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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 June 2019 (27.06.2019)

(10) International Publication Number
WO 2019/126769 A1

(51) International Patent Classification:
*H01L 23/525* (2006.01) *H01L 23/538* (2006.01)

(21) International Application Number:
PCT/US2018/067322

(22) International Filing Date:
21 December 2018 (21.12.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/609,891 22 December 2017 (22.12.2017) US

(71) Applicant: **BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM** [US/US]; 210 West 7th Street, Austin, TX 78701 (US).

(72) Inventors: **SREENIVASAN, Sidlgata, V.**; The University Of Texas At Austin, 10100 Burnet Road, Bldg. 160, Austin, TX 78758 (US). **AJAY, Paras**; The University Of Texas At Austin, 10100 Burnet Road, Bldg. 160, Austin, TX 78758 (US). **SAYAL, Aseem**; The University Of Texas At Austin, 10100 Burnet Road, Bldg. 160, Austin, TX 78758 (US). **ABED, Ovadia**; The University Of Texas At Austin, 10100 Burnet Road, Bldg. 160, Austin, TX 78758 (US). **MCDERMOTT, Mark**; The University Of Texas At Austin, 10100 Burnet Road, Bldg. 160, Austin, TX 78758 (US). **KULKARNI, Jaydeep**; The University Of Texas At Austin, 10100 Burnet Road, Bldg. 160, Austin, TX 78758 (US). **SINGHAL, Shrawan**; The University Of Texas At Austin, 10100 Burnet Road, Bldg. 160, Austin, TX 78758 (US).

(74) Agent: **VOIGT, Robert, A.**; Winstead, P.C., P.O. Box 131851, Dallas, TX 75313 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

(54) Title: NANOSCALE-ALIGNED THREE-DIMENSIONAL STACKED INTEGRATED CIRCUIT

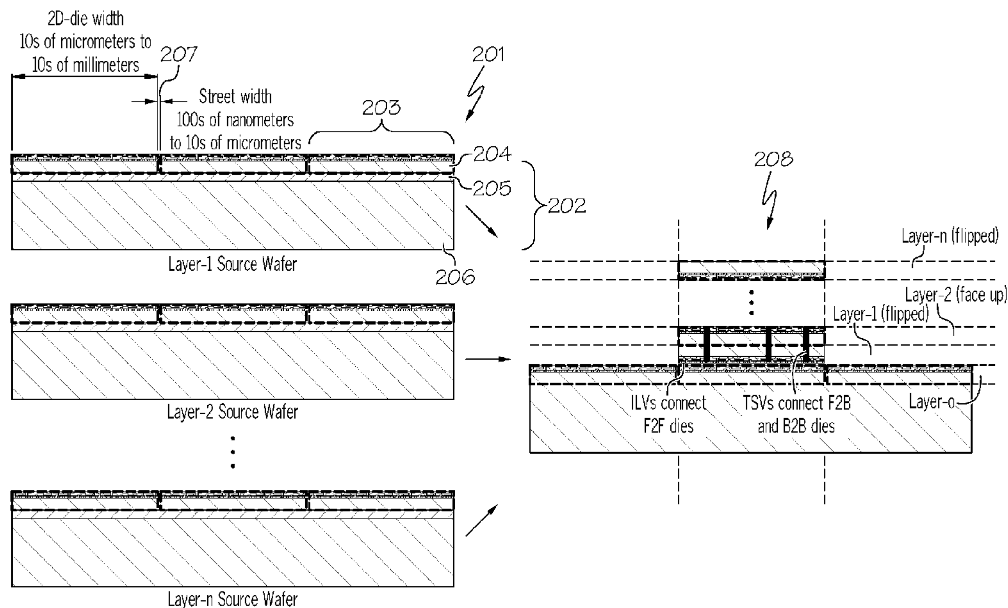



FIG. 2

(57) Abstract: A method for fabricating a three-dimensional (3D) stacked integrated circuit. Pick-and-place strategies are used to stack the source wafers with device layers fabricated using standard two-dimensional (2D) semiconductor fabrication technologies. The source wafers may be stacked in either a sequential or parallel fashion. The stacking may be in a face-to-face, face-to-back, back-to-face or back-to-back fashion. The source wafers that are stacked in a face-to-back, back-to-face or back-to-back fashion may be connected using Through Silicon Vias (TSVs). Alternatively, source wafers that are stacked in a face-to-face fashion may be connected using Inter Layer Vias (ILVs).

WO 2019/126769 A1

[Continued on next page]

KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

**Published:**

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

# NANOSCALE-ALIGNED THREE-DIMENSIONAL STACKED INTEGRATED CIRCUIT

## CROSS REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims priority to U.S. Provisional Patent Application Serial No. 62/609,891, entitled "Nanoscale-Aligned 3D Stacked Integrated Circuit," filed on December 22, 2017, which is incorporated by reference herein in its entirety.

## TECHNICAL FIELD

**[0002]** The present invention relates generally to semiconductor fabrication, and more particularly to a nanoscale-aligned three-dimensional (3D) stacked integrated circuit.

## BACKGROUND

**[0003]** Moore's law is the observation that the number of transistors in a dense integrated circuit doubles about every two years. Two-dimensional (2D) scaling of electronic circuits, as characterized by Moore's law, may now have reached a limit in recent times as feature dimensions have reached atomic scales. For instance, the thickness of high-K capping layers for 10 nm technology nodes is close to 0.5 nm, which is less than the width of two silicon atoms. The metrology precision requirements for multi-patterning technologies (MPT) are close to 0.2 nm which is less than the width of one silicon atom.

**[0004]** In light of these and other limitations, 2D scaling and general top down fabrication have significant challenges in continuing at and beyond the 7 nm node.

SUMMARY

**[0005]** In one embodiment of the present invention, a method for fabricating a three-dimensional (3D) system on a chip (SoC) comprises assembling a Layer-(k) two-dimensional (2D)-die array onto a Layer-(k-1) 2D-die array of a Layer-(k-1) wafer, where the Layer-(k-1) wafer is populated with 2D-dies, and k is a positive integer number greater than 1. The 2D-die array comprises a single 2D-die, a single island of 2D-die that forms a contiguous group of 2D die or multiple islands of 2D die. The method further comprises deploying a fluid allowing lubricated relative motion between the Layer-(k) 2D-die array and the Layer-(k-1) 2D-die array, where the fluid allows precision overlay of the Layer-(k) and Layer-(k-1) 2D-die arrays.

**[0006]** In another embodiment of the present invention, a method for fabricating a three-dimensional (3D) system on a chip (SoC) comprises assembling a Layer-(k) two-dimensional (2D)-die array onto a Layer-(k-1) 2D-die array of a Layer-(k-1) wafer, where the Layer-(k-1) wafer is populated with 2D-dies, and where k is a positive integer number greater than 1. The 2D-die array comprises a single 2D-die, a single island of 2D-die that forms a contiguous group of 2D die or multiple islands of 2D die. The method further comprises providing an encapsulation layer for protecting 2D-dies in each of Layer-(k) wafer and the Layer-(k-1) wafer from etchants used during a pick-and-place process.

**[0007]** In another embodiment of the present invention, a method for fabricating a three-dimensional (3D) system on a chip (SoC) comprises assembling a Layer-(k) two-dimensional (2D)-die array onto a Layer-(k-1) 2D-die array of a Layer-(k-1) wafer, where the Layer-(k-1) wafer is populated with 2D-dies and where k is a positive integer number greater than 1. Furthermore, the 2D-die array comprises a single 2D-die, a single island of 2D-die that forms a contiguous group of 2D die or multiple islands of 2D die. Additionally, the 2D-dies have a thickness less than 10 micrometers.

**[0008]** In a further embodiment of the present invention, an electronic design automation (EDA) methodology for designing three-dimensional (3D) application specific integrated circuit (ASIC) system on a chip (SoC) logic circuitry comprises a combination of software integrated with two-dimensional (2D) EDA solutions, where the software comprises a netlist partitioning algorithm to partition a 3D design netlist into 2D modules. The 2D EDA solutions are used to perform one

or more of the following: synthesis, 3D placement aware synthesis, placement, clock tree synthesis (CTS), routing, design verification and signoff analysis.

**[0009]** The foregoing has outlined rather generally the features and technical advantages of one or more embodiments of the present invention in order that the detailed description of the present invention that follows may be better understood. Additional features and advantages of the present invention will be described hereinafter which may form the subject of the claims of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

**[0010]** A better understanding of the present invention can be obtained when the following detailed description is considered in conjunction with the following drawings, in which:

**[0011]** Figure 1 illustrates an exemplary Layer-k source wafer showing various 2D-die arrangements in accordance with an embodiment of the present invention;

**[0012]** Figure 2 illustrates stacking of Layer-k 2D-die arrays (k>1) onto a Layer-1 2D-die array in accordance with an embodiment of the present invention;

**[0013]** Figures 3A-3B illustrate a cross-section of a Layer-k Silicon-On-Insulator (SOI) wafer with two buried layers in accordance with an embodiment of the present invention;

**[0014]** Figures 4A-4B illustrate another cross-section of a Layer-k SOI wafer in accordance with an embodiment of the present invention;

**[0015]** Figures 5A-5B illustrates a further cross-section of a Layer-k SOI wafer in accordance with an embodiment of the present invention;

**[0016]** Figure 6 is a flowchart of a method for the back-grinding based approach for flipping and bulk-material removal in accordance with an embodiment of the present invention;

**[0017]** Figures 7A-7D depict the cross-sectional views for flipping and bulk-material removal using the steps described in Figure 6 in accordance with an embodiment of the present invention;

**[0018]** Figure 8 is a flowchart of a method for the peel-off based approach for flipping and bulk-material removal in accordance with an embodiment of the present invention;

**[0019]** Figures 9A-9E depict the cross-sectional views for flipping and bulk-material removal using the steps described in Figure 8 in accordance with an embodiment of the present invention;

**[0020]** Figure 10 is a flowchart of a method for overlay and distortion control of multiple packed 2D-dies in accordance with an embodiment of the present invention;

**[0021]** Figures 11A-11B depict the cross-sectional views for providing overlay and distortion control of multiple packed 2D-dies using the steps described in Figure 10 in accordance with an embodiment of the present invention;

**[0022]** Figures 12A-12B illustrate the overlay and distortion control of a single picked 2D-die in accordance with an embodiment of the present invention;

**[0023]** Figure 13 illustrates that Through Silicon Vias (TSVs) are made through the access holes already present in picked-and-placed 2D-dies in accordance with an embodiment of the present invention;

**[0024]** Figures 14A-14C illustrate an exemplary process for temporary attachment and bonding in accordance with an embodiment of the present invention;

**[0025]** Figure 15 illustrates the conventional 2D Application-Specific Integrated Circuit (ASIC) Electronic Design Automation (EDA) flow for logic implementation in accordance with an embodiment of the present invention;

**[0026]** Figure 16 illustrates the 2D single sided static random access memory (SRAM) configuration in accordance with an embodiment of the present invention;

**[0027]** Figure 17 illustrates 3D stand-alone SRAM die stacking in accordance with an embodiment of the present invention;

**[0028]** Figure 18 illustrates a 3D only-bitcell stacked SRAM in accordance with an embodiment of the present invention; and

**[0029]** Figure 19 illustrates the vertical bit-line cross section for the 3D only-bitcell stacked SRAM in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

**[0030]** As stated in the Background section, two-dimensional (2D) scaling and general top down fabrication have significant challenges in continuing at and beyond the 7 nm node.

**[0031]** Embodiments of the present invention address such challenges by scaling in the third ($3^{rd}$) dimension as discussed below.

**[0032]** In one embodiment, the present invention uses source wafers with device layers that were fabricated using standard 2D semiconductor fabrication processes (discussed below in connection with Figure 1) as well as uses the pick-and-place strategies to stack them (source wafers) in a sequential or parallel fashion. Such pick-and-place strategies are discussed in Sreenivasan et al. (WO 2018/119451 A1) (hereinafter referred to as "Sreenivasan et al."), which is hereby incorporated by reference in its entirety. In one embodiment, stacking occurs in a face-to-face (F2F), face-to-back (F2B), back-to-face (B2F) or back-to-back (B2B) fashion. B2F, F2B and B2B can be connected, for example, using Through Silicon Vias (TSVs). F2F can be connected using Inter Layer Vias (ILVs).

**[0033]** A discussion regarding standard semiconductor processes is now deemed appropriate.

**[0034]** A "Layer-0 source wafer," as used herein, refers to a fully populated wafer consisting of transistors and interconnects fabricated using standard 2D fabrication processes. This layer also includes relevant alignment marks and forms the starting layer for the final wafer-scale three-dimensional (3D)-integrated circuit (IC) stack.

**[0035]** A "Layer-k source wafer," as used herein, refers to a fully populated wafer consisting of transistors and interconnects fabricated using standard 2D fabrication processes on a wafer that includes at least one sacrificial layer, such as a buried oxide underneath silicon. This layer also includes relevant alignment marks and is assembled onto layer "k-1" and is part of a 3D-IC stack. The assembly of this layer may be in one step (all 2D die are picked up at once) or in multiple steps where a single 2D-die-array or multiple 2D-die-arrays are picked up from layer "k" wafer and precisely placed onto the layer "k-1" wafer.

**[0036]** In one embodiment, the assembly is performed to achieve sub-50 nm, sub-30 nm, sub-20 nm, sub-10nm or even sub-5nm overlay between each 2D-die of the Layer-(k) wafer and the corresponding 2D-die of the Layer-(k-1) wafer.

**[0037]** Referring to Figure 1, Figure 1 illustrates an exemplary Layer-k source wafer 100 showing various 2D-die arrangements in accordance with an embodiment of the present invention.

**[0038]** Referring to Figure 1, Layer-k source wafer 100 includes a 2D-die array 101 which is a single 2D-die, a 2D-die array 102 which is a contiguous island of 2D-dies and a 2D-die array 103 which is a group of islands.

**[0039]** A "2D-die," as used herein, refers to a single layer of a three-dimensional (3D)-System on a Chip (SoC), where the 3D-SoC includes at least two 2D-die stacked precisely in a three-dimensional arrangement. These 2D-dies are fabricated using standard 2D semiconductor fabrication processes. In one embodiment, the thickness of the 2D-dies may be less than 10 micrometers. Wafers thinned using standard wafer-thinning processes, such as back-grinding, are projected to remain above 15 μm thickness because of defects induced due to the grinding processes. 2D-dies fabricated using a non-grinding-process, however, can be fabricated with thicknesses that are significantly smaller than current thickness limits.

**[0040]** "A 2D-die array," as used herein, refers to a single 2D-die (see 2D-die array 101) or a group of 2D-die that are collectively moved from their source wafer (e.g., Layer-k) and assembled collectively and precisely onto the previous wafer (Layer-(k-1)), where $k > 1$. This 2D-die-array can include a single island of 2D die that form a contiguous group (see 2D-die array 102). Alternatively, the 2D-die-arary can include multiple islands of 2D die, where each island of 2D die forms a contiguous group but the islands are not contiguous (see 2D-die array 103).

**[0041]** "Overlay," as used herein, refers to a vector quantity defined at every point on the wafer. It is the difference between the vector position of points on a substrate geometry and the vector position of the corresponding point in an overlaying pattern. A generally accepted quantifier of overlay is the (Mean + 3*Sigma) value of said overlay vector magnitudes.

**[0042]** "Alignment," as used herein, refers to the set of rigid body errors (translation and rotation) between two overlaying bodies.

**[0043]** Referring to Figure 2, Figure 2 illustrates stacking of Layer-k 2D-die arrays ($k>1$) onto a Layer-1 2D-die array in accordance with an embodiment of the present invention.

**[0044]** As shown in Figure 2, in one embodiment, Layer-1 of the source wafer 201 corresponds to a silicon-on-insulator wafer 202 with three elements 203. In one embodiment, wafer 202 includes a layered silicon 204 – insulator (sacrificial layer) 205 – silicon 206 substrate. In one embodiment, element 203 is a "feedstock," which in its most general form, consists of layers of transistors, interconnects and dielectrics. Furthermore, in one embodiment, element 203, as used herein, may include silicon layer 204 of SOI wafer 202. It may or may not have any functionality in itself, but when assembled together with other elements 203 and possibly additional interconnect and dielectric layers, it could be used to fabricate a working ASIC. Additionally, front-end high-resolution device layers, for which mask cost is high, would reside inside element 203. This is to amortize the cost of expensive masks (for the high-resolution device layers) across the fabrication of a variety of ASIC devices.

**[0045]** In one embodiment, the width of element 203 corresponds to a 2D-die width of tens of millimeters. In one embodiment, the street width or "scribe width" may range from hundreds of nanometers to tens of micrometers. In one embodiment, such a width corresponds to the boundaries 207 of element 203.

**[0046]** Each of the layers of the source wafer shown in Figure 2, such as Layer-2…Layer-n, where n is a positive integer number, are configured similarly as Layer-1 201. As a result, each of these Layers (referred to as simply "Layer-k," where k is a positive integer number) may generally be referred to herein as element 201.

**[0047]** As shown in Figure 2, the layers of the source wafer are stacked in an interweaving fashion (flipped, face up, flipped, face up…) forming a 3D-IC stack 208, which will be discussed in greater detail below.

**[0048]** Furthermore, Figure 2 illustrates that B2F, F2B and B2B can be connected, for example, using Through Silicon Vias (TSVs), and that F2F can be connected using Inter Layer Vias (ILVs). A further description regarding such features, including the Layer-k wafer, is provided below.

**[0049]** In one embodiment, fluid is deployed allowing lubricated relative motion between the Layer-(k) two-dimensional (2D)-die array (e.g., 2D-die array 102) and the Layer-(k-1) 2D-die array (e.g., 2D-die array 102), where the fluid allows precision overlay of the Layer-(k) and Layer-(k-1) 2D-die arrays. In one embodiment, the fluid is a gas, a liquid or a combination

thereof. In one embodiment, such a combination includes disparate gas and liquid portions or portions of homogenously mixed gas and liquid.

**[0050]** In one embodiment, the first layer 2D-die arrays can be on any arbitrary substrate, but subsequent 2D-die arrays (which may be picked-and-placed) need an underlying sacrificial layer as shown in Figures 3A-3B. As a result, in one embodiment, Layer-k 2D-die may need an underlying oxide layer for optimal device functioning (for instance, Fully Depleted (FD)-SOI and Partially Depleted (PD)-SOI). This would necessitate another sacrificial layer at a deeper level for pick-and-place. In one embodiment, these are commercially available through Lapis Semiconductor®.

**[0051]** In one embodiment, the 2D-die width may range from tens of micrometers to tens of millimeters.

**[0052]** Referring now to Figures 3A-3B, Figures 3A-3B illustrate a cross-section of a Layer-k SOI wafer with two buried layers (insulator and sacrificial layers, which can be comprised of the same material, for instance silicon oxide) in accordance with an embodiment of the present invention.

**[0053]** As shown in Figure 3B, the cross-section of a Layer-k SOI wafer 201 illustrates that element 203 may consist of transistors 301, interconnects 302 and dielectrics 303. In one embodiment, element 203 further includes a layer of silicon 304. Furthermore, as discussed above, Layer-k 2D die may need an underlying oxide layer 305 for optimal device performance.

**[0054]** In one embodiment, as shown in Figure 3A, the 2D-die thickness may range from tens of nanometers to tens of micrometers.

**[0055]** Furthermore, in one embodiment, Figure 3A illustrates the boundaries 207 of elements 203.

**[0056]** Alternatively, in one embodiment, Layer-k 2D-die may not need an underlying oxide as shown in Figures 4A-4B. Figures 4A-4B illustrate another cross-section of a Layer-k SOI wafer in accordance with an embodiment of the present invention.

**[0057]** In such an embodiment, a sacrificial layer may need to reside at a deeper level than found in standard PD-SOI wafers for mechanical stability purposes. These are commercially available through multiple sources, for instance, ShinEtsu®.

**[0058]** Furthermore, in one embodiment, the sacrificial oxide (for pick-and-place) is at the same depth as used for standard PD-SOI wafers as shown in Figures 5A-5B. These are available commercially through multiple sources, for instance, Soitec®.

**[0059]** Figures 5A-5B illustrates a further cross-section of a Layer-k SOI wafer in accordance with an embodiment of the present invention.

**[0060]** As shown in Figures 5A-5B, in one embodiment, the 2D-die thickness is approximately 100 nanometers or lower.

**[0061]** A discussion regarding the process and mechanical design concepts for 3D-integrated circuits (ICs) is now deemed appropriate.

**[0062]** In one embodiment, the general applicable assembly sequence is substantially the same as described in Sreenivasan et al. (WO 2018/119451 A1) (hereinafter referred to as "Sreenivasan et al."), which is hereby incorporated by reference in its entirety. For example, the steps are as follows: 1. Etch and encapsulation; 2. Bulk-etch processes (to facilitate subsequent pick-and-place); 3. 2D-die array pickup; 4. Alignment of 2D-die array(s) to product substrate; 5. Temporary attachment and bonding; and 6. Repeat 3-5 until product wafer is fully assembled.

**[0063]** In one embodiment, the assembly sequence for a 3D-IC may require some modifications to steps 2, 4 and 5 as discussed below.

**[0064]** The bulk etch-processes to facilitate subsequent pick-and-place need some modification to account for the type of stacking being done (F2F vs F2B vs B2F vs B2B). With respect to B2F and B2B type stacking, the bulk-etch processes described in Sreenivasan et al. would suffice since the Layer-k wafer does not need to be flipped. However, for F2F and F2B type stacking approaches, in addition to bulk-etch, a wafer flipping step needs to happen. Additionally, for F2F type stacking, a stripping step is needed to selectively remove the encapsulation layer for face-to-face connectivity. This could be done in various ways, depending on the specific nature of encapsulation layers used – for instance, if the encapsulation layer is composed of $Al_2O_3$, then a timed buffered oxide etch might be used. Alternatively, if the encapsulation layer is composed of chemical vapor deposited (CVD) amorphous carbon, an oxygen plasma could be used for the stripping. Alternatively, if the encapsulation layer is composed of multiple layers, for instance $Al_2O_3$ on top of CVD amorphous carbon, then the oxygen plasma step and buffered oxide etch

could be done in sequence. In one embodiment, the encapsulation layer protects the 2D-dies in both the Layer-(k) wafer and the Layer-(k-1) wafer from etchants used during a pick-and-place process. In one embodiment, the encapsulation layer is compatible with existing semiconductor fabrication technologies, such as complementary metal–oxide–semiconductor (CMOS) and III-V semiconductors (e.g., gallium nitride, gallium arsenide). Two different techniques for flipping and bulk-material removal are discussed below in connection with Figures 6, 7A-7D, 8 and 9A-9D.

**[0065]** Figure 6 is a flowchart of a method for the back-grinding based approach for flipping and bulk-material removal in accordance with an embodiment of the present invention. Figures 7A-7D depict the cross-sectional views for flipping and bulk-material removal using the steps described in Figure 6 in accordance with an embodiment of the present invention.

**[0066]** Referring now to Figure 6, in conjunction with Figures 7A-7D, in step 601, an encapsulation layer (not shown) is stripped as shown in Figure 7A. Furthermore as illustrated in Figure 7A, access holes 701 may be used to speed up the etching process. In one embodiment, access holes 701 are used for etchants, such as hydrofluoric acid, to release the 2D-die from the wafer. In one embodiment, access holes 701 are utilized to create conductors that enable Through Silicon Vias (TSVs).

**[0067]** In step 602, Layer-k wafer 201 is flipped and attached to a glass carrier wafer 702 via a laser de-bonding adhesive 703 (commercially available) as shown in Figure 7B.

**[0068]** In step 603, back grinding of Layer-k wafer 201 is performed as shown in Figure 7C.

**[0069]** In step 604, sacrificial layer 205 is etched using an acid, such as hydrofluoric acid (HF).

**[0070]** Figure 8 is a flowchart of a method for the peel-off based approach for flipping and bulk-material removal in accordance with an embodiment of the present invention. Figures 9A-9E depict the cross-sectional views for flipping and bulk-material removal using the steps described in Figure 8 in accordance with an embodiment of the present invention.

**[0071]** Referring now to Figure 8, in conjunction with Figures 9A-9D, in step 801, a timed HF etch is performed on sacrificial layer 205 in such a manner as to form pyramidal pillars (tethers) 901 as shown in Figures 9A and 9B. These pyramidal tethers 901, as will be discussed later, can

facilitate the pick-and-place step. Furthermore, as shown in Figure 9A, access holes 701 may be used to speed up the etching process.

**[0072]** In step 802, the encapsulation layer (not shown) is stripped as shown in Figure 9C.

**[0073]** In step 803, Layer-k wafer 201 is flipped as shown in Figure 9D.

**[0074]** In step 804, the flipped Layer-k wafer 201 is attached to a glass carrier wafer 902 via a laser de-bonding adhesive 903 (commercially available) and silicon and sacrificial layers 206, 205 are peeled off as shown in Figure 9E.

**[0075]** The principles of the present invention also align and provide distortion control of picked 2D-die arrays to the product substrate as discussed below.

**[0076]** In one embodiment, precision alignment can be achieved based on whether single or multiple 2D-dies are being assembled simultaneously, which is distinct from the methods discussed in Sreenivasan et al.

**[0077]** In the case of multiple 2D-dies, the moiré metrology needs to refer to the superstrate and not the individual 2D-dies being picked-and-placed. This would necessitate alignment marks to be patterned on the bottom surface of the superstrate. These marks could be patterned on the absolute corners of the superstrate or could also be distributed areally. Corresponding marks would be needed on the product wafer. Some amount of distortion control of the 2D-dies could be implemented using thermal actuation. Additionally, thermal actuation could be implemented in the wafer chuck as well for added actuation degrees-of-freedom. Observation widows could be made in the superstrate in case the superstrate material is not transparent to the wavelength of light used for metrology (which is generally visible or IR). Alternatively, the superstrate could be constructed out of transparent materials, such as SiC and/or sapphire ($Al_2O_3$), which are commercially available. A discussion of precision alignment involving multiple 2D-dies is discussed below in connection with Figures 10 and 11A-11B.

**[0078]** Figure 10 is a flowchart of a method 1000 for overlay and distortion control of multiple packed 2D-dies in accordance with an embodiment of the present invention. Figures 11A-11B depict the cross-sectional views for providing overlay and distortion control of multiple packed 2D-dies using the steps described in Figure 10 in accordance with an embodiment of the present invention.

**[0079]** Referring to Figure 10, in conjunction with Figures 11A-11B, in step 1001, as the picked 2D-dies 1101 (picked 2D-dies, such as 2D-die arrays 101, 102, 103) are brought close to product wafer 1102, course alignment is first done as shown in Figure 11A. Figure 11A illustrates superstrate 1103 with alignment marks 1104 and observation windows 1105.

**[0080]** In step 1002, fine alignment is performed aligning the alignment marks 1104 and observation window 1105 of superstrate 1103 with the alignment marks 1106 in the substrate. In one embodiment, some amount of distortion control of the 2D-dies could be implemented using thermal actuation via thermal actuators 1107. Additionally, thermal actuation could be implemented in the wafer chuck 1108 as well as for added actuation degrees-of-freedom.

**[0081]** In the case of single 2D-dies, in addition to the method described above, moiré metrology could be conducted using IR-sensitive marks 1201 embedded in the Layer-k and Layer-(k-1) 2D-dies 1101 and an IR-transparent superstrate as shown in Figures 12A-12B. Figures 12A-12B illustrate the overlay and distortion control of a single picked 2D-die in accordance with an embodiment of the present invention.

**[0082]** Referring now to Figure 13, Figure 13 illustrates that Through Silicon Vias (TSVs) are made through the access holes already present in picked-and-placed 2D-dies in accordance with an embodiment of the present invention. As shown in Figure 13, Layer-k, Layer-(k+1) and Layer-(k+2) are arranged in a manner where Layer-(k+1) is flipped and Layer-(k+2) is faced-up. As further shown in Figure 13, Through Silicon Vias are fabricated through field access holes 701.

**[0083]** The density of TSVs that are needed can be as much as 10,000/mm for applications, such as static random-access memory (SRAM) stacking. At this level of TSV density, the diameter of the TSV can be approximately 20 nm to 80 nm. Some or all of these TSVs could potentially be routed through the access holes 701 that already exist in 2D-dies.

**[0084]** Referring now to Figures 14A-14C, Figures 14A-14C illustrate an exemplary process for temporary attachment and bonding in accordance with an embodiment of the present invention.

**[0085]** In one embodiment, temporary attachment may be followed by bonding. In one embodiment, a dynamic air-cushion based “slow landing” approach could be used. Such systems have previously been used in high-precision air-bearing stages, in hard-disk drive

systems, and have been studied for drop skating on solid surfaces. In this approach, a thin layer of UV-curable adhesive could first be dispensed on the edge of the layer-0 2D-die. Said adhesive could be composed of a combination of volatile and non-volatile components, where in the limiting case the adhesive is composed of solely non-volatile components. The liquid comprised of UV-curable adhesive and/or volatile component provides damping thereby substantially minimizing vibrational displacement between the Layer-0 and Layer-1 2D-die. As the superstrate, with attached Layer-1 2D-die is brought in proximity to the Layer-0 die, air-flow through the pressure holes 1401 could be initiated. This would create a bearing composed of air or nitrogen (to obtain an inert environment) around the periphery of the 2D-die. The combined knobs of superstrate z-force and the said above bearing flow rate could be used to control the "soft landing." Simultaneously, coarse alignment corrections could be done as the superstrate 1103 is being urged down. Simultaneously, a second air cushion 1402 is created in between the 2D-dies being stacked. This second air cushion 1402 could provide additional lubrication between 2D-dies during fine-alignment corrections.

**[0086]** Additionally, the outward flow of air from this air cushion would ensure that volatile components in the adhesive (which is on the edge) do not contaminate metal-metal contacts 1403 in the bulk of the 2D-die. Additionally, the flow rate of the second air cushion 1402 could be controlled by varying the topography of the 2D-dies using a superstrate 1103 with z-direction piezoelectric actuators. Such systems have been demonstrated previously. Once the 2D-dies make contact, a blanket UV exposure 1404 could be done to cure the edge placed adhesive. To further secure the 2D-dies, a surface activation of the metal contacts 1403 could be done. Such a process has been shown before for room temperature metal to metal bonding including metals, such as copper, tungsten and aluminum. Surface activation of copper can be achieved using argon ion treatment of the copper surface. In one embodiment, it is assumed that all of the air used in the above air bearings are semiconductor grade clean dry air. Alternatively, if the bearing uses nitrogen, it is also assumed to be semiconductor grade, clean and dry. In one embodiment, the surface activated copper is maintained in an inert environment after the activation process till the bonding step (including transport from tool-tool and in every tool, it is processed in). In one embodiment, vacuum holes 1405 may be used to enable a vacuum based pickup mechanism.

**[0087]** A discussion regarding the design and electronic design automation (EDA)/computer-aided design (CAD) flows required to implement the 3D-IC System on Chip (SoC) is now deemed appropriate. Typically, 2D ASIC SoC comprises of billions of transistors which are placed optimally to meet the performance/speed, area and power specifications. In order to efficiently design 2D ASIC SoC, i.e., meeting design specifications with lower turn-around time (TAT) to market, there exists commercial EDA CAD tools to simplify the design process. However, no such EDA tools exist for 3D-IC ASIC design.

**[0088]** A typical ASIC SoC can be broadly divided into the following segments: logic (CPU, GPU, Modem, etc.), memory/cache (static random access memory (SRAM), embedded dynamic random access memory (eDRAM), etc.), third-party IP blocks, analog IP, IO, etc. 3D SoC design aims to implement same functionality SoC while reducing the foot print and improving its performance in terms of reduced memory access times and latency, higher bandwidth, higher capacity in terms of Mbits/mm$^2$, higher frequency due to shorter interconnect delays, etc.

**[0089]** A typical 3D SoC, which may also be referred to herein as the "Nano-precision aligned 3D Stacked Integrated Circuit (N3SI)" includes $n$ base transistor layers, where $n>1$. In one embodiment, an application specific integrated circuit (ASIC) system on a chip (SoC) with logic and memory circuitry is designed and manufactured in three dimensions using a sub-50 nm overlay pick and place method, which allows precision overlay of the logic and memory circuitry. Each base layer might have $m$ metal layers, where $m>=1$ and may vary for each base layer. The base layers in the 3D stack can be placed in any of the following configurations with respect to each other: face-to-face, face-to-back, back-to-back, etc. The connections across different base layers can be made using Interlayer Via (ILV) if base layers are in face-to-face configuration or using nanoscale Through Silicon Via (nano-TSV) if it is face-to-back or back-to-back configuration. 3D SoC can be designed using combination of any of the following design approaches: 2D logic implementation with 3D memory implementation, 3D logic implementation with 2D memory implementation, 3D logic implementation with 3D memory implementation, etc. The 3D logic implementation can be either performed at the block/partition level or can be performed at the flat level. In the 3D block level logic implementation, partitions are synthesized and routed using 2D tools, but different partitions are placed in different base layers. This approach requires changes only in the top level SoC design, whereas, the block level design of the 3D SoC remains the same as the 2D SoC. Thus, this approach is easier to

implement. In the flat level 3D logic implementation, the partitions are also implemented in 3D, i.e., cells within a partition are placed in multiple base layers. 3D logic implementation and 3D memory design implementation are discussed below, respectively. The area overhead due to TSVs and HF holes can be also optimized by space optimization algorithms.

**[0090]** Electronic Design Automation (EDA) design methodology for 3D-IC logic implementation is now discussed. The conventional 2D ASIC EDA flow for logic implementation is shown in Figure 15 in accordance with an embodiment of the present invention. Synthesis is performed in the front-end design phase, whereas, the backend design phase performs placement, pre-CTS optimization, clock tree synthesis (CTS), route, post route optimizations, signoff analysis and design verification.

**[0091]** The EDA methodology of the present invention for 3D-IC SoC is also similar to the 2D ASIC flow. The methodology attempts to re-use most of the existing commercial 2D EDA tools along with using some of the solutions developed in-house. Such a flow is referred to herein as the "N3SI EDA flow." The following sub-sections describe the N3SI EDA flow design steps.

**[0092] Synthesis for 3D-IC SoC**

**[0093]** The synthesis of 3D-IC SoC makes use of the commercial 2D synthesis tool. In the first pass, the design is synthesized exactly as it is done in 2D SoC. Once the placement is performed, 3D placement aware synthesis is performed. In this synthesis pass, the tool synthesizes the cells more optimally since it has 3D placement information to get accurate interconnect loads and delays. This process flow is also similar to 2D placement aware synthesis, however, placement information in this case is three-dimensional.

**[0094] Placement for 3D-IC SoC**

**[0095]** This section deals with the 3D placement of logic/standard cells. In the methodology of the present invention, design netlist is first partitioned into multiple modules such that each module netlist consists of logic cells, etc. to be placed on different layers of a 3D-IC SoC stack. Then, 2D placement for each module in an assigned layer of 3D stack is performed using commercial 2D EDA tools. The netlist partitioning can be performed using in-house solutions which make use of standard partitioning algorithms, such as FM Min-Cut, Min-Flow, etc. The modules generated in partitioning consist of input/output ports which are not placed only on the

module periphery, but can be placed anywhere in the module. As a result, the in-house developed software uses standard partitioning algorithms to generate the locations of these ports. Multiple modules transfer signals through these ports. These ports can be connected through Interlayer Via (ILV) or nanoscale Through Silicon Via (nano-TSV). The locations of these ports might be constrained based on thermal and mechanical stability of ILVs and TSVs. Once the port locations are decided, the timing budgets and port locations are fed to the 2D placer tool to perform placement of each module independently while making sure that overall timing and performance metrics are met. In order to ensure legal cell placement, placement or routing blockages are formed in the module regions from where TSV or HF holes pass. That is, the in-house developed software uses standard partitioning algorithms to generate placement or routing blockages, such as to avoid (Design Rule Checking) DRC issues at ILV/TSV locations.

**[0096] CTS for 3D-IC SoC**

**[0097]** The clock tree synthesis (CTS) for 3D SoC can be performed using existing 2D EDA placement and route (P&R) tools. Once the design is partitioned and placed into multiple modules, clock tree can be built and optimized for each module separately. However, the challenge with 3D clock tree is to ensure that there is no setup, hold, etc. violations while considering process variations across multiple wafers on which the 3D clock tree might be built. There can be multiple ways to resolve or obviate this problem. One of the possible solutions is to constrain placement of the launch and capture flop on the same layer, i.e., launch and capture flop for data path needs to be placed on the same layer. This can be achieved by the in-house netlist partitioning tool. Another solution is to include the high margin in order to ensure that there are no violations in the worst case process variation scenario.

**[0098] Route for 3D-IC SoC**

**[0099]** The route methodology for the 3D-IC SoC includes 2D routing within each module, and routing across multiple modules using ILV and nano-TSV. The resistance and capacitance values can be determined accurately, and will be discussed next. The 3D-IC routing methodology remains the same as 2D routing. The routing for each module can be implemented using the 2D P&R tool separately. In order to ensure no design rule check (DRC) failures, routing blockages are formed in regions, where ILV and TSV interconnects are placed.

**[00100] Parasitic Extraction for 3D-IC SoC**

**[00101]** The design methodology for parasitic extraction of 3D SoC differs from 2D ASIC. The resistance and capacitance values can vary significantly due to TSVs and ILVs. The commercial EDA tools are not capable of performing 3D extraction. However, embodiments of the present invention utilize the 3D extraction flow which makes use of the existing 2D extractor. In this flow, the layout information for each module or layer is first streamed out. Then, layout/route data for all the modules is streamed into a layout editor tool, such as Virtuoso®. While streaming in, the layout of any specific module can be flipped if required to make it look identical to the 3D SoC stack. Then, the extractor is run on this layout. The resistance and capacitance values obtained have taken into account the 3D layout, considering TSVs and ILVs, and are expected to be accurate.

**[00102]** A 3D design implementation of static random access memory (SRAM) is now discussed. A typical SRAM includes a bit cell array with word and bit lines, sense amplifiers, column and row decoders, timer circuitry, IO, other peripheral circuitry, etc. There are multiple SRAM configurations, such as the butterfly configuration, the single sided configuration, etc. to place SRAM design elements. These configurations differ in implementation complexity, access times, latency, etc. Figure 16 illustrates the 2D single sided SRAM configuration in accordance with an embodiment of the present invention. The SRAM configuration includes basic memory design elements, such as the bit array of SRAM cells 1601, bit lines 1602, word lines 1603, IO cells 1604, timer circuitry 1605, sense amplifiers 1606 and decoders 1607.

**[00103]** Similar to 2D SRAM configurations, 3D SRAM can be designed in multiple configurations as per design needs. The 3D eDRAM is also similar to the 3D SRAM methodology and similar eDRAM configurations can be designed. One of the possible 3D single sided SRAM configurations, shown by Figure 17, is 3D stacked dies of stand-alone SRAM arrays. Figure 17 illustrates 3D stand-alone SRAM die stacking in accordance with an embodiment of the present invention.

**[00104]** In this 3D SRAM configuration, each layer implements self-sustainable 2D single sided SRAM. The data input, power and control signals are fed to each 2D SRAM stacked in a 3D configuration and output data signals are obtained from each layer. Combining the data outputs from all layers make the complete 3D SRAM output. For example, as shown in Figure 17, the 32-bit $D_{in}$ data bus signal is divided into 4 8-bit data bus signals and fed to each of the 4 layers.

The data output $D_{out}$ from each layer comprises of 8 bits, and combining it from 4 layers makes a 32-bit output signal.

**[00105]** One of the other possible 3D single sided SRAM configurations is a 3D only-bitcell stacked SRAM shown in Figure 18 in accordance with an embodiment of the present invention.

**[00106]** In this type of 3D SRAM configuration, base layer, i.e., layer 1, comprises of a bitcell array 1801 with bit lines 1802 and word lines 1803, control and periphery circuitry elements, such as IO cells 1804, timer circuitry 1805, sense amplifiers 1806 and decoders 1807. The stacked 3D layers comprise only a bitcell array, bit lines and word lines. In one embodiment, the control circuitry in the base layer for the 3D SRAM is expected to have more column decoders in comparison to the 2D configuration. Similar to the single sided SRAM design, other 2D SRAM configurations, such as butterfly, etc., can be also implemented in 3D.

**[00107]** In the 3D only-bitcell style stacked SRAM, there are multiple ways to design it as per design specifications. In one of the configurations, each layer contains the bitcell array with the same size as in the 2D SRAM. In the 3D SRAM, the bit line and word line lengths, bandwidth, footprint, etc. remain the same as 2D SRAM, but memory capacity, i.e., array bitcell density, becomes $n$ times, where $n$ is the number of layers. The slight modification to this design configuration would be to add more sense amplifiers in order to increase the memory bandwidth. Figure 19 illustrates the vertical bit-line cross section for the 3D only-bitcell stacked SRAM in accordance with an embodiment of the present invention.

**[00108]** As illustrated in Figure 19, this example does not utilize column decoders to select the bit line layer. However, the memory access time is expected to be reduced because of smaller bit line lengths, which ultimately reduces the time constant $RC$, where $R$ is the resistance and $C$ is the capacitance. In order to select the bit line for the specific layer, a decoder can be added to this design configuration.

**[00109]** Another possible 3D only-bitcell stacked SRAM design configuration reduces the foot print/area while keeping the memory capacity, i.e., array bitcell density, the same. In this configuration, the first base layer includes control circuitry which is identical to the control circuitry used in the 2D SRAM configuration. The footprint of the bit array, which is typically 70% of the SRAM area in the 2D configuration, can be reduced in the 3D configuration. The bitcell array area can be divided by $n$, where $n$ $(n>1)$ is the number of bitcell array layers. In this

configuration, the bit lines and word lines will be smaller in length, with additional column decoders. However, it is expected that this type of memory configuration would result in less memory access times.

**[00110]** By using the principles of the present invention, it is now possible to fabricate a three-dimensional (3D) stacked integrated circuit. In one embodiment, pick-and-place strategies are used to stack the source wafers with device layers fabricated using standard two-dimensional (2D) semiconductor fabrication technologies. The source wafers may be stacked in either a sequential or parallel fashion. The stacking may be in a face-to-face, face-to-back, back-to-face or back-to-back fashion. The source wafers that are stacked in a face-to-back, back-to-face or back-to-back fashion may be connected using Through Silicon Vias (TSVs). Alternatively, source wafers that are stacked in a face-to-face fashion may be connected using Inter Layer Vias (ILVs).

**[00111]** The descriptions of the various embodiments of the present invention have been presented for purposes of illustration, but are not intended to be exhaustive or limited to the embodiments disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the described embodiments. The terminology used herein was chosen to best explain the principles of the embodiments, the practical application or technical improvement over technologies found in the marketplace, or to enable others of ordinary skill in the art to understand the embodiments disclosed herein.

CLAIMS:

1 1. A method for fabricating a three-dimensional (3D) system on a chip (SoC), the method
2 comprising:
3 assembling a Layer-(k) two-dimensional (2D)-die array onto a Layer-(k-1) 2D-die array
4 of a Layer-(k-1) wafer, wherein said Layer-(k-1) wafer is populated with 2D-dies, wherein said k
5 is a positive integer number greater than 1, wherein said 2D-die array comprises a single 2D-die,
6 a single island of 2D-die that forms a contiguous group of 2D die or multiple islands of 2D die;
7 and
8 deploying a fluid allowing lubricated relative motion between said Layer-(k) 2D-die
9 array and said Layer-(k-1) 2D-die array, wherein said fluid allows precision overlay of said
10 Layer-(k) and Layer-(k-1) 2D-die arrays.

1 2. The method as recited in claim 1, wherein said assembly is performed to achieve one of
2 the following: sub-100nm overlay, sub-50nm overlay, sub-30nm overlay, sub-20nm overlay,
3 sub-10nm overlay and sub-5nm overlay between each 2D-die of a Layer-(k) wafer and the
4 corresponding 2D-die of said Layer-(k-1) wafer.

1 3. The method as recited in claim 1, wherein said fluid comprises one of the following: a
2 gas, a liquid and a combination thereof, wherein said combination comprises disparate gas and
3 liquid portions or portions of a homogenously mixed gas and liquid.

1 4. The method as recited in claim 1, wherein said 2D-dies comprise access holes for
2 etchants to release said 2D-dies from said Layer-(k-1) wafer.

1 5. The method as recited in claim 4 further comprising:
2 utilizing said access holes to create conductors that enable through silicon vias in said 3D
3 SoC.

1 6. The method as recited in claim 1, wherein said SoC comprises an application specific
2 integrated circuit (ASIC) system, wherein said ASIC system comprises logic and memory
3 circuitry designed and manufactured in three dimensions (3D) using a pick-and-place method,
4 which allows precision overlay of said logic and memory circuitry.

1 7. The method as recited in claim 6, wherein said ASIC system further comprises:
2 n base layers, wherein said n is greater than 1, wherein one or more of said n base layers
3 are placed in one or more of the following configurations: face-to-face, face-to-back and back-to-
4 back.

1 8. The method as recited in claim 6, wherein said ASIC system is designed using any of the
2 following design approaches: a two-dimensional (2D) logic implementation with a three-
3 dimensional (3D) memory implementation, a 3D logic implementation with a 2D memory
4 implementation, and a 3D logic implementation with a 3D memory implementation.

1 9. The method as recited in claim 6, wherein said ASIC system is used in one or more of the
2 following 3D static random access memory (SRAM) configurations: a 3D stand-alone stacked
3 SRAM and a 3D only-bitcell stacked SRAM.

1 10. A method for fabricating a three-dimensional (3D) system on a chip (SoC), the method
2 comprising:
3 assembling a Layer-(k) two-dimensional (2D)-die array onto a Layer-(k-1) 2D-die array
4 of a Layer-(k-1) wafer, wherein said Layer-(k-1) wafer is populated with 2D-dies, wherein said k
5 is a positive integer number greater than 1, wherein said 2D-die array comprises a single 2D-die,
6 a single island of 2D-die that forms a contiguous group of 2D die or multiple islands of 2D die;
7 and
8 providing an encapsulation layer for protecting 2D-dies in each of Layer-(k) wafer and
9 said Layer-(k-1) wafer from etchants used during a pick-and-place process.

1 11. The method as recited in claim 10, wherein said encapsulation layer is compatible with
2 complementary metal–oxide–semiconductor (CMOS) and/or III-V semiconductors.

1 12. The method as recited in claim 10, wherein said assembly is performed to achieve one of
2 the following: sub-100nm overlay, sub-50nm overlay, sub-30nm overlay, sub-20nm overlay,
3 sub-10nm overlay and sub-5nm overlay between each 2D-die of said Layer-(k) wafer and the
4 corresponding 2D-die of said Layer-(k-1) wafer.

1 13. The method as recited in claim 10, wherein said 2D-dies comprise access holes for
2 etchants to release said 2D-dies from said Layer-(k-1) wafer.

1 14. The method as recited in claim 13 further comprising:
2 utilizing said access holes to create conductors that enable through silicon vias in said 3D
3 SoC.

1 15. The method as recited in claim 10, wherein said SoC comprises an application specific
2 integrated circuit (ASIC) system, wherein said ASIC system comprises logic and memory
3 circuitry designed and manufactured in three dimensions (3D) using said pick-and-place process,
4 which allows precision overlay of said logic and memory circuitry.

1 16. The method as recited in claim 15, wherein said ASIC system further comprises:
2 n base layers, wherein said n is greater than 1, wherein one or more of said n base layers
3 are placed in one or more of the following configurations: face-to-face, face-to-back and back-to-
4 back.

1 17. The method as recited in claim 15, wherein said ASIC system is designed using any of
2 the following design approaches: a two-dimensional (2D) logic implementation with a three-
3 dimensional (3D) memory implementation, a 3D logic implementation with a 2D memory
4 implementation, and a 3D logic implementation with a 3D memory implementation.

1 18. The method as recited in claim 15, wherein said ASIC system is used in one or more of
2 the following 3D static random access memory (SRAM) configurations: a 3D stand-alone
3 stacked SRAM and a 3D only-bitcell stacked SRAM.

1 19. A method for fabricating a three-dimensional (3D) system on a chip (SoC), the method
2 comprising:
3 assembling a Layer-(k) two-dimensional (2D)-die array onto a Layer-(k-1) 2D-die array
4 of a Layer-(k-1) wafer, wherein said Layer-(k-1) wafer is populated with 2D-dies, wherein said k
5 is a positive integer number greater than 1, wherein said 2D-die array comprises a single 2D-die,
6 a single island of 2D-die that forms a contiguous group of 2D die or multiple islands of 2D die,
7 wherein said 2D-dies have a thickness less than 10 micrometers.

1 20. The method as recited in claim 19, wherein said assembly is performed to achieve one of
2 the following: sub-100nm overlay, sub-50nm overlay, sub-30nm overlay, sub-20nm overlay,
3 sub-10nm overlay and sub-5nm overlay between each 2D-die of a Layer-(k) wafer and the
4 corresponding 2D-die of said Layer-(k-1) wafer.

1 21. The method as recited in claim 19, wherein said 2D-dies comprise access holes for
2 etchants to release said 2D-dies from said Layer-(k-1) wafer.

1 22. The method as recited in claim 21 further comprising:
2 utilizing said access holes to create conductors that enable through silicon vias in said 3D
3 SoC.

1 23. The method as recited in claim 19, wherein said SoC comprises an application specific
2 integrated circuit (ASIC) system, wherein said ASIC system comprises logic and memory
3 circuitry designed and manufactured in three dimensions (3D) using a pick-and-place method,
4 which allows precision overlay of said logic and memory circuitry.

1 24. The method as recited in claim 23, wherein said ASIC system further comprises:
2 n base layers, wherein said n is greater than 1, wherein one or more of said n base layers
3 are placed in one or more of the following configurations: face-to-face, face-to-back and back-to-
4 back.

1 25. The method as recited in claim 23, wherein said ASIC system is designed using any of
2 the following design approaches: a two-dimensional (2D) logic implementation with a three-
3 dimensional (3D) memory implementation, a 3D logic implementation with a 2D memory
4 implementation, and a 3D logic implementation with a 3D memory implementation.

1 26. The method as recited in claim 23, wherein said ASIC system is used in one or more of
2 the following 3D static random access memory (SRAM) configurations: a 3D stand-alone
3 stacked SRAM and a 3D only-bitcell stacked SRAM.

1 27. An electronic design automation (EDA) methodology for designing three-dimensional
2 (3D) application specific integrated circuit (ASIC) system on a chip (SoC) logic circuitry
3 comprising:

4 a combination of software integrated with two-dimensional (2D) EDA solutions, wherein
5 said software comprises a netlist partitioning algorithm to partition a 3D design netlist into 2D
6 modules, wherein said 2D EDA solutions are used to perform one or more of the following:
7 synthesis, 3D placement aware synthesis, placement, clock tree synthesis (CTS), routing, design
8 verification and signoff analysis.

1 28. The EDA methodology as recited in claim 27, wherein said algorithm performs one or
2 more of the following: generating locations of ports and generating placement or routing
3 blockages.

1 29. The EDA methodology as recited in claim 28, wherein said locations of said ports are
2 constrained based on thermal and mechanical stability of Inter Layer Vias (ILVs) and Through
3 Silicon Vias (TSVs).

1 30. The EDA methodology as recited in claim 27 further comprising:
2 streaming layout or route data into a layout editor tool, wherein a layout of a specific
3 module is flipped in response to making said layout of said specific module look identical to a
4 3D SoC stack; and
5 obtaining resistance and capacitance values taking into account said layout and
6 considering ILVs and TSVs.

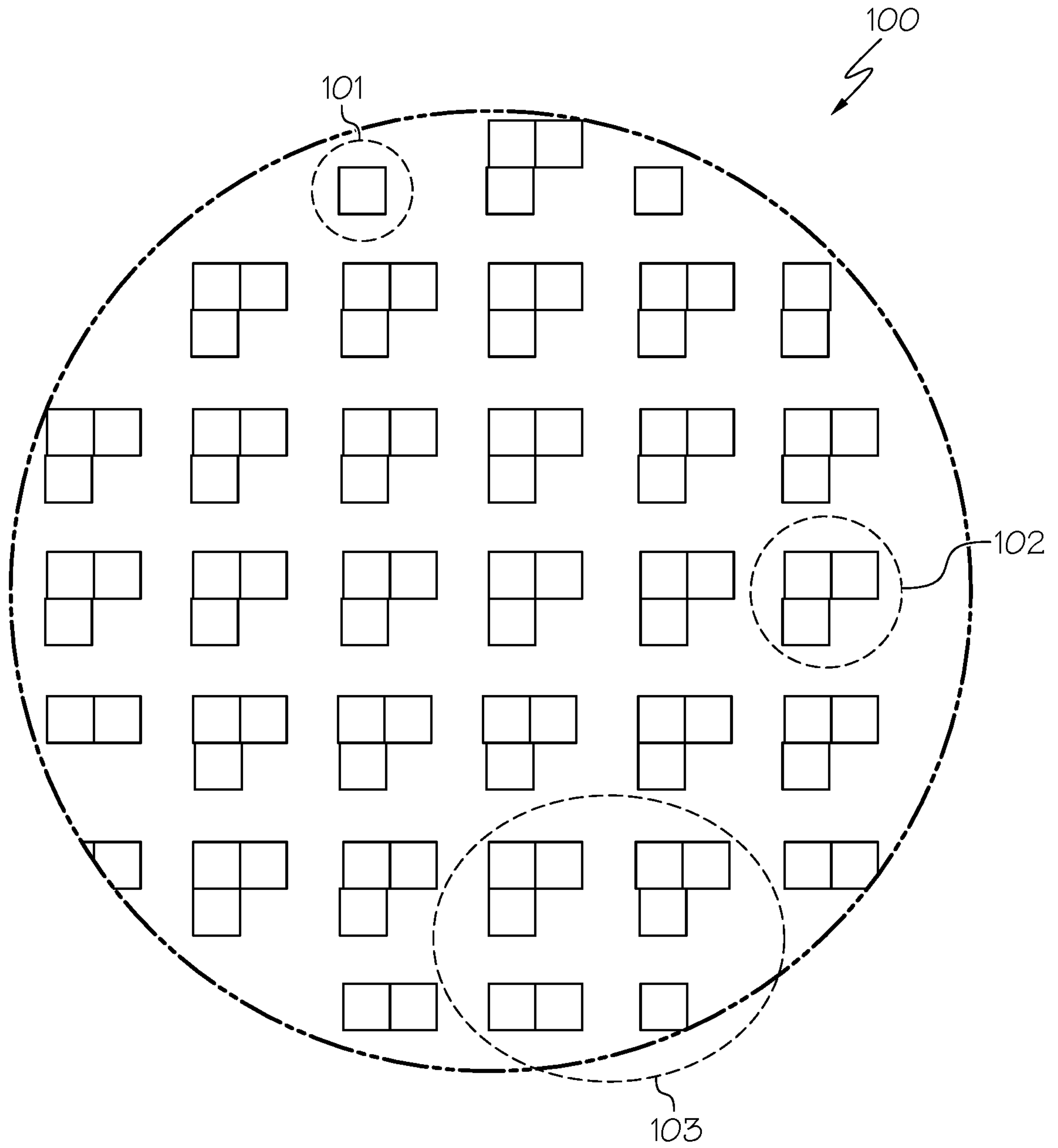



FIG. 1

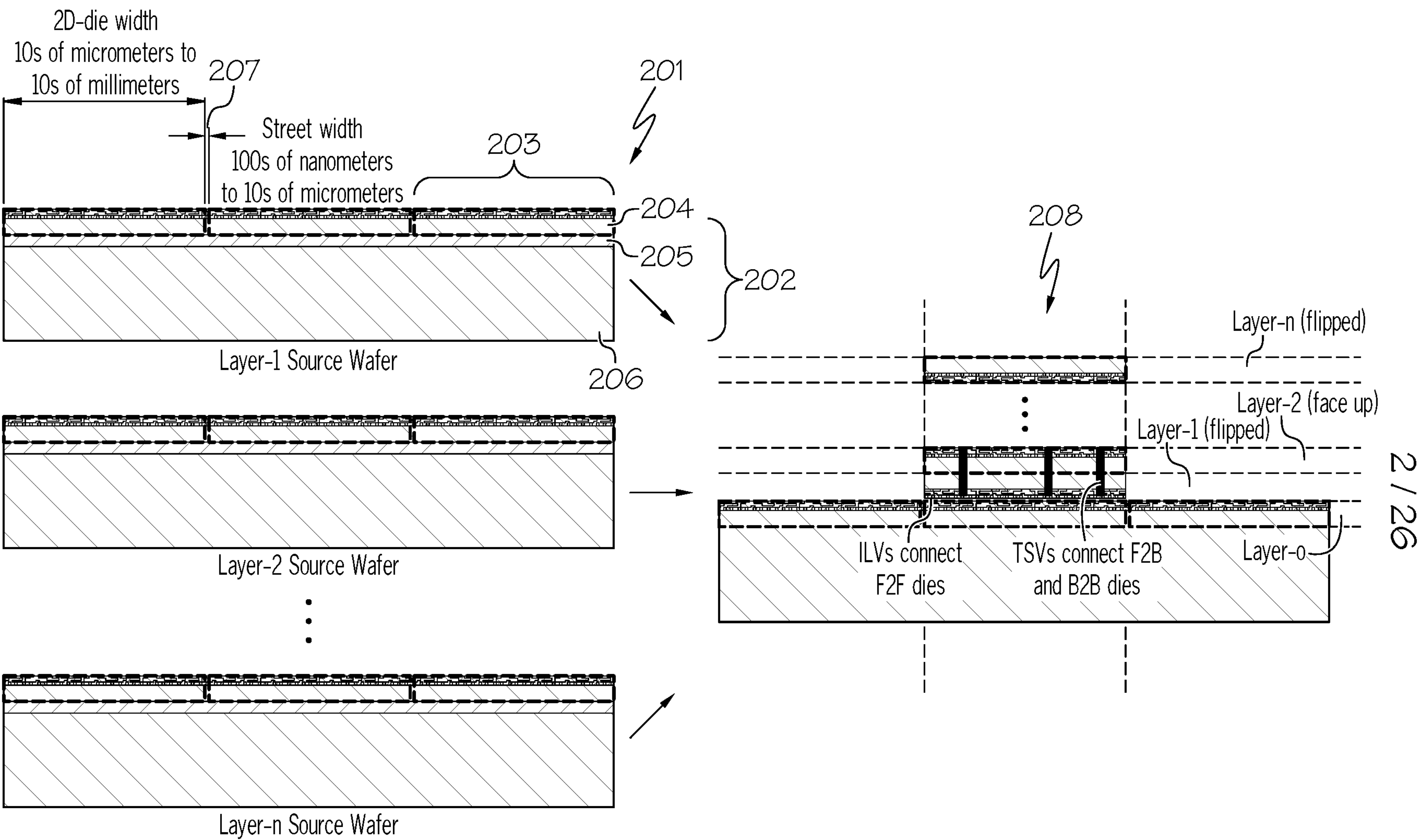



FIG. 2

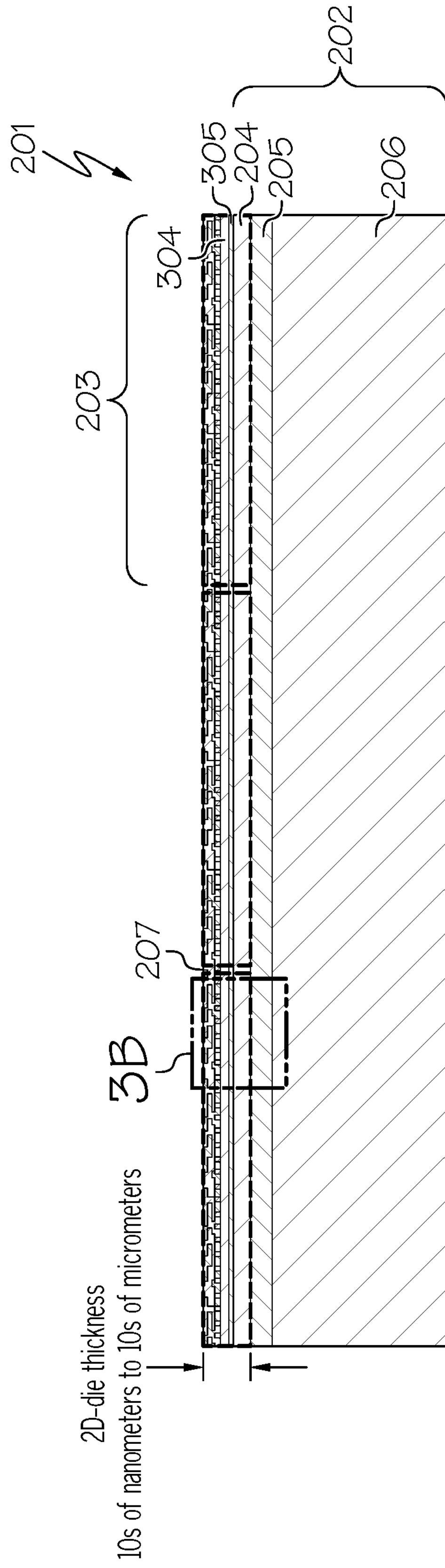



FIG. 3A

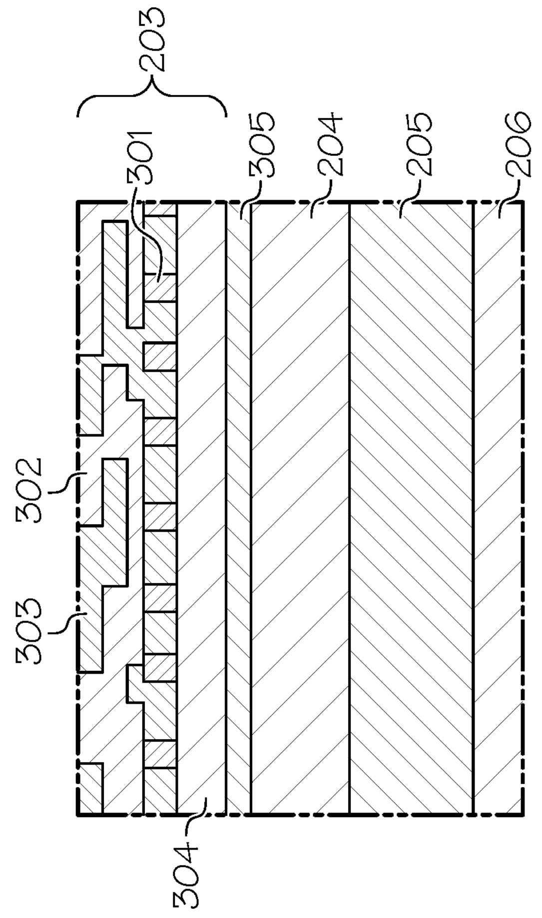



FIG. 3B

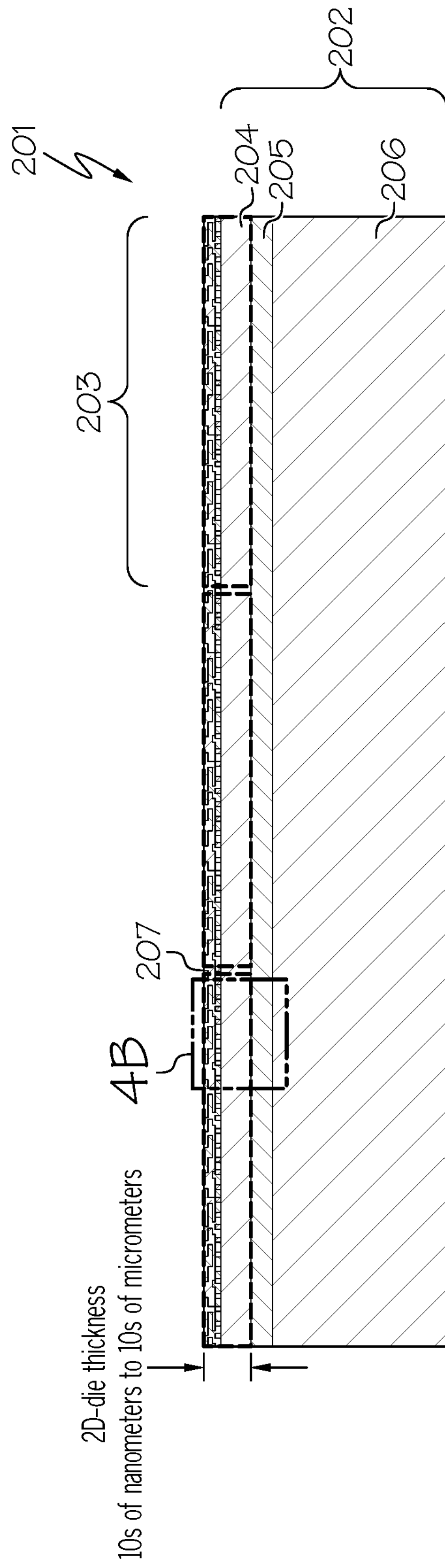



FIG. 4A

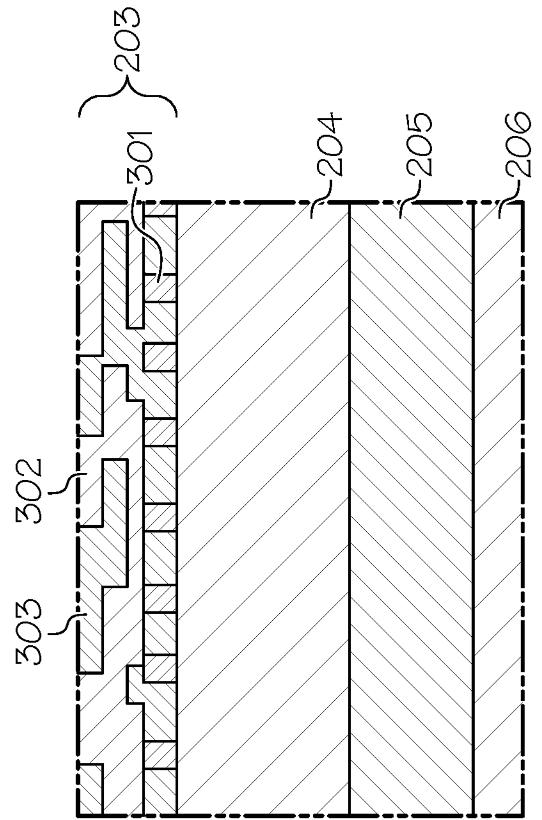



FIG. 4B

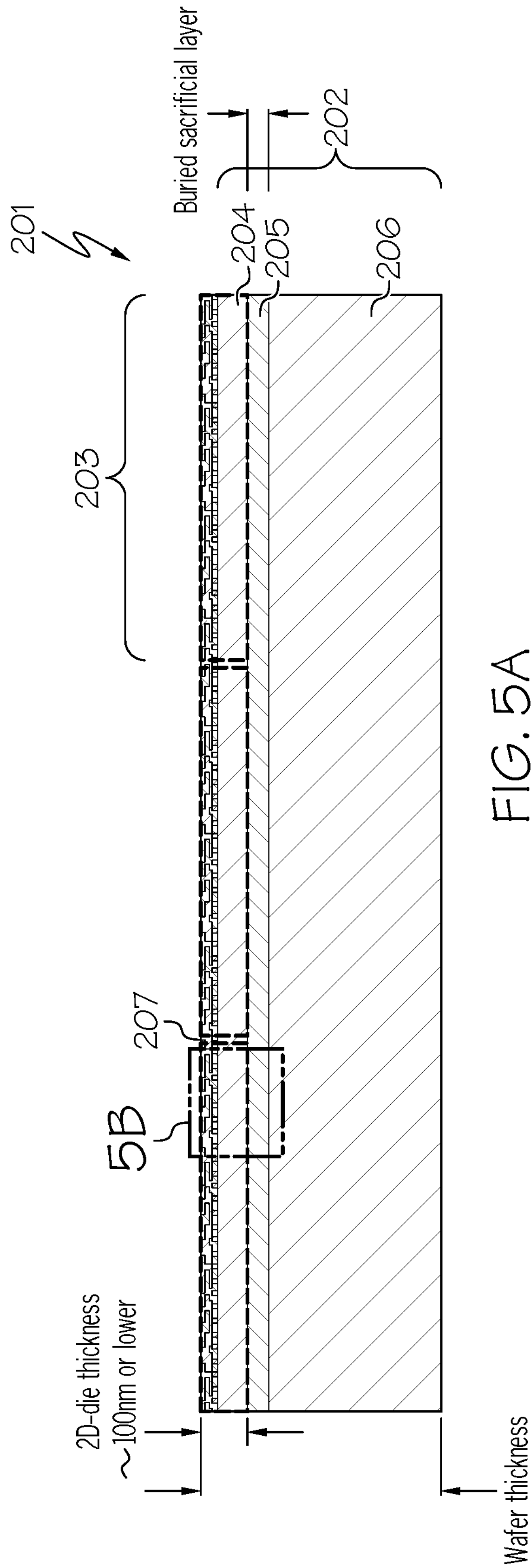



FIG. 5A

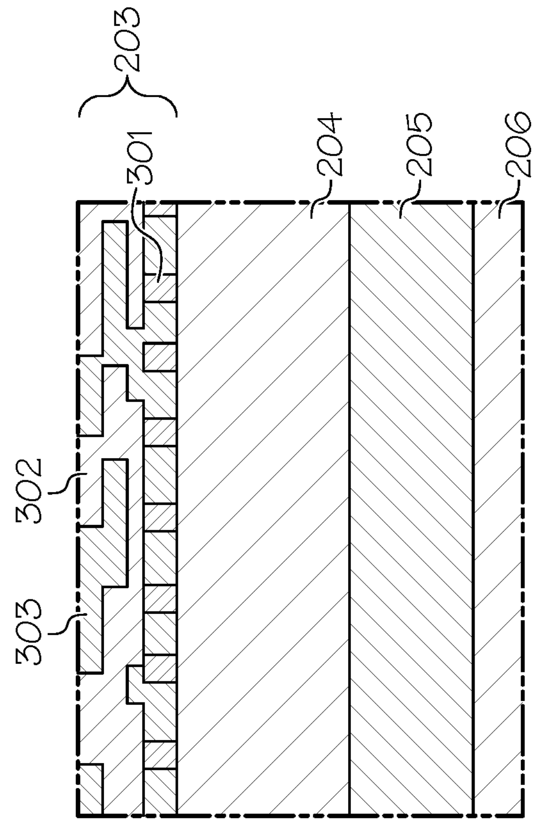



FIG. 5B

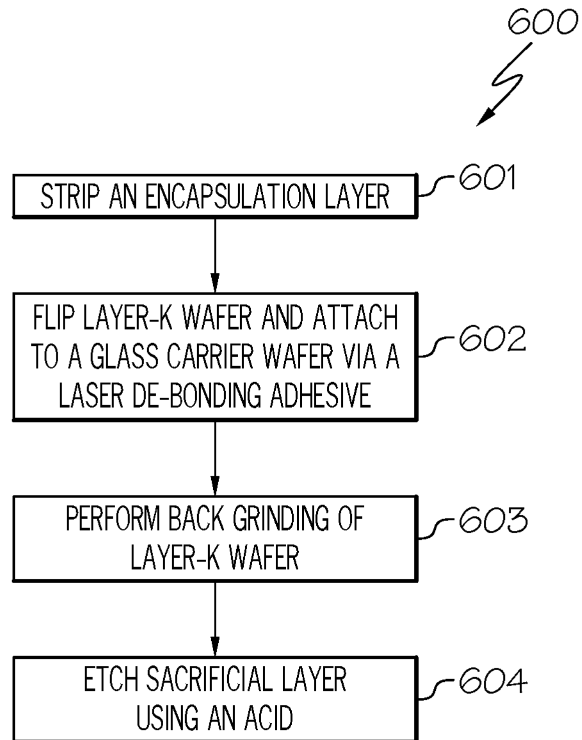



FIG. 6

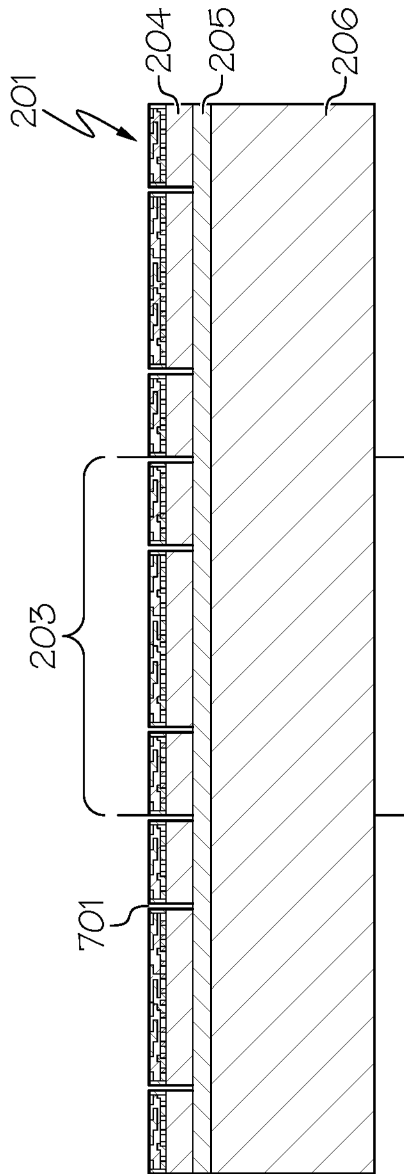



FIG. 7A

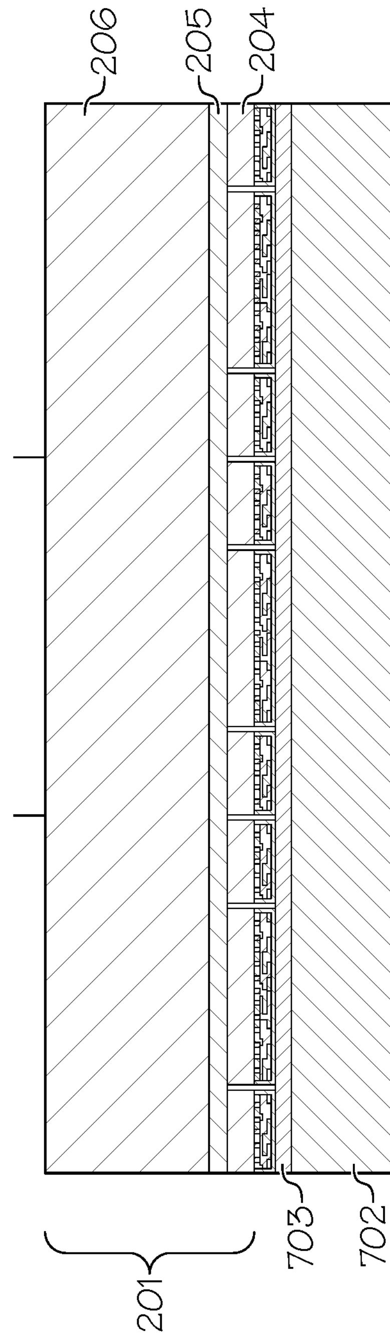



FIG. 7B

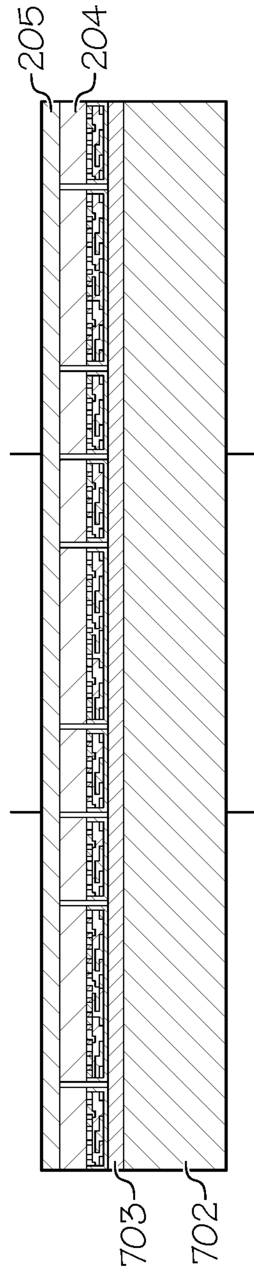



FIG. 7C

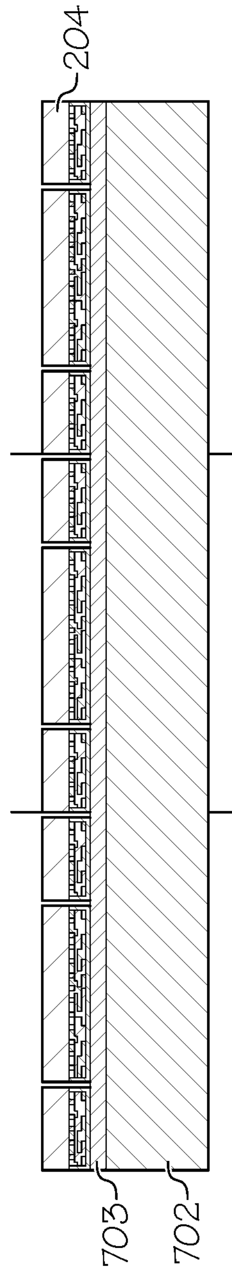



FIG. 7D

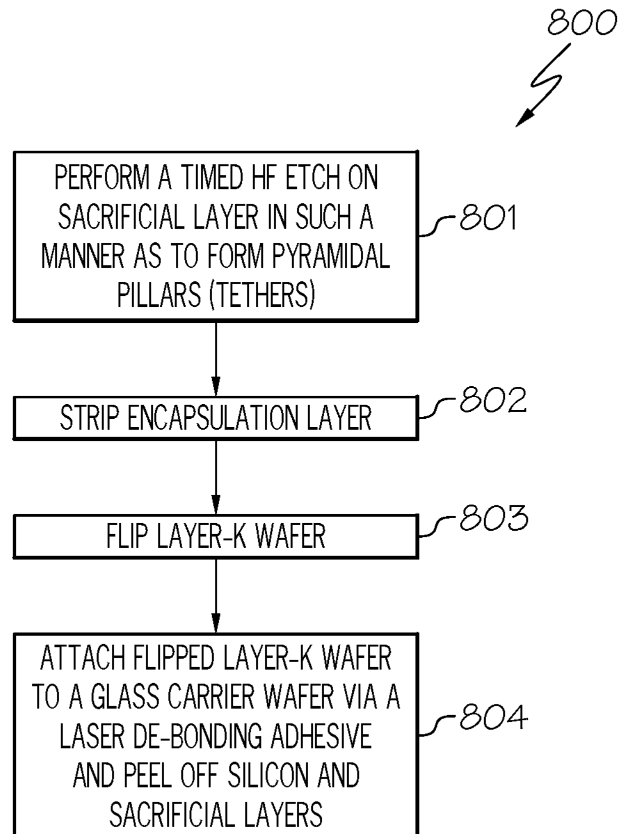



FIG. 8

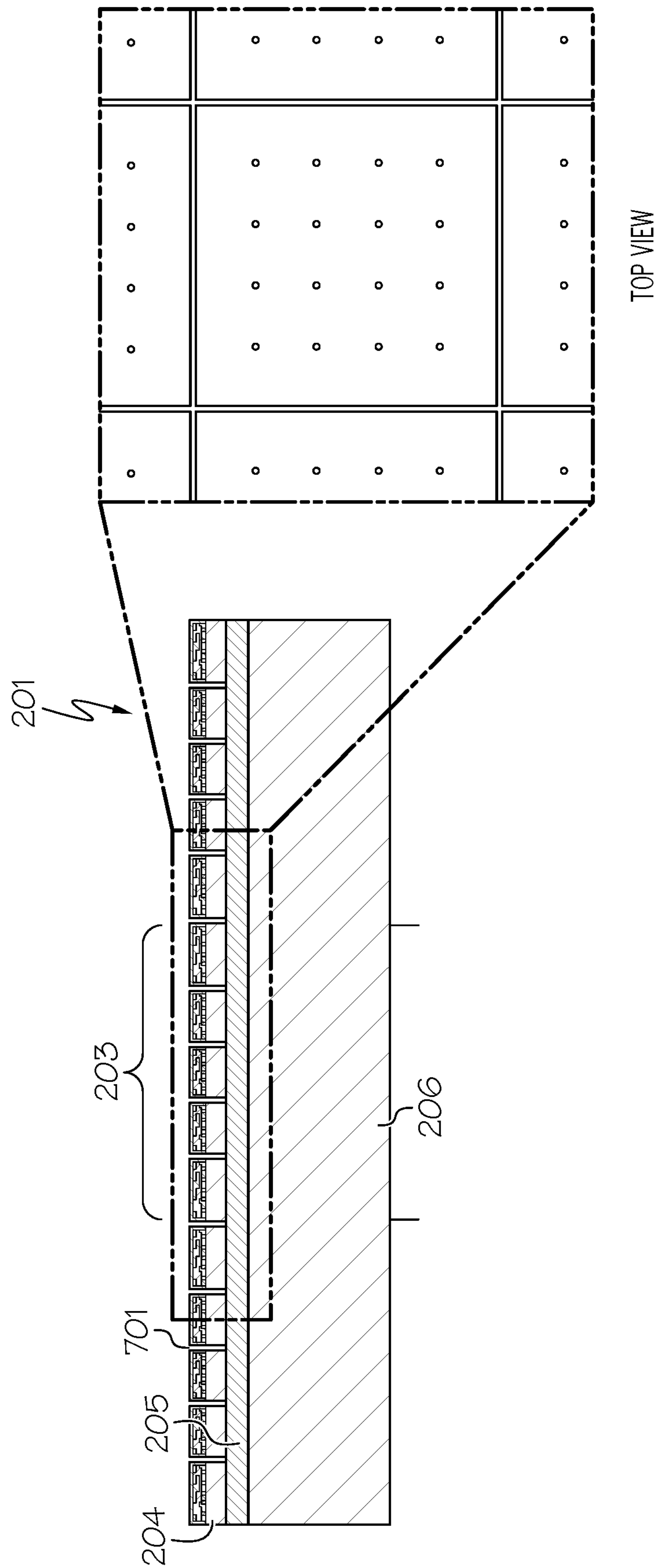



FIG. 9A

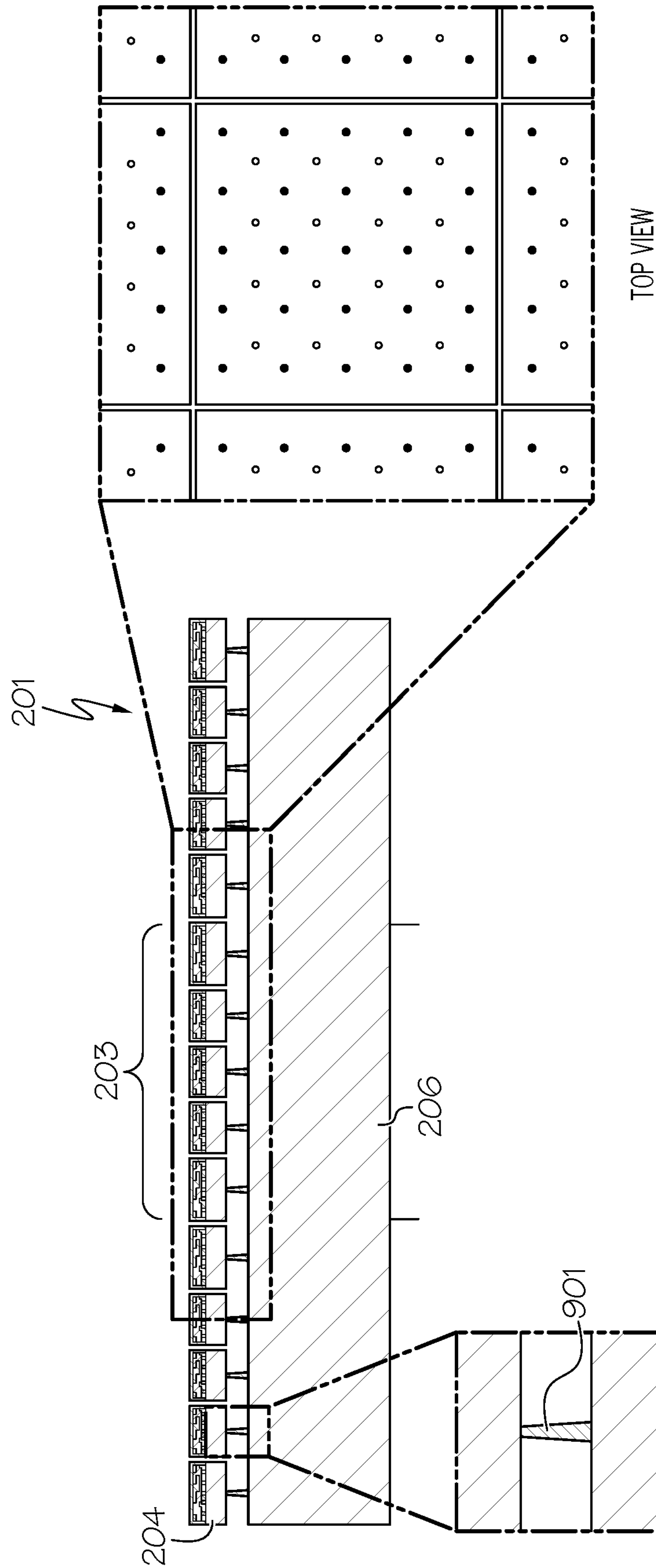



FIG. 9B

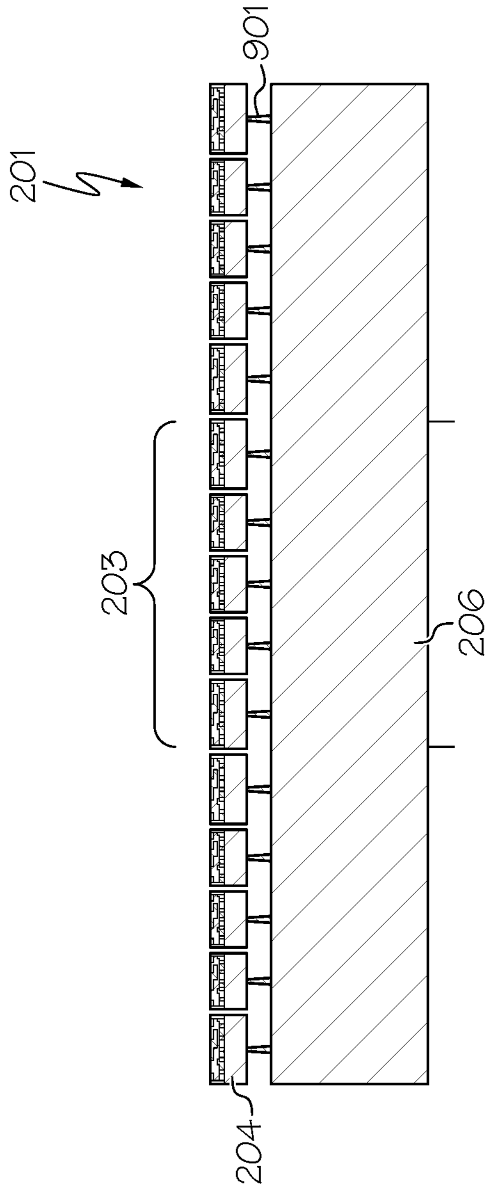



FIG. 9C

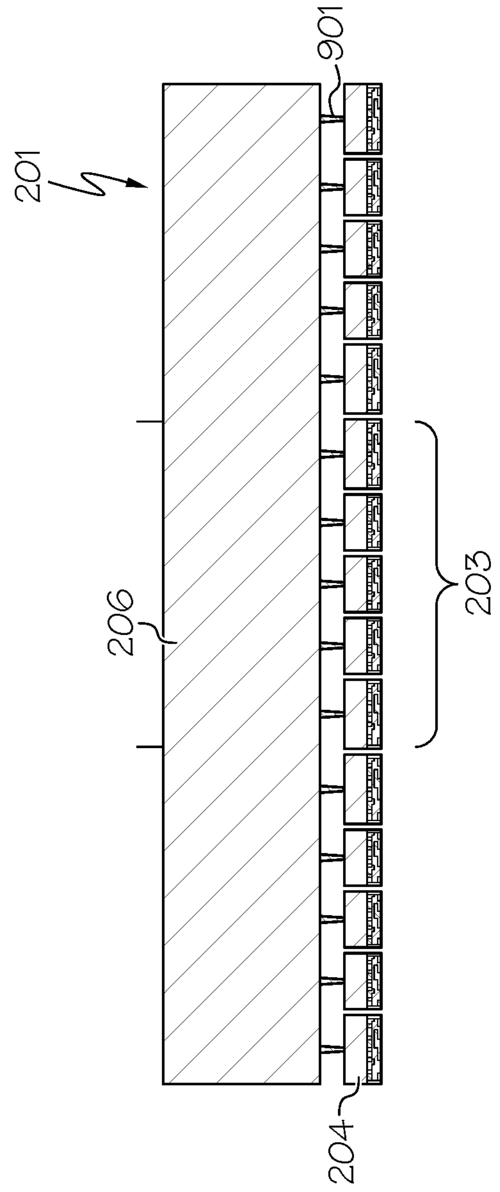



FIG. 9D

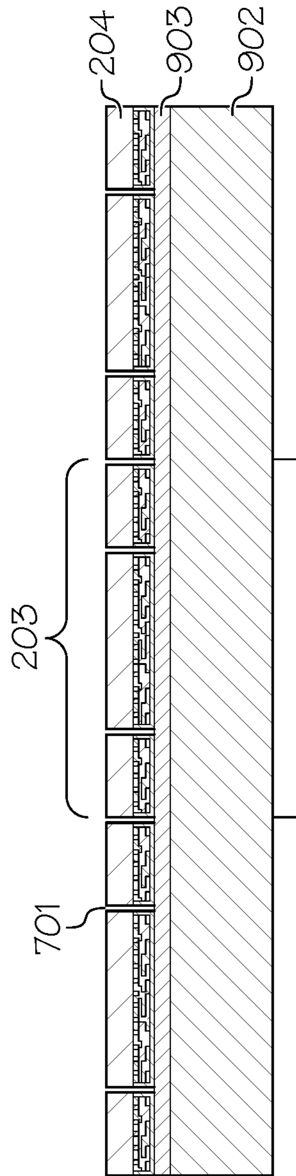



FIG. 9E

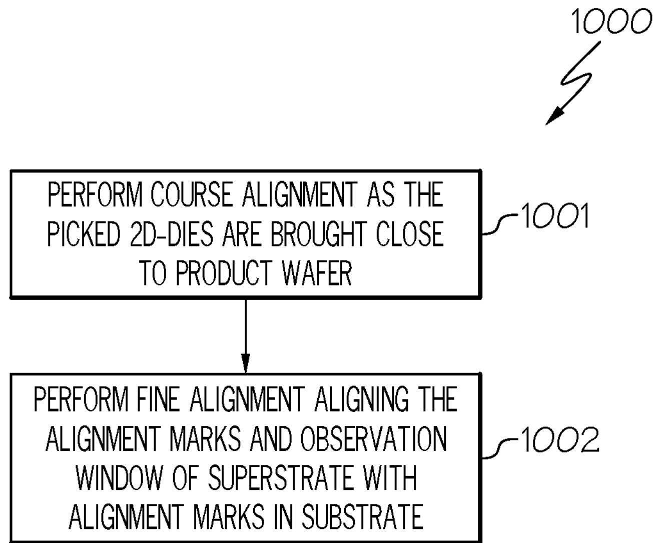



FIG. 10

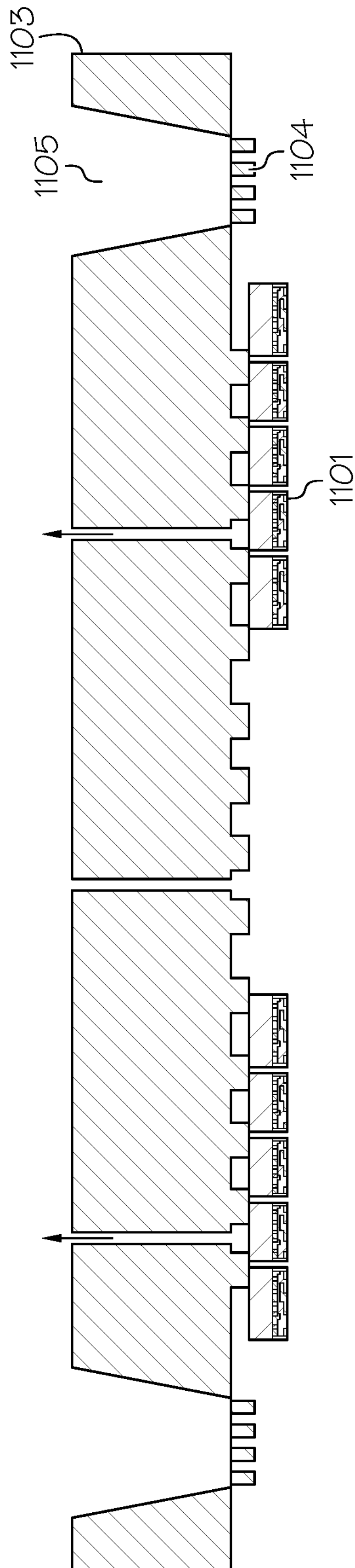



FIG. 11A

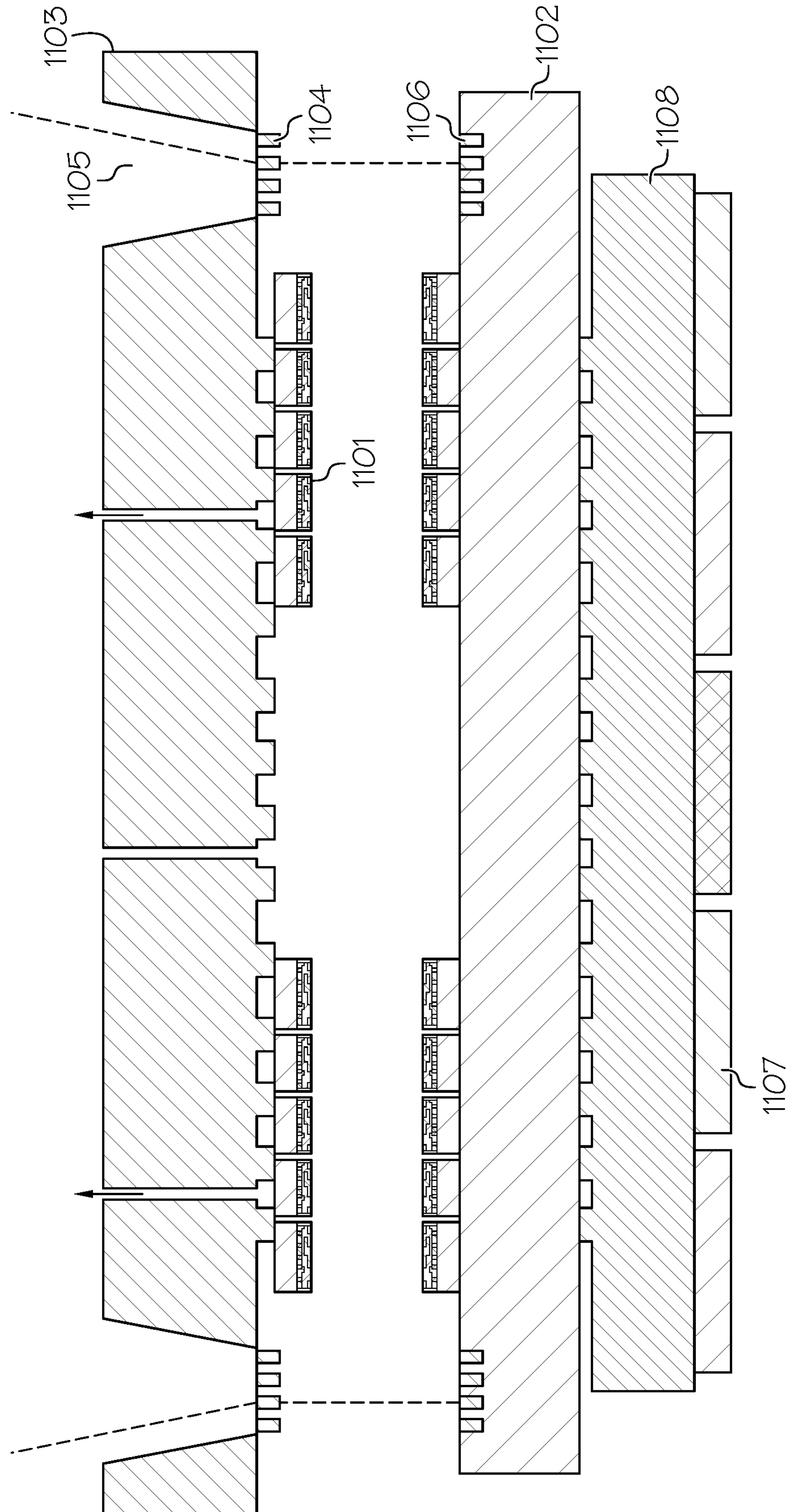
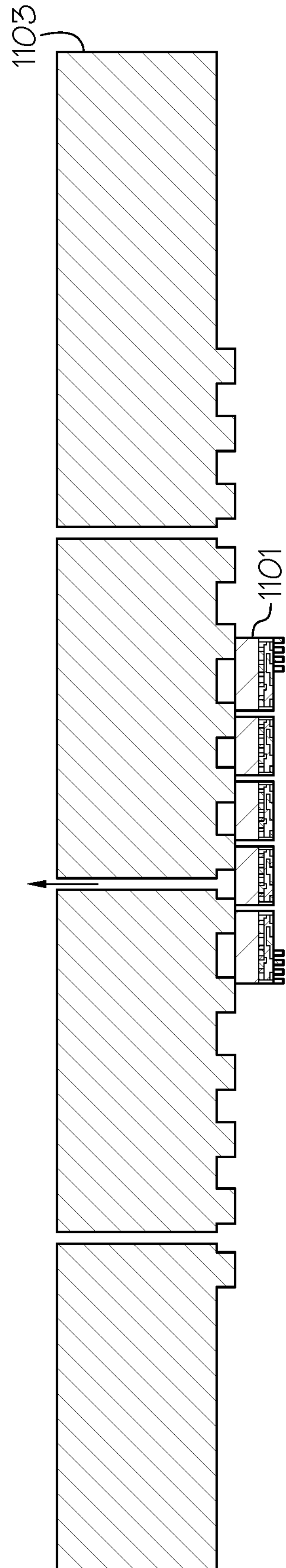



FIG. 11B



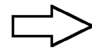



FIG. 12A

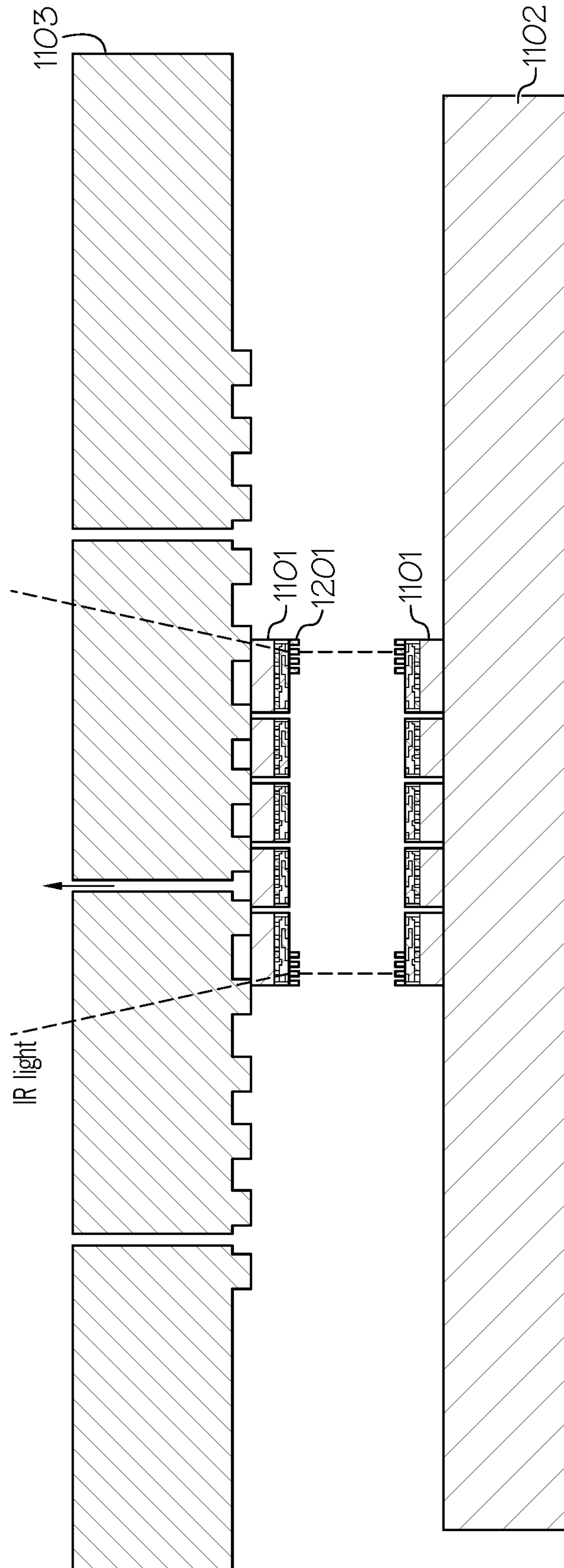



FIG. 12B

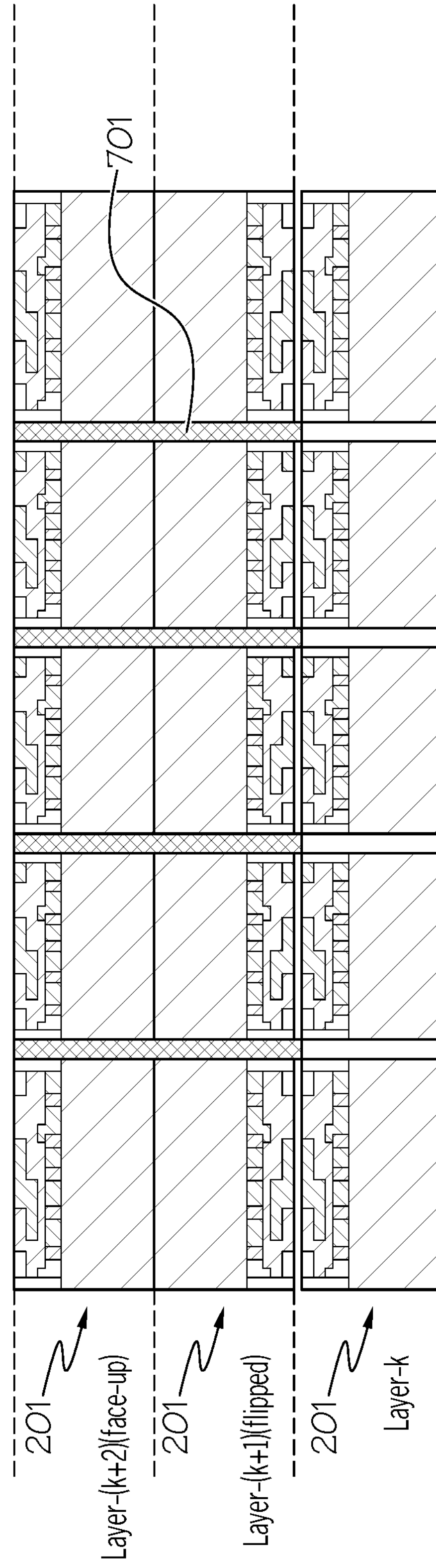



FIG. 13

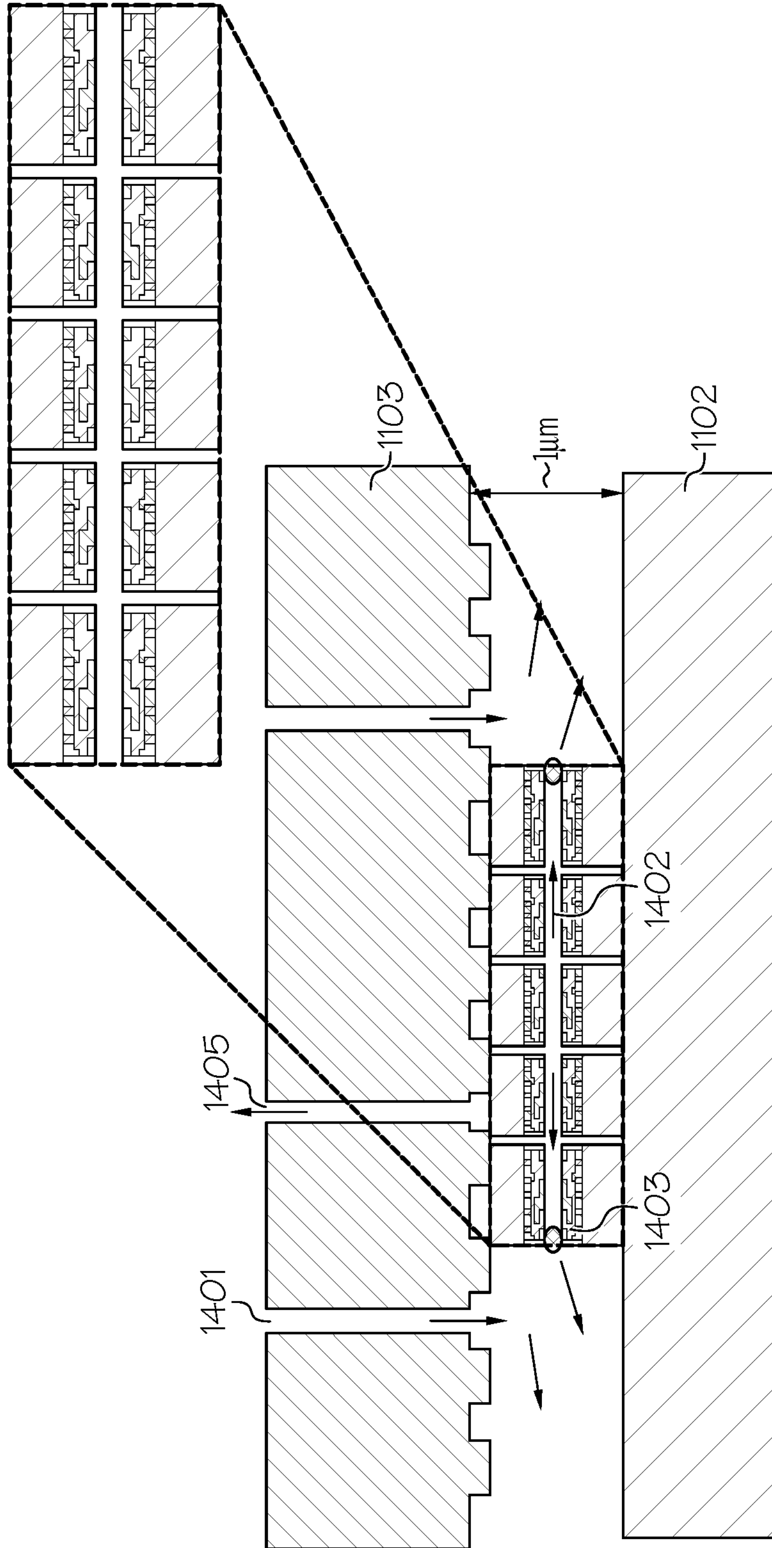



FIG. 14A

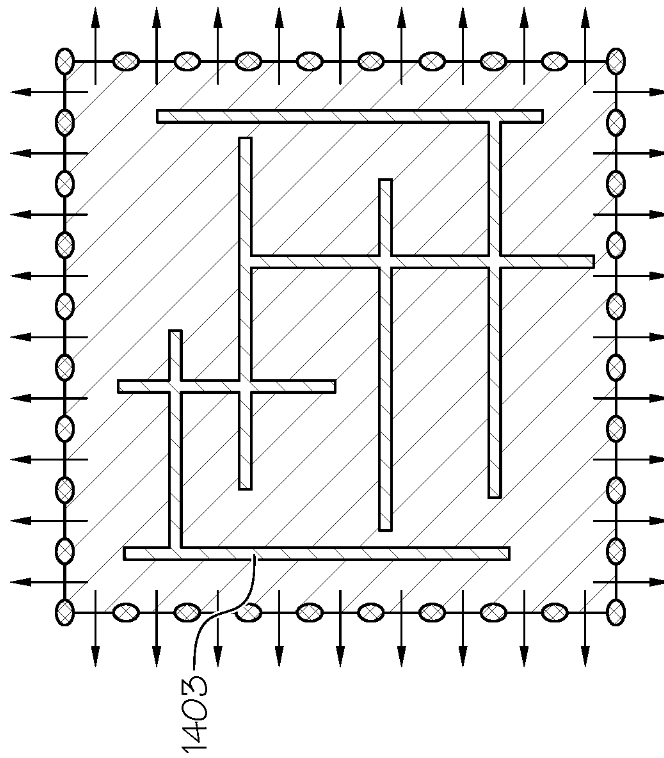



FIG. 14B

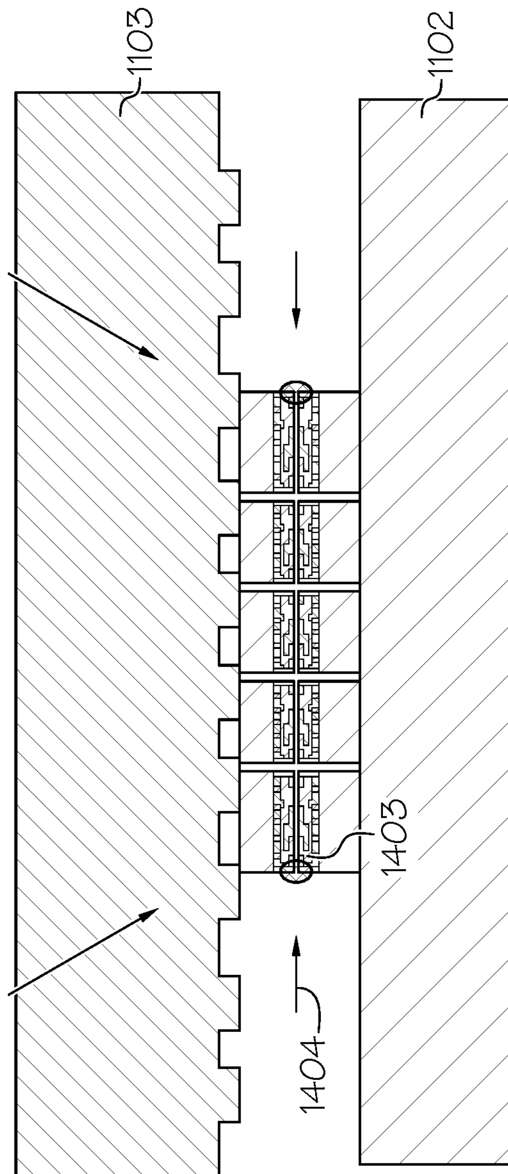



FIG. 14C

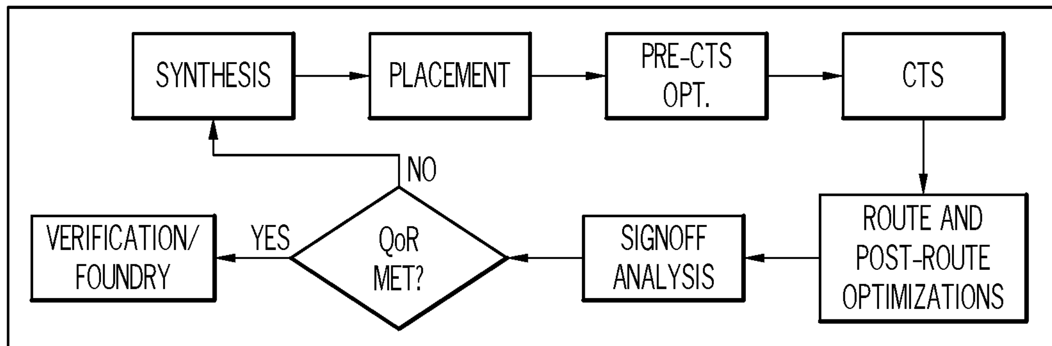



FIG. 15

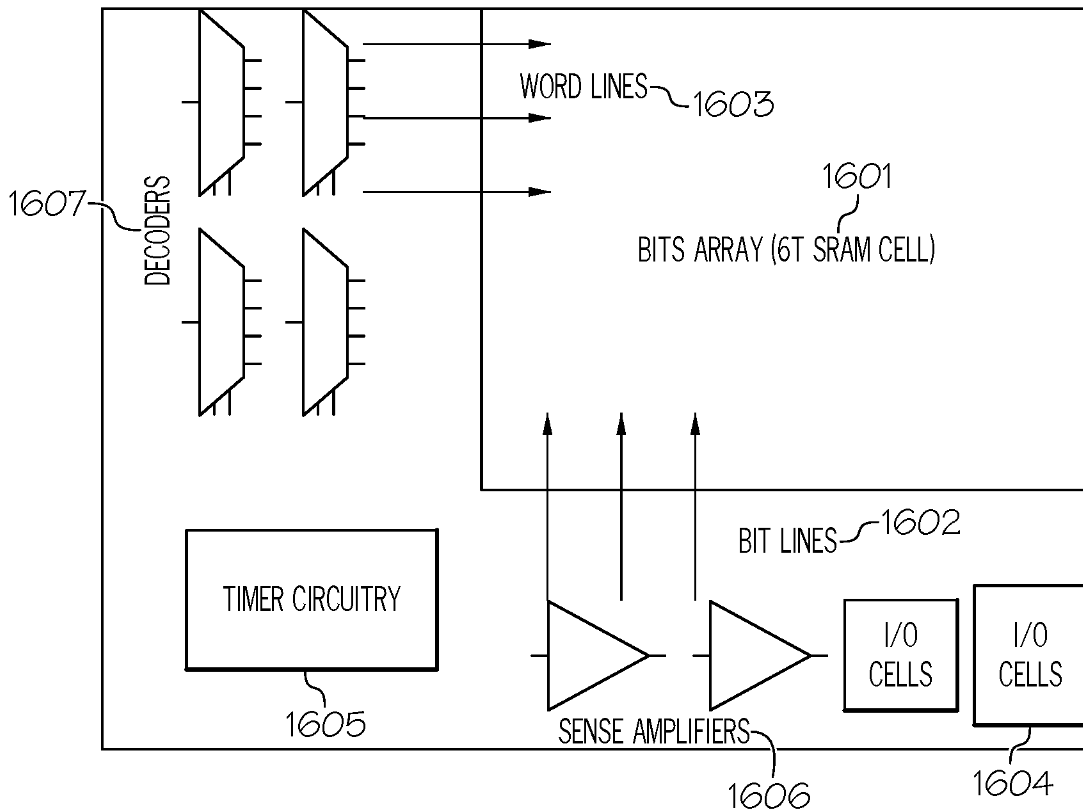



FIG. 16

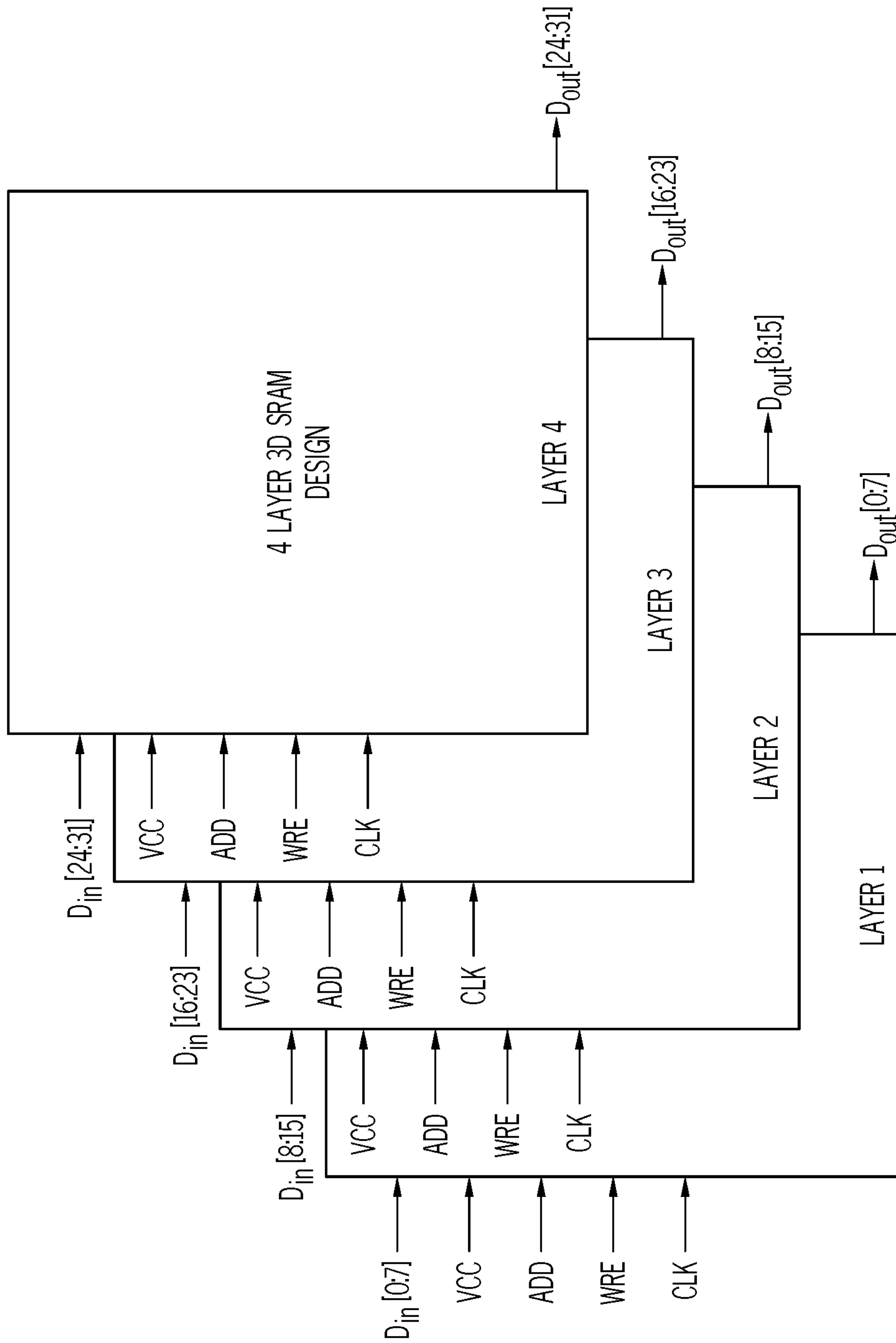



FIG. 17

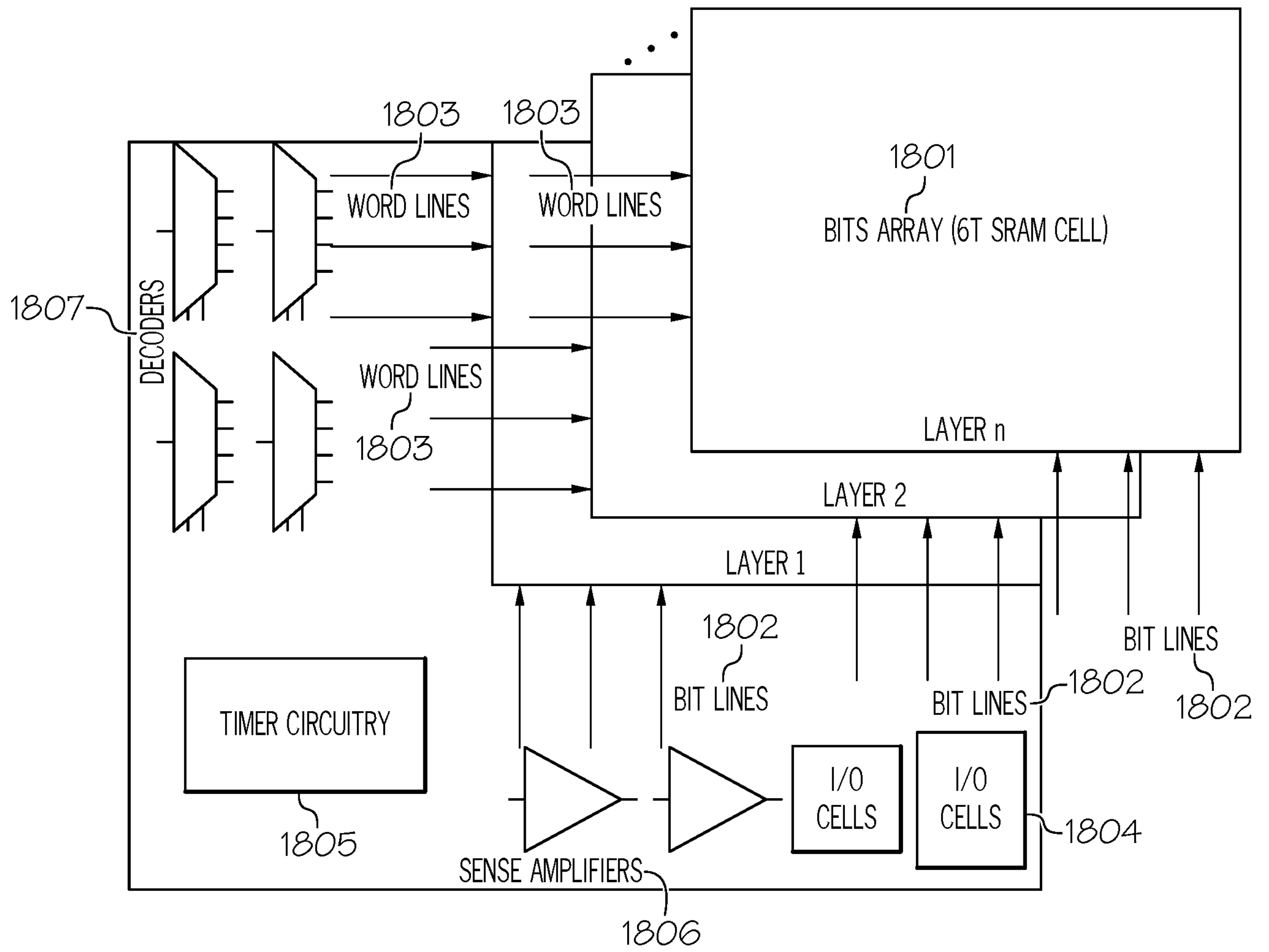



FIG. 18

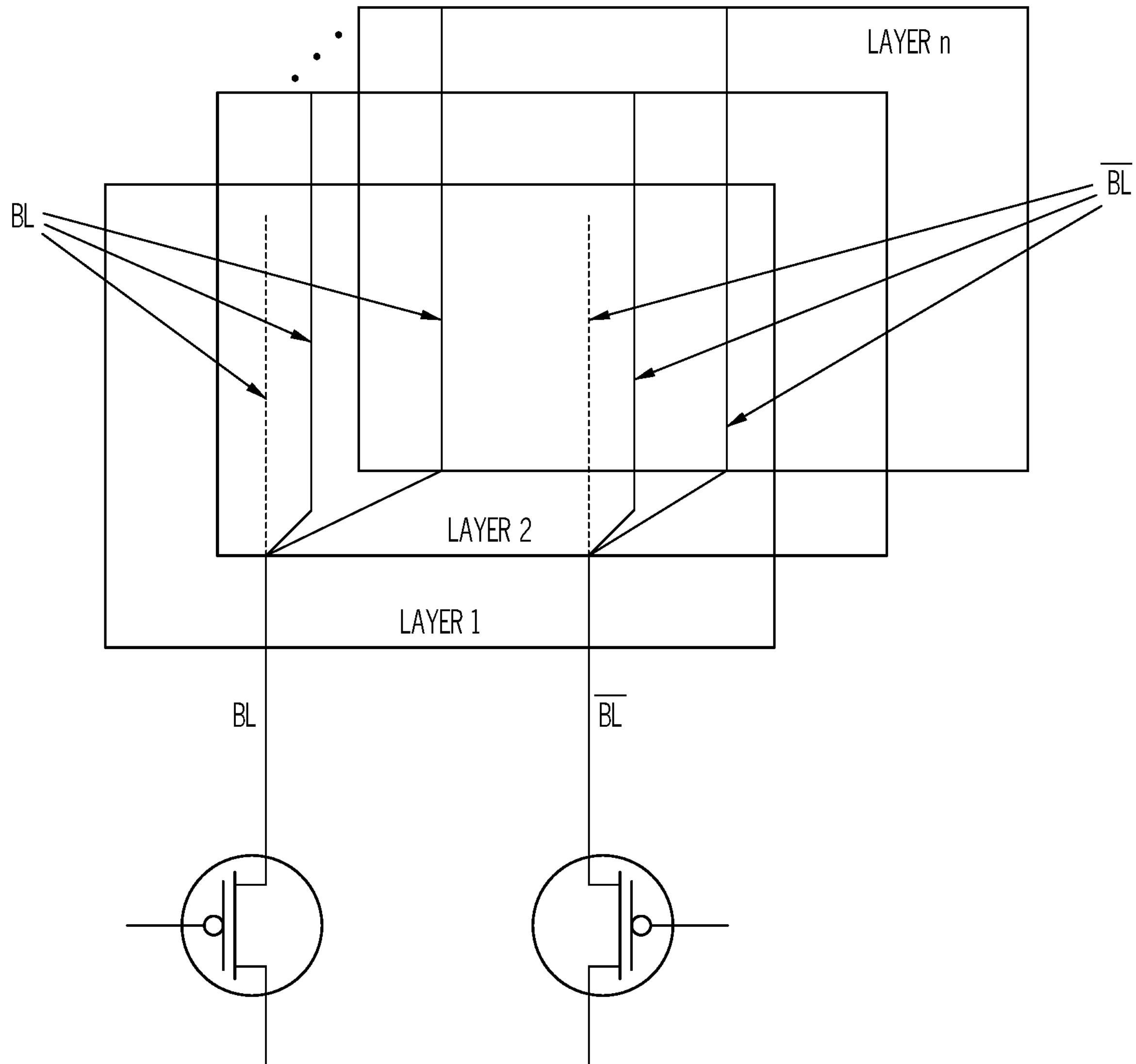



FIG. 19

## INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/067322

**A. CLASSIFICATION OF SUBJECT MATTER**
IPC(8) - H01L 23/525; H01L 23/538 (2019.01)
CPC - H01L 21/563; H01L 23/5383; H01L 23/538 (2019.02)

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 257/678; 257/686; 257/687 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X<br>---<br>Y | US 2016/0232271 A1 (QUALCOMM INCORPORATED) 11 August 2016 (11.08.2016) entire document | 27<br>---<br>28-30 |
| Y | US 2009/0212407 A1 (FOSTER et al) 27 August 2009 (27.08.2009) entire document | 1-26 |
| Y | US 2017/0077007 A1 (UTAC HEADQUARTERS PTE. LTD.) 16 March 2017 (16.03.2017) entire document | 1-18 |
| Y | US 2015/0163926 A1 (MICROCOSM TECHNOLOGY CO., LTD.) 11 June 2015 (11.06.2015) entire document | 1-9, 12, 15-18, 20, 23-26 |
| Y | US 9,640,531 B1 (MONOLITHIC 3D INC) 02 May 2017 (02.05.2017) entire document | 4, 5, 13, 14, 21, 22 |
| Y | US 9,190,392 B1 (SANDIA) 17 November 2015 (17.11.2015) entire document | 6-9, 15-18, 23-26 |
| Y | US 2017/0294415 A1 (MONOLITHIC 3D INC.) 12 October 2017 (12.10.2017) entire document | 19-26 |
| Y | US 2016/0276019 A9 (TAIWAN SEMICONDUCTOR MANUFACTURING CO., LTD.) 22 September 2016 (22.09.2016) entire document | 28-30 |
| Y | US 9,164,969 B1 (GUMASTE et al) 20 October 2015 (20.10.2015) entire document | 30 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 19 February 2019 | 04 MAR 2019 |

| Name and mailing address of the ISA/US | Authorized officer |
|---|---|
| Mail Stop PCT, Attn: ISA/US, Commissioner for Patents<br>P.O. Box 1450, Alexandria, VA 22313-1450<br>Facsimile No. 571-273-8300 | Blaine R. Copenheaver<br>PCT Helpdesk: 571-272-4300<br>PCT OSP: 571-272-7774 |