NON-VOLATILE SEMICONDUCTOR MEMORY DEVICE

Technical Field

The present invention relates to a semiconductor memory device and, particularly, to a technology that can be particularly effectively adapted to a memory system for multi-level data in a nonvolatile semiconductor memory device, such as a technology that can be effectively utilized for a nonvolatile memory device (hereinafter simply referred to as flash memory) capable of electrically erasing a plurality of stored data at one time.

Background Art

In a flash memory, a nonvolatile memory element having a control gate and a floating gate is used as a memory cell, and the memory cell is constituted by a single transistor. In such a flash memory, the programming operation is carried out by applying a voltage of about 4 V(Volt) to the drain region of the nonvolatile memory element as shown in Fig. 21, applying a voltage of about -10 V to a word line to which a control gate CG is connected, discharging the

electric charge from the floating gate FG by using a tunneling current to establish a state of a low threshold voltage (logic "0"). To carry out the erasing operation as shown in Fig. 22, a voltage of about -3 V is applied to the well region, to the drain region and to the source region, a voltage of as high as 10 V is applied to the control gate CG to generate a tunneling current in order to inject negative charge into the floating gate FG and to maintain a high threshold value (logic "1"). Thus, data of one bit is stored in a memory cell.

There has been proposed the concept of a so-called "multi-level" memory to store the data of two or more bits in a memory cell in order to increase the storage capacity. An invention related to the multi-level memory has been disclosed in, for example, Japanese Patent Application H7-14031 (14031/1995).

In such a multi-level memory of the above-mentioned prior application, the consecutive two-bit data "01", "00", "10" and "11" are subjected to the logical conversion, and are stored correspondedly to one of threshold voltages of 1.2 V or lower, 1.6 to 2.3 V, 2.8 to 3.5 V, and 4 V or higher of the memory cell as shown in Fig. 23. In the above-mentioned memory, therefore, it is impossible to decide the

stored data unless the reading operation is executed three times at all times while changing the level of the word line to, e.g., 1.4 V, 2.6 V, 3.7 V, causing a drawback of a long time taken to read the data.

The object of the present invention is to provide a multi-level storage-type nonvolatile semiconductor memory device of which the reading time can be shortened by decreasing the number of times of accesses to the word line when reading the data, and a method of driving such a device.

The above and other objects, and novel features of the present invention will become obvious from the description of the specification and the accompanying drawings.

Disclosure of the Invention

The outline of a representative example of the invention disclosed in this application will be described below.

That is, in a nonvolatile semiconductor memory device in which a plurality of threshold values are set to store multi-level data in a memory cell, each bit of multi-bit data is separately written into a memory cell depending upon an address signal or a control signal and are stored hierarchically. In this

case, data of a plurality of bits may be consecutively written in one memory cell, or after the data have been written in all memory cells bit by bit, the data of remaining bits may be successively overwritten in the memory cells.

Thus, when two bits are stored in one memory cell, the first bit can be read out by accessing the word line only one time and the succeeding bit can be read out by accessing the word lines only twice while changing the level to read the data. Thus, the total number of accesses to the word line is decreased, and the time taken to read data is shortened.

Brief Description of the Drawings

Fig. 1 is an explanatory diagram illustrating the concept (the constitution of the address space) of a data storage system in a multi-level flash memory according to the present invention;

Fig. 2 is a block diagram illustrating an embodiment of the multi-level flash memory according to the present invention;

Fig. 3 is an explanatory diagram illustrating the relationship between the threshold values of memory cells and the stored data in the multi-level flash memory according to the present invention;

Fig. 4 is a diagram illustrating the relationship between the data stored hierarchically in a memory cell and the Z-address for discriminating them in the multi-level flash memory according to the present invention;

Fig. 5 is an explanatory diagram illustrating a change in the threshold value of the memory cell at the time of programming and erasing in the multi-level flash memory according to the present invention;

Fig. 6 is a circuit diagram concretely illustrating a memory array;

Fig. 7 is a circuit diagram concretely illustrating a data latch circuit;

Fig. 8 is a flowchart illustrating a procedure for reading the multi-level flash memory according to the embodiment;

Fig. 9 is a timing chart illustrating the timings of signals in the memory array at the time of reading the multi-level flash memory of the embodiment;

Fig. 10 is a flowchart illustrating a procedure for programming the multi-level flash memory of the embodiment;

Fig. 11 is a timing chart illustrating the timings of signals in the memory array at the time of programming the multi-level flash memory of the

embodiment;

Fig. 12 is a flowchart illustrating a procedure for erasing the data from the multi-level flash memory of the embodiment;

Fig. 13 is a block diagram of a system using the multi-level flash memory of the embodiment;

Fig. 14 is a block diagram schematically illustrating a second embodiment of the multi-level flash memory according to the present invention;

Fig. 15 is a timing chart illustrating a time division method of a third embodiment of the multi-level flash memory according to the present invention;

Fig. 16 is a diagram illustrating the constitution of addresses according to a fourth embodiment in the multi-level flash memory according to the present invention;

Fig. 17 is a circuit diagram illustrating another embodiment of the memory array in the multi-level flash memory according to the present invention;

Fig. 18 is a circuit diagram illustrating a further embodiment of the memory array in the multi-level flash memory according to the present invention;

Fig. 19 is a circuit diagram illustrating a still further embodiment of the memory array in the multi-level flash memory according to the present invention;

Fig. 20 is a circuit diagram illustrating a yet further embodiment of the memory array in the multi-level flash memory according to the present invention;

Fig. 21 is a diagram schematically illustrating the structure of a memory cell used for the flash memory of the embodiment and bias conditions of when a low threshold voltage state (logic "0") is to be established;

Fig. 22 is a diagram schematically illustrating the structure of a memory cell used for the flash memory of the embodiment and bias conditions of when a high threshold voltage state (logic "1") is to be established; and

Fig. 23 is an explanatory diagram illustrating the relationship between the threshold voltage of a 4-value memory cell and the stored two-bit data according to the invention of an earlier application.

Best Mode for Carrying Out the Invention

Fig. 1 illustrates the concept of the address space in a memory capable of storing multilevel values in a memory cell, to which the present invention is applied.

Like the address system in a conventional general semiconductor memory, according to the present

invention, a memory cell can be selected according to the X-address and Y-address. According to the present invention, in addition, each memory is so constituted as to store data of a plurality of bits and, besides, correspondingly a plurality of bits (each bit is expressed by a data of "0" or "1") in a memory cell are selected or designated by an address (hereinafter referred to as Z-address) different from the X-addresses and the Y-addresses (hereinafter referred to as three-dimensional address system). In the three-dimensional address system, three methods for applying the Z-address are conceivable: ① a method in which the Z-address is input sharing the same input pins with the X- and Y-addresses, ② a method in which an independent special-purpose input pin is provided, and ③ a method in which the Z-address is input being included in a command code.

The method ① in which the Z-address is input sharing the input pin with the X- and Y-addresses includes a method in which, when the number of bits is different between the X-addresses and the Y-addresses, the pins not used when the address of the smaller number of bits is input are used for the input of the Z-address; and an address multiplex method in which the X-, Y- and Z-addresses are inputted in a time

division manner through common pins. The pin sharing method includes a method in which the pins for the data and command code are shared for the input of the X-, Y- and Z-addresses. Among the above-mentioned methods, in the method ② in which the independent special-purpose input pin is provided, the signal may be regarded as, not an address signal, but a control signal or a selection signal (the reason is that when the number of bits of data stored in a memory cell is "2", in particular, the address for discriminating it can be composed of only a single bit). Therefore, a signal for discriminating the bit in the memory cell should not be affected by the way it is called but should be judged substantially.

Fig. 2 is a block diagram of an embodiment of the case when the method ②in which the independent special-purpose input pin is employed is applied to a flash memory among the above-mentioned access methods.

The memory of this embodiment is so constituted that the operation mode such as data writing, reading or erasing operation is changed over depending upon a command code input from an external controller, and the input pins of the command code are also used as data input/output pins I/O.

In Fig. 2, reference numeral 10 denotes a memory

array in which memory cells constituted by MOSFETs having a floating gate are arranged in the form of a matrix, 11 denotes a multiplexer connected to input/output pins I/O, 12 denotes a data input buffer which receives write data through the multiplexer 11, 13 denotes a data output buffer which outputs the data read out from the memory array 10 to an external unit through the input/output pins I/O, and 14 denotes a command decode control circuit for generating internal control signals by decoding a command code given from the external unit. The command decode control circuit 14 is equipped with a control function (sequencer) for successively generating and outputting the control signals for the circuits in the memory in order to execute the processing according to the command based upon the decoding result of the command code, and automatically executes the corresponding processing by decoding the command code that is given. Like the control unit in a central processing unit (CPU) of the microprogram type, the control function is constituted by a ROM (read-only memory) in which are stored a series of micro instructions necessary to execute the command (instruction), and generates the start address of the group of micro-instructions corresponding to the command code and gives it to a micro ROM (μROM)

so that a micro program is started.

In Fig. 2, furthermore, reference numeral 15 denotes an X-address decoder which decodes the X-addresses that are input from the external unit, enables a word line in the memory array 10 to assume a selection level, and determines selection the level (Vewl-3, Vrwl-3, Vwvl-3, Vwwl-3, Vevl-3) according to the Z-address that is input, 16 denotes a Y-address decoder which decodes the Y-addresses that are input from the external unit to generate a selection signal, 17 denotes a sense latch circuit equipped with sense amplifiers of which the number corresponds to the number of data lines) and which are connected to the data lines in the memory array 10, detects the levels of the data lines and holds the programming data to drive the data lines to a potential corresponding to the data, and 18 denotes a sequence of column switches that are selectively turned on by a selection signal from the Y-address decoder 15, and connects a data line corresponding to the Y-addresses in the memory array 10 to the data input buffer 12 or to the data output buffer 13.

In addition to the above-mentioned circuits, the multi-level flash memory of this embodiment is equipped with a control signal buffer circuit 19 which

takes in control signals from the external unit and feeds them to the command decode control circuit 14, an internal power source-generating circuit 20 comprising charge pump circuits or the like for generating voltages needed in the chip, such as reading voltages Vrw1-3, programming voltages Vww1-3, erasing voltages Vew1-3 and verifying voltages Vev1-3 based upon the power source voltage Vcc supplied from an external unit, and an all decision circuit 21 for deciding whether the programming or the erasure has finished or not based upon the data in the sense latch sequence 17 at the time of the programming or erasing.

The command decode control circuit 14 further has a function for generating and outputting a ready/busy signal R/B representing whether the device is accessible from the outer side or not to the internal circuits and to the external units.

Moreover, the flash memory of this embodiment shares the external terminals (pins) I/O for the input of the programming data and the commands. For this purpose, the control signal buffer circuit 19 changes over the multiplexer 11 according to the control signals input from the external unit, discriminates the signal input through the input/output pins I/O, and feeds them to the data input buffer 12 or to the

command decode control circuit 14. The control signals inputted to the flash memory of this embodiment from the external unit include a reset signal RES, a chip selection signal CE, a programming control signal WE, an output control signal OE, a command data enable signal CDE representing that the command code and the data are valid, a serial clock SC, etc. Though there is no particular limitation, the flash memory of this embodiment is so constituted as to read the data of a row of word line held in the sense latch circuit 17 in synchronism with the clock SC, send them to an external unit, and to transfer a row of the programming data input from the external unit to the sense latch sequence 17 in synchronism with the clock SC. In this case, no Y-address is required.

Fig. 3 illustrates the relationship (definition) between the stored data (two bits) and the distribution of threshold values of an element of when one of the four threshold values is stored in a memory cell. In this embodiment, the data is written into a memory cell so as to assume any one of the four threshold values, α or lower, α to β, β to γ, or γ or higher. By the programming method that will be described later, a memory cell that has a threshold

value of α or lower is defined as a cell for storing 2-bit data "01", the memory cell that has a threshold value of α to β is defined as a cell for storing 2-bit data "00", the memory cell that has a threshold value of β to γ is defined as a cell for storing a 2-bit data "10", and the memory cell that has a threshold value of γ or higher is defined as a cell for storing a 2-bit data "11".

In Fig. 3, α, β and γ are values of the word line-reading levels at the time of reading the data. Concrete values α, β and γ are determined depending upon the power source voltage. As an example, when the power source voltage is 3.3 V, α is 1.5 V, β is 2.0 V, and γ is 2.5 V. In Fig. 3, furthermore, the hatched areas are inhibition regions which the threshold values should not enter in order to prevent erroneous reading, and $\alpha 0$, $\alpha 1$, $\beta 0$, $\beta 1$, $\gamma 0$ and $\gamma 1$ represent values of reading levels in the verifying operation that is executed during the programming or after the erasing. Out of the 2-bit data "**", the first bit (high-order bit) is the one (hereinafter referred to as the first programmed data) programmed first in time, and the second bit (low-order bit) is the one (hereinafter referred to as the second programmed data) programmed next in time.

When 2-bit data are stored in a memory cell as described above, each of the bits can be selected by using the Z-address. That is, as shown in Fig. 4, when the Z-address is "0", the first programming data is designated among the 2-bit memory data in the memory cell connected to the selected word line and, when the Z-address is "1", the second programming data is designated among the 2-bit memory data in the memory cell connected to the selected word line. When the first programmed data is read out, however, the selected word line assumes the potential β to execute the reading operation one time as will be described later. When the second programmed data is read out, the selected word line assumes the potentials α and γ to execute the reading operation two times.

Next, the method of programming data (method of changing the threshold value) applied to the flash memory of this embodiment will be described with reference to Fig. 5.

As shown in Fig. 5(A), all memory cells are brought into the erasing state (the threshold value has the second smallest value, i.e., about 1.7 V, and the memory data is "00") prior to programming the data. In establishing the erasing state, the method in which first, all memory cells are stored with the

data "11" of the greatest threshold value and, then, the threshold value Vth in all memory cells are returned to the erasing state "00" may be adopted.

Next, the first programming data is written. At this moment, for the memory cells stored with the data "1", a bias is applied to them as shown in Fig. 5(B), and the programming is so effected that the threshold value of the memory cells of 1.7 V which is in the erasing state assumes the highest threshold value of about 2.8 V. At the step where the writing of the first programming data is finished, there exists neither the memory cells having the smallest threshold value nor the memory cells having the second largest threshold value. After the programming, the word line assumes the level $\gamma 1$ to execute the verification. The memory of this embodiment can be used in a manner that the data are read by only writing the first programming data into all memory cells, not writing the second programming data, and setting the level of the word line to β. In this case, the Z-address necessarily assumes "0" (or "1"). When each of the memory regions corresponding to "0" and "1" of Z-address is defined as a plane, the above-mentioned method is the one that uses one plane only.

When two planes are used, the second programming

data is written after the first programming data is written. At this moment as shown in Fig. 5(C), the writing into the memory cell of which the first programming data is "0" is distinguished from the writing into the memory cell of which the first programming data is "1". In writing the first programming data, as described above, a bias is applied to the memory cell of which the data is "1" so as to increase the threshold value (hereinafter referred to as up-writing). In writing the second programming data as shown in Fig. 5(C), on the other hand, the data is written so as to lower the threshold value (hereinafter referred to as down-writing). In the case of the writing of the first programming data, it is possible to increase the threshold values of only the memory cells into which the data "1" among the memory cells of the state "00" is desired to be written, in order to selectively establish the state "11". It is, however, also possible to once increase the threshold values of all memory cells which are in the state "00" and to decrease the threshold value of only those memory cells of which the first programming data is "0".

In this embodiment, furthermore, when the first programming data is "0" and the second programming

data is "1", the bias is applied to the memory cells so as to lower the threshold value, so that the state is changed to a state of the smallest threshold value (the threshold value is about 1.0 V, and the memory data is "01"). When the first programming data is "1" and the second programming data is "0", the bias is applied to the memory cells so as to lower the threshold value, whereby the state is changed to a state of the second largest threshold value (threshold value is about 2.3 V, and the memory data is "10"). In this embodiment, though there is no particular limitation, the programming is, first, so effected that the first programming data is "0" and the second programming data is "1" and, then, the programming is so effected that the first programming data is "1" and the second programming data is "0". The bias voltage or the programming pulse width of when the threshold value is lowered to 1.0 V may be made different from that of when the threshold value is lowered to 2.3 V. This is because the programming characteristics are not the same.

Meanwhile, when the first programming data is "0" and the second programming data is "0" or when the first programming data is "1" and the second programming data is "1", the data line is fed with no

bias that changes the threshold value of the memory cell. After the programming, verification is effected by causing the word line to assume the levels $\alpha 0$ and $\gamma 0$. Table 1 shows how the threshold value of the memory cell changes with the programming operation. The threshold values of each memory cell are Vth1, Vth2, Vth3 and Vth4 (Vth1 < Vth2 < Vth3 < Vth4).

Table 1

First programming data		0	0	1	1
Second programming data		1	0	0	1
Change in threshold value	1st programming	no change	no change	Vth2→ Vth4	Vth2→ Vth4
	2nd programming	Vth2→ Vth1	no change	Vth4→ Vth3	no change
Final threshold value		Vth1	Vth2	Vth3	Vth4

To erase the data as shown in Fig. 5(D), up-writing of memory cells having the smallest threshold value (memory data "01") is effected, and the down-writing of memory cells having the largest threshold value (memory data "11") and memory cells having the second largest threshold value (memory data "10") is effected. Thus, the threshold values of all memory cells are changed into the second smallest state (memory data "00"). Besides, as will be described later, erasing of the second programmed data and erasing of the first programmed data may be performed.

Verification after the erasing is effected by causing the word line to assume the levels $\alpha 1$ and $\beta 0$.

Fig. 6 illustrates a concrete example of the memory array 10. The memory array of this embodiment is constituted by two mats. A sense latch circuit SLT is disposed between the two mats. The data line in the unselected mat is precharged to a half-precharge level which is half the precharge level of the data line in the selected mat. The sense latch circuit is so constituted as to detect the data in a differential way. The memory array is symmetrically with respect to the sense latch circuit SLT. Accordingly, Fig. 6 illustrates the sense latch circuit SLT and only the mat on one side.

The memory array 10 of this embodiment includes, as shown in Fig. 6, a plurality of memory sequences MCC each constituted of n memory cells (MOSFETs having a floating gate) MC1 to MCn arranged in parallel in the direction of column having sources and drains commonly connected are arranged in the row direction (direction of word lines WL) and in the column direction (direction of main data lines DL). In each memory sequence MCC, the drains and sources of n memory cells MC1 to MCn are connected to a common subdata line SDL and to a common subsource line SSL.

The subdata line SDL is connected to the main data line DL through a switching MOSFET Qs1. The subsource line SSL can be connected to a grounding point or to a negative voltage through a switching MOSFET Qw2 and a common source line CSL.

Among the plurality of memory sequences MCC, those arranged in the direction of word lines are formed in the same well region WELL over the semiconductor substrate. To increase the threshold values of the memory cells, a negative voltage such as -3 V is applied to the well region WELL and a voltage such as 10 V is applied to the word lines sharing the well region, to enable the upward writing. To erase the data, all switching MOSFETs Qst1, Qst2 sharing the well region are turned on, and the negative voltage of -3 V is applied to the sources and drains of the memory cells.

To lower the threshold value of the memory cell, on the other hand, a negative voltage such as of -10 V is applied to the word line to which a memory cell to be selected is connected, a voltage such as of 4 V is applied to the main data line DL corresponding to the memory cell to be selected, the switching MOSFET Qs1 on the subdata line SDL to which the selected memory cell is connected is turned on, and a voltage of 4 V

is applied to the drain region. However, the switching MOSFET Qs2 on the subsource line SSL is kept off.

To read the data, voltages such as of 1.5 V, 2.0 V or 2.5 V are applied to the word line to which is connected a memory cell that is to be selected, the main data line DL corresponding to the memory cell to be selected is precharged to a potential such as of 1 V, and the switching MOSFET Qs1 on the subdata line SDL to which the selected memory cell is connected is turned on. The switching MOSFET Qs2 on the subsource line SSL is turned on, and the ground potential is applied thereto. At this moment, when the selected memory cell is on, the electric charge on the corresponding data line is discharged into the common source, and the potential decreases. The potential is amplified through the sense latch circuit SLT connected to an end of the main data line DL. At the time of reading the data, the sense latch circuit SLT detects the level of the data line and holds it. At the time of programming, the sense latch circuit SLT holds the programming data and applies a potential corresponding to the data that is held to the data line.

Fig. 7 concretely illustrates the sense latch

circuits SLT connected to the data lines. As shown, each sense latch circuit SLT has a sense amplifier SA in which the input/output terminals of the two CMOS inverters are mutually connected. Between the sense amplifiers SA and the memory array 10, there are provided precharging MOSFETs Qp0, Qp1, data transfer MOSFETs Qt0, Qt1 capable of connecting/cutting off one input/output terminal of the sense amplifier SA and the main data line DL, data inversion circuits DIC, resetting MOSFETs Qr0, Qr1, a MOSFET Qj for all decision, a column switching MOSFETs Qc, a data filter circuit DFC for recognizing the first programmed data, and the like.

In Fig. 7, the reason why two data line precharging MOSFETs Qp, two data transfer MOSFETs Qt, two data inversion circuits DIC and two resetting MOSFETs Qr are shown is that those connected to oddly-numbered data lines and those connected to evenly-numbered data lines are separately shown. A selection signal from the Y-decoder circuit is applied to the gate of the column switching MOSFET Qc, and the data line is connected to the data input buffer 12 and to the data output buffer 13 through the column switching MOSFET Qc and the input/output line I/O. Though not shown, each sense amplifier SA is connected to the

common power source line and to the ground line via power source feeding/cutting off switching MOSFETs. The sense amplifier SA is activated upon turning these MOSFETs on and is inactivated upon turning these MOSFETs off.

The source terminal of the precharging MOSFET Qp0 or Qp1 is connected to the main data line DL, the drain terminals thereof is connected to the voltage feeding line VL0 or VL1, and the gate terminal thereof is fed with a precharge control signal RPC0 or RPC1. At the time of reading the data, the precharging MOSFET Qp0 or Qp1 is turned on to precharge the main data line DL to a level such as of 1 V. The data transfer MOSFETs Qt0 and Qt1 are connected between the main data line DL in the memory array 10 and the main data lines DLs on the sense amplifier SA side, and the gate terminals thereof are fed with transfer control signals TR0, TR1. After the word line is raised, the data transfer MOSFETs Qt0 and Qt1 are turned on to transmit the level of the data line to the sense amplifier SA. At the time of programming, meanwhile, the data transfer MOSFETs Qt0 and Qt1 are turned on, so that the programmed data held in the sense amplifier SA is transmitted to the main data line DL.

In the data inversion circuit DIC, two MOSFETs

Qd1, Qd2 are connected in series between the main data line DL and the voltage feeding line UP. Control signals PC0, PC1 are applied to the gate terminal of the MOSFET Q1, and the main data line DLs on the sense amplifier SA side is connected to the gate terminal of the MOSFET Qd2. The drain terminals of the resetting MOSFETs Qr0, Qr1 are connected to the main data line DLs, the source terminals thereof are connected to the ground lines GL0, GL1, and the gate terminals thereof are fed with data reset signals RSA0, RSA1. As Qd1 and Qd2 are turned on, the main data line DLs is discharged to the ground potential.

The MOSFET Qj for all decision is connected between the sense lines EC0, EC1 and the ground line GL2, the gate terminal thereof is connected to the main data line DLs on the sense amplifier SA side, and a current sense circuit CSC constituting the all decision circuit 21 is connected to one end of each of the sense lines EC0, EC1. Upon detecting whether the data lines are all at the low level or not, it is decided that the data read out are all "0". That is, when any one of the data read out is "1", the corresponding MOSFET Qj is turned on, and a current flows to the ground point through the sense lines EC0, EC1, and is hence detected by the current sense

circuit CSC.

The data file circuit DFC for recognizing the first programming data is constituted by MOSFETs Qf1, Qf2 connected in series between the voltage feeding line DCP and one input/output terminal of the sense amplifier SA, and a transfer MOSFET Qf3 connected in series with the column switching MOSFET Qc. A control signal DC0 is applied to the gate terminal of the MOSFET Qf1, and a control signal DC1 is applied to the gate terminal of the MOSFET Qf2. The MOSFETs Qf1 to Qf3 are controlled by the control signals and by the data held in the sense amplifier SA, in order to write the second programming data that corresponds to the above-mentioned first programming data. That is, the control signal DC0 is set to the high level and the DC1 is set to the low level to turn the Qf1 on and the Qf2 off, whereby a signal of the high level is supplied to the gate of Qf3 to thereby connect all data lines to the input/output lines I/O. Meanwhile, the control signal DC0 is set to the low level and the DC1 is set to the high level to turn the Qf1 off and the Qf2 on, whereby a signal of a level corresponding to the data held in the sense amplifier SA is fed to the gate of Qf3, and the data lines are selectively connected to the input/output lines I/O.

Next, with reference to a flowchart of Fig. 8 a procedure for reading the data from the memory cell of the embodiment will be described. Fig. 9 illustrates changes in the signal lines in the memory array and in the sense latch circuit of when the data is to be read out. Symbols shown in Fig. 8 correspond to symbols of signals shown in Figs. 6 and 7. R or L of the last letter of some of the symbols is for discriminating whether the signal relates to which mat out of the upper and lower mats on both sides of the sense latch circuit SLT. PRL and PRR represent signals for precharging the data lines, and correspond to the code RPC in Fig. 7. Furthermore, DPB denotes an on/off control signal of the power source for sense latch.

Referring to Fig. 8, the reading operation is started upon the input of a read command. When the command that is input is a read command, an address signal inputted just after or simultaneously with the command (step S1) is fetched. It is then decided whether the data to be read out is the first programmed data or the second programmed data in the memory cell based upon the Z-address among the addresses that are fetched (step S2). When it is decided that the command is for reading the first programmed data, the level of the word line is set to

β which is the intermediate read-out level, the main data line DL is precharged, and the X-address is decoded to raise the word line WL (steps S3, S4). At this moment, furthermore, the switching MOSFET Qs1 of the subdata line is turned on.

Thus, the level of the data line is held or discharged depending upon the first programmed data in the memory cell connected to the selected word line and, hence, the switching MOSFET Qs2 on the subsource line and the data transfer MOSFET Qt are turned on, so that the level of the data line is transmitted to the sense amplifier SA and the power source is supplied to the sense amplifier to activate it (step S5). Thus, the read-out data is amplified and held in the sense amplifier, whereby the column switch Qc is turned on, and the read-out data is sent to the data output buffer and output to an external unit (step S6).

When it is decided that the data to be read out at the step S2 is the second programmed data in the memory cell, the procedure proceeds to a step S7 whereby the level of the word line is set to γ which is the highest read-out level, the main data line DL is precharged and, at the same time, the X-address is decoded to raise the word line WL (step S8). Then, the level of the data line is transmitted to the sense

amplifier SA to activate it and to amplify the data that is read out (step S9). Then, the data inversion circuit DIC is controlled, the data on the data line is inverted by utilizing the data held by the sense amplifier SA and is used for the precharging operation at the next reading time (step S10). Thereafter, the level of the word line is set to α which is the lowest read-out level, and the X-address is decoded to raise the word line WL (step S11, S12). Thereafter, the level of the data line is transmitted to the sense amplifier SA, and the data that is read out is amplified by the sense amplifier (step S13). Next, the data inversion circuit DIC is controlled and the data on the data line is inverted by utilizing the data held in the sense amplifier SA (step S14). The data is amplified again by the sense amplifier SA, whereby the desired second programming data is held by the sense amplifier. Therefore, the column switch Qc is turned on to send the read-out data to the data output buffer and to output it to an external unit (step S15).

In the following Table 2, L and H represent changes in the level of the data line of when the second programmed data (1001) is read out from the memory cells where the data "01", "00", "10" and "11"

are stored at the above-mentioned steps S8 to S15. Here, L represents that the data line is at the low level and H represents that the data line is at the high level. For comparison, in the lowest space, there are shown the levels of the data line of when the first programmed data (0011) are read out from the memory cells stored with the same data by allowing the word line to be at the level β. It will be understood from Table 2 that the data are correctly read out by the control according to the flowchart of Fig. 8.

Table 2

1	Stored data	01	00	10	11
2	read out with γ	L	L	L	H
	data inverted	H	H	H	L
	read out with α	L	H	H	L
	data inverted	H	L	L	H
	read-out data	1	0	0	1
3	read out with β	L	L	H	H

Next, the procedure for writing data into the memory cell will be described with reference to a flowchart of Fig. 10. It is assumed that each memory cell has been erased and has the second lowest threshold value prior to starting the programming. Fig. 11 illustrates changes in the signal lines in the memory array and in the sense latch circuit in programming the data. Symbols shown in Fig. 11

correspond to the symbols of signals shown in Figs. 6 and 7. R or L of the last letter of some of the symbols is for discriminating whether the signal relates to which mat out of the upper and lower mats on both sides of the sense latch circuit. YGi denotes a gate control signal of the column switch Qc, and PWL denotes a signal for precharging the data line and corresponds to the symbol PC in Fig. 7.

As shown in Fig. 10, the programming operation is started upon the receipt of a write command from an external unit. When the command that is input is a write command, an address signal and a programming data inputted just after or simultaneously with the command (step S21) is fetched. Then, based upon the Z-address among the addresses that are fetched, it is decided whether the programming corresponds to the first programming or to the second programming of the memory cell (step S22). When it is decided to be the first programming, the programming data is transferred to the sense latch (step S23). In this case, the data filter circuit DFC is so controlled as to turn Qf1 on and to turn all of Qf3 on, and unconditionally transfers the programming data to the sense amplifier SA. Next, the programming is effected so as to increase the threshold value of the memory cell of

which the programming data is "1" (up-writing)(step S24).

The operation of increasing the threshold value of the memory cell is effected by applying, for example, 10 V to the word line and, for example, -3 V to the well region. In the memory array constituted as in the above-mentioned embodiment, all memory cells in a sector shares the word lines and the well regions and, hence, it is impossible to selectively increase the threshold value of the memory cells. In the above-mentioned up-writing operation (step S24), therefore, the threshold value of all memory cells sharing the word lines in a sector, is once increased. That is, the operation is effected so that the state "00" of the memory cells is shifted to the state "11". Then, by using the data latched by the sense latch SA, the level of the selected word line is so set that a bias voltage shown in Fig. 21 is applied to memory cells that are desired to be placed in the state "00", and the switching MOSFET Qs1 and the data transfer MOSFET Qt are turned on to set the level of the selected data line, thereby discharging the electric charge from the floating gate and lowering the threshold value.

Thereafter, the level of the word line connected

to the memory cell into which the data is written is set to $\beta 0$ to raise the selected word line WL and to execute the verification (step S25). The verification is the same as the above-mentioned operation for reading the first programmed data and will be not described in detail. After the verification, it is decided by the current sense circuit CSC whether the data latched by all sense amplifiers SA are all "0" or not (step S26). When the normally programmed memory cell is verified, the data held by the sense amplifier changes from "1" to "0". The data "1" remains in the sense amplifier corresponding to a memory cell which has not been sufficiently programmed. When all of the data of the sense latches are not "0", the procedure returns to the step S24, where the programming is effected again by using the data held by the sense latches. The programming operation ends when it is decided through the verification operation that the data of the sense latches are all "0".

On the other hand, when it is decided that the programming to be executed at the step S22 is the second programming of the memory cell, the procedure proceeds to a step S27. In this embodiment as described earlier, the second programming must be separately executed depending upon the first

programmed data. At the step S27, therefore, the first programmed data is read out, first. The level of the word line at this moment is β. To read the data, the main data line DL is precharged and, then, the word line WL is raised in the same manner as during the reading operation. Thereafter, the data inversion circuit DIC is controlled to invert the data read out by the sense amplifier SA (step S28).

Next, the data filter circuit DFC is controlled (Qf2 is turned on) so that the sense amplifier SA fetches the input programming data according to the data held by the sense amplifier SA (step S29). At this moment, the data of the logical product (AND) of the inverted read-out data and the input data is held by the sense amplifier SA. The following Table 3 shows changes in the data of when the first programmed data is read out from a memory stored with the data "0011", and the second programming data "1100" is input thereto at the above-mentioned steps S27 to S29. It will be understood from Table 3 that the data "1" remains only in a sense amplifier corresponding to a memory cell of which the data "00" is changed to "01" in the upper part of Fig. 5(C).

Table 3

First programmed data	0	0	1	1
Inverted data	1	1	0	0

Input data	1	0	0	1
Final data	1	0	0	0

Thereafter, the level of the selected word line and the level of the selected data line are so set that a bias voltage shown in Fig. 21 is applied to a memory cell in which the programmed data held by the sense amplifier is "1", and the switching MOSFET Qs1 and the data transfer MOSFET Qt are turned on in order to discharge the electric charge from the floating gate and to lower the threshold value (step S30).

Thereafter, the level of the word line through which programming is performed is set to $\alpha 0$, and the selected word line WL is raised to execute the verification (step S31). After the verification, it is decided by the current sense circuit CSC whether the data of all sense latches are all "0" (step S32). When the memory cell that is normally programmed is verified, the data held by the sense amplifier SA changes from "1" to "0", and the data "1" remains in the sense amplifier SA that corresponds to the memory cell which has not been sufficiently programmed. When the data of the sense latches are not all "0", therefore, the procedure returns to the step S30, where the programming is effected again by using the data held in the sense latches. When it is decided by the verification operation that the data of the sense

latches are all "0", the first programming operation ends.

Next, the procedure proceeds to a step S33, and the second programming data is fetched again by the sense amplifier SA. At this moment, the data filter circuit DFC is so controlled as to turn the Qf1 on and to turn all Qf3 on, and unconditionally transfers the programming data to the sense amplifier. Then, the data inversion circuit DIC is controlled to invert the data read out by the sense amplifier SA (step S34). Then, the data line is precharged by using the data held by the sense amplifier SA to read out the data (step S35). At this moment, the level of the word line is β which is the same level as for reading out the first programmed data. Therefore, the memory cell of which the first programmed data is "0", is turned on, whereby the data line is discharged and the data in the corresponding sense amplifier turns to "0". The following Table 4 shows changes in the data of when the first programmed data is read out from a memory cell stored with the data "0011", and the second programming data "1100" is input thereto at the above-mentioned steps S33 to S35. It will be understood from Table 4 that the data "1" remains only in a sense amplifier corresponding to a memory cell of

which the data "11" is changed to "10" in the lower part of Fig. 5(C).

Table 4

First programmed data	0	0	1	1
Input data	1	0	0	1
Inverted data	0	1	1	0
Final data	0	0	1	0

Thereafter, the level of the selected word line and the level of the selected data line are so set that a bias voltage shown in Fig. 21 is applied to a memory cell in which the programmed data held by the sense amplifier is "1", and the switching MOSFET Qs1 and the data transfer MOSFET Qt are turned on in order to discharge the electric charge from the floating gate and to lower the threshold value (step S36).

Thereafter, the level of the programmed word line is set to $\gamma 0$, and the selected word line WL is raised to execute the verification (step S37). After the verification, it is decided by the current sense circuit CSC whether the data of all sense latches are all "0" (step S38). When the memory cell that is normally programmed is verified, the data held by the sense amplifier SA changes from "1" to "0", and the data "1" remains in the sense amplifier that corresponds to the memory cell which has not been

sufficiently programmed. When the data of the sense latches are not all "0", therefore, the procedure returns to the step S36, where the programming is effected again by using the data held in the sense latches. When it is decided by the verification operation that the data of the sense latches are all "0", the programming operation ends.

In the above description, to write the first programming data (up-writing) in the above-mentioned embodiment, the threshold value of only the memory cell which is desired to be maintained in the state "00" is lowered after the threshold value of all memory cells is once increased. It is, however, also possible to execute the programming operation which selectively increases the threshold value of only the memory cell into which the first programming data "1" is desired to be written by the method described below.

That is, the only thing to be done is that a voltage is applied to the drain of a memory cell of which the threshold voltage is not to be increased, and a channel is formed just under the gate electrode to block the injection of an electric charge into the floating gate. Concretely speaking, a voltage of, for example, 10 V is applied to the word line and a

voltage of, for example, -3 V is applied to the well region. Then, a voltage of, for example, 3 V is applied to the drain of a memory cell which is not desired to be changed from the state "00" to the state "11", and a voltage of, for example, -3 V is applied to the drain of a memory cell which is desired to be changed from the state "00" to the state "11". The level of the word line at the time of verification is $\gamma 1$.

Next, the procedure for erasing the data stored in the memory cell will be described with reference to the flowchart of Fig. 12.

As shown in Fig. 12, the erasing operation is started upon the input of an erase command from an external unit. When the command that is inputted is an erase command, address signals inputted just after or simultaneously with the command (step S41). It is then decided, based on the Z-address among the fetched addresses, whether the data to be erased is the first programmed data or the second programmed data in the memory cell (step S42). When it is decided that it is the first programmed data, the second programmed data is read out from a memory cell designated by the X-address and is output to the outside of the memory (step S43). The second programmed data output to the

external unit is stored in the buffer memory by a controller or the like.

Next, the first programmed data is read out from the memory cell designated by the input address and held in the sense latch (step S44). By using the thus held data, the level of the selected word line and the level of the selected data line are so set that a bias voltage shown in Fig. 21 is applied to a memory cell of which the read-out data is "1", whereby the switching MOSFET Qs1 and the data transfer MOSFET Qt are turned on, discharging the electric charge from the floating gate and lowering the threshold value to effect the erasing (step S45).

Thereafter, the level of the erased word line is set to $\beta 0$ to raise the selected word line WL and to effect the verification (step S46). The verification operation is the same as the above-mentioned operation for reading the data (Fig. 8), and will be not described in detail. After the verification, it is decided by the current sense circuit CSC whether the data of all sense latches are all "0" or not (step S47). When the normally programmed memory cell is verified, the data held in the sense latch changes from "1" to "0", and the data "1" remains in the sense latch corresponding to the memory cell that has not

been properly programmed. When the data of the sense latches are not all "0", the procedure returns to the step S45 where the erasing is effected again by using the data held by the sense latch.

When it is decided through the verification operation that the data of the sense latches SA are all "0", the level of the word line is set to $\alpha 1$ and the selected word line WL is raised to effect the verification (step S48). After the verification, it is decided whether the data of all sense latches are all "1" (step S49). The decision of all "1" is easily effected by an all-decision circuit constituted by a MOSFET Qj provided on the opposite side to the sense latch SA and a current sense circuit CSC, or by providing a similar circuit constituted by a P-channel MOSFET instead of the MOSFET (N-channel) Qj of Fig. 7. When the data of sense latches are not all "1", the procedure proceeds to a step S50, where the programming is so effected, by using the data held by the sense latch, as to increase the threshold value of the memory cell of which the read-out data is "0" and, then, the verification is effected again (step S51). When it is decided through the verification operation that the data in the sense latches are all "0" (step S52), the second programmed data stored in the

external buffer memory is input and written into the memory (step S53). This programming operation is executed according to a procedure similar to the flowchart of Fig. 10.

When it is decided that the data to be erased at the step S42 is the second programmed data of the memory cell, the procedure proceeds to a step S60 for erasing the second programmed data. The processing of erasing the second programmed data is carried out by substantially the same procedure as the processing of erasing the first programmed data at the steps S41 to S51. The difference is that the first programmed data is read out prior to erasing the second programmed data and saved in an external buffer memory, and in erasing the second programmed data, the erasing for increasing the threshold value is effected separately from the erasing for lowering the threshold value depending upon the first programmed data. Since the erasing for increasing the threshold value and the erasing for lowering the threshold value are separately executed depending upon the first programmed data, a control operation similar to the one executed at the steps S27 to S38 in the flowchart for executing the programming of Fig. 10 is adopted.

Fig. 13 illustrates the constitution of a system

using the multi-level flash memory of the above-mentioned embodiment.

Though there is no particular limitation, the system of this embodiment is constituted by a microcomputer (CPU) 110 for controlling the whole system, a read-only memory (ROM) 120 for storing the program and fixed data of the system, a random access memory (RAM) 130 in/from which data can be randomly written/read to provide a working area to the CPU 110, a local memory 140 including multi-level flash memories 141, 142, ... of the above-mentioned embodiment, a memory controller 150 for controlling the read, write and erase of the data in the local memory 140, and a buffer memory 160 used for saving the second programmed data or the first programmed data at the time of erasing or rewriting the data in the local memory 140.

The CPU 110, the ROM 120, the RAM 130 and the memory controller 150 are connected together through a main bus 170 which includes a control bus 171, an address bus 172 and a data bus 173. The flash memories 141, 142, ..., the memory controller 150 and the buffer memory 160 are connected together through a local bus 180 which includes a clock bus 181, an address bus 182 and an I/O bus 183. Though there is

no particular limitation, these ICs and the bus are mounted on a single board such as a printed wiring board.

The memory controller 150 includes an interface circuit 151 for interface with the CPU 110, a command controller 152 which generates and outputs a read-out command, a write command, an erase command and a data transfer clock to the flash memories 141, 142, ... in response to an instruction from the CPU 110, an address controller 153 which converts addresses into physical addresses of the flash memories 141, 142, ... based upon the logical addresses output from the CPU 110, and converts them into the above-mentioned three-dimensional addresses, and a data I/O controller 154 which transfers the data read out from the flash memories 141, 142, ... to the RAM 130, and transfers the programming data from the RAM 130 to the flash memories 141, 142, ...

Figs. 14 to 17 illustrate other embodiments of the system for discriminating the data depending upon the hierarchy in the memory cell (corresponding to the method based on the Z-address of the above-mentioned embodiment).

Among them, Fig. 14 illustrates a constitution for inputting, in a time division manner, the Z-

address together with the data, command, X-address and Y-address through the input/output pins I/O, instead of inputting the Z-address through special-purpose external terminals (see symbol AZ of Fig. 2). The Z-address inputted to the input/output pins I/O is fed to an X-address decoder 15 through a multiplexer 11. Likewise, X-address and Y-address are fed to the X-address decoder 15 and to the Y-address decoder 16 through the multiplexer 11. Command or data and address are distinguished from each other using a control signal CDE from an external unit. The command and data are distinguished from each other in a manner that the command is inputted in synchronism with a write enable signal WE, and the data is inputted in synchronism with a serial clock SC.

Relative to even the address multiplexing method in which the above-mentioned pins are commonly used, there can be contrived a method for separately fetching the X-address and the Y-address at different timings and a method for fetching the Z-address by utilizing a free bit in the X-address as shown in Fig. 15. The method shown in Fig. 16 is to illustrate the address constitution of when the Z-address (1 bit) is to be fetched by using 1 bit between the free two bits of the highest order in the case where the X-address

of 14 bits is to be fetched in two times by using eight input/output pins I/O.

It is also allowable to input the Z-address included in a command instead of the Z-address included in the X-address. Moreover, the read command, write command and erase command may be separately prepared for each of the data depending the hierarchy in the memory cell. Even in this case, as shown in Fig. 14, there can be contrived a method for inputting the address and the command in a time division manner as shown in Fig. 14 and a method for inputting them through separate external terminals as shown in Fig. 2. In addition to the address or command, furthermore, a control signal or a selection signal may be used for discriminating the data depending upon the hierarchy in the memory cell.

Figs. 17 to 20 illustrate other constitutions of the memory array 10. In the memory array of the above-mentioned embodiment (see Fig. 6), the sources of n memory cells MC arranged in the direction of data lines are connected to the common subsource line SSL, the drains are connected to the common subdata line SDL, the subdata line SDL is connected to the data line through the switching MOSFET Qs1, and the subsource line SSL is connected to the common source

line CSL through the switching MOSFET Qs2. In the embodiment of Fig. 17, the memory cells MC are arranged in pairs in the direction of data lines, a common drain CD of each pair is connected to a corresponding main data line DL, and the sources of the memory cells in the direction of the word lines WL are connected to the common source line CSL for every memory row. The sense latch circuit SLT connected to each data line is the same as that of the memory array (see Fig. 6) of the above-mentioned embodiment. In the memory array of this embodiment, the reading/programming operations are carried out in quite the same manner as of the memory array (see Fig. 6) of the above-mentioned embodiment except the control operation by the switching MOSFETs Qs1, Qs2. It is further possible to provide a subdata line and a subsource line so as to be controlled by the switching MOSFETs Qs1 and Qs2.

Fig. 18 is a memory array of the so-called NAND type. N memory cells MC1 to MCn are so arranged in the direction of the main data line DL that their channels are connected in series, one end thereof (lower end in the drawing) is connected to the main data line DL via the switching MOSFET Qs1, and the other (upper end) thereof is connected to the common

source line CSL via the switching MOSFET Qs2. In this embodiment, the potential of the word line other than the selected word line is set to such a high level that the switching MOSFETs are turned on irrespective of the threshold value of the memory cell connected thereto, and the selected word line is set to a potential corresponding to the first programmed data or the second programmed data, thereby reading or programming the data. Thus, the selected memory cell performs the reading/programming operation.

Referring to Fig. 19, the n memory cells MC1 to MCn arranged in the direction of the main data line DL are so arranged that their channels are connected in series, the common sources and the common drains are alternatingly connected to the subdata lines SDL and to the subsource lines SSL, and the subdata lines SDL are connected to the corresponding main data lines DL through the switching MOSFET Qs1. This embodiment is the same as of the memory array (see Fig. 6) of the above-mentioned embodiment except that the mutually adjacent memory cells in the direction of the main data line DL share the source and the drain, and no switching MOSFET Qs2 is used. The reading/programming operations are carried out by the same control except the control of Qs2. It is, however, also possible to

provide Qs2 between the subsource line SSL and the common source line CSL.

Fig. 20 illustrates a modification of the above-mentioned embodiment (Fig. 6) and is the same in that the drains of the n memory cells MC arranged in the direction of the data lines are connected to a common subdata line SDL which is connected to the data line through the switching MOSFET Qs1. In this embodiment, however, the memory cells are connected in pairs in the direction of the word lines, and their common sources are connected to the subsource lines SSL to decrease the number of the subsource lines SSL. The subsource line SSL is connected to the common source line CSL through the switching MOSFET Qs2. The sense latch circuit SLT may be constituted quite in the same manner as that of Fig. 7, and executes the reading/programming operation by the same control.

As described above, in the embodiments, in a nonvolatile semiconductor memory device in which data of multiple levels are stored in memory cells by setting a plurality of threshold values, and the bits of 2-bit data are separately written into one memory cell depending upon the address signal or the control signal. At the time of reading the data, therefore, the first bit (first programmed data) is read out by

accessing the word line only one time, and the second bit (second programmed data) is read out by accessing the word line two times while changing the level. This makes it possible to decrease the total number of accesses to the word line and to shorten the time taken to read the data.

In the above-mentioned embodiments, furthermore, the data inversion circuit is provided between a data line in the memory array and the sense amplifier that amplifies the read-out data, making it possible to efficiently read out the second programmed data.

In the above-mentioned embodiments, furthermore, the data filter circuit is provided between the data input buffer and the sense amplifier to control the input of programming data depending upon the data read out by the sense amplifier, making it possible to efficiently write the second programming data in accordance with the first programming data.

In the above-mentioned embodiments, furthermore, the address signal for discriminating the data depending upon the hierarchy in the memory cell is input in a time division manner together with the command or the data, making it possible to decrease the number of the required external terminals (pins).

In the foregoing the invention accomplished by

the present inventors has been described by way of embodiments, but the invention is in no way limited to them and can be modified in a variety of ways without departing from the spirit and scope of the invention. For example, the above-mentioned embodiments have dealt with the case where 2-bit data is stored in a memory. The invention, however, can also be adapted even to a non-volatile memory in which data of three or more bits is stored in a memory cell.

In the above-mentioned embodiments, furthermore, the memory cell having the second lowest threshold value is regarded as a memory cell of the erased state. However, a threshold value of any other level may be employed to establish the erased state. In the embodiments, furthermore, the threshold value is changed by effecting the programming of a memory cell that corresponds to a sense amplifier which is holding the data "1". It is, however, also possible to change the threshold value by effecting the programming of a memory cell that corresponds to a sense amplifier which is holding the data "0".

Industrial Applicability

The foregoing description has been made relative to the case where the invention accomplished by the

present inventors is applied to a flash memory in the field which is the background of the invention. However, the present invention is in no way limited thereto only but can be utilized for general nonvolatile memory devices using FAMOSs as memory elements as well as to semiconductor devices equipped with memory cells having a plurality of threshold values.

CLAIMS:

1. A method of driving a nonvolatile semiconductor memory device in which memory cells have two or more threshold values and from which data are read out by changing the level of the word lines in two or more steps to store the data of two or more bits in a memory cell, wherein bits of multi-bit data are separately written into a memory cell depending upon an address signal or a control signal, to effect the reading and erasing.

2. A method of driving a nonvolatile semiconductor memory device according to claim 1, wherein said memory cell can be stored with the data of two or more bits, and the erasing state thereof is made to correspond to the second lowest threshold value thereof.

3. A method of driving a nonvolatile semiconductor memory device according to claim 2, wherein the bit data written earlier into the memory cell is made to correspond to a threshold value of the second lowest state and to a threshold value of the highest state.

4. A method of driving a nonvolatile semiconductor

memory device according to claim 3, wherein the bit data written, later, into the memory cell is made to correspond to a threshold value of the lowest state and to a threshold value of the second highest state.

5. A method of driving a nonvolatile semiconductor memory device according to claim 4, wherein in programming the data later into the memory cell, the programming from the threshold value of the second lowest state to the threshold value of the lowest state is effected first, and the programming from the threshold value of the highest state to the threshold value of the second highest state is effected next.

6. A method of driving a nonvolatile semiconductor memory device according to claim 1, 2, 3, 4 or 5, wherein the word line is selected based on a first address signal, a desired memory cell is selected based on a second address signal out of a plurality of memory cells that use the word lines in common, and the bits of a multi-bit data in a memory cell are executed based on a third address signal.

7. A method of driving a nonvolatile semiconductor memory device according to claim 1, 2, 3, 4, 5 or 6,

* C L M S *

wherein in rewriting the first programming data that is programmed first among the bit data stored in the memory cell, the second programmed data written later is read out and stored in a buffer memory, the memory cell is erased, then, a new first programming data is written and, thereafter, the second programming data stored in the buffer memory is written.

8. A method of driving a nonvolatile semiconductor memory device according to claim 1, 2, 3, 4, 5 or 6, wherein in rewriting the second programming data that is programmed later among the bit data stored in the memory cell, the first programmed data written earlier is read out and stored in a buffer memory, the memory cell is erased, then, the first programming data that is stored in the buffer memory is written and, thereafter, a new second programming data is written.

9. A method of driving a nonvolatile semiconductor memory device according to claim 1, 2, 3, 4, 5, 6, 7 or 8, wherein in changing a threshold value of any memory cell from a low state into a high state, the threshold values of all memory cells in the same sector are once increased, and the threshold values of the memory cells that need not be increased are

lowered.

10. A method of driving a nonvolatile semiconductor memory device according to claim 1, 2, 3, 4, 5, 6, 7 or 8, wherein when the memory cells are constituted by MOSFETs having a control gate and a floating gate and when the memory cells sharing the word line are formed in the same well region, the drain voltage of the memory cells of which the threshold value must be increased are made different from that of the memory cells in the same sector of which the threshold value needs not be increased at the time of changing the threshold value of any memory cell from the low state into a high state, and a predetermined bias voltage is applied across the control gate and the well region.

11. A nonvolatile semiconductor memory device in which memory cells have two or more threshold values and are read out by changing the level of the word lines in two or more steps to store the data of two or more bits in a memory cell, wherein the data is hierarchically stored in the memory cell for each of the bits, and the data of each layer is discriminated by an address signal or a control signal.

* C L M S *

12. A nonvolatile semiconductor memory device according to claim 11, wherein a special-purpose terminal is provided for inputting an address signal or a control signal for discriminating hierarchical data in said memory cell.

13. A nonvolatile semiconductor memory device according to claim 11, wherein an address signal or a control signal for discriminating the hierarchical data in said memory cell is input through a common external terminal together with other address signals or data signals in a time division manner.

14. A nonvolatile semiconductor memory device according to claim 11, 12 or 13, wherein a data inversion circuit is provided between a data line in the memory array and a sense amplifier that amplifies the read-out data.

15. A nonvolatile semiconductor memory device according to claim 11, 12, 13 or 14 wherein a data filter circuit is provided between said sense amplifier and the data input buffer so as to control the input of programming data according to the data read out by the sense amplifier.

* C L M S *

16. A data processing system comprising a microcomputer, a nonvolatile semiconductor memory device of claim 11, 12, 13, 14 or 15, a controller for controlling the reading, programming and erasing of said memory device, and a buffer memory device for temporarily storing the data read out from said memory device, wherein said microcomputer and said control device are connected together through a main bus, and said controller and said memory device are connected together through a local bus.