(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 April 2006 (20.04.2006)

PCT

(10) International Publication Number
WO 2006/040165 A2

(51) International Patent Classification:
H01L 21/8246 (2006.01) *H01L 21/336* (2006.01)

(21) International Application Number:
PCT/EP2005/011039

(22) International Filing Date: 13 October 2005 (13.10.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/967,014 15 October 2004 (15.10.2004) US

(71) Applicant *(for all designated States except US)*: **INFINEON TECHNOLOGIES AG** [DE/DE]; St.-Martin-Str. 53, 81669 München (DE).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **VERHOEVEN, Martin** [NL/DE]; Am Jacobstein 30, 01445 Radebeul (DE).

(74) Agent: **EPPING HERMANN FISCHER PATENTANWALTSGESELLSCHAFT MBH**; Ridlerstrasse 55, 80339 München (DE).

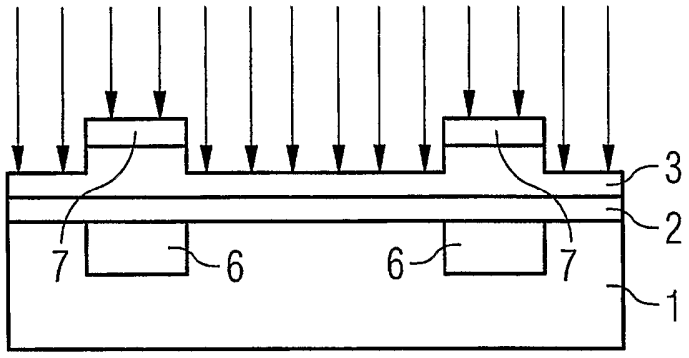
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK,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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHOD FOR PRODUCTION OF CHARGE-TRAPPING MEMORY CELLS



(57) **Abstract:** An oxide layer, a nitride layer, and a layer of amorphous silicon are applied to a surface of a semiconductor substrate. A resist mask is applied and implantations are performed to form doped regions of source and drain and doped regions within the amorphous silicon layer. The resist mask and undoped parts of the amorphous silicon are removed to form a silicon mask. The silicon mask is applied to etch back the nitride layer. After a removal of the silicon mask, the nitride is oxidized to form an oxide-nitride-oxide layer sequence which is laterally restricted to the area above the source/drain regions.

WO 2006/040165 A2

Description

Method for production of charge-trapping memory cells

Technical field

The invention concerns the fabrication of charge-trapping memory cells comprising an oxide-nitride-oxide memory layer sequence and being intended to store two bits of information.

Background of the invention

Non-volatile memory cells that are electrically programmable and erasable can be realized as charge-trapping memory cells, which comprise a memory layer sequence of dielectric materials with a memory layer between confinement layers of dielectric material having a larger energy band gap than the memory layer. The memory layer sequence is arranged between a channel region within a semiconductor body and a gate electrode provided to control the channel by means of an applied electric voltage. Charge carriers moving from source to a drain through the channel region are accelerated and gain enough energy to be able to penetrate the lower confinement layer and to be trapped within the memory layer. The trapped charge carriers change the threshold voltage of the cell transistor structure. Different programming states can be read by applying the appropriate reading voltages. Examples of charge-trapping memory cells are the SONOS memory cells, in which each confinement layer is an oxide and the memory layer is a nitride of the semiconductor material, usually silicon.

Typical applications of memory products require a steady miniaturization of the memory cells. A reduction of the area that is required by an individual memory cell can be obtained by shrinking the cell structure or by an increase of the number of bits that can be stored within one memory cell transistor structure.

A publication by B. Eitan et al., "NROM: a Novel Localized Trapping, 2-Bit Nonvolatile Memory Cell" in IEEE Electron Device Letters, volume 21, pages 543 to 545 (2000), describes a charge-trapping memory cell with a memory layer sequence of oxide, nitride and oxide which is especially adapted to be operated with a reading voltage that is reverse to the programming voltage (reverse read). The oxide-nitride-oxide layer sequence is especially designed to avoid the direct tunneling regime and to guarantee the vertical retention of the trapped charge carriers. The oxide layers are specified to have a thickness of more than 5 nm. Two bits of information can be stored in every memory cell.

In order to provide a better two-bit separation in charge-trapping memory cells, several different structures of an arrangement of separate memory layers of dielectric material or floating gate electrodes at both sides of the gate electrode above the source and drain junctions at the channel ends have been proposed. During the write operation to program the memory cell, channel-hot electrons are injected predominantly in the ONO area just above the pn-junction at the drain. A reversal of the electric voltage between source and drain enables the storage of a second bit at the other channel end.

In the course of further miniaturization of the memory cell, the problem of a precise arrangement and localization of the

memory layer with respect to the gate electrode and the regions of source and drain is of increasing importance. The further shrinking of the cell dimensions will imply a greater difficulty to separate the two bits that are stored in the same memory cell. This derives from the fact that electrons are to some extent injected also in the area between the regions of source and drain. Therefore, memory layer structures have been proposed, in which the memory layer is interrupted above the channel region.

Summary of the invention

It is an object of the present invention to provide an improved fabrication method for charge-trapping memory cells that are intended for two-bit storage.

It is a further object of the invention to provide a method for the fabrication of charge-trapping memory cells with improved two-bit separation that is suitable for a shrinkage of the device structures.

It is still a further object of this invention to provide the aforementioned methods with standard process steps of semiconductor technology.

The method according to this invention comprises the steps of applying an oxide layer, a nitride layer, and a layer of amorphous silicon onto a main surface of a semiconductor substrate, applying a resist mask with openings and performing an implantation of doping atoms to form doped regions of source and drain. By a further implantation step, the layer of amorphous silicon is provided with a dopant in areas lo-

cated above the regions of source and drain. The resist mask and parts of the silicon layer that have not been implanted are subsequently removed and the remaining parts of the silicon layer are used as a silicon mask in further process steps. The nitride layer beneath the layer of amorphous silicon is partly etched back in the areas that are not covered by the silicon. Then, the silicon layer is removed and the nitride is oxidized until only parts of the nitride layer remain within areas above the source and drain regions. In this manner, oxide-nitride-oxide memory layer sequences are formed that are laterally restricted to the areas of source and drain and formed in self-aligned fashion with respect to the source and drain regions.

A preferred alternative comprises a further method step, by which the resist mask is laterally reduced or trimmed between the implantation steps to form the source and drain regions and to form the doped regions within the amorphous silicon layer so that the produced ONO layer slightly extends over the lateral boundaries of the source and drain regions.

These and other objects, features and advantages of the invention will become apparent from the following brief description of the drawings, detailed description and appended claims and drawings.

Brief description of the drawings

Figure 1a shows a cross-section of a first intermediate product of an example of the inventive method after the application of the amorphous silicon layer and resist mask.

Figure 1b shows the cross-section according to figure 1a after the implantation steps.

Figure 1c shows the cross-section according to figure 1b after the removal of the resist mask and non-implanted parts of the silicon layer.

Figure 1d shows the cross-section of figure 1c after the etching of the nitride layer.

Figure 1e shows the cross-section according to figure 1d after an oxidation step.

Figure 1f shows the cross-section of figure 1e after the application of the gate conductor.

Figure 2a shows a cross-section according to figure 1a.

Figure 2b shows the cross-section according to figure 2a after the implantation of the source and drain regions.

Figure 2c shows the cross-section of figure 2b after a pull-back step to widen the openings in the resist mask.

Figure 2d shows the cross-section according to figure 2c after a further implantation step.

Figure 2e shows the cross-section of figure 2d after the removal of the resist mask and non-implanted regions of the silicon layer.

Figure 2f shows the cross-section of figure 2e after an etching of the nitride layer.

Figure 2g shows the cross-section according to figure 2f after an oxidation step.

Figure 2h shows the cross-section according to figure 2g after the application of the gate conductor.

Detailed description

The general method according to this invention is first described with reference to figures 1a to 1f, which show different intermediate products of an example of the method. According to the cross-section shown in figure 1a, a substrate 1 of semiconductor material, preferably silicon, is provided with a layer sequence comprising an oxide layer 2 that is applied on a main surface of the substrate, a nitride layer 3 and a layer of amorphous silicon 4. A resist mask 5 is applied, which has openings in the areas of the regions of source and drain to be formed by a subsequent implantation step.

Figure 1b shows the cross-section of the intermediate product according to figure 1a after the performance of two implantation steps. This is indicated in figure 1b by the arrows pointing downwards into the regions in which a dopant is implanted to form doped regions. A deep implantation forms the regions of source and drain 6. A shallow implantation forms doped regions within the layer of amorphous silicon 4 in areas that are located above the source/drain regions 6. The sequence of implantation steps is not fixed; it is preferred to perform the deep implantation first and the shallow implantation afterwards.

Figure 1c shows a further intermediate product in a cross-section according to figure 1b after the removal of the resist mask 5 and of those parts of the amorphous silicon layer which have not been implanted. The remaining parts of the silicon layer 4, which are doped, form a silicon mask 7 above the areas of the source/drain regions 6.

Figure 1d shows the cross-section according to figure 1c for a subsequent etching step, indicated by the arrows in figure 1d, by which the nitride layer 3 is partly removed in a vertical direction. The silicon mask 7 is applied to restrict the etching to areas between the source/drain regions 6. Above the source/drain regions 6, the nitride layer 3 remains in its original thickness.

Figure 1e shows a further intermediate product after the removal of the silicon mask 7 and an oxidation step to form a second oxide layer 8. This second oxide layer 8 comprises the original oxide layer 2 and parts of the nitride layer 3 which are completely converted into oxide, thus forming the second oxide layer 8. The thickness of the nitride layer 3 and the depth of the etching step shown in figure 1d are adapted so that the oxidation step forms a thorough oxide layer 8 in the areas between the source/drain regions 6, while thin remaining layer parts of the nitride layer 3 are left above the source/drain regions 6. In this way, an oxide-nitride-oxide layer sequence is formed above the source/drain regions 6 in a self-aligned manner with respect to the source/drain regions 6. Thus, the memory layer sequence can be arranged exactly above the source and drain regions and completely interrupted above the channel region provided between source and drain.

Figure 1f shows the cross-section according to figure 1e after the application of a gate conductor 9 to form gate electrodes above the channel regions and wordlines to connect the gate electrodes along rows of memory cell arrays.

Figures 2a to 2h show cross-sections of intermediate products of an alternative of the inventive method, which is especially preferred. It may be desired to have memory layer sequences above the pn junctions of the source and drain regions adjacent to the channel. This can be accomplished by the following method, which comprises an additional method step between the two implantation procedures.

Figure 2a shows the cross-section according to figure 1a, showing that the point of departure is the same as in the general method.

Figure 2b shows the subsequent implantation step to form the source/drain regions 6. This alternative embodiment comprises a further method step after the deep implantation indicated in figure 2b.

Figure 2c shows this further method step, which is a pull-back step to widen the openings of the resist mask 5. This is indicated in figure 2c by the arrows in the form of triangles and the broken lines representing the original contours of the resist mask 5. The larger openings, which are produced in this way, define the area of the later oxide-nitride-oxide layer sequence, which is intended as storage means.

Figure 2d shows the cross-section according to figure 2c for the further implantation step, by which those regions of the

amorphous silicon layer 4 are doped which are left free by the widened openings of the resist mask 5. These doped regions are self-aligned to the source/drain regions 6 at least as far as the pull-back step according to figure 2c can be controlled, but slightly extend over the lateral boundaries of the source/drain regions.

Figure 2e shows the cross-section of figure 2d after the removal of the resist mask 5 and the undoped regions of the layer 4 of amorphous silicon. In this way, a silicon mask 7 is formed, which has slightly smaller openings as compared to the silicon mask 7 which is applied in the first embodiment of the method described above.

Figure 2f shows the cross-section according to figure 2e for the subsequent etching step, by which those parts of the nitride layer 3 which are not covered by the silicon mask 7 are removed to a certain predefined depth.

Figure 2g shows the cross-section according to figure 2f after the removal of the silicon mask 7 and the performance of an oxidation step to form the second oxide layer 8 according to the first embodiment of the method. A comparison between figures 2g and 1e shows the difference in the lateral extension of the formed ONO layer.

Figure 2h shows the cross-section of the product that is obtained after the application of the gate conductor 9. Further standard process steps which are known per se can follow to complete this device.

Although the present invention and its advantages have been described in detail, it should be understood that various

changes, substitutions and alterations can be made herein without departing from the spirit and scope of the invention as defined by the appended claims.

List of reference numerals

1 substrate
2 oxide layer
3 nitride layer
4 layer of amorphous silicon
5 resist mask
6 source/drain region
7 silicon mask
8 second oxide layer
9 gate conductor

We claim:

1. Method for production of charge-trapping memory cells with separate memory layers for two-bit separation, comprising:
providing a semiconductor substrate;
applying an oxide layer on said substrate;
applying a nitride layer on said oxide layer;
applying a layer of amorphous silicon on said nitride layer;
applying a resist mask with openings on said layer of amorphous silicon;
applying said resist mask in a subsequent implantation to form doped regions of source and drain and to provide said layer of amorphous silicon with doped regions that are located above said doped regions of source and drain;
removing said resist mask;
removing part of said layer of amorphous silicon that has not been provided with a doping, to form a silicon mask;
applying said silicon mask to etch back said nitride layer;
removing said silicon mask;
performing an oxidation to convert all of said nitride layer into oxide, except for parts of said nitride layer that are located in areas above said doped regions of source and drain, thus forming an oxide-nitride-oxide layer sequence above these areas; and
applying a gate conductor provided as gate-electrode and wordline.

2. Method according to claim 1, further comprising:
widening the openings of said resist mask after the formation of said doped regions of source and drain and before the formation of said doped regions in said layer of amorphous silicon.

3. Method according to claim 1 or 2, further comprising: providing the layers within said oxide-nitride-oxide layer sequence with thicknesses that are suitable for a storage by charge trapping.

FIG 1A

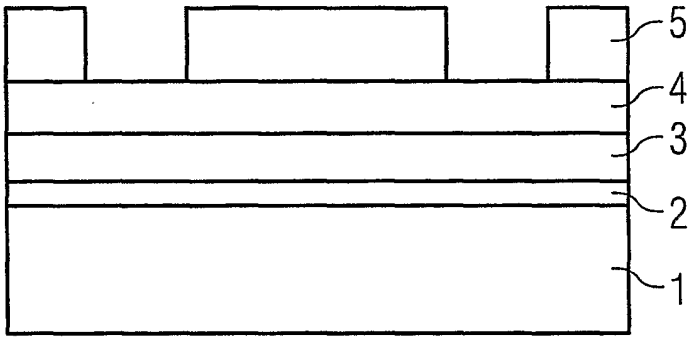


FIG 1B

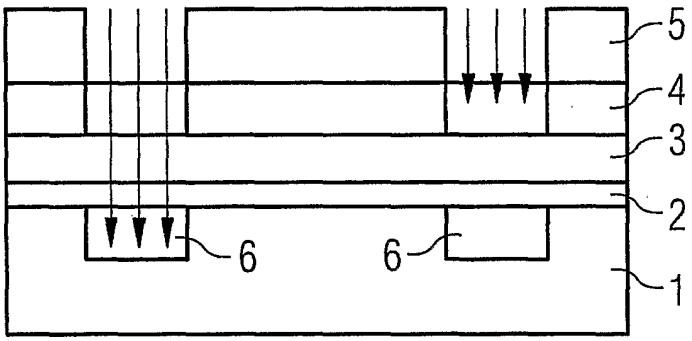


FIG 1C

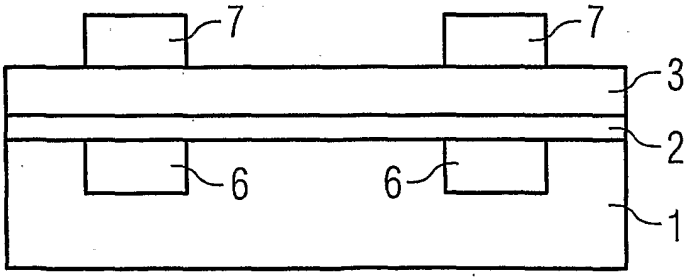


FIG 1D

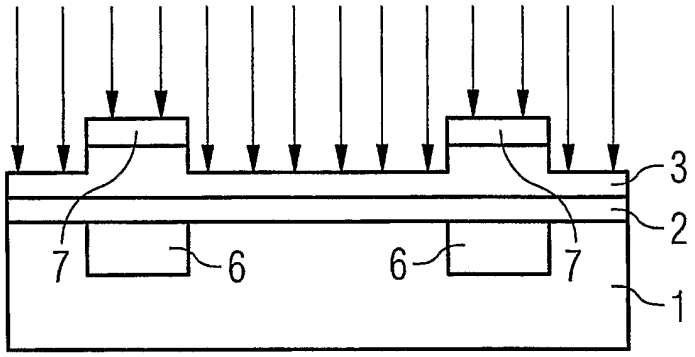


FIG 1E

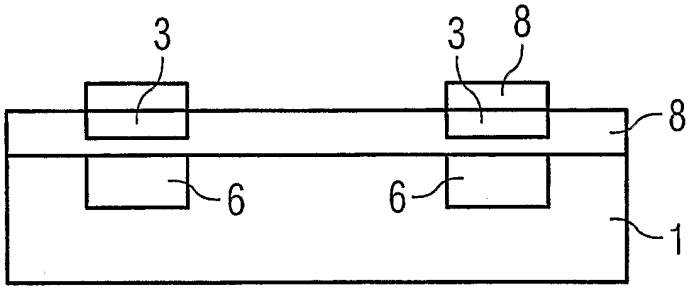


FIG 1F

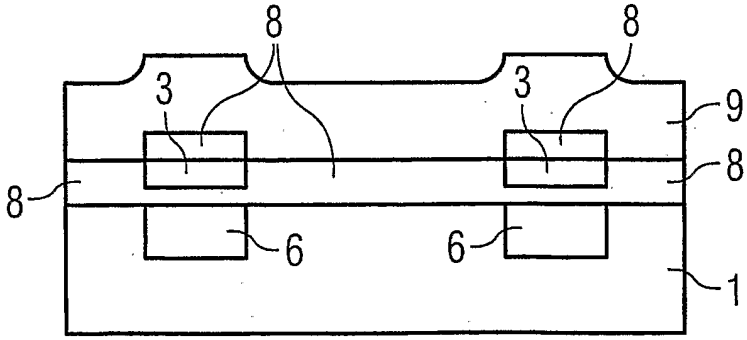


FIG 2A

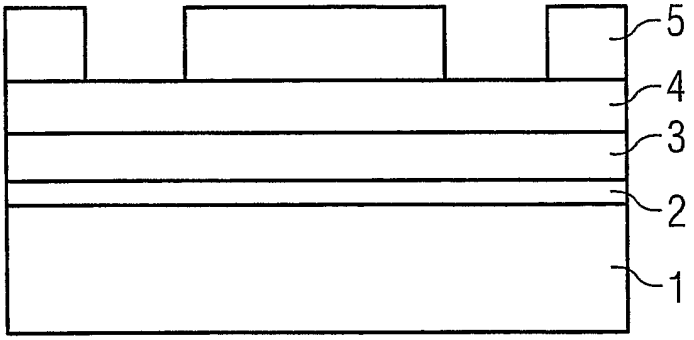


FIG 2B

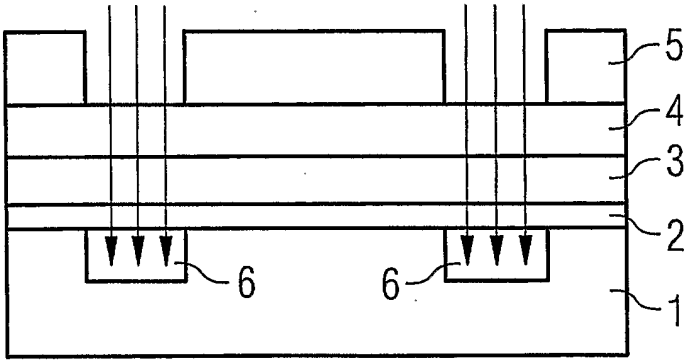


FIG 2C

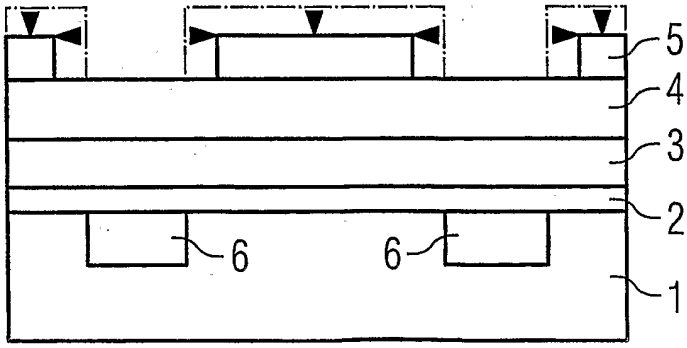


FIG 2D

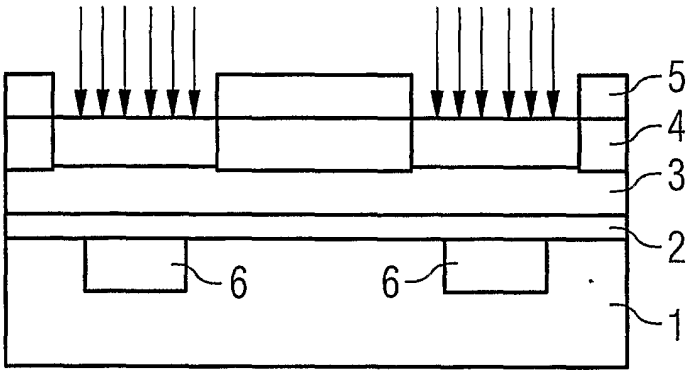


FIG 2E

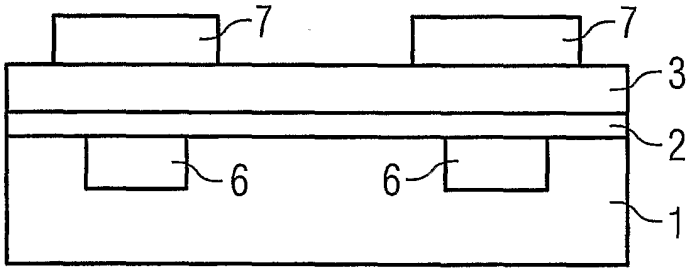


FIG 2F

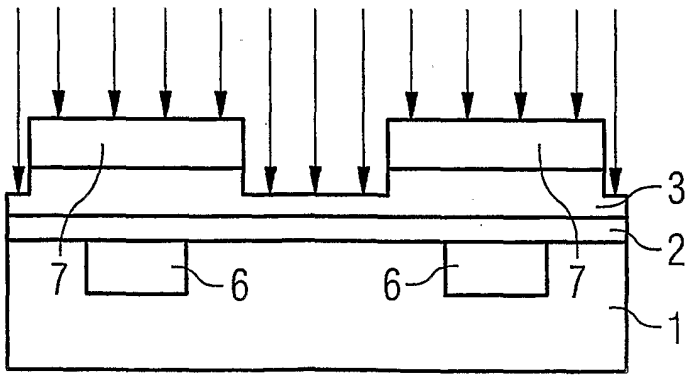


FIG 2G

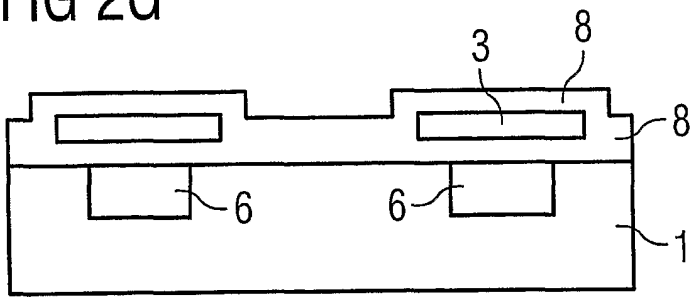


FIG 2H