

METHOD OF MANUFACTURING STACKED CHIP PACKAGES

Claim Of Priority

[0001] Benefit of priority is hereby claimed to U.S. Patent Application Serial Number 11/583,411, filed on October 19, 2006, which application is herein incorporated by reference.

Technical Field

[0002] This application relates generally to semiconductor devices and device fabrication and, more particularly, to embodiments of multi-chip packages.

Background

[0003] An ever present goal in the semiconductor industry has been to decrease the size of devices and to increase the performance of devices. However, both of these goals present large technical hurdles as the two goals are often in conflict with each other.

[0004] As the minimum feature size achievable in semiconductor manufacturing decreases, the capacitive coupling between adjacent metal lines becomes a significant impediment to achieving higher performance. Further, as the minimum feature size decreases, the number of devices potentially achievable in a given area increases, as a second power function. The number of wiring connections is increasing at least as rapidly. In order to accommodate the increased wiring, the chip designer would like to shrink the space between adjacent lines to the minimum achievable dimension. This has the unfortunate effect of increasing the capacitive load.

[0005] One way to accommodate the increased wiring and reduce capacitive load is to substitute lower dielectric constant materials for the insulating material. A common insulating material to date is SiO_2 , which has a dielectric constant of around 4. SiO_2 is now used in most very large scale integrated circuit (VLSI) chips. Another way to accommodate the increased

wiring and reduce capacitive load is to shorten the distance between devices by denser packaging.

[0006] Multi-chip stacking has been proposed in the past; however, one significant concern with stacking chips (i.e. memory chips, logic chips, processor chips, etc.) is the manufacturing yield of the assembly. Because any given wafer will currently have a percentage of defective chips on its surface, when multiple wafers are stacked, for example 10 wafers high, the likelihood of at least one defective chip in any given stack goes up significantly.

Alternatively, stacking individual chips that have been tested and known to be good involves a greater amount of high precision handling when compared to stacking wafers prior to dicing. In either scenario, a high amount of loss due to manufacturing yield is present.

[0007] What are needed are methods and devices that improve the manufacturing yield of large numbers of stacked chips in a multi-chip assembly. What are also needed are improved methods and devices to enhance performance, reduce size, and improve other properties and features of multi-chip assemblies.

Brief Description of the Drawings

[0008] FIG. 1 shows an electronic system according to an embodiment of the invention.

[0009] FIG. 2A-2C show two wafers in various stages of chip assembly according to an embodiment of the invention.

[0010] FIG. 3A-3C show cross section views of wafers in various stages of chip assembly according to an embodiment of the invention.

[0011] FIG. 4 shows a multi-chip assembly according to an embodiment of the invention.

[0012] FIG. 5 shows another multi-chip assembly according to an embodiment of the invention.

Detailed Description

[0013] In the following detailed description of the invention, reference is made to the accompanying drawings that form a part hereof and in which is shown, by way of illustration, specific embodiments in which the invention may be practiced. In the drawings, like numerals describe substantially similar components throughout the several views. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, logical, and electrical changes may be made without departing from the scope of the present invention.

[0014] The terms "wafer" and "substrate" used in the following description include any structure having an exposed surface with which to form an integrated circuit (IC) structure of the invention. The term substrate is understood to include semiconductor wafers. The term substrate is also used to refer to semiconductor structures during processing and may include other layers, such as silicon-on-insulator (SOI), etc. that have been fabricated thereupon. Both wafer and substrate include doped and undoped semiconductors, epitaxial semiconductor layers supported by a base semiconductor or insulator, as well as other semiconductor structures well known to one skilled in the art. The term conductor is understood to include semiconductors and the term insulator or dielectric is defined to include any material that is less electrically conductive than the materials referred to as conductors.

[0015] The term "horizontal" as used in this application is defined as a plane parallel to the conventional plane or surface of a wafer or substrate, regardless of the orientation of the wafer or substrate. The term "vertical" refers to a direction perpendicular to the horizontal as defined above. Prepositions, such as "on," "side" (as in "sidewall"), "higher," "lower," "over," and "under" are defined with respect to the conventional plane or surface being on the top surface of the wafer or substrate, regardless of the orientation of the wafer or substrate.

[0016] Although the terms "memory chip" and "logic chip" are used in the following description, one of ordinary skill in the art will recognize that in

one embodiment, a chip may include both memory circuitry and logic circuitry on the same chip. A chip with both memory circuitry and logic circuitry on the same chip is defined to be both a "memory chip" and a "logic chip" as used in the following description. The following detailed description is, therefore, not to be taken in a limiting sense and the scope of the present invention is defined only by the appended claims, along with the full scope of equivalents to which such claims are entitled.

[0017] An embodiment of an information handling system such as a personal computer is included to show an embodiment of a high-level device application for the present invention. FIG. 1 is a block diagram of an information handling system 1 incorporating at least one multi-chip assembly 4 formed by methods in accordance with one embodiment of the invention. Information handling system 1 is merely one embodiment of an electronic system in which the present invention can be used. Other examples include, but are not limited to, personal data assistants (PDAs), cellular telephones, MP3 players, aircraft, satellites, military vehicles, etc.

[0018] In this example, information handling system 1 comprises a data processing system that includes a system bus 2 to couple the various components of the system. System bus 2 provides communications links among the various components of the information handling system 1 and may be implemented as a single bus, as a combination of busses, or in any other suitable manner.

[0019] Multi-chip assembly 4 is coupled to the system bus 2. Multi-chip assembly 4 may include any circuit or operably compatible combination of circuits. In one embodiment, multi-chip assembly 4 includes a processor 6 that can be of any type. As used herein, "processor" means any type of computational circuit such as, but not limited to, a microprocessor, a microcontroller, a graphics processor, a digital signal processor (DSP), or any other type of processor or processing circuit.

[0020] In one embodiment, a memory chip 7 is included in the multi-chip assembly 4. Those skilled in the art will recognize that a wide variety of memory chips may be used in the multi-chip assembly 4. Acceptable types of memory chips include, but are not limited to, Dynamic Random Access Memory

(DRAMs) such as SDRAMs, SLDRAMs, RDRAMs and other DRAMs.

Memory chip 7 can also include non-volatile memory such as flash memory.

[0021] In one embodiment, additional logic chips 8 other than processor chips are included in the multi-chip assembly 4. An example of a logic chip 8 other than a processor includes an analog to digital converter. Other circuits on logic chips 8 such as custom circuits, an application-specific integrated circuit (ASIC), etc. are also included in one embodiment of the invention.

[0022] Information handling system 1 may also include an external memory 11, which in turn can include one or more memory elements suitable to the particular application, such as one or more hard drives 12, and/or one or more drives that handle removable media 13 such as floppy diskettes, compact disks (CDs), digital video disks (DVDs), and the like.

[0023] Information handling system 1 may also include a display device 9 such as a monitor, additional peripheral components 10, such as speakers, etc. and a keyboard and/or controller 14, which can include a mouse, trackball, game controller, voice-recognition device, or any other device that permits a system user to input information into and receive information from the information handling system 1.

[0024] Figures 2A-2C show selected acts in a process of forming multi-chip packages according to an embodiment of the invention. In Figure 2A, a first wafer 210 and a second wafer 220 are shown. The first wafer 210 includes a number of individual chips 212 and the second wafer includes a number of individual chips 222.

[0025] In Figure 2B, the second wafer 220 is electrically coupled to the first wafer 210 to form a wafer assembly 200. In the operation, each of the multiple individual chips 212 are aligned and coupled with multiple individual chips 222. In wafer form, the joining process between chips 212 and 222 is easier and produces a higher manufacturing yield. Among other factors, the larger size makes handling and connection easier between chips on the first wafer 210 and the second wafer 220. Also the interconnection process joins multiple chips 212, 222 in a single operation.

[0026] In Figure 2B, the second wafer 220 is shown having a thickness 224. In one method of manufacture, the individual chips 222 are partially scored in the wafer 220 prior to attachment to the first wafer 210. In one method the scored side of the second wafer 220 is located at the interface between the first wafer 210 and the second wafer 220 after joining as shown in Figure 2B.

[0027] Figure 2C shows a stage in one method of manufacture where the second wafer 220 is thinned from thickness 224 as shown in Figure 2B to a thickness 226 as shown in Figure 2C. In one method of manufacture, the partial scoring of the second wafer 220 is completed by the thinning process in Figure 2C to allow separation of individual chips 222. Although partial scoring and thinning are discussed as one method to separate individual chips 222 from the second wafer 220, the invention is not so limited. One of ordinary skill in the art, having the benefit of the present disclosure, will recognize that other methods of separating individual chips 222 such as full thickness sawing after wafer connection, etc. are encompassed by the present invention.

[0028] In one embodiment, the second wafer 220 is positioned with individual chips 222 in flip-chip orientation to connect to the individual chips 212 on the first wafer 210. In one embodiment, the first wafer includes substantially all high-yield circuits to increase the likelihood that all or most of the individual chips 212 on the first wafer 210 are free of defects. An example of a high-yield circuit in one embodiment includes only re-distribution plane circuitry on the individual chips 212.

[0029] In one embodiment the second wafer includes memory chips such as dynamic random access memory, although the invention is not so limited. One method includes testing all individual chips 222 prior to connecting the second wafer 220 to the first wafer 210. Defective chips are identified during the testing process. As shown in Figure 2C, after individual chips 222 are coupled to the first wafer 210 and separated, defective chips can be removed from the wafer assembly 200 without disturbing the other chips 222 that remain connected to the first wafer 210. The removal of defective chips leaves gaps in the chips 222 such as first gap 234 and second gap 236. Good chips 230 and 232 can then be inserted into gap 234 and gap 236 to form a complete array of chips

222 without any defective chips present. Where chips 222 were thinned prior to defective chip removal, a number of previously thinned good chips are used to fill gaps such as 234 and 236.

[0030] The process described above and in Figures 2A-2C may be repeated to form multi-wafer stacks. Because defective chips are being removed and replaced as the stack progresses, a final wafer stack includes multiple chip stacks without any defective chips. The assembly process of stacking chips is made easier by stacking wafers, yet the manufacturing yield of a given wafer stack is significantly increased by replacing individual defective chips with good chips along the way during each additional wafer level assembly.

[0031] In one embodiment, a holding wafer is attached to the stack of wafers after the last functional wafer is added in the stack. The first wafer, such as the redistribution circuit wafer, can then be diced or separated by thinning, etc. Later, the holding wafer can be removed or diced, etc. to safely separate the individual multi-chip stacks. In one embodiment, the holding wafer attachment is temporary and the individual multi-chip stacks are later removed from the holding wafer. In one embodiment, the last wafer in the stack prior to the holding wafer is scored deeper than the prior wafers to provide for easier separation with less chip thinning.

[0032] Many different stacks of chips are contemplated, in accordance with embodiments of the invention, using the techniques described above. One contemplated embodiment of a chip stack includes a high-yield redistribution circuit chip on the bottom, with multiple memory chips stacked on top, with a logic chip such as a processor chip stacked on top. In one embodiment, a logic chip in the stack includes both logic circuitry and static random access memory. Other contemplated configurations include the logic chip located adjacent to the redistribution circuit chip. Specific reasons for such a configuration, such as multiple processor embodiments, are discussed below. Although a chip stack including a redistribution circuit chip, at least one memory chip, and a logic chip are described herein, the invention is not so limited. Techniques described above are useful for manufacturing any stack of more than one chip.

[0033] Figures 3A-3C show side views of a chip manufacturing operation according to embodiments of the invention. A first chip 310 is shown oriented in flip-chip orientation in Figure 3A. The first chip 310 includes a number of devices 332 such as transistors, memory cells, etc. The devices 332 are shown formed in a substrate 330 with a thickness 316. Horizontal redistribution circuitry 340, such as metal trace lines, and vias or contacts 338 are shown interconnecting devices 332 within the first chip 310. One or more regions of insulator 334, such as silicon dioxide, may be used to electrically isolate the redistribution circuitry 340.

[0034] In one embodiment, the first chip 310 includes a first vertical communication channel 344 and a second vertical communication channel 346. The first vertical communication channel 344 includes a via, or the like. Several types of communication channel 344 are within the scope of the invention, including metal conductors, optical conductors, etc. The second vertical communication channel 346 illustrates a coaxial conductor embodiment. As shown in Figure 3A, prior to connection to another wafer, the vertical communication channels do not pass through an entire thickness of the first chip 310.

[0035] Figure 3B shows the first chip 310 coupled in flip-chip orientation to a second chip 320 using connection structures 312. In one embodiment, the second chip 320 includes a high-yield chip such as a redistribution circuit chip and the first chip 310 includes a memory chip. However, the description in Figures 3A-3C is applicable to a connecting operation between any two chips in a chip stack.

[0036] In one embodiment, the connection structures 312 include solder structures such as controlled chip collapse connections (C4). Any of a number of pad metallurgies are possible to support the connection structures 312. One pad metallurgy includes TiNiCuAu pads. In one embodiment TiNiCuAu pads are formed using a liftoff process. One advantage of TiNiCuAu pads formed using a liftoff process includes the ability to form smaller contacts, thus permitting a more dense interconnection configuration. Although solder connection structures 312 are described, the invention is not so limited. Other

possible connection structures 312 include gold to gold bonding connection structures, conductive epoxies, conductor filled epoxies, conductive films, etc.

[0037] Figure 3B shows an embodiment where the connection structures 312 leave a gap 314 between the first chip 310 and the second chip 320. In one embodiment the gap 314 is filled with a solid material. In other embodiments the gap 314 is left open. Filled embodiments may use adhesives such as epoxy or other filler material to provide additional mechanical strength to a final multi-chip assembly.

[0038] In embodiments where the gap 314 is left open, a cooling fluid such as a gas or liquid is able to pass between individual chips in a chip stack to enhance cooling during operation. In one electronic system embodiment, a cooling system utilizing the gaps 314 is coupled to the system to cool a chip stack formed according to methods described. In one embodiment, the cooling system uses a gas or a compressed gas to flow through the gaps 314 and cool the individual chips. One compressed gas includes compressed hydrogen. One compressed gas includes compressed helium. In selected embodiments, a composite gas of hydrogen and helium is used for cooling. Liquid cooling fluids contemplated for implementation of embodiments of the invention include volatile liquids and supercritical fluids such as liquid carbon dioxide.

[0039] Figure 3C shows the first chip 310 that has been thinned from thickness 316 as shown in Figure 3B to thickness 318. In addition to completing a scoreline for chip separation as described above, thinning chips in a wafer may be used to expose circuitry for additional chip interconnectivity from the back side of the wafer. Figure 3C shows the first vertical communication channel 344 and the second vertical communication channel 346 as described above. In one embodiment, the thinning process exposes vertical communication channels in the form of through-wafer interconnects to permit through chip interconnectivity.

[0040] One advantage of through chip interconnectivity as opposed to chip edge interconnectivity includes the increased density of interconnectivity that is possible. Chip edge connections between chips in a multi-chip stack are limited by the amount of chip edge real estate. In contrast, through chip

connections can pass through any available location within the plane of each chip. In one embodiment, both chip edge connections and through chip connections are used to communicate between chips in a multi-chip stack. Although through chip interconnections are described, embodiments using only chip edge connections are also within the scope of the invention.

[0041] Figure 4 shows one embodiment of a chip stack 400 formed using methods described above. In one embodiment, a bottom chip 410 includes a high-yield chip such as a redistribution circuitry chip. A number of memory chips 420 are shown located on top of the redistribution circuitry chip 410. A logic chip 430 such as a processor chip is stacked on top of the chip stack 400. In one embodiment, nine memory chips 420 are stacked on top of the bottom chip 410 with a single logic chip 430 stacked on top of the memory chips 420. Other embodiments include multiple redistribution circuitry chips in a single stack 400, multiple logic chips 430 in a single stack 400, and any of a number of different stacking orders for the stack 400. Placing a redistribution circuitry chip first provides a high-yield base wafer that increases manufacturing yield as described above. Placing at least one logic chip 430 on an outside surface of the stack 400 allows for increased cooling of one of the hotter running chips in the stack.

[0042] Figure 5 shows a chip stack 500 including multiple logic chips such as multiple processors. Two or more redistribution circuitry chips 512 are again used as a base, with two or more logic chips 510 mounted on top of the redistribution circuitry chips 512. A number of memory chips 520 are then stacked on top of the two or more logic chips 510.

[0043] While a number of advantages of embodiments of the invention are described, the above lists are not intended to be exhaustive. Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement that is calculated to achieve the same purpose may be substituted for the specific embodiment shown. This application is intended to cover any adaptations or variations of the present invention. It is to be understood that the above description is intended to be illustrative and not restrictive. Combinations of the

above embodiments, and other embodiments, will be apparent to those of skill in the art upon reviewing the above description. The scope of the invention includes any other applications in which the above structures and methods are used. The scope of the invention should be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

What is claimed is:

1. A method of forming a semiconductor structure, comprising:
attaching a first wafer to a second wafer, the first wafer including only redistribution circuitry, and the second wafer including a plurality of semiconductor chips;
physically removing at least one selected semiconductor chip from the surface of the first wafer; and
physically replacing the removed semiconductor chips with at least one replacement semiconductor chip.
2. The method of claim 1, further comprising selecting the plurality of semiconductor chips to include memory chips.
3. The method of claim 2, further comprising selecting the memory chips to include DRAM chips.
4. The method of claim 1, wherein attaching a first wafer to a second wafer includes attaching a first wafer to a second wafer after configuring the second wafer to include logic chips.
5. The method of claim 4, further comprising selecting the logic chips to include static random access memory.
6. The method of claim 1, wherein attaching a first wafer to a second wafer includes coupling input/output regions between the chips using controlled chip collapse connection structures.
7. The method of claim 1, wherein attaching a first wafer to a second wafer includes coupling input/output regions between the chips using gold to gold bonding connection structures.
8. The method of claim 1, wherein attaching a first wafer to a second wafer includes coupling input/output regions between the chips using pad metallurgy limiting lift off connection structures.
9. The method of claim 8, wherein attaching a first wafer to a second wafer includes coupling input/output regions between the chips using TiNiCuAu connection structures.
10. A method of forming a semiconductor structure, comprising:

attaching a first wafer to a second wafer, the first wafer including only redistribution circuitry, and the second wafer including a first plurality of semiconductor chips; physically removing at least one first selected semiconductor chip of the first plurality from the surface of the first wafer; physically replacing the first removed semiconductor chip with at least one first replacement semiconductor chip; attaching a third wafer to the second wafer, the third wafer including a second plurality of semiconductor chips; removing at least one second selected semiconductor chip of the second plurality from the surface of the second wafer; and replacing the second removed semiconductor chips with at least one second replacement semiconductor chip.

11. The method of claim 10, wherein attaching the first wafer to the second wafer includes attaching a first wafer to a second wafer, the second wafer including a first plurality of memory chips.

12. The method of claim 10, wherein attaching the third wafer to the second wafer includes attaching a third wafer to a second wafer, the third wafer including a second plurality of logic chips.

13. The method of claim 10, wherein attaching the third wafer to the second wafer includes attaching a third wafer to a second wafer, the third wafer including a second plurality of logic chips with static random access memory located on the logic chips.

14. The method of claim 10, wherein attaching a first wafer to a second wafer includes coupling input/output regions between the chips using pad metallurgy limiting lift off connection structures.

15. The method of claim 14, wherein attaching a first wafer to a second wafer includes coupling input/output regions between the chips using TiNiCuAu connection structures.

16. A method of forming a semiconductor structure, comprising:
stacking and electrically interconnecting a number of semiconductor wafers, including:
stacking a number of memory wafers and at least one redistribution circuitry wafer on a starting wafer;
attaching at least one logic wafer to the memory wafers;
wherein selected chips in at least one of the number of semiconductor wafers are physically removed and physically replaced during the stacking process; and

separating individual stacks of chips from the number of stacked and interconnected semiconductor wafers, wherein separating individuals stacks of chips includes attaching a temporary holder wafer to the logic wafer.

17. The method of claim 16, wherein stacking a number of memory wafers includes stacking a number of dynamic random access memory wafers.

18. The method of claim 16, wherein stacking a number of memory wafers includes stacking nine memory wafers.

19. The method of claim 16, further comprising thinning at least some of the wafers prior to subsequent wafer attachment.

20. A method of forming a semiconductor structure, comprising:
attaching a first wafer to a second wafer, the first wafer including only redistribution circuitry, and the second wafer including a plurality of semiconductor chips mounted with an active side towards the first wafer;
thinning a backside of the second wafer to a first thickness to expose interconnection circuitry in the plurality of semiconductor chips;
removing selected semiconductor chips from the surface of the second wafer; and
replacing the removed semiconductor chips with replacement semiconductor chips wherein the replacement chips include the first thickness.

21. The method of claim 20, wherein the plurality of semiconductor chips includes memory chips.

22. The method of claim 20, further including attaching a third wafer to the second wafer, the third wafer including a plurality of logic chips.

23. The method of claim 20, further including depositing contact metallurgy on the exposed interconnection circuitry.

24. The method of claim 23, wherein depositing contact metallurgy on the exposed interconnection circuitry includes depositing a gold metallurgy on exposed interconnection circuitry.

25. The method of claim 23, wherein depositing contact metallurgy on the exposed interconnection circuitry includes using a lift off process to deposit contact metallurgy on the exposed interconnection circuitry.

26. The method of claim 25, wherein depositing contact metallurgy on the exposed interconnection circuitry includes depositing a TiNiCuAu metallurgy on exposed interconnection circuitry.

27. A method of forming an electronic system, comprising:
forming a semiconductor structure, including:
stacking and interconnecting a number of semiconductor wafers, including:
 stacking a number of memory wafers on a starting wafer;
 attaching at least one logic wafer to the memory wafers;
 wherein selected chips in one or more of the number of semiconductor wafers are removed and replaced during the stacking process;
 separating individual stacks of chips from the number of semiconductor wafers;
coupling the semiconductor structure to a system circuit, including a display screen and an input device; and
coupling a cooling system to the semiconductor structure.

28. The method of claim 27, wherein coupling the semiconductor structure to a system circuit, including a display screen and an input device includes coupling the semiconductor structure to a system circuit, including a display screen and a keyboard.

29. The method of claim 27, wherein coupling the semiconductor structure to a system circuit, including a display screen and an input device includes coupling the semiconductor structure to a system circuit, including a display screen and a mouse.

30. The method of claim 27, wherein coupling a cooling system to the semiconductor structure includes flowing a cooling fluid through gaps between chips and interconnection structures spaced between the chips.

31. The method of claim 30, wherein flowing a cooling fluid through gaps includes flowing a pressurized gas through gaps.

32. The method of claim 31, wherein the gas includes hydrogen.

33. The method of claim 32, wherein the gas includes helium.

34. The method of claim 30, wherein flowing a cooling fluid through gaps includes flowing a supercritical fluid through gaps.

35. The method of claim 30, wherein flowing a cooling fluid through gaps includes flowing a liquid through gaps.