PROGRAMMING FLASH MEMORY USING DISTRIBUTED LEARNING METHODS

In a memory device including an array of memory cells, each memory cell having more than two possible states, a method for programming a memory cell to a desired state is disclosed. The method comprises a control engine programming a subset (220), wherein the characterization information indicates programming characteristics of a representative memory cell of the array of memory cells (225). The control engine then uses the characterization information to directly program the memory cell to approximately the desired state (230) without performing a program verify operation.
<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM</td>
<td>Armenia</td>
</tr>
<tr>
<td>AT</td>
<td>Austria</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
</tr>
<tr>
<td>BY</td>
<td>Belarus</td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
</tr>
<tr>
<td>CI</td>
<td>Côte d'Ivoire</td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
</tr>
<tr>
<td>CN</td>
<td>China</td>
</tr>
<tr>
<td>CS</td>
<td>CzechoSlovakia</td>
</tr>
<tr>
<td>CZ</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
</tr>
<tr>
<td>EE</td>
<td>Estonia</td>
</tr>
<tr>
<td>ES</td>
<td>Spain</td>
</tr>
<tr>
<td>FI</td>
<td>Finland</td>
</tr>
<tr>
<td>FR</td>
<td>France</td>
</tr>
<tr>
<td>GA</td>
<td>Gabon</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>GE</td>
<td>Georgia</td>
</tr>
<tr>
<td>GN</td>
<td>Guinea</td>
</tr>
<tr>
<td>GR</td>
<td>Greece</td>
</tr>
<tr>
<td>HU</td>
<td>Hungary</td>
</tr>
<tr>
<td>IE</td>
<td>Ireland</td>
</tr>
<tr>
<td>IT</td>
<td>Italy</td>
</tr>
<tr>
<td>JP</td>
<td>Japan</td>
</tr>
<tr>
<td>KE</td>
<td>Kenya</td>
</tr>
<tr>
<td>KG</td>
<td>Kyrgyzstan</td>
</tr>
<tr>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
</tr>
<tr>
<td>KR</td>
<td>Republic of Korea</td>
</tr>
<tr>
<td>KZ</td>
<td>Kazakhstan</td>
</tr>
<tr>
<td>LI</td>
<td>Liechtenstein</td>
</tr>
<tr>
<td>LK</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>LR</td>
<td>Liberia</td>
</tr>
<tr>
<td>LT</td>
<td>Lithuania</td>
</tr>
<tr>
<td>LU</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>LV</td>
<td>Latvia</td>
</tr>
<tr>
<td>MC</td>
<td>Monaco</td>
</tr>
<tr>
<td>MD</td>
<td>Republic of Moldova</td>
</tr>
<tr>
<td>MG</td>
<td>Madagascar</td>
</tr>
<tr>
<td>ML</td>
<td>Mali</td>
</tr>
<tr>
<td>MN</td>
<td>Mongolia</td>
</tr>
<tr>
<td>MR</td>
<td>Mauritania</td>
</tr>
<tr>
<td>MW</td>
<td>Malawi</td>
</tr>
<tr>
<td>MX</td>
<td>Mexico</td>
</tr>
<tr>
<td>NE</td>
<td>Niger</td>
</tr>
<tr>
<td>NL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>NO</td>
<td>Norway</td>
</tr>
<tr>
<td>NZ</td>
<td>New Zealand</td>
</tr>
<tr>
<td>PL</td>
<td>Poland</td>
</tr>
<tr>
<td>PT</td>
<td>Portugal</td>
</tr>
<tr>
<td>RO</td>
<td>Romania</td>
</tr>
<tr>
<td>RU</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>SD</td>
<td>Sudan</td>
</tr>
<tr>
<td>SF</td>
<td>Sweden</td>
</tr>
<tr>
<td>SG</td>
<td>Singapore</td>
</tr>
<tr>
<td>SI</td>
<td>Slovenia</td>
</tr>
<tr>
<td>SK</td>
<td>Slovakia</td>
</tr>
<tr>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>SZ</td>
<td>Swaziland</td>
</tr>
<tr>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>TG</td>
<td>Togo</td>
</tr>
<tr>
<td>TJ</td>
<td>Tajikistan</td>
</tr>
<tr>
<td>TT</td>
<td>Trinidad and Tobago</td>
</tr>
<tr>
<td>UA</td>
<td>Ukraine</td>
</tr>
<tr>
<td>UG</td>
<td>Uganda</td>
</tr>
<tr>
<td>US</td>
<td>United States of America</td>
</tr>
<tr>
<td>UZ</td>
<td>Uzbekistan</td>
</tr>
<tr>
<td>VN</td>
<td>Viet Nam</td>
</tr>
</tbody>
</table>
PROGRAMMING FLASH MEMORY USING DISTRIBUTED LEARNING METHODS

FIELD OF THE INVENTION

The present invention relates generally to memory devices and more particularly to methods for programming memory devices.

BACKGROUND OF THE INVENTION

Nonvolatile semiconductor memory is a fundamental building block for a typical computer system. One type of prior nonvolatile semiconductor memory device is the flash electrically-erasable programmable read-only memory ("flash EEPROM"), and one type of prior flash memory cell comprises a single field effect transistor ("FET") including a select gate, a floating gate, a source, and a drain. Information is stored in the flash cell by altering the amount of charge stored on the floating gate, which causes the threshold voltage \( V_t \) of the flash cell to be varied. The flash memory cell is read by applying a select voltage via a wordline to the select gate. The amount of drain current \( I_D \) that the flash memory cell conducts when the select voltage is applied is determined by the threshold voltage \( V_t \) of the flash memory cell, and the state of the memory cell may be determined by comparing either the threshold voltage \( V_t \), the drain current \( I_D \), or the amount of charge stored on the floating gate to the same characteristic of a reference flash memory cell.

A typical prior flash memory cell stores only one bit of digital data, but flash memory cells that store more than one bit are known in the prior art. The number of bits stored by a flash memory cell depends on 1) the number of different analog states to which a flash memory cell may be placed by programming circuitry and 2) the number of different analog states that can be accurately determined by sensing circuitry.
Theoretically, there can be one analog state for each electron trapped on the floating gate, and a flash memory cell may be placed in a new state by simply trapping another electron on the floating gate; however, programming of a flash memory cell may be affected by environmental considerations and cell-to-cell variations. Environmental considerations similarly affect the operation of sensing a flash memory cell, and the resolution of sensing circuitry may not be fine enough to accurately discriminate between closely spaced analog states. Therefore, each state to which a memory cell may be placed typically corresponds to a range of charge and/or a corresponding range of threshold voltages or drain currents. Device characteristics confine the programming window, which is the total range of threshold voltages (or drain currents) that may be subdivided into two or more analog states, to a finite range such that requiring the discrimination between additional states narrows the range of threshold voltages (or drain currents) that each state may occupy. The "state width" of each analog state thus narrows.

Figure 1 is a flow chart showing an exemplary prior art method for placing a flash memory cell having two possible analog states ("erased" and "programmed") to the programmed state. The flash array is initially erased such that each of the flash memory cells are in the erased state, and a flash memory cell is selected for being placed to the programmed state at process block 5. Typically, several memory cells are programmed in parallel.

A programming pulse, which comprises applying appropriate voltages to the select gate, source, and drain of each selected flash memory cell for a predetermined amount of time, is applied to the selected flash memory cell at process block 10. The duration of the programming pulse (the "pulse width") and the programming voltages determine the amount of charge that is added to the floating gate of the flash memory cell.

For some prior technologies, only a single programming pulse is used to place a memory cell in the programmed state.
The programming method of Figure 1, however, employs a program-verify paradigm that allows greater control of the programming process. At process block 15, a verify operation is performed wherein the state of the selected flash memory cell is sensed and compared to a reference. If the selected flash memory cell is not in the programmed state, programming pulses are applied to the selected flash memory cell until it is successfully placed to the programmed state. The process then ends at process block 20.

When the number of states for a flash memory cell exceeds two, it becomes important to accurately place the flash memory cell in the desired state without "program overshoot," which occurs when the flash memory cell is accidentally placed to a state beyond the desired state. Program-verify placement schemes are therefore desirable. Unfortunately, because state widths are narrowed, programming pulse widths should also be narrowed, which results in more programming pulses being applied and more verify operations being performed. The use of multiple verify operations can therefore lead to substantial overhead that degrades programming performance of the memory device.

SUMMARY AND OBJECTS OF THE INVENTION

Therefore, it is an object of the present invention to provide a method for more quickly placing a memory cell having three or more analog states to a desired state.

This and other objects of the invention are provided by a method for programming a memory cell to a desired state. The method comprises a control engine programming a subset of the array of memory cells. Characterization information is determined from the step of programming the subset, wherein the characterization information indicates programming characteristics of a representative memory cell of the array of memory cells. The control engine then uses the characterization
information to directly program the memory cell to
approximately the desired state without performing a program
verify operation.

Other objects, features, and advantages of the present
invention will be apparent from the accompanying drawings and
from the detailed description which follows below.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example and
not limitation in the figures of the accompanying drawings, in
which like references indicate similar elements, and in which:

FIGURE 1 shows a prior art program-verify programming
method.

FIGURE 2 shows a general nonvolatile memory cell.
FIGURES 3A-3D show various alternative expressions for
the states of a memory device.

FIGURE 4 shows a flash memory cell configured for
programming.

FIGURE 5 shows a family of programming curves wherein
the programming drain voltage is held constant and the
programming gate voltage is varied.

FIGURE 6 shows a family of programming curves wherein
the programming gate voltage is held constant and the
programming drain voltage is varied.

FIGURE 7 shows the effect of impact ionization on the
maximum drain bias voltage of one type of flash cell.

FIGURE 8 shows a memory device that may be
programmed according to the disclosed programming methods.

FIGURE 9 shows an exact placement method of
programming.

FIGURE 10 shows average programming time as a
function of state time using an exact placement method.
FIGURE 11 shows a distributed learning method of programming according to one embodiment.

FIGURE 12 shows a distributed learning method of programming according to an alternative embodiment.

FIGURE 13 shows a predictive learning method of programming according to one embodiment.

FIGURE 14 shows a predictive learning method of programming according to an alternative embodiment.

FIGURE 15 shows a relative placement method of programming.

FIGURE 16 shows a data stream analysis method of programming according to one embodiment.

FIGURE 17 shows a data stream analysis method according to an alternative embodiment.

DETAILED DESCRIPTION

A number of methods for quickly placing a memory cell to one of three or more analog states are disclosed. These methods may be readily applied to memory cells that have only two possible analog states. As will also be discussed, some of the described methods may be used to quickly program memory cells to store analog data wherein the threshold voltage $V_t$ of the memory cell corresponds to an analog voltage. Wherein these methods are described with reference to flash EEPROMs, it should be noted that nonvolatile memory devices other than flash EEPROMs and volatile memory devices such as Dynamic Random Access Memories (DRAM) are capable of storing three or more analog states. Therefore, the disclosed methods may find similar application for memory devices other than flash EEPROMs.

Figure 2 shows a nonvolatile memory cell 25 having a select gate 30, a floating gate 35, a source 40, and a drain 45. Nonvolatile memory cell 25 behaves as a field effect transistor
having a threshold voltage $V_t$ that increases as charge is added to floating gate 35. The memory cell drain current $I_D$ ("cell current") decreases as the threshold voltage $V_t$ and cell charge level increase. The memory cell threshold voltage $V_t$ is related to the cell current $I_D$ by the expression:

$$I_D \propto G_m \times (V_G - V_t) \text{ for } V_D > V_G - V_t$$

$G_m$ is the transconductance of the memory cell; $V_G$ is the memory cell gate voltage; $V_D$ is the memory cell drain voltage; and $V_t$ is the memory cell threshold voltage.

Given this relationship, there are a number of different ways to sense the amount of charge (or "read" the data) stored on of the floating gate of the memory cell, including: sensing the cell current of a memory cell when a constant voltage is applied to the select gate of the memory cell; sensing the amount of voltage required at the select gate to give rise to an expected cell current for the memory cell; sensing a voltage drop across a load that is coupled to the drain of the memory cell when a constant voltage is applied to the select gate of the memory cell, wherein the cell current determines the amount of the voltage drop across the load; and sensing the amount of voltage required at the select gate to give rise to an expected voltage drop across a load that is coupled to the drain of the memory cell. To determine the analog state of the memory cell it is sufficient to compare a characteristic of the memory cell to a known reference.

Defining States For A Memory Cell

Generally, the physical characteristics of a nonvolatile memory cell require a minimum threshold $V_{t\text{min}}$ to which the nonvolatile memory cell may be erased and a maximum threshold $V_{t\text{max}}$ to which the nonvolatile memory cell may be programmed. The minimum threshold voltage $V_{t\text{min}}$ and the maximum voltage $V_{t\text{max}}$ delineate a maximum programming window for the non-volatile memory cell. The minimum threshold voltage $V_{t\text{min}}$ is constrained by erase times and gate
disturb voltages, and the maximum threshold voltage $V_{t\text{max}}$ is constrained by drain disturb voltages and bake charge loss.

Wherein the maximum width of the programming window is determined by physical characteristics of the nonvolatile memory cell, the manner in which states are defined within the programming window is influenced by a number of factors, including the following:

1) temperature fluctuations during programming;
2) programming voltage fluctuations during programming;
3) location of the memory cell within the memory array;
4) random variations of channel length;
5) drain disturb;
6) gate disturb; and
7) the resolution of the circuitry used to sense the state of the nonvolatile memory cell.

Figures 3A-3D show alternative expressions and definitions of states within the same programming window. Figure 3A shows a programming distribution of the number of cells in a given state versus the threshold voltage of that state. As shown, four states, State 0, State 1, State 2, and State 3 are defined within the programming window. For the purposes of illustration, the programming distribution for each state as shown is a bell curve wherein the majority of the cells programmed to a particular state fall within the center of the state. Figure 3A further shows a number of separation ranges are shown between contiguous states. The separation ranges are provided in order to more easily discriminate between states; however, separation ranges are theoretically not required. The state widths and the separation range widths are shown as being equal such that each state and each separation range defines one-seventh of the programming window.
Figure 3B shows a programming distribution of memory cells wherein eight states are defined within the programming window. Again, each state and each separation range is shown as defining an equal range of threshold voltages $V_t$ such that each state and each separation range occupies one-fifteenth of the total programming window.

Wherein the state width and separation range width for Figures 3A and 3B are shown as being equal, state width and separation range width may be defined somewhat more arbitrarily, and other considerations may limit the manner in which states may be defined. For example, for states closer to the edge of the programming window, it may be desirable to provide a larger state width. Figure 3C shows State 0 and State 3 as occupying threshold voltages ranges of 1500 mv, wherein State 1 and State 2 each have a state width of only 500 mv. Figure 3D shows the equivalent state distribution for Figure 3C in terms of the cell current $I_D$.

Figures 3A-3D illustrate possible state distributions for digital data storage applications. The same memory cell having the same programming window may be used to store analog data for applications such as sound recording and playback. For analog storage applications each threshold voltage $V_t$ or cell current $I_D$ within the programming window corresponds directly to an analog input voltage such that sensing the actual $V_t$ of the memory cell allows the direct synthesis of the input voltage by output circuitry. In this manner, sound may be recorded and played back. Examples of prior analog storage architectures may be found in U.S. Patent No. 4,890,259, entitled "High Density Integrated Circuit Analog Signal Recording and Playback System," and U.S. Patent No. 5,126,967, entitled "Writable Distributed Non-Volatile Analog Reference System and Method For Analog Signal Recording and Playback."

Programming Characteristics of Flash Memory
Figure 4 shows a flash memory cell configured for programming by hot electron injection. The select gate 30 of flash memory cell 25 is connected to a programming voltage \( V_G \). A typical programming voltage for prior flash memory cells is 12.0 volts.

Applying the programming voltage \( V_G \) to the select gate 30 switches the FET of the flash memory cell on, causing current to flow from the drain 45 to the source 40. The programming voltage \( V_G \) also creates a "vertical" electric field between the substrate 50 and the floating gate 35. Electron flow in the vertical electric field is depicted as an arrow having its head at floating gate 35 and its tail at substrate 50. This substantially shows the direction of electron flow in the vertical electric field.

As shown, source 40 is coupled to system ground VSS, and drain 45 is coupled to a drain voltage \( V_D \). The difference in potential between the drain 45 and the source 40 creates a "horizontal" electric field that accelerates electrons from the source 40 across the channel towards the drain 45. For one embodiment, it is sufficient for \( V_D \) to be 5-7 volts greater than the voltage at source 40. Electron flow in the horizontal electric field is shown as an arrow having its head at drain 45 and its tail at source 40. This substantially shows the direction of electron flow across the channel. The accelerated or "hot" electrons collide with the lattice structure of the substrate 50, and some of the hot electrons are swept onto the floating gate by the vertical electric field. In this manner, the amount of charge stored on the floating gate may be increased.

The state to which a non-volatile memory is placed is determined by the gate voltage \( V_G \), the drain voltage \( V_D \), the effective channel length \( L_{\text{eff}} \) of the memory cell, temperature, and pulse width, wherein the pulse width is the duration for which the programming gate voltage \( V_G \) and the programming drain voltage \( V_D \) are applied to the memory cell. As will now be
discussed, the programming gate voltage \( V_G \) is of primary significance.

Figure 5 graphs memory cell threshold voltage \( V_t \) versus the log of programming time for different programming gate voltages \( V_G \). The programming gate voltage \( V_G \) determines the relative strength of the vertical electric field, and increasing programming gate voltage \( V_G \) increases the strength of the vertical electric field during programming. Curve 60 shows threshold voltage \( V_t \) given a programming gate voltage \( V_G \) of 8 volts. Curve 65 shows threshold voltage \( V_t \) for a programming gate voltage \( V_G \) of 9 volts. Programming gate voltages \( V_G \) of 10 volts, 11 volts, and 11.5 volts results in curve 70, 75, and 80, respectively.

All five curves 60-80 show the threshold voltage \( V_t \) increasing exponentially in the "linear region" to the left of curve 55. The linear region is so named because when threshold voltage \( V_t \) is plotted on a linear time scale, the threshold voltage \( V_t \) increases linearly with \( V_G \) while programming in the linear region. Thus, the threshold voltage \( V_t \) increases greatly given a small increase in programming pulse duration when programming in the linear region, and precise control of the threshold voltage \( V_t \) is difficult.

Precise control of the threshold voltage \( V_t \) is easier when programming is performed in the "saturated region" to the right of curve 55. As shown, threshold voltage \( V_t \) increases more slowly, logarithmically, with time when the cell is programmed in the saturated region.

Programming of memory cell while operating in the saturated region is slow if the gate voltage \( V_G \) is maintained constant and the total programming pulse duration is increased. For example, given an initial programming pulse of 1\( \mu \)s duration and having a programming gate voltage \( V_G \) of 8.0 volts, the threshold voltage \( V_t \) of the memory cell is approximately 3.7 volts. If the programming gate voltage \( V_G \) is maintained at 8.0
volts, a programming pulse of approximately 10μs duration is required to raise the threshold voltage $V_t$ by 1.0 volt to 4.7 volts.

Programming in the saturated region occurs much more quickly if the gate voltage $V_G$ is increased with each subsequent programming pulse. In fact, as may be seen in Figure 5, increasing the gate voltage $V_G$ while programming in the saturated region results in the threshold voltage $V_t$ increasing by approximately the same amount for a constant pulse width. Thus, given the initial 1μs programming pulse with $V_G$ equal to 8.0 volts, a subsequent 1μs programming pulse with $V_G$ equal to 9.0 volts will raise the $V_t$ of the memory all from 3.7 volts to 4.7 volts, a one-to-one correspondence between the increase in $V_G$ and the increase in $V_t$.

The curves of Figure 5 assume a fixed programming drain voltage $V_D$, which is the source of the horizontal electric field across the channel. Figure 6 shows a family of curves given a constant programming gate voltage $V_G$ and a multiplicity of programming drain voltages $V_D$. As shown, the drain voltage $V_D$ affects the time when the memory cell enters the saturated region of programming. Different channel lengths result in a different family of curves; however, equivalent behavior for memory cells having different effective channel lengths may be achieved by trimming the programming drain voltage $V_D$.

The impact ionization induced Bi-polar turn-on voltage ($V_{Bi}$) limits the maximum drain bias voltage level that can be used while programming via hot electronic injection. Figure 7 illustrates the effect of $V_{Bi}$ upon threshold voltage $V_t$. The programming gate voltage $V_G$ is the same for each of the current $V_D1$, $V_D2$, and $V_D3$. For drain bias voltage levels less than $V_{Bi}$, increasing $V_D$ only affects the linear region of the $V_t$ versus time curve, as shown by the merging of curve $V_D2$ into the curve for $V_D1$. Maximum threshold voltage levels in the saturated region are unaffected by the increase in drain bias voltages levels.
When the drain bias voltage $V_D$ is greater than $V_{Bii}$, the threshold voltage levels in the saturated region are also raised.

**Exemplary Memory Device**

The generalized placement methods described herein are described with respect to an exemplary memory device shown in Figure 8. Memory device 120 is fabricated on a single semiconductor substrate and includes memory array 125, row decoder 130, column decoder 135, sensing circuitry 140, reference array 145, control engine 150, voltage switch 155, and command interface 160. Memory device 120 receives addresses via address lines 165 and receives and outputs data via bi-directional data lines 170. Data is stored using nonvolatile memory cells within memory array 125, wherein memory array 125 may include any type of memory cell with programmable threshold voltages, such as memory cells with trapping dielectrics or floating gates. Wherein memory device 120 is nonvolatile, control engine 140 may further include a write buffer 152 comprising SRAM for temporarily storing data with which to program memory array 125. The maximum allowable power consumption of memory device 120 is a primary factor in determining the maximum number of memory cells that may be programmed at any one time, and write buffer 152 is typically selected to store at least enough data to program a maximum number of cells at a time.

To read data stored in the memory array 125, row decoder 130 and column decoder 135 select a number of memory cells of the memory array 125 in response to a user-provided address received via address lines 165. Row decoder 130 selects the appropriate row of memory array 125, and column decoder 135 selects the appropriate column (or columns) of memory array 125. Sensing circuitry 140 compares the states of the selected memory cells to the states of reference cells of reference array 145. Sensing circuitry 140 may include differential comparators that output digital logic voltage levels in response to the comparisons between memory cells and reference cells. Thus, the analog
states of the memory cells may be expressed and output as digital data. The precise \( V_t/ID \) of a selected memory cell may be similarly determined.

Control engine 150 controls the erasure and programming of memory array 125. For one embodiment, control engine 150 includes a processor that is controlled by microcode stored in on-chip memory. Alternatively, the control engine 150 may be implemented as a state machine or by using combinational logic. Control engine 150 may also be implemented as a semiconductor device that externally controls the operation of memory device 120. The particular implementation of control engine 150 does not affect the described methods of programming of memory cells.

Control engine 150 manages memory array 125 via control of row decoder 130, column decoder 135, sensing circuitry 140, reference cell array 145, and voltage switch 155. Voltage switch 155 controls the various voltage levels necessary to read, program, and erase memory array 125. User commands for reading, erasure, and programming are communicated to control engine 150 via command interface 160. The external user issues commands to command interface 160 via three control pins: output enable OEB, write enable WEB, and chip enable CEB.

**Exact Placement Programming**

Figure 9 is a flow diagram showing an "exact placement" method for programming a number of memory cells in parallel. The method may be alternatively used to sequentially and individually program multiple memory cells. This exact placement method is described in more detail in U.S. Patent No. 5,440,505, which is commonly assigned to Intel Corporation of Santa Clara, California.

The method begins at process block 175 wherein memory array 125 is erased such that all of the memory cells are in a known state prior to programming. The step of erasing is not
required if some other mechanism for ensuring that the states of the selected cells are known prior to programming is provided. At process block 180, the control engine 150 initializes the programming variables including the source voltage $V_S$, the drain voltage $V_D$, the gate voltage $V_G$, and the pulse width $T$. As shown in figure 9, the source voltage $V_S$ is initialized to system ground $V_S$, the drain voltage $V_D$ is initialized to the trimmed drain voltage $V_D_{TRIM}$, the gate voltage $V_G$ is initialized to $V_G_{INITIAL}$, and the pulse width $T$ is initialized to a first pulse width $T_1$. The initial gate voltage $V_G_{INITIAL}$ and the initial pulse width $T_1$ are selected such that application of a single pulse will result in programming each selected memory cell to the saturation region for the initial gate voltage $V_G_{INITIAL}$.

At process block 185, an initial programming pulse is applied to the selected memory cells. A verify operation is undertaken at process block 190 wherein it is determined whether each memory cell in the selected subset of memory cells is at the destination state. If a memory cell is at the destination state, the process for that memory cell ends at process block 205. In order to avoid program overshoot the initial programming voltage $V_G_{INITIAL}$ and the initial pulse with $T_1$ are selected such that an initial pulse will not result in programming to the desired state. Thus, the process continues at process block 195, wherein the control engine 150 reduces the pulse width to a time $T_2$ and increases the gate voltage $V_G$ by a gate step voltage $\Delta V_G$. An additional programming pulse is applied to the memory cell or cells at process block 200, wherein the gate voltage is equal to $V_G + \Delta V_G$, and the pulse width is equal to $T_2$. Another verify step is undertaken at process block 190. If the memory cell is at the destination state, programming for that memory cell ends at process block 205. Any remaining memory cells that have not achieved their destination state repeat process steps 195 and 200. In this manner, the gate voltage is gradually increased for each programming pulse.
The gate step voltage $\Delta V_G$ and the pulse width $T_2$ are selected such that there will be a one-to-one correspondence between an increase in gate voltage $V_G$ and an increase in threshold voltage $V_t$. For example, if the gate step voltage is 300mv, the threshold voltage $V_t$ of a memory cell that is being programmed will be raised by 300mv each time a programming pulse with an increased gate voltage is applied. The gate step voltage $\Delta V_G$ and the pulse width $T_2$ are further selected to result in programming in the saturation region for the programming gate voltage $V_G$.

The relationship between threshold voltage $V_t$ and programming gate voltage $V_G$ follows directly from the curves shown in Figure 5. The amount of reduction from the initial pulse width $T_1$ to the subsequent pulse width $T_2$ depends on state widths and the magnitude of the gate step voltage $\Delta V_G$. The amount of reduction in pulse width tends to increase as state width and gate step voltage $\Delta V_G$ decrease.

The exact placement method shown in Figure 9 is a robust method that is easy to implement. Unfortunately, given a small state width, a large number of programming pulses must be performed. Furthermore, verify operations, which are essentially read operations, must be undertaken after each programming pulse. As shown in Figure 10, the average programming time for programming a logical a byte of memory cells increases as the state width decreases. This average programming time is a function of pulse width, the number of pulses, voltage settling time, verify time, and control engine overhead. Given a fixed programming window, state width must be decreased if additional states are to be discriminated. Therefore, an increase in average programming time is inevitable, and it is desirable to find alternative methods for programming memory cells having multiple analog states.

There are a number of ways to decrease the average programming time. For example, the number of pulses may be
reduced. Alternatively, the number of verify operations may be reduced, or the state machine overhead may be distributed across the entire array.

5 Learning Methods

One manner in which to reduce overall programming time is for control engine 150 to "learn" the programming characteristics of the memory array. According to a first type of learning algorithm, a statistically significant subset of memory cells for the array are programmed using an exact placement method, and control engine 150 derives characterization information regarding the average programming times for the array from the programming of the subset of cells. Characterization information may include the average number of pulses required to achieve specific destination states given the pulse width parameters of the exact placement algorithm. Alternatively, characterization information may simply be the derivation of the V<sub>t</sub>-versus-time characteristics of a memory cell such as the family of curves 60-80 shown in Figure 5. An alternative learning method is predictive in nature, and requires deriving characterization information for each cell during programming. Learning methods assume that environmental variables are held constant during programming.

Distributed Learning Methods

Figure 11 is a flow chart showing a distributed learning method. Prior to programming, the array is erased at process block 215. At process block 220, control engine 150 programs a subset of the array using an exact placement or similar method. At process block 225, the control engine 150 derives characterization information for an average cell of the array. Once the characterization information has been determined the selected cells may be programmed directly to the desired state. By direct programming it is meant that no verify operations are
undertaken to achieve the destination state. This may mean that a single pulse is used to program cells to the destination state, wherein the gate voltage $V_G$ and the pulse width $T$ are selected such that programming occurs in the saturation region. For example, if the family of curves shown in Figure 5 has been derived, a single pulse of 1μs duration with a gate voltage $V_G$ of 11.0 volts will place a cell directly to State 3 as defined by Figure 3C. Alternatively, direct programming may involve the application of a series of pulses as specified by the exact placement method of Figure 10 without performing verify operations. At process block 230, control engine 150 directly programs each of the array cell to its destination state using the characterization information. The distributed learning method ends at step 235.

Figure 12 is a flow chart showing a distributed learning method according to an alternative embodiment. The array is erased by control engine 150 at process block 240. Control engine 150 programs a subset of the array using an exact placement method at process block 245. At process block 250, control engine 150 derives characterization information for an average cell of the array. At process block 255, the control engine programs the cell to a point near the destination state, which provides a guardband against programming overshoot for memory cells that abnormally deviate from the average memory cell. The control engine finishes programming the cell to the designation state using the exact placement method at 260, and the distributed learning method ends at process block 235.

The overall programming time is greatly reduced for both types of distributed learning methods when compared to the previously described exact placement method. Much of this time savings is provided by eliminating verify operations.

**Predictive Learning**

Figure 13 is a flow chart showing a predictive learning method according to one embodiment. Unlike distributed
learning, predictive learning is performed for each individual cell. Thus, the family of programming curves is individually determined for each memory cell.

Control engine 150 begins by erasing the array at process block 270. A first programming pulse is applied to a memory cell by control engine 150 at process block 275. Control engine 150 determines the threshold voltage \( V_t \) after applying the first pulse at process block 280. Such a determination may be made simply by performing a verify operation. Alternatively, the control engine may determine the drain current \( I_D \).

Control engine 150 applies a second pulse to the memory cell at process block 285, and the control engine 150 again determines the threshold voltage \( V_t \) at process block 290. The control engine 150 may subtract the threshold voltage \( V_t \) measured after the first pulse from the threshold voltage \( V_t \) measured after the second pulse to determine a difference voltage \( \Delta V_t \), which the control engine may use to predict the \( V_t \)-versus-time characteristics for the memory cell. Thus, at process block 295, control engine 150 derives characterization information for the cell from the difference voltage \( \Delta V_t \). Control engine 150 may then program the cell directly to the destination state using the characterization information, and the predictive learning method ends at process block 305.

The two characterizing programming pulses are preferably selected to result in saturation region programming. For example, the two characterizing pulses may be performed according to the exact placement method of Figure 9. If the selected memory cells are placed to a known value of \( V_t/I_D \) prior to programming, only one characterizing pulse is required.

Figure 14 is a flow chart showing a predictive learning method according to an alternative embodiment. Control engine 150 erases the array at process block 310 and applies a first programming pulse to a memory cell at process block 315. The control engine 150 measures the threshold voltage \( V_t \) after the
first pulse at process block 320. The control engine 150 then applies a second programming pulse to the memory cell at process block 325 and measures the threshold voltage at process block 330. The control engine 150 derives characterization information from the cell using the difference voltage $\Delta V_t$ at process block 335. Control engine 150 then programs the memory cell directly to a point near the destination state for the memory cell using the derived characterization information such that no verify operation is required. This occurs at process block 340. At process block 345, control engine 150 completes programming of the memory cell using an exact placement method such as that shown in figure 10. The process ends at process block 350.

Predictive learning methods are especially well suited for quickly storing analog voltages in the memory array. For example, wherein the desired final threshold voltage $V_t$ for the memory cell is $V_{t\_target}$, the programming voltage $V_{GC}$ of the characterizing pulse may be given by the following equation:

$$V_{GC} = V_{t\_target} + c$$

wherein $c$ is a constant selected such that the characterizing pulse cannot result in overshoot of $V_{t\_target}$. According to one embodiment $c$ is equal to 3.5 volts.

Once the threshold voltage $V_{tc}$ after the characterization pulse has been determined, the programming voltage $V_{GP}$ of the programming pulse is determined by simply determining the differences between $V_{t\_target}$ and $V_{tc}$, increasing $V_{GC}$ by the same amount. This is shown by the following equation:

$$V_{GP} = V_{GC} + (V_{t\_target} - V_{tc})$$

Distributed learning methods may find similar application when storing analog data in nonvolatile media.

Relative Placement Methods

Exact placement methods and learning methods may be characterized as "absolute placement" methods because the states
are defined in an objective manner using predefined reference values. An alternative type of placement method is a "relative placement" method. Wherein absolute placement methods contemplate well defined state boundaries and separation ranges, relative placement methods require only that states are defined relative to one another. For example, given a four state memory cell wherein the lowest state is State 0 and the highest state is State 3, it is sufficient that a gate voltage associated with each state follow the following order:

\[ V_{G0} < V_{G1} < V_{G2} < V_{G3} \]

wherein \( V_{G0} \) defines threshold voltage for a memory cell in state 0, \( V_{G1} \) defines the threshold voltage for a memory cell in State 1, \( V_{G2} \) defines the threshold voltage for a memory cell in State 2, and \( V_{G3} \) defines the threshold voltage for a memory cell in State 3. Practically speaking, each state must still define a range of voltages, and the range of voltages must be guardbanded such that the largest threshold voltage of one state is not greater than the smallest threshold voltage of the next state.

Programming using such an ordering of states method merely entails applying one pulse at the selected gate voltage for the state to place the memory cells in the desired state. The pulse width should still be selected such that saturation programming occurs for the given value of programming gate voltage.

Figure 15 is a flow chart showing an ordering of states method. Control engine 150 erases the array at processing block 155. The programming variables are initialized at processing block 360, wherein the gate voltage is initialized to be \( V_{G1} \) and the pulse width is T3. The pulse width T3 is selected such that saturation programming occurs for the programming gate voltage of \( V_{G1} \). Control engine 150 applies a pulse to a selected cell or cells at process block 365 to place the cells in State 1. At process block 370 control engine 150 sets the programming voltage to \( V_{G2} \) and the pulse width to T4, wherein T4, is selected to result in saturation programming for a programming gate
voltage of VG2. At process block 375, control engine 150 applies a programming pulse to place selected cells in State 2. At process block 380, control engine 150 sets the programming gate voltage to VG3 and sets the pulse width to T5, wherein T5 is selected to result in saturation programming given the programming gate voltage VG3. A pulse is applied at process block 382 to place selected cells in State 3, and the ordering programming process is completed at process block 385. No verify operations are required.

The process shown in Figure 15 may be formed in a "carry along" manner wherein all cells of the array that are to be programmed receive all programming pulses until they have received the programming pulse that places them to the desired state. For example, the first programming pulse is applied to all the cells of the array that are to be programmed. Those cells that are to be programmed only to the State 1 are deselected to prevent further programming, and all cells that are to be programmed to State 2 and subsequent states receive the second pulse. State 2 cells receive no further programming.

The process of Figure 15 may be alternatively performed such that cells only receive the programming pulse that places them in their destination state. Predictive learning techniques may be applied to ensure that each programming pulse programs the selected memory cells in the saturated region.

Data Stream Analysis Methods

An alternative type of placement method reduces overall programming time by maximizing the number of memory cells that are programmed at any given time. Programming is often performed on a "block" by block basis, wherein a block of memory cells typically includes a single addressable byte or word of data. The programming of a block of memory cells occurs within a single "programming cycle."

Statistically, the data stored in a block of memory cells is random and may include cells that are to remain in the erased
state (State 0) as well as cells that are to be programmed. Therefore, programming by block typically results in less than all of the memory cells of a block being programmed per programming cycle. For memory cells having only two states, only half of the cells of a block are programmed per programming cycle, on average. For memory cells having n states, only 1/n of the cells of a block are programmed to a particular state, on average.

Programming time can be reduced by analyzing the stream of data to be programmed to ensure that up to the maximum number of memory cells (i.e., a full block) are programmed at each programming cycle. Figure 16 if a flow chart of a data stream analysis method according to one embodiment. Control engine 150 erases the array at process block 390. At process block 395, control engine 150 analyzes the set of data to be programmed into the memory array to determine which memory cells are to be programmed to which state and which memory cells are to remain in the erased state. Analysis of the data stream may be performed externally to memory device 120 and the analysis sent to control engine 150 such that control engine 150 may appropriately control programming of memory cell array.

Once the data stream has been analyzed, programming begins at process block 400, wherein control engine 150 programs up to a maximum number of cells having State 1 as their destination state. For example, if the number of cells to be programmed is determined by the number of bits in a block of memory, control engine 150 ensures that a full block of memory cells having State 1 as their destination are programmed at process block 400. If the number of cells having a destination state of State 1 is less than the maximum number, all of the cells are programmed at process block 400. Programming may be performed using any of the previously described methods.

If at process block 405 additional cells have State 1 as their destination state, process block 400 is repeated. Otherwise control
engine 150 ensures that a maximum number of cells having State 2 as their destination state are programmed at process block 410. At process block 415, if additional cells having State 2 as their destination state exist, control engine 150 repeats the step of process block 410 until all such cells are programmed.

At process block 420, control engine 150 ensures that up to a maximum number of cells having State 3 as their destination state are programmed. At process block 425, control engine 150 determines whether additional State 3 cells need to be programmed. If not, the process ends at process block 430.

Figure 17 shows an alternative data stream analysis method wherein memory cells having differing destination states are programmed simultaneously. The programming method of Figure 17 recognizes that cells programmed to a destination state beyond a first program state must "pass through" the first programmed state. The process begins at process block 435 wherein control engine 150 erases the memory cell array. At process block 440, control engine 150 analyzes the set of data to be programmed. Control engine 150 ignores cells that are to remain in the erased state or State 0. At process block 445, control engine 150 programs all cells having a destination state of State 1 or beyond to State 1, wherein up to a maximum number of cells are programmed simultaneously. Thus, some cells having a destination state of State 2, State 3, etc. are initially programmed to State 1. Process block 450 control engine programs all cells having a destination state of State 2 or beyond to State 2. Again, up to a maximum number of cells are programmed simultaneously. At process block 455, control engine 150 programs all cells having a destination state of State 3 or beyond to State 3. Again, up to a maximum number of cells are programmed simultaneously. The process ends at process block 460. If a memory cell may achieve more than four analog states, the process of Figure 17 may be readily modified.
The maximum number of cells that may be programmed is typically defined by the maximum power budget of memory device 120. Programming by data stream analysis depends on which cells are to be programmed, and programming may be independent of the physical correspondence of the array to memory locations defined by the write buffer. Because the block of memory cells that are programmed during a single programming cycle will not have a defined relationship (e.g. each bit is not a bit of a single addressable byte of data), the array decoding of memory device 120 should be modified to allow individual addressing of cells in the array such that a block of memory cells may be programmed per programming cycle without regard to their physical location in the array.

In the foregoing specification the invention has been described with reference to specific exemplary embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention. The specification and drawings are, accordingly, to be regarded in an illustrative rather than restrictive sense.
CLAIMS
What is claimed is:

1. In a memory device including an array of memory cells, each memory cell having more than 2 possible states, a method for programming a memory cell to a desired state, comprising:
   a control engine programming a subset of the array of memory cells;
   determining characterization information from the step of programming the subset, the characterization information indicating programming characteristics of a representative memory cell of the array of memory cells; and
   the control engine using the characterization information to directly program the memory cell to approximately the desired state without performing a program verify operation.

2. The method of claim 1, wherein the step of the control engine directly programming the memory cell comprises:
   the control engine determining a total number of programming pulses to apply to the memory cell to achieve the desired state in response to the characterization information; and
   the control engine applying the plurality of programming pulses to the memory cell.

3. The method of claim 2, wherein the step of the control engine directly programming the memory cell further comprises:
   the control engine determining a pulse width for each of the plurality of programming pulses in response to the characterization information.

4. The method of claim 3, wherein the step of the control engine directly programming the memory cell further comprises:
   the control engine determining programming voltages for each of the plurality of programming pulses in response to the characterization information.

5. The method of claim 4, wherein the memory cell exhibits linear and saturated programming behavior such that there is a linear programming region and a saturated programming region, the steps of the control engine determining the total number of
-26-

programming pulses, the pulse widths of the programming pulses, and the programming voltages of the programming pulses being performed by the control engine such that the step of the control engine applying the plurality of programming pulses to the memory cell results in the memory cell being programmed in the saturated programming region for each programming pulse.

6. The method of claim 1, wherein the step of the control engine directly programming the memory cell to approximately the desired state comprises:

the control engine using the characterization information to determine a programming pulse width and programming voltages of a single programming pulse capable of directly programming the memory cell to approximately the desired state; and

the control engine applying the single programming pulse to the memory cell.

7. The method of claim 6, wherein the memory cell exhibits linear and saturated programming behavior such that there is a linear programming region and a saturated programming region, the step of the control engine using the characterization information to determine a programming pulse width and programming voltages of a single programming pulse being performed by the control engine such that the step of the control engine applying the single of programming pulses to the memory cell results in the memory cell being programmed in the saturated programming region.

8. The method of claim 1, further comprising:

the control engine completing programming of the memory cell to the desired state by applying at least one programming pulse followed by a program verify operation.

9. In a memory device including an array of memory cells, each memory cell having more than 2 possible states, a method for programming a memory cell to a desired state, comprising:
a control engine programming a subset of the array of memory cells
to determine characterization information indicating
representative of programming characteristics for the array of
memory cells; and

the control engine using the characterization information to directly
program the memory cell to the desired state without performing
a program verify operation.

10. The method of claim 9, wherein the step of the control
engine using the characterization information to directly
program the memory cell to the desired state comprises:

the control engine determining a total number of a plurality of
programming pulses to apply to the memory cell to achieve the
desired state in response to the characterization information; and

the control engine applying the plurality of programming pulses to
the memory cell.

11. The method of claim 10, wherein the step of the control
engine using the characterization information to directly
program the memory cell to the desired state further comprises:

the control engine determining a pulse width for each of the
plurality of programming pulses in response to the
characterization information.

12. The method of claim 11, wherein the step of the control
engine using the characterization information to directly
program the memory cell to the desired state further comprises:

the control engine determining programming voltages for each of
the plurality of programming pulses in response to the
characterization information.

13. The method of claim 12, wherein the memory cell exhibits
linear and saturated programming behavior such that there is a
linear programming region and a saturated programming region,
the steps of the control engine determining the total number of
programming pulses, the pulse widths of the programming
pulses, and the programming voltages of the programming
pulses being performed by the control engine such that the step
of the control engine applying the plurality of programming pulses to the memory cell results in the memory cell being programmed in the saturated programming region for each programming pulse.

14. The method of claim 9, wherein the step of the control engine using the characterization information to directly program the memory cell to the desired state comprises:

- the control engine using the characterization information to determine a programming pulse width and programming voltages of a single programming pulse capable of directly programming the memory cell to approximately the desired state; and
- the control engine applying the single programming pulse to the memory cell.

15. The method of claim 14, wherein the memory cell exhibits linear and saturated programming behavior such that there is a linear programming region and a saturated programming region, the step of the control engine using the characterization information to determine a programming pulse width and programming voltages of a single programming pulse being performed by the control engine such that the step of the control engine applying the single of programming pulses to the memory cell results in the memory cell being programmed in the saturated programming region.

16. In a memory device including an array of memory cells, a method for storing an analog voltage in a memory cell by programming a threshold voltage of the memory cell to a desired voltage, comprising:

- a control engine programming a subset of the array of memory cells to determine characterization information from the step of programming the subset, the characterization information indicating programming characteristics of a representative memory cell of the array of memory cells; and
the control engine using the characterization information to
program the threshold voltage of the memory cell directly to the
desired voltage.
17. The method of claim 16, wherein the step of the control
engine using the characterization information to directly
program the memory cell to the desired state comprises:
the control engine determining a total number of a plurality of
programming pulses to apply to the memory cell to achieve the
desired state in response to the characterization information; and
the control engine applying the plurality of programming pulses to
the memory cell.
18. The method of claim 17, wherein the step of the control
engine using the characterization information to directly
program the memory cell to the desired state further comprises:
the control engine determining a pulse width for each of the
plurality of programming pulses in response to the
characterization information.
19. The method of claim 18, wherein the step of the control
engine using the characterization information to directly
program the memory cell to the desired state further comprises:
the control engine determining programming voltages for each of
the plurality of programming pulses in response to the
characterization information.
20. The method of claim 19, wherein the memory cell exhibits
linear and saturated programming behavior such that there is a
linear programming region and a saturated programming region,
the steps of the control engine determining the total number of
programming pulses, the pulse widths of the programming
pulses, and the programming voltages of the programming
pulses being performed by the control engine such that the step
of the control engine applying the plurality of programming
pulses to the memory cell results in the memory cell being
programmed in the saturated programming region for each
programming pulse.
21. The method of claim 16, wherein the step of the control engine using the characterization information to directly program the memory cell to the desired state comprises:
the control engine using the characterization information to
determine a programming pulse width and programming
voltages of a single programming pulse capable of directly
programming the memory cell to approximately the desired
state; and
the control engine applying the single programming pulse to the
memory cell.

22. The method of claim 21, wherein the memory cell exhibits
linear and saturated programming behavior such that there is a
linear programming region and a saturated programming region,
the step of the control engine using the characterization
information to determine a programming pulse width and
programming voltages of a single programming pulse being
performed by the control engine such that the step of the control
ingine applying the single of programming pulses to the
memory cell results in the memory cell being programmed in
the saturated programming region.
FIG. 1

FIG. 2
FIGURE 5
FIG. 7
ERASE ARRAY

INITIALIZE PROGRAMMING VARIABLES:
\( V_S = V_{SS} \);
\( V_D = V_{D\_TRIM} \);
\( V_G = V_{G\_INITIAL} \);
\( T = T_1 \)

APPLY PROGRAMMING PULSE

MEMORY CELLS AT DESTINATION STATE?

YES

NO

\( T = T_2 \);
INCREASE \( V_G \) BY \( \Delta V_G \)

APPLY PROGRAMMING PULSE

END

FIG. 9
ERASE ARRAY 270

APPLY FIRST PULSE TO CELL 275

MEASURE $V_T$ (ID) 280

APPLY SECOND PULSE TO CELL 285

MEASURE $V_T$ (ID) 290

DERIVE CHARACTERIZATION INFORMATION FOR CELL FROM $\Delta V_T$ 295

PROGRAM CELL DIRECTLY TO DESTINATION STATE USING CHARACTERIZATION INFORMATION 300

END 305

FIG. 13
FIG. 14
ERASE ARRAY

INITIALIZE PROGRAMMING VARIABLES:
\( V_S = V_{SS} \)
\( V_D = V_{D, TRIM} \)
\( V_G = V_{G1} \)
\( T = T_3 \)

APPLY PULSE

SET
\( V_G = V_{G2} \)
\( T = T_4 \)

APPLY PULSE

SET
\( V_G = V_{G3} \)
\( T = T_5 \)

APPLY PULSE

END

FIG. 15
ERASE ARRAY

ANALYZE SET OF DATA TO BE PROGRAMMED

PROGRAM UP TO MAXIMUM NUMBER OF CELLS HAVING STATE 1 AS DESTINATION STATE

MORE STATE 1 CELLS?

PROGRAM UP TO MAXIMUM NUMBER OF CELLS HAVING STATE 2 AS DESTINATION STATE

MORE STATE 2 CELLS?

PROGRAM UP TO MAXIMUM NUMBER OF CELLS HAVING STATE 3 AS DESTINATION STATE

MORE STATE 3 CELLS?

END

FIG. 16
ERASE ARRAY

ANALYZE SET OF DATA TO BE PROGRAMMED

PROGRAM ALL CELLS HAVING A DESTINATION STATE OF STATE 1 OR BEYOND TO STATE 1, WHEREIN UP TO A MAXIMUM NUMBER OF CELLS ARE PROGRAMMED SIMULTANEOUSLY

PROGRAM ALL CELLS HAVING A DESTINATION STATE OF STATE 2 OR BEYOND TO STATE 2, WHEREIN UP TO A MAXIMUM NUMBER OF CELLS ARE PROGRAMMED SIMULTANEOUSLY

PROGRAM ALL CELLS HAVING A DESTINATION STATE OF STATE 3 OR BEYOND TO STATE 3, WHEREIN UP TO A MAXIMUM NUMBER OF CELLS ARE PROGRAMMED SIMULTANEOUSLY

END

FIG. 17
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
   IPC(6) : G11C 11/34, 27/00
   US CL : 365/185.03, 45, 185.19

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
   U.S. : 365/185.03, 185.11, 185.18, 185.19, 45

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
   NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
   See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, P</td>
<td>US 5,557,567 A (BERGEMONT et al.) 17 September 1996, column 2, lines 44-52.</td>
<td>1 and 16</td>
</tr>
<tr>
<td>A, P</td>
<td>US 5,487,033 A (KEENEY et al.) 23 January 1996, column 6, lines 41-45 and column 7, lines 12-33.</td>
<td>1 and 16</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:
   *A* document defining the general state of the art which is not considered to be of particular relevance
   *E* earlier document published on or after the international filing date
   *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
   *O* document referring to an oral disclosure, use, exhibition or other means
   *P* document published prior to the international filing date but later than the priority date claimed

I later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

* document member of the same patent family

Date of the actual completion of the international search
30 JANUARY 1997

Date of mailing of the international search report
2 5 FEB 1997

Authorized officer
HUU HOANG

Telephone No. (703) 305-3494

Form PCT/ISA/210 (second sheet) (July 1992)
B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS
search terms: memory or memories, (multi or multiple) (2a) (state# or bit# or threshold or level#), program? (2a)
"subset" (p) (characteriz? or characteristics)