UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

MERCEDES-BENZ USA, LLC,

Petitioner,

v.

DAEDALUS PRIME LLC,

Patent Owner.

---

Case No.: IPR2023-01344

U.S. Patent No. 9,575,895

---

# PATENT OWNER'S RESPONSE

## TABLE OF CONTENTS

**I. INTRODUCTION AND SUMMARY OF ARGUMENTS .. 1**

**II. BACKGROUND, THE '895 PATENT, AND CLAIM CONSTRUCTION ..... 2**

**A. Background ..... 2**

**B. The '895 Patent ..... 3**

**C. Claim Construction and the Level of Ordinary Skill in the Art ..... 5**

**III. THE PETITION FAILS TO DEMONSTRATE THAT ANY CHALLENGED CLAIM IS UNPATENTABLE ..... 10**

**A. The Challenged Claims are Not Unpatentable in View of Sinharoy ..... 14**

1. Overview of Sinharoy ..... 14

2. Claim 1 is not unpatentable in view of Sinharoy ..... 15

3. Claims 11 and 15 are not unpatentable in view of Sinharoy ..... 23

4. The respective dependent claims are not unpatentable in view of Sinharoy ..... 23

5. The further challenges relying on Sinharoy are also unavailing ..... 23

*a) Petitioner has failed to demonstrate that the Xu reference is prior art to the '895 patent* ..... 24

**B. The Challenged Claims are Not Unpatentable in View of Tang ..... 27**

1. Overview of Tang ..... 27

2. Independent claims 1, 11, and 15 are not unpatentable in view of Tang ..... 29

3. The respective dependent claims are not unpatentable in view of Tang ..... 33

**IV. CONCLUSION ..... 33**

## TABLE OF AUTHORITIES

### CASES

*Acceleration Bay LLC v. Activision Blizzard, Inc.*,
908 F.3d 765 (Fed. Cir. 2018)..... 25

*Ancora Technologies Inc. v. Apple Inc.*,
744 F.3d 732 (Fed. Cir. 2014)..... 8

*CFMT, Inc. v. Yieldup Intern. Corp.*,
349 F.3d 1333 (Fed. Cir. 2003)..... 22, 32

*Dynamic Drinkware, LLC v. Nat'l Graphics, Inc.*,
800 F.3d 1375 (Fed. Cir. 2015)..... 22

*Harmonic Inc. v. Avid Tech., Inc.*,
815 F.3d 1356 (Fed. Cir. 2016)..... 22

*Hartness Int'l. Inc. v. Simplimatic Engineering Co.*,
819 F.2d 1100 (Fed. Cir. 1987)..... 23, 24, 33

*In re Gordon*,
733 F.2d 900 (Fed. Cir. 1984)..... 12

*In re Klopfenstein*,
380 F.3d 1345 (Fed. Cir. 2004)..... 25

*In re Magnum Oil Tools Int'l, Ltd.*,
829 F.3d 1364 (Fed. Cir. 2016)..... 22

*In re Peterson*,
315 F.3d 1325 (Fed. Cir. 2003)..... 12

*In re Ratti*,
270 F.2d 810 (CCPA 1959) ..... 11

*In re Warner*,
379 F.2d 1011 (CCPA 1967) ..... 11

*Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.*,
821 F.3d 1359 (Fed. Cir. 2016)..... 11

*Jazz Pharm., Inc. v. Amneal Pharm., LLC*,
895 F.3d 1347 (Fed. Cir. 2018)..... 25

*KSR Int'l v. Teleflex Inc.*,
550 U.S. 398 (2007)..... 11, 12

*Phillips v. AWH Corp.*,
415 F.3d 1303 (Fed. Cir. 2005)..... 6, 8

*SAS Inst. Inc. v. Iancu*,
138 S. Ct. 1348 (2018)..... 10, 11

*Schering Corp. v. Geneva Pharm.*,
339 F.3d 1373 (Fed. Cir. 2003)..... 22, 32

*Thorner v. Sony Comput. Ent. Am. LLC*,
669 F.3d 1362 (Fed. Cir. 2012)..... 5

*Xerox Corp. v. Bytemark, Inc.*,
IPR2022-00624 (PTAB Aug. 24, 2022) ..... 18

**STATUTES**

35 U.S.C. § 102(a)..... 24

35 U.S.C. § 311(b)..... 24

**REGULATIONS**

37 C.F.R. § 42.108(c) ..... 22

37 C.F.R. § 42.65(a) ..... 18

**OTHER AUTHORITIES**

*Changes to the Claim Construction Standard for Interpreting Claims in Trial Proceedings Before the Patent Trial and Appeal Board*, 83 Fed. Reg. 51,340 (Oct. 11, 2018) ..... 5

PTAB Consolidated Trial Practice Guide (Nov. 2019)..... 18, 30

## EXHIBIT LIST

| EXHIBIT NO. | DESCRIPTION |
|---|---|
| 2001 | Transcript of Deposition of Dr. Horst |
| 2002 | Declaration of Michael C. Brogioli, Ph.D. |

## I. INTRODUCTION AND SUMMARY OF ARGUMENTS

Mercedes-Benz USA LLC. ("MB USA") challenges the patentability of claims 1-17 of U.S. Patent No. 9,575,895 (the "'895 patent") fail to demonstrate that these challenged claims are unpatentable.

First, Sinharoy does not teach or suggest a single, that is a solitary, caching agent for the processor, as required by the challenged claims. Instead, Sinharoy describes a processor architecture that includes eight processor cores and a multi-level cache structure. Ex. 1003 at 1. Each core has its own level 1 (L1), level 2 (L2), *id*. at 11, and region of a level 3 (L3) cache. *Id*. at 16-17. Each L2 cache and each L3 cache region has its own controller. *Id*. at 18-19. In the case of cache misses, each separate controller must snoop memory accesses by another controller. *Id*. at 18. In other words, no one cache controller is or acts as *a single*—that is, a solitary—*caching agent for the processor*, as recited by the challenged independent claims.

Second, the petition fails to demonstrate that Tang teaches or suggests a single caching agent for a processor, and certainly not one that includes a plurality of distributed portions *each associated with a corresponding one of the plurality of cores*, as further required by the challenged claims. Tang itself shows, in Figure 10, a pair of cache coherence controllers which communicate with one another via a "coherence link." The presence of and

need for such a coherence link belies any conclusion that the two cache controllers are a single logical unit. If that were the case, Tang would have said so, or at least annotated the illustration of the cache architecture to imply as much. The fact that Tang instead demands a separate coherence link indicates these units are operating independently of one another.

Further, in Tang's design the last level cache ways are shared across multiple cores of the processor and so, contrary to the allegations in the petition, it is not true that the cache controller has partitioned regions or units each associated with a given core. Different parts of the last level cache controller may be associated with different ways of the last level cache, but those ways are not each associated with a given core of the processor and so neither are their associated cache controllers.

These and further reasons why the claims are not unpatentable in view of the cited references are discussed in greater detail below.

## II. BACKGROUND, THE '895 PATENT, AND CLAIM CONSTRUCTION

### A. Background

Modern computer systems include many different components and over time functionalities previously provided by separate integrated circuit ("IC") devices, including input/output ("I/O") functions, have steadily

migrated so as to be incorporated on a single semiconductor die. Ex. 1001 at 1:11-20. While there are advantages achieved in such integrations, complications also arise. For example, whereas IO components were traditionally associated with their own caching agents, meaning that dedicated logic was available to handle cache coherency operations due to IO transactions, the inclusion of such separate caching agents on a single die as IO functions are moved on-chip presents performance challenges for a multiprocessor system. *Id*. at 1:21-36.

### B. The '895 Patent

The '895 patent describes a solution for particular "problems [that] arise once an IO component is integrated on the same chip with a multiprocessor." *Id*. at 1:25-26. In such circumstances,

> Traditional IO integration treats the IO component as a separate caching agent, meaning that dedicated logic is associated with the IO component to handle cache coherency operations. When an IO agent is performing read/write operations to main memory, it has to snoop the CPU side cache to maintain cache coherency. In MP systems, this becomes a major scaling problem.

*Id*. at 1:26-33. The solution described in the '895 patent employs "a CPU caching agent [ ] configured to support both CPU traffic and IO traffic as

well, thereby resolving the scaling issue." *Id*. at 1:62-64; *and see id*. at 2:24-26, 64-66; 3:17-19. The single caching agent can "be distributed such that each of the different portions of the caching agent can be associated with a corresponding core and LLC [(last level cache)] bank or slice." *Id*. at 3:7-10; *and see id*. at 6:32-39.

The use of a single caching agent to support both CPU and IO traffic is captured in each of the independent claims of the '895 patent. For example, claim 1 recites,

> A processor comprising:
>
> a plurality of cores, a shared cache memory, a memory controller to interface with a memory coupled to the processor, an integrated input/output (IO) module to interface between the processor and an IO device coupled to the processor and a caching agent to perform cache coherency operations for the plurality of cores and the IIO module, wherein the processor is to receive an allocation transaction from the IO device and directly store data of the allocation transaction into the shared cache memory, *wherein the caching agent is a single caching agent for the processor* and includes a plurality of distributed portions each associated with a corresponding one of the plurality of cores.

*Id*. at 7:48-61 (emphasis added); *and see id*. at 8:36-41 (claim 8, reciting a "single caching agent," shared by a plurality of cores of a multicore processor and an IO module, and that includes a "plurality of distributed portions each associated with a corresponding one of the plurality of cores"), 9: (claim 15, reciting a "first caching agent" to perform cache coherency operations for a first multicore processor that includes a plurality of cores and a first IO module, and that is a "single caching agent" for the first multicore processor and includes a plurality of distributed portions each associated with a corresponding one of the plurality of cores).

### C. Claim Construction and the Level of Ordinary Skill in the Art

In *inter partes* reviews such as this one, the Board applies the same claim construction standard used to construe a claim in a civil action under 35 U.S.C. § 282(b). *Changes to the Claim Construction Standard for Interpreting Claims in Trial Proceedings Before the Patent Trial and Appeal Board*, 83 Fed. Reg. 51,340 (Oct. 11, 2018). Accordingly, the claims should be construed consistent with their plain and ordinary meaning as understood by a person of ordinary skill in the art when read in the context of the specification and prosecution history. *Thorner v. Sony Comput. Ent. Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012) (citing *Phillips v. AWH Corp.*,

415 F.3d 1303, 1312–13 (Fed. Cir. 2005) (en banc)). "The ordinary and customary meaning of a claim term is the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention." *Phillips* at 1313.

In its Institution Decision, the Board accepted MB USA's proposed definition of a person of ordinary skill in the art as someone with:

> a good working knowledge of computer science or electrical engineering and multicore processor systems. She would have gained this through an undergraduate Bachelor's degree in computer science or electrical engineering, or a comparable field, as well as some experience in the field (*e.g.*, in multicore processor systems). (*Id.*) The more education one has, the less experience needed to attain an ordinary level of skill. (*Id.*) Similarly, more experience in the field may serve as a substitute for formal education. (*Id.*)

Paper 7 at 10-11 (citing Pet. at 8). This definition departs from that offered by Petitioner's own declarant, Dr. Horst, see Ex. 1007 at ¶ 30, who proposed a more precise description of the kind of experience (both in terms of number of years and the prescribed field(s)) such an individual would have. For purposes of this Response, Patent Owner adopts the definition of a person of ordinary skill in the art proposed by Dr. Horst, *see* Ex. 2002 at ¶

25, and urges the Board to do likewise in place of the more nebulous definition proposed in the Petition.

As noted above, each of the independent claims of the ’895 patent recites a “single caching agent,” either “for the processor” (claim 1), “for the multicore processor” (claim 11), or “for the first multicore processor” (claim 15). In the Institution Decision, the Board construed the phrase, “single caching agent” as meaning “one caching agent for the processor, which may be formed from distributed portions.” Paper 7 at 12. Patent Owner urges the Board to reconsider this construction as part of any Final Written Decision inasmuch as it introduces unnecessary obstacles to understanding the claims and conflicts with the understanding of the claims that a person of ordinary skill in the art considering the ’895 patent would arrive at. Ex. 2002 at ¶¶ 30-31.

The ’895 patent uses the term “single” consistent with its “plain and ordinary meaning” of indicating only *one*, and not, for example, one of many. *Id.* at ¶ 30; *and see, e.g.*, Ex. 1001 at 1:62-64 (“***a*** CPU caching agent can be configured to support both CPU traffic and IO traffic as well); 2:24-26, (“In various embodiments, ***this*** caching agent may be ***a combined*** caching agent both for the CPU as well as for an integrated IO agent.”), 64-66 (“[V]ia caching agent 200, cores of the processor and an IIO module can

maintain coherency ***without the need for additional caching agents***.”); 3:16-19 (“***a single caching agent*** per multicore processor socket, where each socket includes multiple cores and an IIO module.”); 6:30-39 (“Note that while distributed caching agents are shown, understand that these distributed portions ***form a single caching agent***, and which is configured to handle cache coherency operations both for the cores as well as an IIO module 750.”); 7:13-17 (“***a single caching agent*** shared by the IIO module and the cores of the processor.”) (emphases added). Thus, understanding each of the independent claims as reciting a solitary, that is one, caching agent for an entire processor (or multicore processor or first multicore processor), both fully comports with the way in which the term is used in the subject patent and the application of widely accepted meanings of commonly understood words. *See Phillips*, 415 F.3d at 1314; *Ancora Technologies Inc. v. Apple Inc.*, 744 F.3d 732, 734 (Fed. Cir. 2014) (“A claim term should be given its ordinary meaning in the pertinent context, unless the patentee has made clear its adoption of a different definition or otherwise disclaimed that meaning.”).

As the Board correctly noted, when referring to Figure 6, the specification explains that the “single caching agent” may be formed from distributed portions. Paper 7 at 12 (referencing Ex. 1001 at 6:35–37). But

this circumstance is already accounted for in claims 1, 11, and 15, each of which recite, "wherein the caching agent is a single caching agent for the [processor/multicore processor/first multicore processor] *and includes a plurality of distributed portions each associated with a corresponding one of the [first] plurality of cores*." Ex. 1001 at 7:58-61; 8:37-39; 9:6-10 (emphases added). Accordingly, it is not necessary to account for such a circumstance as part of the construction of "single caching agent" and a person of ordinary skill in the art would appreciate as much. Ex. 2002 at ¶ 31. Adopting the Board's initial proposed construction would have the claims read as "wherein the caching agent is one caching agent for the [processor/multicore processor/first multicore processor], which may be formed from distributed portions and includes a plurality of distributed portions each associated with a corresponding one of the [first] plurality of cores." Such a construction imparts unnecessary confusion into the meaning of the claim. For example, are the distributed portions that form the one caching agent the same as the plurality of distributed portions each associated with a corresponding one of the plurality of cores, or are they different distributed portions? If they are different, then where is such an embodiment described in the specification? And, if they are the same, why is it necessary to indicate the caching agent is formed of those distributed

portions and also includes them, and how can that be? And, if the one caching agent is already said to include a plurality of distributed portions, why is only possible ("may") the one caching agent might be formed from them?

To avoid such unneeded complications, the Board should revise its initial construction and simply construe the phrase "single caching agent" as meaning *one* caching agent, *i.e.*, one caching agent "for the processor" (claim 1), one caching agent "for the multicore processor" (claim 11), or one caching agent "for the first multicore processor" (claim 15), as appropriate. The claims already specify that the one caching agent includes a plurality of distributed portions, as explained with reference to Fig. 6, so such a circumstance need not be read into the construction of the "single caching agent."

## III. THE PETITION FAILS TO DEMONSTRATE THAT ANY CHALLENGED CLAIM IS UNPATENTABLE

A petition must be judged as it is written. The Board may not overlook flaws in the petition or reformulate its challenges to enhance their suitability. *See SAS Inst. Inc. v. Iancu*, 138 S. Ct. 1348, 1355 (2018) ("[I]t's the petitioner, not the Director, who gets to define the contours of the

proceeding."); *see also id.* ("[T]he petitioner is master of its complaint . . ."); *see also id.* at 1353 (forbidding the Board from "curat[ing]" the petition). Furthermore, because all of the challenges in the petition rely on conclusions of obviousness, the analyses set forth in the petition must include "some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness." *KSR Int'l v. Teleflex Inc.*, 550 U.S. 398, 418 (2007) (*quoting In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006)). The challenger must supply the factual basis for the challenge and "may not resort to speculation, unfounded assumptions or hindsight reconstruction to supply deficiencies in its factual basis." *In re Warner*, 379 F.2d 1011, 1017 (CCPA 1967). Additionally, in cases such as the present, where the obviousness challenges involve a modification or combination of teachings, the Petitioner must also establish that one skilled in the art would have had a reasonable expectation of success to make the modification or combination. *Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.*, 821 F.3d 1359, 1367-68 (Fed. Cir. 2016). The modification or combination must not "require a substantial reconstruction and redesign of the elements" in a reference or "a change in the basic principles under which [a reference] was designed to operate." *In re Ratti*, 270 F.2d 810, 813 (CCPA 1959). Nor can the

modification or combination render a reference unsatisfactory or "inoperable for its intended purpose." *In re Gordon*, 733 F.2d 900, 902 (Fed. Cir. 1984).

Also, the prior art must be considered for all that it teaches, and a *prima facie* case of obviousness may be rebutted "by showing that prior art teaches away from claimed invention in any material respect." *In re Peterson*, 315 F.3d 1325, 1331 (Fed. Cir. 2003).

Last but not least, obviousness cannot be established by hindsight combination to produce the claimed invention, as the Supreme Court has emphasized:

> A factfinder should be aware, of course, of the distortion caused by hindsight bias and must be cautious of arguments reliant upon *ex post* reasoning. *See Graham [v. John Deere Co.]*, 383 U.S. [1,] 36, 86 S. Ct. 684, 15 L. Ed. 2d 545 [(1966)] (warning against a "temptation to read into the prior art the teachings of the invention in issue" and instructing courts to "'guard against slipping into use of hindsight'" (*quoting Monroe Auto Equip. Co. v. Heckethorn Mfg. & Supply Co.*, 332 F.2d 406, 412 (6th Cir. 1964)).

*KSR*, 550 U.S. at 421.

In its petition, MB USA presented two groups of challenges to the claims of the ’895 patent; one based on the teachings of Sinharoy (Ex. 1003), and the other based on the teachings of Tang (Ex. 1004):

| **Ground** | | **References** | **Claims** |
|---|---|---|---|
| 1a | §§ 102/103 | Sinharoy[1] | 1, 3-7, 11-13, 15-17 |
| 1b | § 103 | Sinharoy in view of Xu[2] | 2, 10 |
| 1c | § 103 | Sinharoy in view of Tang[3] | 8, 9, 14 |
| 2a | §§ 102/103 | Tang | 1, 5-9,11-14, 15, 17 |
| 2b | § 103 | Tang in view of Harikumar[4] | 2-4, 10, 16 |

*See* Pet. at 4. As explained below, however, the petition fails to demonstrate that any challenged claim is unpatentable.

---

[1] B. Sinharoy et al., IBM POWER7 multicore server processor, IBM J. Res. & Dev., v. 25, no. 1 (May/June 2011).

[2] Thomas Canhao Xu et al., Explorations of Optimal Core and Cache Placements for Chip Multiprocessor, 2011 NORCHIP (Jan. 12, 2012).

[3] Dan Tang et al., DMA Cache: Using On-Chip Storage to Architecturally Separate I/O Data from CPU Data for Improving I/O Performance, HPCA - 16 2010 The Sixteenth International Symposium on High-Performance Computer Architecture (Jan. 9-14, 2010 Bangalore, India).

[4] US PGPUB 2009/0164739.

### A. The Challenged Claims are Not Unpatentable in View of Sinharoy

#### 1. Overview of Sinharoy

Sinharoy is a technical journal paper that describes IBM's POWER7 server processor; a device with an eight-core processor and a memory subsystem that contains three levels of on-chip cache.[5] Ex. 1003 at Abst.; *and see id.* at 2, 16–19, Fig. 1. According to Sinharoy, in addition to an individual L1 cache (with instruction and data components), each core of the POWER7 processor has a small (256 KB), private L2 cache, and a 4-MB local L3 region of a 32-MB L3 cache. *Id*. at 16-19; and see Ex. 2001 at 14:12-16 ("L1 means level 1. It's the cache that's the closest to the core. And then Level 2 are caches that are shared among the cores, and [ ] Level 3 is the highest level or the last level cache, in Sinharoy.").

Each L2 cache of the Power7 processor has its own "single centralized controller" to manage read/write request operations. Ex. 1003 at 18 ("The private 256-KB POWER7 L2 cache utilizes 128 byte lines and is

[5] A cache is "a high-speed memory that is used to provide memory references more quickly to a processor than if it had to go all the way to the main memory to perform all those cache references -- all those memory references." Ex. 2001 at 11:20-24.

eight-way set associative. It is comprised of a single centralized controller, with two address-hashed cache data arrays."). Likewise, each individual L3 cache region has its own "single centralized controller." *Id*. at 19 ("Each L3 region utilizes 128-byte lines and is eight-way set associative. A given L3 region comprises a single centralized controller, with four address hashed eDRAM cache data banks and four address hashed SRAM directory banks."). Each L3 region controller provides "a single core/L2 dispatch port, a single lateral L3 region dispatch port, dual snoop dispatch ports (for even and odd cache lines), and a single pool of operational resources." *Id*. For reads that miss a core's L2 cache, the core's L3 region is accessed via the core/L2 dispatch port. *Id*. And, "Storage accesses that miss both the L2 cache and the 4-MB local L3 region are broadcast to [a] coherence fabric and snooped by [a] memory controller, other L2 caches, possibly other L3 caches, and by the seven remote 4-MB L3 regions that comprise the remaining 28 MB of the on-chip L3 cache." *Id*.

### 2. Claim 1 is not unpatentable in view of Sinharoy

According to the petition, Sinharoy teaches a "shared cache memory," as recited in claim 1, by describing the "shared 32-MB L3 cache" which is made up of "eight 4-MB L3 regions." Pet. at 12. And, Sinharoy teaches the

recited "caching agent" of claim 1 by describing "centralized controllers for the L2 and L3 caches." *Id*. at 16 (citing Ex. 1003 at 1:18-19). MB USA alleges, "Sinharoy discloses that the caching agent . . . is formed as a single caching agent for the processor [because] [t]he cache controllers for the L2 and L3 caches, including RC and write machines, form a logical single caching controller, with the single cache controller including a plurality of distributed portions each associated with one of the cores." *Id*. at 18 (citing Ex. 1007 at ¶ 65); *and see id*. at 19 (alleging "[E]ach centralized cache controller for a given L3 region is a 'distributed portion' of the caching agent.").

MB USA's characterization of the multiple L2 and L3 cache controllers of the Power7 processor as a "single cache controller" is wrong. Ex. 2022 at ¶¶ 35-39. Sinharoy specifically states that each core's L2 cache has its own controller. Ex. 1003 at 18 ("The private 256-KB L2 cache . . . is comprised of a single centralized controller . . . ."). These are not distributed portions of one cache controller for the entire L2 and L3 cache hierarchy. They are, instead, exactly what Sinharoy says they are: individual controllers for a core's private L2 cache and a person of ordinary skill in the art would appreciate as much. Ex. 2002 at ¶ 36-39.

For example, Sinharoy specifies that "Storage accesses that miss both the L2 cache and the 4-MB local L3 region are broadcast to the coherence fabric *and snooped by* the memory controller, *other L2 caches*, possibly other L3 caches, *and by the seven remote 4-MB L3 regions that comprise the remaining 28 MB of the on-chip L3 cache*." *Id*. at 19 (emphasis added). Thus, a person of ordinary skill in the art would understand that when a cache access by a particular core misses that core's private L2 cache and its allocated portion of the shared L3 cache, that miss is broadcast and must be (and is) snooped by the cache controllers for each of the other cores' respective L2 caches and by the respective, separate cache controllers for the other allocated regions of the shared L3 cache. Ex. 2022 at ¶¶ 36-37. This need for snooping of cache misses by all of the individual cache controllers of other cores' respective L2 and L3 cache controllers indicates that no one single cache controller is or is acting as *a single*—that is, a solitary—*caching agent for the processor*, as recited by claim 1. Instead, each core's L2 cache and allocated portion of the L3 cache has its own controller, which is and acts independently of the other cache controllers of the processor. *Id*.

Petitioner's declarant, Dr. Horst, testifies that, "A cache cannot maintain its own coherence without responding to transactions to other caches. If one cache ignored transactions to other caches, the system would

not operate properly because multiple caches could have copies of the same cache line, but those lines could hold different data. As such, a 'single caching agent' is a necessary single logical component formed by physically distributed portions." Ex. 1007 at ¶ 65. But Dr. Horst offers no explanation as to why such a "single caching agent" arrangement is *necessary*. In fact, Sinharoy itself undermines this statement.

In Sinharoy, it is undisputed that each core's L2 cache and allocated portion of the L3 cache has its own controller. Ex. 1003 at 18, 19. And, it is undisputed that these individual controllers maintain coherency by snooping one another's transactions. *Id*. at 19. Dr. Brogioli explains that it is through these individual snooping actions that overall coherency among the caches is maintained, notwithstanding that the individual cache controllers are and operate independent of one another. Ex. 2002 at ¶ 38. Not only is Dr. Horst's testimony inconsistent with the disclosure of Sinharoy, the fact that it is made without citing any underlying data means it should be afforded little or no weight. 37 C.F.R. § 42.65(a); *and see* PTAB Consolidated Trial Practice Guide ("Consol TPG") at p. 35 (Nov. 2019); *Xerox Corp. v. Bytemark, Inc.*, IPR2022-00624, Paper 9, slip op. at 15-17 (PTAB Aug. 24, 2022) (denying institution of trial where expert's testimony was merely conclusory assertions) (precedential).

At deposition, Dr. Horst agreed that Sinharoy does not explicitly describe a single cache controller for the processor. *See* Ex. 2001 at 15:11 – 16:19. But Dr. Horst alleged that,

> [I]f you go to page 20 at the top, it says,
>
> > "When the associated core and L2 are enabled, a number of heuristics cooperate to give precedence to the associated L2 cache while allowing load and capacity balancing among all eight L3 regions on the chip, thereby causing them to behave as a highly set-associative (64-way) at the present time 32 objective L3 cache."
>
> And the coordinating to do capacity balancing means that there has to be a centralized controller for all the caches to do that load balancing.

*Id*. at 16:20 – 17:8. This is not so.

Dr. Brogioli clarifies that the heuristics referred to in the first paragraph on p. 20 of Sinharoy are explained in the following paragraphs thereof and that those explanations make clear what is going on. Ex. 2022 at ¶ 38. In particular, each core's 4-MB L3 region is treated as a victim cache for the associated core's L2 cache, and the other seven cores' 4-MB L3 regions are treated as victim caches for the subject core's 4-MB L3 region.

*Id*. (discussing Sinharoy at p. 20). That is, each 4-Mbyte localized L3 region serves a dual role: as a victim cache for its associated L2 cache, and as a victim cache for the other seven 4-Mbyte regions. *Id*. The heuristics described by Sinharoy strike a balance between these roles by monitoring and managing cache line allocation. When a localized L3 region evicts a cache line, that L3 region first decides whether to route the line to memory or to another L3 region. If another region is selected and cloned copies exist, the line is merged with the clones. If no clones exist, the L3 region determines which of the seven other regions is the best candidate to store the cache line. But, the targeted local L3 region must then decide whether to accept or reject the cache line. *Id*. The need for such decision making by the candidate localized L3 region demonstrates that there is no single cache controller operating. If it were, the need for such decision making by the candidate localized L3 region would not exist. Instead, the single cache controller would simply enforce the "load balancing" and store the cache line at the identified candidate cache. That does not happen, however and instead it is up to the candidate L3 region to decide whether or not to accept the proposed cache line. Thus, rather than supporting Dr. Horst's premise of a "single" cache controller for the processor, Sinharoy's discussion of load balancing actually demonstrates why a person of ordinary skill in the art

would conclude there is no such single cache controller present. *Id*. The situation is as Sinharoy says, each L3 region, like each L2 cache, has its own cache controller. *Id*.

Further portions of Sinharoy also undercut Dr. Horst's and Petitioner's characterization of the cache controller in Sinharoy. For example, Dr. Horst relies on a statement in Sinharoy that reads, "possibly other L3 caches" may need to snoop L2/L3 misses for his conclusion that there must only be a single logical cache controller. Ex. 2001 at 18:3-9. But he ignores the remainder of that same sentence that specifies the controllers of "the seven remote 4-MB L3 regions that comprise the remaining 28MB of the on-chip L3 cache" must perform such snooping. See Ex. 1003 at 19. Dr. Horst's selective reading of Sinharoy does not support the petition's contention of a single cache controller for the Power7 processor and, in fact, directly conflicts with the actual teachings of Sinharoy.

Petitioner's restatement of the premise that the separate centralized controllers for the L2 and L3 caches described by Sinharoy are distributed portions of a single cache controller as the conclusion of same is both wrong and an insufficient basis for finding that claim 1 of the '895 patent is anticipated by or obvious in view of Sinharoy. In fact, the cache controllers for the L2 and L3 caches are not merely distributed portions of a single

cache controller for the processor, as required by claim 1, and, as Dr. Brogioli has outlined, Sinharoy explains as much. Thus, for at least the foregoing reasons, Sinharoy fails to teach or suggest "*a single caching agent for the processor*," as recited in independent claim 1. Hence, claim 1 is not unpatentable in view of Sinharoy. *Schering Corp. v. Geneva Pharm.*, 339 F.3d 1373, 1377 (Fed. Cir. 2003) (a patent claim is "invalid for anticipation [only] if a single prior art reference discloses each and every limitation" of the claim); *CFMT, Inc. v. Yieldup Intern. Corp*., 349 F.3d 1333, 1342 (Fed. Cir. 2003) (obviousness requires a suggestion of all limitations in a claim); *see also* 37 C.F.R. § 42.108(c); *Harmonic Inc. v. Avid Tech., Inc.*, 815 F.3d 1356, 1363 (Fed. Cir. 2016) ("In an [*inter partes* review], the petitioner has the burden from the onset to show with particularity why the patent it challenges is unpatentable.") (citing 35 U.S.C. § 312(a)(3) (requiring *inter partes* review petitions to identify "with particularity . . . the evidence that supports the grounds for the challenge to each claim")); *Dynamic Drinkware, LLC v. Nat'l Graphics, Inc*., 800 F.3d 1375, 1378 (Fed. Cir. 2015) (the burden of persuasion never shifts to the Patent Owner); *In re Magnum Oil Tools Int'l, Ltd*., 829 F.3d 1364, 1380 (Fed. Cir. 2016) (a petitioner cannot satisfy its burden of proving obviousness by employing "mere conclusory statements").

### 3. Claims 11 and 15 are not unpatentable in view of Sinharoy

Independent claims 11 and 15 also each recite "*a single caching agent for the processor*," and are, therefore, not unpatentable in view of Sinharoy for at least the same reasons as claim 1.

### 4. The respective dependent claims are not unpatentable in view of Sinharoy

Claims 3-7 depend (directly or indirectly) from claim 1, claims 12 and 13 depend from claim 11, and claims 16 and 17 depend from claim 15. Each respective dependent claim is not unpatentable in view of Sinharoy for at least the same reasons as claim 1. *Hartness Int'l. Inc. v. Simplimatic Engineering Co.*, 819 F.2d 1100 (Fed. Cir. 1987) (a claim that depends from a nonobvious independent claim is nonobvious because it contains all of the limitations of that independent claim plus a further limitation).

### 5. The further challenges relying on Sinharoy are also unavailing

The petition alleges that claims 2 and 10 are obvious in view of the combined teachings of Sinharoy and Xu, and that claims 8, 9, and 14 are obvious in view of the combined teachings of Sinharoy and Tang. Pet. at 4.

Each of these challenges would appear to rely on the analysis of Sinharoy teaching *a single caching agent for the processor* refuted above with respect to claim 1. *Id*. at 33-35, 36-42. Claims 2 and 8-10 depend from claim 1 and claim 14 depends from claim 11. Hence, for the same reason that independent claims 1 and 11 are not unpatentable in view of Sinharoy, none of the respective dependent claims 2, 8-10, and 14 are unpatentable in view of Sinharoy even when considered in combination with Xu and/or Tang. *Hartness Int'l.*, 819 F.2d 1100.

*a) Petitioner has failed to demonstrate that the Xu reference is prior art to the '895 patent*

As noted above, the petition cites Xu in combination with Sinharoy with respect to claims 2 and 10 and alleges Xu is prior art to the '895 patent. Pet. at 33 n. 1. However, the petition fails to make a sufficient showing that Xu qualifies as prior art. Therefore, all challenges based on Xu fail.

To be eligible as prior art, Xu must qualify as a "printed publication," specifically that it was publicly available before the invention claimed in the '895 patent. *See* 35 U.S.C. § 311(b) *and* pre-AIA 35 U.S.C. § 102(a). The determination of whether a document is a "printed publication" under 35 U.S.C. § 102 "involves a case-by-case inquiry into the facts and circumstances surrounding the reference's disclosure to members of the

public." *In re Klopfenstein*, 380 F.3d 1345, 1350 (Fed. Cir. 2004). Public accessibility is "the touchstone in determining whether a reference constitutes a 'printed publication.'" *Jazz Pharm., Inc. v. Amneal Pharm., LLC*, 895 F.3d 1347, 1355 (Fed. Cir. 2018). "Whether a reference qualifies as a printed publication under § 102 is a legal conclusion based on underlying fact findings." *Acceleration Bay LLC v. Activision Blizzard, Inc.*, 908 F.3d 765, 772 (Fed. Cir. 2018).

"A reference is considered publicly accessible if it was 'disseminated or otherwise made available to the extent that persons interested and ordinarily skilled in the subject matter or art, exercising reasonable diligence, can locate it.'" *Id*. (citing *Jazz Pharm*., 895 F.3d at 1355–1356). Here, no such showing has been made. Petitioner relies on the MacPherson declaration for the proposition that Xu was available to the public prior to the invention claimed in the '895 patent. Pet. at 33 n. 1. However, the MacPherson Declaration conflicts with the IEEE' own records concerning the manner by which that institution provides public accessibility to its materials.

Ex. 1015 is said to be a copy of the IEEE Xplore Page concerning the Xu article. *See* Pet. at A-1, 33 n.1. Mr. MacPherson explains that IEEE Xplore (a digital library) is the vehicle by which the Institute for Electrical

and Electronic Engineers makes articles such as Xu available to the public. Ex. 1013 at ¶ 7. The IEEE Xplore Page states that Xu was added to the digital library January 12, 2012. Ex. 1014 at p. 1. This is after the effective filing date of the ’895 patent, December 31, 2011. Ex. 1001 at p. 1 (63). Hence, Xu has not been shown to be “publicly available” prior to the invention recited in the ’895 patent.

Petitioner contends that Xu was “published” in 2011, on or about November 14-15 and relies on Mr. MacPherson to say that it was available conference participants at that time. Pet. at 33, n.1. But Petitioner offers no explanation of how or where publication occurred and Mr. MacPherson recites no personal knowledge of any such publication. He attests only to be a custodian of records for the IEEE. *See* Ex. 1013. Indeed, Mr. MacPherson says that he obtained a copy of the Xu article via IEEE Xplore and that this is the manner by which IEEE “make[s such articles] available to the public.” *Id*. at 7, 9. Mr. MacPherson provides no basis for a claim that copies of Xu were made publicly available prior to January 12, 2012, the date it was added to IEEE Xplore, and even if one accepts his statement that “Copies of the conference proceedings were made available no later than the last day of the conference,” *id*. at 11, he offers no testimony that the Xu article was actually included as part of those conference proceedings and makes no

claim that he was actually at the conference and can attest to the contents of those proceeding from personal knowledge. Thus, his statement must be seen as speculative at best.

In short, Petitioner has not shown by a preponderance of evidence that Xu was publicly accessible before the invention claimed in the '895 patent and therefore qualify as prior art under 35 U.S.C. § 102(a). No coy of the 2011 NORCHIP Conference Proceedings was provided, so there is nothing to indicate that Xu was included in same. And, the digital library page that identifies the date on which the Xu article was added to that vehicle for public accessibility states on its face that date was after the invention date. Accordingly, Petitioner's challenges that rely on Xu must fail.

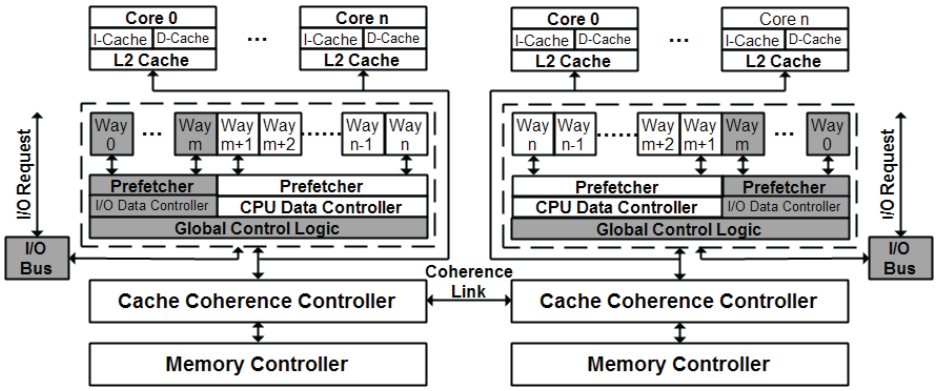
### B. The Challenged Claims are Not Unpatentable in View of Tang

#### 1. Overview of Tang

Tang is an article describing "a DMA cache technique to store I/O data in dedicated on-chip storage." Ex. 1004 at 1, Abst. Two such techniques are described, the first being a "Decoupled DMA Cache (DDC)" design, and the second being a "Partitioned-Based DMA Cache (PBDC)" design. *Id*. The petition alleges the unpatentability of the independent claims of the '895 patent based on Tang's description of the PBDC design. Pet. at 45-55

(addressing claim 1), 55-58 (addressing claim 11), 58-61 (addressing claim 15).

Tang's PBDC design is discussed with reference to Figures 10-12. Tang's Figure 10 is reproduced below, for convenience.




According to Tang, in the PBDC design, several ways of a processor's last level cache ("LLC") are used as a dedicated DMA cache for storing so-called I/O data. Ex. 1004 at 6-7. A "way" is simply a portion of a cache memory. *See* Ex. 1007 at ¶ 115, n. 16. I/O data is distinguished from CPU data in that IO data is "a piece of data produced by I/O devices directly," while CPU data is "1) the data not produced by the I/O devices and 2) the I/O data updated by the CPU." Ex. 1004 at 2 (§ 2). Figure 10, above, shows this division of the cache ways, with $Way_0$ – $Way_m$ being the I/O data ways and $Way_{m+1}$ – $Way_n$ being the CPU data ways. *Id*. at 7 (§ 3.2.2).

As evident from Figure 10, Tang's PBDC design includes two cache coherence controllers and corresponding last level cache controllers ("LLC-

Ctrler"). *Id*. at 6-7; Fig. 10. The LLC-Ctrler "consists of Global Control Logic (GCL), IO Data Controller (IOD-Ctrler), CPU Data Controller (CPUD-Ctrler), two Prefetchers and Configuration Module (CM). *Id*. at 7 (§ 3.2.2). The IOD-Ctrler and the CPUD-Ctrler are responsible for maintaining cache coherence and other things. *Id*. While the CPUD-Ctrler uses the processor's original cache management policies, the IOD-Ctrler uses different cache management policies for the I/O data. *Id*.

### 2. Independent claims 1, 11, and 15 are not unpatentable in view of Tang

The petition alleges that the cache coherence controller and the LLC controller are collectively a "single caching agent for the processor" as recited in the challenged independent claims. Pet. at 53-54. This is refuted by Tang itself. As noted above, Tang shows a pair of cache coherence controllers which communicate with one another via a "coherence link." Ex. 1004 at Fig. 10. If the two coherence controllers were part of the same "single caching agent for the processor" no coherence link would be needed as the two units would be part of the same whole. Ex. 2002 at ¶ 44. The petition, and Petitioner's declarant, allege that these two coherence controllers are part of the same "logical" unit, but Tang makes no such claim. Indeed, Dr. Horst simply states that this must be the case because

there is a need to maintain cache coherence throughout the processor. Ex. 1007 at ¶ 127. But even if such a need exists, there is no evidence that the only way this can be achieved is if the two coherence controllers illustrated by Tang are in fact the same "logical" unit. One would expect that if that were the case, Tang would simply have drawn the two units as a single one, or at least placed them in a common dotted outline so as to show such a relationship. Ex. 2002 at ¶ 44. Instead, Tang took care to demote their separate nature by illustrating them as such and the provision of a coherence link demonstrates that the two operate not as a single unit but instead independently and communicate with one another as to their independent actions. *Id*. In short, nothing in Tang suggests there is a single, that is solitary, caching agent for the processor, Dr. Horst's conclusory testimony to the contrary notwithstanding. *See* Consol TPG at 35.

Even if Tang can be read as disclosing a single caching agent for the processor, it is evident that that caching agent does not include a plurality of distributed portions *each associated with a corresponding one of the plurality of cores*, as further required by the challenged claims. Ex. 2002 at ¶ 45. As noted above, in Tang's PBDC design each LLC controller includes an I/O data controller and a CPU data controller. The I/O data controller is associated with some number of ways of the last level cache while the CPU

data controller is associated with the remaining ways of the last level cache. *Id*. at ¶¶ 45-46; Ex. 1004 at 7. But, as shown in Figure 10, all of the ways of this last level cache are shared across all of the cores of the processor. Indeed, Tang notes that on a particular core's L2 cache miss, the request is provided to the CPU data controller to see if the requested data is present in any of the CPU data ways, while concurrently a probe-read request to the IO data controller to investigate the IO data ways is made. This highlights the shared nature of the LLC cache ways across all of the processor's cores. Ex. 2002 at ¶¶ 45-46.

Thus, the petition fails to adequately address the challenged claims' requirement that the single caching agent includes a plurality of distributed portions *each associated with a corresponding one of the plurality of cores*. The petition alleges, "Because the LLC-Controller and the cache coherence controller (i.e., the caching agent) manage cache coherence protocol for the plurality of cores (core 0 to core n) and the n-way LLC ($Way_0$ to $Way_n$), they have partitioned regions or units (i.e., distributed portions), each associated with a given core." Pet. at 54 (citing Ex. 1007 at ¶129.). This statement is both factually wrong and is simply trying to reformulate the premise of the argument as its conclusion. As demonstrated above and by Dr. Brogioli, in Tang, the different ways of the LLC cache are shared across the cores of the

processor and so it is not true that the cache controller has partitioned regions or units each associated with a given core. Ex. 2002 at ¶ 46. The partitioned units may be associated with different ways of the last level cache, but those ways are not each associated with a given core and so neither are their associated cache controllers. *Id*.

For its part, the Horst Declaration, on which the petition relies, offers nothing more than the petition. Repeating the allegation in the petition, Dr. Horst states, "Because the LLC-Controller and the cache coherence controller (i.e., the caching agent) manage cache coherence protocol for the plurality of cores (core 0 to core n) and the n-way LLC (Way $_0$ to Way $_n$), they necessarily have partitioned regions or units (i.e., distributed portions), each associated with a given core." This is simply restating the argument set forth in the petition without adding any facts or data on which to base such a claim. As such, it is entitled to little or no weight. And, in any event, it is simply incorrect for the reasons discussed above.

For at least the foregoing reasons then, the challenged independent claims are not unpatentable in view of Tang. *Schering Corp.*, 339 F.3d at 1377; *CFMT, Inc*., 349 F.3d at 1342.

### 3. The respective dependent claims are not unpatentable in view of Tang

Claims 5-9 depend from claim 1, claims 13 and 14 depend from claim 11, and claim 17 depends from claim 15. Each respective dependent claim is not unpatentable in view of Tang for at least the same reasons as its respective independent claim. *Hartness Int'l.*, 819 F.2d 1100.

Claims 2-4, 10, and 16 are alleged to be unpatentable in view of the combined teachings of Tang and Harikumar, Pet. at 4, but the petition relies on Tang for disclosing a *single* caching agent that includes a plurality of distributed portions *each associated with a corresponding one of the plurality of cores*. *See Id*. at 70-77. As demonstrated above, tang fails to teach or suggest such features. Hence, the petition has not demonstrated that 2-4, 10, and 16 are unpatentable.

## IV. CONCLUSION

For at least the foregoing reasons, the Board should find that Petitioner has failed to prove the unpatentability of any of the challenged claims.

Respectfully submitted,

Dated: May 28, 2024

*/Tarek N. Fahmi/*
Tarek N. Fahmi
Reg. No. 41,402

Ascenda Law Group, PC
2150 N First St., Suite 420

San Jose, CA 95131
Tel: 866-877-4883
Email: tarek.fahmi@ascendalaw.com

## WORD-COUNT CERTIFICATE

I certify that the foregoing document complies with the type-volume limitation of 37 C.F.R. § 42.24(a) and (b) and contains 7,306 words (including words in drawings created by the Patent Owner or as annotations added by the Patent Owner to images reproduced in this document) based on the word count indicated by Microsoft® Word for Mac v. 16.85, excluding those portions exempted by the rule.

*/Tarek N. Fahmi/*
Tarek N. Fahmi
Reg. No. 41,402

## CERTIFICATE OF SERVICE

The undersigned hereby certifies that a copy of the foregoing:

PATENT OWNER'S RESPONSE

and its accompanying exhibits (if any) were served on May 28, 2024, by filing this document though the PTAB P-TACTS System as well as by delivering a copy via email directed to the attorneys of record for the Petitioner at the following addresses:

Celine J. Crowson
Joseph J. Raffetto
Scott Hughes
Helen Y. Trac
Ryan Stephenson
Nicholas Rotz
Hogan Lovells US LLP
555 13th Street N.W.
Washington, D.C. 20004

Email:
celine.crowson@hogan.com
joseph.raffetto@hoganlovells.com
scott.hughes@hoganlovells.com
helen.trac@hoganlovells.com
ryan.stephenson@hoganlovells.com
nicholas.rotz@hoganlovells.com

The parties have agreed to electronic service in this proceeding.

Respectfully submitted,

Dated: May 28, 2024

*/Tarek N. Fahmi/*
Tarek N. Fahmi
Reg. No. 41,402

Ascenda Law Group, PC
2150 N First St., Suite 420
San Jose, CA 95131
Tel: 866-877-4883
Email: patents@ascendalaw.com