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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 May 2006 (18.05.2006)

PCT

(10) International Publication Number
WO 2006/053258 A2

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

(21) International Application Number:
PCT/US2005/041051

(22) International Filing Date:
10 November 2005 (10.11.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/904,461 11 November 2004 (11.11.2004) US

(71) Applicant *(for all designated States except US)*: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, NY 10504 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **YANG, Haining, S.** [US/US]; 36 Robinson Lane, Wappingers Falls, NY 12590 (US).

(74) Agent: **CAPELLA, Steven**; IBM Corporation, 2070 Route 52, Bldg. 321, Zip 482, Hopewell Junction, NY 12533 (US).

(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, KN, 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.

WO 2006/053258 A2

(54) Title: METHOD TO ENHANCE CMOS TRANSISTOR PERFORMANCE BY INDUCING STRAIN IN THE GATE AND CHANNEL

(57) Abstract: A method of manufacturing complementary metal oxide semiconductor transistors forms different types of transistors such as N-type metal oxide semiconductor (NMOS) transistors and P-type metal oxide semiconductor (PMOS) transistors (first and second type transistors) on a substrate (12). The method forms an optional oxide layer (52) on the NMOS transistors and the PMOS transistors and then covers the NMOS transistors and the PMOS transistors with a hard material (50) such as a silicon nitride layer. Following this, the method patterns portions of the hard material layer (50), such that the hard material layer remains only over the NMOS transistors. Next, the method heats (178, 204) the NMOS transistors and then removes the remaining portions of the hard material layer (50). By creating compressive stress in the gates (22) and tensile stress (70) in the channel regions of the NMOS transistors (NFETs), without creating stress in the gates (20) or channel regions of the PMOS transistors (PFETs), the method improves performance of the NFETs without degrading performance of the PFETs.

METHOD TO ENHANCE CMOS TRANSISTOR PERFORMANCE BY INDUCING STRAIN IN THE GATE AND CHANNEL

BACKGROUND OF THE INVENTION

Field of the Invention

This invention is in the field of using strain engineering to improve CMOS transistor device performance. More specifically, it relates to inducing strain in a transistor channel by modulating the stress in the gate.

Description of the Related Art

Complementary metal oxide semiconductor (CMOS) device performance may be improved or degraded by the stress applied to the channel region. The stress may be applied by bending the wafer or by placing a stressful material nearby. When tensile stress is applied to N-type metal oxide semiconductor (NMOS) along its channel direction, electron mobility is improved resulting in higher on-current and speed. On the other hand, NMOS performance is degraded when the stress is compressive. P-type metal oxide semiconductor (PMOS) device performance may be improved using a compressive stress to enhance hole mobility. Similarly, PMOS performance will be degraded by a tensile stress applied along the channel direction.

SUMMARY OF THE INVENTION

The method of manufacturing complementary metal oxide semiconductor transistors presented herein forms different types of transistors such as N-type metal oxide semiconductor (NMOS) transistors and P-type metal oxide semiconductor (PMOS) transistors (first and second type transistors) on a substrate. The invention forms an optional oxide layer on the NMOS transistors and the PMOS transistors and then covers the NMOS

transistors and the PMOS transistors with a hard material such as a silicon nitride layer. Following this, the invention patterns portions of the silicon nitride layer, such that the silicon nitride layer remains only over the NMOS transistors. Next, the invention heats the NMOS transistors and then removes the remaining portions of the silicon nitride layer.

The optional oxide layer is used as an etch stop layer to control the process of removing the remaining portions of the silicon nitride layer. The heating process creates compressive stress in the gate, which in turn causes tensile stress in channel regions of transistors that were covered by the silicon nitride layer. Thus, the heating process creates tensile stress in channel regions of the NMOS transistors without causing tensile stress in channel regions of the PMOS transistors. More specifically, during the heating process, volume expansion of gate conductors of the NMOS transistors is restricted, resulting in compressive stress in the gate conductors of the NMOS transistors. The compressive stress in the gate conductors of the NMOS transistors causes tensile stress in channel regions of the NMOS transistors.

In another embodiment, the invention again forms N-type metal oxide semiconductor (NMOS) transistors and P-type metal oxide semiconductor (PMOS) transistors on a substrate. However, in this embodiment, the invention first protects the NMOS transistors and then implants ions into the PMOS transistors to render amorphous the PMOS transistors. Then, the invention performs an annealing process to crystallize the PMOS transistors. After this, the invention protects the PMOS transistors with a mask before implanting ions into the NMOS transistors. Then both the NMOS transistors and the PMOS transistors are covered with a rigid layer, and the NMOS transistors and the PMOS transistors are heated. During this heating process, the rigid layer prevents the gate of the NMOS transistors from expanding which creates compressive stress within the gates of the NMOS transistors. Again, this compressive stress within the gates of the NMOS

transistors causes tensile stress within the channel regions of the NMOS transistors. After this, the rigid layer is removed and the remaining structures of the transistor are completed.

By creating compressive stress in the gates and tensile stress in the channel regions of the NMOS transistors (NFETs), without creating stress in the gates or channel regions of the PMOS transistors (PFETs), the invention improves performance of the NFETs without degrading performance of the PFETs.

These and other aspects of invention are described in further detail below.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-9 are schematic cross-sectional diagrams illustrating different stages in a process of manufacturing a field effect transistor according to a first embodiment.

Figures 10-16 are schematic cross-sectional diagrams illustrating different stages in a process of manufacturing a field effect transistor according to a second embodiment.

Figure 17 is a flow diagram illustrating a preferred method of the invention.

Figure 18 is a flow diagram illustrating a preferred method of the invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention and the various features and advantageous details thereof are explained more fully with reference to the non-limiting embodiments that are illustrated in the accompanying drawings and detailed in the following description. It should be noted that the features illustrated in the drawings are not necessarily drawn to scale. Descriptions of well-known components and processing techniques are omitted so as to not unnecessarily obscure the present invention. The examples used herein are intended merely to facilitate an understanding of ways in which the invention may be practiced and to further enable those of skill in the art to practice the invention. Accordingly, the examples should not be construed as limiting the scope of the invention.

As mentioned above, NMOS performance is improved when the channel region is placed under tensile stress and performance is degraded when the stress is compressive; however, PMOS device performance will be degraded by a tensile stress applied along the channel direction. Therefore, the invention provides a manufacturing method that only creates tensile stress in the NMOS devices without creating tensile stress in PMOS devices. More specifically, the invention generates compressive stress in the transistor gate, and tensile stress is induced in the channel due to the proximity between the gate and channel.

A transistor gate stack generally comprises a gate polysilicon and spacers (of oxide and nitride). When the transistor is annealed at an elevated temperature, the polysilicon grains may grow (or become crystalline if the polysilicon is amorphous before anneal) resulting in a volume increase in the gate conductor size. However, if the gate stack is covered with a rigid, hard material during the annealing process, the size of the gate cannot increase and compressive stress is created within the gate.

This compressive stress is generated due to different thermal expansion coefficients among the materials in the gate stack in addition to the

volume change due to crystallization of polysilicon as mentioned above. As discussed in greater detail below, the invention covers the gate stack with a hard layer (such as a silicon nitride layer) prior to annealing the gate stack. This causes compressive stress within the gate stack. The invention uses hard materials such as silicon nitride, silicon carbide, etc. to cover the gate during the annealing process. The invention advantageously uses such rigid films, as compared to, for example, covering the gate stack with an oxide. When oxides and other films that are not as rigid are used, such films may deform and change shape slightly during the annealing process, yielding to the stress in the gate, and not effectively creating stress within the gate stack. When the transistor gate is annealed and covered by a Si_3N_4 layer, the polysilicon volume change and spacer deformation are limited by the Si_3N_4 layer, inducing high stress in the gate stack after anneal. The stress remains in the gate and channel even after Si_3N_4 is removed.

Referring now to the drawings, Figures 1-9 are schematic cross-sectional diagrams illustrating different stages in a process of manufacturing a field effect transistor according to a first embodiment and Figures 10-16 are schematic cross-sectional diagrams illustrating different stages in a process of manufacturing a field effect transistor according to a second embodiment. Many of the processes and materials used to form the transistors that are covered with the inventive rigid layer are well-known to those ordinary skill in the art (for example, see US patent number 5,670,388 which is incorporated herein by reference).

In Fig. 1, polysilicon 10 is deposited on a wafer 12 (such as a silicon wafer) after a shallow trench isolation (STI) region 14 and gate oxide 16 are formed using well-known processing techniques. The polysilicon 10 is patterned to form gate stacks 20, 22 as shown in Fig. 2 using, for example, well-known masking and etching processes. In this example, the gate stack 20 on the left will be used in one type of transistor, such as a P-type transistor

(PFET) while the gate stack 22 on the right will be used in an opposite type of transistor such as an N-type transistor (NFET). In Fig. 3, a sidewall spacer 30 is formed on gate stack 20 and extension/halo implants are made for both NFET and PFET.

In Fig. 4, another sidewall spacer 40 is formed and source/drain ion implantations 42 are made. Note that the gate polysilicon 20, 22 (as well as source/drain regions 42) is rendered amorphous as represented by the different shading in the drawings due to the ion bombardment of the source/drain ion implantation. In this process, crystalline or polycrystalline silicon becomes amorphous silicon that will expand when heated.

In Fig. 5 a rigid (hard) film 50 such as silicon nitride, silicon carbide, etc. is deposited over the wafer 12 using conventional deposition process, such as chemical vapor deposition (CVD) or plasma enhanced CVD process or other suitable process. Prior to forming the rigid film 50, an optional etch stop layer 52 such as SiO_2, etc. can be grown or deposited. The material used for the rigid film 50 can comprise any appropriate material that does not substantially deform when the gate conductor 22 tries to expand during the annealing process that is described below. The thickness of the rigid film 50 and the optional etch stop layer 52 can be any thickness that is appropriate, depending upon the manufacturing process being utilized and the specific design of the transistor involved, so long as the rigid film 50 is thick enough to prevent the gate conductor 22 from expanding significantly during the annealing process. For example, the thickness of rigid layer 50 may be in the range of 500Å to 1500Å and the thickness of the etch stop layer may be in the range of 20Å to 50Å.

In Fig. 6 the rigid film 50 is patterned using well known masking and material removal processes leaving rigid film 50 to cover the NFETs only. In Fig. 7, a thermal anneal is performed to activate the implanted dopants and to

crystallize the amorphous silicon. The anneal temperature may be, for example, in the range of 700C to 1100C. Note NFET gate 22 becomes stressed because it is encapsulated by rigid layer 50 and cannot significantly expand. As amorphous silicon becomes crystalline, its volume expands. However, because the rigid layer 50 prevents the exterior of the NFET gate 22 from increasing in size, stress builds up within the NFET gate 22. This stress remains within the NFET gate 22 even after the rigid layer 50 is removed because the outer portions of the gate polysilicon 22 will retain their shape and size once the temperature lowers below the annealing temperature. This compressive stress within the NFET gate 22 causes tensile stress in NFET channel region 70. Tensile stress along the channel direction enhances electron mobility and hence improves NFET device performance. The same stress will degrade hole mobility and hence degrade PFET performance. Therefore, in Figure 6, the rigid layer 50 was removed from the PFET region before the annealing process, to allow the PFET 20 to freely expand.

In Fig. 8, and the remaining portions of the rigid layer 50 are removed again using well-known material removal processes. If the etch stop layer 52 was utilized, it can now be removed using, for example a cleaning process that utilizes HF containing chemicals. As mentioned above, they compressive stress remains within the gate 22 and therefore tensile stress remains in the channel 70 even after the rigid film 50 is removed. In Fig. 9, silicide regions 65 are formed on top of gates 20, 22 and on the source/drain regions. Self-aligned silicide (Salicides) can be formed at 300C to 700C using Ni or Co. Non-reacted metal is then stripped away from the wafer. Inter-layer dielectrics (ILD) and interconnects are then formed using well-known processing and materials.

By creating compressive stress in the gates and tensile stress in the channel regions of the NMOS transistors (NFETs), without creating stress in

the gates or channel regions of the PMOS transistors (PFETs), the invention improves performance of the NFETs without degrading performance of the PFETs.

Another embodiment is shown in Figures 10-16. More specifically, in Fig. 10, a mask 102, such as a photoresist mask, is patterned and the PFET source/drain implantations 100 are performed while the NFET is covered with photoresist 102. As mentioned, during the implant process, PFET gate 20 is rendered amorphous. Then, in Fig. 11, the mask 102 is stripped and a heating process, such as a rapid thermal anneal (RTA) is performed to crystallize the PFET amorphous silicon 20. This crystallization process of the gate 20 will cause the gate 20 to expand and, because there is no rigid layer over the gate 20, this expansion does not create compressive stress within the gate 20.

In Fig. 12, another photoresist mask 122 is patterned to cover the PFETs and a second ion implantation process is performed on the exposed NFETs to form the source/drain regions 120 and to render amorphous the gate conductor 22. Then, in Fig. 13, the photoresist 122 is again stripped. Note that because the PFETs were protected by a mask 122, only the NFETs have amorphous silicon regions remaining.

In Fig. 14, the rigid layer 50 and the optional oxide layer 52 are formed as discussed above. Then, in Fig. 15, a thermal anneal is performed to activate implanted dopants and to crystallize amorphous silicon. Again, the anneal temperature may be in the range of, for example, 700C to 1100C. Note that only the NFET gate poly 22 becomes compressively stressed because the PFET gate 20 did not contain amorphous state material that was within the gate 22. Then, in Fig. 16, the rigid film 50 and optional oxide film 52 are removed and the wafer is ready for salicidation, as discussed above.

Figure 17 shows the first embodiment in flow chart form. More specifically, in item 170 the method forms different (e.g., opposite) types of transistors such as N-type metal oxide semiconductor (NMOS) transistors and P-type metal oxide semiconductor (PMOS) transistors (first and second type transistors) on a substrate. In item 172, the invention forms an optional oxide layer on the NMOS transistors and the PMOS transistors and then covers the NMOS transistors and the PMOS transistors with a rigid material such as a silicon nitride layer in item 174. Following this, the invention patterns portions of the rigid layer in item 176, such that the rigid layer remains only over the NMOS transistors. Next, the invention heats the NMOS transistors in item 178 and then removes the remaining portions of the rigid layer in item 180.

In the second embodiment shown in flow chart form in Figure 18, the invention again forms N-type metal oxide semiconductor (NMOS) transistors and P-type metal oxide semiconductor (PMOS) transistors on a substrate in item 190. However, in this embodiment, the invention first protects the NMOS transistors in item 192 and then implants ions into the PMOS transistors to render amorphous the PMOS transistors in item 194. Then, the invention performs an annealing process to crystallize the PMOS transistors in item 196. After this, the invention protects the PMOS transistors with a mask in item 198 before implanting ions into the NMOS transistors in item 200. Then, both the NMOS transistors and the PMOS transistors are covered with a rigid layer in item 202 and the NMOS transistors and the PMOS transistors are heated in item 204. During this heating process, the rigid layer prevents the gate of the NMOS transistors from expanding which creates compressive stress within the gates of the NMOS transistors. Again, this compressive stress within the gates of the NMOS transistors causes tensile stress within the channel regions of the NMOS transistors. After this, the rigid layer is removed in item 206 and the remaining structures of the transistor are completed in item 208.

The heating process creates compressive stress in the gate, which in turn causes tensile stress in channel regions of transistors that were covered by the silicon nitride layer. Thus, the heating process creates tensile stress in channel regions of the NMOS transistors without causing tensile stress in channel regions of the PMOS transistors. More specifically, during the heating process, volume expansion of gate conductors of the NMOS transistors is restricted, resulting in compressive stress in the gate conductors of the NMOS transistors. The compressive stress in the gate conductors of the NMOS transistors causes tensile stress in channel regions of the NMOS transistors. By creating compressive stress in the gates and tensile stress in the channel regions of the NMOS transistors (NFETs), without creating stress in the gates or channel regions of the PMOS transistors (PFETs), the invention improves performance of the NFETs without degrading performance of the PFETs.

While the invention has been described in terms of preferred embodiments, those skilled in the art will recognize that the invention can be practiced with modification within the spirit and scope of the appended claims.

CLAIMS

What is claimed is:

1. A method of manufacturing a transistor, said method comprising:

forming a first-type transistor on a substrate (12), said transistor having a gate conductor (22);

covering said transistor with a rigid layer (50); and

heating (178) said transistor to create tensile stress (70) in said transistor.

2. The method according to claim 1, further comprising forming an oxide layer (52) on said transistor prior to forming said rigid layer (50).

3. The method according to claim 1, wherein said rigid layer (50) comprises at least one of silicon nitride and silicon carbide.

4. The method according to claim 1, further comprising implanting ions into a gate (22) of said first-type transistor before covering said first-type transistor with said rigid layer (50).

5. The method according to claim 1, wherein said substrate further includes other transistors that are not covered by said rigid layer (50), and said heating process (178) creates tensile stress (70) in channel regions of said first-type transistor without causing tensile stress in channel regions of other transistors that are not covered by said rigid layer (50).

6. The method according to claim 1, wherein during said heating process (178), volume expansion of gate conductors (22) of first-type transistor is restricted, resulting in compressive stress in said gate conductors (22) of said first-type transistor.

7. The method according to claim 6, wherein said compressive stress in said gate conductors (22) of said first-type transistor causes tensile stress (70) in channel regions of said first-type transistor.

8. A method of manufacturing complementary transistors, said method comprising:

forming first-type transistors having gate conductors (22) and second-type transistors having gate conductors (20) on a substrate (12);

covering said first-type transistors and said second-type transistors with a rigid layer (50);

patterning portions of said rigid layer (50), such that said rigid layer (50) remains only over said first-type transistors; and

heating (178) said first-type transistors.

9. The method according to claim 8, further comprising forming an oxide layer (52) on said first-type transistors and said second-type transistors prior to forming said rigid layer (50) on said first-type transistors and said second-type transistors.

10. The method according to claim 8, wherein said rigid layer (50) comprises at least one of silicon nitride and silicon carbide.

11. The method according to claim 8, wherein said heating process (178) creates tensile stress (70) in channel regions of said first-type transistors covered by said rigid layer (50).

12. The method according to claim 8, wherein said heating process (178) creates tensile stress (70) in channel regions of said first-type transistors without causing tensile stress in channel regions of said second-type transistors.

13. The method according to claim 8, wherein during said heating process (178), volume expansion of gate conductors (22) of said first-type transistors is restricted, resulting in compressive stress in said gate conductors (22) of said first-type transistors.

14. The method according to claim 13, wherein said compressive stress in said gate conductors (22) of said first-type transistors causes tensile stress (70) in channel regions of said first-type transistors.

15. The method of according to any of claims 8 through 14 wherein said first-type transistors are N-type metal oxide semiconductor (NMOS) transistors and said second-type transistors are P-type metal oxide semiconductor (PMOS) transistors.

16. A method of manufacturing complementary transistors, said method comprising:

forming first-type transistors having gate conductors (22) and second-type transistors having gate conductors (20) on a substrate (12);

protecting said second-type transistors with a mask (122);

implanting ions (200) into said first-type transistors;

covering said first-type transistors and said second-type transistors with a rigid layer (50); and

heating (204) said first-type transistors and said second-type transistors.

17. The method according to claim 16, further comprising forming an oxide layer (52) on said first-type transistors and said second-type transistors prior to forming said rigid layer (50) on said first-type transistors and said second-type transistors.

18. The method according to claim 16, wherein said rigid layer (50) comprises at least one of silicon nitride and silicon carbide.

19. The method according to claim 16, wherein said heating process (204) creates tensile stress in channel regions of said first-type transistors.

20. The method according to claim 16, wherein said heating process (204) creates tensile stress in channel regions of said first-type transistors without causing tensile stress in channel regions of said second-type transistors.

21. The method according to claim 16, wherein during said heating process (204), volume expansion of gate conductors (22) of said first-type transistors is restricted, resulting in compressive stress in said gate conductors (22) of said first-type transistors.

22. The method according to claim 21, wherein said compressive stress in said gate conductors (22) of said first-type transistors causes tensile stress in channel regions of said first-type transistors.

23. The method of according to any of claims 16 through 22 wherein said first-type transistors are N-type metal oxide semiconductor (NMOS) transistors and said second-type transistors are P-type metal oxide semiconductor (PMOS) transistors.

Fig.1

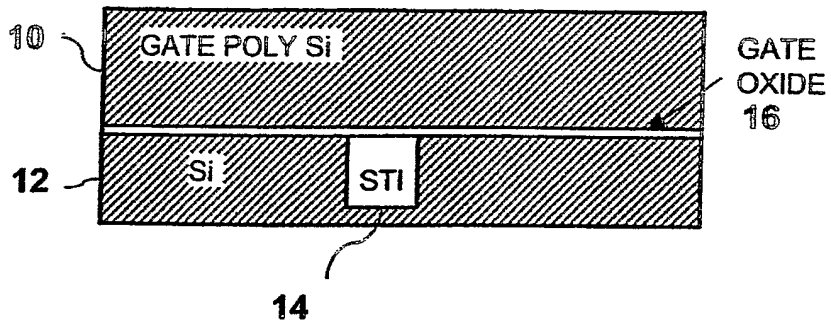


Fig.2

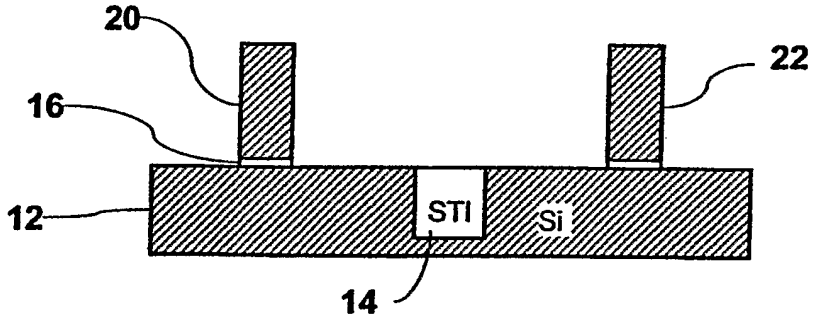


Fig.3

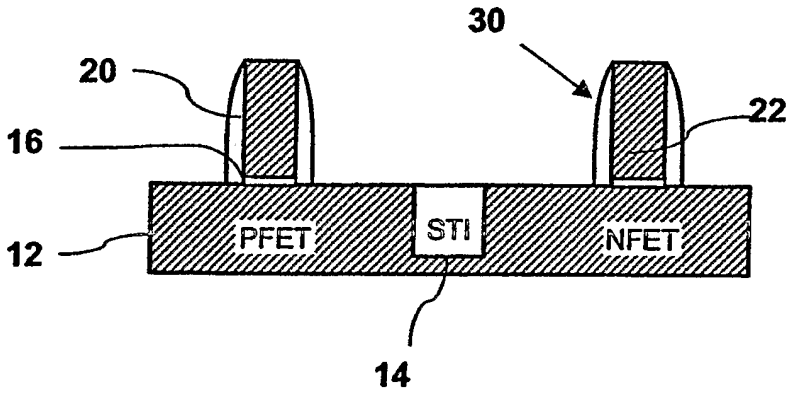
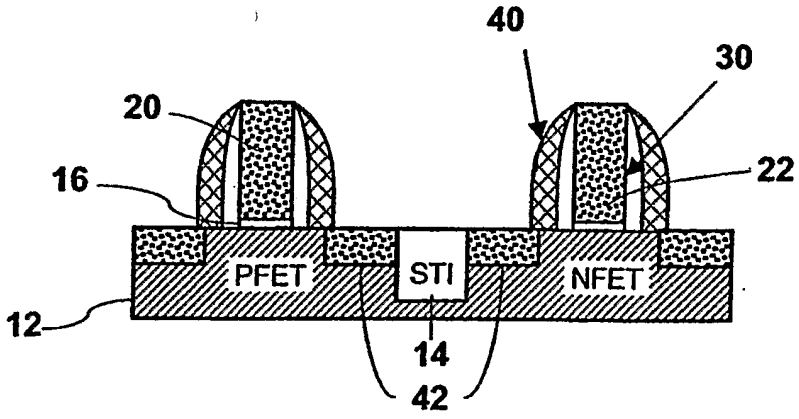


Fig.4



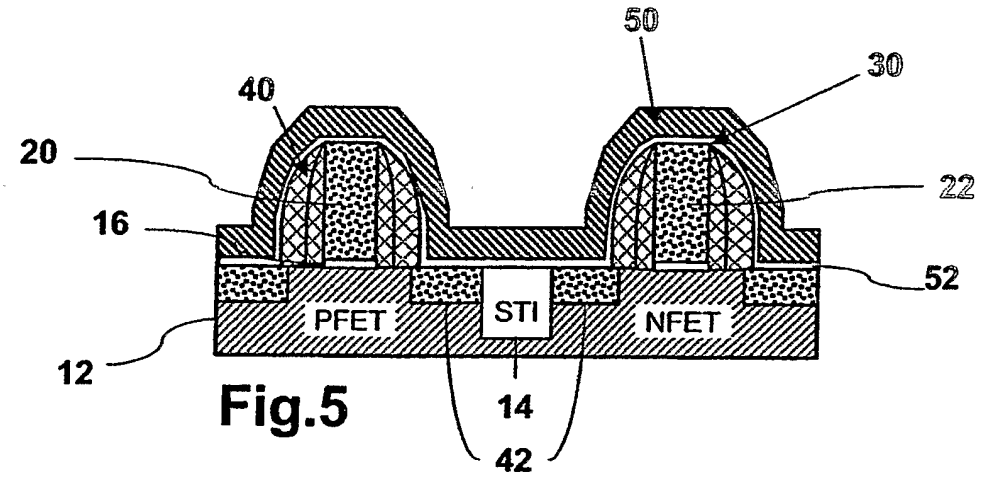


Fig.5

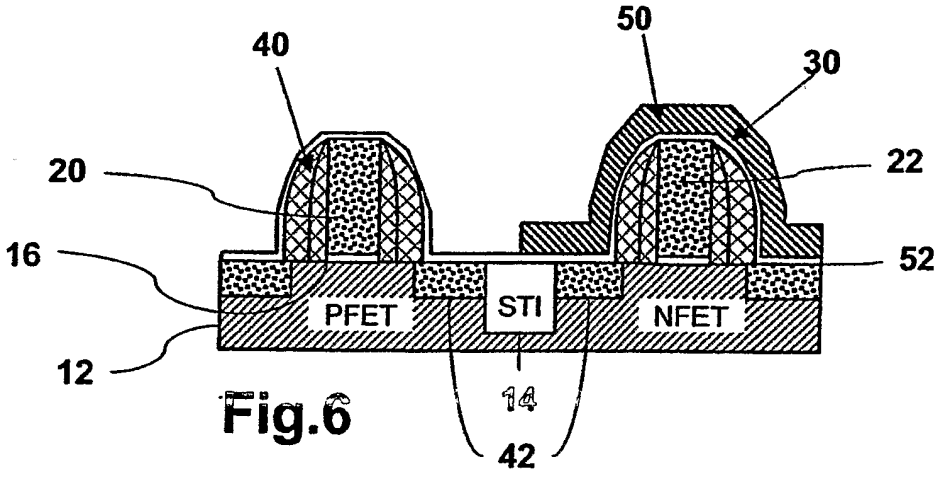


Fig.6

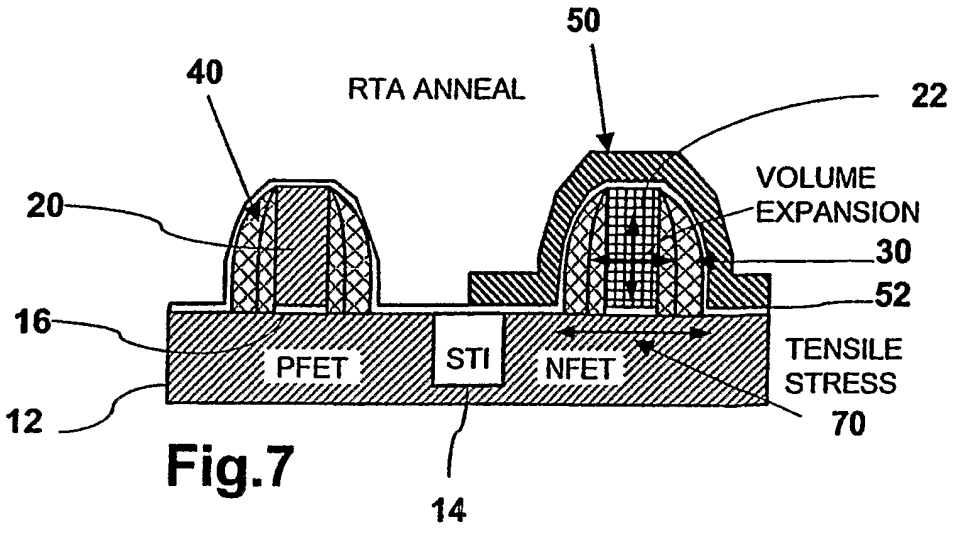


Fig.7

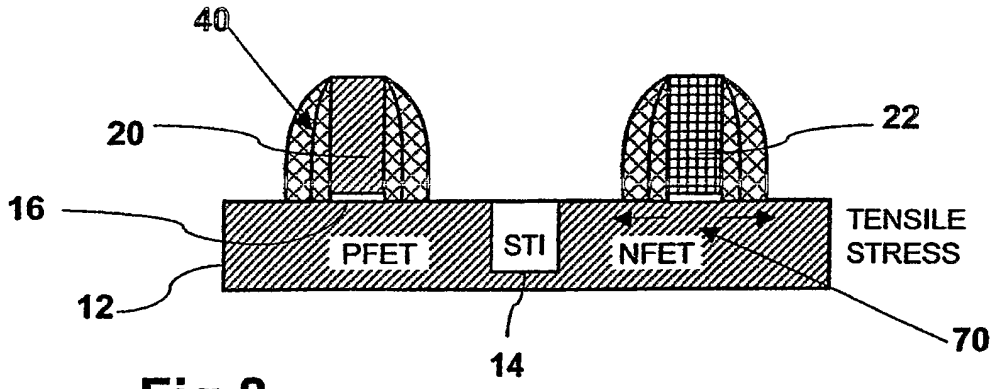


Fig.8

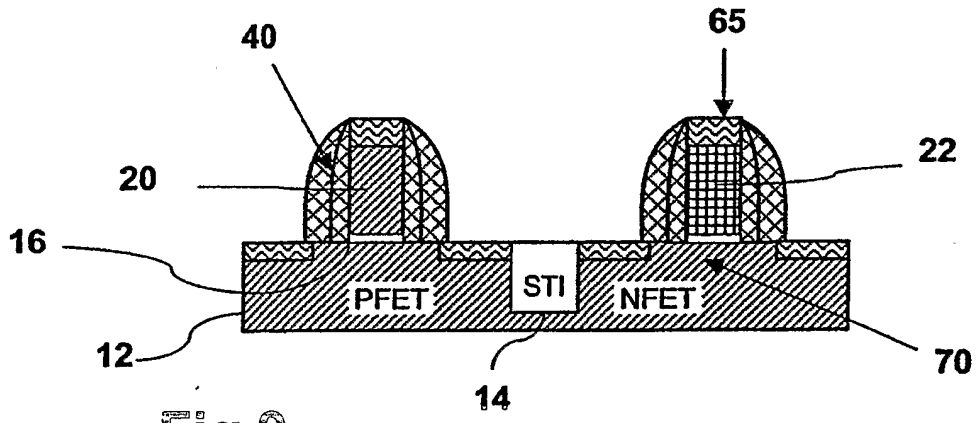


Fig.9

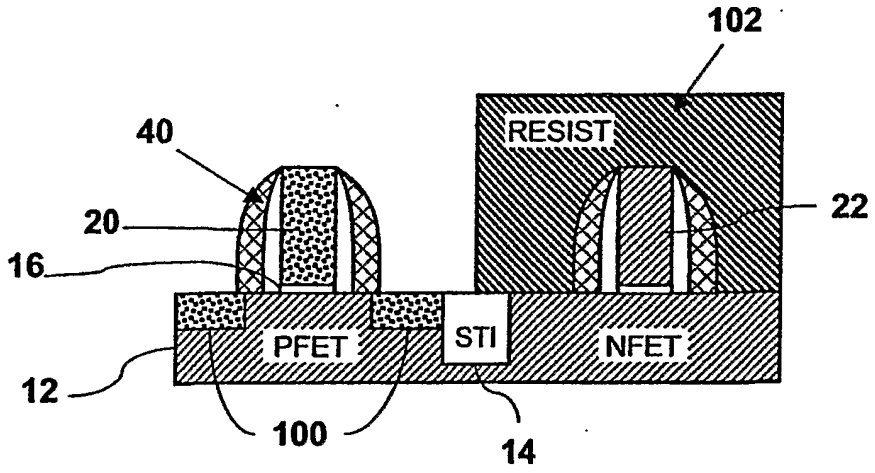


Fig.10

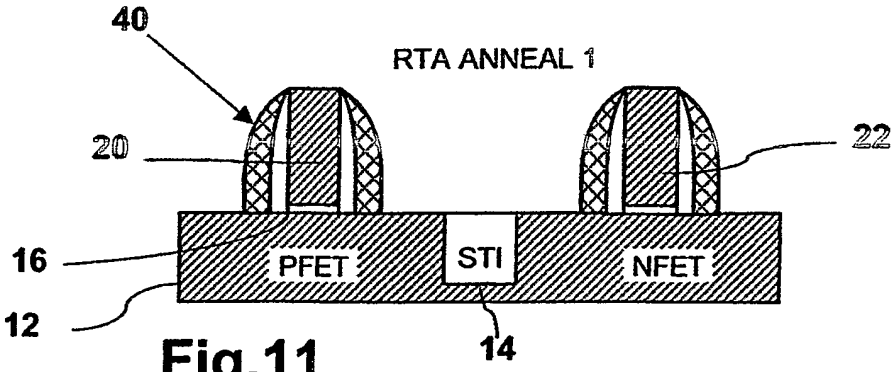


Fig.11.

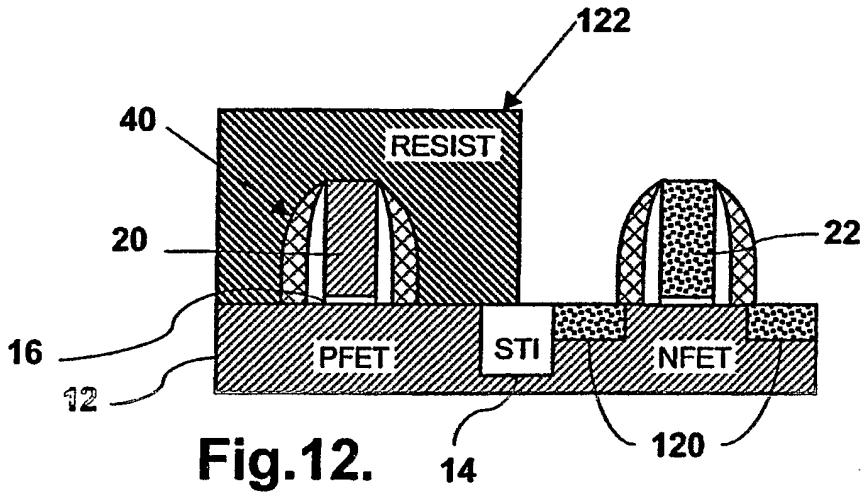


Fig.12.

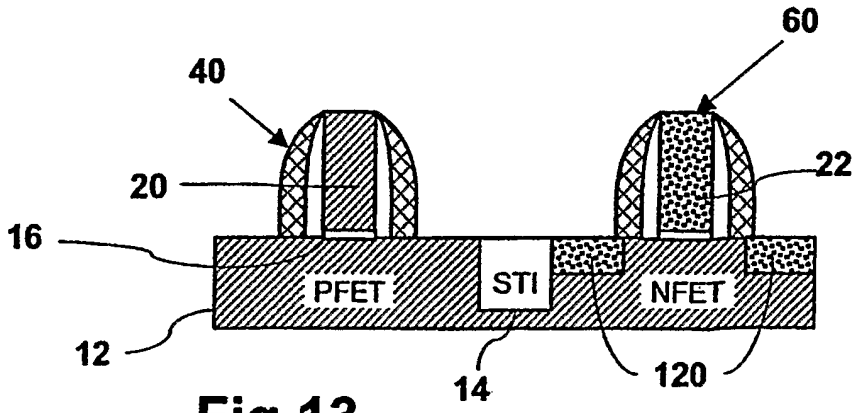


Fig.13.

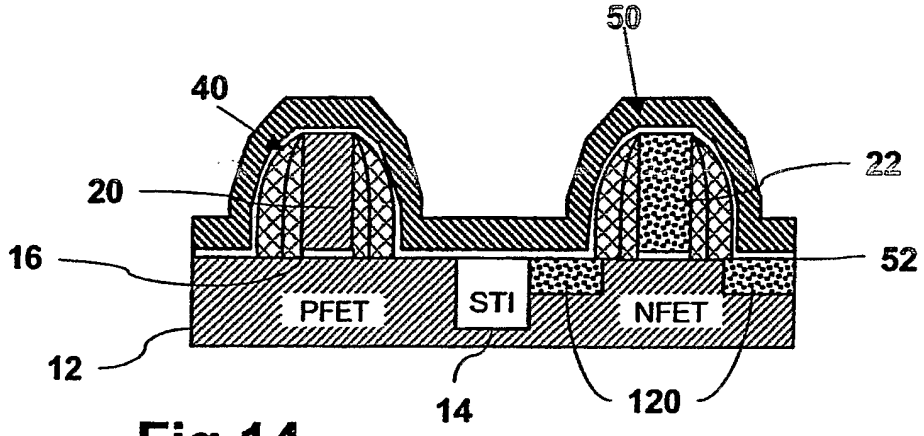


Fig.14

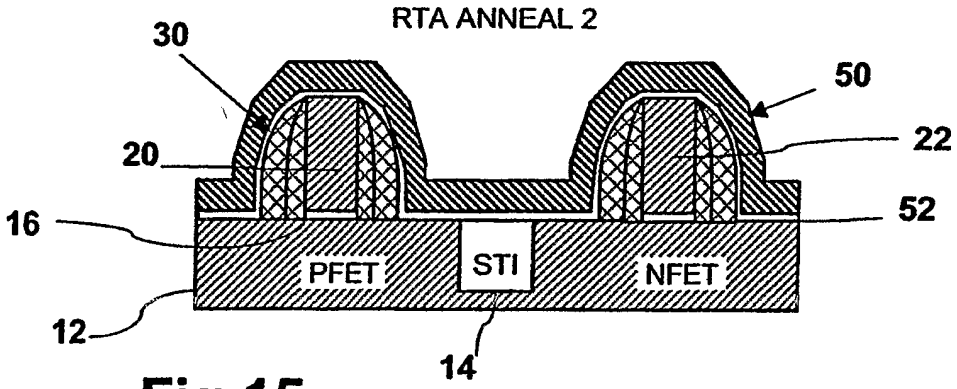


Fig.15

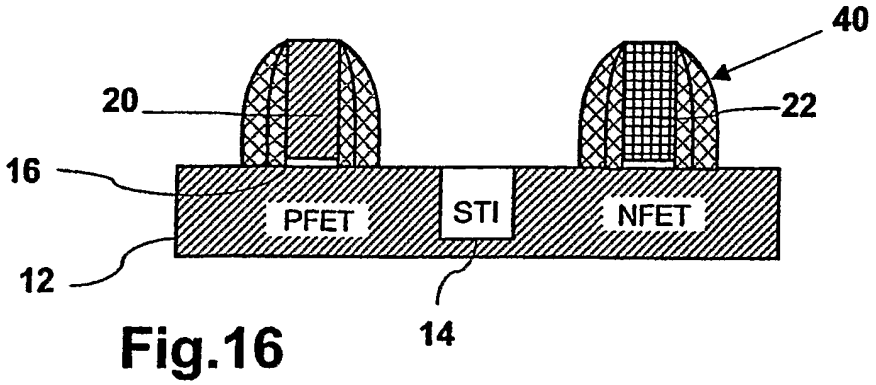


Fig.16

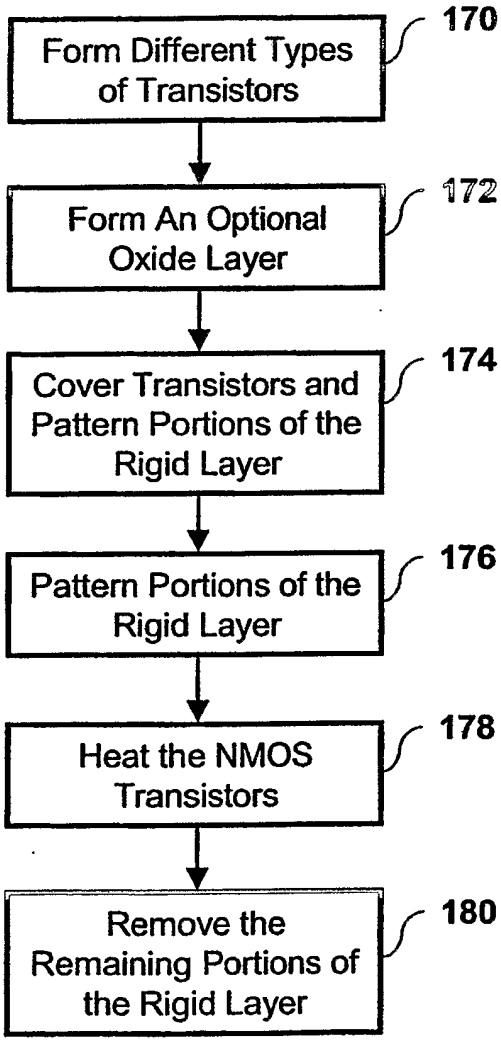

Form Different Types of Transistors
170
Form An Optional Oxide Layer
172
Cover Transistors and Pattern Portions of the Rigid Layer
174
Pattern Portions of the Rigid Layer
176
Heat the NMOS Transistors
178
Remove the Remaining Portions of the Rigid Layer
180

Fig. 17

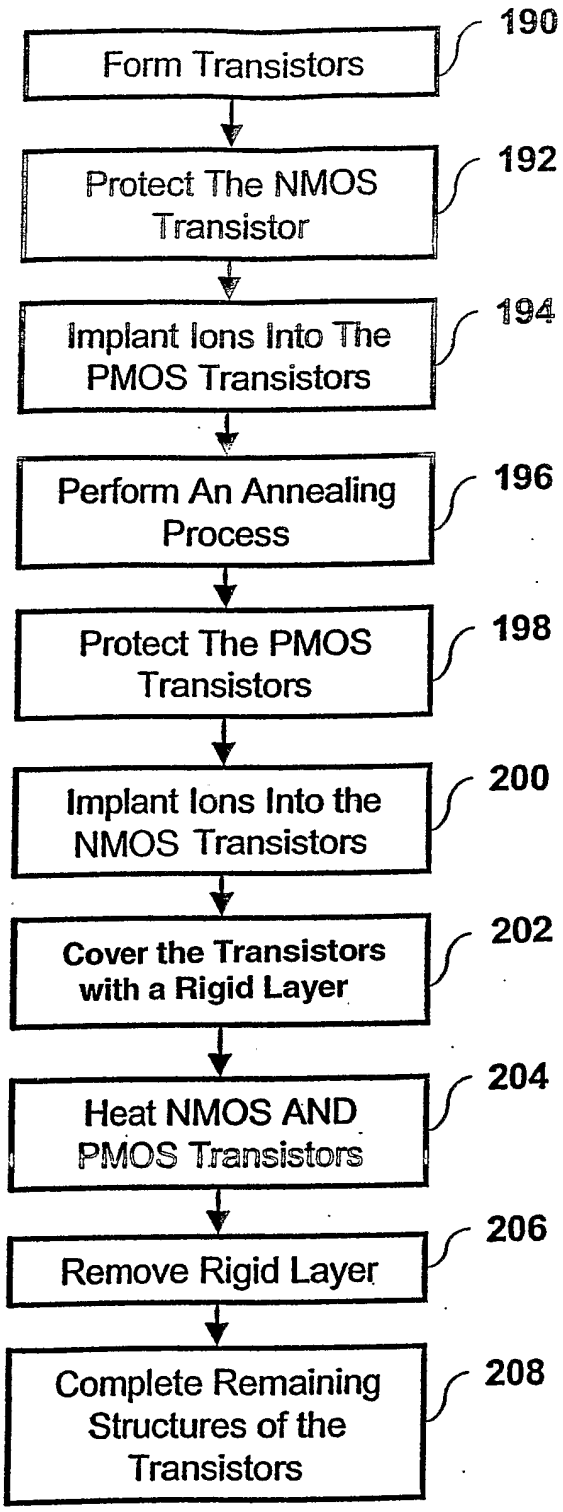

Form Transistors
190
Protect The NMOS Transistor
192
Implant Ions Into The PMOS Transistors
194
Perform An Annealing Process
196
Protect The PMOS Transistors
198
Implant Ions Into the NMOS Transistors
200
Cover the Transistors with a Rigid Layer
202
Heat NMOS AND PMOS Transistors
204
Remove Rigid Layer
206
Complete Remaining Structures of the Transistors
208

Fig. 18