

**IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF TEXAS
MIDLAND/ODESSA DIVISION**

ADVANCED CLUSTER SYSTEMS, INC.,

Plaintiff,

v.

ADVANCED MICRO DEVICES, Inc.,

Defendant.

Civil Action No. 7:24-cv-00244-DC-DTG

**PLAINTIFF ADVANCED CLUSTER SYSTEMS, INC.'S
PRELIMINARY INFRINGEMENT CONTENTIONS**

Plaintiff Advanced Cluster Systems, Inc. (“ACS”) hereby serves on Defendant Advanced Micro Device, Inc.(“AMD”) the following Preliminary Infringement Contentions.¹

The following disclosures are based on information currently available to ACS. Fact and expert discovery in this action has not begun, and ACS reserves the right to modify, amend, or otherwise supplement these disclosures as discovery reveals new documents and information. ACS further reserves the right to modify, amend, or otherwise supplement these disclosures should Defendant alter, supplement, or otherwise clarify documents and information currently available to ACS. ACS notes that Defendants have not yet provided any versions, subroutines and APIs for the source code for the Accused Products. Defendants have not yet produced any relevant and responsive technical documents that show the operation of the Accused Products and have not

¹ ACS anticipates that its contentions will require supplementation because AMD: (1) has not produced any versions, subroutines or APIs for the source code for the Accused Products; (2) has not provided any relevant and responsive technical documents for the Accused Products; and (3) has not searched for or produce documents responsive to ACS’s forthcoming Requests for Production.

produced any documents responsive to ACS's forthcoming Requests for Production. Such responsive documents are relevant to several of the required disclosures below.

I. DISCLOSURES

A. ASSERTED CLAIMS

ACS identifies the following asserted claims (the "Asserted Claims"):

U.S. Patent No.	Asserted Claims
10,333,768	1, 4, 20, 21, 26, 29, 30, 31, 33, 35, 37, 39
11,563,621	1, 3, 4, 6, 7, 8, 9, 11, 13, 15, 16, 17, 21, 23, 24, 26, 27, 28, 29, 30
11,570,034	1, 2, 3, 8, 10, 24, 25, 27, 28, 30
11,811,582	1, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20
12,021,679	1, 2, 3, 8, 10, 17, 18

B. ACCUSED PRODUCTS AND METHODS

Based on presently available information, ACS contends that AMD infringes the Asserted Claims based on the following instrumentalities (the "Accused Products"). Specifically, ACS asserts that the Accused Products meet the limitations of the asserted apparatus claims identified below.

Asserted Patent	Asserted Claims	Accused Product(s)
10,333,768	1, 4, 20, 21, 26 29, 30, 31, 33, 35, 37, 39	All server and workstation products that utilize AMD's accelerator platforms, including at least AMD's MI300X and MI325X Platforms and all products including or related AMD INSTINCT™ Accelerator Qualified Servers https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf incorporating any of the Accused AMD Accelerator Products, as well as any products incorporating those items, and any

Asserted Patent	Asserted Claims	Accused Product(s)
		products that are substantially similar to those items (the “Accused AMD Server Products”). Also, for example, each of AMD’s MI210, MI250, and MI250X Accelerators (“MI200 accelerators”) is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server (<i>e.g.</i>, “MI200 Platform”) such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. All AI accelerator products that utilize AMD’s AI accelerator architectures and are especially adapted for use in server and workstation products, including at least the MI300X, MI325X, MI210, MI250, and MI250X AI accelerator products, all variants thereof, and any products that are substantially similar to those products (the “Accused AMD AI Accelerator Products”) <i>See</i> Exhibit A1 hereto. Each of the Accused AMD AI Accelerator Products and each of the Accused AMD Server Products is configured and operates in a substantially similar manner in relevant part.
11,563,621	1, 3, 4, 6, 7, 8, 9, 11, 13, 15, 16, 17, 21, 23, 24, 26, 27, 28, 29, 30	All server and workstation products that utilize AMD’s accelerator platforms, including at least AMD’s MI300X and MI325X Platforms and all products including or related AMD INSTINCT™ Accelerator Qualified Servers https://www.amd.com/system/files/documents/amd-

Asserted Patent	Asserted Claims	Accused Product(s)
		instinct-server-soln-catalog.pdf) incorporating any of the Accused AMD Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items (the “Accused AMD Server Products”). Also, for example, each of AMD’s MI210, MI250, and MI250X Accelerators (“MI200 accelerators”) is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server (<i>e.g.</i>, “MI200 Platform”) such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. All AI accelerator products that utilize AMD’s AI accelerator architectures and are especially adapted for use in server and workstation products, including at least the MI300X, MI325X, MI210, MI250, and MI250X AI accelerator products, all variants thereof, and any products that are substantially similar to those products (the “Accused AMD AI Accelerator Products”) <i>See</i> Exhibit A2 hereto. Each of the Accused AMD AI Accelerator Products and each of the Accused AMD Server Products is configured and operates in a substantially similar manner in relevant part.
11,570,034	1, 2, 3, 8, 10, 24, 25, 27, 28, 30	All server and workstation products that utilize AMD’s accelerator platforms, including at least AMD’s MI300X and MI325X Platforms and all

Asserted Patent	Asserted Claims	Accused Product(s)
		products including or related AMD INSTINCT™ Accelerator Qualified Servers (https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf) incorporating any of the Accused AMD Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items (the “Accused AMD Server Products”). Also, for example, each of AMD’s MI210, MI250, and MI250X Accelerators (“MI200 accelerators”) is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server (<i>e.g.</i>, “MI200 Platform”) such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. All AI accelerator products that utilize AMD’s AI accelerator architectures and are especially adapted for use in server and workstation products, including at least the MI300X, MI325X, MI210, MI250, and MI250X AI accelerator products, all variants thereof, and any products that are substantially similar to those products (the “Accused AMD AI Accelerator Products”) <i>See</i> Exhibit A3 hereto. Each of the Accused AMD AI Accelerator Products and each of the Accused AMD Server Products is configured and operates in a substantially similar manner in relevant part.

Asserted Patent	Asserted Claims	Accused Product(s)
11,811,582	1, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20	All server and workstation products that utilize AMD's accelerator platforms, including at least AMD's MI300X and MI325X Platforms and all products including or related AMD INSTINCT™ Accelerator Qualified Servers (https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf) incorporating any of the Accused AMD Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items (the "Accused AMD Server Products"). Also, for example, each of AMD's MI210, MI250, and MI250X Accelerators ("MI200 accelerators") is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server (<i>e.g.</i>, "MI200 Platform") such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. All AI accelerator products that utilize AMD's AI accelerator architectures and are especially adapted for use in server and workstation products, including at least the MI300X, MI325X, MI210, MI250, and MI250X AI accelerator products, all variants thereof, and any products that are substantially similar to those products (the "Accused AMD AI Accelerator Products")

Asserted Patent	Asserted Claims	Accused Product(s)
		See Exhibit A4 hereto. Each of the Accused AMD AI Accelerator Products and each of the Accused AMD Server Products is configured and operates in a substantially similar manner in relevant part.
12,021,679	1, 2, 3, 8, 10, 17, 18	All server and workstation products that utilize AMD's accelerator platforms, including at least AMD's MI300X and MI325X Platforms and all products including or related AMD INSTINCT™ Accelerator Qualified Servers https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf) incorporating any of the Accused AMD Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items (the "Accused AMD Server Products"). Also, for example, each of AMD's MI210, MI250, and MI250X Accelerators ("MI200 accelerators") is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server (e.g., "MI200 Platform") such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. All AI accelerator products that utilize AMD's AI accelerator architectures and are especially adapted for use in server and workstation products, including at least the MI300X, MI325X, MI210, MI250, and MI250X AI accelerator products, all variants thereof, and any products that are substantially

Asserted Patent	Asserted Claims	Accused Product(s)
		similar to those products (the “Accused AMD AI Accelerator Products”) <i>See Exhibit A5 hereto. Each of the Accused AMD AI Accelerator Products and each of the Accused AMD Server Products is configured and operates in a substantially similar manner in relevant part.</i>

ACS does not intend for the product names or model numbers in this chart of Accused Products, or in the appended infringement charts in Exhibits A1-A5, to be a conclusive list. By identifying product names and model numbers in this chart of Accused Products, and in the appended infringement charts in Exhibits A1-A5, ACS intends to include all model numbers and configurations for each identified Accused Product. ACS has identified additional model numbers where available based on currently available information. ACS reserves the right to supplement this table of Accused Products as additional information becomes available through discovery in this action.

C. INFRINGEMENT CHARTS

For the Asserted Patent, ACS attaches a claim chart (*see* Exhibits A1-A5) to these Preliminary Infringement Contentions identifying specifically where and how each limitation of each Asserted Claim is found within each Accused Product. The attached claims are exemplary.

ACS has not completed its investigation of the facts relating to this case, has not completed discovery in this action, and has not completed preparation for trial. Therefore, ACS expressly reserves the right to amend, supplement, or otherwise modify the attached charts as additional information becomes available through discovery and/or as necessitated by any claim construction ruling from the Court.

D. INDIRECT INFRINGEMENT

In addition to directly infringing the Asserted Patents by, for example, making, using, selling, offering for sale, and/or importing into the United States the Accused Products, AMD has indirectly infringed the Asserted Patents by actively inducing direct infringement of the Asserted Patents by others.

Defendant actively induces others to infringe the Asserted Patents by marketing and selling the Accused Products, knowing, no later than the date ACS served the original complaint in this case on September 26, 2024, and intending that the Accused Products would be used by customers, users, system builders, partners, and retailers/distributors in a manner that infringes the Asserted Patent.

ACS is informed and believes, thereon alleges that AMD makes, uses, sells, offers for sales and/or imports one or more of the Accused Products for or on behalf of third parties such as customers, users, system builders, partners, and retailers/distributors, knowing and intending that the Accused Products will be used by the third parties in a manner that practices the Asserted Claims as shown in Exhibits A1-A5. For example, AMD published and provided marketing materials, technical specifications, whitepapers, datasheets, user manuals, and development and testing information, and other resources on its website² that instructed and encouraged third parties to integrate the Accused Products into products using AMD's technology that were then made, used, sold, offered for sale and/or imported into the United States.³ AMD also provides a website to inform customers with the servers and workstations that support the Accused AMD Accelerator

² See, e.g., <http://amd.com/>, <https://www.amd.com/en/products/accelerators/instinct.html>, and pages linked thereon.

³ E.g., <https://www.amd.com/en/where-to-buy/accelerators/instinct.html#partner> and <https://www.amd.com/content/dam/amd/en/documents/products/accelerators/instinct/amd-instinct-solution-catalog.pdf>.

Products,⁴ case studies of customer success stories,⁵ and supporting documentation for the Instinct™ Accelerators.⁶ These activities were designed to bring infringing products that incorporate AMD's Accused Products to market in the United States. As a result of AMD's activities, the Accused Products have been used in a manner that directly infringes the Asserted Patents. AMD continues to engage in acts of inducement of infringement of the Asserted Patents.

Defendant induces others to directly infringe in numerous ways. As a non-exhaustive exemplary list, a direct infringer may make Server Products by combining Accused AMD Accelerator Products into an infringing system; use a Server Product owned by the direct infringer, use a Server Product owned by a third party, use a Server Product owned by Defendant; sell and/or offer to sell Server Products made by the direct infringer, sell and/or offer to sell Server Products acquired from another party, including possibly from Defendant; and/or import Server Products into the United States.

Based on currently available information, Defendant provides marketing materials, technical specifications, whitepapers, datasheets, user manuals, development and testing information, and other resources that instruct and encourage customers, users, system builders, and/or retailers/distributors to make, use, sell, offer for sale, and import the server products in an infringing manner in the United States. Defendant instructs and encourage this infringement by, for example, with marketing, training, professional services, sales and distributions, and service

⁴ E.g., https://www.amd.com/en/where-to-buy/accelerators/instinct.html#sortCriteria=%40amd_partner%20ascending.

⁵ E.g., https://www.amd.com/en/resources/case-studies.html#sortCriteria=%40amd_release_date%20descending&f-amd_form_factor=Servers&f-amd_product_type=Accelerators.

⁶ E.g., https://www.amd.com/en/search/documentation/hub.html#sortCriteria=%40amd_release_date%20descending&f-amd_product_type=Accelerators&f-amd_product_brand=Instinct.

and support for the server products. The server products embody or practice the Asserted Claims as set forth in Exhibits A1-A5.

Additionally, Defendant provides marketing materials, technical specifications, whitepapers, datasheets, user manuals, development and testing information, and other resources that instruct and encourage customers, users, system builders, and retailers/distributors to use the Accused Products in an infringing manner in the United States. For example, AMD provides instructions for integrating the Accused Products into other products that infringe the Asserted Patents. These third parties then use the materials and support provided by Defendant to make, use, sell, offer for sale, and/or import into the United States infringing products that use the Accused Products therein. Defendant instructs and encourages this by, for example, providing marketing, training, professional services, sales and distributions, and service and support.

AMD also encourages others to provide, or provides itself, use of Accused Products in a compute as a service model. Their compute as a service model encourages use of an Accused Product by a party different from the owner of the Accused Product. This model encourages many infringing users to share access to and use of the Accused Products. AMD also offers support for Accused Products allowing, and inducing, infringing use to continue.

In view of at least Defendant's knowledge of the Asserted Patents, Defendant knew or should have known that such activities would cause direct infringement of the Asserted Patents by Defendant's customers, end users, system builders, partners, and retailers/distributors of the Accused Products. Defendant's inducement constitutes infringement of the Asserted Patents in violation of 35 U.S.C. § 271(b).

ACS has not completed its investigation of the facts relating to this case, has not completed discovery in this action, and has not completed preparation for trial. Therefore, ACS expressly

reserves the right to amend, supplement, or otherwise modify the attached charts as additional information becomes available through discovery.

E. LITERAL INFRINGEMENT AND INFRINGEMENT UNDER THE DOCTRINE OF EQUIVALENTS

Based on currently available information, ACS alleges that each of the Accused Products literally infringe each of the Asserted Claims. *See supra* §§ I.B, I.C & I.D, and Exhibits A1-A5 hereto.

To the extent any claim limitations are not literally infringed, ACS contends the limitations are infringed under the doctrine of equivalents because the identified features of the Accused Products are insubstantially different from the literal claim elements, for example because they perform substantially the same function in substantially the same way to achieve substantially the same result. ACS has not completed its investigation of the facts relating to this case and has not yet completed discovery in this action. Furthermore, the Asserted Claims have not yet been construed by the Court, nor has Defendant addressed claim construction. Defendant also has not yet set forth its bases, if any, for non-infringement of the Asserted Claims. Accordingly, ACS reserves the right to amend, supplement, or otherwise modify these infringement contentions as additional information becomes available through the claim construction process and discovery, including the right to further articulate its contentions with regards to infringement under the doctrine of equivalents.

F. PRIORITY DATE OF ASSERTED CLAIMS

ACS contends that the Asserted Claims of the Asserted Patents are entitled to a priority date no later than the filing of provisional application No. 60/813,738, which was filed on June 13, 2006, and of provisional application No. 60/850,908, which was filed on October 11, 2006. The

Asserted Patents claim priority to both provisional applications on their faces as related U.S. Applications.

II. DOCUMENT PRODUCTION ACCOMPANYING DISCLOSURE

Concurrently with the service of its foregoing disclosures and contentions, ACS will be producing via a FTP link a copy of the file history of each Asserted Patent and documents evidencing conception and reduction to practice for each claimed invention (ACS_0000167 – ACS_0076737).

These disclosures, contentions, and document production are based on information currently available to ACS. Fact and expert discovery in this action has not started, and ACS reserves the right to modify, amend, or otherwise supplement its disclosures, contentions, and document production as new documents and information are revealed through discovery and/or to the extent Defendant alters, supplements, or otherwise clarifies documents and information currently available to ACS.

Respectfully submitted,

By: /s/ David Lindner

CROWELL & MORING LLP

Jon Gurka (admitted *pro hac vice*)
jgurka@crowell.com
Kainoa Asuego (admitted *pro hac vice*)
kasuega@crowell.com
3 Park Plaza
20th Floor
Irvine, CA 92614
Phone: (949) 798-1331

David Lindner (admitted *pro hac vice*)
dlindner@crowell.com
Andrew McElligott (admitted *pro hac vice*)
amcelligott@crowell.com
455 N Cityfront Plaza Drive, Ste 3600
Chicago, IL 60611
Telephone: (312) 321-4200

Michelle Wang
michellewang@crowell.com
Two Manhattan West
375 Ninth Avenue
New York, NY 10001
Telephone: (212) 233-4000

BARCELÓ, HARRISON & WALKER LLP

Reynaldo C. Barceló (admitted *pro hac vice*)
rey@bhiplaw.com
2901 West Coast Hwy, Suite 200
Newport Beach, CA 92663
Phone: (949) 340-9736

CHERRY JOHNSON SIEGMUND JAMES, PLLC

Mark D. Siegmund (TX Bar No 24117055)
msiegmund@cjsjlaw.com
William D. Ellerman (TX Bar No 24007151)
wellerman@cjsjlaw.com
7901 Fish Pond Rd., 2nd Floor

Waco, TX 76710
Telephone: (254) 732-2242

*Attorneys for Plaintiffs
Advanced Cluster Systems, Inc.*

CERTIFICATE OF SERVICE

I hereby certify that on February 24, 2025, I served the foregoing document (including the accompanying claim chart exhibits A1-5) via e-mail on Defendant's counsel of record at beferguson@winston.com; ccliu@winston.com; tmelsheimer@winston.com; and rmann@winston.com.

Dated: February 24, 2025

/s/ Andrew Greenfeld
Andrew Greenfeld

Exhibit A1
U.S. Pat. No. 10,333,768 in view of
Accused AMD Server Products

U.S. Pat. No. 10,333,768	Accused AMD Server Products
1. A computer cluster comprising:	Each of the Accused AMD Server Products ¹ comprises a computer cluster. For example, as depicted below and as described further herein, AMD's INSTINCT™ MI300X Platform ("MI300X Platform") includes a computer cluster. ²

¹ The Accused AMD Server Products include, but are not limited to, all products including or related to AMD's MI300X and MI325X Platforms (<https://www.amd.com/en/products/accelerators/instinct/mi300/platform.html>), and all products including or related to AMD INSTINCT™ Accelerator Qualified Servers (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>) incorporating any of the Accused Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items.

² "AMD INSTINCT™ MI300X Platform™ Data Sheet" (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); *see also* MI300 introductory video ("Introducing AMD Instinct™ MI300 Series Accelerators"), at <https://www.youtube.com/watch?v=uVdNpk8tvBI>; *see also* "AMD INSTINCT™ MI325X Platform™ Data Sheet" (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

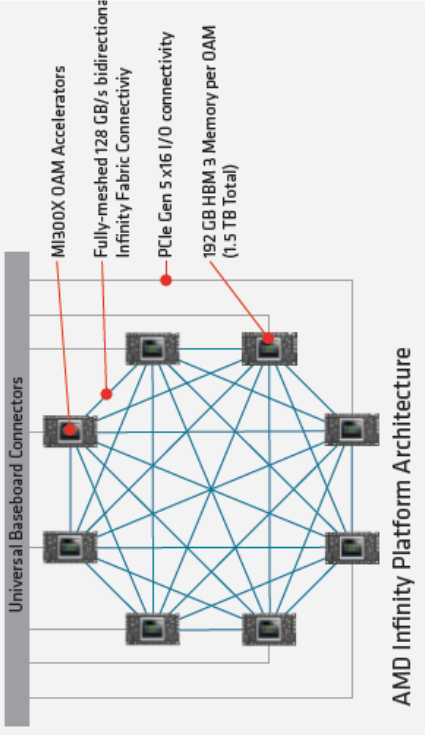
Also, for example, as described further herein, each of the MI210, MI250, and MI250X Accelerators ("MI200 accelerators") is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server ("MI200 Platform") such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. *See e.g.*, "AMD Instinct Accelerator-Powered Servers" (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>); "AMD Instinct MI250X Accelerators" (<https://www.amd.com/en/products/accelerators/instinct/mi250x.html>); "AMD Instinct MI250 Accelerators" (<https://www.amd.com/en/products/accelerators/instinct/mi250.html>); "AMD Instinct MI210 Accelerators" (<https://www.amd.com/en/products/accelerators/instinct/mi210.html>); "AMD CDNA™ 2 Architecture," pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (Infinity Fabric connecting MI200 accelerators both in-package and inter-package). *See also*, "AMD INSTINCT™ MI200 Series Accelerator" (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>); "AMD INSTINCT™ MI210 Accelerator" (<https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/product-briefs/instinct-mi210-brochure.pdf>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 Each of the Accused AMD Server Products includes a plurality of nodes, each of which comprises an Accused AMD Accelerator Product.³ For example, as depicted below, the AMD MI300X Platform comprises eight MI300X accelerators “on an industry-standard universal baseboard” that “are connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”⁴

³ Accused AMD Accelerator Products include, but are not limited to, all products including or relating to AMD’s MI300 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi300.html>) (MI300X, MI325X, MI300A), and MI200 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi200.html>) (MI250X, MI210), as well as any products incorporating those items, and any products that are substantially similar to those items.

⁴ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); *see also* “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, AMD specifies a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. *See* “AMD CDNA™ 2 Architecture,” p. 7 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>). *See also, e.g., id.*, pp. 6 and 8-9 (additional topologies); pp. 3-4 (Infinity Fabric connecting MI200 accelerators both in-package and inter-package).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
a plurality of nodes, wherein each of the plurality of nodes comprises a hardware processor, wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, and	 Each of the Accused AMD Server Products comprises a plurality of nodes, each of which comprises an Accused AMD Accelerator Product. For example, as depicted below, the MI300X Platform comprises eight MI300X accelerators (e.g., nodes) that “are connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”⁵

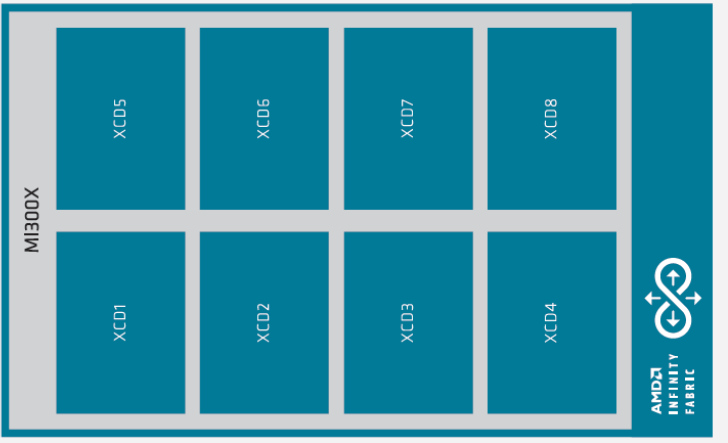
⁵ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, the MI200 series accelerators are designed to be integrated into at least one MI200 Platform that incorporates a plurality of MI200 accelerators according to a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. See e.g., “AMD Instinct Accelerator-Powered Servers” (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>); “AMD CDNA™ 2 Architecture,” p. 7 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>); pp. 3-4 (Infinity Fabric connecting MI200 accelerators both in package and inter-package). See also, e.g., *id.*, pp. 6 and 8-9 (additional topologies).

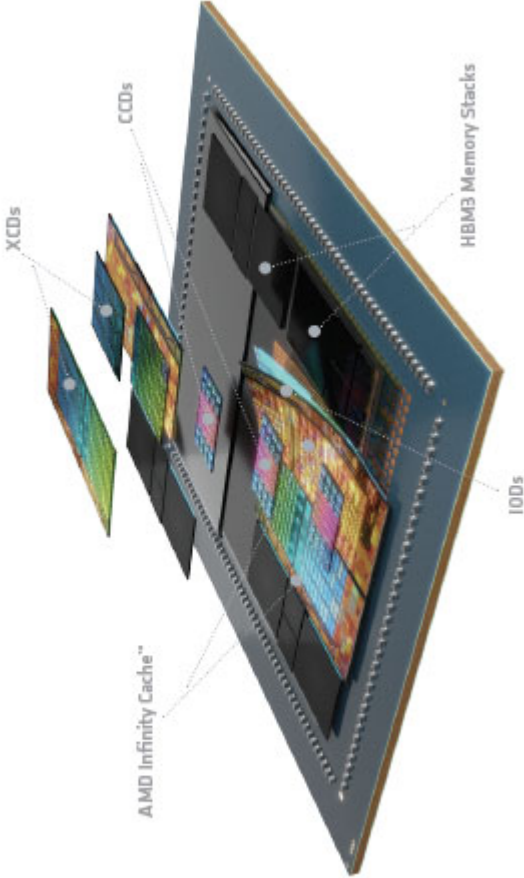
U.S. Pat. No. 10,333,768 wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing the hardware processor to evaluate mathematical expressions; and	Accused AMD Server Products The diagram illustrates the AMD Infinity Platform Architecture. It features a central mesh of nodes, each representing an MI300X OAM Accelerator. These nodes are interconnected via a 'Fully-meshed 128 GB/s bidirectional Infinity Fabric Connectivity'. The architecture also includes 'PCIe Gen 5x16 I/O connectivity' and '192 GB HBM3 Memory per OAM (1.5 TB Total)'. The entire system is connected to a 'Universal Baseboard Connectors' interface. The caption below the diagram reads 'AMD Infinity Platform Architecture'. Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises a hardware processor, wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, and wherein each of the nodes is configured to access a non-transitory computer readable medium comprising program code for a single-node kernel that, when executed, is capable of causing the hardware processor to evaluate mathematical expressions. Specifically, each Accused AMD Accelerator Product comprises a hardware processor that comprises multiple processor cores. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).⁶ The architecture of the MI300X accelerator is depicted in the following figure:⁷
--	---

⁶ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

⁷ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 Each Accused AMD Accelerator Product comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster of each Accused AMD Service Product. For example, each MI300X accelerator in each MI300X Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes the eight MI300X accelerators in the MI300X Platform⁸. This is done,

⁸ Also, for example, each MI200 accelerator in an MI200 Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes at least three MI200 accelerators.

U.S. Pat. No. 10,333,768	Accused AMD Server Products for example, by using AMD's ROCm™ Collective Communication Library (RCCL), APIs, and math libraries.⁹ Each AMD Accelerator Product is configured to access a non-transitory computer readable medium (e.g., memory accessible by a XCD, including local memory, shared memory, and HBM memory, as well as memory external to the XCD).¹⁰ Figure 1. Advanced 3D Package and chiplet-based construction of the AMD MI300 Series processors 
--------------------------	--

⁹ “RCCL documentation” (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) (“RCCL (pronounced ‘Rickel’) is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-all, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations.... [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform.”). RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.* RCCL also includes commands that cause similar operation of a MI200 accelerator in an MI200 Platform. *Id.*

¹⁰ “AMD CDNA™ 3 Architecture,” p. 3 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/white-papers/amd-cdna-3-white-paper.pdf>); *Id.* at p. 2 (“The MI300 Series integrates up to 8 vertically stacked accelerator complex dies (XCD) and 4 I/O dies (IOD) containing system infrastructure, all tied together with the AMD Infinity Fabric™ technology and connecting to 8 stacks of high-bandwidth memory (HBM).”). *See also, e.g.*, “AMD CDNA 2 Architecture,” pp. 4-5 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (GCD accessible memory).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	Each Accused AMD Accelerator Product in each Accused AMD Server Product comprises program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions. For example, when executed, the code stored in the memory accessible by an XCD causes the processor to evaluate mathematical expressions (e.g., using the CUs and/or ROCm™ APIs and/or executing commands in the RCCL library). For example, the XCD are configured, when the program code is executed, to evaluate mathematical expressions, for example square root, sine, cosine, and tangent, as well as matrix operations and/or operations between vectors.¹¹ Each XCD in a MI300X accelerator of the MI300X Platform, is configured to access one or more non-transitory memory devices comprising program code for a single-node kernel that, when executed causes the XCD to produce results of mathematical express evaluation. A first single-node kernel is responsible for interpreting user instructions and distributing a call to at least one of the other MI300X accelerators for execution. The other MI300X accelerators in the cluster are responsible for executing the tasks that are distributed to them by the first-single-node kernel. When the user submits a job to the cluster, the first single-node kernel parses the job and distributes the tasks to the other nodes for execution. The other nodes then execute the tasks and return the results to the first single-node kernel or communicate the results of mathematical expression evaluation to other nodes. The first single-node kernel then collects the results from the other nodes and returns them to the user. Thus, the XCD in a first MI300X accelerator is configured to access executable program code stored in memory

¹¹ See, e.g., “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see, e.g., “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>); see e.g., “RCCL API Library” at nccIAllReduce (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also “RCCL APL Library” at nccIAllGather (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at nccIAllReduce and at nccIAllGather (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	accessible by the XCD, where the executable program code is program code for a single-node kernel that, when executed, causes the processor to interpret user instructions. ¹²
a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture;	Each of the Accused AMD Accelerator Products in each Accused AMD Server Product comprises a mechanism for the nodes to communicate results of mathematical expression evaluation ¹³ with each other using a peer-to-peer architecture (e.g., the “direct connectivity between each of the [MI300X accelerators]” of the MI300X Platform, ¹⁴ along with related RCCL commands ¹⁵). ¹⁶

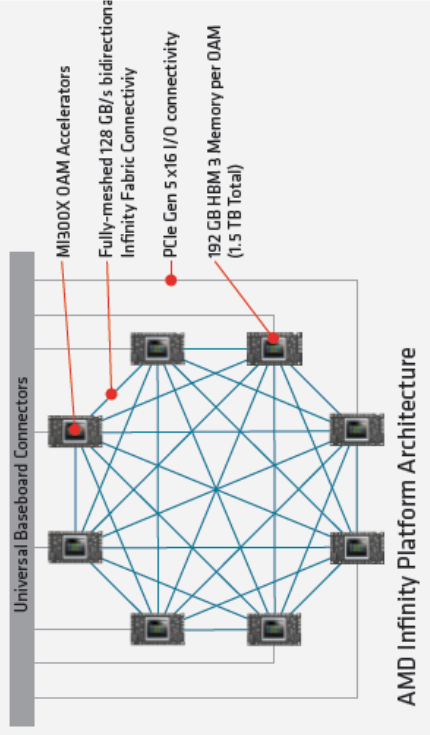
¹² The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable.

¹³ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library), (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>) (Math Library); see also, “roclBLAS” (<https://rocm.docs.amd.com/projects/roclBLAS/en/docs-5.5.1/index.html>); see also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture).

¹⁴ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

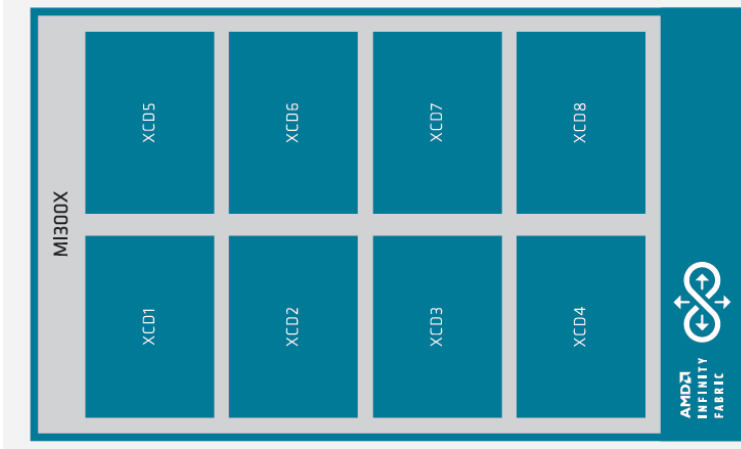
¹⁵ See “AMD Instinct™ MI300X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html>) (confirming MI300X OAM Accelerators support AMD’s ROCm™ software platform); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>) (confirming MI325X OAM Accelerators support AMD’s ROCm™ software platform).

¹⁶ See also, “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”); “AMD Instinct™ MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); see also, “AMD Instinct™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.¹⁷ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, and for the MI300X accelerator nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture.¹⁸
wherein the plurality of nodes comprises: a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node	Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple

¹⁷ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹⁸ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

U.S. Pat. No. 10,333,768 kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and	Accused AMD Server Products core units (“CUs”).¹⁹ The architecture of the MI300X accelerator is depicted in the following figure:²⁰  The diagram illustrates the MI300X accelerator architecture. It features eight compute dies, labeled XCD1 through XCD8, arranged in a 2x4 grid. Each die is represented by a teal square. The dies are interconnected by an AMD Infinity Fabric, which is shown as a network of lines connecting the dies. The AMD logo and the text 'AMD INFINITY FABRIC' are visible in the bottom right corner of the diagram.
--	---

¹⁹ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3–4, 7–9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

²⁰ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3–4, 7–9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	For example, the RCCL includes commands that cause a XCD of a MI300X accelerator in a MI300X Platform to distribute calls to at least one other MI300X accelerator in the MI300X Platform for execution.²¹ The source code for AMD’s sample HelloRccl.cpp program²² includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for a user interface and for a first single-node kernel, where the single node kernel is configured to interpret user instructions and to distribute call to at least one of a plurality of other MI300X nodes for execution.²³ AMD offers its AMD Accelerator Cloud (“AAC”) that “is a private cloud environment that enable customers to remotely access and use AMD EPYC™ CPUs, AMD Instinct™ GPUs and AMD ROCm™ technologies to experience the power of AMD hardware and software — along with all the possibilities that come with it. The AMD Accelerator Cloud makes it easy for prospective buyers to test and evaluate AMD Instinct accelerators. . . .”²⁴ “After being given access to the AMD Accelerator Cloud, users will be able to request access to a variety of node configurations with a range of AMD Instinct accelerator options. The user can select the desired operating system, ROCm version and directly install their preferred applications for testing or development. As we expand the AAC in the coming months, customers will be able to access a broader range of system types and GPUs, including the latest MI200 series accelerators.”²⁵

²¹ “RCCL documentation” (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) (“RCCL (pronounced ‘Rickel’) is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-all, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations.... [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform.”). RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.*

²² See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

²³ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes. *Id.*

²⁴ See “Take a Test Drive in the AMD Accelerator Cloud” (<https://community.amd.com/t5/instinct-accelerators/take-a-test-drive-in-the-amd-accelerator-cloud/ba-p/500901>).

²⁵ *Id.*

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	AMD advertises that its AMD Accelerator Product can be accessed through the cloud through other leading clouding providers, and advertises that such configurations can use AMD Instinct™ MI200 and MI300 series accelerators.²⁶ The computer cluster of each Accused AMD Accelerator Product includes program code to provide the user interface to allow users, including AMD’s own employees, to access the computer cluster through, at least, the cloud. The program code to provide the user interface of the Accused AMD Accelerator Product could also allow a user to connect to the computer cluster of each Accused AMD Accelerator Product through a direct connection.
a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the first mathematical	Each Accused AMD Server Product comprises a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the first mathematical expression evaluation to a third node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).²⁷ The architecture of the MI300X accelerator is depicted in the following figure:²⁸

²⁶ See “Unlock the Power of AMD Instinct™ GPU Accelerators Across Cloud Providers” (<https://community.amd.com/t5/instinct-accelerators/unlock-the-power-of-amd-instinct-gpu-accelerators-across-cloud/ba-p/729070/jump-to/first-unread-message>).

²⁷ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

²⁸ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
expression evaluation to a third node;	For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.²⁹ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for a user interface and for a first single-node kernel, for interpreting user instructions, for distributing calls to at least one of a plurality of other MI300X nodes for execution, for receiving the result of mathematical expression

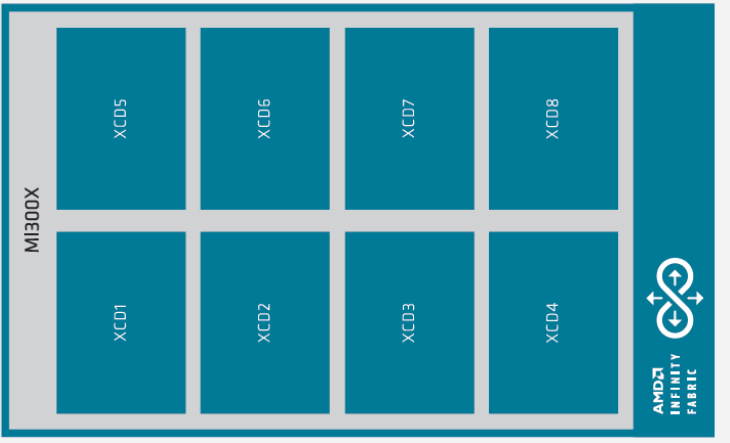
²⁹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	from other nodes, for executing mathematical expression evaluations, and for communicating a result of mathematical expression evaluation to other nodes and to the user interface. ³⁰
wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of the first mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second expression evaluation to the first node;	Each Accused AMD Server Product comprises a third node comprising a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of a first mathematical expression evaluation from the second node, to execute at least a second mathematical expression evaluation using the received result, and to communicate the result of the second mathematical expression evaluation to the first node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).³¹ The architecture of the MI300X accelerator is depicted in the following figure:³²

³⁰ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

³¹ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

³² “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf>) (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 For example, when executed, code stored in the memory accessible by the XCD of the third MI300X accelerator in a MI300X Platform causes the XCD to receive a result of a first mathematical expression evaluation from a second MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the received result (e.g., by evaluating mathematical expressions using the CUs and/or ROCm™ APIs and/or executing commands from the

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	RCCL)³³, and to communicate the result of the second mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.³⁴ Also, for example, at least because of the architecture of the XCD along with related resources of the ROCm™ Software, the XCD in a third MI300X accelerator in a MI300X Platform is configured to receive the result of a first mathematical expression evaluation from the second MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the received result (<i>e.g.</i>, via executable program code, such as RCCL commands), and to communicate the result of the second mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.³⁵ The source code for the AMD's sample HelloRccl.cpp program,³⁶ includes the details of using RCCL commands to receive the result of a first mathematical expression evaluation from a second Accused AMD Accelerator Product in an Accused AMD Server Product, to execute at least a second mathematical expression evaluation using the received result, and to communicate the result of the second mathematical expression evaluation to the first Accused AMD Accelerator in the Accused Server Product.
wherein the first node is configured to return the	A first Accused AMD Accelerator Product in each Accused AMD Server Product is configured to return the result of a second mathematical expression evaluation to the user interface.

³³ See *e.g.*, “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>”).

³⁴ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See *e.g.*, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, *e.g.*, “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at [nccIAllReduce](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html) and at [nccIAllGather](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html).

³⁵ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

³⁶ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
result of the second mathematical expression evaluation to the user interface;	For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform is configured to communicate a result of mathematical expression evaluation³⁷ to other nodes using a peer-to-peer architecture³⁸ and to a user interface. AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp program includes the details for using RCCL commands to return the result of mathematical expression evaluations from a first MI300X node in a MI300X Platform to the user interface.³⁹
wherein one or more of the nodes are configured to: accept user instructions; after accepting user instructions, communicate at least some of the user	One or more of the Accused AMD Accelerator Products in an Accused AMD Server Product are configured to: (1) accept user instructions; (2) after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other, and (3) after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.

³⁷ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software. *Id.* See also, e.g., “RCCL AllReduce” (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>) (Math Library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/index.html>); see also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture).

³⁸ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software. See e.g., “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”).

³⁹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>. The RCCL commands cause similar operation of MI200 accelerator nodes. *Id.* See also, <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
instructions using the mechanism for the nodes to communicate with each other; and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.	For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform is configured to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the MI300X accelerator nodes to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels on one or more MI300X accelerator nodes in an MI300X Platform.⁴⁰ Also, for example, AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp⁴¹ program includes the details for using RCCL commands to cause an Accused AMD Accelerator Product in an Accused AMD Server Product to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the Accused AMD Accelerator Products to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single node-kernels on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.
4. The computer cluster of claim 1, wherein each of the nodes comprises one or more cluster node modules.	Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises one or more cluster node modules. More particularly, each Accused AMD Accelerator Product node establishes intercommunications with other AMD Accelerator Product nodes of the computer cluster to exchange messages using a

⁴⁰ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

⁴¹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products peer-to-peer architecture (e.g., the “direct connectivity between each of the [MI300X accelerators]” of the MI300X Platform,⁴² along with related RCCL commands⁴³).⁴⁴ For example, the ROCm™ Software Platform includes commands that cause an XCD of a MI300X accelerator in a MI300X Platform to establish intercommunication with other nodes of the MI300X Platform and exchange messages using a peer-to-peer architecture. One nonlimiting example of such commands are the commands included in the ROCm™ Collective Communication Library (RCCL)⁴⁵ RCCL is included in the ROCm™ software platform.⁴⁶ Other of the ROCm™ Software Platform libraries and tools can also be used to establish intercommunication with other nodes of the MI300X Platform and exchange messages using a peer-to-peer architecture.⁴⁷
--------------------------	---

⁴² “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>).

⁴³ See “AMD INSTINCT™ MI300X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html>) (confirming MI300X OAM Accelerators support AMD’s ROCm™ software platform).

⁴⁴ See also, “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”); “AMD INSTINCT™ MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi250/mi250x.html>) (confirming MI250X OAM Accelerators support AMD’s ROCm™ software platform); “AMD INSTINCT™ MI200 Series Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi200.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD INSTINCT™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi250/mi250.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD INSTINCT™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi210/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).

⁴⁵ See <https://rocm.docs.amd.com/en/docs-6.1.1/reference/api-libraries.html>; “RCCL documentation” (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) (“RCCL (pronounced ‘Rickel’) is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-allv, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations.... [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform.”).

RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.* See also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). *See also*, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html).

⁴⁶ See <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/amd-rocm-6-brief.pdf>; “AMD CDNA 2 Architecture,” pp. 13 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

⁴⁷ These other ROCm™ Software Platform libraries and tools can also be used to establish intercommunication with other nodes of the MI200 Platform and exchange messages using a peer-to-peer architecture. *Id.*

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	The source code for AMD’s sample HelloRccl.cpp program⁴⁸ includes the details of an example for using RCCL commands to initialize a RCCL communicator on a MI300X Platform system comprising multiple MI300X accelerator nodes to establish intercommunication with other nodes of the MI300X Platform and exchange messages using a peer-to-peer architecture.⁴⁹
20. The computer cluster of claim 1, wherein one or more of the nodes are configured to accept user instructions via one or more of the nodes.	One or more of the Accused AMD Accelerator Products in each Accused AMD Server Product is configured to accept user instructions via one or more of the Accused AMD Accelerator Products in the Accused AMD Server Product. communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other.⁵⁰ AMD offers its AMD Accelerator Cloud (“AAC”) that “is a private cloud environment that enable customers to remotely access and use AMD EPYC™ CPUs, AMD Instinct™ GPUs and AMD ROCm™ technologies to experience the power of AMD hardware and software — along with all the possibilities that come with it. The AMD Accelerator Cloud makes it easy for prospective buyers to test and evaluate AMD Instinct accelerators. . . .”⁵¹ “After being given access to the AMD Accelerator Cloud, users will be able to request access to a variety of node configurations with a range of AMD Instinct accelerator options. The user can select the desired operating system, ROCm version and directly install their preferred applications for testing or development. As we expand the AAC in

⁴⁸ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>; *see also* <https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>.

⁴⁹ *Id.* The HelloRccl.cpp program includes details for using RCCL commands to similarly initialize a RCCL communicator on a MI200 Platform system comprising multiple MI200 accelerator nodes.

⁵⁰ See “AMD Instinct™ MI300X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi300x.html>) (confirming MI300X OAM Accelerators support AMD’s ROCm™ software platform). *See also*, “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”); “AMD Instinct™ MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi250x.html>) (confirming MI250 software platform); “AMD Instinct™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi250.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).
⁵¹ See “Take a Test Drive in the AMD Accelerator Cloud” (<https://community.amd.com/t5/instinct-accelerators/take-a-test-drive-in-the-amd-accelerator-cloud/ba-p/500901>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products the coming months, customers will be able to access a broader range of system types and GPUs, including the latest MI200 series accelerators.”⁵² AMD advertises that its AMD Accelerator Product can be accessed through the cloud through other leading clouding providers, and advertises that such configurations can use AMD Instinct™ MI200 and MI300 series accelerators.⁵³ Each Accused AMD Accelerator Product includes program code to provide the user interface to allow users, including AMD’s own employees, to access the computer cluster through, at least, the cloud. The program code to provide the user interface of the Accused AMD Accelerator Product could also allow a user to connect to the computer cluster of each Accused AMD Accelerator Product through a direct connection.
21. The computer cluster of claim 20, wherein one or more of the nodes are configured to communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other.	One or more of the Accused AMD Accelerator Products in each Accused AMD Server Product is configured to communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other.⁵⁴

⁵² *Id.*

⁵³ See “Unlock the Power of AMD Instinct™ GPU Accelerators Across Cloud Providers” (<https://community.amd.com/t5/instinct-accelerators/unlock-the-power-of-amd-instinct-gpu-accelerators-across-cloud/ba-p/729070/jump-to/first-unread-message>).

⁵⁴ See “AMD Instinct™ MI300X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html>) (confirming MI300X OAM Accelerators support AMD’s ROCm™ software platform). See also, “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”); “AMD Instinct™ MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	The diagram illustrates the AMD Infinity Platform Architecture. It features a central mesh of nodes, each representing an MI300X OAM Accelerator. These accelerators are interconnected via a fully-meshed 128 GB/s bidirectional Infinity Fabric Connectivity. The architecture is connected to Universal Baseboard Connectors and PCIe Gen 5x16 I/O connectivity. Each accelerator contains 192 GB of HBM3 Memory, totaling 1.5 TB per OAM. For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.⁵⁵ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, and for the MI300X accelerator nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture.⁵⁶

⁵⁵ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

⁵⁶ *Id.* RCCL also supports other collective communication functions. See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>). The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

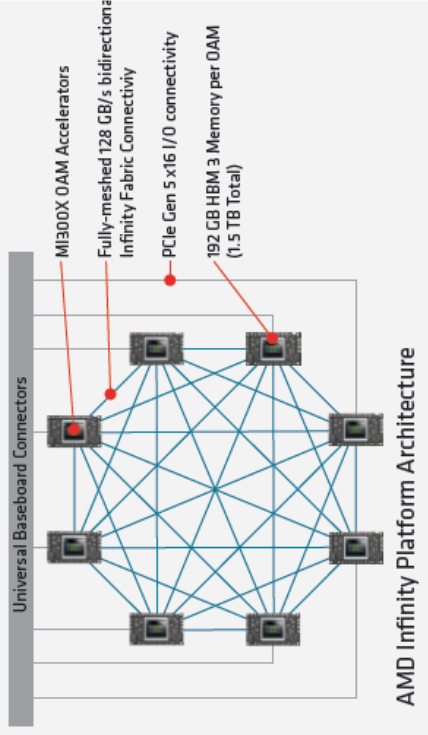
U.S. Pat. No. 10,333,768	Accused AMD Server Products
26. A computer cluster comprising: "	Each of the Accused AMD Server Products⁵⁷ comprises a computer cluster. For example, as depicted below and as described further herein, AMD's INSTINCT™ MI300X Platform ("MI300X Platform") includes a computer cluster.⁵⁸ 

⁵⁷ The Accused AMD Server Products include, but are not limited to, all products including or related to AMD's MI300X and MI325X Platforms

(<https://www.amd.com/en/products/accelerators/instinct/mi300/platform.html>), and all products including or related to AMD INSTINCT™ Accelerator Qualified Servers (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>) incorporating any of the Accused Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items.

⁵⁸ "AMD INSTINCT™ MI300X Platform™ Data Sheet" (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); *see also* MI300 introductory video ("Introducing AMD Instinct™ MI300 Series Accelerators"), at <https://www.youtube.com/watch?v=uVdNpk8tvBI>; *see also* "AMD INSTINCT™ MI325X Platform™ Data Sheet" (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, as described further herein, each of the MI210, MI250, and MI250X Accelerators ("MI200 accelerators") is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server ("MI200 Platform") such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. *See e.g.*, "AMD Instinct Accelerator-Powered Servers" (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>); "AMD Instinct MI250X Accelerators" (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>); "AMD Instinct MI250 Accelerators" (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>); "AMD Instinct MI210 Accelerators" (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>); "AMD CDNA™ 2 Architecture," pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (Infinity Fabric connecting MI200 accelerators both in-package and inter-package). *See also*, "AMD INSTINCT™ MI200 Series Accelerator" (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>); "AMD INSTINCT™ MI210 Accelerator" (<https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/product-briefs/instinct-mi210-brochure.pdf>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products Each of the Accused AMD Server Products includes a plurality of nodes, each of which comprises an Accused AMD Accelerator Product.⁵⁹ For example, as depicted below, the AMD MI300X Platform comprises eight MI300X accelerators “on an industry-standard universal baseboard” that “are connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”⁶⁰  AMD Infinity Platform Architecture Each of the Accused AMD Server Products comprises a plurality of nodes, each of which comprises an Accused AMD Accelerator Product. For example, as depicted below, the MI300X Platform
a plurality of nodes, "wherein one or more of the nodes are configured to	

⁵⁹ Accused AMD Accelerator Products include, but are not limited to, all products including or relating to AMD’s MI300 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi300.html>) (MI300X, MI325X, MI300A), and MI200 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi200.html>) (MI250X, MI250, MI210), as well as any products incorporating those items, and any products that are substantially similar to those items.

⁶⁰ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, AMD specifies a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. See “AMD CDNA™ 2 Architecture,” p. 7 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>). See also, e.g., *id.*, pp. 6 and 8-9 (additional topologies); pp. 3-4 (Infinity Fabric connecting MI200 accelerators both in-package and inter-package).

U.S. Pat. No. 10,333,768 in view of Accused AMD Server Products
ACS's Prelim. Infringement Contentions

24

Exh. A1
February 24, 2025

U.S. Pat. No. 10,333,768 receive: a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; and an instruction from a user interface or a script; and	Accused AMD Server Products comprises eight MI300X accelerators (e.g., nodes) that “are connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”⁶¹ AMD Infinity Platform Architecture Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises a hardware processor, wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes. One or more of the nodes are also configured to receive an instruction from a user interface or a script. Specifically, each Accused AMD Accelerator Product comprises a hardware processor that comprises multiple processor cores. For example, each MI300X accelerator is a node that includes an
--	--

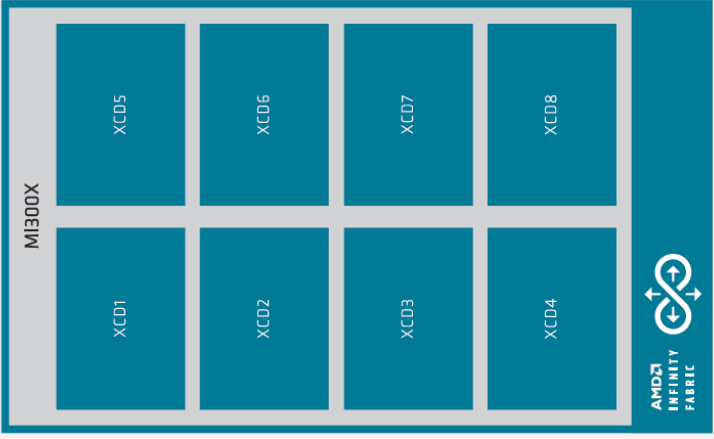
⁶¹ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, the MI200 series accelerators are designed to be integrated into at least one MI200 Platform that incorporates a plurality of MI200 accelerators according to a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. See e.g., “AMD Instinct Accelerator-Powered Servers” (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>); “AMD CDNA™ 2 Architecture,” p. 7 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>); pp. 3-4 (Infinity Fabric connecting MI200 accelerators both in-package and inter-package). See also, e.g., *id.*, pp. 6 and 8-9 (additional topologies).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	accelerator complex die (“XCD”) that contains multiple core units (“CUs”). ⁶² The architecture of the MI300X accelerator is depicted in the following figure: ⁶³

⁶² Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

⁶³ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p.2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 Each Accused AMD Accelerator Product comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster of each Accused AMD Service Product. For example, each MI300X accelerator in each MI300X Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes the eight MI300X accelerators in the MI300X Platform⁶⁴. This is done,

⁶⁴ Also, for example, each MI200 accelerator in an MI200 Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes at least three MI200 accelerators.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	for example, by using AMD's ROCm™ Collective Communication Library (RCCL), APIs, and math libraries.⁶⁵ AMD offers its AMD Accelerator Cloud ("AAC") that "is a private cloud environment that enable customers to remotely access and use AMD EPYC™ CPUs, AMD Instinct™ GPUs and AMD ROCm™ technologies to experience the power of AMD hardware and software — along with all the possibilities that come with it. The AMD Accelerator Cloud makes it easy for prospective buyers to test and evaluate AMD Instinct accelerators. . . ."⁶⁶ "After being given access to the AMD Accelerator Cloud, users will be able to request access to a variety of node configurations with a range of AMD Instinct accelerator options. The user can select the desired operating system, ROCm version and directly install their preferred applications for testing or development. As we expand the AAC in the coming months, customers will be able to access a broader range of system types and GPUs, including the latest MI200 series accelerators."⁶⁷ AMD advertises that its AMD Accelerator Product can be accessed through the cloud through other leading clouding providers, and advertises that such configurations can use AMD Instinct™ MI200 and MI300 series accelerators.⁶⁸ The computer cluster of each Accused AMD Accelerator Product includes program code to provide the user interface to allow users, including AMD's own employees, to access the computer cluster through, at least, the cloud. The program code to provide the user interface of the Accused AMD Accelerator Product could also allow a user to connect to the computer cluster of each Accused AMD Accelerator Product through a direct connection.

⁶⁵ "RCCL documentation" (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) ("RCCL (pronounced 'Rickel') is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-all, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations.... [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform."). RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.* RCCL also includes commands that cause similar operation of a MI200 accelerator in an MI200 Platform. *Id.*

⁶⁶ See "Take a Test Drive in the AMD Accelerator Cloud" (<https://community.amd.com/t5/instinct-accelerators/take-a-test-drive-in-the-amd-accelerator-cloud/ba-p/500901>).

⁶⁷ *Id.*

⁶⁸ See "Unlock the Power of AMD Instinct™ GPU Accelerators Across Cloud Providers" (<https://community.amd.com/t5/instinct-accelerators/unlock-the-power-of-amd-instinct-gpu-accelerators-across-cloud/ba-p/729070/jump-to/first-unread-message>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates that the nodes can receive instructions from a user interface or a script.⁶⁹
a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using asynchronous calls;	Each of the Accused Accelerator Products in each Accused AMD Server Product comprises a mechanism for the nodes to communicate results of mathematical expression evaluation⁷⁰ with each other using asynchronous calls. Universal Baseboard Connectors MI300X OAM Accelerators Fully-meshed 128 GB/s bidirectional Infinity Fabric Connectivity PCIe Gen 5 x16 I/O connectivity 192 GB HBM3 Memory per OAM (1.5 TB Total) AMD Infinity Platform Architecture 71

⁶⁹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>. The RCCL commands cause similar operation of MI200 accelerator nodes. *Id.* See also, <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

⁷⁰ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), *see also*, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-half-library>), (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-half-library>) (Math Library); *see also*, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/index.html>); *see also*, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture).

⁷¹ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi325x-platform-data-sheet.pdf>); “AMD INSTINCT™ MI250 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi250x-platform-data-sheet.pdf>); “AMD INSTINCT™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi250x.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).

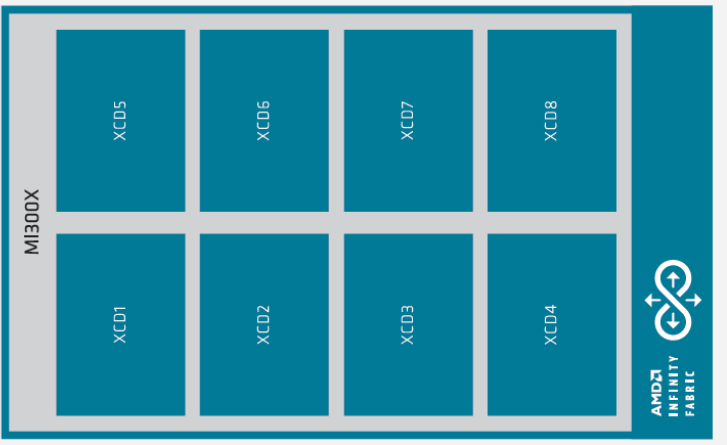
U.S. Pat. No. 10,333,768	Accused AMD Server Products
wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions; wherein the plurality of nodes comprises: a first node comprising a first hardware	For example, AMD provides the ROCm libraries utilizing both collective and point to point commands. AMD provides a sample HelloRccl.cpp program illustrating the use of RCCL commands to communicate results of mathematical expressions.⁷² AMD further provides for programming code to check the operation of the collective and point to point commands.⁷³ Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises a plurality of nodes that are configured to access a non-transitory computer readable medium program for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions, wherein the plurality of nodes comprises a first node, a first hardware processor, and is configured to access a first memory. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).⁷⁴ The architecture of the MI300X accelerator is depicted in the following figure:⁷⁵

⁷² See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

⁷³ *Id.* See also, e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); “HIP Stream Management” (https://rocm.docs.amd.com/projects/HIP/en/latest/reference/hip_runtime_api/modules/stream_management.html#_CPPv420hipStreamSynchronize11hipStream1), (https://rocm.docs.amd.com/projects/HIP/en/docs-5.5.1/doxygen/html/group__stream.html#gabbfb9f573a6ebe8c478605ecb5504a74).

⁷⁴ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

⁷⁵ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p.2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and	 Each AMD Accelerator Product is configured to access a non-transitory computer readable medium (e.g., memory accessible by a XCD, including local memory, shared memory, and HBM memory, as well as memory external to the XCD).⁷⁶

⁷⁶ “AMD CDNA™ 3 Architecture,” p. 3 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/white-papers/amd-cdna-3-white-paper.pdf>); *Id.* at p. 2 (“The MI300 Series integrates up to 8 vertically stacked accelerator complex dies (XCD) and 4 I/O dies (IOD) containing system infrastructure, all tied together with the AMD Infinity Fabric™ technology and connecting to 8 stacks of high-bandwidth memory (HBM).”). *See also, e.g.*, “AMD CDNA 2 Architecture,” pp. 4-5 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (GCD accessible memory).

U.S. Pat. No. 10,333,768	Accused AMD Server Products Figure 1. Advanced 3D Package and chiplet-based construction of the AMD MI300 Series processors Each Accused AMD Accelerator Product in each Accused AMD Server Product comprises program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions. For example, when executed, the code stored in the memory accessible by an XCD causes the processor to evaluate mathematical expressions (e.g., using the CUs and/or ROCm™ APIs and/or executing commands in the RCCL library). For example, the XCD are configured, when the program code is executed, to evaluate mathematical expressions, for example square root, sine, cosine, and tangent, as well as matrix operations and/or operations between vectors.⁷⁷
--------------------------	--

⁷⁷ See, e.g., “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see, e.g., “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>); see e.g., “RCCL API Library” at [nclAllReduce](https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#) (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also “RCCL APL Library” at [nclAllGather](https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#) (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	The RCCL further includes commands that cause a XCD of a MI300X accelerator in a MI300X Platform to distribute calls to at least one other MI300X accelerator in the MI300X Platform for execution.⁷⁸ The source code for AMD’s sample HelloRccl.cpp program⁷⁹ includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for a user interface and for a first single-node kernel, where the single node kernel is configured to interpret user instructions and to distribute call to at least one of a plurality of other MI300X nodes for execution.⁸⁰ AMD offers its AMD Accelerator Cloud (“AAC”) that “is a private cloud environment that enable customers to remotely access and use AMD EPYC™ CPUs, AMD Instinct™ GPUs and AMD ROCm™ technologies to experience the power of AMD hardware and software — along with all the possibilities that come with it. The AMD Accelerator Cloud makes it easy for prospective buyers to test and evaluate AMD Instinct accelerators. . . .”⁸¹ “After being given access to the AMD Accelerator Cloud, users will be able to request access to a variety of node configurations with a range of AMD Instinct accelerator options. The user can select the desired operating system, ROCm version and directly install their preferred applications for testing or development. As we expand the AAC in

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. *See e.g.*, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). *See also, e.g.*, “ROCm

Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at nclAllReduce and at nclAllGather (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

⁷⁸ “RCCL documentation” (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) (“RCCL (pronounced ‘Rickel’) is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-all, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations. . . . [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform.”). RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.*

⁷⁹ *See* <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

⁸⁰ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes. *Id.*

⁸¹ *See* “Take a Test Drive in the AMD Accelerator Cloud” (<https://community.amd.com/t5/instinct-accelerators/take-a-test-drive-in-the-amd-accelerator-cloud/ba-p/500901>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	the coming months, customers will be able to access a broader range of system types and GPUs, including the latest MI200 series accelerators.”⁸² AMD advertises that its AMD Accelerator Product can be accessed through the cloud through other leading clouding providers, and advertises that such configurations can use AMD Instinct™ MI200 and MI300 series accelerators.⁸³ The computer cluster of each Accused AMD Accelerator Product includes program code to provide the user interface to allow users, including AMD’s own employees, to access the computer cluster through, at least, the cloud. The program code to provide the user interface of the Accused AMD Accelerator Product could also allow a user to connect to the computer cluster of each Accused AMD Accelerator Product through a direct connection.
a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the mathematical expression evaluation to a third node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).⁸⁴ The architecture of the MI300X accelerator is depicted in the following figure:⁸⁵	Each Accused AMD Server Product comprises a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the mathematical expression evaluation to a third node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).⁸⁴ The architecture of the MI300X accelerator is depicted in the following figure:⁸⁵

⁸² *Id.*

⁸³ See “Unlock the Power of AMD Instinct™ GPU Accelerators Across Cloud Providers” (<https://community.amd.com/t5/instinct-accelerators/unlock-the-power-of-amd-instinct-gpu-accelerators-across-cloud/ba-p/729070/jump-to/first-unread-message>).

⁸⁴ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3–4, 7–9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

⁸⁵ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-accelerator-data-sheet.pdf>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
communicate a result of mathematical expression evaluation to a third node; wherein	For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.⁸⁶ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for a user interface and for a first single-node kernel, for interpreting user instructions, for distributing calls to at least one of a plurality of other MI300X nodes for execution, for receiving the result of mathematical expression

mi325x-datasheet.pdf. See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

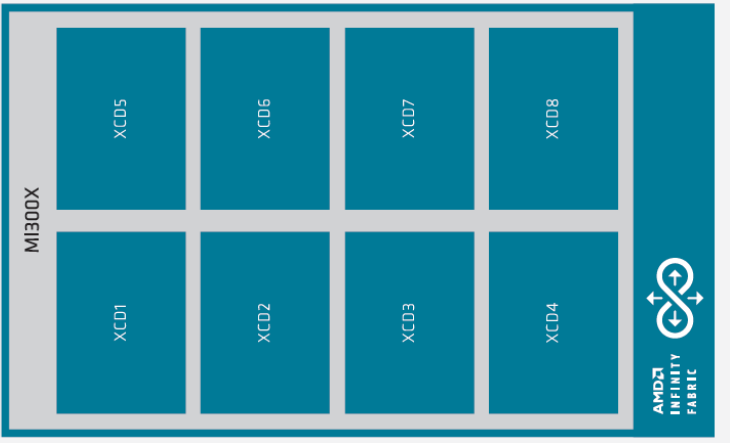
⁸⁶ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	from other nodes, for executing mathematical expression evaluations, and for communicating a result of mathematical expression evaluation to other nodes and to the user interface. ⁸⁷
the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and to communicate the result of the second mathematical expression evaluation to the first node;	Each of the plurality of Accused AMD Accelerator Product nodes in each accused AMD Server Product comprises a third node comprising a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, to execute at least a second mathematical expression evaluation using the received result, and to communicate the result of the second mathematical expression evaluation to the first node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).⁸⁸ The architecture of the MI300X accelerator is depicted in the following figure:⁸⁹

⁸⁷ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

⁸⁸ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

⁸⁹ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 For example, when executed, code stored in the memory accessible by the XCD of the third MI300X accelerator in a MI300X Platform causes the XCD to receive a result of a first mathematical expression evaluation from a second MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the received result (e.g., by evaluating mathematical expressions using the CUs and/or ROCm™ APIs and/or executing commands from the

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	RCCL)⁹⁰, and to communicate the result of the second mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.⁹¹ Also, for example, at least because of the architecture of the XCD along with related resources of the ROCm™ Software, the XCD in a third MI300X accelerator in a MI300X Platform is configured to receive the result of a first mathematical expression evaluation from the second MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the received result (<i>e.g.</i>, via executable program code, such as RCCL commands), and to communicate the result of the second mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.⁹² The source code for the AMD's sample HelloRccl.cpp program,⁹³ includes the details of using RCCL commands to receive the result of mathematical expression evaluation from a second Accused AMD Accelerator Product in an Accused AMD Server Product, to execute at least a second mathematical expression evaluation using the received result, and to communicate the result of the second mathematical expression evaluation to the first Accused AMD Accelerator in the Accused Server Product.
wherein the first node is configured to return the	A first Accused AMD Accelerator Product in each Accused AMD Server Product is configured to return the result of a second mathematical expression evaluation to the user interface or the script.

⁹⁰ See *e.g.*, “RCCL AllReduce” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see *also*, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see *also*, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>”).

⁹¹ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See *e.g.*, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See *also, e.g.*, “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at [nccIAllReduce](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html) and at [nccIAllGather](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html).

⁹² The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

⁹³ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
result of the second mathematical expression evaluation to the user interface or the script;	For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform is configured to communicate a result of mathematical expression evaluation⁹⁴ to other nodes using a peer-to-peer architecture⁹⁵ and to a user interface or the script. AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp program includes the details for using RCCL commands to return the result of mathematical expression evaluations from a first MI300X node in a MI300X Platform to the user interface or the script.⁹⁶
wherein one or more of the nodes are configured to: accept user instructions; after accepting user instructions, communicate at least some of the user	One or more of the Accused AMD Accelerator Products in an Accused AMD Server Product are configured to: (1) accept user instructions; (2) after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other, and (3) after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.

⁹⁴ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software. *Id.* See also, e.g., “RCCL AllReduce” (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>) (Math Library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/index.html>); see also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture).

⁹⁵ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software. See e.g., “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”).

⁹⁶ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>. The RCCL commands cause similar operation of MI200 accelerator nodes. *Id.* See also, <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
instructions using the mechanism for the nodes to communicate with each other; and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.	For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform is configured to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the MI300X accelerator nodes to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels on one or more MI300X accelerator nodes in an MI300X Platform.⁹⁷ Also, for example, AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp⁹⁸ program includes the details for using RCCL commands to cause an Accused AMD Accelerator Product in an Accused AMD Server Product to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the Accused AMD Accelerator Products to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single node-kernels on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.
29. A computer cluster comprising:	Each of the Accused AMD Server Products⁹⁹ comprises a computer cluster. For example, as depicted below and as described further herein, AMD's INSTINCT™ MI300X Platform ("MI300X Platform") includes a computer cluster.¹⁰⁰

⁹⁷ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

⁹⁸ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

⁹⁹ The Accused AMD Server Products include, but are not limited to, all products including or related to AMD's MI300X and MI325X Platforms (<https://www.amd.com/en/products/accelerators/instinct/mi300/platform.html>), and all products including or related to AMD INSTINCT™ Accelerator Qualified Servers (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>) incorporating any of the Accused Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items.

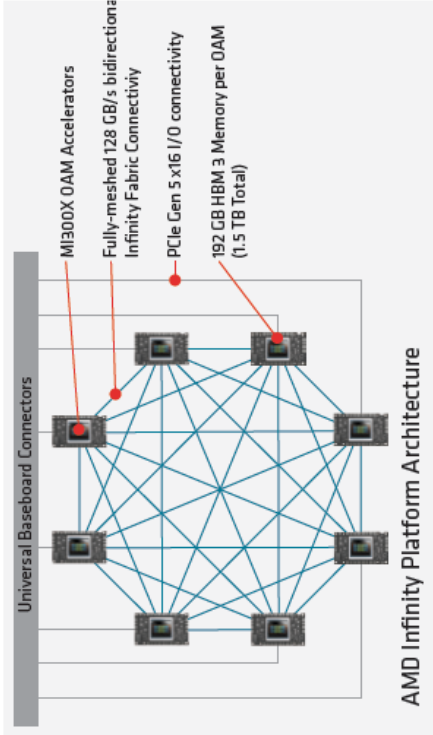
¹⁰⁰ "AMD INSTINCT™ MI300X Platform™ Data Sheet" (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also MI300 introductory video ("Introducing AMD Instinct™ MI300 Series Accelerators"), at

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 Each of the Accused AMD Server Products includes a plurality of nodes, each of which comprises an Accused AMD Accelerator Product.¹⁰¹ For example, as depicted below, the AMD MI300X Platform comprises eight MI300X accelerators “on an industry-standard universal baseboard” that “are

¹⁰¹ <https://www.youtube.com/watch?v=uVdNpk8tvBI>; see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, as described further herein, each of the MI210, MI250, and MI250X Accelerators (“MI200 accelerators”) is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server (“MI200 Platform”) such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster. See e.g., “AMD Instinct Accelerator-Powered Servers” (<https://www.amd.com/system/files/documents/amd-instinct-server-solutions-catalog.pdf>); “AMD Instinct MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>); “AMD Instinct MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>); “AMD Instinct MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>); “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (Infinity Fabric connecting MI200 accelerators both in-package and inter-package). See also, “AMD INSTINCT™ MI200 Series Accelerator” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>); “AMD INSTINCT™ MI210 Accelerator” (<https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/product-briefs/instinct-mi210-brochure.pdf>).

¹⁰¹ Accused AMD Accelerator Products include, but are not limited to, all products including or relating to AMD’s MI300 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi300.html>) (MI300X, MI325X, MI300A), and MI200 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi200.html>) (MI250X, MI250, MI210), as well as any products incorporating those items, and any products that are substantially similar to those items.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”¹⁰²  AMD Infinity Platform Architecture
a plurality of nodes, wherein one or more of the nodes are configured to receive: a command to start a cluster initialization	Each of the Accused AMD Server Products comprises a plurality of nodes, each of which comprises an Accused AMD Accelerator Product. For example, as depicted below, the MI300X Platform comprises eight MI300X accelerators (e.g., nodes) that “are connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”¹⁰³

¹⁰² “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); *see also* “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, AMD specifies a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. *See* “AMD CDNA™ 2 Architecture,” p. 7 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>). *See also*, e.g., *id.*, pp. 6 and 8-9 (additional topologies); pp. 3-4 (Infinity Fabric connecting MI200 accelerators both in-package and inter-package).

¹⁰³ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); *see also* “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, the MI200 series accelerators are designed to be integrated into at least one MI200 Platform that incorporates a plurality of MI200 accelerators according to a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. *See e.g.*,

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	Each Accused AMD Accelerator Product comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster of each Accused AMD Service Product. For example, each MI300X accelerator in each MI300X Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes the eight MI300X accelerators in the MI300X Platform¹⁰⁶. This is

¹⁰⁶ [mi325x-datasheet.pdf](https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf)). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

Also, for example, each MI200 accelerator in an MI200 Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes at least three MI200 accelerators.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	done, for example, by using AMD’s ROCm™ Collective Communication Library (RCCL), APIs, and math libraries.¹⁰⁷ AMD offers its AMD Accelerator Cloud (“AAC”) that “is a private cloud environment that enable customers to remotely access and use AMD EPYC™ CPUs, AMD Instinct™ GPUs and AMD ROCm™ technologies to experience the power of AMD hardware and software — along with all the possibilities that come with it. The AMD Accelerator Cloud makes it easy for prospective buyers to test and evaluate AMD Instinct accelerators. . . .”¹⁰⁸ “After being given access to the AMD Accelerator Cloud, users will be able to request access to a variety of node configurations with a range of AMD Instinct accelerator options. The user can select the desired operating system, ROCm version and directly install their preferred applications for testing or development. As we expand the AAC in the coming months, customers will be able to access a broader range of system types and GPUs, including the latest MI200 series accelerators.”¹⁰⁹ AMD advertises that its AMD Accelerator Product can be accessed through the cloud through other leading clouding providers, and advertises that such configurations can use AMD Instinct™ MI200 and MI300 series accelerators.¹¹⁰ The computer cluster of each Accused AMD Accelerator Product includes program code to provide the user interface to allow users, including AMD’s own employees, to access the computer cluster through, at least, the cloud. The program code to provide the user interface of the Accused AMD Accelerator Product could also allow a user to connect to the computer cluster of each Accused AMD Accelerator Product through a direct connection.

¹⁰⁷ “RCCL documentation” (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) (“RCCL (pronounced ‘Rickel’) is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-all, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations. . . . [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform.”). RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.* RCCL also includes commands that cause similar operation of a MI200 accelerator in an MI200 Platform. *Id.*

¹⁰⁸ See “Take a Test Drive in the AMD Accelerator Cloud” (<https://community.amd.com/t5/instinct-accelerators/take-a-test-drive-in-the-amd-accelerator-cloud/ba-p/500901>).

¹⁰⁹ *Id.*

¹¹⁰ See “Unlock the Power of AMD Instinct™ GPU Accelerators Across Cloud Providers” (<https://community.amd.com/t5/instinct-accelerators/unlock-the-power-of-amd-instinct-gpu-accelerators-across-cloud/ba-p/729070/jump-to/first-unread-message>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates that the nodes can receive instructions from a user interface or a script. ¹¹¹
a mechanism for the nodes to communicate results of mathematical expression evaluation with each other;	Each of the Accused AMD Accelerator Products in each Accused AMD Server Product comprises a mechanism for the nodes to communicate results of mathematical expression evaluation ¹¹² with each other (e.g., using a peer-to-peer architecture; “direct connectivity between each of the [MI300X accelerators]” of the MI300X Platform, ¹¹³ along with related RCCL commands ¹¹⁴). ¹¹⁵

¹¹¹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>. The RCCL commands cause similar operation of MI200 accelerator nodes. *Id.* See also, <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹¹² See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library), (<https://rocm.docs.amd.com/en/latest/index.html>); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.1/index.html>); see also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture).
¹¹³ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

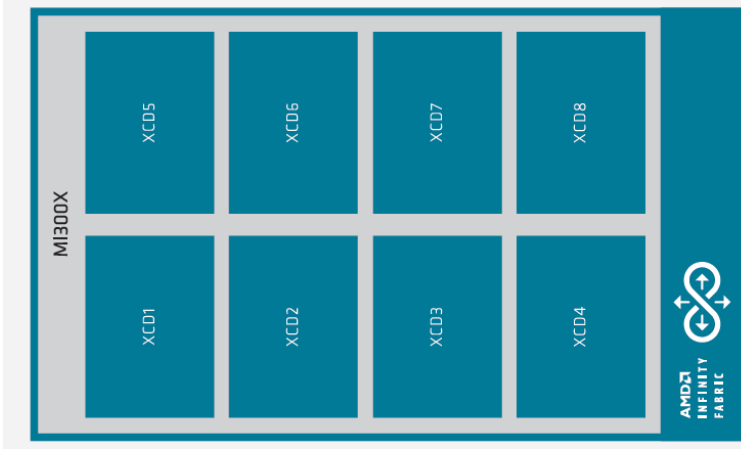
¹¹⁴ See “AMD Instinct™ MI300X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html>) (confirming MI300X OAM Accelerators support AMD’s ROCm™ software platform); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p.2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>) (confirming MI325X OAM Accelerators support AMD’s ROCm™ software platform).

¹¹⁵ See also, “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”); “AMD Instinct™ MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); see also, “AMD Instinct™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).

U.S. Pat. No. 10,333,768	Accused AMD Server Products Universal Baseboard Connectors MI300X OAM Accelerators Fully-meshed 128 GB/s bidirectional Infinity Fabric Connectivity PCIe Gen 5x16 I/O connectivity 192 GB HBM3 Memory per OAM (1.5 TB Total) AMD Infinity Platform Architecture For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.¹¹⁶ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, and for the MI300X accelerator nodes to communicate results of mathematical expression evaluation with each other.¹¹⁷
wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical	Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises a plurality of nodes that are configured to access a non-transitory computer readable medium program for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions, wherein the plurality of nodes comprises a first node, a first hardware processor, and is configured to access a first memory. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple

¹¹⁶ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹¹⁷ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

U.S. Pat. No. 10,333,768 expressions; wherein the plurality of nodes comprises: a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and	Accused AMD Server Products core units (“CUs”).¹¹⁸ The architecture of the MI300X accelerator is depicted in the following figure:¹¹⁹  The diagram illustrates the architecture of the MI300X accelerator. It features eight compute units, labeled XCD1 through XCD8, arranged in a 2x4 grid. Each unit is represented by a teal square. The units are interconnected by an AMD Infinity Fabric, which is shown as a network of lines connecting the units. The AMD logo and the text 'AMD INFINITY FABRIC' are visible in the bottom right corner of the diagram.
--	--

¹¹⁸ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

¹¹⁹ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p.2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products Each AMD Accelerator Product is configured to access a non-transitory computer readable medium (e.g., memory accessible by a XCD, including local memory, shared memory, and HBM memory, as well as memory external to the XCD).¹²⁰ Figure 1. Advanced 3D Package and chiplet-based construction of the AMD MI300 Series processors Each Accused AMD Accelerator Product in each Accused AMD Server Product comprises program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions. For example, when executed, the code stored in the memory accessible by an XCD causes the processor to evaluate mathematical expressions (e.g., using the CUs and/or ROCm™ APIs and/or executing commands in the RCCL library). For example, the XCD are configured, when the program code is executed, to evaluate mathematical expressions, for example
--------------------------	---

¹²⁰ “AMD CDNA™ 3 Architecture,” p. 3 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/white-papers/amd-cdna-3-white-paper.pdf>); *Id.* at p. 2 (“The MI300 Series integrates up to 8 vertically stacked accelerator complex dies (XCD) and 4 I/O dies (IOD) containing system infrastructure, all tied together with the AMD Infinity Fabric™ technology and connecting to 8 stacks of high-bandwidth memory (HBM).”). *See also*, e.g., “AMD CDNA 2 Architecture,” pp. 4-5 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (GCD accessible memory).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	square root, sine, cosine, and tangent, as well as matrix operations and/or operations between vectors.¹²¹ The RCCL further includes commands that cause a XCD of a MI300X accelerator in a MI300X Platform to distribute calls to at least one other MI300X accelerator in the MI300X Platform for execution.¹²² The source code for AMD’s sample HelloRccl.cpp program¹²³ includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for a user interface and for a first single-node kernel, where the single node kernel is configured to interpret user instructions and to distribute call to at least one of a plurality of other MI300X nodes for execution.¹²⁴ AMD offers its AMD Accelerator Cloud (“AAC”) that “is a private cloud environment that enable customers to remotely access and use AMD EPYC™ CPUs, AMD Instinct™ GPUs and AMD ROCm™ technologies to experience the power of AMD hardware and software — along with all the

¹²¹ See, e.g., “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see, e.g., “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>); see e.g., “RCCL API Library” at nccAllReduce (<https://rocm.docs.amd.com/projects/rccl/en/latest/reference/api-library.html#>); see also “RCCL APL Library” at nccAllGather (<https://rocm.docs.amd.com/projects/rccl/en/latest/reference/api-library.html#>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at nccAllReduce and at nccAllGather (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

¹²² “RCCL documentation” (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) (“RCCL (pronounced ‘Rickel’) is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-allv, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations.... [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform.”). RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.*

¹²³ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

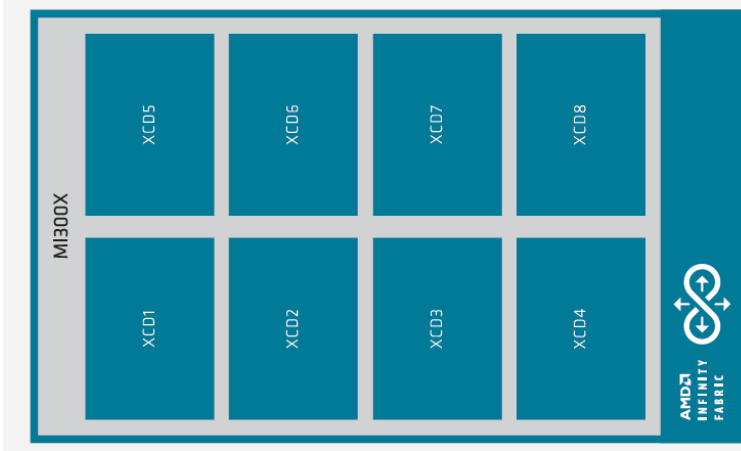
¹²⁴ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes. *Id.*

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	possibilities that come with it. The AMD Accelerator Cloud makes it easy for prospective buyers to test and evaluate AMD Instinct accelerators. . . .”¹²⁵ “After being given access to the AMD Accelerator Cloud, users will be able to request access to a variety of node configurations with a range of AMD Instinct accelerator options. The user can select the desired operating system, ROCm version and directly install their preferred applications for testing or development. As we expand the AAC in the coming months, customers will be able to access a broader range of system types and GPUs, including the latest MI200 series accelerators.”¹²⁶ AMD advertises that its AMD Accelerator Product can be accessed through the cloud through other leading clouding providers, and advertises that such configurations can use AMD Instinct™ MI200 and MI300 series accelerators.¹²⁷ The computer cluster of each Accused AMD Accelerator Product includes program code to provide the user interface to allow users, including AMD’s own employees, to access the computer cluster through, at least, the cloud. The program code to provide the user interface of the Accused AMD Accelerator Product could also allow a user to connect to the computer cluster of each Accused AMD Accelerator Product through a direct connection.
a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the mathematical expression evaluation to a third node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple	Each Accused AMD Server Product comprises a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the mathematical expression evaluation to a third node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple

¹²⁵ See “Take a Test Drive in the AMD Accelerator Cloud” (<https://community.amd.com/t5/instinct-accelerators/take-a-test-drive-in-the-amd-accelerator-cloud/ba-p/500901>).

¹²⁶ *Id.*

¹²⁷ See “Unlock the Power of AMD Instinct™ GPU Accelerators Across Cloud Providers” (<https://community.amd.com/t5/instinct-accelerators/unlock-the-power-of-amd-instinct-gpu-accelerators-across-cloud/ba-p/729070/jump-to/first-unread-message>).

U.S. Pat. No. 10,333,768 communicate a result of mathematical expression evaluation to a third node; wherein	Accused AMD Server Products core units (“CUs”).¹²⁸ The architecture of the MI300X accelerator is depicted in the following figure:¹²⁹  The diagram illustrates the MI300X accelerator architecture. It features eight compute units, labeled XCD1 through XCD8, arranged in a 2x4 grid. Each unit is represented by a teal square. The units are interconnected by a network labeled 'AMD INFINITY FABRIC' at the bottom right, which includes a circular logo with arrows indicating connectivity. The entire grid is enclosed within a larger teal border labeled 'MI300X' at the top left.
---	--

¹²⁸ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

¹²⁹ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

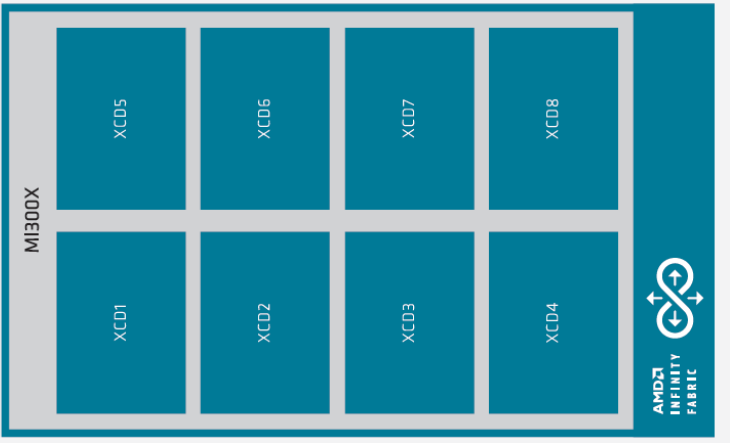
U.S. Pat. No. 10,333,768	Accused AMD Server Products
	For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.¹³⁰ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for a user interface and for a first single-node kernel, for interpreting user instructions, for distributing calls to at least one of a plurality of other MI300X nodes for execution, for receiving the result of mathematical expression from other nodes, for executing mathematical expression evaluations, and for communicating a result of mathematical expression evaluation to other nodes and to the user interface.¹³¹
the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result,	Each of the plurality of Accused AMD Accelerator Product nodes in each accused AMD Server Product comprises a third node comprising a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, to execute at least a second mathematical expression evaluation using the received result, and to communicate the result of the second mathematical expression evaluation to the first node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).¹³² The architecture of the MI300X accelerator is depicted in the following figure:¹³³

¹³⁰ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹³¹ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

¹³² Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

¹³³ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
and communicate the result of the second mathematical expression evaluation to the first node; and	 For example, when executed, code stored in the memory accessible by the XCD of the third MI300X accelerator in a MI300X Platform causes the XCD to receive a result of a first mathematical expression evaluation from a second MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the received result (e.g., by evaluating mathematical expressions using the CUs and/or ROCm™ APIs and/or executing commands from the

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	RCCL)¹³⁴, and to communicate the result of the second mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.¹³⁵ Also, for example, at least because of the architecture of the XCD along with related resources of the ROCm™ Software, the XCD in a third MI300X accelerator in a MI300X Platform is configured to receive the result of a first mathematical expression evaluation from the second MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the received result (e.g., via executable program code, such as RCCL commands), and to communicate the result of the second mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.¹³⁶ The source code for the AMD's sample HelloRccl.cpp program,¹³⁷ includes the details of using RCCL commands to receive the result of mathematical expression evaluation from a second Accused AMD Accelerator Product in an Accused AMD Server Product, to execute at least a second mathematical expression evaluation using the received result, and to communicate the result of the second mathematical expression evaluation to the first Accused AMD Accelerator in the Accused Server Product. A first Accused AMD Accelerator Product in each Accused AMD Server Product is configured to return the result of a second mathematical expression evaluation to the user interface or the script.
wherein the first node is configured to return the	

¹³⁴ See e.g., “RCCL AllReduce” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>”).

¹³⁵ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at [nccIAllReduce](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html) and at [nccIAllGather](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html).

¹³⁶ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

¹³⁷ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
result of the second mathematical expression evaluation to the user interface or the script;	For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform is configured to communicate a result of mathematical expression evaluation¹³⁸ to other nodes using a peer-to-peer architecture¹³⁹ and to a user interface or the script. AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp program includes the details for using RCCL commands to return the result of mathematical expression evaluations from a first MI300X node in a MI300X Platform to the user interface or the script.¹⁴⁰
wherein one or more of the nodes are configured to: accept user instructions; after accepting user instructions, communicate at least some of the user	One or more of the Accused AMD Accelerator Products in an Accused AMD Server Product are configured to: (1) accept user instructions; (2) after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other, and (3) after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.

¹³⁸ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software. *Id.* See also, e.g., “RCCL AllReduce” (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>) (Math Library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/index.html>); see also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture).

¹³⁹ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software. See e.g., “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”).

¹⁴⁰ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>. The RCCL commands cause similar operation of MI200 accelerator nodes. *Id.* See also, <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
instructions using the mechanism for the nodes to communicate with each other; and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.	For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform is configured to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the MI300X accelerator nodes to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels on one or more MI300X accelerator nodes in an MI300X Platform.¹⁴¹ Also, for example, AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp¹⁴² program includes the details for using RCCL commands to cause an Accused AMD Accelerator Product in an Accused AMD Server Product to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the Accused AMD Accelerator Products to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single node-kernels on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.
30. The computer cluster of claim 4, wherein each of the plurality of nodes implements asynchronous calls that enable the single-node kernel to perform computation tasks while the cluster node modules are simultaneously	Each of the Accused AMD Server Products comprises a plurality of nodes, each of which comprises an Accused AMD Accelerator Product. Each of the plurality of nodes implement asynchronous calls that enable the single-node kernel to perform computation tasks while the cluster node modules are simultaneously communicating with one another, for example through the use a peer-to-peer

¹⁴¹ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

¹⁴² See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
communicating with one another.	architecture of (e.g., the “direct connectivity between each of the [MI300X accelerators]” of the MI300X Platform,¹⁴³ along with related RCCL commands¹⁴⁴).¹⁴⁵ The diagram illustrates the AMD Infinity Platform Architecture. It features a central mesh of MI300X OAM Accelerators. These accelerators are connected to a Universal Baseboard Connectors interface. The architecture also includes Fully-meshed 128 GB/s bidirectional Infinity Fabric Connectivity and PCIe Gen 5 x16 I/O connectivity. Additionally, the diagram shows 192 GB HBM3 Memory per OAM, totaling 1.5 TB. AMD provides the ROCm libraries utilizing both collective and point to point commands. For example, AMD provides a sample HelloRccl.cpp program illustrating the use of RCCL commands for the nodes in the system to implement asynchronous calls that enable the single-node kernel to perform

¹⁴³ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

¹⁴⁴ See “AMD Instinct™ MI300X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html>) (confirming MI300X OAM Accelerators support AMD’s ROCm™ software platform); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p.2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>) (confirming MI325X OAM Accelerators support AMD’s ROCm™ software platform). See also e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

¹⁴⁵ See also, “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”); “AMD Instinct™ MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).

U.S. Pat. No. 10,333,768 in view of Accused AMD Server Products

ACS’s Prelim. Infringement Contentions

Exh. A1

February 24, 2025

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	computation tasks while the cluster node modules are simultaneously communicating with one another. ¹⁴⁶
31. The computer cluster of claim 4, wherein intercommunication among the plurality of single-node kernels during thread execution is enabled by the plurality of cluster node modules, and wherein the computer cluster is configured to permit exchange of information between nodes during the course of a parallel computation.	At least because of the architecture of the Accused AMD Accelerator Product nodes along with related resources included in AMD's ROCm™ software, the intercommunication among the plurality of single-node kernels during thread execution is enabled by the plurality of cluster node modules, and the computer cluster is configured to permit exchange of information between nodes during the course of a parallel computation. For example, each Accused AMD Accelerator Product in each Accused AMD Server Product comprises program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions. For example, when executed, the code stored in the memory accessible by an XCD causes the processor to evaluate mathematical expressions (e.g., using the CUs and/or ROCm™ APIs and/or executing commands in the RCCL library). For example, the XCD are configured, when the program code is executed, to evaluate mathematical expressions, for example square root, sine, cosine, and tangent, as well as matrix operations and/or operations between vectors.¹⁴⁷

¹⁴⁶ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

See also, e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); “HIP Stream Management” (https://rocm.docs.amd.com/projects/HIP/en/latest/reference/hip_runtime_api/modules/stream_management.html#), (https://rocm.docs.amd.com/projects/HIP/en/docs-5.5.1/doxygen/html/group__stream.html#gabbfb9f573a6be8c478605ecb5504a74).¹⁴⁷ See, e.g., “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>); see e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at <https://rocm.docs.amd.com/en/docs-5.5.1/api.html>); “ROCm 5.5.1 RCCL API Library” at <https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	For example, the source code for the HelloRccl.cpp program ¹⁴⁸ includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, and for the MI300X accelerator nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture during the course of a parallel computation. ¹⁴⁹
33. The computer cluster of claim 1, wherein the plurality of nodes are configured to permit exchange of information between nodes during the course of parallel computation.	At least because of the architecture of the Accused AMD Accelerator Product nodes along with related resources included in AMD's ROCm™ software, the plurality of nodes are configured to permit exchange of information between nodes during the course of parallel computation. For example, each Accused AMD Accelerator Product in each Accused AMD Server Product comprises program code that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions. For example, when executed, the code stored in the memory accessible by an XCD causes the processor to evaluate mathematical expressions (e.g., using the CUs and/or ROCm™ APIs and/or executing commands in the RCCL library). For example, the XCD are configured, when the program code is executed, to evaluate mathematical expressions, for example square root, sine, cosine, and tangent, as well as matrix operations and/or operations between vectors. ¹⁵⁰

¹⁴⁸ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹⁴⁹ The Hello.Rccl.cpp program includes details for using RCCL commands to similarly initialize a RCCL communicator on a MI200 Platform system comprising multiple MI200 accelerator nodes. *Id.*

RCCL also supports other collective communication functions. See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

¹⁵⁰ See, e.g., “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>); see e.g., “RCCL API Library” at <https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>; “RCCL APL Library” at <https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/API_Reference_Guide.html).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
35. A computer cluster node for evaluating expressions in parallel with other computer cluster	For example, the source code for the HelloRccl.cpp program¹⁵¹ includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, and for the MI300X accelerator nodes to permit exchange of information between the nodes during the course of a parallel computation.¹⁵² Each of the Accused AMD Accelerator Products¹⁵³ (for example, in an Accused AMD Server Product¹⁵⁴) comprises a computer cluster node for evaluating expressions in parallel with other computer cluster nodes. For example, as depicted below and as described further herein, AMD’s INSTINCT™ MI300X Platform (“MI300X Platform”) includes a computer cluster comprising a computer cluster node for evaluating expressions in parallel with other computer cluster nodes.¹⁵⁵

[5.5.1/reference/gpu_libraries/math.html](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html)); “ROCm 5.5.1 RCCL API Library” at ncclAllReduce and at ncclAllGather (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

¹⁵¹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>.

¹⁵² The Hello.Rccl.cpp program includes details for using RCCL commands to similarly initialize a RCCL communicator on a MI200 Platform system comprising multiple MI200 accelerator nodes. *Id.*

RCCL also supports other collective communication functions. See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

¹⁵³ Accused AMD Accelerator Products include, but are not limited to, all products including or relating to AMD’s MI300 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi300.html>) (MI300X, MI300A, MI325X), and MI200 Series Accelerators (<https://www.amd.com/en/products/accelerators/instinct/mi200.html>) (MI250X, MI250, MI210), as well as any products incorporating those items, and any products that are substantially similar to those items.

¹⁵⁴ The Accused AMD Server Products include, but are not limited to, all products including or related to AMD’s MI300X and MI325X Platforms (<https://www.amd.com/en/products/accelerators/instinct/mi300/platform.html>), and all products including or related to AMD INSTINCT™ Accelerator Qualified Servers (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>) incorporating any of the Accused Accelerator Products, as well as any products incorporating those items, and any products that are substantially similar to those items.

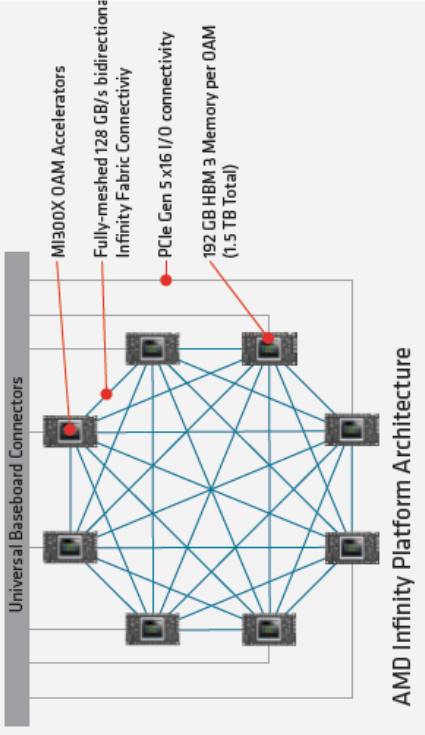
¹⁵⁵ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also MI300 introductory video (“Introducing AMD Instinct™ MI300 Series Accelerators”), at <https://www.youtube.com/watch?v=uVdNpk8tvBI>; see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, as described further herein, each of the MI210, MI250, and MI250X Accelerators (“MI200 accelerators”) is designed to be integrated into at least one AMD INSTINCT™ Accelerator Qualified Server (“MI200 Platform”) such that the MI200 platform incorporates a plurality of MI200 accelerators in a computer cluster comprising a computer cluster node for evaluating expressions in parallel with other computer cluster nodes. See e.g., “AMD Instinct

U.S. Pat. No. 10,333,768	Accused AMD Server Products
nodes, the computer cluster node comprising:	Each of the Accused AMD Server Products includes a plurality of computer cluster nodes, each of which comprises an Accused AMD Accelerator Product. For example, as depicted below, the AMD MI300X Platform comprises eight MI300X accelerators “on an industry-standard universal baseboard” that “are connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”¹⁵⁶

Accelerator-Powered Servers” (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>); “AMD Instinct MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>); “AMD Instinct MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>); “AMD Instinct MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>); “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (Infinity Fabric connecting MI200 accelerators both in-package and inter-package). See also, “AMD INSTINCT™ MI200 Series Accelerator” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>); “AMD INSTINCT™ MI210 Accelerator” (<https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/product-briefs/instinct-mi210-brochure.pdf>).¹⁵⁶ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, AMD specifies a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. See “AMD CDNA™ 2 Architecture,” p. 7 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>). See also, e.g., *id.*, pp. 6 and 8-9 (additional topologies); pp. 3-4 (Infinity Fabric connecting MI200 accelerators both in-package and inter-package).

U.S. Pat. No. 10,333,768	Accused AMD Server Products 
a hardware processor configured to access one or more non-transitory memory devices comprising program code for a single-node kernel that, when executed, causes the hardware processor to interpret user instructions, to evaluate mathematical expressions, and to	Each of the Accused AMD Server Products comprises a plurality of nodes, each of which comprises an Accused AMD Accelerator Product. For example, as depicted below, the MI300X Platform comprises eight MI300X accelerators (e.g., nodes) that “are connected with an AMD Infinity Fabric™ mesh that provides direct connectivity between each of the GPUs over 128 GB/s bidirectional links.”¹⁵⁷

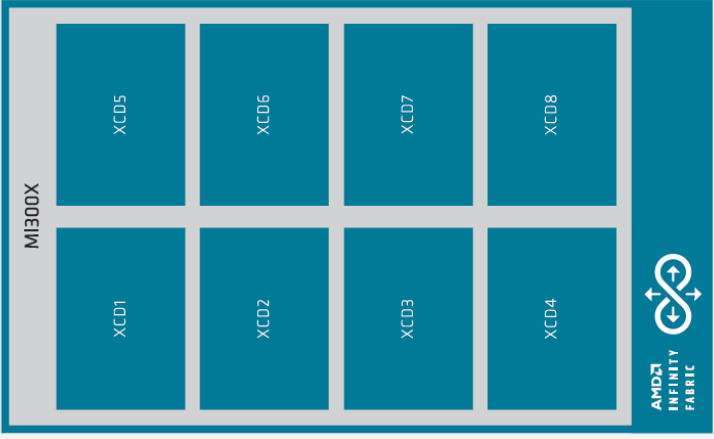
¹⁵⁷ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

Also, for example, the MI200 series accelerators are designed to be integrated into at least one MI200 Platform that incorporates a plurality of MI200 accelerators according to a “Mainstream HPC/ML Topology” that includes four MI200 accelerators connected with an AMD Infinity Fabric™ mesh. See e.g., “AMD Instinct Accelerator-Powered Servers” (<https://www.amd.com/system/files/documents/amd-instinct-server-soln-catalog.pdf>); “AMD CDNA™ 2 Architecture,” p. 7 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>); pp. 3-4 (Infinity Fabric connecting MI200 accelerators both in-package and inter-package). See also, e.g., *id.*, pp. 6 and 8-9 (additional topologies).

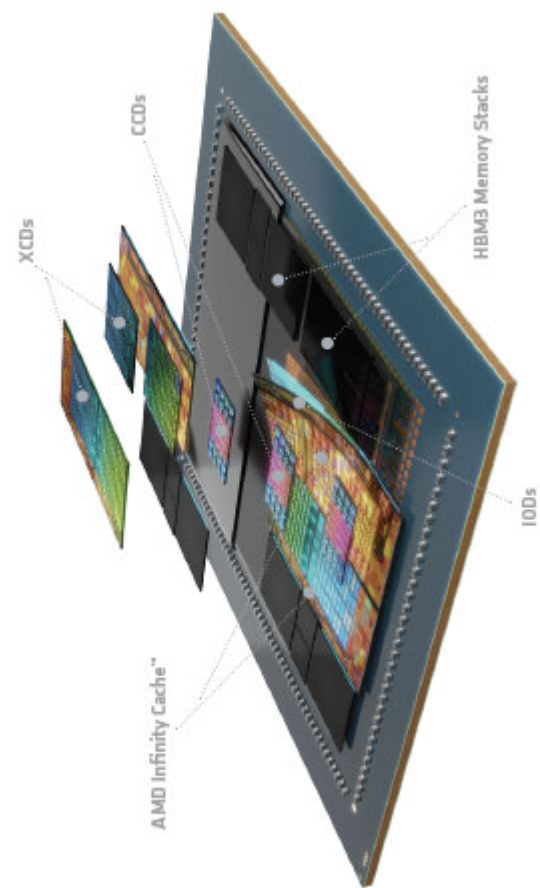
U.S. Pat. No. 10,333,768	Accused AMD Server Products
produce results of mathematical expression evaluation, wherein the hardware processor comprises multiple processor cores; a user connection interface configured to receive a command to start a cluster initialization process for a computer cluster;	Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises a hardware processor configured to access one or more non-transitory memory devices comprising program code for a single-node kernel that, when executed, causes the hardware processor to interpret user instructions to evaluate mathematical expressions, and to produce results of mathematical expression evaluation, wherein the hardware processor comprises multiple processor cores. Specifically, each Accused AMD Accelerator Product comprises a hardware processor that comprises multiple processor cores. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).¹⁵⁸ The architecture of the MI300X accelerator is depicted in the following figure:¹⁵⁹

¹⁵⁸ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

¹⁵⁹ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 Each Accused AMD Accelerator Product comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster of each Accused AMD Service Product. For example, each MI300X accelerator in each MI300X Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes the eight MI300X accelerators in the MI300X Platform.¹⁶⁰ This is

¹⁶⁰ Also, for example, each MI200 accelerator in an MI200 Platform comprises a user connection interface configured to receive a command to start a cluster initialization process for the computer cluster that includes at least three MI200 accelerators.

U.S. Pat. No. 10,333,768	Accused AMD Server Products done, for example, by using AMD’s ROCm™ Collective Communication Library (RCCL), APIs, and math libraries.¹⁶¹ Each AMD Accelerator Product is configured to access a non-transitory computer readable medium (e.g., memory accessible by a XCD, including local memory, shared memory, and HBM memory, as well as memory external to the XCD).¹⁶² Figure 1. Advanced 3D Package and chiplet-based construction of the AMD MI300 Series processors 
--------------------------	--

¹⁶¹ “RCCL documentation” (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) (“RCCL (pronounced ‘Rickel’) is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-all, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations.... [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform.”). RCCL also includes commands that cause similar operation of a MI200 accelerator in an MI200 Platform. *Id.*

¹⁶² “AMD CDNA™ 3 Architecture,” p. 3 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/white-papers/amd-cdna-3-white-paper.pdf>); *Id.* at p. 2 (“The MI300 Series integrates up to 8 vertically stacked accelerator complex dies (XCD) and 4 I/O dies (IOD) containing system infrastructure, all tied together with the AMD Infinity Fabric™ technology and connecting to 8 stacks of high-bandwidth memory (HBM).”). *See also, e.g.*, “AMD CDNA 2 Architecture,” pp. 4-5 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (GCD accessible memory).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	Each Accused AMD Accelerator Product in each Accused AMD Server Product comprises program code for a single-node kernel that, when executed, causes the hardware processor in the Accused AMD Accelerator Product to interpret user instructions, to evaluate mathematical expressions, and to produce result of mathematical expression evaluation (e.g., executable ROCm™ code stored in memory accessible by the XCD in a MI300X node). For example, when executed, the code stored in the memory accessible by an XCD causes the hardware processor to interpret user instructions to evaluate mathematical expressions (e.g., using the CUs and/or executing ROCm™ API and/or commands in the RCCL library), and to produce results of mathematical expression evaluation. For example, the XCD are configured, when the program code is executed, to evaluate mathematical expressions, for example square root, sine, cosine, and tangent, as well as matrix operations and/or operations between vectors.¹⁶³ Each XCD in a MI300X accelerator of the MI300X Platform, is configured to access one or more non-transitory memory devices comprising program code for a single-node kernel that, when executed causes the XCD to produce results of mathematical express evaluation. A first single-node kernel is responsible for interpreting user instructions and distributing call to at least one of the other MI300X accelerators for execution. The other MI300X accelerators in the cluster are responsible for executing the tasks that are distributed to them by the first-single-node kernel. When the user submits a job to the cluster, the first single-node kernel parses the job and distributes the tasks to the other nodes for execution. The other nodes then execute the tasks and return the results to the first single-node kernel or communicate the results of mathematical expression evaluation to other nodes. The first single-

¹⁶³ See, e.g., “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see, e.g., “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>); see e.g., “RCCL API Library” at nclAllReduce (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also “RCCL APL Library” at nclAllGather (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at nclAllReduce and at nclAllGather (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	node kernel then collects the results from the other nodes and returns them to the user. Thus, the XCD in a first MI300X accelerator is configured to access executable program code stored in memory accessible by the XCD, where the executable program code is program code for a single-node kernel that, when executed, causes the processor to interpret user instructions. ¹⁶⁴
a mechanism to communicate results of evaluation with other computer cluster nodes using a peer-to-peer architecture; and	Each of the Accused AMD Accelerator Products in each Accused AMD Server Product comprises a mechanism for the Accused AMD Accelerator Products to communicate results of evaluation ¹⁶⁵ with other Accused AMD Accelerator Products using a peer-to-peer architecture (e.g., the “direct connectivity between each of the [MI300X accelerators]” of the MI300X Platform, ¹⁶⁶ along with related RCCL commands ¹⁶⁷). ¹⁶⁸

¹⁶⁴ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable as described herein with respect to the XCDs in MI200X accelerators.

¹⁶⁵ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library), (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>) (Math Library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>), (<https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/index.html>); see also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture).
¹⁶⁶ “AMD INSTINCT™ MI300X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-platform-data-sheet.pdf>); see also “AMD INSTINCT™ MI325X Platform™ Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-platform-datasheet.pdf>).

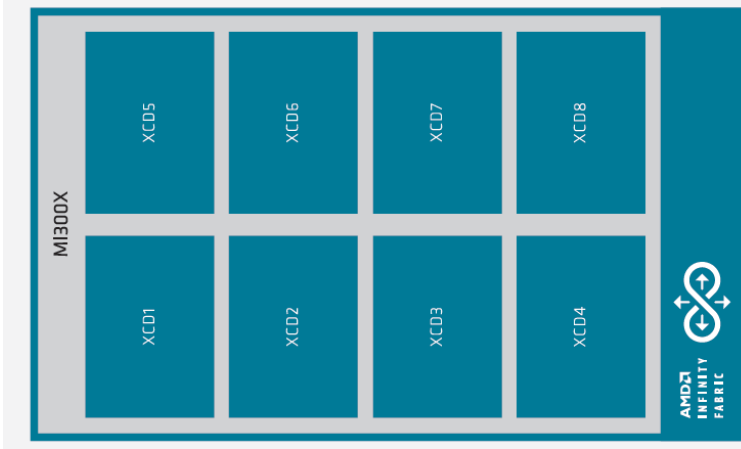
¹⁶⁷ See “AMD Instinct™ MI300X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html>) (confirming MI300X OAM Accelerators support AMD’s ROCm™ software platform); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>) (confirming MI325X OAM Accelerators support AMD’s ROCm™ software platform).

¹⁶⁸ See also, “AMD INSTINCT™ MI200 Series Accelerator Data Sheet” (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/instinct-mi200-datasheet.pdf>) (“peer-to-peer I/O connectivity”); “AMD Instinct™ MI250X Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250x.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); see also, “AMD Instinct™ MI250 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>) (confirming MI250 accelerators support AMD’s ROCm™ software platform); “AMD Instinct™ MI210 Accelerators” (<https://www.amd.com/en/products/accelerators/instinct/mi200/mi210.html>) (confirming MI210 accelerators support AMD’s ROCm™ software platform).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	Universal Baseboard Connectors MI300X OAM Accelerators Fully-meshed 128 GB/s bidirectional Infinity Fabric Connectivity PCIe Gen 5x16 I/O connectivity 192 GB HBM3 Memory per OAM (1.5 TB Total) AMD Infinity Platform Architecture For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.¹⁶⁹ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, and for the MI300X accelerator nodes to communicate results of evaluation with other MI300X accelerator nodes using a peer-to-peer architecture.¹⁷⁰
program code that, when executed, is capable of causing the hardware processor to: receive calls from a second node comprising a second hardware processor configured to access a second memory	Each of the plurality of Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprises program code that, when executed, is capable of causing the hardware processor to: receive calls from a second node comprising a second hardware processor configured to access a second memory comprising program code for a user interface and program code for a second single-node kernel, the second single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple

¹⁶⁹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹⁷⁰ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

U.S. Pat. No. 10,333,768 comprising program code for a user interface and program code for a second single-node kernel, the second single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution;	Accused AMD Server Products core units (“CUs”).¹⁷¹ The architecture of the MI300X accelerator is depicted in the following figure:¹⁷²  The diagram illustrates the MI300X accelerator architecture. It features eight compute dies, labeled XCD1 through XCD8, arranged in a 2x4 grid. The dies are connected by an AMD Infinity Fabric, which is represented by a blue background with a white infinity symbol and the text 'AMD INFINITY FABRIC'. The label 'MI300X' is positioned at the top left of the grid.
--	--

¹⁷¹ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3–4, 7–9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

¹⁷² “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3–4, 7–9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	For example, the RCCL includes commands that cause a XCD of a MI300X accelerator in a MI300X Platform receive calls from a second MI300X accelerator in the MI300X Platform comprising a second MI300X XCD configured to access a second memory comprising program code for a user interface and program code configured to interpret user instructions and distribute calls to at least one of the other MI300X accelerator nodes in the MI300X Platform for execution.¹⁷³ The source code for AMD's sample HelloRccl.cpp program¹⁷⁴ includes using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for a user interface and program code configured to interpret user instructions and distribute calls to at least one of the other MI300X nodes in the MI300X Platform.¹⁷⁵ AMD offers its AMD Accelerator Cloud ("AAC") that "is a private cloud environment that enable customers to remotely access and use AMD EPYC™ CPUs, AMD Instinct™ GPUs and AMD ROCm™ technologies to experience the power of AMD hardware and software — along with all the possibilities that come with it. The AMD Accelerator Cloud makes it easy for prospective buyers to test and evaluate AMD Instinct accelerators. . . ."¹⁷⁶ "After being given access to the AMD Accelerator Cloud, users will be able to request access to a variety of node configurations with a range of AMD Instinct accelerator options. The user can select the desired operating system, ROCm version and directly install their preferred applications for testing or development. As we expand the AAC in the coming months, customers will be able to access a broader range of system types and GPUs, including the latest MI200 series accelerators."¹⁷⁷

¹⁷³ "RCCL documentation" (<https://rocm.docs.amd.com/projects/rccl/en/latest/index.html>) ("RCCL (pronounced 'Rickle') is a stand-alone library that provides multi-GPU and multi-node collective communication primitives optimized for AMD GPUs. It implements routines such as all-reduce, all-gather, reduce, broadcast, reduce-scatter, gather, scatter, all-to-allv, and all-to-all as well as direct point-to-point (GPU-to-GPU) send and receive operations.... [RCCL] supports an arbitrary number of GPUs installed in a single-node or multi-node platform."). RCCL also includes commands that cause similar operation of the GCD of a MI200 accelerator in an MI200 Platform. *Id.*

¹⁷⁴ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹⁷⁵ *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes. *Id.*

¹⁷⁶ See "Take a Test Drive in the AMD Accelerator Cloud" (<https://community.amd.com/t5/instinct-accelerators/take-a-test-drive-in-the-amd-accelerator-cloud/ba-p/500901>).

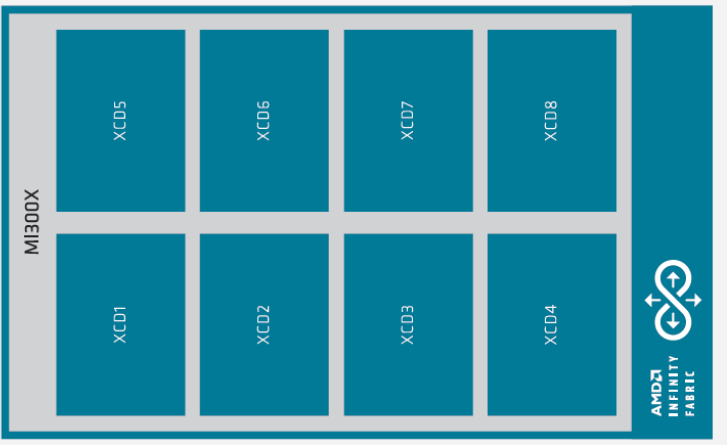
¹⁷⁷ *Id.*

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	AMD advertises that its AMD Accelerator Product can be accessed through the cloud through other leading clouding providers, and advertises that such configurations can use AMD Instinct™ MI200 and MI300 series accelerators.¹⁷⁸ The computer cluster of each Accused AMD Accelerator Product includes program code to provide the user interface to allow users, including AMD’s own employees, to access the computer cluster through, at least, the cloud. The program code to provide the user interface of the Accused AMD Accelerator Product could also allow a user to connect to the computer cluster of each Accused AMD Accelerator Product through a direct connection.
execute, using the hardware processor, at least a first mathematical expression evaluation; and communicate a result of the first mathematical expression evaluation to a third node comprising	Each Accused AMD Server Product comprises program code that, when executed, is capable of causing a hardware processor to execute, using the hardware processor, at least a first mathematical expression evaluation and communicate a result of the first mathematical expression evaluation to a third node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).¹⁷⁹ The architecture of the MI300X accelerator is depicted in the following figure:¹⁸⁰

¹⁷⁸ See “Unlock the Power of AMD Instinct™ GPU Accelerators Across Cloud Providers” (<https://community.amd.com/t5/instinct-accelerators/unlock-the-power-of-amd-instinct-gpu-accelerators-across-cloud/ba-p/729070/jump-to/first-unread-message>).

¹⁷⁹ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

¹⁸⁰ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 For example, AMD provides a sample HelloRccl.cpp program utilizing both collective and point to point RCCL commands.¹⁸¹ Specifically, the source code for the HelloRccl.cpp program includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes, accessing program code for executing mathematical expression evaluations, and for communicating a result of mathematical expression evaluations to other nodes.¹⁸²

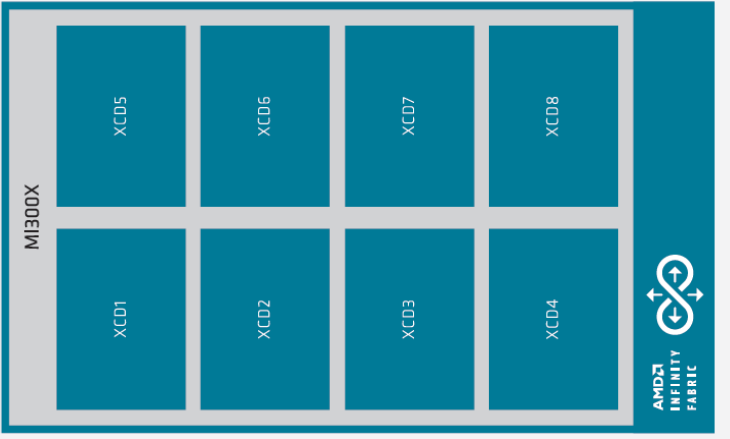
¹⁸¹ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹⁸² *Id.* The RCCL commands cause similar initialization and operation of MI200 accelerator nodes.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the computer cluster node, to execute at least a second mathematical expression evaluation using the result of the first mathematical expression evaluation, and to communicate a result of the second mathematical expression evaluation to the first node;	Each Accused AMD Server Product comprises a third node comprising a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the computer cluster node, to execute at least a second mathematical expression evaluation using the result of the first mathematical expression evaluation, and to communicate the result of the second mathematical expression evaluation to the first node. Each Accused AMD Accelerator Product comprises a hardware processor. For example, each MI300X accelerator is a node that includes an accelerator complex die (“XCD”) that contains multiple core units (“CUs”).¹⁸³ The architecture of the MI300X accelerator is depicted in the following figure:¹⁸⁴

¹⁸³ Also, for example, each MI200 accelerator is a node that includes an MI200 Graphics Compute Die (“GCD”) that contains multiple compute units (CUs). See e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>).

¹⁸⁴ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf>) (“Each OAM module includes: • Eight accelerated compute dies (XCDs) with 38 compute units (CUs), 32 KB of L1 cache per CU, 4 MB shared L2 cache shared across CUs, and 256 MB of AMD Infinity Cache™ shared across 8 XCDs. The compute units support a broad range of precisions for both AI/ML and HPC acceleration, native hardware support for sparsity, and enhanced computational throughput.”); see also “AMD INSTINCT™ MI325X ACCELERATOR Data Sheet,” p. 2 (<https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/product-briefs/instinct-mi325x-datasheet.pdf>). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	 For example, when executed, code stored in the memory accessible by the XCD of the third MI300X accelerator in a MI300X Platform causes the XCD to receive a result of a first mathematical expression evaluation from a first MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the result of the first mathematical expression evaluation (<i>e.g.</i>, by evaluating mathematical expressions using the CUs and/or ROCm™ APIs and/or

U.S. Pat. No. 10,333,768	Accused AMD Server Products executing commands from the RCCL)¹⁸⁵, and to communicate the result of the mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.¹⁸⁶ Also, for example, at least because of the architecture of the XCD along with related resources of the ROCm™ Software platform, the XCD in a third MI300X accelerator in a MI300X Platform is configured to receive the result of a first mathematical expression evaluation from the first MI300X accelerator in the MI300X Platform, to execute at least a second mathematical expression evaluation using the result of the first mathematical expression evaluation (e.g., via executable program code, such as RCCL commands), and to communicate the result of the second mathematical expression evaluation to the first MI300X accelerator in the MI300X Platform.¹⁸⁷ The source code for the AMD's sample HelloRccl.cpp program,¹⁸⁸ includes the details of using RCCL commands to receive the result of first mathematical expression evaluation from a first Accused AMD Accelerator Product in an Accused AMD Server Product, to execute at least a second mathematical expression evaluation using the result of the first mathematical expression evaluation, and to communicate the result of the second mathematical expression evaluation to the first Accused AMD Accelerator in the Accused Server Product.
--------------------------	--

¹⁸⁵ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>”).

¹⁸⁶ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at [nccIAllReduce](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html) and at [nccIAllGather](https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html).

¹⁸⁷ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

¹⁸⁸ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
wherein the user connection interface is configured to return at least one result of mathematical expression evaluation to a user interface or a script;	A user connection interface in a first Accused AMD Accelerator Product in each Accused AMD Server Product is configured to return at least one result of mathematical expression evaluation to a user interface or a script. For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform comprises a user connection interface that is configured to return at least one result of mathematical expression evaluation¹⁸⁹ to a user interface or a script. AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp program includes the details for using RCCL commands to implement a user connection interface configured to return at least one result of mathematical expression evaluation from a MI300X node in a MI300X Platform to the user interface or a script.¹⁹⁰
wherein the computer cluster node is configured to: accept user instructions; after accepting user instructions, communicate at least some of the user instructions using the	One or more of the Accused AMD Accelerator Products in an Accused AMD Server Product are configured to: (1) accept user instructions; (2) after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other, and (3) after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to the single-node kernel on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.

¹⁸⁹ See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software. *Id.* See also, e.g., “RCCL AllReduce” (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>); see also, “ROCm libraries” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>) (Math Library); see also, “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/index.html>); see also, “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) compatible with CDNA2 architecture).<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>.

¹⁹⁰ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl/HelloRccl.cpp>. The RCCL commands cause similar operation of MI200 accelerator nodes. *Id.* See also, <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
mechanism for the nodes to communicate with each other; and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to the single-node kernel.	For example, at least because of the architecture of the XCD along with related resources included in AMD's ROCm™ software, the XCD in a first MI300X accelerator node in a MI300X Platform is configured to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the MI300X accelerator nodes to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels on one or more MI300X accelerator nodes in an MI300X Platform.¹⁹¹ Also, for example, AMD provides a sample RCCL program, HelloRccl.cpp, which demonstrates both collective and point-to-point communication. The source code for the HelloRccl.cpp¹⁹² program includes the details for using RCCL commands to cause an Accused AMD Accelerator Product in an Accused AMD Server Product to accept user instructions, after accepting user instructions, communicate at least some of the user instructions using the mechanism for the Accused AMD Accelerator Products to communicate with each other, and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single node-kernels on one or more Accused AMD Accelerator Products in an Accused AMD Server Product.
37. The computer cluster node of claim 35, wherein the computer cluster node is configured to permit exchange of information with other computer cluster nodes during the course of parallel computation.	The Accused AMD Accelerator Product nodes in each Accused AMD Server Product comprise one or more cluster nodes. At least because of the architecture of the Accused AMD Accelerator Product nodes with related resources included in AMD's ROCm™ software, each of the cluster nodes of the Accused AMD Server Products is configured to permit exchange of information between nodes during the course of parallel computation. For example, each Accused AMD Accelerator Product in each Accused AMD Server Product comprises program code that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions. For example, when executed, the code stored in the memory accessible by an XCD causes the processor to evaluate mathematical expressions (<i>e.g.</i>, using the CUs and/or

¹⁹¹ The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable, at least because of their architecture along with related resources of the ROCm™ software.

¹⁹² See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

U.S. Pat. No. 10,333,768	Accused AMD Server Products
	ROCm™ APIs and/or executing commands in the RCCL library). For example, the XCD are configured, when the program code is executed, to evaluate mathematical expressions, for example square root, sine, cosine, and tangent, as well as matrix operations and/or operations between vectors.¹⁹³ For example, the source code for the HelloRccl.cpp program¹⁹⁴ includes the details of using RCCL commands to initialize a RCCL communicator on a MI300X Platform comprising multiple MI300X accelerator nodes among a plurality of cluster node modules, and permit exchange of information between the cluster node modules during the course of a parallel computation.¹⁹⁵
39. The computer cluster node of claim 35, wherein the hardware processor	The Accused AMD Accelerator Product nodes include XCD or GCD processors that are GPUs. ¹⁹⁶ A GPU is a special purpose microprocessor.

¹⁹³ See, e.g., “ROCm libraries” (<https://rocm.docs.amd.com/en/latest/reference/api-libraries.html#math-apis>) (Math half library); “rocBLAS” (<https://rocm.docs.amd.com/projects/rocBLAS/en/latest/index.html>); see e.g., “RCCL API Library” at nccAllReduce (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>); “RCCL APL Library” at nccAllGather (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>).

The GCDs in MI200 accelerators of an MI200 Platform are similarly configured and operable. See e.g., “ROCm Compatibility Matrix” (<https://rocm.docs.amd.com/en/latest/compatibility/compatibility-matrix.html>) (ROCm 6.2.1 compatible with CDNA2 architecture). See also, e.g., “ROCm Software Groups” (<https://rocm.docs.amd.com/en/docs-5.5.1/reference/all.html>); “rocBLAS API Reference Guide” (https://rocm.docs.amd.com/projects/rocBLAS/en/docs-5.5.1/API_Reference_Guide.html); “Math Libraries,” (https://rocm.docs.amd.com/en/docs-5.5.1/reference/gpu_libraries/math.html); “ROCm 5.5.1 RCCL API Library” at nccAllReduce and at nccAllGather (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

¹⁹⁴ See <https://github.com/ROCm/rccl/blob/release/rocm-rel-6.2/tools/HelloRccl.cpp>; <https://github.com/ROCm/rccl/blob/release/rocm-rel-5.1/tools/HelloRccl/HelloRccl.cpp>.

¹⁹⁵ The HelloRccl.cpp program includes details for using RCCL commands to similarly initialize a RCCL communicator on a MI200 Platform system comprising multiple MI200 accelerator nodes. *Id.*

RCCL also supports other collective communication functions. See e.g., “RCCL API Library” (<https://rocm.docs.amd.com/projects/rccl/en/latest/api-reference/api-library.html#>), (<https://rocm.docs.amd.com/projects/rccl/en/docs-5.5.1/api.html>).

¹⁹⁶ “AMD INSTINCT™ MI300X ACCELERATOR Data Sheet,” p.2, <https://www.amd.com/content/dam/amd/en/documents/instinct-tech-docs/data-sheets/amd-instinct-mi300x-data-sheet.pdf> (architecture for MI300X accelerators); AMD Instinct™ Accelerators <https://www.amd.com/en/products/accelerators/instinct.html> (linking to spec sheets for MI325X and MI300A chips). See also, e.g., “AMD CDNA™ 2 Architecture,” pp. 3-4, 7-9 (<https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf>) (architecture for MI200 accelerators).

U.S. Pat. No. 10,333,768 comprises a special purpose microprocessor.	Accused AMD Server Products
--	-----------------------------