



(51) International Patent Classification:

H01L 21/28 (2006.01) *H01L 29/788* (2006.01)
H01L 21/336 (2006.01) *G11C 16/04* (2006.01)
H01L 29/423 (2006.01)

(21) International Application Number:

PCT/EP2010/064060

(22) International Filing Date:

23 September 2010 (23.09.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

0916776.8 23 September 2009 (23.09.2009) GB
0921013.9 30 November 2009 (30.11.2009) GB

(71) Applicant (for all designated States except US): **UNIVERSITÉ CATHOLIQUE DE LOUVAIN** [BE/BE];
Place de l'Université 1, B-1348 Louvain-La-Neuve (BE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TANG, Xiaohui** [BE/BE]; Rue de Neafmoustier 8, B-1348 Louvain-La-Neuve (BE). **RASKIN, Jean-Pierre** [BE/BE]; Rue de la Brasserie 17, B-5310 Saint-Germain (BE).

(74) Agents: **BIRD, William E.** et al.; Klein Dalenstraat 42A, B-3020 Winksele (BE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

(54) Title: A MEMORY DEVICE AND A METHOD OF MANUFACTURING THE MEMORY DEVICE

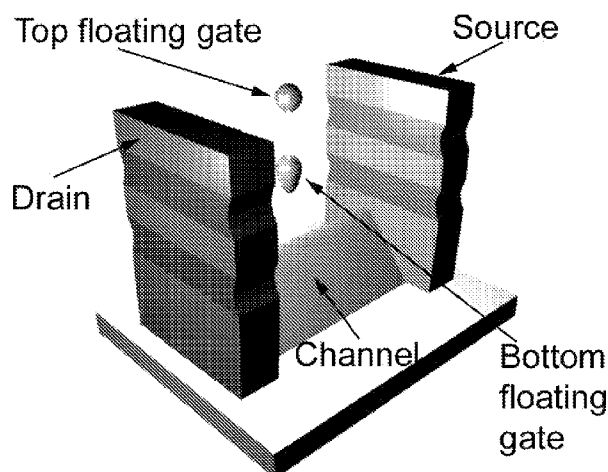


Fig. 10

(57) Abstract: A semiconductor memory device is provided with a source and a drain electrode between which a channel extends, one, two or more nano floating gates overlying said channel, and a control gate. It is manufactured in that an insulating substrate is provided with an alternating layer stack of a first and a second semiconductor material, for instance Si/SiGe/Si/SiGe/Si/BOX or Si/SiGe/Si/BOX stacks. The alternating layer stack is patterned down to the insulating substrate according to a predefined pattern including the source electrode, the drain electrode and a channel region extending between said electrodes. The patterned alternating layer stack is then oxidized, such that the second semiconductor material in the channel region is completely oxidized under certain conditions to separate the channel from the floating gates made of first semiconductor material or the second semiconductor material agglomerates under other conditions to form one of the floating gates. The semiconductor memory device preferably comprises a first and a second floating gate on top of each other, which are self-aligned with said channel. The present technology makes it possible to realize multil-bit memory cells and qubit quantum computers.

A MEMORY DEVICE AND A METHOD OF MANUFACTURING THE MEMORY DEVICE

FIELD OF THE INVENTION

5 The invention relates to a method of manufacturing a semiconductor memory device provided with a source and a drain electrode between which a channel extends, one or more self-aligned floating gates overlying said channel.

 The invention further relates to a semiconductor memory device thus obtained, applications thereof and methods of operating said device.

BACKGROUND OF THE INVENTION

 Semiconductor memory devices are widely used for storage of information. Particularly, flash memory arrays may be employed for this purpose. Two transistor configurations in a memory cell with an access transistor and a memory device, or memory transistor, allow fast random access and low power program and erase by tunneling. The access transistor, typically a MOSFET transistor, which may act as a switch, may help in the selection and/or inhibition of the memory transistor with an array. The use of an access transistor may allow the memory to be fully programmed and erased by Fowler-Nordheim tunneling. A memory transistor typically comprises a source and a drain electrode between which a channel extends. A floating gate overlies the channel, and a control gate again overlies the floating gate.

 The continuous drive to higher densities and smaller device have led to proposals for nanowire memory devices. The term 'nanowire' may denote a wire-like structure of dimensions in the order of magnitude of several, such as 5 nm, to several hundreds of nanometers, such as 500 nm (and may also cover larger or smaller dimensions). For instance, the width of the nanowire may be 10 nm or less. The length of the nanowire may well be larger than its width, the terms width and length being used herein as two mutually perpendicular directions. Typical nanowires exhibit aspect ratios (the ratio between width and length) of more than one, particularly more than two, more particularly more than 5 or up to 1000 or more. Many different types of nanowires are known including semiconductor nanowires (for instance made of silicon, germanium,

InP, GaN etc), metallic nanowires and carbon nanotubes. Particularly though not exclusively the semiconductor nanowires are suitable for use as a channel in a memory transistor.

Single electron memories are an ultimate device for low power consumption, fast access speed and high integration density. In single electron memories, a nanoscale floating gate is charged/discharged by a few (ideally one) electrons based on energy quantization and Coulomb blockade effect. Two main approaches, multiple-dot and single-dot devices, have been proposed to fabricate these memories since Nakazato reported the first single electron memory in *Electron. Lett.*, vol. 29, no. 4, pp. 384–385, 1993.

US Patents Nos. 5,714,766 and 5,801,401 relate to the fabrication of the multiple-dot memory devices. An enormous number of nanocrystals is randomly deposited on an existing MOSFET channel in this method. The nanocrystals function as the floating gates of the memory device to store individual electrons. The enormous number of nanocrystals implies that this device is intrinsically not single electron memory device, even if each nanocrystal stores only one electron. This device presents the advantage of being more immune to spurious discharging as electrons are distributed among isolated nanocrystals. However, the randomness of nanocrystal deposition induces fluctuations in the threshold voltage shift due to the difficulty in maintaining the size of nanocrystals, which are used as the floating gates, to be uniform.

Single-dot memory devices are typically composed of a narrow MOSFET channel with a single nanoscale floating gate embedded in the gate dielectric. Guo in *Science*, vol. 275, no. 5300, pp. 649–651, Jan 1997 and Fujitsu in U.S. Pat. No. 5,886,380 introduced a Si single electron memory operating at room temperature. The operation of this device is based on the effect of charging with one single electron. This is achieved by making the width of a channel equal to or smaller than the Debye screen length. A very high resolution lithography tool is required to define this floating gate and its precise positioning above the channel. This difficulty of integration and complicated fabrication process is a major disadvantage.

A nanowire memory transistor is known from WO2008/129478. This known nanowire transistor teaches the provision of the nanowire into a trench within a substrate.

This leads to a three-dimensional geometry of the memory cell, that would have the advantage that the channel current may flow vertically and that there is no need to scale down the length of the transistor. The nanowires may be grown within trenches surrounded by silicon dioxide which may provide isolation between the bit lines in an array configuration. In order to provide a contact to the top side, and to fill up the trench, the nanowire may be covered by polysilicon or the like. The bottom side of the nanowire would be coupled to the substrate buried in the trench. It is not clear from its disclosure whether this is a single-dot memory device or a multiple-dot memory device. It is however a disadvantage that the manufacture of this nanowire memory transistor is not so simple. In order to make contact to the substrate buried in the trench, the bottom of the trench should be free of an oxide. However, an oxide is typically formed along the trench. Furthermore, if one would grow silicon with the trench, it is not clear how to ensure that the width of the nanowire remains small and how to ensure that growth takes place without grain boundaries and in the desired orientation.

Another nanowire memory transistor is known from an article in Solid State Electronics 44 (2000), 2259-2264 and provides a single-dot memory. This known memory transistor is made on an SOI substrate. Arsenic is implanted in the top silicon film of the SOI substrate. The top silicon film is thereafter protected against oxidation with a protection layer, particularly of silicon nitride. This substrate structure is then patterned according to a predefined pattern. The pattern includes a source and a drain electrode and a channel region in between thereof. Within the pattern, the channel region has a smaller width than the source and the drain electrodes. Within the channel region a floating gate region is present between a first and second path region. The path regions have an even smaller width than the floating gate region. As a result, in the subsequent oxidation step, the top silicon is nearly completely oxidized into silicon oxide in the path regions – a channel remains. In the floating gate region however, also a floating gate remains which is separated from the channel with an intermediate oxide. The larger width of the source and drain regions results therein that these will not be oxidized, except for a minor shell at the sidewall surface. Overall, the nanoscale floating gate is self-aligned with the nanowire channel and the floating gate and the tunnel oxide is formed in one process step.

It is a disadvantage of this known nanowire memory device that scattering effects have been observed in operation of the transistor. This is deemed to be due to the high concentration of arsenic needed to achieve the split of the channel and the floating gate.

5 SUMMARY OF THE INVENTION

It is an object of the present invention to provide an alternative method for the manufacture of a memory device as well to provide a memory device.

According to a first aspect of the invention, a method of manufacturing a semiconductor memory device provided with a source and a drain electrode between
10 which a channel extends, one, two or more floating gates overlying said channel, and a control gate, which method comprises the steps of

providing a substrate with an insulating layer and an alternating layer stack of a first and a second semiconductor material, wherein the second semiconductor material oxidizes selectively with respect to the first semiconductor material, or reverse.

15 patterning the alternating layer stack down to the insulating layer of the substrate according to a predefined pattern including the source electrode, the drain electrode and the channel region extending between said electrodes, and selectively oxidizing said patterned alternating layer stack, such that the second semiconductor material in the channel region is completely oxidized to separate the
20 channel from the floating gate or such that the second semiconductor material agglomerates overlying the channel region as the floating gate.

According to a second aspect of the invention, a semiconductor memory device provided with:

a substrate with an insulating layer, preferably a buried substrate such as a SOI
25 with a buried oxide, and an alternating layer stack of a first and a second semiconductor material, wherein the second semiconductor material oxidizes selectively with respect to the first semiconductor material,

a source and a drain electrode defined in the alternating layer stack, between which electrodes a channel extends in a first layer of the first material of the alternating
30 layer stack ,

a first floating gate that overlies said channel and is defined in a second layer of

the first material of the alternating layer stack, an intermediate layer of the second semiconductor material between said first and second layer of the first semiconductor material being oxidized between said channel and said first floating gate, or wherein the first floating gate is defined in a second layer of the first material of the alternating layer stack, and

a control gate.

According to a third aspect of the invention, an array of memory devices is provided.

The array of memory devices further comprises word lines coupled to the control gates and bit lines coupled to the drain electrodes.

According to a fourth aspect of the invention, storage medium is provided.

The storage medium comprises a memory device according to the invention and/or an array according to the invention, wherein the storage medium preferably has a capacity of 1 Tb or more.

According to a fifth aspect of the invention, an integrated circuit is provided. The Integrated Circuit comprises a storage medium according to the invention and/or a memory device according to the invention and/or an array according to the invention.

According to a sixth aspect of the invention, a memory device is provided. The memory device, such as a non-volatile memory device, according to the invention is comprised in a MOSFET, EEPROM, EPROM, DRAM, SRAM, or speed flash memory.

According to a seventh aspect of the invention, a method of operating memory devices is provided.

According to an eighth aspect of the invention, a memory device obtainable by the present method is provided.

The inventors have found in investigations leading to the invention that the first and second semiconductor material can be chosen such that they can be etched and oxidized selectively, i.e. the second material is etched and oxidized faster than the first material, or vice versa. As a result, a split between floating gate and channel can be reliably made by provision of a layer of the second semiconductor material in between of layers of for example the first semiconductor material. Moreover, the thickness of a tunnel oxide present between the channel and the floating gate can be specified by

choosing the thickness of the said intermediate layer of the second semiconductor material.

It is an advantage of the method of the invention that it allows downscaling of memory devices due to a solution of reliability problems. Moreover, when downscaling, the advantage of self-alignment of the nanoscale floating gate with the nanowire channel - per se known from the prior art in Solid State Electronics as above referred to - is maintained. The performance of the present inversion memory device is improved by canceling arsenic implantation. As the size of the floating gate and the thickness of the tunnel oxide may be down to 10 nm, while semiconductor industry at present is engaged with 32 nm, it is clear that a major downscaling can be obtained.

In a specific embodiment of the device invention, a second floating gate is present, which in an example is defined in a third layer of the first material of the alternating layer stack, an intermediate layer of the second semiconductor material between said second and third layer of the first semiconductor material being oxidized between said first and second floating gate. The resulting memory device with a double floating gate on top of each other. This provides multiple level storage function or long retention time. Specifically, the storage function can be quantized, i.e. have a quantized threshold voltage shift. In one preferred embodiment, the lateral extension of the floating gate is thereto less than 10 nm to guarantee the presence of Coulomb blockade effect at room temperature.

In a further embodiment, the control gate may be located adjacent to the floating gate. Suitably, the control gate constitutes a structure that is present adjacent to the floating gate and the channel. Such a structure will also be referred to as a side control gate. More preferably, the side control gate comprises a first and a second portion present on opposing sides of the channel and the floating gates. The first and second portion are suitably coupled with an interconnect overlying the floating gates. Preferably, the side control gate is embodied in a layer stack that comprises the same layers as the source and the drain. In this manner, it may be manufactured simultaneously with the source and drain, as part of the predefined pattern. Most suitably, the layer stack of the side control gate is silicidized. This again may occur simultaneously with the silicidation of the source and the drain electrode.

Preferably, the first semiconductor material is silicon and the second semiconductor material is SiGe (silicon-germanium) alloy or Germanium, or vice versa. Preferably the germanium content of the SiGe is less than 50%. These materials are well known and accepted in industrial semiconductor processing. Moreover, they may be provided with dopants so as to further optimize the operation of the transistor. For instance, an implantation may be carried out into the source and drain regions to reduce the contact resistance after the provision of the channel and the one or more floating gates. A further advantage is that the upper layer of the alternating layer stack can be covered with a silicon nitride protection layer. This protection layer protects the top side during patterning, inhibits oxidation from the top side during oxidation and is resistant to high temperatures resulting from any subsequent implantation, diffusion and/or other high temperature steps.

According to a further aspect, a method of programming and/or erasing of a memory device is provided. Herein charge is inserted or removed from the floating gates individually by variation of an applied voltage on the control gate.

More specifically, the operation of the double floating gate memory transistor is the following: On writing, a first electron is injected into the first (bottom) floating gate by applying a positive voltage to the control gate. One electron injection is due to the presence of Coulomb blockade effect. When the writing voltage is increased, the first electron stored in the bottom floating gate is pushed into the second (top) floating gate. Simultaneously, a second electron is injected into the first floating gate. Erasing of both floating gates may be carried out simultaneously by application of a higher negative voltage to the control gate. Alternatively, one may stepwise erase the floating gates by application of a relatively low negative voltage, so as to erase the second floating gate. Thereafter, if desired, the first floating gate may be erased by application of a higher negative voltage. On the other hand, single-electron memories with double nano floating-gates show a long data retention time due to the fact that quantum effects existing in additional floating-gate suppress charge leakage.

These and other aspects of the invention will be further elucidated with reference to the Figures, in which:

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagrammatical cross-sectional view illustrating a first embodiment of the memory device of the present invention;

FIG. 2 is a schematic layout for the first embodiment of the memory device;

5 FIG. 2a-2d show cross-sections of a fully self-aligned double floating gate memory device according to the invention;

FIG. 3 is a diagrammatical cross-sectional view showing the result of a first step of the method of the invention, in a first embodiment;

10 FIG. 4A through 4C are cross-sectional views of the floating gate region, the path regions and the source and drain regions obtained after a second step of the method of the invention, in a first embodiment;

FIG. 5 is a cross-sectional view of the floating gate region in the memory device of the invention, after completion of a third step of the method of the invention;

15 FIG. 6 is a cross-sectional view of the path region after completion of the third step of the method of the invention;

FIG. 7 is a cross-sectional view of the source/drain regions after completion of the third step of the method of the invention;

FIG. 8a and 8b are a diagrammatic top view and a SEM image of the device of the invention;

20 FIG. 9 is a graph showing drain current as a function of control gate voltage in the device of the invention for three different states;

FIG. 10 is a schematic bird eye's view of the first embodiment of the invention;

FIG. 11 shows hysteretic behavior of a self-aligned memory device according to the invention;

25 FIG. 12 shows a drain current versus time characteristics of a self-aligned memory device according to the invention; and

FIG. 13 shows time dependence of a memory window width of a self-aligned memory device according to the invention.

30 DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention will be described with respect to particular embodiments

and with reference to certain drawings but the invention is not limited thereto but only by the claims. The drawings described are only schematic and are non-limiting. In the drawings, the size of some of the elements may be exaggerated and not drawn on scale for illustrative purposes. Where the term “comprising” is used in the present description and claims, it does not exclude other elements or steps. Where an indefinite or definite article is used when referring to a singular noun e.g. "a" or "an", "the", this includes a plural of that noun unless something else is specifically stated.

The term “comprising”, used in the claims, should not be interpreted as being restricted to the means listed thereafter; it does not exclude other elements or steps. Thus, the scope of the expression “a device comprising means A and B” should not be limited to devices consisting only of components A and B. It means that with respect to the present invention, the only relevant components of the device are A and B.

Furthermore, the terms first, second, third and the like in the description and in the claims, are used for distinguishing between similar elements and not necessarily for describing a sequential or chronological order. It is to be understood that the terms so used are interchangeable under appropriate circumstances and that the embodiments of the invention described herein are capable of operation in other sequences than described or illustrated herein.

Moreover, the terms top, bottom, over, under and the like in the description and the claims are used for descriptive purposes and not necessarily for describing relative positions. It is to be understood that the terms so used are interchangeable under appropriate circumstances and that the embodiments of the invention described herein are capable of operation in other orientations than described or illustrated herein.

The present invention will be further explained with reference to the figures. In the following, the device with a double floating gate is taken as an example. However, the method and the device of the invention are not limited thereto. The scope of the invention is given by the claims.

It is an objective of the present invention is to build an ultra-highly integrated memory by overcoming drawbacks of a previous memory device. Accordingly, the previous memory with a floating gate is fabricated based on arsenic-assisted etching and oxidation effects, which only can provide one floating gate.

According to the invention, a multiple-level storage function is provided by vertically arranging two floating gates over a channel in order to overcome a limit in ultrahigh integration. Each of the two floating gates may be filled individually. This is performed through changing the level of the control gate voltage in order to inject one
5 electron into each floating gate. At the same time, the additional FG in tunnel oxide results electrons in the MOSFET channel passing through double tunnel junctions. Quantum effects (quantum confinement and Coulomb blockade) induce splitting of an energy band of both FGs into a number of energy levels. It is believed that the band-offset between both FGs prevents an electron in the upper FG from returning to the
10 channel, thereby increasing retention time.

It is an advantage of the present invention that the non-volatile memory consumes an extremely small amount of power. Programming by means of tunneling a single electron into a floating gate does not cost much power. Moreover, the well-controlled thickness of tunnel oxides and energy quantization allows optimization of the
15 programming voltage to a minimum. A further advantage of the reduced dimensions of the memory, .e.g. to below 10 nm floating gate size, which guarantees the presence of Coulomb blockade effect at room temperature, is the resulting increase in access speed. In addition, the presence of the lower floating gate results in a long retention time for a single-level application of single electron memory device with both floating gates.

FIG. 1 is a vertical cross-sectional view schematically illustrating a configuration of a memory device with a first and a second floating gate according to the present invention. Fig. 10 shows a bird's eye perspective of this embodiment. This memory device is a multiple level single electron device, as it provides storage of more than one single electron. As shown in FIG. 1, the multiple-level single electron with two floating
20 gates according to the present invention fundamentally has a structure in which a Si nanowire channel 2 is defined on an insulating layer 1. Preferably, use is made of a substrate with a buried insulating layer, such as silicon-on-insulator (SOI) substrate with a buried oxide. The silicon film of the substrate is by patterning and oxidation converted into the nanowire channel 2. The same top silicon film also is present as part of the
25 source 3 and the drain 4 electrodes. These source and drain electrode 3, 4 are defined in the alternating layer stack. The alternating layer stack is in one example a heterostructure
30

of Si/SiGe/Si/SiGe, etc.. Due to the vertical extension, as compared to the channel that is present in the first Si layer only, the source and drain electrodes 3,4 are elevated. A tunnel oxide is present on top of the channel 2. In this example, the tunnel oxide is constituted by an insulating layer 5 of (SiGe)O₂. Thereon, the first and the second floating gates 6 and 7 are present. The second floating gate 7 overlies the first floating gate 6 and is separated therefrom through an insulating layer 8 of (SiGe)O₂ acting as a second tunnel oxide arranged. An insulating layer 9 overlies the second floating gate 7 to act as the control gate oxide. A control gate 10 is formed on the control gate oxide 9.

It has been found by the inventors that the channel 1 in an example has a substantially triangular structure, which is characterized by a tip on its top side. It is believed that the tip of the channel 1 functions as a field emission tip and facilitates the emission of charge carriers (holes or electrons) into and from the floating gates 6,7, which is considered advantageous.

The manufacturing in accordance with the method comprises three major steps. The results of each step are outlined in Fig. 3, Fig. 4A-Fig 4C and Fig. 5-7, respectively, in cross-sectional diagrammatical drawings. Equal reference numerals in different figures refer to the same or like parts.

Fig. 3 shows a substrate with the alternating layer stack. On top of a buried oxide layer 1 a silicon top layer 2 is provided. This is a structure typically part of a silicon-on-insulator substrate. The buried oxide layer 1 has a thickness in the order of 50-400 nm in this example, such as 100 nm. The further layers of the alternating layer stack include a first layer 13 of the second semiconductor material SiGe, a second layer 14 of the first semiconductor material Si, a second layer 15 of the second semiconductor material SiGe and a third layer 16 of the first semiconductor material Si. The alternating layer stack is covered with a protection layer 17, which is silicon nitride in this example. The layer thicknesses of the first and second layer of second semiconductor material are preferably in the range between 5 and 20 nm. The layer thicknesses of the third and second layer of first semiconductor material are preferably in the range of 10-20 nm. It is preferred that the first layer thickness of the first semiconductor material is higher than that of the other layers. The exact layer thicknesses are subject to a design optimization dependent on the lateral geometries and the desired tunnel oxide thicknesses. The SiGe suitably comprises

less than 50% of Ge, based on its atomic content. A preferred composition is $\text{Si}_{0.7}\text{Ge}_{0.3}$ in this example. A composition of about $\text{Si}_{0.7}\text{Ge}_{0.3}$ with from 50% to 95% Si provides a sufficient selectivity towards oxidization, wherein the composition of about $\text{Si}_{0.7}\text{Ge}_{0.3}$ (from 65% to 75% Si) provides the best results. The first silicon layer may be slightly doped, for instance with p-type dopant to create a channel of the memory device. In the present invention, the layers of the alternating layer stack may be formed by epitaxial growth or by CVD deposition.

Fig. 4A-4C show schematically the result of the patterning step of the alternating layer stack according to a predefined pattern. A suitable predefined pattern is shown in Fig. 2. The pattern defines a source and a drain region 21, 22, as well as a channel region comprising a floating gate region 23 and adjacent path regions 24, 25. As shown in Fig. 2, the width of the source and drain regions 21, 22 is larger than that of the channel region. Within the channel region, the diameter of the floating gate region 23 is larger than that of the path regions 24, 25. The pattern shown in Fig. 2 is evidently merely one example. Instead of the circle, a square or rhombus could be defined. Instead of one channel, two or more channels could be defined between the source and the drain region 21, 22. Instead of one floating gate region, two or more floating gate regions could be defined, each separated by path regions. The dimensions of the source and drain regions 21, 22 may be varied, as long as the size of these regions 21, 22 exceeds the size of the floating gate region 23. A typical diameter of the floating gate region is between 50 and 100 nm, for instance 75 nm, a typical width of the path regions is between 20 and 80 nm, for instance 50 nm, a typical width of the source and drain regions is between 70 and 400 nm, for instance 100 nm.

Fig. 4A shows schematically the result of patterning for the floating gate region 23. Fig. 4B shows schematically the result of patterning for the path regions 24, 25. Fig. 4C shows schematically the result of patterning for the source and drain regions 21, 22. The pattern is defined lithographically, for instance with e-beam lithography, optical lithography and imprint. The pattern is then transferred to the alternating layer stack by means of reactive ion etching or wet etching. Due to the material differences, a difference in etching between the first semiconductor material and the second semiconductor material occurs. Effectively, the SiGe is preferably etched, so as to result in under etching

of the silicon. This under etching already removes a major portion of the SiGe in the path region, whereas it removes merely an edge of the source and drain regions.

Fig. 5-7 shows schematically the result of oxidation of the patterned structures as shown in Fig. 4A-C. The oxidation typically leads to lateral expansion of the regions. As shown in Fig. 5, the oxidation in the floating gate region 23 is effective to fully oxidize the SiGe, as well as a shell of the Si. The SiGe is oxidized into $(\text{SiGe})\text{O}_2$. Herein, also the one or more floating gates are formed. Preferably, the floating gates have a diameter of less than 10 nm. This results in the advantageous electron by electron loading and erasure, as will be explained hereinafter, when discussing the operation of the device. As shown in Fig. 6, the oxidation in the path regions 24, 25 is effective to oxidize both the complete alternating layer stack, except for the first layer. In order that the first (bottom) layer is not completely oxidized, it has preferably a larger thickness than the other layers. However, already due to the under etching of the upper layers, those layers have a relatively large surface area in comparison to the bottom layer. Thus, they have a larger tendency of being oxidized, that is being selectively oxidized. As shown in Fig. 7, the source and the drain regions 21, 22 are provided with an oxide shell, while being otherwise unchanged.

After the processing shown here, further steps may be applied. Suitably the protection layer of Si_3N_4 is removed. A layer of polysilicon or metal material is deposited on the top surface of the structure. The layer is thereafter patterned to define a control gate. The source/drain regions and the control gate may be highly-doped or silicided to obtain a low contact resistance.

Preferably, the width of the control gate is larger than that of the floating gate region, such as from 10% to 100% larger, but the control gate does not cover the source/drain region. It is preferred that the source and drain electrodes and the control gate are silicided in a single step. Preferably, a layer of at least one material from the group of Pt, Er, Ti and Yb is deposited on the surface of the heterostructure. Then, the silicidation step is processed in a rapid thermal annealing system to change the heterostructures of the control gate and the source/drain region into silicide for a low contact resistance. The floating gate region and the path region may be protected against silicidation with a protection layer, for instance the SiO_2 formed during the oxidation step. The silicidation

on the control gate may be limited to contact areas adjacent to the floating gate region, so as to prevent contamination.

Fig. 8 shows a diagrammatical top view of a further embodiment of the device of the invention. Herein the control gate 26 is provided as a side control gate which extends adjacent to the floating gate 23 and suitably also the channel 24,25. Preferably, this side control gate 26 comprises the same materials and structure as the source and drain electrodes 21, 22, as shown in Fig. 7. It can then be patterned simultaneously with the source electrode 21, the drain electrode 22 and the floating gate 23. The length of the side control gate 26 is larger enough to avoid the formation of the tunnel oxide at the second semiconductor material. The cross-section of the side control gate 26 is the same as the Fig. 7 in which the oxide provides a shell, while being other wise unchanged. If a silicidation step is applied, which is preferable, the side control gate 26 is also silicidized and thus changed into the metallic material, with the source electrode 21 and drain electrode 22. The floating gate is unchanged due to the protection of the oxide layer from oxidation process. This further simplifies the fabrication process.

If a memory cell of a memory transistor as shown in this example plus an access transistor is desired, the access transistor may be fabricated in the same manner as the memory transistor, i.e. as a nanowire transistor without floating gates. Specifically, one could define the source electrode, drain electrode, wherein a channel extends between the said electrodes. The width of the channel is the same as that of the path in the memory device. As shown in Fig. 6, the oxidation in the channel of the access transistor is effective to oxidize the complete alternating layer stack, except for the first layer. Alternatively, one could define an access transistor in a more conventional design. One option thereto is to provide a patterned oxide layer on the first layer of silicon. When providing the pattern by etching, this additional oxide layer would act as an etch stop and the first layer of silicon is then maintained. It may be oxidized in the oxidation step and the gate may be applied together with the control gate. Another option is to define such access transistor in a top layer of the alternating layer stack.

The alternating layer stack used in the method of the invention has a same structure as a quantum well. In a further embodiment, therefore, other devices based on the quantum well may be combined in one circuit with the present memory transistor.

The alternating layer stack used in the method of the invention has etching and oxidation selectivity. In a further embodiment, therefore, other devices, such as 3D MEMS, multi-channel MOSFETs and variable sensors may be combined in one circuit with the present memory transistor.

5 The alternating layer stack used in the method of the invention provides a well-controlled capability for both floating gate sizes. In a further embodiment, therefore, other devices, such as qubit quantum computers, are provided with an asymmetric floating gate.

According to the fabrication method of the present invention, a first and a second
10 floating gate are self-aligned with the channel and a first and a second tunnel oxide are formed in one process step. This is beneficial so as to ensure efficient tunneling of charge carriers into the floating gates.

A principle of the operation of the multiple-level memory with two floating gates fabricated by the above-described method will now be described.

15 The condition to observe Coulomb blockade effect at room temperature is $K_B T < q^2/C_\Sigma$. For real device applications, the charging energy q^2/C_Σ should exceed the thermal energy $K_B T$ at least by a factor of three. Here C_Σ is the total capacitance of the floating gate. A first simple model for the floating gate is a silicon conducting sphere which has a self-capacitance of $C_\Sigma = 2\pi\epsilon_{ox} d$ (d is the diameter of the floating gate and ϵ_{ox} is the
20 permittivity of oxide). For the presence of Coulomb blockade effect at room temperature, the diameter of the floating gate is preferably than 8.5 nm from $K_B T = 0.3 q^2/C_\Sigma$. In this case, electrons can be controlled individually at room temperature according to the calculation.

A prior art single electron memory device as published in Solid State Electronics
25 44(2000), 2259-2264 is based on the arsenic-assisted etching and oxidation effects. However, the prior art memory has problems in that the tunnel oxide is thicker than desired and fixed by the arsenic doping concentration and energy. A thick tunnel oxide, corresponds to the range where Fowler Nordheim tunneling occurs and provides a long retention time. However, it needs a high operation voltage to inject electrons from the
30 channel into the floating gate. On the other hand, a very thin tunnel oxide corresponds to the range where direct tunneling occurs and reduces the operation voltage, but the

retention time becomes short due to leakage current. A suitable trade-off is therefore provided in an embodiment with an intermediate tunnel oxide thickness. A suitable range is for instance from 7 nm to 20 nm. Accordingly, a method of the present invention of multiple-level single electron memory with one or more floating gates can well define the tunnel oxide thickness by controlling the thickness of the SiGe layers.

A reading/writing/erasing method will now be described. This method is further illustrated with reference to Fig. 9. This Fig shows a graph of the drain current I_{DS} as a function of the voltage V_{GS} applied to the control gate. State 0 refers to the case of the multiple-level single electron memory according to the invention, to which no data is written; State 1 refers to the case of the multiple-level single electron memory according to the invention, to which data is written using the bottom floating gate. State 2 refers to the case of the multiple-level single electron memory according to the invention, to which data is written using the top and bottom floating gates.

A multiple-level writing method comprises following steps:

in a first step, the bottom floating gate of a typical memory cell as shown in FIG. 1 is charged with one electron. Hereto a source-drain voltage V_{DS} and a positive control gate voltage V_{GS} are applied, with values based on the Fowler-Nordheim tunnel model. In this case, a positive threshold voltage shift is observed in the drain current of an n-type MOSFET and state 1 is obtained as shown in FIG. 9;

in a second step, the top floating gate of the memory device as shown in FIG. 1 is charged by remaining the source-drain voltage V_{DS} and applying a positive control gate voltage higher than V_{GS} , a larger positive threshold voltage is observed as shown in FIG. 9 and state 2 is obtained.

An erasing method is performed, starting from State 2 as shown in Fig.9, by applying a negative gate voltage with an absolute value of V_{GS} to the control gate. Therewith, the electron stored in the bottom floating gate is removed and put back into the channel. Simultaneously, the electron stored in the top floating gate is returned back to the bottom floating gate. Erasing may be alternatively performed, starting from State 2 as shown in Fig. 9, by applying a negative gate voltage larger than the absolute value of V_{GS} to the control gate. Then two electrons stored in the both floating gates can go into the channel.

Performing reading is carried out in the following manner: the top floating gate is charged with a greater positive control gate voltage than that V_{GS} required by the bottom floating gate. Therewith, the threshold voltage variations of the top floating gate and the bottom floating gate become different. Hence, the States can be reviewed.,,. Two electrons
5 can be charged into the top and bottom floating gates, respectively, by applying the different levels of control gate voltage V_{GS} to the control gate, so that two states are defined by a word-line which reduces the number of the word-line and increases the integration density of the memory cell.

Also the memory according to the present invention is operated by one electron
10 and so can drastically shorten the writing and erasing time, compared to existing memory cells, thus a fast operation on the order of ns or faster is possible. Moreover, the memory according to the present invention has no degradation of the MOSFET performance since the channel of the MOSFET can be intrinsic or lowly-doped. Therefore, the present invention is applicable to ultra-high integration, low power consumption and fast access
15 speed flash memories, EEPROMs, EPROMs, DRAMs, SRAMs and others.

In short, the memory according to the present invention is a non-volatile memory which consumes an extremely small amount of power. This is the result of an optimized definition of the tunnel oxide by controlling the thickness of the SiGe layer. The storage in the floating gate may be optimized to store one electron based on the energy
20 quantization and Coulomb blockade effect. Also, the degraded performance of the prior art memory device, caused by arsenic-assisted etching and oxidation effects, is overcome by using a heterostructure, i.e. an alternating layer stack. Therewith, high integration, high performance, low power consumption and fast access speed can be achieved. For instance, a memory capacity of 1 Tb or greater may be obtained with the embodiment
25 having a double floating gate.

Above is an example for a multi-level memory application. The device according to the present invention also can be used for single-level application, to obtain a long retention time, which is described in example 2.

30 Example 1

A commercially available P-type SOI wafer having a resistivity of 15-25 Ω cm is

used as the starting substrate according to this embodiment of the present invention. Silicon overlay is about 50 nm. A layer of $\text{Si}_{0.7}\text{Ge}_{0.3}$ with a thickness of 15 nm and a layer of Si with a thickness of 20 nm are grown by an epitaxial method, so that the heterostructure is composed of a $\text{Si}/\text{Si}_{0.7}\text{Ge}_{0.3}/\text{Si}$ stack (20/15/50 nm) on a 145-nm thick buried oxide. A thin layer of Si_3N_4 with a thickness of 20 nm is deposited on the surface of the heterostructure in order to protect the surface of the top Si from oxidation. The layout as shown in FIG. 2 is transferred into the heterostructure by e-beam lithography and Reactive Ion Etching using the Si_3N_4 layer as a hard mask. The layout is composed of a central circle connected to wider source/drain regions by two narrow paths. Typically, the diameter of the floating gate region is chosen in such a way that after wet oxidation at 750°C for 1 hour, the $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer is totally oxidized due to a fast oxidation rate and a tunnel oxide ($\text{SiGe})\text{O}_2$ is formed. The diameter of the floating gate region is for instance in the order of 50-100 nm, for instance 75 nm. The tunnel oxide separates the channel from the floating gate. The channel has a triangular shape as a result of the patterning and oxidation steps. The diameter of the floating gate obtained from this example is less than 10 nm compatible with the Coulomb blockade effect at room temperature so that the electron can be injected one by one. The tunnel oxide obtained from this example is in the range from 7 nm to 10 nm considered in Fowler-Nordheim tunnel model which not only reduces the control gate voltage to inject one electron from the channel into the floating gate, but also reduces the current leakage from the direct tunnel effect.

Example 2

A fully self-aligned double floating-gate single-electron memory

Introduction

A critical issue of a (nanocrystal) NC FG (floating gate) memory is charge retention since a tunnel oxide between a channel and an NC FG is very thin in such a device. To obtain a long retention time, a double NC FG device is provided. Its working principle can be explained from Fig.10, in which both FGs are made of Si, wherein a top Si FG is charged by one or a few electrons from in this case a MOSFET channel through double tunnel junctions separated by a lower Si FG. The size of the lower FG is preferably below 10 nm, allowing quantum effects (quantum confinement and Coulomb blockade) to induce

splitting of an energy band into a number of energy levels at room temperature. The band-offset between the upper and lower FGs prevents a charge in the upper FG from returning to the channel, thereby resulting in a longer retention time. In other words, the double FGs are analogous to an ‘artificial dipole’, which is embedded in the gate oxide between the control gate and the MOSFET channel. An electric field surrounding the ‘artificial dipole’ is weaker than that surrounding a single FG in a single-electron memory device. Therefore, a charge can be stored in an upper FG for a relative long time. Above an example is provided wherein both FGs are composed of the same material, such as Si or polysilicon. This requires the size of the lower FG to be smaller than 10 nm for compatibility with the quantum effects at room temperature. The second example relates to use of different materials, namely Si and Ge as the upper and lower FGs, respectively. Compared with Si, Ge has a narrower band gap and a smaller effective mass. Therefore, the band-offset between both FGs is created automatically, which will amongst others suppress the charge leak. In this case, the size of the lower FG is preferably not scaled below 10 nm.

For fabrication of a single-electron memory with double FGs, a key technological issue is to perfectly align both FGs and precisely position them on a nanoscale MOSFET channel. Various attempts have been made.

It is proposed to use dopants (phosphorus) for the lower FG. The idea is very promising. However, such a technology is not able to precisely control the implantation of phosphorus into the appointed positions on the Si substrate.

A nonvolatile Si quantum memory fabricated by depositing Si nanocrystals on the top of an amorphous Si layer is reported. Self-aligned double NC FGs are obtained after oxidization. However, this device has multiple-FGs, in which lateral interference between FGs due to Coulomb interaction is one of the largest obstacles for this device application. Although alternative methods, wherein a double FGs structure is composed of nanocrystals or amorphous Si are currently in development and under investigation, the randomness of the deposition therein does not assure the desired sizes and defined positions.

The present invention is therefore the first demonstration of a fully self-aligned double FG single electron memory device. Up to now, the fully self-aligned single-electron

memory devices with only one Si NC FG and one Ge NC FG, have not yet been demonstrated. Here, based on a Si/ Si_{1-x}Ge_x /Si heterostructure and a smart layout, the inventors fabricate a new double FG single-electron memory, in which the upper Si NC FG, lower Ge NC FG and nanoscale MOSFET channel are fully self-aligned.
5 Importantly, the size of each FG is well defined by tuning the thickness of the Si and Si_{1-x}Ge_x alloy layers.

Fully self-aligned double floating-gate single-electron memory fabrication process

10 Si_{0.7}Ge_{0.3} and Si layers with a thickness of 20 nm are successively grown on a p-type silicon-on-insulator (SOI) wafer with 50-nm-thick Si film by epitaxy, so that a heterostructure composed of a Si/Si_{0.7}Ge_{0.3}/Si (20/20/50 nm) stack is formed on a 145-nm-thick buried oxide. It is important to note that the epitaxial method could be replaced by low-cost film deposition techniques. Further typical thicknesses vary from about 10
15 nm to about 50 nm for the heterostructure layers, from about 20 to about 100 nm for the Si-film, and from about 60 nm- 250 nm for the buried oxide.

A layer of Si₃N₄ with a thickness of from 5-30 nm, such as 10 nm is deposited on the surface of the heterostructure as a hard mask.

A pattern as shown in Fig. 2a is then defined on the heterostructure, such as by
20 electron-beam lithography. Thereafter the heterostructure is etched down to the buried oxide such as by Reactive Ion Etching. The pattern formed is composed of a central circle with a diameter of from 25-200 nm, such as 75 nm, connected to wider source/drain regions (from 100- 500 nm wide, such as 200-nm-wide) by narrow paths (from 25- 200 nm wide, such as 50-nm-wide). Then wet oxidation is carried out at elevated temperature,
25 such as at 750 °C, during an adequate time, such as for 75 min. During this step, a Si NC and a Ge NC are formed over a channel having a tip, such as triangular-like channel. The channel is an important feature obtained by the present invention, providing various advantages, as mentioned.

Figure 2b shows a cross-sectional transmission electron micrograph of a device
30 center after wet oxidation, in which the Ge NC is aligned not only with the top Si NC, but also with the bottom Si triangle.

Formation of a control gate oxide and double tunnel oxides is considered due to point defects presented at the $\text{Si}_3\text{N}_4/\text{Si}$ and $\text{Si}/\text{Si}_{0.7}\text{Ge}_{0.3}$ interfaces. These point defects provide reaction sites for the formation of SiO_2 , considered to increase the oxidation rate.

A reason for the Ge NC formation is considered as follows. According to a theory of binary alloy oxidation, oxide growth depends on alloy composition. For a case of the $\text{Si}_{1-x}\text{Ge}_x$ alloy oxidation, having a Si concentration above about 50%, Si is preferentially or selectively oxidized due to the fact that SiO_2 has a larger negative Gibbs free energy than GeO_2 . In the $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer, Si is first oxidized to form SiO_2 . Ge is thus agglomerated at the $\text{SiO}_2/\text{Si}_{0.7}\text{Ge}_{0.3}$ interface, forming a Ge-rich shell. Si next diffuses through the Ge-rich shell to react with oxygen which has diffused through the SiO_2 layer. As oxidation proceeds, Ge agglomerates continuously, and the Ge concentration in the Ge-rich shell increases. On the other hand, the oxygen concentration at the oxidation front decreases with the increase of the SiO_2 thickness. The decreasing oxygen concentration at the oxidation front counteracts the effect of increasing Ge concentration in the Ge-rich shell so that Ge is not oxidized and accumulates to form the Ge NC during the 75-min oxidation process. The self-limiting oxidation process results in a diameter of the Ge NC being equal to a thickness of the initial $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer.

Double tunnel oxides vertically separate the circle region of the device into top Si NC, middle Ge NC and bottom triangular nanowire, which are used as the upper FG, the lower FG and the channel of the present memory device, respectively. A smart layout (see Fig. 2a) restricts both FGs to the central circle since the Si and Ge NCs are completely oxidized in narrow paths (Fig. 2c). The wide source and drain regions are wide enough to avoid the formation of the tunnel oxides at the interfaces (Fig. 2d). Thus, the oxidation process only provides a small influence for the wide source and drain regions. As a result, both FGs are self-aligned to the channel and restricted in the center of the device, while source and drain regions are elevated.

It is worth noting that the diameter of the circle in Fig. 2a is chosen in such a way that after etch and wet oxidation, the dimension of the Si NC is in an example less than 10 nm to guarantee single electron injection at room temperature.

After wet oxidation, in an example the whole structure is uniformly doped by ion implantation, for instance using arsenic at energy of 140 keV and a dose of $2 \times 10^{14} \text{ cm}^{-2}$.

In a next step the doping impurities are subsequently annealed by rapid thermal annealing at 950°C for 60sec. This step leads to a junctionless transistor. More specifically, an n-type Si nanowire channel is connected by an n-type elevated source and elevated drain.

Finally a control gate is formed such as by electron-beam lithography, depositing a conducting layer, such as a platinum (Pt) layer, and followed by a lift-off process. Pt is chosen as control gate material because it has nearly the same work-function as P+ polycrystalline silicon. Similar materials are considered as well.

From Fig. 2b, it is observed that the control gate has an Ω configuration, which covers three edges of the device. Such is advantageous.

To further simplify the fabrication process, also memory devices with two side control gates are provided, offering a reduced number of masks compared with a process for memory devices with a top control gate. The fabrication process of the side control gate memory devices is nearly the same as that in the top control gate, except the implantation energy is changed.

As shown in Fig. 10, the triangular channel has a lower height than the source and drain regions. Therefore if the doping energy is well optimized, such as the As doping energy, it is possible to selectively implant the elevated source and drain only, while the channel remains undoped. In the present case, in an example the doping energy is set to 105 keV relating to its projected range being about 60 nm, which is little bit larger than the thickness sum of Si_3N_4 , Si and $\text{Si}_{0.7}\text{Ge}_{0.3}$ layers. As a result, an n-type transistor with a p-type channel and n+ source/ drain regions is obtained. At the same time, the side gates composed of the heterostructure also are doped by this step. This simplifies the fabrication process and leads to a corresponding reduction in cost.

The present fabrication method shows amongst others the following advantages.

(1) Both FGs and the nanowire MOSFET channel are simultaneously formed in a single process step and fully self-aligned.

(2) Both FGs are made from different materials (Si and Ge) and the band-offset between both FGs is automatically created. Therefore, in an example the lower FG must not be scaled below 10 nm to present quantum effects at room temperature.

(3) The smart layout restricts both FGs in the center of the device and results in elevated source and drain regions.

(4) The fabrication process is simplified. Specifically, for the top control gate device, the device operation is based on a junctionless transistor without the source/drain doping, while, for the side gate device, the control gate and source/drain regions are doped at the same time.

(5) Although both $\text{Si}_{0.7}\text{Ge}_{0.3}$ and Si layers with the identical thickness are grown on the SOI wafer, Figure 2b reveals that the diameter of the Ge NC is 3 times larger than that of the Si NC. This is considered to be attributed to formation of a Ge-rich shell, which acts as a diffusion barrier for the reaction between a $\text{Si}_{1-x}\text{Ge}_x$ alloy and oxygen outside the shell. If both FGs are required to be of the same size, in an example the Si layer must be about 3 times thicker than the $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer. This implies that each FG size and tunnel oxide thickness can be well defined in sub-nanometer-scale precision by controlling the thicknesses of the Si and $\text{Si}_{1-x}\text{Ge}_x$ alloy layers, respectively, and the composition of the $\text{Si}_{1-x}\text{Ge}_x$ alloy. This makes it possible to realize a multilevel bit memory cell with symmetric NC FGs and qubit quantum computers with asymmetric NCs.

In principle the two floating Gates can be interchanged. A consequence that both may have an FG coupling effect.

Further sizes may vary amongst the two FGs. In principle, the memory devices do not directly require that the floating gates be asymmetric.

It is noted that a Ge FG can be charged by an optimized gate voltage. In the second example, a goal is to store an electron in the upper floating gate for a long time (this is an single-level memory application). If the aim is to store an electron in the Ge floating gate for multi-level memory application, there is a need to turn the control gate voltage. It is envisaged to use the configurations Si/Si and Si/Ge as double floating gates. Each configuration has its advantage.

Self-aligned double floating-gate single-electron memory properties

A decisive signature of single electron injection is drain current hysteretic behavior with abrupt decrease, a stepwise decrease of a drain current as a function of gate voltage, where each step corresponds to rejection of one electron from the channel to the upper FG.

Figure 11 demonstrates hysteretic behavior of a self-aligned double floating-gate

single-electron memory at room temperature. When the gate voltage is swept from 0 to 8 V, four current shifts are observed. This suggests that a single electron is injected into the NC FG. The diameter of the upper Si FG obtained from Fig. 2b is 7.5 nm, which diameter is compatible with the Coulomb charging energy at room temperature. According to quantum theory, a condition to observe Coulomb blockade effect in a practical device is that the charging energy ($3q^2/C\Sigma$) is 3 times larger than the thermal energy (kBT). Here $C\Sigma$ is the total capacitance of the upper FG. From Fig. 2b, we can see that the upper FG is a Si conducting sphere, which has a self-capacitance of $C\Sigma = 2\pi\epsilon_0\epsilon_{ox}d$ (d is the diameter of the FG and ϵ_{ox} is the permittivity of SiO_2). For presence of Coulomb blockade effect at room temperature, in an example d must be less than 8.5 nm from $q^2/C\Sigma = 3kBT$. This is consistent with the size of the FG (7.5) from Fig. 2b. Therefore memory operation may be performed by one electron only. It can be found that when the gate voltage is swept reversely, only two back shifts are obtained. This is considered due to band-offset between both FGs prevents the electron stored in the upper FG from returning to the channel.

To further support single electron injection, time evolution of drain current is studied. First a negative pulse of -10 V is applied to the control gate to empty both FGs. Then a constant voltage of 7 V is set and maintained during the measurements. Figure 12 shows a typical drain current-time characteristics. When one electron is injected into the FG, an abrupt reduction of drain current is observed. The quantization of the current steps, $\Delta I = 0.3$ nA, confirms that they stem from single electron injection in a FG. By repeating the same measurement more than 100 times, the average injection time of the first electron is calculated to be 1s at 7 V, which is the nearly same order of magnitude compared to that in the single FG single-electron memory, showing 0.7 s at 5.58 V.

Figure 13 shows time dependence of a memory window width for a double FG single-electron memory. It is clear that after 104 sec, the memory window width has dropped to 0.25 V, but is not zero (0). The electron tunneling probability from the upper FG to the channel is much smaller in the double FG device because electrons are considered to have to overcome multiple barriers (two oxide barriers and the band-offset between both FGs), resulting in less leaking out of electrons from the upper FG, thereby providing a long retention time. It is found that the memory window (> 0.5 V) in a double

FG device is wider than that (< 0.5 V) in a single FG device for the first write/erase cycle. This is considered to relate to more electrons being stored in the lower Ge FG and the multiple barriers ensure the single electron injection from the lower Ge FG to the upper Si FG.

5 The above measurement results indicate that the addition of a lower Ge FG significantly improves charge retention, and at the same time, does not degrade the write/erase speed. It is noted that leak suppression is useful only in a low voltage region, while programming/erasing is performed by a high control gate voltage. It is worth noting that in the present device, a larger threshold voltage of about 4 V is obtained because of a
10 thicker gate oxide between the control gate and the active region. This can be seen from Fig. 2b, in which the gate oxide is composed of 10-nm-thick SiO_2 and 10-nm-thick Si_3N_4 . In an example the distance between the side gate and the device center is larger than 30 nm. To reduce the threshold voltage, the Si_3N_4 layer should be removed in the top control gate device and the side gate should be removed to approach the active region in the side
15 control gate device.

 The present invention therefore provides a fully self-aligned double floating-gate single-electron memory. Device fabrication is based on a $\text{Si}/\text{Si}_{1-x}\text{Ge}_x/\text{Si}$ heterostructure and a smart layout. In an example the device has an upper Si nanocrystal floating gate (7.5 nm) and a lower Ge nanocrystal floating gate (20 nm). Both floating gates and the
20 nanoscale channel of the device are fully self-aligned. The size of each FG can be well defined by modulating the thicknesses of the Si and $\text{Si}_{1-x}\text{Ge}_x$ alloy layers. This makes it possible to realize multilevel bit memory cells with symmetric NC FGs and qubit quantum computers with asymmetric NCs.

Claims

1. A method of manufacturing a semiconductor memory device provided with a source and a drain electrode between which a channel extends, one, two or more floating gates overlying said channel, and a control gate, which method comprises the steps of

5 providing a substrate with an insulating layer and an alternating layer stack of a first and a second semiconductor material, wherein the second semiconductor material oxidizes selectively with respect to the first semiconductor material, or reverse.

10 patterning the alternating layer stack down to the insulating layer of the substrate according to a predefined pattern including the source electrode, the drain electrode and the channel region extending between said electrodes, and

15 selectively oxidizing said patterned alternating layer stack, such that the second semiconductor material in the channel region is completely oxidized to separate the channel from the floating gate or such that the second semiconductor material agglomerates overlying the channel region as the floating gate.

2. The method as claimed in Claim 1, wherein the alternating layer stack comprising at least a first and a second layer of the first semiconductor material and a first layer of the second semiconductor material.

3. The method as claimed in Claim 2, wherein a substrate with a buried oxide layer and a top semiconductor layer is used, the top semiconductor layer constituting the first layer of the first semiconductor material of the alternating layer stack.

25 4. The method as claimed in any of Claims 1-3, wherein the predefined pattern specifies the source and the drain electrode with a larger width than the channel region.

30 5. The method as claimed in Claim 1 or 4, wherein the channel region as specified in the predefined pattern comprises a floating gate portion between first and second path portions, the path portions having a smaller width such that in the oxidation step both the first layer of second material and the second layer of the first material are fully oxidized.

6. The method as claimed in any of Claims 1-5, wherein the alternating layer stack comprises at least two layers of the first semiconductor material and at least one layer of the second semiconductor material, resulting in a first and a second floating gate overlying each other.

7. The method as claimed in Claim 5, wherein the channel region as specified in the predefined pattern comprises a first and a second floating gate portion, each between path portions.

8. The method as claimed in any of Claims 1-7, wherein the first semiconductor material comprises Si and the second semiconductor material comprises SiGe or Ge, or wherein the first semiconductor material comprises SiGe or Ge and the second semiconductor material comprises Si.

9. The method as claimed in any of Claims 1-8, wherein prior to the patterning a protection layer is deposited.

10. The method as claimed in any of Claims 1-9, wherein the control gate is embodied as a side control gate located adjacent to the floating gate and the channel, and wherein the side control gate is preferably specified in the predefined pattern.

11. A semiconductor memory device provided with:

a substrate with an insulating layer, preferably a buried substrate such as a SOI with a buried oxide, and an alternating layer stack of a first and a second semiconductor material, wherein the second semiconductor material oxidizes selectively with respect to the first semiconductor material or reverse,

a source and a drain electrode defined in the alternating layer stack, between which electrodes a channel extends in a first layer of the first material of the alternating layer stack ,

a first floating gate that overlies said channel and is defined in a second layer of

the first material of the alternating layer stack, an intermediate layer of the second semiconductor material between said first and second layer of the first semiconductor material being oxidized between said channel and said first floating gate, or wherein the first floating gate is defined in a first layer of the second material of the alternating layer stack, and
5 a control gate.

12. The semiconductor memory device as claimed in Claim 11, wherein a second floating gate is present, wherein the channel is a nanoscale channel, and wherein floating
10 gates are fully aligned with the nanoscale channel, wherein the floating gates preferably form an artificial dipole.

13. The semiconductor memory device as claimed in Claim 11 or 12, wherein at least the first and/or second floating gates comprise a nanocrystal, preferably both floating
15 gates comprise a nanocrystal.

14. The semiconductor memory device as claimed in any of Claims 11-13, wherein the second floating gate is defined in a third layer of the first material of the alternating layer stack, an intermediate layer of the second semiconductor material between said
20 second and third layer of the first semiconductor material being oxidized between said first and second floating gate, or wherein the second floating gate is defined in a first layer of the second material of the alternating layer stack.

15. The semiconductor memory device as claimed in any of claims 11-14, wherein
25 the channel is a nanowire, and/or wherein the channel has a tip-like structure, wherein the tip points towards the floating gates, such as a triangular-like structure, .

16. The semiconductor memory device as claimed in any of the Claims 11-15, further comprising an access transistor to form a 2-transistor memory cell together with the
30 memory device.

17. An array of memory devices as claimed in any of Claim 11-16, further comprising word lines coupled to the control gates and bit lines coupled to the drain electrodes.

5 18. A storage medium comprising a memory device as claimed in any of Claim 11-16 and/or an array according to claim 17, wherein the storage medium preferably has a capacity of 1 Tb or more.

19. An Integrated Circuit comprising a storage medium according to claim 18 and/or
10 a memory device as claimed in any of Claim 11-16 and/or an array according to claim 17.

20. A memory device, such as a non-volatile memory device, as claimed in any of Claim 11-16 in a MOSFET, EEPROM, EPROM, DRAM, SRAM, quantum computer or speed flash memory.

15

21. A method of operating the memory device as claimed in any of the Claims 11 to 16.

22. The method as claimed in Claim 21, wherein the memory is programmed or
20 erased in that charge is inserted into or removed from the floating gates individually by variation of an applied voltage on the control gate.

23. A memory device obtainable by the method of any of claims 1-10.

1/10

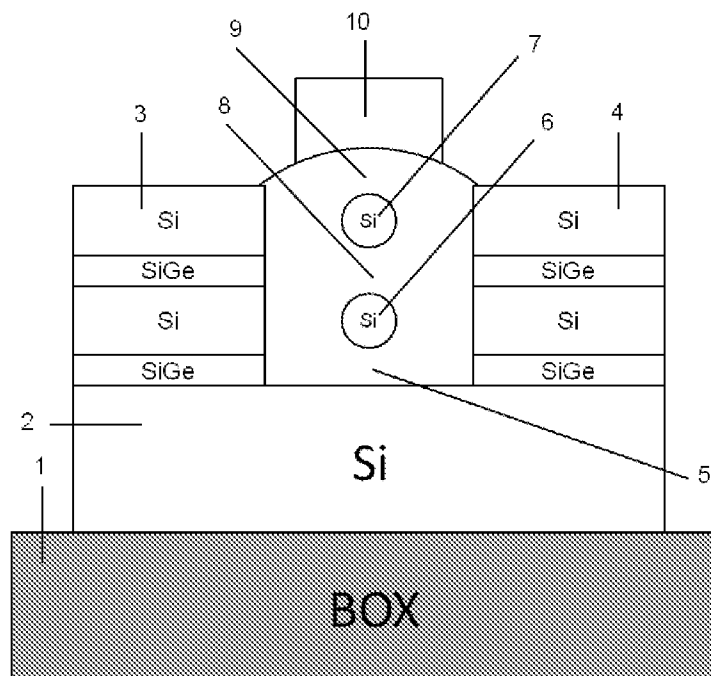


Fig. 1

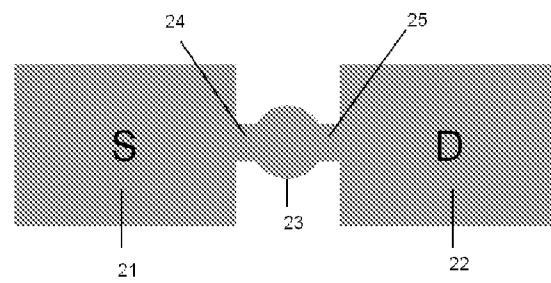


Fig. 2

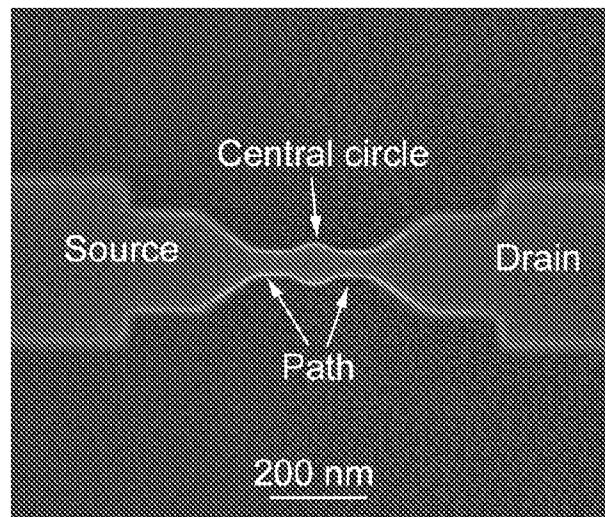


Fig. 2a

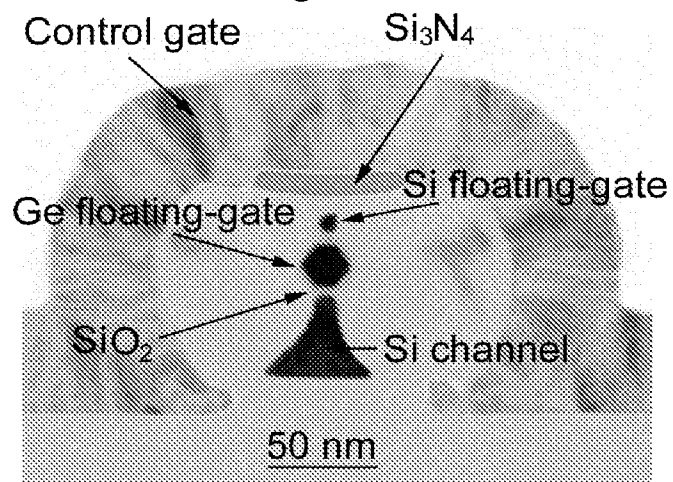


Fig. 2b

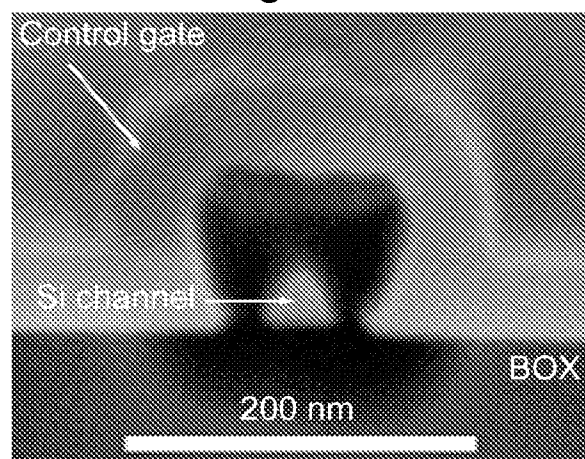


Fig. 2c

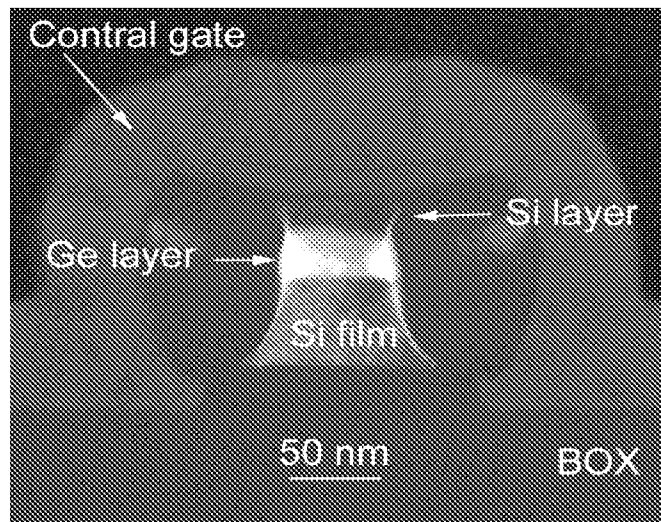


Fig. 2d

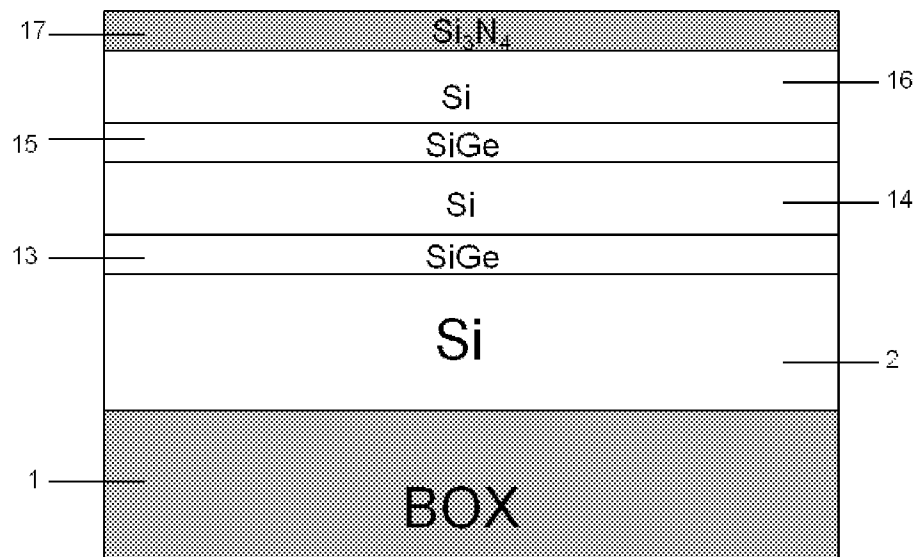


Fig. 3

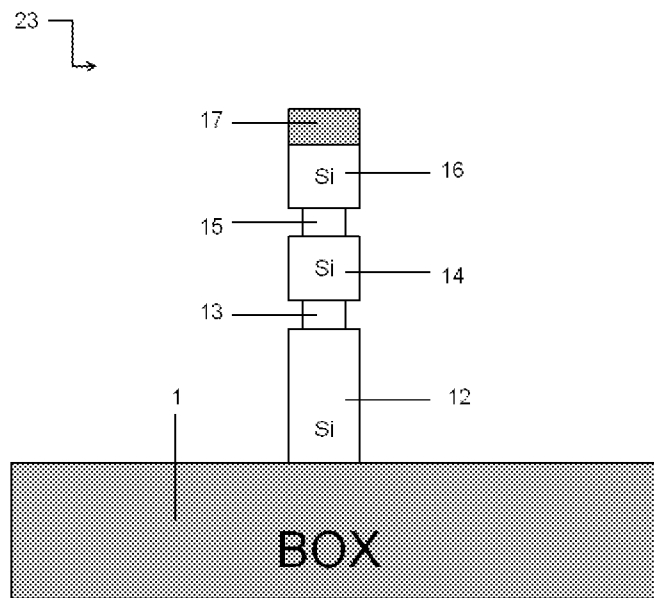


Fig. 4a

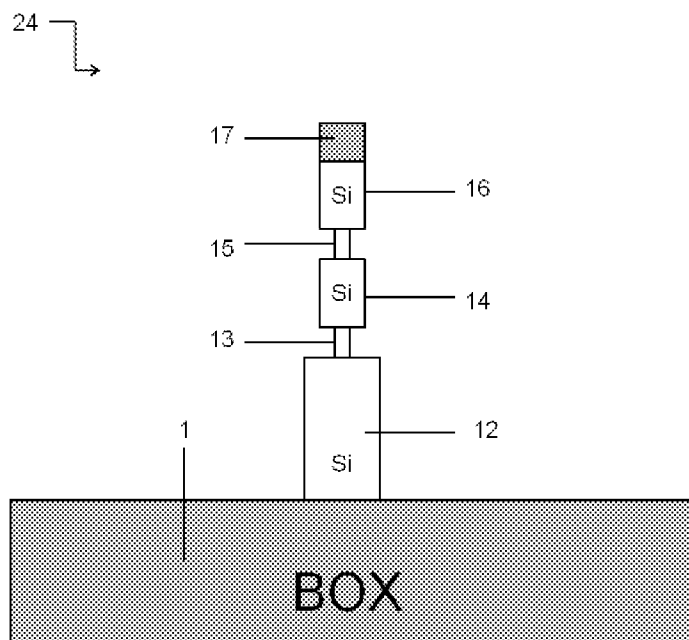


Fig. 4b

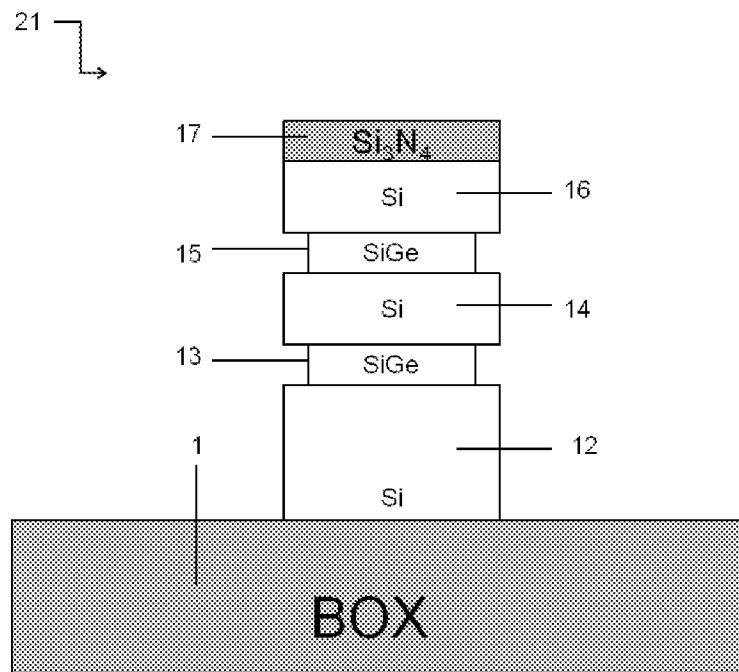


Fig. 4c

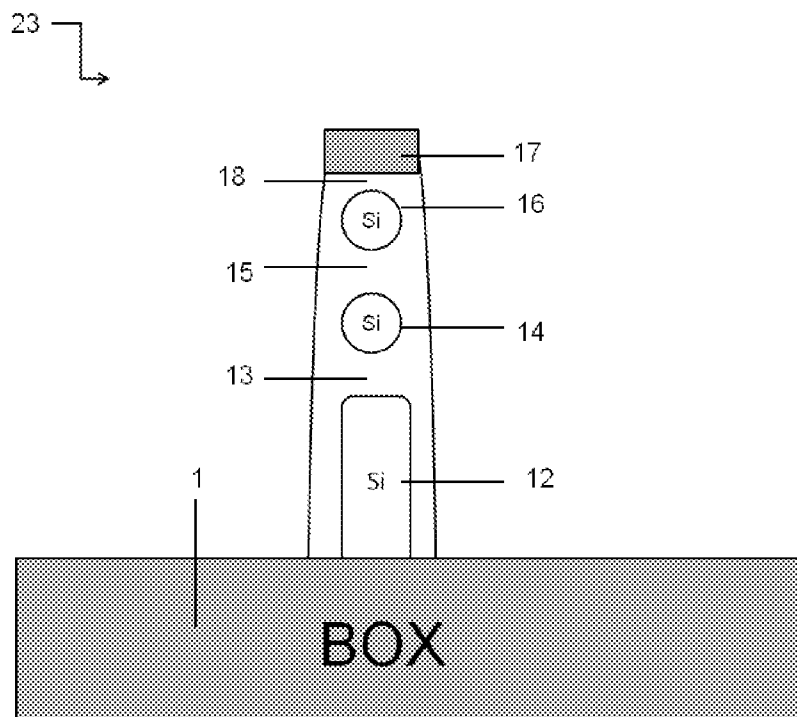


Fig. 5

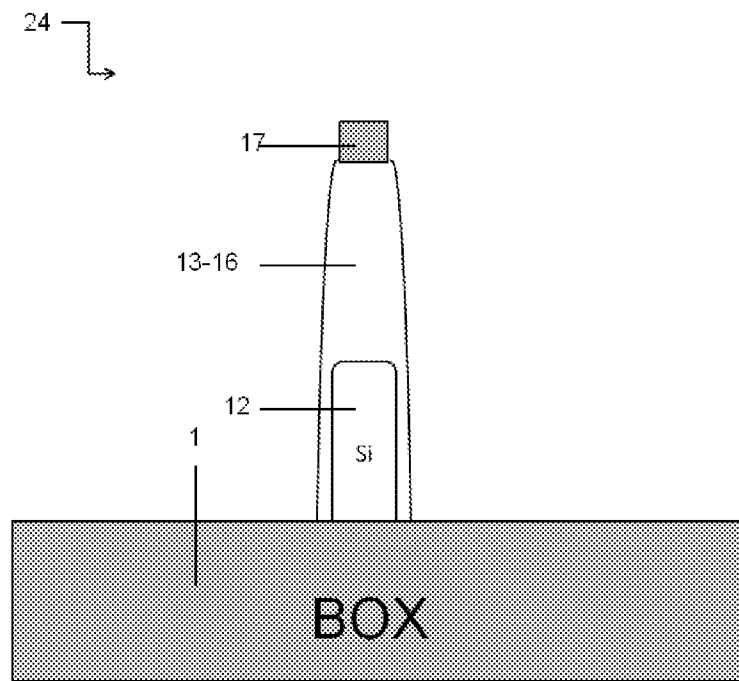


Fig. 6

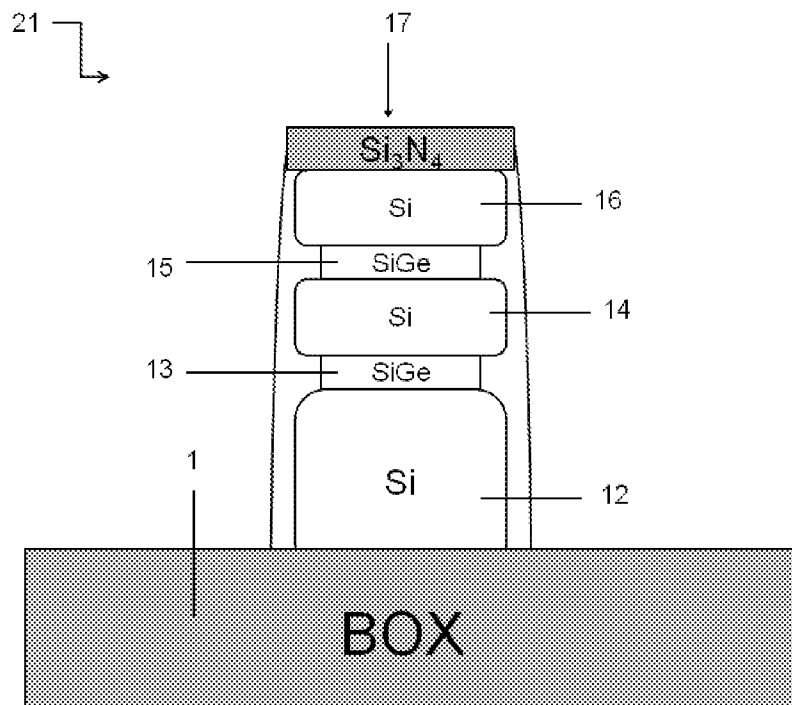


Fig. 7

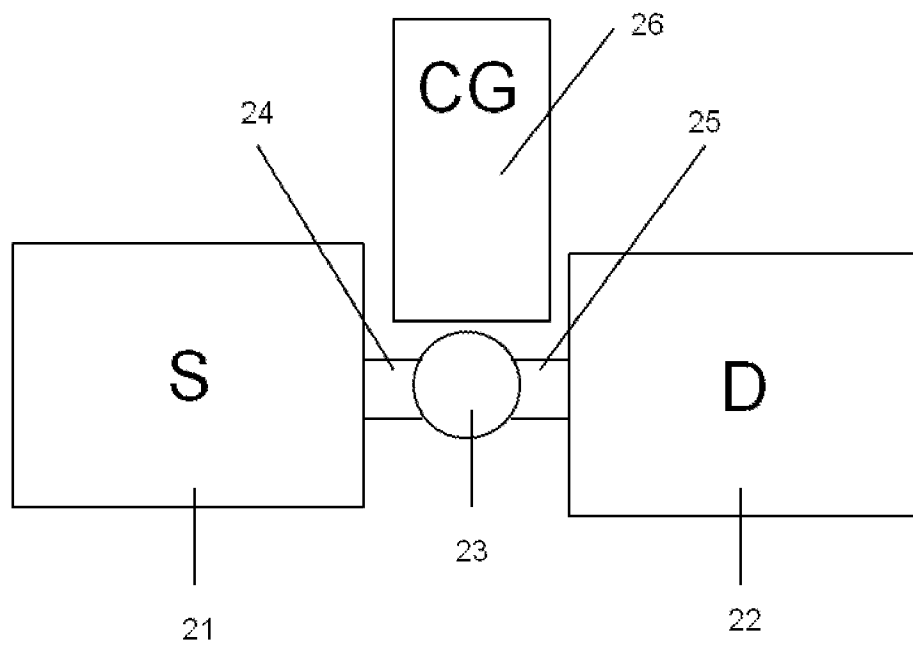


Fig. 8a

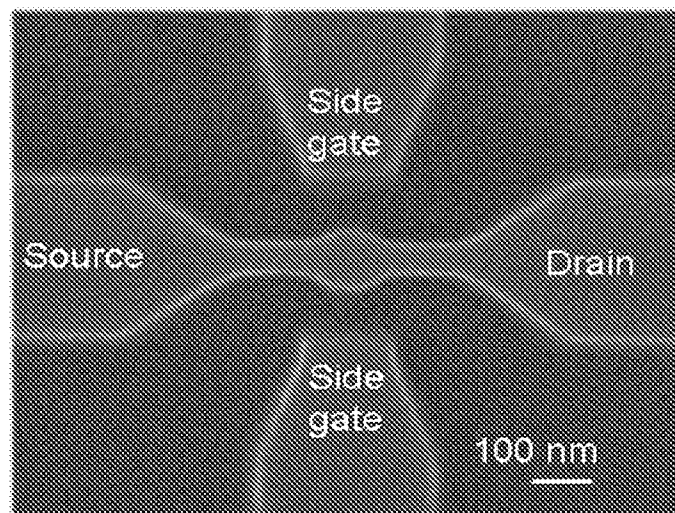


Fig. 8b

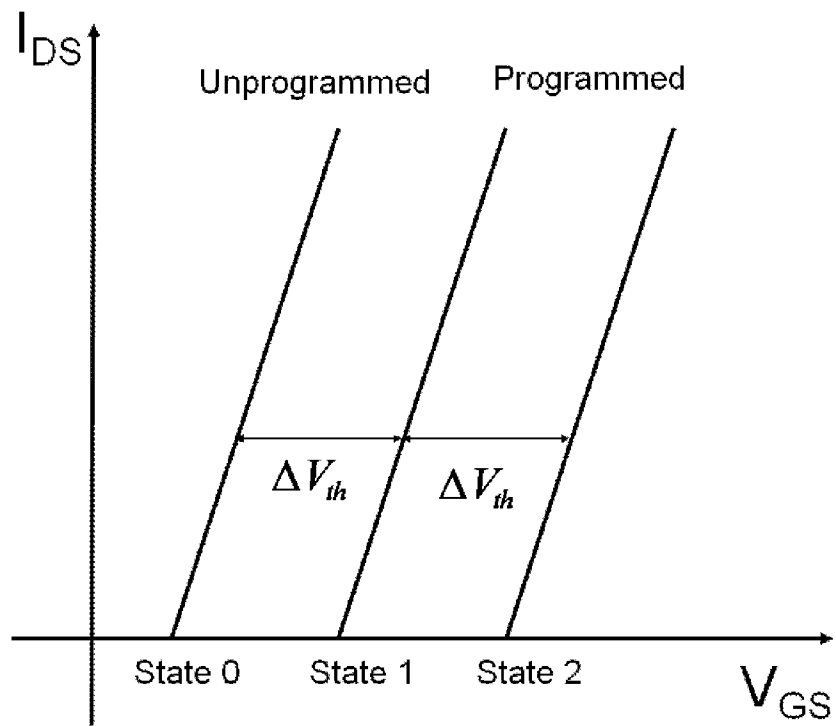


Fig. 9

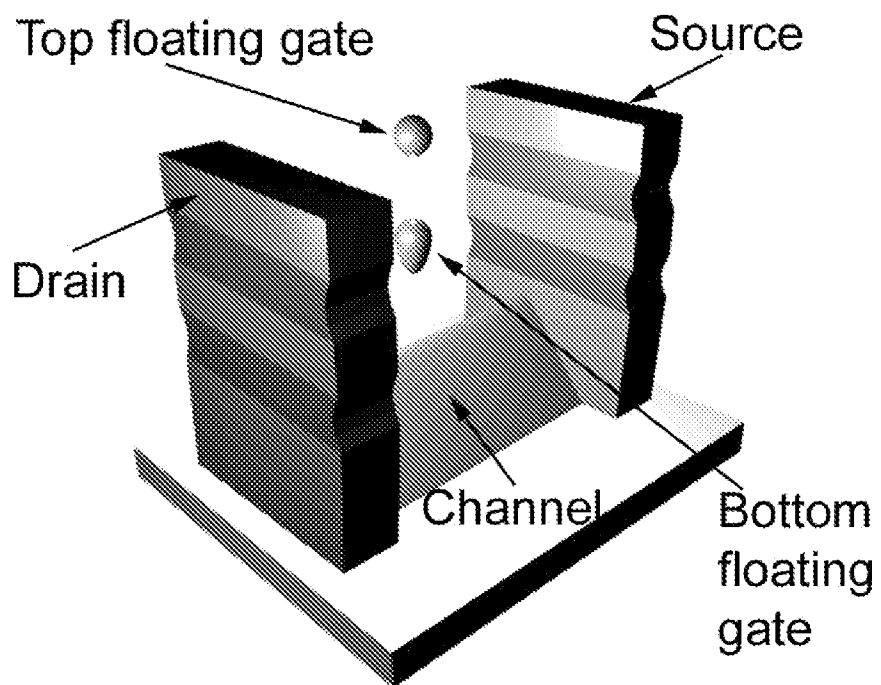


Fig. 10

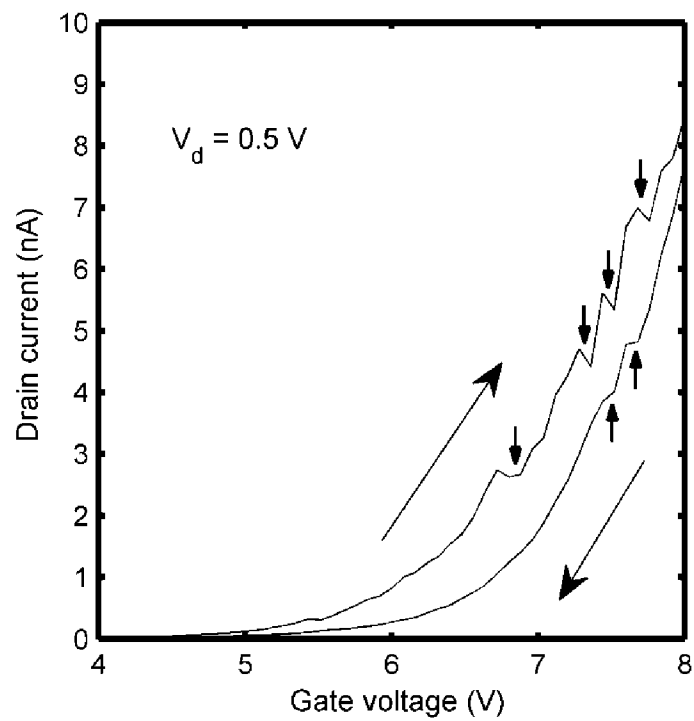


Fig. 11

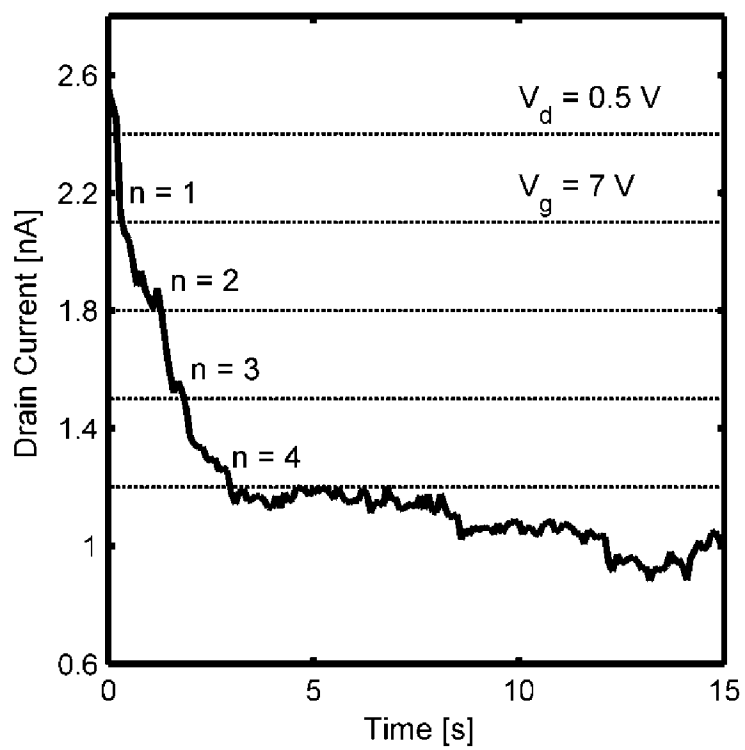


Fig. 12

10/10

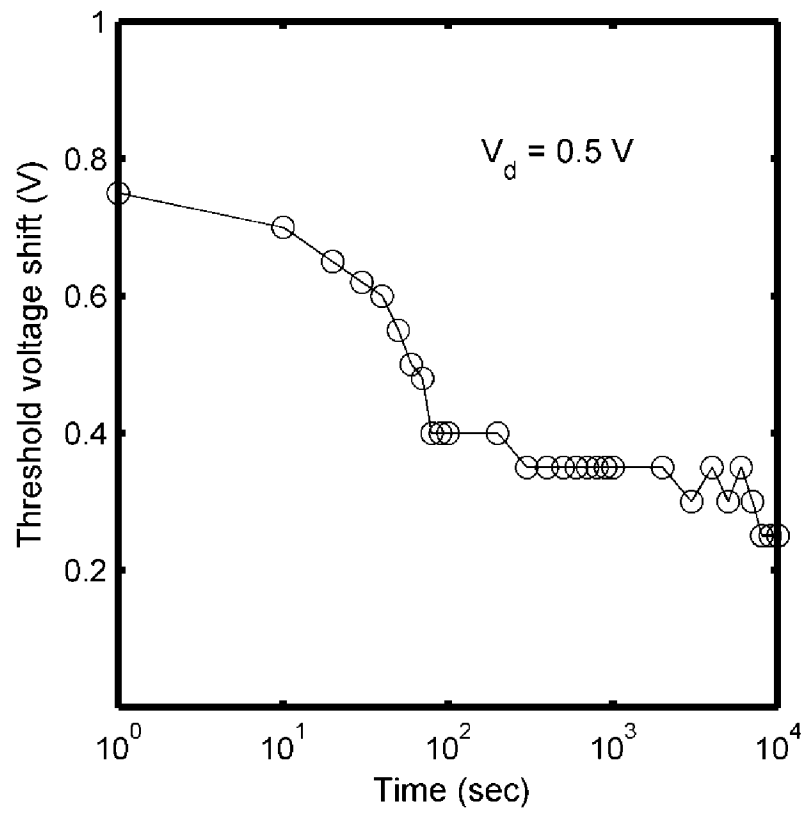


Fig. 13

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2010/064060

A. CLASSIFICATION OF SUBJECT MATTER

INV. H01L21/28

H01L21/336

H01L29/423

H01L29/788

ADD. G11C16/04

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H01L G11C

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TANG X ET AL: "Self-aligned silicon-on-insulator nano flash memory device", SOLID STATE ELECTRONICS, ELSEVIER SCIENCE PUBLISHERS, BARKING, GB, vol. 44, no. 12, 1 December 2000 (2000-12-01), pages 2259-2264, XP004227144, ISSN: 0038-1101, DOI: DOI:10.1016/S0038-1101(00)00221-5 cited in the application	11,13, 15-23
A	page 2259 - page 2263; figures 1-10 ----- -/--	1-10,12, 14



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

19 November 2010

Date of mailing of the international search report

30/11/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Albrecht, Claus

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2010/064060

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2008/069765 A1 (AGENCY SCIENCE TECH & RES [SG]; LO GUO QIANG [SG]; BERA LAKSHMI KANTA) 12 June 2008 (2008-06-12)	11, 13, 15-23
A	paragraph [0034] - paragraph [0050]; figures 1A-4 . -----	1-10, 12, 14
A	US 5 886 380 A (NAKAJIMA ANRI [JP]) 23 March 1999 (1999-03-23) cited in the application column 3, line 1 - column 11, line 6; figures 1A-8 -----	1-23

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2010/064060

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008069765 A1	12-06-2008	US 2008135949 A1	12-06-2008
US 5886380 A	23-03-1999	JP 9312378 A	02-12-1997