

PERFORMANCE AND AREA EFFICIENT SYNAPSE MEMORY CELL STRUCTURE

Technical Field

[0001] The present invention relates to a performance and area efficient synapse memory cell structure. In particular it provides a system and method for operating a synapse memory cell structure.

BACKGROUND

[0002] An analog memory, which has characteristics such that its resistance and capacitance can be controlled continuously, is typically used as a memory device to store a synapse weight value in a neural network system. In order to achieve good performance as neural network system, synapse memory cell preferably has high resolution. Many memory devices have low resolution or a small dynamic range. It is hard to improve them because of their device characteristic and noise margin. Some memory devices have rectangle shape with large aspect ratio that are not suited for integration into an array block. Therefore, there is a need in the art to address the aforementioned problem.

SUMMARY

[0003] Viewed from a first aspect, the present invention provides a synapse memory system comprising: a plurality of synapse memory cells provided at cross points of a plurality of axon lines and a plurality of dendrite lines, each synapse memory cell comprising a plurality of analog memory devices, each synapse memory cell being configured to store a weight value according to an output level of a write signal, the weight value to be stored in each synapse memory cell being one value of at least three different values, the plurality of analog memory devices being combined to constitute each synapse memory cell; a write portion configured to write the weight value to each synapse memory cell, the write portion comprising a write driver and an output controller, the write driver being configured to output the write signal to a subject synapse memory cell, the subject synapse memory cell being one of the synapse memory cells, the subject synapse memory cell being selected to store the weight value, the output controller being configured to control the output level of the write signal of the write driver; and read drivers configured to read the weight value stored in the synapse memory cells.

[0004] Viewed from a further aspect, the present invention provides a device comprising a synapse memory system, wherein the synapse memory system comprises: a plurality of synapse memory cells provided at cross points of a plurality of axon lines and a plurality of dendrite lines, each synapse memory cell comprising a plurality of analog memory devices, each synapse memory cell being configured to store a weight value according to an

output level of a write signal, the weight value to be stored in each synapse memory cell being one value of at least three different values, the plurality of analog memory devices being combined to constitute each synapse memory cell; a write portion configured to write the weight value to each synapse memory cell, the write portion comprising a write driver and an output controller, the write driver being configured to output the write signal to a subject synapse memory cell, the subject synapse memory cell being one of the synapse memory cells, the subject synapse memory cell being selected to store the weight value, the output controller being configured to control the output level of the write signal of the write driver; and read drivers configured to read the weight value stored in the synapse memory cells.

[0005] Viewed from a further aspect, the present invention provides a method for operating a synapse memory system, the synapse memory system comprising: a plurality of synapse memory cells provided at cross points of a plurality of axon lines and a plurality of dendrite lines, each synapse memory cell comprising a plurality of analog memory devices; a write portion comprising a write driver and an output controller; and a set of read drivers; the method comprising: storing, by a synapse memory cell of the system, a weight value according to an output level of a write signal, the weight value stored being one value of at least three different values; writing, by the write portion, the weight value in the synapse memory cell, the writing comprising: outputting, by the write driver, the write signal to the synapse memory cell selected to store the weight value, and controlling, by the output controller, the output level of the write signal of the write driver; and reading, by the read drivers, the weight value stored in the synapse memory cells.

[0006] According to an aspect of the present invention, there is provided a synapse memory system. The system includes plural synapse memory cells, a write portion, and read drivers. The plural synapse memory cells are provided at cross points of plural axon lines and plural dendrite lines. Each synapse memory cell includes plural analog memory devices. Each synapse memory cell is configured to store a weight value according to an output level of a write signal. The weight value to be stored in each synapse memory cell is one value of at least three different values. The plural analog memory devices are combined to constitute each synapse memory cell. The write portion is configured to write the weight value to each synapse memory cell. The write portion includes a write driver and an output controller. The write driver is configured to output the write signal to a subject synapse memory cell. The subject synapse memory cell is one of the synapse memory cells. The subject synapse memory cell is selected to store the weight value. The output controller is configured to control the output level of the write signal of the write driver. The read drivers are configured to read the weight value stored in the synapse memory cells.

[0007] According to another aspect of the present invention, there is a device including the aforementioned synapse memory system.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The present invention will now be described, by way of example only, with reference to preferred embodiments, as illustrated in the following figures:

[0009] FIG. 1 shows a synapse memory having a cross-bar array according to a first exemplary embodiment of the present invention.

[0010] [0001]FIG. 2 shows a synapse memory system according to the first exemplary embodiment.

[0011] FIG. 3 shows a timing chart for illustrating timings of signal outputs.

[0012] [0002]FIG. 4 shows the synapse memory and synapse memory cells each composed of multiple memory devices.

[0013] FIG. 5A shows an exemplary structure of an analog memory composed of an MRAM. FIG. 5B shows the state where the analog memory of FIG. 5A stores the largest (or the smallest) value. FIG. 5C shows the state where the analog memory of FIG. 5A stores the smallest (or the largest) value.

[0014] FIG. 6 shows an exemplary configuration where the analog memory shown in FIGS. 5A to 5C is used for each memory device in the synapse memory cell.

[0015] FIG. 7 shows an example of a device which includes a neuro-synaptic core system.

DETAILED DESCRIPTION

[0016] Hereinafter, exemplary embodiments of the present invention will be described in detail with reference to the attached drawings. It is to be noted that the present invention is not limited to the exemplary embodiments to be given below and may be implemented with various modifications within the scope of the present invention. In addition, the drawings used herein are for purposes of illustration, and do not show actual dimensions.

First exemplary embodiment

[0017] FIG. 1 depicts a synapse memory 100 having a cross-bar array according to a first exemplary embodiment of the present invention.

[0018] [0003]Hardware implementation of a neuromorphic system may include synapse memories as well as neuron bodies and network connection with axons and dendrites. As shown in FIG. 1, the synapse memory 100 may include synapse memory cells 10 arranged/placed at all cross points of all axons (axon lines) 20 and all dendrites (dendrite lines) 30. Each of the synapse memory cells 10 may be configured to store a synapse weight value, which indicates a weight of synapse connection of the corresponding synapse memory cell 10. Note that the axons 20 correspond to respective axons of pre-neurons and the dendrites 30 correspond to respective dendrites of post-neurons.

[0019] The memory device for the synapse memory cell 10 may be a non-volatile random access memory (NVRAM), which is an analog memory. The synapse memory cell 10 will be described in detail later.

[0020] FIG. 2 depicts a synapse memory system 1 according to the first exemplary embodiment.

[0021] As shown in FIG. 2, the synapse memory system 1 may include the synapse memory 100, a write driver 40, a write line selector 45, read drivers 50, a read driver selector 150, a weight evaluator 70, a weight encoder 80, and a write pulse generator 90. Note that the synapse memory system 1 can be a neuromorphic system on silicon. The synapse memory system 1 is an example of the claimed synapse memory system.

[0022] Here, as mentioned above referring to FIG. 1, the synapse memory 100 has the cross-bar array of the axons 20 and the dendrites 30, and includes the synapse memory cells 10 arranged at all cross points of the axons 20 and the dendrites 30. This exemplary embodiment assumes that writing the synapse weight value to the NVRAM is executed by controlling the voltage value.

[0023] The write driver (synapse memory cell driver) 40 is connected to the synapse memory 100. The write driver 40 may be configured to write the synapse weight value to the respective synapse memory cells 10 in response to a learning operation input or a recognition operation input. Specifically, the write driver 40 may apply the voltage to the respective axons 20 selected by the write line selector 45. The write line selector 45 selects the axons 20 to which the write driver 40 applies the voltage. For example, the write line selector 45 switches the selected axons 20 in a predetermined order at a certain timing. In this example, the weight encoder 80 controls the write driver 40 and the write line selector 45 to apply the voltage to the synapse memory cell(s) 10 provided on the axon(s) 20 selected by the write line selector 45 according to the switching timing of the write line selector 45. This enables the single pair of the write driver 40 and the write line selector 45 to write the synapse weight value to all of the synapse memory cells 10 provided on the respective axons 20.

[0024] The dendrites 30 are connected to a dendrite driver (not shown) that selects, from the dendrites 30, a subject dendrite(s) with which the synapse weight value is written in write operation or read in read operation of the synapse weight value. The dendrite driver may selectively apply the voltage to the dendrites 30. This enables the dendrite driver to determine the subject dendrite(s). Specifically, in the write operation of the synapse weight value, the dendrite driver sets one of the dendrites 30 to a ground (GND) state and sets the others of the dendrites 30 to a high impedance (Hi-Z) state. The write line selector 45 selects one of the axons 20 which is connected to the write driver 40 to apply the voltage thereto. This results in the synapse weight value being written in the synapse memory cell 10 on the cross point of the dendrite 30 set to the GND state and the axon 20 applied with the voltage by the write driver 40. The write driver 40 receives a pulse signal for controlling an output timing. The write driver 40 is applied with a driving voltage (VDD). Note that the write driver 40 is configured to control the voltage applied

to the synapse memory cells 10.

[0025] The read drivers 50 are connected to the synapse memory 100. The read drivers 50 may be configured to read the synapse weight value from the synapse memory cells 10 in response to a recognition operation input. Specifically, the dendrite driver (not shown) sets a dendrite(s) 30 with which the synapse weight value is read to the GND state and sets the others of the dendrites 30 to the Hi-Z state. The read driver selector 150 selects or enables the read drivers 50 which correspond to the axon(s) 20 to apply the voltage thereto. This results in the synapse weight value being read from the synapse memory cell 10 on the cross point of the dendrite 30 set to the GND state and the axon 20 applied with the voltage by the read drivers 50. In the read operation of the synapse weight value, a total synapse weight value is read for each dendrite 30 set to the GND state. The total synapse weight value is a sum of the synapse weight values of the respective synapse memory cells 10 which are on the axon 20 being applied with the voltage.

[0026] The weight evaluator 70 may evaluate the current synapse weight value with a neuron output signal from the dendrites 30 and determine the next synapse weight value. Specifically, the weight evaluator 70 compares the neuron output signal obtained from the dendrites 30 with a desired output signal given as a learning operation input, and calculates the next synapse weight value with which the current synapse weight value stored in the synapse memory cell 10 is to be updated.

[0027] The weight encoder 80 may encode the next synapse weight value calculated by the weight evaluator 70. The weight encoder 80 controls the driving voltage of the write driver 40 based on the next synapse weight value calculated by the weight evaluator 70. This enables the write driver 40 to control the voltage applied to the synapse memory cells 10.

[0028] The write pulse generator 90 may conduct a modulation process, such as pulse width modulation and pulse number (or frequency) modulation. In other words, the write pulse generator 90 may generate and output a pulse signal based on the encoded value generated by the weight encoder 80. This pulse signal and the driving voltage controlled by the weight encoder 80 lead the write driver 40 to update the current synapse weight value stored in the synapse memory cell 10.

[0029] To update the current synapse weight value, several methods can be applicable such as (i) a method adding a gap between the current synapse weight value and the next synapse weight value to the current synapse weight value, and (ii) a method resetting the current synapse weight to the minimum value, thus writing the next synapse weight value.

[0030] For example, the method (i) can be applicable to update the synapse weight value by decreasing it in the

following cases. That is to say, the NVRAM used as the synapse memory cell 10 can reduce a stored value by applying a voltage having opposite polarity to a voltage applied for increasing the stored value, and a decreasing amount of the stored synapse weight value can be controlled by a value of the supply voltage.

[0031] In the method (i), several levels as to the output value of the write driver 40 can be set for controlling the synapse weight value. For example, the levels are "driver common level", "set threshold level", and "reset threshold level". The driver common level is a voltage level when the write driver 40 is in an off state (inactive). The synapse memory cell 10 is configured to keep the current weight value if the output of the write driver 40 is between the set threshold level and the reset threshold level or in high impedance (Hi-Z) state and the voltage value applied to the dendrites 30 is at the driver common level. The set threshold level, i.e. a positive threshold level, is a voltage level at or over which the write operation to increase the synapse weight value is executable. The reset threshold level, i.e. a negative threshold level, is a voltage level at or under which the write operation to decrease the synapse weight value is executable. The output value of the write driver 40 is an example of the claimed output level.

[0032] Referring to FIG. 3, the above method will be described in more detail. FIG. 3 is a time chart showing timings of signal outputs from the write driver 40, timings of signal outputs from the read drivers 50, and timings of state transitions of the synapse memory cell 10.

[0033] (1) First, from time t1 to time t2 (period t1-t2), the write driver 40 applies the voltage lower than the reset threshold level to the one of the synapse memory cells 10 (target cell) so that the write driver 40 sets the target cell to a reset state (a certain decremented value). That is to say, the write operation to decrease the synapse weight value is performed. The value to be decremented in the target cell depends on an amount by which the voltage applied by the write driver 40 falls below the reset threshold level. In the time chart, this operation is performed at a state "Reset1".

[0034] (2) Next, from time t2 to time t3 (period t2-t3), the write driver 40 applies the voltage higher than the set threshold level to the target cell so that the write driver 40 sets the target cell to a set state (a certain incremented value). That is to say, the write operation to increase the synapse weight value is performed. The value to be incremented in target cell depends on an amount by which the voltage applied by the write driver 40 exceeds the set threshold level. In the time chart, this operation is performed at a state "Set1".

[0035] (3) Next, from time t4 to time t5 (period t4-t5), the read drivers 50 apply the voltage between the driver common level and the set threshold level to the target cell so that the read drivers 50 set the target cell to a read state. That is to say, the read operation of the synapse weight value is performed. If the target cell read in this state is the same one that was written in the "Set1" state, the value to be read at this point corresponds to the synapse weight value updated at the above step (2). In the time chart, this operation is performed at a state

"Read1".

[0036] (4) Next, from time t6 to time t7 (period t6-t7), the write driver 40 applies the voltage lower than the reset threshold level to the target cell so that the write driver 40 sets the target cell to another reset state to another decremented value. That is to say, the write operation to decrease the synapse weight value is performed. In the time chart, this operation is performed at a state "Reset2".

[0037] (5) Next, from time t8 to time t9 (period t8-t9), the read drivers 50 apply the voltage between the driver common level and the set threshold level to the target cell so that the read drivers 50 set the target cell to another read state. That is to say, the read operation of the synapse weight value is performed. If the target cell read in this state is the same one that was written in the "Reset2" state, the value to be read at this point corresponds to the synapse weight value updated at the above step (4). In the time chart, this operation is performed at a state "Read2".

[0038] (6) In a period other than the periods t1-t2, t2-t3, and t6-t7, the target cell is set to the high impedance (Hi-Z) state against the write driver 40. That is to say, the write operation is not performed in the period. Further, in a period other than the periods t4-t5 and t8-t9, the target cell is set to the high impedance state against the read drivers 50. That is to say, the read operation is not performed in the period.

[0039] When the write driver 40 is in the off state, the applied voltage to the synapse memory cell 10 is between the set threshold level and the reset threshold level or high impedance state and the synapse memory cell 10 keeps the current weight value. This prevents the write operation of the synapse weight value.

[0040] In response to an output signal from the write driver 40, the applied voltage shifts from the driver common level to or above the set threshold level or to or below the reset threshold level.

[0041] When the voltage applied by the write driver 40 exceeds the set threshold level, the synapse weight value is updated to be increased. An amount of the increment depends on an amount by which the voltage applied by the write driver 40 exceeds the set threshold level. That is to say, controlling the exceeding amount by adjusting the voltage applied by the write driver 40, namely by selecting a write level, enables to control the increment amount of the synapse weight value.

[0042] When the voltage applied by the write driver 40 falls below the reset threshold level, the synapse weight value is updated to be decreased. An amount of the decrement depends on an amount by which the voltage applied by the write driver 40 falls below the reset threshold level. That is to say, controlling the falling amount by adjusting the voltage applied by the write driver 40, namely by selecting a write level, enables to control the

decrement amount of the synapse weight value.

[0043] Note that if the applied voltage is in a range between the set threshold level and the reset threshold level, the synapse weight value is maintained. In other words, the write operation on the synapse memory cell 10 is not conducted. The read drivers 50 read the synapse weight value from the synapse memory cells 10 by applying the voltage in this range to the synapse memory cells 10.

[0044] Below a description will be given of the synapse memory cell 10.

[0045] A non-volatile random access memory (NVRAM), which is an analog memory, may be used as the synapse memory cell 10. The analog memory has characteristics such that its resistance and capacitance can be controlled continuously. Examples of the NVRAM include a flash memory, a ferroelectric random access memory (FeRAM), a magnetoresistive random access memory (MRAM), a phase change random access memory (PRAM), and a resistive random access memory (ReRAM).

[0046] Writing data to the NVRAM is executed by applying a voltage or a current to the NVRAM. The NVRAM can store the data using a state change of a corresponding synapse memory cell 10, in response to the applied voltage or current. Note that the state change used for recording data depends on structure of the NVRAM. For example, the state change may include a change in an electrical state, a change in a magnetic state, and a change in phases.

[0047] An extent of the state change may vary depending on the magnitude of the applied voltage or current, i.e. a voltage value or a current value. This enables the NVRAM to store a continuous value (analog value) or a multi-value (discrete value). Note that the NVRAM can store and hold a value being one of at least three different (not digital) values. In other words, the NVRAM can be used as an analog memory. In the example shown in the figure, the synapse memory cell 10 is constituted by the NVRAM, so that controlling the voltage value or the current value to be applied to the NVRAM enables to control the synapse weight value stored in the synapse memory cell 10. Specifically, if the voltage value is used to control write operation, a larger voltage value allows the NVRAM to store a larger synapse weight value. This tuning of the voltage (current) can be done after manufacturing of a synapse memory system 1 (described later). Note that whether to use the voltage value or the current value to control the write operation on the NVRAM can be determined based on a characteristic of the NVRAM, such as a type of change in the state used for data recording, and variation width, speed, or precision of the state change that varies depending on the voltage value or the current value.

[0048] By the way, in order to achieve good performance as a neural network system, the synapse memory cell 100 is desired to have more than enough resolution steps. However, lots of memory devices have considerably low

resolution or small dynamic range and it is difficult to improve them because of device characteristic and noise margin.

[0049] In view of this, the present embodiment combines multiple memory devices with low resolution or small dynamic range (or both) as a single synapse memory (synapse memory cell 10). More preferably, a different magnitude may be assigned for each device so that total dynamic range can be increased.

[0050] FIG. 4 depicts the synapse memory 100 and the synapse memory cells 10 each composed of multiple memory devices.

[0051] In the figure, m memory devices 11 are combined to constitute one synapse memory cell 10. The memory devices 11 included in one synapse memory cell 10 are assumed to have the same device characteristic. Also, each memory device 11 is assumed to have n resolution steps.

[0052] Since one synapse memory cell 10 is composed of m memory devices 11, m lines of dendrites corresponding to the respective memory devices 11 are connected in one synapse memory cell 10. These m lines constitute one set of dendrites 30.

[0053] Each of the m lines constituting one set of dendrites 30 is assigned with a number in the range of 0 to $(m-1)$. When the memory devices 11 constituting the synapse memory cell 10 are to be distinguished from each other, each of the memory devices 11 is hereinafter noted with the number of the line connected to that memory device 11, like the memory device 11(0), the memory device 11(1) ... the memory device 11($m-2$), and the memory device 11($m-1$).

[0054] The resolution of the synapse memory cell 10 as a whole shown in FIG. 4 is determined by a combination of dynamic ranges of the multiple memory devices 11. For example, the resolution of the synapse memory cell 10 as a whole is expressed by an exponential scale using resolution (n) of the memory devices 11 with n as the base. Accordingly, each memory device 11 is assigned with a different magnitude. More specifically, assuming that the resolution of the synapse memory cell 10 is a value of m digits, the memory device 11(0) represents the first digit, the memory device 11(1) represents the second digit ... the memory device 11($m-2$) represents the $(m-1)$ -th digit, and the memory device 11($m-1$) represents the m -th digit.

[0055] To obtain the resolution of the synapse memory cell 10 as a whole, a current from each of the lines constituting one set of dendrites 30 is converted by converters 12($k-1$) ($k = 1, 2, \dots, m$) according to the digit expressed by the corresponding memory device 11. More specifically, a current from the line connected to the memory device 11(0) is multiplied by one (or n^0), a current from the line connected to the memory device 11(1) is

multiplied by n (or n^1), ... a current from the line connected to the memory device 11(m-2) is multiplied by n^{m-2} , and a current from the line connected to the memory device 11(m-1) is multiplied by n^{m-1} . Thus, the output from each line can represent the digit of the corresponding memory device 11(k - 1) ($k = 1, 2, \dots, m$) that forms part of the value of the resolution of the synapse memory cell 10 as a whole.

[0056] Then, an adder 13 adds up the currents from the lines. As described above, the output current from each of the lines has been converted according to the digits of the resolution of the synapse memory cell 10, and thus adding up these currents allows to obtain a current corresponding to the resolution of the synapse memory cell 10 as a whole. Then, a current-to-voltage converter (IVC) 14 converts the output value into a voltage, which is then undergoes analog-to-digital conversion by an analog-to-digital converter (ADC) 15. The thus-obtained digital value is outputted as an output signal from the merged dendrites 35.

[0057] FIG. 7 depicts an example of a device 1010 which includes a neuro-synaptic core system 1000.

[0058] As shown in FIG. 7, the synapse memory system 1 may be provided on the device 1010, such as a computer. In the shown example, multiple neuro-synaptic core systems 1000 are provided on a chip 1001, such as a semiconductor chip. The chip 1001 is then mounted on the device 1010.

[0059] Each of the neuro-synaptic core systems 1000 includes the synapse memory system 1 and a neuron body 700. The neuron body 700 performs neuron operations. Specifically, the neuron body 700 obtains the product-sum value of the synapse weight value and input value based on the total current sensed by the current sensor. The total current sensed by the current sensor may represent the product-sum value of the synapse weight values stored in the synapse memory cells 10 on the cross points of the selected axon(s) 20 and one set of dendrites 30 and the input values. Then, the neuron body 700 outputs neuron output corresponding to the total current value.

[0060] Here, the neuro-synaptic core systems 1000 are connected to each other via the pre-driven axons 25. The synapse memory system 1 is connected to a pre-neuron via the pre-driven axon 25. The neuron body 700 is a part of the post-neuron and its output is pre-driven axon 25 which is connected to synapse memory systems 1 in the same or other neuro-synaptic core systems 1000. The synapse memory system 1 receives an output signal from the pre-neuron to accumulate the weight values of the synapses whose inputs are activated via their axons 20. If the accumulated value of the multiplied value of synapse weight value and input value from the axon 20 (or pre-driven axon 25) reaches a predetermined threshold, the neuron body 700 outputs the output signal as the neuron output of the post-neuron at the pre-driven axon 25.

Second exemplary embodiment

[0061] FIG. 5A shows an exemplary structure of an analog memory composed of the MRAM. FIG. 5B shows the state where the analog memory of FIG. 5A stores a value "0". FIG. 5C shows the state where the analog memory of FIG. 5A stores a value "1".

[0062] The analog memory 16 includes a free layer 161 and pinning layers 162. The free layer 161 is planar and includes fixed regions 161a at ends thereof and a data region 161b between the fixed regions 161a. The pinning layers 162 are deposited on the respective fixed regions 161a at the ends of the free layer 161.

[0063] As excited by the pinning layers 162, the two fixed regions 161a have magnetic fields in opposite directions.

[0064] Due to influence of the fixed regions 161a, the magnet field of the data region 161b may be identical to the magnetic field of either fixed region 161a or may partially include the magnetic fields of both fixed regions 161a. The boundary at which the magnetic field of the data region 161b changes is referred to as a domain wall 161c.

[0065] As the fixed regions 161a are excited by the pinning layers 162, the position of the domain wall 161c moves in the data region 161b between the fixed regions 161a. An analog value is stored according to the position of the domain wall 161c.

[0066] Accordingly, to increase the dynamic range of the analog memory 16 configured as above, it is required to increase, as much as possible, the length on which the domain wall 161c can move. In other words, the length of the data region 161b (the length a in FIG. 5A) needs to be as long as possible. Still in other words, a distance between the two fixed regions 161a needs to be as long as possible.

[0067] Meanwhile, when the domain wall 161c moves between the two fixed regions 161a as shown in FIGS. 5B and 5C in writing a memory, current (or electron movement per unit time) more than a certain threshold amount is brought by the voltage of more than the set threshold level or reset threshold level in FIG. 3. The direction of this electron movement (opposite of the current direction) is shown in FIGS. 5B and 5C. The amount of domain wall 161c movement depends on the current (the amount of electrons moved in a unit time). So the number of recording value steps of the memory within its full range is proportional to the length of the data region 161b where the domain wall 161c can move from that of in FIG. 5B to that of in FIG. 5C. Thus the length of a of the device 16 in FIG. 5A should be maximized. On the other hand, recording current is proportional to the cross-section area of the memory device which is vertical to the flowing current. Thus the width b of the device 16 in FIG. 5A should be minimized to reduce the recording current to reduce the power consumption. For these reasons, the shape of the

analog memory 16 with a large dynamic range and enough resolution of stored value in low power consumption will exhibit large aspect ratio (ratio of a:b of FIG. 5A).

[0068] The structure of the analog memory 16 explained with reference to FIGS. 5A to 5C is by way of example only. However, lots of analog memories actually exhibit large aspect ratio so that they can have a large dynamic range with a lot of small amount of steps.

[0069] Assume that each of the memory devices 11 of the synapse memory cell 10 explained with reference to FIG. 4 is composed of the analog memory 16 explained with reference to FIGS. 5A to 5C.

[0070] FIG. 6 shows an exemplary configuration where the analog memory 16 shown in FIGS. 5A to 5C is used for each memory device 11 in the synapse memory cell 10.

[0071] In the example shown in FIG. 6, multiple memory devices 11 (analog memories 16) in each synapse memory cell 10 are arranged such that their long sides (the sides along the direction of the length a in FIG. 5A) are parallel to each other. This can reduce aspect ratio of the synapse memory cell 10 as a whole (namely the ratio of the longitudinal direction of the memory devices 11 to the direction in which the memory devices 11 are arranged), relative to the aspect ratio of individual memory devices 11. This increases flexibility in the shape of the synapse memory cell 10 and the synapse memory 100 in the configuration where multiple memory devices 11 are combined as one synapse memory cell 10.

Third exemplary embodiment

[0072] As explained with reference to FIG. 2, the synapse memory system 1 may include or external blocks may include peripheral devices connected to the synapse memory 100. These peripheral devices include those provided on the axons 20 side of the synapse memory 100 and those provided on the merged dendrites 35 side of the synapse memory 100.

[0073] Under normal operations of the synapse memory system 1 involving inference (or recognition) and training, a signal output from an anterior neuron body 700 is input via the pre-driven axons 25 to the synapse memory 100, which in turn outputs the signal via the merged dendrites 35 to a posterior neuron body 700.

[0074] On the other hand, under operation of the synapse memory system 1 involving back propagation, a signal output from the posterior neuron body 700 is input via the merged dendrites 35 to the synapse memory 100, which in turn outputs the signal via the pre-driven axons 25 to the anterior neuron body 700.

[0075] For this reason, a group of peripheral devices of almost same size is provided on both of the axons 20 (or pre-driven axons 25) side and the dendrites 30 (or the merged dendrites 35) side of the synapse memory 100. Thus, as the aspect ratio of the synapse memory 100 (namely the ratio of length along its axons 20 to length along its dendrites 30) gets closer to 1:1, it becomes easier to place the peripheral devices relative to the synapse memory 100 and thus becomes easier to design the synapse memory system 1.

[0076] As explained with reference to FIG. 4, when the synapse memory cell 10 is composed of multiple memory devices 11, the synapse memory cell 10 may have large aspect ratio depending on the number of memory devices 11 included in the synapse memory cell 10. In that case, in order to reduce aspect ratio of the synapse memory 100 as a whole, the synapse memory cells 10 need to be spaced with blank gap of a certain distance from each other in vertical direction according to the aspect ratio of each synapse memory cell 10, which results in useless space being created.

[0077] On the other hand, arranging the analog memories 16 as the memory devices 11 in each synapse memory cell 10 as shown in FIG. 6 can reduce the aspect ratio of the synapse memory 100 as a whole while eliminating useless space.

[0078] More specifically, multiple memory devices 11 (analog memories 16) in each synapse memory cell 10 are arranged such that their long sides are parallel to each other. This can make the aspect ratio of the synapse memory cell 10 close to 1:1. Thus, the synapse memory 100 composed of such synapse memory cells 10 can have aspect ratio close to 1:1 with no useless space between the synapse memory cells 10.

CLAIMS

1. A synapse memory system comprising:
a plurality of synapse memory cells provided at cross points of a plurality of axon lines and a plurality of dendrite lines, each synapse memory cell comprising a plurality of analog memory devices each including a plurality of resolution steps, each synapse memory cell being configured to store a weight value according to an output level of a write signal, the weight value to be stored in each synapse memory cell being one value of at least three different values, the plurality of analog memory devices each including the plurality of resolution steps being combined to constitute each synapse memory cell expressed by an exponential scale using the plurality of resolution steps of all of the synapse memory cells; a write portion configured to write the weight value to each synapse memory cell, the write portion comprising a write driver and an output controller, the write driver being configured to output the write signal to a subject synapse memory cell, the subject synapse memory cell being one of the synapse memory cells, the subject synapse memory cell being selected to store the weight value, the output controller being configured to control the output level of the write signal of the write driver;
and read drivers configured to read the weight value stored in the synapse memory cells.
2. The synapse memory system according to claim 1, wherein dynamic ranges of the plurality of analog memory devices are combined to obtain a dynamic range of each synapse memory cell as a whole.
3. The synapse memory system according to claim 2, wherein each of the plurality of analog memory devices is assigned with a different magnitude.
4. The synapse memory system according to any one of the preceding claims, wherein the plurality of synapse memory cells are arranged such that long sides of the plurality of analog memory devices included in each synapse memory cell are parallel to each other.
5. The synapse memory system according to claim 4, wherein aspect ratio of each synapse memory cell is smaller than aspect ratio of each analog memory device.
6. The synapse memory system according to claim 4, wherein aspect ratio of each synapse memory cell is almost 1:1.
7. The synapse memory system according to any one of the preceding claims, wherein each analog memory device comprises a non-volatile random access memory (NVRAM).

8. The synapse memory system according to any one of the preceding claims, wherein each analog memory device comprises a magnetoresistive random access memory (MRAM).
9. A method for operating a synapse memory system, the synapse memory system comprising:
 - a plurality of synapse memory cells provided at cross points of a plurality of axon lines and a plurality of dendrite lines, each synapse memory cell comprising a plurality of analog memory devices each including a plurality of resolution steps, wherein the plurality of analog memory devices each including the plurality of resolution steps is combined to constitute each synapse memory cell expressed by an exponential scale using the plurality of resolution steps of all of the synapse memory cells; a write portion comprising a write driver and an output controller;
 - and a set of read drivers;
 - the method comprising:
 - storing, by a synapse memory cell of the system, a weight value according to an output level of a write signal, the weight value stored being one value of at least three different values;
 - writing, by the write portion, the weight value in the synapse memory cell, the writing comprising:
 - outputting, by the write driver, the write signal to the synapse memory cell selected to store the weight value, and controlling, by the output controller, the output level of the write signal of the write driver;
 - and reading, by the read drivers, the weight value stored in the synapse memory cells.
10. The method of claim 9, further comprising combining dynamic ranges of the plurality of analog memory devices to obtain a dynamic range of each synapse memory cell as a whole.
11. The method of claim 10, further comprising assigning each of the plurality of analog memory devices with a different magnitude.
12. The method of any one of claims 9 to 11, further comprising arranging the plurality of synapse memory cells such that long sides of the plurality of analog memory devices included in each synapse memory cell are parallel to each other.
13. The method of claim 12, wherein an aspect ratio of each synapse memory cell is smaller than aspect ratio of each analog memory device.
14. The method of claim 12, wherein the aspect ratio of each synapse memory cell is almost 1:1.
15. The method of any one of claims 9 to 14, wherein each analog memory device comprises a non-volatile random access memory (NVRAM).

16. The method of any one of claims 9 to 14, wherein each analog memory device comprises a magnetoresistive random access memory (MRAM).