 nm for nonencapsulated cells.

The industry uses TiO_x deposited by chemical vapor deposition (CVD). PECVD SiN_x is very interesting since it also serves as a passivating layer, as explained earlier.

By using double layer coatings with $\lambda/4$ design, with growing indices from air to silicon, the minimum in reflection is broader in wavelength. Evaporated SZn and MgF_2

are used in laboratory high-efficiency cells. The low index of passivating SiO_x in contact with Si degrades the performance of the ARC. The SiO_x layer is then made as thin as is compatible with efficient passivation [63].

7.3.6.2 Texturing

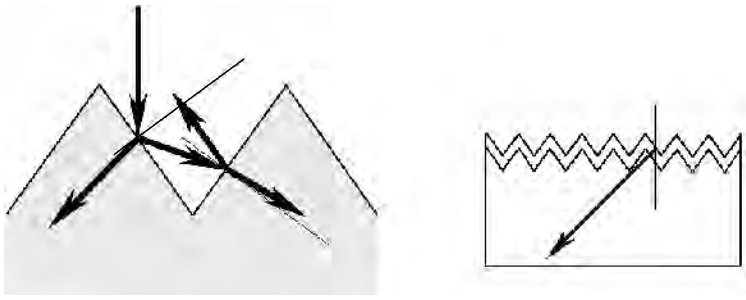
Alkaline (KOH or NaOH-based) solutions etch anisotropically a-Si crystal exposing {1 1 1} planes on which the etching rate is lowest. On [1 0 0]−oriented wafers, randomly distributed, square-base pyramids are formed, whose size is adjusted to a few microns by controlling etching time and temperature. In a textured face, a ray can be reflected toward a neighboring pyramid (Figure 7.5a) and hence absorption is enhanced. Though calculation of reflection requires ray tracing, a rough estimate of near-normal incidence can be derived, by assuming that each ray strikes twice the Si surface so that reflection is the square of the untextured case.

Texturing is incorporated into both industrial and laboratory Si solar cells, and, in combination with antireflection (AR) coating, reduces reflection losses to a few percent. In the latter case, in order to better control the pyramid geometry and to allow delineation of fine features on the surface, photolithographic techniques are used to define inverted or upright pyramids at the desired positions. It has to be noted that in this case the reflectivity is similar to that of a random texture [64].

Light entering the substrate at a textured surface is tilted with respect to the cell normal. This means that photogeneration takes place closer to the collection junction, which is very beneficial for low-diffusion length cells, by enhancing the collection efficiency of medium- to long wavelengths (Figure 7.5a). The effect is equivalent to an increase of the absorption coefficient. As a drawback, textured surfaces present higher SRVs.

7.3.6.3 Light-trapping

Long-wavelength photons are weakly absorbed in silicon, and, unless internal reflectances are high, they will escape the substrate without contributing to photogeneration. The aim of light-trapping or light confinement techniques is to achieve high internal reflectances.



(a) (b)

Figure 7.5 Effects of surface texturing: (a) decreased reflection; and (b) increased photogeneration in the base

Practical back mirrors that are fully compatible with the cell electrical design, can be implemented, such as those schematized in Figure 7.4. A metal can make a good reflector, but Al, especially after heat treatment, gives low reflectance. The Si-oxide-metal structure in Figure 7.4(c) can present a high reflectance by capitalizing on interference effects [65].

At the front, the metal mirror is not applicable because the ray paths must be kept open for the entering light. Still, high front reflectance can be achieved because of total internal reflection [61]. Rays striking the surface at angles larger than the critical air-Si one are totally reflected. Texturing one or both surfaces with macroscopic or microscopic features serves this purpose by tilting the rays. Even in the case of geometric texturing with well-defined surface orientations, after a few internal reflections, the direction of rays inside the wafer is randomly distributed: this is the lambertian case, useful analytic approximation to light-trapping. Bifacial structures in Figure 7.4 can, for the same reason, be very efficient at confining the light [66].

Light-trapping increases the effective thickness of the wafer for absorption. In the geometrical optics regime, it has been shown that for one-side isotropic illumination the maximum enhancement factor (though perhaps not realizable) is 4 $(n_{Si}/n_{air})^2$, that is, each ray traverses 50 times the cell thickness before escaping [67]. The corresponding enhancement in photogeneration will be lower because of the competition of the absorption by free carriers at long wavelength.

Light-trapping is essential for thin cells. Even in the thick PERL design it can suppose around 1 mA·cm^{-2} enhancement in short-circuit current with respect to the case where the internal reflectances were zero.

7.3.7 Performance Comparison

For illustration purposes, Table 7.3 collects relevant parameters of the Auger-limited ideal Si solar cell [30], the best one-sun PERL cell [36] and a typical screen-printed, industrial cell on Cz-Si. The different concepts behind each set of data must be accounted for when comparing the figures. For instance, the ideal cell is assumed as being isotropically illuminated, while measurements are made for near-normal incidence.

Table 7.3 Cell performance (25°C, AM1.5 Global 0.1 W·cm^{-2})

Cell type	Ideal (calculated)	PERL (measured)	Industrial (typical)
Size (cm^2)	–	4	100
Thickness (μm)	80	450	300
Substrate resistivity (Ω·cm)	Intrinsic	0.5	1
Short-circuit current density, J_{SC} (A·cm^{-2})	0.0425	0.0422	0.034
Open-circuit voltage, V_{OC} (V)	0.765	0.702	0.600
Fill factor, *FF*	0.890	0.828	0.740
Efficiency, η (%)	28.8	24.7	15.0

The most striking difference between the best and the ideal solar cell is the difference in design: thick and low injection *versus* thin and high injection. The PERL cell is surely the best design for the currently achievable levels of surface recombination, which limit open-circuit voltage and shift the optimum thickness to high values. The low resistivity follows then from transport considerations for the chosen structure. The very high fill factor of the ideal cell is characteristic of high injection, Auger-limited operation.

Reduction of surface recombination in the best laboratory cells relies on surface passivation and the restriction of very heavily doped regions to a minimum. In the end, this is possible because of the possibility of defining and aligning very small features on the surfaces.

The very heavily doped emitter, along with lower substrate lifetimes, is responsible for the reduced short-circuit current and open-circuit voltage in the industrial cell. The fill factor is affected by the large device area in conjunction with the limitations of the metallization technique, which further reduces the current because of shading.

Continuous improvement in material quality and cost-driven thinning of the substrates will increase the need for industrial cells to implement surface passivation schemes. This will require the refinement of the metallization technique; another issue is that substrate lifetimes in an industrial environment depend on the gettering action of very heavy diffusions, which are not compatible with optimum surface performance. The PERL approach – high-temperature processes and delineation of fine features – is the most successful path to high efficiency, but it is not the only one that can inspire the forthcoming developments in industrial cells. It is worth noting, in this respect, that heterojunction with intrinsic thin-layer (HIT) solar cells have entered the exclusive club of more than 20% efficiency with a different approach (see Section 7.7.2).

7.4 MANUFACTURING PROCESS

7.4.1 Process Flow

Figure 7.6 shows the main steps of a simple process for solar cell fabrication based on screen printing. With more or less minor modifications, this process is currently used by many manufacturers. The main virtues of this 30-year-old PV technology are easy automation, reliability, good usage of materials and high yield. The drawback, as explained in preceding sections, is the efficiency penalty derived from the coarse and aggressive metallization technique. Some specific treatments worked out for mc-Si will be explained in Section 7.6.

Each step is briefly described in the following text with illustrative purposes: values of temperature, time and so on will only be indicative.

1. *Starting material*: The industry uses so-called solar grade Cz-Si wafers, round in origin but very often trimmed to a pseudo-square shape, or multicrystalline square wafers. Wafer dimensions are between 10- and 15-cm side and between 200- and 350-μm thickness. Doping is *p*-type (boron) to a resistivity of around 1 Ω·cm.
2. *Saw damage removal*: The sawing operation leaves the surfaces of "as cut" wafers with a high degree of damage. This presents two problems: the surface region is of

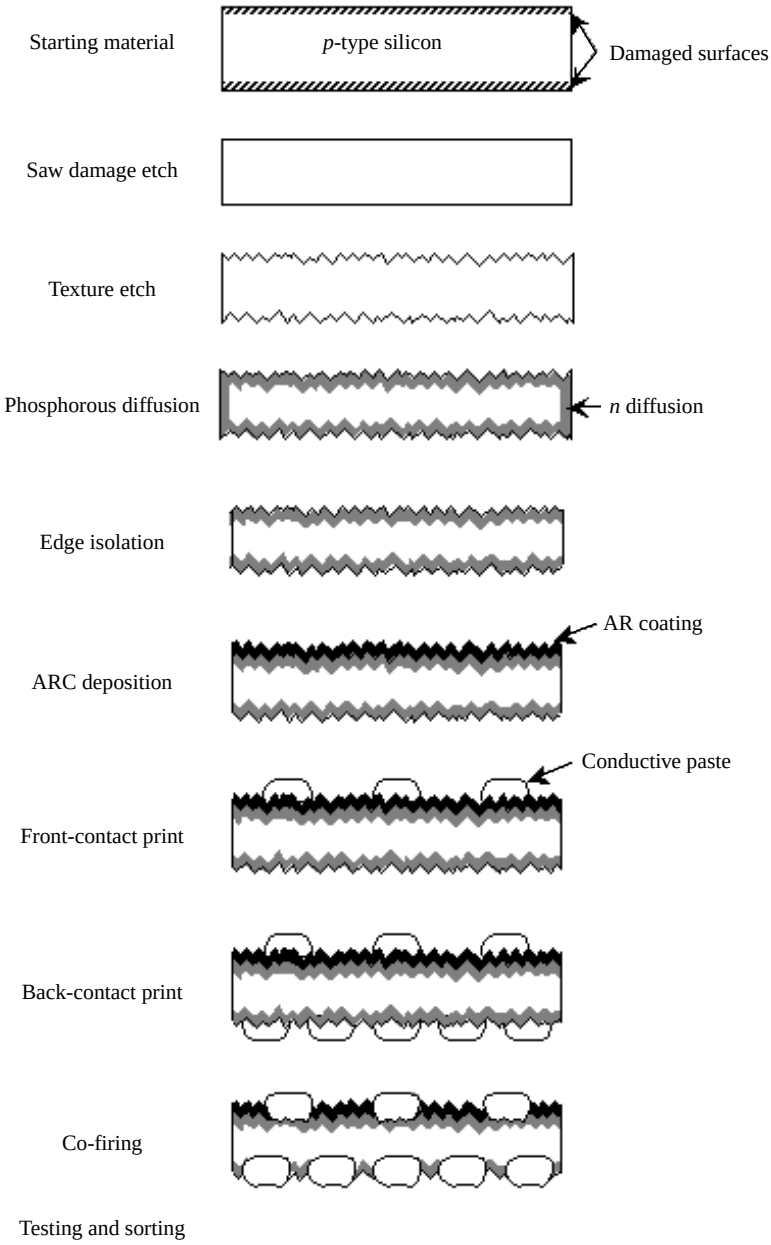


Figure 7.6 A typical processing sequence with schematic illustrations of the resulting structures

a very bad quality and the defects can lead to wafer fracture during processing [68]. For this reason, about ten microns are etched off from each face in alkaline or acid solutions. The wafers, in Teflon cassettes, are immersed in tanks containing the solution under temperature and composition control. Alkaline etches are preferred to acid solutions due to considerations of waste disposal.

Petitioner Canadian Solar (USA) Inc., Ex. 1043, p. 1008

3. *Texturization*: NaOH etching leading to microscopic pyramids is commonly employed. Their size must be optimized, since very small pyramids lead to high reflection, while very large ones can hinder the formation of the contacts. To ensure complete texturing coverage and adequate pyramid size, the concentration, the temperature and the agitation of the solution and the duration of the bath must be controlled (in fact NaOH at a higher concentration and at a higher temperature is commonly used as an isotropic etch for saw damage removal). Alcohol is added to improve homogeneity through an enhancement of the wettability of the silicon surface. Typical parameters are 5% NaOH concentration, 80°C and 15 min [69].
   Texturing alternatives for multicrystalline material are presented in Section 7.6.
4. *Phosphorus diffusion*: Phosphorus is universally used as the *n*-type dopant for silicon in solar cells. Since solid-state diffusion demands high temperature, it is very important that the surfaces are contamination-free before processing. To this end, after the texturing the wafers are subjected to acid etch to neutralize alkaline remains and eliminate adsorbed metallic impurities.
   The industry uses a number of procedures to perform the phosphorus diffusion. The following classification is based on the type of furnace in which the high-temperature step takes place:
   Quartz furnaces: The cells to be diffused, loaded in quartz boats, are placed in a quartz tube with resistance heating and held at the processing temperature (Figure 7.7a). The cells enter and exit the furnace through one end, while gases are fed through the opposite one. Phosphorus itself can be supplied in this way, typically by bubbling nitrogen through liquid $POCl_3$ before injection into the furnace. Solid dopant sources are also compatible with furnace processing. Five to fifteen minutes at temperatures in the range from 900 to 950°C can be considered representative. As suggested in Figure 7.6, both surfaces and the edges of the wafer will be diffused.

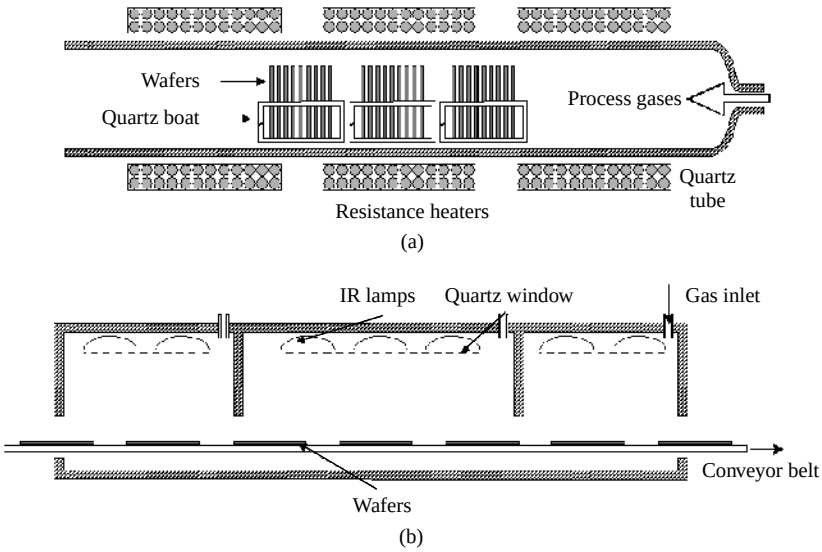


Figure 7.7 (a) A quartz furnace; and (b) a belt furnace for the diffusion of phosphorus

Belt furnaces: In this case, solid phosphorus sources are used: screen printed, spun-on or CVD pastes containing phosphorus compounds are applied on one wafer face and, after drying, placed in a conveyor belt passing through the furnace (Figure 7.7b). The temperature inside it can be adjusted in several zones and, though the furnace is open, gases can be supplied. The temperature cycle undergone by the wafer will mimic the temperature profile along the furnace with the time scale set by the advancing speed of the belt. A cycle begins with several minutes at around 600°C with clean air to burn off the organic materials of the paste, followed by the diffusion step in nitrogen at 950°C for 15 min. Only one face of the wafer is diffused, but the parasitic junction formed at the edges is also present in this technique due to diffusion from the gas phase. Resistance or infrared heating can be employed, the latter offering the possibility of faster heating–cooling rates of the wafer.

The main benefit of the quartz furnace is cleanliness, since no metallic elements are hot and no air flows into the tube. Though this is a batch step, high throughput can be achieved since many wafers can be simultaneously diffused in each tube, commercial furnaces consisting of stacks of four tubes. In a belt furnace, the ambient air can get into the furnace and the hot conveyor belt is a source of metallic impurities. The assets of belt furnaces are found in automation and in-line production, throughput and the ability to implement temperature profiles. New designs try to reconcile the advantages of both types of equipment [70].

After diffusion, an amorphous glass of phospho-silicates remains at the surface that is usually etched off in diluted HF because it can hinder subsequent processing steps.

5. *Junction isolation*: The *n*-type region at the wafer edges would interconnect the front and back contacts: the junction would be shunted by this path translating into a very low shunt resistance. To remove this region, dry etching, low temperature procedures are used.

   For the widely employed etch with plasma, the cells are coin-stacked and placed into barrel-type reactors. In this way, the surfaces are protected and only the edges remain exposed to the plasma. This is obtained by exciting with an RF field a fluorine compound (CF_4, SF_6), which produces highly reactive species, ions and electrons that quickly etch the exposed silicon surface [71]. Though this is a batch step, a large number of wafers can simultaneously be processed allowing high throughput. Laser cutting of the wafer edges is an alternative in industrial use.

6. *ARC deposition*: Titanium dioxide (TiO_2) is often used for creating the antireflecting coating due to its near-optimum refraction index for encapsulated cells. A popular technique is atmospheric pressure chemical vapor deposition (APCVD) from titanium organic compounds and water: the mixture is sprayed from a nozzle on the wafer held at around 200°C and the compound is hydrolyzed on the surface [72]. This process is easily automated in a conveyor-belt reactor. Other possibilities include to spin-on or screen print appropriate pastes.

   Silicon nitride is also used as AR coating material with unique properties that will be described in Section 7.6.

7. *Front contact print and dry*: The requirements for the front metallization are low contact resistance to silicon, low bulk resistivity, low line width with high aspect ratio, good mechanical adhesion, solderability and compatibility with the encapsulating

materials. Resistivity, price and availability considerations make silver the ideal choice as the contact metal. Copper offers similar advantages, but it does not qualify for screen printing because subsequent heat treatments are needed during which its high diffusivity will produce contamination of the silicon wafer.

Screen printing compares very unfavorably with vacuum evaporation in the first three requirements. It has been already commented how this affects the cell design and leads to a significant efficiency gap between laboratory and commercial cells. But throughput and cost compensate for it.

Screen printing will be described in more depth in the following section. It is used to stick a paste containing silver powder to the front face of the wafer in the comblike (fingers plus bus bars) pattern. Automatic screen printers are available that are capable of in-line, continuous operation with high throughput. These machines accept wafers from packs, cassettes or a belt line, place them with sufficient accuracy under the screen and deliver the printed wafers to the belt line. The paste is a viscous liquid due to the solvents it contains; these are evaporated in an in-line furnace at 100 to 200°C. The dried paste is apt for subsequent processing.

8. *Back contact print and dry*: The same operations are performed on the backside of the cell, except that the paste contains both silver and aluminum and the printed pattern is different.

   Aluminum is required because silver does not form ohmic contacts to p-Si, but cannot be used alone because it cannot be soldered. The low eutectic temperature of the Al-Si system means that some silicon will be dissolved and then recrystallize upon cooling in a p-type layer.

   Though in principle a continuous contact will give better electrical performance (lower resistance), most commercial wafers feature a back contact with a mesh structure: apart from paste-saving considerations, this is preferred to a continuous layer because the different expansion coefficients would produce warp of the cell during the subsequent thermal step.

9. *Cofiring of metal contacts*: A high-temperature step is still needed: organic components of the paste must be burnt off, the metallic grains must sinter together to form a good conductor, and they must form an intimate electric contact to the underlying silicon. As Figure 7.6 shows, the front paste is deposited on an insulating layer (the AR coating) and the back contact on the parasitic n-type rear layer.

   Upon firing, the active component of the front paste must penetrate the ARC coating to contact the n-emitter without shorting it: too mild a heat treatment will render high contact resistance, but too high a firing temperature will motivate the silver to reach through the emitter and contact the base. In extreme situations, this renders the cell useless by short-circuiting it. In more benign cases, small shunts appear as a low shunt resistance or dark current components with a high ideality factor that reduces the fill factor and the open-circuit voltage.
   The back paste, in its turn, must completely perforate the parasitic back emitter to reach the base during the firing.
   In order to comply with these stringent requirements, the composition of the pastes and the thermal profile of this critical step must be adjusted very carefully.

10. *Testing and sorting*: The illuminated $I-V$ curve of finished cells is measured under an artificial light source with a spectral content similar to sunlight (a sun simulator)

at a controlled temperature of 25°C. Defective devices are then rejected, and the rest are classified according to their output. The manufacturer establishes a number of classes attending, typically, to the cell current at a fixed voltage near the maximum power point. Modules will subsequently be built with cells of the same class, thus guaranteeing minimal mismatch losses. If, for instance, cell currents within a class must be equal within 5%, the accuracy and stability of the system must be better than that. Automatic testing systems that meet the very demanding requirements of high throughput processing are available.

7.4.2 Screen-printing Technology

Screen printing is a thick-film technology, a terminology that opposes it to the usual microelectronic procedures of evaporation of thin films. It consists in translating a layer of a material in a desired pattern to the surface of the wafer. Though it can be employed for virtually any step in solar cell manufacturing, contact formation constitutes the most demanding, frequent and conspicuous application of screen printing. Screens and pastes are the essential elements of this technology [68].

1. *Screens*: Screens are tight fabrics of synthetic or stainless steel wires stretched on an aluminum frame, as sketched in Figure 7.8. The screen is covered with a photosensitive emulsion, which is treated with photographic techniques in such a way that it is removed from the regions where printing is desired.

   For printing fine and thick layers, as is needed for the front contact of a solar cell, the wires must be very thin and closely spaced [73]. On the other hand, the opening of the reticule must be several times larger than the largest particle contained in the paste to be printed. Screens for solar cell production typically feature 200 wires per inch, wire diameter around 10 μm, mesh opening around 30 μm, corresponding to

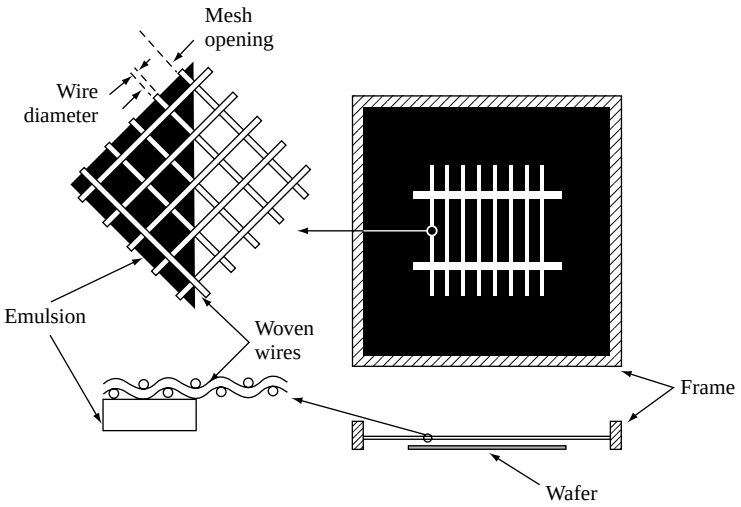


Figure 7.8 A screen for transferring the top contact pattern to a solar cell

nearly 50% open surface, that is, not intercepted by wires and a total thickness (woven wires plus emulsion) of around 100 μm.

2. *Pastes*: The pastes are the vehicles that carry the active material to the wafer surface. Their composition is formulated to optimize the behavior during printing. A paste for the metallic contacts of the solar cell is composed of the following:

   Organic solvents that provide the paste with the fluidity required for printing
   Organic binders that hold together the active powder before its thermal activation
   Conducting material, which is a powder of silver composed of crystallites of a size of tenths of microns; for the *p*-contact, aluminum is also present. This amounts to 60 to 80% in weight of the paste
   Glass frit, 5 to 10% in weight, a powder of different oxides (lead, bismuth, silicon etc.) with a low melting point and high reactivity at the process temperature, that enables movement of the silver grains and etches the silicon surface to allow intimate contact.

   The paste composition is extremely important for the success of the metallization and is critically linked to the heat treatment.

3. *Printing*: Figure 7.9 illustrates the process of printing a paste through the patterned emulsion on a screen. The screen and the wafer are not in contact, but a distance apart called the *snap-off*. After dispensing the paste, pressure is applied to the squeegee, which can be made of metal or rubber: this puts the screen in contact with the wafer. The squeegee is then moved from one side of the screen to the opposite one, dragging and pressing the paste in front of it. When an opening is reached, the paste fills it and sticks to the wafer, remaining there after the squeegee has passed and the screen has elastically retired.

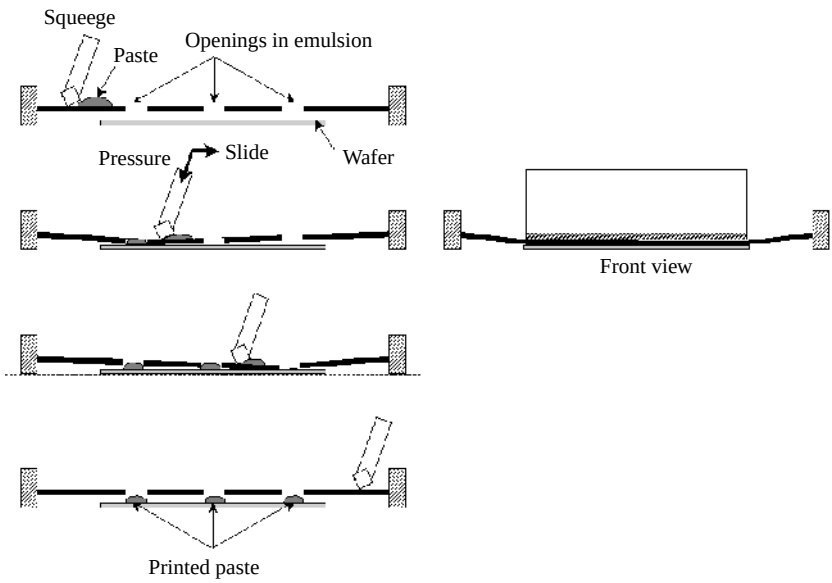


Figure 7.9 Illustration of a printing sequence

The amount of printed paste depends on the thickness of the screen material and the emulsion and the open area of the fabric. It also depends on the printed line width.

The viscous properties are of utmost relevance: when printing, the paste must be fluid enough to fill without voids all the volume allowed by the fabric and the emulsion, but after being printed it must not spread over the surface.

Critical parameters of this process are the pressure applied on the screen, the snap-off distance and the velocity of the squeegee.

4. *Drying*: Solvents are evaporated at 100 to 200°C right after printing so that the wafer can be manipulated without the printed pattern being damaged.

5. *Firing*: Firing of the pastes is usually done as a three-step process in an IR belt furnace. In the first step, when heating up, the organic compounds that bind the powder together are burnt in air. In the next step, the highest temperature between 600 and 800°C is reached and maintained for a few minutes. Higher temperatures are needed if an AR coating must be penetrated; crystal orientation and paste composition must be considered too. In the last step, the wafer is cooled down.

   The phenomena that take place during firing are very complex and not completely understood. The oxides forming the glass frit melt, enabling silver grains to sinter and form a continuous conductor so that the layer can present low sheet resistance. Neither the silver melting point nor the silicon–silver eutectic temperature is reached, sintering consisting of the intimate contact of solid silver crystallites. At the same time, the reactive molten glass etches some silicon and silver grains are allowed to form intimate contact with the substrate. The amount of etched silicon is on the order of 100 nm. When a layer of TiO_2 or SiN is present, the glass frit is able to etch through it. In fact, the quality of the contact improves because of a better homogeneity.

   The picture of the contact after cooling down shows two zones [74]. In the inner one, crystallites of silver are plugged into silicon forming crystalline interfaces and presumably very good electrical contact in a sort of "point contact". These grains are embedded in a compact amorphous glass. The outer zone is more porous and contains silver grains and glass frit: this porosity explains why the resistivity of silver paste is much higher than that of pure silver.

   Besides, the contact resistance of printed contacts is much higher than that of an evaporated contact to *n*-Si of the same doping. It seems that, although enough silver grains make good contact with silicon, not all them are connected to the grains in the outer layer; many remain isolated by the glass.

   When the paste, in the case of the back metallization, contains aluminum as well as silver, the Al-Si eutectic formed and recrystallized ensures a good contact. With dielectric layers the contact appears to be localized as well, and some beneficial role is attributed to the metal atoms in the frit [75].

6. *Limitations and trends in screen printing of contacts*: As explained in former sections, the high contact resistance and the etching action of the glass frit require the front emitters to be highly doped and not very thin if screen printing is used. Only improved paste formulation and processing can overcome this limitation.

   Narrow but thick fingers with good sheet conductance are also needed. Well-defined lines must be much wider than the pitch of the woven fabric; 60-μm lines seem achievable, with 100-μm ones being standard (see Figure 7.10). Incrementing the amount of transferred paste implies increasing the thickness of the emulsion or the