UNITED STATES PATENT AND TRADEMARK OFFICE

---

**BEFORE THE PATENT TRIAL AND APPEAL BOARD**

---

SAMSUNG ELECTRONICS CO., LTD.,
SAMSUNG ELECTRONICS AMERICA, INC., and QUALCOMM, INC.,
Petitioners

v.

DAEDALUS PRIME LLC.
Patent Owner.

**DECLARATION OF TREVOR MUDGE**
**IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW OF**
**U.S. PATENT NO. 10,049,080**

I. INTRODUCTION ..... 1
A. Qualifications ..... 1
B. Materials Considered ..... 3
II. LEGAL STANDARDS ..... 5
A. Anticipation ..... 5
B. Obviousness ..... 6
C. Level of Ordinary Skill in the Art ..... 9
D. Claim Construction ..... 11
III. SUMMARY OF GROUNDS ..... 12
IV. THE ’080 PATENT ..... 12
A. Overview of the ’080 Patent ..... 12
B. Prosecution History of the ’080 Patent ..... 16
V. TECHNOLOGY BACKGROUND ..... 18
A. Relationship between power, voltage, and frequency ..... 19
B. Multi-core processing approaches ..... 20
C. Processor Microarchitectures ..... 24
D. Heterogeneous multi-core processors ..... 25
VI. THE PRIOR ART IN THE APPLIED INVALIDITY GROUNDS ..... 27
A. Mathieson (Ex-1005) ..... 27
B. Carmack (Ex-1006) ..... 30
C. Sutardja – Sutardja ’748 (Ex-1007) and Sutardja ’785 (Ex-1008) ..... 32
D. Rychlik (Ex-1009) ..... 34
VII. CLAIM CONSTRUCTION ..... 35
VIII. DETAILED EXPLANATION OF THE UNPATENTABILITY GROUNDS ..... 36
A. Ground 1: Claims 1-4, 7-12, 15-20, 23-24 are rendered obvious by Sutardja (Ex-1007, incorporating Ex-1008) ..... 36
1. Independent Claim 1 ..... 36



Petitioner Mercedes Ex-1002

a. Element 1[pre]: A multi-core processor comprising: .....36

b. Element 1[a][i]: a first plurality of cores and a second plurality of cores that support a same instruction set, .....37

c. Element 1[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and.....46

d. Element 1[b][i]: power management hardware to, from a state where the first plurality of cores and the second plurality of cores are enabled, disable all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores, .....47

e. Element 1[b][ii]: wherein an operating system to execute on the multi-core processor is to monitor a demand for the multi-core processor and control the power management hardware based on the demand. .....58

2. Dependent Claim 2: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores. .....61

3. Dependent Claim 3: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that consume less power than corresponding logic gates of the first plurality of cores. .....63

4. Dependent Claim 4: The multi-core processor of claim 1, wherein the second plurality of cores each have a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores. .....64

5. Dependent Claim 7: The multi-core processor of claim 1, wherein the first plurality of cores are at a maximum operating frequency in the state. .....65

6. Dependent Claim 8: The multi-core processor of claim 1, wherein ..... 66

a. Element 8[a]: the power management hardware is to enable all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores, ..... 66

b. Element 8[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand. ..... 71

7. Independent Claims 9 and 17: ..... 71

a. Element 9[preamble]: A method comprising: ..... 71

b. Element 17[preamble]: A non-transitory machine readable medium containing program code that when processed by a machine causes a method to be performed, the method comprising: ..... 71

c. Elements 9[a][i] and 17[a][i]: operating a multi-core processor such that a first plurality of cores and a second plurality of cores execute a same instruction set, ..... 72

d. Elements 9[a][ii] and 17[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and ..... 72

e. Elements 9[b][i] and 17[b][i]: disabling with power management hardware, from a state where the first plurality of cores and the second plurality of cores are enabled, all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores, ..... 72

f. Element 9[b][ii] and 17[b][ii]: wherein an operating system executing on the multi-core processor monitors a demand for the multi-core processor and controls the power management hardware based on the demand. ..... 72



Petitioner Mercedes Ex-1002

8. Dependent Claims 10 and 18: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating of the second plurality of cores comprises driving logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores. .....73

9. Dependent Claims 11 and 19: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating of the second plurality of cores comprises driving logic gates that consume less power than corresponding logic gates of the first plurality of cores. .....73

10. Dependent Claims 12 and 20: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the second plurality of cores at a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores. .....73

11. Dependent Claims 15 and 23: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the first plurality of cores at a maximum operating frequency in the state.....73

12. Dependent Claims 16 and 24: The [method of claim 9/non-transitory machine readable medium of claim 17], further comprising.....74

a. Elements 16[a] and 24[a]: enabling, with the power management hardware, all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores, .....74

b. Elements 16[b] and 24[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.....74

B. Ground 2: Claims 5-6, 13-14, and 21-22 are rendered obvious by Sutardja in view of Rychlik.....74

1. Dependent Claims 5, 13, and 21: .....74



Petitioner Mercedes Ex-1002

a. Elements 5[a], 13[a], 21[a]: The [multi-core processor of claim 1/method of claim 9/non-transitory machine readable medium of claim 17], [wherein the power management hardware is to disable/further comprising disabling, with the power management hardware,] an additional core of the second plurality of cores for each continued drop in demand below a next lower threshold until one core of the second plurality of cores remains enabled, and ....74

b. Elements 5[b], 13[b], 21[b]: [lower/lowering] an operating frequency or a supply voltage of the one core of the second plurality of cores as demand drops below a next lower threshold. ....80

2. Dependent Claims 6, 14, 22: [The multi-core processor of claim 5/method of claim 13/The non-transitory machine readable medium of claim 21], [wherein the power management hardware is to raise/further comprising raising, with the power management hardware,] a supply voltage or an operating frequency of said one core in response to higher demand....83

C. Ground 3: Claims 7, 15, and 23 are rendered obvious by Sutardja in view of Carmack ....84

1. Dependent Claims 7, 15, 23: The [multi-core processor of claim 1/method of claim 9/non-transitory machine readable medium of claim 17], wherein [the operating comprises operating] the first plurality of cores [are] at a maximum operating frequency in the state. ....84

D. Ground 4: Claims 1-4, 7-12, 15-20, and 23-24 are rendered obvious by Mathieson in view of Sutardja....86

1. A POSITA would have been motivated to combine Mathieson with Sutardja ....86

2. Independent Claim 1 ....88

a. Element 1[preamble]: A multi-core processor comprising: ....88



Petitioner Mercedes Ex-1002

b. Element 1[a][i]: a first plurality of cores and a second plurality of cores that support a same instruction set, ..... 89

c. Element 1[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and ..... 93

d. Element 1[b][i]: power management hardware to, from a state where the first plurality of cores and the second plurality of cores are enabled, disable all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores, ..... 96

e. Element 1[b][ii]: wherein an operating system to execute on the multi-core processor is to monitor a demand for the multi-core processor and control the power management hardware based on the demand. ..... 104

3. Dependent Claim 2: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores. ..... 107

4. Dependent Claim 3: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that consume less power than corresponding logic gates of the first plurality of cores. ..... 111

5. Dependent Claim 4: The multi-core processor of claim 1, wherein the second plurality of cores each have a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores. ..... 112

6. Dependent Claim 7: The multi-core processor of claim 1, wherein the first plurality of cores are at a maximum operating frequency in the state. ..... 114

7. Dependent Claim 8: The multi-core processor of claim 1, wherein ..... 116



Petitioner Mercedes Ex-1002

a. Element 8[a]: the power management hardware is to enable all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores, .....116

b. Element 8[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand. .....121

8. Independent Claims 9 and 17: .....121

a. Element 9[preamble]: A method comprising: .....121

b. Element 17[preamble]: A non-transitory machine readable medium containing program code that when processed by a machine causes a method to be performed, the method comprising: .....122

c. Elements 9[a][i] and 17[a][i]: operating a multi-core processor such that a first plurality of cores and a second plurality of cores execute a same instruction set, .....122

d. Elements 9[a][ii] and 17[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and .....122

e. Elements 9[b][i] and 17[b][i]: disabling with power management hardware, from a state where the first plurality of cores and the second plurality of cores are enabled, all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores, .....122

f. Element 9[b][ii] and 17[b][ii]: wherein an operating system executing on the multi-core processor monitors a demand for the multi-core processor and controls the power management hardware based on the demand. .....123

9. Dependent Claims 10 and 18: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein

the operating of the second plurality of cores comprises driving logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores. .....123

10. Dependent Claims 11 and 19: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating of the second plurality of cores comprises driving logic gates that consume less power than corresponding logic gates of the first plurality of cores. .....123

11. Dependent Claims 12 and 20: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the second plurality of cores at a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores. .....123

12. Dependent Claims 15 and 23: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the first plurality of cores at a maximum operating frequency in the state.....124

13. Dependent Claims 16 and 24: The [method of claim 9/non-transitory machine readable medium of claim 17], further comprising.....124

a. Elements 16[a] and 24[a]: enabling, with the power management hardware, all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores, .....124

b. Elements 16[b] and 24[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.....124

E. Ground 5: Claims 5-6, 13-14, and 21-22 are rendered obvious by Mathieson/Sutardja in view of Rychlik.....125

1. Dependent Claims 5, 13, and 21: .....125



Petitioner Mercedes Ex-1002

a. Elements 5[a], 13[a], 21[a]: The [multi-core processor of claim 1/method of claim 9/non-transitory machine readable medium of claim 17], [wherein the power management hardware is to disable/further comprising disabling, with the power management hardware,] an additional core of the second plurality of cores for each continued drop in demand below a next lower threshold until one core of the second plurality of cores remains enabled, and ..... 125

b. Elements 5[b], 13[b], 21[b]: [lower/lowering] an operating frequency or a supply voltage of the one core of the second plurality of cores as demand drops below a next lower threshold. ..... 126

2. Dependent Claims 6, 14, 22: [The multi-core processor of claim 5/method of claim 13/The non-transitory machine readable medium of claim 21], [wherein the power management hardware is to raise/further comprising raising, with the power management hardware,] a supply voltage or an operating frequency of said one core in response to higher demand..... 127

IX. CONCLUSION..... 127

# I. INTRODUCTION

1. I have been retained by Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. (collectively, "Samsung" or "Petitioner"), as an independent expert in this proceeding before the Patent Trial and Appeal Board ("PTAB" or "Board"). I understand that Samsung is requesting that the Board institute an *inter partes* review ("IPR") proceeding of U.S. Patent No. 10,049,080 ("the '080 Patent") (Ex-1001), currently assigned to Daedalus Prime LLC ("Patent Owner" or "PO").

2. I am not and have never been an employee of Samsung. I am being compensated at my usual and customary rate of $600 per hour. No part of my compensation depends on the outcome of this proceeding, and I have no other interest in this proceeding.

3. I have been asked to provide my independent analysis of the '080 Patent in light of the prior art publications cited below. I have also been asked to consider the state of the art and prior art available as of December 22, 2011. Based on the prior art discussed in this declaration, it is my opinion that Claims 1-24 of the '080 Patent is unpatentable for the reasons provided below.

## A. Qualifications

4. I am currently a the Bredt Family Professor of Computer Science and Engineering at the University of Michigan, Ann Arbor. All of my opinions stated in

this declaration are based on my own personal knowledge and professional judgment. In forming my opinions, I have relied on my over 45 years of research, academic, industry, and consulting engineering experience in IC (integrated circuit) processing, semiconductor devices, and computer architecture with an emphasis on power and energy control.

5. I am over 18 years of age and, if I am called upon to do so, I would be competent to testify as to the matters set forth herein. I understand that a copy of my current curriculum vitae, which details my education and professional and academic experience, is being submitted by Petitioner as Exhibit 1003. The following provides an overview of some of my experience that is relevant to the matters set forth in this declaration.

6. I received the Ph.D. in Computer Science from the University of Illinois, Urbana. I am the Bredt Family Professor of Computer Science and Engineering at the University. I am the author of numerous papers on computer architecture, programming languages, VLSI design, and computer vision. I have chaired 57 PhD theses in these areas. In 2014 I received the ACM/IEEE CS Eckert-Mauchly Award for "pioneering contributions to low-power computer architecture and its interaction with technology." This is known as the computer architecture community's most prestigious award. I also received the University of Illinois

Distinguished Alumni Award. I am a Life Fellow of the IEEE, a Fellow of the ACM, and a member of the IET and the British Computer Society.

7. Based on my experience and education, I believe that I am qualified to opine as to the knowledge and level of skill of one of ordinary skill in the art at the time of the alleged invention of the ’080 Patent, as well as the state of the art at that time.

**B. Materials Considered**

8. In forming my opinions, I have reviewed the following documents:[1]

| | |
|---|---|
| Ex-1001 | U.S. Patent No. 10,049,080 to George et al. (“the ’080 Patent”) |
| Ex-1002 | Declaration of Dr. Trevor Mudge |
| Ex-1003 | Curriculum Vitae of Dr. Trevor Mudge |
| Ex-1004 | Prosecution History of the ’080 Patent (Application No. 15/431,527) |
| Ex-1005 | U.S. Patent Pub. No. 2011/0213950 to Mathieson et al. (“Mathieson”) |
| Ex-1006 | U.S. Patent Pub. No. 2009/0309243 to Carmack et al. (“Carmack”) |
| Ex-1007 | U.S. Patent Pub. No. 2008/0288748 to Sutardja et al. (“Sutardja ’748”) |
| Ex-1008 | U.S. Patent Pub. No. 2007/0083785 to Sutardja (“Sutardja ’785”) |
| Ex-1009 | U.S. Patent Pub. No. 2011/0145615 to Rychlik et al. (“Rychlik”) |
| Ex-1010 | Prosecution History of U.S. Patent No. 9,569,278 (“the ’278 Patent”) |

[1] Four-digit pin citations that begin with 0 are to the branded numbers added by Samsung in the bottom right corner of the exhibits. All other pin citations are to original page, column, paragraph, or line numbers.

| | |
|---|---|
| Ex-1011 | INTENTIONALLY LEFT BLANK |
| Ex-1012 | INTENTIONALLY LEFT BLANK |
| Ex-1013 | INTENTIONALLY LEFT BLANK |
| Ex-1014 | INTENTIONALLY LEFT BLANK |
| Ex-1015 | INTENTIONALLY LEFT BLANK |
| Ex-1016 | INTENTIONALLY LEFT BLANK |
| Ex-1017 | INTENTIONALLY LEFT BLANK |
| Ex-1018 | INTENTIONALLY LEFT BLANK |
| Ex-1019 | INTENTIONALLY LEFT BLANK |
| Ex-1020 | Claim Mapping Table |
| Ex-1021 | INTENTIONALLY LEFT BLANK |
| Ex-1022 | U.S. Patent Pub. No. 2006/0095807 to Grochowski ("Grochowski") |
| Ex-1023 | U.S. Patent Pub. No. 2012/0317568 to Aasheim ("Aasheim") |
| Ex-1024 | Jeffrey C. Mogul et al., *Operating Systems and Asymmetric Single-ISA CMPs: The Potential for Saving Energy*, Hewlett-Packard Development Company, L.P. (2007) |
| Ex-1025 | Juan Carlos Saez et al., *Operating System Support for Mitigating Software Scalability Bottlenecks on Asymmetric Multicore Processors*, ACM 978-1-4503-004-5/10/05 (2010) |
| Ex-1026 | U.S. Patent No. 7,093,147 to Farkas et al. ("Farkas") |
| Ex-1027 | Charles Lefurgy et al., *Energy Management for Commercial Servers*, Computer 39 (Dec. 2003). |
| Ex-1028 | Yushi Shen et al., *Enabling the New Era of Cloud Computing: Data Security, Transfer, and Management* (Information Science Reference 2014). |

| | |
|---|---|
| Ex-1029 | Stefanos Kaxiras and Margaret Martonosi, *Computer Architecture Techniques for Power-Efficiency*, in Synthesis Lectures on Computer Architecture #4 (Morgan & Claypool 2008). |
| Ex-1030 | Vasanth Venkatachalam and Michael Franz, *Power Reduction Techniques For Microprocessor Systems*, 37 ACM Computing Surveys 195 (2005). |
| Ex-1031 | Euiseong Seo et al., *Energy Efficient Scheduling of Real-Time Tasks on Multicore Processors*, 19 IEEE Transactions on Parallel and Distributed Systems 1540 (Nov. 2008). |
| Ex-1032 | Rakesh Kumar et al., *Single-ISA Heterogeneous Multi-Core Architectures: The Potential for Processor Power Reduction*, Proceedings of the 36th International Symposium on Microarchitecture (MICRO-36 2003), IEEE Computer Society (2003). |
| Ex-1033 | U.S. Patent No. 8,615,647 to Hum et al. ("Hum") |

## II. LEGAL STANDARDS

9. In forming my opinions and considering the subject matter of the '080 Patent and its claims in light of the prior art, I am relying on certain legal principles that counsel in this case explained to me. My understanding of these concepts is summarized below.

10. I understand that earlier publications and patents may act to render a patent unpatentable for one of two reasons: (1) anticipation, and (2) obviousness.

### A. Anticipation

11. It is my understanding that the claims of a patent are anticipated by a prior art reference if each and every element of the claim is found either explicitly or inherently in the reference. I understand that inherency requires a showing that

the missing descriptive matter in the claim is necessarily present in the allegedly anticipating reference, and that it would have been so recognized by a person of ordinary skill in the art ("POSITA").

12. I understand that when a challenged claim covers several structures, either generically or as alternatives, the claim is deemed anticipated if any of the structures within the scope of the claim is found in the prior art reference.

13. Although anticipation typically involves the analysis of a single prior art reference, I understand that additional references may be used to show that the prior art reference has enabling disclosure (i.e., allows a POSITA to make the invention without undue experimentation), to explain the meaning of a term used in the prior art reference, and/or to show that a characteristic is inherent in the prior art reference.

**B. Obviousness**

14. I understand that a claim is invalid as obvious if it would have been obvious to a person of ordinary skill in the art at the time the alleged invention was made. This means that even if all of the elements of the claim cannot be found in a single prior art reference that would anticipate the claim, a person of ordinary skill in the art who was aware of the prior art would have been able to come up with the claimed invention. This may be the case, for example, where the missing element represents only an insubstantial different over the prior art or a reconfiguration of a

known system. I understand that in an obviousness determination, the person of ordinary skill in the art is presumed to have knowledge of all material prior art.

15. I understand that an obviousness analysis requires an understanding of the scope and content of the prior art, any differences between the alleged invention and the prior art, and the level of ordinary skill in evaluating the pertinent art.

16. I understand that when a product is available, design incentives and other market forces can prompt variations of it, either in the same field or a different one. If a person of ordinary skill in the art can implement a predictable variation, obviousness likely bars its patentability. For the same reason, if a technique has been used to improve one device and a person of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique would have been obvious.

17. I understand that whether a prior art reference renders a patent claim unpatentable as obvious is determine from the perspective of a person of ordinary skill in the art at the time of the alleged invention. I have been told that there is no requirement that the prior art contain an express suggestion to combine known elements to achieve the claimed invention, but a suggestion to combine known elements to achieve the claimed invention may come from the prior art, as filtered through the knowledge of one skilled in the art. In addition, I have been told that

the inferences and creative steps a person of ordinary skill in the art would employ are relevant to the determination of obviousness.

18. I understand that one may consider, e.g., whether (1) the change was merely the predictable result of using prior art elements according to their known functions, or whether it was the result of true inventiveness; (2) there is some teaching or suggestion in the prior art to make the modification or combination of elements claimed in the patent; (3) the claimed innovation applies a known technique that had been used to improve a similar device or method in a similar way; (4) the claimed invention would have been obvious to try, meaning that the claimed innovation was one of a relatively small number of possible approaches to the problem with a reasonable expectation of success by those skilled in the art; (5) the invention merely substituted one known element for another known element in order to obtain predictable results; (6) the invention merely applies a known technique to a known device, method, or product to yield predictable results; or (7) known work in one field of endeavor may have prompted variations of it for use in either the same field or a different one based on design incentives or other market forces that would have been predictable to a person of ordinary skill in the art.

19. I further understand that certain factors may support or rebut the obviousness of a claim. I understand that such secondary considerations include, among other things, commercial success of the patented invention, skepticism of

those having ordinary skill in the art at the time of the invention, unexpected results of the invention, any long-felt but unsolved need in the art that was satisfied by the alleged invention, the failure of others to make the alleged invention, praise of the alleged invention by those having ordinary skill in the art, and copying of the alleged invention by others in the field. I understand that there must be a nexus—that is, a connection—between any such secondary considerations and the alleged invention. I also understand that contemporaneous and independent invention by others is a secondary consideration tending to show obviousness.

20. I am not aware of any allegations by the named inventor of the ’080 Patent or any assignee of the ’080 Patent that any secondary considerations tend to rebut the obviousness of the ’080 Patent.

21. Additionally, I understand that in considering obviousness, it is important not to use the benefit of hindsight derived from the patent under consideration.

**C. Level of Ordinary Skill in the Art**

22. When interpreting a patent, I understand that it is important to identify the relevant art pertaining to that patent, as well as the level of ordinary skill in that art at the time of the claimed invention. The “art” is the field of technology to which the patent is related.

23. I am informed and understand that the POSITA is a hypothetical person who is presumed to know the relevant prior art. I understand that the actual inventor's skill is not determinative of the level of ordinary skill. I further understand that factors that may be considered in determining level of skill include: (i) the types of problems encountered in the art; (ii) prior art solutions to those problems; (iii) the rapidity with which innovations are made; (iv) the sophistication of the technology; and (v) the educational level of active workers in the field. I understand that not all such factors may be present in every case, and one or more of them may predominate.

24. I understand that a POSITA is one who is presumed to be aware of all pertinent art, thinks along conventional wisdom in the art, and is a person of ordinary creativity. A POSITA would have had knowledge of circuit boards and microcircuits, including components thereof, and related technologies as of July 23, 2010.

25. A POSITA would have had a bachelor's degree in electrical engineering, computer science, computer engineering, material science, physics, applied physics, or a related field, and at least two years of experience in the research, design, development, or testing of electronic circuits or components or software for controlling electronic circuits or components, or the equivalent, with additional education substituting for experience and vice versa.

26. Based on my education and experience, I would have, at least, met the criteria for a POSITA in December 22, 2011, and I still exceed it today.

**D. Claim Construction**

27. I understand that the United States Patent and Trademark Office interprets claim terms in an *inter partes* review proceeding under the same claim construction standard that is used in a United States federal court. I understand that under this standard, the meaning of claim terms is considered from the viewpoint of a POSITA at the time of the alleged invention.

28. I have been informed that claim terms are generally given their ordinary and customary meaning as understood by one of ordinary skill in the art in light of the specification and the prosecution history pertaining to the patent. I understand, however, that claims terms are generally not limited by the embodiments described in the specification.

29. I understand that in addition to the claims, specification, and prosecution history, other evidence may be considered to ascertain the meaning of claim terms, including textbooks, encyclopedias, articles, and dictionaries. I have been informed that this other evidence is often less significant and less reliable than the claims, specification, and prosecution history.

## III. SUMMARY OF GROUNDS

30. I understand that Petitioner challenges the validity of Claims 1-24 of the ’080 Patent. In particular, I understand that Petitioner asserts the following grounds for challenging the claims of the ’080 Patent:

| Ground | Summary |
|---|---|
| 1 | Claims 1-4, 7-12, 15-20, and 23-24 are rendered obvious by Sutardja alone (Ex-1007, incorporating Ex-1008) |
| 2 | Claims 5-6, 13-14, and 21-22 are rendered obvious by Sutardja in view of Rychlik (Ex-1009) |
| 3 | Claims 7, 15, and 23 are rendered obvious by Sutardja in view of Carmack (Ex-1006) |
| 4 | Claims 1-4, 7-12, 15-20, and 23-24 are rendered obvious by Mathieson (Ex-1005, incorporating Ex-1006) in view of Sutardja |
| 5 | Claims 5-6, 13-14, and 21-22 are rendered obvious by Mathieson/Sutardja in view of Rychlik |

## IV. THE ’080 PATENT

### A. Overview of the ’080 Patent

31. The ’080 Patent is directed to power management in a multicore processor. Ex-1001, 2:19-22, 3:34-62. As the ’080 Patent explains:

> Computing system power consumption is becoming more and more of a concern. As such, a number of different

> power management schemes are incorporated into modern day computing systems.

Ex-1001, 2:19-22. The '080 Patent describes several of these existing and well-known prior art power management schemes:

> Typically, the power management component of the system will scale up the processing performance of the system as the system's workload increases, and, scale down the processing performance of the system as the system's workload decreases. Decreasing the processing performance of the system corresponds to power savings as the power consumption of the system is strongly correlated with its performance capabilities.
>
> A typical way to scale processing performance and power consumption with workload is to ***enable/disable entire cores and raise/lower their supply voltages and operating frequencies in response to system workload***. For example, … under a maximum performance and power consumption state 201 all cores are enabled and each core is provided with a maximum supply voltage and maximum clock frequency. By contrast, under a minimum performance and power consumption state 202 (at which program code can still be executed), only one core is enabled. The single core is provided with a minimum supply voltage and minimum operating frequency.

Ex-1001, 2:23-42. These prior art power management schemes are illustrated in Figure 2 of the '080 Patent.[2] Ex-1001, 3:34-37.

[2] Throughout this Petition, all annotations and emphasis have been added, unless otherwise noted.

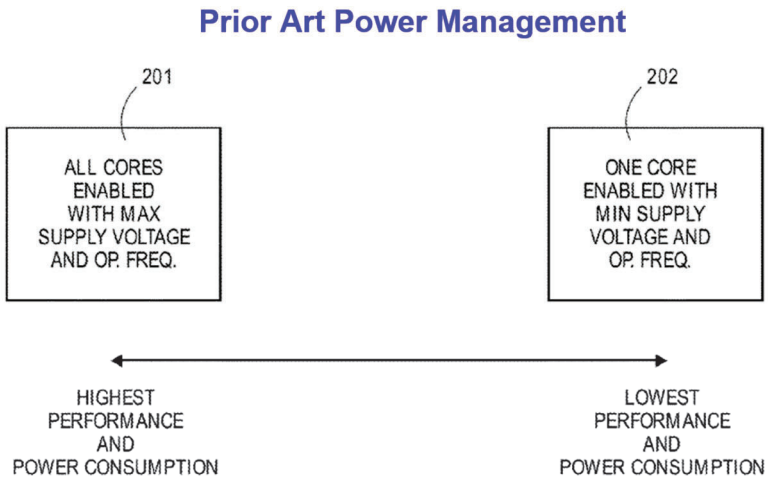



**FIG. 2**

Ex-1001, Figure 2 (annotated).

32. The '080 Patent purports to add to the existing body of prior art power management schemes by relying on power consumption concepts that the '080 Patent refers to as "basic":

> Some basic concepts of electronic circuit power consumption …[include the fact that] the speed of operation of interconnected logic gates [] rises as the width of its driving transistors [] (measured, for each transistor, along the semiconductor surface perpendicular to the direction of current flow) increase…. [as well as the fact that] a circuit that is designed to have relatively wide source/sink transistors, a high supply voltage, short load lines and receive a higher frequency clock signal will operate faster and therefore consume more power than circuits oppositely oriented as to these same parameters.

Ex-1001, 2:43-52, 3:26-33.

33. The ’080 Patent purports to use these “basic concepts” for “a new approach” in which one or more of the cores in a multi-core processor includes narrower transistors than the other cores, and is therefore “designed to be lower performance and therefore consume less power than [the] other cores,” while nevertheless supporting the same instruction set as the other cores. Ex-1001, 3:50-62. As a result, the ’080 Patent asserts that its multicore processor is improved, because:

> [T]he multi-processor is able to entertain a power management strategy that is the same/similar to already existing power management strategies, yet, still achieve an even lower power consumption in the lower/lowest performance/power states. Specifically, [for a] power state … in which only one cores is left operable (the remaining cores are disabled)[,] … if the one remaining operable core is the low power cores, the processor will exhibit even lower power consumption than the prior art low power state.

Ex-1001, 4:20-29.

34. But the concept of a multicore processor that includes both high-power and low-power cores, all of which support the same instruction set, was well-known in the prior art. For example, as I explain in Section V.D. below, Kumar et al. introduced the concept of a multi-core processor with both high-power and low-power cores that all execute the same instruction set—the “single-ISA heterogeneous multi-core architecture”—in 2003, almost a decade before the ’080

Patent's parent application was filed. Ex-1032, 0001; *see also* Section V.D. Indeed, as explained in the Grounds below, all claims of the '080 Patent are obvious in view of the prior art.

**B. Prosecution History of the '080 Patent**

35. I understand the '080 Patent was filed on February 13, 2017, as a continuation of application No. 13/335,257 (now U.S. Patent No. 9,569,278), filed on December 21, 2011.

36. During prosecution, the Examiner rejected pending claims 20, 23-28, 31-36, and 39-43 of the '080 Patent (equivalent to issued claims 1, 4-9, 12-17, and 20-24) as anticipated or obvious over U.S. Patent Pub. No. 2006/0095807 to Grochowski and pending claims 21-22, 29-30, and 37-38 (equivalent to issued claims 2-3, 10-11, and 18-19) as obvious over Grochowski in view of U.S. Patent Pub. No. 2008/0263324 (Sutardja '324) to which Sutardja '748 (Ex-1007) claims priority as a continuation. Ex-1004, 0067-0072. The Applicant was able to overcome these rejections by arguing that Grochowski failed to teach all of the limitations of the independent claims, as illustrated in the below excerpt from the prosecution history of the '080 Patent:

*§§102 and 103 Rejections*

Merely in the interest of compact prosecution, the Applicant notes that the cited portions of Grochowski do not appear to teach or suggest the Applicant's claims. For example, the cited portions of Grochowski do not appear to teach or suggest:

> 20. A multi-core processor comprising:
> a first plurality of cores and a second plurality of cores that support a same instruction set, wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and
> **power management hardware to, from a state where the first plurality of cores and the second plurality of cores are enabled, disable all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores, wherein an operating system to execute on the multi-core processor is to monitor a demand for the multi-core processor and control the power management hardware based on the demand**. (Emphasis added.)

Ex-1004, 0056.

37. Notably, the Applicant did not dispute that Sutardja '324 teaches the added transistor-related limitations of dependent Claims 2-3, 10-11, and 18-19. Ex-1004, 0054-0057. Nor did either the Applicant or the examiner include any discussion of Sutardja '324 with reference to pending Claims 20, 24-28, 32-36, and 40-43 of the '080 Patent (equivalent to issued Claims 1, 5-9, 13-17, and 21-24).

38. As reflected in the file history, the purported novelty of the claims of the '080 Patent lay in the claimed power management hardware. Ex-1004, 0056-0057. However, as explained in the Grounds below, this limitation (indeed all limitations of Claims 1-24) was well-known in the prior art.

39. I understand that Mathieson, Carmack, and Rychlik—all the references relied upon in Grounds 4 and 6, against all Challenged Claims—were not before the PTO during prosecution of the ’080 Patent. Thus, I understand that the Petition presents substantially new arguments that were not considered during prosecution.

40. I also understand that Sutardja ’748 (Ex-1007) is cited on the face of the ’080 Patent, and its parent, Sutardja ’324, was previously considered by the Office. However, Sutardja ’785 (Ex-1008), which both Sutardja ’748 and Sutardja ’324 incorporated by reference, was not before the Office. Furthermore, while the Examiner relied upon Sutardja ’324 in rejecting the dependent claims that issued as claims 2-3, 10-11, and 18-19 of the ’080 Patent, the prosecution history of the ’080 Patent indicates that Applicant did not dispute that Sutardja ’324 taught the limitations added by those claims. Ex-1004, 0054-0057.

## V. TECHNOLOGY BACKGROUND

41. To provide background for the element-by-element analysis of the claims to follow, below I will present an overview of the state of the art existing at the time of the alleged invention relating to processors generally, and as it relates to multi-core devices, and power, voltage and frequency management approaches, among other relevant technology. As I will describe below from a high-level functional perspective, in the next section based on specific prior art, and then in subsequent sections on a limitation-by-limitation basis, all of these technologies and

techniques were well-known to those of ordinary skill in the art at the time of the alleged invention claimed in the ’080 Patent, and a POSITA would have readily understood the combination of elements of Claim 1-24 to have been obvious

**A. Relationship between power, voltage, and frequency**

42. There are two key relationships linking core frequency and voltage to core power consumption. They are: (1) power is proportional to $V^2f$ – where V is the voltage to operate the core, and f is the frequency of operation of the core; and (2) the frequency of operation is directly proportional to the voltage (i.e., f is proportional to V). Thus, higher frequency requires a greater core voltage resulting in the consumption of much more power.

43. Frequency provides a crude measure of performance; for a single processing core (an integrated circuit), improved performance can be achieved by increasing the frequency at which the core operates. Thus, one can readily see that the goals of high performance and low power consumption are in tension with one another. In general, high performance comes at the expense of high levels of power consumption.

44. A well-known and widely used power management technique is thus to operate a core at the lowest voltage and frequency levels sufficient to provide adequate performance for the workload currently executing on the core. This is known as dynamic voltage and frequency scaling (DVFS).

**B. Multi-core processing approaches**

45. "Limiting power consumption presents a critical issue in computing, particularly in battery powered devices such as laptop computers and cell phones. Limiting power in other computer settings—such as server farms—warehouse-sized buildings filled with Internet service providers' servers—is also important."

46. I wrote this in 2000/1 in two review articles, at which time the problem of power consumption was well known and starting to become a design constraint for computer designers. In particular, while manufacturers of processor cores, before about the turn of the century, had been able to increase performance by simply increasing frequency, the fact that power increases as square of the voltage finally caught up with them and meant that increasing frequency quickly became impractical.

47. The solution that began to emerge was to employ multiple processors to do the work of one. The reasoning was that two processors could run at half the frequency, but now the quadratic dependency on supply voltage gave a positive effect, because the voltage supply could be lowered at a quadratic rate. Thus the power consumption of two processors could actually be less than that of one fast processor.

48. Although I have presented a highly simplified version of the underlying physics, the basic argument of using multi-cores to save power without sacrificing

performance holds, and this led the industry to adopt multi-cores. Notably, the single chip multi-core concept had been anticipated by researchers at Stanford University years before the industry adopted this strategy with their Hydra Multiprocessor-on-a-Chip, in February 1998 [report CSL-TR-98-749, Stanford 1998]. In about 2005, Intel and its rivals began to perform a "right turn," to quote their promotional material, which meant they moved away from using a single fast processor and replaced it with multiple slower cores on an integrated circuit to meet power limitations, while at the same time maintaining performance. "Each processor may include on or more cores and each processor/core may perform operations independent of the other processors/cores." Ex-1027, 8:46-48. The background of the '080 Patent reflects this concept, which discusses a scenario with multiple cores, and was widely employed at the date of the patent. The use multiple cores opened up further ways to control power consumption: when not needed some cores could be turned off.

49. Notably, while dynamic power consumption (i.e., power consumption during instruction execution), discussed above, dominates during periods of high load, power is also consumed whenever cores are supplied with electric current, even when those cores are not actively executing instructions. This type of power consumption is referred to as "leakage" power consumption, and can dominate the total power consumed by a multi-core processor during periods of low load. Thus,

a simple and commonly-implemented technique for power conservation in a multi-core processor is to disable one or more cores of the multi-core processor when those cores are not needed. *See, e.g.*, Ex-1031, 0002, 0007 (explaining that "[m]ost commercial multicore processors are designed with the ability to dynamically adjust the number of active cores," and that disabling unnecessary cores may "save a considerable amount of leakage power in low load periods, where the leakage power makes up a large fraction of the total power consumption"). Indeed, as illustrated in the below figure, reducing the number of active cores during periods of low load can offer noticeable power savings.

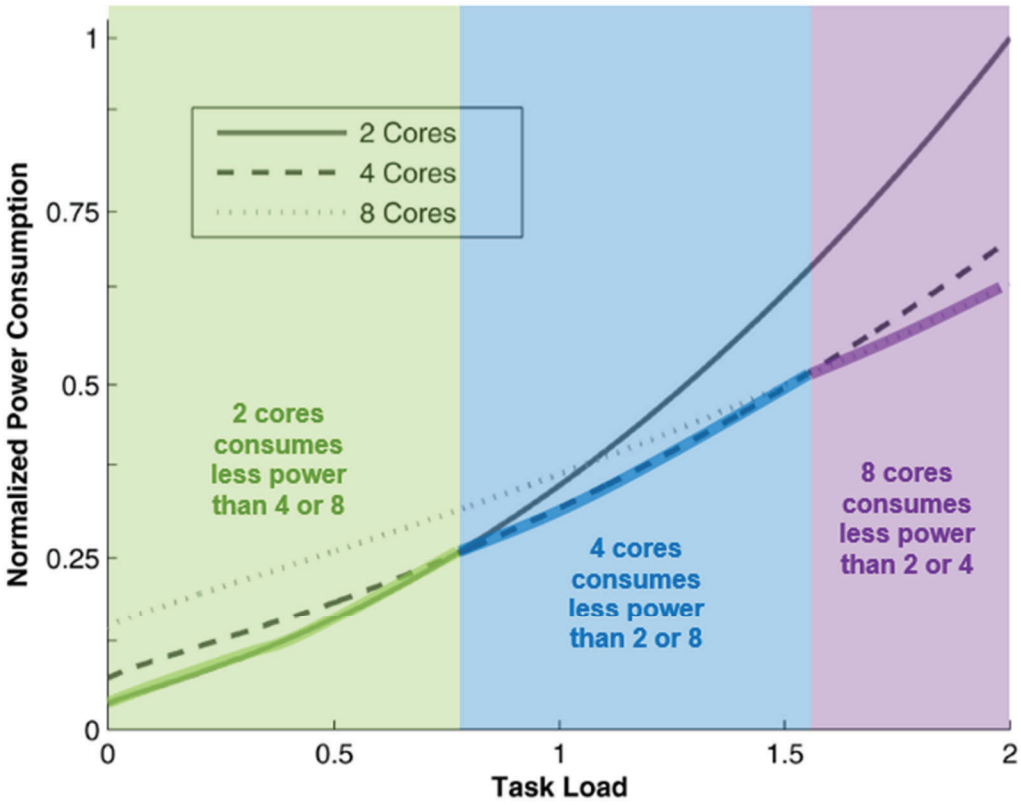



Fig. 4. Expected power consumption as a function of task load for various multicore processors.

Ex-1031, Figure 4.

50. In addition to reducing the number of active cores, DVFS (as described above) may be used on each of the cores in a multi-core processor to further control power and performance. In particular, during periods of low load, when high performance is not needed, power can be reduced not only by turning off a subset of the cores, but also by lowering the supply voltage and operating frequency of the remaining, active cores.

### C. Processor Microarchitectures

51. A processing core is a complex digital circuit that is made up of many transistors. Each transistor operates based on the principles of binary logic, where it can be in one of two states: on or off. The transistors in a core are used to implement logic gates, which are the basic building blocks of digital circuits. Each logic gate consists of one or more transistors that are connected in a specific way to achieve the desired logical operation (e.g., AND, OR, NAND, NOR, etc.). By combining multiple logic gates, a processor can perform complex operations, such as arithmetic and logic operations.

52. The power consumed by a processing core depends on the power consumed by its constituent transistors. There are two main sources of power consumption by transistors: static power consumption and dynamic power consumption. Dynamic power consumption refers to the power consumed by a transistor when it is actively switching signals. This power consumption is proportional to the frequency at which the transistor is switching and the load being driven by the transistor. Static power consumption refers to the power consumed by the transistor even when it is not actively switching signals. This power consumption is due to leakage current that flows even when the transistor is in its "off" state.

53. Because the power consumed by a processor depends on the power consumed by its constituent transistors, cores with different power consumption properties can be formed by using different transistors for the cores. For example, the power consumption of a core can be reduced by decreasing the width of the transistors used to form the core, which results, for example, in decreased leakage power consumption. *See, e.g.*, Ex-1029, 0007 (explaining that "[l]eakage grows with the aspect ratio [(width/length)] of a transistor"). This, however, will come at the expense of decreasing the core's performance. *See, e.g.*, Ex-1030, 0007 (explaining that "reducing the width [of a transistor] increases the transistor's delay").

**D. Heterogeneous multi-core processors**

54. While early multi-core processors typically included identical cores, in 2003, Kumar et al. demonstrated that significant power savings could be achieved by designing a multi-core processor to include cores that differ in their power consumption properties, but nevertheless execute the same instruction set—what Kumar et al. termed a "single-ISA heterogeneous multi-core architecture." Ex-1032, 0001. As Kumar et al. explained:

> The architecture consists of a chip-level multiprocessor with multiple, diverse processor cores. ***These cores all execute the same instruction set, but*** include significantly different resources and ***achieve different performance and energy efficiency on the same application***. During

> an application's execution, the operating system software tries to match the application to the different cores, attempting to meet a defined objective function. For example, it may be trying to meet a particular performance requirement or goal, but doing so with maximum energy efficiency.

Ex-1032, 0002.

55. As mentioned above, the cores in Kumar's heterogeneous multi-core system implemented the ***same*** instruction set architecture (ISA). Implementing an ISA—assembling the basic electronic logic gates—can be done in many ways. For example, a more sophisticated logical implementation may have better performance and consume more power, while a simpler implementation is more power efficient at the expense of performance. But, because the ISAs on both types of Kumar's cores are the same, they can run the same software unchanged and switching between cores during the execution of a program can be seamless.

56. Significant power savings can be achieved through the use of only two different types of cores in a multicore processor—a set of low-power, low-performance cores, together with a set of high-power, high-performance cores. In particular, "[i]n times of light load, activity [may] shift[] to low-power cores," to conserve power, while at the same time, "in times of heavy load, low-power cores can offer additional parallelism without significant increases in power

consumption." Ex-1024, 0003-0004 (describing Kumar et al.'s single-ISA heterogeneous multi-core system).

57. Since Kumar et al. first proposed the use of a single-ISA heterogeneous multi-core processor, the concept has been widely adopted for power conservation purposes. *See, e.g.*, Ex-1005, ¶¶28-29; Ex-1007, ¶¶230, 259. Indeed, the idea was subsequently adopted commercially and branded as Big/Little to indicate the different (heterogeneous) core sizes.

58. And the use of well-known power management techniques within such single-ISA heterogeneous multi-core processing systems (including, for example, core scaling and DVFS, as I described above) opens up a much wider trade-off between high-performance, at the cost of high power, and very low power operating regimes when performance is not the goal. For example, one can not only switch from high-performance high-power cores to lower-performance low-power cores, but further reduce the power/performance by lowering the supply voltage and operating frequency of the low-power cores.

## VI. THE PRIOR ART IN THE APPLIED INVALIDITY GROUNDS

### A. Mathieson (Ex-1005)

59. Mathieson, titled "System and Method for Power Optimization," was filed on May 25, 2010 and published September 1, 2011. Ex-1005, Cover.

60. I understand Mathieson is prior art under at least 35 U.S.C. §§ 102(a) and (e).

61. Like the '080 Patent, Mathieson discloses a multicore processor that includes a first set of cores and a second set of cores, which differ in their levels of power consumption because of differences in their constituent transistors:

> According to various embodiments, the cores 212 included in the first set of cores 210 comprise "fast" cores and the cores 222 included in the second set of cores 220 comprise "slow" cores. For example, the cores 212 may be manufactured using faster transistors that have significant static leakage. In some embodiments, when the computing needs and/or workload of the first set of cores 210 are lowered, then the clock speed is lowered to reduce power. The static leakage is not a significant issue at the high clock speeds required for peak performance. However, at slower clock speeds, the static leakage of the fast transistors can dominate the overall power consumptions.
>
> …
>
> The second set of cores 220 includes one or more "slow" cores 222 constructed from slower transistors that are not capable of operating as quickly as the transistors included in the cores 212 of the first set of cores 210. In some embodiments, the second set of cores 220 has a much lower leakage power loss than the first set of cores 210, but is not capable of achieving the same performance levels as the first set of cores 210.

Ex-1005, ¶¶28-29.

62. As a result, for a range of operating frequencies, Mathieson's first set of cores consume more power than its second set of cores, for the same operating frequency. This is illustrated below, in Figure 6 of Mathieson:

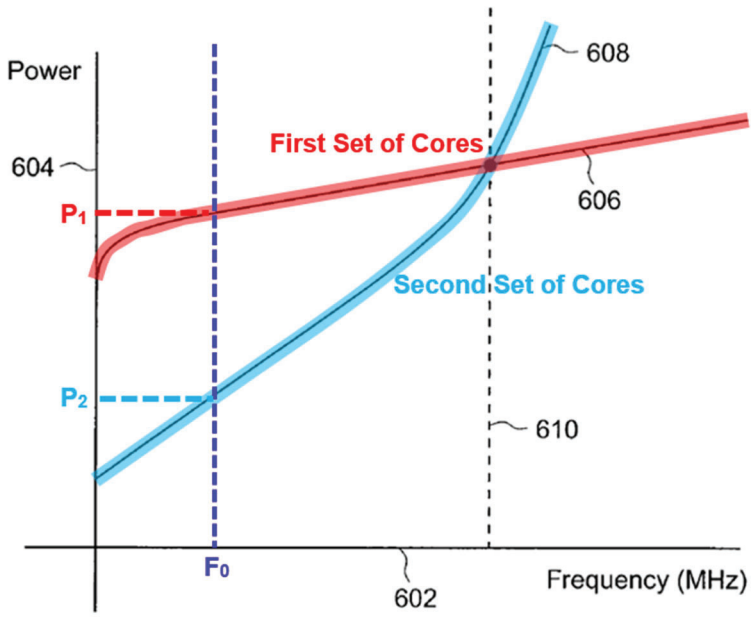



FIG. 6

Ex-1005, Figure 6 (annotated).

63. Because Mathieson's cores consume different levels of power for the same operating parameters, Mathieson teaches that Mathieson's multi-core CPU can conserve power by switching between using the first set of cores and/or the second set of cores, based on the size of the workload. Ex-1005, ¶¶31-32. In particular, from a first mode of operation in which the first set of cores (and optionally the second set of cores) are enabled, Mathieson teaches switching to a second mode of

operation in which only the second set of cores are enabled, for a decrease in workload below a "threshold value." Ex-1005, ¶¶30-32, 57, Fig. 4A.

**B. Carmack (Ex-1006)**

64. Carmack, titled "Multi-core Integrated Circuits Having Asymmetric Performance Between Cores," was filed on June 11, 2008 and published December 17, 2009.

65. I understand that Carmack is prior art under at least 35 U.S.C. §§ 102(a) and (b).

66. I also understand that Carmack belongs to the same patent family as Mathieson, all of Carmack's inventors are listed as inventors on Mathieson, and both Carmack and Mathieson disclose the same subject matter of multicore processors that include cores with different power consumption properties, resulting from transistor differences between the cores. Ex-1005, Cover, ¶¶17-18; Ex-1006, Cover, ¶¶27-29. Indeed, I understand that Mathieson incorporates by reference Carmack's entire disclosure. Ex-1005, ¶1.

67. Like Mathieson, Carmack teaches dynamically selecting between running a set of high-power cores, a set of low-power cores, or both sets of cores, depending on the workload to be processed. This is illustrated in Figure 3 of Carmack, reproduced below.

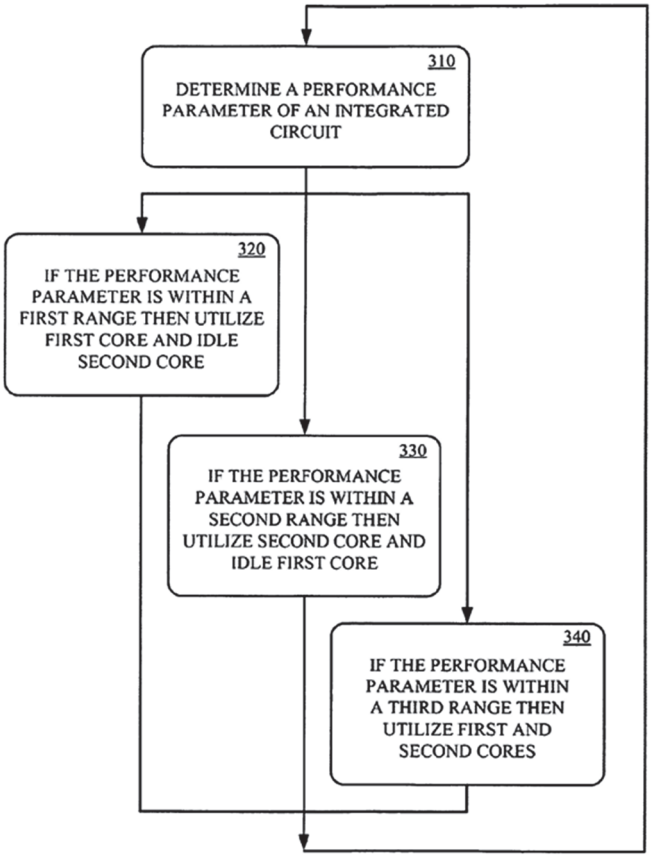



Figure 3

Ex-1006, Figure 3.

68. Carmack also discloses that, when both types of cores are used, the high-power cores may be operating at their maximum frequency:

> When both cores 110, 120 are utilized, the second core 120 would be given a fraction of the total work scaled to its performance advantage over the first core 110. For situations where the overhead of coordinating asymmetric cores becomes too high, the system can lower ***the peak frequency of the faster core 120*** to match the maximum frequency of the slower core 110.

Ex-1006, ¶22.

### C. Sutardja – Sutardja '748 (Ex-1007) and Sutardja '785 (Ex-1008)

69. Sutardja '748 (Ex-1007), titled "Dynamic Core Switching," was filed June 30, 2008 and published November 20, 2008. Ex-1007, Cover. Sutardja '748 was filed as a continuation of application No. 12/145,660, which itself is a continuation-in-part of Sutardja '785 (Ex-1008). *Id*.

70. I understand that Sutardja '748 and Sutardja '785 both qualify as prior art under at least 35 U.S.C. §§ 102(a) and (b).

71. Sutardja '748 incorporates by reference the disclosure of Sutardja '785 in its entirety. Ex-1007, ¶1. Accordingly, this Declaration refers to Sutardja '748, incorporating Sutardja '785, as "Sutardja."

72. Both Sutardja '748 and Sutardja '785 (collectively "Sutardja") disclose the same subject matter of a multicore processor that includes both high-power and low-power cores. Ex-1007, ¶¶212-215; Ex-1008, ¶¶208, 215. And, like Mathieson and the '080 Patent, Sutardja teaches that the differences in power consumption between the cores is due to differences in the transistors used in the cores. Ex-1007, ¶230; Ex-1008, ¶¶217, 225. In particular, Sutardja explains that:

> The transistors 106 of the HP core 24 tend to consume more power during operation in the active state than the transistors 110 of the LP core 28. In some implementations, the transistors 106 may have higher leakage current than the transistors 110. The transistors 106 may have a size that is greater than a size of the transistors 110. Gates of the transistors 106 may be larger

> than gates of the transistors 110. A gate count of the HP core 24 may be greater than the gate count of the LP core 28.

Ex-1007, ¶230.

73. Like Mathieson and the '080 Patent, Sutardja also teaches operating the multicore processor in a high-power mode (in which the high-power cores and, potentially, the low-power cores, are enabled) and a low-power mode (in which only the low-power cores are enabled), based on system loading:

> In the HP mode, the HP core 24 is in the active state and processes threads. The LP core 28 may also operate during the HP mode. In other words, the LP core 28 may be in the active state during all or part of the HP mode.

Ex-1007, ¶223.

> After executing instructions using the HP core 204, system loading may eventually decrease. When the system loading decreases to a predetermined point, the core switching module 212 may transition operation from the HP core 204 back to the LP core 200. As can be appreciated, systems employing core switching may benefit from the low power consumption of the LP core 200 when relatively low processing speed is required. When high speed processing is desired, the HP cores 204 can be used and then turned off when the processing load decreases sufficiently.

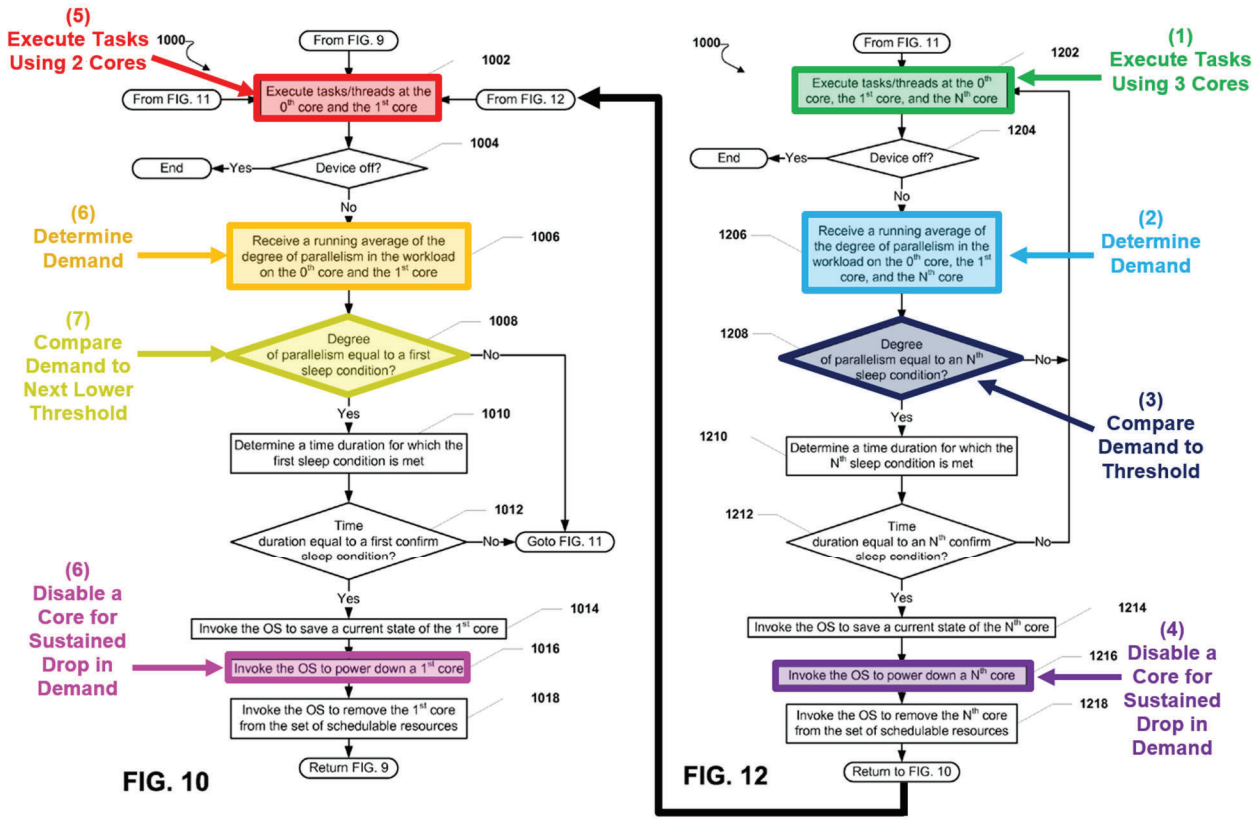
Ex-1007, ¶259.

**D. Rychlik (Ex-1009)**

74. Rychlik, titled "System and Method for Controlling Central Processing Unit Power Based on Inferred Workload Parallelism," was filed on November 11, 2010 and published June 16, 2011. Ex-1009, Cover.

75. I understand that Rychlik is prior art under at least 35 U.S.C. §§ 102(a) and (e).

76. Rychlik teaches a power management technique for a multi-core processor that involves disabling cores one-by-one, in response to falling levels of demand. Ex-1009, ¶¶68-71, 88. Thus, "[u]nlike traditional systems, which power all CPU cores symmetrically, i.e., all on or all off, [Rychlik's] system and method may [be used] to asymmetrically power CPU cores on and off." Ex-1009, ¶96. This is illustrated in Figure 12 of Rychlik, below.




Ex-1009, Figure 10 (annotated).

Rychlik also teaches that this power management technique may be combined with other power management techniques. In particular, Rychlik explains that a "dynamic clock and voltage scaling (DCVS) algorithm" may be used on each core to adjust the frequency and/or voltage of each core in response to changes in demand. Ex-1009, ¶¶34, 100.

## VII. CLAIM CONSTRUCTION

77. I do not believe that any claim term requires explicit construction to resolve the issues presented in this Petition. I apply the plain and ordinary meaning

to each claim term, as that plain and ordinary meaning would have been understood by a POSITA.

78. I reserve the right to offer opinions on any claim constructions proposed in this proceeding or to offer opinions on additional constructions in the district court.

## VIII. DETAILED EXPLANATION OF THE UNPATENTABILITY GROUNDS

### A. Ground 1: Claims 1-4, 7-12, 15-20, 23-24 are rendered obvious by Sutardja (Ex-1007, incorporating Ex-1008)

#### 1. Independent Claim 1

##### a. Element 1[pre]: A multi-core processor comprising:

79. Sutardja teaches the preamble of Claim 1 to the extent it is limiting.

80. Sutardja teaches a multi-core processor in the form of "a multi-core processing system having a LP core and a HP core," as illustrated below in annotated Figure 3B. Ex-1007, ¶221.

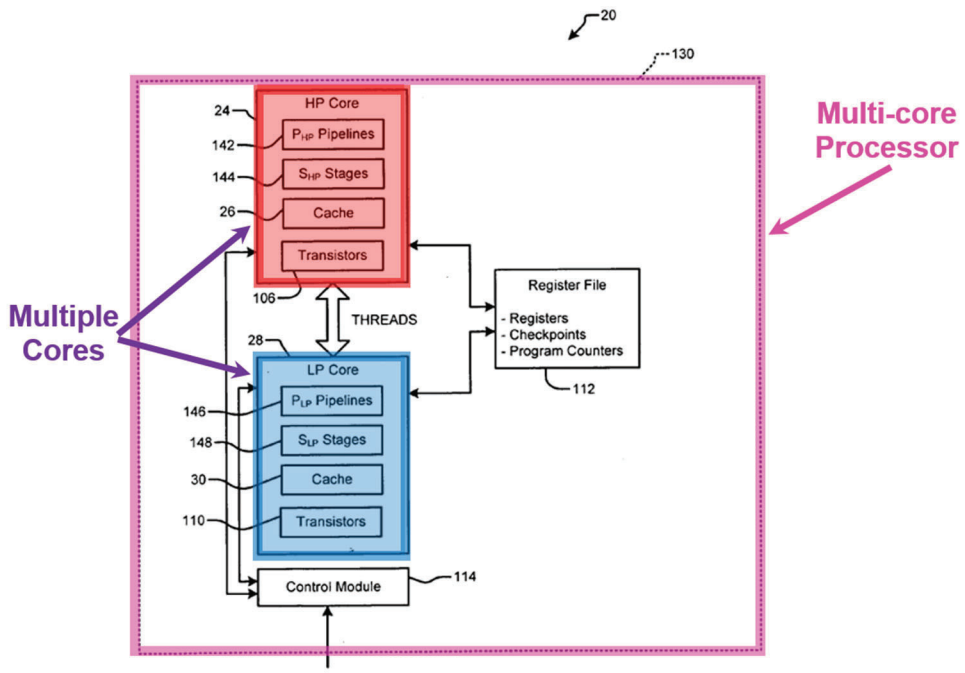



FIG. 3B

Ex-1007, Figure 3B (annotated).

**b. Element 1[a][i]: a first plurality of cores and a second plurality of cores that support a same instruction set,**

81. Sutardja teaches Element 1[a][i].

82. Sutardja discloses two different types of cores: "a high-speed, high-power (HP) core" and "a low-speed, low-power (LP) core." Ex-1007, ¶212. Sutardja explains that both types of cores may support the same instruction set. Ex-1007, ¶213 ("An instruction set architecture (ISA) of the LP core 12 may differ from the ISA of the HP core 16…. Alternatively, ***both may use the same ISAs***."); Ex-1007, ¶225 ("The LP core 28 may use an ISA that is ***the same as*** or a subset of the ISA of the HP core 24.").

83. Sutardja also teaches that the multi-core processor may include a plurality of HP cores and a plurality of LP cores. In particular, Sutardja explains that "the HP core 204 may comprise ***multiple identical cores***." Ex-1007, ¶249. Thus, "[w]hen applications demand still higher performance than that provided by one HP core, ***multiple HP cores may be used.***" Ex-1007, ¶219.

84. In addition to including multiple HP cores in Sutardja's multi-core processor, a POSITA would have been motivated to include multiple LP cores. Indeed, it was well-known to include multiple low-power cores together with multiple high-power cores in a multi-core processor. For example, U.S. Patent Publication No. 2012/0317568 explains that "heterogeneous processing system[s] … with cores that have identical functional equivalence but differing performance/power operating characteristics" were well-known and "[t]ypically … have one or more big cores and ***one or more little cores***." Ex-1023, ¶13. Similarly, *Operating System Support for Mitigating Software Scalability Bottlenecks on Asymmetric Multicore Processors* by Juan Saez et al. explains that:

> An asymmetric multicore processor (AMP) includes cores exposing the same instruction-set architecture, but differing in features, size, speed and power consumption []. A typical AMP would contain ***a number of simple, small and low-power*** **slow** ***cores*** and a few complex, large and high-power *fast* cores.

Ex-1025, at 0001. And, in *Operating Systems and Asymmetric Single-ISA CMPs: The Potential for Saving Energy*, Jeffrey C. Mogul et al. explains that the original model for a "single-Instruction-Set-Architecture" multi-core processor included multiple low-power cores, which serve two power-saving purposes: "In times of light load, activity shifts to low-power cores; in times of heavy load, low-power cores can offer additional parallelism without significant increases in area or power consumption." Ex-1024, 0002-0003.

85. Sutardja also provides explicit motivation for including multiple LP cores. As Sutardja explains, various types of applications may be executed on Sutardja's multi-core processor, including a set of "first applications" "that require relatively low processing speed during operation":

> The first applications 42 may include applications that require relatively low processing speed during operation and may be executed during LP and HP modes. In other words, the first applications 42 may comprise applications that are generally not very processor intensive.

Ex-1007, ¶226.

86. Sutardja also contemplates that sets of "second applications" and "third applications" may execute on Sutardja's multi-core processor, each with their own processing requirements:

> The second applications 44 may comprise applications that require medium processing speed (e.g., typically higher than the first applications 42 and lower than the

> third applications 46 on average). Depending upon the circumstances, the second applications 44 may require low, medium, or high processing speed. The third applications 46 may be applications that tend to require high processing speed and may be best operated in the HP mode. Both HP and LP cores 24 and 28 may be capable of running all three types of applications. Alternatively, some applications may be limited to running using only the HP core 24.

Ex-1007, ¶227.

87. While Sutardja's "first applications" may be executed on Sutardja's LP core, because the processing capacity of the LP core is limited, "if a sufficient number of the first applications 42 are being executed concurrently, the processing system 20 may need to transition to and operate in the HP mode." Ex-1007, ¶226. Such a situation may arise where a large number of first applications are concurrently executing on the LP core, such that the processing capacity of the LP core is not sufficient to meet the performance requirements of all of those concurrently executing applications. In such a scenario, it may be desirable to transfer execution of one or more of the first applications from the LP core to the HP core. Doing so may enable the LP core to meet the performance requirements of the remaining first applications that are continuing to execute on the LP core, with the HP core meeting the performance requirements of the transferred applications.

88. The above-described strategy makes sense for scenarios in which the typical number of concurrently executing first applications is such that the

processing capacity of the LP core alone is sufficient to meet the performance requirements of those executing applications, and therefore the need to transfer execution of first application(s) from the LP core to the HP cores is rare. However, for scenarios in which the transfer of execution of one or more of the first applications from the LP core to the HP cores is routine, use of this strategy is undesirable.

89. First, if a large number of second and/or third applications are also executing on Sutardja's multi-core processor, there may not be sufficient processing capacity available on Sutardja's HP core(s) to accommodate the first applications to be transferred, without negatively impacting the performance of the executing second and/or third applications.

90. Second, routinely operating in Sutardja's HP mode, while nevertheless executing applications that do not require the high speeds offered by the HP cores and that could be executed on an LP core, would frustrate Sutardja's stated purpose of "reduc[ing] power consumption," because spreading the execution of the first applications across an LP core and an HP core would consume more power than using a pair of LP cores. Ex-1007, ¶222; *see also* Ex-1026, 1:65-67 ("Constantly using a processor that uses a lot of power and that exceeds the needs of the application software can lead to significant power waste and costs."). This would be the case, because each of Sutardja's HP cores consumes more power, for the same

operating frequency and voltage, as the LP core. As explained for Element [a][ii], below, the '080 Patent re-states the well-known fact that cores with wider transistors will consume more power: "[the] lower power core will consume less power than [the] higher power core," "for a same applied clock or operating frequency, because of its narrower drive transistor widths." Ex-1001, 4:13-16. A POSITA would recognize that Sutardja's LP core will consume less power than its HP cores for the same reason. In particular, Sutardja explains that:

> The HP core 24 may be fabricated using a high leakage/high speed process. The LP core 28 may be fabricated using a low leakage/low speed process. The HP and LP cores 24 and 28 include transistors 106 and 110, respectively. ***The transistors 106 of the HP core 24 tend to consume more power*** during operation in the active state ***than the transistors 110 of the LP core 28***. In some implementations, the transistors 106 may have higher leakage current than the transistors 110. The transistors 106 may have a size that is greater than a size of the transistors 110. ***Gates of the transistors 106 may be larger than gates of the transistors 110***.

Ex-1007, ¶230. Thus, a POSITA would have understood that executing a set of first applications using a pair of LP cores would consume less power than executing that same set of first applications using an LP core and an HP core.

91. A POSITA seeking to implement Sutardja's teachings in a system in which a large number of applications "that require relatively low processing speed during operation" are expected to operate would therefore have been motivated to

include multiple LP cores capable of handling the execution of such applications without needing to transfer the execution of any of the applications to the HP cores. This may be the case, for example, for a server, to which Sutardja teaches its disclosure may be applied. Ex-1007, ¶220. Indeed, it would have been well-known to a POSITA that server systems typically include multiple "simple" cores together with multiple "high-performance" cores, because "[s]erver workloads can comprise multiple applications with varying computational and performance requirements." Ex-1027, 0005. For instance, it was well-known to developers of server systems that if the tasks to execute on the system "have internal parallelism," power reductions can be achieved by "replac[ing] high-performance processors with a larger number of slower but more energy-efficient processors providing the same throughput." Ex-1027, 0006.

92. A POSITA would have recognized that implementing multiple LP cores would achieve Sutardja's stated purpose of processing applications in a power-efficient manner, by using a core of the minimum power consumption to execute a given task, whenever such a core is able to meet the processing requirements of that task. *See, e.g.*, Ex-1026, 2:17-19 (explaining that, for power conservation purposes, when choosing from among cores offering different levels of performance, "the target processor core … sought for [a] job is the one that consumes a minimum power and still yields acceptable performance."). Moreover, a POSITA would have

had a reasonable expectation of success in doing so, since Sutardja already envisions using multiple HP cores, and there is nothing in Sutardja's disclosure that would preclude using multiple LP cores for the same reasons and in the same manner as multiple HP cores.

93. A POSITA would also have been motivated to include multiple LP cores in Sutardja's multi-core processor, based on Sutardja '785, which is incorporated by reference into Sutardja '748 (collectively, "Sutardja"). Sutardja '785 discloses a multi-processor system that includes multiple primary processors, in the form of a high-power CPU and a high-power GPU (the claimed first plurality of cores), as well as multiple secondary processors, in the form of a low-power CPU and a low-power GPU (the claimed second plurality of cores). Ex-1008, ¶¶215, 223. These multiple primary and secondary processors are illustrated below, in annotated Figure 2A. Ex-1008, ¶¶66, 127-128.

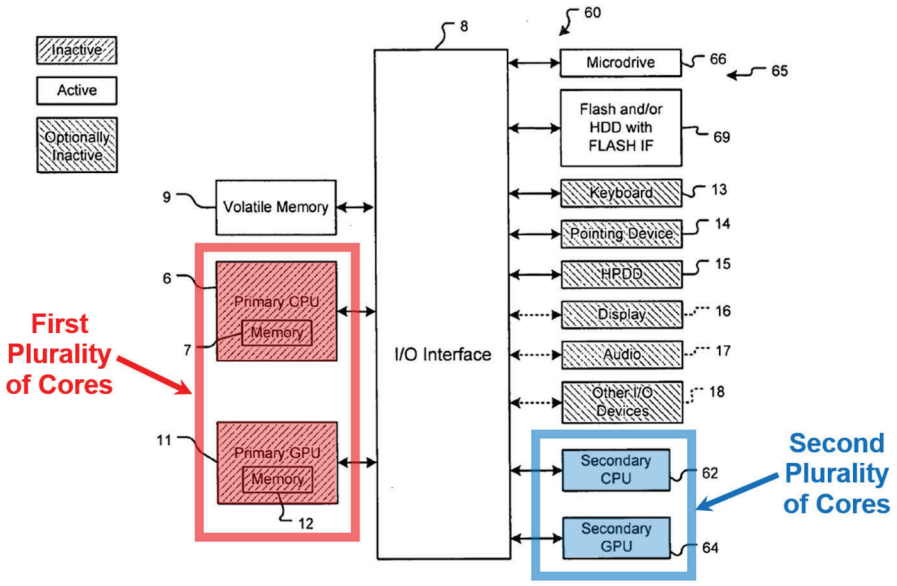



**FIG. 2A**

Ex-1008, Figure 2A (annotated).

94. Sutardja '785 explains that a benefit of using multiple processors, in the form of a CPU and a GPU is that it "increase[s] the speed of graphics processing and performance." Ex-1008, ¶5. A POSITA seeking to implement Sutardja would have had both the teachings of Ex-1007 and incorporated Ex-1008, and would have understood that, by including GPUs and CPUs within each set of HP and LP cores, the benefits of increased speed of graphics processing and performance and power-efficient processing, taught by each reference, could be achieved. Ex-1008, ¶5; Ex-1007, ¶216. As I have explained above, significant power savings can be achieved through the use of different types of cores in a multicore processor. In particular, "[i]n times of light load, activity [may] shift[] to low-power cores," to conserve

power, while at the same time, "in times of heavy load, low-power cores can offer additional parallelism without significant increases in power consumption." Ex-1024, 0003-0004 (describing Kumar et al.'s single-ISA heterogeneous multi-core system); *see also supra* ¶56.

**c. Element 1[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and**

95. Sutardja teaches Element 1[a][ii].

96. Sutardja's LP cores (the claimed second plurality of cores) consume less power, for a same applied operating frequency and supply voltage, than Sutardja's HP cores (the claimed first plurality of cores). As I explained in Section V.C above, the power consumed by a processor depends on its constituent transistors. *See supra* ¶53. Accordingly, cores with different transistors may exhibit different power consumption properties. For instance, all else being equal, the power consumption of a core with narrow transistors will be less than a core with wider transistors. *Id.*; *see also* Ex-1029, 0007 (explaining that the leakage power consumed by a transistor "grows with the aspect ratio [(width/length)] of a transistor"). Notably, the '080 Patent re-states this well-known fact that cores with wider transistors will consume more power, for the same applied operating frequency and voltage: "[the] lower power core will consume less power than [the]

higher power core," "for a same applied clock or operating frequency, because of its narrower drive transistor widths." Ex-1001, 4:13-16. A POSITA would recognize that Sutardja's LP core will consume less power than its HP cores for the same reason. In particular, Sutardja explains that:

> The HP core 24 may be fabricated using a high leakage/high speed process. The LP core 28 may be fabricated using a low leakage/low speed process. The HP and LP cores 24 and 28 include transistors 106 and 110, respectively. ***The transistors 106 of the HP core 24 tend to consume more power*** during operation in the active state ***than the transistors 110 of the LP core 28***. In some implementations, the transistors 106 may have higher leakage current than the transistors 110. The transistors 106 may have a size that is greater than a size of the transistors 110. ***Gates of the transistors 106 may be larger than gates of the transistors 110***.

Ex-1007, ¶230. Indeed, Sutardja contemplates that its HP and LP cores may consume different levels of power, by stating that "the LP and HP cores … have different properties and capabilities (e.g., different processing speeds, ***power consumption***, etc.)." Ex-1007, ¶213. Accordingly, Sutardja's LP cores, with their narrower transistors, consume less power.

**d. Element 1[b][i]: power management hardware to, from a state where the first plurality of cores and the second plurality of cores are enabled, disable all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores,**

97. Sutardja teaches Element 1[b][i].

98. First, Sutardja teaches a state where the first plurality of cores (the HP cores) and the second plurality of cores (the LP cores) are both enabled. As illustrated in Figure 4 below, at step 306 (purple) of method 300, Sutardja's processing system may be in the HP mode, which corresponds to a state in which the HP cores (the claimed first plurality of cores) are enabled. Ex-1007, ¶223 ("In the HP mode, the HP core 24 is in the active state and processes threads."); Ex-1007, ¶245.

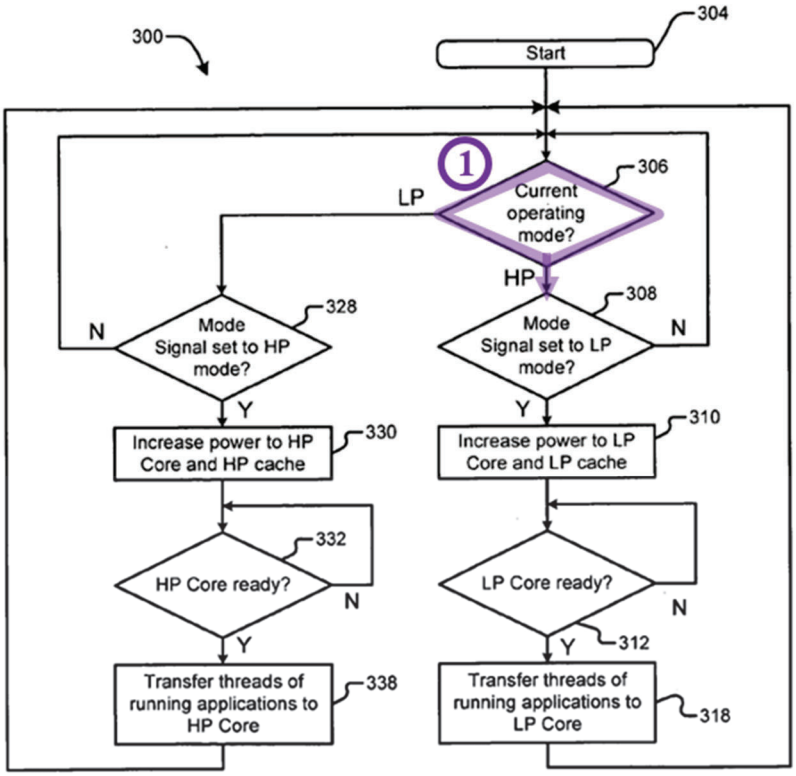



**FIG. 4**

Ex-1007, Figure 4 (annotated).

99. Sutardja explains that the HP mode may also correspond to a state in which both the HP and LP cores are active: "the LP core 28 may be in the active state during all or part of the HP mode." Ex-1007, ¶223. Thus, a POSITA would have understood that step 306 of Figure 4 can correspond to a state in which both the HP and LP cores are enabled. A POSITA would have been motivated to enable both Sutardja's HP and LP cores to operate together, for example, when the workload to be processed by Sutardja's processing system is large, such that the processing resources of Sutardja's HP cores, alone, are not sufficient to achieve the desired performance level. *See, e.g.*, Ex-1024, 0002-0003 (explaining that power can be conserved in a multicore processing system that includes both high-power and low-power cores not only "[i]n times of light load, [when] activity shifts to low-power cores," but also "in times of heavy load, [when] low-power cores can offer additional parallelism without significant increases in area or power consumption"). As another example, a POSITA would have been motivated to enable both Sutardja's HP and LP cores together, by actively operating the HP cores and maintaining the LP cores in "standby" mode, because doing so would "allow quicker transitions to the active state" for the LP cores, when transitioning to the LP mode. Ex-1007, ¶224. In particular, a POSITA would have understood that powering up a core that had previously been powered down would take more time than simply increasing the power being supplied to the core.

100. Second, Sutardja teaches a demand in the form of "system load" and/or "power/performance demand":

> [C]ores are dynamically enabled (i.e., activated) or disabled (i.e., deactivated) based on the system load. The ***system load may be partly determined based on resource utilization and performance demanded*** by the applications.

Ex-1007, ¶218.

> The core profile module 412 may generate an event when the power/performance requirement changes to a value greater or less than a threshold value…. The event is a request for switching cores according to the change in the ***power/performance demand***.

Ex-1007, ¶286.

101. Third, Sutardja teaches disabling all of the first plurality of cores (Sutardja's HP cores) for a drop in demand (system load) below a threshold, without disabling any of the second plurality of cores (Sutardja's LP cores). As illustrated in annotated Figure 4 (reproduced with annotations below), step 308 (green) of method 300 involves determining whether Sutardja's multi-core processor should switch from the HP mode to the LP mode.

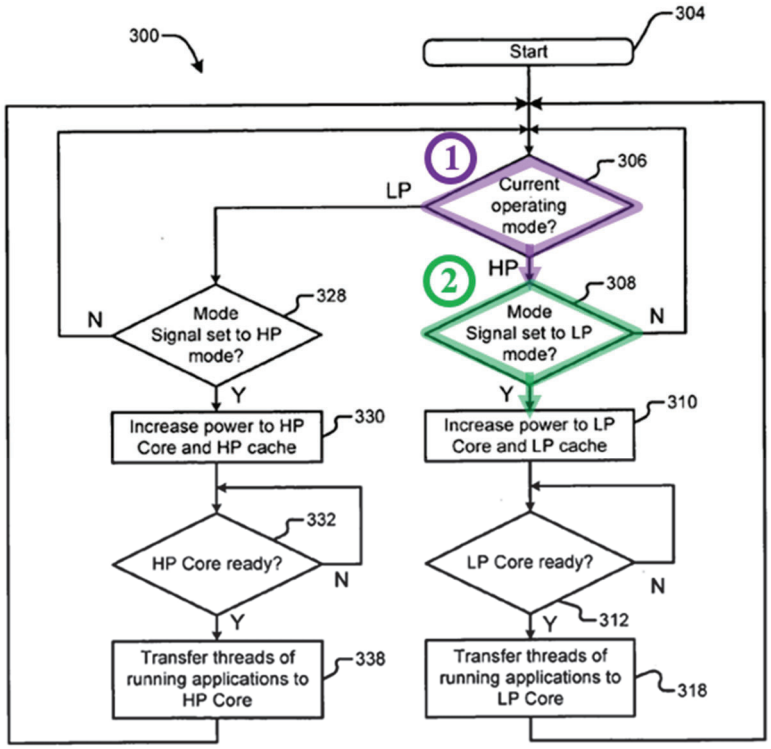



**FIG. 4**

Ex-1007, Figure 4 (annotated); Ex-1007, ¶245 ("If step 306 is equal to HP, control continues with step 308 and determines whether the mode signal is set to the LP mode. If step 308 is true, control continues with step 310 and increases power to the LP core 28 and the cache 30.").

102. Sutardja explains that the decision to switch from the HP mode to the LP mode is based on system loading (the claimed demand):

> After executing instructions using the HP core 204, system loading may eventually decrease. When the system loading ***decreases to a predetermined point***, the core switching module 212 may transition operation from the HP core 204 back to the LP core 200. As can be

> appreciated, systems employing core switching may benefit from the low power consumption of the LP core 200 when relatively low processing speed is required. When high speed processing is desired, the HP core 204 can be used and then turned off when the processing load decreases sufficiently.

Ex-1007, ¶259.

> The core profile module 412 may generate an event when the ***power/performance requirement changes to a value*** greater or ***less than a threshold value***…. The event is a request for switching cores according to the change in the power/performance demand.

Ex-1007, ¶286.

103. Thus, if system loading (the claimed demand) "decreases to a predetermined point" (the claimed threshold), Sutardja's system transitions to the LP mode (as illustrated in steps 310 through 318 (blue) of annotated Figure 4, below). Ex-1007, ¶245.

104. In the LP mode, Sutardja's HP cores are disabled. As Sutardja explains, "[i]n the LP mode, the LP core 28 operates in the active state and processes threads and the HP core 24 is in the inactive state." Ex-1007, ¶224. Sutardja further explains that in "the inactive state," "[t]he inactive core may be completely turned off (no power supplied)." Ex-1007, ¶224. A core that is "completely turned off" is disabled.

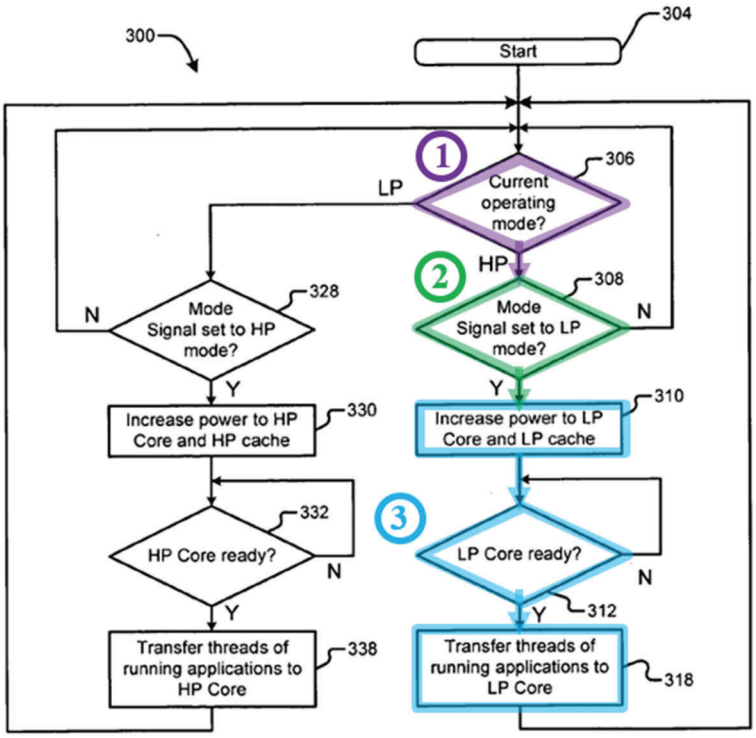



**FIG. 4**

Ex-1007, Figure 4 (annotated).

105. A POSITA would have also understood that LP cores (the claimed second plurality of cores) would remain enabled throughout the transition from the HP mode (during which both the HP and LP cores are enabled) to the LP mode (during which only the LP cores are enabled) (i.e., the transition from the HP mode to the LP mode would occur without disabling any of the LP cores). First, Sutardja discloses that during the transition from the high-power mode to the low-power mode, the low-power cores receive the "threads" and other "information such as

registers, checkpoints and/or program counters" transferred from the high-power cores:

> Operation begins in step 1500. In step 1504, control determines whether the device is operating in a high-power mode. In step 1508, control determines whether ***a transition to low-power mode is requested***. When step 1508 is true, control ***transfers data or graphics threads to the low-power processor or GPU*** in step 1512. In step 1516, control ***transfers information such as registers, checkpoints and/or program counters to the low-power processor or GPU if needed***…. In step 1520, control determines whether the threads and/or other information have been properly transferred to the lower-power processor or GPU. If step 1520 is true, control transitions the high-power processor or GPU to the inactive state.

Ex-1008, ¶237.

106. A POSITA would have understood that sufficient power would need to be supplied to the LP cores (i.e., the LP cores would have to be enabled) in order to receive these threads and information. In particular, a core that is powered down is unable to execute any operations and therefore unable to receive threads and information. Additionally, a POSITA would have recognized that disabling the LP cores momentarily, only to re-enable them, would detrimentally affect performance and power consumption, and would potentially lead to lost threads during the transfer. For example, disabling the LP cores only to re-enable them may cause a delay in processing due to the time needed to power up the previously disabled cores. As another example, threads may be lost if the system attempted to transfer threads

to the LP cores while they were momentarily disabled, because a core that is powered down is unable to execute any operations and therefore unable to receive threads and information. Indeed, Sutardja explains that leaving the LP cores enabled, in a "standby" mode, enables "quicker transitions to the active state" for the LP cores, when transitioning to the LP mode. Ex-1007, ¶224.

107. Finally, Sutardja teaches the claimed power management hardware, by teaching that hardware is used to perform the transition from the HP mode to the LP mode. In particular, Sutardja explains that a "core switching module" is used to transition the system between the HP and LP modes:

> [T]he core switching module 212 copies the state of the LP core 200, activates the HP core 204, and initializes the HP core 204 with the state of the LP core 200. After the HP core 204 is initialized, the core switching module 212 deactivates the LP core 200.

Ex-1007, ¶258.

> If step 256 is true, a control message is sent to the core switching module 212. The core switching module 212 reads the state (pointers, registers, opcodes, operands, program counters (PC), etc.) of the LP core 200 and copies the state or information representative of the state of the LP core 200 into the high-speed memory 210 in step 260. In step 264, the core switching module 212 transitions the HP core 204 to the active state. The core switching module 212 loads the state of the LP core 200 from the high-speed memory 210 into the HP core 204 in step 268. In step 272, the core switching module 212 transitions the LP core 200 to the inactive state.

Ex-1007, ¶261.

> [T]he core switching module 212 reads the state of the HP core 204 and copies the state or information representative of the state of the HP core 204 into the high-speed memory 210 in step 296. Alternatively, the state may be copied directly to the core to be activated. In step 300, the core switching module 212 transitions the LP core 200 to the active state. In step 306, the core switching module 212 loads the state of the HP core 204 from the high-speed memory 210 into the LP core 200.

Ex-1007, ¶263.

108. The hardware controlled by the core switching module and used to transition between the HP and LP modes corresponds to the claimed power management hardware.

109. Sutardja explains that the core switching module "may be integrated with [a] hypervisor module 213," as illustrated in annotated Figure 5, below. Ex-1007, ¶253.

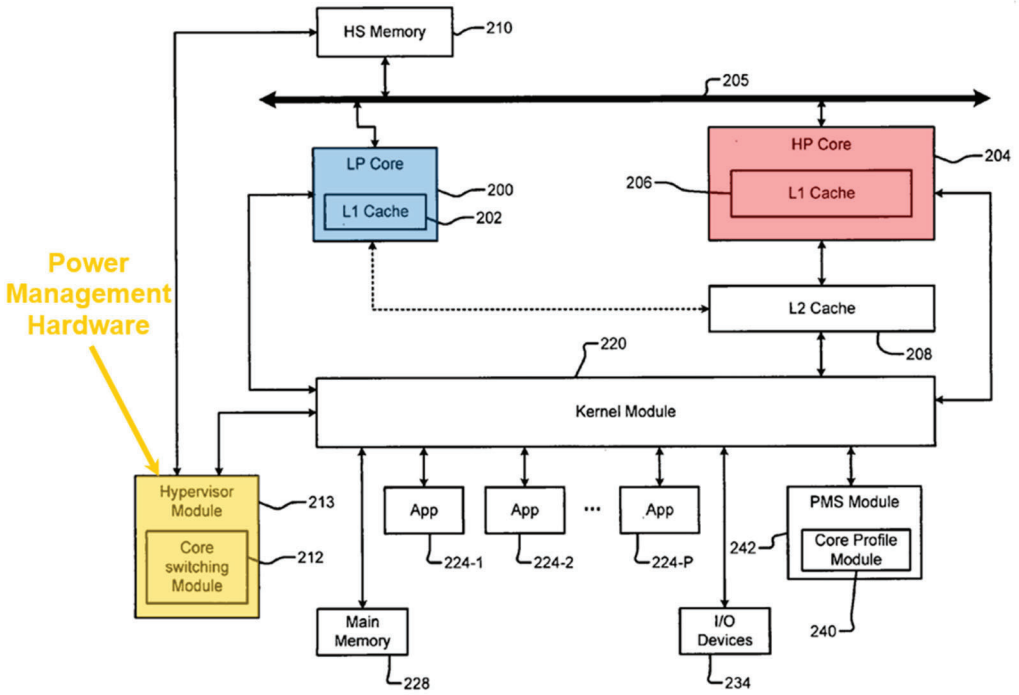



FIG. 5

Ex-1007, Figure 5 (annotated).

110. Sutardja teaches that the hypervisor may "operate[] above the kernel" or it "may operate below the kernel." Ex-1007, ¶254. Because the kernel is the lowest level of the operating system, a POSITA would have understood that a hypervisor that "operate[s] below the kernel" is installed directly on the underlying hardware. Ex-1007, ¶¶249, 254. For example, as explained in *Enabling the New Era of Cloud Computing: Data Security, Transfer, and Management*:

> There are chiefly two kinds of Hypervisor: the bare-metal hypervisor and the hosted hypervisor. For the bare-Metal hypervisor, ***the hypervisor runs directly on the hardware***… and executes commands against the hardware. … The hosted hypervisor is a hypervisor that runs inside the operating system. It is installed and runs as an application on top of an operating system.

Ex-1028, 0003-0004.

111. A POSITA would have understood that the hardware on which the disclosed hypervisor is installed corresponds to the claimed power management hardware.

**e. Element 1[b][ii]: wherein an operating system to execute on the multi-core processor is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.**

112. Sutardja teaches Element 1[b][ii].

113. Sutardja discloses an operating system executing on the multi-core processor. Sutardja explains that:

> The LP and HP cores 200 and 204 may execute the ***same operating system (OS)***. … A kernel module 220, which is a component of the OS, manages system resources. For example only, ***the OS may include Linux® or other suitable operating system***."

Ex-1007, ¶249.

114. As noted above, "[a] kernel module 220 [] is a component of the OS [and] manages system resources." Ex-1007, ¶249. Sutardja's kernel module is illustrated in annotated Figure 5, below.

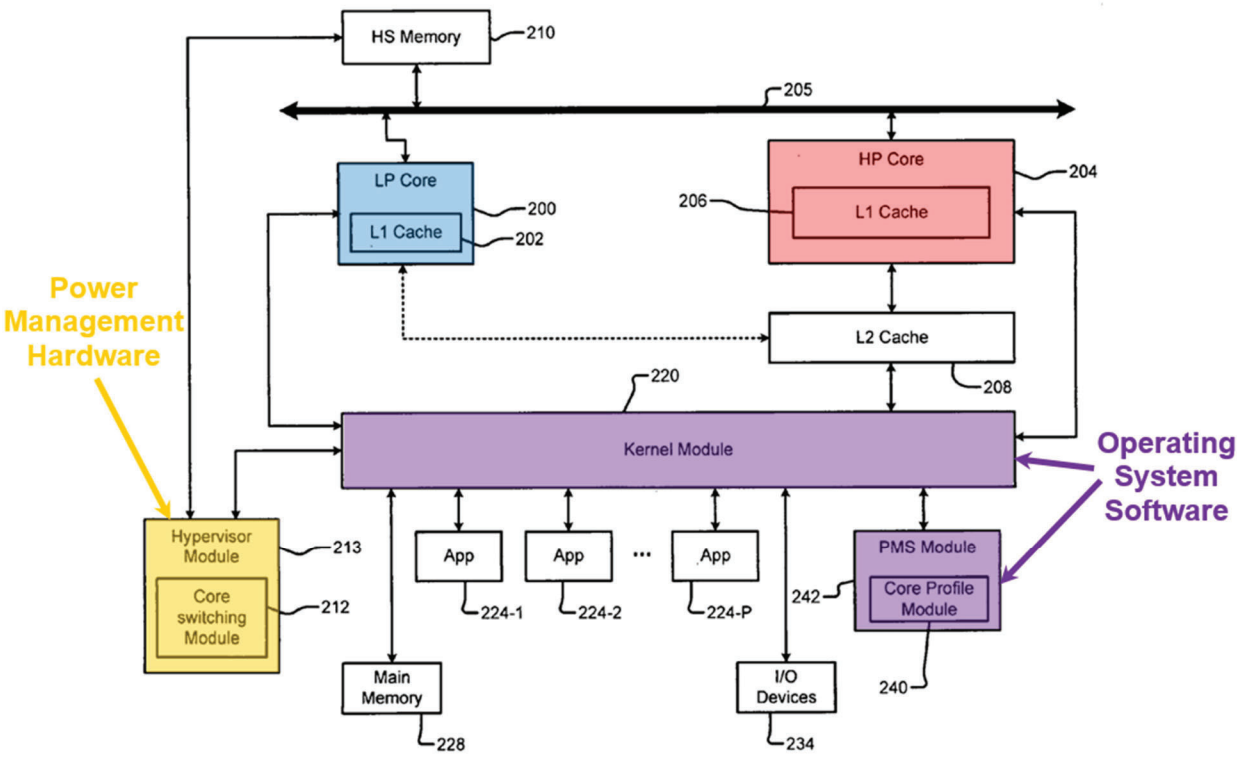



**FIG. 5**

Ex-1007, Figure 5 (annotated).

115. Sutardja's operating system monitors a demand for the multi-core processor. As explained for Element 1[b][i], Sutardja teaches the claimed demand in the form of "system loading" and/or "power/performance demand." *See* Section VIII.A.1.d; Ex-1007, ¶259. As Sutardja explains, a core profile module included in a "power management system (PMS) module" (illustrated in annotated Figure 5 above) monitors this demand:

> A power management system (PMS) module 242 may be provided to manage power consumption and to initiate or control changes in system operation based on the monitored system loading. ***A core profile module 240***

> ***may monitor system loading*** such as core loading, input/output loading, application loading, and/or loading based on other system operating parameters.

Ex-1007, ¶251.

> In step 252, ***the core profile module 240 monitors system loading*** of the LP core 200. In step 256, the core profile module 140 determines whether there is a need to switch to the HP core 204.

Ex-1007, ¶260.

116. Sutardja's "FIG. 5 is a functional block diagram." Ex-1007, ¶195. A POSITA would have understood that functional block diagrams like Figure 5 are used to depict how various software modules function together. Thus, a POSITA would have understood that the PMS Module is executed by the kernel module (part of the Linux OS) to control the functionality of Sutardja's multi-core processor. A POSITA would have considered the software modules executed by the kernel module to control the functionality of Sutardja's multi-core processor to be part of the operating system. Thus, Sutardja discloses an operating system that includes kernel module and PMS module 242.

117. Sutardja further discloses that the operating system (including the kernel module and PMS module) controls the power management hardware (core switching module) based on the demand. For example, Sutardja explains that:

> When the LP core 200 reaches its highest operating speed and additional system loading occurs, the core profile module 240, ***the PMS module 242, and/or the kernel***

> ***module 220*** disable interrupts and ***generate a system call to a hypervisor module 213 or a core switching module 212 to trigger a transition***. The core switching module 212 may be standalone or may be integrated with the hypervisor module 213. The core switching module 212 may transition (switch) the operation from the LP core 200 to the HP core 204 in response to the system call and return control to the kernel module 220 after the transition (switching) is complete.

Ex-1007, ¶253.

> In step 288, ***the core profile module 240 monitors system loading*** of the HP core 204. In step 292, ***the core profile module 240 determines whether there is a need to switch*** to the LP core 200.
>
> If step 292 is false, control returns to step 288. Otherwise ***if step 292 is true***, the core switching module 212 reads the state of the HP core 204 and copies the state or information representative of the state of the HP core 204 into the high-speed memory 210 in step 296. Alternatively, the state may be copied directly to the core to be activated. In step 300, ***the core switching module 212 transitions the LP core 200 to the active state***. In step 306, the core switching module 212 loads the state of the HP core 204 from the high-speed memory 210 into the LP core 200.

Ex-1007, ¶¶262-63.

**2. Dependent Claim 2: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores.**

118. Sutardja teaches Claim 2.

119. Indeed, during prosecution of the ’080 Patent, the Examiner found that the parent application to Sutardja ’748 (which includes the same disclosure as Sutardja ’748) taught Claim 2, and the Applicant did not dispute that finding. Ex-1004, 0054-0057, 0067-0072.

120. Sutardja’s LP and HP cores are comprised of logic gates and these gates are comprised of transistors. Ex-1007, ¶230. Indeed, as I explained in Section V, above, logic gates, which are the basic building blocks of digital circuits, are implemented using transistors. *See supra* ¶51. Sutardja’s “HP and LP cores 24 and 28 include transistors 106 and 110, respectively,” as illustrated in annotated Figure 3B, below. Ex-1007, ¶230.

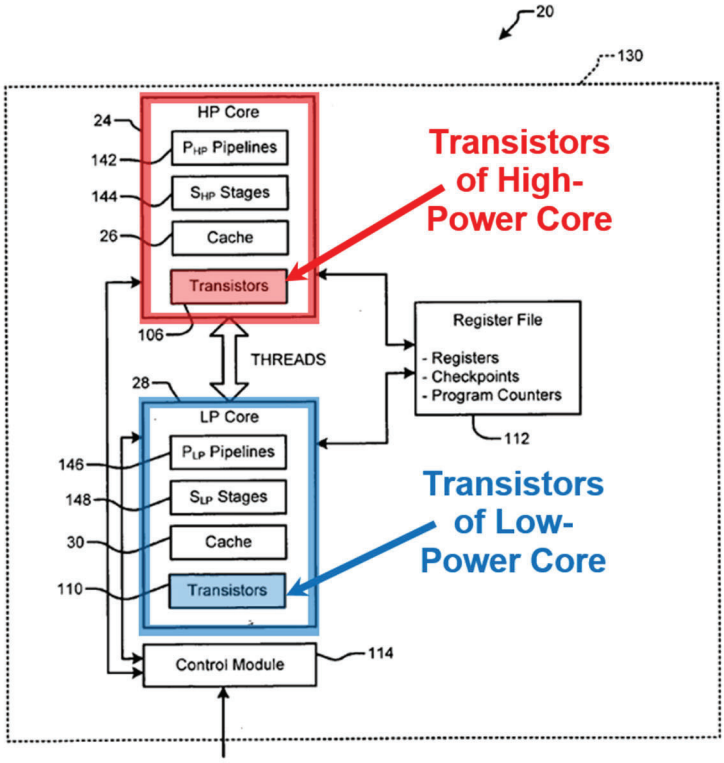



FIG. 3B

Ex-1007, Figure 3B (annotated).

121. Sutardja explains that "[g]ates of the transistors 106 may be larger than gates of the transistors 110." Ex-1007, ¶230. Because the gates of transistors 106 of the HP core are larger than the gates of transistors 110 of the LP core, the gates of transistors 110 of the LP core are therefore narrower than the gates of transistors 106 of the HP core. Sutardja's transistors are logic gate driver transistors, because they are used to implement the logic gates of Sutardja's cores (the fundamental building blocks of the digital circuits within the cores) by driving current through the logic gates. *See supra* ¶51. Thus, Sutardja teaches that the second plurality of cores (Sutardja's LP cores) comprise logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores (Sutardja's HP cores).

**3. Dependent Claim 3: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that consume less power than corresponding logic gates of the first plurality of cores.**

122. Sutardja teaches Claim 3.

123. Indeed, during prosecution of the '080 Patent, the Examiner found that the parent application to Sutardja '748 (which includes the same disclosure as Sutardja '748) taught Claim 3, and the Applicant did not dispute that finding. Ex-1004, 0054-0057, 0067-0072.

124. As previously explained, the cores in Sutardja are comprised of logic gates and these gates are comprised of transistors. Ex-1007, ¶230. Sutardja explains that "[t]he transistors 106 of the HP core 24 tend to consume more power during operation in the active state than the transistors 110 of the LP core 28," and thus, the transistors 110 of the LP core 28 tend to consume less power during operation in the active state than the transistors 106 of the HP core 24. Ex-1007, ¶230. Accordingly, a POSITA would have understood that the logic gates of the LP core (formed from transistors 110) consume less power than the logic gates of the HP core (formed from transistors 106).

**4. Dependent Claim 4: The multi-core processor of claim 1, wherein the second plurality of cores each have a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores.**

125. Sutardja teaches Claim 4.

126. Sutardja provides an example in which the LP core has a maximum operating frequency of 500 MHz, while the HP core has a greater maximum operating frequency, of at least 1 GHz:

> Most mobile computing devices execute applications that have different levels of complexity and/or processing requirements. Some applications can be efficiently executed using a single-core, low-speed processor while other applications can be efficiently executed using a single-core, high-speed processor. For example only, the low-speed processor may operate at speeds that are ***less than or equal to 500 MegaHertz (MHz)*** while the high-

> speed processor may operate at speeds that are ***greater than or equal to 1 GigaHertz (GHz)***.

Ex-1007, ¶210; *see also id*., ¶234.

127. Thus, the second plurality of cores (Sutardja's LP cores) each have a maximum operating frequency (e.g., 500 MHz) that is less than a maximum operating frequency (e.g., greater than 1 GHz) of the first plurality of cores (Sutardja's HP cores).

**5. Dependent Claim 7: The multi-core processor of claim 1, wherein the first plurality of cores are at a maximum operating frequency in the state.**

128. Sutardja teaches Claim 7.

129. Sutardja discloses the voltage and frequency scaling may be applied to the cores. Specifically, Sutardja discloses that the frequency of the HP cores may be adjusted "within a frequency range from low to high":

> Initially, Core 2 [(the HP core)] may operate at a frequency and voltage level specified by the frequency change sequence and the voltage change sequence, respectively. Subsequently, the FCS module 434 ***may change the clock frequency of Core2 within a frequency range from low to high*** and then from high to low according to the performance level demanded by the application.

Ex-1007, ¶304; *see also id.*, ¶301.

130. A POSITA would have been motivated to set the high end of this frequency range to the HP cores' maximum frequency, to enable the HP cores to provide their highest levels of performance, when needed (i.e., when demand is

high). This would enable the HP cores to meet "the performance level demanded by the application[s]" executing on the cores. Ex-1007, ¶304. Indeed, as I explained in Section V above, it was well-known to dynamically scale the frequency at which a core operates based on power and performance considerations. *See supra* ¶¶44, 50.

131. A POSITA also would have recognized that the HP cores may be operating at their maximum operating frequency in the state in which both the HP and LP cores are enabled. In particular, a POSITA would have understood that operating both the HP and LP cores at the same time would consume more power than operating the HP cores alone. Accordingly, a POSITA would have recognized that, when applications are executing only on Sutardja's HP cores and the workload (demand) increases, in order to continue to meet the performance levels required by that rising workload (demand), Sutardja's multi-core system would first increase the operating frequencies of the HP cores to their maximum values, before enabling the LP cores, because doing so would conserve power.

#### 6. Dependent Claim 8: The multi-core processor of claim 1, wherein

##### a. Element 8[a]: the power management hardware is to enable all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores,

132. Sutardja teaches Element 8[a].

133. First, Sutardja teaches a state where the second plurality of cores (Sutardja's LP cores) are enabled, and the first plurality of cores (Sutardja's HP cores) are disabled. As illustrated in Figure 4 (reproduced with annotations below), in step 306 (blue) of method 300, Sutardja's processing system may determine that it is operating in the LP mode, which corresponds to a state in which only the LP cores (the claimed second plurality of cores) are enabled. Ex-1007, ¶224 ("In the LP mode, the LP core 28 operates in the active state and processes threads and the HP core 24 is in the inactive state."); Ex-1007, ¶246.

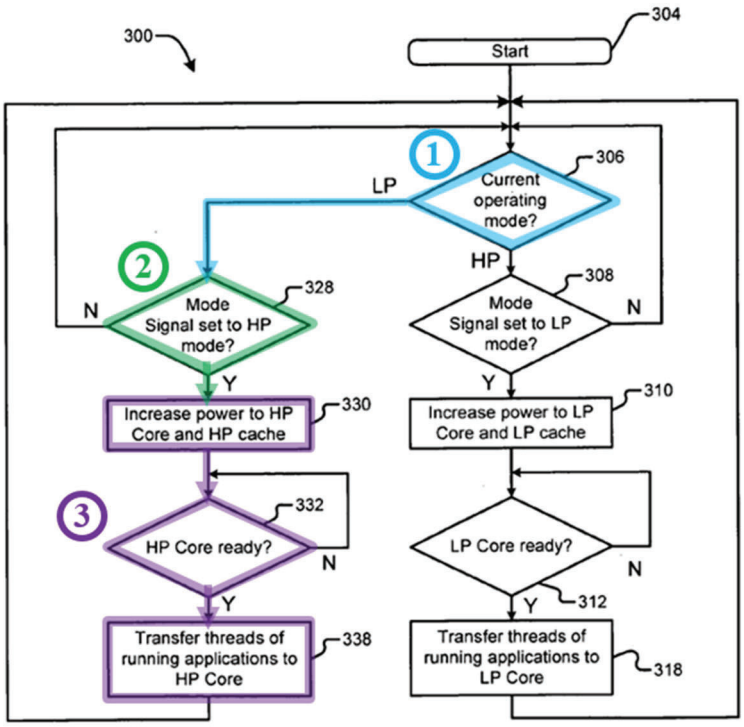



**FIG. 4**

Ex-1007, Figure 4 (annotated).

134. Second, as explained for Element 1[b][i], Sutardja teaches a demand in the form of "system loading" and/or "power/performance demand":

> [C]ores are dynamically enabled (i.e., activated) or disabled (i.e., deactivated) based on the system load. The ***system load may be partly determined based on resource utilization and performance demanded*** by the applications.

Ex-1007, ¶218.

> The core profile module 412 may generate an event when the power/performance requirement changes to a value greater or less than a threshold value…. The event is a request for switching cores according to the change in the ***power/performance demand***.

Ex-1007, ¶286; *see also* Section VIII.A.1.d.

135. Third, Sutardja teaches enabling all of the first plurality of cores (Sutardja's HP cores) for an increase in demand (system load) above the threshold, without disabling any of the second plurality of cores (Sutardja's LP cores). As illustrated in annotated Figure 4, above, step 328 (green) of method 300 involves determining whether Sutardja's multi-core processor should switch from the LP mode to the HP mode.

136. Sutardja explains that the decision to switch from the LP mode to the HP mode (with both the LP and HP cores enabled) is based on system loading (the claimed demand):

> When the LP core 200 reaches its highest operating speed and ***additional system loading occurs***, the core profile

> module 240, the PMS module 242, and/or the kernel module 220 disable interrupts and generate a system call to a hypervisor module 213 or a core switching module 212 to trigger a transition. … The core switching module 212 ***may transition (switch) the operation from the LP core 200 to the HP core 204*** in response to the system call.

Ex-1007, ¶253;

> In step 252, the core profile module 240 ***monitors system loading*** of the LP core 200. In step 256, the core profile module 240 determines whether there is a need to switch to the HP core.

Ex-1007, ¶260;

> The core profile module 412 may generate an event when the ***power/performance requirement changes to a value greater*** or less ***than a threshold value***…. The event is a request for switching cores according to the change in the power/performance demand.

Ex-1007, ¶286.

137. Thus, if system loading (the claimed demand) increases above the capacity of the LP core (i.e., above the claimed threshold), Sutardja's system transitions to the HP mode (as illustrated in steps 330 through 338 (purple) of annotated Figure 4, above). Ex-1007, ¶253; *see also* Ex-1007, ¶286 ("The core profile module 412 may generate an event when the power/performance requirement changes to a value greater or less than a threshold value…. The event is a request for switching cores according to the change in the power/performance demand.").

138. Furthermore, as previously explained, Sutardja discloses that "the LP core [] may be in the active state during all or part of the HP mode." Ex-1007, ¶223; Ex-1008, ¶208. For example, a POSITA would have been motivated to enable both Sutardja's HP and LP cores to operate together when the workload to be processed by Sutardja's system is large, such that the processing resources of Sutardja's HP cores, alone, are not sufficient to achieve the desired performance. *See, e.g.*, Ex-1024, 0002-0003 (explaining that power can be conserved in a multicore processing system that includes both high-power and low-power cores not only "[i]n times of light load, [when] activity shifts to low-power cores," but also "in times of heavy load, [when] low-power cores can offer additional parallelism without significant increases in area or power consumption"). As another example, a POSITA would have been motivated to enable both Sutardja's HP and LP cores together, by actively operating the HP cores and maintaining the LP cores in "standby" mode, because doing so would "allow quicker transitions to the active state" for the LP cores, when transitioning back to the LP mode. Ex-1007, ¶224. Accordingly, a POSITA would have understood that the transition to the HP mode may occur without disabling the LP cores (the claimed second plurality of cores). In particular, a POSITA would have recognized that disabling the LP cores momentarily, only to re-enable them (either in active or standby mode), would detrimentally affect performance and power consumption, and would potentially lead to lost threads and data during the

transition. For example, a core that is powered down is unable to execute any operations and therefore unable to receive threads and information.

139. Finally, as explained for Element 1[b][i], Sutardja also teaches the claimed power management hardware, by teaching that hardware is used to performed the above-described transition. *See* Section VIII.A.1.d.

**b. Element 8[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.**

140. Sutardja teaches Element 8[b] for the same reasons as explained above for Element 1[b][ii]. *See* Section VIII.A.1.e.

**7. Independent Claims 9 and 17:**

**a. Element 9[preamble]: A method comprising:**

141. To the extent the preamble is limiting, Sutardja teaches a method. *See, e.g.*, Ex-1007, Cls. 15-28.

**b. Element 17[preamble]: A non-transitory machine readable medium containing program code that when processed by a machine causes a method to be performed, the method comprising:**

142. To the extent the preamble is limiting, Sutardja teaches a non-transitory machine readable medium containing program code. For example, Sutardja discloses "[a] computer program stored on a computer readable medium and executed by a processor…." Ex-1007, Cl. 43.

**c. Elements 9[a][i] and 17[a][i]: operating a multi-core processor such that a first plurality of cores and a second plurality of cores execute a same instruction set,**

143. Sutardja teaches Elements 9[a][i] and 17[a][i] for the same reasons as explained above for Element 1[a][i]. *See* Section VIII.A.1.b.

**d. Elements 9[a][ii] and 17[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and**

144. Sutardja teaches Elements 9[a][ii] and 17[a][ii] for the same reasons as explained above for Element 1[a][ii]. *See* Section VIII.A.1.c.

**e. Elements 9[b][i] and 17[b][i]: disabling with power management hardware, from a state where the first plurality of cores and the second plurality of cores are enabled, all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores,**

145. Sutardja teaches Elements 9[b][i] and 17[b][i] for the same reasons as explained above for Element 1[b][i]. *See* Section VIII.A.1.d.

**f. Element 9[b][ii] and 17[b][ii]: wherein an operating system executing on the multi-core processor monitors a demand for the multi-core processor and controls the power management hardware based on the demand.**

146. Sutardja teaches Elements 9[b][ii] and 17[b][ii] for the same reasons as explained above for Element 1[b][ii]. *See* Section VIII.A.1.e.

**8. Dependent Claims 10 and 18: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating of the second plurality of cores comprises driving logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores.**

147. Sutardja teaches Claims 10 and 18 for the same reasons as explained above for Claim 2. *See* Section VIII.A.2.

**9. Dependent Claims 11 and 19: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating of the second plurality of cores comprises driving logic gates that consume less power than corresponding logic gates of the first plurality of cores.**

148. Sutardja teaches Claims 11 and 19 for the same reasons as explained above for Claim 3. *See* Section VIII.A.3.

**10. Dependent Claims 12 and 20: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the second plurality of cores at a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores.**

149. Sutardja teaches Claims 12 and 20 for the same reasons as explained above for Claim 4. *See* Section VIII.A.4.

**11. Dependent Claims 15 and 23: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the first plurality of cores at a maximum operating frequency in the state.**

150. Sutardja teaches Claims 15 and 23 for the same reasons as explained above for Claim 7. *See* Section VIII.A.5.

**12. Dependent Claims 16 and 24: The [method of claim 9/non-transitory machine readable medium of claim 17], further comprising**

**a. Elements 16[a] and 24[a]: enabling, with the power management hardware, all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores,**

151. Sutardja teaches Elements 16[a] and 24[a] for the same reasons as explained above for Element 8[a]. *See* Section VIII.A.6.a.

**b. Elements 16[b] and 24[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.**

152. Sutardja teaches Element 16[b] for the same reasons as explained above for Element 8[b]. *See* Section VIII.A.6.b.

**B. Ground 2: Claims 5-6, 13-14, and 21-22 are rendered obvious by Sutardja in view of Rychlik**

**1. Dependent Claims 5, 13, and 21:**

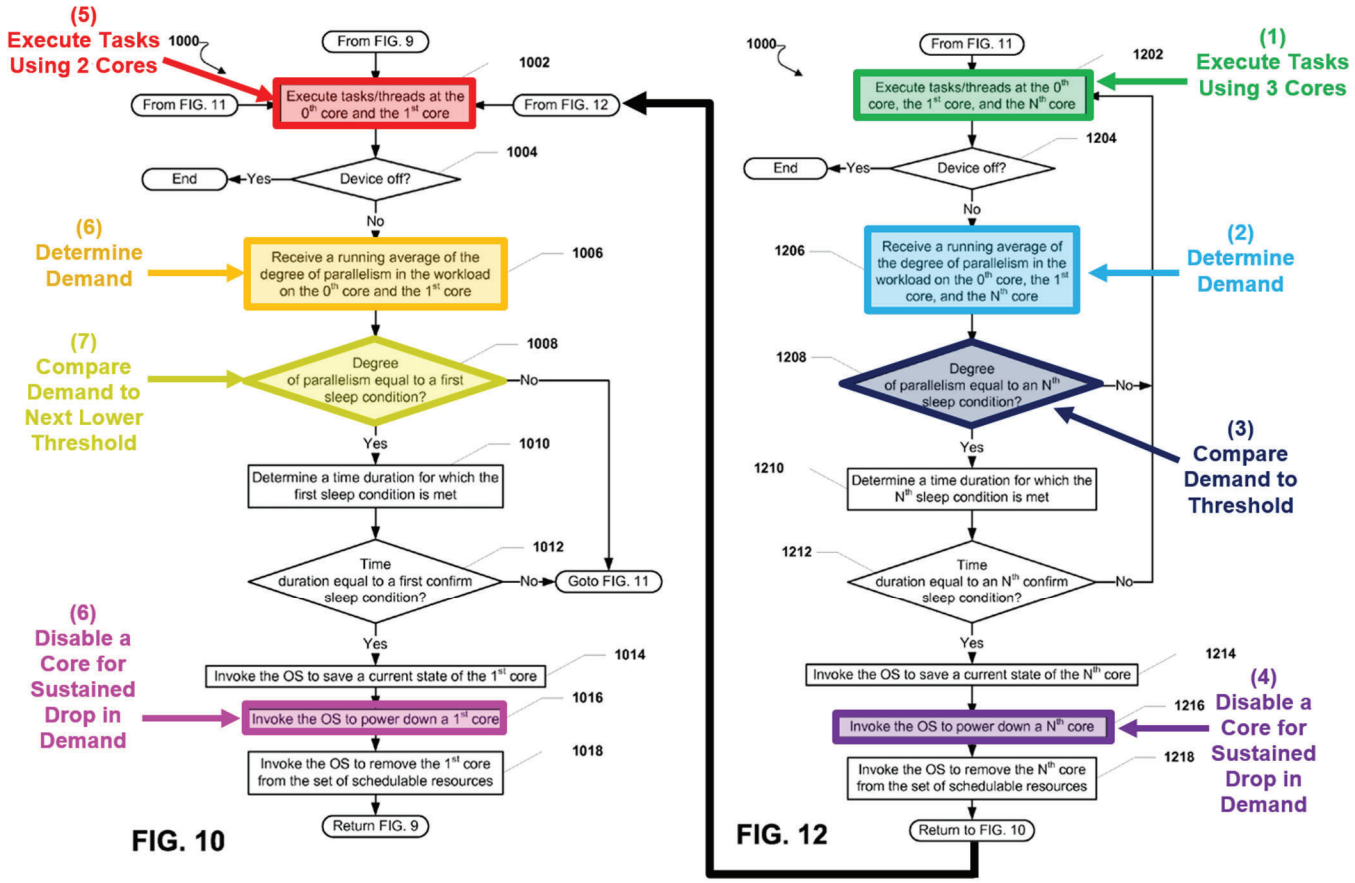
**a. Elements 5[a], 13[a], 21[a]: The [multi-core processor of claim 1/method of claim 9/non-transitory machine readable medium of claim 17], [wherein the power management hardware is to disable/further comprising disabling, with the power management hardware,] an additional core of the second plurality of cores for each continued drop in demand below a next lower threshold until one core of the second plurality of cores remains enabled, and**

153. Sutardja in view of Rychlik teaches Elements 5[a], 13[a], and 21[a].

154. Rychlik discloses a plurality of cores in the form of "[a] multicore CPU 402 [that] may include a zeroth core 410, a first core 412, and an Nth core 414." Ex-1009, ¶34.

155. Rychlik teaches demand for the plurality of cores in the form of a "workload," and explains that the workload is monitored by "a parallelism monitor." Ex-1009, ¶37 ("[A] parallelism monitor 440 may monitor the workload on the cores 410, 412, 414."). The workload (the claimed demand) may take the form of "a total system load," or "a running average of the degree of parallelism in the workload." Ex-1009, ¶38.

156. Rychlik also discloses disabling an additional core of the plurality of cores for each continued drop in demand below a next lower threshold. *See, e.g.*, Ex-1009, ¶96. This process is illustrated in Figures 10 and 12 of Rychlik, annotated below. While Rychlik's method may be applied to a multi-core processor with any number, N, of cores, for ease of explanation only, the annotations and discussion to follow assumes Rychlik's multicore CPU includes N=3 cores.




Ex-1009, Figures 10 and 12 (annotated).

157. As illustrated in Figure 12, above, at step 1202 (green) Rychlik's zeroth core, first core, and Nth (e.g., 3rd) core are enabled:

> At block 1202 of FIG. 12, one or more tasks, threads, or a combination thereof may be executed at the zeroth core, the first core, and the Nth core.

Ex-1009, ¶68.

158. If, at step 1208 (dark blue), the degree of workload parallelism (the claimed demand) satisfies "[an] Nth sleep condition," which may correspond to "the

parallelism [being] less than or equal to [a] ***threshold value***" associated with the Nth sleep condition, then in step 1216 (purple) one of Rychlik's cores is disabled:

> [T]he method may move to block 1206 and the MP controller may receive a running average of the degree of parallelism in the workload on the zeroth core, the first core, and the Nth [core] from the parallelism monitor. In a particular aspect, at any instant in time, the total number of tasks, threads, or a combination thereof, waiting on the ready-to-run queues of an operating system (OS) plus the number of tasks actually running may be an approximation for the degree of parallelism in the workload on the cores.
>
> At decision 1208, the MP controller may determine whether the degree of parallelism is equal to an Nth sleep condition. In a particular aspect, the Nth sleep condition may be a threshold value associated with the degree of parallelism in the workload on the core.

Ex-1009, ¶¶68-69. Assuming the process started with 3 enabled cores, disabling Rychlik's Nth (i.e., 3rd core) would leave Rychlik's zeroth core and first core enabled, as illustrated in step 1002 (red) of Figure 10, above. Ex-1009, ¶71.

159. If, at step 1008 (yellow), the degree of workload parallelism (the claimed demand) satisfies "a first sleep condition," which may correspond to "the parallelism [being] less than or equal to [a] ***threshold value***" associated with the first sleep condition (the claimed next lower threshold), then in step 1016 (pink) one of Rychlik's two remaining cores is disabled, leaving only one of the cores enabled:

> Moving now to block 1002 of FIG. 10, one or more tasks, threads, or a combination thereof may be executed at the zeroth core and the first core. … [T]he method 900 may move to block 1006 and the MP controller may receive a

> running average of the degree of parallelism in the workload on the zeroth core and the first core from the parallelism monitor. …
>
> At decision 1008, the MP controller may determine whether the degree of parallelism is equal to a first sleep condition. In a particular aspect, the first sleep condition may be a threshold value associated with the degree of parallelism in the workload on the core. …
>
> [I]f the degree of parallelism is equal to the first sleep condition, … the method 900 may move to block 1014 and the MP controller may invoke the OS to save a current state of the first core. At block 1016, the MP controller may invoke the OS to power down the first core so that one core, i.e., the zeroth core, is running and executing threads and tasks. Further, at block 1018, the MP controller may invoke the OS to remove the first core from the set of schedulable resources available to the OS.

Ex-1009, ¶¶61-64.

160. Rychlik similarly teaches disabling an additional core each time system load (demand) falls below a next lower threshold load value:

> [O]ne or more cores may be turned on or off based on the load value. It may be determined whether the load value meets a predetermined condition in order to determine whether to turn the one or more cores on or off. For example, ***the load value may be compared to a threshold*** and if the load value exceeds the threshold one or more cores may be turned on or off. For example, if there were two cores operating and the load value indicated that the two cores operating at or near capacity, a third core may be powered on and the workload may be spread between the three cores. Further, ***as the load value decreases below another threshold*** indicating that having a third core operational is wasting power, ***the third core may be powered off*** and the workload may be spread across the

two remaining cores.

Ex-1009, ¶88. A POSITA would have understood that, just like the process illustrated in Figures 10 and 12, as load continued to drop, additional cores would be powered off, until only one of Rychlik's cores remained enabled. In particular, a POSITA would understand that there is nothing special about Rychlik's "third core." Rather, the third core is used in an example of a process that could and should be applied to all of the cores in Rychlik's system to achieve Rychlik's stated purpose of power management in multi-core processors. Indeed, as I explained in Section V, it was well-known to disable cores in response to falling demand in order to conserve power. *See supra* ¶41; Ex-1031, Figure 4.

161. A POSITA would have been motivated to apply Rychlik's teachings to Sutardja's LP cores, to disable an additional core of Sutardja's LP cores (the claimed second plurality of cores), for each continued drop in workload (the claimed demand) below a next lower threshold, until only one of Sutardja's LP cores remained. Both Rychlik and Sutardja relate to the same subject matter—namely, power management in multi-core processors. Ex-1007, ¶212; Ex-1009, ¶88. A POSITA seeking to implement Sutardja's teachings in a system with multiple of Sutardja's LP cores, would understand that the well-known power management technique of processor scaling may be applied to the multiple LP cores to conserve power. In particular, a POSITA would understand that scaling the number of active

cores based on demand, by disabling one or more of those LP cores when demand is low, would save power. *See, e.g.*, Ex-1031, Fig. 4; *see also* Section V.B. Accordingly, a POSITA would have been motivated to search for references, such as Rychlik, that disclose implementation details of processor scaling methods. A POSITA would have been motivated to apply Rychlik's teachings to Sutardja's LP cores, by disabling one or more of Sutardja's LP cores as the demand for processing resources decreases, so as to prevent those cores from "wasting power." Ex-1009, ¶88. A POSITA would have had a reasonable expectation of success in implementing Rychlik's power management technique in Sutardja's multi-core system, because it amounts to use of a known technique (disabling cores based on demand, as taught by Rychlik) to improve a similar system (Sutardja's multi-core system that includes multiple LP cores) in the same way (by reducing power consumption during periods of low demand).

**b. Elements 5[b], 13[b], 21[b]: [lower/lowering] an operating frequency or a supply voltage of the one core of the second plurality of cores as demand drops below a next lower threshold.**

162. Sutardja in view of Rychlik teaches Elements 5[b], 13[b], and 21[b].

163. Rychlik and Sutardja both teach lowering an operating frequency and/or a supply voltage of the one core of the second plurality of cores as demand

drops. Rychlik discloses including a "dynamic clock and voltage scaling (DCVS) algorithm" on each core:

> The multicore CPU 402 may include a zeroth core 410, a first core 412, and an Nth core 414. The zeroth core 410 may include a zeroth dynamic clock and voltage scaling (DCVS) algorithm 416 executing thereon. The first core 412 may include a first DCVS algorithm 417 executing thereon. Further, the Nth core 414 may include an Nth DCVS algorithm 418 executing thereon. In a particular aspect, each DCVS algorithm 416, 417, 418 may be independently executed on a respective core 412, 414, 416.

Ex-1009, ¶34.

164. Rychlik explains that, when only one core is enabled (the claimed one core of the second plurality of cores), the frequency of the core may be raised to accommodate rising levels of demand:

> Under normal operation, in a dual-core example, the system may start out with a DCVS algorithm operating on zeroth CPU core, while a first CPU core remains dormant. In this mode, ***the system may respond to instantaneous transient loads by ramping up the frequency*** of the zeroth CPU core.

Ex-1009, ¶100. A POSITA would have understood that, in an equivalent manner, Rychlik's system would lower the frequency of the core in response to lower levels of demand (i.e., demand falling below a next lower threshold). In particular, a POSITA would have understood that it would waste power to maintain a core with a "ramp[ed] up [] frequency" when the "transient load" has passed, such that a lower

frequency is sufficient to meeting current processing requirements. *See supra* ¶50 (explaining that it was well-known that during periods of low load, when high performance is not needed, power can be reduced by lowering the supply voltage and operating frequency of any active cores).

165. Sutardja also teaches the use of voltage and frequency scaling on Sutardja's LP core (the claimed one core of the second plurality of cores):

> In use, Core1 may be the LP core, and Core2 may be the HP core. The mobile computing device may normally execute most applications using Core1. Core1 may operate at a frequency and voltage level specified by the frequency change sequence and the voltage change sequence, respectively. The frequency change sequence may include control signals that increase or decrease the clock frequency of the active core sequentially as needed. The voltage change sequence may include control signals that increase or decrease the supply voltage of the active core sequentially as needed.
>
> When high performance is required (e.g., when utilization of Core1 is maximum at current clock frequency and supply voltage), the FCS module 434 may vary the clock frequency of Core1 within a frequency range from low to high based on the frequency change sequence generated for Core1 until the highest and/or desired operating frequency of Core1 is reached. Alternatively, or additionally, the VCS module 436 may change the supply voltage of Core1 within a voltage range from low to high based on the voltage change sequency generated for Core1 until the highest and/or desired operating supply voltage of Core1 is reached.

Ex-1007, ¶¶301-302.

166. In particular, Sutardja explains that for low workloads (demand), the frequency and/or voltage of the LP core may be reduced to save power: "The [system] may monitor CPU utilization and determine a lowest possible clock frequency and/or supply voltage for operating the … LP cores … [to] minimize impact on performance based on workload while saving power." Ex-1007, ¶237.

167. As explained above for Elements 5[a], 13[a], and 21[a], a POSITA would have been motivated to implement Rychlik's power management technique in Sutardja's multi-core processor. Section VIII.B.1.a. A POSITA would also have been motivated to apply the voltage/frequency teachings of Rychlik and Sutardja to the combined Sutardja-Rychlik system, in order to conserve power when only one of Sutardja's LP cores remains enabled, and demand continues to fall (i.e., demand drops below the claimed next lower threshold).

**2. Dependent Claims 6, 14, 22: [The multi-core processor of claim 5/method of claim 13/The non-transitory machine readable medium of claim 21], [wherein the power management hardware is to raise/further comprising raising, with the power management hardware,] a supply voltage or an operating frequency of said one core in response to higher demand.**

168. Sutardja in view of Rychlik teaches Claims 6, 14, and 22.

169. As explained for Claims 5, 13, and 21, above, a POSITA would have been motivated to implement Rychlik's power management techniques in Sutardja's multicore processor. *See* Section VIII.B.1. And, as explained for Elements 5[b],

13[b], and 21[b], Rychlik teaches that, when only one core in Rychlik's system is enabled (the claimed one core of the second plurality of cores), the frequency of that core may be raised to accommodate rising levels of demand: "the system may respond to instantaneous transient loads by ramping up the frequency of the [single] CPU core." Ex-1009, ¶100.

### C. Ground 3: Claims 7, 15, and 23 are rendered obvious by Sutardja in view of Carmack

#### 1. Dependent Claims 7, 15, 23: The [multi-core processor of claim 1/method of claim 9/non-transitory machine readable medium of claim 17], wherein [the operating comprises operating] the first plurality of cores [are] at a maximum operating frequency in the state.

170. Sutardja in view of Carmack teaches Claims 7, 15, and 23.

171. Carmack teaches that when first and second cores 110, 120, with differing power consumption properties, are operating together, the second cores 120 (the claimed first plurality of cores) may be operating at their peak (i.e., maximum) frequency:

> When both cores 110, 120 are utilized, the second core 120 would be given a fraction of the total work scaled to its performance advantage over the first core 110. For situations where the overhead of coordinating asymmetric cores becomes too high, ***the system can lower the peak frequency of the faster core 120 to match the maximum frequency of the slower core 110***….

Ex-1006, ¶22. Because Carmack explains that the frequency of the faster core 120 is reduced from its peak value when "the overhead of coordinating asymmetric cores

becomes too high," a POSITA would have understood that for situations where the overhead of coordinating is ***not*** too high Carmack's faster core 120 would be operating at its peak (i.e., maximum) frequency.

172. A POSITA would have been motivated to combine the teachings of Sutardja and Carmack, and would have had a reasonable expectation of success in doing so because each relates to the same well-known subject matter of power management in multi-core processors. Ex-1006, Abstract, ¶14; Ex-1007, ¶259. In particular, both relate to techniques for power reduction in multi-core processors that focus on the inclusion of cores with different power consumption properties within the same processing system. Ex-1006, ¶¶17-18; Ex-1007, ¶212.

173. Sutardja discloses that the operating frequency of its HP cores may be varied from a low value to a high value, "according to the performance level demanded." Ex-1007, ¶304. A POSITA would have been motivated to set the high value to the maximum operating frequency of Sutardja's HP cores, as taught by Carmack, to maintain desired performance levels during periods of high demand. A POSITA would have had a reasonable expectation of success in doing so, because it amounts to a simple substitution of one known element (Sutardja's "high" operating frequency value) for another (Carmack's maximum operating frequency) to obtain predictable results (Sutardja's HP cores operating at their maximum frequency).

### D. Ground 4: Claims 1-4, 7-12, 15-20, and 23-24 are rendered obvious by Mathieson in view of Sutardja

174. Mathieson incorporates, by reference, the disclosure of Carmack. Ex-1005, ¶1. Thus, the Carmack disclosures and teachings discussed below are attributable to Mathieson.[3]

#### 1. A POSITA would have been motivated to combine Mathieson with Sutardja

175. In addition to the reasons provided below for specific claim elements, a POSITA would have been motivated to combine the teachings of Mathieson and Sutardja, and would have had a reasonable expectation of success in doing so because each relates to the same well-known subject matter of power management in multi-core processors. Ex-1005, Abstract (describing "[a] technique for reducing the power consumption required to execute processing operations"), ¶31; Ex-1007, ¶¶259, 278 ("[A] power management system … that optimizes power consumption of the mobile device using dynamic core switching is shown."). In particular, both references teach techniques for reducing power consumption by multi-core processing systems in response to decreased demand/workload. Ex-1005, ¶¶31-32;

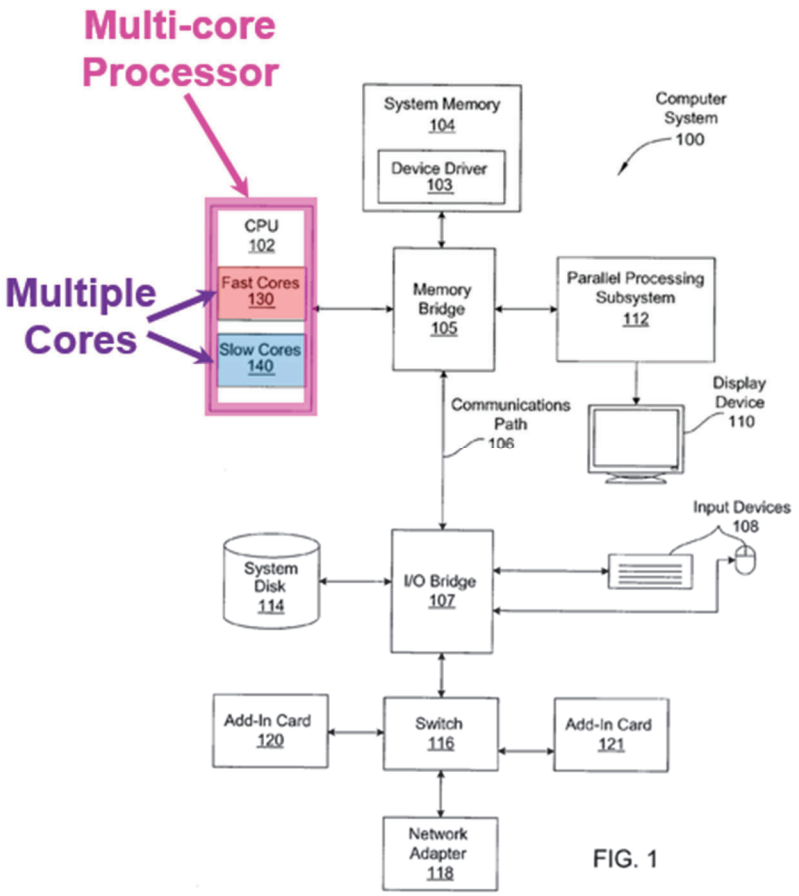
[3] For clarity, Ground 4 relies on Mathieson (Ex-1005) incorporating by reference Carmack (Ex-1006) as a single reference obviousness ground, referred to herein as simply "Mathieson." To the extent that any relied upon teaching is argued to have not been fully incorporated from Ex-1006 into Ex-1005, it would have been obvious to combine any such teaching from Ex-1006 with Ex-1005, for the same reasons provided herein.

Ex-1007, ¶259. To implement their power reduction techniques, both references rely on the inclusion of cores with different power consumption properties within the same processing system. Ex-1005, ¶20 ("The CPU 102 includes one or more 'fast' cores 130 and one or more 'shadow' or slow cores 140…. In some embodiments, the cores 130 are associated with higher performance and higher leakage power than the cores 140."); Ex-1007, ¶212 ("According to the present disclosure, both increased power savings and increased performance may be achieved in a mobile device by using a multi-core processing system with a low-speed, low-power (LP) core and a high-speed, high-power (HP) core."). Furthermore, both references contemplate that the differences in power consumption between the cores may be achieved through differences in the constituent transistors. Ex-1005, ¶28 ("[T]he cores 212 may be manufactured using faster transistors that have significant static leakage."); Ex-1007, ¶230 ("The transistors 106 of the HP core 24 tend to consume more power during operation in the active state than the transistors 110 of the LP core 28. In some implementations, the transistors 106 may have higher leakage current than the transistors 110. The transistors 106 may have a size that is greater than a size of the transistors 110. Gates of the transistors 106 may be larger than gates of the transistors 110."). As explained below, at least because of these similarities, a POSITA seeking to implement Mathieson's teachings would have looked to Sutardja for additional implementation details.

### 2. Independent Claim 1

#### a. Element 1[preamble]: A multi-core processor comprising:

176. Even if the preamble is limiting, Mathieson discloses a multi-core processor in the form of "a central processing unit (CPU) 102," illustrated below in annotated Figure 1. Ex-1005, ¶20.




Ex-1005, Figure 1 (annotated).

177. Mathieson's CPU 102 is a multi-core processor because it includes multiple cores. In particular, Mathieson explains that "[t]he CPU 102 includes one

or more 'fast' cores 130 and one or more 'shadow' or slow cores 140." Ex-1005, ¶20.

**b. Element 1[a][i]: a first plurality of cores and a second plurality of cores that support a same instruction set,**

178. Mathieson and Sutardja teach Element 1[a][i].

179. Mathieson discloses a first plurality of cores in the form of "a first set of cores 210," and a second plurality of cores in the form of "a second set of cores 220," as illustrated in annotated Figure 2, below. Ex-1005, ¶25.

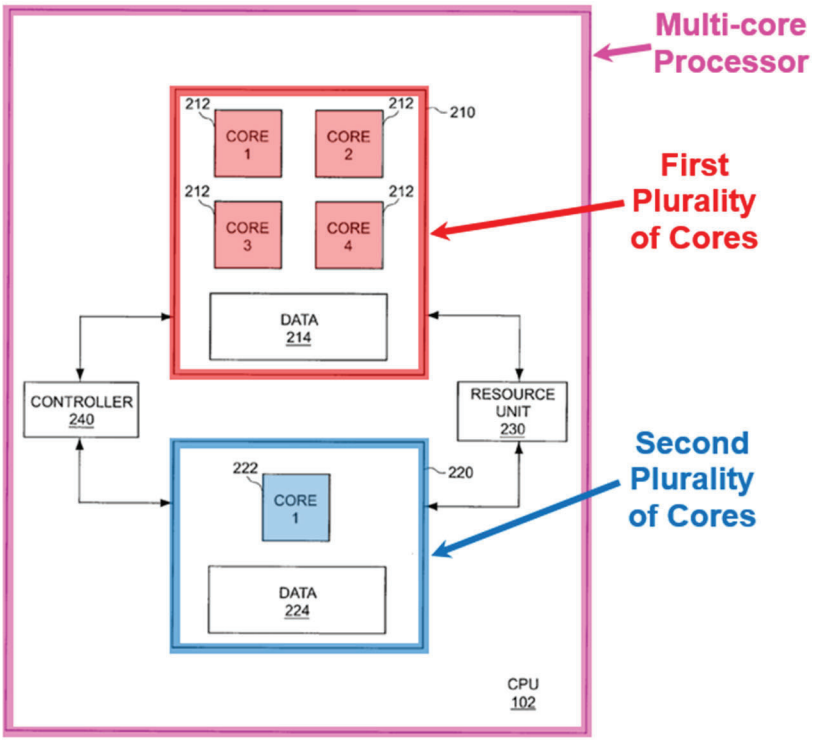



FIG. 2

Ex-1005, Figure 2 (annotated).

180. While Figure 2 illustrates that the first set of cores 210 (the claimed first plurality of cores) includes 4 cores 212, and that the second set of cores 220 (the

claimed second plurality of cores) only includes one core 222, Mathieson discloses that each set of cores 210 and 220 may include any number of cores:

> [T]he first set of cores 210 includes one or more cores 212 and data 214, and the second set of cores 220 includes one or more cores 222 and data 224.

Ex-1005, ¶25;

> According to various embodiments, ***the first set of cores includes N cores and the second set of cores includes M cores***. In one embodiment, N is equal to M. In some embodiments, the first set of cores 210 may include multiple cores, e.g., four cores, and the second set of cores 220 may include a single core 222. In other embodiments, the first set of cores 210 may include a single core and/or the second set of cores 220 may include multiple cores.

Ex-1005, ¶28.

181. Mathieson in view of Sutardja renders obvious that the first plurality of cores and the second plurality of cores support a same instruction set.

182. First, Mathieson explains that "the cores included in the first set of cores 210 may implement substantially ***the same functionality*** as the cores included in the second set of cores 220." Ex-1005, ¶26. Because an instruction set architecture (ISA) specifies the tasks a processor is capable of doing as well as how those tasks gets done, a POSITA would have understood that a straightforward way to enable both of Mathieson's sets of cores 210 and 220 to implement the same functionality, as taught be Mathieson, would be to implement the same ISA (and thereby support

the same instruction set) in each core. This is so because cores that support the same instruction set will be able to perform the same functionality.

183. Second Mathieson explains that "processor state" may be transferred between Mathieson's cores:

> When execution of the processing operations switches from the first set of cores to the second set of cores, in some embodiments, the controller 240 is configured to ***transfer the processor state from the first set of cores to the second set of cores***.

Ex-1005, ¶43.

184. To accomplish this, Mathieson explains that "the processing state may be ***directly transferred*** to the other set of cores via a dedicated bus." *Id*. A POSITA would have understood a straightforward way to ensure the ability to directly transfer the processing state from the first set of cores to the second set of cores, is for both sets of cores to support the same instruction set architecture (ISA). In particular, a POSITA would have understood that, in order for the execution of processing operations to be transferred from a core with a first ISA to a core with a second ISA, the processing state from the first core (associated with the first ISA) would need to be translated into an equivalent form associated with the second ISA. *See, e.g.*, Ex-1033, 6:24-38. Because Mathieson does not suggest the need for any such translation, a POSITA would have understood that Mathieson contemplates that both of Mathieson's sets of cores 210 and 22 implement the same ISA (and accordingly,

the same instruction set), such that the processor state from the first set of cores 210 can be transferred to the second set of cores 220 directly, without translation.

185. Based on these teachings, a POSITA seeking to produce Mathieson's multi-core processor would have been motivated to search for implementation details in similar references that teach cores with different power consumption properties that implement the same functionality. Sutardja is one such reference, which provides express teachings to use the "same instruction set" for cores with different power consumption properties. *See* Section VIII.A.1.b.

186. As explained for Ground 1, Sutardja discloses "a multi-core processing system with a low-speed, low-power (LP) core and a high-speed, high-power (HP) core," where the differences in speed and power consumption may be due to differences in transistor gate size between the cores. Sections VIII.A.1.b, VIII.A.2-3; Ex-1007, ¶¶212, 230. Sutardja explains that "dynamic compatibility" may be achieved between the two types of cores through the use of the same instruction set on each of the cores: "the instruction set of the LP core may be ***the same*** or a subset of the instruction set of the HP core." Ex-1007, ¶270; *see also* Ex-1007, ¶¶213-214; Ex-1008, ¶¶209, 216, 224, 231. A POSITA would have understood that "dynamic compatibility" between Sutardja's LP and HP cores corresponds to the ability to dynamically switch between the two types of cores when executing a set of operations. *See also* Ex-1007, ¶¶269-70 (explaining that the same instruction set is

one way in which "[c]ore switching can be transparent to the OS and the applications"). This would require that both types cores support execution of those operations, and therefore, the functionality associated with those operations.

187. Accordingly, to achieve Mathieson's teaching that both sets of Mathieson's cores implement the same functionality, a POSITA would have been motivated to design both types of cores to support the same instruction set, as taught by Sutardja. To do so, a POSITA would have been motivated to use the same logic design for both types of cores and simply scale the transistor size in one type of core as compared to the other. Not only would the resulting cores with appropriately-scaled transistors exhibit the desired differences in performance and power-consumption, as taught by Mathieson, but the use of the same logic design for the two types of cores would also have the added benefit of ease of manufacturability, ease of design, test, and verification, and other benefits. For example, simply scaling transistor sizes without changing the basic logic design would have simplified the design process by reducing the number of changes to be made during the design and implementation of the different cores.

**c. Element 1[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and**

188. Mathieson teaches Element 1[a][ii].

189. As illustrated in Figure 6 of Mathieson (reproduced with annotations below), for a large frequency range (i.e., up until the frequency associated with line 610), the power consumed by Mathieson's second set of cores (blue) is less than the power consumed by Mathieson's first set of cores (red) for ***the same operating frequency***. For example, as illustrated in annotated Figure 6, below, for frequency $F_0$, the power $P_2$ consumed by Mathieson's second set of cores (the claimed second plurality of cores) is significantly less than the power $P_1$ consumed by Mathieson's first set of cores (the claimed first plurality of cores).

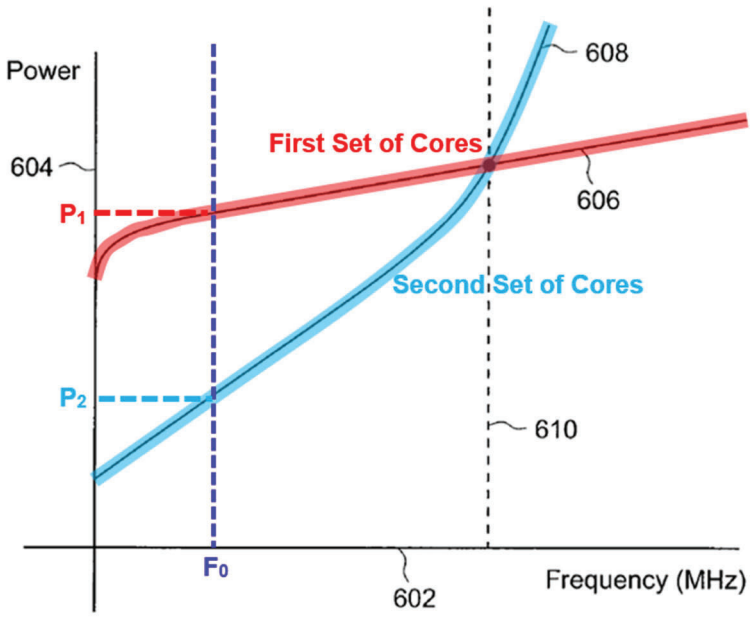



FIG. 6

Ex-1005, Figure 6 (annotated).

190. As Mathieson explains, for "particular operating characteristics," including "operating frequency [and] voltage," either an operating mode in which the first set of cores is enabled or an operating mode in which the second set of cores is enabled may be "more optimal," with "the CPU 102 [] us[ing] less power to operate in [that] mode." Ex-1005, ¶¶30-31, 37.

191. Mathieson explains that the difference in power consumption levels between the two sets of cores 210 and 220, for the same operating characteristics (e.g., frequency and voltage), is due to microarchitectural differences between the cores:

> [T]he cores 212 included in the first set of cores 210 comprise "fast" cores and the cores 222 included in the second set of cores 220 comprise "slow" cores. For example, the cores 212 may be manufactured using ***faster transistors*** that have ***significant static leakage***. In some embodiments, when the computing needs and/or workload of the first set of cores 210 are lowered, then the clock speed is lowered to reduce power. The static leakage is not a significant issue at the high clock speeds required for peak performance. However, ***at slower clock speeds, the static leakage of the fast transistors can dominate the overall power consumption***.

Ex-1005, ¶28.

192. The '080 Patent similarly explains that the difference in power consumption between its two sets of cores arises from transistor differences between the cores:

> By design of the multiprocessor, referring to FIG. 5, the lower power core(s) exhibit inherently lower power consumption (and processing performance) than the higher power core(s). That is, for a same applied clock or operating frequency, ***because of its narrower drive transistor widths***, for example, ***a lower power core will consume less power than a higher power core***.

Ex-1001, 4:13-16.

193. Thus, both Mathieson and the '080 Patent disclose that transistor differences between the two types of cores are responsible for the different power consumption properties of the cores.

**d. Element 1[b][i]: power management hardware to, from a state where the first plurality of cores and the second plurality of cores are enabled, disable all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores,**

194. Mathieson teaches Element 1[b][i].

195. First, Mathieson teaches a state where the first plurality of cores (Mathieson's first set of cores 210) and the second plurality of cores (Mathieson's second set of cores 220) are both enabled. As illustrated in Figure 4A below, at step 402 (purple) of Mathieson's method 400A, CPU 102 is in the first mode of operation, corresponding to a state in which the first set of cores 210 are enabled. Ex-1005, ¶55 ("As shown, the method 400A begins at step 402, where a controller included in the processor causes one or more operations to be executed by a first set of cores.").

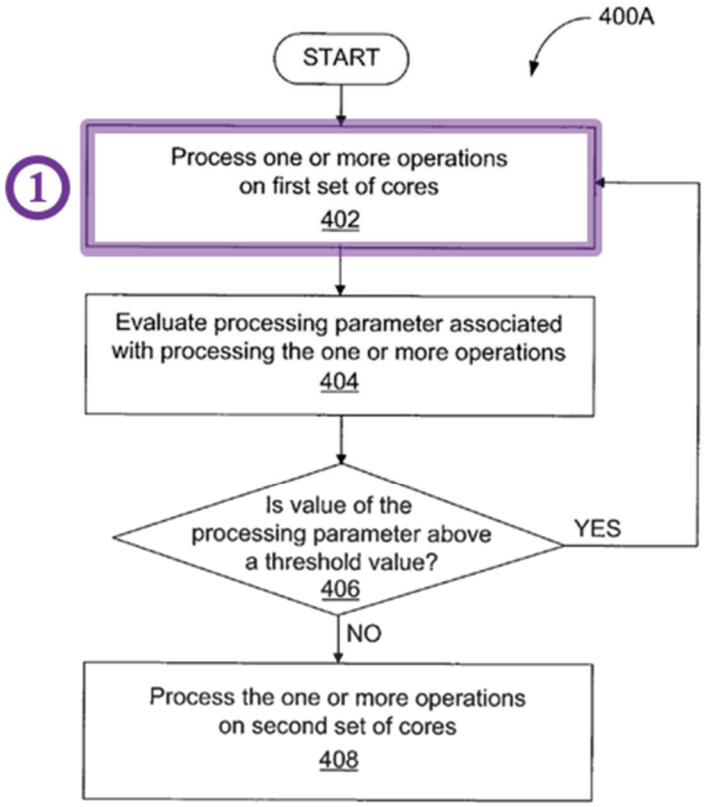



FIG. 4A

Ex-1005, Figure 4A (annotated).

196. Mathieson also teaches that, when the CPU 102 is operating in the first mode of operation, Mathieson's second set of cores may also be operating:

> In the first mode of operation, the first set of cores 210 is enabled and operable and the second set of cores 220 is disabled. In the second mode of operation, the second set of cores 220 is enabled and operable and the first set of cores 210 is disabled…. In some embodiments, ***the CPU 102 may operate in both the first mode and the second mode simultaneously***.

Ex-1005, ¶¶30-31. Thus, a POSITA would have understood that, at step 402 of method 400A, both sets of Mathieson's cores may be enabled and operating together. Indeed, a POSITA would have been motivated to enable both sets of cores to operate together at step 402, for example, when the workload to be processed by CPU 102 is large, such that the processing resources of the first set of cores 210, alone, are not sufficient to achieve the desired performance level. *See, e.g.*, Ex-1024, 0002-0003 (explaining that power can be conserved in a multicore processing system that includes both high-power and low-power cores not only "[i]n times of light load, [when] activity shifts to low-power cores," but also "in times of heavy load, [when] low-power cores can offer additional parallelism without significant increases in area or power consumption").

197. Second, Mathieson teaches a demand in the form of a "workload." Ex-1005, ¶31. For example, Mathieson explains that, while in the first mode of operation, the workload is evaluated to determine whether "***the workload*** has [] decreased to a point where the CPU 102 would use less power to operate in the second mode". *Id*.

198. Third, Mathieson teaches disabling all of the first plurality of cores (Mathieson's first set of cores 210) for a drop in demand (Mathieson's workload) below a threshold without disabling any of the second plurality of cores (Mathieson's second set of cores 220). As illustrated in Figure 4A of Mathieson

(reproduced below with annotations), step 406 (green) of method 400A involves "***evaluating the workload*** [by] determining whether a processing parameter associated with processing the one or more operations is greater than or less than ***a threshold value***." Ex-1005, ¶32, ¶56 ("At step 406, the controller determines whether a value of the processing parameter is above a threshold value.") Fig. 4A.

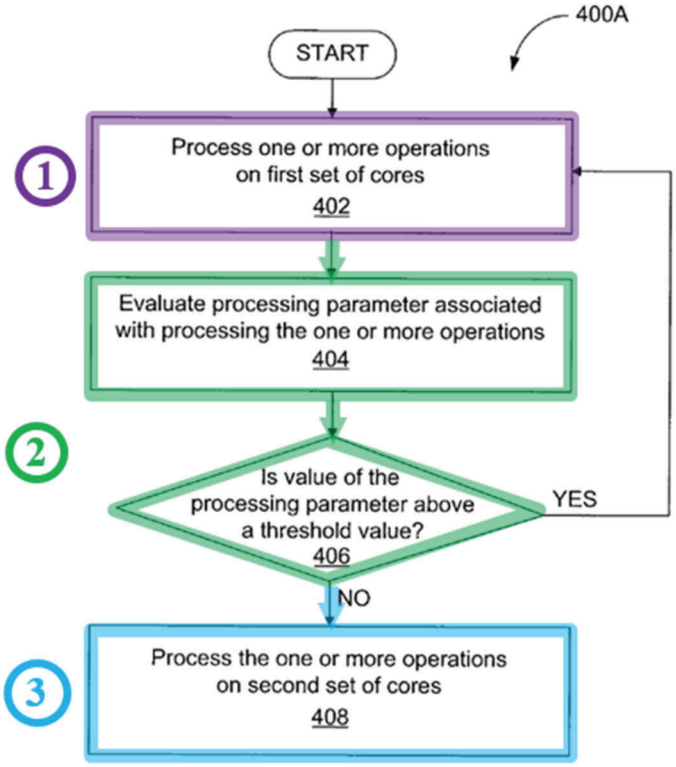



FIG. 4A

Ex-1005, Figure 4A (annotated).

199. If the workload, as determined by the processing parameter, (the claimed demand) is less than the threshold value, CPU 102 switches to the second mode of operation, corresponding to a state in which the first set of cores 210 (the

claimed first plurality of cores) are disabled and the second set of cores 220 (the claimed second plurality of cores) are enabled (step 408 (blue) of annotated Figure 4A, above). Ex-1005, ¶¶30, 57, Figure 4A. In particular, Mathieson explains that:

> At step 406, the controller determines whether a value of the processing parameter is above a threshold value. In some embodiments, determining whether the value of the processing parameter is above the threshold value is determined dynamically at regular time intervals based on the current processing operations being executed by the processor…. If the controller determines that the value of the processing parameter is not above the threshold value, then the method 400A proceeds to step 408.
>
> At step 408, the controller causes one or more operations to be executed by a second set of cores.

Ex-1005, ¶¶56-57.

200. A POSITA would have understood that the second set of cores 220 would remain enabled throughout the switch from the first mode of operation (in which both the first set of cores 210 and the second set of cores 220 are enabled) to the second mode of operation (in which only the second set of cores 220 are enabled). Not only would disabling the second set of cores 220 only to re-enable them detrimentally affect performance and power consumption, it could also potentially lead to lost data during the transition, as explained in Section VIII.A.1.d.

201. Carmack, the entire disclosure of which is incorporated by reference into Mathieson, similarly explains that, from a state in which the first plurality of cores (Carmack's second cores 120) and the second plurality of cores (Carmack's

first cores 110) are enabled, all of the first plurality of cores (Carmack's second cores 120) are disabled for a drop in demand (workload), without disabling any of the second plurality of cores (Carmack's first cores 110).

202. As illustrated in Figure 3 of Carmack (reproduced with annotations below), Carmack teaches using only the second plurality of cores (Carmack's first cores 110) when workload is within a first range, using only the first plurality of cores (Carmack's second cores 120) when workload is within a second range, and using both the first plurality of cores and the second plurality of cores when workload is within a third range, with the workload being determined "periodically." Ex-1006, ¶¶20-21, 23, Fig. 3.

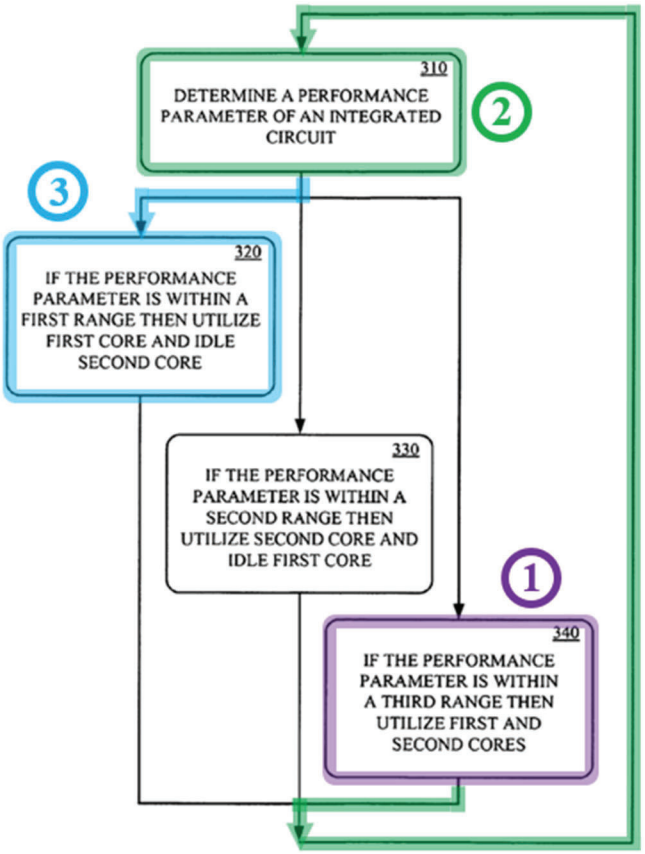



Figure 3

Ex-1006, Figure 3 (annotated).

203. In particular, Carmack explains:

> Referring now to FIG. 3, a method of dynamic operation of asymmetric cores in an integrated circuit, in accordance with another embodiment of the present technology, is shown. At 310, a performance parameter of the integrated circuit is determined. The performance parameter may be determined by the asymmetric core control circuit 130. At 320, a first core 110 of the integrated circuit is utilized and a second core 120 is idled if the performance parameter is within a first predetermined range. At 330, the second core 120 is utilized and the first core 110 is idled if the performance parameter is within a second predetermined range. At 340, both the first and second cores 110, 120 are utilized if the performance parameter is within a third

> predetermined range. The processes 310-340 may be selectively repeated a plurality of times during operation of the integrated circuit 100. In one implementation, the performance parameter is determined periodically.

Ex-1006, ¶21. While Carmack describes the method of Figure 3 in terms of a single first core 110 and a single second core 120, Carmack explains that any number of first and second cores 110, 120 may be used:

> Furthermore, embodiments of the present technology can be scaled to any number (N) of cores of varying mixes of power consumption and performance advantages. … Additionally, it may be possible to use two or more cores in parallel to achieve even higher performance.

Ex-1006, ¶20; *see also id.*, ¶23.

204. Thus, if workload (the claimed demand) is within Carmack's third range at a first measurement time, and then at the next measurement time the workload has fallen below a threshold value such that it lands within Carmack's first range, Carmack's system will transition from a state in which both the first and second cores 110 and 120 are enabled (step 340, purple) to a state in which only the first cores 110 (the claimed second plurality of cores) are enabled (step 320, blue). And, as explained above, a POSITA would have been motivated to keep the first cores 110 (the claimed second plurality of cores) enabled throughout the transition, based on power and performance considerations.

205. Finally, Mathieson also explains that disabling/enabling cores is done using hardware (the claimed power management hardware):

> In one embodiment, the controller saves the processor state to the shared resource 230, ***triggers a hardware mechanism that stops and powers off the first set of cores 210***, and boots the second set of cores 220.

Ex-1005, ¶43.

**e. Element 1[b][ii]: wherein an operating system to execute on the multi-core processor is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.**

206. Mathieson teaches Element 1[b][ii].

207. As explained for Element 1[b][i], Mathieson teaches a demand for the multi-core processor in the form of a "workload" to be processed by CPU 102. Section VIII.D.2.d; Ex-1005, ¶¶30-33.

208. Mathieson also discloses that a controller within CPU 102 (illustrated in annotated Figure 2 below) monitors the workload (the claimed demand):

> In some embodiments, a controller 240 included within the CPU 102 is configured to ***evaluate at least a workload*** associated with one or more operations to be executed by the CPU 102. In some embodiments, the controller is implemented in software and is executed by the CPU 102. Based on the evaluated workload, the controller 240 is able to configure the CPU 102 to operate in a first mode of operation or a second mode of operation.

Ex-1005, ¶30.

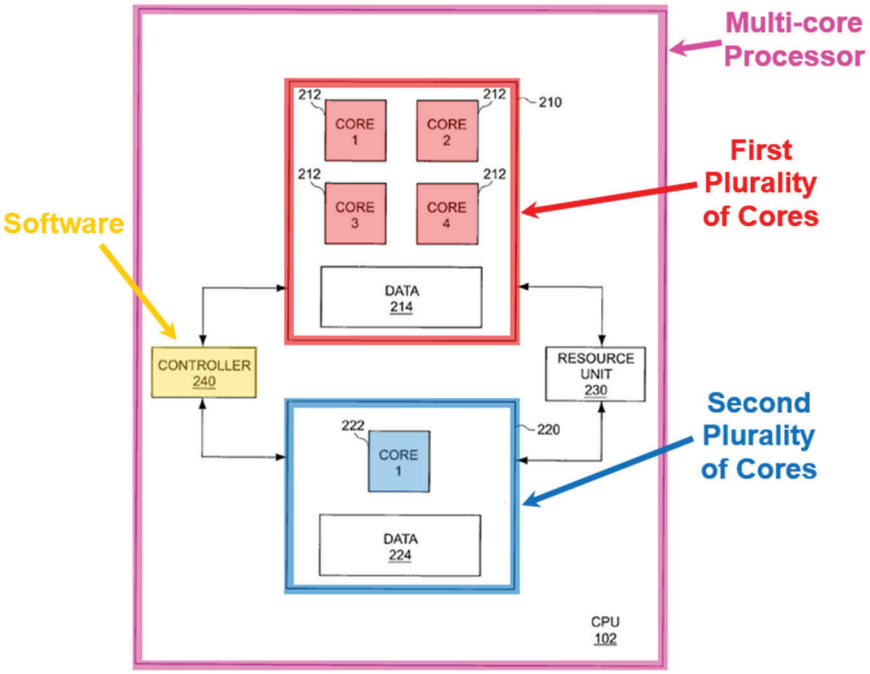



FIG. 2

Ex-1005, Figure 2 (annotated).

209. A POSITA would have viewed Mathieson's controller 240 as being part of an operating system executing on CPU 102. First, as Mathieson explains, "the controller is implemented in software and is executed by the CPU 102." Ex-1005, ¶30. Second, Mathieson's controller 240 both manages the "processing operations" executing on CPU 102, and selects the underlying hardware (e.g., cores 210 and/or 220) on which those operations are executed. Ex-1005, ¶¶35, 62. Similarly, an operating system is software, and the functions performed by an operating system include process and device management. A POSITA would have understood that an operating system, in general, includes a kernel (the lowest level of the operating system) and one or more modules that the kernel executes to control

processes and manage devices within the processor. A POSITA would have recognized that Mathieson's controller corresponds to a software module that may be included as part of an operating system and executed by a kernel. Thus, at least because a POSITA would have recognized that Mathieson's controller performs the functions of an operating system, the POSITA would have viewed controller 240 as being part of an operating system executing on CPU 102.

210. Mathieson also discloses that controller 240 (the claimed operating system) controls the power management hardware based on the demand. First, Mathieson explains that the controller causes CPU 102 to switch between operating modes based on workload (the claimed demand):

> If the controller 240 [] detects that the workload has [] decreased to a point where the CPU 102 would use less power to operate in the second mode, then ***the controller 240 causes*** the CPU 102 to operate in the second mode.

Ex-1005, ¶31;

> Based on the evaluated workload, ***the controller 240 is able to configure*** the CPU 102 to operate in [the] first mode of operation of [the] second mode of operation.

Ex-1005, ¶30.

211. Second, Mathieson explains that the controller 124 causes the CPU 102 to change operating modes by controlling Mathieson's power management hardware:

> When execution of the processing operations switches

> from the first set of cores to the second set of cores, in some embodiments, the controller 240 is configured to transfer the processor state from the first set of cores to the second set of cores. In one embodiment, ***the controller*** saves the processor state to the shared resource 230, ***triggers a hardware mechanism*** that stops and powers off the first set of cores 210, and boots the second set of cores 220.

Ex-1005, ¶43.

**3. Dependent Claim 2: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores.**

212. Mathieson and Sutardja teach Claim 2.

213. The first and second cores of Mathieson's CPU 120 are comprised of logic gates and these gates are comprised of transistors:

> [T]he power consumption associated with the CPU 102 is derived from "dynamic" switching power and "static" leakage power. The switching power loss is based on the charging and discharging of the each transistor and its associated capacitance, and increases with operating frequency and number of gates.,

Ex-1005, ¶27; *see also id.*, ¶36. Moreover, this is fundamental—a POSITA would have understood that different kinds of logic gates, e.g., NOR, NAND, XOR, are the functional building blocks of a processing core, and are formed using transistors.

214. Mathieson discloses that the transistors of the first set of cores 210 (the claimed first plurality of cores) differ from those of the second set of cores 220 (the claimed second plurality of cores):

> [T]he cores 212 included in the first set of cores 210 comprise "fast" cores and the cores 222 included in the second set of cores 220 comprise "slow" cores. For example, the cores 212 may be manufactured using ***faster transistors that have significant static leakage*** [as compared to the cores of the second set of cores 220].

Ex-1005, ¶28.

215. Mathieson teaches that the second plurality of cores (Mathieson's second set of cores 220) comprise logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores (Mathieson's first set of cores 210). A POSITA would have understood that only a small number of design choices affect transistor speed and/or static leakage. A well-known design choice affecting both of these properties is the width of the transistor. *See, e.g.*, Ex-1029, 0015 ("Leakage grows with the aspect ratio [(width/length)] of a transistor."); Ex-1030, 0007 (explaining that "reducing the width [of a transistor] increases the transistor's delay"). In particular, a POSITA would have understood that increasing the width of a transistor would increase both the speed of the transistor and the static leakage of the transistor. *Id*.

216. Thus, a POSITA would have been motivated to make predictable use of transistors with greater widths for Mathieson's first set of cores 210 (the claimed first plurality of cores), as compared with Mathieson's second set of cores 220 (the claimed second plurality of cores), to achieve the properties disclosed by

Mathieson—namely, that the transistors in the first set of cores 210 are faster than the transistors in the second set of cores 220. A POSITA also would have recognized that using transistors with greater widths for Mathieson's first set of cores 210 (the claimed first plurality of cores) would result in Mathieson's first set of cores exhibiting significant static leakage, just as Mathieson discloses. A POSITA would have been particularly motivated to make such predictable use of transistors with greater widths for Mathieson's first set of core 210, because transistor width is one of a finite number of identified factors that affect both transistor speed and static leakage. Thus, a POSITA would have been motivated to use narrower transistors in the second set of cores 220 (the claimed second plurality of cores) that in the first set of cores 210 (the claimed first plurality of cores).

217. To the extent that a POSITA would have sought additional implementation details about the transistors to include in Mathieson's cores, including the widths of the transistors to use in each type of core, this features is also taught by Sutardja. As explained for Ground 1, Sutardja teaches Claim 2. *See* Section VIII.A.2.

218. A POSITA seeking to follow Mathieson's teaching that a first set of cores 210 (the claimed first plurality of cores) include "faster transistors that have significant static leakage," as compared to a second set of cores 220 (the claimed

second plurality of cores),would have been motivated to search for additional references that disclose transistor design details that would achieve such differences.

219. Sutardja is one such reference. Sutardja discloses transistors with the same properties as those of Mathieson's transistors:

> The HP core 24 may be fabricated using a ***high leakage/high speed*** process. The LP core 28 may be fabricated using a ***low leakage/low speed*** process. The HP and LP cores 24 and 28 include transistors 106 and 110, respectively. … In some implementations, ***the transistors 106 may have higher leakage current than the transistors 110***.

Ex-1007, ¶230. Sutardja further discloses that these differences may result from differences in the widths of the constituent transistors: "Gates of the transistors 106 may be larger than gates of the transistors 110." Ex-1007, ¶230.

220. Accordingly, a POSITA would have been motivated to construct the transistors in Mathieson's second set of cores (the claimed second plurality of cores) with narrower gates than the transistors in Mathieson's first set of cores (the claimed first plurality of cores), as taught by Sutardja, to achieve the properties that Mathieson teaches are exhibited by those transistors. Moreover, as explained in Section YY, this feature would have been implemented by applying the combined Mathieson/Sutardja teachings to use the same instruction set and logic design in the two different sets of cores. A POSITA would have had a reasonable expectation of

success in doing so, because Sutardja's transistors exhibit the same speed and leakage properties as Mathieson's transistors.

**4. Dependent Claim 3: The multi-core processor of claim 1, wherein the second plurality of cores comprise logic gates that consume less power than corresponding logic gates of the first plurality of cores.**

221. Mathieson and Sutardja teach Claim 3.

222. As explained for Claim 2, the cores in Mathieson are comprised of logic gates and these gates are comprised of transistors. Ex-1005, ¶¶27, 36. Moreover, a POSITA would have understood that logic gates in a processing core are formed from transistors.

223. Mathieson explains that power consumed by CPU 102 is consumed by its constituent transistors:

> [T]he power consumption associated with the CPU 102 is derived from "dynamic" switching power and "static" leakage power. The switching power loss is based on the charging and discharging of the each transistor and its associated capacitance. … The leakage power loss is based on gate and channel leakage in each transistor….

Ex-1005, ¶27.

224. Mathieson further explains that "at slower clock speeds, the static leakage of the fast transistors [of the first set of cores] can dominate the overall power consumption." Ex-1005, ¶28; *see also id.*, ¶73 ("[T]he lower total power associated with operating the processing complex at lower frequencies using the

slow cores is based on the lower leakage power associated with the slow cores."). Thus, a POSITA would have understood that at lower operating frequencies, the transistors (and the logic gates formed therefrom) of the second set of cores 220 (the claimed second plurality of cores) consume less power than the transistors (and the logic gates formed therefrom) of the first set of cores 210 (the claimed first plurality of cores), because they exhibit lower static leakage.

225. To the extent that a POSITA would have sought additional implementation details about logic gate power consumption, this features is taught by Sutardja. As explained for Ground 1, Sutardja teaches Claim 3. *See* Section VIII.A.3. And, as explained for Claim 2, a POSITA would have been motivated to use Sutardja's transistors in Mathieson's cores. *See* Section VIII.D.2.

**5. Dependent Claim 4: The multi-core processor of claim 1, wherein the second plurality of cores each have a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores.**

226. Mathieson teaches Claim 4.

227. As illustrated in Figure 6 (reproduced with annotations below), Mathieson's second set of cores (the claimed second plurality of cores) have a maximum frequency (labelled $F_2$) that is lower than the maximum frequency (labelled $F_1$) of Mathieson's first set of cores (the claimed first plurality of cores). Ex-1005, Figure 6.

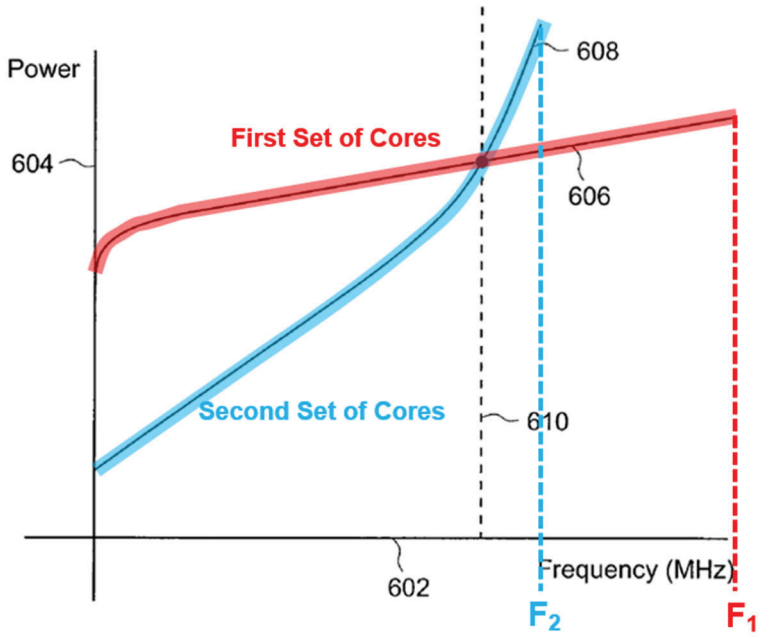



FIG. 6

Ex-1005, Figure 6 (annotated).

228. Although Mathieson does not specify the maximum frequency of the first or second set of cores, it would have been readily apparent to a POSITA that the power usage of the second set of cores 608 becomes exponential after the frequency at line 610 and that, therefore, its maximum frequency is necessarily less than that of the first set of cores for which power usage is only gradually increasing in that range.

229. Carmack, incorporated by reference into Mathieson, similarly discloses that the maximum frequency of the second plurality of cores (Carmack's first cores

110) is less than the maximum frequency of the first plurality of cores (Carmack's second cores 120):

> [T]he system can lower the peak frequency of the faster core 120 to match the maximum frequency of the slower core 110.

Ex-1006, ¶22.

**6. Dependent Claim 7: The multi-core processor of claim 1, wherein the first plurality of cores are at a maximum operating frequency in the state.**

230. Mathieson teaches Claim 7.

231. First, Mathieson teaches that "operating in both the first and second modes simultaneously may result in lower overall power efficiency." Ex-1005, ¶31. Second, Mathieson teaches that the operating frequency of the first set of cores (the claimed first plurality of cores) may be adjusted based on the size of the workload. Ex-1005, ¶31 ("For example, if the CPU 102 is operating in the first mode at high frequency, and the controller 240 detects that the workload has decreased to a point where operating in the first mode at lower frequency would save power, then the controller 240 may decrease the operating frequency of the first set of cores 210."). While Mathieson discloses decreasing operating frequency in response to falling workloads, a POSITA would have understood that Mathieson's system could equivalently increase operating frequency in response to increasing workloads.

232. Based on the above two teachings, a POSITA would have recognized that, when only Mathieson's first set of cores are enabled, it would be more power-efficient to accommodate increasing workloads by increasing the operating frequency of the first set of cores, rather than by enabling Mathieson's second set of cores and operating both sets of cores together. Thus, a POSITA seeking to implement Mathieson's CPU 102 in a power-efficient manner, would have designed the CPU such that, from a state in which only the first set of cores are enabled, when faced with increasing workloads, the operating frequency of the first set of cores would first be increased to its maximum value, and then, only after and the workload had continued to increase beyond the capacity of the first set of cores operating at their maximum frequency, would the second set of cores be enabled for operation together with the first set of cores (the claimed state).

233. Indeed, Carmack, the entire disclosure of which is incorporated by reference into Mathieson, Ex-1005, ¶1, discloses that when both the first and second cores 110, 120 are operating together, the second cores 120 (the claimed first plurality of cores) may be operating at their peak (i.e., maximum) frequency:

> When both cores 110, 120 are utilized, the second core 120 would be given a fraction of the total work scaled to its performance advantage over the first core 110. For situations where the overhead of coordinating asymmetric cores becomes too high, the system can lower the peak frequency of the faster core 120 to match the maximum frequency of the slower core 110….

Ex-1006, ¶22. Because Carmack explains that the frequency of the faster core 120 is reduced from its peak value when "the overhead of coordinating asymmetric cores becomes too high," a POSITA would have understood that for situations where the overhead of coordinating is ***not*** too high, Carmack's faster core 120 would be operating at its peak (i.e., maximum) frequency.

**7. Dependent Claim 8: The multi-core processor of claim 1, wherein**

**a. Element 8[a]: the power management hardware is to enable all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores,**

234. Mathieson teaches Element 8[a].

235. First, Mathieson discloses a state where the second plurality of cores (Mathieson's second set of cores 220) are enabled, and the first plurality of cores (Mathieson's first set of cores 210) are disabled, as explained above, for Element 1[b][i]. *See* Section VIII.D.2.d. As illustrated in Figure 4A (annotated below), in step 408 (blue) of method 400A, CPU 102 is in the second mode of operation, corresponding to a state in which the second set of cores 220 are enabled and the first set of cores 210 are disabled. Ex-1005, ¶30 ("In the second mode of operation, the second set of cores 220 is enabled and operable and the first set of cores 210 is disabled), ¶57 ("At step 408, the controller causes one or more operations to be execute by a second set of cores."), Fig. 4A.

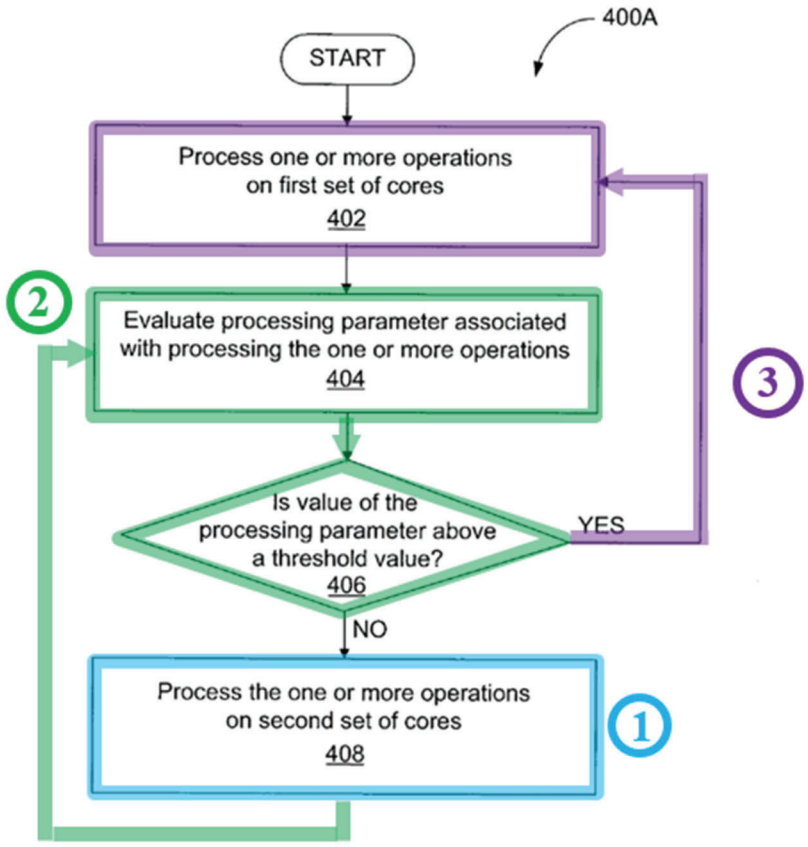



FIG. 4A

Ex-1005, Figure 4A (annotated).

236. Second, as explained for Element 1[b][i], Mathieson teaches (1) the claimed demand in the form of a "workload" that is associated with "a processing parameter," and (2) identifying the claimed increase in demand above the threshold by "determining whether [the] processing parameter … is greater than … a threshold value." *See* Section VIII.D.2.d; Ex-1005, ¶¶31-32.

237. Third, Mathieson teaches enabling all of the first plurality of cores (Mathieson's first set of cores 210) for the increase in demand (workload) above the

threshold (Mathieson's threshold value), without disabling any of the second plurality of cores (Mathieson's second set of cores 220). Notably, while Figure 4A (reproduced above) illustrates method 400A terminating at step 408, a POSITA would have understood that method 400A would repeat during the operation of Mathieson's multi-core processor. Indeed, Mathieson discloses that step 406 (green), during which "the controller determines whether a value of the processing parameter is above a threshold value," occurs "dynamically at regular time intervals." Ex-1005, ¶56. Accordingly, from a state in which only the second set of cores 220 (the claimed second plurality of cores) are enabled, a POSITA would understand that method 400A would return to steps 404 and 406 (green) where the workload, as determined by the processing parameter, (the claimed demand) would again be evaluated and compared against the threshold value (the claimed threshold). Ex-1005, Fig. 4A. If the workload (the claimed demand) is greater than the threshold value (the claimed threshold), CPU 102 switches to the first mode of operation, corresponding to a state in which the first set of cores 210 (the claimed first plurality of cores) are enabled (step 402 (purple) of annotated Figure 4A, above). Ex-1005, ¶56, Fig. 4A.

238. As explained for Element 1[b][i], based on Mathieson's disclosure that "the CPU 102 may operate in both the first mode and the second mode simultaneously," a POSITA would have understood that both sets of cores 210 and

220 may be enabled and operating together at step 402 (purple), or at least would have been motivated to do so. Ex-1005, ¶31; Section VIII.D.2.d; *see also, e.g.*, Ex-1024, 0002-0003 (explaining that power can be conserved in a multicore processing system that includes both high-power and low-power cores not only "[i]n times of light load, [when] activity shifts to low-power cores," but also "in times of heavy load, [when] low-power cores can offer additional parallelism without significant increases in area or power consumption").

239. Furthermore, a POSITA would have understood that the second set of cores 220 (the second plurality of cores) would remain enabled throughout the transition from the second mode of operation (in which only the second set of cores 220 are enabled), to the first mode of operation (in which both the first and second sets of cores 210, 220 are enabled), because disabling the second set of cores 220 only to re-enable them would detrimentally affect performance and power consumption, and could potentially lead to lost data during the transition.

240. Carmack, the entire disclosure of which is incorporated by reference into Mathieson, Ex-1005, ¶1, similarly illustrates that, from a state in which the second plurality of cores (Carmack's first cores 110) are operating, the first plurality of cores (Carmack's second cores 120) are enabled, for an increase in demand (workload), without disabling any of the second plurality of cores (Carmack's first cores 110). In particular, as illustrated in Figure 3 of Carmack (annotated below), if

workload (the claimed demand) is within Carmack's first range at a first measurement time, and then at the next measurement time the workload has increased above a threshold value such that it lands within Carmack's third range, Carmack's system will transition from a state in which only the first cores 110 (the claimed second plurality of cores) are enabled (step 320, blue) to a state in which both the first and second cores 110 and 120 are enabled (step 340, purple). Ex-1006, ¶¶13, 21, Fig. 3.

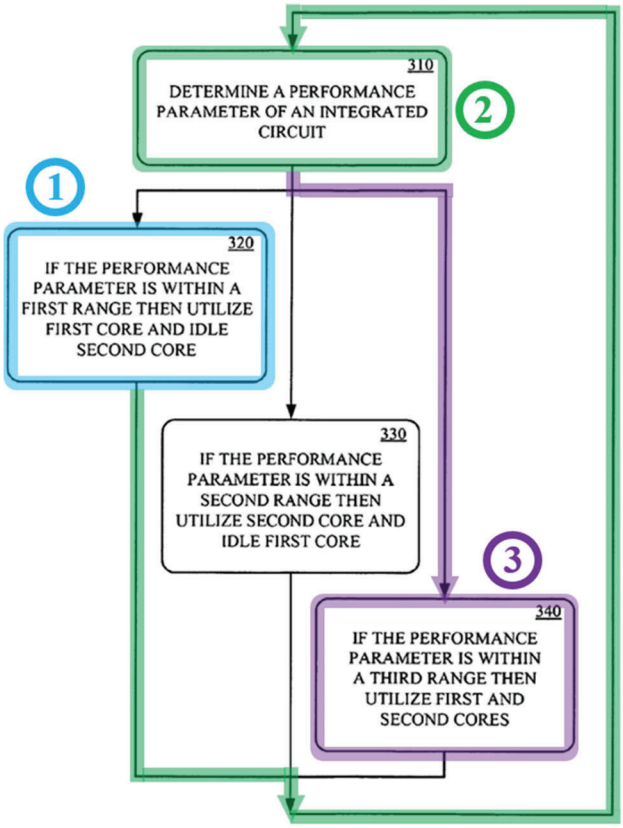


Figure 3

Ex-1006, Figure 3 (annotated).

241. And, as explained above, a POSITA would have been motivated to keep the first cores 110 (the claimed second plurality of cores) enabled through this transition, based on power and performance considerations, as well as to help ensure that data is not lost during the transition. Section VIII.A.1.d.

242. Finally, as explained for Element 1[b][i], Mathieson also discloses that disabling/enabling cores is done using hardware (the claimed power management hardware). Section VIII.D.2.d; Ex-1005, ¶43.

**b. Element 8[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.**

243. Mathieson teaches Element 9[b], for the reasons described above for Element 1[b][ii]. Section VIII.D.2.e.

**8. Independent Claims 9 and 17:**

**a. Element 9[preamble]: A method comprising:**

244. Even if the preamble is limiting, Mathieson teaches a method. For example, Figure 4A of Mathieson presents "method steps for switching between modes of operation of a processing complex." Ex-1005, ¶54, Cls. 1-11.

**b. Element 17[preamble]: A non-transitory machine readable medium containing program code that when processed by a machine causes a method to be performed, the method comprising:**

245. Even if the preamble is limiting Mathieson teaches a non-transitory machine readable medium containing program code. For example, Mathieson discloses "[a] computer-readable medium including instructions that, when executed, cause a processing complex to perform the steps of …." Ex-1005, Cl. 12.

**c. Elements 9[a][i] and 17[a][i]: operating a multi-core processor such that a first plurality of cores and a second plurality of cores execute a same instruction set,**

246. Mathieson teaches Elements 9[a][i] and 17[a][i] for the same reasons as explained above for Element 1[a][i]. *See* Section VIII.D.2.b.

**d. Elements 9[a][ii] and 17[a][ii]: wherein the second plurality of cores consume less power, for a same applied operating frequency and supply voltage, than the first plurality of cores; and**

247. Mathieson teaches Elements 9[a][ii] and 17[a][ii] for the same reasons as explained above for Element 1[a][ii]. *See* Section VIII.D.2.c.

**e. Elements 9[b][i] and 17[b][i]: disabling with power management hardware, from a state where the first plurality of cores and the second plurality of cores are enabled, all of the first plurality of cores for a drop in demand below a threshold without disabling any of the second plurality of cores,**

248. Mathieson teaches Elements 9[b][i] and 17[b][i] for the same reasons as explained above for Element 1[b][i]. *See* Section VIII.D.2.d.

**f. Element 9[b][ii] and 17[b][ii]: wherein an operating system executing on the multi-core processor monitors a demand for the multi-core processor and controls the power management hardware based on the demand.**

249. Mathieson teaches Elements 9[b][ii] and 17[b][ii] for the same reasons as explained above for Element 1[b][ii]. *See* Section VIII.D.2.e.

**9. Dependent Claims 10 and 18: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating of the second plurality of cores comprises driving logic gates that have narrower logic gate driver transistors than corresponding logic gates of the first plurality of cores.**

250. Mathieson teaches Claims 10 and 18 for the same reasons as explained above for claim 2. *See* Section VIII.D.3.

**10. Dependent Claims 11 and 19: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating of the second plurality of cores comprises driving logic gates that consume less power than corresponding logic gates of the first plurality of cores.**

251. Mathieson teaches Claims 11 and 19 for the same reasons as explained above for claim 3. *See* Section VIII.D.4.

**11. Dependent Claims 12 and 20: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the second plurality of cores at a maximum operating frequency that is less than a maximum operating frequency of the first plurality of cores.**

252. Mathieson teaches Claims 12 and 20 for the same reasons as explained above for claim 4. *See* Section VIII.D.5.

**12. Dependent Claims 15 and 23: The [method of claim 9/non-transitory machine readable medium of claim 17], wherein the operating comprises operating the first plurality of cores at a maximum operating frequency in the state.**

253. Mathieson teaches Claims 15 and 23 for the same reasons as explained above for claim 7. *See* Section VIII.D.6.

**13. Dependent Claims 16 and 24: The [method of claim 9/non-transitory machine readable medium of claim 17], further comprising**

**a. Elements 16[a] and 24[a]: enabling, with the power management hardware, all of the first plurality of cores for an increase in demand above the threshold without disabling any of the second plurality of cores,**

254. Mathieson teaches Elements 16[a] and 24[a] for the same reasons as explained above for Element 8[a]. *See* Section VIII.D.7.a.

**b. Elements 16[b] and 24[b]: wherein an operating system is to monitor a demand for the multi-core processor and control the power management hardware based on the demand.**

255. Mathieson teaches Element 16[b] for the same reasons as explained above for Element 8[b]. *See* Section VIII.D.7.b.

### E. Ground 5: Claims 5-6, 13-14, and 21-22 are rendered obvious by Mathieson/Sutardja in view of Rychlik

#### 1. Dependent Claims 5, 13, and 21:

##### a. Elements 5[a], 13[a], 21[a]: The [multi-core processor of claim 1/method of claim 9/non-transitory machine readable medium of claim 17], [wherein the power management hardware is to disable/further comprising disabling, with the power management hardware,] an additional core of the second plurality of cores for each continued drop in demand below a next lower threshold until one core of the second plurality of cores remains enabled, and

256. Mathieson/Sutardja in view of Rychlik teaches Elements 5[a], 13[a], and 21[a].

257. Mathieson discloses that, as part of the power management techniques implemented by Mathieson's multi-core processor, "the number of active cores [may be] increased or decreased.". Ex-1005, ¶76. Based on this teaching, a POSITA would have been motivated to search for references that disclose implementation details relating to enabling/disabling individual cores for power management purposes.

258. Rychlik is one such reference. Just like Mathieson, Rychlik discloses power management techniques in multi-core systems. Ex-1009, Abstract. As explained for Ground 2, Rychlik teaches Elements 5[a], 13[a], and 21[a]. *See* Section VIII.B.1.a.

259. A POSITA would have been motivated to disable an additional core in Mathieson's second set of cores, whenever the workload to be processed by Mathieson's CPU 102 dropped below a next lower threshold, until only one core remained enabled, as taught by Rychlik, because doing so would prevent Mathieson's second set of cores from "wasting power." Ex-1009, ¶88;.

260. A POSITA would have had a reasonable expectation of success in implementing Rychlik's power management technique in Mathieson's multi-core system, because it amounts to use of a known technique (disabling cores based on demand, as taught by Rychlik) to improve a similar system (Mathieson's multi-core system) in the same way (by reducing power consumption during periods of low demand).

**b. Elements 5[b], 13[b], 21[b]: [lower/lowering] an operating frequency or a supply voltage of the one core of the second plurality of cores as demand drops below a next lower threshold.**

261. Mathieson/Sutardja in view of Rychlik teaches Elements 5[b], 13[b], and 21[b].

262. As explained for Ground 2, Rychlik teaches Elements 5[b], 13[b], and 21[b]. *See* Section VIII.B.1.b. And, as explained above for Elements 5[a], 13[a], and 21[a], a POSITA would have been motivated to implement Rychlik's power management technique in Mathieson's multi-core processor. Section VIII.E.1.

**2. Dependent Claims 6, 14, 22: [The multi-core processor of claim 5/method of claim 13/The non-transitory machine readable medium of claim 21], [wherein the power management hardware is to raise/further comprising raising, with the power management hardware,] a supply voltage or an operating frequency of said one core in response to higher demand.**

263. Mathieson/Sutardja in view of Rychlik teaches Claims 6, 14, and 22. As explained for Ground 2, Rychlik teaches the added limitations of these claims. *See* Section VIII.B.2. And, as explained above for Elements 5[a], 13[a], and 21[a], a POSITA would have been motivated to implement Rychlik's power management technique in Sutardja's multi-core processor. Section VIII.E.1.

## IX. CONCLUSION

264. For the reasons set forth above, I believe that Claims 1-24 of the '080 Patent are unpatentable in view of the prior art.

265. In signing this declaration, I recognize that it will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in the case and that cross-examination will take place in the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

266. I reserve the right to supplement my opinions in the future to respond to any arguments that the Patent Owner raises and to take into account new information as it becomes available to me.

267. I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Respectfully Submitted

/Trevor Mudge/

Trevor Mudge

Date: February 14, 2023