

FLASH MEMORIES

edited by
Paolo Cappelletti
Carla Golla
Piero Olivo
Enrico Zanoni

FLASH MEMORIES

By

Paolo Cappelletti

Carla Golla

Piero Olivo

Enrico Zanoni

KLUWER ACADEMIC PUBLISHERS
Boston / Dordrecht / London

Distributors for North, Central and South America:

Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA
Telephone (781) 871-6600
Fax (781) 871-6528
E-Mail <kluwer@wkap.com>

Distributors for all other countries:

Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS
Telephone 31 78 6392 392
Fax 31 78 6546 474
E-Mail <orderdept@wkap.nl>



Electronic Services <<http://www.wkap.nl>>

Library of Congress Cataloging-in-Publication Data

Flash memories / by Paolo Cappelletti ... (et al.).

p. cm.

Includes bibliographical references.

ISBN 0-7923-8487-3

1. Flash memories (computers) I. Cappelletti, Paolo

TK7895.M4F58 1999

004.5--dc21

99-25278

CIP

Copyright © 1999 by Kluwer Academic Publishers, Second Printing 2000.

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, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher, Kluwer Academic Publishers, 101 Philip Drive, Assinippi Park, Norwell, Massachusetts 02061

Printed on acid-free paper.

Printed in the United States of America

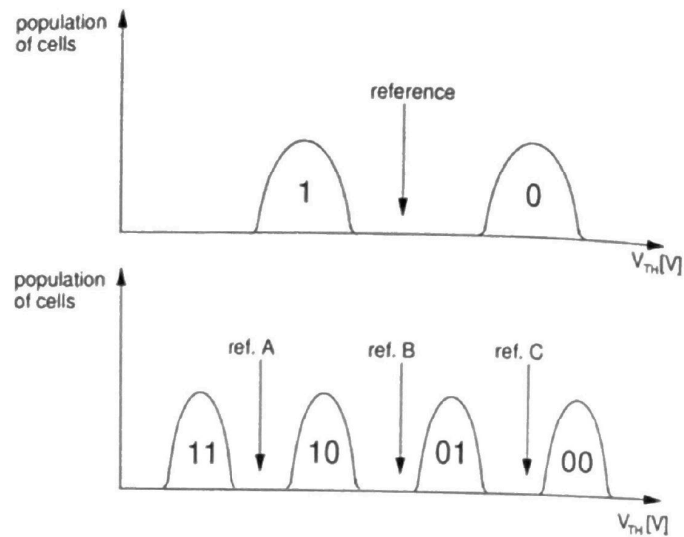


Figure 5.11 Flash cell: threshold distributions.

information, while in the latter 2^n bits of information are stored in a single cell. The main issue with a multilevel Flash memory is technological, since its development demands a great confidence in the process, because a tight distribution of the threshold of the programmed cells is needed.

In Fig. 5.11, threshold distributions for both standard and multilevel Flash memories are shown, together with the criteria used to determine the content of the cell.

No matter how many bits are stored, the complex of activities necessary to detect the cell content is referred to as Read operation; this operation is performed by a set of circuits which form the so-called Read path.

First of all, the address of the data is acquired by the input buffers: these circuits must fulfill requirements such as the noise immunity (to avoid a spurious commutation of the inputs), TTL voltage compliance and third level sensitivity (some input pins are connected not only to the usual buffer, but also to a circuit whose output goes high when the input reaches two or three times the V_{CC} value, in order to perform certain operations).

The input buffers drive the decoding circuits, which are made up of a pre-decoder followed by the decoder (see Section 5.2), whose purpose is to bias the selected cells with the proper read voltage; eight or sixteen cells are simultaneously selected through the column predecoder and decoder, and the proper drain voltage is forced to the cells.

In order to determine the value stored in the cells, a differential sensing is performed, comparing the matrix cells with a known reference. Being a current

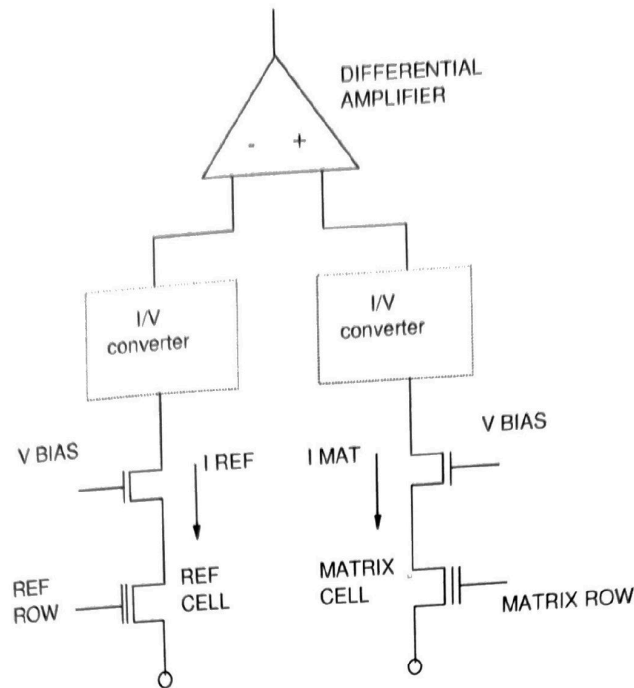


Figure 5.12 Differential sensing scheme.

comparison (a Flash cell sinks more or less current depending on its threshold voltage), a current-to-voltage conversion is performed on both the matrix and the reference branches (where transistors are used as loads), and finally the two voltages are compared by a differential amplifier (Fig. 5.12). It is worth noting that if both cells or both load transistors should vary in strength, this difference would result in a common mode voltage shift but not in a differential voltage variation, thus achieving a correct sensing.

If the cell of the matrix is programmed, it drains less current than the reference cell: the output voltage from the matrix cell is higher than that of the reference cell. If the cell of the matrix is virgin or erased, it has the same current as the reference cell: the cell can be shown virgin (more current than in the reference) if for example the load transistor for reference side is larger than the matrix load (these issues are deeply analyzed in Section 5.4).

The result of the sensing (which is a voltage) passes through a multiplexer to the output buffers, which cause the external data pin to change their state accordingly; the main problem is the noise introduced on V_{CC} and GND due to the large current required to charge and discharge the huge load capacitance of the data output pins.

As to the general issues related to the Read operation, great efforts are spent in developing solutions to reduce as much as possible the delay between

the address transition and the valid data at the output buffers; in other words, to achieve a faster access time.

5.1.4.2 Program Path Building Blocks Description. A Flash device can be programmed on a word or byte basis. As it was pointed out in Section 5.1.3, Program is a very complex and slow operation which is performed, in the actual Flash generation, by an embedded Program/Erase Controller; the user has only to provide the correct instruction, which is composed of a sequence of different commands, followed by the data to be programmed together with their address location. Then, the Program/Erase Controller automatically starts and performs the Program operation. The user or the external microcontroller can check the status of the operation using either the R/B# pin or the status on DQ0-DQ7 (the former can be issued either in "Data Polling - Toggle Bits" or "Status Register" fashion). In some Flash memories the Program can be suspended to read some other locations inside the array.

When the algorithm starts, the content of the memory location to be programmed is compared with the data: if they are valid (since to program means to change cell status from "1" to "0", and the opposite task can only be performed by an Erase operation, it is wrong to try to program a "1" onto a "0"), then a high voltage (HV switch of Fig. 5.10) is driven on the wordline of the cell to be programmed and then a program pulse is applied on the drain of the cell through the V_p generation circuit (Fig. 5.10). At the end of the program pulse, another verify is performed, reading the byte/word and comparing it with the data to be programmed (latched at the beginning of the operation). If comparison fails, another program pulse is applied to the uncompletely programmed bits. Fig. 5.13 shows a simplified flow chart and the timing of the main signals involved. The program efficiency decreases while threshold voltage increases: therefore it is possible to obtain a common value for programmed cells threshold voltages, even if the starting values are different.

5.1.4.3 Erase Path Building Blocks Description. The Erase operation is performed on all the cells inside a sector (thus sharing a common source) and can be divided in three fundamental steps:

1. Program All 0;
2. Erase;
3. Recovery of over-erased cells.

With the first step all the bits of the sector are put in the same condition as for their threshold voltages; an internal counter is used to address all the

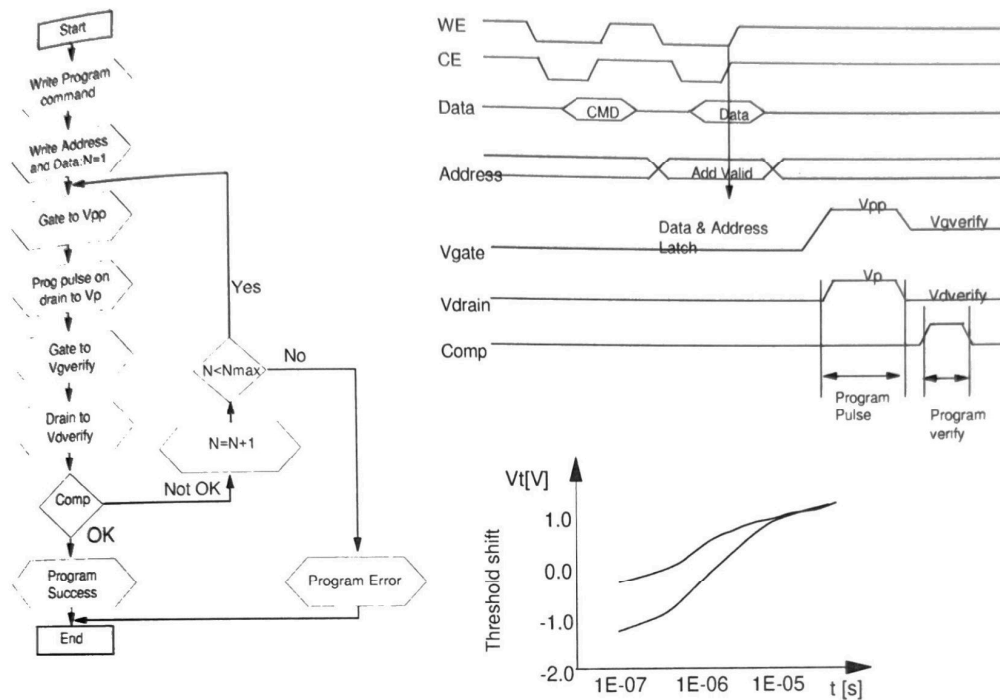


Figure 5.13 Program operation - flowchart.

words inside the sector, since Program operation is performed sequentially on every word.

The second step consists in providing a high voltage on the common source through Source switch keeping all the gates to GND and all the drain floating (this is valid both in a source and a gate erase scheme). If high voltage (V_{PP}) is not available on input pin, an internal circuit will generate it. At the end of erase pulse, a verify is performed on every word, in order to see if all the cells have been erased. If this verify fails, another erase pulse is given.

The third step performs a depletion verify to detect the over-erased bits. If present, these bits are recovered by programming them with a proper algorithm.

A simplified erase flow chart and the erase timing are shown in Fig. 5.14.

5.2 READ PATH: DECODING

As far as information storage or retrieval is concerned, it is necessary to devise a method to address the data inside the memory array. As already pointed out in Section 5.1.2, NOR matrix organization is considered: Flash cells are arranged in rows and columns selected by means of separate address decoding