

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 IPR2023-01344
Patent 9,575,895

**SUPPLEMENTARY DECLARATION OF DR. ROBERT W. HORST
IN SUPPORT OF PETITIONER'S REPLY FOR
U.S. PATENT NO. 9,575,895**

Filed on behalf of Petitioner:
Celine Jimenez Crowson (Reg. No. 40,357)
Joseph Raffetto (Reg. No. 66,218)
Scott Hughes (Reg. No. 68,385)
Nicholas Rotz (Reg. No. 75,959)

HOGAN LOVELLS US LLP
555 13th Street N.W.
Washington, D.C. 20004
Telephone: 202.637.5600
Facsimile: 202.637.5710

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
A. Engagement	1
B. Information Considered.....	1
II. CLAIM CONSTRUCTION.....	2
III. The Challenged Claims Are Unpatentable Over Sinharoy	4
A. Dr. Brogioli’s Argument that L2 and L3 Caches Have Structurally Individual Cache Controllers Is Irrelevant.....	4
B. Dr. Brogioli’s Argument that the Cache Controllers for the L2 and L3 Caches Work Independently from Each Other Is Erroneous Because He Ignores Other Relevant Disclosures in Sinharoy, Such as the Capacity and Load Balancing, the Cache State Table, and the RC Machines Within the L2 and L3 Cache Controllers.	5
IV. The Challenged Claims Are Unpatentable Over Tang.....	11
A. Dr. Brogioli Has Misread and Misinterpreted the Petition, My Prior Declaration, and Figure 10 of Tang.	12
B. A POSITA Would Have Understood the Disclosures in Tang Show that the LLC Controller and Cache Coherence Controller Form a Logical “Single Caching Agent,” Which Has Distributed Portions Each Associated with A Corresponding Core.....	15
V. CERTIFICATION.....	18
<u>APPENDIX A</u>	1

I, Dr. Robert W. Horst, declare:

I. INTRODUCTION

A. Engagement

1. I have been retained on behalf of Mercedes-Benz USA, LLC (“**Mercedes**”) to offer technical opinions relating to U.S. Patent No. 9,575,895 (EX1001) (the “**895 Patent**”) and prior art references relating to its subject matter.¹ In particular, in this Supplementary Declaration, I have been asked to respond to Patent Owner’s expert, Dr. Michael C. Brogioli’s opinions set forth in his Declaration (the “Brogioli Declaration” or EX2002). I hereby incorporate by reference my engagement and background and qualification in my expert declaration in support of Mercedes’s Petition (the “Prior Declaration”).

B. Information Considered

2. In addition to the information considered in my Prior Declaration (EX1007), I have reviewed Patent Owner’s Preliminary Response, the Board’s Institution Decision (“Paper 7”), Patent Owner’s Response (the “POR”), the Brogioli Declaration, my deposition transcript (EX2001), his deposition transcript

¹ For brevity, when providing citations herein, I will generally refer to the name of the reference, rather than the exhibit number. The exhibit numbers associated with each reference can be found in Appendix A.

(the “Brogioli Transcript”), as well as the other materials referenced in Appendix A. I hereby incorporate by reference the opinions set forth in my Prior Declaration into this Supplementary Declaration.

3. My analyses are based on my years of education, research, and work experience, as well as my investigation and study of relevant materials. In my analyses, I have considered the materials that I identify in this Declaration and those listed in Appendix A.

4. My review and assessment of the materials provided in this proceeding is ongoing, and I will continue to consider any new material as it is provided. I reserve the right to review, supplement, and amend my analyses based on new information and on my continuing review of the materials already provided.

II. CLAIM CONSTRUCTION

5. Based on my review of the Institution Decision, I understand that the Board has adopted a construction for the claim term “single caching agent,” which means “one caching agent for the processor, which may be formed from distributed portions.” (Paper 7 at 12.) I agree with this construction, because it is supported by the claim context and the specification of the ’895 patent.

6. Based on my review of Patent Owner’s Response and Brogioli Declaration, I understand that they propose a construction for the term as “a

solitary, that is one, caching agent for an entire processor (or multicore processor or first multicore processor).” (POR, 8, 10.) For the reasons set forth below, in my opinion, Patent Owner’s and Dr. Brogioli’s proposed construction is unnecessary and inconsistent with the claim context and specification of the ’895 Patent.

7. First, as the Board noted in the Institution Decision, the specification of the ’895 Patent discloses “a processor 700 includes a distributed configuration having ‘partitions or slices each including a core 710 and a partition of a caching agent 715 and a LLC 720.’” (’895 Patent, Figure 6, 6:33-35.) “[W]hile distributed caching agents are shown, understand that these distributed portions **form** a single cache agent.” (*Id.*, 6:35-37 (emphasis added); *see also* EX1018, 52:3-53:10 (agreeing that the single caching agent can be “formed by distributed portions”).) Furthermore, based on my review of the ’895 Patent, it does not disclose any specific structures or circuit implementations for the “single caching agent,” except that it discusses the caching agent can be formed from distributed portions and that it “may be logic.” (*See* ’895 Patent, 2:43-44, 6:35-39.) In addition, Claim 1 of the ’895 Patent requires that “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.” (’895 Patent, Claim 1.)

8. Second, based on my review of the ’895 Patent and Brogioli Transcript, nowhere in the ’895 Patent uses the term “solitary,” and Dr. Brogioli

appears to have admitted that this term may be his own word. (See the '895 Patent, EX1018, 46:8-47:8.)

9. Therefore, in my opinion, the Board's adopted construction (i.e., "one caching agent for the processor, which may be formed from distributed portions") is consistent with the claim context and the specification of the '895 Patent and is correct.

III. THE CHALLENGED CLAIMS ARE UNPATENTABLE OVER SINHAROY

10. I have been asked to respond to Dr. Brogioli's opinions regarding the prior art reference *IBM POWER7 multicore server processor* ("Sinharoy") (EX1003) and have reviewed his opinions set forth in the Brogioli Declaration. In sum, his opinions are incorrect and in my opinion, Sinharoy discloses independent claims 1, 11, and 15 of the '895 Patent.

11. Based on my review of the Brogioli Declaration, he appears to have only contended that Sinharoy does not disclose the "single caching agent" element in claim element 1[g] (or similar elements in independent claims 11 and 15). (See Brogioli Declaration ¶¶35-39.) I reserve the right to further supplement my responsive opinions, to the extent that Dr. Brogioli attempts to opine on any other elements not included in his Declaration.

A. Dr. Brogioli's Argument that L2 and L3 Caches Have Structurally Individual Cache Controllers Is Irrelevant.

12. First, Dr. Brogioli argues that because the L2 and L3 caches in Sinharoy have structurally individual cache controllers, there is no single cache controller in Sinharoy. (EX2002, ¶36.) This argument misses the point. In my Prior Declaration, I have stated that the “cache controllers for the L2 and L3 caches in Sinharoy including the pool of RC and write machines in Sinharoy are a ‘caching agent’ and form a **logical ‘single caching agent’** for the processor.” (EX1007, ¶65 (emphasis added).)

B. Dr. Brogioli’s Argument that the Cache Controllers for the L2 and L3 Caches Work Independently from Each Other Is Erroneous Because He Ignores Other Relevant Disclosures in Sinharoy, Such as the Capacity and Load Balancing, the Cache State Table, and the RC Machines Within the L2 and L3 Cache Controllers.

13. Second, Dr. Brogioli incorrectly argues that all L3 misses must be snooped by the non-local L3 regions: “[a] person of ordinary skill in the art would have understood and appreciate that the need to snoop the cache miss by **all of the individual cache controllers of the 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.” (EX2002, ¶37 (emphasis added).) In fact, Sinharoy’s single logical cache controller takes advantage of global knowledge of the remote L3 regions to avoid snooping them when it is known that they could not have modified copies of the cache line. (See EX1003, 1:16-1:19, Table 2.) In my opinion, this remote snooping could not be

avoided if the L3 cache regions had independent cache controllers. (See EX1003, 1:16-1:19, Table 2.)

14. Dr. Brogioli appears to argue that because the L2 and L3 caches in Sinharoy have some snooping operations, the cache controllers for them work independently from each other and cannot form a “single caching agent.” (EX2002, ¶¶36-38.) Based on my review of the Brogioli Declaration and his deposition transcript, in my opinion, he has ignored other key disclosures in Sinharoy, and thus, his opinions are incorrect. If he had reviewed the disclosures in Sinharoy more comprehensively, he would have recognized that the cache controllers in Sinharoy do not work independently from each other but work together and form a single logical caching agent.

15. First, Dr. Brogioli has opined on a sentence from Sinharoy, which reads that “[s]torage 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.” (EX2002, ¶36 (quotation omitted).) While he attempts to provide some explanations about the snooping operation in this scenario, he seemingly ignores the term “possibly other L3 caches” in this sentence. The term “possibly” means that not all of the other L3 cache regions need to be snooped. As I have opined in

my deposition, this disclosure means the cache controllers for the L2 and L3 caches do not work independently from each other. (See EX2001, 18:3-16.) The caches in Sinharoy cannot independently handle read and write misses without coordinating with each other. (See *id.*) To avoid the need of snooping all of the other L3 regions, there has to be “some high-level logic that does that coordination.” (*Id.*, 17:11-14.) A POSITA would have understood that the cache controller for a L3 region has knowledge about whether the lines in its local L3 region are held in any other L2 and L3 caches as well. (See *id.*, 18:3-9.) Such knowledge means that each of them has to have knowledge of what is in the other caches to be able to avoid snooping in other regions. (See *id.*, 18:3-16.)

16. Second, this coordination among the cache controllers for the L2 and L3 is further supported by other disclosures in Sinharoy. For example, the Brogioli Declaration appears to be completely silent about the load balancing disclosures in Sinharoy. Sinharoy discloses that the “capacity, data flow logic, and control flow logic of a 4-MB local L3 region are shared between the L2 cache to which the local L3 region is coupled, and the other seven 4-MB L3 regions on the chip.” (EX1003, 1:19.) Sinharoy then discloses that “when the core and L2 associated with a given L3 region are **disabled, the remaining L3 regions enjoy full access to the L3 region.**” (*Id.* (emphasis added).) A POSITA would have understood that the reason why the remaining L3 regions still enjoy full access to the disabled L3

region is that the other cache controllers still have knowledge about the disabled L3 cache region. Thus, they do not operate independently from each other. Following that, Sinharoy discloses “[w]hen the associated core and L2 are enabled, a number of heuristics² 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) adaptive 32-MB L3 cache.” (*Id.* at 1:20.) As I have explained during my deposition, the load and capacity balancing means that there is a single “centralized controller for all the caches.” (See EX2001, 17:5-8.) The single cache controller includes the distributed portions of cache controllers of the L2 and L3 caches. (See EX1007, ¶¶55-58.)

17. Furthermore, the Brogioli Declaration also ignores Table 2 of Sinharoy, which explains the cache states in L2 and L3 caches. Sinharoy discloses that the “POWER7 L2 and L3 caches support the same 13-cache-state protocol as the POWER6 design point.” (EX1003, 1:17.) “If a given core’s L2 cache or local L3 region owns a copy of the targeted cache line (i.e., holds the line in an M, ME, MU, T, TE, TN, or TEN cache state), the data will be installed into the local L3 region.” (*Id.*) In Table 2, for the cache state “T,” under the “Sharers and scope” column, for

² Sinharoy also discloses the “replacement algorithm serves as a primary heuristic in creating this balance” (Sinharoy, 1:20.) A POSITA would have understood that by using the term “primary heuristics,” there are other heuristics used in Sinharoy.

the “T” state (Description, “Formerly MU, now shared”), Sinharoy specifies “Yes, probably global.” (*Id.*, Table 2.) In my opinion, the “probably global” means that when a cache line is in this state and is potentially shared with other L3 cache regions, accesses to the local L3 regions need to snoop the remote L3 regions. Meanwhile, for the cache state “TEN,” (Description, “Formerly ME, now shared”), it specifies “Yes, local.” (*Id.*)³ In my opinion, the “local” means that the targeted cache line is shared but it is known that the other non-local cores have not modified that cache line, so the local L3 cache region does not need to snoop other regions. The difference between the cache states “T” and “TEN” shows that snooping can be avoided in certain scenarios. At least because a L3 cache region does not need to snoop other L2 caches or L3 cache regions every single time, a POSITA would have understood that the cache controllers for them work together with each other and form a single logical caching agent.

18. Finally, Dr. Brogioli also disregards the disclosures about the read/claim (RC) machines in Sinharoy. A POSITA would have understood that the RC machines in both the L2 and L3 cache controllers show the cache controllers coordinate with each other. Sinharoy discloses that “the read/claim (RC) machines, which have been utilized in prior designs to manage core fetches, read–modify–write

³ Based on Dr. Brogioli’s deposition transcript, I understand that he indicated he is unfamiliar with the TEN cache state. (EX1018, 85:4-16.)

operations on behalf of core stores, and core prefetch operations, have been divided into two classes: 1) a set of eight fully functional RC machines that are incorporated into the L2 cache controller and manage core fetches and stores that hit the L2 cache and 2) a set of 24 more energy-efficient ‘lightweight’ RC machines that are incorporated into the L3 cache controller and manage prefetches and stores that miss the L2 cache.” (EX1003, 1:18.) The L2 RC machines and L3 lightweight RC machines “aggressively reorder storage operations,” thereby “enabling the high volume of system coherence traffic needed to scale robustly to a large number of sockets.” (*Id.*, 1:18-19.) These RC machines are shared across all L3 regions and hence are part of a logical single cache controller. In my opinion, if the L3 regions had separate independent cache controllers, they would each need their own RC machines and there would be no opportunity to globally reorder the storage operations. The RC machines in both the L2 and L3 cache controllers also further support that the L2 and L3 cache controllers work with each other and share information and logic across cores in certain scenarios. Thus, there is not an independent cache controller per core and the shared and portioned parts of the cache controllers form a logical “single caching agent.”

19. In sum, based on a correct and comprehensive read of the disclosures in Sinharoy, a POSITA would have understood that the cache controllers for the L2 and L3 caches do not work independently from each other, as Dr. Brogioli

incorrectly suggests. Instead, they work together to form a logical “single caching agent.” Thus, it is my opinion that Sinharoy discloses independent claim 1, 11, and 15 of the ’895 Patent.

IV. THE CHALLENGED CLAIMS ARE UNPATENTABLE OVER TANG

20. I have also been asked to respond to Dr. Brogioli’s opinions regarding the prior art reference *DMA Cache: Using On-Chip Storage to Architecturally Separate I/O Data from CPU Data for Improving I/O Performance* (“**Tang**”) (EX1004) and reviewed his opinions set forth in the Brogioli Declaration. In short, his opinions are incorrect and in my opinion, Tang discloses independent claims 1, 11, and 15 of the ’895 Patent.

21. First, based on my review of the Brogioli Declaration, he has misread and misinterpreted the Petition, my Prior Declaration, and Figure 10 of Tang to form his opinions. Furthermore, he incorrectly asserts that Tang does not disclose the caching agent which “includes a plurality of distributed portions each associated with a corresponding one of the plurality of cores.”

22. Based on my review of his Declaration, he appears to have only contended that Tang does not disclose the “single caching agent” or the caching agent including “a plurality of distributed portions each associated with a corresponding one of the plurality of cores” in claim element 1[g] (or similar elements in independent claims 11 and 15). (See Brogioli Declaration ¶¶43-46.) I

reserve the right to further supplement my responsive opinions, to the extent that Dr. Brogioli attempts to opine on any other elements not included in his Declaration.

A. Dr. Brogioli Has Misread and Misinterpreted the Petition, My Prior Declaration, and Figure 10 of Tang.

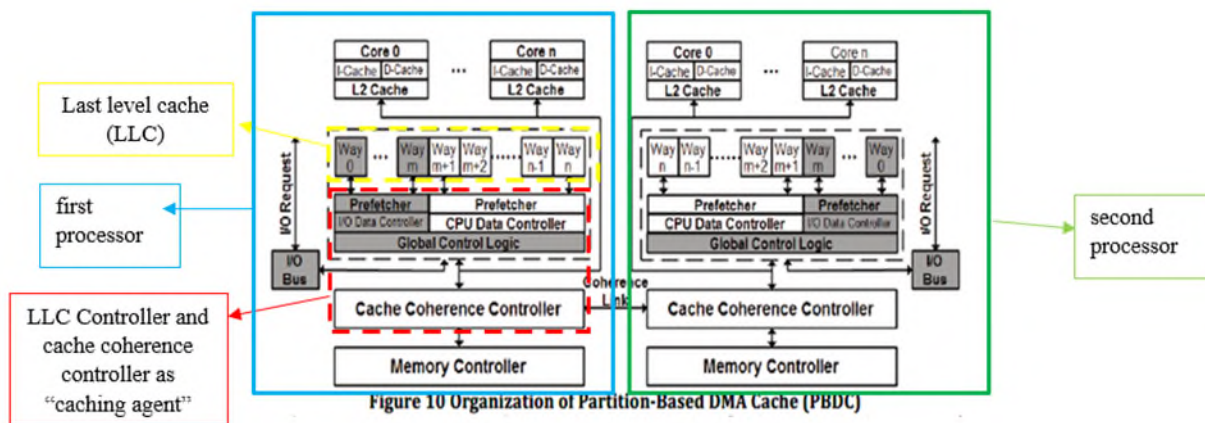
23. First, Dr. Brogioli opines that “the petition alleges that Tang’s cache coherence controller and LLC controller are collectively a ‘single caching agent for the processor’ as recited in the challenged independent claims.” (EX2002, ¶43.) This is consistent with my Prior Declaration. (EX1007, ¶127.) However, following that, he opines that “[i]f the two coherence controllers were part of the same ‘single caching agent for the processor’ no coherence link would be needed” and that the “petition alleges that the two coherence controllers one ‘logical’ unit, but Tang makes no such statement.” (EX2002, ¶44.) Neither the Petition nor my Prior Declaration makes such an argument. (See Pet., 53-55 (noting the LLC controller and the cache coherence controller on the left side of Figure 10 of Tang form a logical single caching agent); EX1007, ¶¶127-29 (same).) Instead, my Prior Declaration has clearly explained that the cache controller and the LLC controller of, for example, just the processor on the left of Figure 10 form a single caching agent for that processor. (EX1007, ¶123.) This can be clearly seen in paragraph 123 of my Prior Declaration, where, for example, it annotates Figure 10 with a red dotted line box around the single caching agent of that one processor. Thus, the

claimed feature of a “single caching agent” for one processor is met by Tang. Furthermore, my Prior Declaration also explains that Tang has a second processor shown on the right side with its own single caching agent comprised of its own cache controller and LLC controller. (EX1007, ¶141.)

24. In addition to misreading the Petition and my Prior Declaration, Dr. Brogioli has also misread Figure 10 of Tang. Dr. Brogioli alleges that there are a pair of cache coherence controllers which communicate with one another via a “coherence link.” (EX2002, ¶44.) His allegation appears to assume that the “pair” of cache coherence controllers are in a **single** processor.

25. However, according to Tang, there are **two** processors in Tang’s Figure 10. Section 3.2 of Tang discloses a partition-based DMA cache (PBDC) design. (EX1004, §3.2.) Tang explicitly discloses that “[f]igure 10 that illustrates the organization of PBDC in a **multiprocessor** system.” (*Id.*, §3.2.1. (emphasis added), Figure 10 (annotated below).) Even Dr. Brogioli admitted in his deposition that Tang discloses a processor on the left side of Figure 10, and another processor on the right side with the same structure. (EX1018, 60:11-63:10 (calling the processor on the right side a “clone” of the left side processor).) He also confirmed that the processor on the left includes an LLC controller and a cache coherence controller, on which I have opined in my Prior Declaration that they form a logical

“single caching agent.” (*Id.*, 61:22-62:15; EX1007, ¶123; EX1004, Figure 10 (annotated below).) This contradicts his arguments in his declaration.



(EX1004, Figure 10 (annotated).)

26. Furthermore, a POSITA would have understood the “coherence link” in Figure 10 connects the two processors together to form a cluster and is not a connection internal to a processor. As such, the presence of the “coherence link” is irrelevant as to whether each processor has its own respective “single caching agent.” The ’895 Patent discloses a “Quick Path Interconnect (QPI)” link can “act as an interface to another agent (such as another processor in a multiprocessor system) coupled to the processor.” (’895 Patent, 6:63-67.) A POSITA would have understood that the QPI link can also serve as a “coherence link” between the different processors. Thus, Dr. Brogioli’s allegation that my Prior Declaration

asserts there are a pair of cache coherence controllers within a single processor in Figure 10 of Tang is incorrect.

B. A POSITA Would Have Understood the Disclosures in Tang Show that the LLC Controller and Cache Coherence Controller Form a Logical “Single Caching Agent,” Which Has Distributed Portions Each Associated with A Corresponding Core.

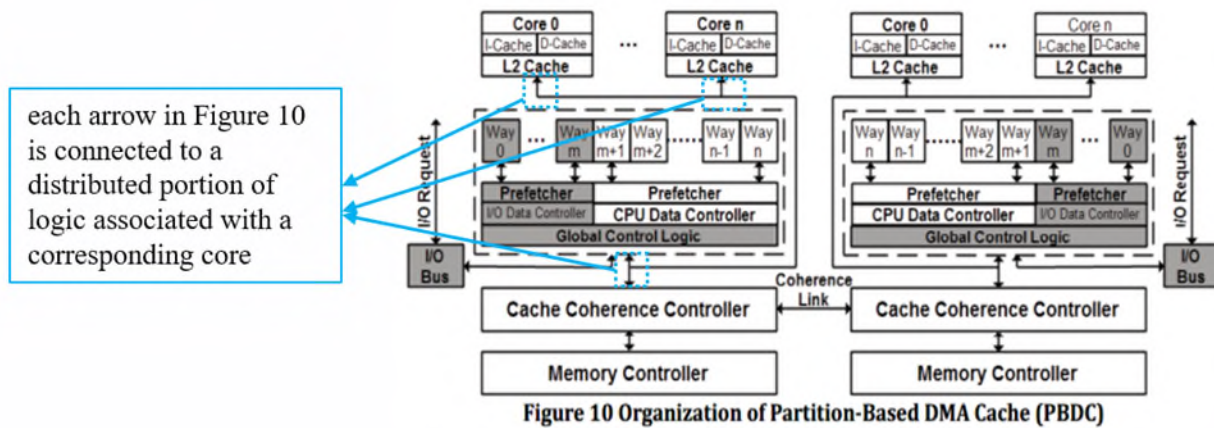
27. Second, Dr. Brogioli then alleges that “in Tang the caching agent does not include a plurality of distributed portions each associated with a corresponding one of the plurality of cores, as required by the challenged claims.” (EX2002, ¶45.) This is incorrect based on Tang’s disclosures, in particular with regard to Figure 10 and Section 3.2.2.

28. As my Prior Declaration has already stated, in the PBDC design, Tang discloses that the LLC is partitioned into “several ways” as the “dedicated DMA Cache.” (EX1007, ¶128.) Tang recognizes that some applications are I/O intensive, and others are not. (EX1004, § 2.2.) It is my opinion that a POSITA would recognize that the I/O-intensive applications may be assigned to one core with non-IO applications assigned to the remaining cores. In a simple case, Core 0 runs all I/O and Core 1 runs everything else. The I/O portion of the cache includes the I/O ways plus the cache ways for the I/O intensive applications, while the remainder of the cache is assigned to the core running other applications. In this case, the cache is partitioned with core 0 associated with the shaded I/O Ways 0-m plus a subset of Ways m+1 through Way n, while Core 1 is associated with the remaining Ways

m+1 through Way n. Tang also discloses that the “partitioned cache approaches are proposed to reduce cache interferences among the threads in multicore platform.” (EX1004, §6.) It is my opinion that a POSITA would have understood that cache partitioning usually refers to the partitioning of shared L2 or L3 caches among a set of programming threads running simultaneously on different cores. A POSITA would have understood that partitioning of an L3 cache to threads running on different cores means that the cache partitions (for example, the ways of the LLC) are each associated with a corresponding core.

29. In addition, according to Figure 10 of Tang, the LLC controller and the cache coherence controller (i.e., the “single caching agent”) include signals and logic (depicted as arrows in Figure 10) that are each connected to core 0 to core n. (EX1004, Figure 10 (annotated below).) The separate arrows to each of the cores 0-n shows that the shared logic must have signals to individual cores for handling, e.g., L2 cache misses, and those signals are sent or received by logic associated with individual cores. As shown in Figure 10, each signal (depicted as a single arrow going into a given core in Figure 10) is associated with logic in a single given core, which is a “distributed portion” of the caching agent. (EX1004, Figure 10 (annotated below).) Thus, because the cache coherence controller sends different signals to the different cores, it has “distributed portions” of logic that connect to these signals. In addition, the cache coherence controller has

“distributed portions” corresponding to each core, allowing it to separately address and send signals to the different cores. (EX1004, Figure 10 (annotated below).)



(EX1004, Figure 10 (annotated).)

30. In sum, the LLC controller and cache coherence controller of Tang, forming a logical “single caching agent,” include a “plurality of distributed portions each associated with a corresponding one of the plurality of cores.” Therefore, it is my opinion that Tang discloses independent claims 1, 11, and 15 of Tang.

V. CERTIFICATION

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true and, further, that these statements were made with the knowledge that willful false statements and the like are punishable by fine, imprisonment, or both, under Section 1001 of Title 18 of the U.S. Code.

Dr. Robert W. Horst

Dated: September , 2024

APPENDIX A

Materials Considered

Exhibit No.	Description
1001	U.S. Patent No. 9,575,895 (<i>Challenged Patent</i>)
1002	File History for the '895 Patent
1003	IBM POWER7 multicore server processor to Sinharoy et al. (" Sinharoy ")
1004	DMA Cache: Using On-Chip Storage to Architecturally Separate I/O Data from CPU Data for Improving I/O Performance to Tang et al. (" Tang ")
1005	Explorations of Optimal Core and Cache Placements for Chip Multiprocessor to Thomas Canhao Xu et al. (" Xu ")
1006	U.S. Patent Application Publication No. 2009/0164739 to Harikumar (" Harikumar ")
1007	Declaration of Dr. Robert Horst
1016	Architectural Implications of Cache Coherence Protocols with Network Applications on Chip Microprocessors to Kyueun Yi et al. (" Yi ")
1018	Deposition Transcript of Dr. Michael C. Brogioli
2001	Deposition Transcript of Dr. Robert Horst
2002	Declaration of Dr. Michael C. Brogioli
	Institution Decision ("Paper 7")
	Patent Owner's Preliminary Response
	Patent Owner's Response ("POR")
	Any additional materials cited in the Supplementary Declaration

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
A. Engagement	1
B. Information Considered.....	1
II. CLAIM CONSTRUCTION.....	2
III. The Challenged Claims Are Unpatentable Over Sinharoy	4
A. Dr. Brogioli’s Argument that L2 and L3 Caches Have Structurally Individual Cache Controllers Is Irrelevant.....	4
B. Dr. Brogioli’s Argument that the Cache Controllers for the L2 and L3 Caches Work Independently from Each Other Is Erroneous Because He Ignores Other Relevant Disclosures in Sinharoy, Such as the Capacity and Load Balancing, the Cache State Table, and the RC Machines Within the L2 and L3 Cache Controllers.	5
IV. The Challenged Claims Are Unpatentable Over Tang.....	11
A. Dr. Brogioli Has Misread and Misinterpreted the Petition, My Prior Declaration, and Figure 10 of Tang.	12
B. A POSITA Would Have Understood the Disclosures in Tang Show that the LLC Controller and Cache Coherence Controller Form a Logical “Single Caching Agent,” Which Has Distributed Portions Each Associated with A Corresponding Core.....	15
V. CERTIFICATION.....	18
<u>APPENDIX A</u>	1

I, Dr. Robert W. Horst, declare:

I. INTRODUCTION

A. Engagement

1. I have been retained on behalf of Mercedes-Benz USA, LLC (“**Mercedes**”) to offer technical opinions relating to U.S. Patent No. 9,575,895 (EX1001) (the “**895 Patent**”) and prior art references relating to its subject matter.¹ In particular, in this Supplementary Declaration, I have been asked to respond to Patent Owner’s expert, Dr. Michael C. Brogioli’s opinions set forth in his Declaration (the “Brogioli Declaration” or EX2002). I hereby incorporate by reference my engagement and background and qualification in my expert declaration in support of Mercedes’s Petition (the “Prior Declaration”).

B. Information Considered

2. In addition to the information considered in my Prior Declaration (EX1007), I have reviewed Patent Owner’s Preliminary Response, the Board’s Institution Decision (“Paper 7”), Patent Owner’s Response (the “POR”), the Brogioli Declaration, my deposition transcript (EX2001), his deposition transcript

¹ For brevity, when providing citations herein, I will generally refer to the name of the reference, rather than the exhibit number. The exhibit numbers associated with each reference can be found in Appendix A.

(the “Brogioli Transcript”), as well as the other materials referenced in Appendix A. I hereby incorporate by reference the opinions set forth in my Prior Declaration into this Supplementary Declaration.

3. My analyses are based on my years of education, research, and work experience, as well as my investigation and study of relevant materials. In my analyses, I have considered the materials that I identify in this Declaration and those listed in Appendix A.

4. My review and assessment of the materials provided in this proceeding is ongoing, and I will continue to consider any new material as it is provided. I reserve the right to review, supplement, and amend my analyses based on new information and on my continuing review of the materials already provided.

II. CLAIM CONSTRUCTION

5. Based on my review of the Institution Decision, I understand that the Board has adopted a construction for the claim term “single caching agent,” which means “one caching agent for the processor, which may be formed from distributed portions.” (Paper 7 at 12.) I agree with this construction, because it is supported by the claim context and the specification of the ’895 patent.

6. Based on my review of Patent Owner’s Response and Brogioli Declaration, I understand that they propose a construction for the term as “a

solitary, that is one, caching agent for an entire processor (or multicore processor or first multicore processor).” (POR, 8, 10.) For the reasons set forth below, in my opinion, Patent Owner’s and Dr. Brogioli’s proposed construction is unnecessary and inconsistent with the claim context and specification of the ’895 Patent.

7. First, as the Board noted in the Institution Decision, the specification of the ’895 Patent discloses “a processor 700 includes a distributed configuration having ‘partitions or slices each including a core 710 and a partition of a caching agent 715 and a LLC 720.’” (’895 Patent, Figure 6, 6:33-35.) “[W]hile distributed caching agents are shown, understand that these distributed portions **form** a single cache agent.” (*Id.*, 6:35-37 (emphasis added); *see also* EX1018, 52:3-53:10 (agreeing that the single caching agent can be “formed by distributed portions”).) Furthermore, based on my review of the ’895 Patent, it does not disclose any specific structures or circuit implementations for the “single caching agent,” except that it discusses the caching agent can be formed from distributed portions and that it “may be logic.” (*See* ’895 Patent, 2:43-44, 6:35-39.) In addition, Claim 1 of the ’895 Patent requires that “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.” (’895 Patent, Claim 1.)

8. Second, based on my review of the ’895 Patent and Brogioli Transcript, nowhere in the ’895 Patent uses the term “solitary,” and Dr. Brogioli

appears to have admitted that this term may be his own word. (See the '895 Patent, EX1018, 46:8-47:8.)

9. Therefore, in my opinion, the Board's adopted construction (i.e., "one caching agent for the processor, which may be formed from distributed portions") is consistent with the claim context and the specification of the '895 Patent and is correct.

III. THE CHALLENGED CLAIMS ARE UNPATENTABLE OVER SINHAROY

10. I have been asked to respond to Dr. Brogioli's opinions regarding the prior art reference *IBM POWER7 multicore server processor* ("Sinharoy") (EX1003) and have reviewed his opinions set forth in the Brogioli Declaration. In sum, his opinions are incorrect and in my opinion, Sinharoy discloses independent claims 1, 11, and 15 of the '895 Patent.

11. Based on my review of the Brogioli Declaration, he appears to have only contended that Sinharoy does not disclose the "single caching agent" element in claim element 1[g] (or similar elements in independent claims 11 and 15). (See Brogioli Declaration ¶¶35-39.) I reserve the right to further supplement my responsive opinions, to the extent that Dr. Brogioli attempts to opine on any other elements not included in his Declaration.

A. Dr. Brogioli's Argument that L2 and L3 Caches Have Structurally Individual Cache Controllers Is Irrelevant.

12. First, Dr. Brogioli argues that because the L2 and L3 caches in Sinharoy have structurally individual cache controllers, there is no single cache controller in Sinharoy. (EX2002, ¶36.) This argument misses the point. In my Prior Declaration, I have stated that the “cache controllers for the L2 and L3 caches in Sinharoy including the pool of RC and write machines in Sinharoy are a ‘caching agent’ and form a **logical ‘single caching agent’** for the processor.” (EX1007, ¶65 (emphasis added).)

B. Dr. Brogioli’s Argument that the Cache Controllers for the L2 and L3 Caches Work Independently from Each Other Is Erroneous Because He Ignores Other Relevant Disclosures in Sinharoy, Such as the Capacity and Load Balancing, the Cache State Table, and the RC Machines Within the L2 and L3 Cache Controllers.

13. Second, Dr. Brogioli incorrectly argues that all L3 misses must be snooped by the non-local L3 regions: “[a] person of ordinary skill in the art would have understood and appreciate that the need to snoop the cache miss by **all of the individual cache controllers of the 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.” (EX2002, ¶37 (emphasis added).) In fact, Sinharoy’s single logical cache controller takes advantage of global knowledge of the remote L3 regions to avoid snooping them when it is known that they could not have modified copies of the cache line. (See EX1003, 1:16-1:19, Table 2.) In my opinion, this remote snooping could not be

avoided if the L3 cache regions had independent cache controllers. (See EX1003, 1:16-1:19, Table 2.)

14. Dr. Brogioli appears to argue that because the L2 and L3 caches in Sinharoy have some snooping operations, the cache controllers for them work independently from each other and cannot form a “single caching agent.” (EX2002, ¶¶36-38.) Based on my review of the Brogioli Declaration and his deposition transcript, in my opinion, he has ignored other key disclosures in Sinharoy, and thus, his opinions are incorrect. If he had reviewed the disclosures in Sinharoy more comprehensively, he would have recognized that the cache controllers in Sinharoy do not work independently from each other but work together and form a single logical caching agent.

15. First, Dr. Brogioli has opined on a sentence from Sinharoy, which reads that “[s]torage 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.” (EX2002, ¶36 (quotation omitted).) While he attempts to provide some explanations about the snooping operation in this scenario, he seemingly ignores the term “possibly other L3 caches” in this sentence. The term “possibly” means that not all of the other L3 cache regions need to be snooped. As I have opined in

my deposition, this disclosure means the cache controllers for the L2 and L3 caches do not work independently from each other. (See EX2001, 18:3-16.) The caches in Sinharoy cannot independently handle read and write misses without coordinating with each other. (See *id.*) To avoid the need of snooping all of the other L3 regions, there has to be “some high-level logic that does that coordination.” (*Id.*, 17:11-14.) A POSITA would have understood that the cache controller for a L3 region has knowledge about whether the lines in its local L3 region are held in any other L2 and L3 caches as well. (See *id.*, 18:3-9.) Such knowledge means that each of them has to have knowledge of what is in the other caches to be able to avoid snooping in other regions. (See *id.*, 18:3-16.)

16. Second, this coordination among the cache controllers for the L2 and L3 is further supported by other disclosures in Sinharoy. For example, the Brogioli Declaration appears to be completely silent about the load balancing disclosures in Sinharoy. Sinharoy discloses that the “capacity, data flow logic, and control flow logic of a 4-MB local L3 region are shared between the L2 cache to which the local L3 region is coupled, and the other seven 4-MB L3 regions on the chip.” (EX1003, 1:19.) Sinharoy then discloses that “when the core and L2 associated with a given L3 region are **disabled, the remaining L3 regions enjoy full access to the L3 region.**” (*Id.* (emphasis added).) A POSITA would have understood that the reason why the remaining L3 regions still enjoy full access to the disabled L3

region is that the other cache controllers still have knowledge about the disabled L3 cache region. Thus, they do not operate independently from each other. Following that, Sinharoy discloses “[w]hen the associated core and L2 are enabled, a number of heuristics² 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) adaptive 32-MB L3 cache.” (*Id.* at 1:20.) As I have explained during my deposition, the load and capacity balancing means that there is a single “centralized controller for all the caches.” (*See* EX2001, 17:5-8.) The single cache controller includes the distributed portions of cache controllers of the L2 and L3 caches. (*See* EX1007, ¶¶55-58.)

17. Furthermore, the Brogioli Declaration also ignores Table 2 of Sinharoy, which explains the cache states in L2 and L3 caches. Sinharoy discloses that the “POWER7 L2 and L3 caches support the same 13-cache-state protocol as the POWER6 design point.” (EX1003, 1:17.) “If a given core’s L2 cache or local L3 region owns a copy of the targeted cache line (i.e., holds the line in an M, ME, MU, T, TE, TN, or TEN cache state), the data will be installed into the local L3 region.” (*Id.*) In Table 2, for the cache state “T,” under the “Sharers and scope” column, for

² Sinharoy also discloses the “replacement algorithm serves as a primary heuristic in creating this balance” (Sinharoy, 1:20.) A POSITA would have understood that by using the term “primary heuristics,” there are other heuristics used in Sinharoy.

the “T” state (Description, “Formerly MU, now shared”), Sinharoy specifies “Yes, probably global.” (*Id.*, Table 2.) In my opinion, the “probably global” means that when a cache line is in this state and is potentially shared with other L3 cache regions, accesses to the local L3 regions need to snoop the remote L3 regions. Meanwhile, for the cache state “TEN,” (Description, “Formerly ME, now shared”), it specifies “Yes, local.” (*Id.*)³ In my opinion, the “local” means that the targeted cache line is shared but it is known that the other non-local cores have not modified that cache line, so the local L3 cache region does not need to snoop other regions. The difference between the cache states “T” and “TEN” shows that snooping can be avoided in certain scenarios. At least because a L3 cache region does not need to snoop other L2 caches or L3 cache regions every single time, a POSITA would have understood that the cache controllers for them work together with each other and form a single logical caching agent.

18. Finally, Dr. Brogioli also disregards the disclosures about the read/claim (RC) machines in Sinharoy. A POSITA would have understood that the RC machines in both the L2 and L3 cache controllers show the cache controllers coordinate with each other. Sinharoy discloses that “the read/claim (RC) machines, which have been utilized in prior designs to manage core fetches, read–modify–write

³ Based on Dr. Brogioli’s deposition transcript, I understand that he indicated he is unfamiliar with the TEN cache state. (EX1018, 85:4-16.)

operations on behalf of core stores, and core prefetch operations, have been divided into two classes: 1) a set of eight fully functional RC machines that are incorporated into the L2 cache controller and manage core fetches and stores that hit the L2 cache and 2) a set of 24 more energy-efficient ‘lightweight’ RC machines that are incorporated into the L3 cache controller and manage prefetches and stores that miss the L2 cache.” (EX1003, 1:18.) The L2 RC machines and L3 lightweight RC machines “aggressively reorder storage operations,” thereby “enabling the high volume of system coherence traffic needed to scale robustly to a large number of sockets.” (*Id.*, 1:18-19.) These RC machines are shared across all L3 regions and hence are part of a logical single cache controller. In my opinion, if the L3 regions had separate independent cache controllers, they would each need their own RC machines and there would be no opportunity to globally reorder the storage operations. The RC machines in both the L2 and L3 cache controllers also further support that the L2 and L3 cache controllers work with each other and share information and logic across cores in certain scenarios. Thus, there is not an independent cache controller per core and the shared and portioned parts of the cache controllers form a logical “single caching agent.”

19. In sum, based on a correct and comprehensive read of the disclosures in Sinharoy, a POSITA would have understood that the cache controllers for the L2 and L3 caches do not work independently from each other, as Dr. Brogioli

incorrectly suggests. Instead, they work together to form a logical “single caching agent.” Thus, it is my opinion that Sinharoy discloses independent claim 1, 11, and 15 of the ’895 Patent.

IV. THE CHALLENGED CLAIMS ARE UNPATENTABLE OVER TANG

20. I have also been asked to respond to Dr. Brogioli’s opinions regarding the prior art reference *DMA Cache: Using On-Chip Storage to Architecturally Separate I/O Data from CPU Data for Improving I/O Performance* (“**Tang**”) (EX1004) and reviewed his opinions set forth in the Brogioli Declaration. In short, his opinions are incorrect and in my opinion, Tang discloses independent claims 1, 11, and 15 of the ’895 Patent.

21. First, based on my review of the Brogioli Declaration, he has misread and misinterpreted the Petition, my Prior Declaration, and Figure 10 of Tang to form his opinions. Furthermore, he incorrectly asserts that Tang does not disclose the caching agent which “includes a plurality of distributed portions each associated with a corresponding one of the plurality of cores.”

22. Based on my review of his Declaration, he appears to have only contended that Tang does not disclose the “single caching agent” or the caching agent including “a plurality of distributed portions each associated with a corresponding one of the plurality of cores” in claim element 1[g] (or similar elements in independent claims 11 and 15). (See Brogioli Declaration ¶¶43-46.) I

reserve the right to further supplement my responsive opinions, to the extent that Dr. Brogioli attempts to opine on any other elements not included in his Declaration.

A. Dr. Brogioli Has Misread and Misinterpreted the Petition, My Prior Declaration, and Figure 10 of Tang.

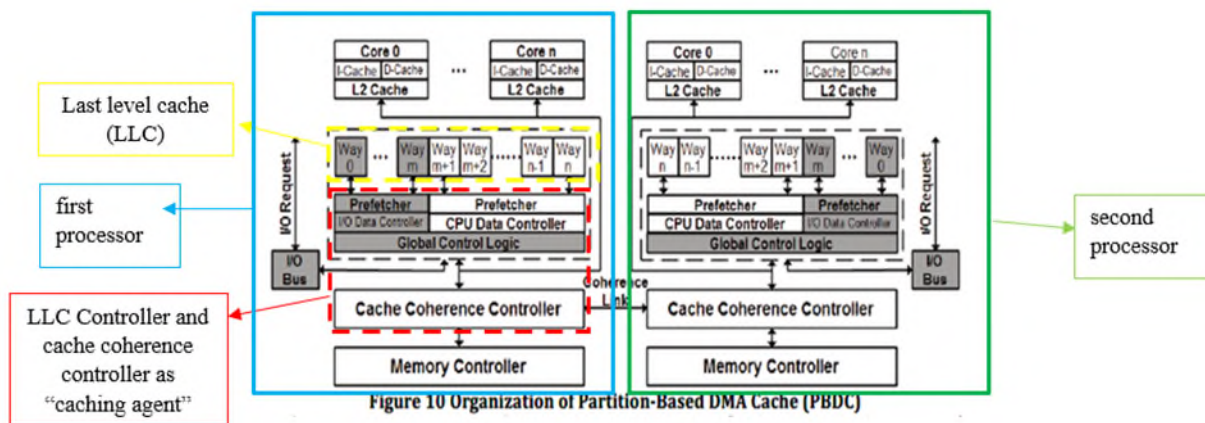
23. First, Dr. Brogioli opines that “the petition alleges that Tang’s cache coherence controller and LLC controller are collectively a ‘single caching agent for the processor’ as recited in the challenged independent claims.” (EX2002, ¶43.) This is consistent with my Prior Declaration. (EX1007, ¶127.) However, following that, he opines that “[i]f the two coherence controllers were part of the same ‘single caching agent for the processor’ no coherence link would be needed” and that the “petition alleges that the two coherence controllers one ‘logical’ unit, but Tang makes no such statement.” (EX2002, ¶44.) Neither the Petition nor my Prior Declaration makes such an argument. (See Pet., 53-55 (noting the LLC controller and the cache coherence controller on the left side of Figure 10 of Tang form a logical single caching agent); EX1007, ¶¶127-29 (same).) Instead, my Prior Declaration has clearly explained that the cache controller and the LLC controller of, for example, just the processor on the left of Figure 10 form a single caching agent for that processor. (EX1007, ¶123.) This can be clearly seen in paragraph 123 of my Prior Declaration, where, for example, it annotates Figure 10 with a red dotted line box around the single caching agent of that one processor. Thus, the

claimed feature of a “single caching agent” for one processor is met by Tang. Furthermore, my Prior Declaration also explains that Tang has a second processor shown on the right side with its own single caching agent comprised of its own cache controller and LLC controller. (EX1007, ¶141.)

24. In addition to misreading the Petition and my Prior Declaration, Dr. Brogioli has also misread Figure 10 of Tang. Dr. Brogioli alleges that there are a pair of cache coherence controllers which communicate with one another via a “coherence link.” (EX2002, ¶44.) His allegation appears to assume that the “pair” of cache coherence controllers are in a **single** processor.

25. However, according to Tang, there are **two** processors in Tang’s Figure 10. Section 3.2 of Tang discloses a partition-based DMA cache (PBDC) design. (EX1004, §3.2.) Tang explicitly discloses that “[f]igure 10 that illustrates the organization of PBDC in a **multiprocessor** system.” (*Id.*, §3.2.1. (emphasis added), Figure 10 (annotated below).) Even Dr. Brogioli admitted in his deposition that Tang discloses a processor on the left side of Figure 10, and another processor on the right side with the same structure. (EX1018, 60:11-63:10 (calling the processor on the right side a “clone” of the left side processor).) He also confirmed that the processor on the left includes an LLC controller and a cache coherence controller, on which I have opined in my Prior Declaration that they form a logical

“single caching agent.” (*Id.*, 61:22-62:15; EX1007, ¶123; EX1004, Figure 10 (annotated below).) This contradicts his arguments in his declaration.



(EX1004, Figure 10 (annotated).)

26. Furthermore, a POSITA would have understood the “coherence link” in Figure 10 connects the two processors together to form a cluster and is not a connection internal to a processor. As such, the presence of the “coherence link” is irrelevant as to whether each processor has its own respective “single caching agent.” The ’895 Patent discloses a “Quick Path Interconnect (QPI)” link can “act as an interface to another agent (such as another processor in a multiprocessor system) coupled to the processor.” (’895 Patent, 6:63-67.) A POSITA would have understood that the QPI link can also serve as a “coherence link” between the different processors. Thus, Dr. Brogioli’s allegation that my Prior Declaration

asserts there are a pair of cache coherence controllers within a single processor in Figure 10 of Tang is incorrect.

B. A POSITA Would Have Understood the Disclosures in Tang Show that the LLC Controller and Cache Coherence Controller Form a Logical “Single Caching Agent,” Which Has Distributed Portions Each Associated with A Corresponding Core.

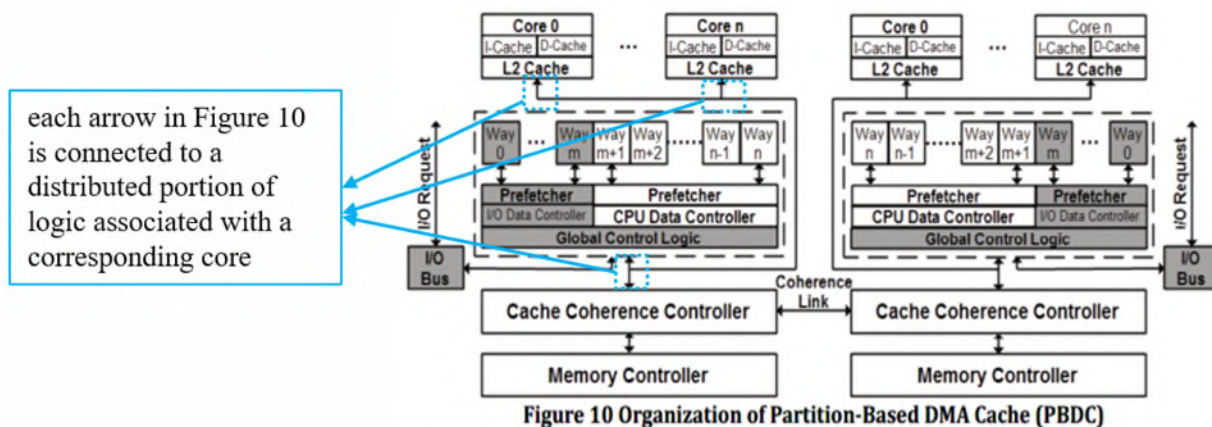
27. Second, Dr. Brogioli then alleges that “in Tang the caching agent does not include a plurality of distributed portions each associated with a corresponding one of the plurality of cores, as required by the challenged claims.” (EX2002, ¶45.) This is incorrect based on Tang’s disclosures, in particular with regard to Figure 10 and Section 3.2.2.

28. As my Prior Declaration has already stated, in the PBDC design, Tang discloses that the LLC is partitioned into “several ways” as the “dedicated DMA Cache.” (EX1007, ¶128.) Tang recognizes that some applications are I/O intensive, and others are not. (EX1004, § 2.2.) It is my opinion that a POSITA would recognize that the I/O-intensive applications may be assigned to one core with non-IO applications assigned to the remaining cores. In a simple case, Core 0 runs all I/O and Core 1 runs everything else. The I/O portion of the cache includes the I/O ways plus the cache ways for the I/O intensive applications, while the remainder of the cache is assigned to the core running other applications. In this case, the cache is partitioned with core 0 associated with the shaded I/O Ways 0-m plus a subset of Ways m+1 through Way n, while Core 1 is associated with the remaining Ways

m+1 through Way n. Tang also discloses that the “partitioned cache approaches are proposed to reduce cache interferences among the threads in multicore platform.” (EX1004, §6.) It is my opinion that a POSITA would have understood that cache partitioning usually refers to the partitioning of shared L2 or L3 caches among a set of programming threads running simultaneously on different cores. A POSITA would have understood that partitioning of an L3 cache to threads running on different cores means that the cache partitions (for example, the ways of the LLC) are each associated with a corresponding core.

29. In addition, according to Figure 10 of Tang, the LLC controller and the cache coherence controller (i.e., the “single caching agent”) include signals and logic (depicted as arrows in Figure 10) that are each connected to core 0 to core n. (EX1004, Figure 10 (annotated below).) The separate arrows to each of the cores 0-n shows that the shared logic must have signals to individual cores for handling, e.g., L2 cache misses, and those signals are sent or received by logic associated with individual cores. As shown in Figure 10, each signal (depicted as a single arrow going into a given core in Figure 10) is associated with logic in a single given core, which is a “distributed portion” of the caching agent. (EX1004, Figure 10 (annotated below).) Thus, because the cache coherence controller sends different signals to the different cores, it has “distributed portions” of logic that connect to these signals. In addition, the cache coherence controller has

“distributed portions” corresponding to each core, allowing it to separately address and send signals to the different cores. (EX1004, Figure 10 (annotated below).)



(EX1004, Figure 10 (annotated).)

30. In sum, the LLC controller and cache coherence controller of Tang, forming a logical “single caching agent,” include a “plurality of distributed portions each associated with a corresponding one of the plurality of cores.” Therefore, it is my opinion that Tang discloses independent claims 1, 11, and 15 of Tang.

V. CERTIFICATION

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true and, further, that these statements were made with the knowledge that willful false statements and the like are punishable by fine, imprisonment, or both, under Section 1001 of Title 18 of the U.S. Code.



Dr. Robert W. Horst

Dated: September 11, 2024

APPENDIX A

Materials Considered

Exhibit No.	Description
1001	U.S. Patent No. 9,575,895 (<i>Challenged Patent</i>)
1002	File History for the '895 Patent
1003	IBM POWER7 multicore server processor to Sinharoy et al. (" Sinharoy ")
1004	DMA Cache: Using On-Chip Storage to Architecturally Separate I/O Data from CPU Data for Improving I/O Performance to Tang et al. (" Tang ")
1005	Explorations of Optimal Core and Cache Placements for Chip Multiprocessor to Thomas Canhao Xu et al. (" Xu ")
1006	U.S. Patent Application Publication No. 2009/0164739 to Harikumar (" Harikumar ")
1007	Declaration of Dr. Robert Horst
1016	Architectural Implications of Cache Coherence Protocols with Network Applications on Chip Microprocessors to Kyueun Yi et al. (" Yi ")
1018	Deposition Transcript of Dr. Michael C. Brogioli
2001	Deposition Transcript of Dr. Robert Horst
2002	Declaration of Dr. Michael C. Brogioli
	Institution Decision ("Paper 7")
	Patent Owner's Preliminary Response
	Patent Owner's Response ("POR")
	Any additional materials cited in the Supplementary Declaration