

INTEGRATED CIRCUIT HAVING MEMORIES AND A SHARED WRITE PORT**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] The present application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/518,988, filed on August 11, 2023, entitled "INTEGRATED CIRCUIT HAVING MEMORIES AND A SHARED WRITE PORT", identified by Docket Number P23-133-US-PSP, the entire contents of which is incorporated herein by reference in its entirety.

[0002] The present application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/602,733, filed on November 27, 2023, entitled "METHOD AND SYSTEM FOR KNOWN-GOOD-DIE TESTABILITY OF FACE-TO-FACE BONDED CHIPLETS", the entire contents of which is incorporated herein by reference in its entirety.

[0003] The present application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/602,737, filed on November 27, 2023, entitled "SYSTEM AND METHOD FOR HAVING CORRECT-BY-CONSTRUCTION TIMING CLOSURE", the entire contents of which is incorporated herein by reference in its entirety.

[0004] The present application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/567,649, filed on March 20, 2024, entitled "ASSEMBLY HAVING A FACE-TO-FACE BONDED CHIPLET", identified by Docket Number P24-052-US-PSP, the entire contents of which is incorporated herein by reference in its entirety.

[0005] The present application claims the benefit and priority to U.S. Provisional Patent Application No. 63/637,742, filed on April 23, 2024, entitled "INTEGRATED CIRCUIT HAVING MICROVAULT MEMORIES", identified by Docket Number P24-081-US-PSP, the entire contents of which is incorporated herein by reference in its entirety.

[0006] The present application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/637,764, filed on April 23, 2024, entitled "FEFET STRUCTURES ON INTEGRATED CIRCUITS", identified by Docket Number P24-082-US-PSP, the entire contents of which is incorporated herein by reference in its entirety.

[0007] The present application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/674,471, filed on July 23, 2024, entitled "SYSTEM, METHOD, AND APPARATUS FOR WAFER-SCALE MEMORY", identified by Docket Number P24-135-US-PSP, the entire contents of which is incorporated herein by reference in its entirety.

BACKGROUND*Relevant Field*

[0008] The present disclosure relates to integrated circuits. More particularly, the present disclosure relates to integrated circuits having multiple modules with separate read address spaces and a shared write address space.

Description of Related Art

[0009] Chiplets refer to miniature chips that are designed to work as a single entity while using advanced packaging technology. These miniaturized chips are created by dividing the larger chip into several smaller chips, each with its own function or capability. The concept originated from the semiconductor industry's need to overcome the physical restrictions of traditional monolithic chip designs and achieve higher levels of integration. The idea behind chiplets is to create a modular system of interconnected and interchangeable chips that can be combined in different configurations to create advanced computing systems with improved performance, power efficiency, and functionality.

[0010] Chiplets can be based on different architectures, such as CPU, GPU, memory, or IO, and can be assembled and stacked in a variety of ways, depending on the specific application requirements. One of the advantages of the chiplet approach is the ability to mix and match different chiplets from different manufacturers to create custom solutions that meet specific computing needs. This approach also allows for faster time to market, reduced development costs, and increased flexibility, as chiplets can be upgraded or replaced without the need for a complete system redesign.

[0011] The use of chiplets may be used in various industries, including consumer electronics, cloud computing, and data centers, where the demand for high-performance computing and energy efficiency is high. Chiplets are expected to play a significant role in the future of computing and are likely to unlock new possibilities for creating more powerful and/or sophisticated electronic devices.

[0012] SUMMARY

[0013] An integrated circuit is disclosed herein that may be part of a semiconductor device. A method of manufacturing or a method of writing and reading data to modules are disclosed herein and can be used with all examples, embodiments, and aspects as described herein.

[0014] In one general aspect, the integrated circuit may include a modules group having a plurality of modules including a first module and a second module. The integrated circuit

may also include a shared write port configured to write to the modules group. The integrated circuit may furthermore include a first read port configured to read from the first module. The integrated circuit may in addition include a second read port configured to read from the second module.

[0015] Implementations may include one or more of the following features. The integrated circuit may have the first and second read ports configured to be inactive when a write operation is applied to the shared write port. The integrated circuit may be disposed on a first semiconductor device and may include: a shared write peripheral configured to write to the modules group via the shared write port, the shared write peripheral disposed on a silicon substrate of the first semiconductor device; a first read peripheral configured to read the first module via the first read port, where the first read peripheral is disposed on the silicon substrate of the first semiconductor device; and a second read peripheral configured to read the second module via the second read port, where the second read peripheral is disposed on the silicon substrate of the first semiconductor device, where the first module and the second module are disposed on a second layer of the first semiconductor device.

[0016] The integrated circuit may be configured such that the first module is a three-dimensional column of memory bit cells disposed between a first surface of the first semiconductor device and the silicon substrate of the first semiconductor device. The integrated circuit may have the first read port configured to be in electrical communication with a second semiconductor device and/or the second semiconductor device configured to be secured to a first surface of the first semiconductor device.

[0017] One or more interposers may be disposed between the first surface of the first semiconductor device and the second semiconductor device. The first read port may include a read address bus that traverses through the second layer of the first semiconductor device. The first read port may include a read data bus that traverses through the second layer of the first semiconductor device. The shared write port may be configured to be in electrical communication with a second semiconductor device. The second semiconductor device may be configured to be secured to a first surface of the first semiconductor device. The integrated circuit may include an electrically connecting interposer disposed between the first surface of the first semiconductor device and the second semiconductor device. The shared write port may include a write address bus that traverses through the second layer of the first semiconductor device. The shared write port may include a write data bus that traverses through the second layer of the first semiconductor device.

[0018] The first and second read ports may be configured to process reads concurrently relative to each other. The shared write port may be configured to write to an address space. The shared write port may be configured to write to the first module via a first portion of the address space and write to the second module via a second portion of the address space. Each module of the plurality of modules may include an independent read port for concurrent reading via a respective independent read port of any of the plurality of modules where the first read port is configured to read from the first module concurrently with another read of the second module using the second read port. The modules group may be configured to have a single write address space configured to write to the plurality of modules. The first read port may be configured to have a first read address space and the second read port is configured to have a second read address space. The first read address space may numerically overlaps with the second read address space. The first read address space may be coextensive with the second read address space and/or contiguous with the second read address space.

[0019] The modules group may be formed on a second layer portion of an integrated circuit device. The integrated circuit may be implemented as a chiplet configured to be face-to-face bonded. The chiplet may include a plurality of input/output (I/O) bonds, where each of the plurality of input/output (I/O) bonds is configured to provide a respective read port for a respective module of the plurality of modules where each of the plurality of input/output (I/O) bonds is configured to interface with a complementary respective input/output (I/O) bonds of a second device as bonded with the chiplet.

[0020] The integrated circuit may be part of a semiconductor device. The semiconductor device may be implemented as a die, a wafer, or a Chiplet die and may be bonded with a second semiconductor device, which is one of a GPU (Graphics Processing Unit), an SoC (System-On-chip), and a GPPU (General Purpose Processing Unit).

[0021] The second semiconductor device may be implemented as a Chiplet die, a wafer, an ASIC (Application Specific Semiconductor Chip), and an FPGA (Field Programmable Gate Array). The semiconductor device may include a plurality of input/output (I/O) bonds where each of the plurality of input/output (I/O) bonds is configured to provide a respective read port for a respective module of the plurality of modules and each of the plurality of input/output (I/O) bonds is configured to electrically connect with a complementary respective input/output (I/O) bonds of the second semiconductor device.

[0022] The Chiplet die has a first side and a second side where the second side is configured to be face-to-face bonded to the second semiconductor device. The Chiplet die and second semiconductor device may be part of a multi-chip package. The semiconductor device

may include an electrically connecting interposer disposed between the Chiplet die and the second semiconductor device. The first read port may be disposed on the Chiplet die. the Chiplet die has a surface and the first module may include: a memory array having a plurality of memory unit cells arranged within the first module; and a read peripheral circuitry configured to read data stored within the memory array via the first read port.

[0023] Footprints are the space that a device can project onto a surface (e.g., how much x-y space a device occupies which may or may not extend into a third dimension (e.g., z-axis). A footprint on the surface of the read peripheral circuitry may overlap with a footprint on the surface of the memory array. The plurality of memory unit cells may be arranged in a three-dimensional configuration. A footprint on the surface of the first read port may overlap a footprint on the surface of the memory array. A footprint on the surface of the read peripheral circuitry may be coextensive with a footprint on the surface of the memory array. A footprint on the surface of the first read port may be coextensive with a footprint on the surface of the memory array. The memory array, the read peripheral circuitry, and the first read port may be stacked vertically and configured to minimize a module footprint on the surface of the Chiplet die.

[0024] In some aspects, write peripheral circuitry may be connected to the shared write port on the surface of the Chiplet die such that: the write peripheral circuitry defines a first footprint on the surface where the first footprint does not overlap with a footprint of the memory array, the first footprint does not overlap with a footprint of the first read port, and the shared write port defines a second footprint on the surface where the second footprint does not overlap with the footprint of the memory array and does not overlap with the footprint of the first read port. In some aspects, the first footprint may overlap the second footprint.

[0025] The plurality of memory unit cells may be formed from at least one non-volatile memory unit cell. These memory unit cells may comprise various materials such as ferroelectric, magnetic, spin orbit torque, spin-transfer-torque, phase-change, and/or anti-ferroelectric materials. One implementation may include organizing the plurality of modules into separate partitions, each with a dedicated read peripheral, where each dedicated read peripheral has an independent clock. Alternatively, each module may have its own dedicated read peripheral. During a reset, the modules group may be configured to process write command only and disable read capability. Conversely, when the modules group is not in the reset, write capability may be disabled. The modules group may include at least two different non-volatile memory technologies.

[0026] One implementation may include an integrated circuit with a dynamic allocation circuitry to allocate memory blocks to the plurality of modules based on the usage of the modules group. Additionally, the integrated circuit may include a write peripheral connected to a dedicated I/O pad enabling data transfer external to the integrated circuit. The integrated circuit may be formed via an additive manufacturing process on a silicon substrate and may be electrically connected to a second semiconductor device with another integrated circuit.

[0027] The integrated circuit may include a read peripheral for at least one of the plurality of modules that is disposed on the silicon substrate or the second semiconductor device. The second semiconductor device may be implemented as a system-on-chip, Chiplet die, wafer, ASIC, or FPGA.

[0028] In one implementation, the plurality of modules is formed from non-volatile memory unit cells arranged in a three-dimensional connectivity fabric vertical to a silicon substrate and the second semiconductor device, and may utilize at least one of a cross-point, 3D NANDs, 3D NORs, 3D ANDs, and a stacked planar layer. The integrated circuit may include a single write peripheral with a dedicated clock or a plurality of clocks which feed to respective modules with decoupled timing relative to other modules. Each of the plurality of clocks are configured for clocking a respective read port of a respective module of the plurality of modules.

[0029] Another implementation may include the modules group formed on a chiplet with a first side and a second side, the second side being configured for bonding to a second semiconductor device. The integrated circuit may also include a decoder circuit, a driver circuit, and a register circuit on a silicon substrate of the chiplet. Alternatively, the modules group may be formed on a second layer of the chiplet. The second semiconductor device may have a plurality of processing elements, where each processing element includes a respective interface to communicate with a respective module of the plurality of modules on the modules group when the second semiconductor device is bonded to the chiplet. The chiplet may also have an interface to the shared write port on the second side to thereby interface with a complementary interface on the second semiconductor device.

[0030] The second semiconductor device may have a network-on-a-chip that is configured to provide inter-element communication for the plurality of processing elements. The plurality of processing elements may include at least one embedded FPGA, or one of a soft processor, a DSP block, an embedded processor, and a microcontroller.

[0031] In yet another implementation, a write operation to the modules group is performed through a priority arbitration circuit that facilitates the modules to be accessed in a

predetermined order. The integrated circuit may include a power gating circuitry that selectively powers down a module of the plurality of modules when not in use. Additionally, the shared write port may be configured to write to a virtual address space that is mapped onto a physical memory space.

[0032] The integrated circuit may also include a control circuit configured to enable the shared write port during write operations and disable the shared write port during read operations to conserve energy. Alternatively, the control circuit may disable the first read port during write operations and enable the first read port during read operations to conserve energy. A power management module may also be included, which selectively powers down the shared write port during read operations to conserve energy. The write circuitry may be configured to dynamically shift power allocation from write operations to read operations, or to enter a sleep mode and power down the shared write port during read operations to conserve energy.

[0033] The non-volatile memory utilized in the plurality of modules may be selected from the group comprising of a FeFET, a FeRAM, a ReRAM, a SOT (Spin Orbit Torque), and a STT (Spin Transfer Torque). The write peripheral for modules group may be implemented on a silicon substrate and disposed between the modules group and the second side of the chiplet. The read peripheral for the first module may be implemented on a silicon substrate and disposed between the modules group and the first side of the chiplet, or alternatively, in a second layer. In some implementations, each module has its own dedicated write peripheral that utilizes a shared clock.

[0034] BRIEF DESCRIPTION OF THE DRAWINGS

[0035] These and other aspects will become more apparent from the following detailed description of the various embodiments of the present disclosure with reference to the drawings wherein:

[0036] Fig. 1 is a block diagram of an integrated circuit that may be part of a semiconductor device such as a chiplet in accordance with an embodiment of the present disclosure;

[0037] Fig. 2 shows a perspective view of an assembly having the integrated circuit of Fig. 1 implemented on a semiconductor device that is electrically connected to another device to form the assembly in accordance with an embodiment of the present disclosure;

[0038] Fig. 3 shows a block diagram illustrating the memory address space of the integrated circuit of Fig. 1 in accordance with an embodiment of the present disclosure;

[0039] Fig. 4 shows a block diagram illustrating the memory address space with the signal interfaces of the integrated circuit of Fig. 1 in accordance with an embodiment of the present disclosure;

[0040] Fig. 5 shows an illustration of an integrated circuit that may be part of a semiconductor device such as a chiplet in accordance with an embodiment of the present disclosure; and

[0041] Fig. 6 shows a perspective of an assembly having the integrated circuit of Fig. 1 implemented on a semiconductor device that is electrically connected to a system-on-a-chip in accordance with an embodiment of the present disclosure.

[0042] **DETAILED DESCRIPTION**

[0043] Fig. 1 shows a block diagram of an integrated circuit 100 that may be packaged as a bondable chiplet (e.g., face-to-face chiplet bondable) in accordance with an embodiment of the present disclosure. The integrated circuit (IC) 100 includes a modules group 106 consisting of modules 108, 110, 112, and 114. The IC 100 also features a shared write port 102, configured to write to the modules group 106 using a write peripheral 104. Additionally, it includes read peripherals 116, 118, 120, and 122 and read ports 124, 126, 128, and 130, configured to read from the modules 108, 110, 112, 114. The IC 100 also includes an interlock 132.

[0044] The write port 102 may be configured to provide a single write address space for all of the modules group 106 where each of the modules 108, 110, 112, 114 has a dedicated read port 124, 126, 128, 130 respectively. The integrated circuit 100 may be packaged as part of a chiplet configured to be electrically connected to another integrated circuit device (e.g., another chiplet, or IC package, with or without electrical contacts, electrical bumps, etc.). The chiplet may be electrically connected to another device including, for example, by bonding, soldering, wafer-to-wafer bonding, face-to-face chiplet bonding, chiplet-to-wafer bonding, chiplet-to-interposer bonding, and/or may be connected together with an interposer or other interfacing technology. None, one, or more of interposers may be used or other interfacing technologies that are common to heterogeneous 3D system-in-package solutions may be utilized in electrically connecting a chiplet to another device.

[0045] Each read port (124, 126, 128, 130) in the chiplet may feature electrical contacts on a side of the chiplet or on multiple sides of the chiplet. The read ports 124, 126, 128, 130 may use multi-cycle pipelined circuitry. Upon bonding to another device (e.g., wafer, chiplet, chip, SOC, package, FPGA, etc.), the electrical contacts may line up in a manner that provides dedicated access to specific modules of the modules 108, 110, 112, 114. For instance, a

processing/computing element may have exclusive access to module 108 via the read port 124, which may contain the neural network weights in a registry file. Similarly, a different processing/computing element may have exclusive read access to module 110 via the read port 126, which includes a different registry file. In this specific embodiment, this arrangement of the electrical contacts ensures that each computing/processing element has the dedicated access it needs to carry out its specific computation efficiently thereby providing a compact, modular, and scalable system that allows different processing elements to maintain dedicated access to specific modules 108, 110, 112, 114. Without dedicated access, different processing elements might have to queue up to use the same resource which would slow down overall processing speed. By providing dedicated access, the proposed chiplet ensures that each processing element can operate at its maximum capability without interference from other computing elements in this specific embodiment.

[0046] The write peripheral 104 is a peripheral circuitry responsible for processing and writing data into the memory cells found within the modules 108, 110, 112, 114. The write peripheral 104 may include dedicated contacts so that a chip electrically connected (e.g., bonded) to a chiplet of the integrated circuit can access the write port 102 via a shared write logic system. This system involves utilizing a shift register-based, different voltage design, preferably high voltage design, that has a shared write address and data components. This shared write logic system is designed to be accessed via a bonded chip, another bonded chiplet, and/or via other circuitry in the same package as the integrated circuit 100. A shift register could allow the system to move data through a series of stages, with each subsequent stage receiving the data from the previous stage. By utilizing a shift register, the system can increase the data throughput while maintaining a low rate of data transfers. The shared write address space refers to the location where data is written in the chiplet.

[0047] In another embodiment, an interlock 132 may disable the read ports 124, 126, 128, 130 while data is being written to the modules group 106 via the write port 102. Likewise, the interlock 132 may disable the write port 102 when read operations are being carried out on the read ports 124, 126, 128, 130. The written data can later be accessed concurrently by all processing elements that need to read the data via a respective one of the read ports 124, 126, 128, 130. This ensures that all processing elements have the most commonly used data available to them without regard to other reads being concurrently carried out by other processing elements.

[0048] The write peripheral 104 circuit includes a write driver. This unit receives the data to be written and converts it into suitable signals that can change the state of the memory

cells. Depending on the type of memory technology used, these signals could involve voltage levels, current pulses, or other types of energy. The shared write logic system may be high voltage due to the specific voltage requirements of the chiplet. The write driver must provide enough power to reliably change the state of the memory cells, but it must also operate within suitable parameters to avoid causing damage or unnecessary wear.

[0049] The write peripheral 104 circuit may also feature a data buffer or write buffer. This component temporarily stores the data to be written, allowing the write operation to be performed at an optimal pace. By balancing the speed of incoming data with the speed at which the memory cells can be written, the write buffer helps prevent data loss and optimizes system performance.

[0050] The write peripheral 104 may also include, in some embodiments, a write control unit that orchestrates the sequence of operations in the write process. It generates control signals to activate the write driver at the appropriate times, controls the flow of data from the write buffer, and coordinates the timing of the write operations. By synchronizing these various activities, the write control unit ensures efficient and reliable write operations.

[0051] The write peripheral 104 may also include data encoding mechanisms to improve reliability and data integrity. For example, before the data is written to the memory cells, these mechanisms encode it in a way that allows potential errors to be detected, and in some cases, corrected when the data is later read. This can be helpful in systems where data integrity has a higher priority, such as in servers or scientific research devices.

[0052] The write peripheral 104 may also include a timing unit that serves as the system's heartbeat, supplying clock signals that synchronize the operation of the system's various components. In some systems, it may include components like oscillators, clock generators, or phase-locked loops. The timing unit may ensure that all operations occur at the suitable time relative to each other.

[0053] The IC 100 may be implemented as a face-to-face bonded chiplet, with modules 108, 110, 112, and 114 formed from a non-volatile memory. In some specific embodiments, the IC 100 may also feature a dynamic allocation circuitry to allocate memory blocks to the modules group 106 based on the usage of the modules group 106 (e.g., each module 108, 110, 112, and 114 may include dynamic allocation circuitry for dynamically allocating a range of read locations for a respective processing element).

[0054] The IC 100 features a plurality of clocks, with each clock of the plurality of clocks feeding a respective module of the plurality of modules, providing each respective module with decoupled timing relative to the other modules of the plurality of modules. The

modules group 106 may be arranged in any topology known to one of ordinary skill in the relevant art. Bit-cell density can be up to 10 times more dense than embedded SRAM cells in the modules group 106.

[0055] The IC 100 may be formed on a chiplet that includes a first side and a second side, with the second side configured for bonding to a second semiconductor device. The IC 100 may include a high voltage write logic adjacent to the first side of the chiplet. A decoder circuitry, a driver circuitry, and a register circuitry may be formed on the silicon substrate portion of the chiplet, while the modules group 106 is formed on a second layer portion of the chiplet. The second semiconductor device may comprise a plurality of processing elements. Each processing element includes a respective interface to communicate with a respective module of the plurality of modules on the modules group 106 when the second semiconductor device is bonded to the chiplet.

[0056] The silicon substrate traditionally serves as the initial stage of IC fabrication, focusing on the creation of active components, particularly transistors. Techniques like diffusion, ion implantation, oxidation, and material deposition are employed to fashion the intricate structures of transistors. These processes operate at small scales. The application of photolithography, etching, and implantation techniques enables the definition of transistor structures with precision. The silicon substrate is used to establish the fundamental building blocks for signal processing, amplification, and control within the IC. This layer is sometimes called the Front-End-Of-The-Line (“FEOL”).

[0057] Next in the manufacturing process, a second layer may be added that traditionally takes on the role of interconnect fabrication, facilitating the electrical connections between various IC components. This phase traditionally focused on the creation of passive components, including interconnects, vias, and metal-insulator-metal (MIM) capacitors. The second layer processes typically differ from the processes used on the silicon substrate in terms of precision and scale. The interconnects are formed by depositing and patterning metal layers, typically aluminum or copper, to construct the wiring network. Dielectric layers, such as silicon dioxide or low-k dielectrics, are introduced to insulate the interconnects and prevent signal interference between different wiring layers. The second layer’s traditional function is to establish the necessary interconnections that enable the routing and distribution of electrical signals throughout the IC. However, as described herein, circuitry may be utilized within this second layer (sometimes referred to as Back-End-Of-The-Line (“BEOL”).

[0058] Alternate embodiments of the IC 100 may be implemented as a stacked die, a monolithic design, TSVs, or silicon through vias. In a stacked die design, several dies may be

stacked on top of each other, with each die performing different functions, such as memory and processing. The stacked die may communicate through wire bonds, microbumps, or bump-less bonds. In a monolithic design, the various functions and modules of the IC 100 may be integrated onto a single die, forming a more compact and power-efficient design.

[0059] Additionally, the IC 100 may include one or more interlocks 132 to manage conflicts in reading and writing data. The modules group 106 may be formed from a variety of non-volatile or semi-volatile (e.g., very long refresh periods) memory technologies, such as Static Random-Access Memory (SRAM), Ferroelectric Field Effect Transistor (FeFET), Ferroelectric Random Access Memory (FeRAM), Resistive Random Access Memory (ReRAM), Spin-Orbit Torque (SOT) Memory, Spin Transfer Torque (STT) Memory, charge trap, floating gate memories, and/or Schottky diodes.

[0060] The modules group 106 may utilize a Static Random-Access Memory (SRAM) topology. The SRAM topology may employ a cross-coupled flip-flop structure (e.g., latching flip-flops), ensuring the stored data remains intact as long as power is supplied. Thus, in some specific embodiments, the modules group 106 may utilize heterogeneous types of memory including volatile and non-volatile memory types.

[0061] The modules group 106 may utilize a Flash Memory Topology. Flash memory is a non-volatile memory technology used in applications where data persistence is needed, such as solid-state drives (SSDs) and USB flash drives. The flash memory topology disclosed herein features a matrix of memory cells, each consisting of a floating-gate transistor or charge trap device. The modules group 106 may also use wear-leveling techniques to prolong the lifespan of the memory cells.

[0062] The modules group 106 may utilize a Ferroelectric Random-Access Memory (FeRAM) topology. The FeRAM topology utilizes a ferroelectric material capable of retaining polarization states. One such memory topology may, in specific embodiments, utilize a FeFET to retain state information and program the ferroelectric material. These ferroelectric materials may be used to retain state information and act as a memory bit cell.

[0063] The modules group 106 may utilize a Phase Change Memory (PCM) Topology, which is a non-volatile memory technology that utilizes reversible phase changes in materials to store data. The PCM topology may include any phase change material, for example a chalcogenide alloy or a chalcogenide glass housed within a memory cell.

[0064] The modules group 106 may utilize a Resistive Random-Access Memory (ReRAM) Topology, which is a non-volatile memory technology based on resistive switching

phenomena. The ReRAM topology may utilize a thin-film material that exhibits reversible changes in resistance upon the application of electrical stimuli.

[0065] The modules group 106 may utilize a Spin-Orbit Torque (SOT) Magnetic Random-Access Memory (MRAM) topology. SOT-MRAM is a type of non-volatile memory that utilizes spin-orbit torque to switch the magnetic state of a storage element. The SOT-MRAM topology may incorporate a magnetic tunnel junction (MTJ) structure and leverages the spin-orbit coupling effect to write and read data. The magnetic tunnel junction may have a dielectric layer between a magnetic fixed layer and a magnetic free layer. Writing may be done by switching magnetization of the free magnetic layer by injecting an in-plane current in an adjacent SOT layer. Reading may be done by putting current into the magnetic tunnel junction. The SOT-MRAM can optimize the spin-orbit materials by using current-driven switching schemes while minimizing write energy consumption, in some specific embodiments.

[0066] The modules group 106 may utilize a Spin Transfer Torque (STT) Magnetic Random-Access Memory Topology. The STT-MRAM is another type of non-volatile memory that relies on spin transfer torque to manipulate the magnetic state of a storage element. The STT-MRAM topology can use a magnetic tunnel junction (MTJ) structure, where the magnetization orientation determines the stored data. Additionally, the orientation of a magnetic layer in a magnetic tunnel junction or spin valve can be changed using a spin-polarized current, for example.

[0067] The IC 100 may include a single write peripheral 104 with a dedicated clock, or each module 108, 110, 112, 114 may have its own dedicated write peripheral utilizing a shared clock (not shown in Fig. 1). Additionally, the modules group 106 may be organized into separate partitions, each with a dedicated read peripheral 116, 118, 120, 122 having an independent clock.

[0068] Another possible embodiment of the IC 100 includes an interface (e.g., the same, different, higher or lower voltage) to enable data transfer external to the packaging of the IC 100. The IC 100 may also include an integrated microcontroller unit (MCU) or a digital signal processor (DSP) for processing data within the IC in yet additional embodiments.

[0069] Fig. 2 shows a perspective view of an assembly 200 of the integrated circuit 212 of Fig. 1 implemented on a chiplet 230 that is bonded to a second device 226 in accordance with an embodiment of the present disclosure. The integrated circuit 212 is the circuitry within the chiplet 230. The second device 226 may be a chiplet, semiconductor wafer, semiconductor package, encased circuitry, etc. For example, the second device 226 may be an AI accelerator such that each processing unit has read access to one module (or a predetermined set) of the

modules group 236. In yet another embodiment, the second device 226 may be a network controller where there is an offload circuit to read the data from each of the modules to process incoming/outgoing packets, etc. The assembly 200 includes a modules group 236 having a plurality of modules, including a first module 232 and a second module 234. Fig. 2 shows several modules, however, for clarity, only modules 232, 234 have reference numbers. The integrated circuit 212 further comprises a shared write port 222. The shared write port 222 interfaces into the write peripheral 202.

[0070] Although the second device 226 may use the shared write port 222 via an address & data bus with a clock and a enable signal to write data to any modules within the modules group 236, other ways of writing data may be considered. For example, serial connections, parallel connections, various buses, or ports, may be used, such as a DDR (Double Data Rate) Interface, an SRAM (Static Random-Access Memory) Interface, a NAND Flash Memory Interface, a NOR Flash Memory Interface, an HBM (High Bandwidth Memory) Interface, a GDDR (Graphics Double Data Rate) Interface, an NVMe (Non-Volatile Memory Express) Interface, SPI, I2C, etc. Each of the modules has a read port with a read address 218 (to send an address to a module 234) and read data 220 (which is the data read from the module 232).

[0071] The modules group 236 is formed on a chiplet 230 having two sides including a surface 228 that can be bonded to and complement a second device 226. The chiplet 230 may be formed by forming circuitry on a silicon substrate 204 and then by adding a second layer 206. In other embodiments, these layers may be reversed and/or other layers may be added, removed, etc. The read address 218 and read data 220 are used for reading the module 232.

[0072] Although the second device 226 may use an address and data bus with a clock and an enable signal to read data from the module 232, other ways of reading data may be considered. For example, serial connections, parallel connections, various buses, or ports, may be used, such as a DDR (Double Data Rate) Interface, a SRAM (Static Random-Access Memory) Interface, a NAND Flash Memory Interface, a NOR Flash Memory Interface, an HBM (High Bandwidth Memory) Interface, a GDDR (Graphics Double Data Rate) Interface, an NVMe (Non-Volatile Memory Express) Interface, SPI, I2C, etc.

[0073] All of the read ports (e.g., 218 and 220) are configured to be inactive when a write operation is applied to the shared write port 222. The read ports may also be configured to process reads concurrently with each other. The shared write port 222 is configured to write to an address space, where the shared write port 222 is configured to write to the first module

232 via a first portion of the address space and write to the second module 234 via a second portion of the address space. Each module of the plurality of modules group 236 includes an independent read port for concurrent reading via a respective independent read port of any of the plurality of modules.

[0074] Each read port for a respective module may include contacts for circuitry found within the second device 226 to interface via metallic contacts. Thus, there may be metallic contacts on the top layer 208 that are configured to interface with metallic contacts on the surface 228 of the chiplet 230 such that the metallic contacts allow for a read space that is coextensive with a read space of a module of the modules group 236. The read spaces of the modules group 236 may all be coextensive with each other (as is described with reference to Figs. 3 and 4).

[0075] In one embodiment, the read peripheral for the first module 232 is implemented on a silicon substrate 204 (sometimes referred to as a Front-end-of-the-line). The second layer 206 (sometimes called the Back-end-of-the-line) may be built next in the manufacturing process on top of the silicon substrate 204 (and any circuitry) and may contain the respective memory bit cells. In an alternative embodiment, the read peripheral for the first module 232 is implemented in the second layer 206 and is disposed between the modules group 236 and the surface 228 of the chiplet 230.

[0076] The modules group 236 may be configured to process write commands only during reset. The write commands may be “slow write” commands. That is, the modules group 236 may have very low write speeds relative to its read speed. The write logic may be frozen (or disabled) when the modules group 236 are used for reading data. In some specific embodiments, the integrated circuit 212 provides functionality to allocate memory blocks to the modules group 236 based on the usage of the modules group 236. In other embodiments, the memory addresses are fixed along with the allocation. The integrated circuit 212 may be implemented as a face-to-face bonded chiplet 230. The face-to face bonding may be bump-less wafer bonding.

[0077] The modules group 236 can have a single write peripheral 202. In other embodiments, each module of the modules group 236 may have a dedicated write peripheral that utilizes a shared clock. In yet other embodiments, the modules group 236 may also be organized into separate partitions each with partition having a dedicated read peripheral, where each dedicated read peripheral has an independent clock. The partitions may be one, two, or more modules of the modules group 236.

[0078] The write peripheral 202 circuitry's overall architecture may include a series of different components, including write drivers, address decoders, sense amplifiers, data input latches, data bus, etc. and/or some combination thereof. Write drivers or write buffers, may be tasked with transferring data onto the memory cell. They may enhance the input signal to achieve a level appropriate for the memory cell. Address decoders may be used to interpret the memory address that is fed as an input where the data needs to be written. By activating the specific row and column of the memory array linked to that address, they may be used to select the target memory cell. Sense amplifiers may be used to identify and boost the signal from the memory cells during read operations, and also participate in refreshing the memory cell after data write operations. The write operation is instigated by a write enable signal. When a write command is initiated, this signal propels the write drivers and decoders into the writing process. Data input latches may be used as temporary storage units, retaining the data set to be written into the memory until the write operation is implemented. A data bus with a transmission route, can be used to facilitate the movement of data from the data input latches to the memory cells.

[0079] A write operation to the modules group 236 may be performed through a priority arbitration circuit that facilitates the modules to be accessed in a predetermined order, and the shared write port 222 may be configured to write to a virtual address space that is mapped onto a physical memory space. The integrated circuit 212 may include a high voltage write logic used within the write peripheral 202, and the second semiconductor device 226 may comprise a plurality of processing elements, whereby each processing element includes a respective interface to communicate with a respective module of the modules group 236. Furthermore, the chiplet 230 may include an interface to the shared write port 222 on the surface 228 to interface with a complementary interface on the second semiconductor device 226.

[0080] The integrated circuit 212 may also include a power gating circuitry that selectively powers down a module of the modules group 236 when not in use. Additionally, the integrated circuit 212 may have the write peripheral 202 of the modules group 236 connected to a dedicated I/O pad to enable data transfer external to the package of the integrated circuit.

[0081] The integrated circuit 212 may utilize multiple modules of the modules group 234 grouped together. These modules may be synchronized with one another in specific embodiments. In some cases, all the modules are synchronized, while in other instances, only specific modules are to be synchronized. For instance, the circuit on a second device 226 may need to synchronize with a specific module when reading data from one of the modules in the module group 236.

[0082] To synchronize the modules, the integrated circuit 212 may use various timing technologies. In some cases, a plurality of clocks may feed each respective module of the modules group 236, thereby allowing each module to have decoupled timing relative to the other modules in the group. This decoupling ensures that any delay in one module will not affect the functioning of other modules. It is worth noting that the clocks used may or may not need to be synchronized. In some cases, a common clock can be used to synchronize the modules. In yet other embodiments, the clock signal or signals may be provided by the second device 226.

[0083] In alternative embodiments, other synchronization techniques can be used, such as phase comparison of the clock signals or a phase-locked loop (PLL) synchronization method. Another embodiment for synchronizing the modules in the IC could use delay-locked loop (DLL) synchronization. In this method, a delay element is added to the clock signal path, and the output is compared to the input clock signal. The feedback loop adjusts the delay element until the output of the DLL matches the input, resulting in synchronization of the clock signals.

[0084] In another embodiment, the integrated circuit 212 could use a combination of different synchronization techniques to achieve synchronization between the modules of the modules group 236. For example, some modules may use PLL synchronization while others use clock delay lines or DLL synchronization, depending on their specific requirements. Additionally, the integrated circuit 212 can also use redundant synchronization techniques to ensure reliability and redundancy in case one method fails. For example, the integrated circuit 212 could use both PLL synchronization and DLL synchronization simultaneously, so that if one method fails, the other can still maintain synchronization.

[0085] Fig. 3 shows a block diagram 300 illustrating the memory address space of the integrated circuit of Fig. 1 in accordance with an embodiment of the present disclosure. The memory address space includes a write address space 316 and read data address spaces 310, 312, 314.

[0086] The write address space 316 consists of various units where data, e.g., weights, and/or instructions can be stored. These units are referred to as memory addresses. The module group 302 includes multiple memory modules 304, 306, 308. The write address space 316 may be distributed among the memory modules 304, 306, 308 such that the write address space 316 spans from 0 to $N*M-1$. As shown in Fig. 3, the modules group 302 has N memory modules 304, 306, 308, where N is a positive integer, and each module has a memory size of M . The

total number of unique write memory addresses in the write address space will be $N*M$, which can be referenced by an integer from 0 to $N*M-1$.

[0087] Starting at 0, memory addresses of the write address space 316 are ordered sequentially up to $N*M-1$. In other words, the first address is 0 and the final address is $N*M-1$, encompassing a total of $N*M$ addresses. This ordering can be linear (each address increases by one) or some other specified pattern depending on the implementation.

[0088] The write memory addressing can be implemented in a variety of ways based on the system architecture. One method used in a specific embodiment is to use the base and limit registers. The base register holds the smallest legal physical write memory address, and the limit register specifies the size of the range. Therefore, to generate a logical address, you would add the base to the relative address. In other embodiments, a memory addressing scheme may be used where the base used is set to be 0. Yet additional write addressing techniques will be appreciated by one of ordinary skill in the relevant art.

[0089] For any device that writes to the modules group 302, each memory module can possess a unique set of write memory addresses such all memory addresses within the modules group 302 is unique with respect to writing data. e.g., the first module starting at 0 and the last one ending at $N*M-1$. This allocation, in some embodiments, may be dependent on the memory management system of the device writing data to the modules 304, 306, 308, which could range from simple fixed partitioning schemes to more complex dynamic partitioning models.

[0090] For instance, in a straightforward linear model where each module (304, 306, or 308) has an equal size of M addresses, the first module 304 would possess write addresses 0 to $M-1$, the second module would have write addresses M to $2M-1$, the third module would have write addresses $2M$ to $3*M-1$, and so forth. The N th module 308, therefore, would possess write addresses from $(N-1)*M$ to $N*M-1$.

[0091] It is contemplated that one of ordinary skill in the relevant art may use other implementations of write memory addresses from 0 to $N*M-1$ that depends on various factors such as the hardware architecture, operating system, memory management schemes, and the nature of the programs being run on the system, etc.

[0092] The modules group 302 has different read data address spaces 310, 312, 314. These read address spaces 310, 312, 314 may have overlapping address spaces, may have contiguous address spaces, or may have coextensive address spaces. The read address spaces 310, 312, 314 may be independent relative to each other. The system includes three independent read address spaces, labeled as read address spaces 310, 312, and 314. Each of

these read address spaces is distinct from the others, meaning that reads can be performed in each space without affecting the others.

[0093] The read address spaces 310, 312, 314 may be defined as contiguous blocks of memory addresses, each with its own starting address and ending address. In modules group 302, each read address space 310, 312, 314 may have a range of addresses that corresponds to values from 0 to M-1, where M is a maximum value determined by the size of the modules 304, 306, 308 being used.

[0094] In one embodiment, by allowing one processing unit to interface with each read address space 310, 312, 314, the concurrent reads may be implemented as described herein. The independence of the read address spaces 310, 312, 314 ensures that each processing unit can access its desired data without causing any interference or conflict with other processing units.

[0095] Fig. 4 shows a block diagram illustrating the memory address space with the signal interfaces of the integrated circuit of Fig. 1 in accordance with an embodiment of the present disclosure. The signals used in Fig. 4 may be used with any embodiment described herein. However, one of ordinary skill in the relevant art will appreciate that different signaling schemes may be used.

[0096] The modules group 402 includes modules 404, 406, 408 that share a common write peripheral 411. The write peripheral 411 includes a write address bus that includes the address of the data being written, a write data bus that includes the data, a write clock cause the writes to occur (e.g., either on a leading or trailing edge of the clock signal, etc.). The writes only occur if the write enable signal indicates a write should occur. Any logic may be used, e.g., high voltage may correspond to 1 and a low voltage may correspond to 0, or vice versa. In some embodiments, the write peripheral 411 may be on the chiplet 230 and in other embodiments, the write peripheral 411 is on the second device 226.

[0097] The modules group 402 comprises modules 404, 406, 408 where each has a respective read peripheral 410, 412, 414. Each of the read peripheral 410, 412, 414 has a read address bus to send an address for reading, a read data bus to receive the data, a read clock which is the clock used to control the timing of the output of the digital data, and an output enable that is a precondition to outputting data. Any logic may be used, e.g., high voltage may correspond to 1 and a low voltage may correspond to 0, or vice versa. In yet additional embodiments, multi-bit or analog data storage may be used. In some embodiments, one or more of the read peripherals 410, 412, 414 may be on the chiplet 230 and in other embodiments, one or more of the read peripherals 410, 412, 414 are on the second device 226.

[0098] Fig. 5 shows an illustration of an integrated circuit 500 that may be part of a semiconductor device such as a chiplet in accordance with an embodiment of the present disclosure. The integrated circuit 500 may be disposed on a semiconductor device, such as a chiplet, that has a silicon substrate 506 and a second layer portion 508. Within the integrated circuit 500, there may be an array section forming the module 502 where a three-dimensional column array of memory bit cells 522 has the necessary components to store memory in non-volatile, semi-volatile, or another memory format as described herein.

[0099] The integrated circuit 500 may include a modules group having a plurality of modules including a first module and a second module, etc. even though only a single module 502 is shown. The memory bit cells 522 are written to by the shared write port 512, 516, which includes both a write address bus line 512 and a write data bus 516. These buses run through the second layer 508 and can be connected to a second semiconductor device via an interposer. The second device has electrical contacts that complement those on the surface 518, allowing it to be electrically coupled to the write address and data buses. The memory bit cells 522 can be read from via the read port 524, 526, which includes a read address bus line 524 and a read data bus 526. Both of these buses can also run through the second layer 508 to the second semiconductor device coupled to the surface 518, which also has complementary electrical contacts to allow it to be electrically coupled to the read address and data buses.

[00100] Various kinds of memory technologies may be used for the memory bit cells 522, such as a vertical connectivity fabric structure formed from non-volatile memory unit cells arranged in a three-dimensional column array 522. The memory bit cells 522 may utilize one or more of a cross-point, 3D NANDs, 3D NORs, 3D ANDs, and/or a stacked planar layer.

[00101] In some embodiments, the integrated circuit 500 is electrically connected to a second semiconductor device (not shown in Fig. 5) comprising another integrated circuit, which may be a system-on-chip or a Field-Programmable-Gate-Array. In some embodiments, the memory bit cells 522 may be formed from various non-volatile memory types, such as FeFET, FeRAM, ReRAM, SOT, or STT. Additionally, alternatively, or optionally, the memory bit cells may be formed from non-volatile memory unit cells having 2-terminal devices, 3-terminal devices, or 4-terminal devices.

[00102] For example, the memory unit bit cells 522 may be formed from ferroelectric materials, such as a ferroelectric tunnel junction, a diode, a capacitor, a single-gate transistor, or a dual-gate transistor. Alternatively, the memory unit bit cells 522 may be formed from memristive materials, such as at least one ReRAM, or magnetic materials, such as at least one spin-orbit-torque device or at least one spin-transfer-torque device. Moreover, the non-volatile

memory unit cells 522 may also be formed from phase-change materials or anti-ferroelectric materials.

[00103] In some alternative embodiments, the non-volatile memory unit cells 522 can be formed from other types of materials, such as phase change materials, anti-ferroelectric materials, or multi-bit PCM materials. The non-volatile unit cells can be formed utilizing different structures, such as resistive random-access memory (RRAM) technology, magnetic random-access memory (MRAM) technology, or ferroelectric random-access memory (FRAM) technology.

[00104] Moreover, in some implementations, 3D NAND technology may be utilized to form the memory unit bit cells 522. For example, the memory unit bit cells 522 may be formed from stacked memory layers where each layer includes a plurality of memory cells that can be accessed using shared bit lines. In such a case, the read port 524, 526 may be coupled to the bit lines, and the write port 512, 516 may be coupled to the word lines that control the access to each layer.

[00105] In another embodiment, the 3D connectivity fabric structure can be built with stacked layers of either NAND gates, NOR gates, or AND gates, and in some cases, different types of logic gates may be combined to optimize the structure's functionality. In addition, the 3D connectivity fabric structure may be formed utilizing through-silicon-via (TSV) technology, which allows the vertical interconnection of the different layers of the structure.

[00106] Additionally, the non-volatile memory unit cells may include 2-terminal devices, such as a capacitive or a memristive device with or without an additional selector device such as a diode in series, 3-terminal devices, such as a floating-gate transistor, a transistor with an access gate, or 4-terminal devices, such as a transistor with two access gates. The type and configuration of the non-volatile memory unit cells 522 may depend on the specific application requirements, including the speed, power consumption, and reliability of the circuit. The memory unit cell may include or be a single ferroelectric transistor or 6T SRAM cell. The memory unit cell may be a combination of many different devices, including, but not limited to, one or more of a transistor, a memristor, a capacitor, etc.

[00107] In some embodiments of the present disclosure, a ferroelectric material can be utilized to form the non-volatile memory unit cells 522. The ferroelectric material may be implemented as any kind of device, including, but not limited to, a thin-film device, such as a ferroelectric tunnel junction, a capacitor, a single-gate transistor, or dual-gate transistors, etc.

[00108] In another embodiment, the non-volatile memory unit cells 522 may be formed from a memristive material, such as a Metal Oxide Memristor (MOM), Conductive-Bridging

RAM (CBRAM), or valence change memory (VCM), each of which provides different benefits regarding power consumption, speed, endurance, etc.

[00109] Moreover, in some embodiments, the non-volatile memory unit cells 522 may be formed from a magnetic material, such as spin-orbit-torque (SOT) devices, spin-transfer-torque (STT) devices, or perpendicular magnetic tunnel junctions (p-MTJ).

[00110] In one embodiment, the modules group may include many modules where each of which can be accessed through dedicated read ports 524, 526 with a dedicated read peripheral 520 while sharing the same write port 512, 516 and shared write peripheral 510. The shared write port 512, 516 can be configured to selectively write to one or more of the plurality of modules within the modules group including the memory bit cells 522. Each of the modules may have the same or different sizes, and different module sizes may be configured to optimize the utilization of the memory array with different operating scenarios, etc.

[00111] Furthermore, the integrated circuit 500 may be formed utilizing different manufacturing processes and techniques, which include but not limited to, a CMOS or Bipolar-CMOS-DMOS (BCD) process, a silicon-on-insulator (SOI) process, a FinFET process, a silicon germanium (SiGe) process, a gallium arsenide (GaAs) process, etc.

[00112] Fig. 6 shows a perspective of an assembly 600 having the integrated circuit of Fig. 1 implemented on a semiconductor device, such as the chiplet 230, that is electrically connected to a system-on-a-chip ("SOC") 610 in accordance with an embodiment of the present disclosure. The semiconductor device, in this embodiment, is the chiplet 230 that is electrically connected to a system-on-a-chip ("SOC") 610.

[00113] Referring to Fig. 6, the SOC 610 includes a silicon substrate 602 on which a plurality of processing elements is formed, including a processing element 606. The processing elements can communicate with each other through a Network-on-Chip ("NOC") 604, which is a communication fabric that directs data transfer between the processing elements. The communication fabric can take various forms, including buses, switches, NOCs, etc. The NOC 604 in the SOC 610 directs data traffic between the various nodes (e.g., the processing element 606) and links, which provide the communication paths between the nodes.

[00114] The plurality of processing elements including the processing element 606, which can be any suitable type of processors capable of executing instructions, including microprocessors, graphics processing units (GPUs), digital signal processors (DSPs), or application-specific integrated circuits (ASICs).

[00115] Additionally, the SOC 610 may comprise various modules, such as module 232, which are grouped together to provide memory functionality to the assembly 600 as described

herein. The modules in modules group 236 can be coupled to a respective processing element provide it readable memory. In some embodiments, the coupling between the module (e.g., module 232) and the processing element (e.g., 606) can be achieved through interconnects on the silicon substrate 602.

[00116] After the circuitry is formed on the silicon substrate 602, a second layer 608 can be disposed on top of the substrate. The second layer 608 can be any suitable material, such as an insulating material, a metal, a dielectric, or interconnect layer, and it may be bonded to the chiplet 230. The bonding can be done using any suitable technique, including but not limited to, adhesives, soldering, or welding, etc.

[00117] In general, the assembly 600 provides a means of integrating the chiplet 230, which can include the integrated circuit of Fig. 1, with the SOC 610. Integrating the chiplet 230 provides various advantages, such as enhanced functionality, higher performance, and lower power consumption. Moreover, the integration of the chiplet 230 with the SOC 610 can be accomplished in various ways, depending on the particular application and design objectives of the system.

[00118] The assembly 600 can incorporate various variations and modifications, depending on the specific requirements of the system. For example, the processing elements formed on the silicon substrate 602 can vary in their number, type, and arrangement. Similarly, the modules in modules group 236 can vary in their number, type, and function.

[00119] Furthermore, the second layer 608 can be modified to include additional functionality. For instance, the second layer 608 can include passive components, such as resistors, capacitors, and inductors, or active components, such as transistors or diodes. Incorporating these components in the second layer 608 can further enhance the functionality and performance of the system.

[00120] In another variation, the assembly 600 can incorporate a heterogeneous integration approach, where the chiplet 230 is fabricated using a different technology than that used for the SOC 610. This approach allows for the optimal use of different fabrication technologies for different parts of the system, resulting in improved performance and reduced power consumption.

[00121] Various alternatives and modifications can be devised by those skilled in the art without departing from the disclosure. Accordingly, the present disclosure is intended to embrace all such alternatives, modifications, and variances. Additionally, while several embodiments of the present disclosure have been shown in the drawings and/or discussed herein, it is not intended that the disclosure be limited thereto, as it is intended that the

disclosure be as broad in scope as the art will allow and that the specification be read likewise. Therefore, the above description should not be construed as limiting, but merely as exemplifications of particular embodiments. And those skilled in the art will envision other modifications within the scope and spirit of the claims appended hereto. Other elements, steps, methods, and techniques that are insubstantially different from those described above and/or in the appended claims are also intended to be within the scope of the disclosure.

[00122] The embodiments shown in the drawings are presented only to demonstrate certain examples of the disclosure. And the drawings described are only illustrative and are non-limiting. In the drawings, for illustrative purposes, the size of some of the elements may be exaggerated and not drawn to a particular scale. Additionally, elements shown within the drawings that have the same numbers may be identical elements or may be similar elements, depending on the context.

[00123] Where the term "comprising" is used in the present description and claims, it does not exclude other elements or steps. Where an indefinite or definite article is used when referring to a singular noun, e.g., "a," "an," or "the," this includes a plural of that noun unless something otherwise is specifically stated. Hence, the term "comprising" should not be interpreted as being restricted to the items listed thereafter; it does not exclude other elements or steps, and so the scope of the expression "a device comprising items A and B" should not be limited to devices consisting only of components A and B. This expression signifies that, with respect to the present disclosure, the only relevant components of the device are A and B.

[00124] Furthermore, the terms "first," "second," "third," and the like, whether used in the description or in the claims, are provided for distinguishing between similar elements and not necessarily for describing a sequential or chronological order. It is to be understood that the terms so used are interchangeable under appropriate circumstances (unless clearly disclosed otherwise) and that the embodiments of the disclosure described herein are capable of operation in other sequences and/or arrangements than are described or illustrated herein.

[00125] Each of the characteristics and examples described herein, and combination thereof, may be said to be encompassed by the present disclosure. The present disclosure is thus drawn to the following non-limiting numbered aspects:

[00126] 1. An integrated circuit comprising: a modules group having a plurality of modules including a first module and a second module; a shared write port configured to write to the modules group; a first read port configured to read from the first module; and a second read port configured to read from the second module.

[00127] 2. The integrated circuit according to aspect 1, wherein the first and second read ports are configured to be inactive when a write operation is applied to the shared write port.

[00128] 3. The integrated circuit according to aspect 1, wherein the first and second read ports are configured to process reads concurrently relative to each other.

[00129] 4. The integrated circuit according to aspect 1, wherein the shared write port is configured to write to an address space, the shared write port is configured to write to the first module via a first portion of the address space and write to the second module via a second portion of the address space.

[00130] 5. The integrated circuit according to aspect 1, wherein each module of the plurality of modules includes an independent read port for concurrent reading via a respective independent read port of any of the plurality of modules whereby the first read port is configured to read from the first module concurrently with another read of the second module using the second read port.

[00131] 6. The integrated circuit according to aspect 1, wherein the modules group is configured to have a single write address space configured to write to the plurality of modules.

[00132] 7. The integrated circuit according to aspect 6, wherein the first read port is configured to have a first read address space and the second read port is configured to have a second read address space.

[00133] 8. The integrated circuit according to aspect 7, wherein the first read address space numerically overlaps with the second read address space.

[00134] 9. The integrated circuit according to aspect 7, wherein the first read address space is coextensive with the second read address space.

[00135] 10. The integrated circuit according to aspect 7, wherein the first read address space is contiguous with the second read address space.

[00136] 11. The integrated circuit according to aspect 1, wherein the modules group is formed on a second layer portion of an integrated circuit device.

[00137] 12. The integrated circuit according to aspect 1, wherein the integrated circuit is implemented as a chiplet configured to be face-to-face bonded.

[00138] 13. The integrated circuit of aspect 12, wherein the chiplet further comprises a plurality of input/output (I/O) bonds, wherein each of the plurality of input/output (I/O) bonds is configured to provide a respective read port for a respective module of the plurality of modules, wherein each of the plurality of input/output (I/O) bonds is configured to interface

with a complementary respective input/output (I/O) bonds of a second device as bonded with the chiplet.

[00139] 14. A semiconductor device, comprising: the integrated circuit according to aspect 1.

[00140] 15. The semiconductor device according to aspect 14, wherein the semiconductor device is implemented as a die.

[00141] 16. The semiconductor device according to aspect 14, wherein the semiconductor device is implemented as a wafer.

[00142] 17. The semiconductor device according to aspect 14, wherein the semiconductor device is implemented as a Chiplet die.

[00143] 18. The semiconductor device according to aspect 17, wherein the Chiplet die is bonded with a second semiconductor device.

[00144] 19. The semiconductor device according to aspect 18, wherein in the second semiconductor device is selected from a group consisting of a GPU (“Graphics Processing Unit”), an SoC (“System-On-chip”), and a GPPU (“General Purpose Processing Unit”).

[00145] 20. The semiconductor device according to aspect 18, wherein the second semiconductor device is implemented as a Chiplet die.

[00146] 21. The semiconductor device according to aspect 18, wherein the second semiconductor device is implemented as a wafer.

[00147] 22. The semiconductor device according to aspect 18, wherein the second semiconductor device is an ASIC (“Application Specific Semiconductor Chip”).

[00148] 23. The semiconductor device according to aspect 18, wherein the second semiconductor device is an FPGA (“Field Programmable Gate Array”).

[00149] 24. The semiconductor device according to aspect 18, further comprises a plurality of input/output (I/O) bonds, wherein each of the plurality of input/output (I/O) bonds is configured to provide a respective read port for a respective module of the plurality of modules, wherein each of the plurality of input/output (I/O) bonds is configured to electrically connect with a complementary respective input/output (I/O) bonds of the second semiconductor device.

[00150] 25. The semiconductor device according to aspect 18, wherein the Chiplet die has a first side and a second side, wherein the second side is configured to be face-to-face bonded to the second semiconductor device.

[00151] 26. The semiconductor device according to aspect 25, wherein the Chiplet die and second semiconductor device are part of a multi-chip package.

[00152] 27. The semiconductor device according to aspect 18, further comprising an electrically connecting interposer disposed between the Chiplet die and the second semiconductor device.

[00153] 28. The semiconductor device according to aspect 17, wherein the first read port is disposed on the Chiplet die, the Chiplet die having a surface, the first module comprising: a memory array comprising a plurality of memory unit cells arranged within the first module; and a read peripheral circuitry configured to read data stored within the memory array via the first read port.

[00154] 29. The semiconductor device according to aspect 28, wherein a footprint on the surface of the read peripheral circuitry overlaps with a footprint on the surface of the memory array.

[00155] 30. The semiconductor device according to aspect 28, wherein a footprint on the surface of the first read port overlaps a footprint on the surface of the memory array.

[00156] 31. The semiconductor device according to aspect 28, wherein a footprint on the surface of the read peripheral circuitry is coextensive with a footprint on the surface of the memory array.

[00157] 32. The semiconductor device according to aspect 28, wherein a footprint on the surface of the first read port is coextensive with a footprint on the surface of the memory array.

[00158] 33. The semiconductor device according to aspect 28, wherein the memory array, the read peripheral circuitry, and the first read port are stacked vertically and are configured to minimize a module footprint on the surface of the Chiplet die.

[00159] 34. The semiconductor device according to aspect 28, further comprising a write peripheral circuitry connected to the shared write port on the surface of the Chiplet die, wherein: the write peripheral circuitry defines a first footprint on the surface, the first footprint does not overlap with a footprint of the memory array and the first footprint does not overlap with a footprint of the first read port, and the shared write port defines a second footprint on the surface, the second footprint does not overlap with the footprint of the memory array and does not overlap with the footprint of the first read port.

[00160] 35. The semiconductor device according to aspect 28, wherein the plurality of memory unit cells is formed from at least one non-volatile memory unit cell.

[00161] 36. The semiconductor device according to aspect 34, wherein the first footprint overlaps the second footprint.

- [00162] 37. The semiconductor device according to aspect 28, wherein the memory unit cells include a ferroelectric material.
- [00163] 38. The semiconductor device according to aspect 28, wherein the memory unit cells include at least one of a magnetic material, a spin orbit torque material, a spin-transfer-torque material, a phase-change material, and an anti-ferroelectric material.
- [00164] 39. The semiconductor device according to any one of aspects 28-38, wherein the plurality of memory unit cells is arranged in a three-dimensional configuration.
- [00165] 40. The integrated circuit according to aspect 1, wherein the plurality of modules is organized into separate partitions each with a dedicated read peripheral.
- [00166] 41. The integrated circuit according to aspect 40, wherein each dedicated read peripheral has an independent clock.
- [00167] 42. The integrated circuit according to aspect 1, wherein the plurality of modules each has a dedicated read peripheral.
- [00168] 43. The integrated circuit according to aspect 1, wherein the modules group is configured to process write command only during a reset whereby read capability is disabled during the reset.
- [00169] 44. The integrated circuit according to aspect 43, wherein the modules group is configured to disable write capability when the modules group is not in the reset.
- [00170] 45. The integrated circuit according to aspect 1, wherein the modules group includes at least two different non-volatile memory technologies.
- [00171] 46. The integrated circuit according to aspect 1, further comprising a dynamic allocation circuitry to allocate memory blocks to the plurality of modules based on a usage of the modules group.
- [00172] 47. The integrated circuit according to aspect 1, further comprising a write peripheral of the modules group connected to a dedicated I/O pad to enable data transfer external a package of the integrated circuit.
- [00173] 48. The integrated circuit according to aspect 1, wherein the integrated circuit is formed via an additive manufacturing process on a silicon substrate.
- [00174] 49. The integrated circuit according to aspect 48, wherein the integrated circuit is electrically connected to a second semiconductor device comprising another integrated circuit.
- [00175] 50. The integrated circuit according to aspect 49, further comprising a read peripheral for at least one of the plurality of modules disposed on the silicon substrate.

- [00176] 51. The integrated circuit according to aspect 49, further comprising a read peripheral for at least one of the plurality of modules disposed on the second semiconductor device.
- [00177] 52. The integrated circuit according to aspect 1, wherein the integrated circuit is electrically connected to a second semiconductor device comprising another integrated circuit.
- [00178] 53. The integrated circuit according to aspect 52, wherein the second semiconductor device is a system-on-chip.
- [00179] 54. The integrated circuit according to aspect 52, wherein the second semiconductor device is implemented as a Chiplet die.
- [00180] 55. The integrated circuit according to aspect 52, wherein the second semiconductor device is implemented as a wafer.
- [00181] 56. The integrated circuit according to aspect 52, wherein the second semiconductor device is an ASIC (“Application Specific Semiconductor Chip”).
- [00182] 57. The integrated circuit according to aspect 52, wherein the second semiconductor device is an FPGA (“Field Programmable Gate Array”).
- [00183] 58. The integrated circuit according to aspect 52, wherein the plurality of modules is formed from non-volatile memory unit cells arranged in a three-dimensional connectivity fabric vertical to a silicon substrate and the second semiconductor device.
- [00184] 59. The integrated circuit according to aspect 52, wherein the plurality of modules is formed utilizing at least one of a cross-point, 3D NANDs, 3D NORs, 3D ANDs, and a stacked planar layer.
- [00185] 60. The integrated circuit according to aspect 1, wherein the plurality of modules has a single write peripheral.
- [00186] 61. The integrated circuit according to aspect 60, wherein the single write peripheral has a dedicated clock.
- [00187] 62. The integrated circuit according to aspect 1, further comprising a plurality of clocks, wherein each clock of the plurality of clocks feeds a respective module of the plurality of modules whereby each respective module has decoupled timing relative to other modules of the plurality of modules, wherein each of the plurality of clocks is configured for clocking a respective read port of a respective module of the plurality of modules.
- [00188] 63. The integrated circuit according to aspect 1, wherein the modules group is formed on a chiplet having a first side and a second side, the second side being configured for bonding to a second semiconductor device.

- [00189] 64. The integrated circuit according to aspect 63, further comprising a decoder circuit, a driver circuit, and a register circuit on a silicon substrate of the chiplet.
- [00190] 65. The integrated circuit according to aspect 64, wherein the modules group is formed on a second layer of the chiplet.
- [00191] 66. The integrated circuit according to aspect 63, the second semiconductor device comprising a plurality of processing elements, whereby each processing element includes a respective interface to communicate with a respective module of the plurality of modules on the modules group when the second semiconductor device is bonded to the chiplet.
- [00192] 67. The integrated circuit according to aspect 63, the chiplet comprising an interface to the shared write port on the second side to thereby interface with a complementary interface on the second semiconductor device.
- [00193] 68. The integrated circuit according to aspect 63, the second semiconductor device comprising a network-on-a-chip.
- [00194] 69. The integrated circuit according to aspect 68, the second semiconductor device comprising a plurality of processing elements, wherein the network-on-a-chip is configured to provide inter-element communication for the plurality of processing elements.
- [00195] 70. The integrated circuit according to aspect 69, wherein the plurality of processing elements includes at least one embedded FPGA.
- [00196] 71. The integrated circuit according to aspect 69, wherein the plurality of processing elements includes at least one of a soft processor, a DSP block, an embedded processor, and a microcontroller.
- [00197] 72. The integrated circuit according to aspect 1, wherein a write operation to the modules group is performed through a priority arbitration circuit that facilitates the modules to be accessed in a predetermined order.
- [00198] 73. The integrated circuit according to aspect 1, further comprising a power gating circuitry that selectively powers down a module of the plurality of modules when not in use.
- [00199] 74. The integrated circuit according to aspect 1, wherein the shared write port is configured to write to a virtual address space that is mapped onto a physical memory space.
- [00200] 75. The integrated circuit according to aspect 1, further comprising a control circuit configured to enable the shared write port during write operations and disable the shared write port during read operations thereby conserving energy.

[00201] 76. The integrated circuit according to aspect 1, further comprising a control circuit configured to disable the first read port during write operations and enable the first read port during read operations thereby conserving energy.

[00202] 77. The integrated circuit according to aspect 1, further comprising a power management module configured to selectively power down the shared write port during read operations thereby conserving energy.

[00203] 78. The integrated circuit according to aspect 1, further comprising write circuitry configured to dynamically shift power allocation from write operations to read operations thereby conserving energy.

[00204] 79. The integrated circuit according to aspect 1, further comprising write circuitry configured to enter a sleep mode and power down the shared write port during read operations thereby conserving energy.

[00205] 80. The integrated circuit according to aspect 1, wherein the plurality of modules is formed from a non-volatile memory selected from the group consisting of a FeFET, a FeRAM, a ReRAM, a SOT (Spin Orbit Torque), and a STT (Spin Transfer Torque).

[00206] 81. The integrated circuit according to aspect 1, wherein a write peripheral for modules group is implemented on a silicon substrate.

[00207] 82. The integrated circuit according to aspect 81, wherein the integrated circuit is formed within a chiplet having a first side and a second side, wherein the second side is configured to be bonded to another integrated circuit disposed within a packaging, wherein the write peripheral is disposed between the modules group and the second side of the chiplet.

[00208] 83. The integrated circuit according to aspect 1, wherein a read peripheral for the first module is implemented on a silicon substrate.

[00209] 84. The integrated circuit according to aspect 83, wherein the integrated circuit is formed within a chiplet having a first side and a second side, wherein the second side is configured to be bonded to another integrated circuit disposed within a packaging, wherein the read peripheral is disposed between the modules group and the first side of the chiplet.

[00210] 85. The integrated circuit according to aspect 1, wherein a read peripheral for the first module is implemented in a second layer.

[00211] 86. The integrated circuit according to aspect 85, wherein the integrated circuit is formed within a chiplet having a first side and a second side, wherein the second side is configured to be bonded to another integrated circuit disposed within a packaging, wherein the read peripheral is disposed between the modules group and the first side of the chiplet.

[00212] 87. The integrated circuit according to aspect 1, wherein each module has its own dedicated write peripheral.

[00213] 88. The integrated circuit according to aspect 87, wherein each dedicated write peripheral is configured to utilize a shared clock.

[00214] 89. The integrated circuit as in aspects 1-27, 30, 32, 35-40, 43-46, 48-49, 52-60, 62-80, or 87-88, wherein the integrated circuit is disposed on a first semiconductor device, the integrated circuit further comprising: a shared write peripheral configured to write to the modules group via the shared write port, the shared write peripheral disposed on a silicon substrate of the first semiconductor device; a first read peripheral configured to read the first module via the first read port, wherein the first read peripheral is disposed on the silicon substrate of the first semiconductor device; and a second read peripheral configured to read the second module via the second read port, wherein the second read peripheral is disposed on the silicon substrate of the first semiconductor device, wherein the first module and the second module are disposed on a second layer of the first semiconductor device.

[00215] 90. The integrated circuit according to aspect 89, wherein the first module is a three-dimensional column of memory bit cells disposed between a first surface of the first semiconductor device and the silicon substrate of the first semiconductor device.

[00216] 91. The integrated circuit according to aspect 89, wherein the first read port is configured to be in electrical communication with a second semiconductor device, the second semiconductor device configured to be secured to a first surface of the first semiconductor device.

[00217] 92. The integrated circuit according to aspect 91, further comprising an interposer disposed between the first surface of the first semiconductor device and the second semiconductor device.

[00218] 93. The integrated circuit according to aspect 89, wherein the first read port includes a read address bus that traverses through the second layer of the first semiconductor device.

[00219] 94. The integrated circuit according to aspect 89, wherein the first read port includes a read data bus that traverses through the second layer of the first semiconductor device.

[00220] 95. The integrated circuit according to aspect 89, wherein the shared write port is configured to be in electrical communication with a second semiconductor device, the second semiconductor device configured to be secured to a first surface of the first semiconductor device.

[00221] 96. The integrated circuit according to aspect 95, further comprising an electrically connecting interposer disposed between the first surface of the first semiconductor device and the second semiconductor device.

[00222] 97. The integrated circuit according to aspect 89, wherein the shared write port includes a write address bus that traverses through the second layer of the first semiconductor device.

[00223] 98. The integrated circuit according to aspect 89, wherein shared write port includes a write data bus that traverses through the second layer of the first semiconductor device.

[00224] 99. A method, comprising: forming an integrated circuit of any one of aspects 1-98.

[00225] 100. A method of using the integrated circuit of any one of aspects 1-98, the method comprising: writing to the modules group; and reading from the modules group.

What is Claimed is:

1. An integrated circuit comprising:
 - a modules group having a plurality of modules including a first module and a second module;
 - a shared write port configured to write to the modules group;
 - a first read port configured to read from the first module; and
 - a second read port configured to read from the second module.
2. The integrated circuit according to claim 1, wherein the first and second read ports are configured to be inactive when a write operation is applied to the shared write port.
3. The integrated circuit according to claim 1, wherein the first and second read ports are configured to process reads concurrently relative to each other.
4. The integrated circuit according to claim 1, wherein the shared write port is configured to write to an address space, the shared write port is configured to write to the first module via a first portion of the address space and write to the second module via a second portion of the address space.
5. The integrated circuit according to claim 1, wherein each module of the plurality of modules includes an independent read port for concurrent reading via a respective independent read port of any of the plurality of modules whereby the first read port is configured to read from the first module concurrently with another read of the second module using the second read port.
6. The integrated circuit according to claim 1, wherein the modules group is configured to have a single write address space configured to write to the plurality of modules.
7. The integrated circuit according to claim 6, wherein the first read port is configured to have a first read address space and the second read port is configured to have a second read address space.
8. The integrated circuit according to claim 7, wherein the first read address space numerically overlaps with the second read address space.
9. The integrated circuit according to claim 7, wherein the first read address space is coextensive with the second read address space.
10. The integrated circuit according to claim 7, wherein the first read address space is contiguous with the second read address space.

11. The integrated circuit according to claim 1, wherein the modules group is formed on a second layer portion of an integrated circuit device.
12. The integrated circuit according to claim 1, wherein the integrated circuit is implemented as a chiplet configured to be face-to-face bonded.
13. The integrated circuit of claim 12, wherein the chiplet further comprises a plurality of input/output (I/O) bonds, wherein each of the plurality of input/output (I/O) bonds is configured to provide a respective read port for a respective module of the plurality of modules, wherein each of the plurality of input/output (I/O) bonds is configured to interface with a complementary respective input/output (I/O) bonds of a second device as bonded with the chiplet.
14. A method of forming an integrated circuit, the method comprising:
 - forming a modules group having a plurality of modules including a first module and a second module;
 - forming a shared write port configured to write to the modules group;
 - forming a first read port configured to read from the first module; and
 - forming a second read port configured to read from the second module.