

EXHIBIT 803-A-07

AMD's Implementation of HBM and/or Infinity Fabric ("AMD Product Art")

AMD's Implementation of HBM and/or Infinity Fabric in the following products:

- AMD Radeon R9 Fury Series, including the Fiji GPU (e.g., R9 Fury, R9 Fury X, R9 Nano, R9 Fury X2)
- AMD Radeon Vega Series, including the Vega GPU (e.g., RX Vega 56, RX Vega 64, Radeon VII)
- AMD Instinct MI25, including the Vega GPU
- AMD EPYC 7001 Series (Naples) (e.g., EPYC 7251, 7281, 7301, 7351, 7351P, 7401, 7401P, 7451, 7501, 7551, 7551P, 7601, 7371, 7571)

as shown in the references listed below, anticipates and/or renders obvious (alone or in combination with other references), including at least under Complainant's apparent infringement theory claims 1-3, 7-10 of U.S. Patent No. 11,841,803 ("the '803 patent") under 35 U.S.C. §§ 102 and/or 103.

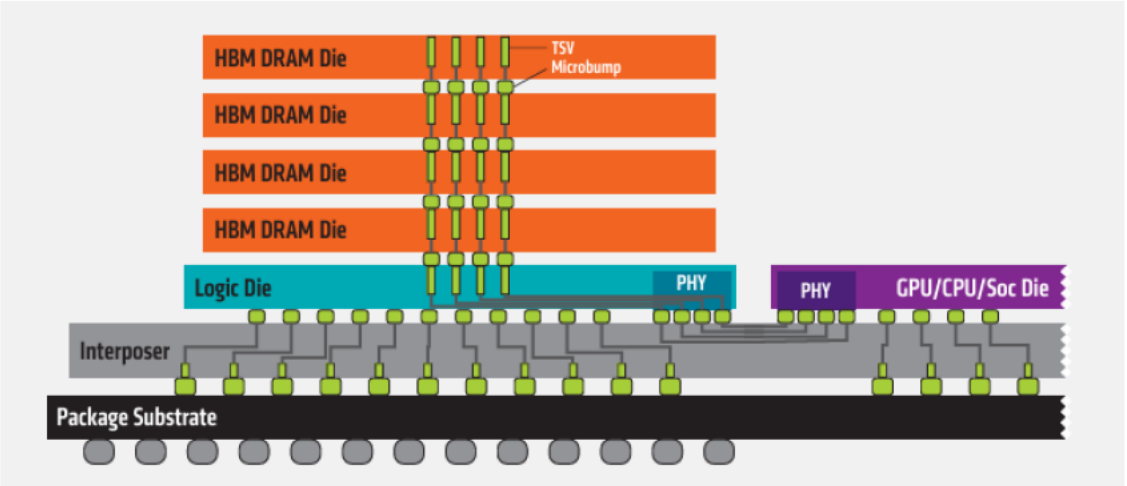
- NVIDITC-000133188: AMD, *High Bandwidth Memory (HBM) Whitepaper*, AMD.com (archived Nov. 24, 2018), <https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf>
- NVIDITC-001015674: AMD, *High Bandwidth Memory (HBM)*, AMD.com (archived Oct. 19, 2018), <https://web.archive.org/web/20181019185554/https://www.amd.com/en/technologies/hbm>
- NVIDITC-000133172: Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/>
- NVIDITC-000135018: Timothy Prickett Morgan, *AMD's Long Road from Naples to Milan Centers on Rome*, The Next Platform (Nov. 6, 2018), <https://www.nextplatform.com/2018/11/06/amds-long-road-from-naples-to-milan-centers-on-rome/>
- NVIDITC-000134966: Timothy Prickett Morgan, *AMD Doubles Down – And Up – With Rome Epyc Server Chips*, The Next Platform (Aug. 7, 2019), <https://www.nextplatform.com/2019/08/07/amd-doubles-down-and-up-with-rome-epyc-server-chips/>
- NVIDITC-001016951: AMD Newsroom, *Vega: AMD's New Graphics Architecture for Virtually Unlimited Workloads*, AMD (Jan. 5, 2017)
- NVIDITC-001015664: Usman Pirzada, *AMD's Infinity Fabric Detailed –The Innovative, Real-World Implementation of The Company's 'Perfect Lego' Philosophy*, WCCFTech (Jan. 16, 2017)
- NVIDITC-001015651: *AMD Radeon R9 Fury X Detailed*, Geeks3D (June 18, 2015)
- NVIDITC-001015605: *AMD Radeon (TM) R9 Nano, World's Smallest and Most Power-Efficient Enthusiast Graphics Card, Brings 4K Gaming to the Living Room*, AMD (Aug. 27, 2015)

NVIDIA has charted the portions of the reference that correspond to each element, including at least under Complainant's apparent interpretation of the claims as reflected in Complainant's infringement contentions. Where the chart below states that the reference "discloses," "teaches," or "suggests" a claim limitation, such disclosure, teaching, or suggestion may be express, implied, or inherent. The cited evidence is merely illustrative, and NVIDIA reserves the right to cite alternative or additional evidence.

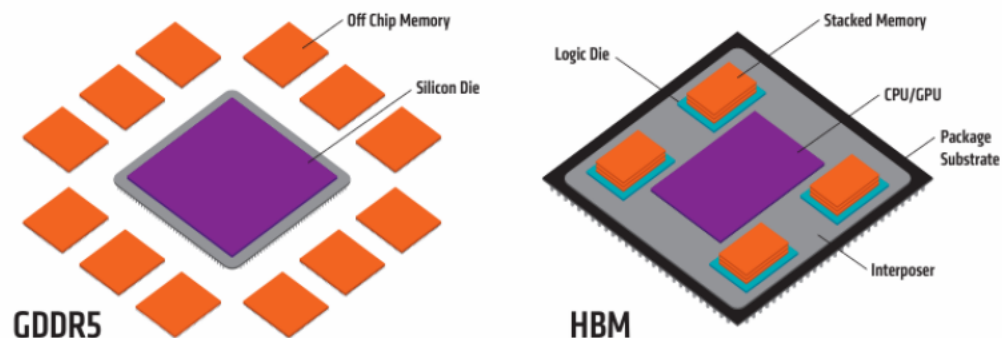
No part of this exemplary chart construes, or is intended to construe, the specification, file history, or claims of the asserted patents. In addition, nothing stated in this chart shall be treated as an admission or suggestion that NVIDIA agrees with Complainant regarding either the scope of any of the asserted claims or Complainant's apparent interpretation of the scope of the asserted claims, or that NVIDIA's accused products meet any limitations of the claims.

To the extent that Complainant contends that AMD Product Art does not disclose one or more limitations of the asserted claims, it would have been obvious to combine the teachings of AMD Product Art with (1) the knowledge of one of ordinary skill in the art, and/or (2) the teachings of any of the prior art set forth in NVIDIA's other invalidity charts and cover pleading, to show all the limitations of the asserted claims, including for the reasons set forth in NVIDIA's Invalidity Contention Cover Pleading, which are incorporated herein.

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
1[Pre]: A system, comprising:	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See evidence cited for '803 patent claims 1-3, 7-10.</i> This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.

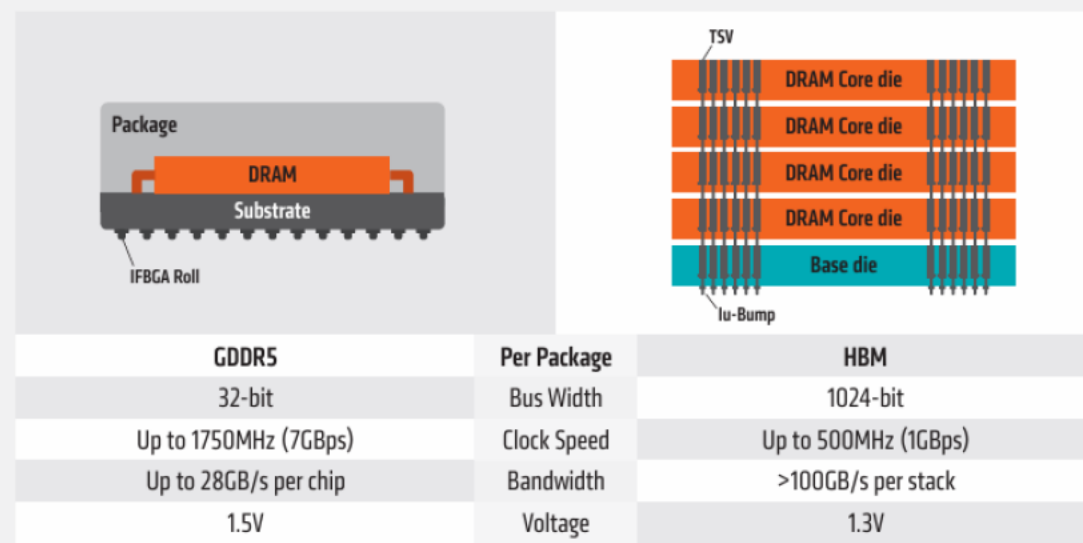
U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
1[A]: a central processing unit (CPU) communicably coupled to a first graphics processing unit (GPU) chiplet of a GPU chiplet array,	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> AMD, <i>High Bandwidth Memory (HBM) Whitepaper</i>, AMD.com (archived Nov. 24, 2018), https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf  AMD, <i>High Bandwidth Memory (HBM) Whitepaper</i>, AMD.com (archived Nov. 24, 2018), https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf

HBM vs GDDR5:
HBM shortens your information commute



AMD, *High Bandwidth Memory (HBM) Whitepaper*, AMD.com (archived Nov. 24, 2018), <https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf>

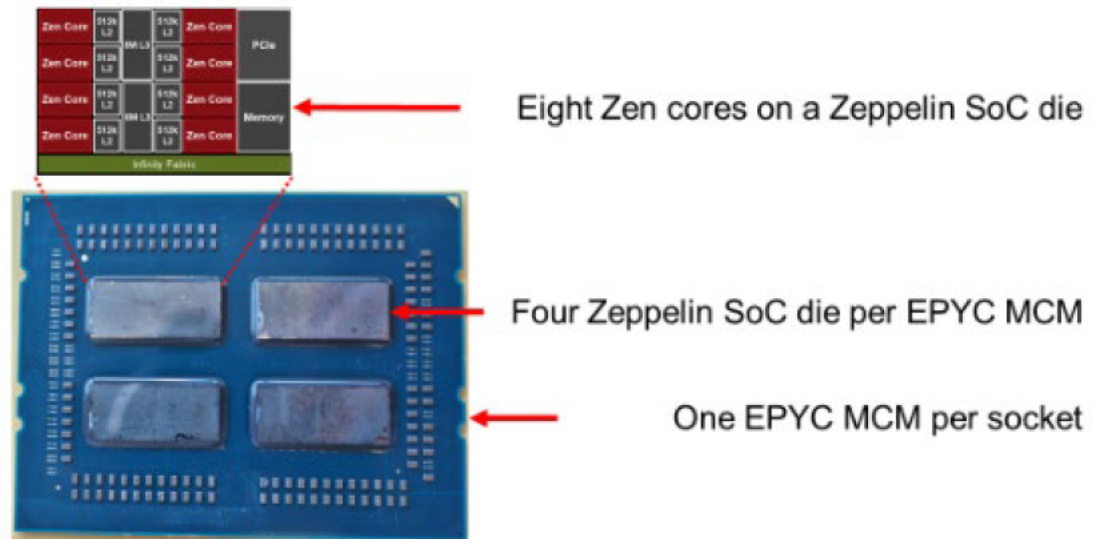
HBM vs GDDR5: Compare side by side



AMD, *High Bandwidth Memory (HBM) Whitepaper*, AMD.com (archived Nov. 24, 2018), <https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf> ("HBM is a new type of memory chip with low power consumption and ultra-wide communication lanes. It uses vertically stacked memory chips interconnected by microscopic wires called "through-silicon vias," or TSVs.").

AMD, *High Bandwidth Memory (HBM)*, AMD.com (archived Oct. 19, 2018), <https://web.archive.org/web/20181019185554/https://www.amd.com/en/technologies/hbm> ("HBM's vertical stacking and fast information transfer open the door for truly exciting performance in innovative form factors. And GPU applications are just the start – look for HBM's superior power efficiency and space savings to spark industry-wide innovation.").

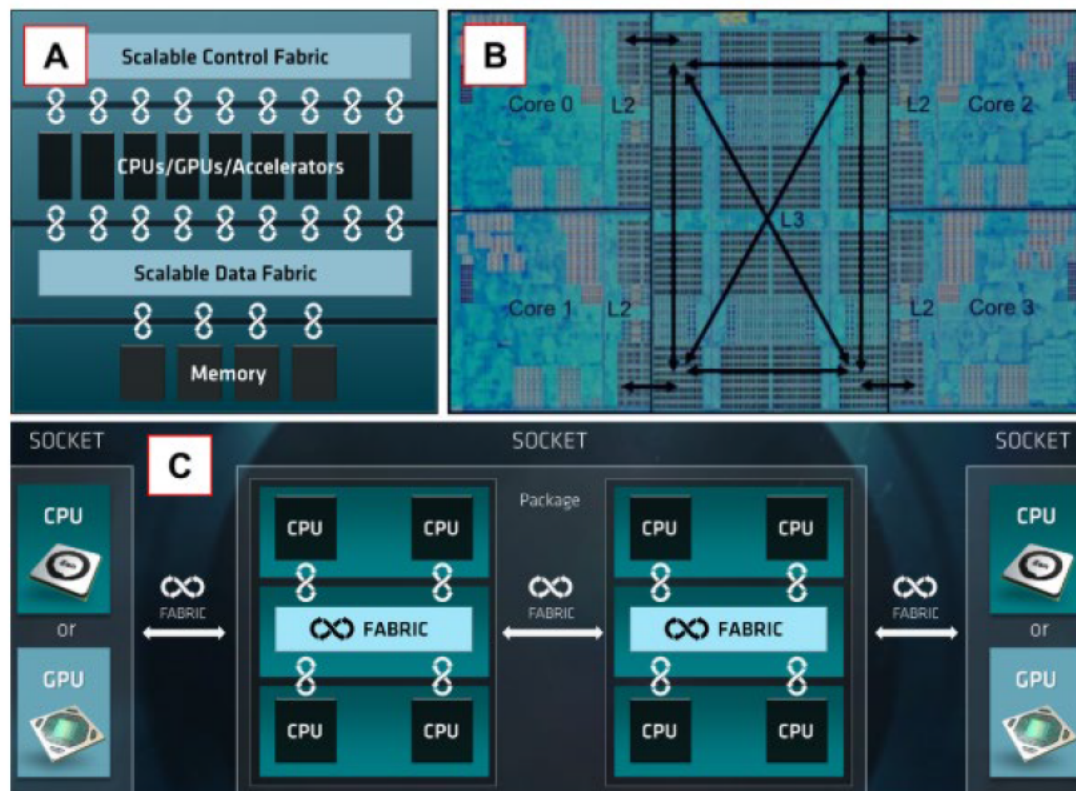
U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	AMD, <i>High Bandwidth Memory (HBM)</i>, AMD.com (archived Oct. 19, 2018), https://web.archive.org/web/20181019185554/https://www.amd.com/en/technologies/hbm ("HBM is a new type of CPU/GPU memory ("RAM") that vertically stacks memory chips, like floors in a skyscraper. In doing so, it shortens your information commute. Those towers connect to the CPU or GPU through an ultra-fast interconnect called the "interposer." Several stacks of HBM are plugged into the interposer alongside a CPU or GPU, and that assembled module connects to a circuit board. Though these HBM stacks are not physically integrated with the CPU or GPU, they are so closely and quickly connected via the interposer that HBM's characteristics are nearly indistinguishable from on-chip integrated RAM."). AMD, <i>High Bandwidth Memory (HBM)</i>, AMD.com (archived Oct. 19, 2018), https://web.archive.org/web/20181019185554/https://www.amd.com/en/technologies/hbm ("First on-die GPU: AMD's Accelerated Processing Units (APUs) were the first to explore the integration of a GPU with the CPU, eliminating the need for a bulky external GPU for compact or inexpensive PCs."). Paul Teich, <i>The Heart Of AMD's Epyc Comeback Is Infinity Fabric</i>, The Next Platform (July 12, 2017), https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/



Source: AMD

AMD Epyc deconstructed

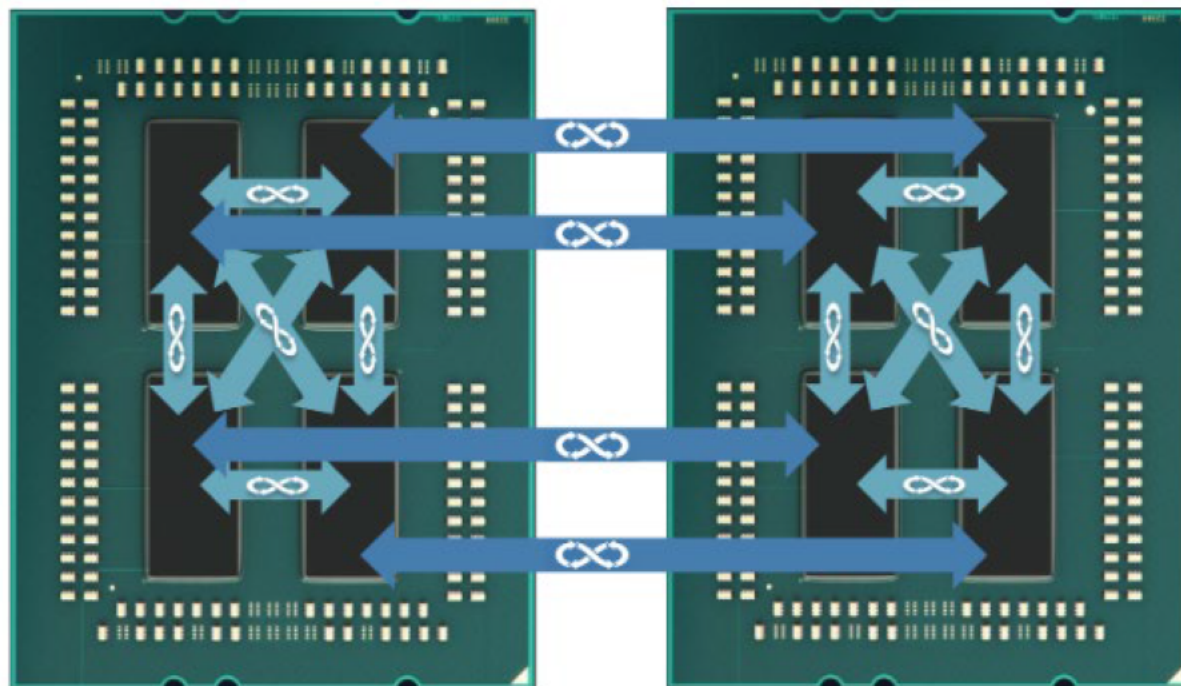
Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/>



Source: AMD

AMD Infinity Fabric high-level block diagram and detail

Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/>

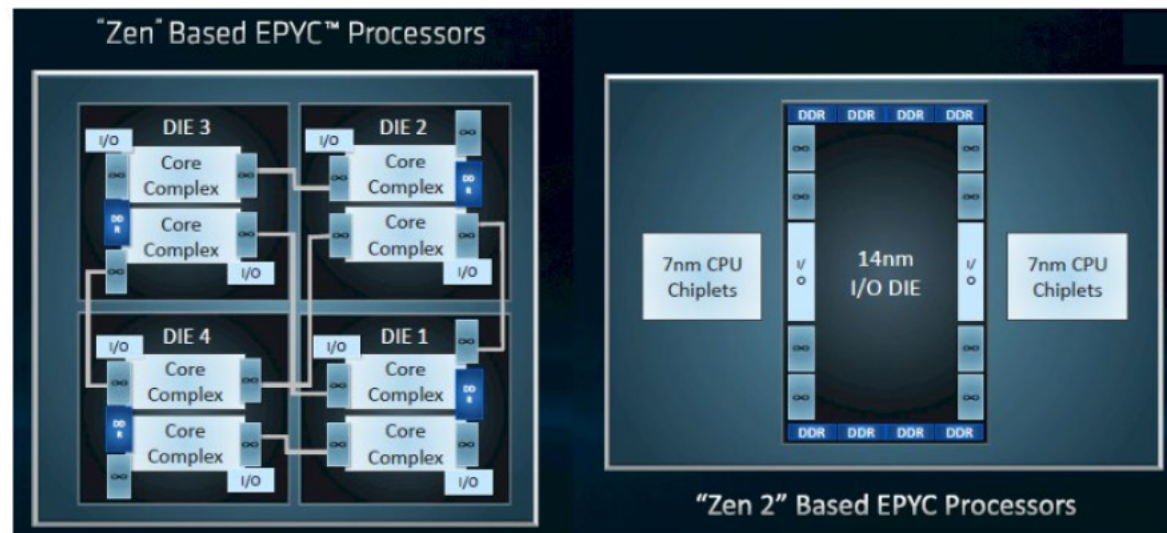


Source: AMD

AMD Infinity Fabric connecting Zeppelin die on a MCM and between MCMs

Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/> ("AMD designed Epyc as a modular architecture around its "Zeppelin" processor die with its eight "Zen" architecture cores. To allow multi-die scalability, AMD first reworked its HyperTransport socket-to-socket system I/O architecture for use on a chip, across a multi-chip module (MCM), and for inter-socket connectivity. AMD has named this coherent connection the Infinity Fabric.").

Timothy Prickett Morgan, *AMD's Long Road from Naples to Milan Centers on Rome*, The Next Platform (Nov. 6, 2018), <https://www.nextplatform.com/2018/11/06/amds-long-road-from-naples-to-milan-centers-on-rome/>



Timothy Prickett Morgan, *AMD's Long Road from Naples to Milan Centers on Rome*, The Next Platform (Nov. 6, 2018), <https://www.nextplatform.com/2018/11/06/amds-long-road-from-naples-to-milan-centers-on-rome/> ("The compute core blocks, on the other hand, scale down really well from 14 nanometers to 7 nanometers. In this case, it looks like four distinct blocks of cores, comprised of eight total eight-core chiplets based on the Zen 2 cores, are on the Rome package and they link to this I/O hub chip through an improved Infinity Fabric, itself based on PCI-Express 4.0 signaling.").

Timothy Prickett Morgan, *AMD Doubles Down – And Up – With Rome Epyc Server Chips*, The Next Platform (Aug. 7, 2019), <https://www.nextplatform.com/2019/08/07/amd-doubles-down-and-up-with-rome-epyc-server-chips/>

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	 The diagram illustrates the evolution of AMD's processor architectures. It is divided into three main sections: 'Traditional Monolithic', '1st Gen EPYC', and '2nd Gen EPYC'. 1. 'Traditional Monolithic' shows a single large die with a central core array (green squares) and peripheral I/O blocks (blue rectangles). 2. '1st Gen EPYC' shows a multi-die architecture with four dies (Die 0, Die 1, Die 2, Die 3) connected by an Infinity Fabric network (blue lines). Each die contains a core array and I/O blocks. 3. '2nd Gen EPYC' shows a more complex multi-die architecture with a central '14nm I/O Die' and multiple '7nm CCD' dies connected by an Infinity Fabric network. Arrows indicate the progression from traditional monolithic to 1st Gen EPYC and then to 2nd Gen EPYC. This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.

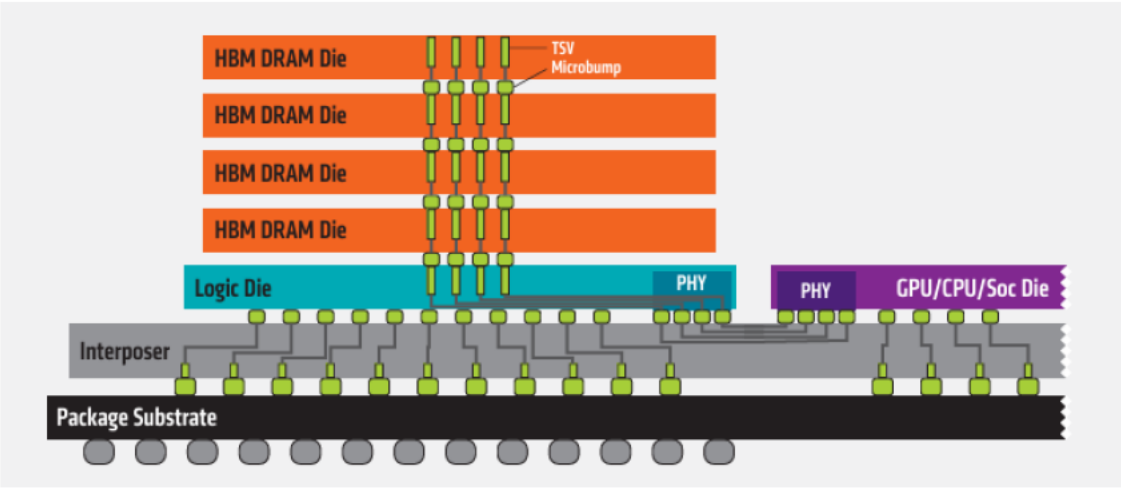
1[B]: wherein the GPU chiplet array includes: the first GPU chiplet communicably coupled to the CPU via a bus; and

Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation.

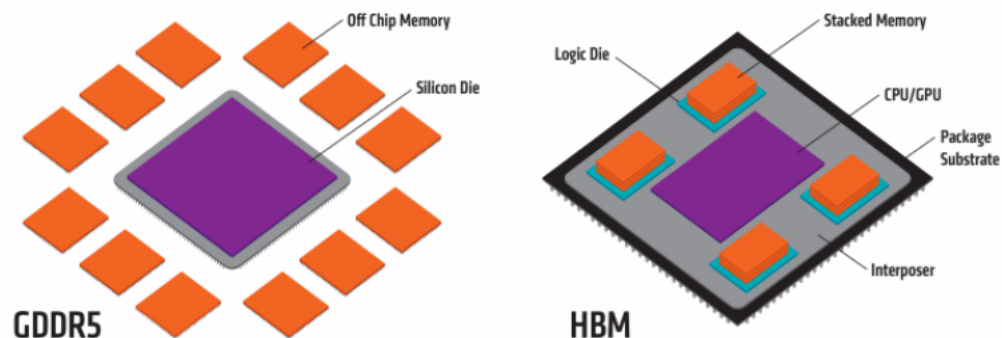
See, e.g.,

See evidence cited for '803 patent limitation 1[A].

This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.

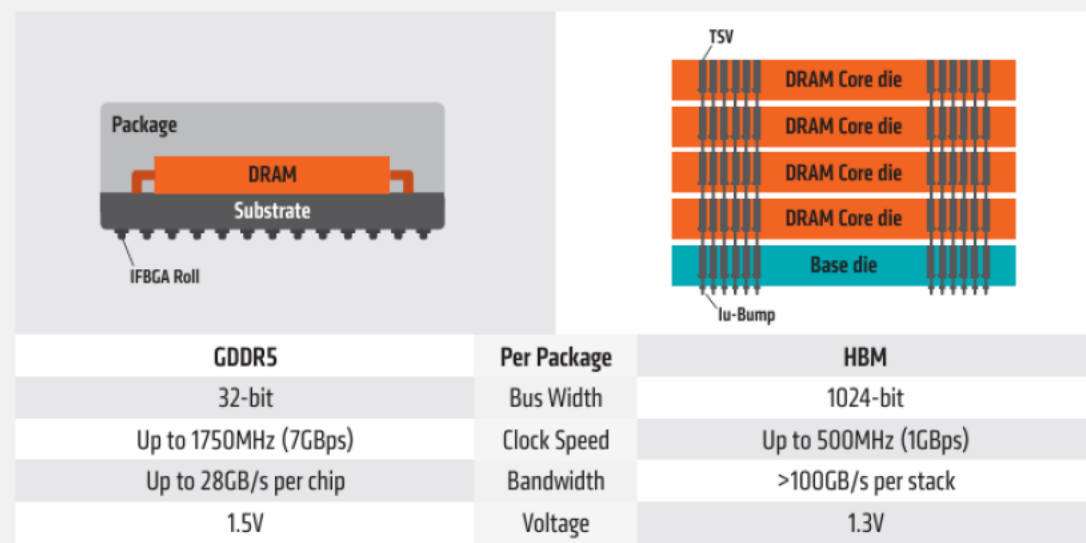
U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
1[C]: a second GPU chiplet communicably coupled to the first GPU chiplet via a passive crosslink,	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> AMD, <i>High Bandwidth Memory (HBM) Whitepaper</i>, AMD.com (archived Nov. 24, 2018), https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf  AMD, <i>High Bandwidth Memory (HBM) Whitepaper</i>, AMD.com (archived Nov. 24, 2018), https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf

HBM vs GDDR5:
HBM shortens your information commute



AMD, *High Bandwidth Memory (HBM) Whitepaper*, AMD.com (archived Nov. 24, 2018), <https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf>

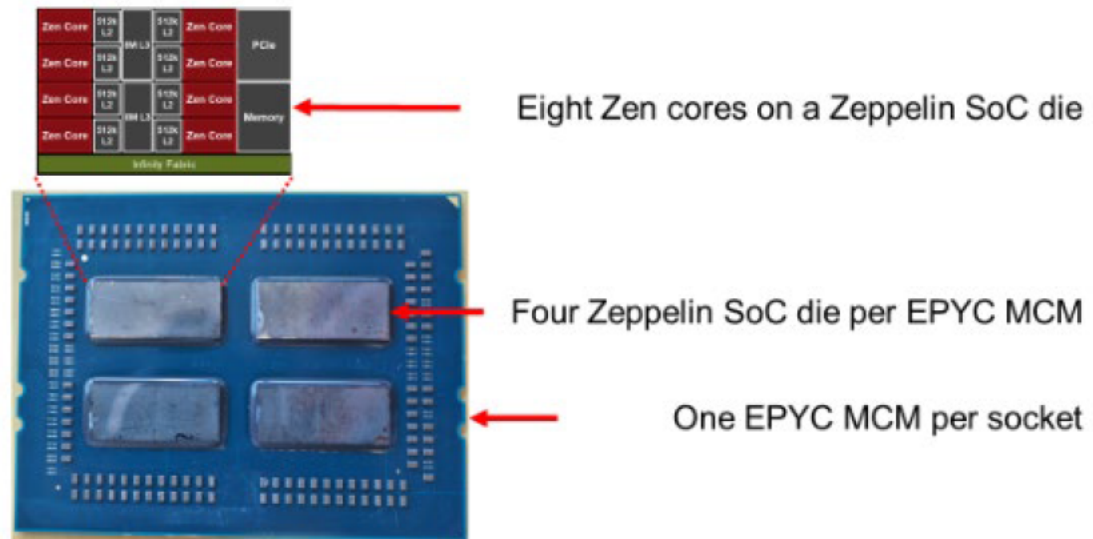
HBM vs GDDR5: Compare side by side



AMD, *High Bandwidth Memory (HBM) Whitepaper*, AMD.com (archived Nov. 24, 2018), <https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf> ("HBM is a new type of memory chip with low power consumption and ultra-wide communication lanes. It uses vertically stacked memory chips interconnected by microscopic wires called "through-silicon vias," or TSVs.").

AMD, *High Bandwidth Memory (HBM)*, AMD.com (archived Oct. 19, 2018), <https://web.archive.org/web/20181019185554/https://www.amd.com/en/technologies/hbm> ("HBM's vertical stacking and fast information transfer open the door for truly exciting performance in innovative form factors. And GPU applications are just the start – look for HBM's superior power efficiency and space savings to spark industry-wide innovation.").

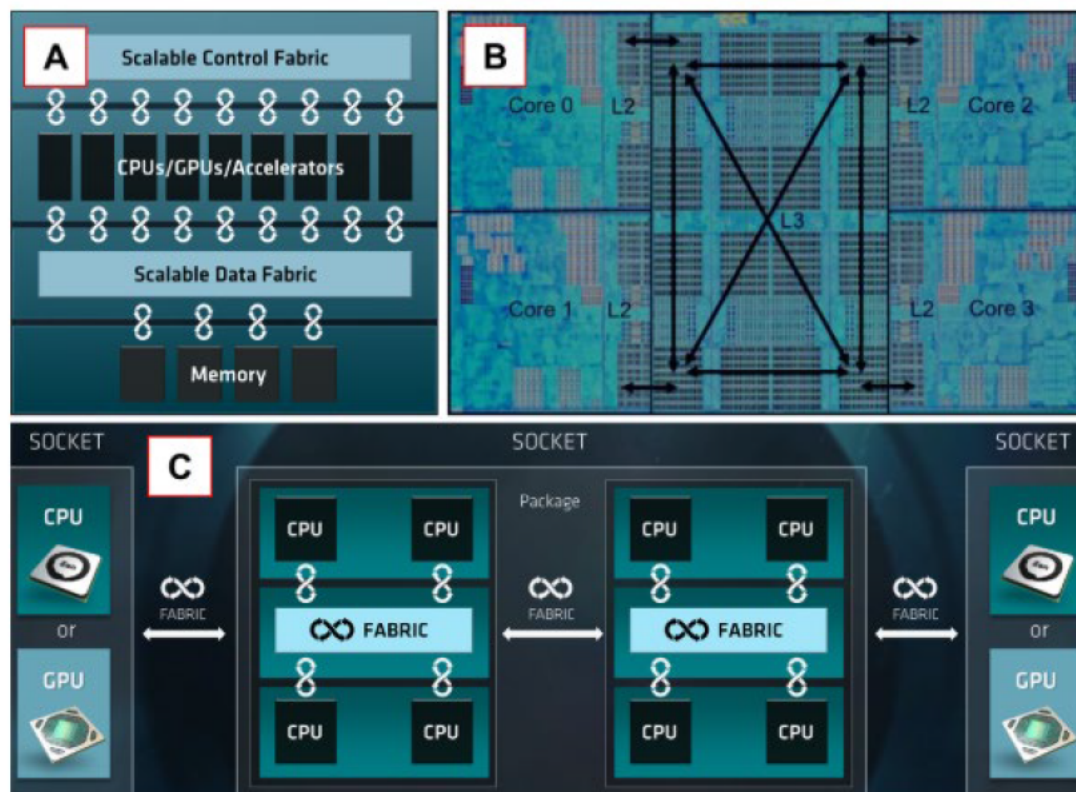
U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	AMD, <i>High Bandwidth Memory (HBM)</i>, AMD.com (archived Oct. 19, 2018), https://web.archive.org/web/20181019185554/https://www.amd.com/en/technologies/hbm ("HBM is a new type of CPU/GPU memory ("RAM") that vertically stacks memory chips, like floors in a skyscraper. In doing so, it shortens your information commute. Those towers connect to the CPU or GPU through an ultra-fast interconnect called the "interposer." Several stacks of HBM are plugged into the interposer alongside a CPU or GPU, and that assembled module connects to a circuit board. Though these HBM stacks are not physically integrated with the CPU or GPU, they are so closely and quickly connected via the interposer that HBM's characteristics are nearly indistinguishable from on-chip integrated RAM."). Paul Teich, <i>The Heart Of AMD's Epyc Comeback Is Infinity Fabric</i>, The Next Platform (July 12, 2017), https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric



Source: AMD

AMD Epyc deconstructed

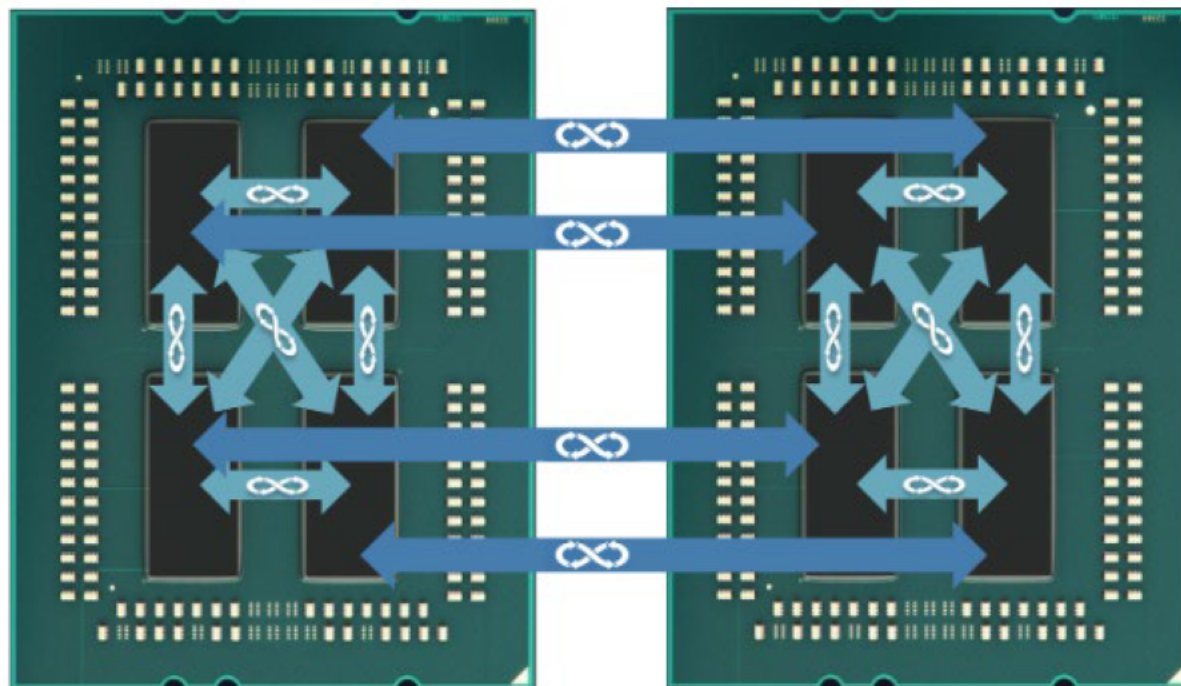
Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/>



Source: AMD

AMD Infinity Fabric high-level block diagram and detail

Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/>



Source: AMD

AMD Infinity Fabric connecting Zeppelin die on a MCM and between MCMs

Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/> <https://bytebridge.medium.com/the-critical-role-of-high-bandwidth-memory-hbm-in-modern-gpus-63b0164f29bc> ("AMD designed Epyc as a modular architecture around its "Zeppelin" processor die with its eight "Zen" architecture cores. To allow multi-die scalability, AMD first reworked its HyperTransport socket-to-socket system I/O architecture for use on a chip, across a multi-chip module (MCM), and for inter-socket connectivity. AMD has named this coherent connection the Infinity Fabric.").

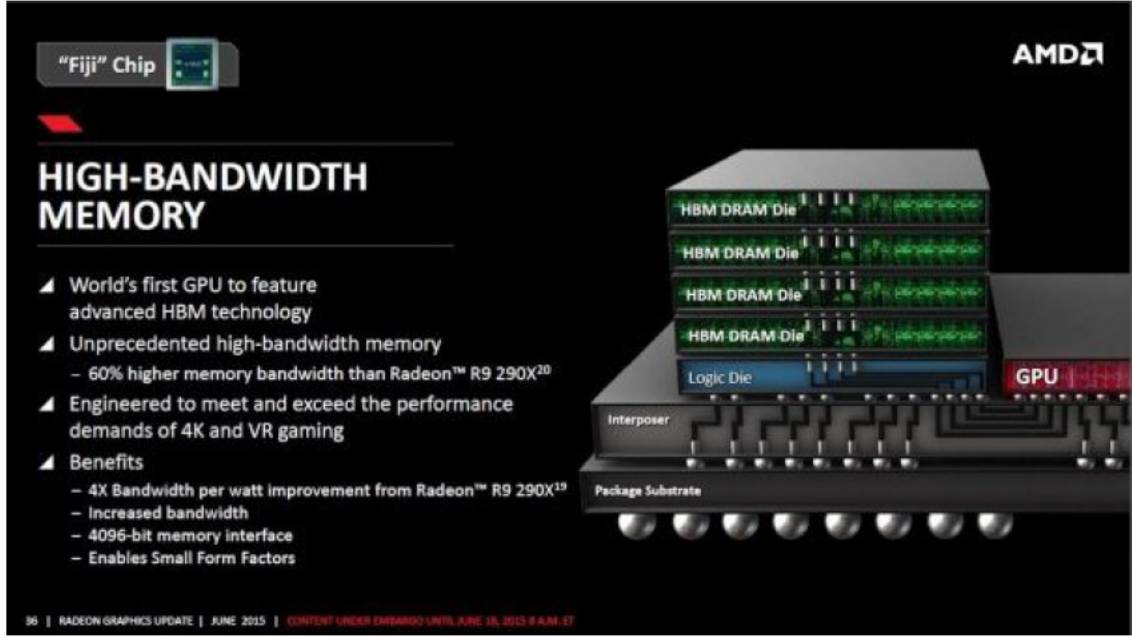
U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	Paul Teich, <i>The Heart Of AMD's Epyc Comeback Is Infinity Fabric</i>, The Next Platform (July 12, 2017), https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/ ("Starting with AMD's Ryzen desktop processor and Epyc server architecture, AMD will implement their scalable Infinity Fabric across all its SoC and MCM products. Think of Infinity Fabric as a superset of HyperTransport, AMD's previous socket-to-socket interconnect architecture, now managed by the HyperTransport Consortium. Infinity Fabric is a coherent high-performance fabric that uses sensors embedded in each die to scale control and data flow from die to socket to board-level."). Timothy Prickett Morgan, <i>AMD's Long Road from Naples to Milan Centers on Rome</i>, The Next Platform (Nov. 6, 2018), https://www.nextplatform.com/2018/11/06/amds-long-road-from-naples-to-milan-centers-on-rome/ The diagram is divided into two main sections. The left section, titled "'Zen' Based EPYC™ Processors", shows a 2x2 grid of four dies labeled DIE 1, DIE 2, DIE 3, and DIE 4. Each die contains two 'Core Complex' blocks and 'I/O' blocks at the corners. The right section, titled "'Zen 2' Based EPYC Processors", shows a central '14nm I/O DIE' flanked by two '7nm CPU Chiplets'. The I/O die has 'DDR' memory blocks on its top and bottom edges and 'I/O' blocks on its left and right edges. Timothy Prickett Morgan, <i>AMD's Long Road from Naples to Milan Centers on Rome</i>, The Next Platform (Nov. 6, 2018), https://www.nextplatform.com/2018/11/06/amds-long-road-from-naples-to-milan-centers-on-rome/ ("The compute core blocks, on the other hand, scale down really well from 14 nanometers to 7 nanometers. In this case, it looks like four distinct

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	blocks of cores, comprised of eight total eight-core chiplets based on the Zen 2 cores, are on the Rome package and they link to this I/O hub chip through an improved Infinity Fabric, itself based on PCI-Express 4.0 signaling.”). Timothy Prickett Morgan, <i>AMD Doubles Down – And Up – With Rome Epyc Server Chips</i>, The Next Platform (Aug. 7, 2019), https://www.nextplatform.com/2019/08/07/amd-doubles-down-and-up-with-rome-epyc-server-chips/  The diagram illustrates the evolution of AMD's server chip architecture. It is divided into three main sections: 'Traditional Monolithic', '1st Gen EPYC', and '2nd Gen EPYC'. 1. 'Traditional Monolithic' shows a single large die with a central core array (green squares) and peripheral I/O and DRAM blocks (blue and grey rectangles). 2. '1st Gen EPYC' shows a multi-die architecture with four dies (Die 0, Die 1, Die 2, Die 3) connected by an Infinity Fabric network. Each die contains a core array and I/O/DRAM blocks. 3. '2nd Gen EPYC' shows a more complex multi-die architecture with a central '14nm I/O Die' and multiple '7nm CCD' dies. The dies are interconnected via an Infinity Fabric network, and the central die handles I/O and DRAM access for the core dies. Timothy Prickett Morgan, <i>AMD Doubles Down – And Up – With Rome Epyc Server Chips</i>, The Next Platform (Aug. 7, 2019), https://www.nextplatform.com/2019/08/07/amd-doubles-down-and-up-with-rome-epyc-server-chips/



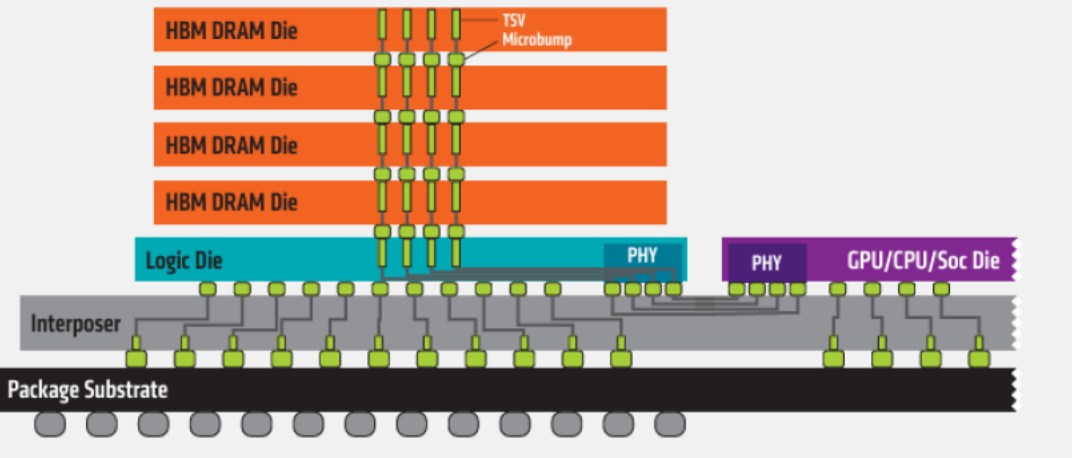
Timothy Prickett Morgan, *AMD Doubles Down – And Up – With Rome Epyc Server Chips*, The Next Platform (Aug. 7, 2019), <https://www.nextplatform.com/2019/08/07/amd-doubles-down-and-up-with-rome-epyc-server-chips/> (“The basic architecture of the Rome multichip module (MCM) was unveiled last November. Rather than having I/O and memory controllers on each chiplet and using PCI-Express ports to create the Infinity Fabric interconnect lashing the dies on the MCM together as was the case with the Naples processor, this time around AMD created a central I/O and memory controller die, etched in the same class of 14 nanometer processes at GlobalFoundries as was the Naples chip, and then hung chiplets containing just core complexes (what AMD calls CCXs). The number of chiplets for these compute cores was not divulged last November, but we now know that each core complex has eight Zen 2 cores and there are eight of them on the Rome package for a total of 64 cores.”).

Timothy Prickett Morgan, *AMD Doubles Down – And Up – With Rome Epyc Server Chips*, The Next Platform (Aug. 7, 2019), <https://www.nextplatform.com/2019/08/07/amd-doubles-down-and-up-with-rome-epyc-server-chips/> (“The block diagram further above that shows the complexes seems to imply that there were only four memory controllers on the Naples and Rome chips, but each DDR block is actually a pair of controllers, for a total of eight controllers. The Rome design has 128 lanes of PCI-Express 4.0 I/O that can reach out to

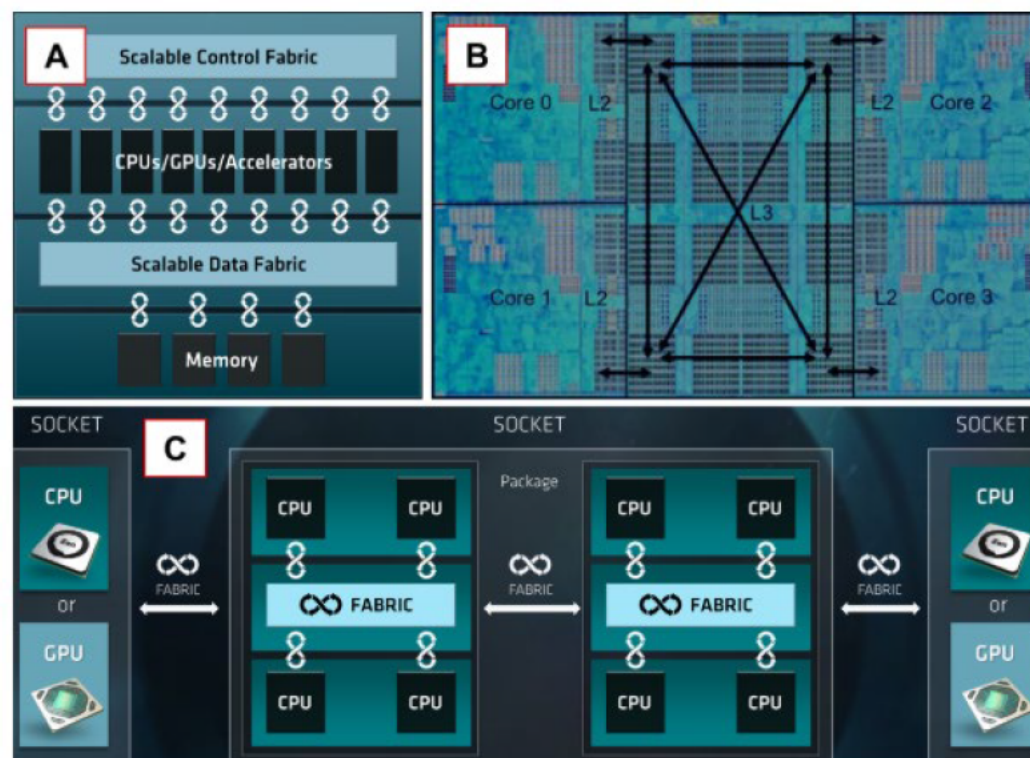
U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	peripheral devices and be used as the NUMA interconnect in a two-socket machine; Naples used PCI-Express 3.0 controllers – 128 lanes of them as well – which ran at half the speed and therefore had half the bandwidth. The NUMA coupling with Naples and Rome eats up 64 of the 128 lanes, so a two-socket box with two processors still only has 128 PCI-Express lanes exposed to the outside world. (Technically, there are two extra lanes that are used for Infinity Fabric control. Count these if you feel like it.) Because all of the chiplets hang off the same I/O and memory hub, which has a fast transport of its own on the die, there is one fewer NUMA hop to get from one chiplet to another, greatly simplifying the NUMA structure of the socket and reducing the overhead of sharing memory across the chiplets in a socket and across sockets.”). NVIDITC-001015651: AMD Radeon R9 Fury X Detailed, Geeks3D (June 18, 2015)  The diagram illustrates the AMD Radeon R9 Fury X architecture. It features a stack of four HBM DRAM Dies on top of a Logic Die, which is mounted on a Package Substrate. The Logic Die is connected to a GPU. The diagram also shows an Interposer and a Package Substrate. The AMD logo is in the top right corner. "Fiji" Chip HIGH-BANDWIDTH MEMORY ▲ World's first GPU to feature advanced HBM technology ▲ Unprecedented high-bandwidth memory – 60% higher memory bandwidth than Radeon™ R9 290X¹⁰ ▲ Engineered to meet and exceed the performance demands of 4K and VR gaming ▲ Benefits – 4X Bandwidth per watt improvement from Radeon™ R9 290X¹⁰ – Increased bandwidth – 4096-bit memory interface – Enables Small Form Factors 36 RADEON GRAPHICS UPDATE JUNE 2015 CONTENT UNDER EMBARGO UNTIL JUNE 18, 2015 8 A.M. ET

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	NVIDITC-001015651: AMD Radeon R9 Fury X Detailed, Geeks3D (June 18, 2015) ("The Radeon R9 Fury X is based on a new GPU codenamed Fiji (XT version for the Fury X). The R9 Fury X is the first graphics card with High-Bandwidth Memory or HBM technology. HBM will replace GDDR5 memory in the near future."). NVIDITC-001015605: <i>AMD Radeon (TM) R9 Nano, World's Smallest and Most Power-Efficient Enthusiast Graphics Card, Brings 4K Gaming to the Living Room</i>, AMD (Aug. 27, 2015) ("AMD (NASDAQ: AMD) continues to push the boundaries of graphics card design, today announcing its category-creating AMD Radeon R9 Nano, the fastest Mini ITX graphics card ever to enable 4K gaming in the living room through ultra-quiet, ultra-compact PC designs. First previewed to gamers around the world during the Electronic Entertainment Expo (E3) in Los Angeles in June 2015, the AMD Radeon R9 Nano graphics card is based on the graphics chip codenamed "Fiji," and is the third "Fiji"-based product to launch this summer alongside the AMD Radeon™ R9 Fury and R9 Fury X graphics cards. The AMD Radeon™ R9 Fury graphics family, based on the "Fiji" chip, marks a turning point in PC gaming with the implementation of High-Bandwidth Memory (HBM) to deliver extreme energy efficiency and performance for ultra-high resolutions, unparalleled VR experiences, smoother gameplay, with the Radeon R9 Nano revolutionizing form-factors for enthusiasts everywhere."). This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
1[D]: wherein the passive crosslink is dedicated for inter-chiplet communications.	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See evidence cited for '803 patent limitations 1[A]-1[C].</i>

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
2. The system of claim 1, wherein the passive crosslink comprises a passive interposer die.	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See evidence cited for '803 patent limitation 1[D].</i> This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
3[A]: The system of claim 1, wherein the first GPU chiplet comprises a first PHY region including conductor structures for chiplet-to-chiplet communications,	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> AMD, <i>High Bandwidth Memory (HBM) Whitepaper</i>, AMD.com (archived Nov. 24, 2018), https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	 <i>See also</i> evidence cited for '803 patent claims 1-2. This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
3[B]: and further wherein the second GPU chiplet comprises a second PHY region including conductor structures for chiplet-to-chiplet communications.	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See</i> evidence cited for '803 patent limitation 3[A]. This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
7. The system of claim 1, wherein the passive crosslink communicably couples all GPU chiplets in the GPU chiplet array.	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See evidence cited for '803 patent limitation 1[D].</i> This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
8[A]: The system of claim 1, further comprising: a first cache memory hierarchy at the first GPU chiplet,	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> Paul Teich, <i>The Heart Of AMD's Epyc Comeback Is Infinity Fabric</i>, The Next Platform (July 12, 2017), https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/



Source: AMD

AMD Infinity Fabric high-level block diagram and detail

Paul Teich, *The Heart Of AMD's Epyc Comeback Is Infinity Fabric*, The Next Platform (July 12, 2017), <https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/> ("Infinity Fabric's System Data Fabric (SDF) connects the L3 caches to memory and to the configurable I/O lanes (described below). SDF uses the configurable I/O lanes for memory coherent communications between compute elements on a single die, between different dies mounted on an MCM, and between sockets in a NUMA server design. Using Infinity Fabric,

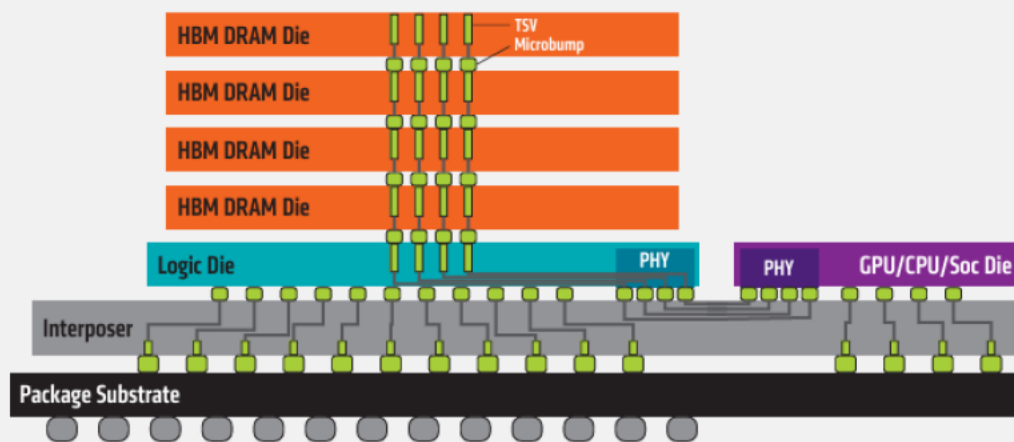
U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	AMD can connect multiple dies on a printed circuit board and enable the MCM to act as a single monolithic compute node.”). Paul Teich, <i>The Heart Of AMD's Epyc Comeback Is Infinity Fabric</i>, The Next Platform (July 12, 2017), https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/ (“The figure above shows how AMD describes Infinity Fabric scaling: SCF connects compute elements SDF enables coherent communications among caches and between caches and memory. There are four fully interconnected core complexes on a Zeppelin die (see below for much more detail on Zeppelin). Each Infinity Fabric on-die link is four-bit point to point. Both SCF and SDF scale between die on an MCM and between sockets.”) Paul Teich, <i>The Heart Of AMD's Epyc Comeback Is Infinity Fabric</i>, The Next Platform (July 12, 2017), https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/ (“AMD redesigned HyperTransport as a software-defined interconnect fabric, and as such Infinity Fabric has separate and scalable control and data paths. In software defined networking (SDN), these paths are referred to as planes – control planes and data planes, to be specific – and the architectural implementation is similar.”). Paul Teich, <i>The Heart Of AMD's Epyc Comeback Is Infinity Fabric</i>, The Next Platform (July 12, 2017), https://www.nextplatform.com/2017/07/12/heart-amds-epyc-comeback-infinity-fabric/ (“The figure above shows another rendition of a fully connected MCM highlighting that Infinity Fabric directly connects a die on one MCM to its matching die on the other MCM. This creates a sparsely connected hypercube where memory attached to one die is at most only two “hops” (intermediary memory controllers attach to the SDF, the memory controllers are in different die) away from any part of system memory. This interconnect topology is analogous to the eight-socket Opteron servers built with HyperTransport a dozen years ago as we pointed out in our original coverage of Epyc back in June, but now it is collapsed into a higher-performance two-socket design.”) Timothy Prickett Morgan, <i>AMD's Long Road from Naples to Milan Centers on Rome</i>, The Next Platform (Nov. 6, 2018), https://www.nextplatform.com/2018/11/06/amds-long-road-from-naples-to-milan-centers-on-rome/ (“The way the architecture works is that PCI-Express

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	4.0 gives Infinity Fabric twice the bandwidth, and the change in topology by moving to a central I/O and memory hub eliminates one layer of NUMA hopping, which in turn improves the latency of the in-socket interconnect. To put it simply, more code will run better on Rome than did on Naples, and that is because of the underlying hardware architecture but also because of the work that has been done with the Linux and Windows Server schedulers to be aware of the Naples and Rome chip NUMA topologies.”). Timothy Prickett Morgan, <i>AMD Doubles Down – And Up – With Rome Epyc Server Chips</i>, The Next Platform (Aug. 7, 2019), https://www.nextplatform.com/2019/08/07/amd-doubles-down-and-up-with-rome-epyc-server-chips/  The diagram illustrates the AMD Epyc server chip architecture. It features a central green box representing the Infinity Fabric, which connects to various components. On the left, under 'COMPUTE', it lists: Up to 2X AMD "Zen" x86 cores (up to 64 cores/128 threads), Up to 4X shared L3 cache (256MB) (Up to 2X L3 cache per core (16MB per 4 cores)), Reduced System Diameter (NUMA domain), and TDP range: 120W-225W. Below this, under 'MEMORY', it lists: 8 channel DDR4 with ECC up to 3200 MHz, RDIMM, LRDIMM, 3DS, NVDIMM, and 2 DIMMs/channel capacity of 4TB/socket*. On the right, under 'PERFORMANCE', it lists: +4x Peak TFLOPS/Socket and +2X increased perf/socket. Below this, under 'INTEGRATED I/O – NO CHIPSET', it lists: 128 lanes PCIe® Gen3 & Gen4 **, Used for PCIe, SATA, and Coherent Interconnect, and Up to 32 SATA or NVMe devices. At the bottom right, under 'SECURITY', it lists: Dedicated Security Subsystem, Hardware Root-of-Trust, and Additional Security Features. The central green box is labeled with components: DBR4 Memory Controllers, PCIe/SATA, Server Controller Hub, and AMD Secure Processor. Surrounding this central box are eight blue boxes, each labeled '8 Cores + L3'. NVIDITC-001016951: AMD Newsroom, Vega: AMD's New Graphics Architecture for Virtually Unlimited Workloads, AMD (Jan. 5, 2017) (“The world's most advanced GPU memory architecture: The Vega architecture enables a new memory hierarchy for GPUs. This radical new approach comes in the form of a new high-bandwidth cache and its controller. The cache features leading-edge HBM2 technology which is capable of transferring terabytes of data every second, doubling the bandwidth-per-pin over the previous generation HBM technology.

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	HBM2also enables much greater capacity at less than half the footprint of GDDR5 memory. Vega architecture is optimized for streaming very large datasets and can work with a variety of memory types with up to 512TB of virtual address space.”). <i>See also</i> evidence cited for '803 patent claim 1[C]. This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
8[B]: wherein a first level of the first cache memory hierarchy is coherent within the first GPU chiplet; and	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> NVIDITC-001015664: Usman Pirzada, <i>AMD's Infinity Fabric Detailed –The Innovative, Real-World Implementation of The Company's 'Perfect Lego' Philosophy</i>, WCCFTech (Jan. 16, 2017) (“Infinity Fabric is a coherent implementation which means that cache coherency is maintained across multiple processors externally and scaling up cores, in a CPU or a GPU, is not a problem and only limited by the bandwidth of the transport itself (which we have mentioned above). This philosophy is in contrast to Intel's stricter vision of a tailor-made design. It also allows AMD to scale up and down designs within a matter of hours rather than months without spending additional human capital or resources. This flexibility will allow it to serve a larger number of custom clients than anyone else.”). <i>See also</i> evidence cited for '803 patent limitation 8[A]. This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
8[C]: a second cache memory hierarchy at the second GPU chiplet,	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See</i> evidence cited for '803 patent limitation 8[A]. This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
8[D]: wherein a first level of the second cache memory hierarchy is coherent within the second GPU chiplet.	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See</i> evidence cited for '803 patent limitation 8[B]. This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
9[A]: The system of claim 8, further comprising: a unified cache memory including both a last level of the first cache memory hierarchy and a last level of the second cache memory hierarchy,	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See</i> evidence cited for '803 patent claim 8.

U.S. Patent No. 11,841,803	AMD's Implementation of HBM and Infinity Fabric ("AMD Product Art")
	This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
9[B]: wherein the unified cache memory is coherent across all GPU chiplets of the GPU chiplet array.	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> <i>See evidence cited for '803 patent claim 8.</i> This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.
10. The system of claim 1, further comprising: a plurality of conductive pillars coupling a circuit board substrate to a first non-PHY region of the first GPU chiplet and a second non-PHY region of the second GPU chiplet.	Including at least under Complainant's apparent infringement theory, AMD Product Art discloses and/or renders obvious this limitation. <i>See, e.g.,</i> AMD, <i>High Bandwidth Memory (HBM) Whitepaper</i>, AMD.com (archived Nov. 24, 2018), https://web.archive.org/web/20181124220319/https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf



See also evidence cited for '803 patent claim 1.

This limitation also would have been obvious to a person of ordinary skill in the art based on (1) AMD Product Art alone; (2) AMD Product Art in combination with the general knowledge of a person of ordinary skill in the art; and/or (3) AMD Product Art in combination with the teachings with respect to this claim element as detailed in NVIDIA's other invalidity charts and cover pleading.