US 20070263449A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0263449 A1**

Yu et al. (43) **Pub. Date: Nov. 15, 2007**

(54) **METHOD AND APPARATUS FOR PROGRAMMING FLASH MEMORY**

(75) Inventors: **Erwin E. Yu**, San Jose, CA (US); **Ebrahim Abedifard**, Sunnyvale, CA (US); **Frederick T. Jaffin**, Gilroy, CA (US); **Uday Chandrasekhar**, Sunnyvale, CA (US)

Correspondence Address:
DICKSTEIN SHAPIRO LLP
1825 EYE STREET, NW
WASHINGTON, DC 20006 (US)

(73) Assignee: **MICRON TECHNOLOGY, INC.**

(21) Appl. No.: **11/433,562**

(22) Filed: **May 15, 2006**

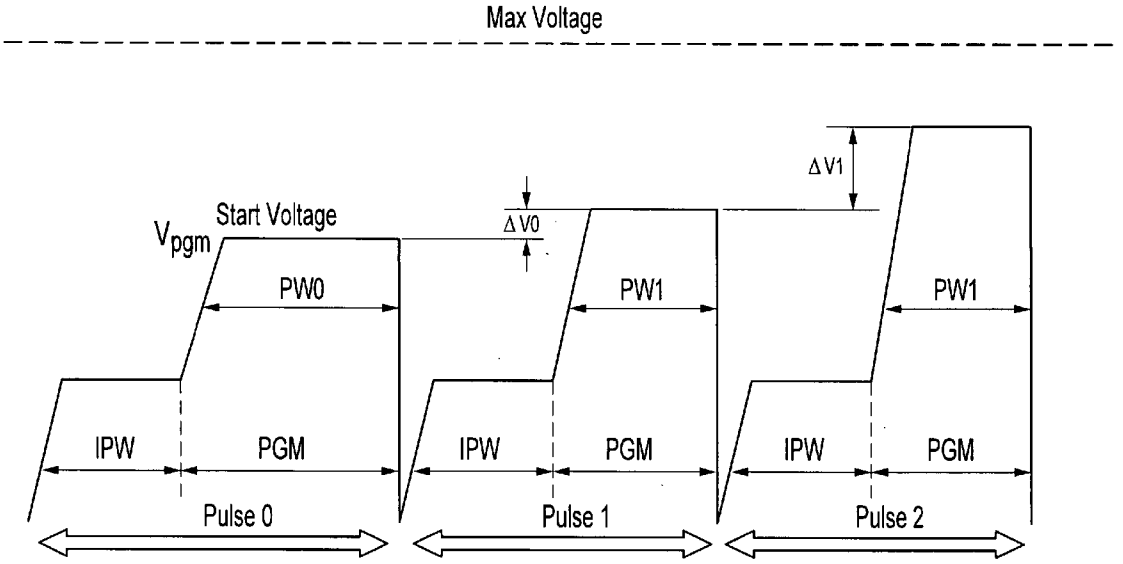
Publication Classification

(51) **Int. Cl.**
G11C 16/04 (2006.01)
G11C 16/06 (2006.01)
G11C 11/34 (2006.01)

(52) **U.S. Cl.** **365/185.19**; 365/185.22; 365/185.17; 365/185.28

(57) **ABSTRACT**

A method and apparatus that provides the ability to control programming pulses having different widths and/or voltages in a flash memory device. The widths and/or voltage levels of programming pulses are set to achieve programming of all memory cells of an array using a minimum number of programming pulses.



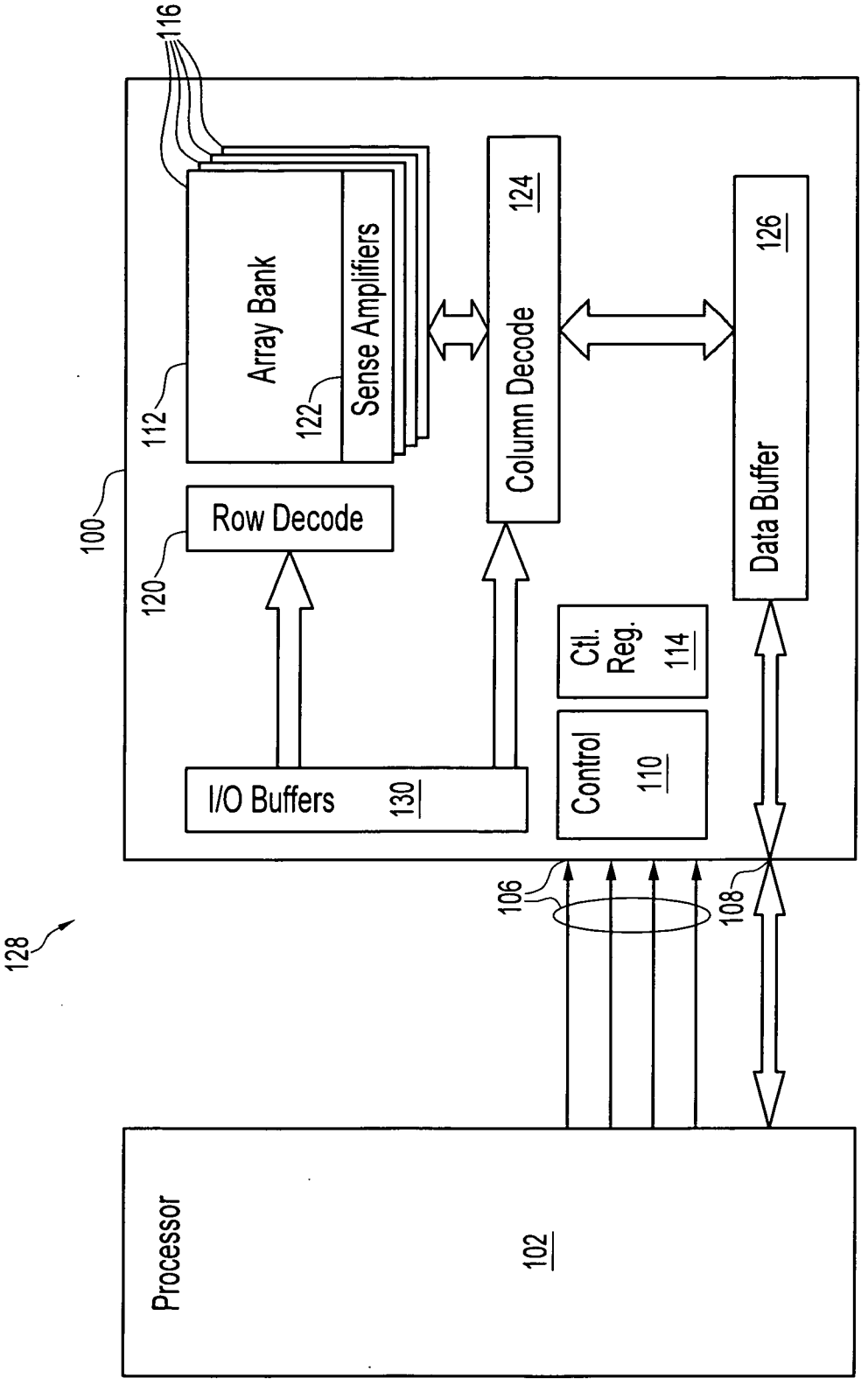

116
Array Bank
Sense Amplifiers
112
122
124
Column Decode
126
Data Buffer
100
Row Decode
120
Ctl. Reg. 114
Control 110
I/O Buffers 130
106
108
128
Processor
102

FIG. 1

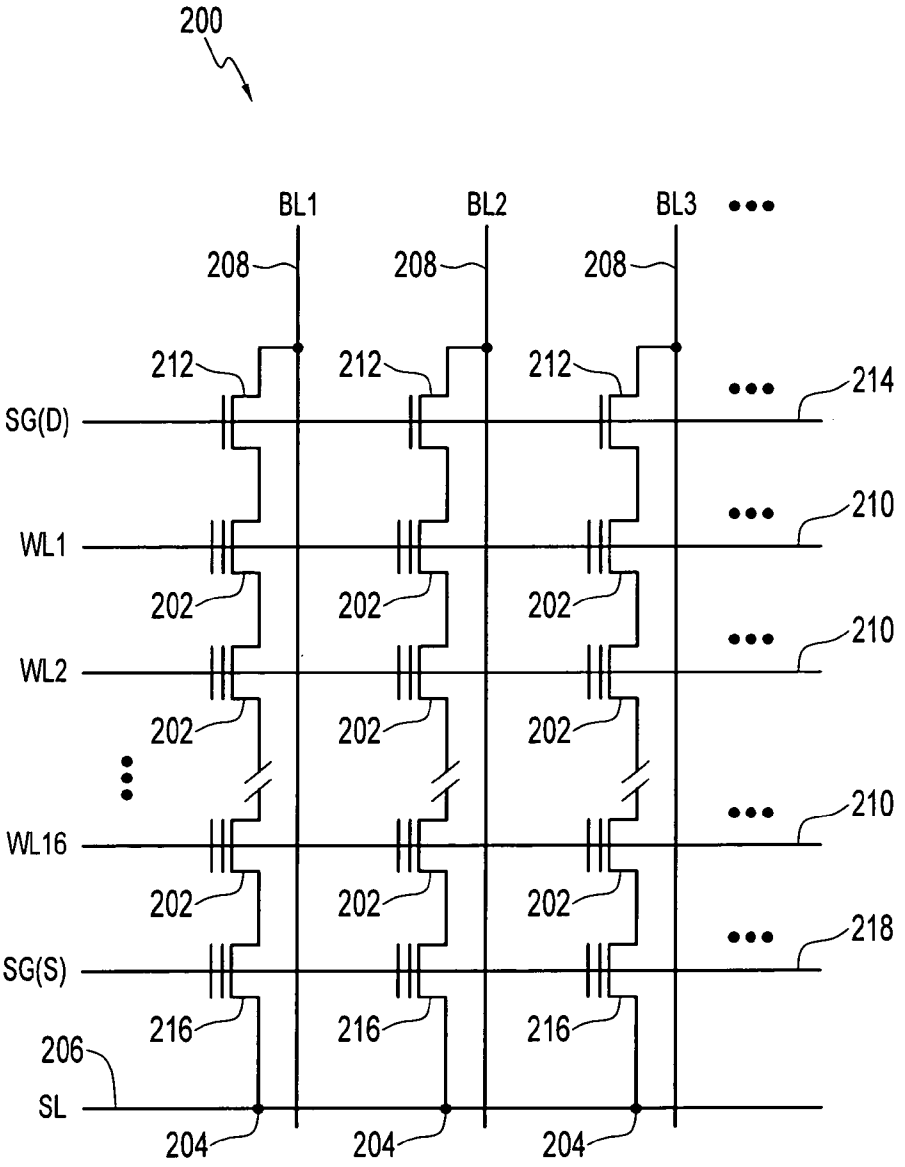

200
BL1
BL2
BL3
208
212
214
SG(D)
210
WL1
202
WL2
WL16
SG(S)
218
216
206
SL
204

FIG. 2

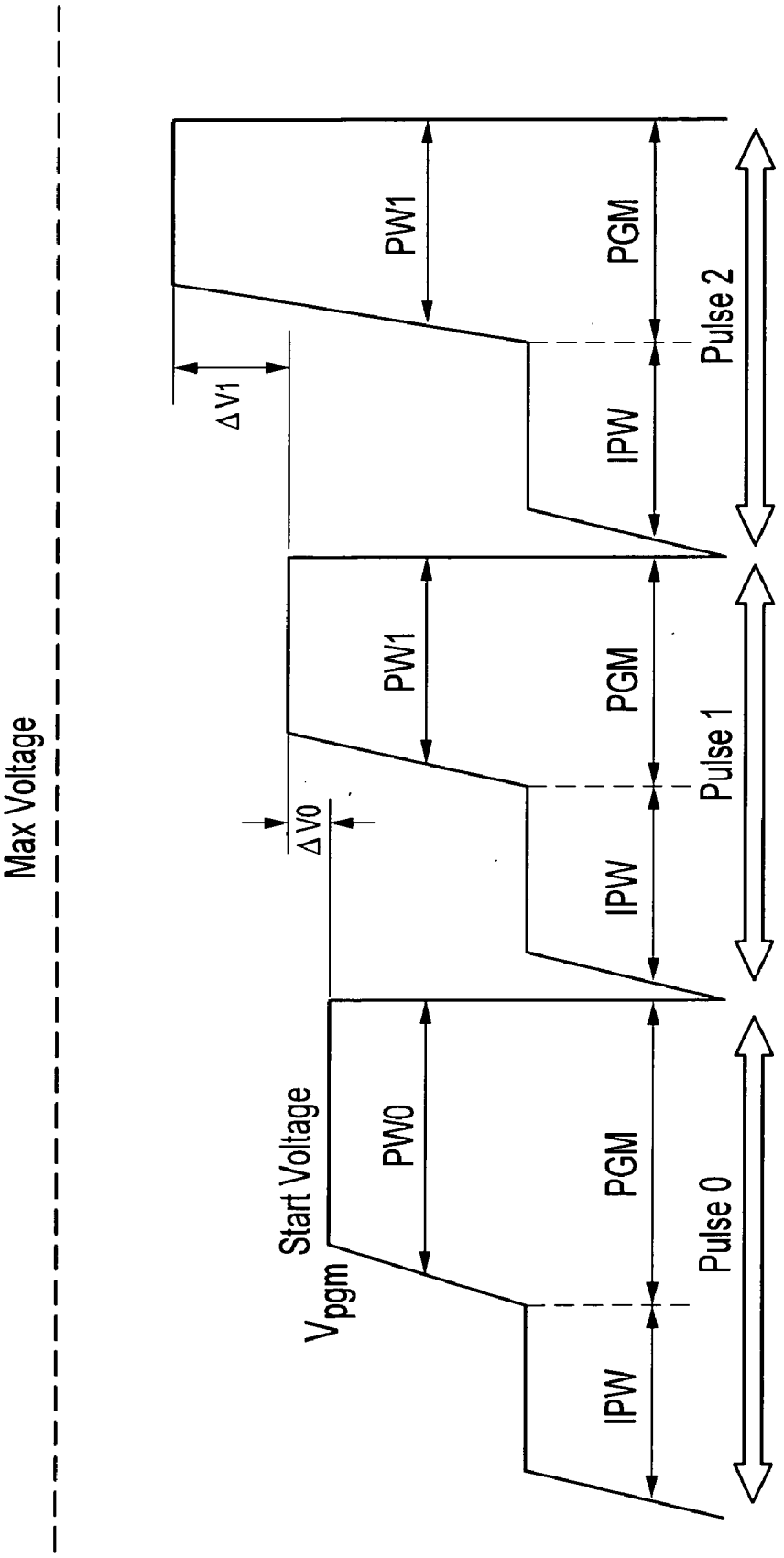

Max Voltage
Start Voltage
V_{pgm}
PW1
PGM
IPW
Pulse 2
ΔV1
PW1
PGM
IPW
Pulse 1
ΔV0
PW0
PGM
IPW
Pulse 0

FIG. 3

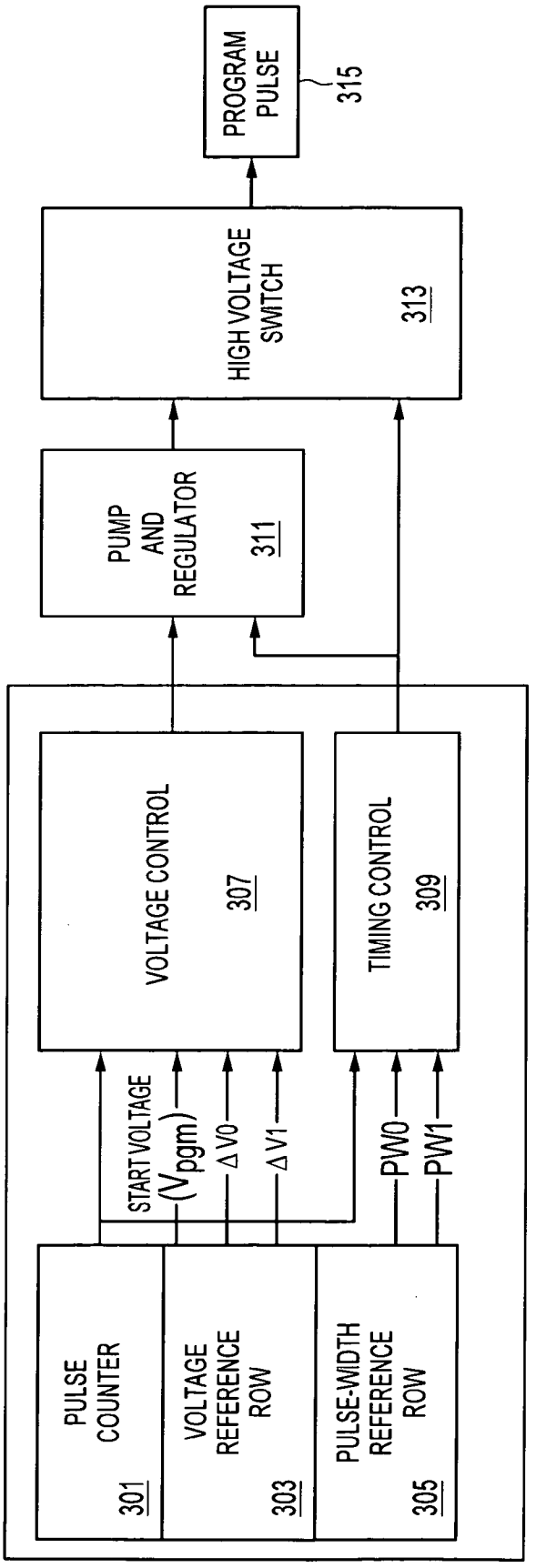

400
PROGRAM PULSE
315
HIGH VOLTAGE SWITCH
313
PUMP AND REGULATOR
311
VOLTAGE CONTROL
307
TIMING CONTROL
309
START VOLTAGE (V_{pgm})
ΔV0
ΔV1
PW0
PW1
PULSE COUNTER
301
VOLTAGE REFERENCE ROW
303
PULSE-WIDTH REFERENCE ROW
305

FIG. 4

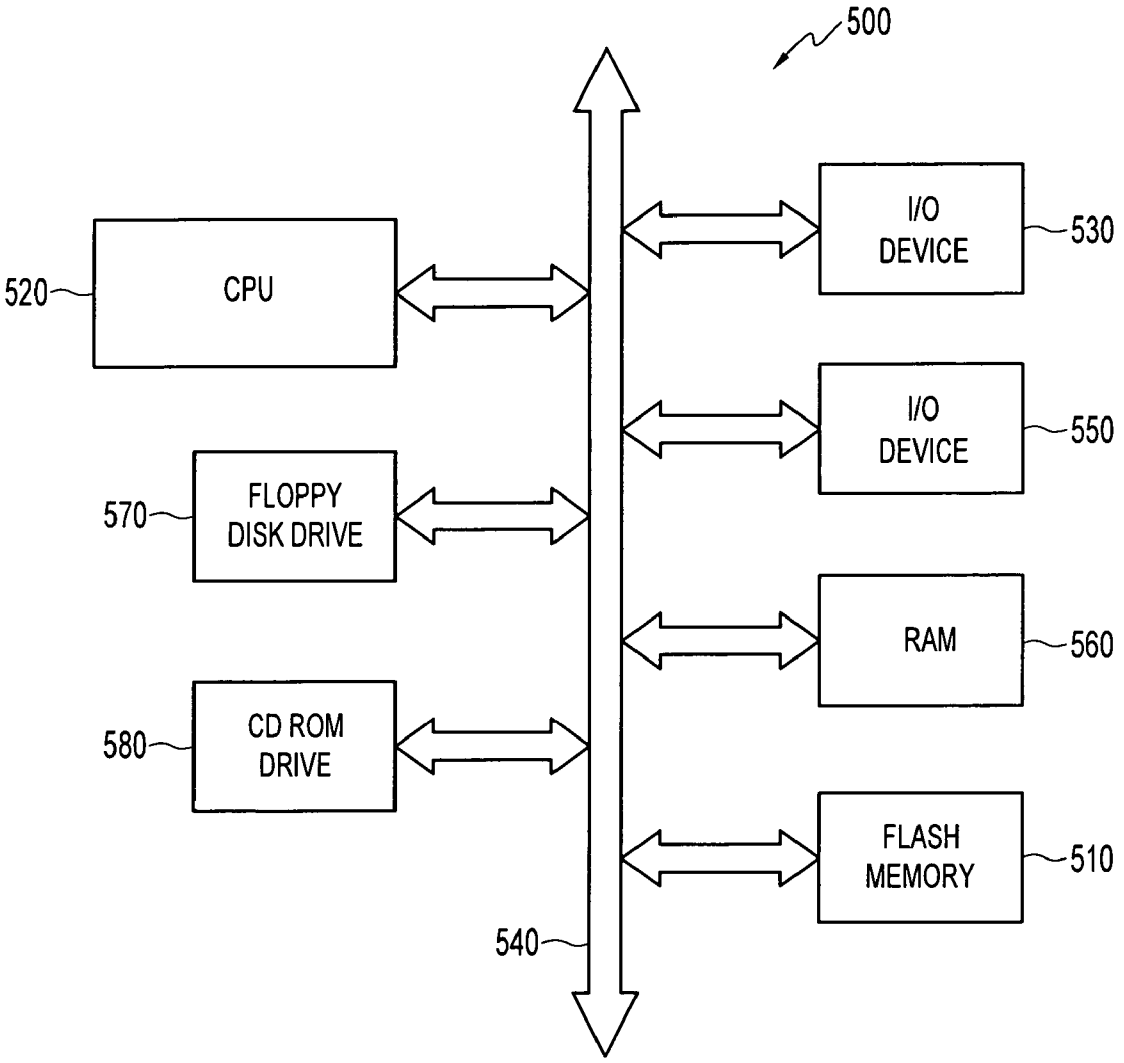

500
CPU
520
I/O DEVICE
530
I/O DEVICE
550
FLOPPY DISK DRIVE
570
RAM
560
CD ROM DRIVE
580
FLASH MEMORY
510
540

FIG. 5

METHOD AND APPARATUS FOR PROGRAMMING FLASH MEMORY

FIELD OF THE INVENTION

[0001] The present invention relates to a method and apparatus for programming memory and, more particularly, to a method and apparatus for programming a flash memory device.

BACKGROUND OF THE INVENTION

[0002] A nonvolatile memory is a type of memory that retains stored data when power is removed. There are various types of nonvolatile memories including e.g., read only memories (ROMs), erasable programmable read only memories (EPROMs), and electrically erasable programmable read only memories (EEPROMs). One type of EEPROM device is a flash EEPROM device (also referred to as "flash memory").

[0003] Each nonvolatile memory device has its own unique characteristics. For example, the memory cells of an EPROM device are erased using an ultraviolet light, while the memory cells of an EEPROM device are erased using an electrical signal. In a conventional flash memory device blocks of memory cells are simultaneously erased (what has been described in the art as a "flash-erasure"). The memory cells in a ROM device, on the other hand, cannot be erased at all. EPROMs, and EEPROMs, including flash memory, are commonly used in computer systems that require reprogrammable nonvolatile memory.

[0004] Two common types of flash memory architectures are the "NAND" and "NOR" architectures, so called for the resemblance which the basic memory cell configuration of each architecture has to a basic NAND or NOR gate circuit, respectively. In the NOR architecture, the floating gate memory cells of the memory array are arranged in a matrix. The gates of each floating gate memory cell of the array matrix are connected by rows to word lines and their drains are connected to bit lines. The source of each floating gate memory cell is typically connected to a common source line. The NOR architecture floating gate memory array is accessed by a row decoder activating a row of floating gate memory cells by selecting the word line connected to their gates. The data values of memory cells in a selected row are placed on the bit lines based on the application of a current from the connected source line to the connected bit lines.

[0005] A NAND array architecture also arranges its array of floating gate memory cells in a matrix such that the gates of each floating gate memory cell are connected by rows to word lines. However, each memory cell is not directly connected to a source line and a bit line. Instead, the memory cells of the array are arranged together in strings, typically of 8, 16, 32, or more, where the memory cells in the string are connected together in series, source to drain, between a common source line and a bit line. The NAND architecture floating gate memory array is then accessed by a row decoder activating a row of floating gate memory cells by selecting the word line connected to their gates. In addition, the word lines connected to the gates of the unselected memory cells of each string are also driven. However, the unselected memory cells of each string are typically driven by a higher gate voltage so as to operate them as pass transistors, allowing them to pass current in a manner that is unrestricted by their stored data values. Current then flows from the bitline to the source line through the channel of each memory cell of the connected string, restricted only by the memory cells of each string that are selected to be read. Thereby, the current encoded stored data values of the row of selected memory cells are placed on the bit lines.

[0006] Generally, in a flash memory device, a charged floating gate represents one logic state, e.g., a logic "0", while a non-charged floating gate represents the opposite logic state e.g., a logic "1". A memory cell of a flash array is programmed by placing the floating gate into one of these charged states. Charges may be injected or written onto the floating gate by any number of methods, including e.g., avalanche injection, channel injection, Fowler-Nordheim tunneling, and channel hot electron (CHE) injection. The floating gate may be discharged or erased by any number of methods including e.g., Fowler-Nordheim tunneling.

[0007] During a typical erase operation using Fowler-Nordheim tunneling, charges stored on the floating gate are driven out of the floating gate, through the tunnel oxide, and into the source region. This tunneling of electrons may be achieved by applying a relatively low positive voltage (e.g., approximately 5 volts) to the source region and a relatively large negative voltage (e.g., −8 to −12 volts) to the control gate. The substrate is usually grounded and the drain region is usually left floating. These voltages create an electric field between the floating gate and the source region, which induces the electrons previously stored on the floating gate to tunnel through the tunnel oxide to the source region.

[0008] Programming of a flash device is typically achieved by biasing with a series of pulses the drain region to a first voltage, relative to the source region, and biasing the control gate to a second positive voltage which is greater than the first voltage. In the absence of any stored charge on the floating gate, this biasing causes the formation of an inversion-layer channel of electrons at the face of the substrate, between the source and drain regions. The drain-to-source voltage accelerates the electrons through the channel to the drain region where they acquire sufficiently large kinetic energy. The larger positive bias on the control gate also establishes an electrical field in a tunneling oxide layer which separates the floating gate from the channel region. This electric field attracts the electrons and accelerates them toward the floating gate, which is between the control gate and the channel region, by a process known as tunneling. The floating gate then accumulates and traps the accumulated charge. The negative charge that accumulates on the floating gate reduces the strength of the electric field in the tunneling oxide layer to the point where it is no longer capable of accelerating the electrons from the drain side of the channel region.

[0009] The accumulation of a large quantity of trapped charge (electrons) on the floating gate will cause the effective threshold voltage of the field effect transistor comprising the source region, drain region, channel region and control gate to increase. If this increase is sufficiently large, the field effect transistor will remain in a nonconductive off state when a predetermined "read" voltage is applied to the control gate during a read operation. In the programmed state, the flash device may be said to be storing a logic 0. Once programmed, the flash device retains its higher threshold voltage even when its power supply is interrupted or

turned off for long periods of time. However, over programming of a flash memory cell can occur when the voltage of the memory cell is driven above the optimal threshold voltage. Over programming can lead to the inaccurate or even inability to read a memory cell during a read cycle.

[0010] Reading the flash device is achieved by applying a predetermined read voltage to the control gate, typically via a word line connecting a row of identical flash devices or cells, and applying a positive bias to the drain region, typically via a bit line connecting a column of identical flash cells. If the flash device is programmed, it will not conduct drain current. However, if the flash device has not been programmed (or has been erased), it is conductive. In this conductive state, the flash device may be said to be storing a logic 1. Thus, by monitoring the bit line current, the programmed state (i.e., 1 or 0) of the cells of the flash device can be determined.

[0011] As previously noted, pulses are applied to the lines of a memory array during programming. In order to program an individual memory cell, the voltage of the memory cell is raised (i.e., shifted), by the applied bias and bias pulses, to a programming threshold voltage V_{TH}. Memory cells of an array vary in the speed in which they are programmed. Accordingly, the threshold voltage shift of memory cells occurs at varying speeds even though a pulse of the same voltage and width is applied to all of the memory cells of the array. The distribution of the speeds at which the memory cells are programmed follows a typical bell curve. Accordingly, there are memory cells that are slow to program and fast program, as well as those that are neither fast nor slow (i.e., program at an average speed). The wide distribution of programming times may lead to over programming of some of the memory cells. Accordingly, there is a need and desire for a method which provides a tighter programming distribution by allowing more programming control options for a flash memory device.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The foregoing and other advantages and features of the invention will be more readily understood from the following detailed description of the invention provided below with reference to the accompanying drawings, in which:

[0013] FIG. **1** illustrates an exemplary simplified block diagram of system containing a flash memory device;

[0014] FIG. **2** illustrates an exemplary view of a memory array structure in a NAND type flash memory device;

[0015] FIG. **3** illustrates exemplary pulses applied to a word line during programming of a flash memory cell according to an embodiment of the invention;

[0016] FIG. **4** illustrates a block diagram illustrating a voltage and timing control circuit according to an embodiment of the invention; and

[0017] FIG. **5** illustrates a processor system incorporating a flash memory device constructed in accordance with an embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0018] In the following detailed description, reference is made to the accompanying drawings, which form a part hereof and show by way of illustration specific embodiments in which the invention may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention, and it is to be understood that other embodiments may be utilized, and that structural, logical, and electrical changes may be made without departing from the spirit and scope of the present invention.

[0019] The invention, in various exemplary embodiments, relates to a method and apparatus for controlling the electrical characteristics of the programming pulses used to program a flash memory device.

[0020] In one embodiment of the invention, the widths and voltage levels of programming pulses are controllably set to achieve memory cell programming using a required number of pulses. Pulse widths are controlled independently of each other and are controllably set to best suit a particular flash memory device. Furthermore, controlling the width and voltage of the programming pulses can lead to a reduction of program cycle time and threshold distribution while the likelihood that a memory cell becomes over programmed decreases.

[0021] FIG. **1** illustrates an exemplary system **128** incorporating a flash memory device **100** according to an embodiment of the present invention. The flash memory device **100** is connected to a host **102**, which is typically a processor, other processing device or memory controller. The flash memory device **100** is connected to a control bus **106** and an address/data bus **108** that are each connected to the processing device **102** to allow memory read and write accesses. It is noted that in alternative embodiments, the address/data bus **108** can be divided into separate buses. Internal to the flash memory device **100**, a control state machine **110** directs internal operations; manages the flash memory array **112** and updates RAM control registers and non-volatile erase block management registers **114**. The registers **114** (which may include tables) are utilized by the control state machine **110** during operation of the flash memory **100**.

[0022] The flash memory array **112** contains a sequence of memory banks **116** or segments, each bank **116** being organized logically into a series of erase blocks. Memory access addresses are received on the address/data bus **108** and are divided into row and column address portions. On a read access the row address is latched and decoded by row decode circuit **120**, which selects and activates a row page of memory cells and the other memory cells in their associated strings across a selected memory bank and communicates with I/O buffers **130**. The bit values encoded in the output of the selected row of memory cells are connected to a global bit line and detected by sense amplifiers **122** associated with the memory bank. The column address for the access is latched and decoded by the column decode circuit **124** which communicates with I/O buffers **130**. The output of the column decode circuit **124** selects the desired column data from the sense amplifier outputs **122** and is connected to the data buffer **126** for transfer from the memory device **100** through the address/data bus **108**. On a write access, the row decode circuit **120** selects the row page and the column decode circuit **124** selects the write sense amplifiers **122**. Data values to be written are connected from the data buffer **126** to the write sense amplifiers **122** selected by the column decode circuit **124** and are then written to the

selected floating gate memory cells of the memory array **112**. The written memory cells are then reselected by the row and column decode circuits **120**, **124** and sense amplifiers **122** so that they can be read to verify that the correct values have been programmed into the selected memory cells.

[**0023**] FIG. **2** shows an exemplary diagram of a NAND flash memory array **200**. The invention may be implemented on a NAND flash memory array **200**, among others, but the description of the invention in connection with NAND flash memory should not be considered limiting. For example, the embodiments described herein may similarly be implemented in a NOR or other type of flash memory device.

[**0024**] The NAND flash memory array **200** comprises a sequence of floating gate memory cell strings **204**. Each floating gate memory cell string **204** contains 16 NMOS floating gate memory cells **202** that are connected drain to source in a series chain. Word lines (WL**1**-WL**16**) **210** that span across multiple strings **204** are connected to the control gates of every floating gate memory cell **202** to control their operation. In operation, a particular word line **210** selects the individual floating gate memory cell **202** in the string **204** to be written to or read from, while the other word lines operate the remaining floating gate memory cells **202** in the string **204** in the pass through mode. Each string **204** of floating gate memory cells **202** are connected to a source line (SL) **206** by a source select gate **216** and to a respective bit line (BL**1**-BLN) **208** by a drain select gate **212**. The source select gates **216** are controlled by a source select gate control line (SG(S)) **218** connected to their control gates. The drain select gates **212** are controlled by a drain select gate control line (SG(D)) **214**.

[**0025**] The present invention provides the ability to vary the width of the first and subsequent pulses applied to the word lines WL **210**, as well as the ability to vary the voltage level of each program pulse for particular memory specifications. Some of the characteristics of the programming pulses of each die are set by product engineers, for example, during sorting and/or testing of the die to meet the required programming specifications. The dies are programmed with pulse voltage and width values in accordance on the die's programming speed.

[**0026**] As programming pulses are applied to the control gate of a memory cell **202**, the threshold voltage of the memory cell **202** rises from an initial voltage V_I towards the program threshold voltage V_{TH} (i.e., the threshold voltage of the memory cell shifts with the application of each programming pulse). Once the program threshold voltage V_{TH} is reached, the memory cell **202** is programmed.

[**0027**] FIG. **3** illustrates exemplary pulses applied to the word lines WL **210**, for example, of a memory array **200** during memory cell **202** programming. The word lines WL **210** are coupled to the control gates of the memory cell transistors in respective rows of the array. The pulses are applied to a specific word line WL **210** which is decoded from a supplied address. Each pulse applied to the word lines **210** for programming is composed of an inhibit pulse IPW and a program pulse PGM. The width and height of each inhibit pulse IPW is the same for each pulse (i.e., PULSE **0**, PULSE **1**, PULSE **2**) applied. According to an exemplary embodiment of the invention, the width PW**0** of the program pulse PGM of the first pulse PULSE **0**, that is applied to a word line WL **210** is wider than the width PW**1** of the succeeding program pulses PGM. The width PW**0** of the first programming pulse of PULSE **0** is set based on the implementation of the memory array **200** (e.g., the characteristics of the programming pulses of each die are set by product engineers during sorting and/or testing of the die to meet the required programming specifications). The width PW**0** is selected based on the implementation to provide a voltage shift of the memory cell **202** during the application of the first pulse PULSE **0** which is sufficient to reduce the threshold voltage shift required by the succeeding pulses. In the illustrated embodiment, the width PW**0** of the first pulse PULSE **0** is larger than the width PW**1** of the subsequent pulses. The exact pulse widths PW**0**, PW**1** are selected based on the program timing specifications of the memory. The advantages of using an independently varied width for the programming pulse of the first pulse PULSE **0**, is that it provides the flexibility to apply the most efficient program pulse width for the first pulse PULSE **0** and the shortest possible widths for the subsequent program pulses.

[**0028**] As previously noted, some memory cells **202** are very fast to program and require the application of only one programming pulse to reach the programming threshold voltage V_{TH}, whereas the memory cells **202** that are slow to program require the application of multiple pulses in order to reach the program threshold voltage V_{TH}. Accordingly, the voltage level of the first programming pulse PULSE **0** is controllably set such that the threshold voltage shift of the fast to program memory cells **202** will decrease once a saturation point (e.g., acceptable program threshold voltage V_{TH}) is reached. The width of the first programming pulse PULSE **0** may also be controllably set to be wide enough to cause a tangible threshold voltage shift on the slower to program memory cells **202** while maintaining a short program cycle. After the first pulse PULSE **0** is applied to the memory cell **202**, a verification of the threshold voltage is performed. The cells which are determined to be programmed during the verification will not receive subsequent programming pulses.

[**0029**] The voltage of the first programming PULSE **1** is controllably set such that the second pulse PULSE **1** has voltage higher than that of the first pulse PULSE **0** (e.g., the voltage ΔV_0 increase may be approximately 250 mV). The fast and average to program memory cells **202** will program during the first and second pulse applications. During the application of these programming pulses, the threshold voltage of the slow to program memory cells **202** will continue to shift toward the programming threshold voltage V_{TH}. After the application of the first and second programming pulses PULSE **0**, PULSE **1**, the threshold voltage shift of the memory cells **202** should be approximately equal to the voltage increment of each subsequent programming pulse (e.g., ΔV_1). The voltage increase between the second and third programming pulses PULSE **1**, PULSE **2** can be as high as 1V, for example. The voltage of the third programming pulse PULSE **2** is varied such that slower to program cells are programmed quickly by increasing the voltage increment ΔV_1 in order to accelerate the programming time. If the acceptable program threshold voltage level is, for example, between 1V-3V then the voltage level of the first and second programming pulses PULSE **0**, PULSE **1** can be controllably set to program the fast and average cells to a program threshold voltage V_{TH} of 1V and the third pulse voltage controllably set with ΔV_1=1V so that the slower to program memory cells **202** will be programmed to a pro-

gramming threshold voltage V_{TH} between 1V-2V. This leads to a tight programming distribution (i.e., faster programming time).

[0030] According to an exemplary embodiment of the invention, the voltage of the first pulse PULSE **0** is V_{pgm}. The voltage of the pulses PULSE **1**, PULSE **2** subsequent to the first pulse PULSE **0** may be set based on the voltages of the programming pulses PGM. The voltage of the second pulse PULSE **1** is ΔV_0 larger than V_{pgm} of PULSE **0** and the voltage of the third pulse PULSE **2** is greater than the voltage of the second pulse PULSE **1** by ΔV_1. Although three pulses are illustrated in FIG. **3**, the invention is not limited to the use of three programming pulses.

[0031] Once the threshold voltage of a memory cell **202** reaches a certain threshold voltage level V_x, the threshold voltage of the memory cell **202** will shift with the application of a programming pulse by approximately the voltage level as the change in voltage of the programming pulse from the preceding program pulse (e.g., ΔV_0, ΔV_1). For example, if the voltage change ΔV_0 of the second program pulse PGM is 250 mV, then with application of PULSE **1**, the voltage level of the memory cell **202** will shift approximately 250 mV towards the threshold voltage V_{TH}. However, before the voltage level V_x is reached, there is no direct correlation between the voltage shift of the memory cell **202** and the voltage of the programming pulse (i.e., programming of the memory cell does not follow a linear trend). The memory cells **202** which program at approximately an average speed or faster are more likely to reach this memory cell voltage level V_x with the application of the initial programming voltage V_{pgm} of the first pulse; however, the slower to program memory cells do not reach this voltage level V_x with the application of the initial programming voltage V_{pgm} (i.e., PULSE **0** illustrated in FIG. **3**).

[0032] After each pulse PULSE **0**, PULSE **1**, PULSE **2** is applied to a selected word line WL **210** of the memory array **220**, a read is initiated (i.e., a verify read) and the data read from the array is compared with the data programmed into the array. If there is a verify fail, one or more of the memory cells has not been programmed (i.e., V_{TH} has not been reached) and subsequent pulses (e.g., PULSE **1**, PULSE **2**) are applied, usually having higher voltage levels, until the threshold voltage V_{TH} is met or a predetermined maximum number of pulses is reached for all memory cells. As previously described, a count is maintained of the program pulses PGM applied to the memory cell and a predetermined number of pulses may be applied to the memory cell during programming to reach the threshold voltage V_{TH}. If the predetermined number of pulses is reached and the read verify does not verify that programming has occurred for the array, a programming failure occurs. The predetermined maximum number of pulses can be set individually for each memory device.

[0033] FIG. **4** illustrates an exemplary block diagram of a circuit **400**, according to one embodiment of the invention, for controlling the pulse width PW**0**, PW**1** and voltage variation ΔV_0, ΔV_1 used to implement variable programming of memory cells in a flash array according to an exemplary embodiment. The circuit **400** illustrated in FIG. **4** may be part of a programming control circuit or other type of control circuit that is either built in on the memory device or connected to the memory device. A pulse counter **301** keeps track of the number of programming of pulses applied to a memory cell. As previously noted, after each pulse is applied to the memory cell array, a read verify is initiated. If the pulse counter **301** has reached the predetermined maximum number of pulses, no additional pulses will be generated by the circuit **400** to program the memory cell. A voltage reference ROM **303** determines the voltage differentials ΔV_0, ΔV_1 that are used by the voltage control block **307** to control generation of the pulses PULSE **0**, PULSE **1**, PULSE **2** in accordance with an exemplary embodiment of the invention. The voltage differentials ΔV_0, ΔV_1, the initial programming voltage V_{pgm}, shown in FIG. **4** as the start voltage, and the number of pulses are used by the voltage control block **307** to generate control signals that determine the voltage levels of the program pulses. Based on the number of pulses, the initial programming voltage V_{pgm}, and the voltage differentials ΔV_0, ΔV_1, the voltage control block **307** regulates a pump and regulator **311** to set the appropriate pulse voltage levels.

[0034] A pulse-width reference ROM **305** provides the timing information (i.e., pulse widths PW**0**, PW**1**) for the programming pulses of pulses PULSE **0**, PULSE **1**, PULSE **2** illustrated in FIG. **3**. Using the number of pulses and the pulse widths PW**0**, PW**1**, the timing control block **309** generates signals to control when the pump and regulator **311** is turned on and discharged. Timing control **309** also controls the activation of the high voltage switch **313** to pass the pulses (e.g., PULSE **0**, PULSE **1**, PULSE **2**) and thus controls the width of the pulses. The voltage switch **313** outputs the program pulses **315** to be applied to the memory cells.

[0035] The independently set pulses help shift the memory cell's voltage to the threshold programming voltage level more quickly and using a minimum number of pulses. The flexibility provided by the ability to set the width and voltage parameters of the program pulses leads to the reduction of program cycle time and threshold distribution. Additionally, the independently set pulse width and voltages decrease the likelihood that any memory cell becomes over programmed.

[0036] FIG. **5** is a block diagram of a processor system **500** utilizing a memory device, e.g., a flash memory device **510**, constructed in accordance with the present invention. That is, the memory device **510** is programmed by individually tailored programming pulses in accordance with the embodiments of the invention. The processor system **500** may be a computer system, a process control system or any other system employing a processor and associated memory. The system **500** includes a central processing unit (CPU) **520**, e.g., a microprocessor, that communicates with the flash memory **510** and an I/O device **530** over a bus **540**. It must be noted that the bus **540** may be a series of buses and bridges commonly used in a processor system, but for convenience purposes only, the bus **540** has been illustrated as a single bus. A second I/O device **550** is illustrated, but is not necessary to practice the invention. The processor system **500** also includes random access memory (RAM) device **560** and may include a read-only memory (ROM) device (not shown), and peripheral devices such as a floppy disk drive **570** and a compact disk (CD) ROM drive **580** that also communicate with the CPU **520** over the bus **540** as is well known in the art.

[0037] The processes and devices described above illustrate preferred methods and typical devices of many that

could be used and produced. The above description and drawings illustrate embodiments, which achieve the objects, features, and advantages of the present invention. However, it is not intended that the present invention be strictly limited to the above-described and illustrated embodiments. For example, the feature of increasing voltage increments of the programming pulses described herein maybe disabled for a particular implementation or the pulse characteristics, including pulse width, may be the same for the second and further subsequent programming pulses. Any modification of the present invention that comes within the spirit and scope of the following claims should be considered part of the present invention.

What is claimed as new and desired to be protected by Letters Patent of the United States is:

1. A method of programming a memory cell of a flash memory array, said method comprising:

applying a first programming pulse, having a first pulse characteristic, to said memory cell;

determining if said memory cell is programmed; and

if said memory cell is not programmed, applying a second programming pulse, having a second pulse characteristic, to said memory cell,

wherein said first and second pulse characteristics are different.

2. The method of claim 1, wherein said first and second pulse characteristics include at least pulse width.

3. The method of claim 1, wherein said first and second pulse characteristics include at least pulse amplitude.

4. The method of claim 1, wherein said first and second pulse characteristics include pulse width and pulse amplitude.

5. The method of claim 1, further comprising reading stored values which define said first and second pulse characteristics and forming said first and second pulses using said stored values.

6. The method of claim 1, further comprising if said memory cell is not programmed, applying a third programming pulse having a third pulse characteristic.

7. The method of claim 6, wherein said second and third pulse characteristics are different.

8. The method of claim 6, wherein said second and third pulse characteristics are the same.

9. The method of claim 2, wherein the width of said first program pulse is wider than the width of said second programming pulse.

10. The method of claim 3, wherein the first program pulse has the lowest amplitude and subsequently applied programming pulses have increasing amplitudes.

11. The method of claim 1, wherein the flash memory device is one of a NAND and NOR flash memory device.

12. The method of claim 6, wherein the width of said first program pulse is wider than the width of said second and third programming pulses.

13. The method of claim 6, wherein the amplitudes of said first, second and third programming pulses are different.

14. The method of claim 13, wherein the amplitude of the third programming pulse is the larger than the first and second programming pulses and the amplitude of the first programming pulse is smaller than the amplitude of the second programming pulse.

15. A method of operating a non-volatile memory device comprising a plurality of floating gate memory cells, where the memory cells are coupled as a plurality of strings, said method comprising the acts of:

selecting a floating gate memory cell to be programmed;

applying a first programming pulse, having a first pulse characteristic, to said memory cell;

determining if said memory cell is programmed; and

if said memory cell is not programmed, applying a second programming pulse, having a second pulse characteristic, to said memory cell,

wherein said first and second pulse characteristics are different.

16. The method of claim 15, wherein said first and second pulse characteristics include at least pulse width.

17. The method of claim 15, wherein said first and second pulse characteristics include at least pulse amplitude.

18. The method of claim 15, wherein said first and second pulse characteristics include pulse width and pulse amplitude.

19. The method of claim 15, further comprising reading stored values which define said first and second pulse characteristics and forming said first and second pulses using said stored values.

20. The method of claim 16, wherein the width of said first program pulse is wider than the width of said second programming pulse.

21. The method of claim 18, wherein the first program pulse has a lower amplitude than the second programming pulse.

22. The method of claim 15, wherein the flash memory device is one of a NAND and NOR flash memory device.

23. A method of programming an array of memory cells of a flash memory device, said method comprising:

storing first pulse characteristics for a first programming pulse;

storing second pulse characteristics for a second programming pulse; and

using said stored first and second stored pulse characteristics to provide first and second sequential programming pulses to be applied said memory cells,

wherein said first and second pulse characteristics different.

24. The method of claim 23, further comprising:

storing third pulse characteristics for a third programming pulse; and

using said stored third pulse characteristics to provide a third sequential programming pulse.

25. The method of claim 24, wherein said first, second and third pulse characteristics are different.

26. The method of claim 23, wherein the width of said first programming pulse is wider than the width of said second programming pulse.

27. The method of claim 23, wherein the first programming pulse has a lower amplitude than the second programming pulse.

28. The method of claim 23, wherein the flash memory device is one of a NAND and NOR flash memory device.

29. A flash memory device comprising:

a plurality of flash memory cells; and

a control circuit coupled to said memory cells for supplying a plurality of programming pulses to said memory cells, wherein the width of a first programming pulse is different from the width of subsequent programming pulses.

30. The flash memory device of claim 29, wherein said control circuit comprises:

a pulse counter;

a voltage pump;

a voltage control circuit which receives a pulse count from the pulse counter and regulates said voltage pump to output an appropriate voltage level; and

a timing control circuit which receives the pulse count from the pulse counter and regulates the pulse width of said programming pulses by controlling when said pump is activated and deactivated; and

a voltage switch controlled by said timing control circuit to output said at least one programming pulse.

31. The flash memory device of claim 29, wherein the voltage levels of said plurality of programming pulses are different.

32. The flash memory device of claim 30, wherein the width of a first programming pulse is wider than the width of subsequent programming pulses

33. The flash memory device of claim 30, wherein a voltage level of a first programming pulse is less than the voltage level of subsequent programming pulses.

34. The flash memory device of claim 29, wherein said flash memory device is one of a NAND and NOR flash memory device.

35. The flash memory device of claim 29, wherein said plurality of programming pulses includes three programming pulses.

36. The flash memory device of claim 35, wherein the width of said first program pulse is wider than the width of said second and third programming pulses.

37. The flash memory device of claim 35, wherein the amplitudes of said first, second and third programming pulses are different.

38. The flash memory device of claim 35, wherein the amplitude of the third programming pulse is the larger than the first and second programming pulses and the amplitude of the first programming pulse is smaller than the amplitude of the second programming pulse.

39. A system comprising:

a processor; and

a memory device coupled to said processor, said memory device comprising:

a plurality of memory cells; and

a control circuit coupled to said memory cells for supplying a plurality of programming pulse to said memory cells, wherein the width of a first programming pulses is different from the width of subsequent programming pulses.

40. The system of claim 39, wherein said control circuit comprises:

a pulse counter;

a voltage control circuit which receives a pulse count from the pulse counter and regulates a voltage pump to output an appropriate voltage level; and

a timing control circuit which receives the pulse count from the pulse counter and regulates when said pump is activated and discharged; and

a voltage switch which is controlled by said timing control circuit to output said at least one programming pulse.

41. The flash memory device of claim 39, wherein said at least one programming pulse is a plurality of programming pulses.

42. The flash memory device of claim 40, wherein the width of said first programming pulse is wider than the other pulses of the plurality programming pulses.

43. The flash memory device of claim 39, wherein said flash memory device is one of a NAND and NOR flash memory device.

44. A device comprising:

a storage device for storing voltage levels and pulse widths of a plurality of programming pulses;

a circuit for using said stored voltage levels and pulse widths to produce said programming pulses used to program a flash memory device,

wherein at least two of said voltage levels are different and at least two of said pulse widths are different.

45. The device according to claim 44, wherein a first programming pulse of said plurality of programming pulses has a wider pulse width than the subsequent programming pulses.

46. The device according to claim 44, wherein a first programming pulse of said plurality of programming pulses has a lower voltage level than the subsequent programming pulses.

47. The device according to claim 44, wherein said flash memory device is one of a NAND and NOR flash memory device.

48. A method of programming a memory cell of a flash memory array, said method comprising:

applying first programming pulse to a plurality of memory cells;

determining whether said memory cells are programmed;

for each of said memory cells that were not programmed, applying a second programming pulse, having a higher voltage level than said first programming pulse, to said memory cells;

determining whether said memory cells are programmed after the application of said second pulse; and

for each of said memory cell that are not programmed, applying a third programming pulse, having a higher voltage level than said first programming pulse.

49. The method of claim 48, wherein said second programming pulse has a narrow pulse width than the first programming pulse.

50. The method of claim 48, wherein said third programming pulse has a pulse width less than the second programming pulse.

* * * * *