US008436404B2

(12) **United States Patent**
Bohr et al.

(10) **Patent No.: US 8,436,404 B2**
(45) **Date of Patent: May 7, 2013**

(54) **SELF-ALIGNED CONTACTS**

(75) Inventors: **Mark T. Bohr**, Aloha, OR (US); **Tahir Ghani**, Portland, OR (US); **Nadia M. Rahhal-Orabi**, Hillsboro, OR (US); **Subhash M. Joshi**, Hillsboro, OR (US); **Joseph M. Steigerwald**, Forest Grove, OR (US); **Jason W. Klaus**, Portland, OR (US); **Jack Hwang**, Portland, OR (US); **Ryan Mackiewicz**, Beaverton, OR (US)

(73) Assignee: **Intel Corporation**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 71 days.

(21) Appl. No.: **12/655,408**

(22) Filed: **Dec. 30, 2009**

(65) **Prior Publication Data**

US 2011/0156107 A1 Jun. 30, 2011

(51) **Int. Cl.**
H01L 29/78 (2006.01)
H01L 21/336 (2006.01)

(52) **U.S. Cl.**
USPC **257/288**; 257/382; 257/E29.116; 257/E21.575; 438/299

(58) **Field of Classification Search** ..... None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,381,302	A *	1/1995	Sandhu et al.	361/305
5,385,866	A *	1/1995	Bartush	438/595
6,054,355	A *	4/2000	Inumiya et al.	438/296
6,184,129	B1 *	2/2001	Hwang et al.	438/653
6,207,514	B1 *	3/2001	Furukawa et al.	438/299
6,294,449	B1 *	9/2001	Wu et al.	438/597
6,310,367	B1 *	10/2001	Yagishita et al.	257/190
6,607,950	B2 *	8/2003	Henson et al.	438/197
6,774,441	B2 *	8/2004	Maki et al.	257/384
6,797,556	B2 *	9/2004	Murthy et al.	438/231
7,541,239	B2 *	6/2009	Curello et al.	438/230
7,544,594	B2 *	6/2009	Change	438/585
7,875,519	B2 *	1/2011	Rachmady et al.	438/299
2002/0192911	A1	12/2002	Parke	
2003/0141554	A1 *	7/2003	Uehara et al.	257/382

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2011/090571	A2	7/2011
WO	2011/090571	A3	11/2011

OTHER PUBLICATIONS

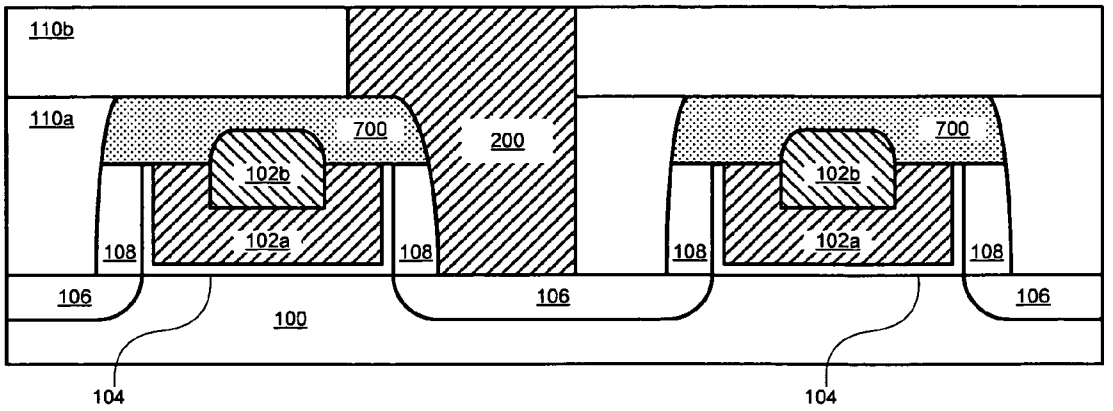
International Search Report and Written Opinion received for PCT Patent Application No. PCT/US2010/059302, mailed on Sep. 9, 2011, 11 pages.

Primary Examiner — Thomas L Dickey
Assistant Examiner — Joseph Schoenholtz
(74) *Attorney, Agent, or Firm* — Blakely, Sokoloff, Taylor & Zafman LLP

(57) **ABSTRACT**

A transistor comprises a substrate, a pair of spacers on the substrate, a gate dielectric layer on the substrate and between the pair of spacers, a gate electrode layer on the gate dielectric layer and between the pair of spacers, an insulating cap layer on the gate electrode layer and between the pair of spacers, and a pair of diffusion regions adjacent to the pair of spacers. The insulating cap layer forms an etch stop structure that is self aligned to the gate and prevents the contact etch from exposing the gate electrode, thereby preventing a short between the gate and contact. The insulator-cap layer enables self-aligned contacts, allowing initial patterning of wider contacts that are more robust to patterning limitations.

30 Claims, 17 Drawing Sheets



TSMC EXHIBIT 1009

U.S. PATENT DOCUMENTS

2003/0156380	A1	8/2003	Fazan et al.	
2005/0167763	A1*	8/2005	Kuan et al.	257/396
2005/0221548	A1*	10/2005	Doyle et al.	438/197
2006/0073699	A1*	4/2006	Lee et al.	438/639
2007/0104862	A1*	5/2007	Lai	427/97.4
2008/0166841	A1	7/2008	Hsieh	
2009/0026552	A1	1/2009	Zhang et al.	
2009/0242936	A1*	10/2009	Cheng et al.	257/190
2009/0311859	A1*	12/2009	Bonilla et al.	438/637

* cited by examiner

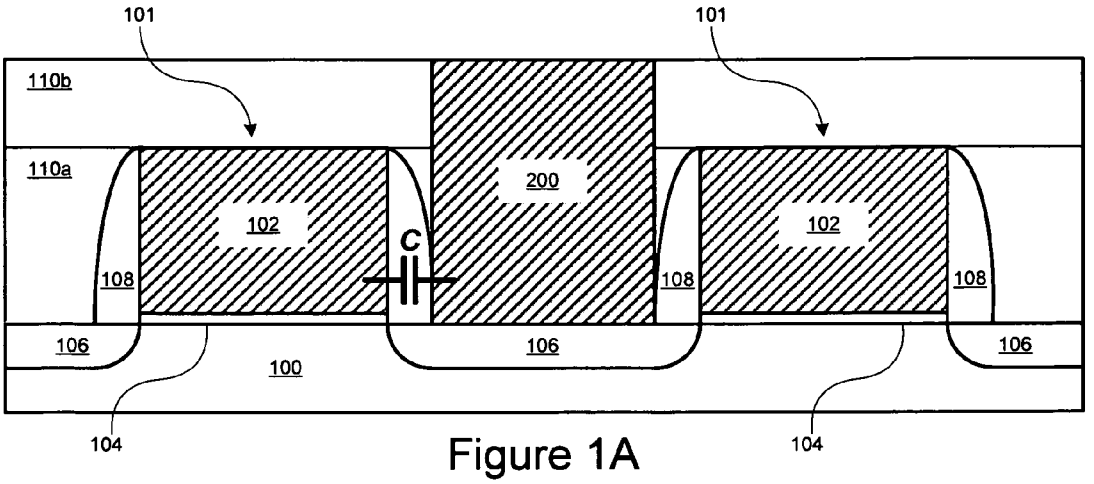


Figure 1A

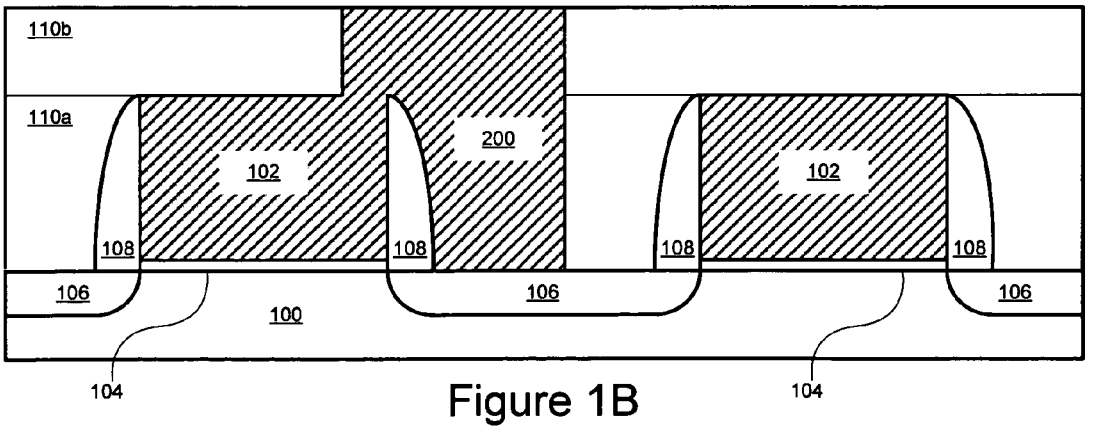


Figure 1B

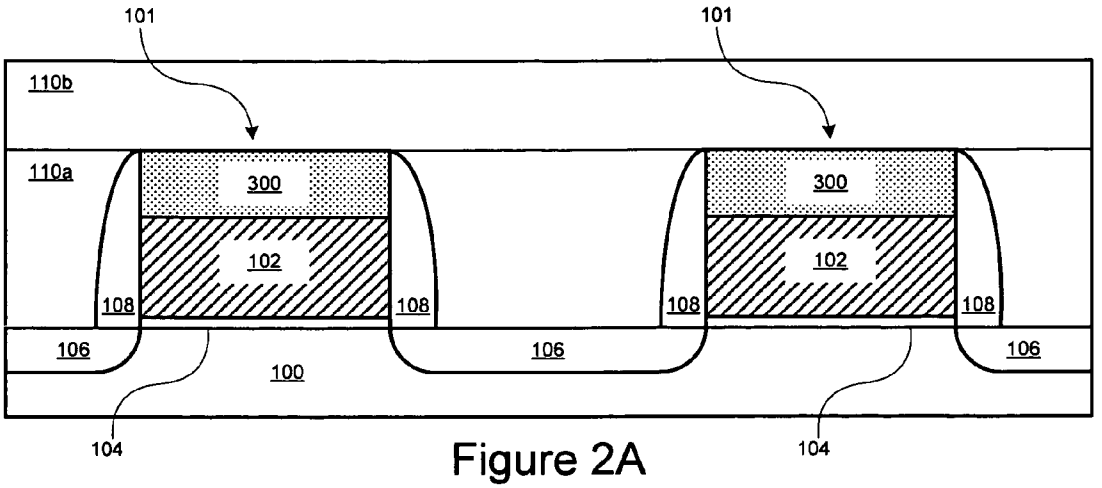


Figure 2A

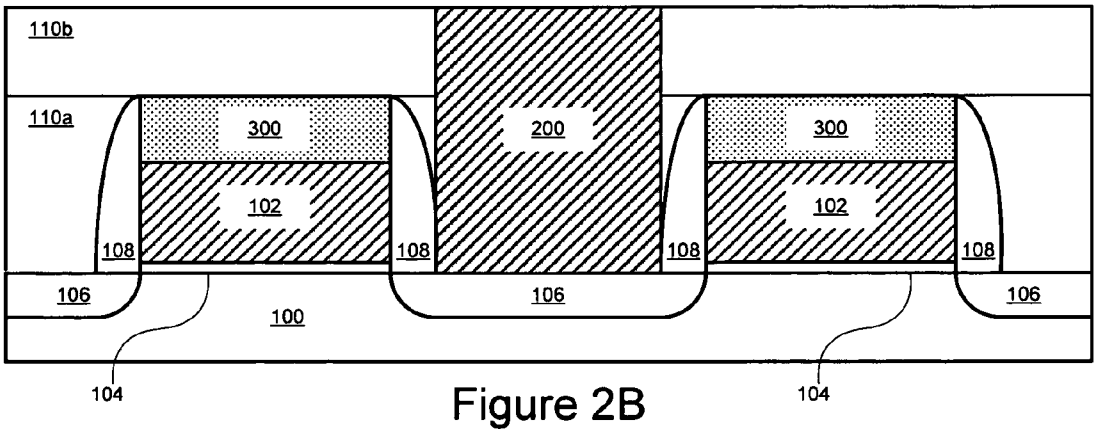


Figure 2B

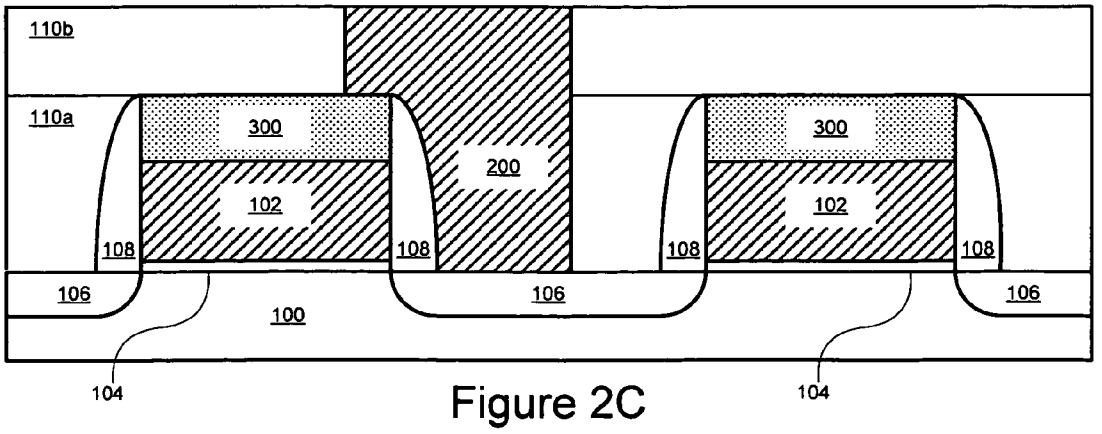


Figure 2C

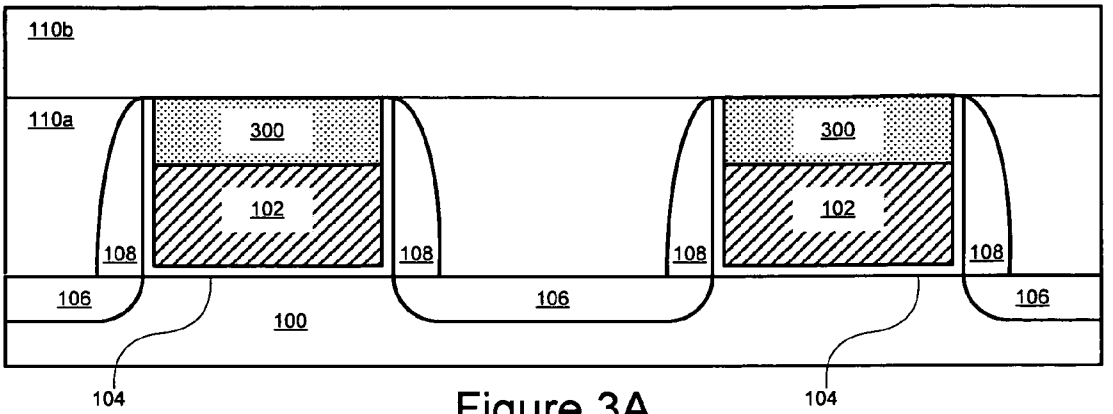


Figure 3A

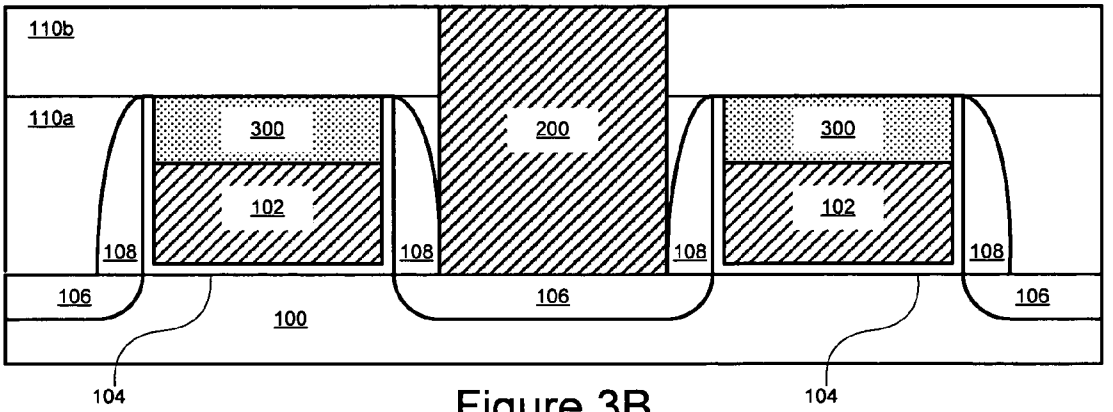


Figure 3B

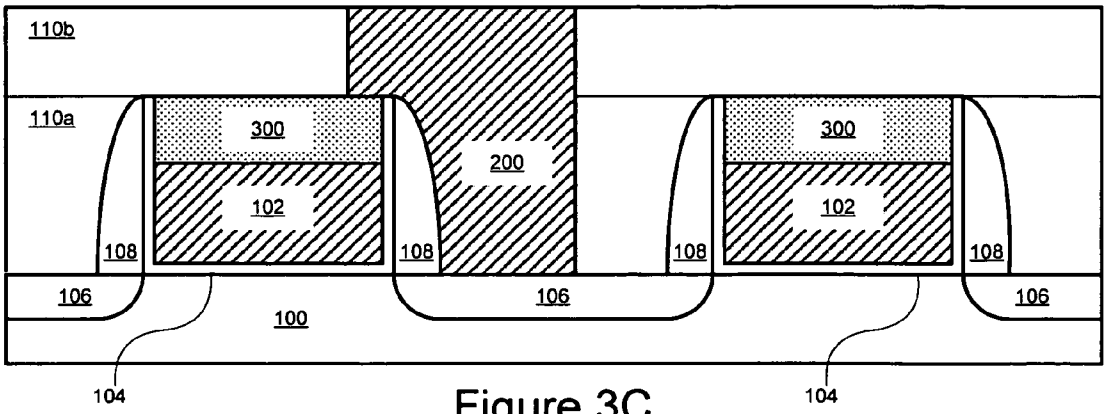


Figure 3C

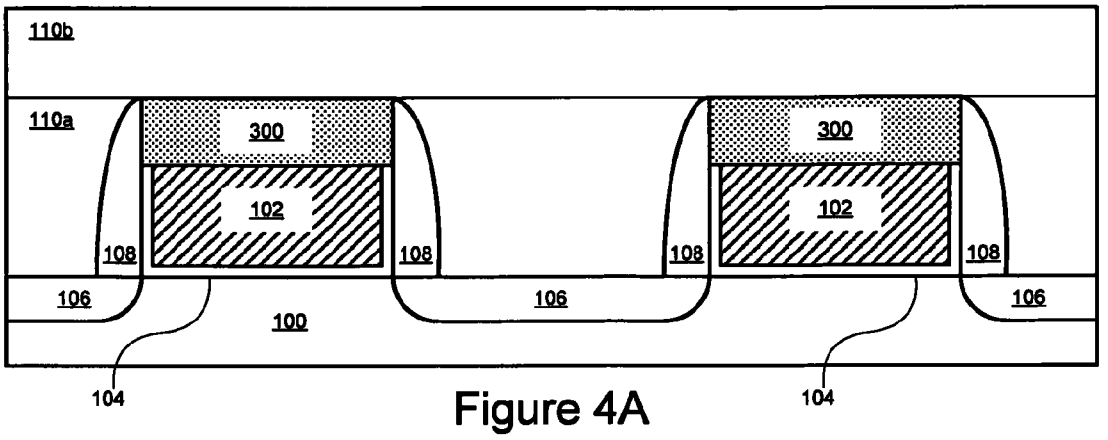


Figure 4A

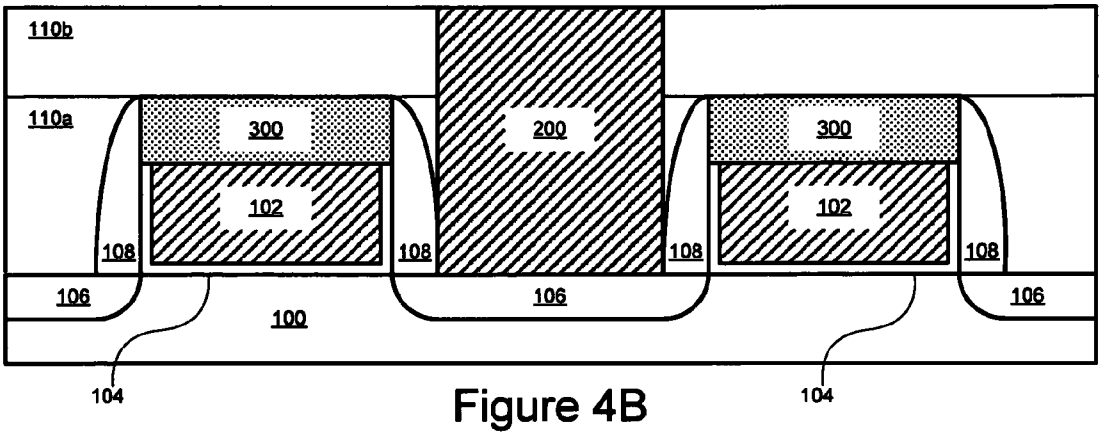


Figure 4B

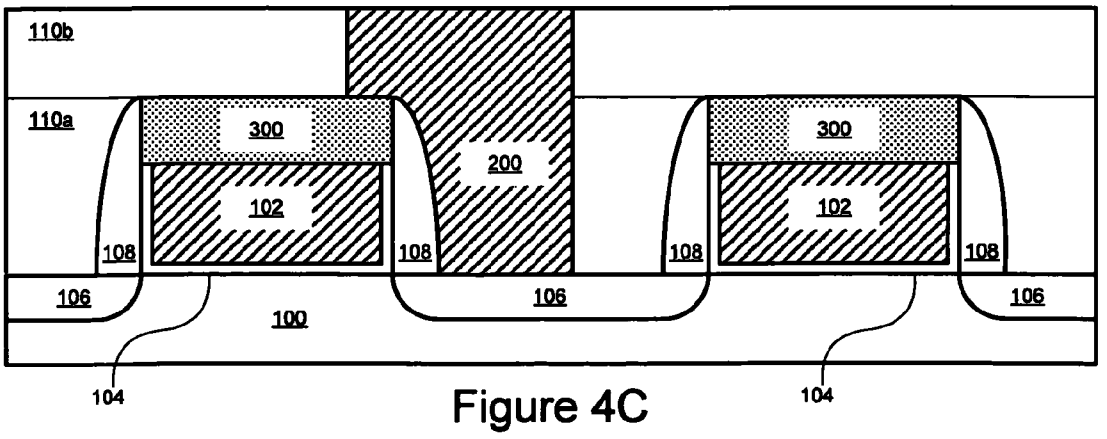


Figure 4C

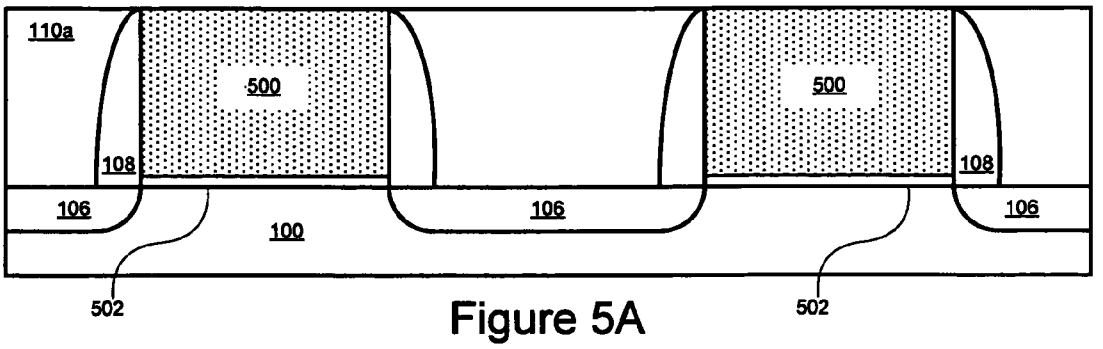


Figure 5A

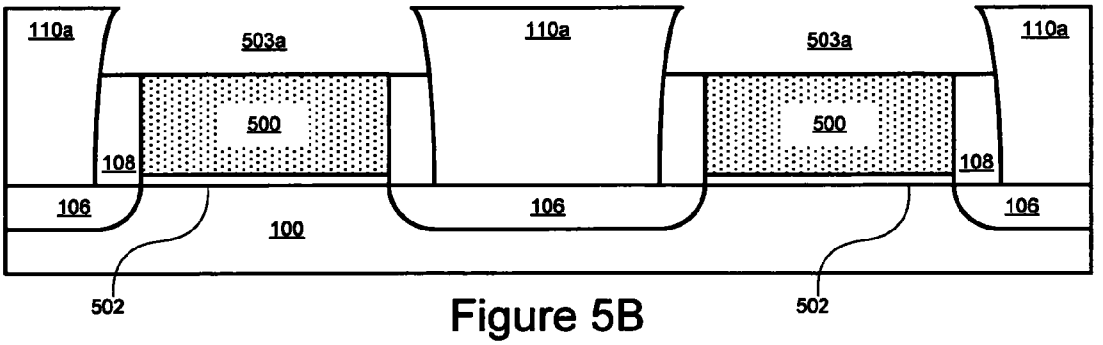


Figure 5B

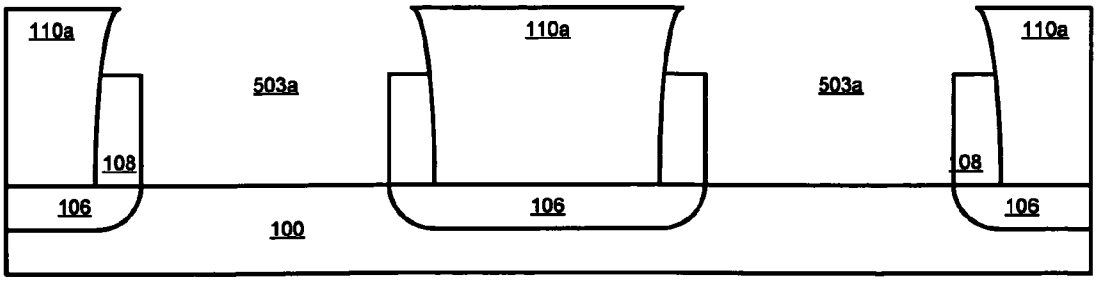


Figure 5C

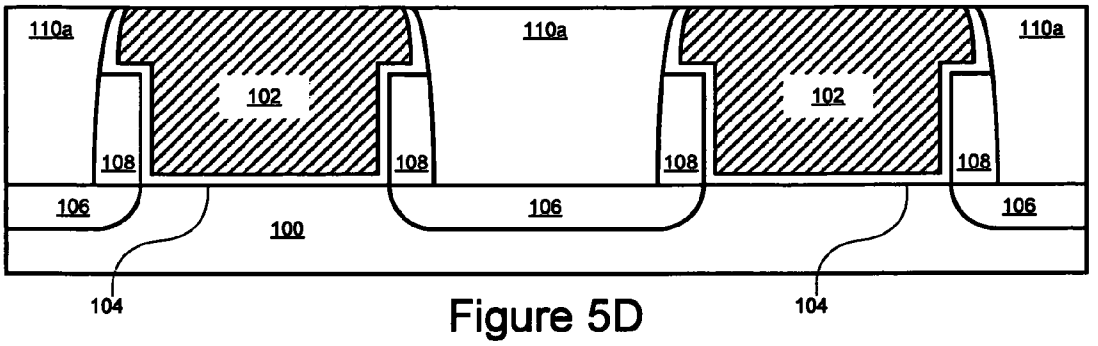


Figure 5D

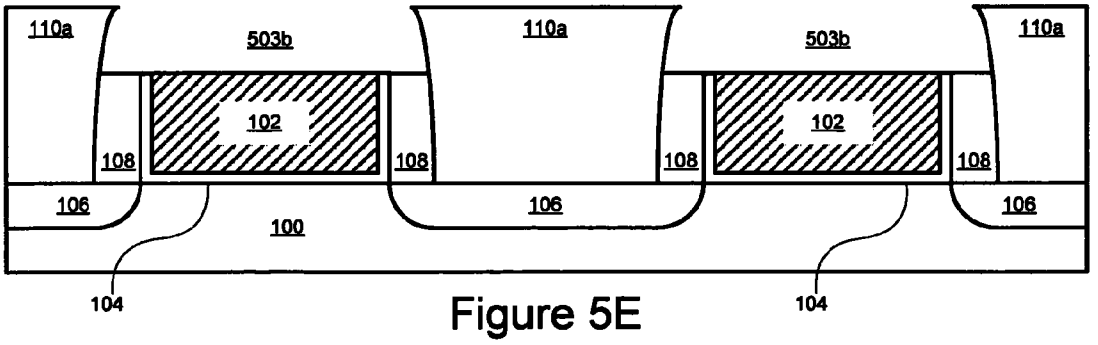


Figure 5E

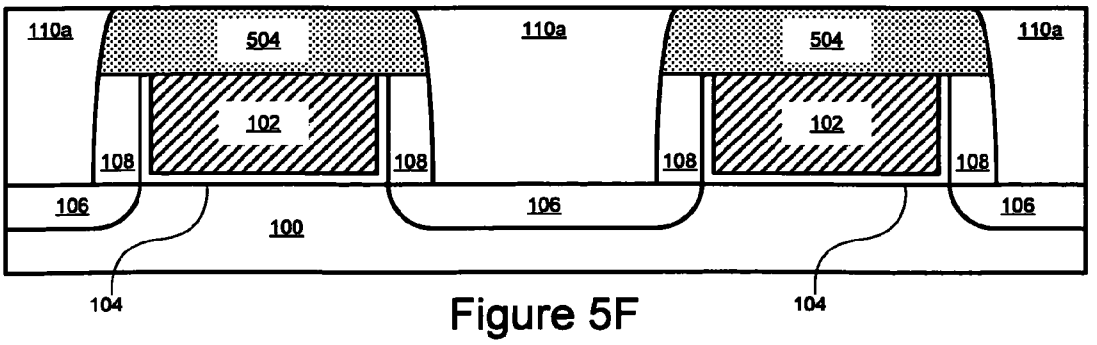


Figure 5F

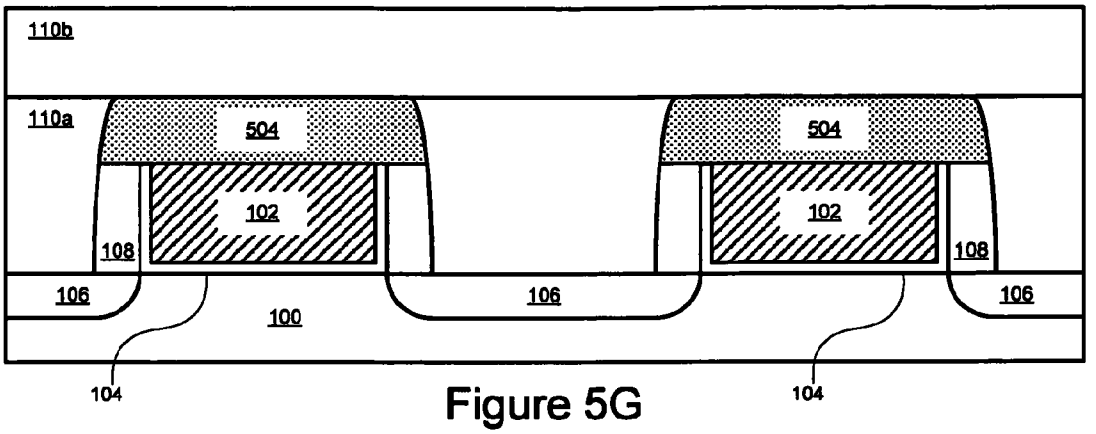


Figure 5G

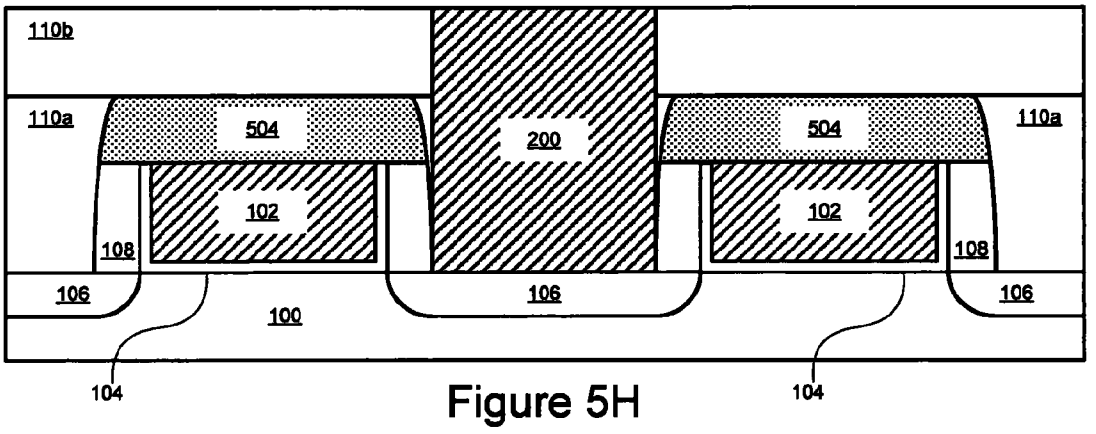


Figure 5H

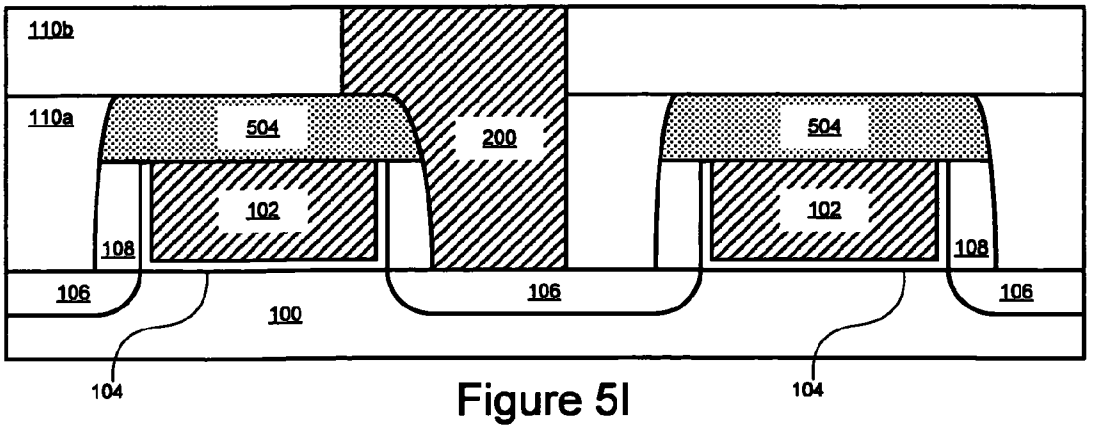


Figure 5I

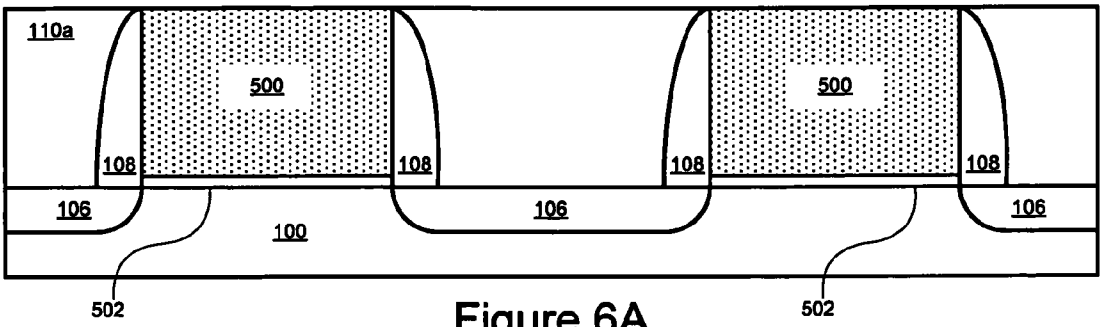


Figure 6A

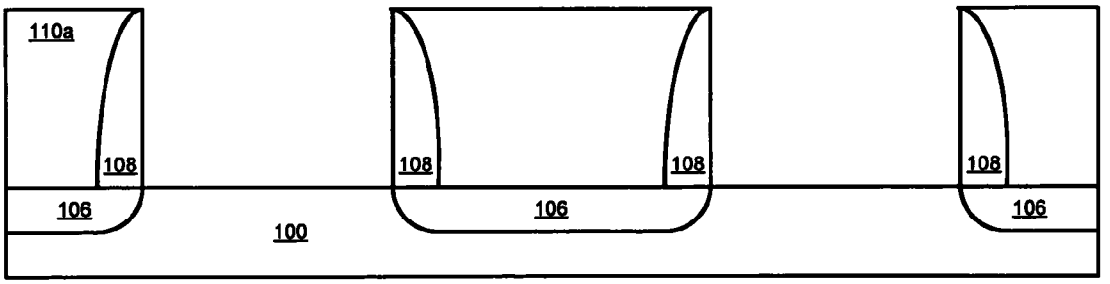


Figure 6B

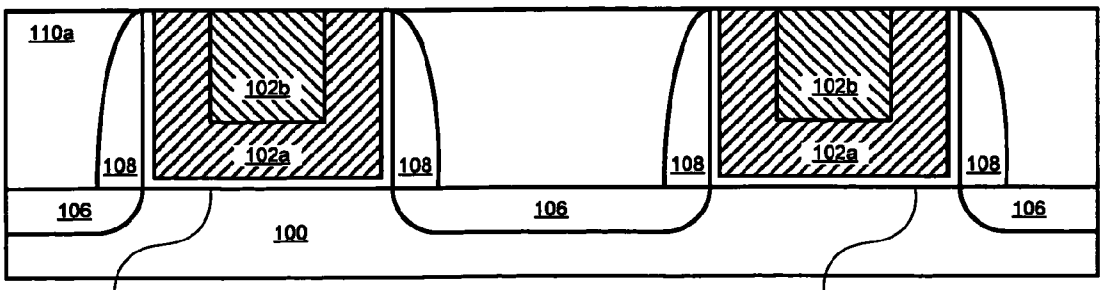


Figure 6C

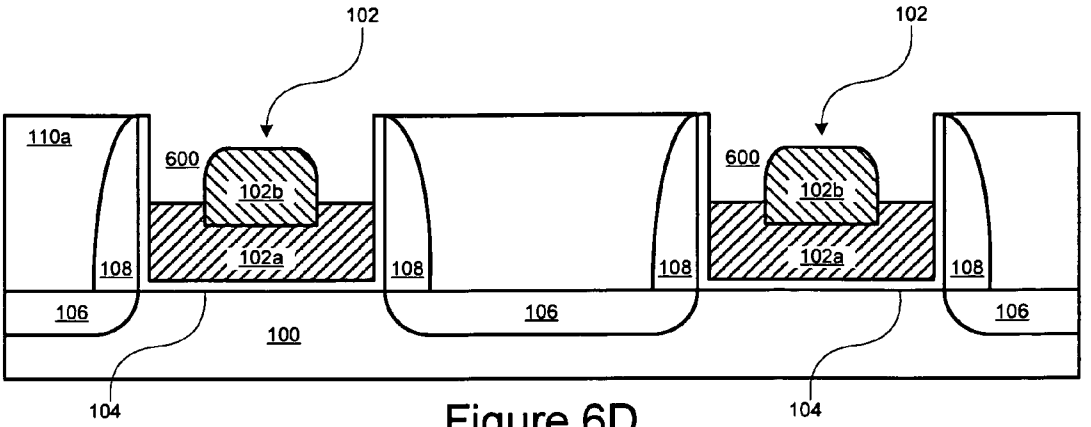


Figure 6D

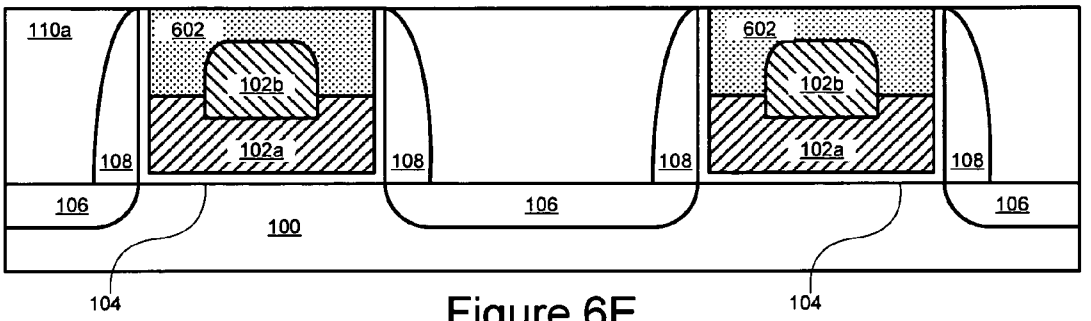


Figure 6E

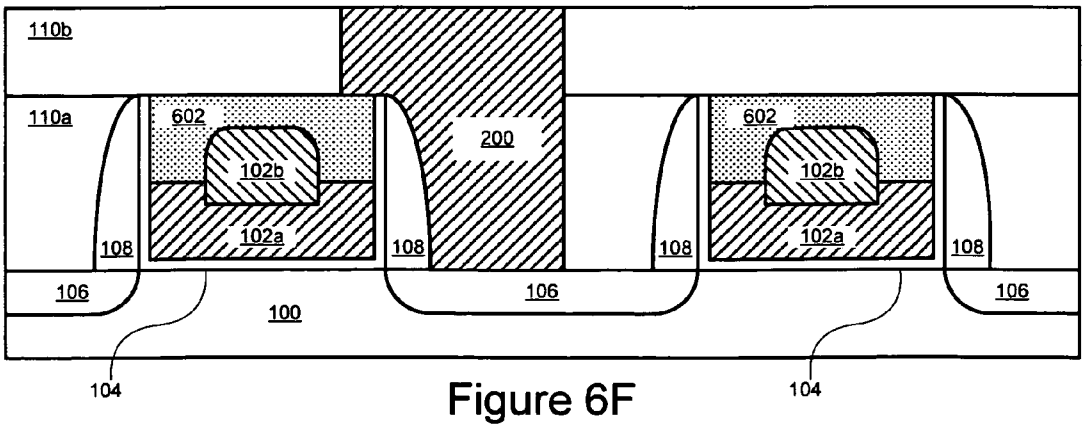


Figure 6F

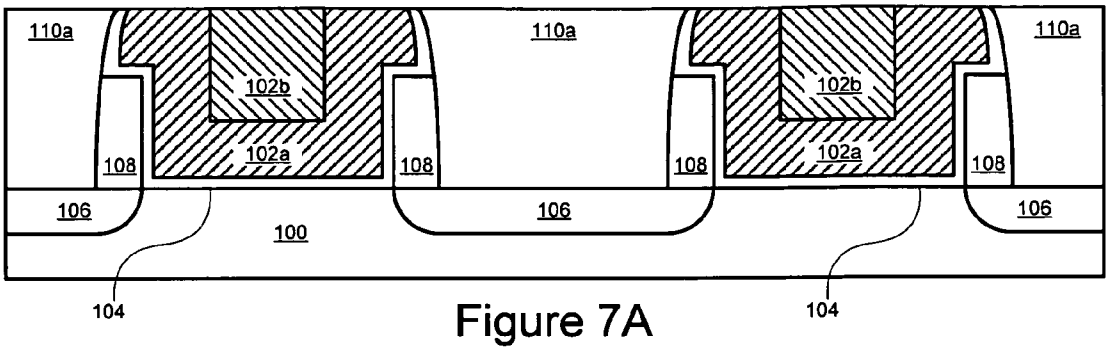


Figure 7A

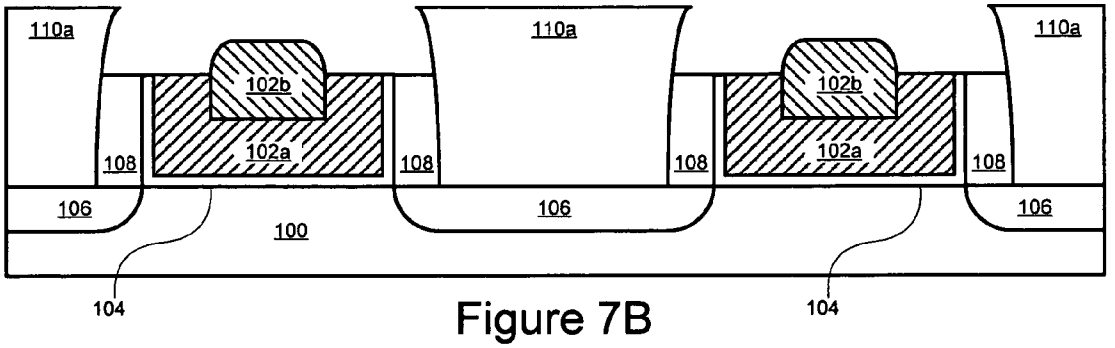


Figure 7B

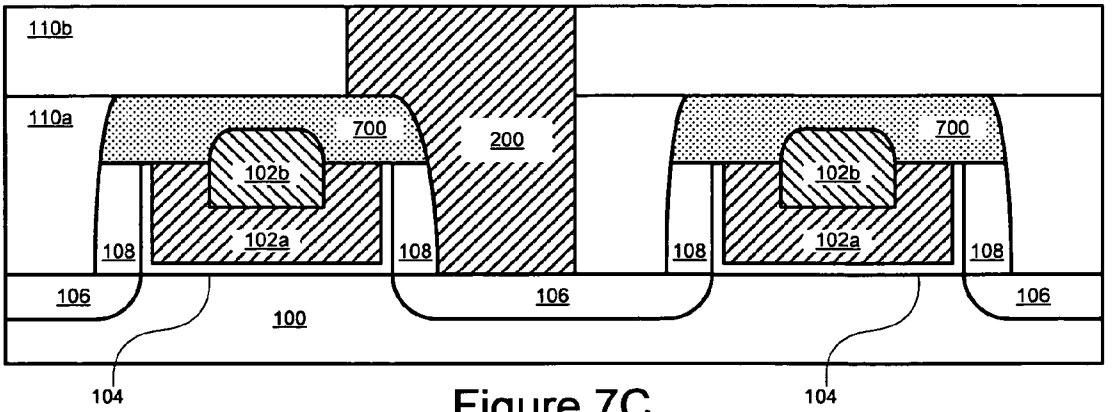


Figure 7C

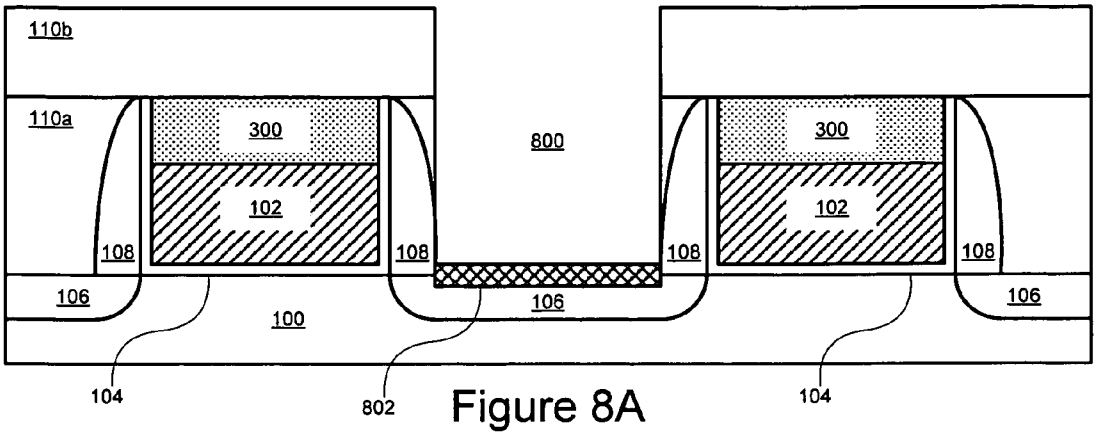


Figure 8A

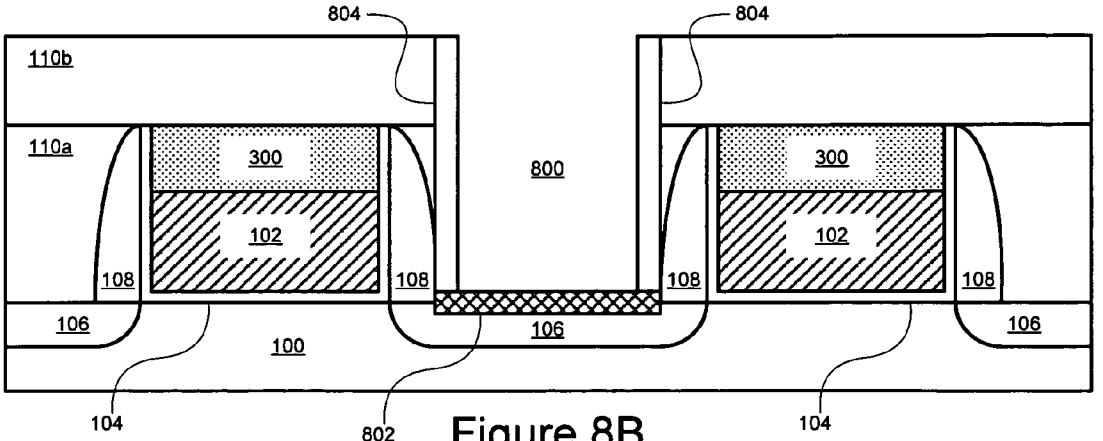


Figure 8B

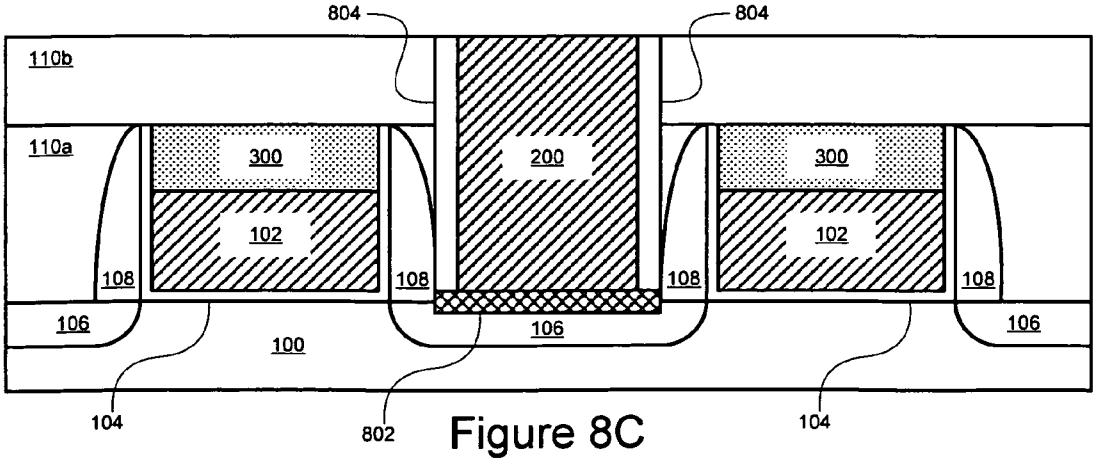


Figure 8C

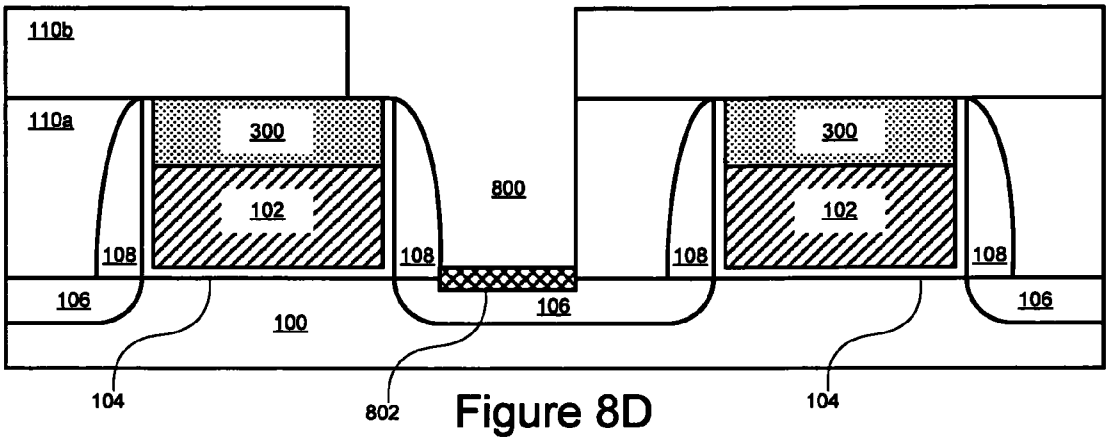


Figure 8D

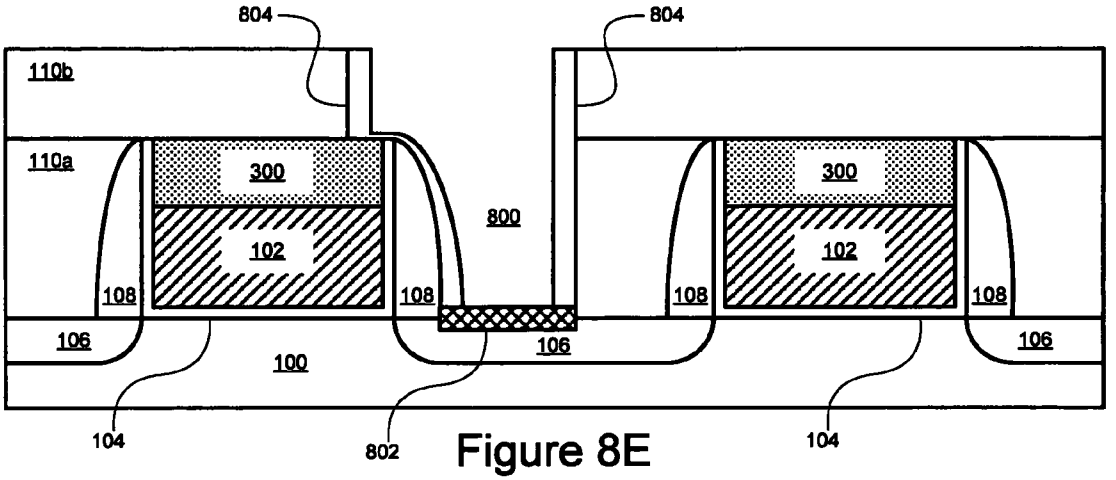


Figure 8E

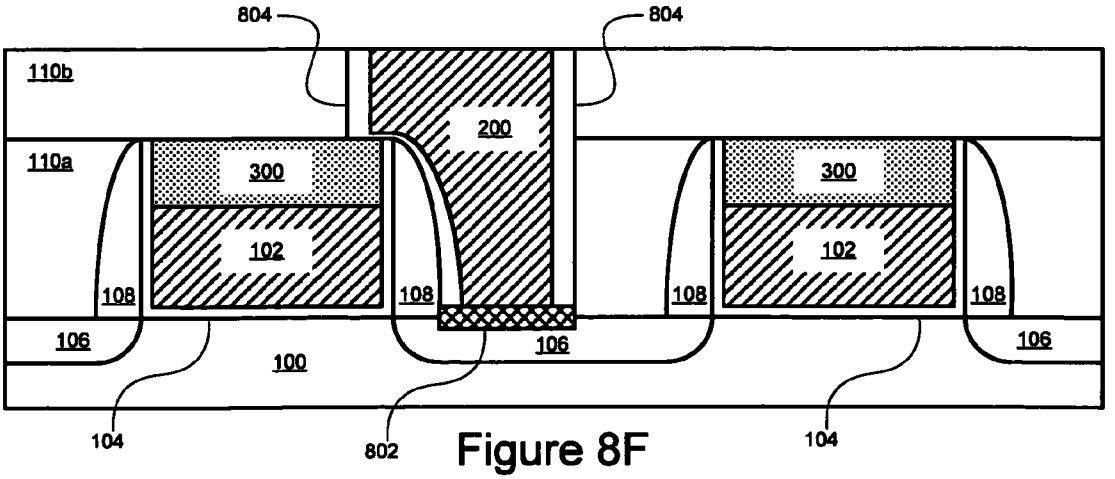


Figure 8F

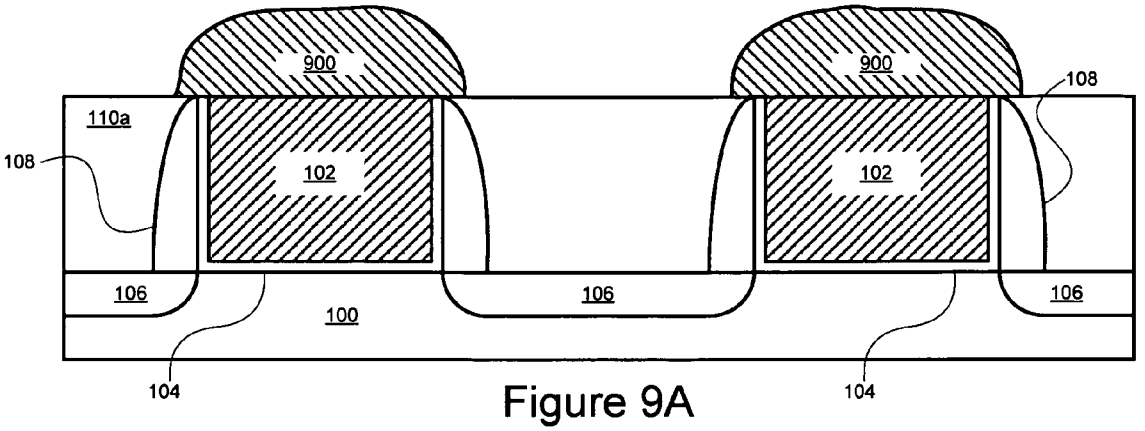


Figure 9A

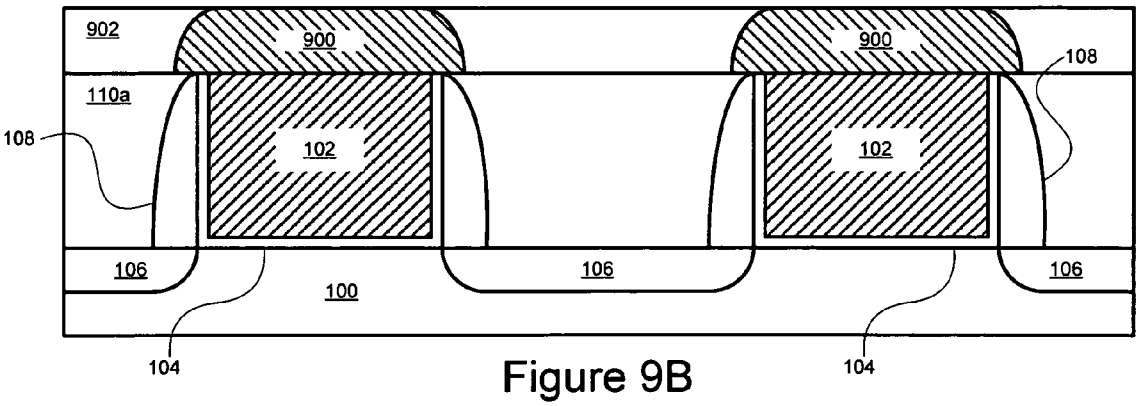


Figure 9B

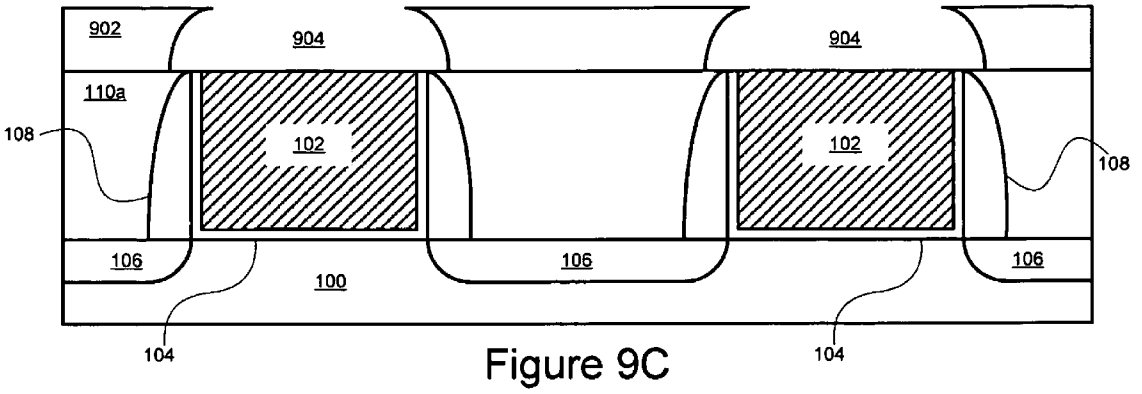


Figure 9C

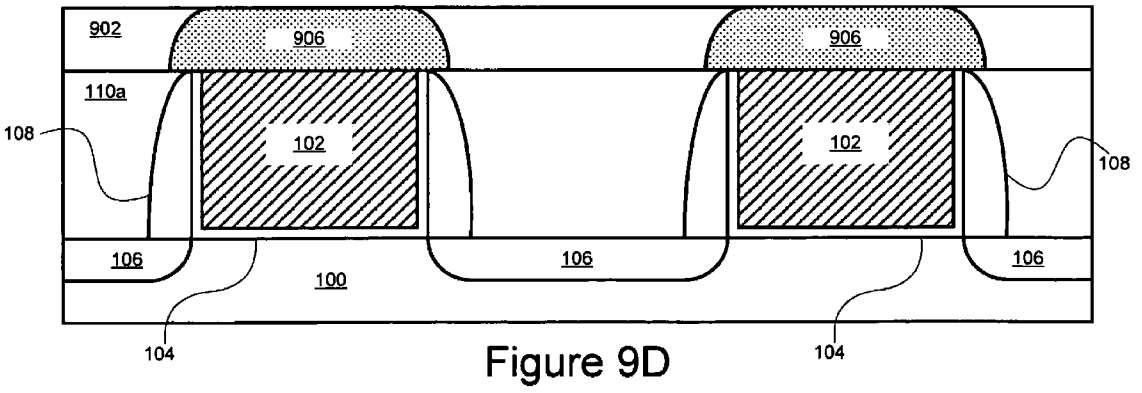


Figure 9D

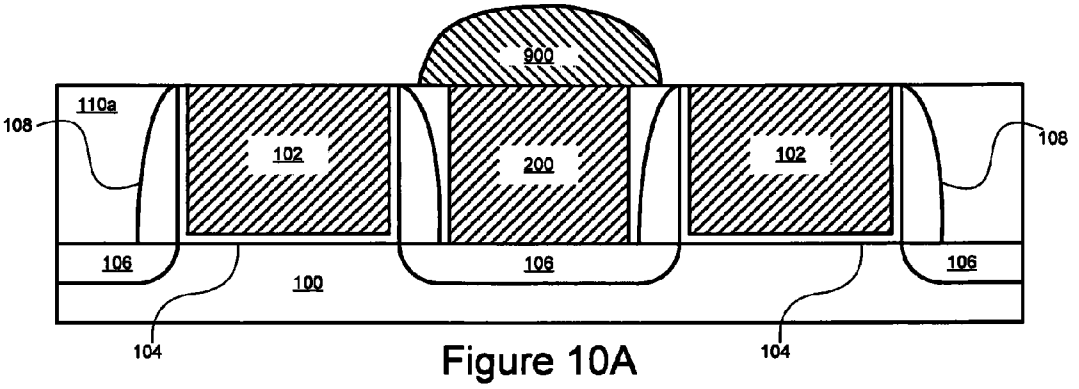


Figure 10A

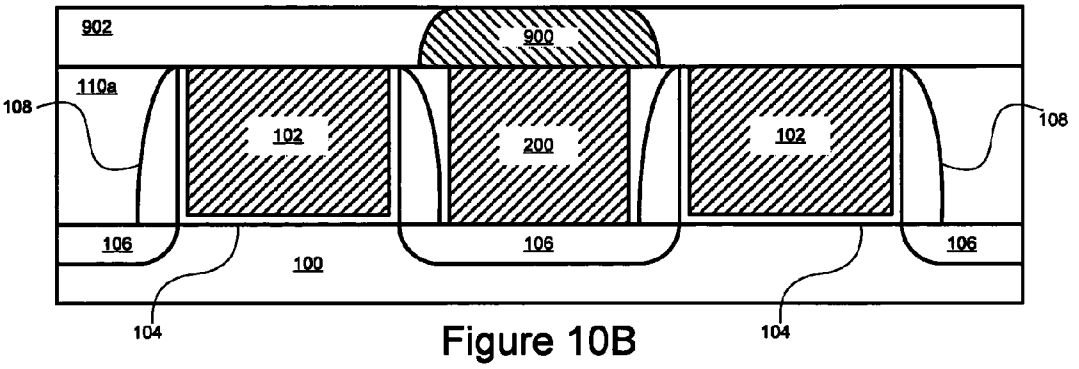


Figure 10B

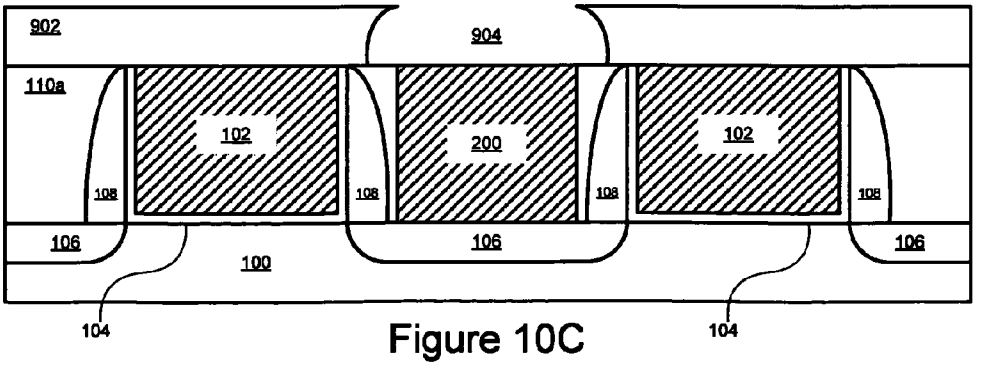


Figure 10C

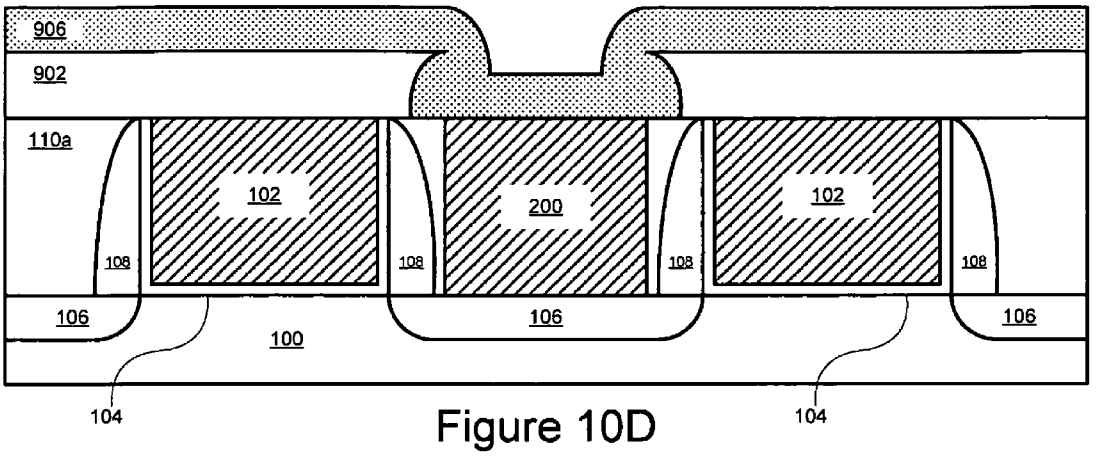


Figure 10D

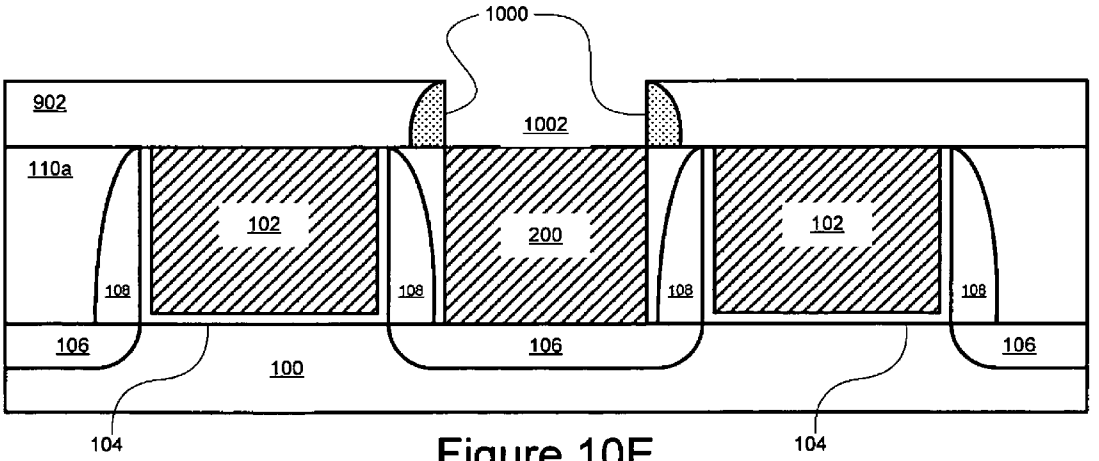


Figure 10E

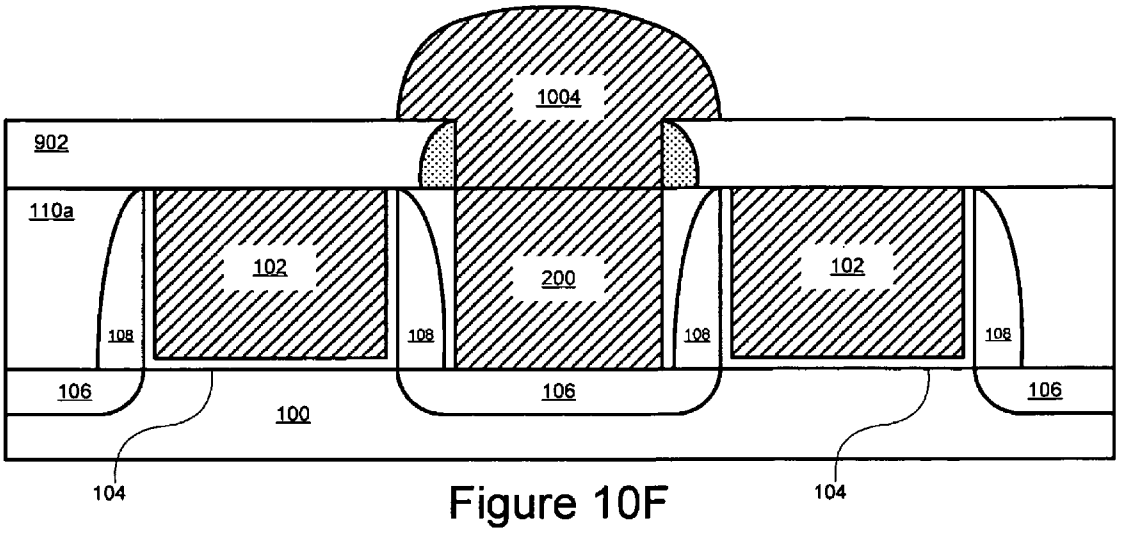


Figure 10F

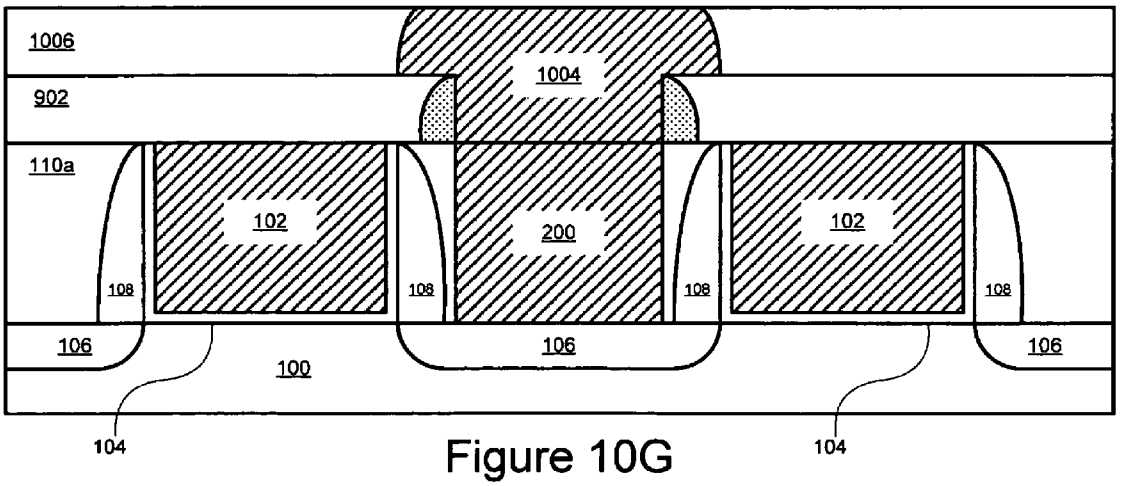


Figure 10G

SELF-ALIGNED CONTACTS

BACKGROUND

Metal-oxide-semiconductor (MOS) transistors, such as MOS field effect transistors (MOSFET), are used in the manufacture of integrated circuits. MOS transistors include several components, such as a gate electrode, gate dielectric layer, spacers, and diffusion regions such as source and drain regions. An interlayer dielectric (ILD) is typically formed over the MOS transistor and covers the diffusion regions.

Electrical connections are made to the MOS transistor by way of contact plugs that are typically formed of a metal such as tungsten. The contact plugs are fabricated by first patterning the ILD layer to form vias down to the diffusion regions. The patterning process is generally a photolithography process. Next, metal is deposited in the vias to form the contact plugs. A separate contact plug is formed down to the gate electrode using the same or a similar process.

One problem that can occur during the fabrication of a contact plug is the formation of a contact-to-gate short. A contact-to-gate short is a short circuit that occurs when the contact plug is misaligned and comes into electrical contact with the gate electrode. One conventional approach to preventing contact-to-gate shorts is by controlling registration and critical dimensions (CDs). Unfortunately, for transistors with gate pitches (gate length+space) at or below 100 nanometers (nm), CD control for gate and contact dimensions needs to be less than 10 nm and the registration control between gate and contact layers also needs to be less than 10 nm to deliver a manufacturable process window. Thus, the likelihood of a contact shorting to a gate is very high. This problem becomes more prevalent as transistor gate pitch dimensions are scaled down further because the critical dimensions become much smaller.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1**A illustrates a substrate and two conventional MOS transistors with a correctly aligned trench contact.

FIG. **1**B illustrates a misaligned trench contact formed to a diffusion region of the MOS transistors, resulting in a contact-to-gate short.

FIG. **2**A illustrates a substrate and two MOS transistors having insulator-cap layers atop their respective metal gate electrodes in accordance with one implementation of the invention.

FIG. **2**B illustrates a correctly aligned trench contact formed between two MOS transistors of the invention having insulator-cap layers.

FIG. **2**C illustrates a misaligned trench contact formed between two MOS transistors of the invention having insulator-cap layers, where the misalignment does not result in a contact-to-gate short.

FIGS. **3**A to **3**C illustrate an insulator-cap layer formed after a replacement metal gate process, in accordance with an implementation of the invention.

FIGS. **4**A to **4**C illustrate an insulator-cap layer formed after a replacement metal gate process, in accordance with another implementation of the invention.

FIGS. **5**A to **5**I illustrate a fabrication process for an insulator-cap layer that extends over the spacers of a MOS transistor, in accordance with an implementation of the invention.

FIGS. **6**A to **6**F illustrate a fabrication process for a metal gate electrode having a stepped profile, in accordance with an implementation of the invention.

FIGS. **7**A to **7**C illustrate MOS transistors having both metal gate electrodes with stepped profiles and insulator-cap layers that extend over the spacers, in accordance with an implementation of the invention.

FIG. **8**A to **8**F illustrate contact sidewall spacers in accordance with an implementation of the invention.

FIGS. **9**A to **9**D illustrate a fabrication process to form an insulating-cap atop a metal gate electrode in accordance with an implementation of the invention.

FIGS. **10**A to **10**G illustrate a fabrication process to form a metal stud and insulating spacers atop a trench contact in accordance with an implementation of the invention.

DETAILED DESCRIPTION

Described herein are systems and methods of reducing the likelihood of contact-to-gate shorts during the fabrication of metal-oxide-semiconductor (MOS) transistors. In the following description, various aspects of the illustrative implementations will be described using terms commonly employed by those skilled in the art to convey the substance of their work to others skilled in the art. However, it will be apparent to those skilled in the art that the present invention may be practiced with only some of the described aspects. For purposes of explanation, specific numbers, materials and configurations are set forth in order to provide a thorough understanding of the illustrative implementations. However, it will be apparent to one skilled in the art that the present invention may be practiced without the specific details. In other instances, well-known features are omitted or simplified in order not to obscure the illustrative implementations.

Various operations will be described as multiple discrete operations, in turn, in a manner that is most helpful in understanding the present invention, however, the order of description should not be construed to imply that these operations are necessarily order dependent. In particular, these operations need not be performed in the order of presentation.

FIG. **1**A illustrates a substrate **100** and two MOS transistors **101**. The MOS transistors **101** include gate electrodes **102**, gate dielectric layers **104**, and spacers **108**. Diffusion regions **106** are formed in the substrate **100**. Interlayer dielectrics (ILD), such as ILD layers **110***a* and **110***b*, are deposited in the regions between and around the two MOS transistors **101**.

FIG. **1**A also illustrates a trench contact **200** that is formed through the ILD layers **110***a/b* down to the diffusion region **106**. The trench contact **200** is typically formed using a photolithography patterning process followed by a metal deposition process. Photolithography patterning processes and metal deposition processes are well known in the art. The photolithography patterning process etches a trench opening through the ILD layers **110***a/b* down to the diffusion region **106**. The metal deposition process, such as electroplating, electroless plating, chemical vapor deposition, physical vapor deposition, sputtering, or atomic layer deposition, fills the trench opening with a metal such as tungsten or copper. A metal liner is often deposited prior to the metal, such as a tantalum or tantalum nitride liner. A planarization process, such as chemical-mechanical polishing (CMP), is used to remove any excess metal and complete the fabrication of the trench contact **200**.

It should be noted that in alternate implementations of the invention, via contacts may be used instead of trench contacts. Thus, the contact opening may be either a trench shape or a via shape, depending on the patterning process used or the needs of a particular integrated circuit process. The implementations of the invention described herein will refer to

contact trench openings and trench contacts, but it should be noted that via openings and via contacts (also known as contact plugs or via plugs) can be used instead of contact trench openings and trench contacts in any of these implementations.

As integrated circuit technology advances, transistor gate pitches progressively scale down. This gate pitch scaling has resulted in a number of new, problematic issues, one of which is increased parasitic capacitance (denoted by the "C" in FIG. **1**A) caused by relatively tight spacing between the trench contact **200** and the diffusion region **106** on one side and the gate electrode **102** on the other. The spacers **108** tend to provide the bulk of the separation between the trench contact **200**/diffusion region **106** and the gate electrodes **102**. Conventional spacer materials, such as silicon nitride, do little to reduce this parasitic capacitance. Unfortunately, parasitic capacitance degrades transistor performance and increases chip power.

Another problematic issue caused by gate pitch scaling is the formation of contact-to-gate (CTG) shorts. The fabrication process for the trench contact **200** is designed to prevent the trench contact **200** from coming into physical contact with the metal gate electrode **102**. When such contact occurs, a CTG short is created that effectively ruins the MOS transistor. CTG shorts have become a major yield limiter as transistor gate pitches have scaled down below 100 nanometers (nm).

Current methods to reduce CTG shorts include controlling registration and patterning contacts with smaller critical dimensions. However, as gate pitch has scaled down, the registration requirements are becoming very difficult to meet with existing technology. For instance, transistors with gate pitches at or below 100 nm require CD control and layer registration control of less than 10 nm to deliver a manufacturable process window. Thus, the likelihood of a contact shorting to a gate is very high.

FIG. **1**B illustrates what happens when the trench contact **200** is misaligned. The same photolithography processes are used, but as shown, the trench contact **200** is formed at a location that is not completely within the area between the two spacers **108**. The misalignment causes the trench contact **200** to be in physical contact with one of the gate electrodes **102**, thereby creating a contact-to-gate short.

In accordance with implementations of the invention, an insulator-capped gate electrode may be used to minimize the likelihood of contact-to-gate shorts. In one implementation, the insulator-cap layer is formed atop the gate electrode **102** and within the spacers **108** of the MOS transistor **101**. In some implementations of the invention, the insulator-cap can consume a significant portion of the volume that exists between the spacers. For instance, the insulator-cap can consume anywhere from 10% to 80% of the volume that exists between the spacers, but will generally consume between 20% and 50% of that volume. The gate electrode and gate dielectric layer consume the majority of the remaining volume. Materials that may be used to form the insulator-cap are described below.

FIG. **2**A illustrates an insulator-capped metal gate electrode in accordance with one implementation of the invention. A substrate **100** is shown in FIG. **2**A upon which MOS transistors **101** are formed. The substrate **100** may be a crystalline semiconductor substrate formed using a bulk silicon substrate or a silicon-on-insulator substructure. In other implementations, the semiconductor substrate may be formed using alternate materials, which may or may not be combined with silicon, that include but are not limited to germanium, indium antimonide, lead telluride, indium arsenide, indium phosphide, gallium arsenide, gallium antimonide, or other Group III-V materials. Although a few examples of materials from which the substrate may be formed are described here, any material that may serve as a foundation upon which a semiconductor device may be built falls within the spirit and scope of the present invention.

Each MOS transistor **101** can be a planar transistor, as shown in FIG. **2**A, or can be a nonplanar transistor, such as a double-gate or trigate transistor. Although the implementations described herein illustrate planar transistors, the invention is not limited to planar transistors. Implementations of the invention may also be used on nonplanar transistors, including but not limited to FinFET or trigate transistors. Each MOS transistor **101** includes a gate stack formed of three layers: a gate dielectric layer **104**, a gate electrode layer **102**, and an insulator-cap layer **300**. The gate dielectric layer **104** may be formed of a material such as silicon dioxide or a high-k material. Examples of high-k materials that may be used in the gate dielectric layer **104** include, but are not limited to, hafnium oxide, hafnium silicon oxide, lanthanum oxide, lanthanum aluminum oxide, zirconium oxide, zirconium silicon oxide, tantalum oxide, titanium oxide, barium strontium titanium oxide, barium titanium oxide, strontium titanium oxide, yttrium oxide, aluminum oxide, lead scandium tantalum oxide, and lead zinc niobate. In some embodiments, the gate dielectric layer **104** may have a thickness between around 1 Angstrom (Å) and around 50 Å. In further embodiments, additional processing may be performed on the gate dielectric layer **104**, such as an annealing process to improve its quality when a high-k material is used.

The gate electrode layer **102** is formed on the gate dielectric layer **104** and may consist of at least a P-type workfunction metal or an N-type workfunction metal, depending on whether the transistor is to be a PMOS or an NMOS transistor. In some implementations, the gate electrode layer **102** may consist of two or more metal layers, where at least one metal layer is a workfunction metal layer and at least one metal layer is a fill metal layer.

For a PMOS transistor, metals that may be used for the gate electrode include, but are not limited to, ruthenium, palladium, platinum, cobalt, nickel, and conductive metal oxides, e.g., ruthenium oxide. A P-type metal layer will enable the formation of a PMOS gate electrode with a workfunction that is between about 4.9 eV and about 5.2 eV. For an NMOS transistor, metals that may be used for the gate electrode include, but are not limited to, hafnium, zirconium, titanium, tantalum, aluminum, alloys of these metals, and carbides of these metals such as hafnium carbide, zirconium carbide, titanium carbide, tantalum carbide, and aluminum carbide. An N-type metal layer will enable the formation of an NMOS gate electrode with a workfunction that is between about 3.9 eV and about 4.2 eV.

The insulator-cap layer **300** is formed on the gate electrode layer **102** and may be formed of materials that include, but are not limited to, silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, other nitride materials, other carbide materials, aluminum oxide, other oxide materials, other metal oxides, boron nitride, boron carbide, and other low-k dielectric materials or low-k dielectric materials doped with one or more of carbon, nitrogen, and hydrogen. The insulator-cap layer **300** is described in more detail below.

A pair of spacers **108** brackets the gate stack. The spacers **108** may be formed from a material such as silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, and silicon oxynitride. Processes for forming spacers are well known in the art and generally include deposition and etching process steps.

Diffusion regions **106** are formed within the substrate **100** adjacent to the gate stacks of the MOS transistors **101**. For each MOS transistor **101**, one adjacent diffusion region **106** functions as a source region and the other adjacent diffusion region **106** functions as a drain region.

The diffusion region **106** may be formed using methods or processes that are well known in the art. In one implementation, dopants such as boron, aluminum, antimony, phosphorous, or arsenic may be implanted into the substrate **100** to form the diffusion regions **106**. In another implementation, the substrate **100** may first be etched to form recesses at the locations of the diffusion regions **106**. An epitaxial deposition process may then be carried out to fill the recesses with a silicon alloy such as silicon germanium or silicon carbide, thereby forming the diffusion regions **106**. In some implementations the epitaxially deposited silicon alloy may be doped in situ with dopants such as boron, arsenic, or phosphorous. In further implementations, alternate materials may be deposited into the recesses to form the diffusion regions **106**.

One or more ILD layers **110***a/b* are deposited over the MOS transistors **101**. The ILD layers **110***a/b* may be formed using dielectric materials known for their applicability in integrated circuit structures, such as low-k dielectric materials. Examples of dielectric materials that may be used include, but are not limited to, silicon dioxide (SiO_2), carbon doped oxide (CDO), silicon nitride, organic polymers such as perfluorocyclobutane or polytetrafluoroethylene, fluorosilicate glass (FSG), and organosilicates such as silsesquioxane, siloxane, or organosilicate glass. The ILD layers **110***a/b* may include pores or other voids to further reduce their dielectric constant.

Fabrication of a trench contact **200**, also referred to as contact patterning, involves at least a photolithography process and an etching process. The photolithography process forms a photoresist hard mask that defines the location of the trench contact **200**. The process begins by depositing a photoresist material on the ILD layer **110***b*. The deposited photoresist layer is exposed to ultraviolet radiation through a patterned optical mask, wherein the pattern defines the trench contact **200**. The photoresist layer is then developed to create a photoresist hard mask layer that includes an opening where the trench contact **200** is to be formed. It should be noted that photolithography processes are well known in the art and this description is simply a brief overview of a typical photolithography process. Many intermediate steps, such as baking and alignment steps, have been omitted.

Once the photoresist hard mask is in place defining the trench contact **200**, an etching process is carried out. The etchant etches portions of the ILD layer **110***a/b* that are left exposed by openings in the photoresist hard mask, such as the opening for the trench contact **200**. The etchant therefore etches a trench opening down to the diffusion region **106**. The etching process used may be a conventional chemical wet etch process or a plasma dry etch process. The etching process is carried out for a duration of time, denoted as T_{ETCH}, that is sufficient to etch the ILD layer **110** all the way down to the diffusion region **106**. The etched trench opening is then filled with one or more metals, as described above, to form the trench contact **200**.

In accordance with implementations of the invention, the insulator-cap layer **300** has a thickness that is sufficient to protect the metal gate electrode **102** from being exposed during fabrication of the trench contact **200** should the contact trench opening be aligned over the insulator-cap layer. Furthermore, the insulator-cap layer **300** has a thickness that is sufficient to electrically isolate the metal gate electrode **102** from the trench contact **200** after the trench contact **200** is formed. In one implementation of the invention, this thickness can range from 5 nm to 50 nm. In another implementation, the height of the insulator-cap layer can account for 20% to 80% of the overall height of the gate stack. The etching process used to form the contact trench opening is selective to the insulator-cap layer **300**. This means the wet or dry etch chemistry will etch the material of the ILD layer **110***a/b* but will selectively stop and self align to the insulator-cap layer **300** and the sidewall spacers **108**.

In accordance with implementations of the invention, the insulator-cap layer **300** also has a thickness that is sufficient to withstand the etching process for the entirety of T_{ETCH} without exposing the underlying metal gate electrode **102**. Stated differently, the insulator-cap layer **300** has an initial thickness sufficient to withstand the etching process for a duration of time needed to etch the ILD layer **110***a/b* all the way down to the diffusion region **106** without any portion of the insulator-cap layer **300** being reduced to a thickness that would permit electrical conductivity between the metal gate electrode **102** and the subsequently formed trench contact **200**. After the etching process, the combination of the insulator-cap layer **300** and the spacers **108** electrically isolates the metal gate electrode **102** from the trench contact **200**, thereby eliminating CTG shorts.

There are several different ways to form the insulator-cap layer **300** of the invention. In one implementation of the invention, where the gate electrode **102** is formed using a gate-first process, a blanket dielectric layer is initially deposited on a substrate. Next, a blanket electrode layer is deposited atop the dielectric layer. Finally, a blanket insulator layer is formed atop the electrode layer. The deposition processes that are used to deposit the dielectric layer, the electrode layer, and the insulator layer are well known in the art and may include, but are not limited to, processes such as electroplating, electroless plating, chemical vapor deposition, atomic layer deposition, physical vapor deposition, and sputtering. The three layers are then etched using conventional patterning processes, such as photolithography processes, to form a gate stack consisting of a gate dielectric layer **104**, a gate electrode layer **102**, and an insulator-cap layer **300**. Spacers **108** and diffusion regions **106** are then formed on opposing sides of the gate stack. An ILD layer **110***a* is deposited over the gate stack, the spacers **108**, and the diffusion region **110**. A trench contact **200** may then be formed as described above.

In an alternate implementation of a gate-first process, a blanket dielectric layer and a blanket electrode layer may be deposited and patterned to form a gate stack that consists of the gate dielectric layer **104** and the gate electrode **102**. A pair of spacers **108** and diffusion regions **106** may be formed on either side of the gate stack. Next, an etching process may be carried out to recess the metal gate electrode **102** within the spacers **108**, thereby reducing the thickness of the metal gate electrode **102**. The recessing of the metal gate electrode **102** results in the formation of a trench between the spacers **108** where the bottom surface of the trench corresponds to the top surface of the recessed metal gate electrode **102**. The metal etch process is followed by an insulator material deposition process that deposits a blanket layer of insulator material and fills the trench between the spacers **108**. A polishing process, such as a chemical mechanical planarization process, is used to polish down the insulator material layer and substantially remove any insulator material that is outside of the spacers **108**. The removal of this excess insulator material yields an insulator-cap layer **300** that is substantially contained within the spacers **108**.

In another implementation of the invention, a gate-last process, such as a replacement metal gate process, is used to form the gate electrode. In this implementation, a blanket dielectric layer and a blanket dummy electrode layer may be initially deposited and patterned to form a gate stack that consists of the gate dielectric layer **104** and a dummy gate electrode (not shown). It should be noted that the term "dummy" is used to indicate that this layer is sacrificial in nature. The materials used in dummy layers may or may not be the same materials that are used in non-dummy layers. For instance, the dummy electrode layer may consist of polysilicon, which is used in real gate electrodes. A pair of spacers **108** and diffusion regions **106** may be formed on either side of the gate stack. Next, the dummy gate electrode may be etched out to form a trench between the spacers **108** and atop the gate dielectric layer **104**. An electrode metal layer may then be deposited to fill the trench. The electrode metal layer may be polished down to remove metal outside of the spacers **108** and to confine the electrode metal to the trench between the spacers **108**, thereby forming a metal gate electrode **102**.

As described above, an etching process is carried out to recess the metal gate electrode **102** within the spacers **108**. The recessing of the metal gate electrode **102** results in the formation of a trench between the spacers **108**. An insulator material deposition process fills the trench and a polishing process is used to polish down the insulator material layer and substantially remove any insulator material that is outside of the spacers **108**. This yields an insulator-cap layer **300** that is substantially contained within the spacers **108**.

FIG. **2**B illustrates a trench contact **200** that is correctly aligned between two MOS transistors having insulator-cap layers **300**. In this instance the insulator-cap **300** is not used.

FIG. **2**C illustrates a misaligned trench contact **200** formed between two MOS transistors having insulator-cap layers **300**. As shown, a portion of the misaligned trench contact **200** is situated directly over the gate electrode **102**. Unlike the prior art transistors shown in FIG. **1**B, however, a CTG short is avoided due to the use of the insulator-cap layer **300**. The insulator-cap layer **300** electrically isolates the metal gate electrode **102** from the misaligned trench contact **200**, allowing the trench contact **200** to be "self-aligned".

FIGS. **3**A to **3**C illustrate a slight variation on the transistors of FIG. **2**A. In FIG. **3**A, a different implementation of a replacement metal gate process is used to form the transistors. In this implementation, a blanket dummy dielectric layer and a blanket dummy electrode layer are deposited on a substrate. Here, the dummy electrode layer may consist of polysilicon and the dummy dielectric layer may consist of silicon dioxide, both of which are used in real gate electrodes and real gate dielectric layers. These two dummy layers are etched to form a gate stack that consists of a dummy gate dielectric layer and a dummy gate electrode layer. Spacers **108** and diffusion regions **106** are then formed on opposing sides of the gate stack. An ILD layer **110***a* is deposited over the gate stack, spacers **108**, and diffusion regions **106**. The ILD layer **110***a* is planarized to expose the dummy electrode layer.

Next, the dummy electrode layer and the dummy gate dielectric layer are removed using one or more etching processes. The removal of the dummy layers produces a trench between the spacers **108**. The substrate **100** forms a bottom surface of the trench. A new high-k gate dielectric layer **104** is deposited into the trench using a chemical vapor deposition process or an atomic layer deposition process. The high-k gate dielectric layer **104** is deposited along the bottom and sidewalls of the trench, thereby forming a "U" shaped gate dielectric layer **104**, as shown in FIG. **3**A. Next, a metal gate electrode layer **102** is deposited atop the high-k gate dielectric layer **104**. Processes for forming the metal gate electrode **102** are well known in the art.

In accordance with implementations of the invention, the final metal gate electrode **102** does not fill the trench in its entirety. In one implementation, the metal gate electrode **102** may initially fill the trench in its entirety, but a subsequent etching process may be used to recess the metal gate electrode **102**. In another implementation, the metal gate electrode deposition process only partially fills the trench with the metal gate electrode **102**. In both implementations, a trench remains above the final metal gate electrode **102** between the spacers **108**.

Finally, an insulator material deposition process is used to deposit a blanket layer of insulator material that fills the trench between the spacers **108**. A polishing process, such as a chemical mechanical planarization process, is then used to polish down the insulator material layer and remove substantially any insulator material that is outside of the spacers **108**. The removal of this excess insulator yields an insulator-cap layer **300** that is substantially confined within the spacers **108**. As shown in FIG. **3**A, the insulator-cap **300** is also confined within the sidewall portions of the gate dielectric layer **104**.

FIG. **3**B illustrates a trench contact **200** that is correctly aligned between two MOS transistors having insulator-cap layers **300**. FIG. **3**C illustrates a misaligned trench contact **200** formed between two MOS transistors having insulator-cap layers **300**. Again, a portion of the misaligned trench contact **200** is situated directly over the gate electrode **102**. A CTG short is avoided due to the use of the insulator-cap layer **300**, which electrically isolates the metal gate electrode **102** from the misaligned trench contact **200**.

FIGS. **4**A to **4**C illustrate a slight variation on the transistors of FIG. **3**A. In FIG. **4**A, a replacement gate process is used again to form transistors having a "U" shaped gate dielectric layer **104**. The gate electrode layer **102** and the gate dielectric layer **104** are initially formed using the same processes detailed above for FIG. **3**A. Unlike FIG. **3**A, in this implementation, both the "U" shaped gate dielectric layer **104** and the metal gate electrode **102** are recessed prior to fabrication of the insulator-cap layer **300**. One or more etching processes may be used to recess both structures. The insulator-cap **300** is then formed using the same process described above for FIG. **3**A and is situated atop both the gate electrode **102** and portions of the gate dielectric layer **104**, as shown in FIG. **4**A. FIG. **4**B illustrates a trench contact **200** that is correctly aligned between two MOS transistors having insulator-cap layers **300**. FIG. **4**C illustrates a misaligned trench contact **200** formed between two MOS transistors having insulator-cap layers **300**. Again, a portion of the misaligned trench contact **200** is situated directly over the gate electrode **102**. A CTG short is avoided due to the use of the insulator-cap layer **300**, which electrically isolates the metal gate electrode **102** from the misaligned trench contact **200**.

FIGS. **5**A to **5**F illustrate the fabrication of an alternate insulator-cap layer that may be used with a MOS transistor. Initially, FIG. **5**A illustrates two MOS transistors that include a dummy gate electrode **500** and a dummy gate dielectric layer **502**. Also shown are a pair of spacers **108** that are generally formed of silicon nitride.

In accordance with implementations of the invention, one or multiple etching processes are carried out to partially recess both the dummy gate electrode layer **500** and the spacers **108**. This dual recess is shown in FIG. **5**B. The etch chemistry used to recess the dummy gate electrode **500** may differ from the etch chemistry used to recess the spacers **108**. The etching processes used may be wet etches, dry etches, or

a combination. When the dummy gate electrode **500** and the spacers **108** have been recessed, a trench **503***a* is formed within the ILD layer **110***a* where the top surfaces of the dummy gate electrode **500** and the spacers **108** form the bottom of the trench.

Moving to FIG. **5**C, one or more etching processes are carried out to completely remove the dummy gate electrode **500** as well as the dummy gate dielectric **502**. Etching processes to completely remove the dummy gate electrode **500** and dummy gate dielectric are well known in the art. Again, these etches may be wet, dry, or a combination. As shown in FIG. **5**C, the trench **503***a* is now much deeper and has a cross-section profile that is relatively wide at the top of the trench **503***a* and relatively narrow at the bottom of the trench **503***a*. The dummy gate electrode **500** and dummy gate dielectric **502** are removed in their entirety, thereby exposing the top of the substrate **100**.

In FIG. **5**D, a gate dielectric layer **104** and a metal gate electrode layer **102** are deposited in the trench **503***a*. A conformal deposition process, such as a CVD or an ALD process, is generally used for the deposition of the gate dielectric layer **104**, resulting in a conformal dielectric layer **104** that covers the sidewalls and bottom surface of the trench **503***a*. The metal gate electrode layer **102** fills the remainder of the trench **503***a*. In some implementations of the invention, the metal gate electrode layer **102** may consist of two or more layers of metal, for instance, a work function metal layer and a fill metal layer.

In a replacement metal gate process flow, it is very challenging to fill narrow gate trenches with metal gate materials, particularly with transistors having gate widths at or below 22 nm. The process flow described here in FIGS. **5**A to **5**D enhances the intrinsic fill characteristics by widening the trench openings at the top without affecting the narrow trench widths at the bottom. Thus, the cross-section profile of the trench **503***a*, with its relatively wide opening at the top, results in an improved metal gate electrode deposition with fewer voids or other defects.

Next, the metal gate electrode layer **102** and the gate dielectric layer **104** are recessed as shown in FIG. **5**E, forming a trench **503***b*. Again, one or more etching processes, either wet or dry, may be used to recess both the gate electrode layer **102** and the gate dielectric layer **104**. The etch processes used must be selective to the ILD layer **110***a*. The metal gate electrode **102** is recessed until its top surface is even with or below the top surfaces of the spacers **108**. Although portions of the metal gate electrode **102** are on top of the spacers **108** in FIG. **5**D, it is important that no portion of the metal gate electrode **102** remain above the top of the spacers **108** after the recessing of the metal gate **102** in FIG. **5**E. This is because any portion of the metal gate electrode **102** that remains atop the spacers **108** may end up forming a CTG short with a misaligned trench contact.

Moving to FIG. **5**F, an insulator material deposition process fills the trench **503***b* and a polishing process is used to polish down the insulator material layer and substantially remove any insulator material that is outside of the trench **503***b*. This yields an insulator-cap layer **504** that is substantially contained within the trench **503***b*. The insulator-cap layer **504** has the appearance of a mushroom top as it extends laterally above the spacers **108**. The insulator-cap layer **504** improves contact-to-gate margin by extending over the gate spacer **108**. The insulator-cap layer **504** may be formed of materials that include, but are not limited to, silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, other nitride materials, other carbide materials, aluminum oxide, other oxide materials, other metal oxides, and low-k dielectric materials.

FIG. **5**G illustrates the deposition of an additional ILD layer **110***b* that covers the insulator-cap layers **504** and sits atop the first ILD layer **110***a*. FIG. **5**H illustrates a trench contact **200** that has been fabricated down to the diffusion region **106** through the ILD layers **110***a* and **110***b*. The trench contact **200** of FIG. **5**H has been correctly aligned between the spacers **108** of adjacent transistors.

FIG. **5**I illustrates a trench contact **200** that is misaligned. As shown, even though the trench contact **200** is situated on top of the metal gate electrode **102**, the insulating-cap layer **504** protects the metal gate electrode **102** and prevents a CTG short from forming by electrically isolating the metal gate electrode **102** from the misaligned trench contact **200**.

Another advantage provided by the insulating-cap layer **504** addresses the parasitic capacitance issue discussed above in relation to FIG. **1**A. Parasitic capacitance issues are caused by the relatively tight spacing between the trench contact **200** and the diffusion region **106** on one side and the gate electrode **102** on the other side. The spacers **108** tend to provide the bulk of the separation between the trench contact **200**/diffusion region **106** and the gate electrodes **102**, but conventional spacer materials, such as silicon nitride, do little to reduce this parasitic capacitance. Nevertheless, silicon nitride is still used because the etching process that creates a contact trench opening for the trench contact **200** is selective to silicon nitride.

In accordance with this implementation of the invention, materials other than silicon nitride may be used in the spacers **108**. Here, the laterally extending insulating-cap layer **504** protects the underlying spacers **108** during etching processes used to fabricate the trench contact **200**. These etching processes are generally anisotropic processes, therefore, the etch chemistry need only be selective to the insulating-cap layer **504**. The insulating-cap layer **504** can then shield the underlying spacers **108**. So with an anisotropic process, the use of the insulating-cap layer **504** means the etch chemistry does not necessarily need to be selective to the material used in the spacers **108**. This removes any constraints on the choice of spacer material and enables the use of materials that are optimized for capacitance. For instance, materials such as silicon oxynitride (SiON), carbon-doped silicon oxynitride (SiOCN), or low-k dielectric materials may be used in the spacers **108** to reduce issues with parasitic capacitance.

FIGS. **6**A to **6**F illustrate the formation of a stepped metal gate electrode in conjunction with an insulating-cap layer in accordance with an implementation of the invention. Initially, FIG. **6**A illustrates two MOS transistors that include a dummy gate electrode **500** and a dummy gate dielectric layer **502**. Moving to FIG. **6**B, one or more etching processes are carried out to completely remove the dummy gate electrode **500** as well as the dummy gate dielectric **502**. Etching processes to completely remove the dummy gate electrode **500** and dummy gate dielectric are well known in the art. The dummy gate electrode **500** and dummy gate dielectric **502** are removed in their entirety, thereby exposing the top of the substrate **100**.

FIG. **6**C illustrates the deposition of dual metal gate electrode layers, a conformal metal gate electrode layer **102***a* and a second metal layer **102***b* that may or may not be conformal. The initial metal gate electrode layer **102***a* may be deposited using a conformal deposition process such as chemical vapor deposition or atomic layer deposition. Other processes, such as physical vapor deposition or sputtering, may also be used. The second metal gate electrode **102***b* is deposited using a conventional deposition process such as chemical vapor

deposition, atomic layer deposition, physical vapor deposition, sputtering, or even processes such as electroplating or electroless plating since a conformal layer is not needed for layer **102***b*.

The initial metal gate electrode layer **102***a* is typically a workfunction metal layer and can be formed using any of the workfunction metals described above. The second metal gate electrode layer **102***b* may be a second workfunction metal layer or it may be a low resistance fill metal layer such as aluminum, tungsten, or copper. In accordance with implementations of the invention, the metal used in the metal gate electrode **102***a* has different etch properties than the metal used in the metal gate electrode **102***b*.

Moving to FIG. **6**D, the dual metal gate electrode layers **102***a* and **102***b* are etched and recessed to form trenches **600** in which insulating cap layers may be fabricated. In accordance with an implementation of the invention, the etching process removes a larger portion of metal layer **102***a* than metal layer **102***b*. This yields a stepped or bulleted profile for the metal gate electrode **102**, as shown in FIG. **6**D. A middle portion of the overall metal gate electrode **102** is relatively thicker than the outer edge portions of the overall metal gate electrode **102**. Stated differently, a middle portion of the metal gate electrode **102** has a relatively larger height than side portions of the metal gate electrode **102**. This stepped profile for the metal gate electrode **102** provides advantages as explained below in FIG. **6**F.

In one implementation, a single etching process is used that etches the metal gate electrode layer **102***a* at a faster rate than the metal gate electrode layer **102***b*. In other words, the etch chemistry is more selective to the metal gate electrode **102***b*. In another implementation, two etching processes may be used, one for metal layer **102***a* and another for metal layer **102***b*. If two etching processes are used, a larger portion of metal layer **102***a* must be removed relative to metal layer **102***b*. Thus in one implementation, the first of the two etching processes may be selective to the metal layer **102***b* and the second of the two etching processes may be selective to the metal layer **102***a*. The etching processes used may be wet etch, dry etch, or a combination of both. It will be appreciated by those of ordinary skill in the art that for almost any arbitrary pair of metals used in metal layers **102***a* and **102***b*, it is possible to find a wet or dry chemical etch that will differentiate between the two metals.

As shown in FIG. **6**E, an insulator material deposition process fills the trenches **600** and a polishing process is used to polish down the insulator material layer and substantially remove any insulator material that is outside of the trench **600**. This yields an insulator-cap layer **602** that is substantially contained within the trench **600**. The insulator-cap layer **602** is relatively thick at its outer edges and relatively thin at its middle portion due to the stepped profile of the metal gate electrode **102**. The insulator-cap layer **602** may be formed of materials that include, but are not limited to, silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, other nitride materials, other carbide materials, aluminum oxide, other oxide materials, other metal oxides, and low-k dielectric materials.

FIG. **6**F illustrates a trench contact **200** that is misaligned. As shown, even though the trench contact **200** is situated on top of the metal gate electrode **102**, the insulating-cap layer **602** protects the metal gate electrode **102** and prevents a CTG short from forming by electrically isolating the metal gate electrode **102** from the misaligned trench contact **200**. The stepped profile of the metal gate electrode **102** provides at least two advantages. First, the stepped profile causes the thick portion of the insulator-cap layer **602** to be positioned between the metal gate electrode **102** and the trench contact **200**, thereby providing strong electrical isolation. Second, the stepped profile allows the middle portion of the metal gate electrode **102** to remain thick, thereby lowering the electrical resistance of the metal gate electrode **102** by increasing its metal content. In various implementations of the invention, the stepped profile may be optimized by trying to maximize the volume or width of the middle portion of the metal gate electrode **102** while maintaining its electrical isolation from misaligned trench contact **200**. In some implementations, this may be done by increasing the size or thickness of the metal gate electrode **102***b*. In further implementations, this may be done by using more than two metal gate electrode layers to more finely tailor the stepped profile.

In accordance with another implementation of the invention, FIGS. **7**A to **7**C illustrate the fabrication of a MOS transistor that combines the wide insulator-cap layer **504** of FIG. **5**F with the stepped profile metal gate electrode **102** of FIGS. **6**D to **6**F. Starting with the structure shown in FIG. **5**C, dual metal gate electrode layers are deposited as shown in FIG. **7**A. One layer is a conformal metal gate electrode layer **102***a* and the other layer is a second metal layer **102***b* that may or may not be conformal. The initial metal gate electrode layer **102***a* is typically a workfunction metal layer and the second metal gate electrode layer **102***b* may be a second workfunction metal layer or it may be a fill metal layer. In accordance with implementations of the invention, the metal used in the metal gate electrode **102***a* has different etch properties than the metal used in the metal gate electrode **102***b*.

Moving to FIG. **7**B, the dual metal gate electrode layers **102***a* and **102***b*, as well as the gate dielectric layer **104**, are etched and recessed. The etch process is selective to the metal gate electrode **102***b*. This yields a stepped profile for the metal gate electrode **102**, as shown in FIG. **7**B. A middle portion of the overall metal gate electrode **102** is relatively thicker than the outer edge portions of the overall metal gate electrode **102**.

An insulating material is then deposited and planarized to form insulator-cap layers **700** atop each metal gate electrode **102**. This is shown in FIG. **7**C. Also shown is a misaligned trench contact **200**. The stepped profile of the metal gate electrode **102** allows the thick portion of the insulator-cap layer **700** to electrically isolate the metal gate electrode **102** from the trench contact **200**. The stepped profile also allows a middle portion of the metal gate electrode **102** to remain thick, thereby reducing electrical resistance. In this implementation, the insulating-cap layer **700** extends over the recessed spacers **108**, thereby protecting the spacers during the trench contact **200** etch process and allowing a material to be used in the spacers **108** that is optimized for reducing parasitic capacitance between the trench contact **200** and the metal gate electrode **102**.

FIGS. **8**A to **8**F illustrate another implementation of the invention in which contact sidewall spacers are used to reduce CTG shorts and to improve parasitic capacitance issues. FIG. **8**A illustrates a contact trench opening **800** that has been etched through ILD layers **110***a* and **110***b* down to the diffusion region **106**. As explained above, photolithography patterning and etching processes are used to form the contact trench opening **800**.

Also shown in FIG. **8**A is a silicide layer **802** that has been formed at the bottom of the contact trench opening **800**. To fabricate the silicide layer **802**, a conventional metal deposition process, such as a sputtering deposition process or an ALD process, may be used to form a conformal metal layer along at least the bottom of the contact trench opening **800**. Often the metal will deposit on the sidewalls of the contact

trench opening **800** as well. The metal may include one or more of nickel, cobalt, tantalum, titanium, tungsten, platinum, palladium, aluminum, yttrium, erbium, ytterbium, or any other metal that is a good candidate for a silicide. An annealing process may then be carried out to cause the metal to react with the diffusion region **106** and form a silicide layer **802**. Any unreacted metal may be selectively removed using known processes. The silicide layer **802** reduces the electrical resistance between the later formed trench contact **200** and the diffusion region **106**.

FIG. **8**B illustrates a pair of contact sidewall spacers **804** that are formed along the sidewalls of the contact trench opening **800**, in accordance with an implementation of the invention. The contact sidewall spacers **804** may be formed using deposition and etching processes similar to the fabrication of gate spacers **108**. For instance, a conformal layer of an insulating material may be deposited within the contact trench opening **800**, resulting in the insulating material being deposited along the sidewalls and bottom surface of the contact trench opening **800**. The insulating material may be silicon oxide, silicon nitride, silicon oxynitride (SiON), carbon-doped silicon oxynitride (SiOCN), any other oxide, any other nitride, or any low-k dielectric material. Next, an anisotropic etching process is used to remove the insulating material from the bottom of the contact trench opening **800**, as well as from other areas such as the surface of the ILD layer **110***b*. This yields the contact sidewall spacers **804** that are shown in FIG. **8**B.

As will be appreciated by those of skill in the art, a separate patterning process may be used to form vias down to the metal gate electrodes **102** in order to form gate contacts. This separate patterning process will typically involve coating the wafer with a sacrificial photo-definable resist layer, etching the gate contacts, and then removing the photoresist with a wet or dry cleaning process or some combination thereof. This separate patterning process is generally carried out after the contact trench opening **800** has been formed, which means first the resist coating and then the wet or dry clean chemistry enters the contact trench opening **800** and can degrade the silicide layer **802**. Therefore, in accordance with an implementation of the invention, the conformal layer of insulating material used to form the spacers **804** is deposited before the patterning process for the gate contacts. The conformal layer remains in place to protect the silicide layer **802** until after the gate contacts have been patterned. Then the anisotropic etch described above may be carried out to etch the conformal layer and form the spacers **804**.

It should be noted that the silicide layer **802** is formed prior to fabrication of the contact sidewall spacers **804**, which is when the contact trench opening **800** is at its largest width. By forming the silicide layer **802** before forming the contact sidewall spacers **804**, a relatively wider silicide layer **802** can be formed to provide better electrical resistance properties, such as lower intrinsic contact resistance. If the contact sidewall spacers **804** are formed first, then less of the diffusion region **106** would be exposed for the silicide fabrication process, yielding a relatively shorter silicide layer.

A metal deposition process is then carried out to fill the contact trench opening **800** and form the trench contact **200**, as shown in FIG. **8**C. As noted above, the metal deposition process can be any metal deposition process, such as electroplating, electroless plating, chemical vapor deposition, physical vapor deposition, sputtering, or atomic layer deposition. The metal used may be any metal that provides suitable contact properties, such as tungsten or copper. A metal liner is often deposited prior to the metal, such as a tantalum or tantulum nitride liner. A CMP process is used to remove any excess metal and complete the fabrication of the trench contact **200**.

The contact sidewall spacers **804** provide an additional layer of protection between the gate electrodes **102** and the trench contact **200**. The final trench contact **200** has a relatively narrower width than trench contacts **200** formed using conventional processes, thereby reducing the likelihood of CTG shorts. And the additional layer of insulation between the gate electrodes **102** and the trench contact **200** reduces parasitic capacitance.

FIGS. **8**D to **8**F illustrate the fabrication of contact sidewall spacers **804** when the contact is misaligned. FIG. **8**D illustrates a misaligned contact trench opening **800** that has been etched through ILD layers **110***a* and **110***b* down to the diffusion region **106**. The insulating-cap layer **300** protects the metal gate electrode **102** from being exposed during this etching process, in accordance with an implementation of the invention. Also shown in FIG. **8**D is a silicide layer **802** that has been formed at the bottom of the contact trench opening **800**. Fabrication processes for the silicide layer **802** were provided above.

FIG. **8**E illustrates a pair of contact sidewall spacers **804** that are formed along the sidewalls of the contact trench opening **800**, in accordance with an implementation of the invention. The contact sidewall spacers **804** may be formed by depositing and etching a conformal layer of an insulating material, as explained above.

A metal deposition process is then carried out to fill the contact trench opening **800** and form the trench contact **200**, as shown in FIG. **8**F. Here again, the contact sidewall spacers **804** provide an additional layer of protection between the gate electrodes **102** and the trench contact **200**. The contact sidewall spacers **804** provide more separation between the final trench contact **200** and the metal gate electrodes **102**, thereby reducing the likelihood of CTG shorts. And the additional layer of insulation between the gate electrodes **102** and the trench contact **200** reduces parasitic capacitance.

FIGS. **9**A to **9**D illustrate another process for forming an insulating-cap layer in accordance with an implementation of the invention. FIG. **9**A illustrates two MOS transistors having metal gate electrodes **102** and gate dielectric layer **104**. The gate electrode layer **102** may include two or more layers (not illustrated), such as a workfunction metal layer and a fill metal layer. Although the gate dielectric layer **104** shown corresponds to a replacement-metal gate process, the following process may also be used with transistors formed using a gate-first approach.

A metal-cap **900** is formed atop the metal gate electrode **102**, as shown in FIG. **9**A. In accordance with implementations of the invention, the metal-cap **900** is formed using a selective deposition process. Some selective deposition processes include, but are not limited to, electroless plating and chemical vapor deposition. Metals that may be selectively deposited include, but are not limited to, cobalt, nickel, platinum, copper, polysilicon, tungsten, palladium, silver, gold, and other noble metals. As will be appreciated by those of skill in the art, the choice of whether an electroless process or a CVD process is used will depend on the composition of the metal gate electrode **102** and the specific metal that is used in the metal-cap **900**. In one example, if the top portion of the metal gate electrode **102** consists of copper metal, then cobalt metal can be electrolessly deposited on the copper. In another example, tungsten or polysilicon can be deposited by CVD on almost any metal that is used in the metal gate electrode **102**. In another example, if the top portion of the metal gate electrode **102** consists of a noble metal, then most metals may be

deposited using an electroless process on the noble metal. As will be appreciated by those of ordinary skill in the art, in general, electroless processes require a noble metal for both the substrate metal and the metal to be deposited. Therefore combinations of metals such as cobalt, nickel, copper, platinum, palladium, gold, and silver are possible.

Moving to FIG. **9**B, an ILD layer **902** is blanket deposited over the ILD **110***a* and the metal-caps **900**. A CMP process is then used to plagiarize both the ILD layer **902** and the metal-caps **900** and cause their top surfaces to be substantially even. This is done to expose the top surface of the metal-caps **900** after the ILD deposition.

Next, as shown in FIG. **9**C, an etching process is used to remove the metal-caps **900** from within the ILD layer **902**. In one implementation, a wet etch chemistry may be applied to remove the metal-caps **900**. In accordance with implementations of the invention, the etch chemistry that is used must be selective to both the ILD layer **902** and the metal gate electrode **102**. This enables the metal-caps **900** to be removed with minimal impact to the ILD layer **902** and the metal gate electrode **102**. The removal of the metal-caps **900** yields voids **904** within the ILD layer **902**.

Moving to FIG. **9**D, an insulating layer, such as a silicon nitride layer, may be deposited and planarized to fill in the voids **904**, thereby forming self-aligned insulating-cap layers **906**. This insulating layer is generally deposited as a blanket layer that fills the voids **904** and covers the ILD layer **902**. A planarization process is then used to remove any excess material that is outside of the voids **904**. This confines the insulating material to the voids **904**, thereby forming insulating-cap layers **906**. The insulator-cap layers **906** may be formed of materials that include, but are not limited to, silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, other nitride materials, other carbide materials, aluminum oxide, other oxide materials, other metal oxides, and low-k dielectric materials. The only constraint is that the material used in the insulator-cap layers **906** be dissimilar to the material used in the ILD layer **902**.

FIGS. **10**A to **10**G illustrate a process for forming a self-aligned metal stud atop the trench contact **200** and a pair of insulating spacers that further insulate the metal stud from the metal gate electrodes **102**, in accordance with an implementation of the invention. FIG. **10**A illustrates two MOS transistors having metal gate electrodes **102** and gate dielectric layer **104**. A trench contact **200** is formed between the two MOS transistors.

A metal-cap **900** is formed atop the trench contact **200**, as shown in FIG. **10**A. In accordance with implementations of the invention, the metal-cap **900** is formed using a selective deposition process. As noted above, selective deposition processes include, but are not limited to, electroless plating and chemical vapor deposition. The same metals and processes described above for use with the metal gate electrode **102** may also be used here with the trench contact **200**. The selective deposition process used and the metal used in the metal-cap **900** will depend on the metal that is used in the trench contact **200**.

In accordance with implementations of the invention, a selective deposition process is chosen that will deposit metal on only the trench contact **200** and not on the metal gate electrode **102**. This can be accomplished by using different types of metals in the trench contact **200** and the metal gate electrode **102**. For example, if aluminum is used in the metal gate electrode **102** and a noble metal is used in the trench contact **200**, then a selective deposition process can be used to deposit the metal-cap **900** on only the noble metal in the trench contact **200**. The same combinations of noble metals described above will work here as well. In some implementations of the invention, when an active metal such as aluminum, tungsten, molybdenum, titanium, tantalum, titanium nitride, or polysilicon is used in the metal gate electrode **102**, then a noble metal such as cobalt, nickel, copper, platinum, palladium, gold, and silver may be used in the trench contact **200**.

Moving to FIG. **10**B, an ILD layer **902** is blanket deposited over the ILD **110***a* and the metal-cap **900**. A CMP process is then used to planarize both the ILD layer **902** and the metal-cap **900** and cause their top surfaces to be substantially even. This is done to expose the top surface of the metal-cap **900** after the ILD deposition.

Next, as shown in FIG. **10**C, an etching process is used to remove just the metal-cap **900** from within the ILD layer **902**. The etch chemistry that is used must be selective to both the ILD layer **902** and the trench contact **200**. This enables the metal-cap **900** to be removed with minimal impact to the ILD layer **902** and the trench contact **200**. The removal of the metal-cap **900** yields a void **904** within the ILD layer **902**.

Moving to FIG. **10**D, an insulating layer **906** may be blanket deposited over the ILD layer **902** and within the void **904**. The insulating layer **906** may be formed of materials that include, but are not limited to, silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, other nitride materials, other carbide materials, aluminum oxide, other oxide materials, other metal oxides, and low-k dielectric materials, including materials that are the same or similar to the material used in the ILD layer **902**.

Next, an etching process, such as an anisotropic etching process is applied to etch down the insulating layer **906** and form spacers **1000**. This is shown in FIG. **10**E. The etching process also creates a trench **1002** between the two spacers **1000**.

Moving to FIG. **10**F, a metal deposition process is used to deposit a self-aligned metal stud **1004** in the trench **1002** between the spacers **1000** and atop the trench contact **200**. In some implementations this metal deposition process may be another selective deposition process, while in other implementations this metal deposition process need not be a selective process. Finally, as shown in FIG. **10**G, an insulating layer may be deposited and planarized to form an ILD layer **1006**. The top of the metal stud **1004** is also planarized to be even with the ILD layer **1006**. In accordance with implementations of the invention, the self aligned metal stud **1004** is prevented from shorting to the gate by the spacers **1000**.

Thus, implementations of the invention are described here that form etch stop structures that are self aligned to the gate, preventing the contact etch from exposing the gate electrode to cause shorting between the gate and contact. A contact to gate short is prevented even in the case of the contact pattern overlaying the gate electrode. Implementations of the invention also address problems such as parasitic capacitance between trench contacts and gate electrodes, dielectric breakdown or direct shorts from contact to gate, and degradation of contact silicide during gate contact patterning.

Accordingly, the use of an insulator-cap layer enables self-aligned contacts, which offer a robust manufacturable process. The invention allows initial patterning of wider contacts which is more robust to patterning limitations. The wider contacts are also desirable for a silicide-through-contact process flow. Not only does this eliminate a major yield limiter in contact-to-gate shorts, but it also alleviates major constraints for contact patterning and allows for more variability. From a lithography perspective, the use of an insulator-cap layer increases the registration window and allows for more critical dimension variability. From an etch perspective, the use of an

insulator-cap layer makes the fabrication process for MOS transistors more tolerant to different profiles, different critical dimensions, and over-etching of the ILD during trench contact formation.

The above description of illustrated implementations of the invention, including what is described in the Abstract, is not intended to be exhaustive or to limit the invention to the precise forms disclosed. While specific implementations of, and examples for, the invention are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the invention, as those skilled in the relevant art will recognize.

These modifications may be made to the invention in light of the above detailed description. The terms used in the following claims should not be construed to limit the invention to the specific implementations disclosed in the specification and the claims. Rather, the scope of the invention is to be determined entirely by the following claims, which are to be construed in accordance with established doctrines of claim interpretation.

The invention claimed is:

1. A transistor comprising:

a substrate;

a pair of spacers on the substrate;

a gate dielectric layer on the substrate and between the pair of spacers;

a gate electrode layer on the gate dielectric layer and between the pair of spacers, the gate dielectric layer on the surface of the substrate and between the gate electrode and the pair of spacers;

an insulating cap layer on the gate electrode layer and between the pair of spacers and directly on said gate dielectric layer between the gate electrode layer and the pair of spacers; and

a pair of diffusion regions adjacent to the pair of spacers.

2. The transistor of claim **1**, wherein a height of the combination of the gate dielectric layer, the gate electrode layer, and the insulating cap layer does not exceed a height of the pair of spacers.

3. The transistor of claim **1**, wherein the insulating cap layer comprises silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, or aluminum oxide.

4. The transistor of claim **1**, wherein the insulating cap layer comprises a nitride material, a carbide material, an oxide material, a metal oxide material, or a low-k dielectric material.

5. The transistor of claim **1**, wherein the insulating cap layer comprises boron nitride or boron carbide.

6. The transistor of claim **1**, wherein the insulating cap layer comprises a low-k dielectric material doped with one or more of carbon, nitrogen, and hydrogen.

7. A transistor comprising:

a substrate;

a pair of spacers on the substrate;

a gate dielectric layer on the substrate and between the pair of spacers;

a gate electrode layer on the gate dielectric layer and between the pair of spacers, the gate dielectric layer on the surface of the substrate and between the gate electrode layer and the pair of spacers;

an insulating cap layer sited atop the gate electrode layer that extends laterally over the top surfaces of the pair of spacers and directly on said gate dielectric layer between the gate electrode layer and the pair of spacers;

an interlayer dielectric layer adjacent to said pair of spacers and said insulating cap layer and having a top surface substantially planar with a top surface of said insulating cap layers and

a pair of diffusion regions adjacent to the pair of spacers.

8. The transistor of claim **7**, wherein the insulating cap layer comprises silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, or aluminum oxide.

9. The transistor of claim **7**, wherein the insulating cap layer comprises a nitride material, a carbide material, an oxide material, a metal oxide material, or a low-k dielectric material.

10. The transistor of claim **7**, wherein the insulating cap layer comprises boron nitride or boron carbide.

11. The transistor of claim **7**, wherein the insulating cap layer comprises a low-k dielectric material doped with one or more of carbon, nitrogen, and hydrogen.

12. The transistor of claim **7**, wherein the pair of spacers comprises silicon oxynitride, carbon-doped silicon oxynitride, or a low-k dielectric material.

13. A transistor comprising:

a substrate;

a pair of spacers on the substrate;

a gate dielectric layer on the substrate and between the pair of spacers;

a gate electrode layer having a stepped profile on the gate dielectric layer, wherein a middle portion of the gate electrode layer has a relatively larger height than side portions of the gate electrode layer wherein the gate dielectric layer is formed along the surface of the substrate and along sidewalls of the pair of spacers and wherein said gate dielectric layer along the pair of spacers has a height above said middle portion of said gate electrode layer;

an insulating cap layer on the gate electrode layer and between the pair of spacers; and

a pair of diffusion regions adjacent to the pair of spacers.

14. The transistor of claim **13**, wherein the insulating cap layer comprises silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, or aluminum oxide.

15. The transistor of claim **13**, wherein the insulating cap layer comprises a nitride material, a carbide material, an oxide material, a metal oxide material, or a low-k dielectric material.

16. The transistor of claim **13**, wherein the insulating cap layer comprises boron nitride or boron carbide.

17. The transistor of claim **13**, wherein the insulating cap layer comprises a low-k dielectric material doped with one or more of carbon, nitrogen, and hydrogen.

18. The transistor of claim **13**, wherein the gate electrode layer comprises two metal layers, wherein a first metal layer comprises the middle portion of the gate electrode layer having the relatively larger height and a second metal layer comprises the side portions of the gate electrode layer.

19. A transistor comprising:

a substrate;

a pair of spacers on the substrate;

a gate dielectric layer on the substrate and between the pair of spacers;

a gate electrode layer having a stepped profile on the gate dielectric layer, wherein a middle portion of the gate electrode layer has a relatively larger height than side portions of the gate electrode layer wherein the gate

dielectric layer is formed along the surface of the substrate and of between the gate electrode and the pair of spacers;

an insulating cap layer sited atop the gate electrode layer that extends laterally over the top surfaces of the pair of spacers and directly on said gate dielectric layer between the gate electrode layer and the pair of spacers;

an interlayer dielectric layer adjacent to said pair of spacers and said insulating cap layer and having a top surface substantially planar with a top surface of said insulating cap layer; and

a pair of diffusion regions adjacent to the pair of spacers.

20. The transistor of claim **19**, wherein the insulating cap layer comprises silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, or aluminum oxide.

21. The transistor of claim **19**, wherein the insulating cap layer comprises a nitride material, a carbide material, an oxide material, a metal oxide material, or a low-k dielectric material.

22. The transistor of claim **19**, wherein the insulating cap layer comprises boron nitride or boron carbide.

23. The transistor of claim **19**, wherein the insulating cap layer comprises a low-k dielectric material doped with one or more of carbon, nitrogen, and hydrogen.

24. The transistor of claim **19**, wherein the gate electrode layer comprises two metal layers, wherein a first metal layer comprises the middle portion of the gate electrode layer having the relatively larger height and a second metal layer comprises the side portions of the gate electrode layer.

25. A transistor comprising:

a substrate;

a pair of spacers on the substrate;

a gate dielectric layer on the substrate and between the pair of spacers;

a gate electrode layer on the gate dielectric layer and between the pair of spacers;

an insulating cap layer on the gate electrode layer and between the pair of spacers;

wherein said gate dielectric layer is between said pair of spacers and said gate electrode layer and wherein said gate dielectric layer is also between said pair of spacers and said insulating cap layer wherein the gate dielectric layer between said insulating cap layer and said pair of spacers has a top surface substantially planar with a top surface of said insulating cap layer; and

a pair of diffusion regions adjacent to the pair of spacers.

26. The transistor of claim **25**, wherein a height of the combination of the gate dielectric layer, the gate electrode layer, and the insulating cap layer does not exceed a height of the pair of spacers.

27. The transistor of claim **25**, wherein the insulating cap layer comprises silicon nitride, silicon oxide, silicon carbide, silicon nitride doped with carbon, silicon oxynitride, or aluminum oxide.

28. The transistor of claim **25**, wherein the insulating cap layer comprises a nitride material, a carbide material, an oxide material, a metal oxide material, or a low-k dielectric material.

29. The transistor of claim **25**, wherein the insulating cap layer comprises boron nitride or boron carbide.

30. The transistor of claim **25**, wherein the insulating cap layer comprises a low-k dielectric material doped with one or more of carbon, nitrogen, and hydrogen.

* * * * *