IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

MONTEREY RESEARCH, LLC, Plaintiff, v. QUALCOMM INCORPORATED, QUALCOMM TECHNOLOGIES, INC., and QUALCOMM CDMA TECHNOLOGIES ASIA-PACIFIC PTE LTD., Defendants.	)	C.A. No. 19-cv-2083-CFC **JURY TRIAL DEMANDED**

FIRST AMENDED COMPLAINT FOR PATENT INFRINGEMENT

Plaintiff Monterey Research, LLC ("Monterey"), for its First Amended Complaint for Patent Infringement against Defendants Qualcomm Incorporated ("Qualcomm Inc."), Qualcomm Technologies, Inc. ("QTI"), and Qualcomm CDMA Technologies Asia-Pacific Pte Ltd. ("QCTAP") (collectively, "Qualcomm" or "Qualcomm Defendants") alleges as follows:

INTRODUCTION

1. Monterey is an intellectual property and technology licensing company. Monterey's patent portfolio comprises over 2,700 active and pending patents worldwide, including approximately 2,000 active United States patents. Monterey's patent portfolio stems from technology developed from a number of leading high-technology companies, including Cypress Semiconductor Corporation, Advanced Micro Devices, Fujitsu, NVX Corporation, Ramtron, and Spansion. Those companies developed key innovations that have greatly enhanced the capabilities of computer systems, increased electronic device processing power, and reduced electronic device power consumption. Among other things, those inventions produced significant technological advances, including smaller, faster, and more efficient semiconductors and integrated circuits.

2. The Qualcomm Defendants, jointly and severally, have infringed and continue to infringe Monterey's patents. Moreover, despite Monterey notifying them of infringement, the Qualcomm Defendants have thus far refused to license those patents and, instead, have continued to make, use, sell, offer to sell, and/or import Monterey's intellectual property within the United States without Monterey's permission.

NATURE OF THE CASE

3. This action arises under 35 U.S.C. § 271 for Qualcomm's infringement of Monterey's United States Patent Nos. 6,459,625 ("the '625 patent"); 6,534,805 ("the '805 patent"); 6,642,573 ("the '573 patent"); 6,651,134 ("the '134 patent"); 6,680,516 ("the '516 patent"); 6,765,407 ("the '407 patent"); 7,572,727 ("the '727 patent"); and 7,977,797 ("the '797 patent") (collectively, "the Patents-in-Suit").

THE PARTIES

4. Plaintiff Monterey is a Delaware limited liability company with offices in New Jersey and California. Monterey maintains a registered agent for service in Delaware: Intertrust Corporate Services Delaware Ltd. located at 200 Bellevue Parkway, Suite 210, Wilmington, Delaware 19808.

5. Defendant Qualcomm Inc. is a Delaware corporation with a principal place of business at 5775 Morehouse Dr., San Diego, California, 92121. Qualcomm Inc. is a publicly traded company and is the parent corporation of defendants QTI and QCTAP. Qualcomm Inc. may be served through its registered agent for service, The Prentice-Hall Corporation System, Inc., 251 Little Falls Drive, Wilmington, Delaware 19808.

6. Defendant QTI is a Delaware corporation with a principal place of business at 5775 Morehouse Dr., San Diego, California, 92121. QTI is a wholly-owned subsidiary of Qualcomm Inc. Qualcomm Inc.'s semiconductor research and engineering business is conducted wholly or

in part through the actions of QTI. Qualcomm Inc. controls and directs the actions of QTI, and therefore both directs QTI to infringe and itself infringes Monterey's patents. QTI may be served through its registered agent for service, Corporation Service Company, 251 Little Falls Drive, Wilmington, Delaware 19808.

7. Defendant QCTAP is a corporation organized under the laws of Singapore, with corporate offices at 6 Serangoon North Avenue 5, #03-04, Singapore 554910, Singapore. Defendant QCTAP is a wholly-owned subsidiary of Qualcomm Inc. QCTAP is responsible, among other things, for accepting orders and sending invoices to certain customers in the United States for Qualcomm products.

8. Qualcomm Inc. exercises control over QTI and QCTAP, and acts collectively with QTI and QCTAP to infringe Monterey's patents by making, using, selling, offering for sale, and/or importing products (including importing products made by a patented process) throughout the United States, including within this District. Qualcomm's customers incorporate those products into downstream products that are made, used, sold, offered for sale, and/or imported throughout the United States, including within this District. Those downstream products include, but are not limited to, smartphones, tablets, televisions, smartwatches, and other products that include Qualcomm semiconductor devices and integrated circuits.

JURISDICTION AND VENUE

9. This Court has jurisdiction over the subject matter of this action under 28 U.S.C. §§ 1331 and 1338(a) at least because this action arises under the patent laws of the United States, including 35 U.S.C. § 271 *et seq*.

10. Personal jurisdiction exists over each Qualcomm Defendant.

11. Personal jurisdiction exists over Qualcomm Inc. and QTI at least because each is a Delaware corporation organized under the laws of the State of Delaware. Each also has a

registered agent for service of process in Delaware. In addition, Qualcomm Inc. and QTI have each committed, aided, abetted, contributed to and/or participated in the commission of acts of infringement giving rise to this action within the State of Delaware by, *inter alia*, directly and/or indirectly making, using, selling, offering for sale, importing products and/or practicing methods that practice one or more claims of the Patents-in-Suit. Furthermore, Qualcomm Inc. and QTI have transacted and conducted business in the State of Delaware and with Delaware residents by making, using, selling, offering to sell, and/or importing (including importing products made by a patented process) products and instrumentalities that practice one or more claims of the Patents-in-Suit. Among other things, Qualcomm Inc. and QTI, directly and/or through intermediaries, use, sell, ship, distribute, import into, offer for sale, and/or advertise or otherwise promote their products throughout the United States, including in the State of Delaware. *See, e.g.*, www.qualcomm.com. At least for those reasons, Qualcomm Inc. and QTI have the requisite minimum contacts within the forum such that the exercise of jurisdiction over Qualcomm Inc. and QTI would not offend traditional notions of fair play and substantial justice.

12. Personal jurisdiction exists over QCTAP at least because it has committed, aided, abetted, contributed to and/or participated in the commission of acts of infringement giving rise to this action within the State of Delaware by, *inter alia*, directly and/or indirectly making, using, selling, offering for sale, importing products and/or practicing methods that practice one or more claims of the Patents-in-Suit. Furthermore, QCTAP transacted and conducted business in the State of Delaware and with Delaware residents with respect to the products and instrumentalities accused of infringing the Patents-in-Suit. Among other things, QCTAP, directly and/or through intermediaries, uses, sells, ships, distributes, imports into, offers for sale, and/or advertises or otherwise promotes its products throughout the United States, including in the State of Delaware.

See, e.g., www.qualcomm.com. For example, QCTAP develops products for sale in the United States, including in the State of Delaware, and tests and verifies products developed in the United States before selling them in the United States, including in the State of Delaware. *See, e.g.*, https://www.qualcomm.com/news/releases/2008/10/30/qualcomm-opens-asia-pacific-test-center-excellence-singapore. As an additional example, QCTAP contracts with and is responsible for accepting orders and sending invoices to customers in the United States. *See, e.g.*, *Tessera Inc. v. Motorola, Inc. et al*, No. 12-cv-692, slip op. at 3 (N.D. Ca. Aug. 7, 2013). At least for those reasons, QCTAP has the requisite minimum contacts within the forum such that the exercise of jurisdiction over QCTAP would not offend traditional notions of fair play and substantial justice.

13. Venue is proper in this Court pursuant to 28 U.S.C. §§ 1391(b) and (c) and 1400(b). Qualcomm Inc. and QTI reside in this district and have committed acts of infringement in this district. Venue is proper with respect to QCTAP at least because QCTAP is a foreign corporation, has committed acts of infringement in this district, and venue is proper in any district in which QCTAP is subject to personal jurisdiction. Venue is further proper based on the facts alleged in the preceding paragraphs, which Monterey incorporates by reference as if fully set forth herein.

THE PATENTS-IN-SUIT

14. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

A. U.S. Patent No. 6,459,625

15. The ’625 patent, titled “Three Metal Process for Optimizing Layout Density,” was duly and properly issued by the USPTO on October 1, 2002. A true and correct copy of the ’625 patent is attached hereto as Exhibit A.

16. Monterey is the owner and assignee of the ’625 patent; owns all right, title, and interest in the ’625 patent; and holds the right to sue for and recover damages for infringement

thereof, including past infringement.

B. U.S. Patent No. 6,534,805

17. The ’805 patent, titled “SRAM Cell Design,” was duly and properly issued by the USPTO on March 18, 2003. On October 14, 2014, the USPTO issued an Ex Parte Reexamination Certificate for the ’805 patent, which confirmed the patentability of the ’805 patent. A true and correct copy of the ’805 patent and the Ex Parte Reexamination Certificate for the ’805 patent is attached hereto as Exhibit B.

18. Monterey is the owner and assignee of the ’805 patent; owns all right, title, and interest in the ’805 patent; and holds the right to sue and recover damages for infringement thereof, including past infringement.

C. U.S. Patent No. 6,642,573

19. The ’573 patent, titled “Use of High-K Dielectric Material in Modified ONO Structure for Semiconductor Devices,” was duly and properly issued by the USPTO on November 4, 2003. A true and correct copy of the ’573 patent is attached hereto as Exhibit C.

20. Monterey is the owner and assignee of the ’573 patent; owns all right, title, and interest in the ’573 patent; and holds the right to sue and recover damages for infringement thereof, including past infringement.

D. U.S. Patent No. 6,651,134

21. The ’134 patent, titled “Memory Device with Fixed Length Non Interruptible Burst,” was duly and properly issued by the USPTO on November 18, 2003. A true and correct copy of the ’134 patent is attached hereto as Exhibit D.

22. Monterey is the owner and assignee of the ’134 patent; owns all right, title, and interest in the ’134 patent; and holds the right to sue and recover damages for infringement thereof, including past infringement.

E. U.S. Patent No. 6,680,516

23. The '516 patent, titled "Controlled Thickness Gate Stack," was duly and properly issued by the United States Patent and Trademark Office ("USPTO") on January 20, 2004. On December 12, 2006, the USPTO issued a Certificate of Correction for the '516 patent. A true and correct copy of the '516 patent and the Certificate of Correction is attached hereto as Exhibit E.

24. Monterey is the owner and assignee of the '516 patent; owns all right, title, and interest in the '516 patent; and holds the right to sue and recover damages for infringement thereof, including past infringement.

F. U.S. Patent No. 6,765,407

25. The '407 patent, titled "Digital Configurable Macro Architecture," was duly and properly issued by the USPTO on July 20, 2004. A true and correct copy of the '407 patent is attached hereto as Exhibit F.

26. Monterey is the owner and assignee of the '407 patent; owns all right, title, and interest in the '407 patent; and holds the right to sue and recover damages for infringement thereof, including past infringement.

G. U.S. Patent No. 7,572,727

27. The '727 patent, titled "Semiconductor Formation Method that Utilizes Multiple Etch Stop Layers," was duly and properly issued by the USPTO on August 11, 2009. A true and correct copy of the '727 patent is attached hereto as Exhibit G.

28. Monterey is the owner and assignee of the '727 patent; owns all right, title, and interest in the '727 patent; and holds the right to sue and recover damages for infringement thereof, including past infringement.

H. U.S. Patent No. 7,977,797

29. The '797 patent, titled "Integrated Circuit with Contact Region and Multiple Etch

Stop Insulation Layer," was duly and properly issued by the USPTO on July 12, 2011. A true and correct copy of the '797 patent is attached hereto as Exhibit H.

30. Monterey is the owner and assignee of the '797 patent; owns all right, title, and interest in the '797 patent; and holds the right to sue and recover damages for infringement thereof, including past infringement.

FACTUAL BACKGROUND

31. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

32. The Patents-in-Suit stem from the research and design of innovative and proprietary technology developed by leading high-technology companies, including Cypress Semiconductor Corporation ("Cypress").[1] Cypress is an American multinational company and pioneer of cutting-edge semiconductor technology. Founded in 1982, Cypress has made substantial investments in researching, developing, and manufacturing high-quality semiconductor devices, integrated circuits, and products containing the same.

33. The Patents-in-Suit are directed to inventive technology relating to semiconductor devices, integrated circuits, and/or products containing the same.

34. The Qualcomm Defendants work closely with their customers, OEMs, foundry suppliers, distributors, and/or other third parties to make, use, sell, offer to sell, and/or import semiconductor devices, integrated circuits, and/or products containing the same. Among other things, the Qualcomm Defendants optimize their manufacturing process for their customers and optimize their products for integration into downstream products. The Qualcomm Defendants' affirmative acts in furtherance of the manufacture, use, sale, offer to sell, and importation of their

[1] Other leading high-technology companies that contributed to inventions disclosed in the Patents-in-Suit include Advanced Micro Devices, Inc. ("AMD") and Spansion LLC ("Spansion").

products in and/or into the United States include, but are not limited to, any one or combination of: (i) designing specifications for manufacture of their products; (ii) collaborating on, encouraging, and/or funding the development of processes for the manufacture of their products; (iii) soliciting and/or sourcing the manufacture of their products; (iv) licensing, developing, and/or transferring technology and know-how to enable the manufacture of their products; (v) enabling and encouraging the use, sale, or importation of their products in the United States; and (vi) advertising their products and/or downstream products incorporating them in the United States.

35. The Qualcomm Defendants also provide marketing and/or technical support services for their products from their facilities in the United States. For example, Qualcomm maintains a website that advertises their products, including identifying the applications for which they can be used and specifications for their products. *See, e.g.*, www.qualcomm.com. Qualcomm's publicly-available website also contains user manuals, product documentation, and other materials related to their products. *See, e.g.*, www.qualcomm.com. For example, Qualcomm provides development content for specific chip products and applications; catalogs of hardware, software, and tools documentation; relevant support articles; various software code and tools; and case-specific technical assistance. *See, e.g.*, www.qualcomm.com.

QUALCOMM'S PRE-SUIT KNOWLEDGE OF MONTEREY'S PATENTS AND CHARGE OF INFRINGEMENT

36. Before filing this action, Monterey, through its agent IPValue Management, Inc. ("IPValue"), notified Qualcomm about the Patents-in-Suit and Qualcomm's infringement thereof. Among other things, Monterey, through its agent IPValue, identified the Patents-in-Suit to Qualcomm; alleged that Qualcomm infringed the Patents-in-Suit, including identifying exemplary infringing products; and offered to license the Patents-in-Suit to Qualcomm. For example:

a. On January 31, 2018, Monterey sent a letter to Qualcomm, notifying Qualcomm of their infringement of certain Monterey patents, including the '625, '516, '805, '407, '727, and '797 patents. Among other things, Monterey identified representative Qualcomm products that utilize those patents, expressly charged that Qualcomm and their customers infringed those patents, and explained that Qualcomm required a license from Monterey. Monterey identified IPValue as Monterey's appointed agent and requested a meeting with Qualcomm.

b. On May 14, 2018, IPValue met in-person with Qualcomm and presented Qualcomm an overview of Monterey's patent portfolio. Among other things, IPValue further explained Monterey's patent portfolio's relevance to Qualcomm and further explained that Qualcomm required a license from Monterey.

c. On July 17, 2018, IPValue again met in-person with Qualcomm and presented Qualcomm with detailed infringement claim charts of certain Monterey patents. Among other things, IPValue's presentations identified specific Monterey patents including the '625, '516, '805, '407, '727, and '797 patents (as well as exemplary patent claims); identified representative Qualcomm products that utilize those patents; identified where every element of each of those exemplary patent claims was found in the representative Qualcomm products; expressly charged that Qualcomm and their customers infringed those patents; and explained that Qualcomm required a license from Monterey.

d. On July 24, 2018, IPValue, on behalf of Monterey, emailed copies of those infringement claim charts to Qualcomm.

e. On October 9, 2018, IPValue met a third time in-person with Qualcomm and presented Qualcomm with additional infringement claim charts of certain Monterey patents. Among other things, IPValue's presentation identified specific Monterey patents including the

'573 and '134 patents (as well as exemplary patent claims); identified representative Qualcomm products that utilize those patents; identified where every element of each of those exemplary patent claims was found in the representative Qualcomm products; expressly charged that Qualcomm and their customers infringed those patents; and explained that Qualcomm required a license from Monterey.

f. On October 16, 2018, IPValue, on behalf of Monterey, emailed Qualcomm copies of the '573 infringement claim chart.

g. On October 17, 2018, IPValue, on behalf of Monterey, emailed Qualcomm copies of the '134 infringement claim chart.

h. On November 13, 2018, IPValue met for a fourth time in-person with Qualcomm, and once again offered to license the Patents-in-Suit to Qualcomm.

i. Monterey continued to contact Qualcomm in the succeeding months to no avail.

j. Despite the numerous meetings and related prior and subsequent communications, at no time during any of those meetings, or at any time prior to Monterey's filing of this Complaint, did Qualcomm deny infringing any element of any claim of the Patents-in-Suit, nor did Qualcomm identify any alleged prior art to any of the Patents-in-Suit.

37. Despite Monterey's repeated efforts—which have continued for well over a year—Qualcomm still has not engaged in any meaningful discussions to end their infringement of the Patents-in-Suit and has not taken a license to them. Instead, Qualcomm continues to knowingly, intentionally, and willfully infringe Monterey's patents directly, contributorily, and by inducement, to obtain their significant benefits without a license from Monterey.

COUNT ONE
INFRINGEMENT OF THE '625 PATENT

38. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

39. Monterey is the assignee and lawful owner of all right, title, and interest in and to the ’625 patent.

40. The ’625 patent is valid and enforceable.

41. The ’625 patent is directed to memory device layout, and particularly to systems for optimizing layout density in the periphery area of a memory device using a three-metal or more interconnect process.

42. The ’625 patent explains that metallization can involve depositing a thin film of conductive metal on a memory device such that the electrical components are formed and electrically connected with the conductive metal. A periphery area of a memory device can include, for example, a plurality of electrical components such as transistors, resistors, capacitors and diodes formed in the silicon substrate during fabrication. Some types of previously known memory used a two-metal layer metallization process to electrically connect the electrical components in the periphery area of the memory. A problem with this two-metal layer metallization method was layout area consumed by the periphery area. The layout area for the traditional electrical connection systems and methods increased the size of the periphery area on the memory. The area on the memory that is not consumed by the periphery area can be, for example, reserved for the core cell area, allowing more core memory cells to be fabricated on the memory. It is therefore desirable to minimize the amount of periphery area consumed, thereby increasing the amount of information stored in the memory.

43. The ’625 patent teaches, among other things, how to selectively place and electrically connect a plurality of electrical components to form sub-circuits and selectively

electrically connect the sub-circuits, including using three or more metal layers, resulting in, among other things, minimizing the layout area of the sub-circuits in the periphery area.

44. Qualcomm products use three or more metal layers in their memory device's periphery area. This has enabled Qualcomm to, among other things, decrease their memory device's periphery area. Memory devices containing an infringing periphery area are found integrated in Qualcomm products, including their system-on-chip semiconductor products. Specifically, at least the Qualcomm products which are manufactured with a 7 nm process node have infringing three metal layer or more metallization in their memory device's periphery area and other infringing features that use the technology of the '625 patent.

45. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the '625 patent under 35 U.S.C. § 271(a), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing in or into the United States without authorization products covered by one or more claims of the '625 patent, including, but not limited to, products that use three or more metal layers in their integrated memory device's periphery area, such as the SM8150 semiconductor device and other products in the Snapdragon 855 series product family; other Qualcomm 7 nm process node semiconductor devices, integrated circuits, and products; and all other semiconductor devices, integrated circuits, and products with similar integrated memory devices containing a periphery area which uses the infringing technology ("the Accused '625 Products").

46. As one non-limiting example, Qualcomm infringes claim 10 of the '625 patent. For example, the SM8150 Snapdragon 855 semiconductor device contains:

a. a plurality of sub-circuits in a periphery area of a memory device (e.g., sub-circuits in the periphery of the SRAM of the SM8150 Snapdragon 855), wherein each sub-circuit

includes at least one electrical circuit with a plurality of circuit components (e.g., electrical circuit with a plurality of circuit components in the periphery of the SRAM of the SM8150 Snapdragon 855);

b. a first metal interconnect layer that partially connects the circuit components, wherein first metal layer lines are oriented in substantially one direction (e.g., metal layer connecting circuit components in the periphery of the SRAM of the SM8150 Snapdragon 855);

c. a second metal interconnect layer that completes the connection of the circuit components, and where the second metal interconnect layer lines are fabricated substantially perpendicular to the first metal layer lines (e.g., metal layer perpendicular to the first that completes the connection of the circuit components in the periphery of the SRAM of the SM8150 Snapdragon 855); and

d. a third metal interconnect layer that connects the plurality of sub-circuits, wherein the third metal interconnect layer lines are fabricated substantially parallel to the first metal layer lines (e.g., metal layer substantially parallel to the first that connects the plurality of sub-circuits in the periphery of the SRAM of the SM8150 Snapdragon 855).

47. Claim 10 of the ’625 patent applies to each Accused ’625 Product at least because each of those products contains the same or similar three or more metal layers in their memory device’s periphery area as the Qualcomm SM8150 Snapdragon 855.

48. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

49. Qualcomm has known of the ’625 patent and their infringement of that patent since at least as early as January 31, 2018.

50. Qualcomm, knowing their products infringe the '625 patent and with the specific intent for others to infringe the '625 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the '625 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import in or into the United States without authorization the Accused '625 Products, as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or other third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm's website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm's products that infringe the '625 patent. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/prod_brief_qcom_sd855_0.pdf. Additional non-limiting examples include the materials found on Qualcomm's website at www.qualcomm.com/products/snapdragon-855-mobile-platform.

51. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '625 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '625 Products, which constitute a material part of the invention of the '625 patent, knowing the Accused '625 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use. *See, e.g.*, https://www.qualcomm.com/support and

https://www.qualcomm.com/system/files/document/files/prod_brief_qcom_sd855_0.pdf.

52. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

53. Qualcomm's infringement of the '625 patent has been knowing, deliberate, and willful, since at least as early as January 31, 2018, the date of Monterey's letter to Qualcomm and therefore the date on which Qualcomm knew of the '625 patent and that their conduct constituted and resulted in infringement of the '625 patent. Monterey continued to put Qualcomm on notice of the '625 patent and Qualcomm's infringement thereof, including without limitation through communications on July 17, 2018; July 24, 2018; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable '625 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or even though Qualcomm otherwise knew or should have known that their actions constituted an unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm's conduct in light of these circumstances is egregious. Qualcomm's knowing, deliberate, and willful infringement of the '625 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

COUNT TWO
INFRINGEMENT OF THE '805 PATENT

54. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

55. Monterey is the assignee and lawful owner of all right, title, and interest in and to the '805 patent.

56. The ’805 patent is valid and enforceable.

57. The ’805 patent is generally directed to static random access memory (“SRAM”) cell design, particularly to optimizing SRAM cell design using a simpler geometric layout.

58. As semiconductor structure size continued to shrink with time, one exemplary issue with the prior art of the ’805 patent was increased difficulties in manufacturing. Specifically, the then-existing memory cells contained complex geometric designs which required numerous processing steps and larger cell sizes. Generally, more processing steps lead to increased manufacturing costs and reduced profits.

59. The ’805 patent teaches, among other things, an improved memory cell layout which allows the features to be arranged in such a way as to minimize cell size. For example, the single local interconnect layer of the ’805 patent allows for a thinner product and fewer processing steps.

60. Qualcomm products use SRAM with a six-transistor (“6T”) and/or eight-transistor (“8T”) cell design. Qualcomm’s 6T and 8T SRAM contain a single local interconnect layer. This has resulted in, among other things, Qualcomm’s ability to decrease the size of their SRAM area and to decrease the number of manufacturing steps.

61. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the ’805 patent under 35 U.S.C. § 271(a), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing in or into the United States without authorization products covered by one or more claims of the ’805 patent, including, but not limited to, all Qualcomm devices incorporating SRAM with a 6T and/or 8T cell design, such as the MSM8974 semiconductor device and other products in the Snapdragon 800, Snapdragon S4 Play, Snapdragon S4 Plus, Snapdragon S4 Pro, Snapdragon 200, Snapdragon

205, Snapdragon 208, Snapdragon 210, Snapdragon 212, Snapdragon 400, Snapdragon 410, Snapdragon 412, Snapdragon 415, Snapdragon 425, Snapdragon 427, Snapdragon 430, Snapdragon 435, Snapdragon 429, Snapdragon 439, Snapdragon 450, Snapdragon 600, Snapdragon 610, Snapdragon 615, Snapdragon 616, Snapdragon 617, Snapdragon 625, Snapdragon 626, Snapdragon 650, Snapdragon 652, Snapdragon 653, Snapdragon 630, Snapdragon 636, Snapdragon 660, Snapdragon 632, Snapdragon 670, Snapdragon 675, Snapdragon 665, Snapdragon 710, Snapdragon 712, Snapdragon 730, Snapdragon 730G, Snapdragon 801, Snapdragon 805, Snapdragon 808, Snapdragon 810, Snapdragon 820, Snapdragon 821, Snapdragon 835, Snapdragon 845, Snapdragon 850, Snapdragon 855, and Snapdragon 8cx series product families; and all other semiconductor devices, integrated circuits, and products with similar infringing technology ("the Accused '805 Products").

62. As one non-limiting example, Qualcomm infringes claim 8 of the '805 patent. For example, the MSM8974 Snapdragon 800 semiconductor device contains:

a. a memory cell (e.g., SRAM cell of the MSM8974 Snapdragon 800) comprising a plurality of substantially oblong active regions (e.g., N-type and/or P-type diffusion areas of the MSM8974 Snapdragon 800) formed in a semiconductor substrate and arranged substantially in parallel with one another, and a plurality of substantially oblong local interconnects (e.g., structures formed at the polysilicon layer on top of the substrate of the MSM8974 Snapdragon 800) above said substrate that extend only partially across the memory cell and are arranged substantially in parallel with one another and substantially perpendicular to said active regions; and

b. a single local interconnect layer (e.g., metal 1 ("M1") layer of the MSM8974 Snapdragon 800) comprising local interconnects (e.g., structures formed at the M1

layer of the MSM8974 Snapdragon 800) corresponding to bitlines (e.g., those formed at the metal 2 ("M2") layer of the MSM8974 Snapdragon 800) and a global word-line (e.g., those formed at the metal 3 ("M3") layer of the MSM8974 Snapdragon 800).

63. Claim 8 of the '805 patent applies to each Accused '805 Product at least because each of those products contain the same or similar structures as the Qualcomm MSM8974 Snapdragon 800.

64. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

65. Qualcomm has known of the '805 patent and their infringement of that patent since at least as early as January 31, 2018.

66. Qualcomm, knowing their products infringe the '805 patent and with the specific intent for others to infringe the '805 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the '805 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import in or into the United States without authorization the Accused '805 Products, as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm's website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm's products that infringe the '805 patent. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/media/documents/files/snapdragon-800-product-brief.pdf.

Additional, non-limiting examples include the materials found on Qualcomm's websites at www.qualcomm.com/products/snapdragon-processors-800.

67. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '805 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '805 Products, which constitute a material part of the invention of the '805 patent, knowing the Accused '805 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/media/documents/files/snapdragon-800-product-brief.pdf.

68. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

69. Qualcomm's infringement of the '805 patent has been knowing, deliberate, and willful, since at least as early as January 31, 2018, the date of Monterey's letter to Qualcomm and therefore the date on which Qualcomm knew of the '805 patent and that their conduct constituted and resulted in infringement of the '805 patent. Monterey continued to put Qualcomm on notice of the '805 patent and Qualcomm's infringement thereof, including without limitation through communications on July 17, 2018 and July 24, 2018; through the complaint filed on November 1, 2019; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable '805 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or

even though Qualcomm otherwise knew or should have known that their actions constituted an unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm's conduct in light of these circumstances is egregious. Qualcomm's knowing, deliberate, and willful infringement of the '805 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

COUNT THREE
INFRINGEMENT OF THE '573 PATENT

70. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

71. Monterey is the assignee and lawful owner of all right, title, and interest in and to the '573 patent.

72. The '573 patent is valid and enforceable.

73. The '573 patent is generally directed to semiconductor devices including a modified oxide-nitride-oxide ("ONO") structure comprising a high-K dielectric material.

74. The '573 patent explains that as device dimensions continue to be reduced, the electrical thickness of the ONO layer must be reduced accordingly. Previously, that had been accomplished by scaling down the physical thickness of the ONO layer. However, as the ONO layer was made physically thinner, leakage current through the ONO layer could increase, and the charge trapping ability of the nitride layer could be reduced, limiting the scaling down of the total physical thickness of the ONO layer.

75. Consequently, the '573 patent describes, among other things, the use of a mid-K or a high-K dielectric material in a modified ONO structure having reduced dimensions without creation of interface states coming from contamination which could provide charge leakage paths and without sacrificing the charge trapping ability of the modified ONO structure. It thus provided

advantages such as (1) reduction of equivalent oxide thickness of ONO in next generation devices; (2) high-K film devices are expected to have improved data retention and reliability; and (3) the high-K dielectric material layer replacing one of both silicon dioxide layers allows fabrication of an ONO layer which is physically thicker, resulting in fewer charge leakage paths within the modified ONO structure.

76. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the ’573 patent under 35 U.S.C. § 271(a) and/or 35 U.S.C. § 271(g), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing (including importing products made by a patented process) in or into the United States without authorization products covered by one or more claims of the ’573 patent, including, but not limited to, products with high-k dielectrics, such as the MSM8994 semiconductor device and other products in the Snapdragon 810 and Snapdragon 808 series product families; other Qualcomm 20 nm process node semiconductor devices, integrated circuits, and products; and all other semiconductor devices, integrated circuits, and products with similar infringing technology (“the Accused ’573 Products”).

77. As one non-limiting example, Qualcomm infringes claim 1 of the ’573 patent. For example, the MSM8994 Snapdragon 810 semiconductor device comprises a semiconductor device comprising a modified ONO structure:

a. wherein the modified ONO structure comprises a bottom dielectric material layer, a nitride layer on the bottom dielectric material layer, and a top dielectric material layer on the nitride layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprises a composite dielectric material, wherein the composite dielectric material

comprises elements of at least one mid-K or high-K dielectric material (e.g., modified ONO structure of the MSM8994 Snapdragon 810);

b. wherein each mid-K or high-K dielectric material independently comprises at least one of hafnium oxide (HfO2), zirconium oxide (ZrO2), barium titanate (BaTiO3), cerium oxide (CeO2), lanthanum oxide (La2O3), lanthanum aluminum oxide (LaAlO3), lead titanate (PbTiO3), strontium titanate (SrTiO3), lead zirconate (PbZrO3), tungsten oxide (WO3), yttrium oxide (Y2O3), bismuth silicon oxide (Bi4Si2O12), barium strontium titanate (BST) (Ba1-xSrxTiO3), PMN (PbMgxNb1-xO3), PZT (PbZrxTi1-xO3), PZN (PbZnxNb1-xO3), and PST (PbScxTa1-xO3) (e.g., dielectric material of the MSM8994 Snapdragon 810 comprising hafnium oxide).

78. Claim 1 of the ’573 patent applies to each Accused ’573 Product at least because each of those products contain the same or similar structures as the Qualcomm MSM8994 Snapdragon 810.

79. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

80. Qualcomm has known of the ’573 patent and their infringement of that patent since at least as early as October 9, 2018.

81. Qualcomm, knowing their products infringe the ’573 patent and with the specific intent for others to infringe the ’573 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the ’573 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import (including import products made by a patented process) in or into the United States without authorization the Accused ’573

Products as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or other third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm's website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm's products that infringe the '573 patent. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf. Additional, non-limiting examples include the materials found on Qualcomm's websites at https://www.qualcomm.com/products/snapdragon-processors-810.

82. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '573 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '573 Products, which constitute a material part of the invention of the '573 patent, knowing the Accused '573 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf.

83. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

84. Qualcomm's infringement of the '573 patent has been knowing, deliberate, and willful, since at least as early as October 9, 2018, the date on which Qualcomm was presented with

the ’573 claim chart, knew of the ’573 patent, and knew that their conduct constituted and resulted in infringement of the ’573 patent. Monterey continued to put Qualcomm on notice of the ’573 patent and Qualcomm’s infringement thereof, including without limitation through communications on October 16, 2018; through the complaint filed on November 1, 2019; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable ’573 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or even though Qualcomm otherwise knew or should have known that their actions constituted an unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm’s conduct in light of these circumstances is egregious. Qualcomm’s knowing, deliberate, and willful infringement of the ’573 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

COUNT FOUR
INFRINGEMENT OF THE ’134 PATENT

85. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

86. Monterey is the assignee and lawful owner of all right, title, and interest in and to the ’134 patent.

87. The ’134 patent is valid and enforceable.

88. The ’134 patent generally concerns memory devices, and is more specifically related to non-interruptible burst read and write access features, as described in JEDEC standards JESD79-3F DDR3 SDRAM, JESD79-4A DDR4 SDRAM, JESD209-3 LPDDR3, JESD209-4 LPDDR4, and similar versions of the JEDEC DDRx standards.

89. The ’134 patent provides a faster and more efficient way for burst read and write access over conventional DRAM devices existing when the patent was filed in early 2000. A conventional DRAM may need an interrupt to perform data refreshes. Prior to the ’134 patent, DRAM memory devices had a burst mode that had the possibility of needing to continually perform interrupts to perform data refreshes.

90. The ’134 patent teaches, among other things, a fixed burst memory that can have non-interruptible bursts, hide required DRAM refreshes inside a known fixed burst length, free up the address and control busses for multiple cycles, and operate at higher frequencies without needing interrupts to perform refreshes of data.

91. Qualcomm products use memory devices that are compliant with the JEDEC standards JESD79-3F DDR3 SDRAM, JESD79-4A DDR4 SDRAM, JESD209-3 LPDDR3, JESD209-4 LPDDR4 and similar versions of the JEDEC DDRx standards that incorporate the innovations of the ’134 patent’s non-interruptible fixed burst length.

92. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the ’134 patent under 35 U.S.C. § 271(a), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing in or into the United States without authorization products covered by one or more claims of the ’134 patent, including, but not limited to, products that comply with the JEDEC standards JESD79-3F DDR3 SDRAM, JESD79-4A DDR4 SDRAM, JESD209-3 LPDDR3, JESD209-4 LPDDR4 and similar versions of the JEDEC DDRx standards that use non-interruptible burst read or write operations, such as the APQ8064E Snapragon 600E semiconductor device and other products in the Snapdragon 400, Snapdragon 410, Snapdragon 412, Snapdragon 415, Snapdragon 425, Snapdragon 427, Snapdragon 430, Snapdragon 435, Snapdragon 429, Snapdragon 439,

Snapdragon 450, Snapdragon 600, Snapdragon 600E, Snapdragon 602, Snapdragon 610, Snapdragon 615, Snapdragon 616, Snapdragon 617, Snapdragon 625, Snapdragon 626, Snapdragon 650, Snapdragon 652, Snapdragon 653, Snapdragon 630, Snapdragon 636, Snapdragon 660, Snapdragon 632, Snapdragon 670, Snapdragon 675, Snapdragon 665, Snapdragon 710, Snapdragon 712, Snapdragon 730, Snapdragon 730G, Snapdragon 800, Snapdragon 801, Snapdragon 805, Snapdragon 808, Snapdragon 810, Snapdragon 820, Snapdragon 821, Snapdragon 835, Snapdragon 845, Snapdragon 850, Snapdragon 855, Snapdragon Wear 2100 Platform, Snapdragon Wear 2500 Platform, Snapdragon Wear 3100 Platform, Smart Display Platform 200, QCS603, QCS605, Atlas VI, and Atlas 7 product families; other semiconductor devices, integrated circuits, and products that are compliant with JESD79-3F DDR3 SDRAM, JESD79-4A DDR4 SDRAM, JESD209-3 LPDDR3, JESD209-4 LPDDR4 or similar versions; and all other semiconductor devices, integrated circuits, and products with similar infringing technology ("the Accused '134 Products").

93. As one non-limiting example, Qualcomm infringes claim 1 of the '134 patent since the APQ8064E Snapdragon 600E semiconductor device contains DDR3 SDRAM memory controllers that operate in conformance with JEDEC's DDR3 SDRAM standard. For example, the APQ8064E Snapdragon 600E contains a circuit comprising:

a. a memory comprising a plurality of storage elements (e.g., banks of storage elements of the APQ8064E Snapdragon 600E);

b. each configured to read and write data in response to an internal address signal (e.g., stored bits of memory bank addressed and defined by internal addresses of the APQ8064E Snapdragon 600E);

c. a logic circuit configured to generate a predetermined number of said

internal address signals (e.g., generating addresses based on bank addresses, row addresses, and column addresses of the APQ8064E Snapdragon 600E) in response to an external address signal (e.g., read or write signals of the APQ8064E Snapdragon 600E), a clock signal (e.g., clock signal of the APQ8064E Snapdragon 600E), and one or more control signals (e.g., control signal of the APQ8064E Snapdragon 600E);

d. wherein said generation of said predetermined number of internal address signals is non-interruptible (e.g., burst reads or writes cannot be terminated or interrupted in the APQ8064E Snapdragon 600E).

94. Claim 1 of the ’134 patent applies to each Accused ’134 Product at least because each of those products either complies with the same JEDEC JESD79-3F DDR3 SDRAM standard, or similar versions of the JEDEC standard, including but not limited to JESD79-4A DDR4 SDRAM, JESD209-3 LPDDR3, and JESD209-4 LPDDR4, which result in infringing features (e.g., non-interruptible burst oriented read or write operations of the Accused ’134 Products) found in the JESD79-3F DDR3 SDRAM standard.

95. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

96. Qualcomm has known of the ’134 patent and their infringement of that patent since at least October 9, 2018.

97. Qualcomm, knowing their products infringe the ’134 patent and with the specific intent for others to infringe the ’134 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the ’134 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import in or into the United States

without authorization the Accused '134 Products, as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or other third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm's website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm's products that infringe the '134 patent. *See, e.g.*, https://www.qualcomm.com/support and https://developer.qualcomm.com/download/sd600/snapdragon-600-device-spec.pdf. Additional non-limiting examples include the materials found on Qualcomm's website at https://www.qualcomm.com/media/documents/files/snapdragon-600e-embedded-platform-product-brief.pdf.

98. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '134 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '134 Products, which constitute a material part of the invention of the '134 patent, knowing the Accused '134 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use. *See, e.g.*, https://www.qualcomm.com/support and https://developer.qualcomm.com/download/sd600/snapdragon-600-device-spec.pdf.

99. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

100. Qualcomm's infringement of the '134 patent has been knowing, deliberate, and

willful, since at least as early as October 9, 2018, the date on which Qualcomm was presented with the '134 patent infringement claim chart, knew of the '134 patent, and knew that their conduct constituted and resulted in infringement of the '134 patent. Monterey continued to put Qualcomm on notice of the '134 patent and Qualcomm's infringement thereof, including without limitation through communications on October 17, 2018; through the complaint filed on November 1, 2019; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable '134 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or even though Qualcomm otherwise knew or should have known that their actions constituted an unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm's conduct in light of these circumstances is egregious. Qualcomm's knowing, deliberate, and willful infringement of the '134 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

COUNT FIVE
INFRINGEMENT OF THE '516 PATENT

101. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

102. Monterey is the assignee and lawful owner of all right, title, and interest in and to the '516 patent.

103. The '516 patent is valid and enforceable.

104. The '516 patent is directed to semiconductor structures, particularly to a semiconductor gate stack and related features. A gate stack can include, for example, a gate insulating layer, a gate layer, a metallic layer—which may optionally be separated from the gate

layer by a barrier layer—and an etch-stop layer.

105. The ’516 patent explains that as the size of a semiconductor element is reduced, it does not necessarily follow that the thickness of specific layers that form that element can be equally reduced. By way of non-limiting example, the thickness of a nitride layer necessary for forming a self-aligned contact (type of contact) may still need to be at least 800 angstroms. This could lead to, among other things, designs for devices that require large aspect ratios for contact vias, and such vias may not be properly filled. An aspect ratio can be, for example, the ratio of the gate stack height to the via width.

106. The ’516 patent teaches, among other things, a gate stack height of at most 2700 angstroms and a via width of at most 0.12 microns. A gate stack with a controlled thickness can, for example, help to avoid forming contact vias with a large aspect ratio. The ’516 patent further teaches, among other things, a gate stack which can include a nitride layer that may be used for forming self-aligned contacts “SAC,” which may be used in designs to significantly reduce device size.

107. Qualcomm products use a semiconductor gate stack structure having a controlled thickness, and the Qualcomm products use vias to connect to certain portions of the semiconductor structure. Specifically, at least the Qualcomm products which are manufactured with a 28 nm or smaller process node have infringing gate stacks of at most 2700 angstroms, via widths of at most 0.12 microns, and other infringing features that use the technology of the ’516 patent.

108. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the ’516 patent under 35 U.S.C. § 271(a), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing in or into the United States without authorization products covered by one or more claims of the ’516

patent, including, but not limited to, products with a gate stack structure with a controlled thickness, such as the MSM8994 semiconductor device and other products in the Snapdragon 810, Snapdragon S4 Plus, Snapdragon S4 Pro, Snapdragon 200, Snapdragon 205, Snapdragon 208, Snapdragon 210, Snapdragon 212, Snapdragon 400, Snapdragon 410, Snapdragon 412, Snapdragon 415, Snapdragon 425, Snapdragon 427, Snapdragon 430, Snapdragon 435, Snapdragon 429, Snapdragon 439, Snapdragon 450, Snapdragon 600, Snapdragon 610, Snapdragon 615, Snapdragon 616, Snapdragon 617, Snapdragon 625, Snapdragon 626, Snapdragon 650, Snapdragon 652, Snapdragon 653, Snapdragon 630, Snapdragon 636, Snapdragon 660, Snapdragon 632, Snapdragon 670, Snapdragon 675, Snapdragon 665, Snapdragon 710, Snapdragon 712, Snapdragon 730, Snapdragon 730G, Snapdragon 800, Snapdragon 801, Snapdragon 805, Snapdragon 808, Snapdragon 820, Snapdragon 821, Snapdragon 835, Snapdragon 845, Snapdragon 850, Snapdragon 855, and Snapdragon 8cx series product families; other Qualcomm 28 nm and smaller process node semiconductor devices, integrated circuits, and products; and all other semiconductor devices, integrated circuits, and products with similar infringing technology ("the Accused '516 Products").

109. As one non-limiting example, Qualcomm infringes claim 5 of the '516 patent. For example, the MSM8994 Snapdragon 810 semiconductor device contains:

a. a semiconductor substrate (e.g., silicon substrate of the MSM8994 Snapdragon 810);

b. a gate layer, a metallic layer, an etch-stop layer, and an insulating layer (e.g., transistor gate stack of the MSM8994 Snapdragon 810);

c. a via with a via width of at most 0.12 micron (e.g., via to a portion of a transistor of the MSM8994 Snapdragon 810);

d. and a gate stack height of at most 2700 angstroms (e.g., transistor gate stack of the MSM8994 Snapdragon 810).

110. Claim 5 of the ’516 patent applies to each Accused ’516 Product at least because each of those products contain the same or similar structures as the Qualcomm MSM8994 Snapdragon 810.

111. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

112. Qualcomm has known of the ’516 patent and their infringement of that patent since at least as early as January 31, 2018.

113. Qualcomm, knowing their products infringe the ’516 patent and with the specific intent for others to infringe the ’516 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the ’516 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import in or into the United States without authorization the Accused ’516 Products, as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or other third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm’s website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm’s products that infringe the ’516 patent. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf. Additional, non-limiting examples include the materials found on Qualcomm’s websites at

www.qualcomm.com/products/snapdragon-processors-810.

114. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '516 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '516 Products, which constitute a material part of the invention of the '516 patent, knowing the Accused '516 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf.

115. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

116. Qualcomm's infringement of the '516 patent has been knowing, deliberate, and willful, since at least as early as January 31, 2018, the date of Monterey's letter to Qualcomm and therefore the date on which Qualcomm knew of the '516 patent and that their conduct constituted and resulted in infringement of the '516 patent. Monterey continued to put Qualcomm on notice of the '516 patent and Qualcomm's infringement thereof, including without limitation through communications on July 17, 2018; July 24, 2018; through the complaint filed on November 1, 2019; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable '516 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or even though Qualcomm otherwise knew or should have known that their actions constituted an

unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm's conduct in light of these circumstances is egregious. Qualcomm's knowing, deliberate, and willful infringement of the '516 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

COUNT SIX
INFRINGEMENT OF THE '407 PATENT

117. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

118. Monterey is the assignee and lawful owner of all right, title, and interest in and to the '407 patent.

119. The '407 patent is valid and enforceable.

120. The '407 patent is generally directed to programmable digital circuit architecture, and particularly to programmable digital devices which are configurable to perform any one of various digital functions, by changing the contents of a register.

121. The '407 patent explains that microcontrollers or controllers have been utilized in various applications for many years. Microcontrollers are frequently found in, for example: appliances, computers and computer equipment, automobiles, environmental control, aerospace, and thousands of other uses. Prior to the '407 patent, field programmable gate arrays ("FPGA") were utilized in several microcontroller applications. FPGAs are highly inefficient with respect to chip area, increasing their cost. Moreover, FPGAs need to have their look-up tables re-programmed in order to enable them to implement a new digital function, which is a time consuming task. FPGAs are not ideally suited for microcontroller applications, since, for example, microcontroller applications are very cost-sensitive. A FPGA is not able to realize the number of digital functions that are demanded by certain microcontroller applications within these strict cost

constraints.

122. The ’407 patent teaches, among other things, a programmable digital circuit block that can be programmed to perform a variety of predetermined digital functions upon being configured with a single register write operation. This solution allows the configuration of the programmable digital circuit block to be determined by a small number of configuration registers, providing much flexibility. In particular, the configuration of the programmable digital circuit block is fast and easy since changes in configuration are accomplished by changing the contents of the configuration registers, whereas the contents are generally a small number of configuration data bits.

123. Qualcomm products use an array of programmable digital circuit blocks, such as ARM cores. For example, the ARM core can be found in a number of different Qualcomm Snapdragon products, as it functions as a processing core of the Snapdragon system-on-chip. The ARM core is programmable to perform a variety of predetermined digital functions by changing the contents of a register.

124. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the ’407 patent under 35 U.S.C. § 271(a), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing in or into the United States without authorization products covered by one or more claims of the ’407 patent, including, but not limited to, products with an ARM core, such as the MSM8994 semiconductor device and other products in the Snapdragon 810, Snapdragon 410, Snapdragon 412, Snapdragon 415, Snapdragon 425, Snapdragon 427, Snapdragon 430, Snapdragon 435, Snapdragon 429, Snapdragon 439, Snapdragon 450, Snapdragon 610, Snapdragon 615, Snapdragon 616, Snapdragon 617, Snapdragon 625, Snapdragon 626, Snapdragon 650,

Snapdragon 652, Snapdragon 653, Snapdragon 630, Snapdragon 636, Snapdragon 660, Snapdragon 632, Snapdragon 670, Snapdragon 675, Snapdragon 665, Snapdragon 710, Snapdragon 712, Snapdragon 730, Snapdragon 730G, Snapdragon 808, Snapdragon 835, Snapdragon 845, Snapdragon 850, Snapdragon 855, and Snapdragon 8cx series product families; other Qualcomm semiconductor devices, integrated circuits, and products containing ARM A53 and newer cores; and all other semiconductor devices, integrated circuits, and products with similar ARM or ARM-derived cores ("the Accused '407 Products").

125. As one non-limiting example, Qualcomm infringes claim 8 of the '407 patent. For example, the MSM8994 Snapdragon 810 semiconductor device contains:

a. an array of programmable digital circuit block (e.g., ARM cores of the MSM8994 Snapdragon 810);

b. where each programmable digital circuit block is configurable to perform a predetermined digital function (e.g., operating in AArch32 or AArch64 execution state in the MSM8994 Snapdragon 810);

c. upon being configured with a single register write operation (e.g., writing RMR register in the MSM8994 Snapdragon 810).

126. Claim 8 of the '407 patent applies to each Accused '407 Product at least because each of those products contain infringing ARM A53 or newer cores, and/or contain ARM or ARM-derived cores containing similar infringing functionality.

127. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

128. Qualcomm has known of the '407 patent and their infringement of that patent since at least as early as January 31, 2018.

129. Qualcomm, knowing their products infringe the '407 patent and with the specific intent for others to infringe the '407 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the '407 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import (including import products made by a patented process) in or into the United States without authorization the Accused '407 Products, as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or other third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm's website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm's products that infringe the '407 patent. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf. Additional non-limiting examples include the materials found on Qualcomm's websites at www.qualcomm.com/products/snapdragon-processors-810.

130. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '407 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '407 Products, which constitute a material part of the invention of the '407 patent, knowing the Accused '407 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use. *See, e.g.*,

https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf.

131. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

132. Qualcomm's infringement of the '407 patent has been knowing, deliberate, and willful, since at least as early as January 31, 2018, the date of Monterey's letter to Qualcomm and therefore the date on which Qualcomm knew of the '407 patent and that their conduct constituted and resulted in infringement of the '407 patent. Monterey continued to put Qualcomm on notice of the '407 patent and Qualcomm's infringement thereof, including without limitation through communications on July 17, 2018; July 24, 2018; through the complaint filed on November 1, 2019; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable '407 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or even though Qualcomm otherwise knew or should have known that their actions constituted an unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm's conduct in light of these circumstances is egregious. Qualcomm's knowing, deliberate, and willful infringement of the '407 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

COUNT SEVEN
INFRINGEMENT OF THE '727 PATENT

133. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

134. Monterey is the assignee and lawful owner of all right, title, and interest in and to

the ’727 patent.

135. The ’727 patent is valid and enforceable.

136. The ’727 patent is generally directed to semiconductor formation, particularly to semiconductor contact formation using multiple etch stop layers and sub interlevel dielectric layers.

137. The ’727 patent explains that semiconductor integrated circuit manufacturing efforts are usually complicated by ever increasing demands for greater functionality. Integrated circuits with more components typically provide greater functionality. One traditional focus for achieving greater densities was directed towards reducing the size of individual components (e.g., transistors). Semiconductor contact formation processes usually include the creation of a contact void for deposition of the contact layer. The smaller the contact void, the more compact the contact and the greater the possible component density. However, decreases in contact sizes can be limited by contact void creation processes (e.g., lithographic etching processes). Standard lithographic etching and removal processes traditionally had difficulty producing relatively small contact voids. Complex processes that attempted to create smaller voids were often cost prohibitive or nonfeasible.

138. The ’727 patent teaches, among other things, how to use multiple etch stop sublayers and dielectric layers to form contacts which allow semiconductor devices to be densely packed. The ’727 patent process allows for the ability to precisely form semiconductor contact regions in a convenient and efficient manner. One exemplary benefit of the ’727 patent contact formation process is that the process facilitates the fabrication of contact regions with a relatively small substrate coupling area and relatively large metal layer coupling area. The relatively small substrate coupling area can permit multiple active regions in an integrated circuit to be arranged

relatively close to one another, while the relatively large metal layer coupling area facilitates avoidance of critical dimension issues. A conducting material's critical dimension may limit how small a contact can be and still operate reliably.

139. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the '727 patent under 35 U.S.C. § 271(a) and/or 35 U.S.C. § 271(g), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing (including importing products made by a patented process) in or into the United States without authorization products covered by one or more claims of the '727 patent, including, but not limited to, products such as the MSM8994 semiconductor device and other products in the Snapdragon 810 and Snapdragon 808 series product families; other Qualcomm 20 nm and smaller process node semiconductor devices, integrated circuits, and products; and all other semiconductor devices, integrated circuits, and products with similar infringing technology ("the Accused '727 Products").

140. As one non-limiting example, Qualcomm infringes claim 1 of the '727 patent. For example, the method used to produce Qualcomm's MSM8994 Snapdragon 810 semiconductor device performs the steps of:

a. forming a gate region (e.g., transistor gate of the MSM8994 Snapdragon 810) and source and drain regions (e.g., source and drain of a transistor in the MSM8994 Snapdragon 810);

b. depositing a multiple etch stop layer comprising (e.g., two layer etch stop layer comprised of silicon nitride in the MSM8994 Snapdragon 810);

c. depositing a first etch stop layer over the substrate (e.g., first silicon nitride etch stop layer in the MSM8994 Snapdragon 810);

d. depositing a sub interlevel dielectric layer over the first etch stop layer (e.g., oxide dielectric layer in the MSM8994 Snapdragon 810);

e. depositing a second etch stop layer over the dielectric layer (e.g., second silicon nitride etch stop layer in the MSM8994 Snapdragon 810) wherein said second etch stop layer has similar selectivity characteristics as the first etch stop layer (e.g., both etch stop layers being comprised of silicon nitride in the MSM8994 Snapdragon 810);

f. depositing a second sub interlevel dielectric layer over the second etch stop layer (e.g., oxide dielectric layer in the MSM8994 Snapdragon 810);

g. forming a contact region in the multiple etch stop layer by selectively removing some of said multiple etch stop layer (e.g., etching into said multiple etch stop layers to form a contact void in the MSM8994 Snapdragon 810) and forming a sub spacer region (e.g., space in the contact void of the MSM8994 Snapdragon 810), such that the substrate coupling area of the contact region (e.g., bottom of the contact region connecting to the silicon substrate in the MSM8994 Snapdragon 810) is smaller than a metal layer coupling area of said contact region (e.g., top of the contact region connecting to the metal layer in the MSM8994 Snapdragon 810); and

h. depositing electrically conductive material in said contact region (e.g., depositing tungsten in the contact region of the MSM8994 Snapdragon 810).

141. Claim 1 of the ’727 patent applies to each Accused ’727 Product at least because each of those products were produced by the same 20 nm process node, or produced by other technology nodes which result in infringing features.

142. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

143. Qualcomm has known of the ’727 patent and their infringement of that patent since

at least as early as January 31, 2018.

144. Qualcomm, knowing their products infringe the '727 patent and with the specific intent for others to infringe the '727 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the '727 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import (including import products made by a patented process) in or into the United States without authorization the Accused '727 Products, as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or other third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm's website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm's products that infringe the '727 patent. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf. Additional non-limiting examples include the materials found on Qualcomm's websites at www.qualcomm.com/products/snapdragon-processors-810.

145. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '727 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '727 Products, which constitute a material part of the invention of the '727 patent, knowing the Accused '727 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity

of commerce suitable for substantial noninfringing use. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf.

146. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

147. Qualcomm's infringement of the '727 patent has been knowing, deliberate, and willful, since at least as early as January 31, 2018, the date of Monterey's letter to Qualcomm and therefore the date on which Qualcomm knew of the '727 patent and knew that their conduct constituted and resulted in infringement of the '727 patent. Monterey continued to put Qualcomm on notice of the '727 patent and Qualcomm's infringement thereof, including without limitation through communications on July 17, 2018; July 24, 2018; through the complaint filed on November 1, 2019; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable '727 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or even though Qualcomm otherwise knew or should have known that their actions constituted an unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm's conduct in light of these circumstances is egregious. Qualcomm's knowing, deliberate, and willful infringement of the '727 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

COUNT EIGHT
INFRINGEMENT OF THE '797 PATENT

148. Monterey incorporates by reference the preceding paragraphs as if fully set forth herein.

149. Monterey is the assignee and lawful owner of all right, title, and interest in and to the '797 patent.

150. The '797 patent is valid and enforceable.

151. The '797 patent is generally directed to a semiconductor device structure, particularly to a semiconductor device with multiple etch stop layers and sub interlevel dielectric layers.

152. The '797 patent explains that semiconductor integrated circuit manufacturing efforts are usually complicated by ever increasing demands for greater functionality. Integrated circuits with more components typically provide greater functionality. One traditional focus for achieving greater densities was directed towards reducing the size of individual components (e.g., transistors). Semiconductor contact formation processes usually include the creation of a contact void for deposition of the contact layer. The smaller the contact void, the more compact the contact and the greater the possible component density. However, decreases in contact sizes can be limited by contact void creation processes (e.g., lithographic etching processes). Standard lithographic etching and removal processes traditionally had difficulty producing relatively small contact voids. Complex processes that attempted to create smaller voids were often cost prohibitive or nonfeasible.

153. The '797 patent teaches, among other things, to use multiple etch stop sublayers and dielectric layers to form contacts which allow semiconductor devices to be densely packed. The '797 patent allows for the ability to precisely form semiconductor contact regions in a convenient and efficient manner. One exemplary benefit of the '797 patent is that the structure facilitates the fabrication of contact regions with a relatively small substrate coupling area and relatively large metal layer coupling area. The relatively small substrate coupling area can permit

multiple active regions in an integrated circuit to be arranged relatively close to one another, while the relatively large metal layer coupling area facilitates avoidance of critical dimension issues. A conducting material's critical dimension may limit how small a contact can be and still operate reliably.

154. Qualcomm products use a semiconductor multiple etch stop layer structure. The products further use multiple sub interlevel dielectric layers, a contact region, and a spacer region. Those infringing features were used in, for example, Qualcomm's 20 nm process semiconductor devices, in order to, among other things, more densely pack semiconductor features into a semiconductor device.

155. Qualcomm has directly infringed, and continues to directly infringe, one or more claims of the '797 patent under 35 U.S.C. § 271(a), either literally and/or under the doctrine of equivalents, by, among other things, making, using, selling, offering to sell, and/or importing in or into the United States without authorization products covered by one or more claims of the '797 patent, including, but not limited to, products such as the MSM8994 semiconductor device and other products in the Snapdragon 810 and Snapdragon 808 series product families; other Qualcomm 20 nm and smaller process node semiconductor devices, integrated circuits, and products; and all other semiconductor devices, integrated circuits, and products with similar infringing technology ("the Accused '797 Products").

156. As one non-limiting example, Qualcomm infringes claim 1 of the '797 patent. For example, the MSM8994 Snapdragon 810 semiconductor device contains:

a. a substrate (e.g., transistor gate of the MSM8994 Snapdragon 810) a contact region that provides an electric path to and from the substrate (e.g., tungsten contact of the MSM8994 Snapdragon 810), and a spacer region such that the substrate coupling area of the

contact region (e.g., bottom of the contact region connecting to the silicon substrate in the MSM8994 Snapdragon 810) is smaller than a metal layer coupling area of said contact region (e.g., top of the contact region connecting to the metal layer in the MSM8994 Snapdragon 810);

b. a multiple etch stop layer (e.g., two layer etch stop layer comprised of silicon nitride and sub interlevel dielectric layers in the MSM8994 Snapdragon 810) that provides electrical insulation between other regions of the integrated circuit (e.g., insulating sub interlevel oxide dielectric layer of the MSM8994 Snapdragon 810);

c. wherein said multiple etch stop layer includes a first etch stop layer on the substrate (e.g., first silicon nitride etch stop layer of the MSM8994 Snapdragon 810);

d. a sub interlevel dielectric layer over the first etch stop layer (e.g., oxide dielectric layer of the MSM8994 Snapdragon 810);

e. a second etch stop layer over the dielectric layer (e.g., second silicon nitride etch stop layer of the MSM8994 Snapdragon 810) wherein said second etch stop layer has similar selectivity characteristics as the first etch stop layer (e.g., both etch stop layers being comprised of silicon nitride in the MSM8994 Snapdragon 810); and

f. a second sub interlevel dielectric layer over the second etch stop layer (e.g., oxide dielectric layer of the MSM8994 Snapdragon 810).

157. Claim 1 of the ’797 patent applies to each Accused ’797 Product at least because each of those products contain the same or similar structures as the Qualcomm MSM8994 Snapdragon 810.

158. Monterey has complied with the requirements of 35 U.S.C. § 287(a) at least because Monterey provided Qualcomm with written notice of the infringement as discussed above.

159. Qualcomm has known of the ’797 patent and their infringement of that patent since

at least as early as January 31, 2018.

160. Qualcomm, knowing their products infringe the '797 patent and with the specific intent for others to infringe the '797 patent, has induced infringement of, and continues to induce infringement of, one or more claims of the '797 patent under 35 U.S.C. § 271(b), either literally and/or under the doctrine of equivalents, by, among other things, actively inducing others, including their customers, to make, use, sell, offer to sell, and/or import in or into the United States without authorization the Accused '797 Products, as well as products containing the same. Qualcomm knowingly and intentionally instructs their customers, OEMs, foundry suppliers, distributors, and/or other third parties to infringe at least through user manuals, product documentation, and other materials, such as those located on Qualcomm's website at www.qualcomm.com. For example, Qualcomm provides data sheets, development content, diagrams, white papers, and software instructing customers on uses of Qualcomm's products that infringe the '797 patent. *See, e.g.*, https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf. Additional non-limiting examples include the materials found on Qualcomm's websites at www.qualcomm.com/products/snapdragon-processors-810.

161. Qualcomm has contributed to the infringement of, and continues to contribute to the infringement of, one or more claims of the '797 patent under 35 U.S.C. § 271(c), either literally and/or under the doctrine of equivalents, by, among other things, selling, offering to sell, and/or importing in or into the United States the Accused '797 Products, which constitute a material part of the invention of the '797 patent, knowing the Accused '797 Products to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use. *See, e.g.*,

https://www.qualcomm.com/support and https://www.qualcomm.com/system/files/document/files/snapdragon_product_brief_810_0.pdf.

162. Monterey has sustained and is entitled to recover damages as a result of Qualcomm's past and continuing infringement.

163. Qualcomm's infringement of the '797 patent has been knowing, deliberate, and willful, since at least as early as January 31, 2018, the date of Monterey's letter to Qualcomm and therefore the date on which Qualcomm knew of the '797 patent and that their conduct constituted and resulted in infringement of the '797 patent. Monterey continued to put Qualcomm on notice of the '797 patent and Qualcomm's infringement thereof, including without limitation through communications on July 17, 2018; July 24, 2018; through the complaint filed on November 1, 2019; and yet again through this amended complaint. Qualcomm nonetheless has committed—and continues to commit—acts of direct and indirect infringement despite knowing that their actions constituted infringement of the valid and enforceable '797 patent, despite a risk of infringement that was known or so obvious that it should have been known to Qualcomm, and/or even though Qualcomm otherwise knew or should have known that their actions constituted an unjustifiably high risk of infringement of that valid and enforceable patent. Qualcomm's conduct in light of these circumstances is egregious. Qualcomm's knowing, deliberate, and willful infringement of the '797 patent entitles Monterey to increased damages under 35 U.S.C. § 284 and to attorney fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

RELIEF REQUESTED

Wherefore, Monterey respectfully requests that this Court enter judgment against Qualcomm as follows:

A. that Qualcomm has infringed each of the Patents-in-Suit;

B. that Qualcomm's infringement of each Patent-in-Suit is and has been willful;

C. that Monterey be awarded damages adequate to compensate it for the patent infringement that has occurred, together with pre-judgment interest, post-judgment interest, and costs;

D. that Monterey be awarded an accounting and additional damages for any infringing sales not presented at trial;

E. that Monterey be awarded all other damages permitted by 35 U.S.C. § 284, including without limitation increased damages up to three times the amount of compensatory damages found;

F. that this is an exceptional case and that Monterey be awarded its costs and reasonable attorneys' fees incurred in this action as provided by 35 U.S.C. § 285;

G. that Qualcomm as well as their officers, directors, agents, employees, representatives, attorneys, and all others acting in privity or in concert with them, their subsidiaries, divisions, successors and assigns be permanently enjoined from further infringement of each of the Patents-in-Suit;

H. that, in the event a permanent injunction preventing further infringement of each of the Patents-in-Suit is not granted, Monterey be awarded a compulsory ongoing licensing fee for any such further infringement; and

I. such other relief as this Court deems just and proper.

DEMAND FOR JURY TRIAL

Monterey hereby demands trial by jury on all claims and issues so triable.

Dated: February 14, 2020

Respectfully submitted,

FARNAN LLP

/s/ Michael J. Farnan

Brian E. Farnan (Bar No. 4089)
Michael J. Farnan (Bar No. 5165)
919 North Market St., 12th Floor
Wilmington, DE 19801
Telephone: 302-777-0300
Facsimile: 302-777-0301
bfarnan@farnanlaw.com
mfarnan@farnanlaw.com

Attorneys for Plaintiff Monterey Research, LLC

Of Counsel:

Jonas McDavit
DESMARAIS LLP
230 Park Avenue
New York, NY 10169
Tel: (212) 351-3400

EXHIBIT A

US006459625B1

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

(10) **Patent No.: US 6,459,625 B1**
(45) **Date of Patent: Oct. 1, 2002**

(54) **THREE METAL PROCESS FOR OPTIMIZING LAYOUT DENSITY**

(75) Inventors: **Colin S. Bill**, Cupertino; **Jonathan S. Su**, Gabriel; **Ravi P. Gutala**, San Jose, all of CA (US)

(73) Assignee: **Advanced Micro Devices, Inc.**, Sunnyvale, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/767,341**

(22) Filed: **Jan. 23, 2001**

Related U.S. Application Data

(60) Provisional application No. 60/185,149, filed on Feb. 25, 2000.

(51) **Int. Cl.**[7] **G11C 16/04**; G11C 5/02

(52) **U.S. Cl.** **365/185.33**; 365/63; 365/52

(58) **Field of Search** 365/185.05, 161, 365/164, 185.01, 185.33, 63, 52

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,590,072 A * 12/1996 Choi 365/185.01

6,034,882 A * 3/2000 Johnson et al. 365/103

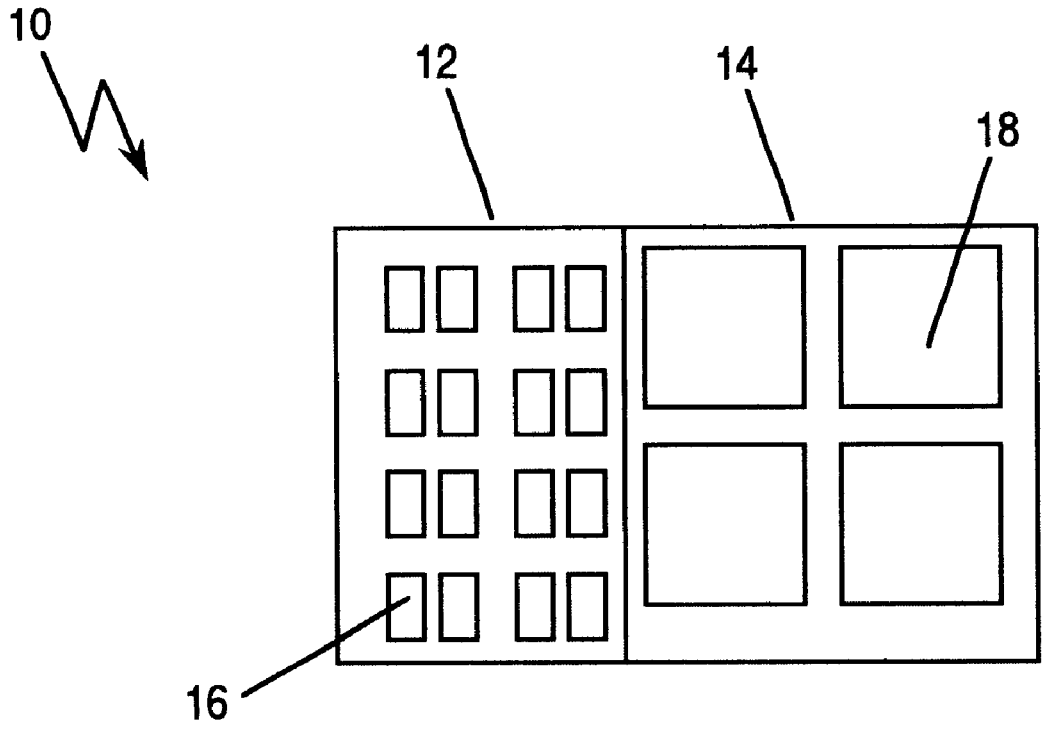
* cited by examiner

Primary Examiner—David Nelms
Assistant Examiner—Thong Le

(57) **ABSTRACT**

The present invention discloses a method and system to optimize electrical interconnection of electrical components in a periphery area of a memory device thereby minimizing the periphery area. The periphery area is divided into a plurality of sub-circuits formed by selectively electrically connecting the electrical components. Electrical interconnection of the electrical components to form the sub-circuits is accomplished using a first metal layer and a second metal layer. The first metal layer is formed to create a plurality of first metal layer lines that are oriented to extend in substantially one direction on the memory device. The second metal layer is formed to create a plurality of second metal layer lines that are oriented to extend substantially perpendicular to the first metal layer lines. The plurality of sub-circuits are electrically interconnected using a third metal layer that is formed to create a plurality of third metal layer lines that are oriented to extend substantially parallel to the first metal layer lines.

14 Claims, 4 Drawing Sheets



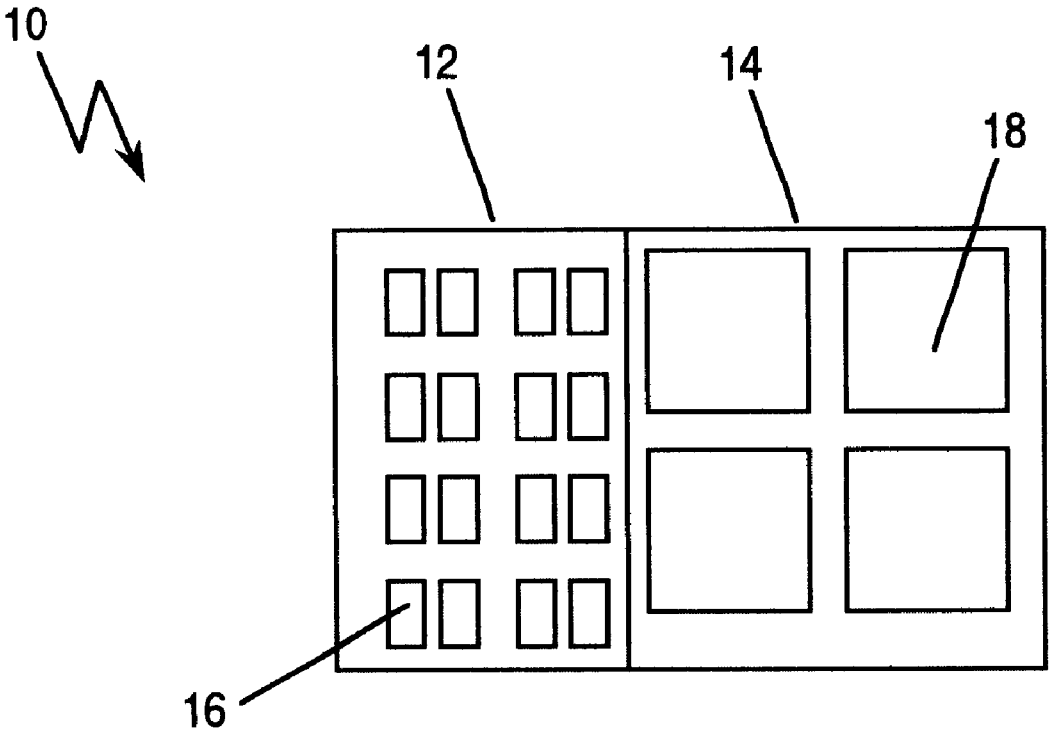


Fig. 1

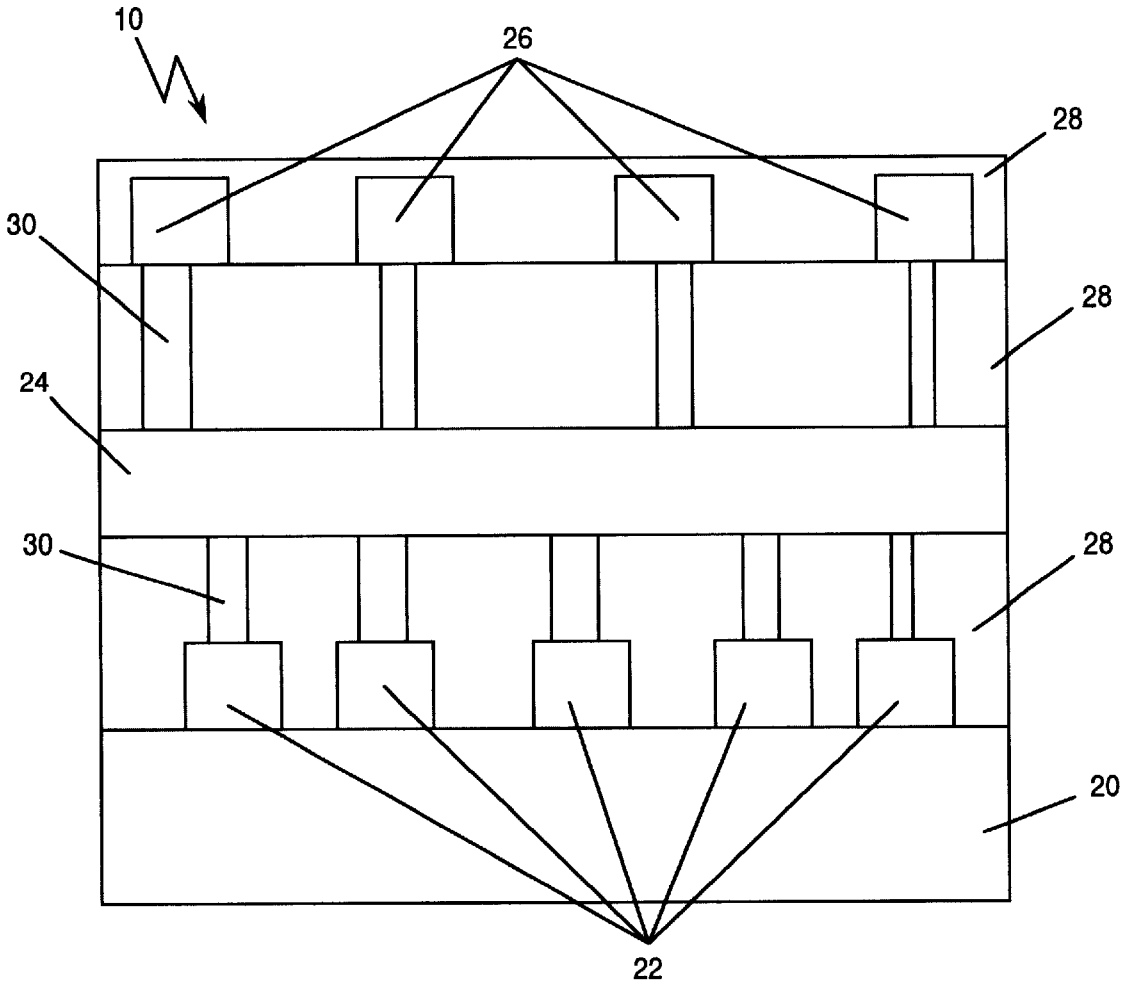


Fig. 2

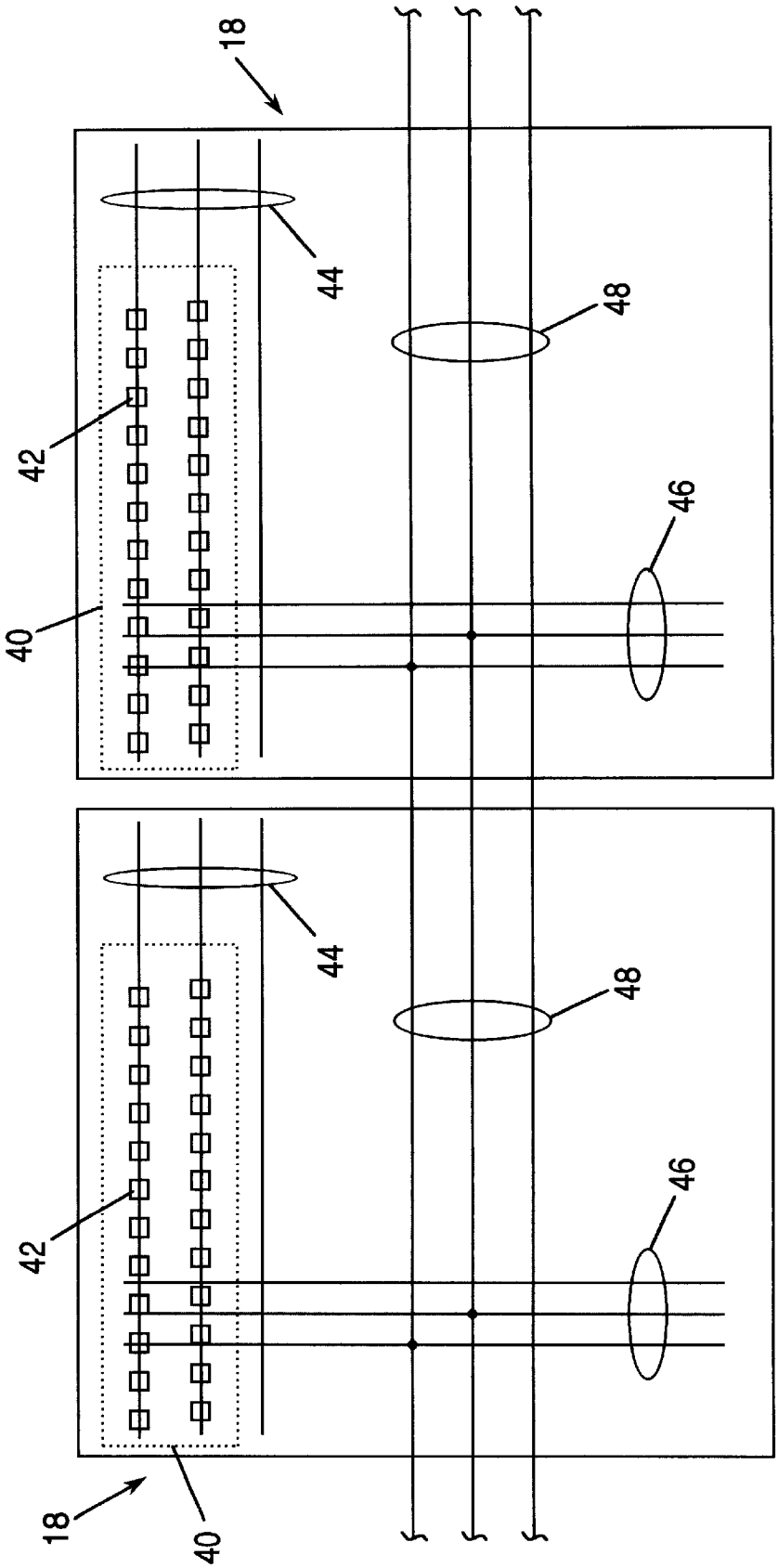


Fig. 3

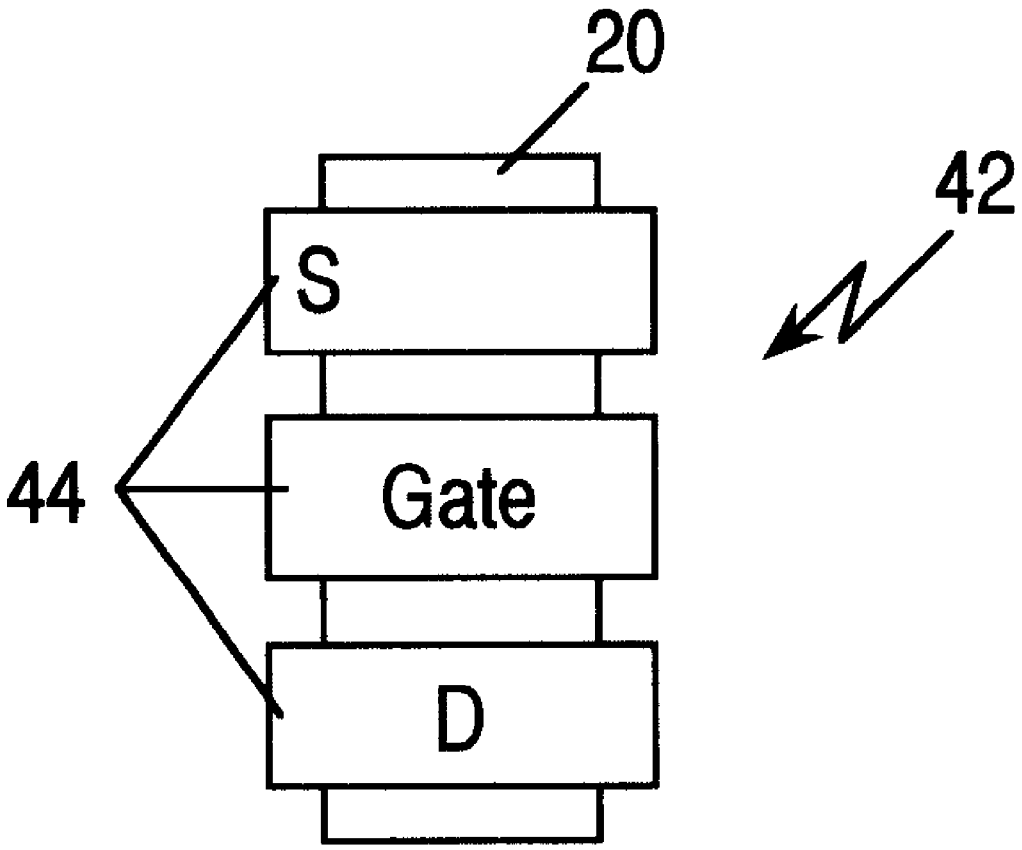


Fig. 4

THREE METAL PROCESS FOR OPTIMIZING LAYOUT DENSITY

This application claims the benefit under 35 U.S.C. §119(e) of Provisional U.S. patent application Ser. No. 60/185,149, filed Feb. 25, 2000.

FIELD OF INVENTION

The present invention relates generally to non-volatile memory devices and, more particularly, to methods and systems for optimization of layout density in a periphery area using a three-metal interconnection process in flash electrically erasable programmable read-only memory (EEPROM) devices.

BACKGROUND OF THE INVENTION

Flash memories are popular memory storage devices because they store information in the absence of continuous power and are capable of being constructed in a very compact form. Flash memory is typically constructed by fabricating a plurality of electrical components in a silicon substrate. It is desirable to place as many of the electrical components as possible in the available area on the silicon substrate to optimize functionality and economical manufacture. The density or layout area of the electrical components depends on the physical size of the electrical components and the electrical connections between components. As the size of the electrical components decreases due to technological advances, more components can be placed in the available area on the flash memory. However, more electrical components require more electrical connections that can offset the layout area gained by the smaller component size.

Flash memory devices include two functional areas, a core cell area to perform memory functions and a periphery area to perform logic functions. As known in the art, the core cell area includes rows and columns of electrical components that are floating-gate transistors formed in the silicon substrate during fabrication. The floating-gate transistors located in the core cell area of the flash memory are typically referred to as core memory cells. The rows of core memory cells within the core cell area are typically electrically connected to form wordlines and the columns of core memory cells within the core cell area are typically electrically connected to form bitlines. As known in the art, the wordlines and bitlines are used to provide predetermined operational voltages to erase, read and write the core memory cells within the core cell area.

In addition to the core cell area, the flash memory also has a periphery area that includes a plurality of electrical components such as transistors, resistors, capacitors and diodes formed in the silicon substrate during fabrication. As known in the art, the resistors, capacitors and diodes may be formed during fabrication to create electrical components such as bipolar and field-effect transistors. The electrical components are electrically connected to form integrated circuits that perform logic functions within the flash memory to support operations such as the read, write and erase of the core memory cells. Part of the formation of the electrical connections of the electrical components occurs during a fabrication process known in the art as metallization.

Generally, metallization involves depositing a thin film of conductive metal on the flash memory such that the electrical components are formed and electrically connected with the conductive metal. In addition to forming portions of the electrical components, the conductive metal electrically

connects the electrical components in a predetermined configuration, thereby "wiring" the electrical components to create the integrated circuits. The conductive metal that electrically connects the electrical components is routed on the flash memory and contributes to the layout area consumed.

Known prior art flash memory uses a two-metal layer metallization process to electrically connect the electrical components in the periphery area of the flash memory. A first layer of metal is typically used to form portions of the electrical components and to electrically connect the electrical components to form a plurality of sub-circuits that perform predetermined logic functions during operation. A second layer of metal is typically used to electrically connect one sub-circuit with another and to electrically connect the sub-circuits with the core memory cells in the core area such that operations can be performed within the flash memory.

A known problem with this method and system of electrical connection is layout area consumed by routing channels of the first and second layer of metal between the sub-circuits in the periphery area. In addition, layout area is consumed for routing channels that are used to route the first and second metal layers between the electrical components that form the sub-circuits. The layout area for the routing channels required by existing electrical connection systems and methods increases the size of the periphery area on the flash memory. The area on the flash memory that is not consumed by the periphery area can be reserved for the core cell area, allowing more core memory cells to be fabricated on the flash memory. It is therefore desirable to minimize the amount of periphery area consumed, thereby increasing the amount of information stored in the flash memory. That is, the ratio of the core area to the periphery area can be maximized.

To that end, a need exists for flash memory with an improved method and system of interconnection of electrical components to minimize the area consumed in the periphery area of the flash memory.

SUMMARY OF THE INVENTION

The present invention discloses a method and system of optimizing layout area consumed in a periphery area of a flash memory. The flash memory includes a core cell area and the periphery area. Within the core cell area, the flash memory includes a plurality of core memory cells; and, within the periphery area, the flash memory includes a plurality of sub-circuits. Generally, the core cell area provides memory-related functions in the flash memory and the periphery area supports the memory-related functions by providing logic-related functions in the flash memory.

The core cell area and the periphery area share the available area on the flash memory. As such, a smaller periphery area is desirable, thereby increasing the area available for the core cell area. In the preferred embodiment, selectively placing and electrically connecting a plurality of electrical components to form the sub-circuits and selectively electrically connecting the sub-circuits with the core memory cells mininimizes the layout area of the sub-circuits in the periphery area.

The preferred sub-circuits include the electrical components such as transistors, resistors, capacitors and diodes that are electrically connected with a first metal layer, a second metal layer and a third metal layer. The electrical components are electrically connected to form the sub-circuits by the first metal layer and the second metal layer. The sub-circuits are electrically interconnected with each other and

with the core memory cells in the core cell area by the third metal layer. Those skilled in the art would understand that, typically, the majority of electrical components in the periphery area used to create the integrated circuits are transistors; however, other electrical components such as diodes and resistors could also be utilized.

The circuit layout of the sub-circuits is such that the transistors are oriented to form a plurality of rows of transistors wherein each transistor has a drain and a source oriented along an axis parallel with the rows of transistors. In another preferred embodiment of the present invention, the drain and the source of the transistors are not oriented along an axis parallel with the rows of transistors.

The first metal layer is applied to the periphery area of the flash memory during fabrication to form and partially interconnect the electrical components in a predetermined circuit configuration. The first metal layer comprises a plurality of first metal lines that provide interconnecting surface "wiring" for the predetermined circuit configuration. The layout in the periphery area of the first metal lines is oriented to extend along an axis substantially parallel to the rows of transistors. The second metal layer also provides surface "wiring" of the electrical components to complete the predetermined circuit configuration and form the sub-circuits. The second metal layer is also applied to the periphery area of the flash memory during fabrication. The layout of the second metal layer on the flash memory is deposited to form a plurality of second metal lines that are oriented to extend along an axis substantially perpendicular to the first metal lines.

The sub-circuits are selectively electrically interconnected and electrically connected with the core memory cells in the core cell area by the third metal layer. The third metal layer is also applied to the periphery area of the flash memory during the fabrication process and is adapted to form a plurality of third metal lines. The third metal lines provide surface "wiring" to electrically connect the sub-circuits with the core memory cells and are oriented to extend along an axis substantially parallel to the first metal lines.

During the fabrication process, the periphery area consumed by the transistors, the first metal layer, the second metal layer and the third metal layer is optimized in the preferred embodiment. The orientation of the transistors uniformly in rows allows the spacing between the transistors in the rows to be minimized without causing short circuits or undesirable leakage currents while still allowing electrical connection of the transistors. In addition, the combination of the first metal layer and the second metal layer to form and electrically connect the electrical components to create the sub-circuits also minimizes the periphery area consumed.

The electrical connections with the transistors are typically located directly below the first metal lines of the first metal layer. Since the first metal lines are substantially straight, additional spacing between the rows of transistors to allow for electrical connection of the transistors to the first metal layer is minimized. The second metal lines of the second metal layer provide additional electrical connections to complete the sub-circuits, thereby minimizing bends in the first metal lines and repositioning of the electrical components under the first metal lines. The second metal lines are also substantially straight and orthogonally pass over the first metal lines, thereby minimizing noise and allowing efficient electrical connections that further minimize consumption of the periphery area. The third metal layer provides electrical connection of the first metal layer and the second metal layer with the core memory cells. As such, the third metal lines can be substantially straight and be routed on top of the sub-circuits such that consumption of the periphery area is minimized.

These and other features and advantages of the invention will become apparent upon consideration of the following detailed description of the presently preferred embodiments of the invention, viewed in conjunction with the appended drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** represents a block diagram of a portion of a preferred flash memory incorporating an embodiment of the present invention.

FIG. **2** generally illustrates a cross-sectional view of a portion of the preferred flash memory.

FIG. **3** illustrates a top view of two sub-circuits of the preferred flash memory.

FIG. **4** illustrates the preferred source/drain orientation in relation to the first metal layer lines.

DETAILED DESCRIPTION OF THE INVENTION

The exemplary embodiments of the invention are set forth below with reference to specific configurations, and those skilled in the art would recognize that various changes and modifications can be made on the specific configurations while remaining within the scope of the claims. The invention may be used with any type of memory device; however, the preferred embodiment of the invention is designed for a flash memory.

FIG. **1** illustrates a general block diagram of a preferred flash memory **10** incorporating an embodiment of the present invention. The flash memory **10** includes a core cell area **12** and a periphery area **14**. As known in the art, within the core cell area **12**, the flash memory **10** includes a plurality of core memory cells **16**. Within the periphery area **14**, the flash memory **10** includes a plurality of sub-circuits **18** that are made up of various types of circuit components such as transistors, capacitors and resistors. Generally, the core cell area **12** performs memory-related functions in the flash memory **10**, and the periphery area **14** supports the memory-related functions by performing logic-related functions in the flash memory **10**. Although not illustrated in FIG. **1**, the periphery area **14** typically includes electronic circuits such as decoder circuits, state machine circuitry, voltage regulator circuits and addressing circuits as known in the art.

In order to understand the method and system of optimizing the layout in the periphery area **14**, an understanding of the operation and configuration of the periphery area **14** is helpful. Generally, during operation of the preferred flash memory **10**, the sub-circuits **18** within the periphery area **14** execute a plurality of functions to assist in performing memory operations such as read, write and erase. The sub-circuits **18** are integrated electric circuits that occupy different physical areas in the periphery area **14** of the flash memory **10**, and are identified based on the function the particular sub-circuit **18** performs. Those skilled in the art would recognize that the sub-circuits **18** in the periphery area **14** of the flash memory **10**, as briefly set forth above, can contain many types of logic circuits and control circuits that perform a variety of functions in the flash memory **10**. The present invention relates to methods of optimally placing and electrically interconnecting transistors that are used

in the sub-circuits **18** to the other respective sub-circuits **18**, as well as interconnecting the various electrical components used in a particular sub-circuit **18** with one another.

The physical size of the periphery area **14** occupied by the sub-circuits **18** is a predetermined area based on the physical size of the particular sub-circuits **18**, which is determined on a circuit-by-circuit basis during the design phase of the flash memory **10**. The design phase occurs prior to fabrication of the flash memory **10** and involves circuit design, dimensioning and circuit layout of the sub-circuits **18** on the surface of the flash memory **10**. The circuit layout area of a particular sub-circuit **18** is preferentially designed using just enough area for the physical dimensions of the electrical components and interconnections as well as providing enough spacing between components and the interconnections to avoid electrical short circuits and problems with leakage current. For the best layout efficiency of the layout of transistors used in the sub-circuits **18**, the transistors should have their source/drain geometries spaced apart by their minimum source/drain spacing rule. In addition, in the preferred embodiment, all of the transistors in the sub-circuits **18** are placed in rows while obeying the minimum source/drain spacing rules of the particular transistors.

FIG. **2** generally illustrates a cross-sectional view of a portion of a preferred flash memory **10**. As known in the art, the preferred flash memory **10** includes a substrate **20** where the electrical components of the flash memory **10** are formed during fabrication. In addition, the preferred flash memory **10** includes a first metal interconnect layer (M1) **22**, a second metal interconnect layer (M2) **24** and a third metal interconnect layer (M3) **26**, which are separated and encapsulated by a plurality of dielectric layers **28**. The dielectric layers **28**, which can be made using any dielectric material commonly used in the art such as oxide, silicon nitride or a polyamide film, provides an insulating layer between the first metal interconnect layer (M1) **22** and the second metal interconnect layer (M2) **24** in the preferred embodiment. In addition, the dielectric layer **28** is also used to provide an insulating layer between wiring lines formed in the respective first (M1), second (M2) and third (M3) metal interconnect layers **22**, **24**, **26** as set forth in detail below. One dielectric layer **28** is also deposited over the third metal interconnect layer (M3) **26** to protect the third metal interconnect layer (M3) **26**.

A plurality of contact holes **30**, which are also referred to in the art as vias, are created during fabrication to penetrate the dielectric layers **28** in predetermined locations of the flash memory **10**. The contact holes **30** provide electrical interconnections between the first metal interconnect layer (M1) **22**, the second metal interconnect layer (M2) **24** and the third interconnect layer **26** at the predetermined locations of the flash memory **10**. As known in the art, once the contact holes **30** are formed, the contact holes **30** are filled with a conductive material such as copper, tungsten or any other suitable conductive material. To that end, a conductive path is created between the respective metal interconnect layers **22**, **24**, **26** with the contact holes **30** at predetermined locations in the preferred flash memory **10**.

FIG. **3** generally represents the top view of two sub-circuits **18** within the periphery area **14** of the flash memory **10**. As previously set forth, the sub-circuits **18** include a plurality of electrical circuits **40**, which generally include a plurality of transistors **42** and a plurality of other circuit components (not shown). Although not illustrated, those skilled in the art would recognize that other circuit components, such as resistors, capacitors and diodes may also be used in the electrical circuits **40**. However, since a majority of the circuit components used in the preferred flash memory **10** are transistors **42**, the present invention is directed to methods and systems that provide optimal transistor **42** layout and interconnection.

In the preferred embodiment, the individual components of the electrical circuits **40** in each sub-circuit **18** are electrically interconnected with one another in a predetermined manner to form the sub-circuits **18** by the first metal interconnect layer (M1) **22** and the second metal interconnect layer (M2) **24**. The circuit component layout of the sub-circuits **18** is preferentially designed such that the transistors **42** of the electrical circuits **40** are positioned to form rows of transistors **42**, wherein each transistor **42** is positioned such that its drain and source are oriented along a common horizontal axis with wiring lines formed by the first metal interconnect layer (M1) **22** as set forth in detail below.

The first metal interconnect layer (M1) **22** is deposited on the periphery area **14** of the flash memory **10** during fabrication to partially interconnect the individual circuit components of each respective electrical circuit **40** in predetermined circuit configurations. Referring to FIGS. **2** and **3**, the first metal interconnect layer (M1) **22** is fabricated to form a plurality of first metal layer lines **44** that provide interconnecting surface "wiring" of the components of the electrical circuits **40**. In the preferred embodiment, the first metal layer lines **44** are formed to extend along an axis substantially parallel to the rows of transistors **42** in the electrical circuits **40**. Those skilled in the art would recognize that occasional deviation from the axis parallel with the rows of transistors **42** will occur to perform necessary electrical connections but, preferentially, the first metal layer lines **44** are formed parallel with the rows of transistors **42**. The electrical circuits **40** are partially electrically interconnected once the first metal interconnect layer (M1) **22** is deposited during fabrication to form the first metal layer lines **44**. As set forth below, the remaining electrical interconnection of the electrical circuits **40** in the sub-circuits **18** is accomplished using the second metal interconnect layer (M2) **24**.

The second metal interconnect layer (M2) **24** also provides surface "wiring" of the electrical circuits **40** to complete the predetermined circuit configuration and formation of the sub-circuits **18**. The second metal interconnect layer (M2) **24** is fabricated on the periphery area **14** of the flash memory **10** during fabrication in a predetermined configuration. Referring to FIGS. **2** and **3**, the second metal interconnect layer (M2) **24** is fabricated to form a plurality of second metal layer lines **46** that are oriented substantially perpendicular to the first metal layer lines **44**. Those skilled in the art would again recognize that occasional deviation from the axis perpendicular with the first metal layer lines **44** will occur to facilitate efficient electrical connections but, preferentially, the second metal layer lines **46** are formed to run perpendicular to the orientation of the first metal layer lines **44**.

As set forth above, the electrical circuits **40** of the sub-circuits **18** are electrically interconnected with the first metal layer lines **44** and the second metal layer lines **46**. As such, each circuit component and transistor **42** in the electrical circuits **40** are interconnected in a predetermined configuration using the first metal layer lines **44** and the second metal layer lines **46**. The contact holes **30** are used to electrically connect the second metal layer lines **46** to respective first metal layer lines **44** and circuit components and transistors **42** on the substrate **20**. Utilizing both the first metal interconnect layer (M1) **22** and the second metal

interconnect layer (M2) **24** within the boundary of the sub-circuits **18** provides the greatest opportunity to place transistors **42** at their minimum spacing and yet have the metal interconnect consuming no extra area than that needed by the transistors **42**. As such, this layout scheme optimizes the area consumed by the transistors **42**.

The third metal interconnect layer (M3) **26** is used to electrically interconnect one respective sub-circuit **18** to another respective sub-circuit **18** in the preferred flash memory **10**. The third metal interconnect layer (M3) **26** is fabricated to form a plurality of third metal layer lines **48** that electrically interconnect the respective sub-circuit **18** to each other. The third layer metal lines **48** are formed perpendicular to the second metal layer lines **46** in the preferred flash memory **10**. Those skilled in the art would recognize that occasional deviation from the axis perpendicular to the second metal layer lines **46** will occur to facilitate electrical connection but, preferentially, the third metal layer lines **48** are formed perpendicular to the second metal layer lines **46**.

Although not illustrated, in the preferred embodiment, the sub-circuits **18** are also electrically interconnected with the core memory cells **16** in the core cell area **12** by the third metal layer lines **48**. The third metal layer lines **48** therefore provide the surface "wiring" that electrically connect the sub-circuits **18** with the core memory cells **16** and are generally oriented to extend along an axis parallel to the first metal layer lines **44**. As illustrated in FIG. **2**, the third metal layer lines **50** are connected with the first and second metal layer lines **44**, **46** using contact holes **30** to reach the other layers. Although not illustrated, in the preferred embodiment of the present invention, those skilled in the art should recognize that the contact holes **30** can travel through more than one metal interconnect layer.

During the fabrication process, the periphery area **14** consumed by the electrical circuits **40** is optimized by using the preferred embodiment of the present invention. As previously set forth, the electrical circuits **40** include uniformly oriented rows of transistors **42** and are oriented such that the drain of a respective transistor **42** is adjacent to the source of the next transistor **42** in the rows of transistors **42**. When the transistors **42** of the electrical circuits **40** are oriented in rows, the layout area between respective electrical circuits **40** is minimized without causing short circuits or undesirable leakage currents while still providing electrical interconnection of the electrical circuits **40**. Referring to FIG. **4**, which is an enlarged view of one of the transistors **42** illustrated in FIG. **3**, the source, gate and drain of the transistors **42** used in the electrical circuits **40** are oriented substantially parallel with the first metal layer lines **44**.

Using the first metal layer lines **44** that are formed by the first metal interconnect layer (M1) **22** and the second metal layer lines **46** that are formed by the second metal interconnect layer (M2) **24** to electrically interconnect the electrical circuits **40** to form the sub-circuits **18** minimizes the periphery area **14** consumed in the preferred flash memory **10**. As previously set forth, the first metal layer lines **44** are oriented to extend substantially parallel with the rows of transistors **42** in the periphery area **14**. In addition, the electrical connections with the drains, sources and control gates of the transistors **42** in the electrical circuits **40** are located directly below the first metal layer lines **44**. Since the first metal layer lines **44** are formed substantially parallel with the transistors **42**, additional layout area between the rows of transistors **42** is further minimized.

The second metal layer lines **46** formed by the second metal interconnect layer **34** provide additional electrical interconnections to complete the sub-circuits **18**, thereby minimizing bends required in the first metal layer lines **44** and repositioning of the electrical components under the first metal layer lines **44**. Prior art electrical interconnection connection of the sub-circuits was completed using the first metal interconnect layer, thereby requiring several bends in the first metal layer lines formed by the first metal interconnect layer or increasing the spacing between the electrical components used in the periphery area. The second metal layer lines **46** formed by the second metal interconnect layer (M2) **24** are also substantially straight and oriented substantially perpendicular to the first metal layer lines **44** in the preferred embodiment. As such, the second metal layer lines **46** uniformly pass over the first metal layer lines **44**, thereby minimizing noise created by parasitic capacitance and allowing efficient electrical connections that further minimize consumption of the space available in the periphery area **14**.

The electrical interconnection of the sub-circuits **18** and the core memory cells **16** by the third metal interconnect layer (M3) **26** is also optimized in the preferred embodiment. The third metal interconnect layer (M3) **26** provides electrical interconnection of the first metal interconnect layer **22**, the second metal interconnection layer **24** and therefore connecting the sub-circuits **18** with the core memory cells **16**. As such, the third metal layer lines **48** that are formed by the third metal interconnect layer (M3) **26** are substantially straight and routed on top of the sub-circuits **18** such that consumption of the periphery area **14** is minimized. In addition, the third metal layer lines **48** are sufficiently separated from the first metal layer lines **44** and perpendicularly pass over the second metal layer lines **46** such that noise created by parasitic capacitance is minimized.

In the prior art, the second metal interconnect layer is used to electrically connect the first metal interconnect layer with the core memory cells by routing the second metal lines of the second metal interconnect layer in routing channels between the sub-circuits. The routing of the second metal lines of the second metal interconnect layer between the sub-circuits consumes layout area between the sub-circuits, thereby enlarging the periphery area. In the present invention, no additional layout area is required between the sub-circuits **18** for the third metal interconnect layer (M3) **26** and consumption of the periphery area **14** is further minimized.

While the invention has been described in its currently best known modes of operation and embodiments, other modes and embodiments of the invention will be apparent to those skilled in the art and it is the following claims, including all equivalents, that are intended to define the spirit and scope of the invention.

What is claimed is:

1. A method of electrically interconnecting a periphery area in a flash memory, comprising:

providing a plurality of sub-circuits fabricated in a periphery area of a silicon substrate, wherein each sub-circuit includes at least one electrical circuit having a plurality of circuit components;

partially electrically interconnecting said circuit components with a first metal interconnect layer including a plurality of first metal layer lines that are oriented substantially in one direction;

completing the electrical interconnection of said circuit components in each respective sub-circuit with a second metal interconnect layer including a plurality of

second metal layer lines that are oriented substantially perpendicular to said first metal layer lines; and

electrically interconnecting each respective sub-circuit with a predetermined number of other sub-circuits with a third metal interconnect layer including a plurality of third metal layer lines.

2. The method of claim **1**, further comprising the step of depositing a dielectric layer between said first, second and third metal interconnect layers.

3. The method of claim **1**, wherein a plurality of contact holes are used to electrically interconnect said first metal interconnect layer, said second metal interconnect layer and said third metal interconnect layer at predetermined locations.

4. The method of claim **1**, wherein said circuit components include a plurality of transistors that are arranged in rows substantially parallel with said first metal layer lines.

5. The method of claim **4**, wherein said transistors are at least spaced apart from each other by at least the minimum source/drain spacing rule of said transistors.

6. A method of optimizing circuit layout in a flash memory, comprising:

forming a plurality of sub-circuits in a periphery area of a silicon substrate, wherein each sub-circuit includes at least one electric circuit having a row of transistors;

depositing a first metal interconnect layer on said substrate to partially electrically interconnect the transistors of the sub-circuits, wherein said first metal interconnect layer forms a plurality of first metal layer lines that are oriented to extend substantially in the same direction as said row of transistors;

depositing a second metal interconnect layer above said first metal interconnect layer to complete the electrical interconnection of said respective transistors within said sub-circuits, wherein said second metal interconnect layer forms a plurality of second metal lines oriented to extend substantially perpendicular to said first metal layer lines; and

depositing a third metal interconnect layer above said second metal interconnect layer to electrically interconnect the respective sub-circuits, wherein said third metal layer forms a plurality of metal layer lines oriented to extend substantially parallel to said first metal layer lines.

7. The method of claim **6**, further comprising the step of depositing a dielectric layer between said first, second and third metal interconnect layers.

8. The method of claim **6**, wherein a plurality of contact holes are used to electrically interconnect said first metal interconnect layer, said second metal interconnect layer and said third metal interconnect layer at predetermined locations.

9. The method of claim **6**, wherein said transistors are at least spaced apart from each other by at least the minimum source/drain spacing rule of said transistors.

10. An electrical interconnection system to optimize layout of a periphery area in a memory device, comprising:

a plurality of sub-circuits in a periphery area of a silicon substrate, wherein each of said sub-circuits includes at least one electric circuit with a plurality of circuit components;

a first metal interconnect layer comprising a plurality of first metal layer lines that partially electrically connect the circuit components of said electric circuits, wherein said first metal layer lines are fabricated to be oriented to extend substantially in one direction;

a second metal interconnect layer comprising a plurality of second metal layer lines that complete the electrical connection of said circuit components of said electric circuits, wherein said second metal layer lines are fabricated to be oriented to extend substantially perpendicular to said first metal layer lines; and

a third metal interconnect layer comprising a plurality of third metal layer lines that electrically interconnect said plurality of sub-circuits, wherein said third metal layer lines are fabricated to be oriented to extend substantially parallel to said first metal layer lines.

11. The electrical interconnection system of claim **10**, wherein said plurality of electrical components comprise a plurality of transistors, a plurality of diodes and a plurality of resistors.

12. The electrical interconnection system of claim **11**, wherein said plurality of transistors are oriented to form a plurality of rows of transistors that run substantially parallel with said plurality of first metal layer lines.

13. The electrical interconnection system of claim **11**, wherein each of said plurality of transistors within said plurality of rows of transistors are positioned such that a drain and a source of each of said plurality of transistors is similarly oriented along an axis substantially parallel with said plurality of first metal lines.

14. The electrical interconnection system of claim **11**, wherein said transistors are spaced apart by at least the minimum source/drain spacing rule of said transistors.

* * * * *

EXHIBIT B

US006534805B1

(12) **United States Patent**
Jin

(10) **Patent No.: US 6,534,805 B1**
(45) **Date of Patent: Mar. 18, 2003**

(54) **SRAM CELL DESIGN**

(75) Inventor: **Bo Jin**, Campbell, CA (US)

(73) Assignee: **Cypress Semiconductor Corp.**, San Jose, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/829,510**

(22) Filed: **Apr. 9, 2001**

(51) **Int. Cl.**[7] **H01L 27/10**; H01L 21/84

(52) **U.S. Cl.** **257/206**; 257/211; 257/369; 438/153

(58) **Field of Search** 257/204, 206, 257/211, 369, 390, 393; 438/152, 153, 238

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,621,232 A	*	4/1997	Ohno	257/288
5,654,915 A		8/1997	Stolmeijer et al.	365/156
5,804,477 A	*	9/1998	Lien	438/210
6,103,579 A	*	8/2000	Violette	438/279
6,150,685 A	*	11/2000	Ashida et al.	257/296

FOREIGN PATENT DOCUMENTS

JP	11-195716	* 7/1999	H01L/21/8244
JP	2000243858	* 9/2000	H01L/21/8244
JP	2001168211	* 6/2001	H01L/21/8244

OTHER PUBLICATIONS

Ueshima et al., "A 5–μm^2 Full–CMOS Cell for High–Speed SRAMs Utilizing an Optical–Proximity–Effect Correction (OPC) Technology," 1996, pp. 146–147.

Woo et al., "A High Performance 3.97 μm^2 CMOS SRAM Technology Using Self–Aligned Local Interconnect and Copper Interconnect Metallization," 2 pgs.

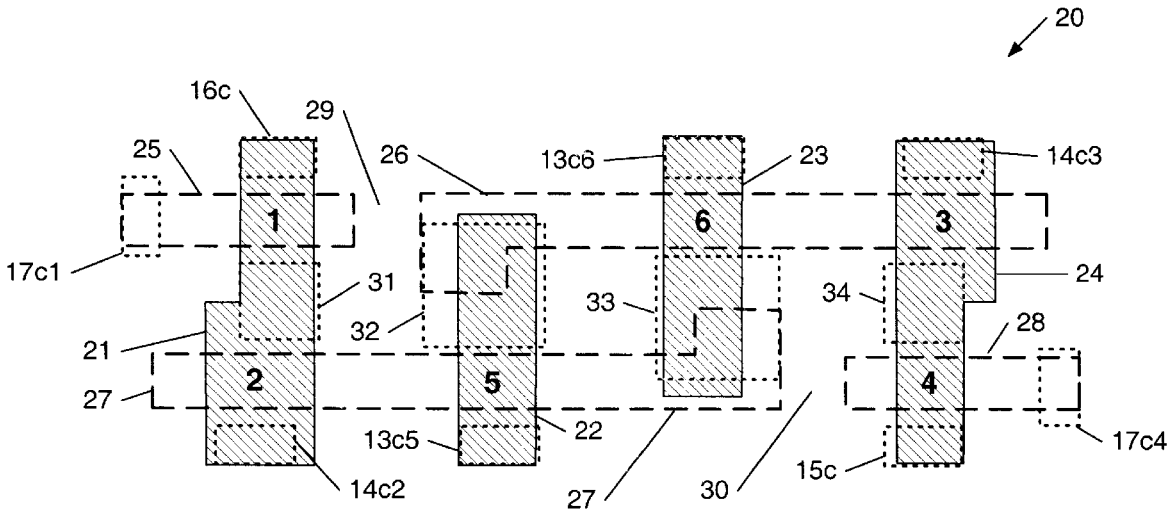
* cited by examiner

Primary Examiner—Mary Wilczewski
Assistant Examiner—Toniae M. Thomas
(74) *Attorney, Agent, or Firm*—Kevin L. Daffer; Conley, Rose & Tayon P.C.

(57) **ABSTRACT**

An embodiment of a memory cell includes a series of four substantially oblong parallel active regions, arranged side-by-side such that the inner active regions of the series include source/drain regions for p-channel transistors, and the outer active regions include source/drain regions for n-channel transistors. Another embodiment of the memory cell includes six transistors having gates substantially parallel to one another, where three of the gates are arranged along a first axis and the other three gates are arranged along a second axis parallel to the first axis. In another embodiment, the memory cell may include substantially oblong active regions arranged substantially in parallel with one another, with substantially oblong local interconnects arranged above and substantially perpendicular to the active regions. A method for fabricating a memory cell may include forming substantially oblong active regions within a semiconductor substrate, and forming substantially oblong local interconnects above and perpendicular to the active regions.

10 Claims, 3 Drawing Sheets



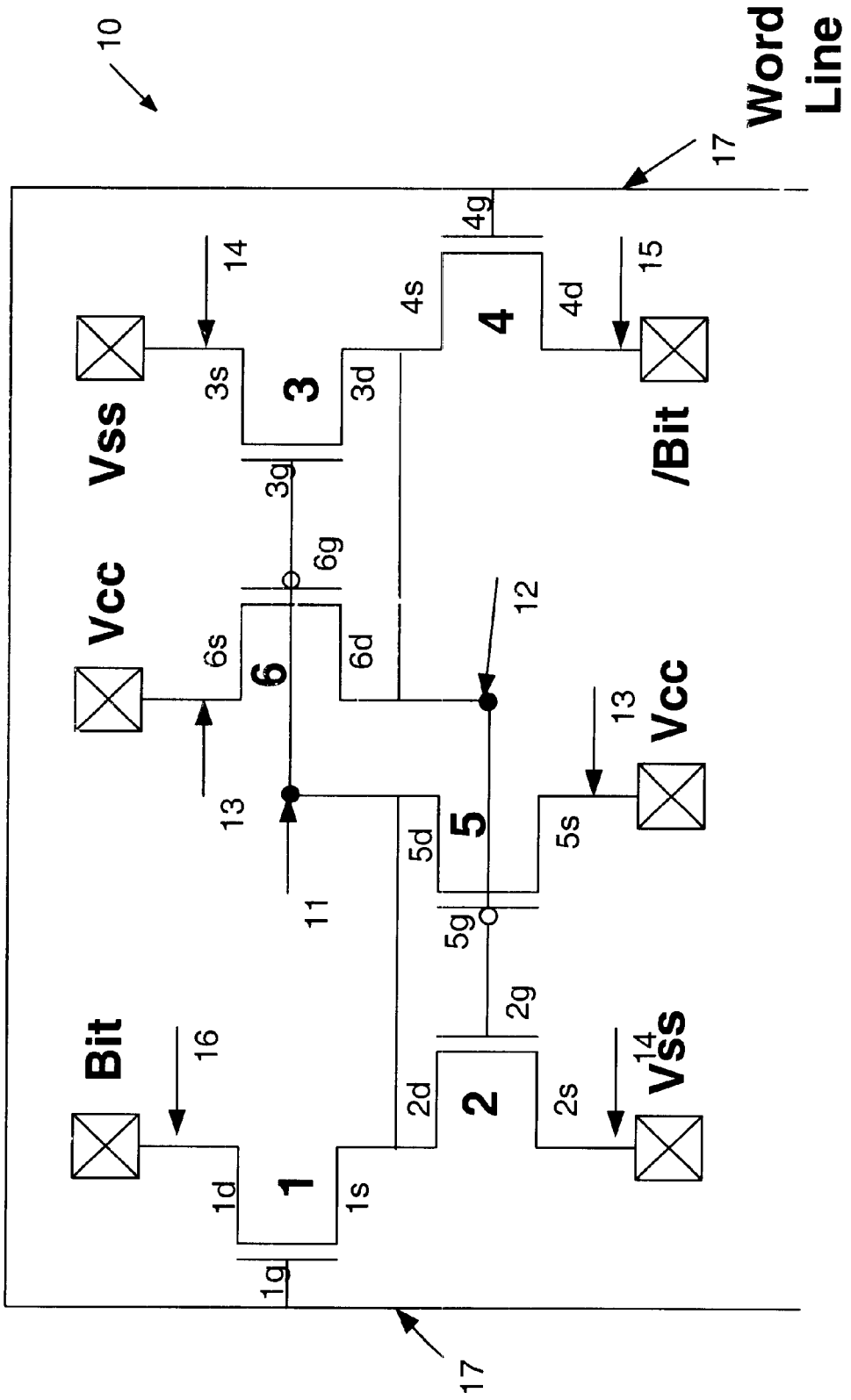

10
Word Line
17
4g
14
15
4s
4d
4
3s
3
3d
Vss
/Bit
3g
6g
Vcc
12
6s
6
6d
13
Vcc
13
5d
5
5s
11
5g
2g
Bit
16
14
Vss
2d
2
2s
1d
1
1s
1g
17

FIG. 1

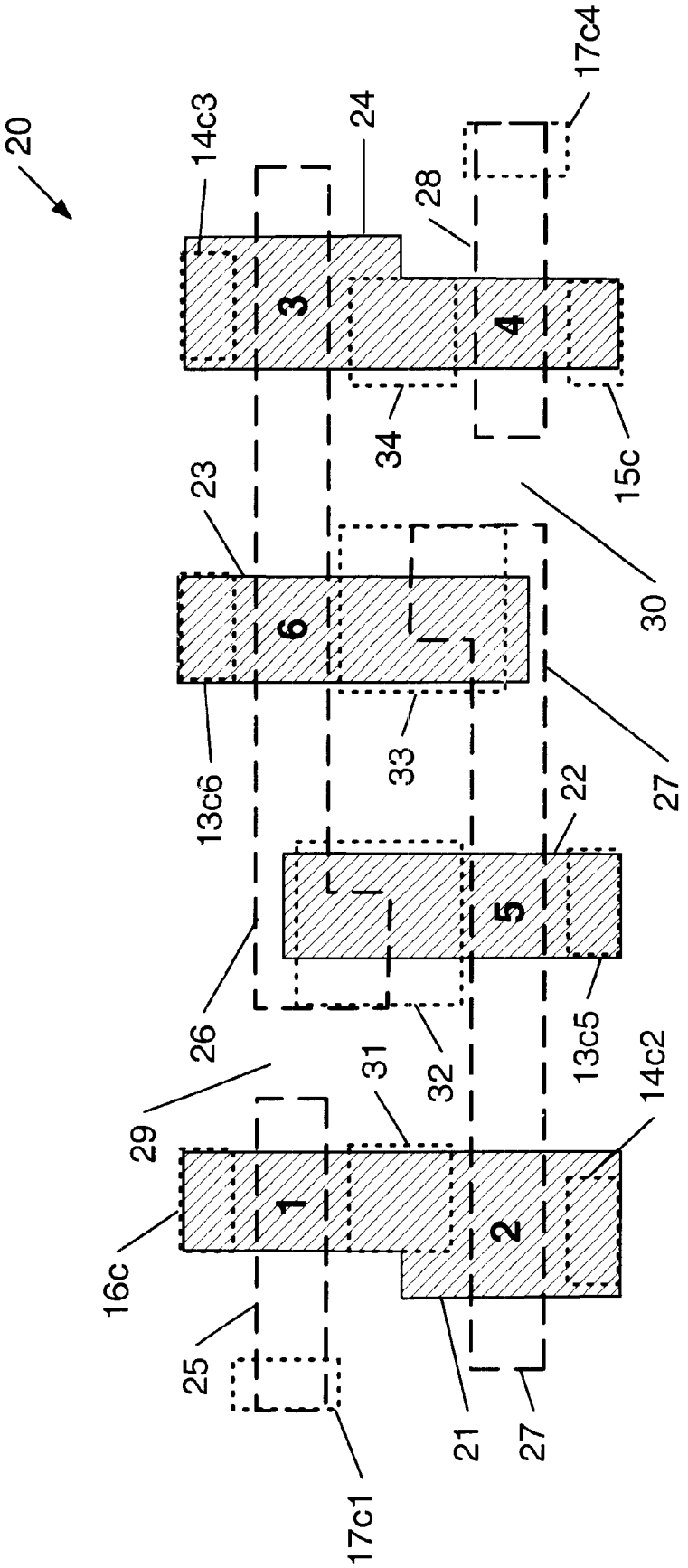

20
14c3
24
17c4
28
3
4
34
15c
23
6
30
13c6
33
27
22
5
26
13c5
31
32
14c2
29
1
2
16c
25
21
27
17c1

FIG. 2

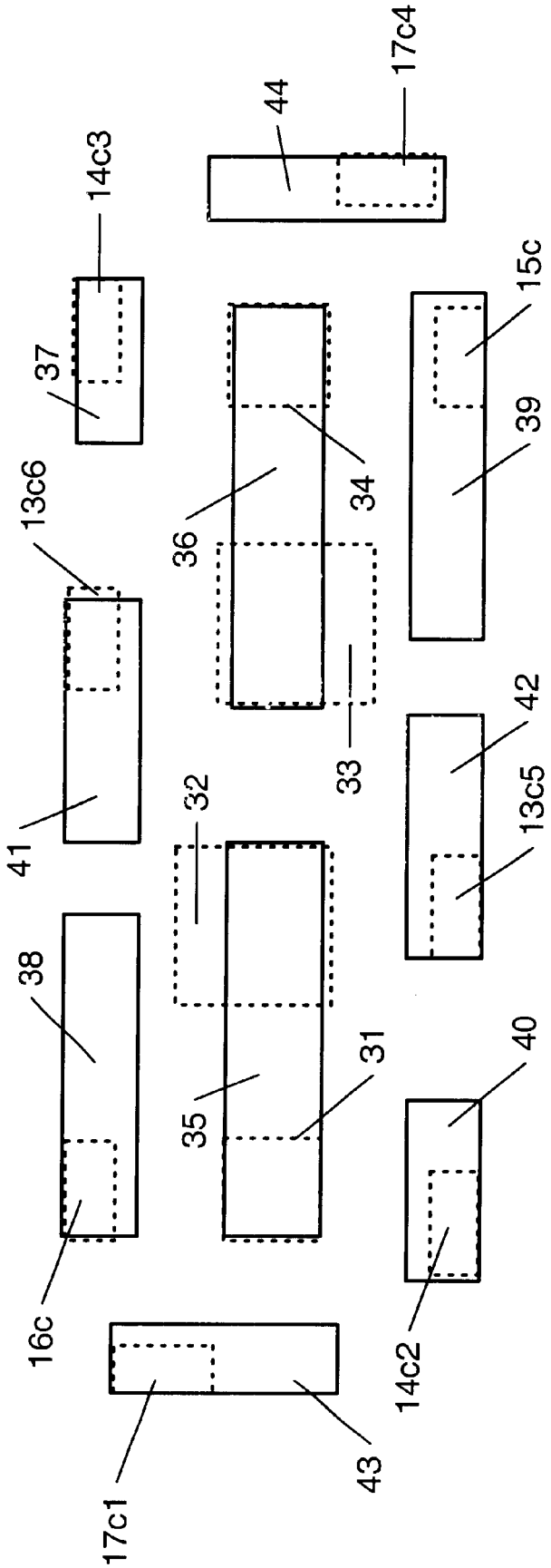

17c4
44
14c3
15c
37
39
13c6
34
36
33
42
41
32
13c5
38
31
40
35
16c
14c2
43
17c1

FIG. 3

SRAM CELL DESIGN

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to semiconductor memory device fabrication, and more particularly to an improved Static Random Access Memory (SRAM) cell design and method of manufacture.

2. Description of the Related Art

The proliferation of computers and other microprocessor-based devices has contributed to an increasing demand for semiconductor memory. Microprocessors are present not only in computers, but in a diverse range of products including automobiles, cellular telephones and kitchen appliances. A conventional microprocessor executes a sequence of instructions and processes information. Frequently, both the instructions and the information reside in semiconductor memory. Therefore, an increased requirement for memory has accompanied the microprocessor boom.

There are various types of semiconductor memory, including Read Only Memory (ROM) and Random Access Memory (RAM). ROM is typically used where instructions or data must not be modified, while RAM is used to store instructions or data which must not only be read, but modified. ROM is a form of non-volatile storage—i.e., the information stored in ROM persists even after power is removed from the memory. On the other hand, RAM storage is generally volatile, and must remain powered-up in order to preserve its contents.

A conventional semiconductor memory device stores information digitally, in the form of bits (i.e., binary digits). The memory is typically organized as a matrix of memory cells, each of which is capable of storing one bit. The cells of the memory matrix are accessed by wordlines and bitlines. Wordlines are typically associated with the rows of the memory matrix, and bitlines with the columns. Raising a wordline activates a given row; the bitlines are then used to read from or write to the corresponding cells in the currently active row. Memory cells are typically capable of assuming one of two voltage states (commonly described as "on" or "off"). Information is stored in the memory by setting each cell in the appropriate logic state. For example, to store a bit having the value 1 in a particular cell, one would set the state of that cell to "on;" similarly, a 0 would be stored by setting the cell to the "off" state. (Obviously, the association of "on" with 1 and "off" with 0 is arbitrary, and could be reversed.)

The two major types of semiconductor RAM, Static Random Access Memory (SRAM) and Dynamic Random Access Memory (DRAM), differ in the manner by which their cells represent the state of a bit. In an SRAM, each memory cell includes transistor-based circuitry that implements a bistable latch. A bistable latch relies on transistor gain and positive (i.e. reinforcing) feedback to guarantee that it can only assume one of two states—"on" or "off." The latch is stable in either state (hence, the term "bistable"). It can be induced to change from one state to the other only through the application of an external stimulus; left undisturbed, it will remain in its original state indefinitely. This is just the sort of operation required for a memory circuit, since once a bit value has been written to the memory cell, it will be retained until it is deliberately changed.

In contrast to the SRAM, the memory cells of a DRAM employ a capacitor to store the "on"/"off" voltage state representing the bit. A transistor-based buffer drives the capacitor. The buffer quickly charges or discharges the capacitor to change the state of the memory cell, and is then disconnected. Ideally, the capacitor then holds the charge placed on it by the buffer and retains the stored voltage level.

DRAMs have at least two drawbacks compared to SRAMs. The first of these is that leakage currents within the semiconductor memory are unavoidable, and act to limit the length of time the memory cell capacitors can hold their charge. Consequently, DRAMs typically require a periodic refresh cycle to restore sagging capacitor voltage levels. Otherwise, the capacitive memory cells would not maintain their contents. Secondly, changing the state of a DRAM memory cell requires charging or discharging the cell capacitor. The time required to do this depends on the amount of current the transistor-based buffer can source or sink, but generally cannot be done as quickly as a bistable latch can change state. Therefore, DRAMs are typically slower than SRAMs. DRAMs offset these disadvantages by offering higher memory cell densities, since the capacitive memory cells are intrinsically smaller than the transistor-based cells of an SRAM.

As microprocessors have become more sophisticated, greater capacity and speed are demanded from the associated memory. SRAMs are widely used in applications where speed is of primary importance, such as cache memory supporting the Central Processing Unit (CPU) in a personal computer. Like most semiconductor devices, SRAMs are fabricated en masse on semiconductor wafers.

Fabrication of a metal-oxide-semiconductor (MOS) integrated circuit involves numerous processing steps. A gate dielectric, typically formed from silicon dioxide ("oxide"), is formed on a semiconductor substrate which is doped with either n-type or p-type impurities. Conductive regions and layers of the device may be isolated from one another by an interlevel dielectric. For each MOS field effect transistor (MOSFET) being formed, a gate conductor is formed over the gate dielectric, and dopant impurities are introduced into the substrate to form a source and drain. Frequently, the integrated circuit will employ a conducting layer to provide a local interconnect function as well. A pervasive trend in modern integrated circuit manufacture is to produce transistors that are as fast as possible and thus have feature sizes as small as possible. Many modern day processes employ features, such as gate conductors and interconnects, which have less than 1.0 μm critical dimension. As feature size decreases, the sizes of the resulting transistor and the interconnect between transistors also decrease. Fabrication of smaller transistors allows more transistors to be placed on a single monolithic substrate, thereby allowing relatively large circuit systems to be incorporated on a single, relatively small die area.

However, integrated circuits become increasingly difficult to manufacture as their dimensions are reduced. Integrated circuits with complex geometries may be particularly difficult to manufacture as dimensions are reduced. Consequently, integrated circuit designs without complex geometries are preferable. Further, reducing the number of steps in an integrated circuit's manufacturing process flow is desired. Reducing the number of processing steps often results in higher profits. Clearly, it would be desirable to have an improved circuit design and method of manufacture to facilitate fabrication of smaller and faster SRAMS.

SUMMARY OF THE INVENTION

The problems outlined above may be addressed by an improved circuit design and method of fabrication disclosed

herein for an integrated circuit, specifically a semiconductor memory device. In the embodiments considered herein, the semiconductor memory device is a static random access memory (SRAM) device, but it is believed that principles disclosed herein are applicable to other types of integrated circuits as well. For example, any device requiring local interconnection of multiple active regions and gates may be suitable.

A memory cell is disclosed herein including a series of four substantially oblong parallel active regions. The active regions are arranged such that the inner active regions comprise source/drain regions for p-channel transistors, while the outer active regions comprise source/drain regions for n-channel transistors. Substantially oblong polysilicon structures may be arranged above and substantially perpendicular to the active regions. Substantially oblong local interconnects may also be arranged above and substantially perpendicular to the active regions. Each active region may include source/drain regions for no more than two transistors. Source/drain contacts to the source/drain regions of the transistors may include at least one shared contact, such that the shared contact is connected to a polysilicon structure as well as an inner source/drain region. A shared contact may be connected to a source/drain contact using a local interconnect. In an embodiment, the local interconnect is dielectrically spaced above the substrate. In an alternate embodiment, the local interconnect may have an upper surface substantially commensurate with the upper surface of at least one respective contact.

A memory cell including six transistors with gates that are substantially parallel to one another is also disclosed. Three of the gates are arranged along a first axis, and the other three are arranged along a second axis parallel to the first axis. Two of the gates along an axis may be arranged within a single polysilicon structure. Of these two, one may be a gate for a p-channel transistor and the other may be a gate for an n-channel transistor. The third gate along an axis may be arranged within another polysilicon structure. This second polysilicon structure may be electrically coupled to a respective local wordline. Each of the two local wordlines may be electrically coupled to a global wordline, which in an embodiment comprises metal. Also included in the memory cell may be a shared contact arranged between the axes and in contact with a source/drain region of a p-channel transistor along one axis and a polysilicon structure along the other axis. In an embodiment, the memory cell may also include an active region substantially perpendicular to the axes and electrically coupled to a bitline where the bitline extends across the entire length of the memory cell. The bitline may be substantially parallel to the active region, and the length of the bitline may be less than a third of the width of the cell.

In an embodiment, a memory cell is disclosed having substantially oblong active regions arranged substantially in parallel with one another within a semiconductor substrate. The memory cell also has multiple local interconnects arranged above and substantially perpendicular to the active regions, where the interconnects are also substantially oblong and in parallel with one another. In an embodiment, the memory cell may also include substantially square local interconnects such that all interconnects are either substantially oblong or substantially square. In an embodiment, the memory cell may also include a shared contact that is electrically coupled to an active region and a polysilicon structure abutting said active region.

Also disclosed herein is a method of fabricating a memory cell including forming substantially oblong active regions arranged substantially in parallel with one another within a semiconductor substrate. The method also includes forming substantially oblong local interconnects arranged above the semiconductor substrate where the interconnects are substantially in parallel with one another and substantially perpendicular to the active regions. In an embodiment, forming the local interconnects may include etching a trench through a dielectric material and depositing a conductive material into the trench. The method may include forming substantially oblong polysilicon structures arranged above the semiconductor substrate where the polysilicon structures are substantially in parallel with one another and substantially perpendicular to the active regions. Forming the polysilicon structures may include forming an access polysilicon structure for each of two access transistor gates within the memory cell, where the access polysilicon structures do not extend across the entire memory cell. In an embodiment, forming the memory cell may include forming a global wordline, where the wordline is dielectrically spaced above the active regions and where the wordline is electrically coupled to the access polysilicon structures.

The improved circuit design and method of fabrication disclosed herein may provide numerous advantages. This circuit design may be improved because the memory cell layout may allow the features to be arranged in such a way as to minimize cell size. Another advantage of the improved circuit design is the substantially parallel features that reduce manufacturing complexities, particularly in photolithography. As a result of the substantially parallel layout, reducing feature sizes to increase device speeds and/or to minimize memory cell size may be facilitated. In addition, the substantially parallel layout may change the aspect ratio of the memory cell such that the bitlines may be reduced in length, thus advantageously decreasing bitline resistivity and increasing memory cell performance. Furthermore, this circuit design may be improved because of the symmetrical the layout design, which may improve noise margins. Yet another advantage of the improved circuit design is the elimination of polysilicon wordlines that traverse the entirety of the memory cell. Elimination of such polysilicon wordlines may minimize cell size by reducing the density of features required on the polysilicon layer of the cell. Reducing the amount of polysilicon in the wordlines may also result in increased use of a metal layer to perform the wordline function, thus advantageously decreasing wordline resistivity and increasing memory cell performance. A further advantage of the improved circuit design is that the improved polysilicon layer may partially perform local interconnecting functions. Therefore, the subsequent local interconnect layer may be greatly simplified and the local interconnects may also be arranged substantially in parallel. The improved layout may further enable the use of a trench local interconnect layer, thus reducing the number of processing steps.

BRIEF DESCRIPTION OF THE DRAWINGS

Other objects and advantages of the invention will become apparent upon reading the following detailed description and upon reference to the accompanying drawings in which:

FIG. **1** shows the transistor configuration of an embodiment of an improved SRAM memory cell;

FIG. **2** represents a layout of the active regions and polysilicon structures for the embodiment of FIG. **1**; and

FIG. **3** represents a layout of the local interconnect for the regions and structures shown in FIG. **2**.

While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof are shown by way of example in the drawings and will herein be described in detail. It should be understood, however, that the drawings and detailed description thereto are not intended to limit the invention to the particular form disclosed, but on the contrary, the intention is to cover all modifications, equivalents and alternatives falling within the spirit and scope of the present invention as defined by the appended claims.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

A circuit diagram for an improved double wordline SRAM memory cell is shown in FIG. **1**. Generally stated, SRAM memory cells may be formed by interconnecting two CMOS inverters together so that the input of a first inverter is tied to the output of a second inverter and vice versa to form a positive feedback orientation. Such configuration of two inverters is commonly referred to as a bi-stable latch as described above. The inverters include transistors commonly referred to as latch transistors. In the embodiment of FIG. **1**, transistors **2** and **5** form the first inverter, transistors **3** and **6** form the second inverter, the node representing the input of the first inverter is labeled **12**, and the node representing the input of the second inverter is labeled **11**. Thus, the memory cell **10** comprises four latch transistors **2**, **3**, **5**, and **6** and two access transistors **1** and **4**, each of which has a drain, source and gate. The latch transistors **2**, **3**, **5**, and **6** include a pair of n-channel pull-down transistors **2** and **3** and a pair of p-channel load transistors **5** and **6**.

The inverters are connected as follows to form the bi-stable latch of FIG. **1**. A drain **2***d* of pull-down transistor **2** is coupled to a drain **5***d* of load transistor **5** at node **11** and a drain **3***d* of a pull-down transistor **3** is coupled to a drain **6***d* of load transistor **6** at node **12**. These nodes **11** and **12** store opposite logic states (i.e., one is a logic "1" while the other is a logic "0"). Sources **5**s and **6**s of transistors **5** and **6** are coupled to a common power line **13** (hereinafter referred to as a "Vcc line") while sources **2***s* and **3***s* of pull-down transistors **2** and **3** are coupled to a common ground line **14** (hereinafter referred to as a "Vss line"). Gates **2***g* and **5***g* of transistors **2** and **5** are coupled together and connected to the node **12** and gates **3***g* and **6***g* of transistors **3** and **6** are coupled together and connected to node **11**.

Such interconnections create positive feedback, which allows the memory cell to store data as either a "high" or "low" input (i.e., a logic "1" or "0"). Data is stored in these memory cells during a "write cycle" and that data is subsequently read during a "read" cycle. The n-channel access transistors **1** and **4** are coupled to the memory cell **10** to allow communication between the cell **10** and an external device through a pair of complementary bitlines **15** and **16**. With respect to the access transistors **1** and **4**, each source **1***s* and **4***s* is coupled to nodes **11** and **12**, respectively. A drain **1***d* of access transistor **1** is coupled to a bitline **16**, referred to as "Bit," which operates as a data line to read data from and write data into the memory cell **10**. A drain **4***d* of a second access transistor **4** is similarly coupled to a complementary bitline **15** called "/Bit," In addition, both gates **1***g* and **4***g* are coupled to wordline **17**.

Applying a positive voltage to the wordline **17** turns on both access transistors **1** and **4**, thus accessing memory cell **10**. This allows one of the two bitlines **15** and **16** to sense the contents of the memory cell **10** based on the voltage at either node **11** or **12**. For example, if node **11** is at a high (Vcc) voltage and node **12** is at the ground potential (Vss), when the wordline **17** is brought to a high voltage, the pull-down transistor **3** and the access transistor **4** are both turned on and will thus pull the bitline "/Bit" **15** down toward the ground potential Vss. Moreover, the load transistor **5** and the access transistor **1** are also tuned on; thus the bitline "Bit" **16** will be pulled up towards the Vcc potential. Thus the state of the cell **10** (either "1" or "0") can be determined by sensing the difference in potential between the bitlines **15** and **16**.

Conversely, writing a "1" or a "0" into the cell **10** can be accomplished by forcing the bitline **15** or the bitline **16** to either Vcc or Vss and then raising the wordline **17**. The potential placed on either the bitline "/Bit" **15** or the bitline "Bit" **16** will then be transferred to the node **11** or **12**, respectively, forcing the cell **10** into either a corresponding "**1**" state or a "**0**" state.

Shown in FIG. **2** is an embodiment of a layout **20** that may be used to form in silicon the memory cell **10** represented in FIG. **1**. (Elements appearing in more than one figure retain the same item numbers throughout the figures.) FIG. **2** presents a top-down view of an exemplary memory cell layout **20**. Layout **20** illustrates the active regions, isolation regions, polysilicon structures, and contact structures that may be used to form the typical metal oxide semiconductor (MOS) transistors, NMOS and PMOS, used in a typical CMOS SRAM. In the embodiment of FIG. **2**, NMOS transistors **1**–**4** are formed within active regions **21** and **24**, and PMOS transistors **5** and **6** are formed within active regions **22** and **23**. The active regions are formed within a semiconductor substrate. The semiconductor substrate may preferably be a silicon substrate doped n-type and p-type in the vicinity of the p-channel transistors and the n-channel transistors, respectively. More specifically, the semiconductor substrate may include n-type and p-type well regions formed in a monocrystalline silicon substrate, or in an epitaxial silicon layer grown on a monocrystalline silicon substrate.

Active regions, i.e., areas where active transistors are to be formed, are labeled **21**–**24** and are arranged side-by-side and substantially parallel to each other. Diffusion regions are also to be formed within the active regions **21**–**24**. For example, diffusion regions may be lightly doped drain regions and heavily doped source/drain regions formed in active regions adjacent to the transistor gate structures. Dielectric isolation regions such as **29** and **30** separate active regions from one another. Isolation regions may be formed by a number of techniques such as shallow trench isolation (STI), recessed oxide isolation (ROI), or local oxidation of silicon (LOCOS). Isolation regions may therefore be field oxide regions, which serve to isolate separate active regions on the semiconductor layer from one another.

In the embodiment of FIG. **2**, NMOS active regions **21** and **24** are utilized for the formation of two transistors each, a pass transistor and a latch transistor. Polysilicon structures **25** and **27** are arranged above active region **21** to form gates of pass transistor **1** and latch transistor **2**, respectively. Similarly, above active region **24**, polysilicon structures **28** and **26** are arranged to form gates of pass transistor **4** and latch transistor **3**, respectively. Consequently, active regions **21** and **24** each have two gate conductors arranged above them. In this embodiment, no active region has more than two gate conductors arranged above it, and therefore no active region forms more than two transistors.

In the embodiment of FIG. **2**, the active regions are substantially oblong, and in some cases may be substantially rectangular as well. For example, PMOS active regions,

such as active regions **22** or **23** shown in FIG. **2**, may have a length that is substantially constant across the width of the region, as well as a width that is substantially constant along the length of the region. However, if an NMOS active region is forming an access transistor and a latch transistor, it may have some variation in width although the length may be substantially constant across the width of the region. For example, active regions **21** and **24** as shown in FIG. **2** are each forming access transistors **1** and **4**, respectively, and latch transistors **2** and **3**, respectively. By design, access transistors frequently have widths that are smaller than those of adjacent latch transistors. The active region is thus designed to ensure the stability of the SRAM. This design feature is commonly referred to as the "beta ratio." A beta ratio is defined as the width of the latch transistor divided by the width of the pass transistor. To ensure circuit stability, the beta ratio should be >1. In an embodiment, the beta ratio is approximately 1.5. Therefore, in an embodiment, the width of the access transistor is approximately ⅔ the width of the latch transistor. Consequently, an NMOS active region may be considered to be substantially oblong if the length of the region is substantially constant and if the width of the region varies by approximately ⅓ or less along the length of the region. Further, an NMOS active region may be considered to be substantially oblong if the length of the region is substantially constant and the width of the region by design varies only with the respective widths of the access and latch transistors. In an embodiment, "substantially oblong" may refer to any region or structure having a length that is greater than or equal to approximately three times its maximum width. Active regions as described above are oblong with respect to, for example, the markedly "L-shaped" regions formed in layouts for which two transistors are arranged at right angles to each other.

Each transistor includes a gate electrode formed above an active region, arranged between a pair of source/drain regions, and separated from the substrate by a relatively thin dielectric. In a preferred embodiment, gate electrodes are arranged within polysilicon structures **25–28** to form transistors **1–6** as shown in FIG. **2**. The polysilicon may be deposited by, for example, using chemical vapor deposition (CVD) of silicon from a silane source. However, the gate electrodes may comprise any suitable conductive material such as polysilicon, aluminum, or copper. Therefore structures **25–28** are not limited to polysilicon. For example, the gate electrodes may include multiple layers of material, such as a doped polysilicon and a silicide. A silicide may be formed from a polysilicon layer upon which a layer of refractory metal such as cobalt or titanium has been formed. Upon heating the refractory metal, a reaction between the polysilicon and the cobalt or titanium may result in the formation of a silicide such as cobalt silicide or titanium silicide. In an embodiment, the width of the gate electrode (or channel length of the transistor) may be approximately 0.12 microns, but may also be larger or smaller depending on the transistor that is being formed.

Many SRAMs are single wordline cells, meaning that there is only one local wordline arranged within each cell. It therefore follows that a double wordline cell has two local wordlines arranged within each cell. The two local wordlines are coupled either within the cell or outside the cell. Frequently, single wordline cells and double wordline cells have some similarities. Both cells may have wordlines that are polysilicon and that extend continuously from one side of the cell to the other. However, the embodiment of FIG. **2** presents a split double wordline cell. That is, the local wordlines of the embodiment of FIG. **2** do not extend continuously from one side of the memory cell to the other. For example, polysilicon structures **25** and **28** each comprise local wordlines and each couples to global wordline **17** (not shown). However, neither polysilicon structure extends continuously from one side of the memory cell to the other. Thus, polysilicon structures **25** and **28** are coupled together and to global wordline **17** outside of cell layout **20**. Polysilicon structures **25** and **28** are coupled such that they rise and fall in potential together.

The split double wordline of the embodiment of FIG. **2** may allow the memory cell size to be reduced while also improving memory cell performance. Each wordline-coupled polysilicon structure **25** and **28** may be electrically coupled to a polysilicon structure arranged within an immediately adjacent cell. However, an entire row of memory cells would not be coupled together by one or two continuous polysilicon wordlines as is frequently found in SRAM circuits. For example, a continuous conductive wordline is frequently formed when the gate electrodes are formed. Because this type of continuous wordline is eliminated in the embodiment of FIG. **2**, the density of the features required by the polysilicon layer of the memory cell is reduced. Consequently, the die size required to accommodate the polysilicon layer may be reduced, and the layout of the remaining polysilicon features may be improved to provide a more manufacturable memory cell. Further, the performance of the memory cell may be improved, as memory cell addressing times may no longer be limited by the resistivity of continuous polysilicon wordlines. As noted above, split local wordline polysilicon structures **25** and **28** are each electrically coupled to the global wordline **17**. Except for those areas where global wordline **17** is electrically coupled to polysilicon structures **25** and **28**, global wordline **17** may be a metal dielectrically spaced from the polysilicon structures.

Conductive regions and layers of the memory cell may be isolated from one another by dielectrics. Examples of dielectrics may include silicon dioxide (SiO_2), tetraethylorthosilicate glass (TEOS), silicon nitride (Si_xN_y), silicon oxynitride ($SiO_xN_y(H_z)$), and silicon dioxide/silicon nitride/silicon dioxide (ONO). The dielectrics may be grown or may be deposited by physical deposition such as sputtering or by a variety of chemical deposition methods and chemistries such as chemical vapor deposition. Additionally, the dielectrics may be undoped or may be doped, for example with boron, phosphorus, boron and phosphorus, or fluorine, to form a doped dielectric layer such as borophosphosilicate glass (BPSG), phosphosilicate glass (PSG), and fluorinated silicate glass (FSG).

The coupling of global wordline **17** to the gates of access transistors **1** and **4** may require the use of contacts **17***c***1** and **17***c***4**, respectively, as shown in FIG. **2**. At various stages in the fabrication of semiconductor devices, it may be necessary to form openings in a dielectric layer to allow for contact to underlying regions or layers. Generally, an opening through a dielectric exposing a diffusion region or an opening through a dielectric layer between polysilicon and an interconnect or first metal layer is called a "contact opening" or a "contact hole." If the opening is substantially oblong, it may be referred to as a "trench." An opening in other dielectric layers such as an opening through an intermetal dielectric layer is commonly referred to as a "via." For purposes of this disclosure, henceforth "contact opening" may be used to refer to a contact opening and/or a via. Similarly, "trench" may be used to refer to any substantially oblong contact opening and/or via. A contact opening or trench may expose a diffusion region within the silicon

substrate such as a source or drain, or may expose some other layer or structure such as an underlying metallization layer, a local interconnect layer, or a gate structure. Conductive contact structures may be formed above the source/drain regions, and interconnects may overlie the contact structures and may connect neighboring contact structures. These contact structures to diffusion regions may be isolated from an adjacent gate structure by a dielectric spacer. The dielectric spacer may also isolate the gate from the diffusion region.

Contacts **17***c***1** and **17***c***4** could be used to electrically couple polysilicon structures **25** and **28**, respectively, to a dielectrically spaced global wordline **17**. Because global wordline **17** is dielectrically spaced, it may be metal. Metal global wordlines have significantly lower resistivity than the polysilicon global wordlines or the silicided global wordlines used in some memory cells. For example, the Rs of polysilicon is approximately 50 ohms/square and the Rs of silicide is approximately 20 ohms/square. However, metals have Rs values that are significantly smaller, for example, less than approximately 0.2 ohms/square. Consequently, metal global wordlines have significantly lower resistivities than polysilicon or silicide global wordlines. Therefore, metal global wordlines may be much longer than global wordlines of polysilicon or silicide, yet still have lower resistivities. As such, metal global wordlines may significantly reduce memory cell addressing times.

For example, in the embodiment of FIG. **2**, the cell aspect ratio is approximately 1:3.5. An aspect ratio as used herein generally describes the ratio between the height and width of a semiconductor feature. In the case of the cell aspect ratio, the ratio is taken between the height of the memory cell and the width of the memory cell. That is, the memory cell of the embodiment of FIG. **2** has a width (along the wordline direction) that is approximately 3.5 times its height (along the bitline direction). Further, the global wordline traversing the memory cell of the embodiment of FIG. **2** is longer than, for example, a global wordline of a memory cell with an equivalent area and an aspect ratio of 1:1. However, despite being longer, the metal global wordline of the cell of FIG. **2** would have a lower resistance than a polysilicon or silicide global wordline of a memory cell with a 1:1 aspect ratio due to the much lower metal Rs value. Further, the bitline traversing the memory cell of the embodiment of FIG. **2** would be shorter than the bitline of a cell with an equivalent area and an aspect ratio of 1:1. Consequently, the embodiment of FIG. **2** may provide significantly faster memory cell addressing times.

A double wordline may advantageously simplify cell layout by allowing all transistor gates to be arranged substantially parallel to one another. Such arrangement may eliminate active regions that are arranged one perpendicular to another and/or continuous active regions that have 90 degree "knees." Furthermore, double wordlines may also eliminate similar arrangement of the polysilicon structures. Thus, the double wordlines may facilitate the use of substantially oblong features for the active regions and the polysilicon structures. That is, the active regions may be arranged substantially parallel to one another, and the polysilicon structures may also be arranged substantially parallel to one another. Such arrangement may facilitate the future reduction of critical dimensions(CDs) for these layers since eliminating complicated geometries aids manufacturability for the photolithography process, as discussed above.

If active regions are arranged such that complex geometries are eliminated, it follows that the isolation regions are also arranged such that complex geometries are eliminated. Elimination of complex geometries may also reduce manufacturing complexities for other manufacturing processes. For example, a shallow trench isolation process ("STI") is more robust with the elimination of complex geometries. Shallow trench isolation is primarily used for isolating active regions, and is rapidly replacing local oxidation of silicon, or LOCOS, isolation structures. STI processes do not exhibit the lateral extension of oxide into the active region of the device, known as the "bird's beak", that is common with LOCOS processes. In contrast to a LOCOS process, the STI process involves patterning the semiconductor substrate, etching shallow trenches into the substrate, filling the trenches with dielectric, and removing the dielectric from the substrate so that the remaining dielectric has an upper surface approximately commensurate with the upper surface of the semiconductor substrate. That is, the STI process results in an essentially planar upper surface. Thus, STI processes further enable the reduction of CDs for polysilicon structures. For example, photolithography processes require substantially planar upper surfaces to pattern sub-micron features due to the very small depth of focus. Thus, STI processes may enable production of smaller, and therefore faster, transistors. However, the STI process is also more robust if the geometry is not complex. As geometries continue to shrink, the STI geometry also shrinks thus making the required patterning, etching, and dielectric fill more difficult. If the STI geometry is complex, the required processing may be much more difficult. Thus, simplified STI geometries are preferred.

Turning now to FIG. **2**, bitline "Bit" **16** is coupled to the drain of transistor **1** by contact **16***c*, while bitline "/Bit" **15** is coupled to the drain of transistor **4** by contact **15***c*. The sources of transistors **2** and **3** are coupled to Vss by contacts **14***c***2** and **14***c***3**, respectively, while the sources of transistors **5** and **6** are coupled to Vcc by contact **13***c***5** and **13***c***6**, respectively. Bitlines **15** and **16**, Vcc **13**, and Vss **14** are also dielectrically spaced from the polysilicon structures. The bitlines, Vss, and Vcc may be contacted to each memory cell using what is commonly referred to as the first metal layer. The "first metal layer" refers to the first conductive layer above the local interconnect layer, and may be a misnomer in those cases where the local interconnect layer utilizes a metal.

Contact regions **31**–**34** as shown in FIG. **2** may be used for the local interconnections of gates and drains as described in the discussion of FIG. **1** above. Many types of contacts may be used; e.g., self-aligned contacts (SAC) or shared contacts may be used. Further, contacts may be substantially rectangular or substantially square. For example, SAC may be substantially square. Contacts **31** and **34** represent openings in the dielectric that expose portions of active regions **21** and **24**, respectively. In particular, contact **31** allows for a contact to the source of transistor **1** and the drain of transistor **2** and contact **34** allows for contact to the source of transistor **4** and the drain of transistor **3**. Contacts **32** and **33** represent openings in the dielectric, which expose portions of active regions **22** and **23**, respectively, as well as exposing portions of polysilicon structures **26** and **27**, respectively. As such, contacts **32** and **33** are shared contacts. That is, contacts **32** and **33** allow for contact not only to diffusion regions, but also to polysilicon structures. For example, as shown in FIG. **2**, contact **32** allows for a contact to the drain of transistor **5**, but it also allows for a contact to the gates of transistors **3** and **6** via polysilicon structure **26**. Polysilicon structure **26** is connected to the gates of transistors **3** and **6** as discussed above, and contact **32** allows for a contact to polysilicon structure **26**. In particular, contact **32** allows for a contact to

a portion of polysilicon structure **26** which may be wider than the remainder of polysilicon structure **26**. Although polysilicon structure **26** has a wider portion, it is nonetheless considered to be substantially oblong. A polysilicon structure may be considered to be substantially oblong if the length of the polysilicon structure is greater than about three times the width of the polysilicon structure. Furthermore, a polysilicon structure may be considered to be substantially oblong despite having a substantially wider region if the wider region solely accommodates a contact region. Similarly, contact **33** is a shared contact which allows for contact to the drain of transistor **6** as well as polysilicon structure **27** and hence gates of transistors **2** and **5**. Consequently, shared contacts **32** and **33** reduce the number of contacts needed by the subsequent local interconnect. Shared contacts such as **32** and **33** may be substantially rectangular.

Local interconnections are generally used for short runs relative to much longer metal conductors used for global connections. Thus, the term "local interconnect" may refer to the function of connecting features within a circuit, or it may refer to a distinct process layer that exclusively performs such short connections. Therefore, a process layer may perform local interconnecting functions, yet not be termed a "local interconnect layer." That is, a polysilicon layer may partially provide local interconnect in addition to providing transistor gates. The polysilicon layer may be configured such that multiple gates are connected, thus forming a local interconnect.

For example, polysilicon structures **26** and **27** provide local interconnect because they each form multiple gates. Polysilicon structure **26** is arranged above active region **23** and active region **24** to form gates of PMOS latch transistor **6** and NMOS latch transistor **3**, respectively. Polysilicon structure **27** is arranged above active region **21** and active region **22** to form gates of NMOS latch transistor **2** and PMOS latch transistor **5**, respectively. Therefore, polysilicon structures **26** and **27** each include two gates. Thus, polysilicon structures **26** and **27** may each perform a local interconnecting function because they each connect two separate gate conductors together. As a result, the subsequent local interconnect layer of the embodiment of FIG. **2** may be simplified because polysilicon structures **26** and **27** are performing a local interconnecting function that is frequently provided by a local interconnect layer. As will be discussed in more detail below, the embodiment of FIG. **2** utilizes an improved, simplified local interconnect layer layout as a result of the local interconnecting function provided by polysilicon structures **26** and **27**.

FIG. **3** illustrates a local interconnect layer which may be used in conjunction with the layout shown in FIG. **2**. Contacts **31–34** may be formed through a dielectric material arranged above the topography of the features shown in FIG. **2**. The dielectric material is preferably deposited using physical vapor deposition (PVD), chemical vapor deposition (CVD), plasma enhanced chemical vapor deposition (PECVD), low pressure chemical vapor deposition (LPCVD), or spin on deposition techniques. In the vapor depositions, a silane or TEOS source, for example, may be used to produce the dielectric. In the spin on depositions, a liquid material of, for example, silicon (i.e., silicates, siloxanes, or silsesquioxanes) or TEOS may be spin-on deposited and subsequently cured. The dielectric material preferably has a thickness of approximately 6000–8000 angstroms, but other thicknesses may be suitable depending on the particular process used. Contacts **31–34** are also shown in FIG. **2**. In the completed circuit, contacts **31** and **32** are connected to local interconnect region **35** which electrically connects the source of transistor **1** to the drains of transistors **2** and **5** and the gates of transistors **3** and **6** as shown in FIGS. **1** and **2**. Thus, local interconnect region **35** completes the connections necessary to form node **11** as illustrated in FIG. **1**. Similarly, local interconnect region **36** and contacts **33** and **34** will complete the connections necessary to form node **12** as illustrated in FIG. **1**. It can be seen from FIG. **3** that local interconnect structures **35** and **36** are substantially rectangular and require connection to the structures of the previous layers in only two areas. This is an improvement over local interconnect structures which require multi-limbed, non-oblong interconnect structures and three connections to the structures of previous layers. The local interconnect of the embodiment of FIG. **3** is simplified as a result of the interconnect provided by the polysilicon structures **26** and **27**.

In one embodiment of the local interconnect, the dielectric material is arranged upon the topography and planarized using, for example, chemical mechanical polishing, or "CMP." The dielectric is then etched to form contact openings **31–34**. Next, an adhesion or "glue" layer may be formed by blanket depositing an adhesion material onto the sidewalls and bottom of openings **31–34**. Suitable adhesion materials include titanium (Ti), titanium-tungsten (TiW), titanium nitride (TiN), or tungsten silicides (WSi_x). Adhesion layers may be used in those cases where conductive materials may adhere poorly to the dielectric material. Adhesion layers may be used, for example, to compensate for poor adhesion characteristics of tungsten and some tungsten alloys to silicon dioxide dielectric films. After an adhesion layer has been formed, a conductive material layer may then be deposited. In the case where tungsten, "W," is the conductive material deposited, this step is referred to as W plug. However, any suitable conducting material may be applied in a manner appropriate to the material. Next, the contact plugs are planarized such that they are substantially commensurate to an uppermost surface of the dielectric material. This planarization step may be accomplished via CMP. Next, an additional conducting layer is deposited. This conducting layer is commonly referred to as the local interconnect layer. The local interconnect layer may be made from a material having higher resistivity than the metals or conducting materials used for global interconnects. Suitable local interconnect materials may include polysilicon, doped polysilicon, refractory metal, silicide, or combinations of these. The local interconnect layer is subsequently patterned and etched, thus completing an embodiment of a local interconnect process flow.

In another local interconnect embodiment, the dielectric material is also arranged upon the topography and planarized. However, the dielectric is then etched to form trench openings that will suffice not only as contacts, but also as the local interconnects themselves. That is, a single trench etched through the dielectric opens up contact regions **31** and **32** as well as interconnect region **35** as shown in FIG. **3**. Similarly, an additional trench may be etched into the dielectric for contact regions **33** and **34** as well as interconnect region **36**. Other openings in the dielectric, such as Vss contact opening **14***c***3**, may be also etched into the dielectric at this time. Next, an adhesion layer may be formed by blanket depositing an adhesion material onto the sidewalls and bottom of the trenches and contact openings. After the adhesion layer has been formed, a conductive material layer may then be deposited. This conductive material is planarized such that the uppermost surfaces of the trenches and contacts are substantially commensurate to an uppermost

surface of the dielectric material. Again, this planarization step may be accomplished via CMP. At this point, the local interconnect layer process for the preferred embodiment is complete. Consequently, multiple processing steps are eliminated as compared to the above-described formation of local interconnects dielectrically spaced above the substrate. In particular, eliminated are the deposition of an additional conductive layer, the patterning of the additional conductive layer, and the etching of the additional conductive layer.

Returning to FIG. **3**, in addition to local interconnect structures **35** and **36** and their respective contacts, other interconnect structures of interest are shown. Interconnect structures **38** and **39** correspond to bitlines "Bit" and "/Bit" and contacts **16***c* and **15***c*, respectively. It is noted that the contacts shown correspond to those shown in FIG. **2**. Similarly, interconnect structures **37** and **40** correspond to Vss and contacts **14***c***3** and **14***c***2**, respectively, while interconnect structures **41** and **42** correspond to Vcc and contacts **13***c***6** and **13***c***5**, respectively. Following the interconnect layer, dielectric material may be arranged upon the interconnect layer, and contacts may subsequently be formed through this dielectric material. Another conducting layer, typically some type of metal, may be formed above the dielectric material and patterned and etched. This metal layer may be referred to as a first metal layer. Frequently, this metal layer contacts the bitlines, Vcc, and Vss globally across several memory cells. Yet another dielectric material may be arranged upon this first metal layer, and upon the dielectric material may be deposited a second metal layer. This second metal layer may be used as a global wordline. For example, interconnect structures **43** and **44** correspond to global wordline **17** and contacts **17***c***1** and **17***c***4**.

It will be appreciated to those skilled in the art having the benefit of this disclosure that the embodiments described are believed applicable to semiconductor memories. Furthermore, although illustrated with reference to SRAMs, the system and method disclosed herein may be adapted to other types of memory devices. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense as to possibly numerous architectures, circuitry, and methodologies which fall within the spirit and scope of the present invention.

What is claimed is:

1. A memory cell comprising a series of four substantially oblong active regions formed within a semiconductor substrate and arranged side-by-side with long axes substantially parallel, wherein each of the inner active regions of the series comprises a pair of source/drain regions for a respective p-channel transistor, and each of the outer active regions of the series comprises a pair of source/drain regions for a respective n-channel transistor.

2. The memory cell as recited in claim **1**, further comprising a plurality of substantially oblong polysilicon structures arranged above and substantially perpendicular to the active regions.

3. The memory cell as recited in claim **1**, wherein each active region comprises source/drain regions for no more than two transistors.

4. The memory cell as recited in claim **2**, further comprising source/drain contacts to the source/drain regions of transistors, wherein at least one of the source/drain contacts comprises a shared contact to one of the inner active regions and one of the polysilicon structures.

5. The memory cell as recited in claim **4**, further comprising a series of substantially oblong local interconnects arranged substantially perpendicular to the active regions, wherein the shared contact is connected to another of the source/drain contacts by one of the local interconnects.

6. The memory cell as recited in claim **5**, wherein the local interconnects are dielectrically spaced above the semiconductor substrate.

7. The memory cell as recited in claim **5**, wherein the local interconnects have an upper surface that is substantially coplanar with an upper surface of the source/drain contacts.

8. A memory cell comprising a plurality of substantially oblong active regions formed in a semiconductor substrate and arranged substantially in parallel with one another, and a plurality of substantially oblong local interconnects above said substrate that extend only partially across the memory cell and are arranged substantially in parallel with one another and substantially perpendicular to said active regions.

9. The memory cell as recited in **8**, further comprising substantially square local interconnects above said substrate, wherein all local interconnects within the cell are either substantially oblong or substantially square.

10. The memory cell as recited in claim **8**, further comprising a shared contact to one of the active regions and a polysilicon structure, wherein the polysilicon structure is arranged substantially perpendicular to said active region and a portion of the polysilicon structure abuts a portion of said active region.

* * * * *

US006534805C1

(12) EX PARTE REEXAMINATION CERTIFICATE (10327th)

United States Patent

Jin

(10) Number: US 6,534,805 C1

(45) Certificate Issued: Oct. 14, 2014

(54) **SRAM CELL DESIGN**

(75) Inventor: **Bo Jin**, Campbell, CA (US)

(73) Assignee: **Cypress Semiconductor Corporation**, San Jose, CA (US)

Reexamination Request:
No. 90/011,833, Aug. 2, 2011

Reexamination Certificate for:

Patent No.:	**6,534,805**
Issued:	**Mar. 18, 2003**
Appl. No.:	**09/829,510**
Filed:	**Apr. 9, 2001**

(51) **Int. Cl.**
H01L 27/10 (2006.01)

(52) **U.S. Cl.**
USPC **257/206**; 257/211; 257/369; 438/153

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

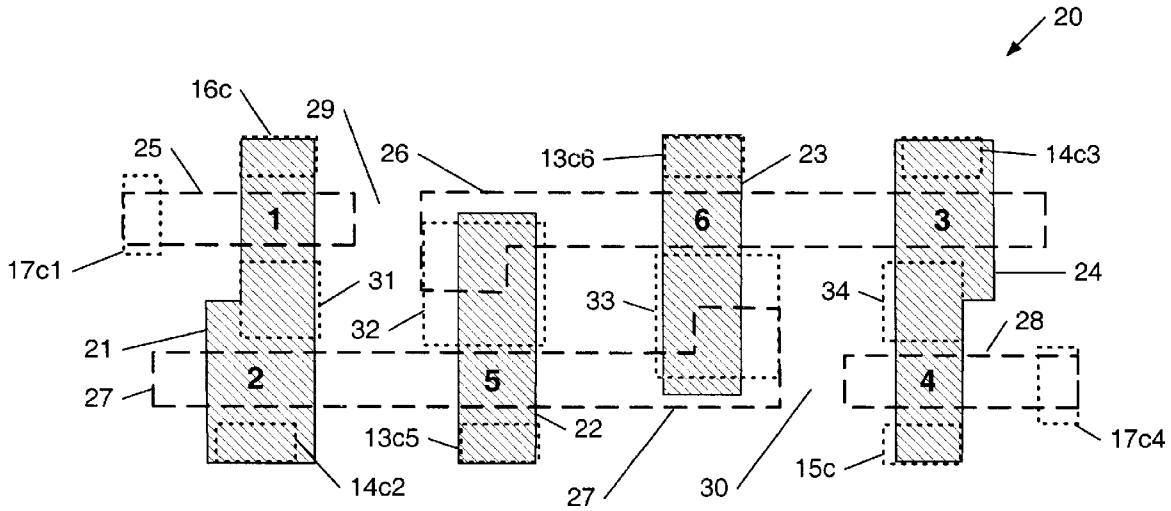
(56) **References Cited**

To view the complete listing of prior art documents cited during the proceeding for Reexamination Control Number 90/011,833, please refer to the USPTO's public Patent Application Information Retrieval (PAIR) system under the Display References tab.

Primary Examiner — Leonardo Andujar

(57) **ABSTRACT**

An embodiment of a memory cell includes a series of four substantially oblong parallel active regions, arranged side-by-side such that the inner active regions of the series include source/drain regions for p-channel transistors, and the outer active regions include source/drain regions for n-channel transistors. Another embodiment of the memory cell includes six transistors having gates substantially parallel to one another, where three of the gates are arranged along a first axis and the other three gates are arranged along a second axis parallel to the first axis. In another embodiment, the memory cell may include substantially oblong active regions arranged substantially in parallel with one another, with substantially oblong local interconnects arranged above and substantially perpendicular to the active regions. A method for fabricating a memory cell may include forming substantially oblong active regions within a semiconductor substrate, and forming substantially oblong local interconnects above and perpendicular to the active regions.



EX PARTE REEXAMINATION CERTIFICATE ISSUED UNDER 35 U.S.C. 307

THE PATENT IS HEREBY AMENDED AS INDICATED BELOW.

Matter enclosed in heavy brackets [ ] appeared in the patent, but has been deleted and is no longer a part of the patent; matter printed in italics indicates additions made to the patent.

AS A RESULT OF REEXAMINATION, IT HAS BEEN DETERMINED THAT:

Claims **1-6** are cancelled.

Claim **8** is determined to be patentable as amended.

Claims **9-10**, dependent on an amended claim, are determined to be patentable.

New claims **11-61** are added and determined to be patentable.

Claim **7** was not reexamined.

8. A memory cell comprising a plurality of substantially oblong active regions formed in a semiconductor substrate and arranged substantially in parallel with one another, and a plurality of substantially oblong local interconnects above said substrate that extend only partially across the memory cell and are arranged substantially in parallel with one another and substantially perpendicular to said active regions*; and*

a single local interconnect layer comprising local interconnects corresponding to bitlines and a global wordline.

11. The memory cell of claim 8, wherein each local interconnect of the single local interconnect layer is substantially oblong.

12. The memory cell of claim 8, the single local interconnect layer comprising local interconnects corresponding to common power and common ground.

13. The memory cell of claim 12, wherein each local interconnect of the single local interconnect layer is substantially oblong.

14. The memory cell of claim 8, comprising:

a first contact to one of the substantially oblong active regions;

a shared contact to another one of the substantially oblong active regions and a polysilicon structure;

a first substantially oblong local interconnect that connects the first contact and the shared contact,

wherein the single local interconnect layer comprises the first substantially oblong local interconnect.

15. The memory cell of claim 8, comprising:

a first contact to one of the substantially oblong active regions;

a shared contact to another one of the substantially oblong active regions and a polysilicon structure;

a first substantially oblong local interconnect that connects the first contact and the shared contact,

wherein the first substantially oblong local interconnect is formed from a trench opening as a contact to one of the substantially oblong active regions.

16. A memory cell, comprising:

a series of four substantially oblong active regions formed within a semiconductor substrate and arranged side-by-side with long axes substantially parallel,

wherein each inner active region of the series comprises a pair of source/drain regions for a respective p-channel transistor, and

wherein each outer active region of the series comprises a pair of source/drain regions for a respective n-channel transistor; and

a single local interconnect layer comprising local interconnects corresponding to bitlines and a global wordline.

17. The memory cell of claim 16, wherein each local interconnect of the single local interconnect layer is substantially oblong.

18. The memory cell of claim 16, the single local interconnect layer comprising local interconnects corresponding to common power and common ground.

19. The memory cell of claim 18, wherein each local interconnect of the single local interconnect layer is substantially oblong.

20. The memory cell of claim 16, comprising:

a first contact to one of the outer active regions;

a shared contact to one of the inner active regions and a polysilicon structure;

a first substantially oblong local interconnect that connects the first contact and the shared contact,

wherein the single local interconnect layer comprises the first substantially oblong local interconnect.

21. The memory cell of claim 16, comprising:

a first contact to one of the outer active regions;

a shared contact to one of the inner active regions and a polysilicon structure;

a first substantially oblong local interconnect that connects the first contact and the shared contact,

wherein the first substantially oblong local interconnect is formed from a trench opening as a contact to one of the active regions.

22. The memory cell of claim 16, comprising:

a plurality of substantially oblong polysilicon structures arranged above and substantially perpendicular to the active regions.

23. The memory cell of claim 16, comprising:

source/drain contacts to the source/drain regions of transistors,

wherein at least one of the source/drain contacts comprises a shared contact to one of the inner active regions and one of the polysilicon structures.

24. The memory cell of claim 23, comprising:

a series of substantially oblong local interconnects arranged substantially perpendicular to the active regions,

wherein the shared contact is connected to another of the source/drain contacts by one of the substantially oblong local interconnects.

25. The memory cell of claim 24, wherein the substantially oblong local interconnects are dielectrically spaced above the semiconductor substrate.

26. The memory cell of claim 24, wherein the substantially oblong local interconnects have an upper surface that is substantially coplanar with an upper surface of the source/drain contacts.

27. A memory cell, comprising:

a series of four substantially oblong active regions formed within a semiconductor substrate and arranged side-by-side with long axes substantially parallel,

wherein each inner active region of the series comprises a pair of source/drain regions for a respective p-channel transistor, and

wherein each outer active region of the series comprises a pair of source/drain regions for a respective n-channel transistor;

a first contact to one of the outer active regions;

a shared contact to both a polysilicon structure and to one of the inner active regions; and

a substantially oblong local interconnect that connects the first contact and the shared contact,

wherein the substantially oblong local interconnect overlaps both the polysilicon structure and said one of the inner active regions.

28. The memory cell of claim 27, wherein the substantially oblong local interconnect is dielectrically spaced above the semiconductor substrate.

29. A memory cell, comprising:

a series of four substantially oblong active regions formed within a semiconductor substrate and arranged side-by-side with long axes substantially parallel,

wherein each inner active region of the series comprises a pair of source/drain regions for a respective p-channel transistor, and

wherein each outer active region of the series comprises a pair of source/drain regions for a respective n-channel transistor;

a first contact to one of the outer active regions;

source/drain contacts to the source/drain regions of transistors, wherein at least one of the source/drain contacts comprises a shared contact to both a polysilicon structure and to one of the inner active regions; and

a substantially oblong local interconnect that connects the first contact and the shared contact,

wherein the substantially oblong local interconnect overlaps both the polysilicon structure and said one of the inner active regions, and

wherein the substantially oblong local interconnect has an upper surface that is substantially coplanar with an upper surface of the source/drain contacts.

30. A memory cell, comprising:

a plurality of substantially oblong active regions formed in a semiconductor substrate and arranged substantially in parallel with one another;

a plurality of substantially oblong local interconnects above said substrate that extend partially across the memory cell and are arranged substantially in parallel with one another and substantially perpendicular to the active regions;

a first contact to one of the active regions;

a shared contact to both a polysilicon structure and to another one of the active regions; and

a substantially oblong local interconnect that connects the first contact and the shared contact,

wherein the substantially oblong local interconnect overlaps both the polysilicon structure and said another one of the active regions.

31. The memory cell of claim 30, comprising:

substantially square local interconnects above the substrate,

wherein local interconnects within the cell are one of substantially oblong and substantially square.

32. The memory cell of claim 30,

wherein the polysilicon structure is arranged substantially perpendicular to said another one of the active regions, and

wherein a portion of the polysilicon structure abuts a portion of the said another one of the active regions.

33. A memory cell, comprising:

a series of four substantially oblong active regions formed within a semiconductor substrate and arranged side-by-side with long axes substantially parallel,

wherein each inner active region of the series comprises a pair of source/drain regions for a respective p-channel transistor,

wherein each outer active region of the series comprises a pair of source/drain regions for a respective n-channel transistor, and

wherein a width varies along a length of each of the active regions.

34. The memory cell of claim 33, wherein the length varies along the width of each of the active regions.

35. The memory cell of claim 33, comprising:

a plurality of substantially oblong polysilicon structures arranged above and substantially perpendicular to the active regions.

36. The memory cell of claim 35, comprising:

source/drain contacts to the source/drain regions of transistors,

wherein at least one of the source/drain contacts comprises a shared contact to one of the inner active regions and one of the polysilicon structures.

37. The memory cell of claim 36, comprising:

a series of substantially oblong local interconnects arranged substantially perpendicular to the active regions,

wherein the shared contact is connected to another of the source/drain contacts by one of the local interconnects.

38. The memory cell of claim 37, wherein the local interconnects are dielectrically spaced above the semiconductor substrate.

39. The memory cell of claim 37, wherein the substantially oblong local interconnects have an upper surface that is substantially coplanar with an upper surface of the source/drain contacts.

40. A memory cell, comprising:

a series of four substantially oblong active regions formed within a semiconductor substrate and arranged side-by-side with long axes substantially parallel,

wherein each inner active region of the arrangement comprises a pair of source/drain regions for a respective p-channel transistor,

wherein each outer active region of the arrangement comprises a pair of source/drain regions for a respective n-channel transistor, and

wherein a width varies along at least a portion of a length and the length varies along at least a portion of the width of each of the four substantially oblong active regions.

41. The memory cell of claim 40, comprising:

a plurality of substantially oblong polysilicon structures arranged above and substantially perpendicular to the active regions.

42. The memory cell of claim 41, comprising:

source/drain contacts to the source/drain regions of transistors,

wherein at least one of the source/drain contacts comprises a shared contact to one of the inner active regions and one of the polysilicon structures.

43. The memory cell of claim 42, comprising:

a series of substantially oblong local interconnects arranged substantially perpendicular to the active regions,

wherein the shared contact is connected to another of the source/drain contacts by one of the local interconnects.

44. The memory cell of claim 43, wherein the local interconnects are dielectrically spaced above the semiconductor substrate.

45. The memory cell of claim 43, wherein the substantially oblong local interconnects have an upper surface that is substantially coplanar with an upper surface of the source/drain contacts.

46. A memory cell, comprising

a series of four substantially oblong active regions formed in a semiconductor substrate and arranged substantially in parallel with one another,

wherein a width varies along a length of each of the active regions; and

a plurality of substantially oblong local interconnects that above the substrate and extend partially across the memory cell and are arranged substantially in parallel with one another and substantially perpendicular to the active regions.

47. The memory cell of claim 46, wherein the length varies along the width of each of the active regions.

48. The memory cell of claim 46, comprising:

substantially square local interconnects above the substrate,

wherein all local interconnects within the cell are either substantially oblong or substantially square.

49. The memory cell of claim 46, comprising:

a shared contact to one of the active regions and a polysilicon structure,

wherein the polysilicon structure is arranged substantially perpendicular to the active region, and

wherein a portion of the polysilicon structure abuts a portion of the active region.

50. A memory cell, comprising:

a series of four substantially oblong active regions formed in a semiconductor substrate and arranged substantially in parallel with one another,

wherein a width varies along at least a portion of a length and the length varies along at least a portion of the width of each of the plurality of active regions; and

a plurality of substantially oblong local interconnects above the substrate that extend partially across the memory cell and are arranged substantially in parallel with one another and substantially perpendicular to the active regions.

51. The memory cell of claim 50, comprising:

substantially square local interconnects above the substrate,

wherein local interconnects within the cell are one of substantially oblong and substantially square.

52. The memory cell of claim 50, comprising:

a shared contact to one of the active regions and a polysilicon structure,

wherein the polysilicon structure is arranged substantially perpendicular to the active region, and

wherein a portion of the polysilicon structure abuts a portion of the active region.

53. A memory cell, comprising:

a series of four substantially oblong active regions formed within a semiconductor substrate and arranged side-by-side with long axes substantially parallel,

wherein each inner active region of the series comprises a pair of source/drain regions for a respective p-channel transistor, and

wherein each outer active region of the series comprises a pair of source/drain regions for a respective n-channel transistor;

a first metal layer above the semiconductor substrate,

wherein the first metal layer contacts bitlines, common power, and common ground globally across a plurality of memory cells; and

a second metal layer above the first metal layer,

wherein the second metal layer is configured as a global wordline.

54. The memory cell of claim 53, comprising:

a plurality of substantially oblong polysilicon structures arranged above and substantially perpendicular to the active regions.

55. The memory cell of claim 54, comprising:

source/drain contacts to the source/drain regions of transistors,

wherein at least one of the source/drain contacts comprises a shared contact to one of the inner active regions and one of the polysilicon structures.

56. The memory cell of claim 55, comprising:

a series of substantially oblong local interconnects arranged substantially perpendicular to the active regions,

wherein the shared contact is connected to another of the source/drain contacts by one of the local interconnects.

57. The memory cell of claim 56, wherein the local interconnects are dielectrically spaced above the semiconductor substrate.

58. The memory cell of claim 56, wherein the substantially oblong local interconnects have an upper surface that is substantially coplanar with an upper surface of the source/drain contacts.

59. A memory cell, comprising:

a plurality of substantially oblong active regions formed in a semiconductor substrate and arranged substantially in parallel with one another;

a plurality of substantially oblong local interconnects above the substrate and extending partially across the memory cell and are arranged substantially in parallel with one another and substantially perpendicular to the active regions;

a first metal layer above the semiconductor substrate,

wherein the first metal layer contacts bitlines, common power, and common ground globally across a plurality of memory cells; and

a second metal layer above the first metal layer,

wherein the second metal layer is configured as a global wordline.

60. The memory cell of claim 59, comprising:

substantially square local interconnects above the substrate,

wherein local interconnects within the cell are one of substantially oblong and substantially square.

61. The memory cell of claim 59, comprising:

a shared contact to one of the active regions and a polysilicon structure,

wherein the polysilicon structure is arranged substantially perpendicular to the active region, and

wherein a portion of the polysilicon structure abuts a portion of the active region.

* * * * *

EXHIBIT C

US006642573B1

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

(10) **Patent No.: US 6,642,573 B1**
(45) **Date of Patent: Nov. 4, 2003**

(54) **USE OF HIGH-K DIELECTRIC MATERIAL IN MODIFIED ONO STRUCTURE FOR SEMICONDUCTOR DEVICES**

(75) Inventors: **Arvind Halliyal**, Cupertino, CA (US); **Mark T. Ramsbey**, Sunnyvale, CA (US); **Wei Zhang**, Sunnyvale, CA (US); **Mark W. Randolph**, San Jose, CA (US); **Fred T. K. Cheung**, San Jose, CA (US)

(73) Assignee: **Advanced Micro Devices, Inc.**, Sunnyvale, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/097,912**

(22) Filed: **Mar. 13, 2002**

(51) **Int. Cl.**[7] **H01L 29/788**; H01L 21/8247

(52) **U.S. Cl.** **257/316**; 438/261; 438/785; 438/954

(58) **Field of Search** 257/316; 438/261, 438/954, 763, 785

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,043,940	A	*	8/1991	Harari	365/185.03
5,120,672	A	*	6/1992	Mitchell et al.	438/261
5,888,870	A	*	3/1999	Gardner et al.	438/261
6,008,091	A	*	12/1999	Gregor et al.	438/261
6,013,553	A		1/2000	Wallace et al.	438/287
6,163,049	A		12/2000	Bui	257/321
6,171,900	B1	*	1/2001	Sun	438/240
6,242,306	B1		6/2001	Pham et al.	438/262
6,309,927	B1		10/2001	Au et al.	438/261
6,438,030	B1		8/2002	Hu et al.	365/185.18
6,461,905	B1		10/2002	Wang et al.	438/183
2002/0106536	A1		8/2002	Lee et al.	428/702
2002/0115252	A1		8/2002	Haukka et al.	438/240
2002/0142624	A1		10/2002	Levy et al.	438/786
2002/0190311	A1		12/2002	Blomme et al.	257/321

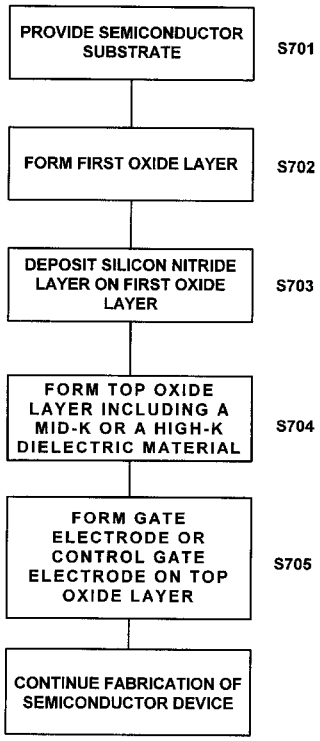
* cited by examiner

Primary Examiner—Richard Booth
(74) *Attorney, Agent, or Firm*—Renner, Otto, Boisselle & Sklar, LLP

(57) **ABSTRACT**

A process for fabrication of a semiconductor device including a modified ONO structure, comprising forming the modified ONO structure by providing a semiconductor substrate; forming a first dielectric material layer on the semiconductor substrate; depositing a silicon nitride layer on the first dielectric material layer; and forming a top dielectric material layer, wherein at least one of the bottom dielectric material layer and the top dielectric material layer comprise a mid-K or a high-K dielectric material. The semiconductor device may be, e.g., a SONOS two-bit EEPROM device or a floating gate flash device including the modified ONO structure.

22 Claims, 4 Drawing Sheets



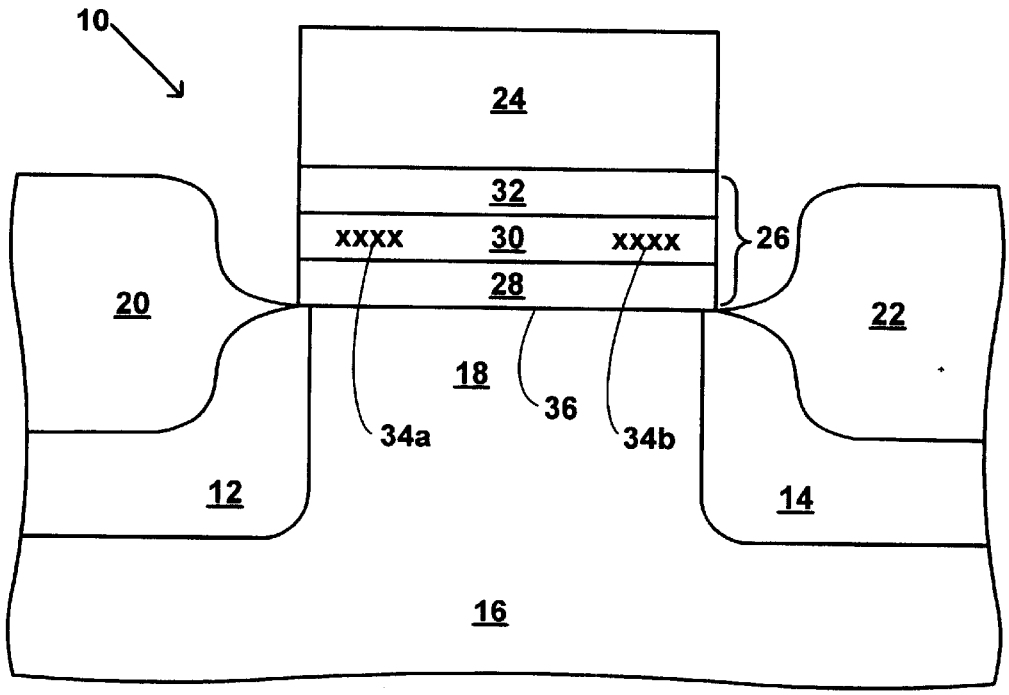

10
24
32
xxxx
30
xxxx
26
28
20
22
18
36
34a
34b
12
14
16

Fig. 1

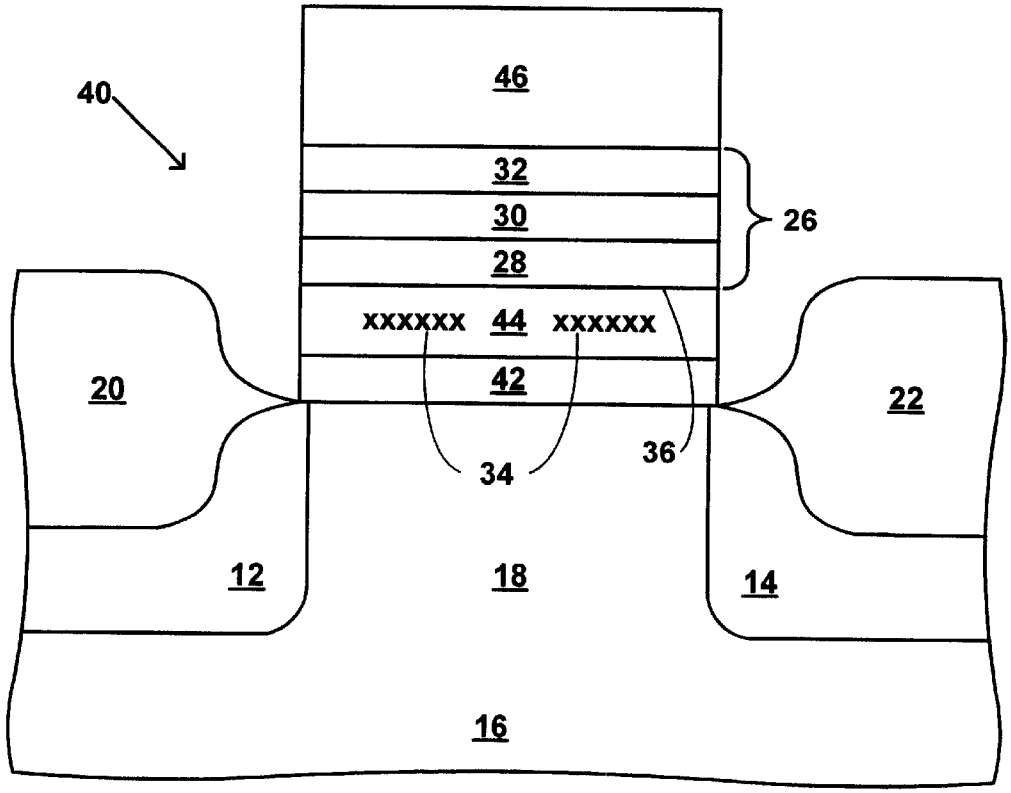

40
46
32
30
26
28
xxxxxx
44
xxxxxx
20
42
22
34
36
12
18
14
16

Fig. 2

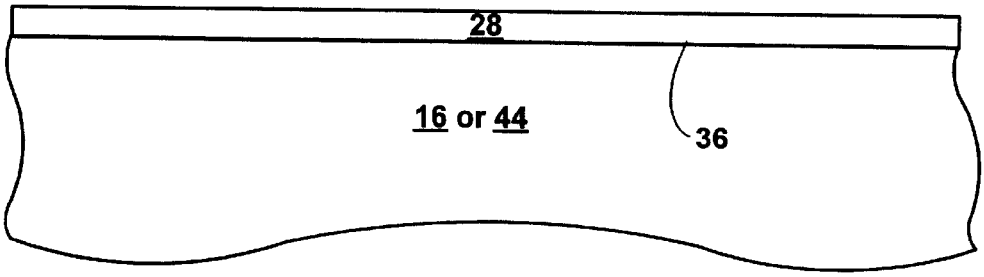


Fig. 3

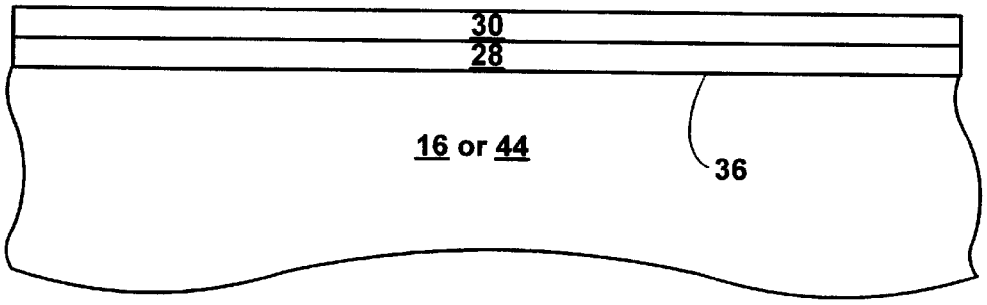


Fig. 4

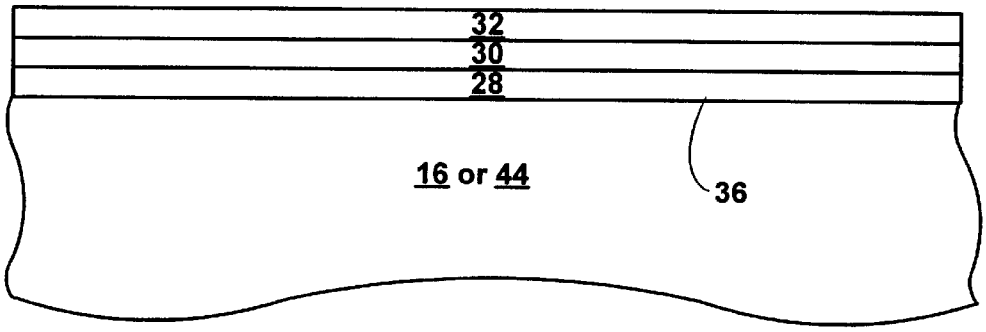


Fig. 5

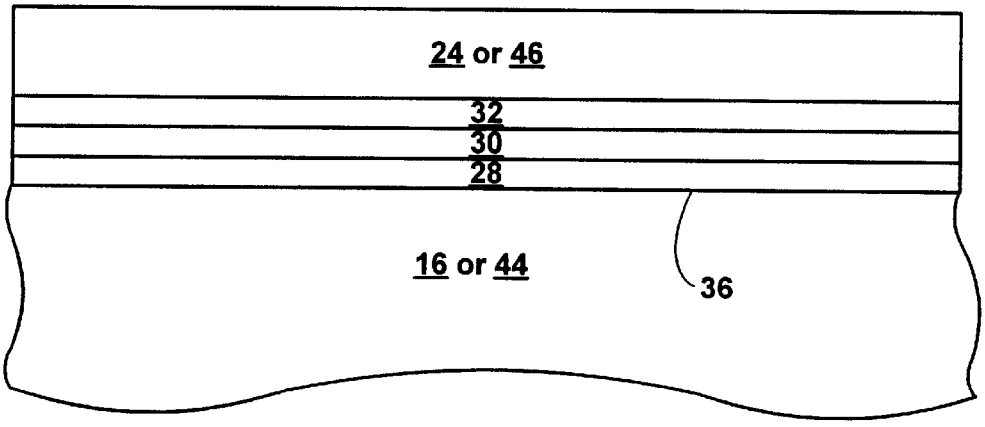


Fig. 6

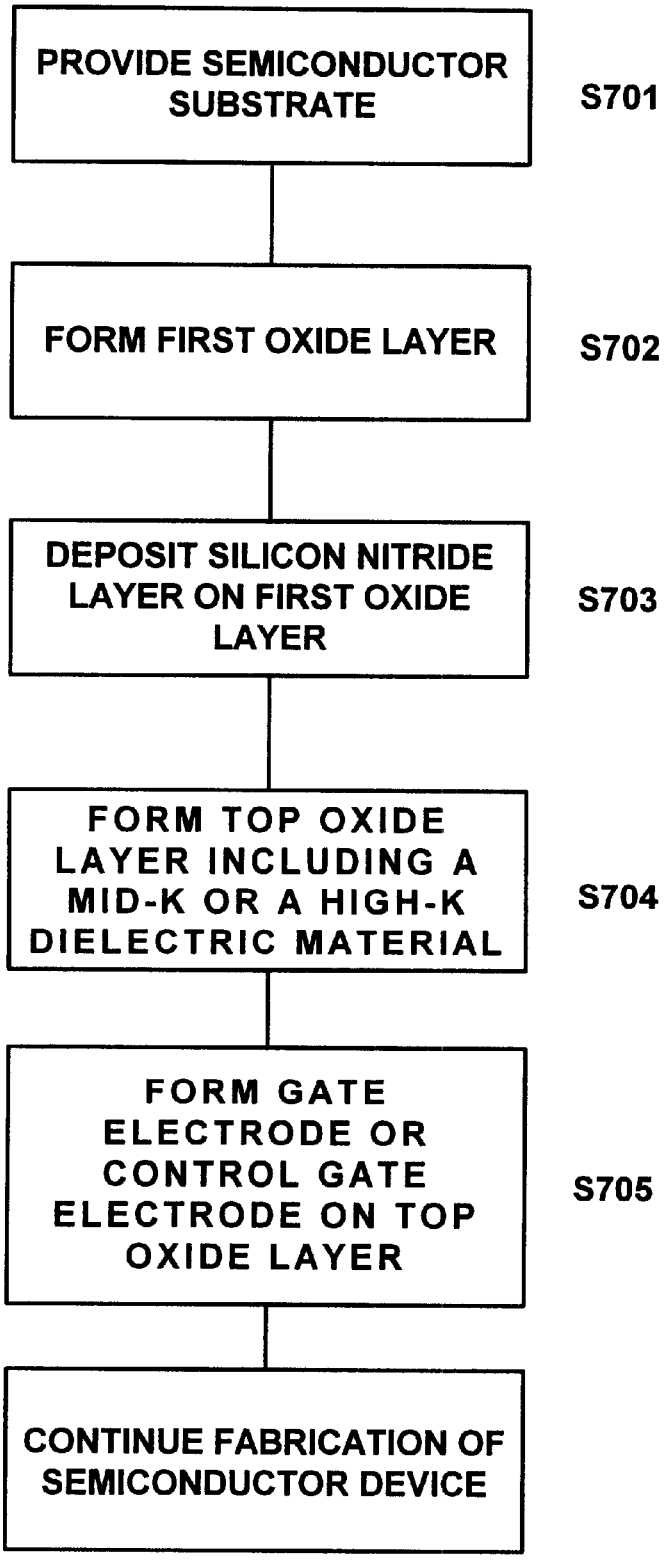

PROVIDE SEMICONDUCTOR SUBSTRATE
S701
FORM FIRST OXIDE LAYER
S702
DEPOSIT SILICON NITRIDE LAYER ON FIRST OXIDE LAYER
S703
FORM TOP OXIDE LAYER INCLUDING A MID-K OR A HIGH-K DIELECTRIC MATERIAL
S704
FORM GATE ELECTRODE OR CONTROL GATE ELECTRODE ON TOP OXIDE LAYER
S705
CONTINUE FABRICATION OF SEMICONDUCTOR DEVICE

Fig. 7

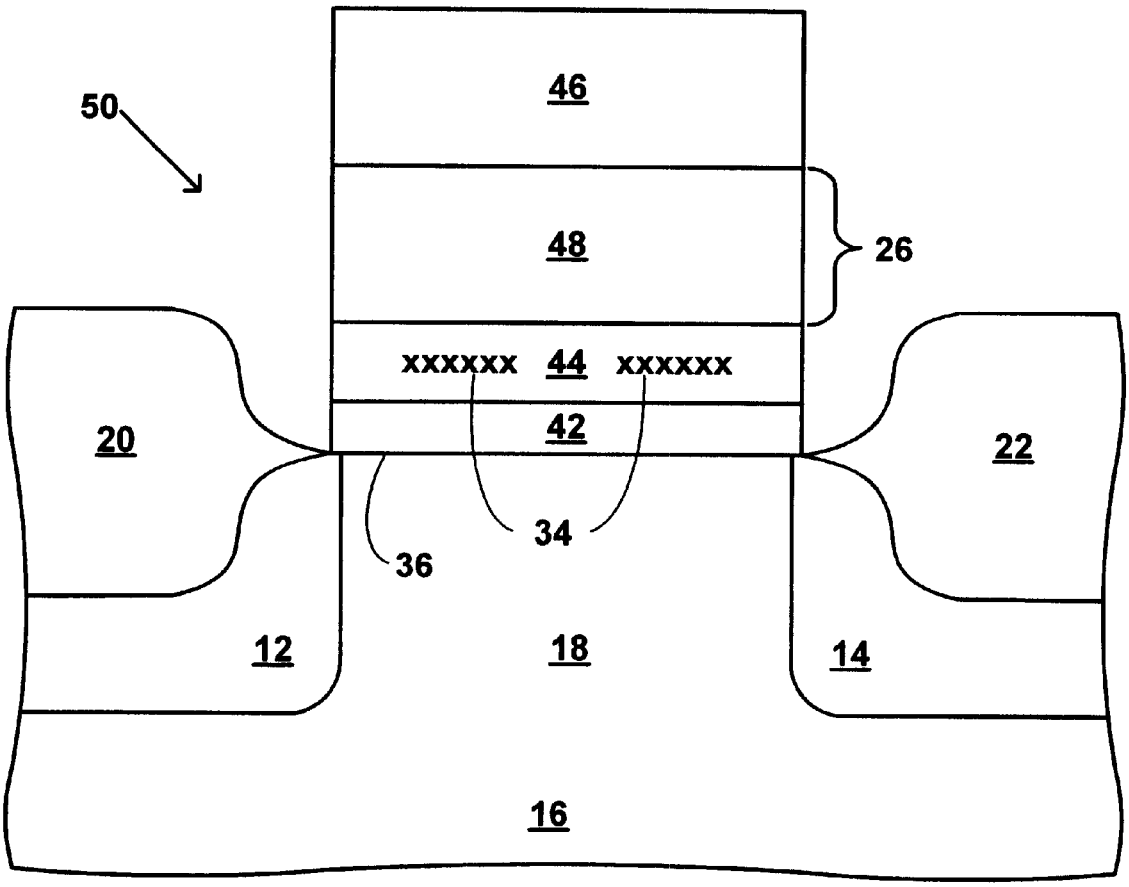


Fig. 8

USE OF HIGH-K DIELECTRIC MATERIAL IN MODIFIED ONO STRUCTURE FOR SEMICONDUCTOR DEVICES

TECHNICAL FIELD

The present invention relates to a process for preparation of a semiconductor device including forming a modified ONO structure. The modified ONO structure comprises a high-K dielectric material.

BACKGROUND ART

Non-volatile memory devices are currently in widespread use in electronic components that require the retention of information when electrical power is terminated. Non-volatile memory devices include read-only-memory (ROM), programmable-read-only memory (PROM), erasable-programmable-read-only memory (EPROM), and electrically-erasable-programmable-read-only-memory (EEPROM) devices. EEPROM devices differ from other non-volatile memory devices in that they can be electrically programmed and erased. Flash EEPROM devices are similar to EEPROM devices in that memory cells can be programmed and erased electrically. However, flash EEPROM devices enable the erasing of all memory cells in the device using a single electrical current pulse.

Product development efforts in EEPROM device technology have focused on increasing the programming speed, lowering programming and reading voltages, increasing data retention time, reducing cell erasure times and reducing cell dimensions. One important dielectric material for the fabrication of the EEPROM is an oxide-nitride-oxide (ONO) structure. One EEPROM device that utilizes the ONO structure is a silicon-oxide-nitride-oxide-silicon (SONOS) type cell. A second EEPROM device that utilizes the ONO structure is a floating gate flash memory device, in which the ONO structure is formed over the floating gate, typically a polysilicon floating gate.

In SONOS devices, during programming, electrical charge is transferred from the substrate to the silicon nitride layer in the ONO structure. Voltages are applied to the gate and drain creating vertical and lateral electric fields, which accelerate the electrons along the length of the channel. As the electrons move along the channel, some of the electrons gain sufficient energy to jump over the potential barrier of the bottom silicon dioxide layer and become trapped in the silicon nitride layer. Electrons are trapped near the drain region because the electric fields are the strongest near the drain. Reversing the potentials applied to the source and drain will cause electrons to travel along the channel in the opposite direction and be injected into the silicon nitride layer near the source region. Because silicon nitride is not electrically conductive, the charge introduced into the silicon nitride layer tends to remain localized. Accordingly, depending upon the application of voltage potentials, electrical charge can be stored in discrete regions within a single continuous silicon nitride layer.

Non-volatile memory designers have taken advantage of the localized nature of electron storage within a silicon nitride layer and have designed memory circuits that utilize two regions of stored charge within an ONO layer. This type of non-volatile memory device is known as a two-bit EEPROM, which is available under the trademark MIRRORBIT™ from Advanced Micro Devices, Inc., Sunnyvale, Calif. The MIRRORBIT™ two-bit EEPROM is capable of storing twice as much information as a conventional

EEPROM in a memory array of equal size. A left and right bit is stored in physically different areas of the silicon nitride layer, near left and right regions of each memory cell. Programming methods are then used that enable two bits to be programmed and read simultaneously. The two-bits of the memory cell can be individually erased by applying suitable erase voltages to the gate and to either the source or drain regions.

As device dimensions continue to be reduced, the electrical thickness of the ONO layer must be reduced accordingly. Previously, this has been accomplished by scaling down the physical thickness of the ONO layer. However, as the ONO layer is made physically thinner, leakage current through the ONO layer may increase, and the charge trapping ability of the nitride layer may be reduced, which limits the scaling down of the total physical thickness of the ONO layer.

A floating gate flash device includes a floating gate electrode upon which electrical charge is stored. The floating gate electrode is formed on a tunnel oxide layer which overlies a channel region residing between the source and drain regions in a semiconductor substrate. The floating gate electrode together with the source and drain regions form an enhancement transistor. Typically, the floating gate electrode may be formed of polysilicon.

In a floating gate flash device, electrons are transferred to the floating gate electrode through a dielectric layer overlying the channel region of the enhancement transistor. The electron transfer is initiated by either hot electron injection, or by Fowler-Nordheim tunneling. In either electron transfer mechanism, a voltage potential is applied to the floating gate electrode by an overlying control gate electrode. The control gate electrode is capacitively coupled to the floating gate electrode, such that a voltage applied on the control gate electrode is coupled to the floating gate electrode. The floating gate flash device is programmed by applying a high positive voltage to the control gate electrode, and a lower positive voltage to the drain region, which transfers electrons from the channel region to the floating gate electrode.

The control gate electrode is separated from the floating gate electrode by an interpoly dielectric layer, typically an oxide-nitride-oxide stack, i.e., an ONO structure or layer. However, as device dimensions continue to be reduced, the electrical thickness of the interpoly dielectric layer between the control gate electrode and the floating gate electrode must be reduced accordingly. Previously, this has been accomplished by scaling down the physical thickness of the ONO layer. However, as the ONO layer is made physically thinner, leakage current through the ONO layer may increase, which limits the scaling down of the total physical thickness of the ONO layer.

Some of the improvements in devices can be addressed through development of materials and processes for fabricating the ONO layer. Recently, development efforts have focused on novel processes and materials for use in fabrication of the ONO layer. While the recent advances in EEPROM technology have enabled memory designers to double the memory capacity of EEPROM arrays using two-bit data storage, numerous challenges exist in the fabrication of material layers within these devices. In particular, the ONO layer must be carefully fabricated to avoid an increase in the leakage current. Accordingly, advances in ONO fabrication and materials technology are needed to ensure proper charge isolation in ONO structures used in MIRRORBIT™ two-bit EEPROM devices and in floating gate flash devices.

DISCLOSURE OF INVENTION

In one embodiment, the present invention relates to a semiconductor device including a modified ONO structure, wherein the modified ONO structure comprises a bottom dielectric material layer, a silicon nitride layer on the bottom dielectric material layer, and a top dielectric material layer on the silicon nitride layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprises a composite dielectric material, and in which the composite dielectric material comprises elements of at least one mid-K or high-K dielectric material.

In another embodiment, the present invention relates to a non-volatile memory cell including a) a substrate comprising a source region, a drain region, and a channel region positioned therebetween; b) a floating gate positioned above the channel region and separated from the channel region by a tunnel dielectric material layer; and c) a control gate positioned above the floating gate and separated from the floating gate by an interpoly dielectric layer, the interpoly dielectric layer comprising a bottom dielectric material layer adjacent to the floating gate, a top dielectric material adjacent to the control gate, and a center layer comprising silicon nitride and positioned between the bottom dielectric material layer and the top dielectric material layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprises a mid-K or high-K dielectric material.

In another embodiment, the present invention relates to a non-volatile memory cell including a) a substrate comprising a source region, a drain region, and a channel region positioned therebetween; b) a charge storage layer comprising a modified ONO structure, the modified ONO structure having a bottom dielectric material layer adjacent the channel region, a top dielectric material layer, and a center charge storage layer comprising silicon nitride and positioned between the bottom dielectric material layer and the top dielectric material layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprises a mid-K or high-K dielectric material; and c) a gate capacitively coupled to the channel region through the charge storage layer.

In one embodiment, the present invention relates to a process for fabrication of a semiconductor device comprising a non-volatile memory cell having a modified ONO structure, comprising forming the modified ONO structure by steps comprising providing a semiconductor substrate; forming a first dielectric material layer on the semiconductor substrate; depositing a silicon nitride layer on the first dielectric material layer; and depositing a top dielectric material layer on the silicon nitride layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprises a mid-K or high-K dielectric material.

In another embodiment, the present invention relates to a process for fabrication of a semiconductor device, the device including a two-bit EEPROM device including a modified ONO structure, comprising forming the modified ONO structure by steps comprising providing a semiconductor substrate; forming a tunnel dielectric material layer overlying the semiconductor substrate; depositing a silicon nitride layer overlying the tunnel dielectric material layer; and depositing a top dielectric material layer overlying the silicon nitride layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprise a mid-K or a high-K dielectric material.

In another embodiment, the present invention relates to a process for fabrication of a semiconductor device, the device including a floating gate flash structure comprising a modified ONO structure, comprising forming the modified ONO structure by steps comprising providing a semiconductor substrate having a floating gate electrode; forming a bottom dielectric material layer overlying the floating gate electrode; depositing a silicon nitride layer overlying the tunnel dielectric material layer; and depositing a top dielectric material layer overlying the silicon nitride layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprise a mid-K or a high-K dielectric material.

Thus, in the present invention, by use of a mid-K or a high-K dielectric material instead of silicon dioxide for the top oxide layer, a modified ONO structure may be fabricated having reduced dimensions without creation of interface states coming from contamination which could provide charge leakage paths within the modified ONO structure and without sacrificing the charge trapping ability of the modified ONO structure in the two-bit EEPROM device. The present invention provides advantages such as (1) reduction of equivalent oxide thickness of ONO in next generation devices; (2) high-K film devices are expected to have improved data retention and reliability; and (3) the high-K dielectric material layer replacing one of both silicon dioxide layers allows fabrication of an ONO layer which is physically thicker, resulting in fewer charge leakage paths within the modified ONO structure. A variety of mid-K, high-K or composite dielectric materials may be used for replacement of at least one of the bottom or top oxide layers of an ONO structure, to obtain the modified ONO structure. These dielectric materials may be formed in a nanolaminate, allowing for exact selection of composition, thickness and K value of the modified ONO structure. Thus, the present invention provides an advance in ONO fabrication technology, and ensures proper charge isolation in modified ONO structures used in MIRRORBIT™ two-bit EEPROM devices, and ensures proper dielectric separation of the control gate electrode from the floating gate electrode in a floating gate flash device, while at the same time providing distinct process and economic advantages.

Although described herein in terms of MIRRORBIT™ two-bit EEPROM devices and floating gate flash devices, the present invention is broadly applicable to fabrication of any semiconductor device that includes an ONO structure.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** schematically illustrates, in cross-section, a portion of a semiconductor substrate containing a two-bit EEPROM transistor which incorporates a modified ONO structure fabricated in accordance with the invention.

FIG. **2** schematically illustrates, in cross-section, a portion of a semiconductor device containing a floating gate flash transistor which incorporates a modified ONO structure fabricated in accordance with the invention.

FIGS. **3–6** illustrate, in cross-section, process steps for the fabrication of a modified ONO structure and a gate structure thereover in accordance with the invention.

FIG. **7** is a schematic flow diagram showing the steps of the present invention.

FIG. **8** schematically illustrates, in cross-section, a portion of a floating gate flash memory device fabricated in accordance with another embodiment of the present invention.

It should be appreciated that for simplicity and clarity of illustration, elements shown in the Figures have not necessarily been drawn to scale. For example, the dimensions of some of the elements are exaggerated relative to each other

for clarity. Further, where considered appropriate, reference numerals have been repeated among the Figures to indicate corresponding elements.

MODES FOR CARRYING OUT THE INVENTION

Referring first to FIG. **1**, there is schematically shown in cross-section a transistor **10** suitable for use in a two-bit EEPROM device, such as the MIRRORBIT™ device. The transistor **10** includes source/drain regions **12** and **14** located in a semiconductor substrate **16** and separated by a channel region **18**. First and second bit line oxide regions **20** and **22** overlie source/drain regions **12** and **14**, respectively. A gate electrode **24** overlies the channel region **18** and is separated therefrom by a modified ONO structure **26**. The gate electrode **24** and the modified ONO structure **26** form a stacked-gate structure. The modified ONO structure **26** includes a first or tunnel dielectric material layer **28**, a silicon nitride layer **30** and a top dielectric material layer **32**, in that order, as shown in FIG. **1**. At least one of the top dielectric material layer **32** and the bottom dielectric material layer **28** comprises a mid-K or high-K dielectric material, which replaces the conventional silicon dioxide material in these layers. Thus, the top dielectric material layer **32** may comprise a mid-K or high-K dielectric material, the bottom dielectric material layer **28** may comprise a mid-K or high-K dielectric material, or both the top and the bottom layers **32** and **28** may comprise a mid-K or high-K dielectric material.

Referring next to FIG. **2**, there is schematically shown in cross-section a transistor **40** suitable for use in a floating gate flash EEPROM device. The transistor **40** includes source/drain regions **12** and **14** located in a semiconductor substrate **16** and separated by a channel region **18**. First and second bit line oxide regions **20** and **22** overlie source/drain regions **12** and **14**, respectively. A tunnel dielectric material layer **42** overlies the channel region **18**. A floating gate electrode **44** overlies the tunnel dielectric material layer **42**. In one embodiment, the tunnel dielectric material layer **42** is silicon dioxide. In another embodiment, the tunnel dielectric material layer **42** comprises a mid-K or a high-K dielectric material. In another embodiment, the tunnel dielectric material layer **42** is a standard-K dielectric material other than silicon dioxide, such as aluminum oxide or silicon oxynitride.

The floating gate electrode **44** is separated from a control gate electrode **46** by an interpoly dielectric layer which comprises a modified ONO structure **26**. The control gate electrode **46**, the modified ONO structure **26** and the floating gate electrode **44** form a floating gate flash memory structure. The modified ONO structure **26** in FIG. **2** is substantially the same as the modified ONO structure **26** shown in FIG. **1**, except that the first dielectric material layer **28** in FIG. **2** may be referred to as a bottom dielectric material layer **28**. Like the modified ONO structure shown in FIG. **1**, the modified ONO structure shown in FIG. **2** includes a mid-K or high-K dielectric material in at least one of the top dielectric material layer **32** or the bottom dielectric material layer **28**, or both, in which the mid-K or high-K dielectric material replaces the conventional silicon dioxide.

As used herein, the term "standard-K dielectric material" refers to a dielectric material having a K up to about 10. Such standard-K dielectric materials include, for example, silicon dioxide, which has a K of about 4, silicon oxynitride, which has a K of about 4–8 depending on the relative content of oxygen and nitrogen, and silicon nitride, which has a K of about 6–9, and aluminum oxide, which has a K of about 10.

As used herein, the term "mid-K dielectric material" refers to a dielectric material having a K in the range from greater than 10 to about 20. Such mid-K dielectric materials include, for example, composite materials such as hafnium silicate, which has a K of about 14, and hafnium silicon oxynitride, which has a K of about 16, depending on the relative content of oxygen and nitrogen, and hafnium silicon nitride, which has a K of about 18.

As used herein, the term "high-K dielectric material" refers to a dielectric material having a K of about 20 or more. Such high-K dielectric materials include, for example, HfO_2, ZrO_2 and others, some of which are identified more fully below. In general, the term "high-K dielectric material" encompasses binary, ternary and higher oxides and any ferroelectric material having a K of about 20 or more.

As used herein, the term "composite dielectric material" refers to a dielectric material comprising the elements of at least two other dielectric materials. A composite dielectric material may have a K value of a standard-K, mid-K or high-K dielectric material, depending on the elements combined to form the composite dielectric material. As described in more detail below, a composite dielectric material may be formed by co-deposition of its component elements, or by sequential deposition followed by a treatment step, e.g., thermal treatment, to combine the elements to form the composite dielectric material.

Regarding dielectric constants, or K values, silicon dioxide (sometimes simply referred to as "oxide") has a K value of approximately 4, while other dielectric materials have higher K values. Silicon nitride ("nitride"), for example, has a K of about 6 to 9 (depending on formation conditions). Much higher K values of, for example, 20 or more can be obtained with various transition metal oxides including hafnium oxide (HfO_2), zirconium oxide, (ZrO_2), tantalum oxide (Ta_2O_5), barium strontium titanate ("BST"), lead zirconate titanate ("PZT"), and others described more fully below. Using a high-K dielectric material for full or partial replacement of at least one of the top oxide layer 32 and the bottom oxide layer 28 to form the modified ONO structure 26 allows a low electrical thickness to be achieved even with a physically thick layer. For example, a high-K dielectric material gate dielectric with a K of 40 and a thickness of 100 angstroms is substantially electrically equivalent to a silicon dioxide gate dielectric (K about 4) having a thickness of about 10 angstroms. The electrically equivalent thickness of high-K materials may be referred to in terms of the equivalent oxide thickness. Thus, the high-K dielectric material gate dielectric with a K of 40 having a given physical thickness has an equivalent oxide thickness which is approximately 1/10 the given physical thickness. For higher-K dielectric materials, thicker gate dielectric layers can be formed while maintaining equivalent oxide thickness values lower than are possible with very thin oxide layers. In this way, the reliability problems associated with very thin dielectric layers may be avoided while transistor performance is increased.

Approximate K-values or, in some cases, a range of K-values, are shown below in Table 1 for several exemplary dielectric materials. It is understood that the present invention is not limited to the specific dielectric materials disclosed herein, but may include any appropriate standard-K, mid-K, high-K and composite dielectric materials which are known and are compatible with the remaining elements of the semiconductor device with which the dielectric materials are to be used.

TABLE 1

Dielectric Material	Approximate Dielectric Constant (K.) (Relative Permittivity)
silicon dioxide	4
silicon nitride	6–9
silicon oxynitride	4–8
aluminum oxide	10
zirconium silicate	12
hafnium silicate	15
lanthanum oxide, La_2O_3	20–30
hafnium oxide, HfO_2	40
zirconium oxide, ZrO_2	25
cerium oxide, CeO_2	26
bismuth silicon oxide, $Bi_4Si_2O_{12}$	35–75
titanium dioxide, TiO_2	30
tantalum oxide, Ta_2O_5	26
tungsten oxide, WO_3	42
yttrium oxide, Y_2O_3	20
$LaAlO_3$	25
BST ($Ba_{1-x}Sr_xTiO_3$)	~20—~200
$PbTiO_3$	~20—~200
$BaTiO_3$	~20—~200
$SrTiO_3$	~20—~200
$PbZrO_3$	~20—~200
PST ($PbSc_xTa_{1-x}O_3$)	~200—~3000
PZN ($PbZn_xNb_{1-x}O_3$)	~200—~5000
PZT ($PbZr_xTi_{1-x}O_3$)	~100—~1000
PMN ($PbMg_xNb_{1-x}O_3$)	~200—~5000

It is noted that the K-values, or relative permittivity, for both standard-K and high-K dielectric materials may vary to some degree depending on the exact nature of the dielectric material and on the process used to deposit the material. Thus, for example, differences in purity, crystallinity and stoichiometry, may give rise to variations in the exact K-value determined for any particular dielectric material.

As used herein, when a material is referred to by a specific chemical name or formula, the material may include non-stoichiometric variations of the stoichiometrically exact formula identified by the chemical name. For example, hafnium oxide, when stoichiometrically exact, has the chemical formula HfO_2. As used herein, the term "hafnium oxide" may include variants of stoichiometric HfO_2, which may be referred to as Hf_xO_y, in which either of x or y vary by a small amount. For example, in one embodiment, x may vary from about 0.75 to about 1.5, and y may vary from about 1.5 to about 3. In another embodiment, x may vary from about 0.9 to about 1.2, and y may vary from about 1.8 to about 2.2. Such variations from the exact stoichiometric formula fall within the definition of hafnium oxide. Similar variations from exact stoichiometry are included when the chemical formula for a compound is used. For example, again using hafnium oxide as an example, when the formula HfO_2 is used, Hf_xO_y as defined above, is included within the meaning. Thus, in the present disclosure, exact stoichiometry is intended only when such is explicitly so stated. As will be understood by those of skill in the art, such variations may occur naturally, or may be sought and controlled by selection and control of the conditions under which materials are formed.

Here and in all numerical values in the specification and claims, the limits of the ranges and ratios may be combined.

The following description of the devices and processes of the present invention are given in the context of a modified ONO structure suitable for use in a two-bit EEPROM device, such as the MIRRORBIT™ two-bit EEPROM device. It is to be understood that, while the present invention is discussed herein in that context, that this is merely exemplary and is not intended to limit the scope of the present invention. The modified ONO structure fabricated by the presently disclosed method is applicable to any semiconductor device in which an ONO structure may be included, and is particularly applicable also to the floating gate flash device described above with reference to FIG. **2**.

Referring to FIG. **1**, the modified ONO structure **26** includes a first dielectric material layer **28** (tunnel dielectric material layer **28** in a two-bit EEPROM device or bottom dielectric material layer **28** in the floating gate device) overlying the channel region **18**. The first dielectric material layer is conventionally silicon dioxide, but in the present invention the silicon dioxide may be fully or partially replaced with a mid-K or a high-K dielectric material. A layer **30** typically comprising silicon nitride overlies the first dielectric material layer **28**. A top dielectric material layer **32** which may comprise a mid-K or a high-K dielectric material overlies the silicon nitride-comprising layer **30**.

In the operation of the exemplary two-bit EEPROM transistor **10** shown in FIG. **1**, voltages are applied to the gate electrode **24** and to the source/drain regions **12** and **14**. The applied voltages cause electrical charge from the source/drain regions **12** and **14** to propagate across the channel region **18**. Once the charge encounters a sufficiently strong vertical field, the charge is either injected or tunnels from the channel region **18** into the silicon nitride charge storage layer **30**. For example, depending upon the particular voltage levels applied to the control-gate electrode **24** and to the source/drain regions **12** and **14**, electrical charges **34***a*, **34***b* are transferred into the layer **30** and are localized to regions in proximity to either the source/drain region **12**, or the source/drain region **14**. The electrical charges **34***a* and **34***b* each represent a single bit stored in the non-volatile memory cell.

Those skilled in the art will recognize that for proper functioning of a two-bit EEPROM device, the electrical charges **34***a* and **34***b* should remain isolated in the regions of the layer **30** to which each is initially introduced. The proper maintenance of the electrical charges **34***a*, **34***b* in localized regions of the layer **30** is needed for the proper performance of the two-bit EEPROM device. In particular, the quality of the modified ONO structure **26** should be such that charge leakage paths are minimized at the interface between the layer **30** and each of the tunnel dielectric material layer **28** and the top dielectric material layer **32**, at least one of which includes a mid-K or high-K dielectric material in the present invention. Additionally, the each of the bottom dielectric material layer **28** and the top dielectric material layer **32** should be of sufficient density that charge trapping sites are minimized within the high-K dielectric material.

Referring to FIG. **2**, the modified ONO structure **26** includes a first dielectric material layer **28** (tunnel dielectric material layer **28** in a two-bit EEPROM device or bottom dielectric material layer **28** in the floating gate device) overlying the floating gate electrode **44**. The first dielectric material layer is conventionally silicon dioxide, but in the present invention the silicon dioxide may be fully or partially replaced with a mid-K or a high-K dielectric material. A layer **30** typically comprising silicon nitride overlies the first dielectric material layer **28**. A top dielectric material layer **32** which may comprise a mid-K or a high-K dielectric material overlies the silicon nitride-comprising layer **30**.

In operation of the floating gate flash device **40** shown in FIG. **2**, electrons are transferred to the floating gate electrode **44** through the modified ONO structure interpoly dielectric layer **26** from the channel region **18** of the enhancement transistor. A voltage potential is applied to the floating gate

electrode **44** by the overlying control gate electrode **46**, which is capacitively coupled to the floating gate electrode **44** through the intervening modified ONO structure **26**. The floating gate flash device **40** is programmed by applying a high positive voltage to the control gate electrode **46** and a lower positive voltage to the drain region **14**, which transfers electrons from the channel region **18** to the floating gate electrode **44**. The electrons are stored as a charge **34** in the floating gate electrode **44**, as shown in FIG. **2**.

It will be recognized that for proper operation of the floating gate flash device **40**, the modified ONO structure interpoly dielectric layer **26** must provide effective dielectric separation between the control gate electrode **46** and the floating gate electrode **44**. Any reduction in the electrical thickness of the layer **30** results in a reduction of the overall electrical thickness of the interpoly dielectric layer.

In accordance with the present invention, charge leakage within the modified ONO structure **26** is minimized by forming a physically thick top and/or bottom mid-K or high-K dielectric material layer **32** and/or **28** having a low equivalent oxide thickness. The reduced charge leakage and improved floating gate or two-bit EEPROM performance obtained by the present invention can be better understood following a description of a fabrication process for the modified ONO structure carried out in accordance with the invention.

The following description of the present invention follows with reference to FIGS. **3–6**. FIG. **7** is a schematic flow diagram showing the steps of the process of the present invention. The following description of the process refers to FIGS. **3–6** sequentially and with reference to FIG. **7** generally. The present invention can be carried out in a cluster tool.

In the first step of the present invention, shown schematically in FIG. **7** as Step **S701**, a semiconductor substrate is provided. The semiconductor substrate may be any appropriately selected semiconductor substrate known in the art. In one embodiment, the semiconductor substrate is a bulk silicon substrate. In one embodiment, the semiconductor substrate is a silicon-on-insulator semiconductor substrate. In another embodiment, the semiconductor substrate is a p-doped silicon substrate. Suitable semiconductor substrates include, for example, bulk silicon semiconductor substrates, silicon-on-insulator (SOI) semiconductor substrates, silicon-on-sapphire (SOS) semiconductor substrates, and semiconductor substrates formed of other materials known in the art. The present invention is not limited to any particular type of semiconductor substrate.

In a floating gate flash device, the semiconductor device provided in the first step of the method comprises not only a silicon substrate **16** but also a tunnel oxide layer **42** and a floating gate electrode **44** which have been formed on the silicon substrate **16**. The tunnel oxide layer **42** and the floating gate electrode **46** may be formed appropriately by any process and from any material known in the art. In an embodiment in which the floating gate electrode **44** is formed of polysilicon, it has a silicon surface as described herein. Thus, the following description is applicable generally to a floating gate flash device, as well as to the two-bit EEPROM described above. Thus, the silicon substrate shown in FIGS. **3–6** may be identified by the reference number **16** or **44**, as appropriate.

Referring to FIG. **3**, in the second step of the present invention, shown schematically in FIG. **7** as Step **S702**, a first dielectric material layer **28** is formed on an upper silicon surface **36** of the semiconductor substrate **16**. In one embodiment, e.g., a floating gate flash device, the upper silicon surface **36** is the upper surface of a polysilicon floating gate electrode, as shown in FIGS. **2** and **3**. In one embodiment, e.g., a two-bit EEPROM device, the semiconductor substrate **16** is a single crystal silicon substrate. The substrate **16** may comprise other structural elements of a semiconductor device. It may, for example, be doped selectively to form a source region and a drain region.

In one embodiment, the silicon surface **36** previously has been processed to remove contaminants and native oxide. A suitable pre-clean procedure includes cleaning the silicon surface **36** with a dilute solution of hydrofluoric acid or any standard cleaning procedure used in the semiconductor industry.

In one embodiment, the bottom dielectric material layer **28** comprises both a high-K dielectric material and a standard-K dielectric material. In one embodiment, the bottom dielectric material layer **28** comprises a mid-K dielectric material. In one embodiment, the bottom dielectric material layer **28** comprises a composite dielectric material, which comprises a composite of elements of, or a reaction product of, two or more dielectric materials, at least one of which is a high-K dielectric material. In one embodiment, the composite dielectric material of which the bottom dielectric material layer **28** is formed is a mid-K dielectric material. The mid-K dielectric material may be a composite of a high-K dielectric material and a standard-K dielectric material. Thus, in one embodiment, the high-K dielectric material completely replaces the silicon dioxide bottom oxide layer of a conventional ONO structure. In another embodiment, the high-K dielectric material is, in essence, added to or combined with, the silicon dioxide bottom oxide layer of a conventional ONO structure to form the bottom dielectric material layer **28** of the modified ONO structure. In another embodiment, the bottom dielectric material layer **28** includes a composite dielectric material comprising elements of at least one high-K dielectric material, which replaces the silicon dioxide layer of a conventional ONO structure.

The first dielectric material layer **28** may be formed either by a growth process (e.g., oxidation of the silicon surface if the first dielectric material layer is to be a conventional oxide layer) or by a deposition process. The following description of methods of forming the bottom dielectric material layer **28** apply equally to forming the top dielectric material layer **32**, which is described in more detail below.

In one embodiment, the first dielectric material layer **28** is formed by a deposition process. In one embodiment, the mid-K or high-K dielectric material may be deposited by chemical vapor deposition (CVD). The CVD method may be any appropriate CVD method known in the art for deposition of a high-K material. In one embodiment, the first dielectric material layer **28** is formed by an ALCVD process. In one embodiment, the first dielectric material layer **28** is formed by an MOCVD process. In one embodiment, the first dielectric material layer **28** is formed by an RTCVD process. In one embodiment, the RTCVD deposition is carried out in the same RTP apparatus as that in which the other steps of the process of the present invention are carried out. In one embodiment, the RTP apparatus is part of a single-wafer cluster tool. In one embodiment, the RTCVD process is carried out at a temperature of about 700° C. to about 800° C.

The first oxide can also be formed by depositing the dielectric material layer in a batch furnace by an LPCVD process. In other embodiments, the CVD method may be PECVD or MLD.

The mid-K or high-K dielectric material may be formed by reacting a suitable metal-containing gas, e.g., hafnium tetra-t-butoxide with a suitable oxygen-containing gas, e.g., oxygen (O_2) or nitrous oxide (N_2O) in, e.g., the CVD apparatus.

In an embodiment in which the CVD is a RTCVD, the RTCVD mid-K or high-K dielectric material deposition may be carried out in three steps including an initial temperature ramp, a deposition step, and cool-down step. In one embodiment, the total reaction time is about 1–3 minutes. In another embodiment, the mid-K or high-K dielectric material deposition step is completed in about 2 minutes.

In one embodiment, the mid-K or high-K dielectric material includes at least one of hafnium oxide (HfO_2), zirconium oxide (ZrO_2), tantalum oxide (Ta_2O_5), barium titanate ($BaTiO_3$), titanium dioxide (TiO_2), cerium oxide (CeO_2), lanthanum oxide (La_2O_3), lanthanum aluminum oxide ($LaAlO_3$), lead titanate ($PbTiO_3$), strontium titanate ($SrTiO_3$), lead zirconate ($PbZrO_3$), tungsten oxide (WO_3), yttrium oxide (Y_2O_3), bismuth silicon oxide ($Bi_4Si_2O_{12}$), barium strontium titanate (BST) ($Ba_{1-x}Sr_xTiO_3$), PMN ($PbMg_xNb_{1-x}O_3$), PZT ($PbZr_xTi_{1-x}O_3$), PZN ($PbZn_xNb_{1-x}O_3$), and PST ($PbSc_xTa_{1-x}O_3$). In addition to the foregoing mid-K or high-K dielectrics, other mid-K or high-K dielectric materials, for example, ferroelectric high-K dielectric materials such as lead lanthanum titanate, strontium bismuth tantalate, bismuth titanate and barium zirconium titanate may be suitably used in the present invention. Other high-K dielectric materials known in the art, including, for example binary and ternary oxides having K values of about 20 or higher, also may be used in the present invention.

In one embodiment, the mid-K or high-K material is a mid-K or high-K material other than tantalum oxide (Ta_2O_5). Tantalum oxide has been found, in some embodiments, to exhibit an undesirably high leakage current.

Thus, for example, in an embodiment in which hafnium oxide is the mid-K or high-K dielectric material, the hafnium may be supplied to a suitable CVD apparatus in the form of a hafnium-containing vapor or gas such as hafnium tetra-t-butoxide, and the oxygen is supplied in gaseous form as oxygen, O_2 or nitrous oxide, N_2O. When a suitable thickness of hafnium oxide has been deposited, the flow of the hafnium-containing vapor or gas and oxygen-containing gas are stopped.

As noted above, in an embodiment in which a plurality of dielectric materials (standard-K, mid-K, high-K or composite) are deposited, alternating sub-layers of each dielectric material may be deposited, or a composite dielectric material which comprises the plurality of dielectric materials may be deposited in a single step. Thus, a first dielectric material precursor may be provided either simultaneously with or sequentially with a second dielectric precursor material. Either the first or the second dielectric precursor material may be a standard-K dielectric precursor or a high-K dielectric precursor, or both may be high-K precursor materials. The process of alternating sub-layer deposition may be continued until a suitable, selected number of sub-layers of desired composition and thickness has been deposited.

In one embodiment, a silicon containing gas, such as SiH_4, and a source of oxygen, may be provided together with or alternating with the high-K precursor vapor, during the step of forming the high-K dielectric material sub-layers, in an embodiment in which a composite dielectric material comprising both a high-K dielectric material and a standard-K dielectric material is to be produced. The composite dielectric material may be a mid-K dielectric material.

In one embodiment, a nitrogen-containing gas, such as ammonia (NH_3) may be provided along with the silicon-containing gas, oxygen-containing gas and high-K precursor.

In one embodiment, the process is carried out for a period of time and at gas flow rates sufficient to form a mid-K or high-K dielectric material-comprising layer having a thickness of about 25 to about 300 Å. In another embodiment, the process is carried out for a period of time and at gas flow rates sufficient to form a mid-K or high-K dielectric material-comprising layer having a thickness of about 50 to about 200 Å.

In one embodiment, a high-K dielectric material precursor is introduced in the CVD apparatus at a flow rate of about 1 standard liter per minute (slpm) and either oxygen or nitrous oxide is introduced at a flow rate of about 1 slpm. Suitable flow rates of any other gases or vapors provided to the CVD apparatus may be determined by those of skill in the art, based on the composition of the dielectric material layer **28** (and/or layer **32**) which is desired.

In one embodiment, the CVD is ALCVD, atomic layer CVD. ALCVD may be used to deposit a dielectric material in layers as thin as a molecular monolayer, which may also be referred to as a nanolayer. Formation of such nano layers allows formation of a nano-laminate structure of any selected dielectric materials. The nano-laminate structure provides for deposition of a plurality of different high-K dielectric materials as sub-layers or nano-layers. The sub-layers may be deposited with single dielectric materials or with a plurality of simultaneously formed dielectric materials. The nano-laminates may thus form a composite high-K dielectric material layer. In one embodiment, sequentially deposited, different dielectric materials may be annealed subsequently to form a composite dielectric material which comprises the elements of the sequentially deposited, different dielectric materials. The conditions of deposition may be suitably selected to provide a nano-lamninate or composite high-K dielectric material layer having a controlled physical thickness, composition and K value.

Thus, for example, ALCVD may be used to deposit alternating monolayers of hafnium oxide and silicon dioxide, in a partial replacement of the top silicon dioxide of a conventional ONO structure with a high-K dielectric material. The alternating layers may be retained, or the structure may be annealed to cause reaction or combination of the elements to form a composite dielectric material which would include the elements Hf/Si/O, i.e., $HfSiO_4$, hafnium silicate. Depending on the relative amounts of hafnium oxide and silicon dioxide deposited, the composite dielectric material may have a formula which varies from the above stoichiometry. Of course, it will be recognized that while the composite material includes the elements of the individual dielectric materials deposited, the exact stoichiometry may vary widely from this example.

As another example, a plurality of high-K dielectric materials may be deposited to replace either or both silicon dioxide layers of the conventional ONO structure in forming the modified ONO structure of the present invention, either simultaneously or in sequentially deposited sub-layers of selected thickness. A combination of high-K dielectric materials may be selected in order to obtain a desired characteristic such as K value, physical thickness, equivalent oxide thickness, or a selected combination of these or other features. The use of ALCVD, with its capability of depos-

iting a molecular monolayer of selected dielectric materials provides a wide range of possible structures for the modified ONO structure.

In another embodiment, the high-K bottom dielectric material layer **28** (and/or the top dielectric material layer **32**) may be formed by means of a low-pressure-chemical-vapor-deposition (LPCVD) process. In this alternative embodiment, the high-K dielectric material can be formed in a batch deposition apparatus. In one embodiment, the LPCVD process is carried out at an absolute pressure of about 200 to about 500 millitorr (mtorr), at temperatures of about 700–800° C. using a mid-K or high-K dielectric material precursor and either oxygen or nitrous oxide.

Suitable dielectric precursor materials are known in the art. For example, for hafnium oxide, hafnium tetra-t-butoxide has been mentioned above. For zirconium oxide, a suitable precursor is zirconium tetra-t-butoxide.

In one embodiment, the first dielectric material layer **28** (when silicon dioxide) is formed by in-situ steam generation (ISSG) oxidation of the silicon surface **36** of the semiconductor substrate **16**. The ISSG oxidation of the silicon surface **36** may be carried out, for example, in a rapid thermal process (RTP) apparatus. The RTP apparatus may be any such apparatus known in the art. In one embodiment, the RTP apparatus is part of a single-wafer cluster tool.

In one embodiment, the ISSG oxidation of the silicon surface **36** is carried out by placing the wafer in the RTP apparatus and flowing a mixture of oxygen-containing gas and hydrogen-containing gas to the chamber at suitable flow rates and pressure. The temperature of the RTP can be in the range from about 800° C. to about 1150° C. The flow rates and temperature may be suitably selected to provide rapid oxidation of the silicon surface **36**, to form an oxide layer of desired thickness.

For the ISSG process, any of the commercially available RTP systems can be utilized. Details of a suitable ISSG process may be found in commonly assigned, copending application U.S. application Ser. No. 10/036,757, filed Dec. 31, 2001. The disclosure of U.S. application Ser. No. 10/036,757 is hereby incorporated herein by reference for its teachings relating to ISSG.

In an alternate embodiment, the first dielectric material layer **28** (when silicon dioxide) may be grown by thermally oxidizing the silicon surface **36** at an elevated temperature in the presence of dry molecular oxygen. In one embodiment, the thermal oxidation is carried out at a temperature in the range of about 900° C. to about 1100° C. The thermal oxidation process may be carried out in either a batch-type thermal oxidation furnace, or alternatively, in a single-wafer oxidation apparatus. In one embodiment, the thermal oxidation is carried out in the same RTP apparatus as that in which the other steps of the present process are carried out. In one embodiment, the RTP apparatus is part of a single-wafer cluster tool.

In one embodiment, the first dielectric material layer **28**, when it is a high-K dielectric material, has a thickness in the range from about 20 to about 300 angstroms (Å), and in another embodiment, the first dielectric material layer **28** has a thickness in the range from about 50 to about 200 Å.

In the third step of the present invention, shown schematically in FIG. **7** as Step **S703**, a nitride layer **30** is formed on the first oxide layer **28**. As shown in FIG. **4**, after forming the first oxide layer **28**, the nitride layer **30** is deposited on the first oxide layer **28**. The nitride layer **30** may be formed by any method known in the art, as appropriate to the device being fabricated.

In one embodiment, the nitride layer **30** is silicon nitride. In another embodiment, the nitride layer **30** is silicon-rich silicon nitride. In other embodiments, the nitride layer **30** may be another suitable charge-storing nitride, such as silicon oxynitride.

In one embodiment, the nitride layer **30** is formed by means of a rapid-thermal-chemical-vapor-deposition (RTCVD) process. In one embodiment, the RTCVD process is carried out at a temperature of about 700° C. to about 800° C. The silicon nitride material may be formed by reacting a suitable nitrogen-containing gas, e.g., ammonia (NH_3) with a suitable silicon-containing gas, e.g., dichlorosilane ($SiCl_2H_2$) or silane (SiH_4).

In one embodiment, the process is carried out for a period of time and at gas flow rates sufficient to form a silicon nitride layer having a thickness of about 50 to about 200 angstroms. In another embodiment, the process is carried out for a period of time and at gas flow rates sufficient to form a silicon nitride layer having a thickness of about 100 angstroms.

In one embodiment, ammonia is introduced in the RTCVD apparatus at a flow rate of about 1 standard liter per minute (slpm) and either dichlorosilane or silane is introduced at a flow rate of about 30 to about 50 standard-cubic-centimeters-per-minute (sccm). The RTCVD process is carried out in three steps including an initial temperature ramp, a deposition step, and cool-down step. In one embodiment, the total reaction time is about 1–3 minutes. In another embodiment, the silicon nitride deposition step is completed in about 2 minutes.

In another embodiment, the silicon nitride layer 30 may be formed by means of a low-pressure-chemical-vapor-deposition (LPCVD) process. In this alternative embodiment, the silicon nitride can be formed in a batch deposition apparatus. In one embodiment, the LPCVD process is carried out at an absolute pressure of about 200 to about 500 millitorr (mtorr), at temperatures of about 700–800° C. using ammonia and either dichlorosilane or silane gas.

In the fourth step of the present invention, shown schematically in FIG. **7** as Step **S704**, a top dielectric material layer **32** is formed on the silicon nitride layer **30**. In one embodiment, the top dielectric material layer **32** is a mid-K or a high-K dielectric material. In another embodiment, when the bottom dielectric material **28** comprises a mid-K or a high-K dielectric material, the top dielectric material layer **32** is silicon dioxide. In another embodiment, as noted above, both the top dielectric material layer **32** and the bottom dielectric material layer **28** comprise a mid-K or a high-K dielectric material. The mid-K or high-K dielectric material may be formed by an appropriate method known in the art, such as any of those disclosed above with respect to formation of the bottom dielectric material layer **28**. As shown in FIG. **5**, after depositing the silicon nitride-comprising layer **30**, the top dielectric material layer **32**, is formed.

In one embodiment, the top dielectric material layer **32** comprises both a high-K dielectric material and a standard-K dielectric material. In one embodiment, the top dielectric material layer **32** comprises a mid-K dielectric material. In one embodiment, the top dielectric material layer **32** comprises a composite dielectric material, which comprises a composite of elements of, or a reaction product of, two or more dielectric materials, at least one of which is a high-K dielectric material. In one embodiment, the composite dielectric material of which the top dielectric material

layer **32** is formed is a mid-K dielectric material. The mid-K dielectric material may be a composite of a high-K dielectric material and a standard-K dielectric material. Thus, in one embodiment, the high-K dielectric material completely replaces the silicon dioxide top oxide layer of a conventional ONO structure. In another embodiment, the high-K dielectric material is, in essence, added to or combined with, the silicon dioxide top oxide layer of a conventional ONO structure to form the top dielectric material layer **32** of the modified ONO structure. In another embodiment, the top dielectric material layer **32** includes a composite dielectric material comprising elements of at least one high-K dielectric material, which replaces the silicon dioxide layer of a conventional ONO structure.

In one embodiment, an important feature of the invention includes the sequential formation of the silicon nitride layer **30** and the high-K top dielectric material layer **32** in the absence of exposure of the layer **30** to ambient atmosphere. Following the deposition of the silicon nitride layer **30** onto the first dielectric material layer **28**, the top dielectric material layer **32** comprising a mid-K or high-K dielectric material may be formed, e.g., by RTCVD, MOCVD or ALCVD in a single wafer cluster tool, without the necessity of being transferred to a separate oxide deposition chamber, which would entail either maintenance of vacuum conditions without exposing the substrate to ambient atmosphere, or application of a positive-pressure inert gas atmosphere during wafer transfer. Thus, the present invention provides distinct process and economic advantages in formation of a modified ONO structure.

As shown in FIG. **6**, following formation of the modified ONO structure **26** in accordance with the present invention, shown schematically in FIG. **7** as Step S**705**, a layer forming a gate electrode **24**, in the case of the two-bit EEPROM, is formed on the top dielectric material layer **32**. In the case of the floating gate flash device, a control gate electrode **46** is formed on the top dielectric material layer **32**. The stacked-gate structures shown in FIGS. **1** and **2** are completed by depositing the layer of gate forming material overlying the top dielectric material layer **32**. A lithographic patterning and etching process then may be carried out to define the gate electrode **24** (or control gate electrode **46**) and the modified ONO structure **26**. Those skilled in the art will recognize that various gate-forming materials can be used to fabricate the gate electrode **24**. For example, the gate electrode **24** and the control gate electrode **46** can be formed with polycrystalline silicon, amorphous silicon, a refractory metal silicide, a metal, and the like.

Following formation of the gate electrode **24**, or the control gate electrode **46**, as appropriate, and the modified ONO structure **26**, fabrication of the semiconductor device continues, as indicated in the final step of FIG. **7**.

For example, fabrication of the semiconductor device may include annealing the device in order to densify the silicon dioxide and/or mid-K or high-K dielectric material layers and/or to further form the composite dielectric material of either or both of the bottom dielectric material layer **28** and the top dielectric material layer **32**, in which the dielectric material layer comprises a mid-K or a high-K dielectric material.

FIG. **8** shows an embodiment of the present invention in which a floating gate flash memory device, having a modified ONO structure as described herein, has been annealed to form a composite dielectric material of the combined layers **28**, **30** and **32**, in which the modified ONO structure includes at least one of the top dielectric material layer **28** and the bottom dielectric material layer **32** were originally comprised of a mid-K or a high-K dielectric material, and the other layers are as described herein.

In this embodiment, a non-volatile memory cell **50** (here a floating gate memory cell) is formed which includes a) a substrate **16** including a source region **12**, a drain region **14**, and a channel region **18** positioned therebetween; b) a floating gate **44** positioned above the channel region **18** and separated from the channel region **18** by a tunnel dielectric film **42**; and c) a control gate **46** positioned above the floating gate **44** and separated from the floating gate **44** by an interpoly dielectric layer **26**. In this embodiment, the interpoly dielectric layer **26** includes a single layer **48** adjacent to both the floating gate **42** and the control gate **46**. In this embodiment, the single layer **48** is a dielectric material which is a metal silicate, a metal aluminate or a metal mixed-aluminate/silicate. Thus, for example, using hafnium as the exemplary metal, the dielectric material forming the layer **48** may be hafnium silicate ($HfSiO_4$), hafnium aluminate ($HfAl_2O_5$) or a hafnium mixed-aluminate/silicate, $HfO_2/SiO_2/Al_2O_3$, which may have a formula such as $Hf_2Si_2Al_2O_{11}$. The single dielectric material comprises the elements of the high-K dielectric material of which at least one of the top and bottom dielectric material layers were formed, the silicon nitride layer **30**, and any silicon dioxide which may have formed either of the top or bottom dielectric material layer which was oxide, rather than a mid-K or high-K dielectric material.

Suitable metals for the metal silicate, metal aluminate or metal mixed-aluminate/silicate include, for example, hafnium, zirconium, yttrium, cerium, tantalum, titanium, lanthanum, tungsten, bismuth, barium, strontium, scandium, niobium or lead, or mixtures thereof. Other metals which, when combined with silicon dioxide or aluminum oxide, or a mixture thereof, yield a material having a K value greater than about 10 may be suitable. The metal silicate, metal aluminate or metal mixed-aluminate/silicate substantially should not react with silicon (or polysilicon) at temperatures of about 600–800° C.

There has been disclosed in accordance with the invention a process for fabricating an ONO floating-gate electrode in both a MIRRORBIT™ two-bit EEPROM device and a floating gate flash device, both of which provide the advantages set forth above (as appropriate). As noted above, the process of the invention is also applicable to other semiconductor devices which include an ONO structure.

Although the invention has been described and illustrated with reference to specific illustrative embodiments thereof, it is not intended that the invention be limited to those illustrative embodiments. Those skilled in the art will recognize that variations and modifications can be made without departing from the spirit of the invention. For example, the thicknesses of the individual layers making up the modified ONO structure can be varied from that described herein. In another variation, the composition of the other layers of the ONO structure or the stacked gate structure may be varied from, e.g., silicon dioxide or silicon nitride to other materials. It is therefore intended to include within the invention all such variations and modifications that fall within the scope of the appended claims and equivalents thereof.

INDUSTRIAL APPLICABILITY

Thus, in accordance with the present invention, a semiconductor device including a modified ONO structure is provided which may be fabricated without creation of inter-

face states that could provide charge leakage paths within the modified ONO structure. The present invention can be carried out in a cluster tool. The present invention provides advantages such as (1) formation of a cleaner interface between layers of the modified ONO structure, resulting in fewer interface states that could provide charge leakage paths; (2) use of a high-K dielectric material, which allows formation of a physically thicker ONO structure having a lower equivalent oxide thickness; and (3) an efficient process which may be carried out in a single device, such as a cluster tool. Thus, the present invention provides an advance in dielectric fabrication technology, and ensures proper charge storage and isolation in modified ONO structures used in MIRRORBIT™ two-bit EEPROM devices, and ensures proper dielectric separation of the control gate electrode from the floating gate electrode in floating gate flash devices, while at the same time providing distinct process and economic advantages. Although described in terms of, and particularly applicable to, two-bit EEPROM devices, the present invention is broadly applicable to fabrication of any semiconductor device including a modified ONO structure.

What is claimed is:

1. A semiconductor device comprising a modified ONO structure, wherein the modified ONO structure comprises a bottom dielectric material layer, a nitride layer on the bottom dielectric material layer, and a top dielectric material layer on the nitride layer, in which at least one of the bottom dielectric material layer and the top dielectric material layer comprises a composite dielectric material, wherein the composite dielectric material comprises elements of at least one mid-K or high-K dielectric material,

wherein each mid-K or high-K dielectric material independently comprises at least one of hafnium oxide (HfO_2), zirconium oxide (ZrO_2), barium titanate ($BaTiO_3$), cerium oxide (CeO_2), lanthanum oxide (La_2O_3), lanthanum aluminum oxide ($LaAlO_3$), lead titanate ($PbTiO_3$), strontium titanate ($SrTiO_3$), lead zirconate ($PbZrO_3$), tungsten oxide (WO_3), yttrium oxide (Y_2O_3), bismuth silicon oxide ($Bi_4Si_2O_{12}$), barium strontium titanate (BST) ($Ba_{1-x}Sr_xTiO_3$), PMN ($PbMg_xNb_{1-x}O_3$), PZT ($PbZr_xTi_{1-x}O_3$), PZN ($PbZn_xNb_{1-x}O_3$), and PST ($PbSc_xTa_{1-x}O_3$).

2. A non-volatile memory cell comprising:

a) a substrate comprising a source region, a drain region, and a channel region positioned therebetween;

b) a floating gate positioned above the channel region and separated from the channel region by a tunnel dielectric material layer; and

c) a control gate positioned above the floating gate and separated from the floating gate by an interpoly dielectric layer, the interpoly dielectric layer comprising a modified ONO structure having a bottom dielectric material layer adjacent to the floating gate, a top dielectric material layer adjacent to the control gate, and a center layer comprising a nitride and positioned between the bottom dielectric material layer and the top dielectric material layer, wherein at least one of the bottom dielectric material layer and the top dielectric material layer comprises a mid-K or high-K dielectric material, and at least one of the bottom dielectric material layer or the top dielectric material layer comprises a composite dielectric material including elements of a high-K dielectric material and at least one additional dielectric material,

wherein each mid-K or high-K dielectric material independently comprises at least one of hafnium oxide (HfO_2), zirconium oxide (ZrO_2), barium titanate ($BaTiO_3$), cerium oxide (CeO_2), lanthanum oxide (La_2O_3), lanthanum aluminum oxide ($LaAlO_3$), lead titanate ($PbTiO_3$), strontium titanate ($SrTiO_3$), lead zirconate ($PbZrO_3$), tungsten oxide (WO_3), yttrium oxide (Y_2O_3), bismuth silicon oxide ($Bi_4Si_2O_{12}$), barium strontium titanate (BST) ($Ba_{1-x}Sr_xTiO_3$), PMN ($PbMg_xNb_{1-x}O_3$), PZT ($PbZr_xTi_{1-x}O_3$), PZN ($PbZn_xNb_{1-x}O_3$), and PST ($PbSc_xTa_{1-x}O_3$).

3. The non-volatile memory cell of claim **2**, wherein both of the bottom dielectric material layer and the top dielectric material layer comprise a composite dielectric material including elements of a high-K dielectric material and at least one additional high-K dielectric material.

4. A non-volatile memory cell comprising:

a) a substrate comprising a source region, a drain region, and a channel region positioned therebetween;

b) a charge storage layer comprising a modified ONO structure, the modified ONO structure having a bottom dielectric material layer adjacent the channel region, a top dielectric material layer, and a center charge storage layer comprising a nitride and positioned between the bottom dielectric material layer and the top dielectric material layer, wherein at least one of the bottom dielectric material layer and the top dielectric material layer comprises a mid-K or high-K dielectric material, wherein each mid-K or high-K dielectric material independently comprises at least one of hafnium oxide (HfO_2), zirconium oxide (ZrO_2), barium titanate ($BaTiO_3$), cerium oxide (CeO_2), lanthanum oxide (La_2O_3), lanthanum aluminum oxide ($LaAlO_3$), lead titanate ($PbTiO_3$), strontium titanate ($SrTiO_3$), lead zirconate ($PbZrO_3$), tungsten oxide (WO_3), yttrium oxide (Y_2O_3), bismuth silicon oxide ($Bi_4Si_2O_{12}$), barium strontium titanate (BST) ($Ba_{1-x}Sr_xTiO_3$), PMN ($PbMg_xNb_{1-x}O_3$), PZT ($PbZr_xTi_{1-x}O_3$), PZN ($PbZn_xNb_{1-x}O_3$), and PST ($PbSc_xTa_{1-x}O_3$); and

c) a gate capacitively coupled to the channel region through the charge storage layer,

and wherein at least one of the bottom dielectric material layer or the top dielectric material layer comprises a composite dielectric material including elements of the high-K dielectric material and at least one additional dielectric material.

5. The non-volatile memory cell of claim **4**, wherein both of the bottom dielectric material layer and the top dielectric material layer comprise a composite dielectric material including elements of a high-K dielectric material and at least one additional high-K dielectric material.

6. The device of claim **1**, wherein the device is a two-bit EEPROM device.

7. The device of claim **1**, wherein the device is a floating gate flash memory device.

8. The device of claim **1**, wherein the nitride layer comprises silicon nitride.

9. The device of claim **1**, wherein both of the bottom dielectric material layer and the top dielectric material layer comprise a composite dielectric material.

10. The device of claim **1**, wherein the composite dielectric material comprises at least one of hafnium oxide or zirconium oxide or cerium oxide.

11. The device of claim **1**, wherein the composite dielectric material is a mid-K dielectric material comprising at least one of silicon or aluminum.

12. The device of claim **2**, wherein the center layer comprises silicon nitride.

13. The device of claim **2**, wherein both of the bottom dielectric material layer and the top dielectric material layer comprise a mid-K or high-K dielectric material.

14. The device of claim **2**, wherein the mid-K or high-K dielectric material comprises at least one of hafnium oxide or zirconium oxide or cerium oxide.

15. The device of claim **2**, wherein the mid-K or high-K dielectric material is a mid-K dielectric material comprising at least one of silicon or aluminum.

16. The device of claim **4**, wherein the center charge storage layer comprises silicon nitride.

17. The device of claim **4**, wherein both of the bottom dielectric material layer and the top dielectric material layer comprise a mid-K or high-K dielectric material.

18. The device of claim **4**, wherein the mid-K or high-K dielectric material comprises at least one of hafnium oxide or zirconium oxide or cerium oxide.

19. The device of claim **4**, wherein the mid-K or high-K dielectric material is a mid-K dielectric material comprising at least one of silicon or aluminum.

20. The device of claim **1**, wherein the composite dielectric material comprises at least one of hafnium oxide or zirconium oxide or cerium oxide and at least one of silicon or aluminum.

21. The non-volatile memory cell of claim **2**, wherein the composite dielectric material comprises at least one of hafnium oxide or zirconium oxide or cerium oxide and at least one of silicon or aluminum.

22. The non-volatile memory cell of claim **4**, wherein the composite dielectric material comprises at least one of hafnium oxide or zirconium oxide or cerium oxide and at least one of silicon or aluminum.

* * * * *

EXHIBIT D

US006651134B1

(12) **United States Patent**
Phelan

(10) **Patent No.: US 6,651,134 B1**
(45) **Date of Patent: Nov. 18, 2003**

(54) **MEMORY DEVICE WITH FIXED LENGTH NON INTERRUPTIBLE BURST**

(75) Inventor: **Cathal G. Phelan**, Mountain View, CA (US)

(73) Assignee: **Cypress Semiconductor Corp.**, San Jose, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/504,344**

(22) Filed: **Feb. 14, 2000**

(51) **Int. Cl.**[7] **C06F 12/00**

(52) **U.S. Cl.** **711/104**; 711/105; 711/167; 711/169; 710/35; 365/233; 365/238.5

(58) **Field of Search** 711/104–105, 169, 711/167; 365/233, 238.5; 710/35

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,651,138 A	*	7/1997	Le et al.	711/154
5,729,504 A	*	3/1998	Cowles	365/236
5,805,928 A	*	9/1998	Lee	710/35
5,936,975 A	*	8/1999	Okamura	714/719
5,966,724 A	*	10/1999	Ryan	711/105
6,085,261 A	*	7/2000	McIntyre, Jr. et al.	710/35
6,289,138 B1	*	9/2001	Yip et al.	382/307

OTHER PUBLICATIONS

Understanding Burst Modes in Synchronous SRAMs, Cypress Semiconductor Corp., Jun. 30, 1999.

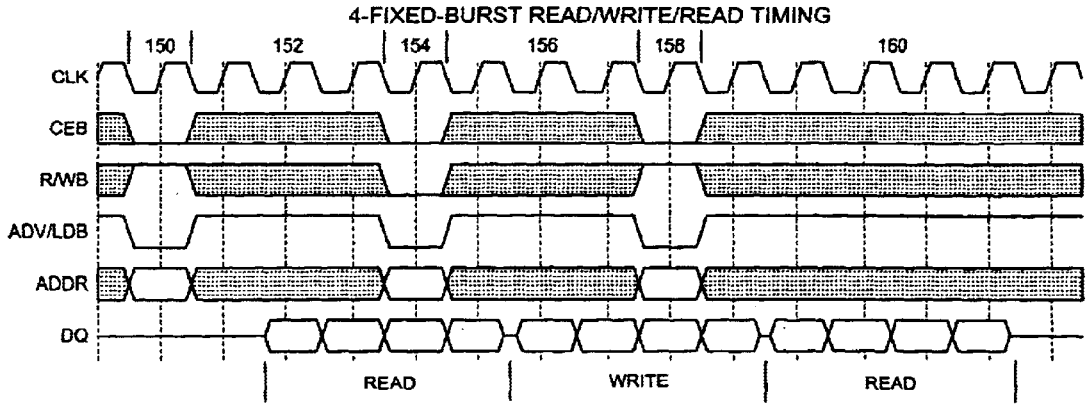
* cited by examiner

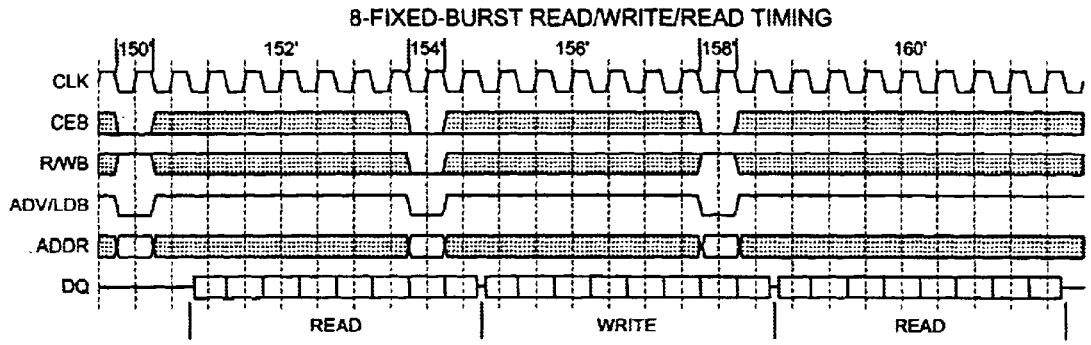
Primary Examiner—Donald Sparks
Assistant Examiner—Medhi Namazi
(74) *Attorney, Agent, or Firm*—Christopher P. Maiorana, P.C.; Robert M. Miller

(57) **ABSTRACT**

An integrated circuit comprising a memory and a logic circuit. The memory may comprise a plurality of storage elements each configured to read and write data in response to an internal address signal. The logic circuit may be configured to generate a predetermined number of the internal address signals in response to (i) an external address signal, (ii) a clock signal and (iii) one or more control signals. The generation of the predetermined number of internal address signals may be non-interruptible.

21 Claims, 3 Drawing Sheets





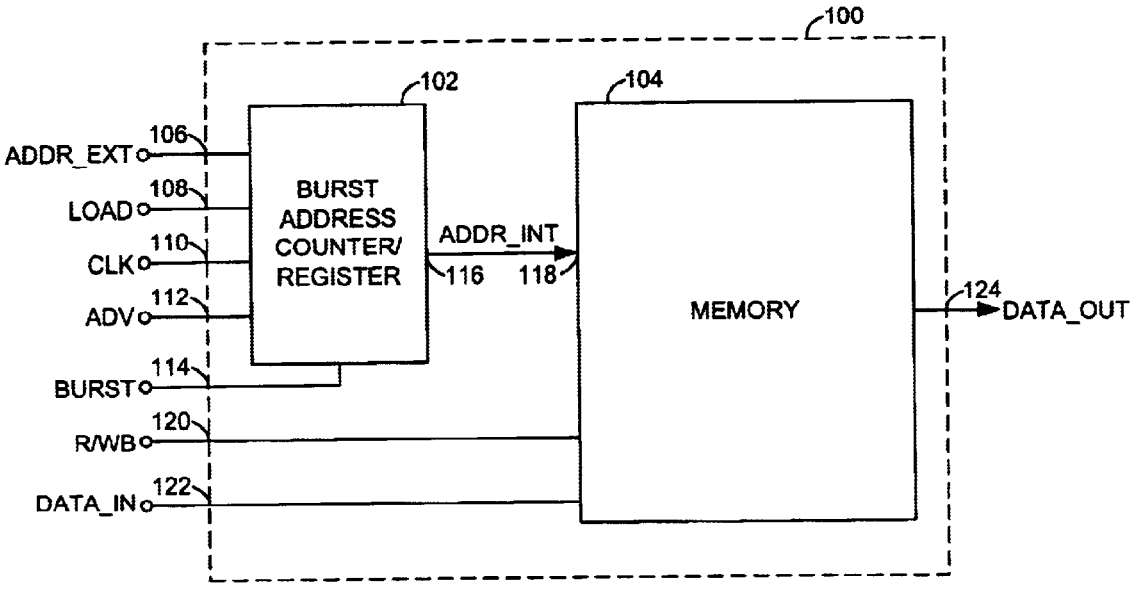

100
102
104
ADDR_EXT
106
LOAD
108
CLK
110
ADV
112
BURST
114
R/WB
120
DATA_IN
122
BURST ADDRESS COUNTER/ REGISTER
ADDR_INT
116
118
MEMORY
124
DATA_OUT

FIG. 1

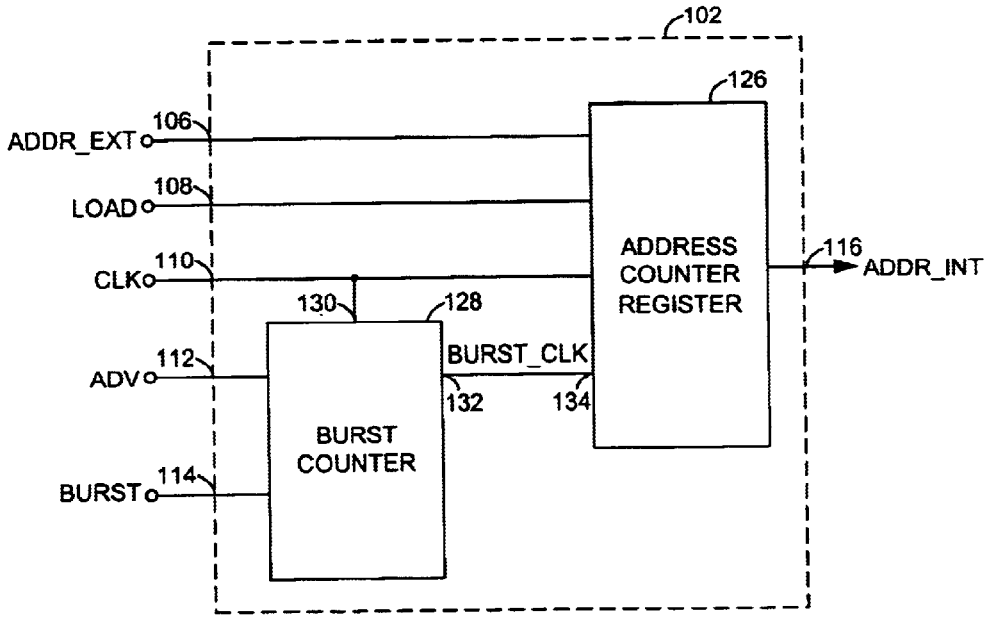

102
126
ADDR_EXT
106
LOAD
108
CLK
110
ADV
112
BURST
114
130
128
BURST_CLK
132
134
BURST COUNTER
ADDRESS COUNTER REGISTER
116
ADDR_INT

FIG. 2

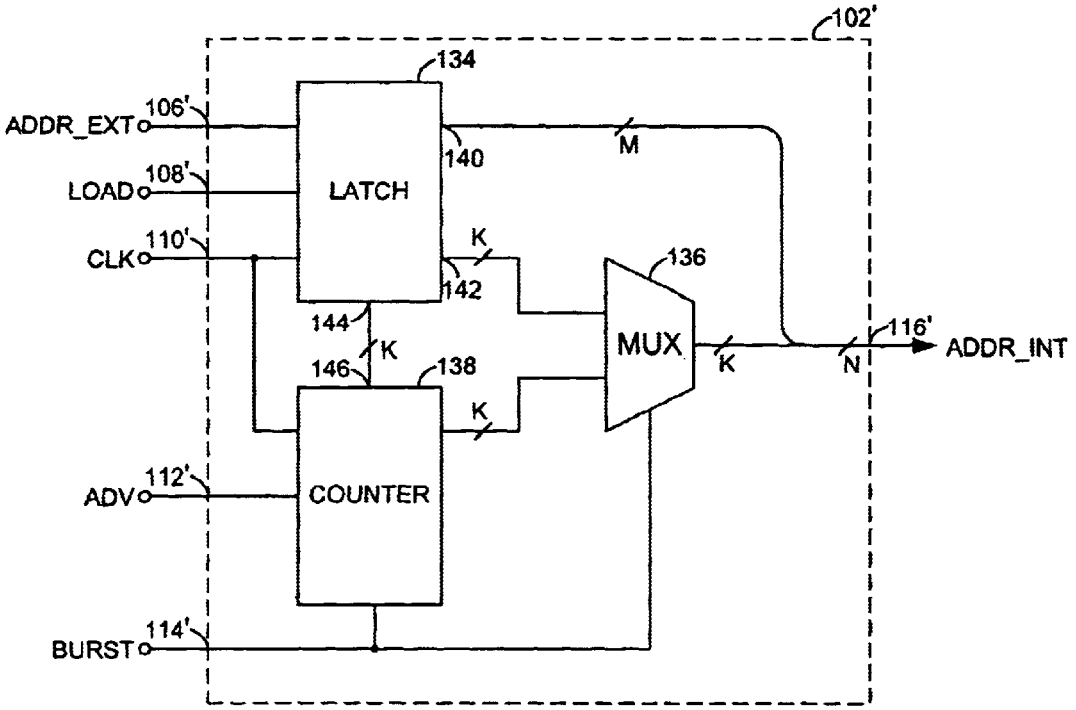

102'
134
ADDR_EXT
106'
LOAD
108'
CLK
110'
LATCH
140
M
K
142
136
144
K
146
138
MUX
K
N
116'
ADDR_INT
K
ADV
112'
COUNTER
BURST
114'

FIG. 3

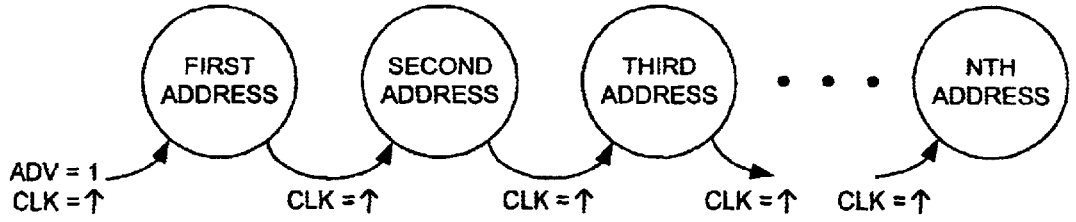

FIRST ADDRESS
SECOND ADDRESS
THIRD ADDRESS
NTH ADDRESS
ADV = 1
CLK = ↑
CLK = ↑
CLK = ↑
CLK = ↑
CLK = ↑

FIG. 4

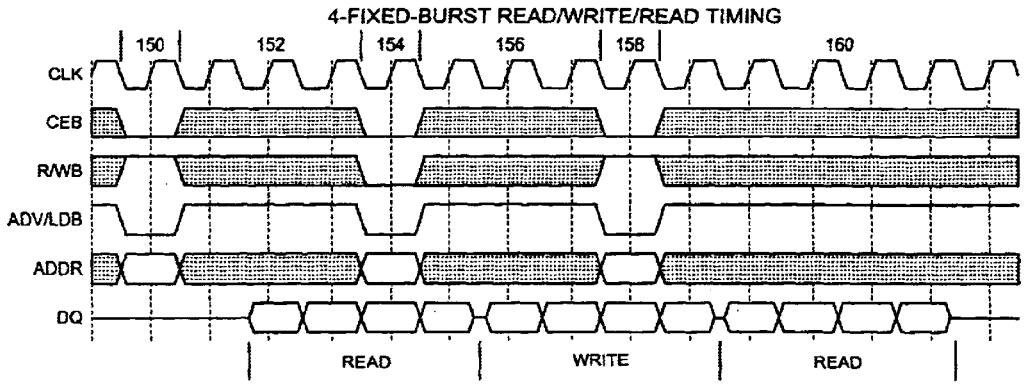


FIG. 5A

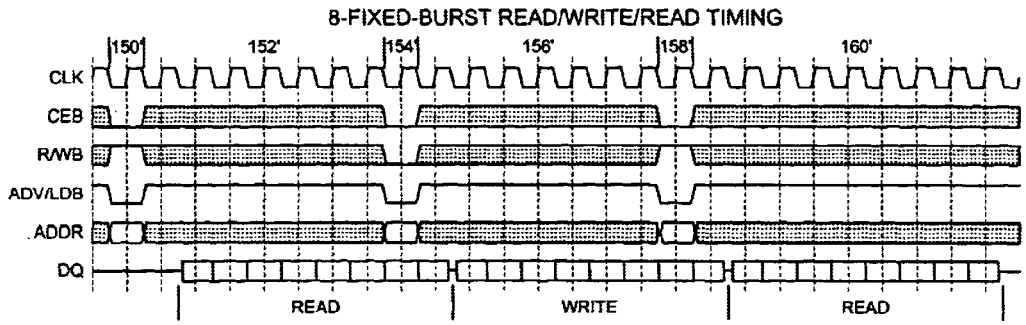


FIG. 5B

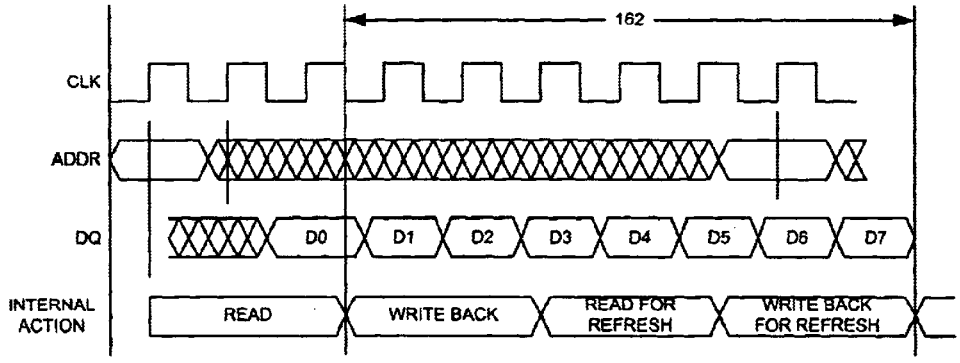


FIG. 6

MEMORY DEVICE WITH FIXED LENGTH NON INTERRUPTIBLE BURST

FIELD OF THE INVENTION

The present invention relates to memory devices generally and, more particularly, to a memory device that transfers a fixed number of words of data with each access.

BACKGROUND OF THE INVENTION

A synchronous Static Random Access Memory (SRAM) can provide data from multiple address locations using a single address. Accessing multiple locations in response to a single address is called a burst mode access. A memory device that provides a burst mode can reduce activity on the address and control buses. The burst mode of a conventional synchronous SRAM can be started and stopped in response to a control signal.

A conventional Dynamic Random Access Memory (DRAM) preserves data during periodic absences of power by implementing a memory cell as a capacitor and an access transistor. Since the charge on the capacitor will slowly leak away, the cells need to be "refreshed" once every few milliseconds. Depending on the frequency of accesses, a conventional DRAM can need an interrupt to perform data refreshes. Using a DRAM in a burst application is difficult because of the need to refresh. Completely hiding refresh cycles (e.g., refreshing data without the need for interrupts) in a DRAM cannot happen with conventional memory devices due to architecture choices that have been made. Data word bursts can be interrupted while in progress since conventional architectures support both burst and single access modes. Conventional DRAM access takes about 10 ns to get data, but nearly 20 ns to complete writeback and equalization. The addition of another 20 ns for a refresh results in a total access of 40 ns.

Since the data burst transfers of conventional memories can be interrupted and single accesses made, the amount of time that the data, address and control busses are not in use can vary. The variability of bus availability complicates the design of systems with shared data, address and control busses.

It would be desirable to have a memory device that has a fixed burst length.

SUMMARY OF THE INVENTION

The present invention concerns an integrated circuit comprising a memory and a logic circuit. The memory may comprise a plurality of storage elements each configured to read and write data in response to an internal address signal. The logic circuit may be configured to generate a predetermined number of the internal address signals in response to (i) an external address signal, (ii) a clock signal and (iii) one or more control signals. The generation of the predetermined number of internal address signals may be non-interruptible.

The objects, features and advantages of the present invention include providing a fixed burst memory that may (i) give network customers who typically burst large data lengths the ability to set a fixed burst length that suits particular needs; (ii) have non-interruptible bursts; (iii) free up the address bus and control bus for a number of cycles; (iv) provide programmability for setting the burst length by using DC levels [Vss or Vcc] on external pins; (v) hide required DRAM refreshes inside a known fixed burst length of data words; and/or (vi) operate at higher frequencies without needing interrupts to perform refreshes of data.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other objects, features and advantages of the present invention will be apparent from the following detailed description and the appended claims and drawings in which:

FIG. **1** is a block diagram illustrating a preferred embodiment of the present invention;

FIG. **2** is a detailed block diagram illustrating a circuit **102** of FIG. **1**;

FIG. **3** is a detailed block diagram of a circuit **102'** illustrating an alternative embodiment of the circuit **102** of FIG. **1**;

FIG. **4** is a flow diagram illustrating an example burst address sequence;

FIGS. **5**A and **5**B are diagrams illustrating example operations of a 4 word (FIG. **5**A) and an 8 word (FIG. **5**B) fixed burst access in accordance with the present invention; and

FIG. **6** is a diagram illustrating an example operation where a burst length may be long enough to include a writeback and a refresh cycle.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring to FIG. **1**, a block diagram of a circuit **100** is shown in accordance with a preferred embodiment of the present invention. The circuit **100** may be implemented, in one example, as a fixed burst memory. The circuit **100** may be configured to transfer a fixed number of words of data with each access (e.g., read or write). A number of words transferred as a group is called a burst. The circuit **100** generally comprises a circuit **102** and a memory array (or circuit) **104**. The circuit **102** may be implemented, in one example, as a burst address counter/register. The memory array **104** may be implemented, in one example, as a static random access memory (SRAM), a dynamic random access memory (DRAM), or other appropriate memory to meet the design criteria of a particular implementation.

The circuit **102** may have an input **106** that may receive a signal (e.g., ADDR_EXT), an input **108** that may receive a signal (e.g., LOAD), an input **110** that may receive a signal (e.g., CLK), an input **112** that may receive a signal (e.g., ADV), and an input **114** that may receive a signal (e.g., BURST). The circuit **102** may have an output **116** that may present a signal (e.g., ADDR_INT) to an input **118** of the memory **104**. The memory **104** may have an input **120** that may receive a signal (e.g., R/Wb), an input **122** that may receive a signal (e.g., DATA_IN) and an output **122** that may present a signal (e.g., DATA_OUT). The various signals are generally "on" (e.g., a digital HIGH, or 1) or "off" (e.g., a digital LOW, or 0). However, the particular polarities of the on (e.g., asserted) and off (e.g., de-asserted) states of the signals may be adjusted (e.g., reversed) accordingly to meet the design criteria of a particular implementation.

The signal ADDR_EXT may be, in one example, an external address signal. The signal ADDR_EXT may be n-bits wide, where n is an integer. The signal CLK may be a clock signal. The signal R/Wb may be a control signal that may be in a first state or a second state. When the signal R/Wb is in the first state, the circuit **100** will generally read data from the memory circuit **104** for presentation as the signal DATA_OUT. When the signal R/Wb is in the second state, the circuit **100** will generally store data received as the signal DATA_IN.

The signal LOAD may be, in one example, an address load control signal. The circuit **100** may be configured to

load an initial address, presented by the signal ADDR_EXT, in response to the signal LOAD. The initial address may determine the initial location where data transfers to and from the memory **104** will generally begin.

The signal ADV may be, in one example, used as a control signal. The circuit **100** may be configured to transfer a fixed number of words to or from the memory **104** in response to the signals ADV, CLK and R/Wb. When the signal ADV is asserted, the circuit **100** will generally begin transferring a predetermined number of words. The transfer is generally non-interruptible. In one example, the signal ADV may initiate the generation of a number of addresses for presentation as the signal ADDR_INT.

The signals ADV and LOAD may be, in one example, a single signal (e.g., ADV/LDb). The signal ADV/LDb may be a control signal that may be in a first state or a second state. When the signal ADV/LDb is in the first state, the circuit **102** will generally load an address presented by the signal ADDR_EXT as an initial address. When the signal ADV/LDb is in the second state, the circuit **102** may be configured to generate the signal ADDR_INT as a fixed number of addresses in response to the signal CLK. The signal ADDR_INT may be, in one example, an internal address signal. The signal ADDR_INT may be n-bits wide. Once the circuit **102** has started generating the fixed number of addresses, the circuit **102** will generally not stop until the fixed number of addresses has been generated (e.g., a non-interruptible burst).

The signal BURST may be, in one example, a configuration signal for programming the fixed number of addresses that the circuit **102** may generate in response to the signals CLK and ADV/LDb. The signal BURST may be generated, in one example, by (i) using bond options, (ii) voltage levels applied to external pins, or (iii) other appropriate signal generation means.

When the memory **104** is implemented as a DRAM, the circuit **100** may be configured to hide required DRAM refreshes (e.g., refreshes may occur without affecting external environment) inside a known fixed burst length of data words. The fixed burst length may allow the circuit **100** to operate at higher frequencies than a conventional DRAM without needing interrupts to perform refreshes of data. In one example, the fixed burst length may be four or eight words. However, the burst length may be set to whatever length is necessary to meet the design criteria of a particular application. For example, the burst length may be programmed, in one example, to allow both writeback and refresh to occur within a single access. The fixed burst-length may be set, in one example, longer or shorter depending upon a frequency or technology to be used.

The circuit **100** may be configured to provide a fixed burst length that may suit the requirements of network customers who typically burst large data lengths. By providing a fixed burst length, the circuit **100** may allow shared usage of data, address and control busses. A fixed length non-interruptible burst generally frees up the address bus and control bus for a known number of cycles. The address and control busses may be shared by a number of memory devices. The circuit **100** may provide a more reliable and/or accurate burst than is possible with multiple chips.

Referring to FIG. **2**, a detailed block diagram illustrating implementation of the circuit **102** is shown. The circuit **102** may comprise an address counter register **126** and a burst counter **128**. The address counter register **126** generally receives the signals ADDR_EXT, LOAD, and CLK. The address counter register **126** may be configured to present the signal ADDR_INT. The signal ADV and the signal BURST may be presented to a burst counter **128**. The signal CLK may be presented at an input **130** of the burst counter **128**. The burst counter **128** may have an output **132** that may present a signal (e.g., BURST_CLK) at an input **134** of the circuit **126**. An initial address may be loaded into the address counter register **126** by presenting the initial address in the signal ADDR_EXT and asserting the signal LOAD. The circuit **126** may be configured to increment an address in response to the signal BURST_CLK. When the signal ADV is asserted, the burst counter **128** will generally present the signal BURST_CLK in response to the signal CLK. The signal BURST_CLK generally contains a number of pulses that has been programmed by the signal BURST.

Referring to FIG. **3**, a detailed block diagram illustrating an alternative embodiment of the circuit **102** is shown. The circuit **102'** may comprise a latch **134**, a multiplexer **136** and a counter **138**. The signals ADDR_EXT, LOAD and CLK may be presented to the latch **134**. The latch **134** may have an output **140** that may present a portion (e.g., m bits, where m is an integer smaller than n) of the signal ADDR_EXT as a portion of the signal ADDR_INT, an output **142** that may present a second portion (e.g., k bits, where k is an integer smaller than n) of the signal ADDR_EXT to a first input of the multiplexer **136**, and an output **144** that may present the second portion of the signal ADDR_EXT to an input **146** of the counter **138**.

The signals ADV, CLK and BURST may be presented to inputs of the counter **138**. The counter **138** may be configured to generate a number of addresses in response to the signals CLK, BURST and ADV. The number of addresses generated by the counter **138** may be programmed by the signal BURST.

The signal BURST may be presented to a control input of the multiplexer **136**. The multiplexer **136** may select between a number of signals from the latch **134** and a number of signals from the counter **138** to be presented as a second portion of the signal ADDR_INT in response to the signal BURST.

Referring to FIG. **4**, a flow diagram illustrating an example burst address sequence is shown. When the signal ADV is asserted, the circuit **100** will generally generate a number of address signals, for example, N where N is an integer. The address signals may be generated, in one example, on a rising edge of the signal CLK. The address signals will generally continue to be generated until the Nth address signal is generated.

Referring to FIGS. **5**A and **5**B, timing diagrams illustrating example operations for a four word (FIG. **5**A) and an eight word (FIG. **5**B) fixed burst memory in accordance with the present invention are shown. The timing diagrams generally illustrate externally measurable signals for four and eight word fixed burst read/write architectures. In general, an operation (e.g., read or write) of the circuit **100** begins with loading an initial address (e.g., portions **150**, **154**, and **158** of FIG. **5**A; portions **150'**, **154'**, and **158'** of FIG. **5**B). Starting with the initial address, a fixed number of words are generally transferred (e.g., line DQ of FIGS. **5**A and **5**B). During the transfer of the fixed number of words, the address and control buses (e.g., ADDR, CE, R/W, etc.) are generally available to other devices (e.g., portions **152**, **156**, and **160** of FIG. **5**A; portions **152'**, **156'**, and **160'** of FIG. **5**B). In one example, the control and address bus activity may be one-fourth (FIG. **5**A) or one-eighth (FIG. **5**B) the data bus activity (e.g., compare line ADDR with line DQ of FIGS. **5**A and **5**B). The reduced bus activity may be an effect of the

architecture. The data bus may be, in one example, active nearly 100% of the time (e.g., line DQ of FIGS. **5**A and **5**B) In one example, there may be no inefficiencies switching from read to write to read etcetera (e.g., see labels under line DQ of FIGS. **5**A and **5**B).

Referring to FIG. **6**, a timing diagram illustrating a fixed burst length long enough to hide a writeback and a refresh cycle is shown. Internally the action being performed may completely hide DRAM refresh activity inside nominal external activities. A portion **162** illustrates that refresh activity (e.g., writeback, read for refresh, and writeback for refresh) may be completed within the time of the burst transfer. When a fixed burst long enough to completely hide refresh activity is provided, there may be no penalty for using DRAM instead of SRAM for the memory **104**.

While the invention has been particularly shown and described with reference to the preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made without departing from the spirit and scope of the invention.

What is claimed is:

1. A circuit comprising:

a memory comprising a plurality of storage elements each configured to read and write data in response to an internal address signal; and

a logic circuit configured to generate a predetermined number of said internal address signals in response to (i) an external address signal, (ii) a clock signal and (iii) one or more control signals, wherein said generation of said predetermined number of internal address signals is non-interruptible.

2. The circuit according to claim **1**, wherein said predetermined number of internal address signals is determined by a fixed burst length.

3. The circuit according to claim **1**, wherein said predetermined number of internal address signals is at least 4.

4. The circuit according to claim **1**, wherein said predetermined number of internal address signals is 8.

5. The circuit according to claim **2**, wherein said fixed burst length is programmable.

6. The circuit according to claim **5**, wherein said fixed burst length is programmed by bond options.

7. The circuit according to claim **5**, wherein said fixed burst length is programmed by voltage levels on external pins.

8. The circuit according to claim **1**, wherein said memory comprises a static random access memory.

9. The circuit according to claim **1**, wherein said memory comprises a dynamic random access memory.

10. The circuit according to claim **9**, wherein said predetermined number of internal address signals is chosen to provide time for at least one writeback or refresh cycle.

11. The circuit according to claim **1**, wherein said predetermined number of internal address signals is chosen to meet predetermined criteria for sharing address and control busses.

12. The circuit according to claim **1**, wherein said logic circuit comprises a counter configured to generate said predetermined number of internal address signals.

13. The circuit according to claim **1**, wherein said external address signal comprises an initial address for data transfers to and from said memory.

14. A memory device according to claim **1**, wherein said circuit is an integrated circuit.

15. The circuit according to claim **1**, further comprising address and control busses configured to present said external address signal and said one or more control signals, wherein said busses are freed up during the generation of said predetermined number of internal address signals.

16. A circuit comprising:

means for reading data from and writing data to a plurality of storage elements in response to a plurality of internal address signals; and

means for generating a predetermined number of said internal address signals in response to (i) an external address signal, (ii) a clock signal and (iii) one or more control signals, wherein said generation of said predetermined number of internal address signals is non-interruptible.

17. A method of providing a fixed burst length data transfer comprising the steps of:

accessing a memory in response to a plurality of internal address signals; and

generating a predetermined number of said internal address signals in response to (i) an external address signal, (ii) a clock signal and (iii) a control signal, wherein said generation of said predetermined number of internal address signals is non-interruptible.

18. The method according to claim **17**, further comprising the step of programming said predetermined number.

19. The method according to claim **18**, wherein said programming step is performed using bond options.

20. The method according to claim **18**, wherein said programming step is performed using voltage levels.

21. The method according to claim **17**, further comprising the step of selecting said predetermined number to provide time for at least one writeback or refresh cycle.

* * * * *

EXHIBIT E

US006680516B1

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

(10) **Patent No.: US 6,680,516 B1**
(45) **Date of Patent: Jan. 20, 2004**

(54) **CONTROLLED THICKNESS GATE STACK**

(75) Inventors: **Alain Blosse**, Belmont, CA (US); **Krishnaswamy Ramkumar**, San Jose, CA (US)

(73) Assignee: **Cypress Semiconductor Corp.**, San Jose, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/313,267**

(22) Filed: **Dec. 6, 2002**

(51) **Int. Cl.**[7] **H01L 29/76**
(52) **U.S. Cl.** **257/412**; 257/413; 438/197
(58) **Field of Search** 257/412, 413; 438/197

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,541,830 B1 * 4/2003 Iyer 257/412
6,552,401 B1 * 4/2003 Dennison 257/375

OTHER PUBLICATIONS

Encyclopedia of Chemical Technology, Kirk–Othmer, vol. 14, pp. 677–709 (1995).
Diaz, C.H., H. Tao, Y. Ku, A. Yen, and K. Young, 2001. "An Experimentally Validated Analytical Model For Gate Line–Edge Roughness (LER) Effects on Technology Scaling", IEEE Electron Device Letters, 22(6)287–289.

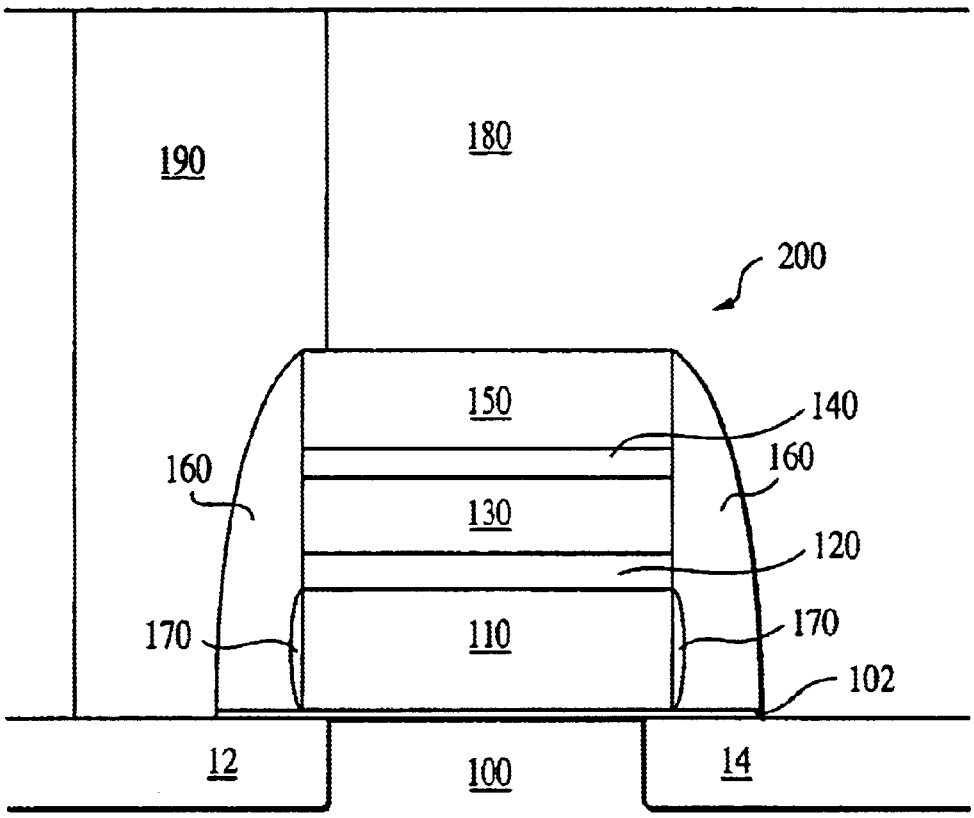
* cited by examiner

Primary Examiner—Mark V. Prenty
(74) *Attorney, Agent, or Firm*—Sonnenschein Nath & Rosenthal LLP

(57) **ABSTRACT**

A semiconductor structure, comprises a semiconductor substrate, a gate layer on the semiconductor substrate, a metallic layer on the gate layer, and an etch-stop layer on the metallic layer. A distance between the substrate and a top of the etch-stop layer is a gate stack height, and the gate stack height is at most 2700 angstroms. In addition, the etch-stop layer has a thickness of at least 800 angstroms.

20 Claims, 6 Drawing Sheets



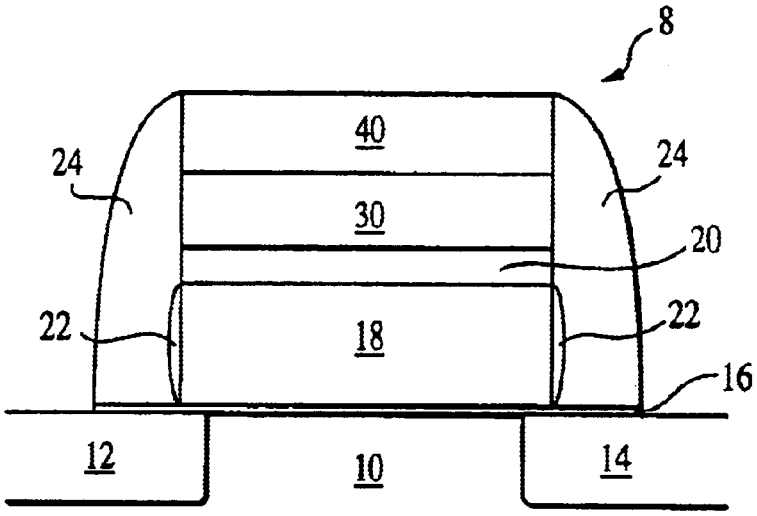


FIG. 1

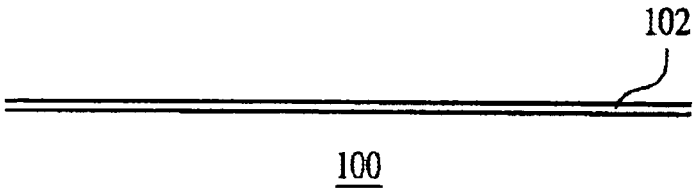


FIG. 2

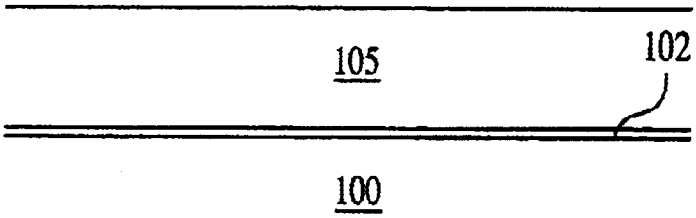


FIG. 3

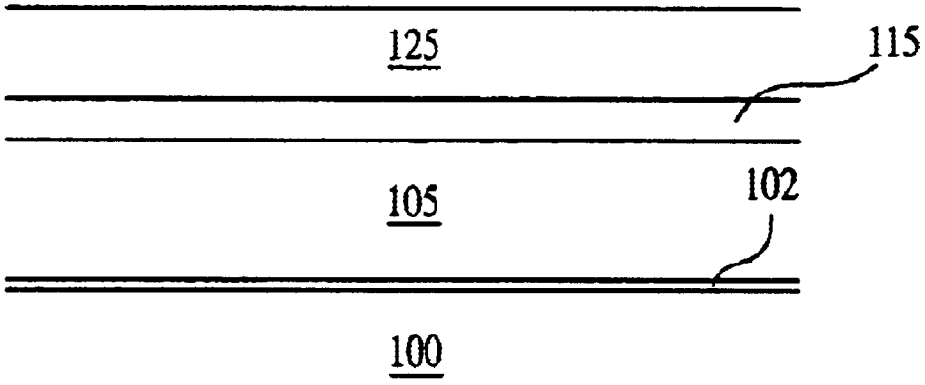


FIG. 4

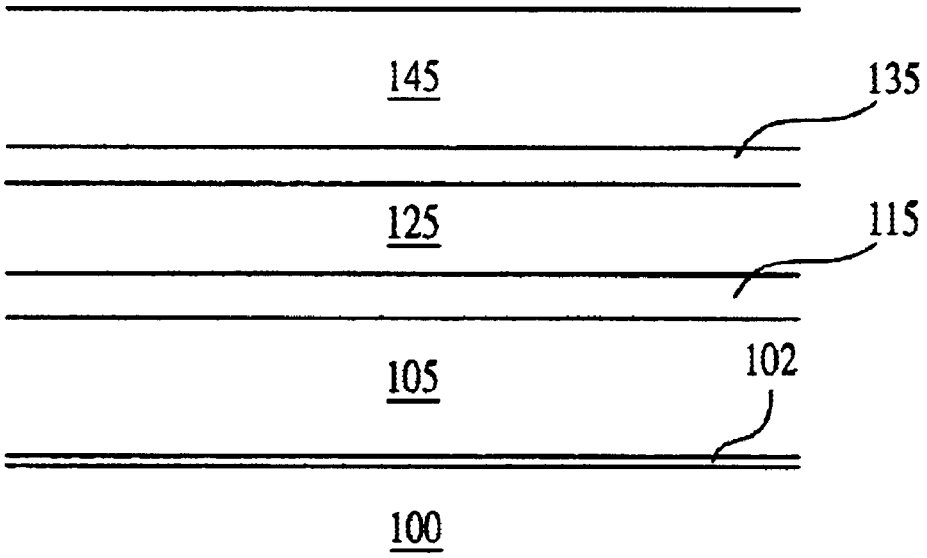


FIG. 5

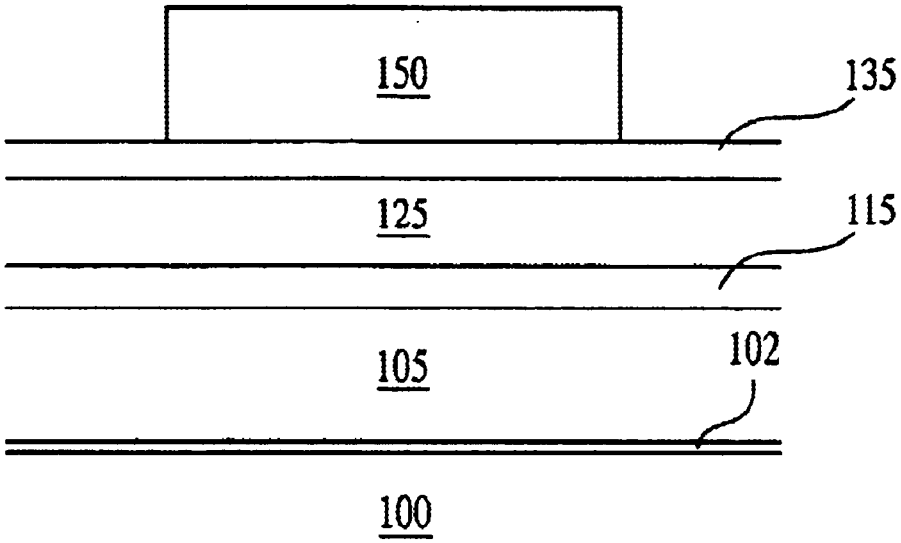


FIG. 6

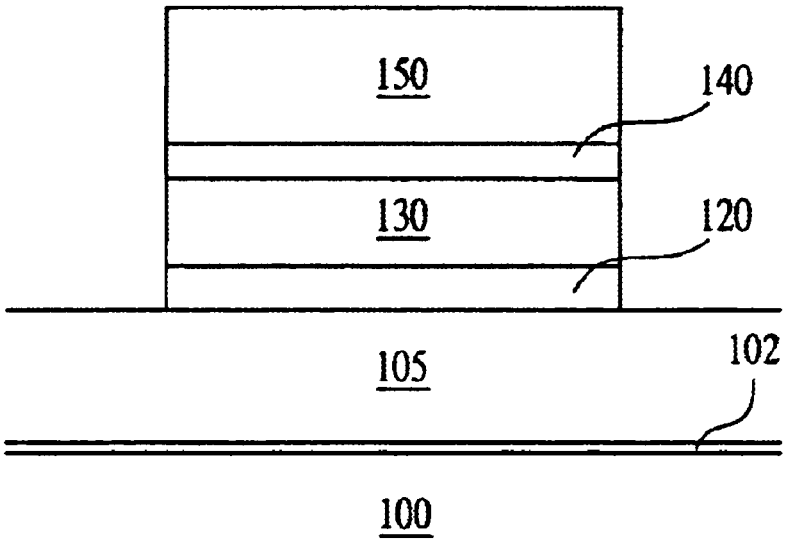


FIG. 7

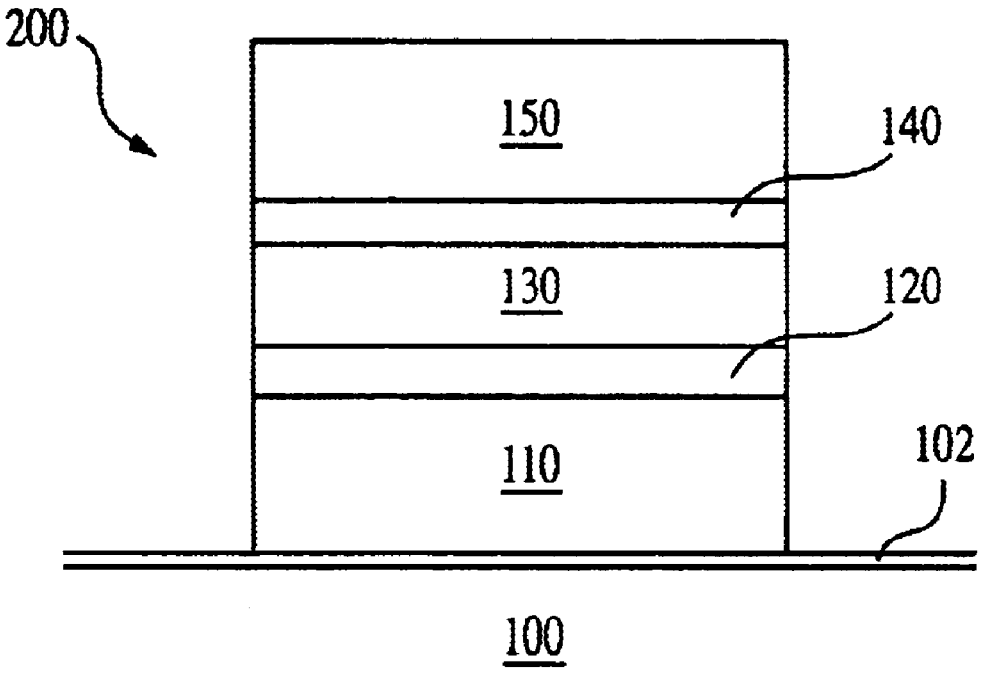


FIG. 8

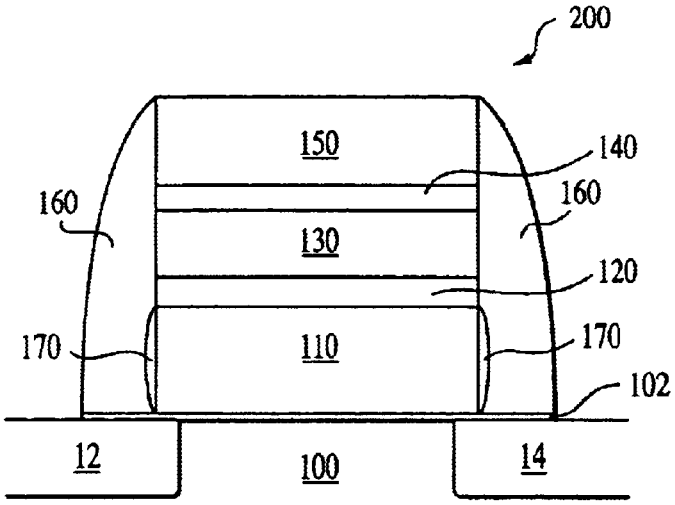

200
150
140
160
160
130
120
170
110
170
102
12
100
14

FIG. 9

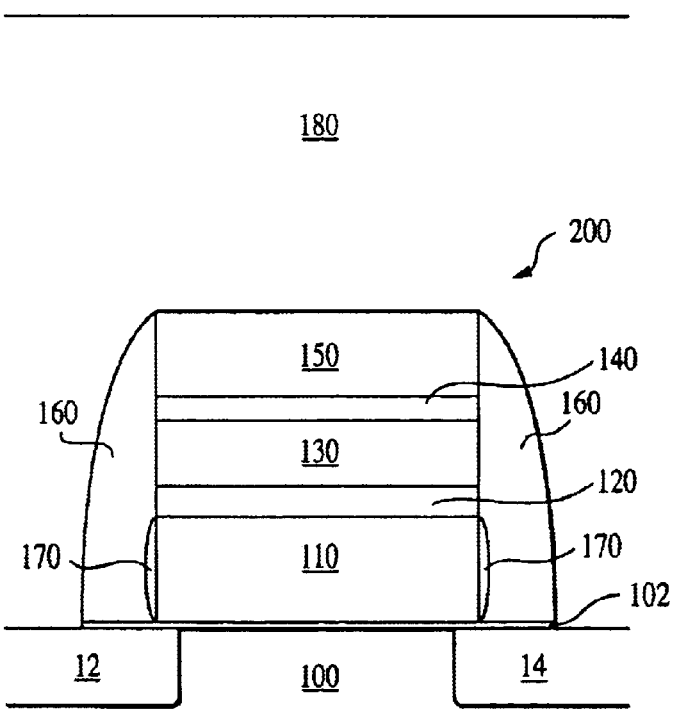

180
200
150
140
160
160
130
120
170
110
170
102
12
100
14

FIG. 10

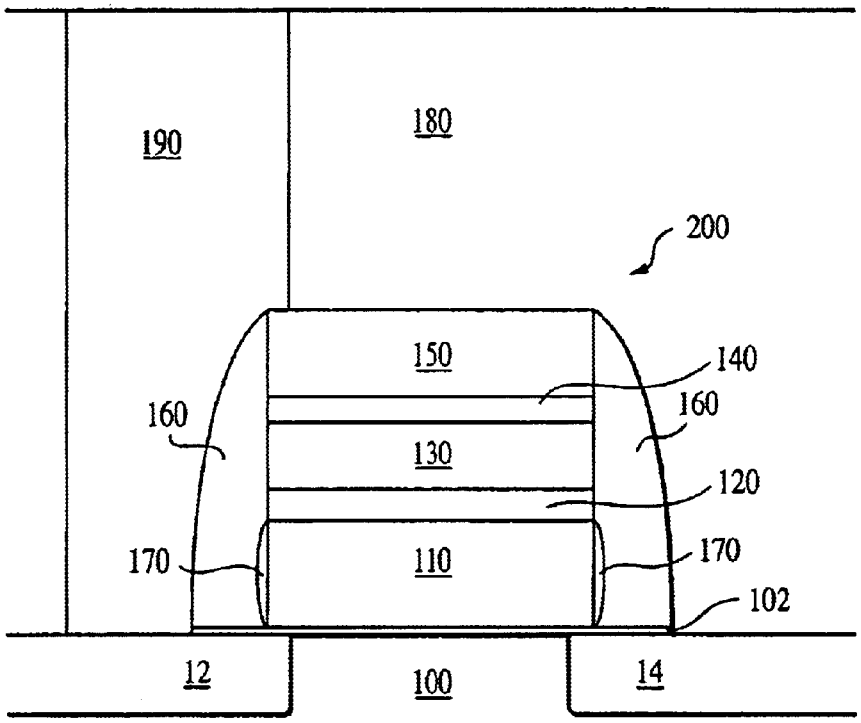


FIG. 11

CONTROLLED THICKNESS GATE STACK

BACKGROUND

Modern integrated circuits are constructed with up to several million active devices, such as transistors and capacitors, formed in and on a semiconductor substrate. Interconnections between the active devices are created by providing a plurality of conductive interconnection layers, such as polycrystalline silicon and metal, which are etched to form conductors for carrying signals. The conductive layers and interlayer dielectrics are deposited on the silicon substrate wafer in succession, with each layer being, for example, on the order of 1 micron in thickness.

A gate structure is an element of a transistor. FIG. **1** illustrates an example of a gate stack **8**. A semiconductor substrate **10** supports a gate insulating layer **16**, which overlaps doped regions (source/drain regions) in the substrate (**12** and **14**), and the gate insulating layer supports a gate **18**, which is typically polycrystalline silicon. On the gate is a metallic layer **30**. The metallic layer may be separated from the gate by one or more other layers, such as nitrides, oxides, or silicides, illustrated collectively as barrier layer **20**. The metallic layer may in turn support one or more other layers (collectively **40**), such as nitrides, oxides, or silicides. Oxide **22** may be formed on the sides of the gate to protect the gate oxide at the foot of the gate stack; and insulating spacers **24** may be formed on either side of the gate stack. Furthermore, contacts to the source/drain regions in the substrate, and to the gate structure, may be formed.

Self-aligned contacts (SAC) allow the design of a semiconductor device to have a distance between the gate and the via contact to the substrate, to be at most one-half the minimum gate width; the contact may even be designed to overlay the gate. Typically, SAC uses a nitride layer on the gate stack, together with spacers that include nitride, to prevent a misaligned contact from electrically contacting the gate itself. If the nitride were not present, then the etch used to form the hole which will become the contact would pass through the dielectric layer all the way to the gate. When present, the nitride layer and spacers act as an etch stop, preventing misalignment from forming a hole all the way to the gate, and therefore allowing design of the device to have a much smaller average distance between the contact and the gate.

The nitride layer on the gate stack has at least a thickness of 800 angstroms when used for forming SAC. If used only for other purposes, such as an etch-stop layer or a hard mask, a thickness of less than 800 angstroms is used. Also, the thickness of at least 800 angstroms is the thickness after the dielectric layer has been formed; the nitride layer is usually thicker when originally formed, allowing for a loss of about 500 angstroms during the gate etch (i.e. thickness for the hard mask function), and a loss of about 200 angstroms during nitride spacer formation.

There is an ongoing need to reduce the size of the elements within integrated circuits and semiconductor structures. As the size of an element is reduced, it does not necessarily follow that the thickness of specific layers that form that element can be equally reduced: for example, the thickness of a nitride layer necessary for forming SAC may still need to be at least 800 angstroms. This could lead to designs for devices that require large aspect ratios for contact vias; such vias may not be properly filled.

BRIEF SUMMARY

In a first aspect, the present invention is a semiconductor structure, comprising a semiconductor substrate, a gate layer on the semiconductor substrate, a metallic layer on the gate layer, and an etch-stop layer on the metallic layer. A distance between the substrate and a top of the etch-stop layer is a gate stack height, and the gate stack height is at most 2700 angstroms. In addition, the etch-stop layer has a thickness of at least 800 angstroms.

In a second aspect the present invention is a semiconductor structure, comprising a semiconductor substrate, a gate layer on the semiconductor substrate, a metallic layer on the gate layer, an etch-stop layer on the metallic layer, an insulating layer on the etch-stop layer and on the substrate, and a via through the insulating layer on the substrate. An area of contact between the via and the substrate has a via width, and the via width is at most 0.12 micron. In addition, a distance between the substrate and a top of the etch-stop layer has a gate stack height, and the gate stack height is at most 2700 angstroms.

In a third aspect, the present invention is a method of making a semiconductor structure, comprising forming a gate layer on a semiconductor substrate, forming a metallic layer on the gate layer, and forming an etch-stop layer having a thickness of at least 1500 angstroms on the metallic layer. A distance between the substrate and a top of the etch-stop layer is a gate stack height, and the gate stack height is at most 2700 angstroms.

The term "distance" means the length between the closest edges of the two objects.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** shows a gate stack structure.

FIGS. **2–7** illustrate a method of forming the structure of FIG. **8**.

FIG. **8** shows a gate stack of the present invention.

FIGS. **9** and **10** show further processing of the gate stack of FIG. **8**.

FIG. **11** shows the gate stack of FIG. **8** after further processing.

DETAILED DESCRIPTION

The present invention includes a gate stack that has a controlled thickness, to avoid the necessity of forming contact vias having a large aspect ratio. The gate stack includes a nitride layer that may be used to for forming SAC, and may be used in designs have a significantly reduced device size.

Referring to FIG. **2**, a gate insulating layer **102** is on a semiconductor substrate **100**. The semiconductor substrate may be a conventionally known semiconductor material. Examples of semiconductor materials include silicon, gallium arsenide, germanium, gallium nitride, aluminum phosphide, and alloys such as $Si_{1-x}Ge_x$ and $Al_xGa_{1-x}As$, where $0 \leqq x \leqq 1$. Preferably, the semiconductor substrate is silicon, which may be doped or undoped. The gate insulating layer **102** may be a conventionally known insulating material. For example, the gate insulating layer may contain silicon oxide or silicon oxynitride.

Referring to FIG. **3**, a gate layer **105** may be formed on the gate insulating layer. The gate layer preferably has a thickness of 400–1200 angstroms, more preferably a thickness of 600–1000 angstroms, most preferably a thickness of 750–850 angstroms. The gate layer may contain a variety of semiconducting materials. Typically, a gate layer contains polycrystalline silicon (poly) or amorphous silicon. The gate layer may be doped with one type of dopant (P^+ or N^+), or it may contain both types of dopants in discrete regions. A split gate is a gate layer containing both P^+ and N^+ doping regions.

In the case of a split gate, those regions of the gate that are P^+ doped (such as with B or BF_2^+) are over N^- doped channel regions of the substrate, forming a PMOS device; those regions of the gate that are N^+ doped (such as with As^+ or phosphorus$^+$) are over P^- doped channel regions of the substrate, forming an NMOS device. The P^+ and N^+ doping regions of the gate are separated by a region which is on an isolation region of the substrate; this isolation region has a width of at most 0.4 microns, more preferably at most 0.36 microns. The doping of the regions of the gate is preferably carried out after forming the gate, by masking and doping each region separately, or by an overall doping of the gate with one dopant type, and then masking and doping only one region with the other dopant type (counter doping).

Referring to FIG. **4**, a barrier layer **115** may optionally be formed on the gate layer. The barrier layer preferably has a thickness of 30–120 angstroms, more preferably 50–100 angstroms, most preferably 60–80 angstroms. The optional barrier layer may contain a variety of materials, including nitrides, silicides, and oxides, and is preferably a conductive material. For example, the barrier layer may contain refractory silicides and nitrides. Preferably, the barrier layer contains silicon nitride, or a nitride or suicide of a metal such as tantalum, titanium, niobium or tungsten, for example tungsten nitride.

Referring still to FIG. **4**, a metallic layer **125** may be formed on the gate layer, or the barrier layer **115**, if it is present. Preferably, the metallic layer has a thickness of 200–600 angstroms, more preferably 300–500 angstroms, most preferably 325–450 angstroms. The metallic layer **125** may contain a variety of metal-containing materials. For example, a metallic layer may contain aluminum, copper, tantalum, titanium, tungsten, or alloys or compounds thereof. Preferably, the metallic layer comprises tungsten or titanium. The metallic layer may be formed, for example, by physical vapor deposition (PVD) of the metal, or by low pressure chemical vapor deposition (LPCVD) of a mixture of a metal halide and hydrogen.

Referring to FIG. **5**, a barrier layer **135** may optionally be formed on the metallic layer. The formation of the second optional barrier layer may be performed as described for the first optional barrier layer **115**, and this layer may be formed of the same materials, and to the same thicknesses.

Referring still to FIG. **5**, an etch-stop layer **145** may be formed on the metallic layer by a variety of methods, including chemical vapor deposition (CVD). Preferably, the etch-stop layer is a nitride layer. More preferably, the etch-stop layer is silicon nitride formed by PECVD. The etch-stop layer may vary in composition, so that the top of the etch-stop layer is anti-reflective, for example so that the top of the etch-stop layer is silicon rich silicon nitride, or silicon oxynitride; this layer may also act as a hard mask to protect the etch-stop layer during subsequent etches. Alternatively, a separate anti-reflective layer (ARC) may be formed.

Preferably, the etch-stop layer is formed rapidly at a relatively low temperature. For example, if the gate layer contains both P^+ and N^+ doping regions, diffusion of the dopants may occur if the wafer is maintained at sufficiently high temperatures for a prolonged period of time. Thus, it is desirable that any high temperature processing is performed only for relatively short periods of time. Likewise, it is desirable that any lengthy processing is carried out at relatively low temperatures. Preferably, the etch-stop layer is formed at a temperature of at most 750° C., if the atmosphere is substantially devoid of oxygen, or in a reducing environment (hydrogen rich). Under typical conditions, a temperature of at most 600° C. is preferred, at most 450° C. is more preferred. A temperature of at least 350° C. is preferred, such as 400° C. The depositing of the etch-stop layer is preferably carried out at a temperature and for a time that does not result in substantial diffusion between the P^+ region and the N^+ region in a split gate.

Preferably, the etch-stop layer has a thickness of at least 800 angstroms, more preferably a thickness of at least 1100 angstroms, most preferably a thickness of at least 1200, after etching of the gate layer, and after formation of gate spacers. About 500 angstroms of etch-stop may be lost during the gate layer etch, and about 200 angstroms of etch-stop may be lost during the spacer formation. Preferably, at least 1500 angstroms thickness of etch-stop are deposited, more preferably at least 1800 angstroms thickness of etch-stop are deposited, most preferably 2100 angstroms thickness of etch-stop are deposited. Preferably, after the gate layer etch and after spacer formation (or, alternatively, after the dielectric layer is formed), the etch-stop layer has a thickness of 800–1800 angstroms, more preferably a thickness of 1100–1500 angstroms, most preferably a thickness of 1200–1400 angstroms. Similarly, the thickness deposited would preferably be these same ranges, with an additional 700 angstroms added to accommodate loss during the gate layer etch and spacer formation, when material are used which may result in a loss of the etch-stop layer at these points in the process.

"Substantial diffusion between the P^+ region and the N^+ region" of the gate means that the threshold voltage (V_T) one or both of the PMOS device or NMOS device changes by more than 20 mV, more preferably 10 mV, even more preferably 5 mV. In order to determine if a nitride deposition results in substantial diffusion between the P^+ region and the N^+ region of a particular split gate, a single PMOS device or NMOS device is formed, with the other, part of the gate forming a comparatively very large reservoir of the opposite doping type, separated by an isolation region of the same size as the actual device. The PMOS device or NMOS device is formed using the nitride deposition, and using BTBAS (bis-(t-butylamino)silane) and ammonia at about 550° C. for spacer deposition, and low temperature selective oxidation, as used in the example.

Referring to FIGS. **6–8**, each layer may be patterned to form the gate stack. The patterning may be accomplished, for example, by conventional photolithographic and etching techniques. Referring to FIG. **6**, the etch-stop layer may be etched to form a patterned etch-stop layer **150**, for example by forming a patterned photoresist on etch-stop layer **145** (FIG. **5**) and then etching the exposed portions of the layer. The etch-stop etching may be carried out by conventional etching techniques, for example by exposure to a plasma formed from a mixture of CF_4 and CHF_3.

Referring to FIG. **7**, the patterned etch-stop layer may be used as a hard mask for the etching of the metallic layer **125** (FIG. **6**) to form a patterned metallic layer **130**. Referring to

FIG. **8**, the patterned etch-stop layer and the patterned metallic layer may be used as a hard mask for the etching of the gate layer **105** (FIG. **7**) to form patterned gate layer **110**. The gate etching may be carried out by conventional gate etch techniques, for example by exposure to a plasma formed from chlorine, hydrobromic acid and/or oxygen.

FIG. **8** thus illustrates a gate stack **200** which may be formed on a semiconductor wafer. Semiconductor substrate **100** supports a gate insulating layer **102**, which in turn supports a gate layer **110**. The gate layer supports a metallic layer **130**, which may optionally be separated from the gate layer by barrier layer **120**. The metallic layer may optionally support a barrier layer **140**. The etch-stop layer **150** is on the metallic layer **130**, or optionally on the layer **140** above the metallic layer.

FIGS. **9–11** illustrate further processing of the gate structure. As shown in FIG. **9**, sidewall oxide regions **170** on gate layer **110**, and spacers **160** (preferably containing nitride), may be formed on the sides of the stack. Furthermore, as shown in FIG. **10**, a dielectric layer **180** maybe formed on the etch-stop layer.

Next, as shown in FIG. **11**, a via **190** formed through the dielectric to the substrate, may be formed. This via may be lined and filled to form a via-contact, for example with TiN and tungsten, respectively. Other processing may include forming contacts to the gate itself.

If the aspect ratio of the via is too large, then it may not be possible to properly fill the via with a conductor. The area of contact between the via and the substrate is know as the via width; the distance between the substrate and the top of the etch-stop layer is the gate stack height. The aspect ratio of the via, which is the ratio of the gate stack height to the via width, is preferably at most 5, more preferably at most 4, and most preferably at most 3.

Preferably, the gate stack height is at most 2700 angstroms, more preferably at most 2600 angstroms, most preferably at most 2500 angstroms, such as 2430 angstroms. The via width is preferably at most 0.12 microns, more preferably at most 0.1 microns, such as 0.05–0.1 microns.

Other processing may be used to complete formation of semiconductor devices from the semiconductor structure. For example, source/drain regions **12**, **14** may be formed in the substrate, additional dielectric layers may be formed on the substrate, and contacts and metallization layers may be formed on these structures. These additional elements may be formed before, during, or after formation of the gate stack.

The related processing steps, including the etching of the gate stack layers and other steps such as polishing, cleaning, and deposition steps, for use in the present invention are well known to those of ordinary skill in the art, and are also described in Encyclopedia of Chemical Technology, Kirk-Othmer, Volume 14, pp. 677–709 (1995); Semiconductor Device Fundamentals, Robert F. Pierret, Addison-Wesley, 1996; Wolf, Silicon Processing for the VLSI Era, Lattice Press, 1986, 1990, 1995 (vols 1–3, respectively), and Microchip Fabrication 4rd. edition, Peter Van Zant, McGraw-Hill, 2000.

The semiconductor structures of the present invention may be incorporated into a semiconductor device such as an integrated circuit, for example a memory cell such as an SRAM, a DRAM, an EPROM, an EEPROM etc.; a programmable logic device; a data communications device; a clock generation device; etc. Furthermore, any of these semiconductor devices may be incorporated in an electronic device, for example a computer, an airplane or an automobile.

EXAMPLE

Example 1

Formation of a Gate Structure

The following detailed steps were used to form the gate stack having a split gate:

Nitrogen-containing gate oxide formation
Poly deposition - single amorphous gate deposit
Mask for P-doping
P^+ poly implantation
N-well implantation
P-channel implantation
P^+ poly implantation strip resist
Mask for N-doping
P-well implantation
N^+ poly implantation
N-channel implantation
N^+ poly implantation strip resist
Tungsten gate pre-clean
Tungsten PVD, sputtering (nitrogen + argon, then argon only)
Nitride - PECVD
Deposit ARC and Resist
Etch mask for nitride
Nitride etch - ARC, silicon nitride, and partial tungsten etch
Remove resist
Tungsten and Poly etch
Post-poly etch clean
Selective oxidation
N^+ source/drain extension implant
Stripping & cleaning
P^+ source/drain extension implant
Stripping & cleaning
Nitride spacer deposition (BTBAS chemistry)
Spacer etch
Post-spacer etch clean
N^+ source/drain implant
Stripping & cleaning
P^+ source/drain implant
Stripping & cleaning
Nitride-poly cut mask etch and clean
Dielectric deposition/planarization/mask for contacts
self-aligned contact (SAC) etch
SAC etch clean

Bottom anti-reflective coating (BARC) was etched under the following conditions: CF_4 flow rate of 100 sccm, Ar flow rate of 100 sccm, power of 600 W, bias of 75 W, a pressure of 16 mTorr, and a temperature of 60° C. Temperature is controlled, for example, by He backside cooling during the BARC etch, as well as in subsequent steps.

The resist was then trimmed under the following conditions: HBr flow rate of 160 sccm, O_2 flow rate of 28 sccm, power of 400 W, a pressure of 8 mTorr, a temperature of 60° C., and for a time of 10 seconds. Alternatively, the resist was trimmed under the following conditions: HBr flow rate of 169 sccm, O_2 flow rate of 19 sccm, power of 400 W, a pressure of 8 mTorr, a temperature of 60° C., and for a time of 5 seconds.

Etching of the nitride was carried out with a plasma at a pressure of 30 mTorr, at a power of 500 W, a bias of 100 W, and at a temperature of 60–65° C. The gas composition is CHF_3 at 275 sccm and CF_4 at 300 sccm. An overetch of 20% was used.

The tungsten was partially etched under the following conditions: NF_3 flow rate of 10 sccm, Cl_2 flow rate of 25 sccm, O_2 flow rate of 5 sccm, Ar flow rate of 50 sccm, N_2 flow rate of 30 sccm, He flow rate of 150 sccm, power of 800 W, bias of 60 W, a pressure of 4 mTorr, a temperature of 60° C., and for a time of 10 seconds.

Resist material was removed by ashing (for example at 80° C. with a mixture of CF_4 and O_2), and the stack was cleaned by treating the wafer with EKC265™ (EKC, Hayward, Calif.; a mixture of 2-(2 aminoethoxy) ethanol, hydroxylamine and catechol) by spinning with spraying (using a spray tool) at 65 or 70° C. for 10 minutes, then 2 minutes at 20° C., followed by rinsing with deionized water, to prevent undesirable oxidation of the tungsten. This clean may be used for any stripping and cleaning step where tungsten or tungsten nitride is exposed to prevent undesirable oxidation. Also, the clean may be carried out with downstream plasma ashing under the following conditions, followed by washing with water: step 1: CF_4 flow rate of 50 sccm, H_2O flow rate of 160 sccm, N_2/H_2 flow rate of 1400 sccm, power of 1050 W, bias of 100 W, a pressure of 750 mTorr, a temperature of 80° C., and for a time of 30 seconds; step 2: NF_3 flow rate of 40 sccm, H_2O flow rate of 170 sccm, O_2 flow rate of 170 sccm, bias of 150 W, a pressure of 250 mTorr, a temperature of 80° C., and for a time of 120 seconds.

The tungsten was then etched under the following conditions: NF_3 flow rate of 15 sccm, Cl_2 flow rate of 25 sccm, O_2 flow rate of 5 sccm, Ar flow rate of 50 sccm, N_2 flow rate of 30 sccm, He flow rate of 150 sccm, power of 800 W, bias of 35 W, a pressure of 4 mTorr, and a temperature of 60° C. An overetch of the tungsten was carried out for 5 seconds. The system was then pumped down for 20 seconds.

The poly was then etched under the following conditions: HBr flow rate of 250 sccm, He(80%)/O_2(20%) flow rate of 12 sccm, power of 450 W, a bias of 40 W, a pressure of 25 mTorr, and a temperature of 60° C. A poly overetch was carried out under the following conditions: HBr flow rate of 150 sccm, He(80%)/O_2(20%) flow rate of 8 sccm, He flow rate of 100 sccm, power of 200W, a bias of 70W, a pressure of 70 mTorr, a temperature of 60° C., and for a time of 63 seconds. Alternatively, the poly overetch was carried out under the following conditions: HBr flow rate of 150 sccm, He(80%)/O_2(20%) flow rate of 13 sccm, He flow rate of 200 sccm, power of 250 W, a bias of 60 W, a pressure of 80 mTorr, a temperature of 60° C., and for a time of 53 seconds. Cleaning may be carried out as described above, or for example, by downstream, followed by rinsing with water (for example with deionized water for 7 cycles), under the following conditions: CF_4 flow rate of 40 sccm, O_2 flow rate of 1000 sccm, H_2O flow rate of 200 sccm, N_2 flow rate of 150 sccm, power of 1700 W, a pressure of 700 mTorr, a temperature of 70° C., and for a time of 80 seconds.

The exposed sides of the poly were covered with a layer of oxide about 50–70 angstroms thick by the selective oxidation. This was carried out by exposing the stack to a mixture of hydrogen and oxygen (10% steam) at a temperature of 750° C. to selectively oxidize the poly relative to the tungsten and tungsten nitride.

BTBAS was used to form a nitride layer for spacer formation under the following conditions: BTBAS flow rate of 50 sccm, NH_3 flow rate of 100 sccm, a pressure of 150 mTorr, and a temperature of 550° C.

Etching of the nitride (nitride-poly cut mask etch and clean) was carried out with a plasma at a pressure of 35 mT, at a power of 280 W, and a temperature of 15°C. The gas composition for the main etch was CHF_3 at 30 sccm, Ar at 60 sccm, and O_2 at 10 sccm. The clean was carried out with plasma ashing in two steps, followed by a solvent clean:

Step 1:

pressure of 2 mTorr, temperature of 185°C., microwave power of 800 W, gas: O_2 at 3750 sccm, N_2 at 375 sccm;

Step 2:

same values, except a temperature of 200° C. and microwave power of 1400 W.

Etching to form contacts (SAC etch) was carried out with a plasma at a pressure of 55 mTorr, a power of 500 W, a temperature of 35° C., with the magnet at 20 Gauss, a gas of CF_4 at 5 sccm, CHF_3 at 10 sccm, $C_2H_2F_4$ at 10 sccm, and Ar at 90 sccm, as the ARC etch; and as the main etch a pressure of 55 mTorr, a power of 500 W, a tempature of 35° C., with the magnet at 25 Gauss, a gas of CHF_3 at 80 sccm, $C_2H_2F_4$ at 8 sccm, and Ar at 90 sccm. The clean was carried out with plasma ashing in two steps, followed by a solvent clean:

Step 1:

pressure of 400 mTorr, temperature of 20+/−5° C., RF power of 420 W, gas: O_2 at 400 sccm.

Step 2:

pressure of 750 mTorr, temperature of 20+/−5° C., RF power of 420 W, gas: N_2 at 400 sccm, H_2 at 400 sccm, and NF_3 at 5 sccm; or alternatively:

pressure of 750 mTorr, temperature of 40+/−5° C., RF power of 350 W, gas: CF_4 at 20 sccm, N_2/5% H_2 at 200 sccm, and O_2 at 500 sccm.

SAC etch clean was carried out using EKC 265™, with a spray tool: temperature of 70° C. for 10 minutes, and an extra 2 minutes at 20° C., followed by rinsing with deionized water and then spin drying in N_2; then washed with H_2SO_4 at 150° C. twice for 10 minutes each and then spin drying in N_2.

In the stack, the silicon nitride layer had a thickness of 1300 angstroms (although the actual amount deposited was greater since silicon nitride is lost during the poly etch and during spacer etch), the tungsten layer had a thickness of 325 angstroms, the tungsten nitride layer had a thickness of 75 angstroms, and the poly layer had a thickness of 735 angstroms. The contacts having a width of 0.13 microns at the top, and a width of 0.05 microns at the bottom.

Patent application Ser. No. 10/314,380, entitled "MULTI-LAYERED GATE STACK" by Saurabh Dutta Chowdhury, Cypress ref. PM02012, filed on the same day as the present application, is hereby incorporated by reference.

What is claimed is:

1. A semiconductor structure, comprising:

a semiconductor substrate,

a gate layer, on the semiconductor substrate,

a metallic layer, on the gate layer, and

an etch-stop layer, on the metallic layer,

wherein a distance between the substrate and a top of the etch-stop layer is a gate stack height,

the gate stack height is at most 2700 angstroms,

the etch-stop layer has a thickness of at least 800 angstroms,

the gate layer comprises a P^+ region and an N^+ region, and the P^+ and N^+ regions are separated by a region which is on an isolation region of the substrate having a width of at most 0.4 microns, and

the etch-stop layer comprises nitride.

2. The semiconductor structure of claim **1**, wherein the etch-stop layer has a thickness of at least 1100 angstroms.

3. The semiconductor structure of claim **1**, wherein the metallic layer comprises tungsten.

4. An integrated circuit comprising a plurality of semiconductor structures of claim **1**,

wherein the semiconductor structures further comprise a dielectric layer on the substrate, and a via contact through the dielectric layer to the substrate, and

wherein the average distance between the via contact and the gate layer in the plurality of semiconductor structures is at most one-half the minimum gate layer width.

5. A semiconductor structure, comprising;

a semiconductor substrate,

a gate layer, on the semiconductor substrate,

a metallic layer, on the gate layer,

an etch-stop layer, on the metallic layer,

an insulating layer, on the etch-stop layer, and on the substrate, and

a via, through the insulating layer, on the substrate,

wherein an area of contact between the via and the substrate has a via width,

the via width is at most 0.12 micron,

a distance between the substrate and a top of the etch-stop layer has a gate stack height, and

the gate stack height is at most 2700 angstroms.

6. The semiconductor structure of claim **5**, wherein the etch-stop layer has a thickness of at least 800 angstroms.

7. The semiconductor structure of claim **3**, wherein the etch stop layer comprises nitride.

8. The semiconductor structure of claim **7**, wherein the etch stop layer has a thickness of at least 1100 angstroms.

9. The semiconductor structure of claim **7**, wherein the gate layer comprises a P^+ region and an N^+ region, and

wherein the P^+ and N^+ regions are separated by a region which is on an isolation region of the substrate having a width of at most 0.4 microns.

10. The semiconductor structure of claim **7**, wherein the metallic layer comprises tungsten.

11. An integrated circuit comprising a plurality of semiconductor structures of claim **5**, wherein the average distance between the via and the gate layer in the plurality of semiconductor structures is at most one-half the minimum gate layer width.

12. A method of making a semiconductor structure, comprising:

forming a gate layer, on a semiconductor substrate,

forming a metallic layer, on the gate layer, and

forming an etch-stop layer having a thickness of at least 1500 angstroms, on the metallic layer,

forming an insulating layer, on the etch-stop layer, and on the substrate, and

forming a via, through the insulating layer, to the substrate,

wherein a distance between the substrate and a top of the etch-stop layer is a gate stack height, and

the gate stack height is at most 2700 angstroms,

an area of contact between the via and the substrate has a via width,

the via width is at most 0.12 micron, and

the etch-stop layer has a thickness of at least 800 angstroms after forming the insulating layer, and

the etch stop layer comprises nitride.

13. The method of claim **12**, wherein the forming of the etch-stop layer is forming the etch-stop layer having a thickness of at least 1800 angstroms.

14. The method of claim **12**, wherein

the forming of the gate layer is forming the gate layer having a thickness of 400–1200 angstroms, and

the forming of the metallic layer is forming the metallic layer having a thickness of 200–600 angstroms.

15. A method of making a semiconductor device, comprising:

forming a semiconductor structure by the method of claim **12**, and

forming a semiconductor device from the semiconductor structure.

16. A method of making an electronic device, comprising:

forming a semiconductor device by the method of claim **13**, and

forming an electronic device comprising the semiconductor device.

17. An integrated circuit comprising a plurality of semiconductor structures, wherein the semiconductor structures comprise:

a semiconductor substrate,

a gate layer, on the semiconductor substrate,

a metallic layer, on the gate layer,

an etch-stop layer, on the metallic layer,

a dielectric layer on the substrate, and a via contact through the dielectric layer to the substrate,

wherein a distance between the substrate and a top of the etch-stop layer is a gate stack height,

the gate stack height is at most 2700 angstroms, and

the etch-stop layer has a thickness of at least 800 angstroms, and

the average distance between the via contact and the gate layer in the plurality of semiconductor structures is at most one-half the minimum gate layer width.

18. The integrated circuit of claim **17** wherein the etch-stop layer of the semiconductor structures comprises nitride.

19. The integrated circuit of claim **17** wherein the etch-stop layer of the semiconductor structures has a thickness of at least 1100 angstroms.

20. The integrated circuit of claim **17** wherein the metallic layer of the semiconductor structures comprises tungsten.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,680,516 B1
APPLICATION NO. : 10/313267
DATED : January 20, 2004
INVENTOR(S) : Alain Blosse et al.

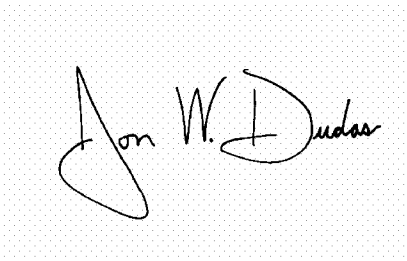
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In the claims:

Column 9, line 20, please delete "3" and insert --5--.
Column 10, line 21, please delete "13" and insert --15--.

Signed and Sealed this

Twelfth Day of December, 2006



JON W. DUDAS
Director of the United States Patent and Trademark Office

EXHIBIT F

US006765407B1

(12) **United States Patent**
Snyder

(10) **Patent No.: US 6,765,407 B1**
(45) **Date of Patent: Jul. 20, 2004**

(54) **DIGITAL CONFIGURABLE MACRO ARCHITECTURE**

(75) Inventor: **Warren Snyder**, Snohomish, WA (US)

(73) Assignee: **Cypress Semiconductor Corporation**, San Jose, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/272,231**

(22) Filed: **Oct. 15, 2002**

Related U.S. Application Data

(63) Continuation of application No. 09/909,045, filed on Jul. 18, 2001, now Pat. No. 6,507,214.
(60) Provisional application No. 60/243,708, filed on Oct. 26, 2000.

(51) **Int. Cl.**[7] **H03K 7/38**
(52) **U.S. Cl.** **326/38**; 326/40; 326/37
(58) **Field of Search** 326/37–41

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,202,687 A		4/1993	Distinti	341/158
5,426,378 A		6/1995	Ong	326/39
5,608,892 A	*	3/1997	Wakerly	711/118
5,703,871 A		12/1997	Pope et al.	370/248
5,828,693 A		10/1998	Mays et al.	375/202
5,880,598 A		3/1999	Duong	326/41
5,996,032 A	*	11/1999	Baker	710/62
6,018,559 A		1/2000	Azegami et al.	377/79
6,144,327 A		11/2000	Distinti et al.	341/126
6,204,687 B1	*	3/2001	Schultz et al.	326/40
6,304,101 B1		10/2001	Nishihara	326/41
6,396,302 B2	*	5/2002	New et al.	326/38

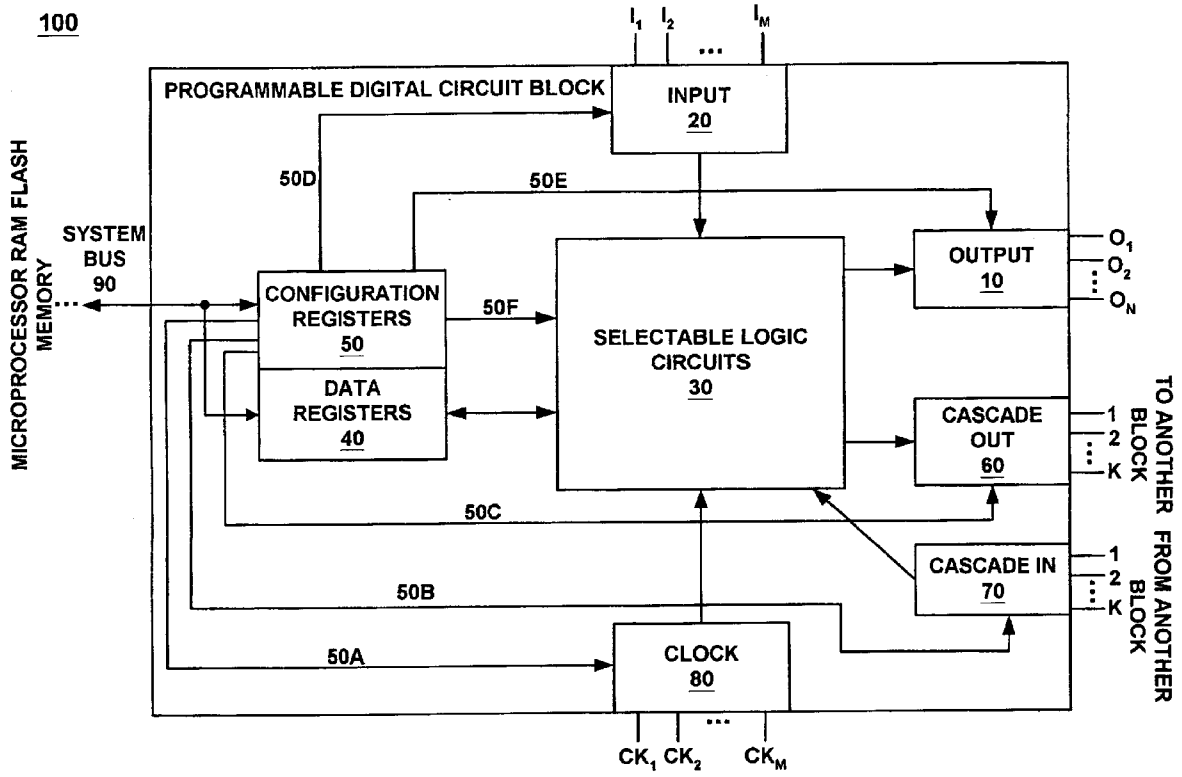
* cited by examiner

Primary Examiner—Vibol Tan
(74) *Attorney, Agent, or Firm*—Wagner, Murabito & Hao LLP

(57) **ABSTRACT**

A new digital configurable macro architecture is described. The digital configurable macro architecture is well suited for microcontroller or controller designs. In particular, the foundation of the digital configurable macro architecture is a programmable digital circuit block. In an embodiment, programmable digital circuit blocks are 8-bit circuit modules that can be programmed to perform any one of a variety of predetermined digital functions by changing the contents of a few registers therein, unlike a FPGA which is a generic device that can be programmed to perform any arbitrary digital function. Specifically, the circuit components of the programmable digital circuit block are designed for reuse in several of the predetermined digital functions such that to minimize the size of the programmable digital circuit block.

20 Claims, 9 Drawing Sheets



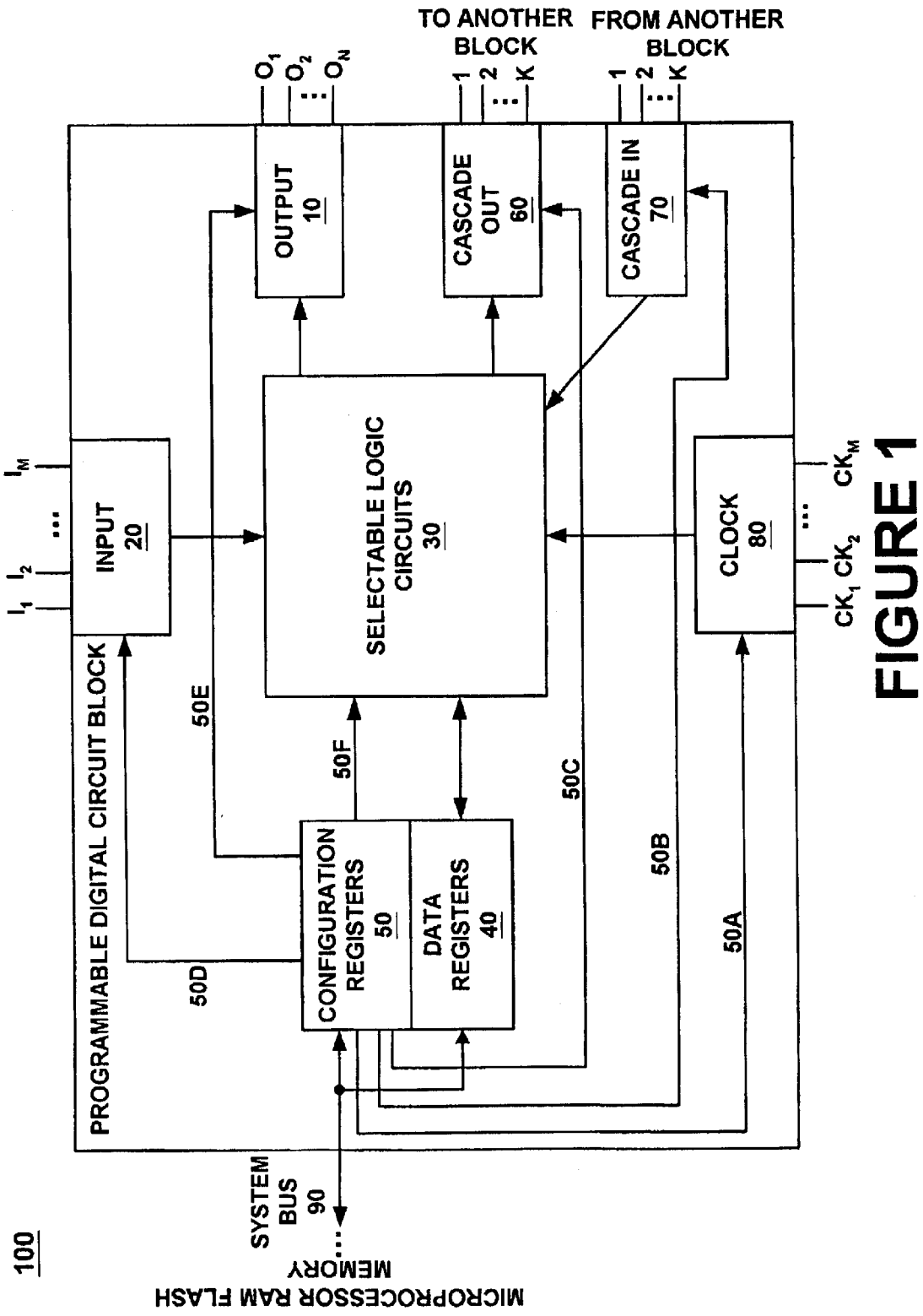

TO ANOTHER BLOCK
FROM ANOTHER BLOCK
O_1 O_2 ... O_N
1 2 ... K
1 2 ... K
OUTPUT 10
CASCADE OUT 60
CASCADE IN 70
SELECTABLE LOGIC CIRCUITS 30
INPUT 20
CLOCK 80
I_M ... I_2 I_1
CK_M ... CK_2 CK_1
PROGRAMMABLE DIGITAL CIRCUIT BLOCK
CONFIGURATION REGISTERS 50
DATA REGISTERS 40
50A
50B
50C
50D
50E
50F
SYSTEM BUS 90
MICROPROCESSOR RAM FLASH MEMORY
100

FIGURE 1

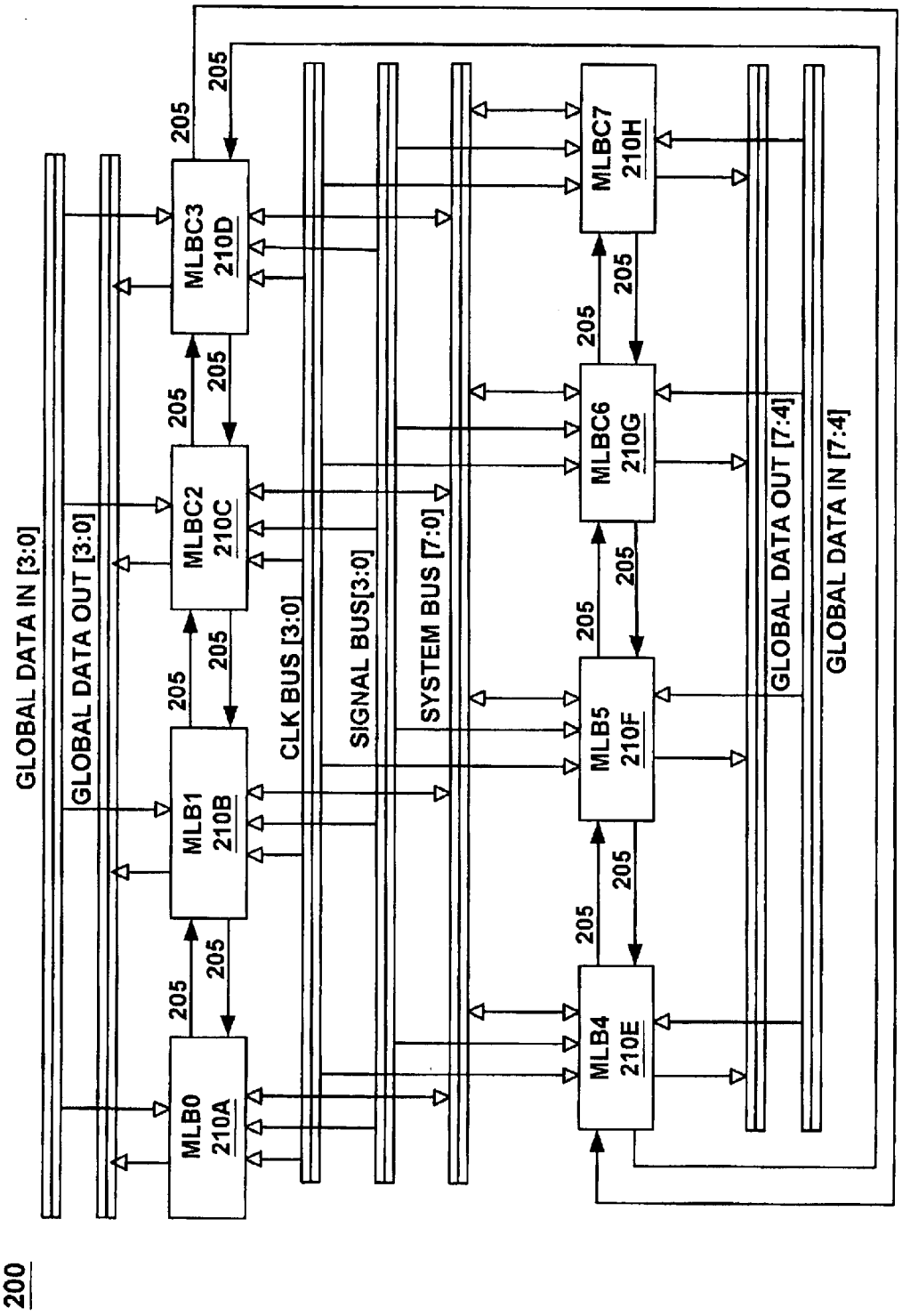

200
GLOBAL DATA IN [3:0]
GLOBAL DATA OUT [3:0]
CLK BUS [3:0]
SIGNAL BUS[3:0]
SYSTEM BUS [7:0]
GLOBAL DATA OUT [7:4]
GLOBAL DATA IN [7:4]
MLB0 210A
MLB1 210B
MLBC2 210C
MLBC3 210D
MLB4 210E
MLB5 210F
MLBC6 210G
MLBC7 210H
205

FIGURE 2

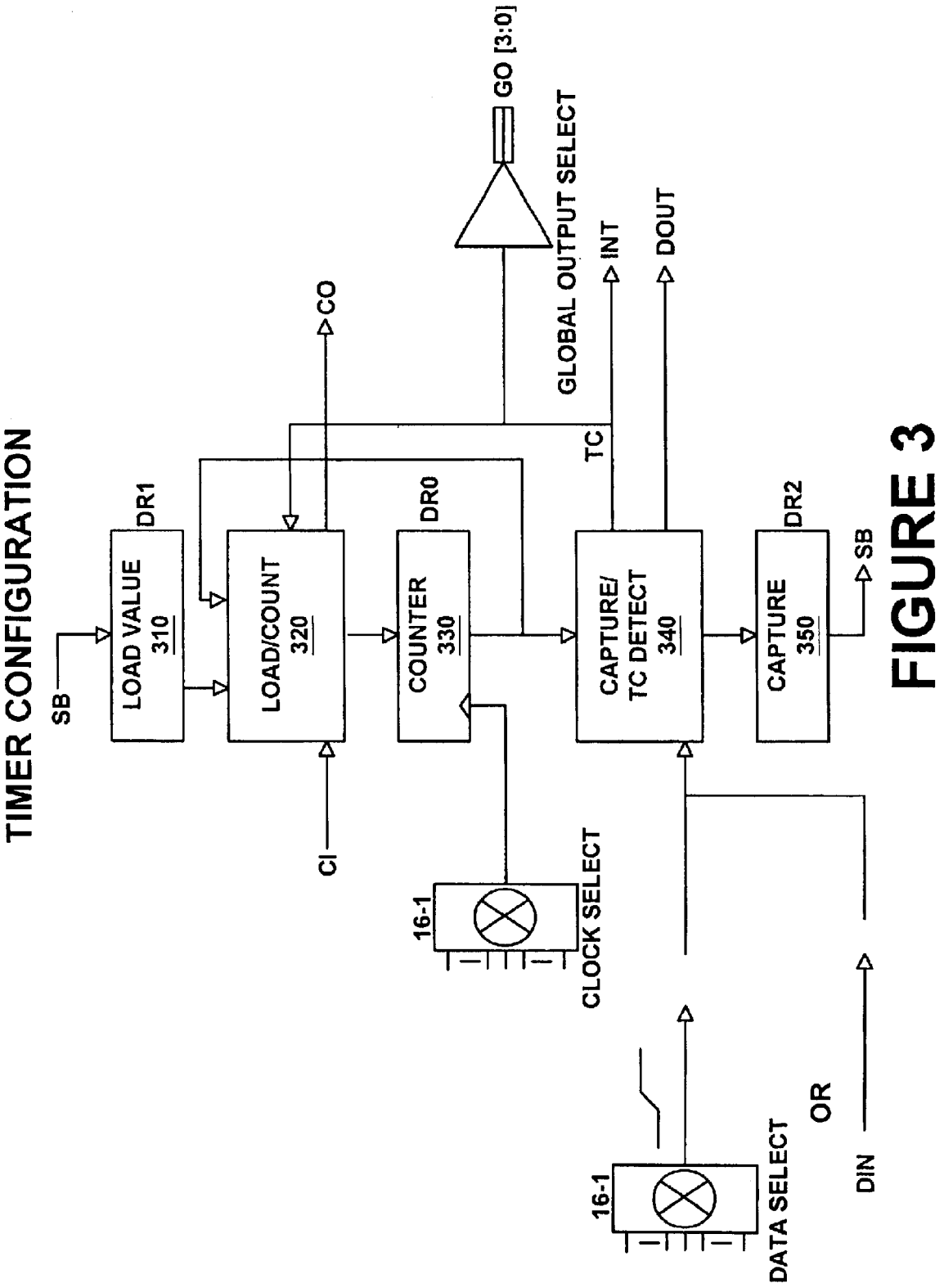

TIMER CONFIGURATION
SB
DR1
LOAD VALUE
310
LOAD/COUNT
320
CI
CO
DR0
COUNTER
330
16-1
CLOCK SELECT
GO [3:0]
GLOBAL OUTPUT SELECT
TC
INT
DOUT
CAPTURE/
TC DETECT
340
DR2
CAPTURE
350
SB
16-1
DATA SELECT
OR
DIN

FIGURE 3

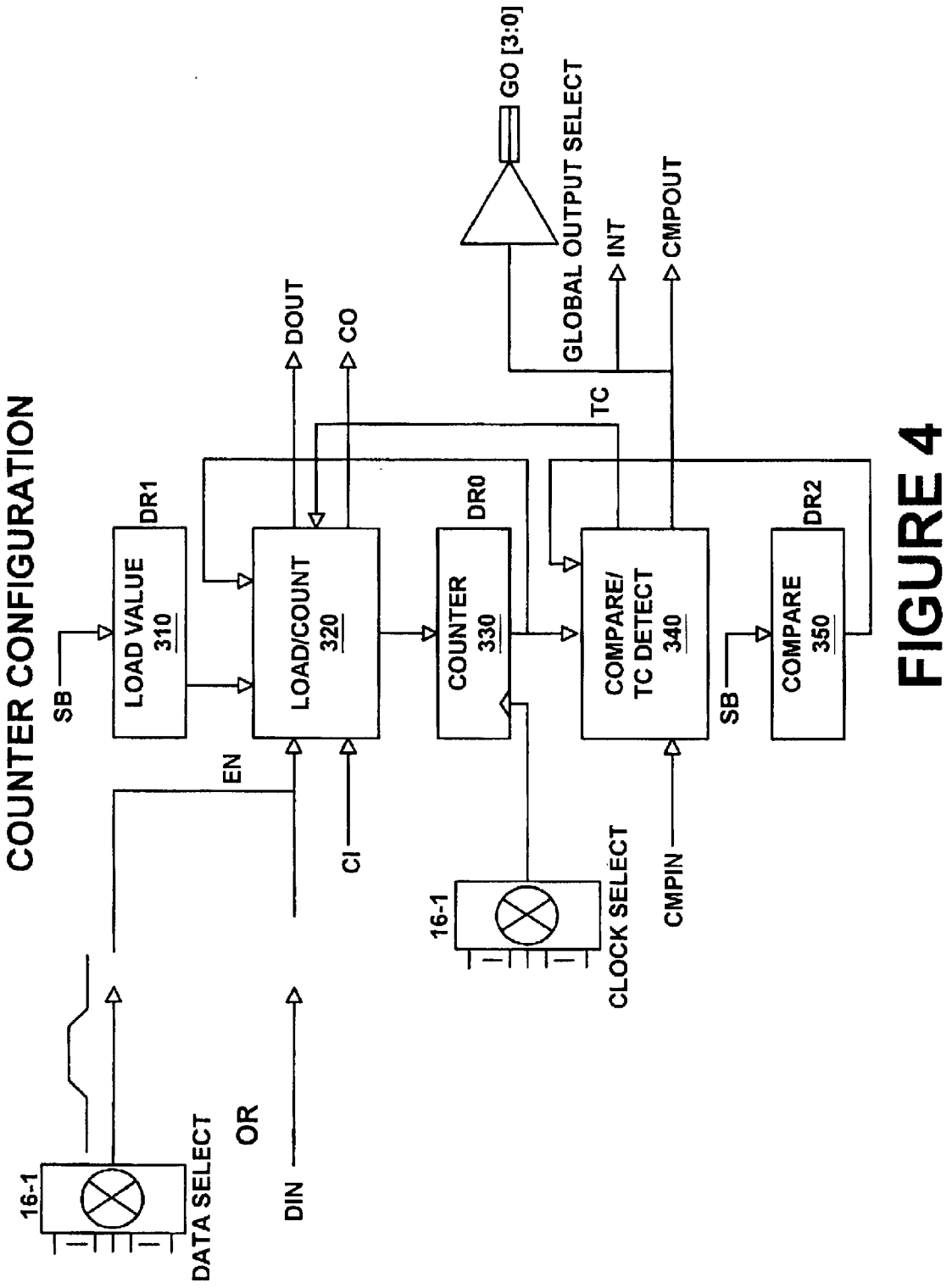

COUNTER CONFIGURATION
GO [3:0]
GLOBAL OUTPUT SELECT
INT
CMPOUT
DOUT
CO
TC
DR1
DR0
DR2
SB
LOAD VALUE 310
LOAD/COUNT 320
COUNTER 330
COMPARE/ TC DETECT 340
COMPARE 350
SB
EN
CI
16-1
CLOCK SELECT
CMPIN
16-1
DATA SELECT
OR
DIN

FIGURE 4

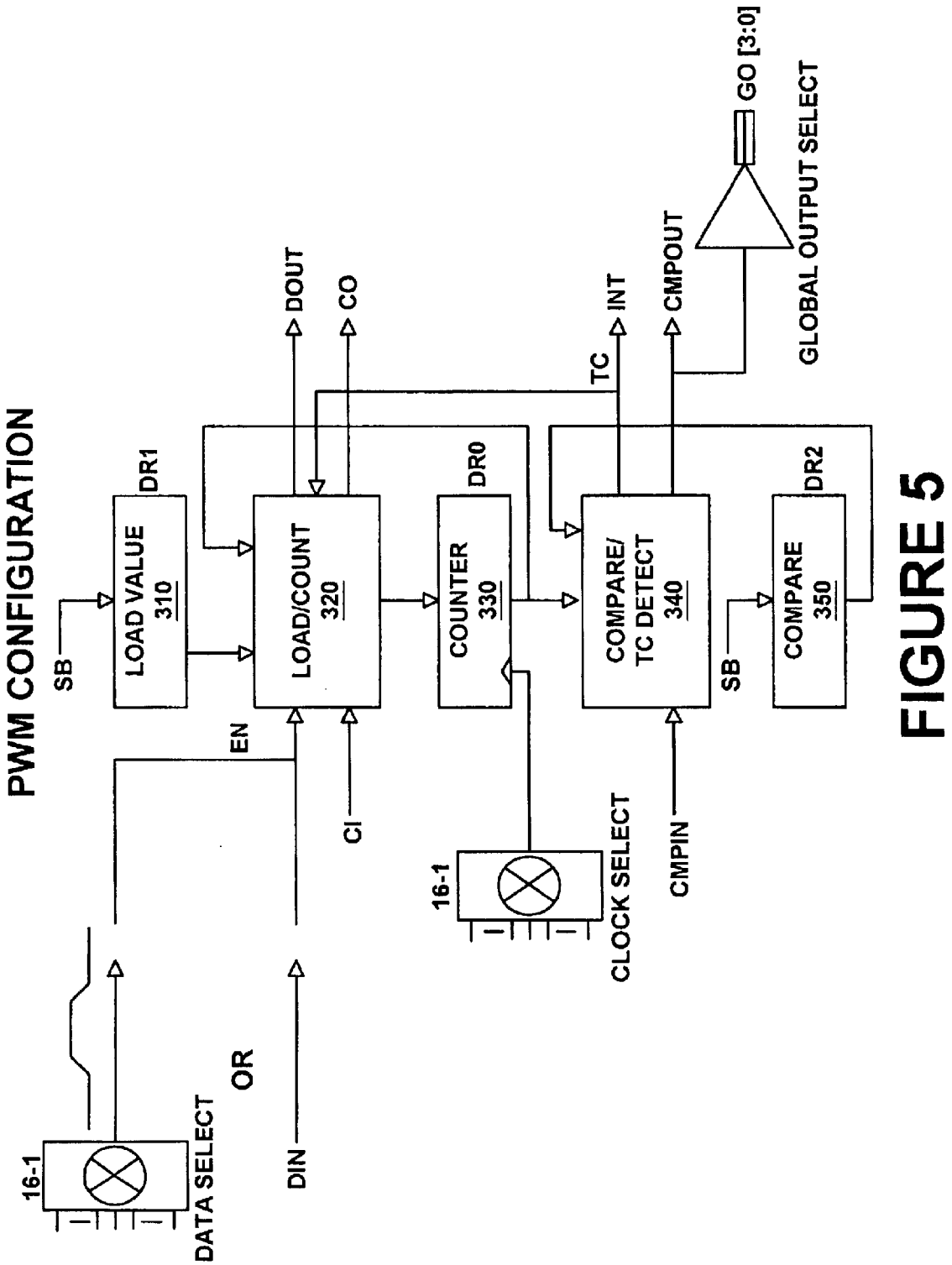

PWM CONFIGURATION
SB
LOAD VALUE 310
DR1
LOAD/COUNT 320
DOUT
CO
COUNTER 330
DR0
COMPARE/ TC DETECT 340
TC
INT
CMPOUT
GO [3:0]
GLOBAL OUTPUT SELECT
SB
COMPARE 350
DR2
EN
CI
16-1
CLOCK SELECT
CMPIN
16-1
DATA SELECT
OR
DIN

FIGURE 5

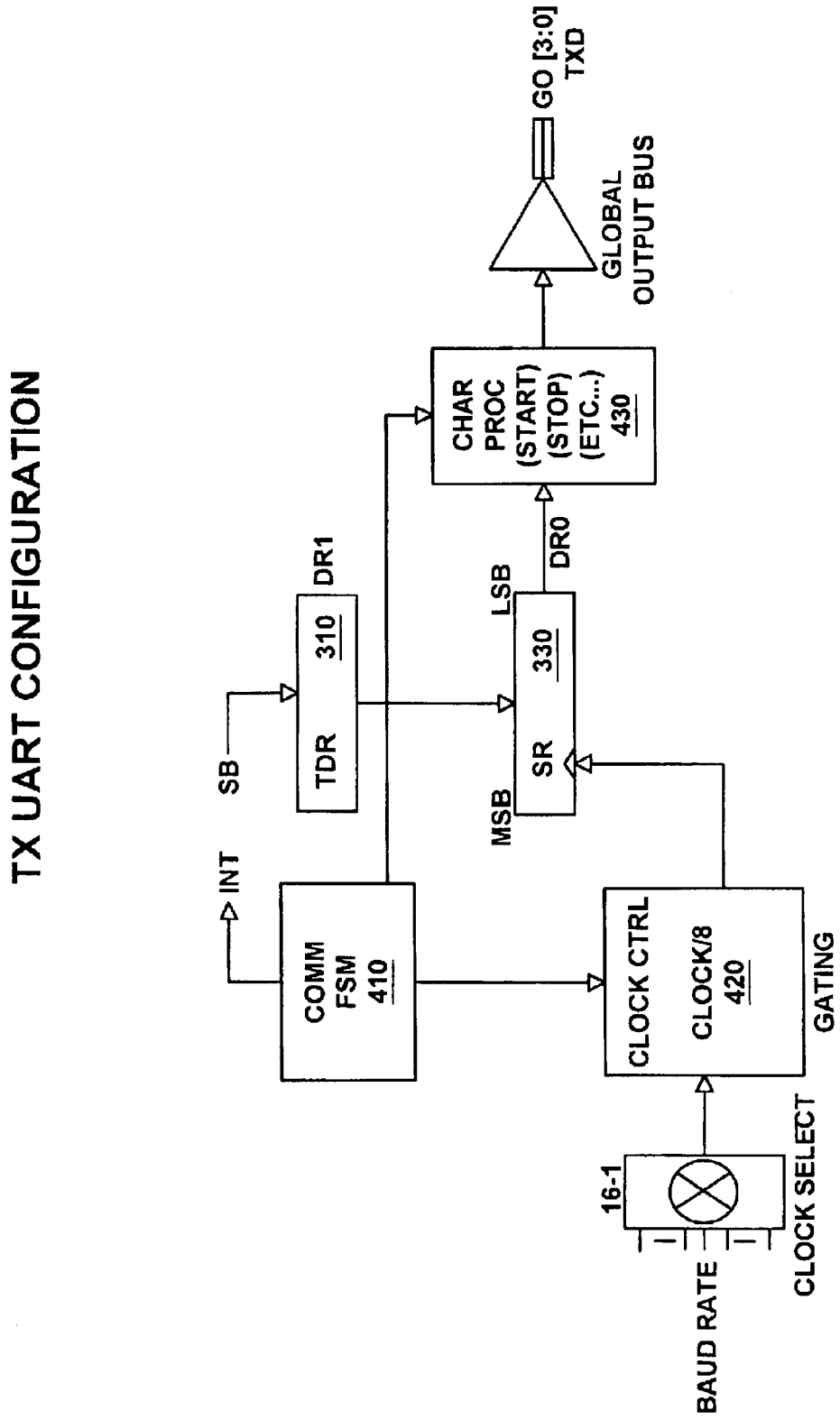

TX UART CONFIGURATION
GO [3:0]
TXD
GLOBAL
OUTPUT BUS
CHAR
PROC
(START)
(STOP)
(ETC...)
430
DR1
310
TDR
SB
LSB
330
SR
MSB
DR0
INT
COMM
FSM
410
CLOCK CTRL
CLOCK/8
420
GATING
16-1
BAUD RATE
CLOCK SELECT

FIGURE 6

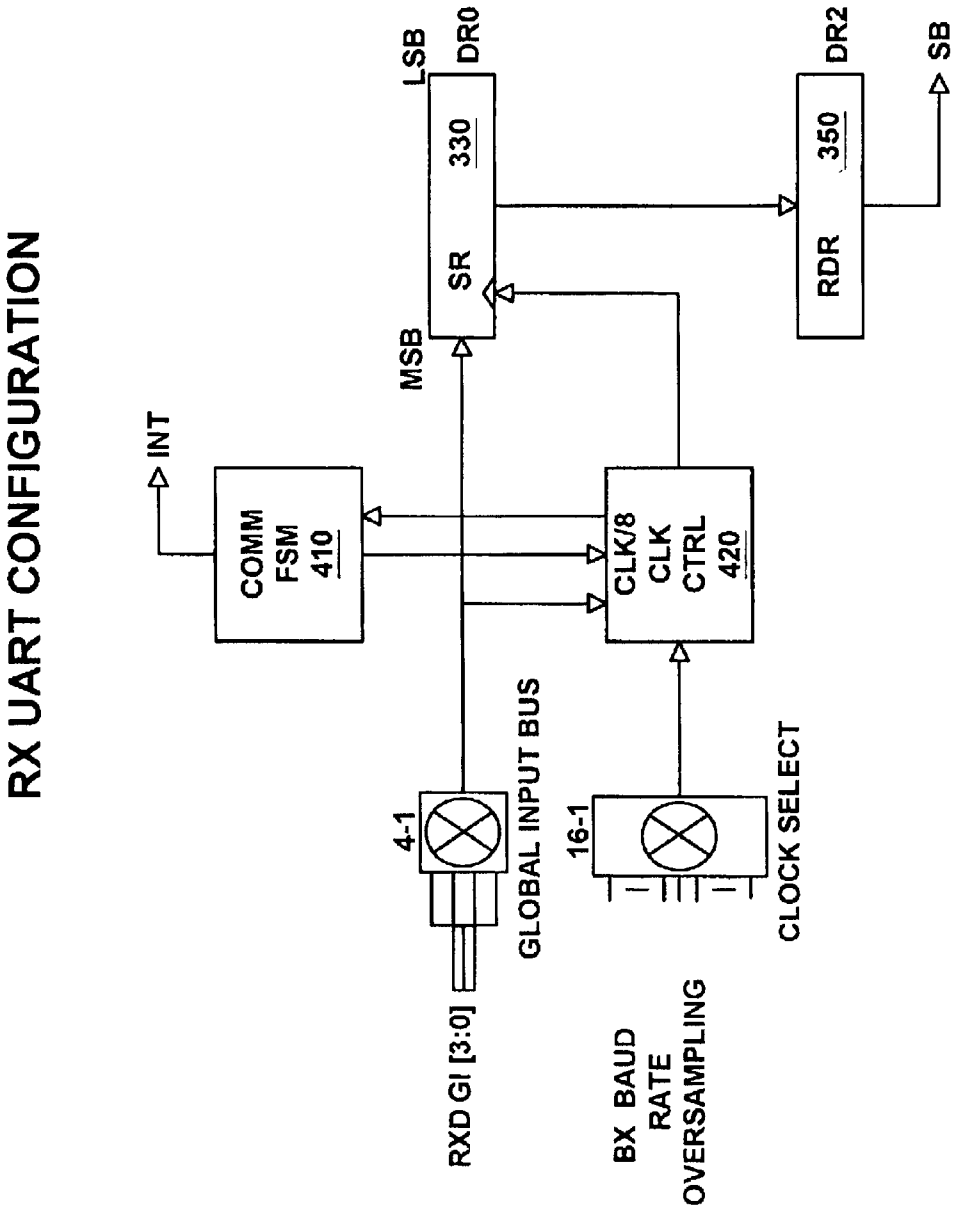

RX UART CONFIGURATION
LSB
DR0
330
SR
MSB
DR2
350
RDR
SB
INT
COMM FSM 410
CLK/8 CLK CTRL 420
4-1
16-1
GLOBAL INPUT BUS
CLOCK SELECT
RXD GI [3:0]
BX BAUD RATE OVERSAMPLING

FIGURE 7

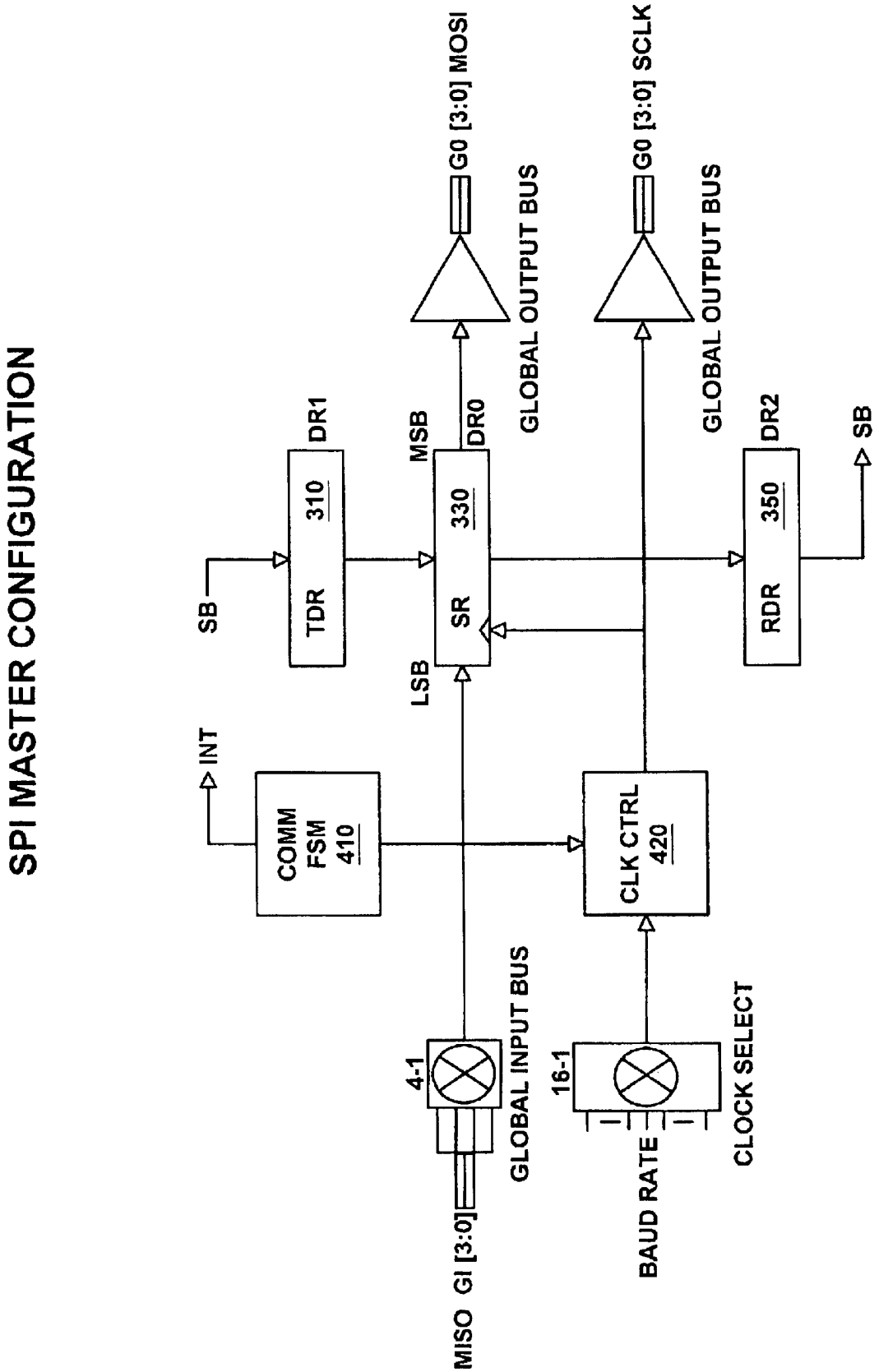

SPI MASTER CONFIGURATION
G0 [3:0] MOSI
G0 [3:0] SCLK
GLOBAL OUTPUT BUS
GLOBAL OUTPUT BUS
DR1
MSB
DR0
DR2
SB
SB
TDR 310
SR 330
RDR 350
LSB
INT
COMM FSM 410
CLK CTRL 420
4-1
16-1
GLOBAL INPUT BUS
CLOCK SELECT
MISO GI [3:0]
BAUD RATE

FIGURE 8

SPI SLAVE CONFIGURATION

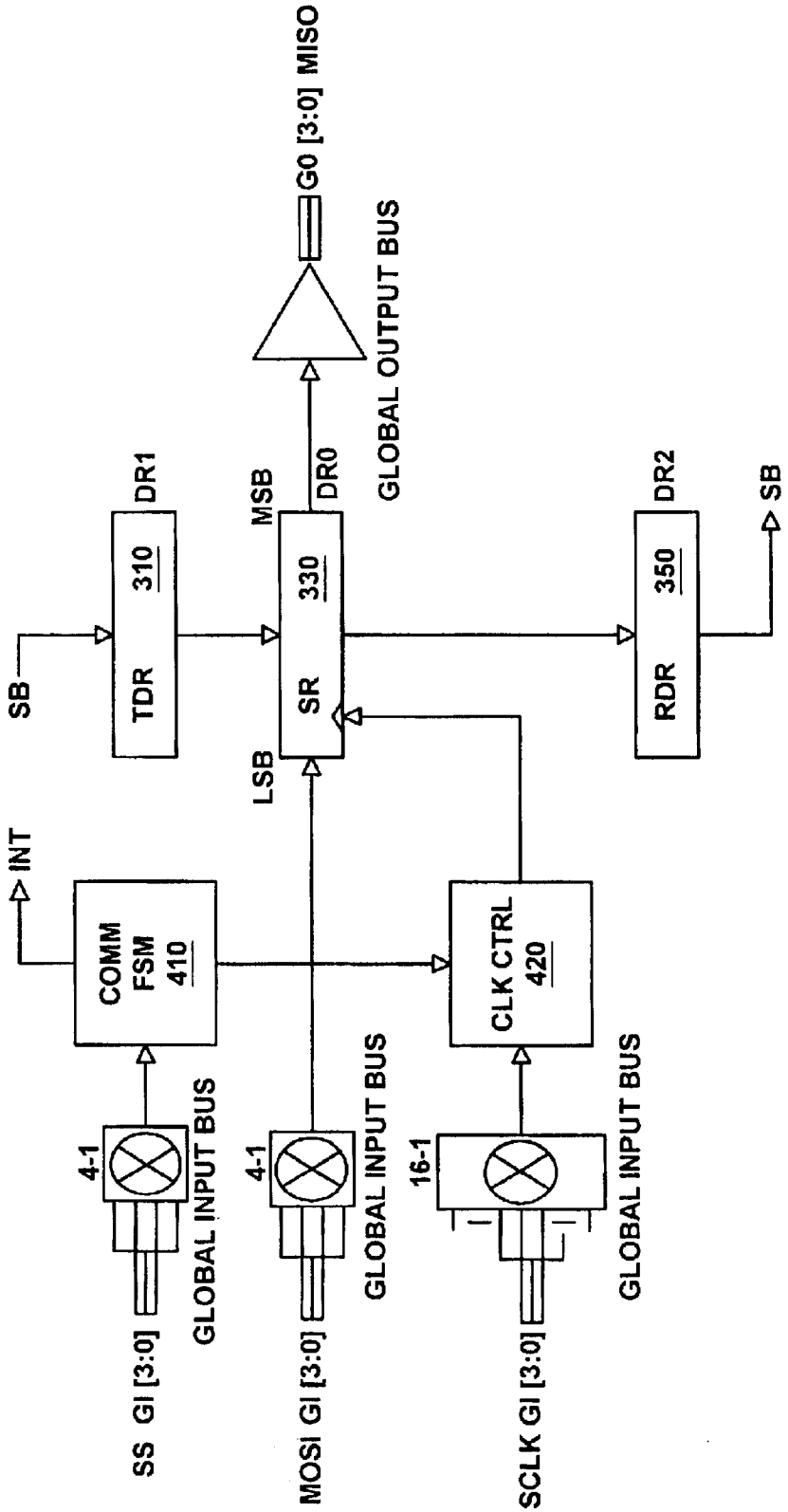


FIGURE 9

DIGITAL CONFIGURABLE MACRO ARCHITECTURE

RELATED U.S. APPLICATION

This patent application is a continuation of application Ser. No. 09/909,045, filed Jul. 18, 2001, now U.S. Pat. No. 6,507,214, entitled "DIGITAL CONFIGURABLE MACRO ARCHITECTURE", by Snyder, which claims priority to the copending provisional patent application, Serial No. 60/243,708, entitled "Advanced Programmable Microcontroller Device," with filing date Oct. 26, 2000, and assigned to the assignee of the present application, which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention generally relates to programmable digital circuits. More particularly, the present invention relates to the field of programmable digital circuits which are dynamically configurable to any one of various digital functions by changing the contents of configuration registers.

2. Related Art

Microcontrollers or controllers have been utilized in various applications for many years. Primarily, microcontrollers are used in control-oriented applications that are interrupt-driven, sensing and controlling external events. Microcontrollers are frequently found in: appliances (e.g., microwave oven, refrigerator, television, VCR, stereo), computers and computer equipment (e.g., laser printers, modems, disk drives), automobiles (e.g., engine control, diagnostics, climate control), environmental control (e.g., greenhouse, factory, home), aerospace, and thousands of other uses.

The Field Programmable Gate Array (FPGA) has become very popular in recent years, even being utilized in several microcontroller applications. One reason for its popularity is the shortage in design cycle time that may be achieved by using programmable devices. Typically, FPGAs offer the highest logic capacity. FPGAs can be programmed to realize different digital functions. In particular, many FPGAs have programmable look-up tables to realize different digital functions. Typically, a FPGA contains from a few to tens of thousands of programmable logic blocks and an even greater number of flip-flops, each programmable logic block having a look-up table, multiplexors, and flip-flops. Most FPGAs do not provide 100% interconnect between programmable logic blocks.

However, FPGAs are highly inefficient with respect to chip area, increasing their cost. Typically, less than half of the logic resources in the FPGA are used to realize a digital function. Moreover, FPGAs need to have their look-up tables re-programmed in order to enable them to implement a new digital function, which is a time consuming task.

FPGAs are not ideally suited for microcontroller applications. Microcontroller applications are very cost-sensitive. A FPGA is not able to realize the number of digital functions that are demanded by today's microcontroller applications within these strict cost constraints.

SUMMARY OF THE INVENTION

A new digital configurable macro architecture is described. The digital configurable macro architecture is

well suited for microcontroller or controller designs. In particular, the foundation of the digital configurable macro architecture is a programmable digital circuit block. In an embodiment, programmable digital circuit blocks are 8-bit circuit modules that can be programmed to perform any one of a variety of predetermined digital functions by changing the contents of a few registers therein, unlike a FPGA which is a generic device that can be programmed to perform any arbitrary digital function. Specifically, the circuit components of the programmable digital circuit block are designed for reuse in several of the predetermined digital functions such that to minimize the size of the programmable digital circuit block. The programmable digital circuit blocks can be configured, for example, as timers, counters, serial communication ports, cyclic redundancy generators/checkers (CRC), or pseudo random sequence generators (PRS). The user selects the digital function that is needed and configures the programmable digital circuit block accordingly.

The programmable digital circuit blocks can be configured to coupled in series or in parallel to handle more complex digital functions. For example, a 24-bit timer can be designed by coupling three 8-bit programmable digital circuit blocks that have been individually configured as 8-bit timers. Additionally, a first programmable digital circuit block that is configured as a CRC generator can feed a second programmable digital circuit block that is configured as a serial output communication port.

More importantly, the configuration of the programmable digital circuit block is determined by its small number of configuration registers. This provides much flexibility. In particular, the configuration of the programmable digital circuit block is fast and easy since changes in configuration are accomplished by changing the contents of the configuration registers, whereas the contents are generally a small number of configuration data bits. Thus, the programmable digital circuit block is dynamically configurable from one predetermined digital function to another predetermined digital function for real-time processing.

These and other advantages of the present invention will no doubt become apparent to those of ordinary skill in the art after having read the following detailed description of the preferred embodiments which are illustrated in the drawing figures.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of this specification, illustrate embodiments of the invention and, together with the description, serve to explain the principles of the present invention.

FIG. **1** illustrates a programmable digital circuit block in accordance with an embodiment of the present invention.

FIG. **2** illustrates a block diagram of an exemplary programmable digital device having a plurality of programmable digital circuit blocks in accordance with an embodiment of the present invention.

FIG. **3** illustrates a block diagram of a timer configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

FIG. **4** illustrates a block diagram of a counter configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

FIG. **5** illustrates a block diagram of a pulse width modulator (PWM) configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

FIG. **6** illustrates a block diagram of a UART transmitter configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

FIG. **7** illustrates a block diagram of a UART receiver configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

FIG. **8** illustrates a block diagram of a SPI Master configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

FIG. **9** illustrates a block diagram of a SPI Slave configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

The drawings referred to in this description should not be understood as being drawn to scale except if specifically noted.

DETAILED DESCRIPTION OF THE INVENTION

Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings. While the invention will be described in conjunction with the preferred embodiments, it will be understood that they are not intended to limit the invention to these embodiments. On the contrary, the invention is intended to cover alternatives, modifications and equivalents, which may be included within the spirit and scope of the invention as defined by the appended claims. Furthermore, in the following detailed description of the present invention, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be recognized by one of ordinary skill in the art that the present invention may be practiced without these specific details. In other instances, well known methods, procedures, components, and circuits have not been described in detail as not to unnecessarily obscure aspects of the present invention.

FIG. **1** illustrates a programmable digital circuit block **100** in accordance with an embodiment of the present invention. The programmable digital circuit block **100** is the foundation of a new digital configurable macro architecture of the present invention. The digital configurable macro architecture is well suited for microcontroller or controller designs.

The design of the programmable digital circuit block **100** in the digital configurable macro architecture was developed after examining and studying conventional microcontrollers to determine the types of digital functions that were implemented within various conventional microcontrollers. It was discovered that there were not very many different types of digital functions demanded in microcontroller applications. Furthermore, it was determined that these different types of digital functions had many circuit components in common. Moreover, it was determined that the digital functions were generally implemented as 8-bit or multiples of 8-bits because their length was generally based on the length of standard buses. This led to the development of the programmable digital circuit blocks **100**, the building block of the digital configurable macro architecture.

In an embodiment, the programmable digital circuit block **100** is an 8-bit circuit module that can be programmed to perform any one of a variety of predetermined digital functions (which are useful in microcontroller applications) by changing the contents of a few configuration registers **50** therein, unlike a FPGA which is a generic device that can be programmed to perform any arbitrary digital function. Specifically, the circuit components of the programmable digital circuit block **100** are designed for reuse in several of the predetermined digital functions such that to minimize the size of the programmable digital circuit block **100**. Hence, the programmable digital circuit block **100** is highly efficient in terms of die area. In an embodiment, the programmable digital circuit block **100** can be configured as a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), a dead zone delay, a UART (universal asynchronous receiver-transmitter) transmitter, a UART (universal asynchronous receiver-transmitter) receiver, a SPI (serial peripheral interface) Master, or a SPI (serial peripheral interface) Slave.

In another embodiment, the programmable digital circuit block **100** can be configured as a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), or a dead zone delay, whereas the digital communication functions (e.g., UART and SPI) are eliminated to further reduce the size of the programmable digital circuit block **100**. In particular, the user selects the digital function that is needed and configures the programmable digital circuit block **100** accordingly. It should be understood that the programmable digital circuit block **100** can be designed to implement other digital functions.

A design can have an array of programmable digital circuit blocks **100** which can be configured to coupled together in series or in parallel to handle more complex digital functions or to increase precision. For example, a 24-bit timer can be designed by coupling three 8-bit programmable digital circuit blocks **100** that have been individually configured as 8-bit timers. Similarly, the 8-bit timer can be extended to 16- or 32-bit digital functions by coupling multiple programmable digital circuit blocks **100** together. Additionally, a first programmable digital circuit block that is configured as a CRC generator can feed a second programmable digital circuit block that is configured as a serial output communication port, reducing device programming and increasing performance.

More importantly, the configuration of the programmable digital circuit block **100** is determined by its configuration registers **50**. The programmable digital circuit block **100** generally has one or more configuration registers **50**. This provides much flexibility. In particular, the configuration of the programmable digital circuit block **100** is fast and easy to configure and re-configure since changes in configuration are accomplished by changing the contents of the configuration registers **50**, whereas the contents are generally a small number of configuration data bits. Thus, the programmable digital circuit block **100** is dynamically configurable from one predetermined digital function to another prede-

termined digital function for real-time processing. In contrast, FPGAs need to have their look-up tables re-programmed in order to have them implement a new digital function, a time-consuming task that is not done in real-time processing.

Referring to FIG. **1**, in an embodiment the programmable digital circuit block **100** includes one or more configuration registers **50**, one or more data registers **40**, a plurality of selectable logic circuits **30**, one or more configurable inputs **20**, one or more configurable outputs **10**, one or more cascade outputs **60**, one or more cascade inputs **70**, a clock input **80**, and a system input **90**. It should be understood that the programmable digital circuit block **100** can have other designs Including lengths other than 8-bits.

The configuration registers **50** are programmed via the system bus **90**. Any device, such as a microprocessor using data stored in a RAM or flash memory, can program (or write to) the configuration registers. The configuration registers **50** receive and store a plurality of configuration data corresponding to any one of the plurality of predetermined digital function described above. The programmed configuration registers **50** configure the programmable digital circuit block **100** to perform any one of the predetermined digital functions based on the configuration data. Moreover, the configuration registers **50** can be dynamically programmed with the configuration data for real-time processing. In addition, the configuration data includes (1) bits for indicating one of the predetermined digital functions and configuring the selectable logic circuits **30**, (2) bits for configuring and selecting the configurable inputs **20** and the configurable outputs **10** and the clock input **80**, (3) bits for indicating the mode of the predetermined digital function (e.g., parity, no parity, etc.), (4) bits for indicating the length of the predetermine digital function if several programmable digital circuit block **100** are coupled together (e.g., 8-bit, 16-bit, 24-bit, etc.), and (5) bits for indicating and configuring the interface between adjacent programmable digital circuit blocks **100** that are coupled together (e.g., configuring and selecting the cascade inputs **70** and the cascade outputs **60** for serial or parallel interfacing).

In general, the number of bits in the configuration data is sufficiently small to enable the configuration registers **50** to be programmed on-the-fly so that the programmable digital circuit block **100** can be dynamically configured and interfaced. Thus, the programmable digital circuit blocks **100** can be configured as a timer for a first length of time, re-configured as a counter for a second length of time, re-configured as a PWM for a third length of time, and so on, for real-time processing. For example, it is possible for a single register write to configure the programmable digital circuit block **100** from a timer to a PWM or to a counter or to a CRC generator or etc.

The connections **50**A–**50**F between the configuration registers **50** and other components of the programmable digital circuit block **100** enable the configuration registers **50** to properly configure the programmable digital circuit block **100** to any one of the predetermined digital functions and to properly interface the programmable digital circuit block **100** with other programmable digital circuit blocks in series or in parallel.

Continuing with FIG. **1**, the selectable logic circuits **30** are tailored such that they have a minimum set of circuit resources that can be programmed by the configuration registers **50** to implement any one of a variety of predetermined digital functions, unlike the FPGA where a substantial amount of circuit resources may remain unused. In particular, the design and structure of the selectable logic circuits **30** are dependent on the predetermined digital functions such that to minimize the size of the programmable digital circuit block **100**. The fixed number of digital functions for the programmable digital circuit block **100** substantially influences the design of the programmable digital circuit block **100**, providing cost savings and improving performance. The configuration registers **50** configure and select any of the selectable logic circuits **30** to perform one of the predetermined digital functions based on the configuration data. More importantly, the selectable logic circuits **30** are reused in several of the predetermined digital functions as will be illustrated below, ensuring the size efficiency of the programmable digital circuit block **100**. In an embodiment, the selectable logic circuits **30** include a plurality of logic gates.

Moreover, the selectable logic circuits **30** realize any one of the variety of predetermined digital functions by using the data registers **40** to receive data, load data, capture data, etc. Thus, the data registers **40** are also reused in several of the predetermined digital functions as will be illustrated below.

Again referencing FIG. **1**, the cascade outputs **60** and the cascade inputs **70** are selected and configured according to the configuration data. The cascade outputs **60** allow the programmable digital circuit block **100** to output signals for directly interfacing with adjacent or neighboring programmable digital circuit blocks. The cascade inputs **70** allow the adjacent or neighboring programmable digital circuit blocks to send signals that directly interface and are received by the programmable digital circuit block **100**. Specifically, the cascade outputs **60** and the cascade inputs **70** enable multiple programmable digital circuit blocks to seamlessly interface to. handle more complex digital functions or to increase precision as described above (e.g., 32-bit timer, CRC generator and SPI Master, 24-bit counter, etc.).

FIG. **2** illustrates a block diagram of an exemplary programmable digital device **200** having a plurality of programmable digital circuit blocks **210**A–**210**H in accordance with an embodiment of the present invention. The plurality of programmable digital circuit blocks **210**A–**210**H includes a first group and a second group. The first group includes the programmable digital circuit blocks **210**A–**210**B and **210**E–**210**F. Moreover, each programmable digital circuit block of the first group can be configured as a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), or a dead zone delay. The second group includes the programmable digital circuit blocks **210**C–**210**D and **210**G–**210**H. Moreover, each programmable digital circuit block of the second group can be configured as a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generatbr/checker (CRC), a pseudo random sequence generator (PRS), a dead zone delay, a UART (universal asynchronous receiver-transmitter) transmitter, a UART (universal asynchronous receiver-transmitter) receiver, a SPI (serial peripheral interface) Master, or a SPI (serial peripheral interface) Slave.

As illustrated in FIG. **2**, adjacent or neighboring programmable digital circuit blocks are interfaced via cascade lines **205** (input or output) as described above. The cascade lines **205** enable the programmable digital circuit blocks **210**A–**210**H to seamlessly interface to handle more complex digital functions or to increase precision. For example, a 32-bit counter can be designed by coupling four 8-bit programmable digital circuit blocks that have been individually configured as 8-bit counters. Similarly, the 8-bit counter can be extended to 16- or 24-bit digital functions by coupling multiple programmable digital circuit blocks together. Additionally, a first programmable digital circuit block that is configured as a CRC generator can feed a second programmable digital circuit block that is configured as a serial output communication port, reducing device programming and increasing performance.

Moreover, the exemplary programmable digital device **200** includes a signal bus for digitized analog signals, a clock bus, a system bus for programming the programmable digital circuit blocks **210**A–**210**H, and a plurality of global data buses for transmitting data to/from the programmable digital circuit blocks **210**A–**210**H.

FIG. **3** illustrates a block diagram of a timer configuration of a programmable digital circuit block in accordance with an embodiment of the present invention. FIG. **4** illustrates a block diagram of a counter configuration of a programmable digital circuit block in accordance with an embodiment of the present invention. FIG. **5** illustrates a block diagram of a pulse width modulator (PWM) configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

As illustrated in FIGS. **3**–**5**, the selectable logic circuits **320** and **340** are reused for the timer, counter, and PWM configurations. Moreover, the first data register **310**, the second data register **330**, and the third data register **350** of the programmable digital circuit block are reused for the timer, counter, and PWM configurations. In essence, the configuration data loaded onto the configuration registers determines how the data registers **310**, **330**, and **350** are to be used, what operation is to be performed on the data by the selectable logic circuits **320** and **340**, where the input data is selected from (e.g., system bus (SB), signal bus, global bus, etc.), where the output data is transmitted, what clock signal is to be used, what are the cascade inputs (e.g., DIN, Cl, etc.) from other programmable digital circuit blocks, what are the cascade outputs (e.g., DOUT, CO, etc.) to other programmable digital circuit blocks, when to generate an interrupt (INT), and what is the data flow within the programmable digital circuit block so that the programmable digital circuit block can properly perform any one of the predetermined digital functions.

FIG. **6** illustrates a block diagram of a UART transmitter configuration of a programmable digital circuit block in accordance with an embodiment of the present invention. FIG. **7** illustrates a block diagram of a UART receiver configuration of a programmable digital circuit block in accordance with an embodiment of the present invention. FIG. **8** illustrates a block diagram of a SPI Master configuration of a programmable digital circuit block in accordance with an embodiment of the present invention. FIG. **9** illustrates a block diagram of a SPI Slave configuration of a programmable digital circuit block in accordance with an embodiment of the present invention.

As illustrated in FIGS. **6**–**9**, the selectable logic circuits **410** and **420** are reused for the UART transmitter, the UART receiver, the SPI Master, and the SPI Slave configurations. Moreover, the first data register **310**, the second data register **330**, and the third data register **350** of the programmable digital circuit block are reused in several of the UART transmitter, the UART receiver, the SPI Master, and the SPI Slave configurations. However, the selectable logic circuit **430** is used in the UART transmitter configuration of FIG. **6** since the UART protocol requires that particular protocol bits (e.g., start bits, stop bits, etc.) to be generated by the UART transmitter.

FIGS. **3**–**9** illustrate that the programmable digital circuit block can be configured fast and easily. Furthermore, FIGS. **3**–**9** illustrate that the programmable digital circuit block is highly efficient in terms of die area.

The foregoing descriptions of specific embodiments of the present invention have been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise forms disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application, to thereby enable others skilled in the art to best utilize the invention and various embodiments with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto and their equivalents.

What is claimed is:

1. A programmable digital device comprising:

a programmable digital circuit block that is configurable to perform any one of a plurality of predetermined digital functions upon being configured with a single register write operation.

2. The programmable digital device as recited in claim **1** wherein said programmable digital circuit block is configurable into a serial arrangement.

3. The programmable digital device as recited in claim **1** wherein said programmable digital circuit block is configurable into a parallel arrangement.

4. The programmable digital device as recited in claim **1** wherein said predetermined digital functions include a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), a dead zone delay, a UART transmitter, a UART receiver, a SPI Master, and a SPI Slave.

5. The programmable digital device as recited in claim **1** wherein said predetermined digital functions include a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), and a dead zone delay.

6. The programmable digital device as recited in claim **1** wherein said predetermined digital functions are 8-bit predetermined digital functions.

7. The programmable digital device as recited in claim **1** wherein said programmable digital circuit block further comprises:

a configuration register for receiving and storing a plurality of configuration data corresponding to any of said plurality of predetermined digital functions; and

a plurality of selectable logic circuits which perform any of said plurality of predetermined digital functions, wherein said predetermined digital functions determine size and arrangement of said selectable logic circuits.

8. A programmable digital device comprising:

an array of programmable digital circuit blocks, each programmable digital circuit block is configurable to perform any one of a plurality of predetermined digital functions upon being configured with a single register write operation.

9. The programmable digital device as recited in claim **8** wherein each programmable digital circuit block is configurable into a serial arrangement.

10. The programmable digital device as recited in claim **8** wherein each programmable digital circuit block is configurable into a parallel arrangement.

11. The programmable digital device as recited in claim **8** wherein said predetermined digital functions include a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), a dead zone delay, a UART transmitter, a UART receiver, a SPI Master, and a SPI Slave.

12. The programmable digital device as recited in claim **8** wherein said predetermined digital functions include a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), and a dead zone delay.

13. The programmable digital device as recited in claim **8** wherein said predetermined digital functions are 8-bit predetermined digital functions.

14. The programmable digital device as recited in claim **8** wherein each programmable digital circuit block further comprises:

a configuration register for receiving and storing a plurality of configuration data corresponding to any of said plurality of predetermined digital functions; and

a plurality of selectable logic circuits which perform any of said plurality of predetermined digital functions, wherein said predetermined digital functions determine size and arrangement of said selectable logic circuits.

15. A method of configuring a programmable digital circuit block, comprising:

selecting one a plurality of predetermined digital functions;

performing a single register write operation to provide to said programmable digital circuit block a plurality of configuration data corresponding to said selected one of said predetermined digital functions; and

configuring said programmable digital circuit block using said configuration data.

16. The method as recited in claim **15** wherein said configuring includes configuring said programmable digital circuit block into a serial arrangement.

17. The method as recited in claim **15** wherein said configuring includes configuring said programmable digital circuit block into a parallel arrangement.

18. The method as recited in claim **15** wherein said predetermined digital functions include a timer, a counter, a pulse width modulator (PWM) a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), a dead zone delay, a UART transmitter, a UART receiver, a SPI Master, and a SPI Slave.

19. The method as recited in claim **15** wherein said predetermined digital functions include a timer, a counter, a pulse width modulator (PWM), a cyclic redundancy generator/checker (CRC), a pseudo random sequence generator (PRS), and a dead zone delay.

20. The method as recited in claim **15** wherein said predetermined digital functions are 8-bit predetermined digital functions.

* * * * *

EXHIBIT G

US007572727B1

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

(10) **Patent No.: US 7,572,727 B1**
(45) **Date of Patent: Aug. 11, 2009**

(54) **SEMICONDUCTOR FORMATION METHOD THAT UTILIZES MULTIPLE ETCH STOP LAYERS**

(75) Inventors: **Wenmei Li**, Sunnyvale, CA (US); **Angela T. Hui**, Fremont, CA (US); **Dawn Hopper**, San Jose, CA (US); **Kouros Ghandehari**, Fremont, CA (US)

(73) Assignee: **Spansion LLC**, Sunnyvale, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 147 days.

(21) Appl. No.: **10/934,828**

(22) Filed: **Sep. 2, 2004**

(51) **Int. Cl.**
H01L 21/4763 (2006.01)

(52) **U.S. Cl.** **438/624**; 438/638; 438/639; 438/740

(58) **Field of Classification Search** 438/639
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,912,188	A *	6/1999	Gardner et al.	438/740
6,078,073	A	6/2000	Habu et al.	
6,083,822	A *	7/2000	Lee	438/624
6,365,504	B1 *	4/2002	Chien et al.	438/624
6,441,418	B1 *	8/2002	Shields et al.	257/296
2002/0008323	A1 *	1/2002	Watanabe et al.	257/758
2005/0003656	A1 *	1/2005	Chung	438/639

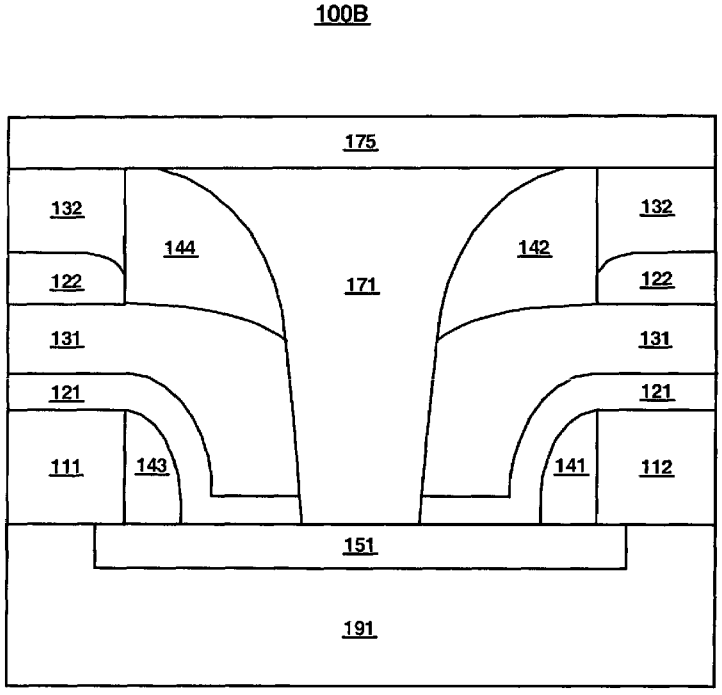
* cited by examiner

Primary Examiner—Stephen W Smoot

(57) **ABSTRACT**

The present invention is a semiconductor contact formation system and method. Contact insulation regions are formed with multiple etch stop sublayers that facilitate formation of contacts. This contact formation process provides relatively small substrate connections while addressing critical lithographic printing limitation concerns in forming contact holes with small dimensions. In one embodiment, a multiple etch stop insulation layer comprising multiple etch stop layers is deposited. A contact region is formed in the multiple etch stop insulation layer by selectively removing (e.g., etching) some of the multiple etch stop insulation layer. In one embodiment, a larger portion of the multiple etch stop insulation layer is removed close to the metal layer and a smaller portion is removed closer to the substrate. The different contact region widths are achieved by performing multiple etching processes controlled by the multiple etch stop layers in the multiple etch stop insulation layer and spacer formation to shrink contact size at a bottom portion. Electrical conducting material (e.g., tungsten) is deposited in the contact region.

15 Claims, 11 Drawing Sheets



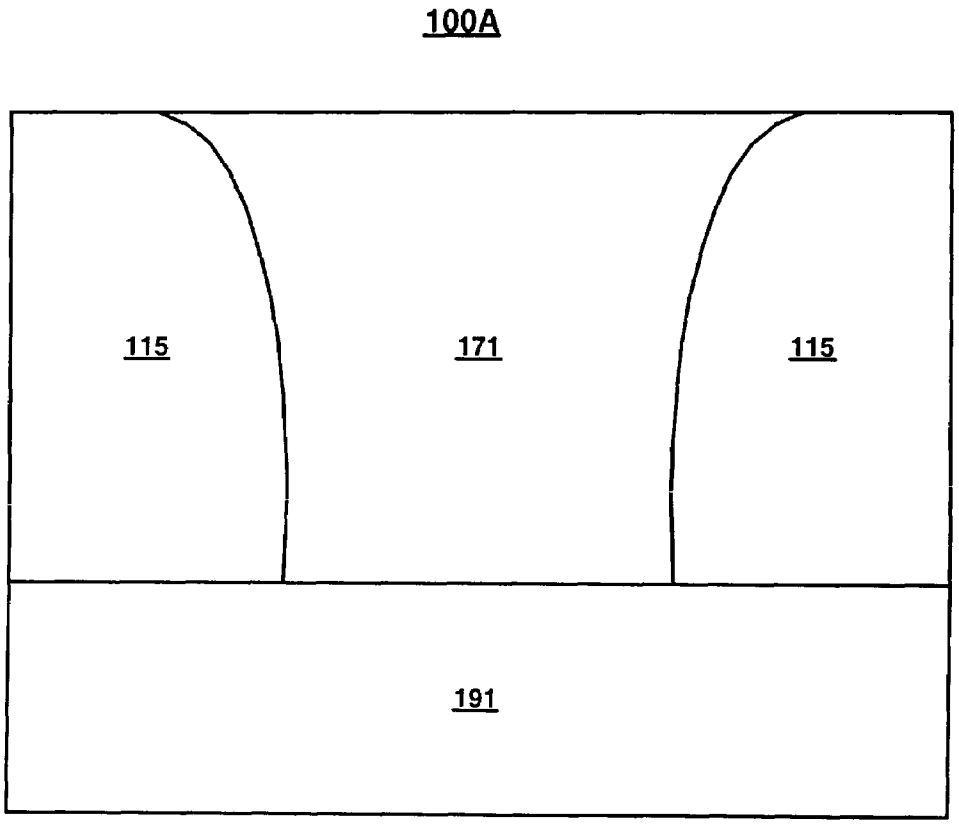


FIG. 1A

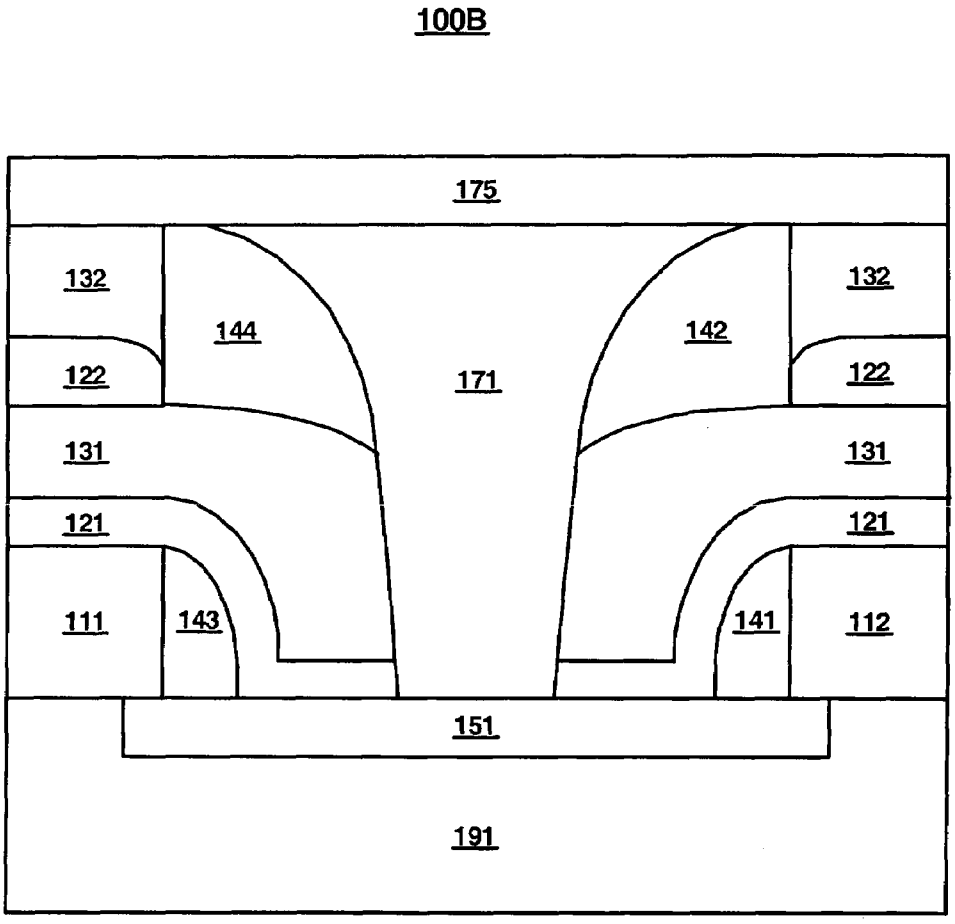

100B
175
132
144
142
132
122
171
122
131
131
121
121
111
143
141
112
151
191

FIG. 1B

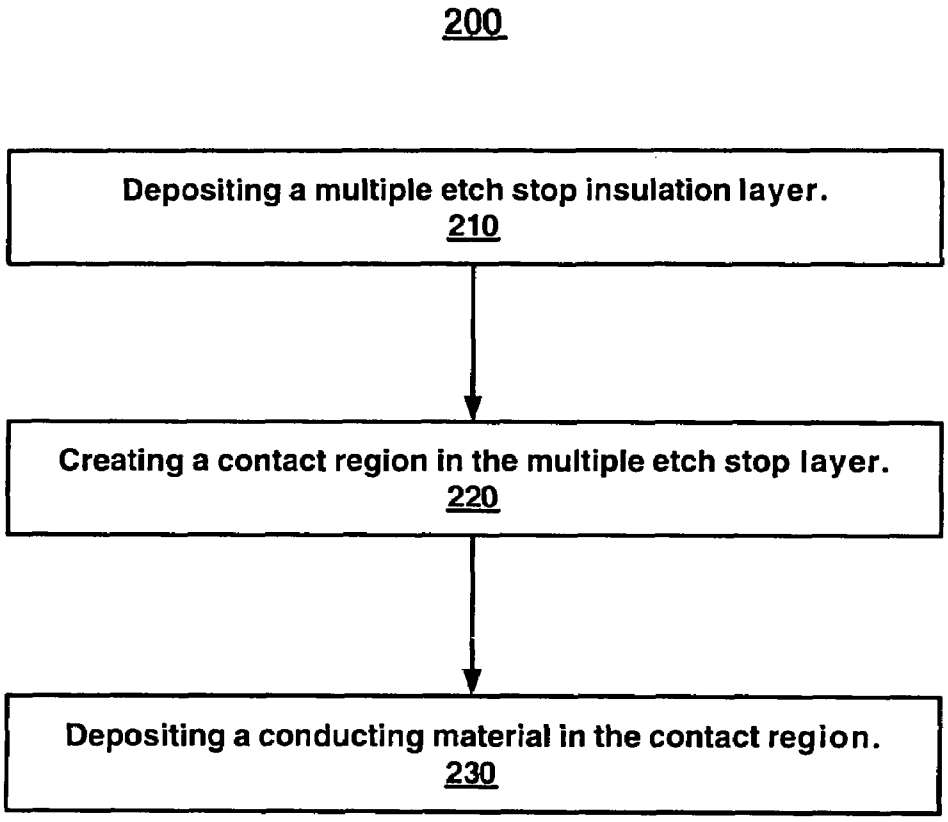

200
Depositing a multiple etch stop insulation layer.
210
Creating a contact region in the multiple etch stop layer.
220
Depositing a conducting material in the contact region.
230

FIG. 2A

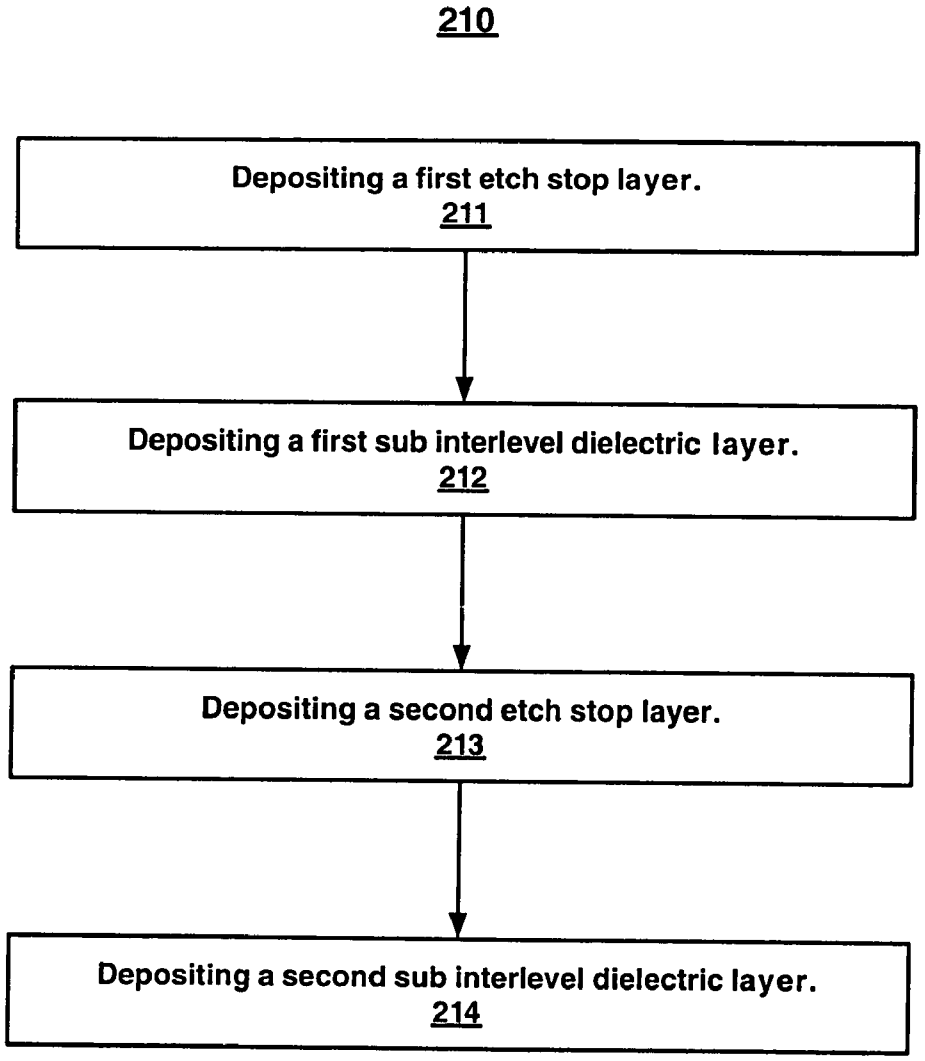

210
Depositing a first etch stop layer.
211
Depositing a first sub interlevel dielectric layer.
212
Depositing a second etch stop layer.
213
Depositing a second sub interlevel dielectric layer.
214

FIG. 2B

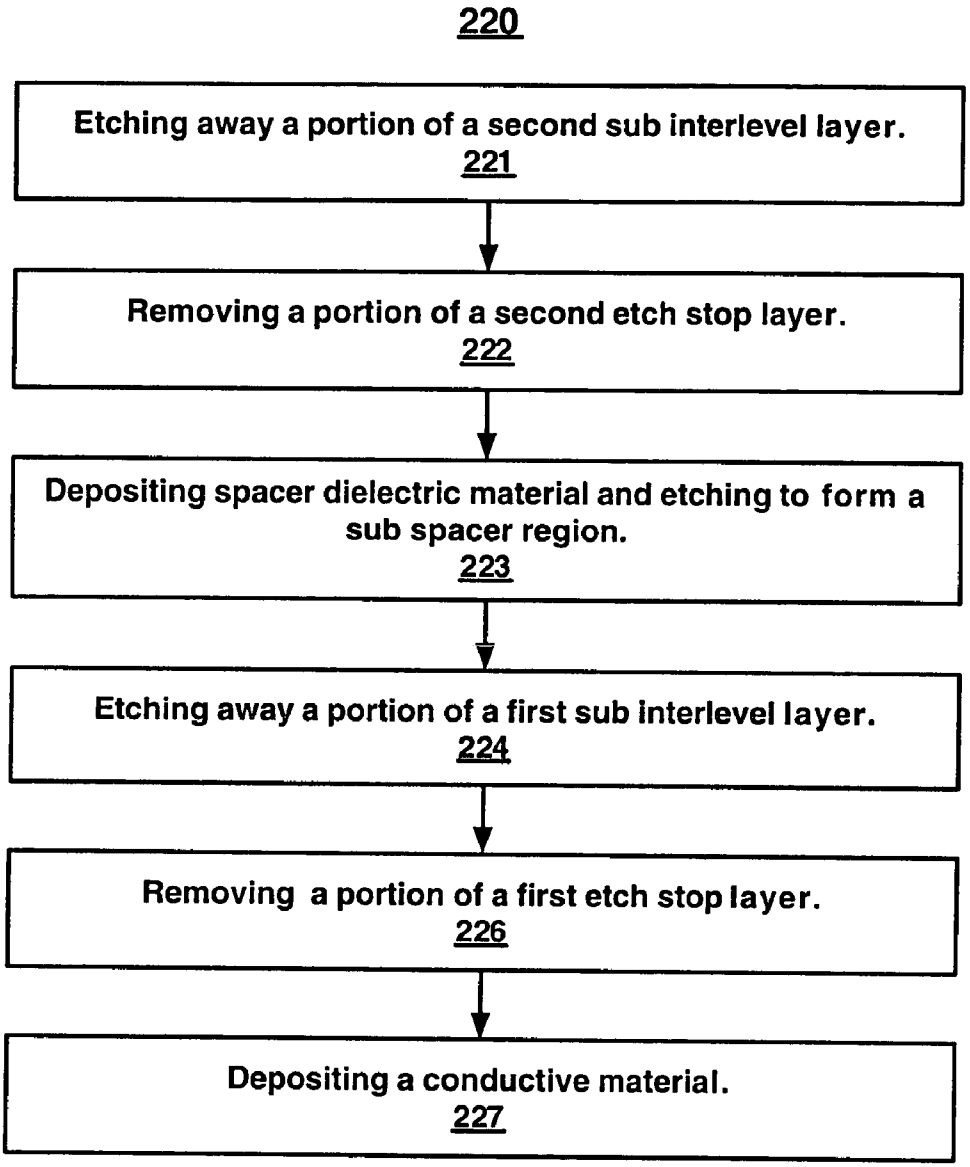

220
Etching away a portion of a second sub interlevel layer.
221
Removing a portion of a second etch stop layer.
222
Depositing spacer dielectric material and etching to form a sub spacer region.
223
Etching away a portion of a first sub interlevel layer.
224
Removing a portion of a first etch stop layer.
226
Depositing a conductive material.
227

FIG. 2C

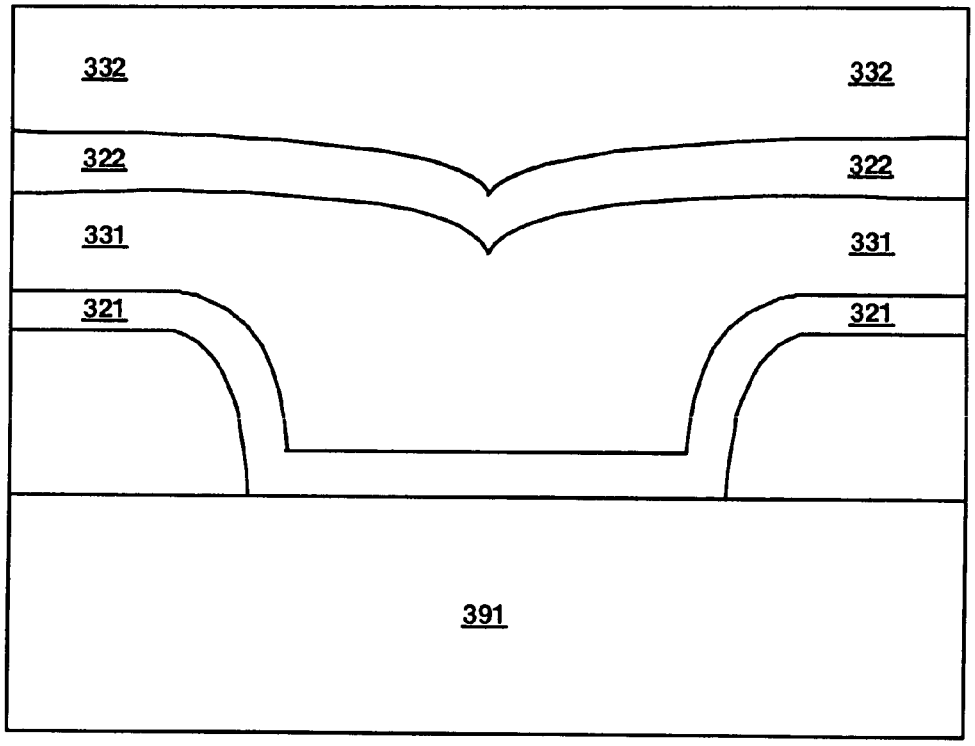


FIG. 3A

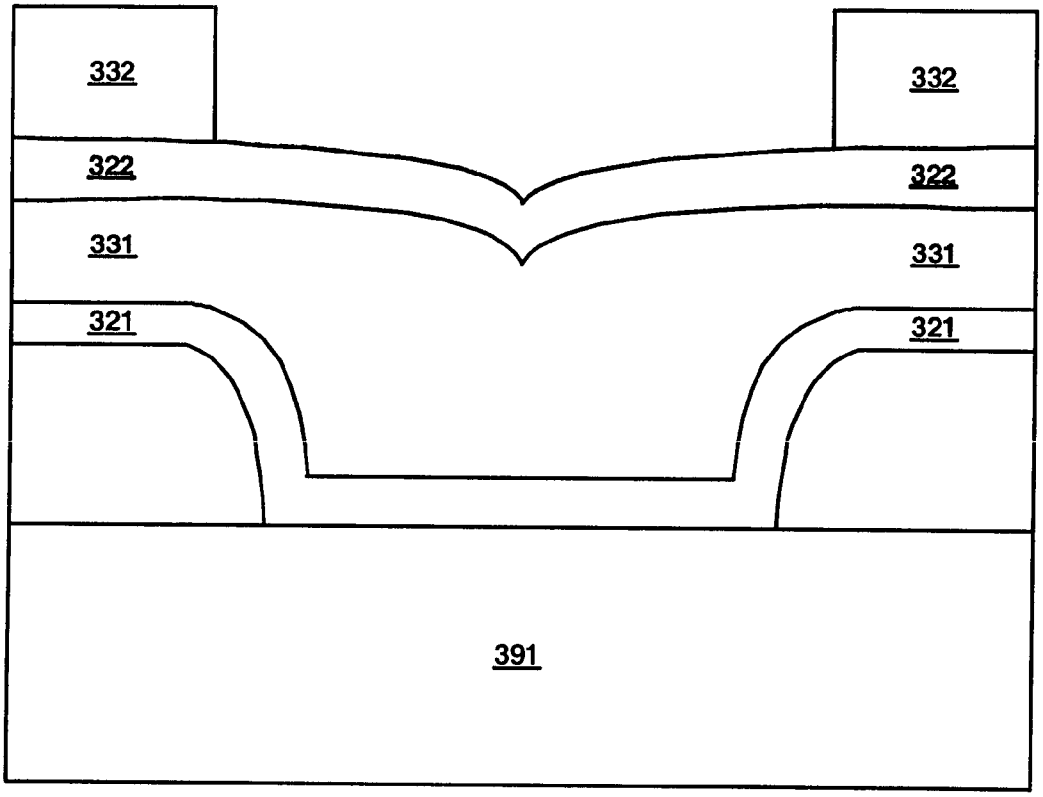


FIG. 3B

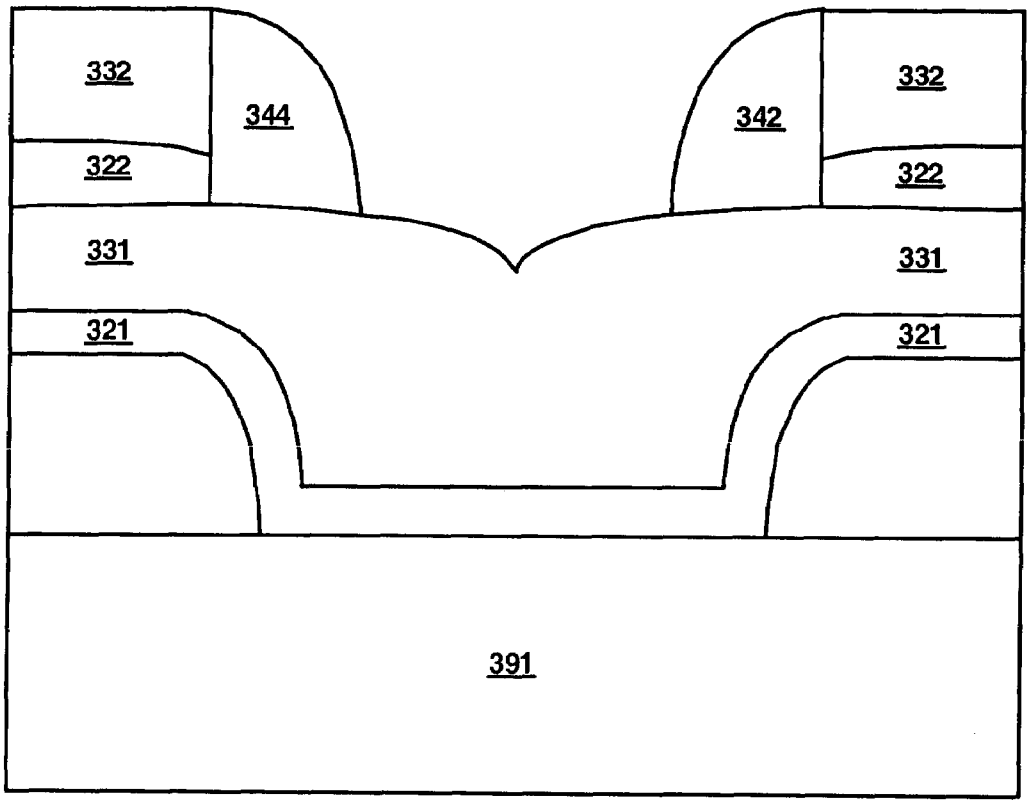


FIG. 3C

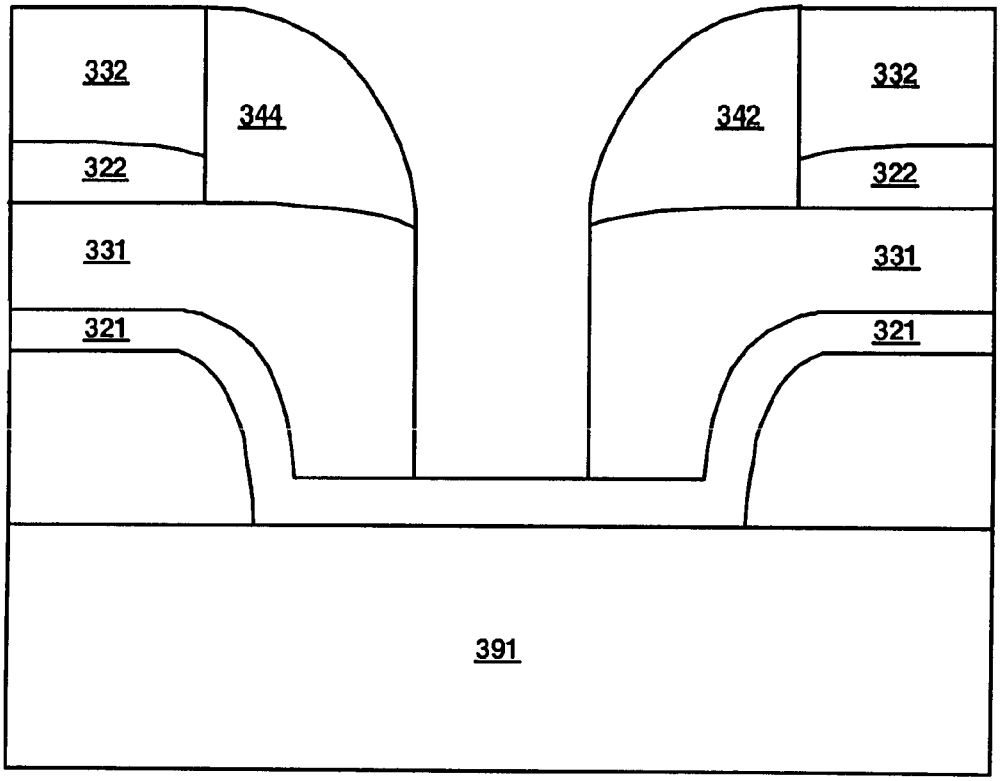


FIG. 3D

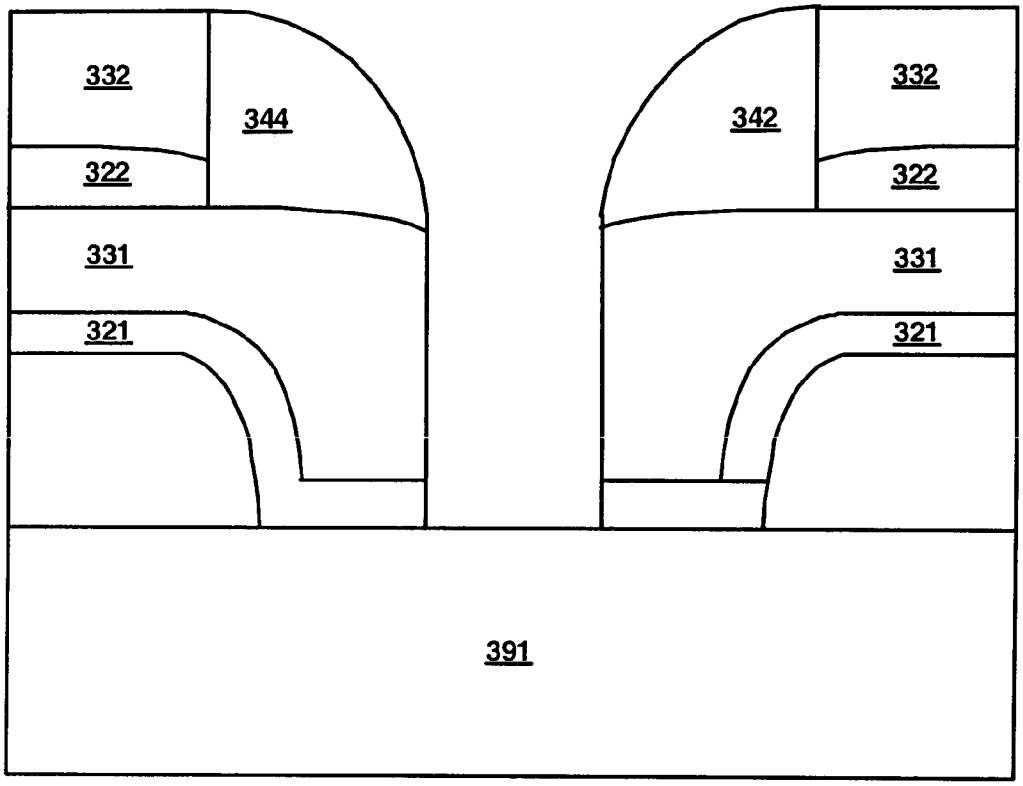


FIG. 3E

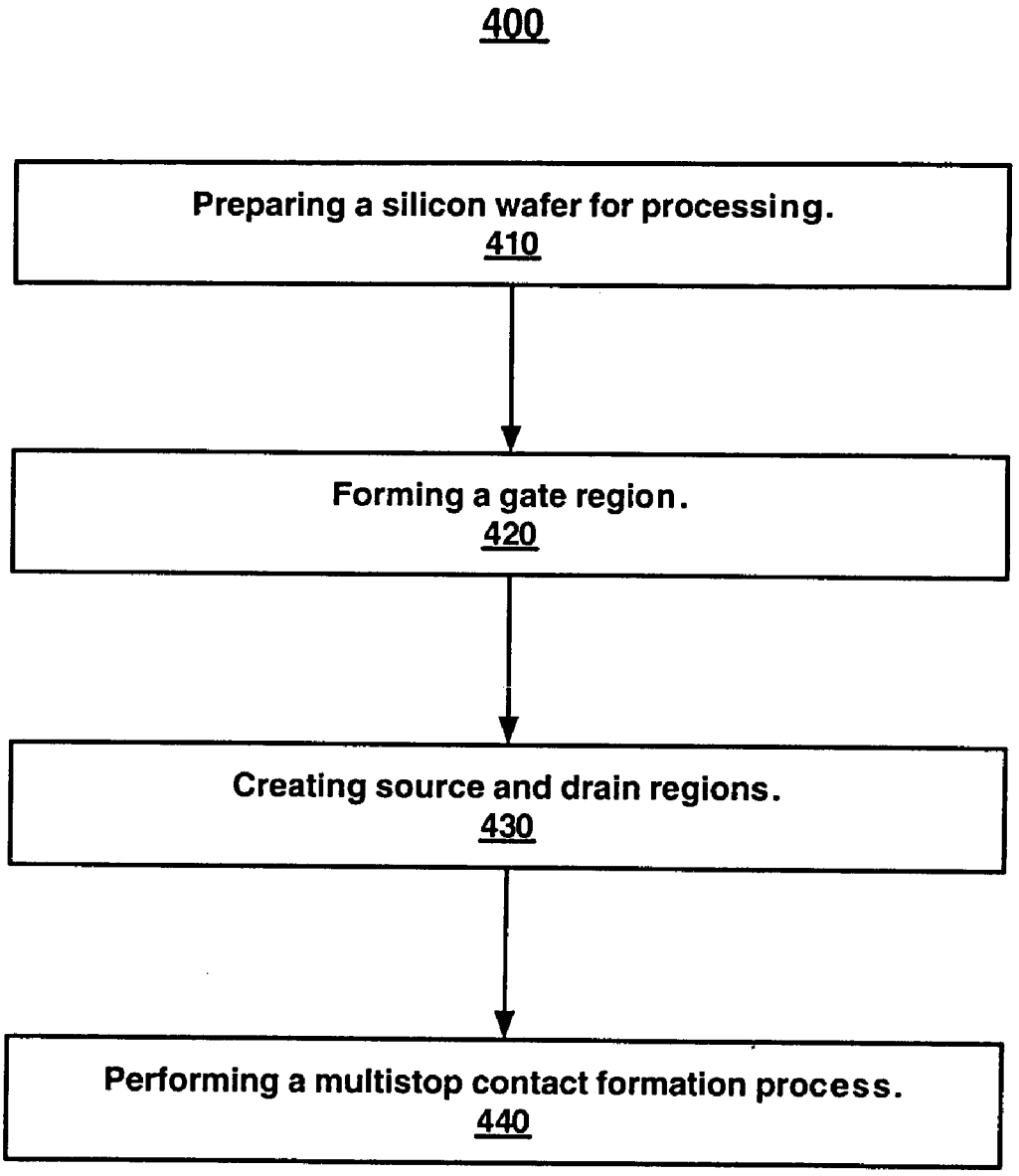

400
Preparing a silicon wafer for processing.
410
Forming a gate region.
420
Creating source and drain regions.
430
Performing a multistop contact formation process.
440

FIG. 4

SEMICONDUCTOR FORMATION METHOD THAT UTILIZES MULTIPLE ETCH STOP LAYERS

TECHNICAL FIELD

The present claimed invention relates to the field of semiconductor contact fabrication. More particularly, the present invention relates to a semiconductor contact fabrication system and method that utilizes multiple etch stop layers.

BACKGROUND ART

Electronic systems and circuits have made a significant contribution towards the advancement of modern society and are utilized in a number of applications to achieve advantageous results. Numerous electronic technologies such as digital computers, calculators, audio devices, video equipment, and telephone systems have facilitated increased productivity and reduced costs in analyzing and communicating data, ideas and trends in most areas of business, science, education and entertainment. Frequently, electronic systems designed to provide these results include integrated circuits. Integrated circuits typically include contact regions for conducting electricity (e.g., between active components) and it is often very difficult to achieve optimized results within requisite narrow tolerances when attempting to fabricate precise contact regions that operate properly.

Semiconductor integrated circuit manufacturing efforts are usually complicated by ever increasing demands for greater functionality. More complicated circuits are usually required to satisfy the demand for greater functionality. For example, there is usually a proportional relationship between the number of components included in an integrated circuit and the functionality. Integrated circuits with more components typically provide greater functionality. However, including more components within an integrated circuit often requires the components to be densely packed in relatively small areas and reliably packing a lot of components in relatively small areas of an integrated circuit (IC) is usually very difficult.

One traditional focus for achieving greater densities has been directed towards reducing the size of individual components (e.g., transistors). The components of an integrated circuit are usually fabricated on a single silicon substrate and maintaining both the integrity of the system as a whole as well as the individual basic device characteristics is very important for proper operation. Proper relational characteristics are very helpful in achieving these objectives and without them there is a tendency for detrimental interactions to occur. Thus, it is important for integrated circuit fabrication technologies to provide an advantageous balance between component integrity and increased component density.

Semiconductor contact formation processes usually include the creation of a contact void for deposition of the contact layer. The contact void creation typically determines the contact configuration. The smaller the void the more compact the contact and the greater the possible component density. However, decreases in contact sizes are usually limited by contact void creation processes (such as lithographic etching processes). Standard lithographic etching and removal processes traditionally have difficulty producing relatively small contact voids. Complex processes that attempt to create smaller voids are often cost prohibitive or nonfeasible.

While decreasing the size of a contact usually permits greater component densities, there are usually physical limitations on how small the contact can become and still operate properly. It is important for contacts to be formed in a manner that ensures proper operation without defects. Interconnection phenomenon such as electromigration can cause reliability problems as the dimension of the contact becomes very small. For example, electromigration can cause discontinuities in conducting materials if the dimensions are too small. Thus, most conducting materials have a critical dimension (CD) that limits how small a contact can be and still operate reliably. Fabrication of small contacts with desirable CD characteristics can be challenging.

It is also important to maintain adequate insulation around the contacts. Without proper component insulation there is a tendency for detrimental interactions between component parts to occur that hinder proper and reliable operation. For example, placement of more components in smaller spaces by reducing the separation between adjacent component parts often increases the probabilities of failures associated with leakage currents. It is also desirable for integrated circuit component formation processes to be efficient and low cost. While introduction of complex and complicated lithographic techniques may attempt to provide small size components, these advance techniques usually consume significant resources and are very expensive. Standard lithographic techniques are usually more efficient and do not require extensive retooling efforts. Therefore, the ability to precisely form semiconductor contact regions in a convenient and efficient manner is often very important.

SUMMARY OF THE INVENTION

The present invention is a semiconductor contact formation system and method. In one embodiment, a contact formation process forms contact insulation regions comprising multiple etch stop sublayers that facilitate formation of contacts. This contact formation process provides relatively small substrate connections while addressing critical dimension concerns in coupling to metal layers. The integrated circuit formation process also facilitates the creation of compact high density components (e.g., flash memory cells) that operate reliably. In one embodiment, a multiple etch stop contact formation process is implemented in which a multiple etch stop insulation layer comprising multiple etch stop layers is deposited. A contact region is formed in the multiple etch stop insulation layer by selectively removing (e.g., etching) some of the multiple etch stop insulation layer in which a larger portion of the multiple etch stop insulation layer is removed close to the metal layer and a smaller portion is removed closer to the substrate. In one exemplary implementation, the different contact region widths are achieved by performing multiple etching processes controlled by the multiple etch stop layers in the multiple etch stop insulation layer. Electrical conducting material (e.g., tungsten) is deposited in the contact region.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1**A is a block diagram of an integrated circuit with a substrate contact in accordance with one embodiment of the present invention.

FIG. **1**B is a block diagram showing a multiple etch stop insulation layer in accordance with one embodiment of the present invention.

FIG. **2**A is a flow chart of a contact formation process in accordance with one embodiment of the present invention.

FIG. **2**B is a flow chart of one embodiment of a present invention multiple etch stop insulation layer deposition process.

FIG. **2**C is a flow chart of a multiple etch stop contact formation process of yet another embodiment of the present invention.

FIG. **3**A is an illustration of one embodiment of a wafer after performing a multiple etch stop insulation layer deposition process.

FIG. **3**B is an illustration of a wafer after a portion of a sub interlevel dielectric layer is etched away in accordance with one embodiment of the present invention.

FIG. **3**C is an illustration of one embodiment of a wafer after a portion of etch stop layer is removed and a spacer is formed in accordance with the present invention.

FIG. **3**D is an illustration of one embodiment of a wafer after another portion of sub interlevel dielectric layer is etched away in one exemplary embodiment.

FIG. **3**E is all illustration of one embodiment of a wafer after a substrate protective etch stop layer is removed.

FIG. **4** is a flow chart of an integrated circuit method including a contact formation process in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Reference will now be made in detail to the preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings. While the invention will be described in conjunction with the preferred embodiments, it will be understood that they are not intended to limit the invention to these embodiments. On the contrary, the invention is intended to cover alternatives, modifications and equivalents, which may be included within the spirit and scope of the invention as defined by the appended claims. Furthermore, in the following detailed description of the present invention, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be obvious to one ordinarily skilled in the art that the present invention may be practiced without these specific details. In other instances, well known methods, procedures, components, and circuits have not been described in detail as not to unnecessarily obscure aspects of the current invention.

A present invention contact system and method includes contact insulation regions comprising multiple etch stop sublayers. In one embodiment of the present invention, the contact regions can be characterized by a relatively small substrate coupling area and relatively large metal layer coupling area. The relatively small substrate coupling area permits multiple active regions of an integrated circuit to be arranged relatively close to one another to achieve higher circuit density.

FIG. **1**A is a block diagram of integrated circuit **100**A, an integrated circuit with a substrate contact in accordance with one embodiment of the present invention. Integrated circuit comprises substrate **191**, a multiple etch stop insulation layer **115**, and a contact region **171**. In one exemplary implementation of the present invention multiple etch stop insulation layer **115** is an interlevel dielectric layer. Substrate **191** provides an electrical well for integrated circuit **100**. Contact region **171** is coupled to substrate **191** and provides an electrical path to and from substrate **191**. Multiple etch stop insulation layer **115** is coupled to contact region **171** and comprises a plurality of sublayers including multiple etch stop layers. Multiple etch stop insulation layer provides electrical insulation between other regions of integrated circuit **100** (e.g., not shown) and isolates guidance of electrical current flow in contact region **171**.

FIG. **1**B is a block diagram of integrated circuit **100**B, one embodiment of integrated circuit **100**A in accordance with the present invention. In integrated circuit **100**B multiple etch stop insulation layer **115** comprises etch stop layer **121**, sub interlevel dielectric layer **131**, etch stop layer **122**, sub interlevel dielectric layer **132** and spacer regions **141** through **144**. In one exemplary implementation, integrated circuit **100**B includes a gate region **111** and gate region **112** and a source or drain region (e.g., **151**). Gate region **111** is coupled to substrate **191**, spacer region **143** and etch stop layer **121** which is coupled to sub interlevel dielectric layer **131**. Sub interlevel dielectric layer **131** is coupled to spacer region **144** and etch stop layer **122** which in turn is coupled to sub interlevel dielectric layer **132**. Contact region **171** is coupled to substrate **191**, sub interlevel dielectric layer **131** and spacer region **144**.

The components of device **100**B cooperatively operate to provide an active device. Gate regions **111** and **112** control the flow of electricity between source and drain regions (e.g., region **151** and another similar region not shown). Contact region **171** conducts electrical current flow to and/or from regions of substrate **191**. For example, contact region **171** can conduct electrical current flow to and/or from a source or drain region (e.g., **151**). Sub interlevel dielectric layer **131**, sub interlevel dielectric layer **132** and spacer regions **141** through **144** provide electrical insulation between contact region **171** other regions of device **100** (e.g., gates **111** and **112**) and isolate guidance of electrical current flow in contact region **171**. Etch stop layer **121** and etch stop layer **122** facilitate definition of the configuration of contact region **171**.

In one embodiment of device **100**B, contact region **171** has a relatively small substrate coupling area and relatively large metal layer coupling area (e.g., coupled to metal layer **175**). For example, the substrate coupling area width can be controlled by spacer width **142** and **144** to the range of 0.06 μm to 0.13 μm and the metal layer coupling area width can be 0.16 μm to 0.3 μm. The relatively small substrate coupling area permits multiple active regions (e.g., gate regions **111** and **112**) to be arranged relatively close to one another while the relatively large metal layer coupling areas of device facilitate avoidance of critical dimension (CD) issues.

FIG. **2**A is a flow chart of contact formation process **200**, one embodiment of the present invention. Contact formation process **200** facilitates the fabrication of contact regions with a relatively small substrate coupling area and relatively large metal layer coupling area. The relatively small substrate coupling area permits multiple active regions in an integrated circuit to be arranged relatively close to one another while the relatively large metal layer coupling area facilitate avoidance of CD issues.

At step **210** a multiple etch stop insulation layer comprising multiple etch stop layers is deposited. In one embodiment of the present invention, the multiple etch stop insulation layer (e.g., **115**) is an interlevel dielectric layer. The multiple etch stop insulation layer permits etch flexibility with the combination of spacer formation **142** and **144** to create a small contact formation. In one exemplary implementation, the multiple etch stop insulation layer is made smooth and level (e.g., polished by a CMP process). In one embodiment of the present invention, step **210** includes a multiple etch stop insulation layer deposition process.

FIG. **2**B is a flow chart of one embodiment of multiple etch stop insulation layer deposition process **210** and FIG. **3**A is an illustration of one embodiment of a wafer after performing multiple etch stop insulation layer deposition process **210**. In step **211** etch stop layer **321** is deposited (e.g., 200 to 500 Å thick). At step **212** sub interlevel dielectric layer **331** is depos-

ited (1,000 to 2000 Å). Etch stop layer **322** (e.g., 200 to 500 Å thick) is deposited at step **213**. In step **214**, another sub interlevel dielectric layer **332** is deposited (e.g., 10 KÅ+1 KÅ). In one exemplary implementation the etch stop layers are nitride or SiON and the sub interlevel dielectric layers are oxide. In one embodiment an ARC layer is deposited (e.g., a ARC film with Si_3N_4, SiON, or Si_3N_4/OX).

Referring to FIG. **2**A again, a contact region is created in the multiple etch stop insulation layer at step **220**. In one embodiment of the present invention, a multiple stop etch process creates a contact region in the multiple etch stop insulation layer. In one exemplary implementation, the multiple etch stop insulation layers creates etching processes with non-lithographic spacer formation using multiple steps so that smaller contact holes (e.g., 0.06 μm to 0.13 μm) can be created close to the substrate coupling area and larger contact top (e.g., 0.16 μm to 0.18 μm) can be created closer to a metal layer coupling area. In one embodiment the contact region is created by a multiple etch stop contact formation process.

FIG. **2**C is a flow chart of multiple etch stop contact formation process **220**A in accordance with one embodiment of the present invention. In step **221***a* portion (e.g., 0.16 μm to 0.18 μm width region) of sub interlevel dielectric layer **332** is etched away. FIG. **3**B is an illustration of one embodiment of a wafer after a portion of sub interlevel dielectric layer **332** is etched away in step **221**. A portion (e.g., 0.16 μm to 0.18 μm width region) of etch stop layer **322** is removed in step **222**. In step **223** sub spacer regions **342** and **344** are formed. FIG. **3**C is an illustration of one embodiment of a wafer after a portion of etch stop layer **322** is removed in step **222** and sub spacer regions **342** and **344** are formed in step **223**. In step **224** a portion (e.g., 0.06 μm to 0.13 μm width region) of sub interlevel dielectric layer **331** is etched. FIG. **3**D is an illustration of one embodiment of a wafer after a portion of sub interlevel dielectric layer **331** is etched away in step **224**. A portion (e.g., 0.06 μm to 0.13 μm width region) of etch stop layer **321** is removed in step **226**. FIG. **3**E is an illustration of one embodiment of a wafer after step **226** is performed. In step **227** a conductive material is deposited in the void left after said etching and removing of the portions of the first sub interlevel dielectric layer, the first etch stop layer, the second sub interlevel dielectric layer, and the second etch stop layer.

Referring now to step **230** shown in FIG. **2**A, a conducting material (e.g., tungsten) is deposited in the contact region. In one exemplary implementation, the conducting material is part of a metal layer deposition. In one embodiment of the present invention, a plurality of metal layers are deposited and each of the respective metal layers are separated by insulating layers. The metal layers can selectively couple integrated circuit components formed on the wafer to each other and external components.

In one embodiment, a present invention contact formation process is included in an integrated circuit fabrication process. In one exemplary implementation, a present invention contact formation process is utilized to provide electrical contacts to a source and drain region. For example, contacts are formed to couple a flash memory cell source and drain region to word and bit lines. FIG. **4** is a flow chart of integrated circuit formation process **400**, including a contact formation process in accordance with one embodiment of the present invention.

In step **410**, a silicon wafer substrate is prepared for processing. In one embodiment of the present invention, the wafer surface is made smooth and level, for example by chemical mechanical polishing (CMP). An oxide pad layer and a subsequent protective layer of nitride are deposited on the surface. In one exemplary implementation, additional polishing is performed to provide a smooth and level surface after the protective oxide and nitride layers are added.

In step **420** a gate region is formed. In one embodiment of the present invention, forming a gate region comprises depositing a gate insulation layer, depositing a control gate layer, and removing the gate insulation layer and the control gate layer from non gate region areas. In one embodiment of the present invention, a floating gate is formed in step **420**. An insulating layer (e.g., oxide) is deposited and a floating gate area is created in the insulating layer. For example, a floating gate area is etched in the insulating layer and a charge trapping material (e.g., a polysilicide) is deposited in the floating gate area. Excess charge trapping material is removed and additional insulating material deposited. A control gate material (e.g., polysilicon) is deposited on top of the insulating material. The materials deposited during the gate formation process are removed (e.g., etched) from areas not included in the gate (e.g., areas above a source and drain). In one exemplary implementation, a sidewall spacer material is deposited on the sides of the gate area and excess sidewall spacer material is removed.

Source and drain regions are created in step **430**. In one embodiment of the present invention, a source and drain formation process is performed. The source and drain area are prepared for implantation and diffusion. For example, excess material from the gate formation process and the protective layer materials over the source and drain areas are removed. In one exemplary implementation, dopants (e.g., arsenic, phosphorus, boron, etc.,) are introduced into the substrate in the source and drain regions by implantation and/or diffusion. In one embodiment, the source and drain include lightly doped region, extensions and/or halos.

In step **440**, a multiple etch stop contact formation process (e.g., contact formation process **200**) is performed. A multiple etch stop insulation layer comprising multiple etch stop layers is deposited. A contact region is formed in the multiple etch stop insulation layer by selectively removing (e.g., etching) some of the multiple etch stop insulation layer. A substrate coupling area of the contact region is smaller than a metal layer coupling area of the contact region. Electrical conducting material (e.g., tungsten) is deposited in the contact region.

Thus, the present invention facilitates precise formation of semiconductor contact regions in a convenient and efficient manner. Utilization of a multi etch stop process in formation of the semiconductor contact regions enables simple lithographic processes to provide contacts with relatively small substrate patterns and relatively large metal layer coupling patterns. This contact formation process provides contacts with small CD features. The shallow junctions also enable reductions in space between integrated circuit components permitting realization of increased component density (e.g., a larger number of components concentrated in smaller areas).

The foregoing descriptions of specific embodiments of the present invention have been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise forms disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application, to thereby enable others skilled in the art to best utilize the invention and various embodiments with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the Claims appended hereto and their equivalents.

What is claimed is:

1. A contact formation process comprising;
preparing a silicon wafer substrate;
forming a gate region;
creating source and drain regions;
depositing a multiple etch stop insulation layer, wherein depositing said multiple etch stop insulation layer comprises:
depositing a first etch stop layer directly on said silicon wafer substrate in a contact region;
depositing a first sub interlevel dielectric layer over said first etch stop layer;
depositing a second etch stop layer over said first sub interlevel dielectric layer, wherein said second etch stop layer has similar selectivity characteristics as said first etch stop layer and wherein said first sub interlevel dielectric layer is between said first etch stop layer and said second etch stop layer; and
depositing a second sub interlevel dielectric layer over said second etch stop layer;
forming said contact region in said multiple etch stop insulation layer by selectively removing some of said multiple etch stop insulation layer and forming a sub spacer region such that a substrate coupling area of said contact region is smaller than a metal layer coupling area of said contact region; and
depositing electrical conducting material in said contact region.

2. A contact formation process of claim 1 wherein said first etch stop layer is in a range of about 300 to 800 Å thick, said first sub interlevel dielectric layer is in a range of about 1,000 to 2,000 Å thick, said second etch stop layer is in a range of about 300 to 800 Å thick and said second sub interlevel dielectric layer is in a range of about 10 KÅ±1 KÅ thick.

3. A contact formation process of claim 1 wherein said forming said contact region in said multiple etch stop insulation layer comprises:
etching away a portion of said second sub interlevel dielectric layer;
removing a portion of said second etch stop layer;
depositing spacer dielectric material and etching to form said sub spacer region;
etching away a portion of said first sub interlevel dielectric layer; and
removing a portion of said first etch stop layer.

4. A contact formation process of claim 3 wherein an area of said portion of said first sub interlevel dielectric layer etched away is smaller than an area of said portion of said second sub interlevel dielectric layer that is etched away due to sub spacer formation.

5. A contact formation process of claim 4 wherein said area of said portion of said first sub interlevel dielectric layer etched away is a range of about 0.06 μm to 0.13 μm wide and said area of said portion of said second sub interlevel dielectric layer that is etched away is in the range of about 16 μm to 0.18 μm wide.

6. A contact formation process of claim 1 further comprising depositing an anti reflective coating layer.

7. A contact formation process of claim 1 wherein said forming said gate region comprises:
depositing a gate insulation layer;
depositing a control gate layer; and
removing said gate insulation layer and said control gate layer from non gate region areas.

8. A contact formation process of claim 6 further comprising forming a floating gate.

9. A contact formation process comprising:
depositing a multiple etch stop insulation layer comprising a first etch stop layer and a second etch stop layer wherein said first etch stop layer and said second etch stop layers have similar selectivity characteristics, said first etch stop layer formed in an area directly next to a substrate corresponding to a contact region and under a first sub interlevel dielectric layer and said second etch stop layer formed under a second interlevel dielectric layer, wherein said first sub interlevel dielectric layer is between said first etch stop layer and said second etch stop layer, wherein said multiple etch stop insulation layer is formed utilizing a lithography process;
creating said contact region in said multiple etch stop insulation layer, wherein said creating of said contact region includes forming sub-spacer regions in removed portions of said second sub interlevel dielectric layer and said second etch stop layer, wherein a non-lithography spacer formation process is also utilized to achieve a contact bottom; and
depositing a conducting material in said contact region.

10. A contact formation process of claim 9 wherein depositing said multiple etch stop insulation layer comprises:
depositing said first etch stop layer;
depositing said first sub interlevel dielectric layer;
depositing said second etch stop layer, and
depositing said second sub interlevel dielectric layer.

11. A contact formation process of claim 9 comprising:
etching away a portion of said second sub interlevel dielectric layer;
removing a portion of said second etch stop layer;
depositing spacer dielectric material and etching to form said sub-spacer regions;
etching away a portion of said first sub interlevel dielectric layer; and
removing a portion of said first etch stop layer.

12. A contact formation process of claim 11 wherein said first etch stop layer and second etch stop layer are nitride.

13. A contact formation process of claim 11 wherein said sub spacer regions comprise nitride or SiON.

14. A contact formation process of claim 9 wherein said contact bottom dimension has a range of about 0.06 μm to 0.13 μm.

15. A contact formation process comprising:
depositing a first etch stop layer directly on a substrate in a contact region;
depositing a first sub interlevel dielectric layer over said first etch stop layer;
depositing a second etch stop layer over said first sub interlevel dielectric layer, wherein said second etch stop layer has similar selectivity characteristics as said first etch stop layer and wherein said first sub interlevel dielectric layer is between said first etch stop layer and said second etch stop layer;
depositing a second sub interlevel dielectric layer over said second etch stop layer;
creating said contact region in said first etch stop layer, said first sub interlevel dielectric layer, second etch stop layer, and said second sub interlevel dielectric layer, wherein said creating of said contact region comprises:
etching away a portion of said second sub interlevel dielectric layer;

removing a portion of said second etch stop layer;

depositing spacer dielectric material;

etching said spacer dielectric material to form a first sub-spacer region and a second sub-spacer region, wherein said first and second sub spacer regions define an exposed portion of said first sub interlevel dielectric layer that is in the range of about 0.06 μm to 0.13 μm wide;

etching away said exposed portion of said first sub interlevel dielectric layer; and

removing a portion of said first etch stop layer; and

depositing a conducting material in said contact region, wherein said conducting material forms a contact bottom that is in the range of about 0.06 μm to 0.13 μm wide.

* * * * *

EXHIBIT H

US007977797B2

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

(10) **Patent No.: US 7,977,797 B2**
(45) **Date of Patent: Jul. 12, 2011**

(54) **INTEGRATED CIRCUIT WITH CONTACT REGION AND MULTIPLE ETCH STOP INSULATION LAYER**

(75) Inventors: **Wenmei Li**, Sunnyvale, CA (US); **Angela T. Hui**, Fremont, CA (US); **Dawn Hopper**, San Jose, CA (US); **Kouros Ghandehari**, Fremont, CA (US)

(73) Assignee: **Spansion LLC**, Sunnyvale, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/539,480**

(22) Filed: **Aug. 11, 2009**

(65) **Prior Publication Data**

US 2009/0294969 A1 Dec. 3, 2009

Related U.S. Application Data

(62) Division of application No. 10/934,828, filed on Sep. 2, 2004, now Pat. No. 7,572,727.

(51) **Int. Cl.**
H01L 23/522 (2006.01)

(52) **U.S. Cl.** **257/760**; 257/774; 257/E23.145

(58) **Field of Classification Search** 438/639
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,912,188	A	6/1999	Gardner et al.	
6,078,073	A	6/2000	Habu et al.	
6,083,822	A	7/2000	Lee	
6,365,504	B1	4/2002	Chien et al.	
6,441,418	B1 *	8/2002	Shields et al.	257/296
7,361,587	B1	4/2008	Li et al.	
2002/0008323	A1	1/2002	Watanabe et al.	
2005/0003656	A1 *	1/2005	Chung	438/639

* cited by examiner

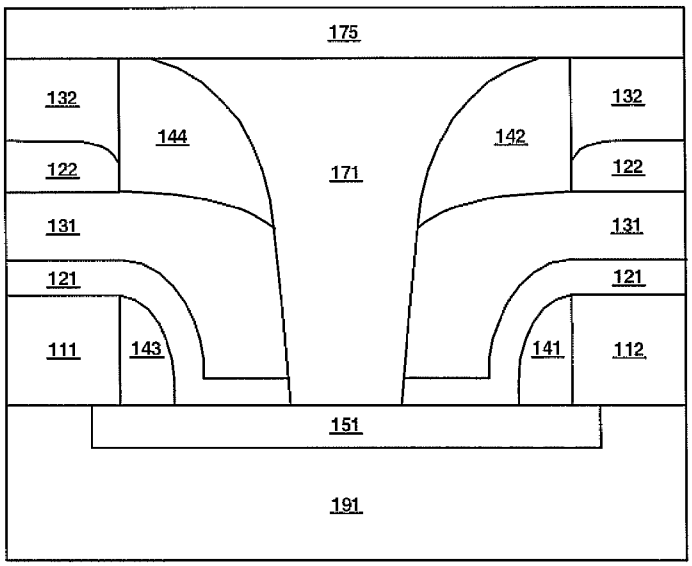
Primary Examiner — Stephen W Smoot

(57) **ABSTRACT**

The present invention is a semiconductor contact formation system and method. Contact insulation regions are formed with multiple etch stop sublayers that facilitate formation of contacts. This contact formation process provides relatively small substrate connections while addressing critical lithographic printing limitation concerns in forming contact holes with small dimensions. In one embodiment, a multiple etch stop insulation layer comprising multiple etch stop layers is deposited. A contact region is formed in the multiple etch stop insulation layer by selectively removing (e.g., etching) some of the multiple etch stop insulation layer. In one embodiment, a larger portion of the multiple etch stop insulation layer is removed close to the metal layer and a smaller portion is removed closer to the substrate. The different contact region widths are achieved by performing multiple etching processes controlled by the multiple etch stop layers in the multiple etch stop insulation layer and spacer formation to shrink contact size at a bottom portion. Electrical conducting material (e.g., tungsten) is deposited in the contact region.

15 Claims, 11 Drawing Sheets

100B



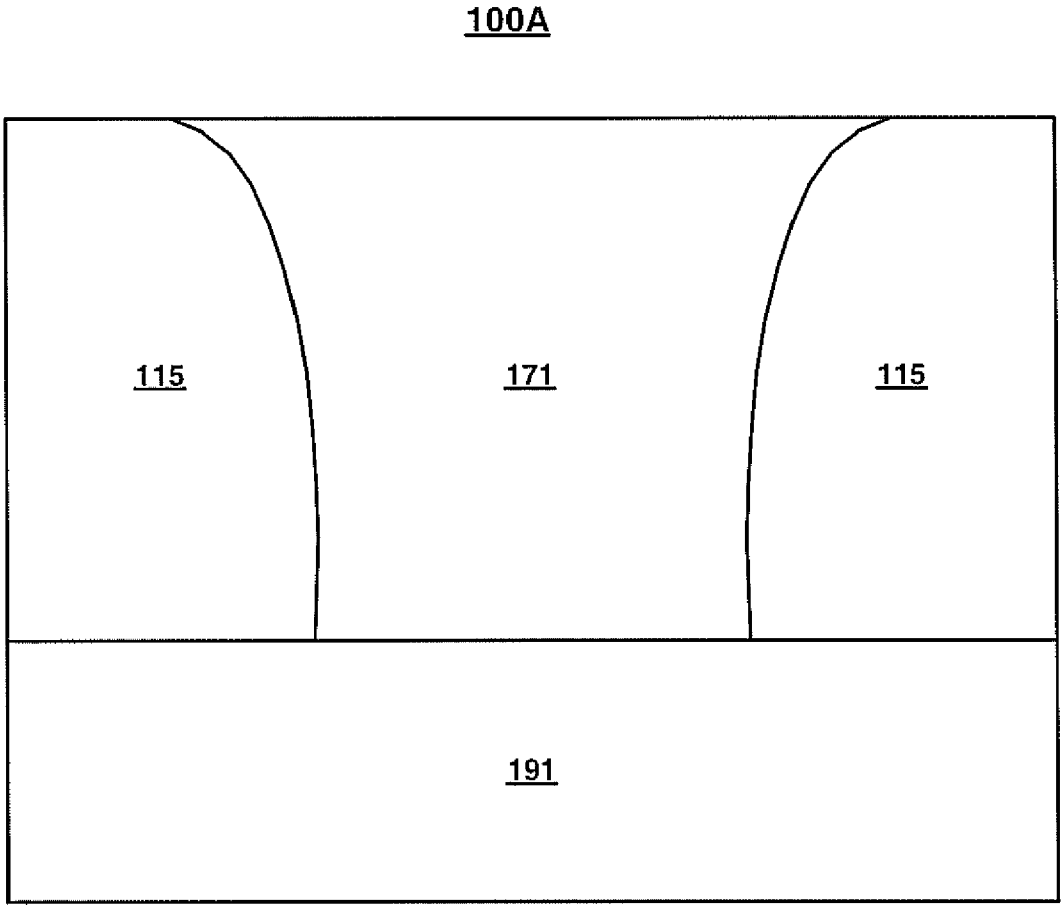


FIG. 1A

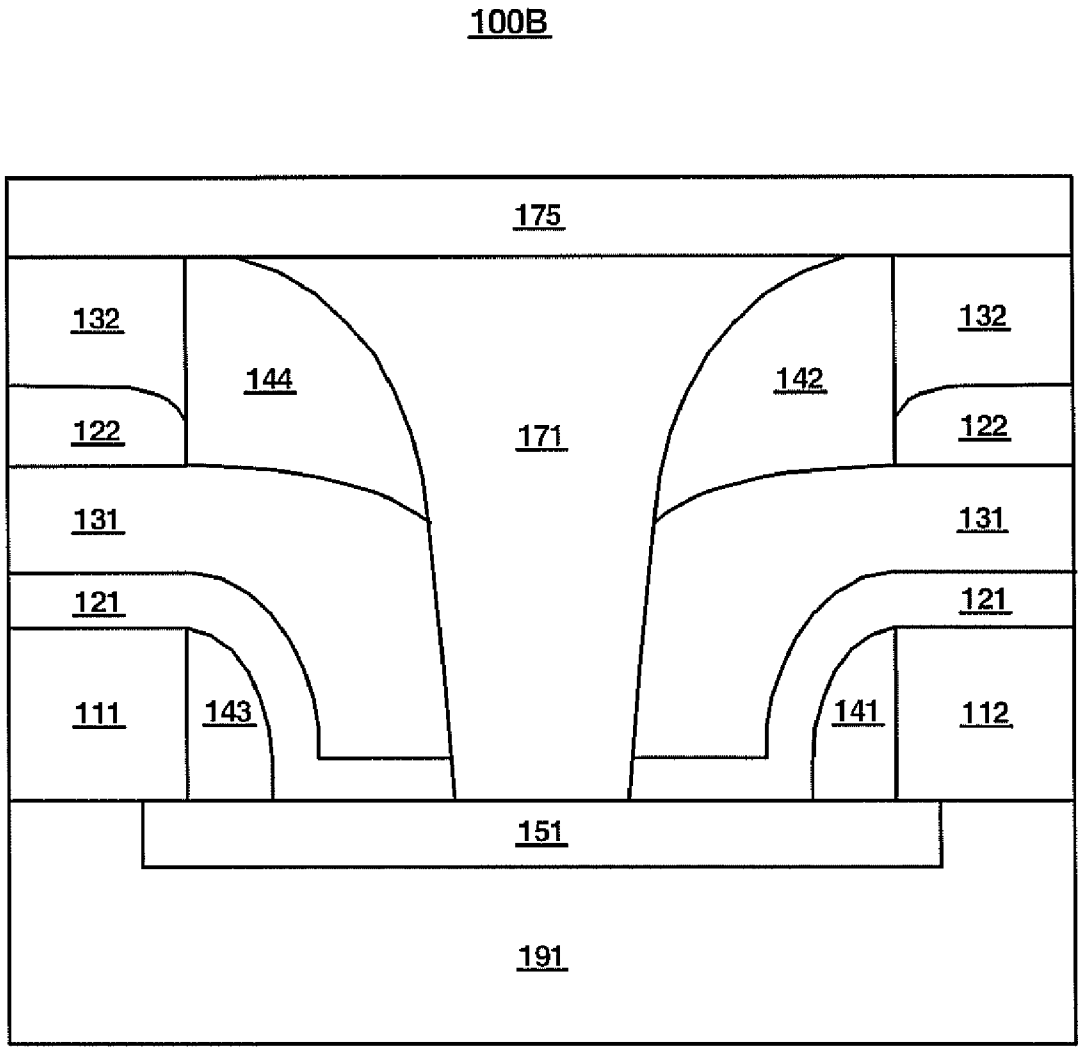

100B
175
132
144
142
132
122
171
122
131
131
121
121
111
143
141
112
151
191

FIG. 1B

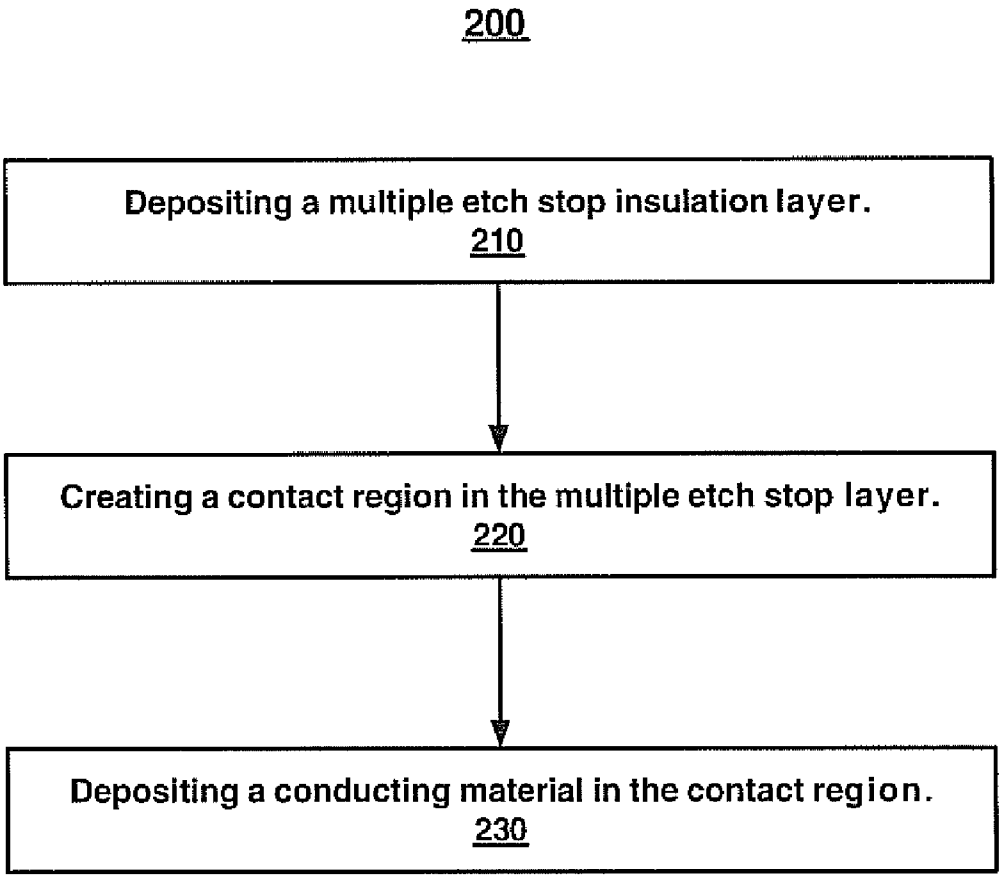

200
Depositing a multiple etch stop insulation layer.
210
Creating a contact region in the multiple etch stop layer.
220
Depositing a conducting material in the contact region.
230

FIG. 2A

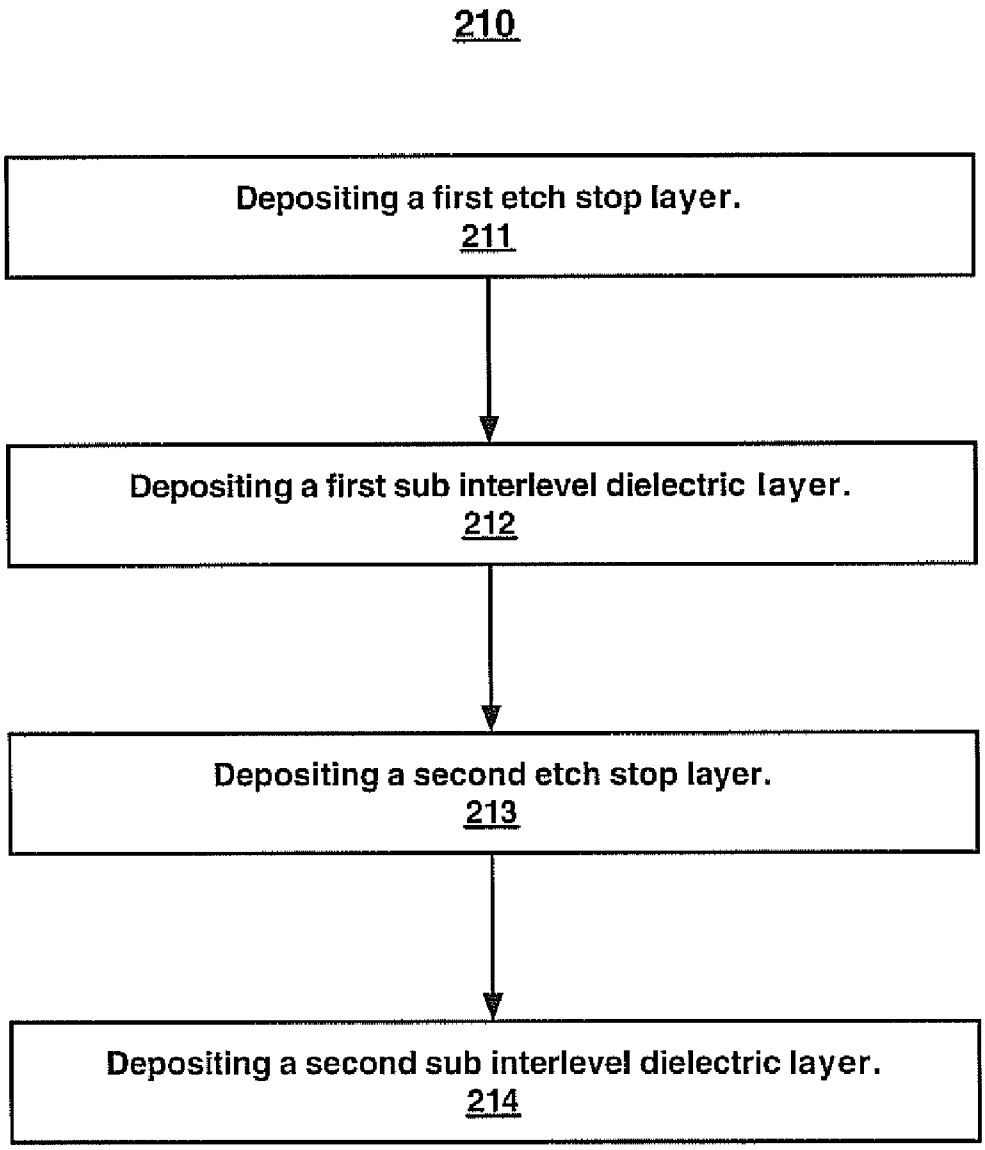

210
Depositing a first etch stop layer.
211
Depositing a first sub interlevel dielectric layer.
212
Depositing a second etch stop layer.
213
Depositing a second sub interlevel dielectric layer.
214

FIG. 2B

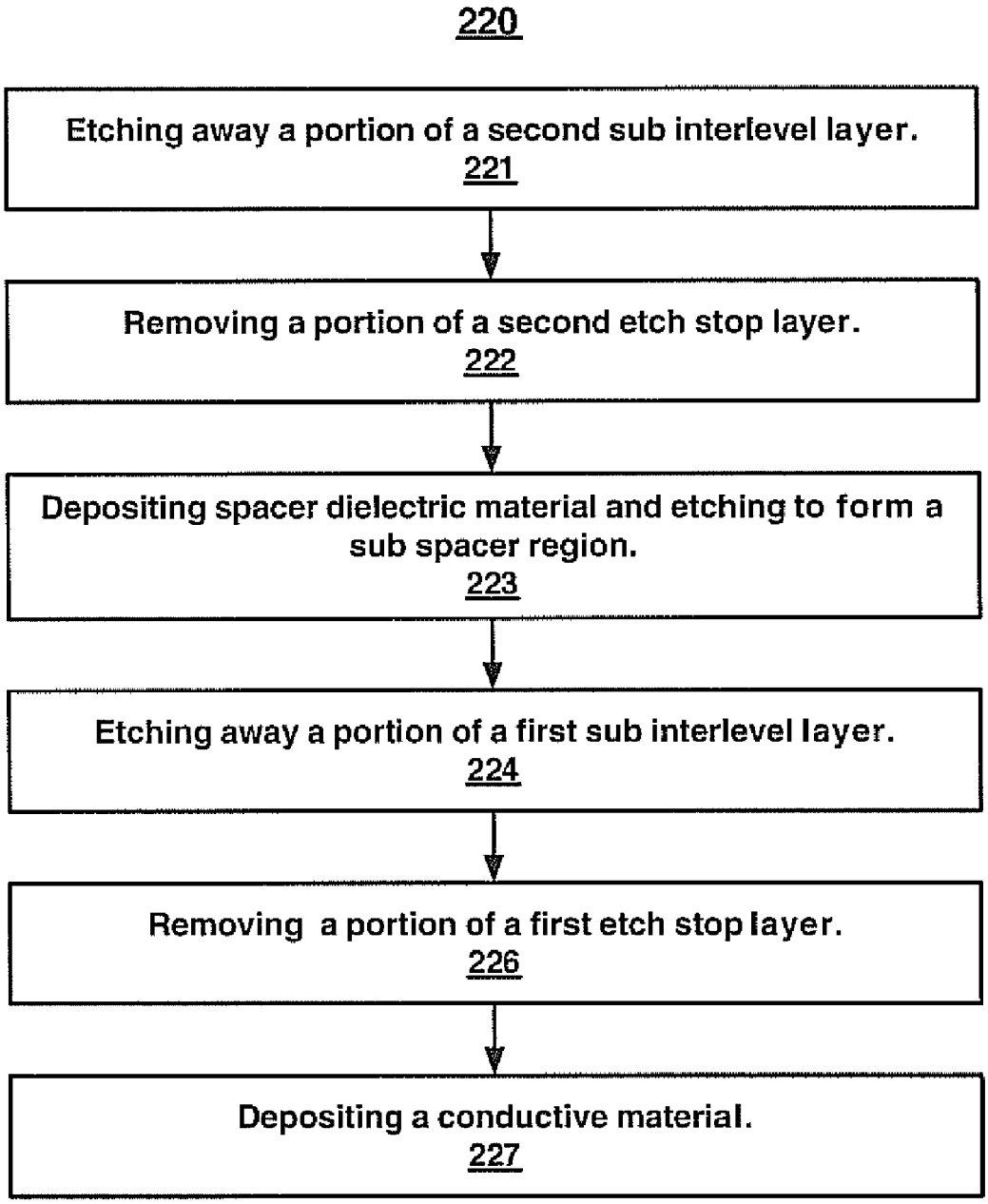

220
Etching away a portion of a second sub interlevel layer.
221
Removing a portion of a second etch stop layer.
222
Depositing spacer dielectric material and etching to form a sub spacer region.
223
Etching away a portion of a first sub interlevel layer.
224
Removing a portion of a first etch stop layer.
226
Depositing a conductive material.
227

FIG. 2C

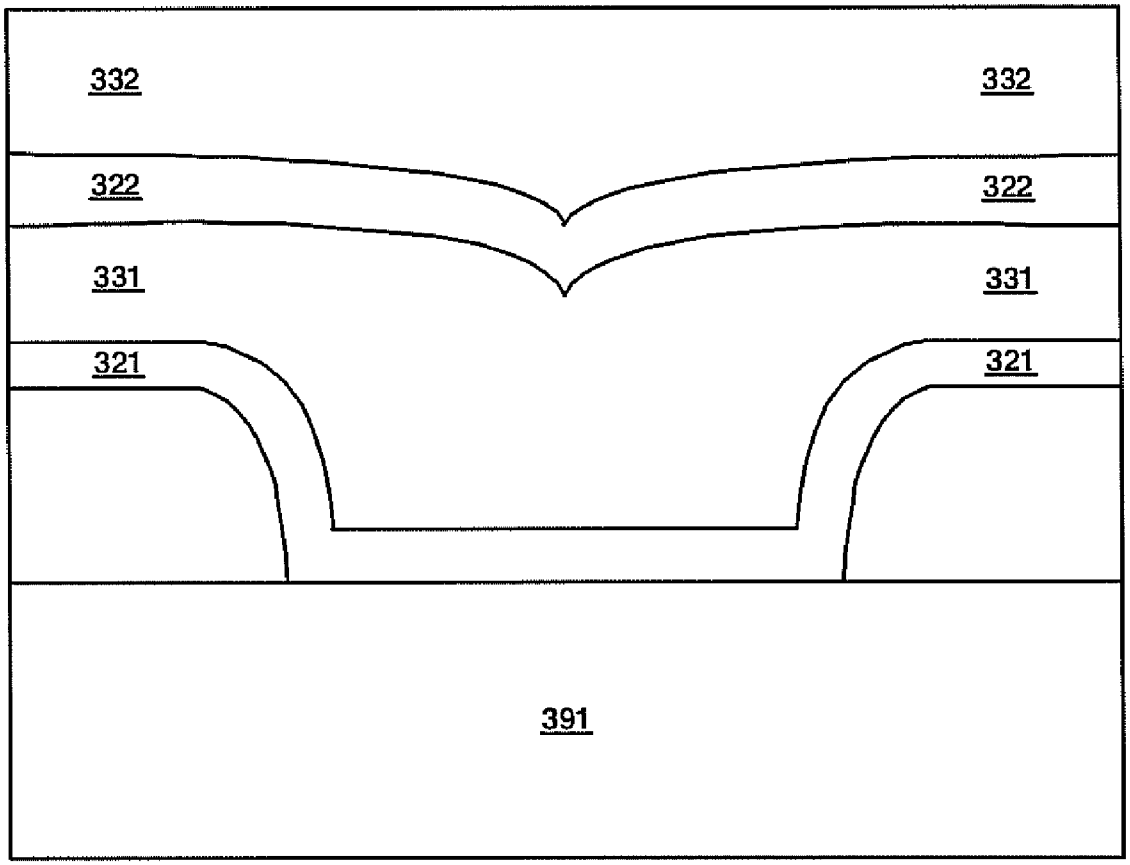


FIG. 3A

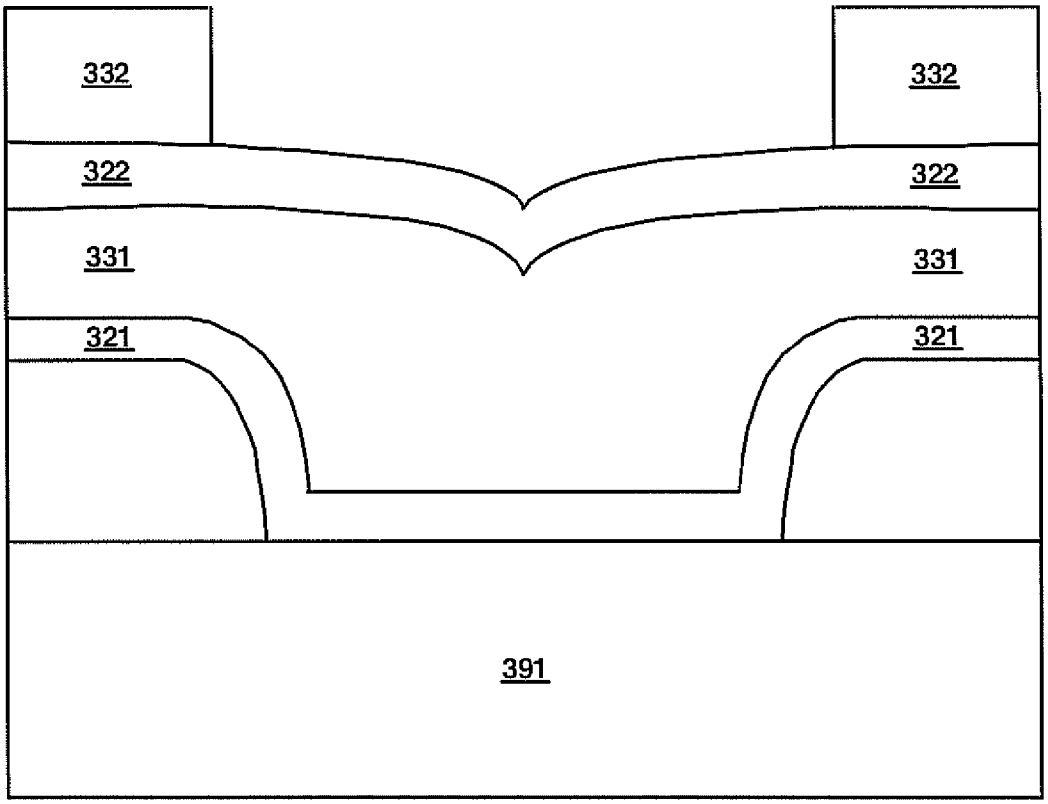


FIG. 3B

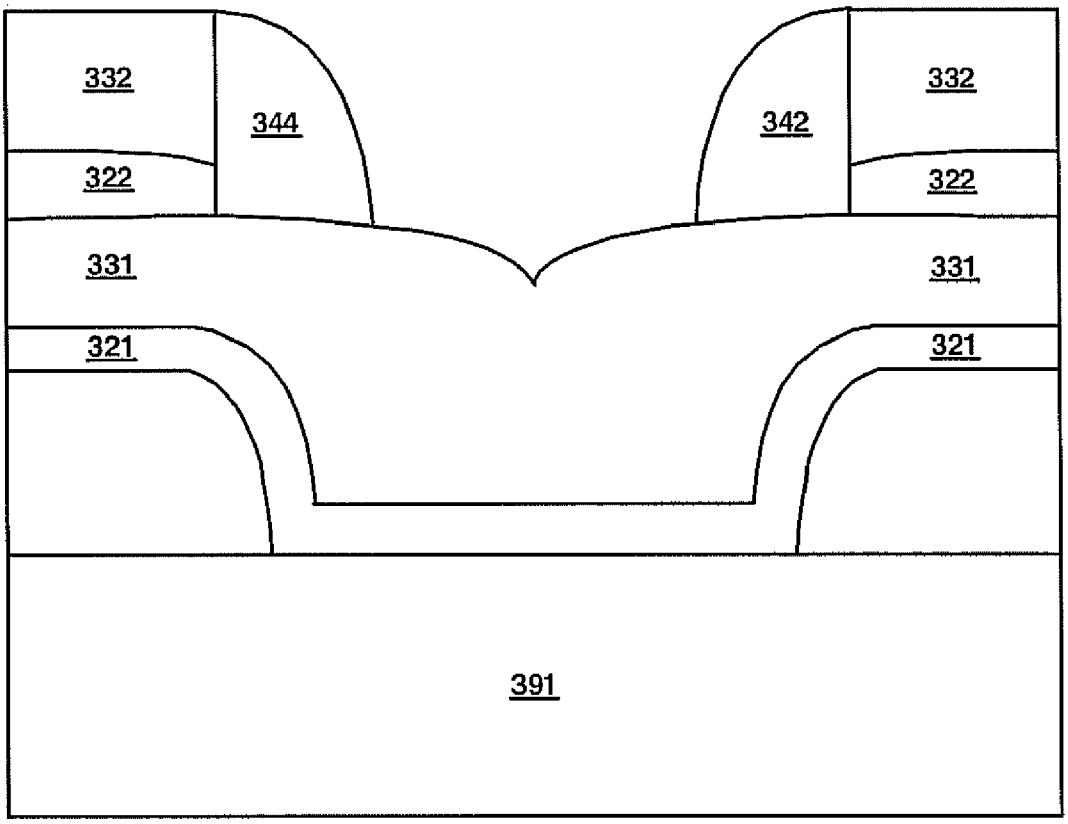


FIG. 3C

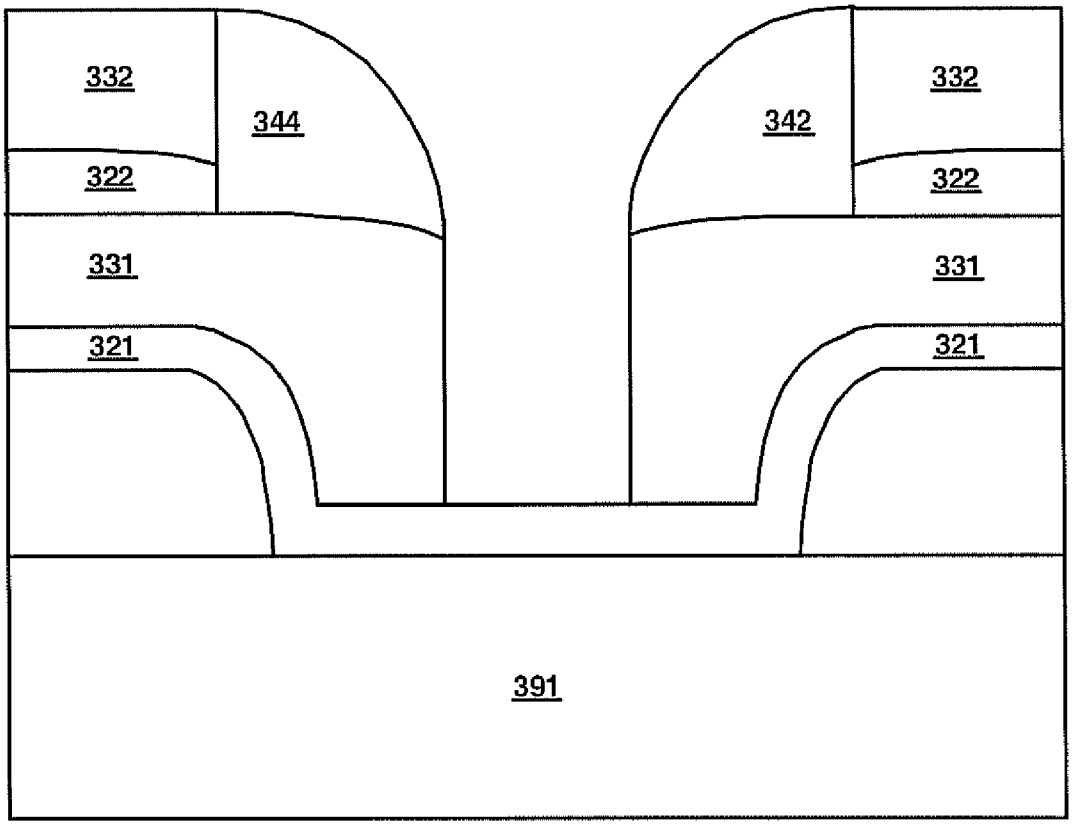


FIG. 3D

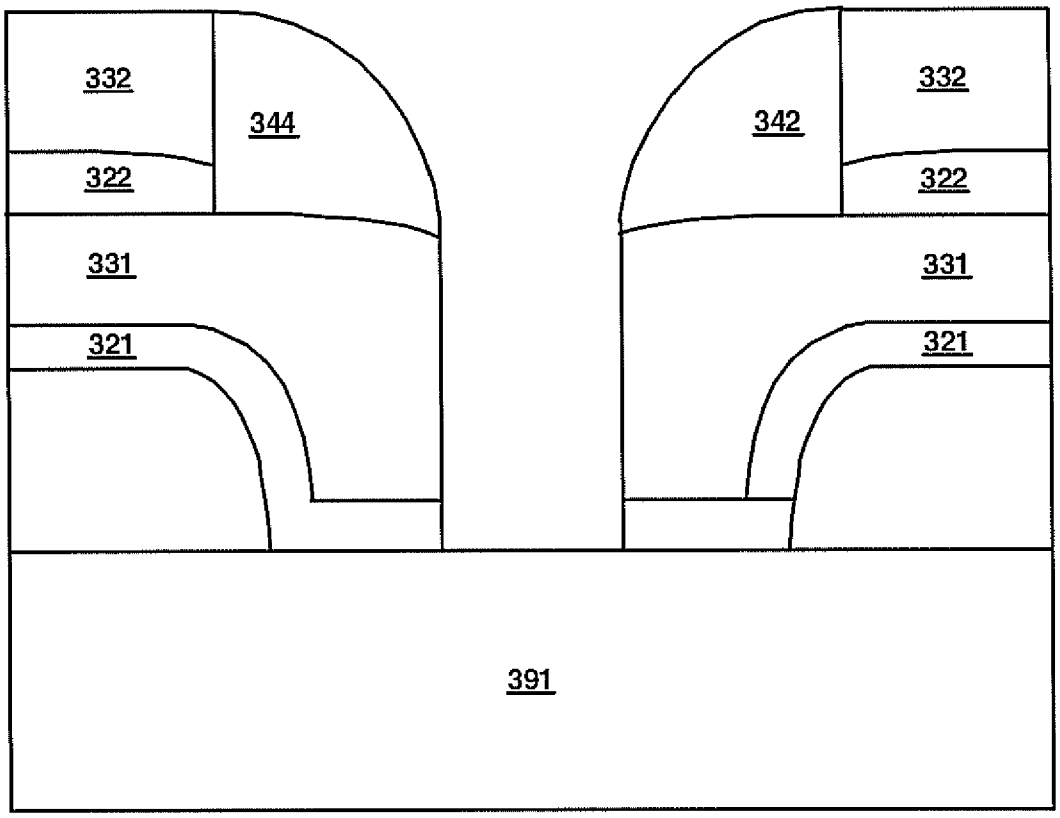


FIG. 3E

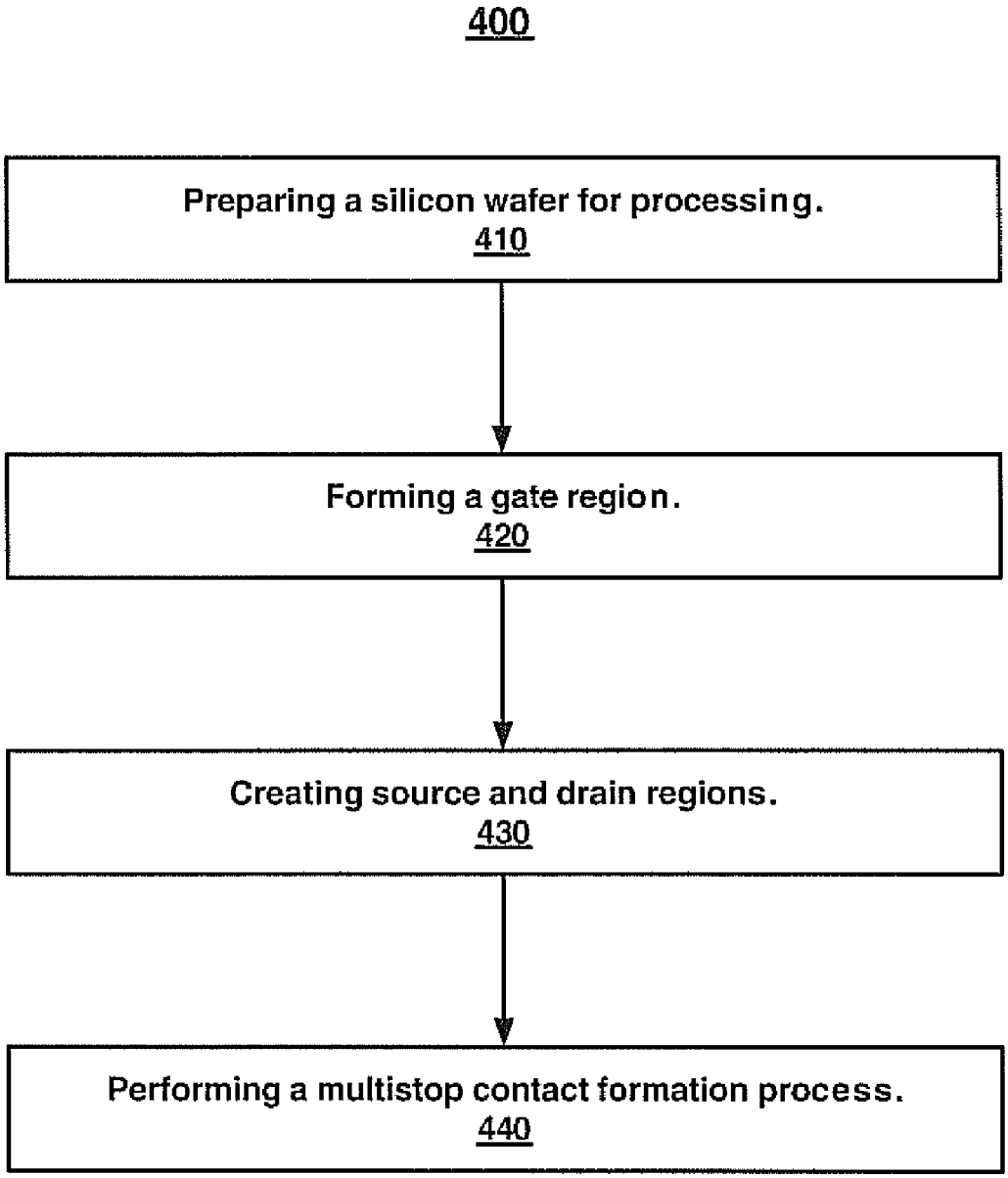

400
Preparing a silicon wafer for processing.
410
Forming a gate region.
420
Creating source and drain regions.
430
Performing a multistop contact formation process.
440

FIG. 4

INTEGRATED CIRCUIT WITH CONTACT REGION AND MULTIPLE ETCH STOP INSULATION LAYER

RELATED APPLICATIONS

This application is a divisional of and claims the benefit of application Ser. No. 10/934,828 (now U.S. Pat. No. 7,572,727) entitled A SEMICONDUCTOR CONTACT FORMATION SYSTEM AND METHOD, filed Sep. 2, 2004, which is incorporated herein by reference.

TECHNICAL FIELD

The present claimed invention relates to the field of semiconductor contact fabrication. More particularly, the present invention relates to a semiconductor contact fabrication system and method that utilizes multiple etch stop layers.

BACKGROUND ART

Electronic systems and circuits have made a significant contribution towards the advancement of modern society and are utilized in a number of applications to achieve advantageous results. Numerous electronic technologies such as digital computers, calculators, audio devices, video equipment, and telephone systems have facilitated increased productivity and reduced costs in analyzing and communicating data, ideas and trends in most areas of business, science, education and entertainment. Frequently, electronic systems designed to provide these results include integrated circuits. Integrated circuits typically include contact regions for conducting electricity (e.g., between active components) and it is often very difficult to achieve optimized results within requisite narrow tolerances when attempting to fabricate precise contact regions that operate properly.

Semiconductor integrated circuit manufacturing efforts are usually complicated by ever increasing demands for greater functionality. More complicated circuits are usually required to satisfy the demand for greater functionality. For example, there is usually a proportional relationship between the number of components included in an integrated circuit and the functionality. Integrated circuits with more components typically provide greater functionality. However, including more components within an integrated circuit often requires the components to be densely packed in relatively small areas and reliably packing a lot of components in relatively small areas of an integrated circuit (IC) is usually very difficult.

One traditional focus for achieving greater densities has been directed towards reducing the size of individual components (e.g., transistors). The components of an integrated circuit are usually fabricated on a single silicon substrate and maintaining both the integrity of the system as a whole as well as the individual basic device characteristics is very important for proper operation. Proper relational characteristics are very helpful in achieving these objectives and without them there is a tendency for detrimental interactions to occur. Thus, it is important for integrated circuit fabrication technologies to provide an advantageous balance between component integrity and increased component density.

Semiconductor contact formation processes usually include the creation of a contact void for deposition of the contact layer. The contact void creation typically determines the contact configuration. The smaller the void the more compact the contact and the greater the possible component density. However, decreases in contact sizes are usually limited by contact void creation processes (such as lithographic etching processes). Standard lithographic etching and removal processes traditionally have difficulty producing relatively small contact voids. Complex processes that attempt to create smaller voids are often cost prohibitive or nonfeasible.

While decreasing the size of a contact usually permits greater component densities, there are usually physical limitations on how small the contact can become and still operate properly. It is important for contacts to be formed in a manner that ensures proper operation without defects. Interconnection phenomenon such as electromigration can cause reliability problems as the dimension of the contact becomes very small. For example, electromigration can cause discontinuities in conducting materials if the dimensions are too small. Thus, most conducting materials have a critical dimension (CD) that limits how small a contact can be and still operate reliably. Fabrication of small contacts with desirable CD characteristics can be challenging.

It is also important to maintain adequate insulation around the contacts. Without proper component insulation there is a tendency for detrimental interactions between component parts to occur that hinder proper and reliable operation. For example, placement of more components in smaller spaces by reducing the separation between adjacent component parts often increases the probabilities of failures associated with leakage currents. It is also desirable for integrated circuit component formation processes to be efficient and low cost. While introduction of complex and complicated lithographic techniques may attempt to provide small size components, these advance techniques usually consume significant resources and are very expensive. Standard lithographic techniques are usually more efficient and do not require extensive retooling efforts. Therefore, the ability to precisely form semiconductor contact regions in a convenient and efficient manner is often very important.

SUMMARY OF THE INVENTION

The present invention is a semiconductor contact formation system and method. In one embodiment, a contact formation process forms contact insulation regions comprising multiple etch stop sublayers that facilitate formation of contacts. This contact formation process provides relatively small substrate connections while addressing critical dimension concerns in coupling to metal layers. The integrated circuit formation process also facilitates the creation of compact high density components (e.g., flash memory cells) that operate reliably. In one embodiment, a multiple etch stop contact formation process is implemented in which a multiple etch stop insulation layer comprising multiple etch stop layers is deposited. A contact region is formed in the multiple etch stop insulation layer by selectively removing (e.g., etching) some of the multiple etch stop insulation layer in which a larger portion of the multiple etch stop insulation layer is removed close to the metal layer and a smaller portion is removed closer to the substrate. In one exemplary implementation, the different contact region widths are achieved by performing multiple etching processes controlled by the multiple etch stop layers in the multiple etch stop insulation layer. Electrical conducting material (e.g., tungsten) is deposited in the contact region.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1**A is a block diagram of an integrated circuit with a substrate contact in accordance with one embodiment of the present invention.

FIG. **1**B is a block diagram showing a multiple etch stop insulation layer in accordance with one embodiment-of the present invention.

FIG. **2**A is a flow chart of a contact formation process in accordance with one embodiment of the present invention.

FIG. **2**B is a flow chart of one embodiment of a present invention multiple etch stop insulation layer deposition process.

FIG. **2**C is a flow chart of a multiple etch stop contact formation process of yet another embodiment of the present invention.

FIG. **3**A is an illustration of one embodiment of a wafer after performing a multiple etch stop insulation layer deposition process.

FIG. **3**B is an illustration of a wafer after a portion of a sub interlevel dielectric layer is etched away in accordance with one embodiment of the present invention.

FIG. **3**C is an illustration of one embodiment of a wafer after a portion of etch stop layer is removed and a spacer is formed in accordance with the present invention.

FIG. **3**D is an illustration of one embodiment of a wafer after another portion of sub interlevel dielectric layer is etched away in one exemplary embodiment.

FIG. **3**E is an illustration of one embodiment of a wafer after a substrate protective etch stop layer is removed.

FIG. **4** is a flow chart of an integrated circuit method including a contact formation process in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Reference will now be made in detail to the preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings. While the invention will be described in conjunction with the preferred embodiments, it will be understood that they are not intended to limit the invention to these embodiments. On the contrary, the invention is intended to cover alternatives, modifications and equivalents, which may be included within the spirit and scope of the invention as defined by the appended claims. Furthermore, in the following detailed description of the present invention, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be obvious to one ordinarily skilled in the art that the present invention may be practiced without these specific details. In other instances, well known methods, procedures, components, and circuits have not been described in detail as not to unnecessarily obscure aspects of the current invention.

A present invention contact system and method includes contact insulation regions comprising multiple etch stop sublayers. In one embodiment of the present invention, the contact regions can be characterized by a relatively small substrate coupling area and relatively large metal layer coupling area. The relatively small substrate coupling area permits multiple active regions of an integrated circuit to be arranged relatively close to one another to achieve higher circuit density.

FIG. **1**A is a block diagram of integrated circuit **100**A, an integrated circuit with a substrate contact in accordance with one embodiment of the present invention. Integrated circuit comprises substrate **191**, a multiple etch stop insulation layer **115**, and a contact region **171**. In one exemplary implementation of the present invention multiple etch stop insulation layer **115** is an interlevel dielectric layer. Substrate **191** provides an electrical well for integrated circuit **100**. Contact region **171** is coupled to substrate **191** and provides an electrical path to and from substrate **191**. Multiple etch stop insulation layer **115** is coupled to contact region **171** and comprises a plurality of sublayers including multiple etch stop layers. Multiple etch stop insulation layer provides electrical insulation between other regions of integrated circuit **100** (e.g., not shown) and isolates guidance of electrical current flow in contact region **171**.

FIG. **1**B is a block diagram of integrated circuit **100**B, one embodiment of integrated circuit **100**A in accordance with the present invention. In integrated circuit **100**B multiple etch stop insulation layer **115** comprises etch stop layer **121**, sub interlevel dielectric layer **131**, etch stop layer **122**, sub interlevel dielectric layer **132** and spacer regions **141** through **144**. In one exemplary implementation, integrated circuit **100**B includes a gate region **111** and gate region **112** and a source or drain region (e.g., **151**). Gate region **111** is coupled to substrate **191**, spacer region **143** and etch stop layer **121** which is coupled to sub interlevel dielectric layer **131**. Sub interlevel dielectric layer **131** is coupled to spacer region **144** and etch stop layer **122** which in turn is coupled to sub interlevel dielectric layer **132**. Contact region **171** is coupled to substrate **191**, sub interlevel dielectric layer **131** and spacer region **144**.

The components of device **100**B cooperatively operate to provide an active device. Gate regions **111** and **112** control the flow of electricity between source and drain regions (e.g., region **151** and another similar region not shown). Contact region **171** conducts electrical current flow to and/or from regions of substrate **191**. For example, contact region **171** can conduct electrical current flow to and/or from a source or drain region (e.g., **151**). Sub interlevel dielectric layer **131**, sub interlevel dielectric layer **132** and spacer regions **141** through **144** provide electrical insulation between contact region **171** other regions of device **100** (e.g., gates **111** and **112**) and isolate guidance of electrical current flow in contact region **171**. Etch stop layer **121** and etch stop layer **122** facilitate definition of the configuration of contact region **171**.

In one embodiment of device **100**B, contact region **171** has a relatively small substrate coupling area and relatively large metal layer coupling area (e.g., coupled to metal layer **175**). For example, the substrate coupling area width can be controlled by spacer width **142** and **144** to the range of 0.06 μm to 0.13 μm and the metal layer coupling area width can be 0.16 μm to 0.3 μm. The relatively small substrate coupling area permits multiple active regions (e.g., gate regions **111** and **112**) to be arranged relatively close to one another while the relatively large metal layer coupling areas of device facilitate avoidance of critical dimension (CD) issues.

FIG. **2**A is a flow chart of contact formation process **200**, one embodiment of the present invention. Contact formation process **200** facilitates the fabrication of contact regions with a relatively small substrate coupling area and relatively large metal layer coupling area. The relatively small substrate coupling area permits multiple active regions in an integrated circuit to be arranged relatively close to one another while the relatively large metal layer coupling area facilitate avoidance of CD issues.

At step **210** a multiple etch stop insulation layer comprising multiple etch stop layers is deposited. In one embodiment of the present invention, the multiple etch stop insulation layer (e.g., **115**) is an interlevel dielectric layer. The multiple etch stop insulation layer permits etch flexibility with the combination of spacer formation **142** and **144** to create a small contact formation. In one exemplary implementation, the multiple etch stop insulation layer is made smooth and level (e.g., polished by a CMP process). In one embodiment

of the present invention, step **210** includes a multiple etch stop insulation layer deposition process.

FIG. **2**B is a flow chart of one embodiment of multiple etch stop insulation layer deposition process **210** and FIG. **3**A is an illustration of one embodiment of a wafer after performing multiple etch stop insulation layer deposition process **210**. In step **211** etch stop layer **321** is deposited (e.g., 200 to 500 Å thick). At step **212** sub interlevel dielectric layer **331** is deposited (1,000 to 2000 Å). Etch stop layer **322** (e.g., 200 to 500 Å thick) is deposited at step **213**. In step **214**, another sub interlevel dielectric layer **332** is deposited (e.g., 10K Å±1K Å). In one exemplary implementation the etch stop layers are nitride or SiON and the sub interlevel dielectric layers are oxide. In one embodiment an ARC layer is deposited (e.g., a ARC film with Si_3N_4, SiON, or Si_3N_4/OX).

Referring to FIG. **2**A again, a contact region is created in the multiple etch stop insulation layer at step **220**. In one embodiment of the present invention, a multiple stop etch process creates a contact region in the multiple etch stop insulation layer. In one exemplary implementation, the multiple etch stop insulation layers creates etching processes with non-lithographic spacer formation using multiple steps so that smaller contact holes (e.g., 0.06 μm to 0.13 μm) can be created close to the substrate coupling area and larger contact top (e.g., 0.16 μm to 0.18 μm) can be created closer to a metal layer coupling area. In one embodiment the contact region is created by a multiple etch stop contact formation process.

FIG. **2**C is a flow chart of multiple etch stop contact formation process **220**A in accordance with one embodiment of the present invention. In step **221** a portion (e.g., 0.16 μm to 0.18 μm width region) of sub interlevel dielectric layer **332** is etched away. FIG. **3**B is an illustration of one embodiment of a wafer after a portion of sub interlevel dielectric layer **332** is etched away in step **221**. A portion (e.g., 0.16 μm to 0.18 μm width region) of etch stop layer **322** is removed in step **222**. In step **223** sub spacer regions **342** and **344** are formed. FIG. **3**C is an illustration of one embodiment of a wafer after a portion of etch stop layer **322** is removed in step **222** and sub spacer regions **342** and **344** are formed in step **223**. In step **224** a portion (e.g., 0.06 μm to 0.13 μm width region) of sub interlevel dielectric layer **331** is etched. FIG. **3**D is an illustration of one embodiment of a wafer after a portion of sub interlevel dielectric layer **331** is etched away in step **224**. A portion (e.g., 0.06 μm to 0.13 μm width region) of etch stop layer **321** is removed in step **226**. FIG. **3**E is an illustration of one embodiment of a wafer after step **226** is performed. In step **227** a conductive material is deposited in the void left after said etching and removing of the portions of the first sub interlevel dielectric layer, the first etch stop layer, the second sub interlevel dielectric layer, and the second etch stop layer.

Referring now to step **230** shown in FIG. **2**A, a conducting material (e.g., tungsten) is deposited in the contact region. In one exemplary implementation, the conducting material is part of a metal layer deposition. In one embodiment of the present invention, a plurality of metal layers are deposited and each of the respective metal layers are separated by insulating layers. The metal layers can selectively couple integrated circuit components formed on the wafer to each other and external components.

In one embodiment, a present invention contact formation process is included in an integrated circuit fabrication process. In one exemplary implementation, a present invention contact formation process is utilized to provide electrical contacts to a source and drain region. For example, contacts are formed to couple a flash memory cell source and drain region to word and bit lines. FIG. **4** is a flow chart of integrated circuit formation process **400**, including a contact formation process in accordance with one embodiment of the present invention.

In step **410**, a silicon wafer substrate is prepared for processing. In one embodiment of the present invention, the wafer surface is made smooth and level, for example by chemical mechanical polishing (CMP). An oxide pad layer and a subsequent protective layer of nitride are deposited on the surface. In one exemplary implementation, additional polishing is performed to provide a smooth and level surface after the protective oxide and nitride layers are added.

In step **420** a gate region is formed. In one embodiment of the present invention, forming a gate region comprises depositing a gate insulation layer, depositing a control gate layer, and removing the gate insulation layer and the control gate layer from non gate region areas. In one embodiment of the present invention, a floating gate is formed in step **420**. An insulating layer (e.g., oxide) is deposited and a floating gate area is created in the insulating layer. For example, a floating gate area is etched in the insulating layer and a charge trapping material (e.g., a polysilicide) is deposited in the floating gate area. Excess charge trapping material is removed and additional insulating material deposited. A control gate material (e.g., polysilicon) is deposited on top of the insulating material. The materials deposited during the gate formation process are removed (e.g., etched) from areas not included in the gate (e.g., areas above a source and drain). In one exemplary implementation, a sidewall spacer material is deposited on the sides of the gate area and excess sidewall spacer material is removed.

Source and drain regions are created in step **430**. In one embodiment of the present invention, a source and drain formation process is performed. The source and drain area are prepared for implantation and diffusion. For example, excess material from the gate formation process and the protective layer materials over the source and drain areas are removed. In one exemplary implementation, dopants (e.g., arsenic, phosphorus, boron, etc.,) are introduced into the substrate in the source and drain regions by implantation and/or diffusion. In one embodiment, the source and drain include lightly doped region, extensions and/or halos.

In step **440**, a multiple etch stop contact formation process (e.g., contact formation process **200**) is performed. A multiple etch stop insulation layer comprising multiple etch stop layers is deposited. A contact region is formed in the multiple etch stop insulation layer by selectively removing (e.g., etching) some of the multiple etch stop insulation layer. A substrate coupling area of the contact region is smaller than a metal layer coupling area of the contact region. Electrical conducting material (e.g., tungsten) is deposited in the contact region.

Thus, the present invention facilitates precise formation of semiconductor contact regions in a convenient and efficient manner. Utilization of a multi etch stop process in formation of the semiconductor contact regions enables simple lithographic processes to provide contacts with relatively small substrate patterns and relatively large metal layer coupling patterns. This contact formation process provides contacts with small CD features. The shallow junctions also enable reductions in space between integrated circuit components permitting realization of increased component density (e.g., a larger number of components concentrated in smaller areas).

The foregoing descriptions of specific embodiments of the present invention have been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise forms disclosed, and obviously many modifications and variations are possible

in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application, to thereby enable others skilled in the art to best utilize the invention and various embodiments with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the Claims appended hereto and their equivalents.

What is claimed is:

1. An integrated circuit comprising

a substrate for providing an electrical well for said integrated circuit;

a contact region coupled to said substrate, said contact region provides an electrical path to and from said substrate;

a spacer region such that a substrate coupling area of said contact region is smaller than a metal layer coupling area of said contact region; and

a multiple etch stop insulation layer coupled to said contact region, wherein said multiple etch stop insulation layer provides electrical insulation between other regions of said integrated circuit and isolates guidance of electrical current flow in said contact region, and wherein said multiple etch stop insulation layer includes:

a first etch stop layer directly on said substrate in said contact region;

a first sub interlevel dielectric layer over said first etch stop layer;

a second etch stop layer over said first sub interlevel dielectric layer, wherein said second etch stop layer has similar selectivity characteristics as said first etch stop layer and wherein said first sub interlevel dielectric layer is between said first etch stop layer and said second etch stop layer; and

a second sub interlevel dielectric layer over said second etch stop layer.

2. An integrated circuit of claim **1** wherein said first etch stop layer protects removal of a substrate material by an etch process.

3. An integrated circuit of claim **1** wherein said first sub interlevel dielectric layer insulates said contact region.

4. An integrated circuit of claim **1** wherein said second etch stop layer protects lower layers during an etching process.

5. An integrated circuit of claim **1** wherein said second sub interlevel dielectric layer insulates said contact region.

6. An integrated circuit of claim **1** wherein said spacer region insulates said contact region.

7. An integrated circuit of claim **1** wherein said substrate coupling area of said contact region permits multiple active regions to be arranged relatively close to one another while a metal layer coupling area of said contact region is larger than said substrate coupling area of said contact region to provide better process window with desired critical dimension (CD).

8. An integrated circuit of claim **1** wherein said first etch stop layer is in a range of about 300 to 800 Å thick, said first sub interlevel dielectric layer is in a range of about 1,000 to 2,000 Å thick, said second etch stop layer is in a range of about 300 to 800 Å thick and said second sub interlevel dielectric layer is in a range of about 10K Å±1K Å thick.

9. An integrated circuit of claim **1** wherein an area of a first space in a portion of said first sub interlevel dielectric layer for said contact region is in a first range of about 0.06 μm to 0.13 μm wide and an area of a second space in said second sub interlevel dielectric layer for said contact region is in a second range of about 0.16 μm to 0.18 μm wide.

10. An integrated circuit of claim **1** further comprising depositing an anti reflective coating layer.

11. An integrated circuit comprising:

a multiple etch stop insulation layer comprising a first etch stop layer and a second etch stop layer wherein said first etch stop layer and said second etch stop layers have similar selectivity characteristics, said first etch stop layer formed in an area directly next to a substrate corresponding to a contact region and under a first sub interlevel dielectric layer and said second etch stop layer formed under a second interlevel dielectric layer, wherein said first sub interlevel dielectric layer is between said first etch stop layer and said second etch stop layer, wherein said multiple etch stop insulation layer is formed utilizing a lithography process;

a contact region in said multiple etch stop insulation layer, wherein creating of said contact region includes forming sub-spacer regions in removed portions of said second sub interlevel dielectric layer and said second etch stop layer, wherein a non-lithography spacer formation process is also utilized to achieve a contact bottom, such that a substrate coupling area of said contact region is smaller than a metal layer coupling area of said contact region.

12. An integrated circuit of claim **11** wherein said first etch stop layer is in a range of about 300 to 800 Å thick, said first sub interlevel dielectric layer is in a range of about 1,000 to 2,000 Å thick, said second etch stop layer is in a range of about 300 to 800 Å thick and said second sub interlevel dielectric layer is in a range of about 10K Å±1K Å thick.

13. An integrated circuit of claim **11** further comprising a floating gate.

14. An integrated circuit of claim **11** further comprising depositing an anti reflective coating layer.

15. An integrated circuit of claim **11** wherein a bottom of said contact region is a first range of about 0.06 μm to 0.13 μm wide and a top of said contact region is in a second range of about 0.16 μm to 0.18 μm wide.

* * * * *