US 20130049824A1

(54) **3D CHIP STACK SKEW REDUCTION WITH RESONANT CLOCK AND INDUCTIVE COUPLING**

(75) Inventors: **JAE-JOON KIM,** YORKTOWN HEIGHTS, NY (US); **YU-SHIANG LIN,** ELMSFORD, NY (US); **LIANG-TECK PANG,** WHITE PLAINS, NY (US); **JOEL A. SILBERMAN,** SOMERS, NY (US)

(73) Assignee: **INTERNATIONAL BUSINESS MACHINES CORPORATION,** ARMONK, NY (US)

## Publication Classification

(57) **ABSTRACT**

There is provided a clock distribution network for synchronizing global clock signals within a 3D chip stack having two or more strata. The clock distribution network includes a plurality of clock distribution circuits, each being arranged on a respective one of the two or more strata for providing the global clock signals to various chip locations. Each of the plurality of clock distribution circuits includes a resonant circuit for providing stratum-to-stratum coupling for the clock distribution network. The resonant circuit includes at least one capacitor and at least one inductor.
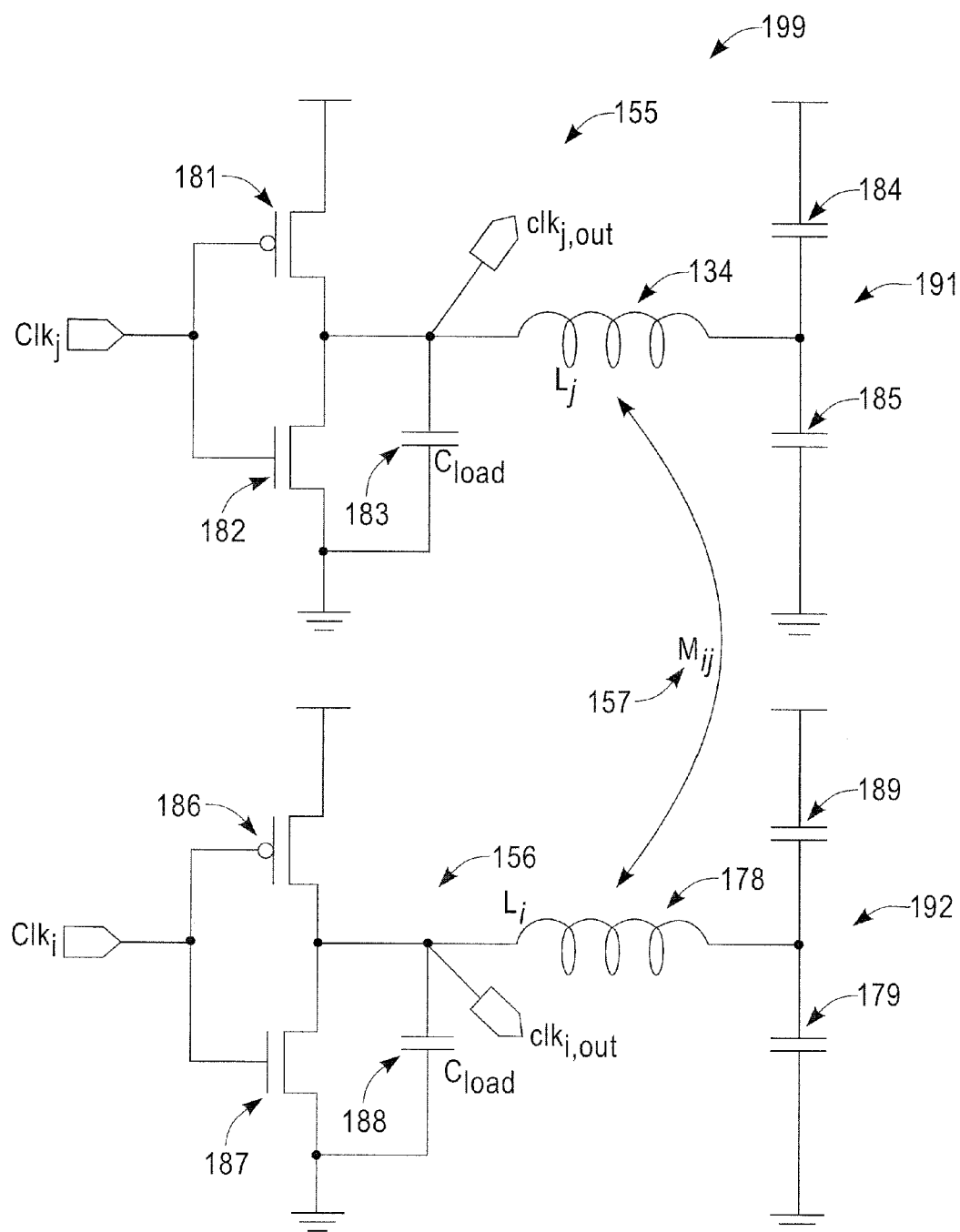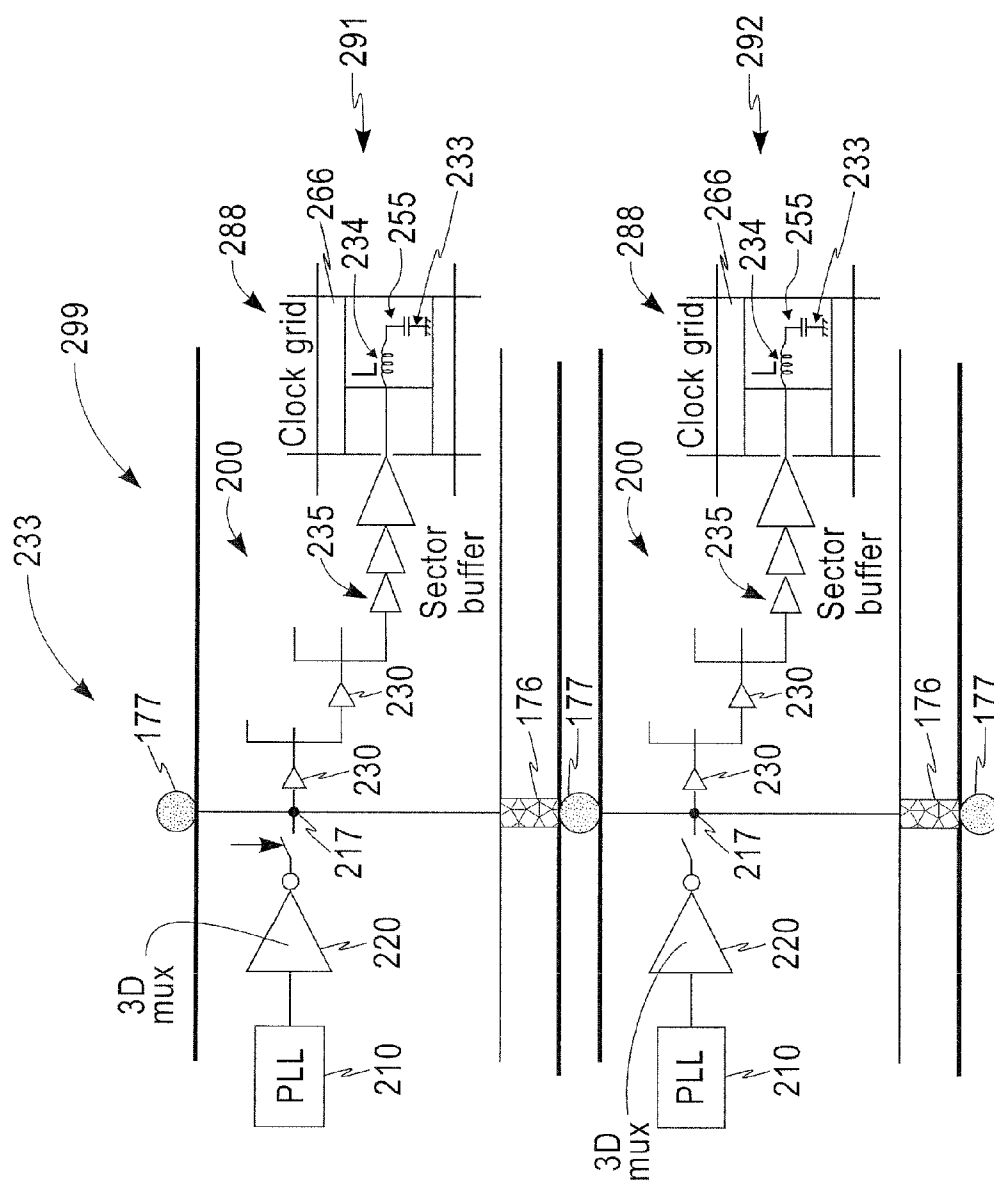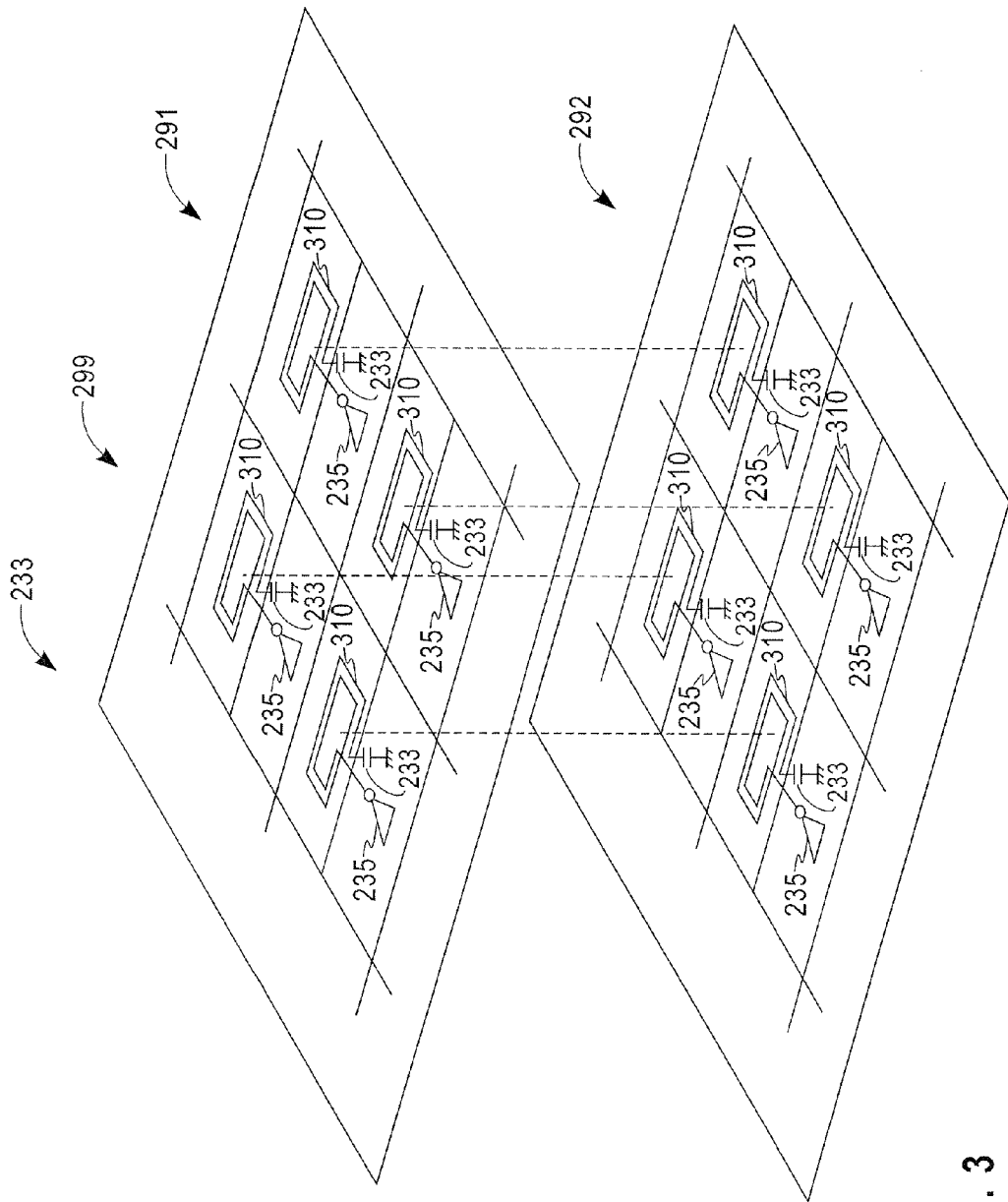
**FIG. 1**

**FIG. 2**

FIG. 3

235 — 299

310

Front side ⟶

176

310

Back side ⟶

Front side ⟶

291

177     177

235     310     292

**FIG. 4**

Start                                    500

Arrange each clock distribution circuit in a set of clock
distribution circuits on a respective one of the strata for
providing the global clock signals to various chip locations.          510

Configure a resonant circuit included in each clock distribution
circuit to provide stratum-to-stratum coupling for the clock
distribution network (e.g., by aligning the coils of inductors
on different strata to minimize skew).          520

Enhance the coupling by adding a back-side inductor having
inductor coils aligned with the inductor coils on the front-sides.          530
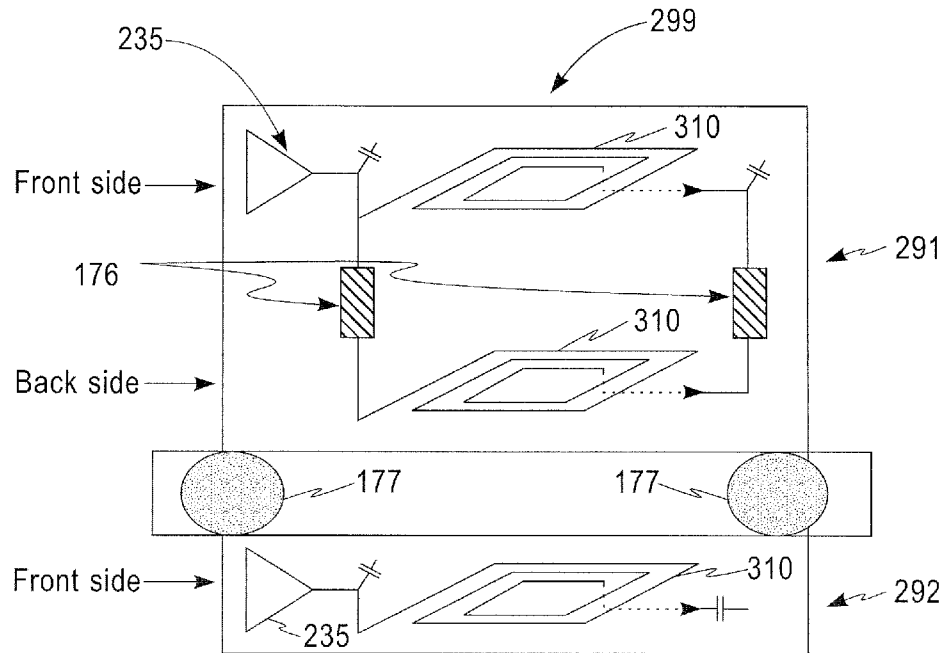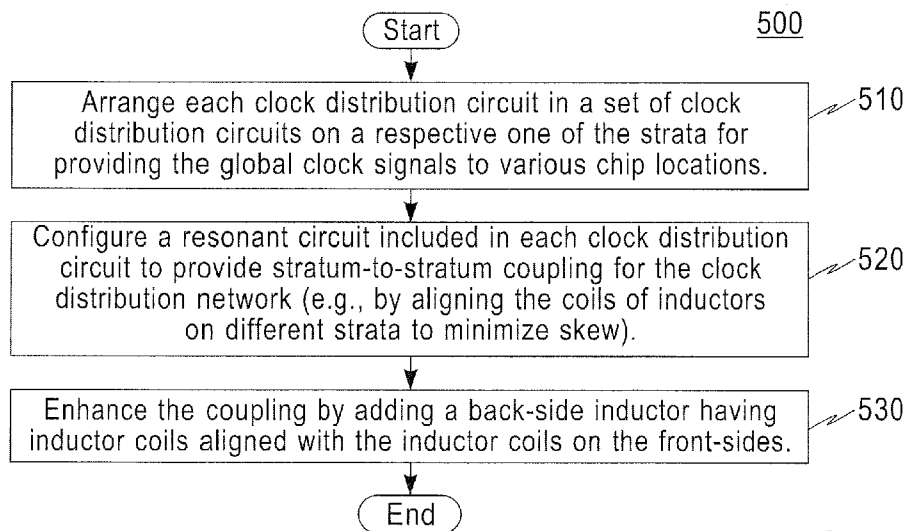
End
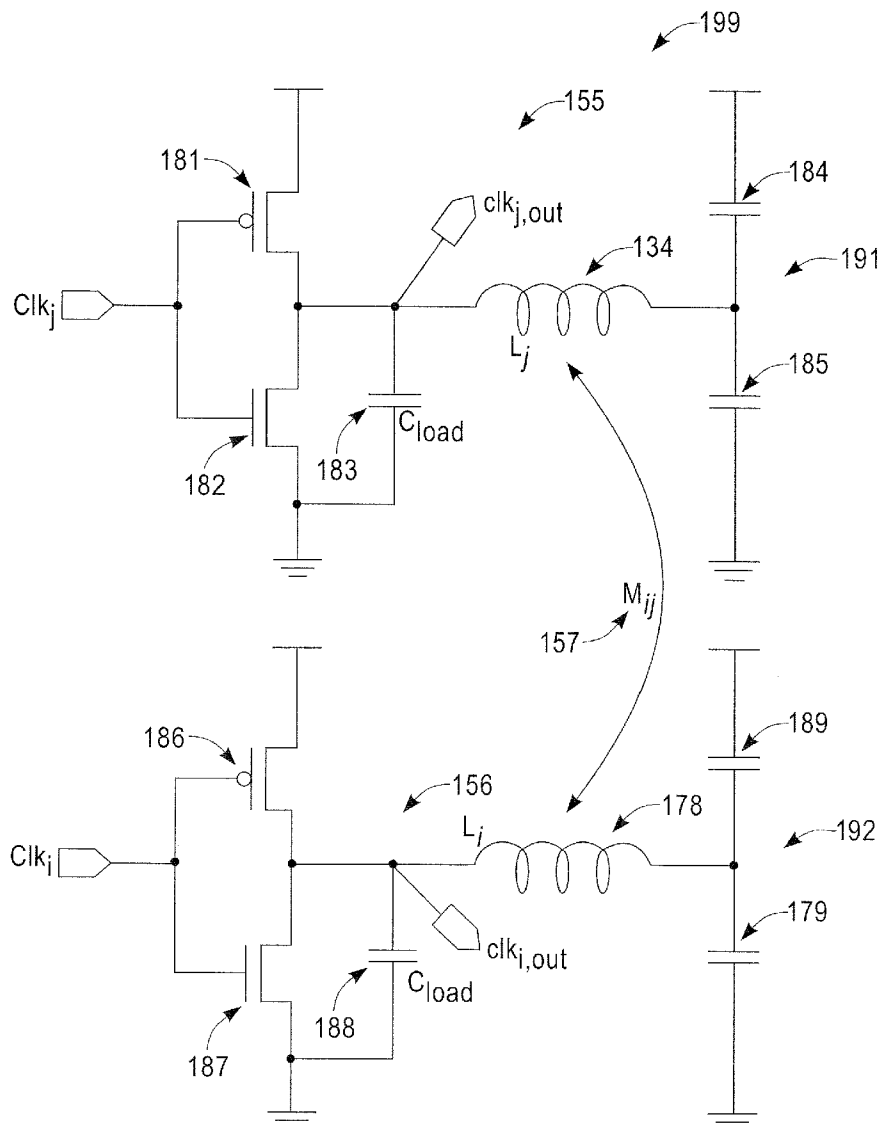
**FIG. 5**

# 3D CHIP STACK SKEW REDUCTION WITH RESONANT CLOCK AND INDUCTIVE COUPLING

## CROSS-REFERENCE TO RELATED APPLICATIONS

[0001]  This application is related to the following commonly assigned applications, all concurrently filed herewith and incorporated herein by reference: Attorney Docket Number YOR920110384US1 (163-430), entitled "PROGRAMMING THE BEHAVIOR OF INDIVIDUAL CHIPS OR STRATA IN A 3D STACK OF INTEGRATED CIRCUITS"; Attorney Docket Number YOR920110385US1 (163-431), entitled "SYNCHRONIZING GLOBAL CLOCKS IN 3D STACKS OF INTEGRATED CIRCUITS BY SHORTING THE CLOCK NETWORK"; Attorney Docket Number YOR920110387US1 (163-433), entitled "3D INTEGRATED CIRCUIT STACK-WIDE SYNCHRONIZATION CIRCUIT"; Attorney Docket Number YOR920110388US1 (163-434), entitled "CONFIGURATION OF CONNECTIONS IN A 3D STACK OF INTEGRATED CIRCUITS"; Attorney Docket Number YOR920110390US1 (163-435), entitled "3D INTER-STRATUM CONNECTIVITY ROBUSTNESS"; Attorney Docket Number YOR920110391US1 (163-436), entitled "AC SUPPLY NOISE REDUCTION IN A 3D STACK WITH VOLTAGE SENSING AND CLOCK SHIFTING"; and Attorney Docket Number YOR920110449US1 (163-441), entitled "VERTICAL POWER BUDGETING AND SHIFTING FOR 3D INTEGRATION".

## BACKGROUND

[0002]  1. Technical Field
[0003]  The present invention relates generally to integrated circuits and, in particular, to 3D chip stack skew reduction with resonant clock and inductive coupling.
[0004]  2. Description of the Related Art
[0005]  Three-dimensional (3D) stacked chips include two or more electronic integrated circuit chips stacked one on top of the other. Three-dimensional integration offers the advantages of heterogeneous design and higher input/output (I/O) density for designers. However, clock distribution across the chip can be challenging considering wafer-to-wafer variation and even stacking chips with different technologies.

## SUMMARY

[0006]  According to an aspect of the present principles, there is provided a clock distribution network for synchronizing global clock signals within a 3D chip stack having two or more strata. The clock distribution network includes a plurality of clock distribution circuits, each being arranged on a respective one of the two or more strata for providing the global clock signals to various chip locations. Each of the plurality of clock distribution circuits includes a resonant circuit for providing stratum-to-stratum coupling for the clock distribution network. The resonant circuit includes at least one capacitor and at least one inductor.
[0007]  According to another aspect of the present principles, there is provided a method for synchronizing global clock signals in a clock distribution network in a 3D chip stack having two or more strata. The method includes arranging each of a plurality of clock distribution circuits on a respective one of the two or more strata to provide the global

clock signals to various chip locations. The method further includes configuring a respective resonant circuit included in each of the plurality of clock distribution circuits to provide stratum-to-stratum coupling for the clock distribution network. The resonant circuit includes at least one capacitor and at least one inductor.

[0008]  According to yet another aspect of the present principles, there is provided a clock distribution network for synchronizing global clock signals within a 3D chip stack having two or more strata. The clock distribution network includes a plurality of clock grids, each being arranged on a respective one of the two or more strata for providing the global clock signals to various chip locations. The clock distribution network further includes a plurality of buffered clock trees, each being arranged on the respective one of the two or more strata for driving a respective one of the plurality of clock grids on a same one of the two or more strata and having at least a root and a plurality of clock buffers. The clock distribution network also includes a plurality of multiplexers, each being arranged on the respective one of the two or more strata. The plurality of multiplexers are for providing a same single clock source to the root of each of the plurality of buffered clock trees. Each of the plurality of clock grids includes at least one sector having a resonant circuit therein for providing stratum-to-stratum coupling for the clock distribution network.

[0009]  According to still another aspect of the present principles, there is provided a method for synchronizing global clock signals in a clock distribution network in a 3D chip stack having two or more strata. The method includes arranging each of a plurality of clock grids on a respective one of the two or more strata to provide the global clock signals to various chip locations. The method further includes arranging each of a plurality of buffered clock trees on the respective one of the two or more strata to drive a respective one of the plurality of clock grids on a same one of the two or more strata. Each of the plurality of buffered clock trees has at least a root and a plurality of clock buffers. The method also includes arranging each of a plurality of multiplexers on the respective one of the two or more strata. The plurality of multiplexers are configured to provide a same single clock source to the root of each of the plurality of buffered clock trees. The method additionally includes configuring a resonant circuit included in each of the plurality of clock grids to provide stratum-to-stratum coupling for the clock distribution network.

[0010]  These and other features and advantages will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

## BRIEF DESCRIPTION OF DRAWINGS

[0011]  The disclosure will provide details in the following description of preferred embodiments with reference to the following figures wherein:

[0012]  FIG. 1 shows resonant circuits 155 and 156 in a 3D chip stack 199, in accordance with an embodiment of the present principles;

[0013]  FIG. 2 shows a clock distribution circuit 233 with resonant circuits 255 for a 3D chip stack 299, in accordance with an embodiment of the present principles;

2

[0014] FIG. 3 shows an alignment 300 between the coils 310 of the inductors 234 of the resonant circuits 255 of FIG. 2, in accordance with an embodiment of the present principles;

[0015] FIG. 4 shows a coupling enhancement 400 for the resonant circuits 255 of FIG. 2, in accordance with an embodiment of the present principle; and

[0016] FIG. 5 shows a method 500 for reducing skewing of global clock signals within a 3D chip stack that includes two or more strata, in accordance with an embodiment of the present principles.

## DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0017] The present principles are directed to 3D chip stack skew reduction with resonant clock and inductive coupling.

[0018] FIG. 1 shows resonant circuits 155 and 156 in a 3D chip stack 199, in accordance with an embodiment of the present principles. Stack 199 includes a stratum-0 191 and a stratum-1 192.

[0019] Regarding stratum-0 191, a clock signal elk is applied to the respective gates of a p-channel MOSFET 181 and an n-channel MOSFET 182. A source of the p-channel MOSFET 181 is connected to a voltage or current source (not shown). A drain of the p-channel MOSFET 181 is connected to the drain of the n-channel MOSFET 182, one side of a capacitor $C_{load}$ 183, and one side of an inductor $L_j$ 134, forming an output resonant clock node $clk_{j,out}$. The source of the n-channel MOSFET 182 is connected to ground and to the other side of capacitor $C_{load}$ 183. The other side of inductor $L_j$ 134 is connected to one side of a capacitor 184 and one side of a capacitor 185. The other side of capacitor 184 is connected to a voltage or current source (not shown). The other side of capacitor 185 is connected to ground. Inductor $L_j$ 134 and capacitor $C_{load}$ 183 form a resonant circuit 155 on stratum-0 191.

[0020] Regarding stratum-1 192, a clock signal $clk_i$ is applied to the respective gates of a p-channel MOSFET 186 and an n-channel MOSFET 187. A source of the p-channel MOSFET 186 is connected to a voltage or current source (not shown). A drain of the p-channel MOSFET 186 is connected to the drain of the n-channel MOSFET 187, one side of a capacitor $C_{load}$ 188, and one side of an inductor $L_i$ 178, forming an output resonant clock node $clk_{i,out}$. The source of the n-channel MOSFET 187 is connected to ground and to the other side of capacitor $C_{load}$ 188. The other side of inductor $L_i$ 178 is connected to one side of a capacitor 189 and one side of a capacitor 179. The other side of capacitor 189 is connected to a voltage or current source (not shown). The other side of capacitor 179 is connected to ground. Inductor $L_i$ 178 and capacitor $C_{load}$ 188 form a resonant circuit 156 on stratum-1 192.

[0021] A mutual coupling $M_{ij}$ 157 is obtained between the inductors $L_j$ 134 and $L_i$ 178. The mutual coupling $M_{ij}$ 157 assists in "pulling in" the phase difference between the two resonant clock nodes $clk_{i,out}$ and $clk_{i,out}$ through injection locking. Advantageously, the mutual coupling $M_{ij}$ 157 eliminates the need for physically connecting the clock network through the chip stack 199. Such an approach, when contrasted to an approach involving shorting of the clock network using chip-to-chip interconnects, provides reduced input/output (I/O) overhead, better power efficiency, reduced alternat-

ing current (AC) supply noise, and offers 180 degree clock distribution in neighboring stratums by simply changing the direction of the current flow.

[0022] As will be appreciated by one skilled in the art, aspects of the present invention may be embodied as a system, method or computer program product. Accordingly, aspects of the present invention may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to herein as a "circuit," "module" or "system." Furthermore, aspects of the present invention may take the form of a computer program product embodied in one or more computer readable medium(s) having computer readable program code embodied thereon.

[0023] Any combination of one or more computer readable medium(s) may be utilized. The computer readable medium may be a computer readable signal medium or a computer readable storage medium. A computer readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples (a non-exhaustive list) of the computer readable storage medium would include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing. In the context of this document, a computer readable storage medium may be any tangible medium that can contain, or store a program for use by or in connection with an instruction execution system, apparatus, or device.

[0024] A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part of a carrier wave. Such a propagated signal may take any of a variety of forms, including, but not limited to, electro-magnetic, optical, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that can communicate, propagate, or transport a program for use by or in connection with an instruction execution system, apparatus, or device.

[0025] Program code embodied on a computer readable medium may be transmitted using any appropriate medium, including but not limited to wireless, wireline, optical fiber cable, RF, etc., or any suitable combination of the foregoing.

[0026] Computer program code for carrying out operations for aspects of the present invention may be written in any combination of one or more programming languages, including an object oriented programming language such as Java, Smalltalk, C++ or the like and conventional procedural programming languages, such as the "C" programming language or similar programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network

(LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider).

[0027] Aspects of the present invention are described below with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0028] These computer program instructions may also be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

[0029] The computer program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other devices to cause a series of operational steps to be performed on the computer, other programmable apparatus or other devices to produce a computer implemented process such that the instructions which execute on the computer or other programmable apparatus provide processes for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0030] The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of code, which comprises one or more executable instructions for implementing the specified logical function (s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

[0031] Reference in the specification to "one embodiment" or "an embodiment" of the present principles, as well as other variations thereof, means that a particular feature, structure, characteristic, and so forth described in connection with the embodiment is included in at least one embodiment of the present principles. Thus, the appearances of the phrase "in one embodiment" or "in an embodiment", as well any other variations, appearing in various places throughout the specification are not necessarily all referring to the same embodiment.

[0032] It is to be appreciated that the use of any of the following "/", "and/or", and "at least one of", for example, in the cases of "A/B", "A and/or B" and "at least one of A and B", is intended to encompass the selection of the first listed option (A) only, or the selection of the second listed option (B) only, or the selection of both options (A and B). As a further example, in the cases of "A, B, and/or C" and "at least one of A, B, and C", such phrasing is intended to encompass the selection of the first listed option (A) only, or the selection of the second listed option (B) only, or the selection of the third listed option (C) only, or the selection of the first and the second listed options (A and B) only, or the selection of the first and third listed options (A and C) only, or the selection of the second and third listed options (B and C) only, or the selection of all three options (A and B and C). This may be extended, as readily apparent by one of ordinary skill in this and related arts, for as many items listed.

[0033] It is to be further appreciated that while one or more embodiments described herein may refer to the use of Silicon with respect to a chip or a through via, the present principles are not limited to using only chips or vias made from Silicon and, thus, chips or vias made from other materials including but not limited to Germanium and Gallium Arsenide may also be used in accordance with the present principles while maintaining the spirit of the present principles. Moreover, it is to be further appreciated that while one or more embodiments described herein may refer to the use of C4 or micro C4 (uC4) connections, the present principles are not limited to solely using C4 or micro C4 connections and, thus, other types of connections may also be used while maintaining the spirit of the present principles.

[0034] It is to be understood that the present invention will be described in terms of a given illustrative architecture having a wafer; however, other architectures, structures, substrate materials and process features and steps may be varied within the scope of the present invention.

[0035] It will also be understood that when an element as a layer, region or substrate is referred to as being "on" or "over" another element, it can be directly on the other element or intervening elements may also be present. In contrast, when an element is referred to as being "directly on" or "directly over" another element, there are no intervening elements present. It will also be understood that when an element is referred to as being "connected" or "coupled" to another element, it can be directly connected or coupled to the other element or intervening elements may be present. In contrast, when an element is referred to as being "directly connected" or "directly coupled" to another element, there are no intervening elements present.

[0036] A design for an integrated circuit chip of photovoltaic device may be created in a graphical computer programming language, and stored in a computer storage medium (such as a disk, tape, physical hard drive, or virtual hard drive such as in a storage access network). If the designer does not fabricate chips or the photolithographic masks used to fabricate chips, the designer may transmit the resulting design by physical means (e.g., by providing a copy of the storage medium storing the design) or electronically (e.g., through the Internet) to such entities, directly or indirectly. The stored design is then converted into the appropriate format (e.g., GDSII) for the fabrication of photolithographic masks, which

typically include multiple copies of the chip design in question that are to be formed on a wafer. The photolithographic masks are utilized to define areas of the wafer (and/or the layers thereon) to be etched or otherwise processed.

[0037] Methods as described herein may be used in the fabrication of integrated circuit chips. The resulting integrated circuit chips can be distributed by the fabricator in raw wafer form (that is, as a single wafer that has multiple unpackaged chips), as a bare die, or in a packaged form. In the latter case the chip is mounted in a single chip package (such as a plastic carrier, with leads that are affixed to a motherboard or other higher level carrier) or in a multichip package (such as a ceramic carrier that has either or both surface interconnections or buried interconnections). In any case the chip is then integrated with other chips, discrete circuit elements, and/or other signal processing devices as part of either (a) an intermediate product, such as a motherboard, or (b) an end product. The end product can be any product that includes integrated circuit chips, ranging from toys and other low-end applications to advanced computer products having a display, a keyboard or other input device, and a central processor.

[0038] FIG. 2 shows a clock distribution circuit 233 with resonant circuits 255 for a 3D chip stack 299, in accordance with an embodiment of the present principles. The clock distribution network 233 includes clock trees 200. Each stratum of a 3D chip stack 299 includes a respective one of the clock trees 200. Stack 299 includes a stratum-0 291 and a stratum-1 292.

[0039] The clock trees 200 have a single clock source 210 (e.g., a phase locked loop (PLL)), selectable using a 3D mux 220, for driving the root 217 of the clock trees in all strata. The trees 200 include clock relay buffers 230 and sector buffers 235. Each clock grid 288 is driven by a respective sector buffer 235. The clock relay buffers 230 relay the global clock signals. The clock source is selected by a 3D mux 220 from one of the strata (strata 291 and 292) and connected using through-Silicon vias (TSVs) 276 and micro C4 connections (μC4) 277. The balanced clock trees 200 distribute the clock to macro clock pins (not shown).

[0040] A resonant (LC tank) circuit 255 is added to each sector 266 of a respective clock grid 288 that is driven by the sector buffer 235. The LC tank circuit 255 includes a capacitor (C) 233 and an inductor (L) 234.

[0041] FIG. 3 shows an alignment 300 between the coils 310 of the inductors 234 of the resonant circuits 255 of FIG. 2, in accordance with an embodiment of the present principles. The coils 310 of the inductors 234 of the LC tank circuits 255 are aligned for the corresponding sector in each stratum to force the inductors 234 to resonant at the same phase, thus minimizing skew. The alignment of the 310 coils of the inductors 234 eliminates the need of physically connecting the clock network through the stack 299. Moreover, the coupled inductors help "pull in" the phase difference between two or more resonant nodes, as explained above

[0042] FIG. 4 shows a coupling enhancement 400 for the resonant circuits 255 of FIG. 2, in accordance with an embodiment of the present principle. When stacked with a front-to-back or back-to-back configuration, the coupling can be enhanced by a back side inductor 434 which is formed by the re-distribution layer (RDL) metal layers. The term "front" refers to the side of the chip that has the active electronics thereon.

[0043] The sector buffers 235 that drive the sectors within which the coils 310 are located are connected using TSVs 176 and/or μC4 277.

[0044] FIG. 5 shows a method 500 for reducing skewing of global clock signals within a 3D chip stack that includes two or more strata, in accordance with an embodiment of the present principles.

[0045] At step 510, each clock distribution circuit in a set of clock distribution circuits is arranged on a respective one of the two or more strata for providing the global clock signals to various chip locations.

[0046] At step 520, a respective resonant circuit included in each clock distribution circuit is configured to provide stratum-to-stratum coupling for the clock distribution network. The resonant circuit includes at least one capacitor and at least one inductor.

[0047] It is to be appreciated that step 520 may involve, e.g., aligning the coils of inductors on the different strata to minimize signal skew. For example, the coils of co-located (essentially the same position, but on different strata) inductors may be aligned.

[0048] Moreover, it is to be appreciated that the two or more strata may be configured with a front-to-back configuration or a back-to-back configuration, and the inductors coils that are aligned to form the coupling are located on respective frontsides of the two or more strata. In such a case, at step 530, the coupling is enhanced by adding a back-side inductor having inductor coils aligned with the inductor coils on the frontsides. For example, the back-side inductor is formed by the register transfer level metal layer.

[0049] Having described preferred embodiments of a system and method (which are intended to be illustrative and not limiting), it is noted that modifications and variations can be made by persons skilled in the art in light of the above teachings. It is therefore to be understood that changes may be made in the particular embodiments disclosed which are within the scope of the invention as outlined by the appended claims. Having thus described aspects of the invention, with the details and particularity required by the patent laws, what is claimed and desired protected by Letters Patent is set forth in the appended claims.

1-12. (canceled)

13. A method for synchronizing global clock signals in a clock distribution network in a 3D chip stack having two or more strata, the method comprising:

arranging each of a plurality of clock distribution circuits on a respective one of the two or more strata to provide the global clock signals to various chip locations;

configuring a respective resonant circuit included in each of the plurality of clock distribution circuits to provide stratum-to-stratum coupling for the clock distribution network, the resonant circuit including at least one capacitor and at least one inductor.

14. The method of claim 13, wherein said configuring step comprises aligning inductor coils included within the resonant circuit on at least two of the two or more strata with respect to each other to form a coupling there between to reduce skewing of the global clock signals with respect to the various chip locations.

15. The method of claim 14, wherein the two or more strata are configured with a front-to-back configuration or a back-to-back configuration, the inductor coils that are aligned to form the coupling are located on respective front-sides of the two or more strata, and the method comprises enhancing the

coupling by adding a back-side inductor having inductor coils aligned with the inductor coils on the front-sides.

16. The method of claim 15, wherein the back-side inductor is formed by a re-distribution layer.

17-24. (canceled)

25. A method for synchronizing global clock signals in a clock distribution network in a 3D chip stack having two or more strata, the method comprising:

arranging each of a plurality of clock grids on a respective one of the two or more strata to provide the global clock signals to various chip locations;

arranging each of a plurality of buffered clock trees on the respective one of the two or more strata to drive a respec-

tive one of the plurality of clock grids on a same one of the two or more strata, each of the plurality of buffered clock trees having at least a root and a plurality of clock buffers;

arranging each of a plurality of multiplexers on the respective one of the two or more strata, the plurality of multiplexers being configured to provide a same single clock source to the root of each of the plurality of buffered clock trees; and

configuring a resonant circuit included in each of the plurality of clock grids to provide stratum-to-stratum coupling for the clock distribution network.

*   *   *   *   *