

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

REALTEK SEMICONDUCTOR CORP.,
Petitioner,
v.

ADVANCED MICRO DEVICES, INC.
Patent Owner.

Case No. IPR2023-00687

U.S. Patent No. 8,854,381

**DECLARATION OF DR. MASSOUD PEDRAM IN SUPPORT OF
PETITIONER'S OPPOSITION TO PATENT OWNER'S
CONTINGENT MOTION TO AMEND**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. QUALIFICATIONS AND PERSONAL EXPERIENCE	2
III. LEGAL PRINCIPLES	2
IV. SUBSTITUTE CLAIMS 28 AND 29 IMPERMISSIBLY ADD NEW MATTER AND SHOULD BE REJECTED UNDER 35 U.S.C. §112	3
A. Proposed Claim 28 Introduces New Matter and Does Not Satisfy The Written Description Requirement of 35 U.S.C. §112	3
B. Proposed Claim 29 Introduces New Matter and Does Not Satisfy The Written Description Requirement of 35 U.S.C. §112	7
V. SUBSTITUTE CLAIM 28 IS OBVIOUS IN LIGHT OF JIAO	9
A. Claim 28	9
1. Claim 28 (Preamble): The computing system of claim 15, wherein the shader core comprises:	9
2. Claim elements 28[a]-28[c]: multiple processing elements; a first plurality of processing elements configured to execute the first task based on a first set of state data associated with the first task; a second plurality of processing elements configured to execute the second task based on a second set of state data associated with the second task, ...	9
3. Claim element 28[d]: wherein each of the plurality of tasks is decomposable into two or more related threads	10
4. Claim element 28[e]: and wherein at least one processing element is configured to receive tasks from each engine.	11
VI. SUBSTITUTE CLAIMS 28 AND 29 ARE OBVIOUS OVER WILT	12
A. Claim 28	12
1. Claim 28 (Preamble): The computing system of claim 15, wherein the shader core comprises:	12
2. Claim elements 28[a]-28[c]: multiple processing elements; a first plurality of processing elements configured to execute the first task based on a first set of state data associated with the first task; a second plurality of processing elements configured	

	to execute the second task based on a second set of state data associated with the second task.....	12
3.	Claim element 28[d]: wherein each of the plurality of tasks is decomposable into two or more related threads	13
4.	Claim element 28[e]: and wherein at least one processing element is configured to receive tasks from each engine.	14
B.	Claim 29	15
1.	Claim 29 (Preamble). The computing system of claim 15, ...	15
2.	Claim element 29[a]: wherein the plurality of engines comprises a first queue configured to receive the first task and a second queue configured to receive the second task, ...	15
3.	Claim element 29[b]: wherein the plurality of engines are arranged in parallel with one another, ...	16
4.	Claim element 29[c]: wherein each engine is configured to receive at least one of the plurality of tasks via a distinct command buffer for each engine, ...	16
5.	Claim element 29[d]: wherein each of the plurality of tasks is decomposable into two or more related threads	17
6.	Claim element 29[e]: wherein each engine is configured to load state data associated with at least one of the plurality of tasks.	17
VII.	DECLARATION	18

I. INTRODUCTION

1. I have been engaged by Realtek Semiconductor Corp. (“Petitioner”) to provide my expert opinion on invalidity related issues in connection with Patent Owner’s Contingent Motion to Amend in IPR2023-00687 (“MTA”) before the Patent Trial and Appeal Board (“Board”). I was asked to provide my opinions in this expert declaration (“Declaration”). I intend to provide testimony regarding my opinions if called to testify at a deposition.

2. This declaration addresses my opinions related to the validity of the substitute amended claims proposed by Patent Owner, namely claims 28 and 29, which would be substituted for claims 16 and 17, respectively.

3. I am being compensated for my time working in connections with this Declaration. However, my compensation is not affected by or contingent upon the substance of my opinions or on the ultimate outcome of this action or Petitioner’s Opposition to Patent Owner’s Contingent Motion to Amend.

4. The opinions expressed in this Declaration are my own and are based on my personal knowledge, my education, experience, and training, and on my understanding of the information and documents referenced in this Declaration. If called to testify in this case, I anticipate that my testimony will concern the matters addressed in this Declaration. My testimony may be affected by the positions that Advanced Micro Devices, Inc. (“Patent Owner”) and/or its expert takes on these and

other related topics after this Declaration is issued, either in the form of expert declarations or deposition testimony, or any ruling that the Board may make, and therefore I reserve the right to provide testimony regarding those issues.

II. QUALIFICATIONS AND PERSONAL EXPERIENCE

5. My qualifications and personal experience are substantially the same as presented in my Declaration in support of the Petition. See, Ex. 1003, Section II, and Ex. 1007.

III. LEGAL PRINCIPLES

6. I continued to adopt and apply the legal principles discussed in Section III of my Declaration in support of the Petition. See, Ex. 1003, Section III.

7. In addition, I understand that certain additional legal issues can be raised in response to Patent Owner's contingent Motion to Amend that were not allowed as part of the Petition. Accordingly, the following additional legal principles have been provided to me from counsel, which I have read and adopted as my own.

8. It is my understanding that an amendment to a claim must not enlarge the scope of the claims of the patent or introduce new matter." I have been informed that new matter is any addition to the claims that lacks written description support in the original application. I further understand that the disclosure in the original application must reasonably convey to those skilled in the

art that the inventor had possession of the claimed subject matter as of the filing date to satisfy the written description requirement. “Possession” of the claimed subject matter can and must be shown by descriptive means such as words, structures, and figures that set forth fully the claimed invention. I understand that it is not sufficient for purposes of the written description requirement that the disclosure, when combined with the knowledge in the art, would lead one to speculate as to modifications that the inventor might have envisioned, but failed to disclose. In other words, it is not enough to show that the proposed substitute claim would have been obvious based on the disclosure in the patent. Rather, the written description must actually or inherently disclose all of the elements of the proposed substitute claim. I understand inherently in this context to mean that the written description must necessarily disclose the proposed substitute claim.

IV. SUBSTITUTE CLAIMS 28 AND 29 IMPERMISSIBLY ADD NEW MATTER AND SHOULD BE REJECTED UNDER 35 U.S.C. §112

A. Proposed Claim 28 Introduces New Matter and Does Not Satisfy The Written Description Requirement of 35 U.S.C. §112

9. AMD proposes to substitute claim 28 for claim 16, which depends from claim 15. Substitute claim 28 is as follows:

28. The computing system of claim 15, wherein the shader core comprises:

- [a] multiple processing elements;
- [b] a first plurality of processing elements configured to

- execute the first task based on a first set of state data associated with the first task; and
- [c] a second plurality of processing elements configured to execute the second task based on a second set of state data associated with the second task,
- [d] wherein each of the plurality of tasks is decomposable into two or more related threads
- [e] and *wherein at least one processing element is configured to receive tasks from each engine.*

10. Claim elements 28[b] and [c] are recited in current claim 16, whereas claim elements 28[a], [d] and [e] are added.

11. Claim 28 pertains to the “shader core,” as defined in the ’381 patent as “an array of processing elements.” Ex. 1001, 1:59-61. However, the patent lacks explicit details regarding the specific arrangement of these processing elements within the shader core, particularly in relation to the claimed “plurality of engines.” Thus, in my opinion and as explained in more detail below, proposed claim element 28[e] introduces new matter and lacks written description support.

12. Proposed claim element 28[e] mandates that “at least one” processing element is “configured to” receive multiple tasks from “each engine.” However, in my opinion, this level of detail is not disclosed in the ’381 patent. There are no specific configurations shown in the figures or disclosed in the text regarding the relationship between any processing element and any of the disclosed engines. In

fact, there is no detail provided at all regarding the configuration of the processing elements in the '381 patent.

13. As stated above, the '381 patent describes the shader core as “an array of processing elements.” Ex. 1001, 1:59-61. The '381 patent treats the “array of processing elements” within the shader core as a black box, with no disclosure of any specific arrangement that the patentee may have possessed at the time of the invention. The '381 patent figures provide no help. For example, figures 3A, 3B, and 4 have only a box with annotation “shader core.” No processing elements are illustrated or mentioned.

14. Figure 2 includes an illustration of processing elements, but without any configuration details, either, as to the processing elements themselves, their relationship to each other, their configuration within the overall shader core, and their relationship to any of the components either inside or outside of the shader core. The only detail shown in Figure 2 is the “scalability” dots showing that there can be any number of processing elements in the rows and columns. A POSITA reading these figures would not be able to know any specific configuration for any processing engine or any subset of processing engines, within the shader core or to component outside the shader core, beyond speculation. Although many configurations can be assumed or dreamed up, none are inherent or necessarily

disclosed. The corresponding descriptions for the figures do not provide any additional detail.

15. Further, Dr. Hart's color-coding exercise to illustrate possible ways that the processing elements may receive tasks from the engines is just that, an exercise. Dr. Hart is not showing inherent or necessarily disclosed configurations. Any number of other possible configurations or task flow, which is really what Dr. Hart is showing, can be imagined by a POSITA. For example, Figure 2 of the '381 patent shows a system that has different types of engines, including a 3D engine (202D) and a low latency 3D engine (202E). Consistent with that disclosure, a POSITA would consider that the patentee may have intended configuring the shader core where certain of the processing elements handle low latency 3D tasks while others do not. Consequently, those processing elements would not receive tasks from each engine. Alternatively, based on the '381 patent disclosure that the system needs to be able to run both graphics processing tasks while running general compute tasks, a POSITA would consider that the patentee may have intended for a subset of the processing elements handle only general compute tasks and may not receive a task from each engine. Further, because spatial and temporal partitioning are not required by the claims, a POSITA may not consider them used at all and configured the processing elements without partitioning where each engines has a dedicated processing element. And so on. It simply is not true

that a POSITA would have considered Dr. Hart's speculative configurations to be inherent or necessarily disclosed in the '381 patent.

16. Instead, all that Dr. Hart has shown is that these configurations are possible. At best, it shows that the system may be capable of such configurations (although without knowing how the shader core was intended to be configured even that is speculation). Perhaps, Dr. Hart is arguing that these configurations are obvious, but, even if true, my understanding is that is not enough to prevent it from being new matter. Further, it is not clear to me that such configurations are obvious. In my opinion, Dr. Hart is merely starting from the amended claim he wants and then using them as a roadmap to propose how the specification could allow for those new limitations. In my opinion, that is not evidence of what the patentee intended.

17. Thus, it is my opinion that proposed claim 28 adds new matter and lacks written description.

B. Proposed Claim 29 Introduces New Matter and Does Not Satisfy The Written Description Requirement of 35 U.S.C. §112

18. AMD proposes to substitute claim 29 for claim 17, which depends from claim 15. Substitute claim 29 is as follows:

29. The computing system of claim 15,

- [a] wherein the plurality of engines comprises a first queue
configured to receive the first task and a second queue
configured to receive the second task,
- [b] wherein the plurality of engines are arranged in parallel
with one another,
- [c] wherein each engine is configured to receive at least one of
the plurality of tasks via a distinct command buffer for
each engine,
- [d] wherein each of the plurality of tasks is decomposable into
two or more related threads, and
- [e] *wherein each engine is configured to load state data
associated with at least one of the plurality of tasks.*

19. Claim element 29[a] is current claim 17, whereas claim elements 29[b], [c], [d] and [e] are added.

20. Like claim element 28[e], the application for the '381 patent does not have written description for proposed claim element 29[e]. Thus, in my opinion, claim 29 introduces new matter and lacks written description support.

21. Specifically, the application for the '381 patent fails to disclose a setup where each engine is specifically designed to load state data pertaining to at least one task from the plurality of tasks. The '381 specification only mentions that the collective plurality of engines is responsible for loading state data, with no mention of each individual engine performing this action. The engines are treated as a group with respect to loading state data, without any detail for any one engine.

For example, I find only two references to the loading of state data by engines.

Once in the Abstract that refers to the “plurality” of virtual engines loading state data, and once at column 3, lines 3-6, that has the same disclosure.

22. Without any detail in the application for the '381 patent, Dr. Hart again can only speculate what the patentee intended with respect to “each” engine configurations. Thus, the claimed element “each engine is configured to load state data” is neither expressly or inherently disclosed, nor is it necessarily present in the disclosure. In my opinion, it has no written description and is new matter.

V. SUBSTITUTE CLAIM 28 IS OBVIOUS IN LIGHT OF JIAO

A. Claim 28

1. Claim 28 (Preamble): *The computing system of claim 15, wherein the shader core comprises:*

23. For all of the reasons provided in my Declarations associated with the Petition and Petitioner’s Reply, the claim elements of claim 15 are obvious in light of Jiao. *See, e.g.,* Ex. 1003 ¶¶ 91-112.

2. Claim elements 28[a]-28[c]: *multiple processing elements; a first plurality of processing elements configured to execute the first task based on a first set of state data associated with the first task; a second plurality of processing elements configured to execute the second task based on a second set of state data associated with the second task,*

24. For all of the reasons provided in my Declarations associated with the Petition and Petitioner’s Reply with respect to claim 16, Jiao renders obvious the

limitations of claim elements 28[a]-[c]. *See, e.g.*, Ex. 1003 ¶ 113. For example, Jiao discloses a shader core having multiple processing elements, which I identified as the plurality of EUs 420. *Id.* Each EU is the claimed “processing element.” *Id.* Each EU is identical in design and configured to execute multiple threads simultaneously or substantially simultaneously based on respective state data including constants, rendering context information, and graphics data. *Id.* Thus, they are configured such that each can be executing a different task. *Id.*

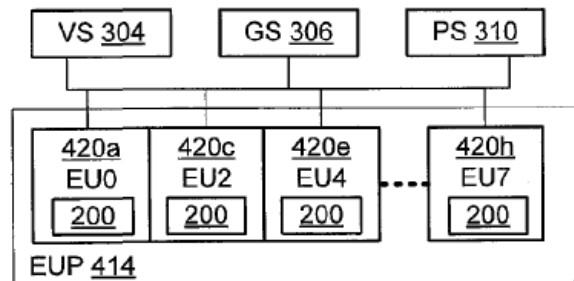
3. Claim element 28[d]: *wherein each of the plurality of tasks is decomposable into two or more related threads*

25. For all of the reasons provided in my Declarations associated with the Petition and Petitioner’s Reply, Jiao disclose and/or renders obvious the claimed plurality of tasks. *See, e.g.*, Ex. 1003 ¶¶ 91-113.

26. Whether or not a task is decomposable into one thread, or two or more threads is merely a matter of complexity of the task. Simple tasks may decompose into a single thread, while most and more complex threads will decompose into two or more threads. Because Jiao’s “tasks” can encompass a very wide range of programmable operations to meet the needs of different computer programs (e.g., *id.* at 5:47-51), it is inherent, or at least obvious to a POSITA, that each of a plurality of “tasks” is decomposable into two or more related threads. *See also* Pet. 43; Ex. 1003 at ¶ 103.

4. **Claim element 28[e]: *and wherein at least one processing element is configured to receive tasks from each engine.***

27. In my opinion, this is expressly disclosed in Jiao. At a minimum, Figure 4 and its corresponding description renders it obvious. Figure 4 (as shown below) expressly identifies a configuration in which each processing element (420a-420h) can receive a task from each engine (VS 304, GS 306, and PS 310).



28. The corresponding description of Figure 4 further supports what is express in the figure. Jiao discloses that as “various shader tasks are generated by the vertex shader 304, geometry shader 306, and pixel shader 310, they are delivered to the respective execution units 420 to be carried out.” Ex. 1005 at 9:35-39. Because the EU pool control unit assigns tasks from *each* of the plurality of engines to *each* of the Execution Units (e.g., eight EUs), and each Execution Unit is a “processing element,” it is obvious that at least one processing element (EU) is configured to receive tasks from each engine. Any lack of clarity in the description is cleared up with the express configuration disclosed in Figure 4.

29. Based on the above, it is my opinion that Jiao's disclosure, when taken as a whole, at least renders obvious substitute claim 28.

VI. SUBSTITUTE CLAIMS 28 AND 29 ARE OBVIOUS OVER WILT

A. Claim 28

1. **Claim 28 (Preamble):** *The computing system of claim 15, wherein the shader core comprises:*

30. For all of the reasons provided in my Declarations associated with the Petition and Petitioner's Reply, the claim elements of claim 15 are obvious in light of Wilt. *See, e.g.,* Ex. 1003 ¶¶ 64-77.

2. **Claim elements 28[a]-28[c]:** *multiple processing elements; a first plurality of processing elements configured to execute the first task based on a first set of state data associated with the first task; a second plurality of processing elements configured to execute the second task based on a second set of state data associated with the second task*

31. For all of the reasons provided in my Declarations associated with the Petition and Petitioner's Reply with respect to claim 16, Wilt renders obvious claim elements 28[a]-28[c]. *See, e.g.,* Ex. 1003 ¶¶ 78-81.

32. With respect to 28[a], claim elements 28[b] and 28[c] inherently already claims "multiple processing elements," which were recited in claim 16 that I addressed in my Declarations. *See, e.g.,* Ex. 1003 ¶¶ 78-81.

3. Claim element 28[d]: *wherein each of the plurality of tasks is decomposable into two or more related threads*

33. As explained above with respect to Jiao, tasks with two or more related threads are obvious to a POSITA. See discussion of Jiao and claim element 28[d] above. Like Jiao, Wilt's disclosed "tasks" encompass a very wide range of programmable operations. It is therefore at least obvious to a POSITA, that when a processing "task" is assigned to a multithreaded processing "core" with tens or hundreds of threads for execution, such as disclosed by Wilt, that each of a plurality of "tasks" is decomposable into two or more related threads.

34. Regardless, whether a task is decomposable into one thread or "two or more" threads is merely a matter of degree or complexity of the task at hand and trivially obvious. A POSITA would have understood that while some tasks may be very simple and decompose into a single thread, most tasks will be more complex and decompose into two or more threads. Ex. 2007, 11:7-16. A POSITA would have understood that graphics processing, as described in Wilt, would never be comprised of *only* single threaded tasks, which is why the processing cores in Wilt are designed to execute tasks using tens or hundreds of threads. *See also* Ex. 1004 at 5:59-6:6 (describing embodiments where a single instruction results in parallel execution of a large number of threads). *See also id.* at 3:60-4:25, 4:61-5:4, 6:13-27, 8:15-29, 10:21-51).

4. Claim element 28[e]: *and wherein at least one processing element is configured to receive tasks from each engine.*

35. As I explained in my previous Declarations, the Wilt system includes a front end 212, work distribution unit 210, and core interface 303 within each PPU that can collectively function as a single “engine,” while the set of cores 208 present in and across all the PPUs collectively constitute the “shader core.” This configuration is illustrated in Figure 2-3 of Ex. 1004. Wilt further reveals that communication among PPUs is facilitated through the bidirectional communication path 113 and the memory bridge 105, as detailed in Ex. 1004, 4:7-23 and Figure 3. The work distribution unit 210 in Wilt is configured to implement various algorithms for distributing tasks among the cores. For instance, each core 208 signals readiness when it can accept a new task, and the work distribution unit assigns the task to a core that is ready. If no core is ready, the work distribution unit can hold the task. Wilt explicitly mentions that alternative algorithms for task distribution can be employed, and that the specific method used by the work distribution unit is not crucial to the invention.

36. Given that the bidirectional communication path 113 and memory bridge 105 connect each engine within every PPU, it can be inferred from Wilt's explicit disclosure that if a PPU has no cores ready to accept a task, the work distribution unit 210 may route the task back through the communication path and

memory bridge to be assigned to a different PPU with available cores. *Id.*; Ex. 2007, 42:21-45:6. Consequently, each core 208 in a PPU is designed to receive tasks not only from the associated engine of that PPU but also from every engine across all other PPUs.

37. This bidirectional task distribution mechanism is further implied by Wilt's description of how each PPU 202 is equipped with a host interface for system-wide communication via communication path 113. The Host interface 206 can generate and receive packets for transmission via the communication path, indicating the seamless exchange of information within the system. Ex. 1004 at 3:49-67. Thus, Wilt, at least, suggests that “at least one processing element is configured to receive tasks from each engine.”

B. Claim 29

1. Claim 29 (Preamble). *The computing system of claim 15,*

38. For all of the reasons provided in my Declarations associated with the Petition and Petitioner's Reply, the claim elements of claim 15 are obvious in light of Wilt. *See, e.g.*, Ex. 1003 ¶¶ 64-77.

2. Claim element 29[a]: *wherein the plurality of engines comprises a first queue configured to receive the first task and a second queue configured to receive the second task,*

39. Claim element 29[a] was recited in original claim 17. For all of the reasons provided in my Declarations associated with the Petition and Petitioner's

Reply, the claim elements of claim 17 are obvious in light of Wilt. *See, e.g.,* Ex. 1003 ¶¶ 82-85.

3. **Claim element 29[b]: *wherein the plurality of engines are arranged in parallel with one another,***

40. Claim element 29[b] is expressly disclosed in Wilt. I identified that each PPU has a front end 212, work distribution unit 210, and core interface 303 that collectively serve as a single “engine” within each PPU. This configuration is illustrated in Figure 2-3 of Ex. 1004. The PPUs (ranging from PPU 0 to PPU U-1) in Wilt are structured in a parallel arrangement, with these one or more PPUs constituting part of the "parallel processing subsystem" as depicted in Figure 2 and discussed in 2:27-28, 3:36-48 of Ex. 1004. The presence of a single "engine" within each PPU indicates that the multiple engines in Wilt are aligned in parallel configuration.

4. **Claim element 29[c]: *wherein each engine is configured to receive at least one of the plurality of tasks via a distinct command buffer for each engine,***

41. Wilt's host interface 206 within each PPU meets the requirement of the claimed “distinct command buffer” for each “engine.” The ’381 patent articulates the function of the “command buffer” as receiving tasks from the “scheduling module” and supplying these tasks to the “engines,” as described in Ex. 1001, 2:8-9, 10:25-39 and represented in Figure 4. Wilt mirrors this concept by disclosing that within each PPU 202, the host interface 206 collects all incoming

packets (designated as the claimed “tasks”) from the application software running on the CPU 102 (representing the claimed “scheduling module”) via the communication path 113, as detailed in Ex. 1004, 3:49-67. Notice that “packets (or other signals)” in this excerpt refer to data packets and control signals for the communication path 113 (which is a PCI-E link in a disclosed embodiment in Wilt.) A POISTA would understand that these packets contain the “commands related to [defining] processing tasks.” Subsequently, the host interface dispatches the received tasks to the “engine” within each PPU, namely the front end unit 212, work distribution unit 210, and core interface 303 of each PPU. Ex. 1004, 3:60-61 and 8:15-21. As a result, the interplay between the host interface and the corresponding “engine” in each PPU, as summarized above, either expressly reveals or renders obvious the presence of the claimed “wherein each engine is configured to receive at least one of the plurality of tasks via a distinct command buffer for each engine.”

5. Claim element 29[d]: *wherein each of the plurality of tasks is decomposable into two or more related threads*

42. See discussion above with respect to claim element 28[d].

6. Claim element 29[e]: *wherein each engine is configured to load state data associated with at least one of the plurality of tasks*

43. Wilt employs the term “state parameters” to denote state data, as highlighted in the document (Ex. 1004 at 10:25-32). According to Wilt's operation,

tasks are first acquired by the host interface 206 in each PPU, which subsequently transfers these tasks through the front end unit 212, work distribution unit 210, and the core interface unit 303 to the “shader core.” This process involves loading the state parameters (state data) related to the task by the core interface unit 303 before routing the tasks to the cores 208 for processing, Ex. 1004 at 8:15-21. Hence, when considering that each “engine” encompasses the front end unit 212, work distribution unit 210, and core interface unit 303 within every PPU, it is evident that every “engine” is programmed to load state data associated with at least one of the plurality of tasks it receives.

VII. DECLARATION

44. In signing this declaration, I understand that this declaration will be filed as evidence in a contested case before the patent Trial and Appeal Board of the USPTO.

45. I am aware that willful false statements and the like made by me in connection with this declaration are punishable by fine or imprisonment, or both (18 U.S.C. § 1001) and may jeopardize the validity of this Opposition and/or the Petition. All statements herein made by me of my own knowledge are true and all statements made on information and belief are believed to be true. I declare under penalty of perjury that the foregoing is true and correct.

Executed this 10th day of April, 2024 at Los Angeles, California.

Respectfully Submitted,



Massoud Pedram, Ph.D.