UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

SINGULAR COMPUTING LLC,
Patent Owner.

Case No. IPR2023-00395
Patent No. 10,754,616

DECLARATION OF PAUL FRANZON, PH.D.

Google Exhibit 1003
Google v Singular

TABLE OF CONTENTS

I. PERSONAL AND PROFESSIONAL BACKGROUND 1
II. MATERIALS REVIEWED AND CONSIDERED 6
III. MY UNDERSTANDING OF PATENT LAW 14
A. Anticipation 16
B. Obviousness 16
IV. PERSON OF ORDINARY SKILL IN THE ART ("POSA") 19
V. THE '616 PATENT 21
A. Specification of the '616 Patent 21
B. Prosecution History of the '616 Patent 24
C. Priority Dates 25
D. The Challenged Claims 27
1. Reduced Bitwidth Floating-Point Numbers 29
2. Parallel Processing Elements and Applications 34
VI. PRIOR-ART REFERENCES 39
A. Public Accessibility of Non-patent Prior Art Exhibits 39
B. Stuttard (EX1006) 41
C. Shirazi (EX1036) 41
VII. GROUND 1A: CLAIMS 1-5 ARE OBVIOUS OVER STUTTARD IN COMBINATION WITH SHIRAZI 42
A. The Stuttard-Shirazi Processor 42
1. Stuttard (EX1006) 42
2. Shirazi (EX1036) 45
3. Stuttard-Shirazi Combination 48
B. Claim 1 63
1. [1PRE]: "A computing system, comprising:" 63
2. [1A]: "a host computer;" 64
3. [1B]: "a computing chip comprising:" 66

4. [1C]: "a processing element array comprising" "first" and "second" "edge processing element[s] positioned at a first edge of the processing element array" and "first" and "second" "interior processing element[s] positioned at [first and second] location[s] in the interior of the processing element array".....66
5. [1D]: "a first processing element connection connecting the first edge processing element with the first interior processing element; a second processing element connection connecting the second edge processing element with the second interior processing element;" 69
6. [1E]: "an input-output unit connected to the first" and "second edge processing element[s];" 71
7. [1F]: Each Processing Element of Limitation [1C] Has a "memory local to" It.....80
8. [1G]: Each Processing Element of Limitation [1C] Comprises an "arithmetic unit" 81
9. [1H]: "a host connection at least partially connecting the input-output unit with the host computer;".....82
10. [1I]: "wherein the first, second, third and fourth arithmetic units each comprises a corresponding multiplier circuit adapted to receive as a first input to the corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits;".....85

C. Claim 2 89

1. [2A]: "the computing chip further comprises a control unit connected to the processing element array,".....89
2. [2B]: "the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and" 90
3. [2C]: "to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction." 92

D. Claim 3: "the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory." 93

E. Claim 4: "said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image." 96

F. Claim 5: "the width of each [first/second] binary mantissa is no more than 10 bits plus a [first/second] sign bit" 103

VIII. GROUND 1B: CLAIMS 7-14 ARE OBVIOUS OVER STUTTARD IN COMBINATION WITH SHIRAZI 104

A. Scaling the Stuttard-Shirazi Processor 104

1. 5,000+ Processing Elements 104

2. Obvious Ways to Implement the Scaled Stuttard-Shirazi Processor 109

3. Reasonable expectation of success 118

B. Claim 7 128

1. [7PRE] "A computing system, comprising:" [7A] "a host computer;" [7B] "a computing chip comprising:" 128

2. [7C]: "a processing element array comprising a plurality of first processing elements, wherein the plurality of first processing elements is no less than 5000 in number, wherein each of a first subset of the plurality of first processing elements is positioned at a first edge of the processing element array, and wherein each of a second subset of the plurality of first processing elements is positioned in the interior of the processing element array;" 129

3. [7D]: "an input-output unit connected to each of the first subset of the plurality of first processing elements;" 131

4. [7E]: "a plurality of processing element connections, each processing element connection connecting one of the plurality of first processing elements with another of the plurality of first

processing elements, wherein each of the plurality of first processing elements is connected to at least one other of the plurality of first processing elements by at least one of the plurality of processing element connections;” 132

5. [7F]: “a plurality of memory units, wherein each of the plurality of first processing elements is associated with a corresponding one of the plurality of memory units, and wherein each of the plurality of memory units is local to its associated one of the plurality of first processing elements;”.....134
6. [7G]: “a plurality of first arithmetic units, wherein each of the plurality of first processing elements has positioned therein at least one of the plurality of first arithmetic units;” 135
7. [7H]: “a host connection at least partially connecting the input-output unit with the host computer;”.....135
8. [7I]: “wherein the plurality of arithmetic units each comprises a first corresponding multiplier circuit adapted to receive as a first input to the first corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the first corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits;”.....136

C. Claim 8 137

1. [8A]: “the computing chip further comprises a plurality of second processing elements, wherein the plurality of second processing elements each comprises a second corresponding multiplier circuit adapted to receive as inputs to the second corresponding multiplier circuit two floating point values each of width at least 32 bits;”.....137
2. [8B]: “wherein, other than the plurality of second processing elements, the computing chip has no other processing element that comprises a multiplier circuit adapted to receive as inputs to the multiplier circuit two floating point values each of width at least 32 bits; and” 140

3. [8C]: "wherein the plurality of first processing elements is greater in number, by at least 100, than the plurality of second processing elements." 141

D. Claim 9: "the plurality of memory units each comprises a corresponding register." 141

E. Claim 10: "said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image." 143

F. Claim 11: "an aggregate maximum bandwidth of the plurality of processing element connections exceeds a maximum bandwidth of the host connection." 143

G. Claim 12: "the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction." 145

H. Claim 13: "the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory." 147

I. Claim 14: "the width of each [first/second] binary mantissa is no more than 10 bits plus a [first/second] sign bit." 148

IX. GROUND 1C: CLAIMS 16-22, 24 ARE OBVIOUS OVER STUTTARD IN COMBINATION WITH SHIRAZI 148

A. Two Connected Stuttard-Shirazi Chips on a Circuit Board 148

1. [16PRE], [16A]-[16H], [16M] 153

2. [16I], [16K]-[16M]: "an external edge processing element external to the computing chip" comprising "a fifth arithmetic unit" comprising a "multiplier circuit adapted to receive" as inputs a first and second floating point value each having a "binary mantissa of width no more than 11 bits" and a "binary

exponent of width at least 6 bits"; and "a fifth memory local to the external edge processing element" 157

3. [16J] "an external input-output unit external to the computing chip;" 158

4. [16N] "wherein the first edge processing element is connected with the external edge processing element through the input-output unit and the external input-output unit." 159

B. Claim 17: "the first edge processing element is connected with the external edge processing element using a torus interconnect." 160

C. Claim 18 165

1. [18A]: "the processing element array further comprises a plurality of first additional processing elements and a plurality of first additional arithmetic units, each of the plurality of first additional processing elements having positioned therein at least one of the plurality of first additional arithmetic units;" 165

2. [18B]: "wherein the plurality of first additional arithmetic units each comprises a second corresponding multiplier circuit adapted to receive" As Third and Fourth Inputs Floating-Point Values of the Same Bitwidths As Recited in [1I] 166

3. [18C]: "wherein the computing chip further comprises a plurality of second additional processing elements" Having the Same Limitations As Limitation [8A]'s "second processing elements" 166

4. [18D]-[18E] 168

D. Claims 19-22 169

E. Claim 24: "The computing system of claim 18, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit, wherein the width of each third binary mantissa is no more than 10 bits plus a third sign bit, and wherein the width of each fourth binary mantissa is no more than 10 bits plus a fourth sign bit." 171

X. CONCLUSION 172

XI. CLAIM LISTING 173

I, Paul Franzon, declare:

1. I have been retained by Wolf, Greenfield & Sacks, P.C., counsel for Petitioner Google LLC, to assess claims 1-5, 7-14, 16-22 and 24 (the "challenged claims") of U.S. Patent No. 10,754,616 ("the '616 patent"). I am being compensated for my time at my standard rate of $500.00 per hour, plus actual expenses. My compensation is not dependent in any way upon the outcome of the *inter partes* review of the '616 patent.

I. PERSONAL AND PROFESSIONAL BACKGROUND

2. I completed my Ph.D. in Electrical and Electronic Engineering in 1989 from the University of Adelaide in Australia. I obtained two additional degrees from the University of Adelaide: a Bachelor of Engineering in Electrical and Electronic Engineering in 1984 and a Bachelor of Science in Physics and Mathematics in 1983.

3. I am currently the Cirrus Logic Distinguished Professor and Director of Graduate Programs in the Department of Electrical and Computer Engineering at North Carolina State University ("NCSU") in Raleigh, North Carolina. I have been affiliated with NCSU in various roles since 1989.

4. I am a Fellow of the IEEE. I have over 30 years' experience designing computing devices including microprocessors, parallel processors, ASICs (Application-specific integrated circuits), programmable logic devices,

computer interfaces and peripherals. I also have over 30 years of experience in the design and fabrication of processors.

5. From 1985 to 1989, I was doing the research for my PhD at the University of Adelaide and AT&T Bell Laboratories. The title of my PhD thesis was “Fault Tolerance in VLSI”. My thesis discussed 3-D chip technologies, including a flip-chip device comprised of a hybrid wafer scale integration (or HWSI). At the University, I contributed to the design of a programmable parallel signal processing chip, the “Transform and Filter Brick” and analyzed the design for defect tolerance through reconfigurability. While I was a consultant at AT&T Bell Laboratories I designed a reconfigurable array processor of signal processing filter elements. Several publications arose in whole or in part from my PhD. In my PhD thesis I addressed multiple integration issues and technologies, including 3D integration using through silicon vias, and reviewed multiple processor types being designed by other groups leveraging wafer scale integration and 3D technologies.

6. I joined North Carolina State University in 1989. From 1989 to 1994 I taught the graduate course “ECE 521 Computer Design and Technology. This covered the architecture and design of microprocessors, including pipelined designs and cache designs. In 1995, I taught ECE 218, Computer Organization and Microprocessors, which covered the architecture and programing of the M68000 microprocessor. From 1996 to 2011 I taught “Design of Complex Digital

Systems." That was an undergraduate course (renumbered several times) covering the design of non-pipelined microprocessors, including arithmetic and logic units. Since 1997 I have been teaching ASIC design. This course covers advanced digital design techniques using Verilog RTL and synthesis tools. It includes teaching modules on arithmetic design including adders and multipliers.

7. Over 1990 – 2000 I was the Principal Investigator of several projects investigating the design of processors and led or contributed to the design of several processors, many of which were built. These included a disaggregated superscalar CPU, a heterogeneous multiprocessor, a video processor, an encryption engine, and the "Dynamically Programmable Cache."

8. Over (roughly) 1993 to 1997 I was the principal investigator in a DARPA funded project entitled "Methodology, Tools and Demonstration of MCM System Optimization." As part of that project, we designed a disaggregated superscalar RISC processor. We also designed and built an encryption processor.

9. In 1995 and 1996 I was a consultant at DCT Inc. was responsible for designing video processors that were compatible with various standards including H.261, MPEG2 and H.263.

10. Over (roughly) 1997 to 1999 I had a joint project with HRL Labs entitled "Dynamically Programmable Cache" (DPC). In this project we designed a hybrid FPGA/cache. In the DPC we merged cache with FPGA, placing a row of

FPGA elements between multiple rows of cache memory and repeating. Together with the CPU that the cache served it functioned as a heterogeneous multiprocessor.

11. Since 2000, I have been heavily involved in the design and fabrication of 3D processors integrated using through silicon vias and other technologies. These included a synthetic aperture radar processor, a 3D memory, an FFT processor, a heterogeneous microprocessor stack, that included a cache hierarchy. Recently, my group designed and fabricated a multiprocessor chiplet set intended for integration on top of a silicon interposer substrate. In that microprocessor chip chiplet set, we redesigned a RISC V core so that extension instructions could be defined and run on other chiplets, acting as co-processors. Some work also continued on the DPC.

12. Over (roughly) 2001 to 2003 I was the NCSU PI of a project entitled "network processor design for just-in-time our optical networking protocols." As part of that project, we build a network processor using FPGAs. Part of this project continued from 2003 to 2005, under DARPA sponsorship, and with a focus on network security.

13. Over (roughly) 2011 to 2017 I was the PI of an Intel funded project entitled "modular, testable, tightly coupled 3D implementation of a heterogeneous processor." As part of that project, we built a two-course stack of heterogeneous

micro processes using advanced 3D integration technologies. This project included the design investigation of the 3D x86 heterogeneous core stack.

14. Over (roughly) 2011 to 2016 I was the PI of the DARPA project entitled "3D enabled customized embedded computer." As part of that project, we designed and built a SIMD processor aimed at supporting a wide range of applications with a focus on low power consumption.

15. Since 2017 I led other projects focused on 2.5 D and 3D processor design. As part of the DARPA funded project entitled "2.5 D extendable processor" my team designed and built a disaggregated multiple chip processor that included a RISC-V CPU tightly coupled with instruction controlled array co-processors aimed at accelerating machine learning tasks. We also designed a 3D programmable processor aimed at machine learning acceleration using an array of the SIMD processes designed in an earlier project.

16. Over my career I have built and tested many integrated circuits ranging using processes from at least the following semiconductor processing nodes: 5 µm, 3 µm, 250 nm, 180 nm, 130 nm, 55 nm, 45 nm, 32 nm, and 28 nm. I have also used public domain 45 nm and 15 nm processing kits in my classes.

17. I have authored numerous peer reviewed publications relating to processor design. I have also supervised graduate students performing work relating to processor design. I am also the named inventor on numerous patents and

patent applications, a number of which relate to the design and fabrication of computing devices. My CV, which is EX1004, contains a complete list of my publications and patents.

II. MATERIALS REVIEWED AND CONSIDERED

18. My findings, as explained below, are based on my years of education, research, experience, and background in the field of computer processor and co-processor design and implementation, as well as my investigation and study of relevant materials for this declaration. When developing the opinions set forth in this declaration, I assumed the perspective of a person having ordinary skill in the art, as set forth in Section IV below. In forming my opinions, I have studied and considered the materials identified in the list below.

Exhibit	Description
1001	U.S. Patent No. 10,754,616
1002	Prosecution History of U.S. Patent No. 10,754,616
1005	Declaration of Dr. Mary K. Bolin
1006	PCT Publication WO2005/037326A2 ("Stuttard")
1007	Schmidt et al., "KPROC - An Instruction Systolic Architecture for Parallel Prefix Applications," Parallel and Distributed Computing Practices Vol. 3, No. 2, pp. 163–178 (2000) ("Schmidt-2000")
1008	Schimmler et al., "A bit-serial floating-point unit for a massively parallel system on a chip," in Parallel Algorithms and Applications, 19:2-3, 79-95 (2004)
1009	Schmidt et al., "A Parallel Accelerator Architecture for Multimedia Video Compression," in Euro-Par'99, Lecture Notes in Computer Science 1685, Springer Verlag, 1999, pp. 950–960 ("Schmidt-1999")
1010	Schimmler et al., "The Instruction Systolic Array – Implementation of a Low-Cost Parallel Architecture as Add-On Board for Personal Computers," High Performance Computing and Networking, Lecture

Exhibit	Description
	Notes in Computer Science 797, Springer Verlag, 1994, pp. 487–488 ("Schimmler-1994")
1011	Schimmler et al., "The Instruction Systolic Array in Image Processing Applications," Vision Systems: Sensors, Sensor systems, and Components, Europto 96, pp. 136–144 ("Schimmler-1996")
1012	Yeh et al., "Parallax: An architecture for real-time physics," the 34th International Symposium on Computer Architecture (ISCA), 2007, pp. 232-243 ("Yeh-ISCA07")
1013	Yeh et al., "The Art of Deception: Adaptive Precision Reduction for Area Efficient Physics Acceleration," 40th IEEE/ACM International Symposium on Microarchitecture (MICRO'07), 2007, pp. 394-406 ("Yeh-Micro07")
1014	Yeh et al., "Fool me twice: Exploring and exploiting error tolerance in physics-based animation" (2007)
1015	U.S. Patent No. 5,355,508 ("Kan")
1016	G.E. Moore, "Cramming More Components onto Integrated Circuits," Proceedings of the IEEE, Vol. 86, No. 1 (January 1998)
1017	IEEE SSCS Newsletter 36-37 (reprinting Moore, "Progress in Digital Integrated Electronics," International Electron Devices Meeting, IEEE, 1975, pp. 11-13).
1018	*Microprocessor Report*, "Mendocino Improves Celeron," Volume 12, Number 11, August 24, 1998
1019	Sim et. al, "MIMD-SIMD hybrid system—towards a new low cost parallel system," Parallel Computing 29 (2003) pp. 21-36 ("Sim")
1020	Taveniku et. al. "The VEGA Moderately Parallel MIMD, Moderately Parallel SIMD, Architecture for High Performance Array Signal Processing," Proceedings of IPPS/SPDP'98 (IEEE Computer Society), pp 226-232 ("Taveniku").
1021	The National Technology Roadmap for Semiconductors, 1997 Edition ("NTRS-1997")
1022	International Technology Roadmap for Semiconductor, 1999 Edition ("ITRS-1999")
1023	International Technology Roadmap for Semiconductors, 2000 Update ("ITRS-2000") (from Exhibit B, Folders 1-2 of the Declaration of Nathaniel E Frank-White, EX1029)
1024	International Technology Roadmap for Semiconductor, 2002 Update ("ITRS-2002") (from Exhibit B, Folders 3-7 of the Declaration of Nathaniel E Frank-White, EX1029)

Exhibit	Description
1025	International Technology Roadmap for Semiconductors, 2004 Update ("ITRS-2004") (from Folders 8-11 of the Declaration of Nathaniel E Frank-White, EX1029)
1026	International Technology Roadmap for Semiconductors, 2007 Edition ("ITRS-2007") (from Folders 12-13 of the Declaration of Nathaniel E Frank-White, EX1029)
1027	International Technology Roadmap for Semiconductors, 2009 Edition ("ITRS-2009") (from Folders 14-15 of the Declaration of Nathaniel E Frank-White, EX1029)
1028	International Technology Roadmap for Semiconductors, 2010 Update ("ITRS-2010") (from Folder 16 of the Declaration of Nathaniel E Frank-White, EX1029)
1029	Declaration of Nathaniel E Frank-White (November 30, 2022)
1030	International Technology Roadmap for Semiconductor, 2001 Edition ("ITRS-2001")
1031	Declaration of Nathaniel E Frank-White (November 30, 2022)
1032	Declaration of Nathaniel E Frank-White (November 30, 2022)
1033	Gargini, "The International Technology Roadmap for Semiconductors (ITRS): Past, present and future," GaAs IC Symposium. IEEE Gallium Arsenide Integrated Circuits Symposium. 22nd Annual Technical Digest 2000
1034	Tong et al., "Reducing Power by Optimizing the Necessary Precision/Range of Floating-Point Arithmetic," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 8, No. 3, June 2000 ("Tong") (from pages 6-19 of the Declaration of Gerard P. Grenier, EX1035)
1035	Declaration of Gerard P. Grenier regarding Tong et al., Reducing Power by Optimizing the Necessary Precision/Range of Floating-Point Arithmetic, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 8, No. 3, June 2000
1036	Shirazi et al., "Quantitative Analysis of Floating Point Arithmetic on FPGA Based Custom Computing Machines," Proceedings of the IEEE Symposium on FPGAs for Custom Computing Machines, April 19-21, 1995 ("Shirazi") (from pages 6-13 of Declaration of Gordon MacPherson, EX1037)
1037	Declaration of Gordon MacPherson regarding Shirazi et al., Quantitative Analysis of Floating Point Arithmetic on FPGA Based Custom Computing Machines, Proceedings of the IEEE Symposium

Exhibit	Description
	on FPGAs for Custom Computing Machines, April 19-21, 1995
1038	Altera Product Catalog (2008) (from pages 82-153 of Declaration of Nathaniel E Frank-White, EX1047)
1039	Stackhouse et al., "A 65 nm 2-Billion Transistor Quad-Core Itanium Processor," in IEEE Journal of Solid-State Circuits, vol. 44, no. 1, pp. 18-31, Jan. 2009 ("Intel-Itanium")
1040	Dell White Paper, "Performance Analysis of Intel® Pentium® M and Pentium 4 Processors," July 2005. ("Pentium-4") (from pages 61-65 of Declaration of Nathaniel E Frank-White, EX1047)
1041	The Guru of 3D Report on Radeon HD Series 5700 ("Radeon-5700") (from pages 10-11 of Declaration of Nathaniel E Frank-White, EX1047)
1042	"Intel Core 2 Extreme QX9650 – Penryn Ticks Ahead," Anandtech.com (Oct. 29, 2007) (from pages 30-40 of Declaration of Nathaniel E Frank-White, EX1047)
1043	TSMC Report (2008), "TSMC Ramps Most Advanced Available Process Technology to Volume Production" (pr.tsmc.com/english/news/1526) ("TSMC-40nm")
1044	"Updated CPU Cheatsheet – Seven Years of Covert CPU Operations," Anandtech.com (Aug. 28, 2004) (from pages 42-49 of Declaration of Nathaniel E Frank-White, EX1047)
1045	Natarajan et al., "A 32nm logic technology featuring 2nd-generation high-k + metal-gate transistors, enhanced channel strain and 0.171μm2 SRAM cell size in a 291Mb array," 2008 IEEE International Electron Devices Meeting, 2008
1047	Declaration of Nathaniel E Frank-White (December 10, 2022) (742000)
1048	Declaration of Nathaniel E Frank-White (December 12, 2022) (744136)
1049	PCT Publication WO2002/065700A2
1050	UK Patent Appl. Publ. No. GB 2,415,561
1051	UK Patent Appl. Publ. No. GB 2,390,506
1052	"Intel's 32nm Update: The Follow-on to Core i7 and More," Anandtech.com (Feb. 11, 2009) (from pages 21-28 of Declaration of Nathaniel E Frank-White, EX1047)
1053	"AMD Evergreen: 40nm DirectX 11 GPU Silicon Arrives," Anandtech.com (June 3, 2009) (from pages 16-19 of Declaration of Nathaniel E Frank-White, EX1047)
1054	U.S. Patent Appl. Publ. No. 2007/0258641A1
1055	U.S. Provisional Application No. 61/218691
1058	U.S. Patent Appl. Publ. No. 2003/0033509A1

Exhibit	Description
1059	U.S. Patent Appl. Publ. No. 2008/0288823A1
1063	U.S. Patent Appl. Publ. No. 2009/0148068A1
1066	U.S. Patent Appl. Publ. No. 2009/0114952A1
1067	U.S. Patent Appl. Publ. No. 2006/0088958A1
1068	U.S. Patent Appl. Publ. No. 2008/0023700A1
1069	U.S. Patent Appl. Publ. No. 2009/0150112A1
1070	U.S. Patent Appl. Publ. No. 2005/0237327A1
1075	U.S. Patent No. 5,226,142
1076	U.S. Patent No. 6,681,059
1077	U.S. Patent No. 7,243,165 ("Kravec")
1078	U.S. Patent No. 7,072,872 ("Caid")
1080	U.S. Patent No. 6,963,891
1081	U.S. Patent No. 6,825,857
1082	U.S. Patent No. 6,353,867
1083	U.S. Patent No. 6,477,610
1084	U.S. Patent No. 5,581,773
1087	U.S. Patent No. 6,700,142
1088	VSI AllianceTM: Virtual Component Interface Standard Version 2 (OCB 2 2.0) (April 2001)
1089	"Port," in Microsoft Computer Dictionary, p. 412 (5th. ed. 2002)
1092	Aty et al., "High-Speed, Area-Efficient FPGA-Based Floating-Point Multiplier," Proceedings of the 12 IEEE International Conference on Fuzzy Systems, December 11, 2003 ("Aty") (from p. 14-17 of Declaration of Gordon MacPherson, EX1093)
1093	Declaration of Gordon MacPherson regarding Aty et al., High-Speed, Area-Efficient FPGA-Based Floating-Point Multiplier, Proceedings of the 12 IEEE International Conference on Fuzzy Systems, December 11, 2003
1094	Chen et al., "A cost effective 32nm high-K/ metal gate CMOS technology for low power applications with single-metal/gate-first process," 2008 Symposium on VLSI Technology, 2008, pp. 67-68
1095	Prosecution History of PCT/GB2004004377
1096	U.S. Patent No. 4,450,402
1097	Schiml et al., "A 0.13 /spl mu/m CMOS platform with Cu/low-k interconnects for system on chip applications," 2001 Symposium on VLSI Technology. Digest of Technical Papers (IEEE Cat. No.01 CH37184), 2001, pp. 101-102
1098	Sudha et al., "An Efficient Digital Architecture for Principal

Exhibit	Description
	Component Neural Network and its FPGA Implementation," IETE Journal of Research, Vol. 53, No. 5, September-October 2007, ISSN 0377-2063, pp. 425-431 ("Sudha")
1100	Oh et al., "A fully pipelined single-precision floating-point unit in the synergistic processor element of a CELL processor," in IEEE Journal of Solid-State Circuits, vol. 41, no. 4, pp. 759-771, April 2006,
1103	Prabhu, "SOMGPU: An unsupervised pattern classifier on Graphical Processing Unit," 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), 2008, pp. 1011-1018
1104	Poli et al., "Processing Neocognitron of Face Recognition on High Performance Environment Based on GPU with CUDA Architecture," 2008 20th International Symposium on Computer Architecture and High Performance Computing, 2008, pp. 81-88
1105	Ligon et al., "A re-evaluation of the practicality of floating-point operations on FPGAs," Proceedings. IEEE Symposium on FPGAs for Custom Computing Machines (Cat. No.98TB100251), 1998, pp. 206-215
1106	Held, "From a Few Cores to Many: A Tera-scale Computing Research Overview," white paper, Intel, 2006 (from pages 67-78 of Declaration of Nathaniel E Frank-White, EX1047)
1107	Reddaway et al., "Ultra-high performance, low-power, data parallel radar implementations," IEEE International Radar Conference, 2005., 2005, pp. 822-826
1108	Lokhmotov et al., "Revisiting SIMD Programming," Languages and Compilers for Parallel Computing: 20th International Workshop, LCPC 2007, Urbana, IL, USA, October 11-13, 2007, Revised Selected Papers. Springer-Verlag, Berlin, Heidelberg, 32–46
1109	U.S. Patent No. 4,534,010
1110	Shahbahrami et al., "Optimization of Content-Based Image Retrieval Functions," 2008 Tenth IEEE International Symposium on Multimedia, 2008, pp. 607-612
1111	Noumsi et al., "Acceleration of a content-based image-retrieval application on the RDISK cluster," Proceedings 20th IEEE International Parallel & Distributed Processing Symposium, 2006
1112	Proceedings 20th IEEE International Parallel & Distributed Processing Symposium, 2006 (TOC)

Exhibit	Description
1113	Pengdong et al., "Performance Comparison between Color and Spatial Segmentation for Image Retrieval and Its Parallel System Implementation," 2008 International Symposium on Computer Science and Computational Technology, 2008, pp. 539-543
1114	Bradley et al., "Exploiting Loop-Level Parallelism for SIMD Arrays Using OpenMP," 89-100 (2007) (attachment 13c to Declaration of Dr. Mary K. Bolin, EX1005)
1115	Haran et al., "22 nm technology compatible fully functional 0.1 μm2 6T-SRAM cell," 2008 IEEE International Electron Devices Meeting, 2008, pp. 1-4
1116	Tong et al., Sean & Hahn, Scott. (2008). Operating System Support for Shared-ISA Asymmetric Multi-core Architectures.
1117	Borkar, "Thousand Core Chips A Technology Perspective," Proceedings of the 44th ACM/IEEE Design Automation Conference, pp. 746-749 (2007) (attachment 11b to Declaration of Dr. Mary K. Bolin, EX1005)
1118	"ClearSpeed kicks it up a notch," Arstechnica.com (Oct. 11, 2004) (from pages 5-6 of Declaration of Nathaniel E Frank-White, EX1123)
1119	U.S. Patent No. 7,447,873
1120	Manocha, "General-purpose computations using graphics processors," in Computer, vol. 38, no. 8, pp. 85-88, Aug. 2005
1122	U.S. Patent Appl. Publ. No. 2009/0183167A1
1123	Declaration of Nathaniel E Frank-White (December 12, 2022)
1124	Burger et al., "Composable Multicore Chips," *in* Multicore Processors and Systems. Integrated Circuits and Systems (Springer 2009)
1125	Roma et al., "In the development and evaluation of specialized processors for computing high-order 2-D image moments in real-time," *in* Proceedings Fifth IEEE International Workshop on Computer Architectures for Machine Perception, pp. 170-179 (2000)
1126	"Intel Developer Forum 22nm News Facts," Intel Corporation (Sept. 22, 2009) (from page 80 of Declaration of Nathaniel E Frank-White, EX1047)
1127	Declaration of Gordon MacPherson regarding EX1013
1128	Declaration of Gordon MacPherson regarding EX1039
1129	Declaration of Gordon MacPherson regarding EX1045
1130	Declaration of Gordon MacPherson regarding EX1094
1131	Declaration of Gordon MacPherson regarding EX1097
1133	Declaration of Gordon MacPherson regarding EX1100
1134	Declaration of Gordon MacPherson regarding EX1110

Exhibit	Description
1135	Declaration of Gordon MacPherson regarding EX1115
1136	Declaration of Gordon MacPherson regarding EX1125
1137	TSMC Press Release, "TSMC Ramps 40nm Volume Production to Promote Innovation as Foundries Assume a Larger Role for $300 Billion Industry" (Nov. 17, 2008)
1138	Thornton et al., "Tradeoff analysis of integer multiplier circuits implemented in FPGAs," in 1999 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM 1999). Conference Proceedings (Cat. No.99CH36368), 1999, pp. 301-304

III. MY UNDERSTANDING OF PATENT LAW

19. In developing my opinions, I discussed various relevant legal principles with Petitioner's attorneys. Though I do not purport to have prior knowledge of such principles, I understood them when they were explained to me and have relied upon such legal principles, as explained to me, in the course of forming the opinions set forth in this declaration. My understanding in this respect is as follows:

20. I understand that "*inter partes* review" (IPR) is a proceeding before the United States Patent & Trademark Office for evaluating the patentability of an issued patent's claims based on prior-art patents and printed publications.

21. I understand that, in this proceeding, Petitioner has the burden of proving that the challenged claims of the '616 patent are unpatentable by a preponderance of the evidence. I understand that "preponderance of the evidence" means that a fact or conclusion is more likely true than not true.

22. I understand that, in IPR proceedings, claim terms in a patent are given their ordinary and customary meaning as understood by a person of ordinary skill in the art ("POSA") in the context of the entire patent and the prosecution history pertaining to the patent. If the specification provides a special definition for a claim term that differs from the meaning the term would otherwise possess, the specification's special definition controls. If a claim element is expressed as a

"means" for performing a specified function, I understand that it covers the corresponding structure described in the specification and equivalents of the described structure. I have applied these standards in preparing the opinions in this declaration.

23. I understand that determining whether a particular patent or printed publication constitutes prior art to a challenged patent claim can require determining the effective filing date (also known as the priority date) to which the challenged claim is entitled. I understand that for a patent claim to be entitled to the benefit of the filing date of an earlier application to which the patent claims priority, the earlier application must have described the claimed invention in sufficient detail to convey with reasonable clarity to the POSA that the inventor had possession of the claimed invention as of the earlier application's filing date. I understand that a disclosure that merely renders the claimed invention obvious is not sufficient written description for the claim to be entitled to the benefit of the filing date of the application containing that disclosure.

24. I understand that for an invention claimed in a patent to be patentable, it must be, among other things, new (novel—*i.e.*, not anticipated) and not obvious from the prior art. My understanding of these two legal standards is set forth below.

A. Anticipation

25. I understand that, for a patent claim to be "anticipated" by the prior art (and therefore not novel), each and every limitation of the claim must be found, expressly or inherently, in a single prior-art reference. I understand that a claim limitation is disclosed for the purpose of anticipation if a POSA would have understood the reference to disclose the limitation based on inferences that a POSA would reasonably be expected to draw from the explicit teachings in the reference when read in light of the POSA's knowledge and experience.

26. I understand that a claim limitation is inherent in a prior art reference if that limitation is necessarily present when practicing the teachings of the reference, regardless of whether a person of ordinary skill recognized the presence of that limitation in the prior art.

B. Obviousness

27. I understand that a patent claim may be unpatentable if it would have been obvious in view of a single prior-art reference or a combination of prior-art references.

28. I understand that a patent claim is obvious if the differences between the subject matter of the claim and the prior art are such that the subject matter as a whole would have been obvious to a person of ordinary skill in the relevant field at

the time the invention was made. Specifically, I understand that the obviousness question involves a consideration of:

- the scope and content of the prior art;
- the differences between the prior art and the claims at issue;
- the knowledge of a person of ordinary skill in the pertinent art; and
- if present, objective factors indicative of non-obviousness, sometimes referred to as "secondary considerations."

29. I understand that such secondary considerations include: (a) commercial success of a product due to the merits of the claimed invention; (b) a long-felt, but unmet need for the invention; (c) failure of others to find the solution provided by the claimed invention; (d) deliberate copying of the invention by others; (e) unexpected results achieved by the invention; (f) praise of the invention by others skilled in the art; (g) the taking of licenses under the patent by others; and (h) evidence that the patentee proceeded contrary to the accepted wisdom of the prior art. I understand that secondary considerations are relevant where there is a connection, or nexus, between them and the claimed invention. I have reviewed Section VIII of the Petition regarding the alleged secondary considerations in the patent owner's prior IPRs involving related patents. I have relied on the Petition's analysis that the Patent Owner's alleged evidence of secondary considerations was too speculative and failed to establish the required nexus with any challenged claim, and therefore I have not considered the alleged

secondary considerations in forming my opinions regarding obviousness of the ’616 patent’s claims. Moreover, even if the Patent Owner established secondary considerations of non-obviousness for the claims, any such evidence would not outweigh the overwhelming evidence that the claimed invention would have been obvious that I discuss below in Sections VII-VIII.

30. I understand that in order for a claimed invention to be considered obvious, a POSA must have had a reason for combining teachings from multiple prior-art references (or for altering a single prior-art reference, in the case of obviousness in view of a single reference) in the fashion proposed.

31. I further understand that in determining whether a prior-art reference would have been combined with other prior art or with other information within the knowledge of a POSA, the following are examples of approaches and rationales that may be considered:

- combining prior-art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- use of a known technique to improve similar devices in the same way;
- applying a known technique to a known device ready for improvement to yield predictable results;
- applying a technique or approach that would have been “obvious to try,” *i.e.*, choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success.

- known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art;

- some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior-art reference or to combine prior-art reference teachings to arrive at the claimed invention. I understand that this teaching, suggestion or motivation may come from a prior-art reference or from the knowledge or common sense of one of ordinary skill in the art.

32. I understand that for a single reference or a combination of references to render the claimed invention obvious, a POSA must have been able to arrive at the claimed invention by altering or combining the applied references.

IV. PERSON OF ORDINARY SKILL IN THE ART ("POSA")

33. I have been informed and understand that for purposes of assessing whether prior-art references disclose every element of a patent claim (thus "anticipating" the claim) and/or would have rendered the claim obvious, the patent and the prior-art references must be assessed from the perspective of a person having ordinary skill in the art ("POSA") to which the patent is related, based on the understanding of that person at the time of the patent claim's priority date. I have been informed and understand that a POSA is presumed to be aware of all pertinent prior art and the conventional wisdom in the art, and is a person of ordinary creativity. I have applied this standard throughout my declaration.

34. The ’616 patent involves technology in the field of computing. I have been asked to provide my opinions as to the state of the art in this field by 2009. I use this timeframe because the face of the ’616 patent indicates an earliest claimed priority date of the June 19, 2009 filing date of a provisional application. EX1001, (60). Unless I specify otherwise, whenever I offer an opinion in this declaration about the knowledge of a POSA, the manner in which a POSA would have understood the claims of the ’616 patent or its description, the manner in which a POSA would have understood the prior art, or what a POSA would have been led to do based on the prior art, I am referencing the 2009 timeframe, even if I do not say so specifically in each case.

35. I understand that the Patent Owner may attempt to prove that the alleged invention recited in the challenged claims was conceived at some time prior to the earliest claimed priority date on the face of the patent (June 19, 2009). At the time of this declaration, I am unaware of the Patent Owner having alleged any earlier conception date or produced any evidence to establish any earlier conception date.

36. In my opinion, a POSA in 2009 or 2010 would have had at least a bachelor’s degree in Electrical Engineering, Computer Engineering, Applied Mathematics, or the equivalent, and at least two years of academic or industry experience in computer architecture. More education could substitute for

experience, and vice versa. This person would have been capable of understanding and applying the teachings of the prior-art references discussed in this declaration, and the teachings of the ’616 patent to the extent of the written description and enablement it provides.

37. By 2009, I held a Ph.D. in Electrical Engineering, and I had 20+ years of experience in computer architecture, including the design of microprocessors and parallel processing systems. Therefore, I was a person of more than ordinary skill in the art during the relevant timeframe. However, I worked with many people who fit the characteristics of the POSA, and I am familiar with their level of skill. When developing the opinions set forth in this declaration, I assumed the perspective of a person having ordinary skill in the art, as set forth above.

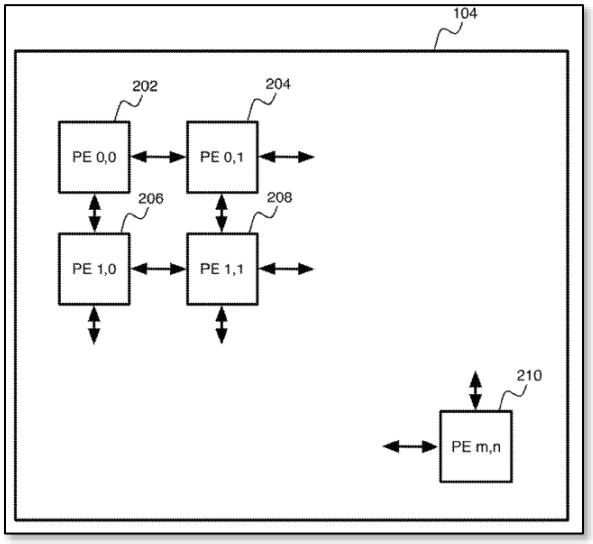
V. THE ’616 PATENT

A. Specification of the ’616 Patent

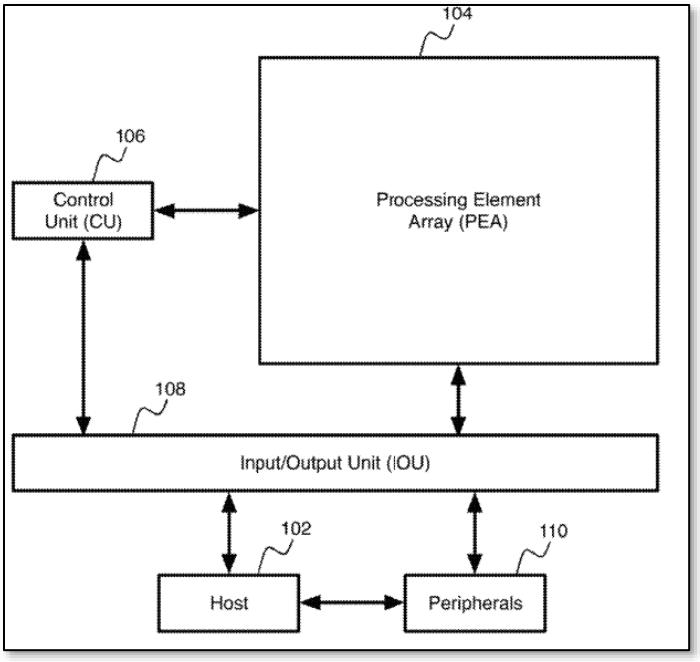
38. The ’616 patent describes a decades-old architecture that was well-known for facilitating parallel computations: an array of interconnected processing elements (PE’s) each having a local memory and an arithmetic unit configured to perform desired arithmetic operations. The ’616 patent itself acknowledges and discusses many such examples, including single-instruction multiple-data (SIMD) arrays, multiple-instruction multiple-data (MIMD) arrays and massively-parallel

computing systems that used such architectures dating back to the 1980s and 1990s (and before). ’616 patent, 3:20-5:36.

39. The ’616 patent describes the use of low-precision high-dynamic range (LPHDR) PEs using this well-known architecture of interconnecting PEs into an array of processors to perform parallel computations, as illustrated in Fig. 2 of the ’616 patent reproduced below. ’616 patent, 9:39-56.



40. The ’616 patent utilizes this PE array in a conventional computer system with a host computer that utilizes the PE array for parallel processing, as illustrated in FIG. 1 of the ’616 patent reproduced below. ’616 patent, 8:12-9:38.



41. Contrary to the ’616 specification’s allegation that it was “commonly believed” among persons of ordinary skill in the art that LPHDR was either “not practical” or “not useful,” (’616 patent, 6:57-66), low precision processing elements with high dynamic range were well-known as were the benefits of LPHDR arithmetic in minimizing resource consumption—including reduced area and power consumption as discussed in §V.D.1 below in connection with the challenged independent claims and further in my analysis of the grounds of unpatentability of the challenged claims in §VIII-IX.

42. The bulk of the ’616 specification describes logarithmic and analog embodiments of LPHDR processing elements in the above architecture (11:57-24:61), which the inventor’s use of simulation on a conventional computer to show that example applications could tolerate low precision computations was not

implemented at the time. ’616 patent, 17:41-23:42. As discussed in further detail below, the claims of the ’616 patent are not directed to these techniques, but are instead drawn to processing elements with floating-point multipliers the ’616 patent mentions once to compare the relatively large size of such multipliers with logarithmic and analog implementations.

43. The ’616 patent’s specification has one isolated statement that “***[i]f*** [some LPHDR] values were represented and manipulated using the methods of floating point arithmetic, they would have binary mantissas of no more than 10 bits plus a sign bit and binary exponents of at least 5 bits plus a sign bit. However, the circuits to multiply and divide these floating point values would be relatively large.” ’616 patent, 5:56-6:6.[1] The specification thus criticizes this option, but then spends the rest of its 30 columns describing logarithmic and analog embodiments.

B. Prosecution History of the ’616 Patent

44. I have reviewed the prosecution history of the ’616 patent (EX1002). The examiner allowed the claims of the ’616 patent without substantively discussing any prior art. Based on my review of the file history and the patent, the examiner allowed the claims of the ’616 patent as originally filed, providing only a

[1] All emphases in this declaration are added unless otherwise indicated.

brief summary of three prior art patents the examiner indicated were "not relied upon" but "considered pertinent to the applicant's disclosure" without any analysis of how they pertain to the claims. *See* EX1002, 137-152 (originally-filed claims), 9-19 (Notice of Allowance and brief summary of prior art patents not relied upon).

45. The '616 patent claims priority through a chain of "continuation" applications. I understand from counsel that in the file histories of the applications to which the '616 patent claims priority, the examiner also allowed the claims in those patents without substantively discussing any prior art. I understand from counsel that in some cases, the examiner issued "double patenting" rejections over claims in previous patents in the priority chain, which the applicant overcame by filing terminal disclaimers or filing a new set of claims.

C. Priority Dates

46. The '616 patent claims priority to a 2009 provisional application, Provisional Application No. 61/218,619 (EX1055) (the "'619 Provisional"). See '616 patent, code (60). In my opinion, the '619 Provisional does not describe at least claims 7-15, which recite "no less than 5000" processing elements.

47. The '619 Provisional disclosed only that in "certain embodiments" the number of "LPHDR arithmetic elements…significantly exceeds (e.g., by at least 20 more than three times) the number of arithmetic elements" of "traditional precision," which discloses neither a number of LPHDR elements, nor what

number of traditional-precision elements the number of LPHDR elements "significantly exceeds." EX1055, 8. In my opinion, this disclosure is not a full, clear, concise and exact description of an array of processing elements numbering "no less than 5000," which I understand from counsel would be required to satisfy the written description requirement. In my opinion, a POSA reading the '619 Provisional would not immediately discern that the "no less than 5000 limitation" was described in the provisional.

48. Although it would have been obvious to a POSA to integrate as many processing elements as semiconductor processing permitted (as I discuss in §VIII.A.1 below), I understand from counsel that obviousness does not satisfy the written description requirement.

49. Thus, in my opinion, independent claim 7 and its respective dependent claims (8-15) are not described in the '619 Provisional and so, I understand form counsel, the '616 patent is not entitled to claim priority to the provisional application's 2009 filing date. I understand that the next-earliest application to which the '616 patent claims priority was filed June 15, 2010. See '616 patent, code (63). I have been informed by counsel that continuation applications share the same specification so that the '616 patent would have the same disclosure as the June 2010 application and therefore would have contained the first mention 5000 processing elements.

D. The Challenged Claims

50. Challenged independent claims 1, 7 and 16 each recite a decades-old processing element (PE) array architecture that was well-known for facilitating parallel computations as discussed above. Further limitations recite conventional computer system components including a "*computing chip*" for the PE array, a "*host computer*," an "*input-output unit*," and a "*host connection*," as also discussed above.

51. The recited PEs include respective multiplier circuits adapted to receive floating-point values having a "*binary mantissa of width no more than 11 bits*" and a "*binary exponent of width at least 6 bits*." Such floating-point formats (in which a binary number can be represented as a fractional "mantissa" (greater than 1 and less than 2) scaled by a power of two (the power being the "exponent")) were also decades old.

52. A simple example of a floating-point representation of a decimal number using base ten would represent the number 3,046,000 as 3.046×10^6. In a typical binary (base-2) floating-point representation, one bit is allocated to indicate the number's sign (positive or negative), some bits are allocated to specify the "exponent" (the number's order of magnitude in base two), and other bits are allocated to specify the mantissa. *See, e.g.*, EX1013 ("Yeh-Micro07"), 395-396 ("The IEEE standard single precision FP representation has a single sign bit (S),

eight exponent bits (E), and 23 mantissa bits (M)…. Precision reduction, as used in this study, refers to the removal of less significant bits from the mantissa….”); Tong (EX1034), 274 (“There are two different IEEE standards for FP computation. IEEE 754 is a binary standard that requires the implied radix (base) to be two. IEEE 854 allows either radix 2 or radix 10 representation. By far, IEEE 754 is more popular and most desktop microprocessors support the IEEE 754 standard.”).

53. It was well known that using reduced-bitwidth floating-point formats (e.g., formats in which the bitwidth of the mantissa is reduced below the IEEE standard single precision format) had benefits. Tong (EX1034) demonstrated that for a variety of image and signal processing applications there was “no noticeable degradation in accuracy” (Tong, 278) when the mantissa of floating-point operands was reduced to 11 bits, with multiple applications (including floating discrete cosine transform (DCT) tolerating even further reduction in the bit-width of the mantissa (e.g., down to as low as 5 bits), but that the exponent bits could not be reduced significantly from 8-bits. Tong, 273, 278-279 (“we can reduce, but not reduce dramatically, the number of exponent bits but the fine precision of the 23-bit mantissa is not essential”).

54. Claim 7 recites that there are “no less than 5000” PEs in the array. The ’616 patent describes no implementation of 5000 processors integrated on a single chip. The ’616 patent only references 5000 PEs once (’616 patent, 28:46-

51), among a generalized, boilerplate description of "certain embodiments," all but one of which have significantly fewer than 5000 PEs (*id.*, 28:13-51). Furthermore, the '616 patent does not describe any number of PEs as being critical, nor does the '616 patent disclose that qualitatively different functionality could be performed or that integrating 5000 PEs on a chip produced any new or unexpected results, only the expected result of increasing computational performance.

1. Reduced Bitwidth Floating-Point Numbers

55. As noted in §V.D above, the '616 patent claims recite multiplier circuits using floating-point values of particular bitwidths that POSAs in the scientific and computing world have been using for years before the '616 patent was filed. As I discuss in more detail in Sections VII-VIII below, these limitations are claimed together with phrases from distant and unrelated portions of the specification, such as a processing element array on a chip (cols. 8-10) with 5000 processing elements (col. 28). None of the conventional add-ons from the specification make Singular's purported invention patentable over the prior art, as I explain in Sections VIII-IX below.

56. The '616 patent claims a system comprising at least one computing chip having arithmetic units comprising a multiplier circuit "adapted to receive as" inputs floating point values having a "binary mantissa of width no more than 11 bits" and a "binary exponent width of at least 6 bits." *See, e.g.*, '616 patent, claim

1. The number of mantissa bits in a floating point value relates to accuracy and can be impacted by the ability to represent numbers that differ from each other by small amounts—*e.g.*, fractional values. For example, in a decimal (base-10) representation, the number 2.13 (represented using two fractional digits) is more precise than if rounded to 2.1 (using just one fractional digit), which is more precise than if rounded to 2 (using no fractional digits).

57. The number of exponent bits in a floating point number relates to the dynamic range of the number, which refers to the range of numerical values that a signal can represent or a unit can operate on or produce. For example, a signal that can represent any numerical value between 1 and 100 has a higher dynamic range than one that can only represent numerical values between 1 and 10.

58. In a typical binary (base-2) floating-point representation, one bit is allocated to indicate the number's sign (positive or negative), some bits are allocated to specify the "exponent" (the number's order of magnitude in base two), and other bits are allocated to specify the mantissa. *See, e.g.*, EX1013 ("Yeh-Micro07"), 395-396 ("The IEEE standard single precision FP representation has a single sign bit (S), eight exponent bits (E), and 23 mantissa bits (M)…. Precision reduction, as used in this study, refers to the removal of less significant bits from the mantissa…."); Tong (EX1034), 274 ("There are two different IEEE standards for FP computation. IEEE 754 is a binary standard that requires the implied radix

(base) to be two. IEEE 854 allows either radix 2 or radix 10 representation. By far, IEEE 754 is more popular and most desktop microprocessors support the IEEE 754 standard.").

59. "In video and imaging systems, color is generally represented as vector coordinates in a multi-dimensional 'color space' having three or more color channels. Common examples include the well-known class[] of RGB." EX1054, [0003]. With RGB, the "color spaces specify pixel values using coordinates that represent intensities of red, green and blue light, respectively." Ex1053, [0003]. "High dynamic range (HDR) imaging presents a more versatile and natural image representation in line with the human vision system." EX1054, [0005]. Thus, "HDR images can present a higher dynamic range than the traditional 8-bit, 10-bit and 112-bit representation to achieve a far higher image quality." EX1054, [0005]. "The pixel value in an HDR image is commonly represented as floating point data or integer data." EX1054, [0005]. So, for example, when using the IEEE's 32-bit floating point format, an HDR pixel would require 96-bits totals: 32-bits for each of the red, green, and blue components. However, it was also known that the benefits of HDR imaging could be achieved using fewer precision bits for each of the RGB components.

60. EX1054, for example, describes the "Radiance RGBE" format, which is "widely used for carrying HDR image data." EX1054, [0047]. The RGBE

format typically represents each of the red, green, and blue components of individual pixels in an image as a floating point number, where each component has only 8 bits for their respective mantissas, and a common 8-bit exponent (i.e., effectively a 16-bit floating point format for each of the RGB components), rather 32-bits for each component. *See* EX1054, [0047]-[0050] ("Radiance RGBE is a widely used format for carrying HDR image data. In an RGBE image, each pixel is represented by 32 bits: one group of bits (e.g., one byte [i.e., 8 bits])) for a red mantissa (R), one group of bits (e.g., one byte) for a green mantissa (G), one group of bits (e.g., one byte) for a blue mantissa (B), and the remaining bits (e.g., one byte) for a common exponent (E)...."). Thus, the RGBE format corroborates that POSAs understood the benefits to imaging applications in representing image data with floating point numbers that had a high dynamic range (e.g., eight exponent bits) but fewer precision bits than the IEEE standard's 23 mantissa bits (e.g., eight mantissa bits for each of the RGB components).

61. Tong (EX1034), for example, noted that "wide dynamic range" was recognized as "a desirable feature," and explained that "[i]t has long been known that many… applications can get by with less precision" over that range. Tong, 273. Tong states that its experimental results show that "programs such as speech recognition and image processing use significantly less power with our reduced bitwidth FP representation than with an IEEE-standard FP representation." Tong,

273. Tong also notes that "[i]t has long been known that many such signal processing applications can get by with less precision/range than full FP," i.e., than with the "IEEE-standard FP representation." Tong, 273. Tong's own experiments demonstrated that "the FP format provides essential dynamic range… but the fine precision of the 23-bit mantissa is not essential." Tong, 279.

62. Tong demonstrated that "reduction in the mantissa bitwidth is the most effective means of reducing power in FP [floating point] datapath elements," such as multipliers. Tong, 277. Tong describes an experiment "[t]o determine the impact of different mantissa and exponent bitwidths" by "emulat[ing] in software different bitwidth FP [floating point] units." Tong, 278.

63. The authors used these emulated FP units on "a set of five signal processing applications" and ran these programs "across a range of mantissa bitwidths." Tong, 278. This experiment showed that "[n]one of the workloads display a noticeable degradation in accuracy when the mantissa bitwidth is reduced from 23 to 11 bits," and that for two of the applications, "the accuracy does not change significantly with as few as 5 mantissa bits." Tong, 278. As Tong explains,

> The reason behind this result is that many programs dealing with human interfaces process sensory data with intrinsically low resolutions. Raw input data with 4-10 bits of precision is rather common in these applications. While intermediate results often require more dynamic range than is available with small bitwidth fixed-point computation, the programs do not require vastly more precision. This is different from scientific programs such as largescale computational fluid dynamics or electrical circuit simulation, which not only require a huge amount of

> precision and dynamic range but also delicate rounding modes to preserve the accuracy of the results. ***What is quite clear from these experiments is that the FP format provides essential dynamic range*** (we can reduce, but not reduce dramatically, the number of exponent bits) ***but the fine precision of the 23-bit mantissa is not essential*** (half as many bits often suffice)

Tong, 278-279.

64. Tong was not alone in this teaching, as I discuss in Section VII.A.3.a below. The challenged claims of the '616 patent encompass this prior-art concept, as I explain in Sections VIII-IX below.

2. Parallel Processing Elements and Applications

65. As discussed in §V.D above, the claims of the '616 patent recite "a processing element array comprising" a plurality of processing elements. *See* claims 1, 7, and 16. The specification admits that computers "have been developed to obtain higher performance," such as "other kinds of computers have been developed to attain higher performance. These machines include single instruction stream/multiple data stream (SIMD) designs, multiple instruction stream/multiple data stream (MIMD) designs, reconfigurable architectures such as field programmable gate arrays (FPGAs), and graphics processing unit designs (GPUs) which, when applied to general purpose computing, may be viewed as single instruction stream/multiple thread (SIMD) designs." '616 patent, 3:19-27. These admitted prior art computing devices are types of parallel processing systems, well-known in the prior art as architectures are designed to improve the

performance and efficiency of parallel computing tasks by allowing multiple processing elements to work together on the same problem.

66. In a SIMD architecture, for example, a single instruction is applied to multiple data elements simultaneously. '616 patent, 3:28-29. This means that all the PEs in a SIMD system (or a subset of the PEs) execute the same instruction at the same time, but on different data. This can be useful for tasks like image processing, where the same operation needs to be performed on large amounts of data. *See, e.g.*, EX1063, [0024] ("Many operations in image processing are fundamentally Single Instruction Multi Data (SIMD) in nature, where local regions of an image all have identical operations performed on them.").

67. In a MIMD architecture, multiple instructions are applied to multiple data elements simultaneously. This means that each processing element in a MIMD system can execute its own instruction on its own data, allowing for greater flexibility and more complex computation. *See, e.g.*, EX1015 ("Kan"), 1:59-62 ("[E]each processor of the MIMD type parallel data processing unit operates independently, and consequently, can each perform advanced processings."). MIMD systems are often used for tasks like scientific simulations and high-performance computing. EX1015 ("Kan"), 2:8-20 ("[D]iscrete models for solving simultaneous equations such as those for expressing the analysis of networks"… do not require large-scale parallelism, but do require complex processing.

Therefore, medium-scale MIMD type parallel processing is suitable for these models…..").

68. Applications with high levels of data parallelism can take advantage of a SIMD processor to process large amounts of data quickly and efficiently. EX1015 ("Kan"), 1:52-56 ("SIMD type parallel data processing unit has only a single instruction stream with the result that control is simple, thus making large-scale parallel processings easy and extremely high-speed processings possible."). Examples of applications with high levels of data parallelism include image and video processing, scientific simulations, and machine learning algorithms. On the other hand, applications with low levels of data parallelism may not benefit as much from this approach. These applications may have a smaller amount of data to process, or the data may be structured in a way that makes it difficult to divide and distribute among multiple processing elements. Examples of applications with low levels of data parallelism include some types of database queries and certain types of business analytics.

69. Overall, the level of data parallelism in an application depends on the specific nature of the data being processed and the algorithms being used. Some applications may have a mix of both high and low levels of data parallelism, depending on the specific tasks they are performing. *See, e.g.*, EX1015 ("Kan"), 2:7-56. Some of the example applications I mentioned above may have a mix of

both high and low levels of data parallelism, like scientific simulations, image and video processing, and machine learning algorithms. For instance, a scientific simulation may have high levels of data parallelism when running complex calculations on a large dataset, but low levels of data parallelism when performing more basic operations or when visualizing the results. An image or video processing application may have high levels of data parallelism when performing complex operations on large datasets, but low levels of data parallelism when performing more basic operations or when displaying the results. A machine learning algorithm may have high levels of data parallelism when training a model on a large dataset, but low levels of data parallelism when making predictions on individual data points. A database application may have high levels of data parallelism when performing complex queries on a large dataset, but low levels of data parallelism when processing small amounts of data or when performing sequential operations.

70. As noted above, image processing applications that often have a mix of both high and low levels of data parallelism include image recognition, image restoration, and image compression. An image recognition application, for instance, may have high levels of data parallelism when training a model on a large dataset of images, but low levels of data parallelism when making predictions on individual images. EX1011, 1115 (“OCR and on-line handwritten recognition are

computationally expensive. Training time is a major bottleneck for improving handwriting recognition (3+ weeks, depending on the language).”). The training phase typically occurs by dividing the dataset into smaller subsets and distributing them among multiple processing elements that work simultaneously on different parts of the dataset, which can help to speed up the training process. However, the same image processing application may also computations with low levels of data parallelism when displaying the results or making predictions on individual images, which are tasks that can typically be performed by a single, or small set, or more-complex processing element (such as when multiple images are being predicted and/or displayed). Likewise, identifying an image within a large database of images is a task that can benefit from data parallelism, to speed up the process of identifying an image and reduce the amount of time and resources required to do so. EX1111, 1-2 (“Content Based Image Retrieval (CBIR) is a technique that allows one to find out images of a database that are (at least) partly similar to a given reference image…. [T]he only solution to improve the performance of current CBIR systems is to accelerate the algorithm using special-purpose implementations. ***Accelerating the application on a parallel machine is the most natural choice***, ….”).

71. As I discussed above (§38), many of the same applications that benefit from parallel processing can also use reduced-bitwidth high dynamic range

floating point numbers because the impact of the low precision on the overall accuracy of the results is often acceptable, in particular when the range of values may be more important than the precision of the calculations. Thus, POSAs recognized the benefits of combining the benefits of reduced-bitwidth floating-point formats with parallel processing computing devices, as I explain in, for example, Section VII.A.3 below. As I also explain in those sections, the challenged claims of the '616 patent also encompass this prior-art concept.

VI. PRIOR-ART REFERENCES

A. Public Accessibility of Non-patent Prior Art Exhibits

72. My opinions regarding public accessibility of the following exhibits are discussed in my Appendix attached hereto in the sections of my Appendix summarized in the table below:

Exhibit	Appendix Section
1007	§II.A.1
1009	§II.A.2
1010	§II.A.3
1011	§II.A.4
1012	§II.C.1
1013	§II.C.2
1014	§II.C.3
1016	§II.C.4
1017	§II.C.5
1018	§II.C.6
1019	§II.C.7

1020	§II.C.8
1021	§II.B.1
1022	§II.B.2
1023	§II.B.3
1024	§II.B.5
1025	§II.B.6
1026	§II.B.7
1027	§II.B.8
1030	§II.B.4
1034	§II.C.43
1036	§II.C.44
1038	§II.C.9
1039	§II.C.10
1040	§II.C.11
1041	§II.C.12
1042	§II.C.13
1043	§II.C.14
1044	§II.C.15
1045	§II.C.16
1088	§II.C.19
1092	§II.C.45
1094	§II.C.20
1097	§II.C.21
1098	§II.C.22
1100	§II.C.23
1103	§II.C.24
1104	§II.C.25
1105	§II.C.26
1106	§II.C.27

1107	§II.C.28
1108	§II.C.29
1110	§II.C.30
1111	§II.C.31
1113	§II.C.32
1114	§II.C.33
1115	§II.C.34
1116	§II.C.35
1117	§II.C.36
1118	§II.C.37
1120	§II.C.38
1124	§II.C.39
1125	§II.C.40
1126	§II.C.41
1137	§II.C.42
1138	§II.C.46

B. Stuttard (EX1006)

73. According to the face of the document, Stuttard is an International Patent Application Publication published on April 28, 2005, titled "Unified SIMD Processor." EX1006, (54). I explain in Sections VII-IX below (Grounds 1A-1C) my opinion that claims 1-5, 7-14, 16-22, and 24 would have been obvious over Stuttard in combination with Shirazi (EX1036).

C. Shirazi (EX1036)

74. According to the face of the document, Shirazi is a paper published by the IEEE in 1995, by Nabeel Shirazi, Al Walters, and Peter Athanas, titled

"Quantitative Analysis of Floating Point Arithmetic on FPGA Based Custom Computing Machines." *See* EX1036, 155 ("© 1995 IEEE"). In my Appendix to this declaration, I discuss my opinions regarding public accessibility of the Shirazi reference before the '616 patent's claimed priority date. I explain in Sections VII-IX below (Grounds 1A-1C) my opinion that claims 1-5, 7-14, 16-22, and 24 would have been obvious over Shirazi in combination with Stuttard.

VII. GROUND 1A: CLAIMS 1-5 ARE OBVIOUS OVER STUTTARD IN COMBINATION WITH SHIRAZI

75. I have been asked to provide my opinion concerning whether claims 1-5, 7-14 and 16-22 and 24 of the '616 patent would have been obvious to a POSA in light of the prior-art references identified in the Petition. For the reasons explained below in Sections VII-IX, it is my opinion that each of claims 1-5, 7-14 and 16-22 and 24 would have been obvious to a POSA.

A. The Stuttard-Shirazi Processor

1. Stuttard (EX1006)

76. Stuttard discloses a "Multi-Threaded Array Processing (MTAP) architecture" with "parallel data processing capability" for "high performance, high data rate processing." Stuttard, 8:23-9:14.

77. As shown in Figure 7 (annotated below), each MTAP "Processor Core" is a "data processor compris[ing] a single control unit" (purple) and two

execution units: a SIMD "poly" execution unit (blue) and a scalar "mono" execution unit (red). Stuttard, 2:29-34.

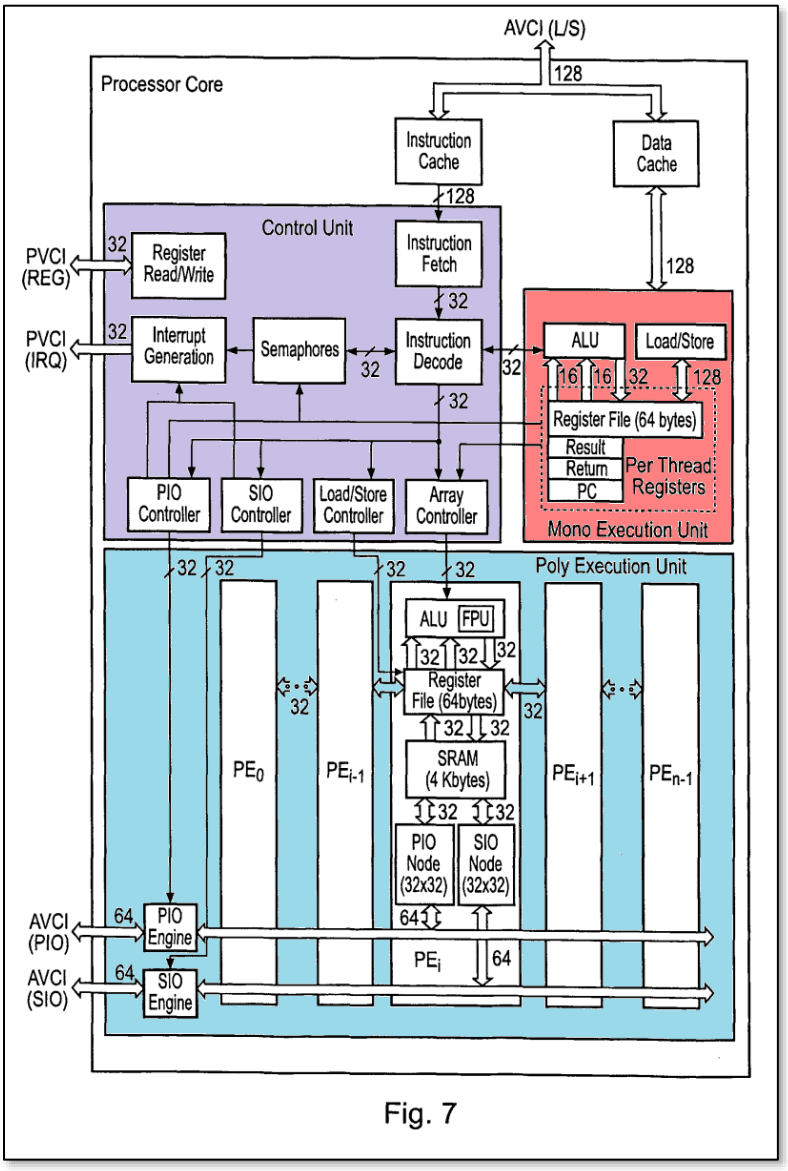


Fig. 7

78. The poly unit comprises an array of processing elements (PEs) and "gets its performance largely from the number of PEs brought to bear on the problem rather than the processing power of the individual PEs." Stuttard, 7:19-21. Thus, the number of PEs in the poly unit is designed "to be scaled to meet the

needs of the application." Stuttard, 10:5-6. For example, the poly unit "may consist of tens, hundreds or even thousands of PEs." Stuttard, 10:12-14. Stuttard (12:15-16) similarly discloses that "the [MTAP] core can be scaled…to 256 or even thousands of PEs" (i.e., a single MTAP core on the chip can be scaled to thousands of PEs").

79. The mono unit is for handling "scalar and non-parallel" data (Stuttard, 10:8-10), and has its own circuitry "for accelerating software implementations of floating point operations" (Stuttard, 16:13-20).

80. Each PE in the poly unit includes an arithmetic logic unit (ALU) with a floating-point unit (FPU) and local memory (e.g., SRAM). Stuttard, Figure 7 (annotating an exemplary PE blue below with an ALU and FPU), 7:14-24 ("each PE has an ALU"), 12:7-11 (exemplary "MTAP core with 64 PEs" with each PE having "an FPU extension"); 13:21-14:4 (noting the "instruction set" for "both mono and **poly execution units**" includes "floating point adds, subtracts, multiplies, divides"). The mono unit likewise includes an ALU capable of "floating point operations." Stuttard, 16:13-20, Figure 7 (annotating the mono execution unit red below); *see also* 13:21-14:4 (noting the "instruction set" for "both **mono** and poly **execution units**" includes "floating point adds, subtracts, multiplies, divides").

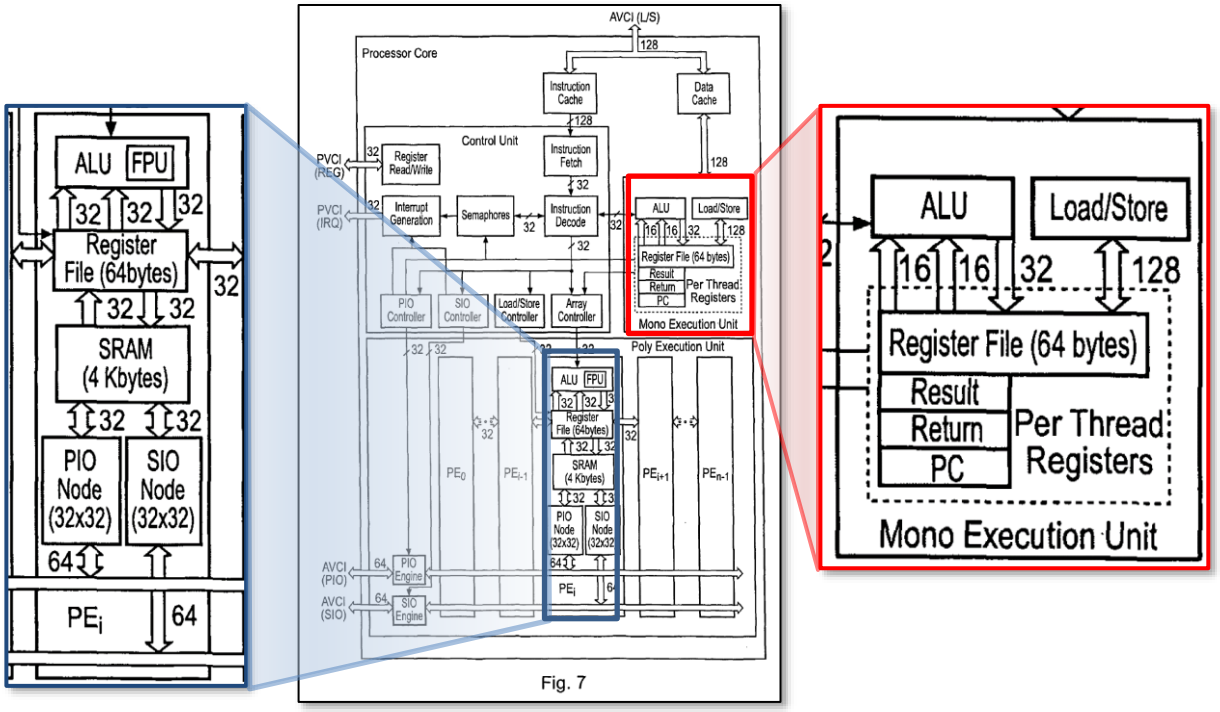


Fig. 7

81. Stuttard teaches that the "mono execution unit is likely to require a wider and more complex ALU" than the poly unit's ALUs. Stuttard, 7:17-24. This is because the "poly execution unit gets its performance largely from the number of PEs brought to bear on the problem, rather than the power of the individual PEs," whereas the mono unit "is only instantiated once" and does not need to be as "efficiently implemented" (e.g., can occupy more die area) compared to the poly PEs. Stuttard, 7:19-24; *see also* VIII.A below (discussing POSAs reason to have a chip with 5000+ poly PEs and several mono units on the same chip).

2. Shirazi (EX1036)

82. Shirazi teaches that "[m]any algorithms rely on floating point arithmetic for the[ir] dynamic range" and "require millions of calculations per

second." Shirazi, 155. For example, 2-D FFTs (Fast Fourier Transforms) and FIR (finite impulse response) filtering computations are commonly used in "[i]mage and digital signal processing applications [that] typically require high calculation throughput." Shirazi, 155. Stuttard teaches that these applications are, therefore, "candidates for acceleration using custom computing machines (CCMs) [] tailored for the application" (Shirazi, 155)—i.e., the type of machine Stuttard describes. Circuitry to operate on IEEE-standard full-precision 32-bit floating-point values, however, required "excessive area" on a chip compared to circuitry adapted to "custom formats, derived for individual applications." Shirazi, 155. Each of Shirazi's two disclosed floating-point formats has at least a 6-bit exponent and no more than an 11-bit mantissa. Specifically, Shirazi's two custom formats are illustrated below (with annotations that I added below): (1) an 18-bit format developed for "2-D FFT application," having one sign bit, 7 exponent bits, and 10 mantissa bits; and (2) a 16-bit format developed for "the FIR filter application," having one sign bit, 6 exponent bits, and 9 mantissa bits.

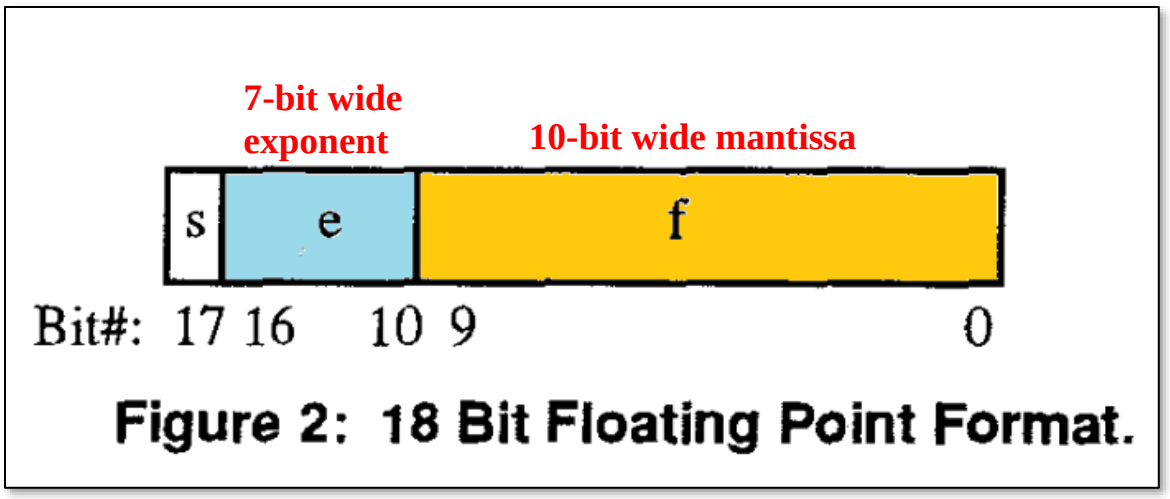


Figure 2: 18 Bit Floating Point Format.

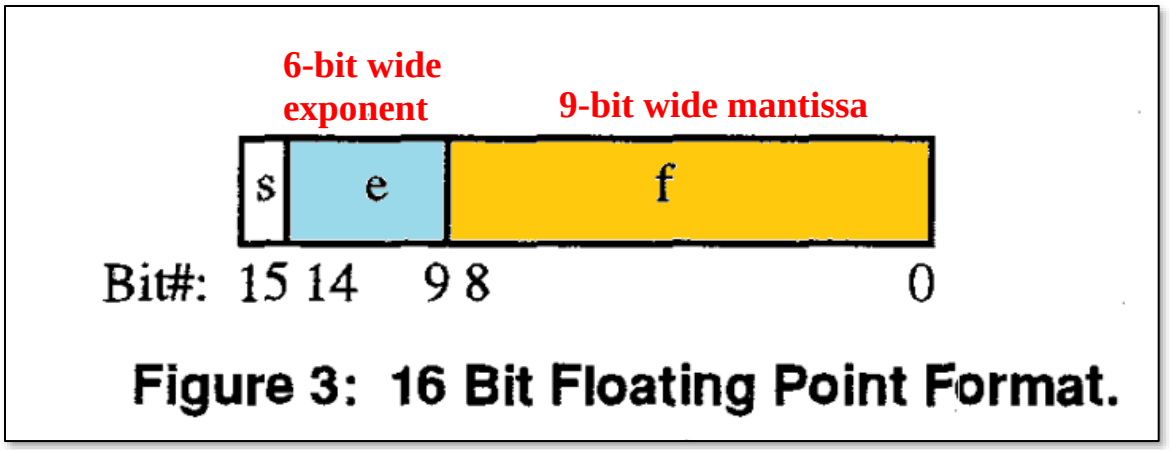


Figure 3: 16 Bit Floating Point Format.

Shirazi, 156.

83. Shirazi teaches that these custom floating-point formats provided sufficient "dynamic range" for applications like 2D-FFT and FIR, respectively, while also "minimiz[ing] area" on the chip compared to circuitry for higher-precision values. Shirazi, 156, 161.

3. Stuttard-Shirazi Combination

a. Reduced-bitwidth Poly PEs

84. In my opinion, a POSA had reasons to adapt the PEs in Stuttard's poly unit to perform floating-point math (e.g., floating point multiplication) using Shirazi's custom reduced-bitwidth floating-point values.

85. As I discussed above in Section VII.A.2, Shirazi taught that its floating-point formats having reduced-bitwidth mantissas provided sufficient precision for common high-data rate applications, while also allowing the arithmetic circuitry to be implemented with reduced "area" on the chip compared to using full-precision circuits. Shirazi, 155-56. Specifically, Shirazi noted that while "conventional" "floating point operators require excessive area," "custom formats…can be implemented on a fraction of a single FPGA." Shirazi, Abstract ("Many algorithms rely on floating point arithmetic for the dynamic range of representations and require millions of calculations per second. Such computationally intensive algorithms are candidates for acceleration using custom computing machines (CCMs) being tailored for the application. Unfortunately, ***floating point operators require excessive area*** (or time) for conventional implementations. Instead, ***custom formats, derived for individual applications***, are feasible on CCMs, ***and can be implemented on a fraction of a single FPGA***.").

86. Shirazi was not alone in this teaching, as this was well-known background knowledge for POSAs. For example, EX1034 ("Tong") similarly taught that "[i]t has **long been known** that many…applications can get by **with less precision**" (273) and, thus the IEEE's single-precision (32-bit) format often "is not essential" (279); *see* Appendix (opinions regarding public accessibility of EX1034). Tong also taught that minimizing the number of precision (mantissa) bits reduces power consumption of each processing element because a "reduced precision FP unit" has a "small size" compared to higher-precision units (282).

87. As another example, EX1092 ("Aty"), is a 2003 paper entitled "Highspeed, Area-Efficient FPGA-Based Floating-point Multiplier," and the paper explained that "[t]he floating-point multiplier, presented here [*i.e.*, in the paper], is implemented for real-time signal processing." Aty, 274. Aty showed that one of the floating-point formats used in its multiplier was a 16-bit floating-point format with 9 mantissa bits and 6 exponent bits. Aty, 276 Tables 2-4 (annotated below). *See also see* Appendix (opinions regarding public accessibility of Aty).

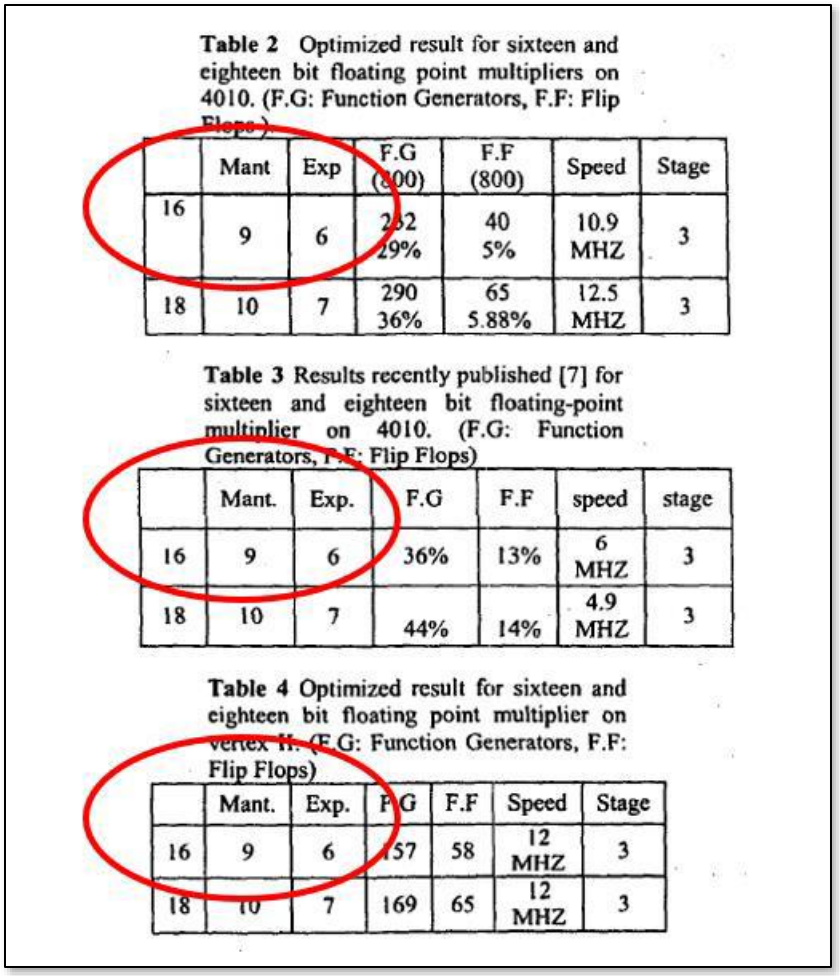


Table 2 Optimized result for sixteen and eighteen bit floating point multipliers on 4010. (F.G: Function Generators, F.F: Flip Flops).

	Mant	Exp	F.G (800)	F.F (800)	Speed	Stage
16	9	6	232 29%	40 5%	10.9 MHZ	3
18	10	7	290 36%	65 5.88%	12.5 MHZ	3

Table 3 Results recently published [7] for sixteen and eighteen bit floating-point multiplier on 4010. (F.G: Function Generators, F.F: Flip Flops)

	Mant.	Exp.	F.G	F.F	speed	stage
16	9	6	36%	13%	6 MHZ	3
18	10	7	44%	14%	4.9 MHZ	3

Table 4 Optimized result for sixteen and eighteen bit floating point multiplier on vertex II. (F.G: Function Generators, F.F: Flip Flops)

	Mant.	Exp.	F.G	F.F	Speed	Stage
16	9	6	157	58	12 MHZ	3
18	10	7	169	65	12 MHZ	3

Aty, Tables 2-4 (annotated)

88. As another explain, EX1098 ("Sudha") is a 2007 paper about a type of "***neural network***" known as a "Principal Component Neural Network (PCNN)." Sudha, 425 ("The neural network that extracts principal components from the input data set is called Principal Component Neural Network (PCNN)."). Sudha explained that its "hardware design of a 64-input 16-output PCNN was coded in [the] Verilog hardware description language. Each parameter in the design is represented ***as a floating point number with an 8-bit mantissa and an 8-bit***

exponent." *Id.*, 428. *See also* Appendix (opinions regarding public accessibility of Sudha).

89. As another example, EX1105 ("Ligon") is a 1998 paper about designing floating-point formats for use in FPGAs. Ligon disclosed the "use of reduced precision floating-point formats to conserve chip area" and recognized that the reduced precision was sufficient "[f]or a broad range of applications" and "retain[ed] the benefit of a large dynamic range." Ligon, 207. *See also* Appendix (opinions regarding public accessibility of Ligon).

90. As another example, EX1125 ("Roma") is a 2000 paper about an "architecture for computing 2-D gray level image moments…by adopting systolic processing techniques and floating point arithmetic" that is "well suited to implement" based on "Digital Signal Processors (DSP), Field Programmable Date [sic: Gate] Arrays (FPGA) and Application Specific Integrated Circuits (ASIC." Roma, 170. *See also* Appendix (opinions regarding public accessibility of Roma). Roma (at 171-174) teaches that determining image moments is useful in pattern-recognition applications (Roma 171) and that "[e]xperimental results show…th[e following] floating-point precision…seems to be enough": "1 sign bit (*s*), 8 bits for the exponent (*e*) and 10 bits used for the fractional part (*f*) of the mantissa," as illustrated in Roma's Figure 8 (copied below):

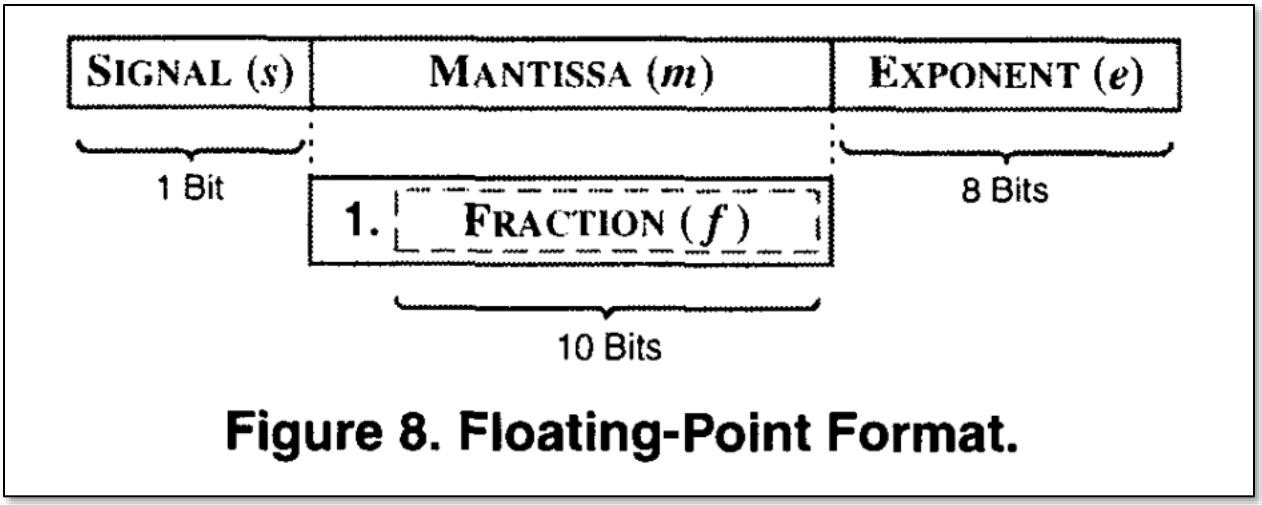


Figure 8. Floating-Point Format.

91. Thus, Shirazi taught (and Tong, Aty, Sudha, Ligon, and Roma and others corroborate my opinion) that custom floating-point formats with less precision than the IEEE's full-precision format (i.e., fewer than 23 mantissa bits) provided a desirable solution for applying smaller-sized arithmetic units to various digital-signal and image processing applications, which reduced power consumption and allowed for more processing elements on a single chip to achieve higher parallelism and speed. *See, e.g.*, Tong, 280, 278-79 (teaching several signal processing applications that had no "noticeable degradation in accuracy" when using 11 (or fewer) mantissas bits, and that six-plus exponent bits "preserver[ed] a reasonable degree of accuracy").

92. As additional corroboration that POSAs had reason to customize Stuttard's poly PEs for applications that were known to work well with reduced-bitwidth formats, EX1106 (an Intel white paper from 2005) taught that "processing

elements **dedicated to a particular workload**…will perform better for that task than a general-purpose execution unit." EX1106, 6. *See* Appendix (opinions regarding public accessibility of EX1106). Thus, POSAs would have recognized the benefit in customizing a Stuttard-based processor to have reduced-bitwidth PEs in the poly execution unit dedicated for use in applications where the reduced-bitwidth floating-point computations were sufficient for the needs of particular applications.

93. For example, as I discussed above in Section VII.A.1, Stuttard's MTAP processor architecture is for use in "high performance, high data rate processing" applications. Stuttard, 8:24-28. POSAs would have understood this architecture to be well-suited for signal-processing applications like the 2-D FFTs and FIR filtering computations – common to many image processing (and other) applications – that Shirazi taught were compatible with reduced-precision floating-point formats. Shirazi, 155-56. For evidence corroborating that POSAs understood FFT and FIR computations were well-suited to processor architectures like Stuttard's, EX1080 at 1:55-2:9 taught that "[t]he FFT is widely used in real time digital signal processing requiring fast execution" and it is "important that the FFT be able to make effective use of these [SIMD] kinds of architectures." Stuttard's processor is an example of a SIMD architecture. *See, e.g.*, Stuttard, Title ("Unified SIMD Processor"), 4:20-6:27 (noting Stuttard's processor includes a

"SIMD part comprising an array of processing elements adapted to operate on said parallel data [i.e., the poly unit] and a scalar part adapted to operate on said scalar data [i.e., the mono unit].").

94. As additional corroboration of my opinion that POSAs understood that Stuttard's processor was well-suited to the types of computations Shirazi taught were well-suited to reduced-bitwidth floating point units (e.g., FFTs), I further refer to EX1107 and EX1108. EX1107 is a 2005 paper that described using MTAP-based processors (i.e., like Stuttard's processor) from Stuttard's Applicant ("ClearSpeed Technology") to perform FFT computations. *See* EX1107, 823-825. *See* Appendix (opinions regarding public accessibility of EX1107). Similarly, EX1108 is a 2008 paper co-authored by two of Stuttard's inventors – Neil Hickey and David Stuttard, *see* Stuttard, (75) – which corroborates my opinion that it was known to program MTAP architectures to compute **"2D and 3D FFTs**." EX1108, 32-34, 44. *See* Appendix (opinions regarding public accessibility of EX1108).

95. As I discussed above (VII.A.1 above), Stuttard's discloses that the poly unit "gets its performance largely from **the number of PEs** brought to bear on the problem rather the processing power of the individual PEs," which "need to be efficiently implemented." Stuttard, 7:19-24. As I discussed above (with corroboration), a POSA would have understood that one known way to "efficiently implement[]" processing elements – e.g., to gain the efficiency of reducing

physical area of each PE to allow more on a chip and reduce the power requirement of each PE – would have been to reduce the precision of the floating-point circuitry in each PE. Thus, Stuttard's disclosure at 7:9-24, for example, would have further motivated POSAs to pursue techniques like Shirazi's reduced-bitwidth floating-point formats, which would have allowed for more PEs on the chip. Shirazi, 155-56.

96. Thus, in my opinion, a POSA had reasons to adapt the floating-point units in the PEs in Stuttard's poly unit to have circuitry that was adapted to accept as inputs, and operates on (e.g., perform floating-point multiplication of), floating-point values having bit-widths no greater than one of Shirazi's reduced-precision floating-point formats discussed above (§VII.A.2).

97. In my opinion, POSAs would have reasonably expected success in this combination (which I will refer to as the "Stuttard-Shirazi combination"). This is because Shirazi teaches that its "custom formats…are feasible on CCMs" (custom computing machines). Shirazi, 155-56. POSAs would have understood that Stuttard's MTAP architecture is an example of a custom computing machine: Stuttard expressly teaches that the "micro-architecture" of processing elements in the poly unit do not have to be identical processing elements in the mono unit, e.g., "the poly processing elements could have an 8-bit ALU, while the mono execution unit has a 32-bit ALU." Stuttard, 3:15-18; *see also* 9:13-14 ("The processor is

designed in a ***highly modular fashion*** which allows ***many details of a specific implementation to be easily customized*** for the target application."). Because Stuttard teaches that the "micro-architecture" of the poly and mono execution units can differ, POSAs would have understood that the precision of the floating point units in the poly and mono units (a micro-architecture design choice) were also capable of differing using ordinary skill. Indeed, Stuttard reinforces this at 24:28-29 by describing simulation software for "the core" that allowed the "number of PEs, size of memory, and **other options**" to be "parametrized" (i.e., a changeable value in the design simulation). Thus, POSAs would have reasonable expected success in using only ordinary skill to customize the floating-point units in Stuttard's poly unit for the target applications Shirazi and Stuttard both describe (e.g., FFT computations), by adapting these PEs to receive as inputs floating-point values in one of Shirazi's reduced-precision formats and no wider than the bitwidths used in those floating point values (because it would have been inefficient for the PEs to be adapted to receive as inputs values with greater bit-widths than the maximum bit-widths of Shriazi's formats).

b. Two-Dimensional PE Array

98. Stuttard's Figure 4 (which I have annotated below) illustrates one "Mono" execution unit and several PEs in the "Poly" execution unit. Stuttard, 10:12-19, 12:15-17.

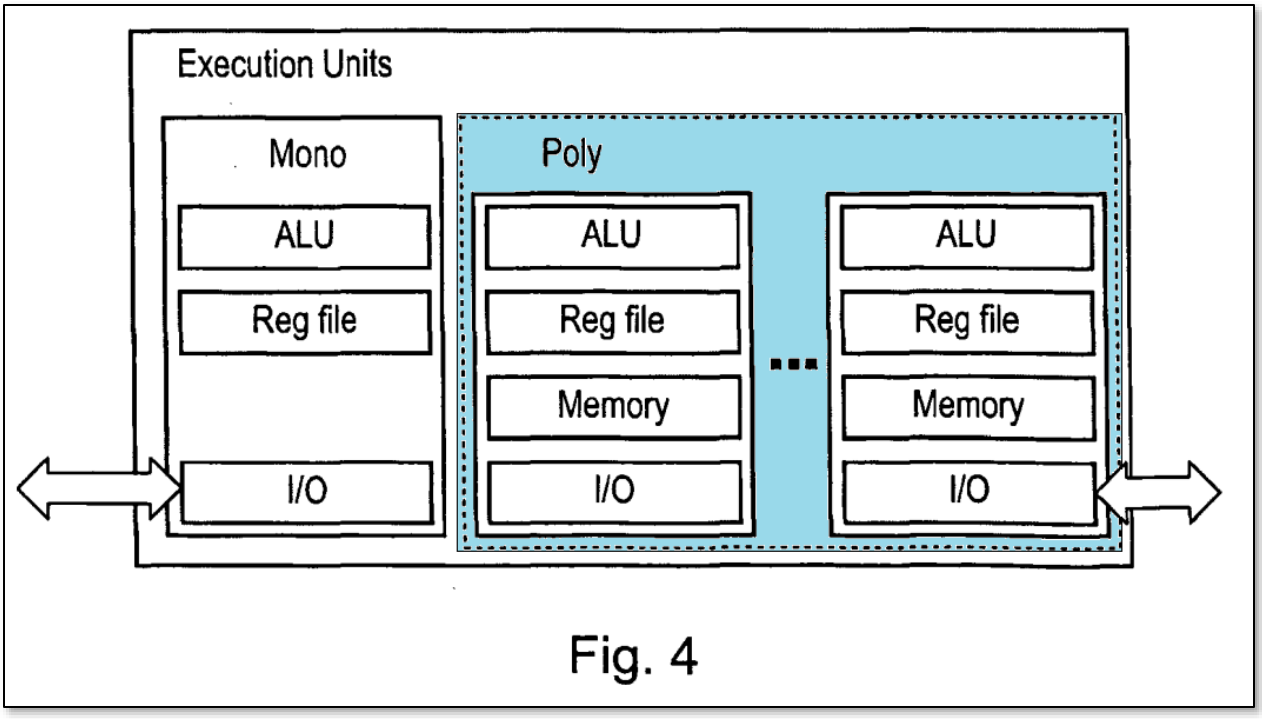


Fig. 4

99. Stuttard's Figures 4 and 7 illustrate poly PEs side-by-side because each PE communicates with its left and right neighbors in what Stuttard calls a "swazzle" path. Stuttard, 20:31-21:2. However "[t]his is a conceptual view, which reflects the programming model" for an MTAP processor, and a POSA would have understood that the side-by-side illustration in Figures 4 and 7 does not limit the physical layout of the processor on an actual, fabricated computing chip. Stuttard, 13:2-5. For example, the mono unit is arbitrarily shown beside the poly unit in Figure 4 but above the poly unit in Figure 7, which demonstrates that neither Figure 4 nor Figure 7 is an actual, physical layout of a computing chip based on Stuttard's teachings. A POSA would have understood that the physical layout of an actual computing chip was a matter of design choice and, as discussed below,

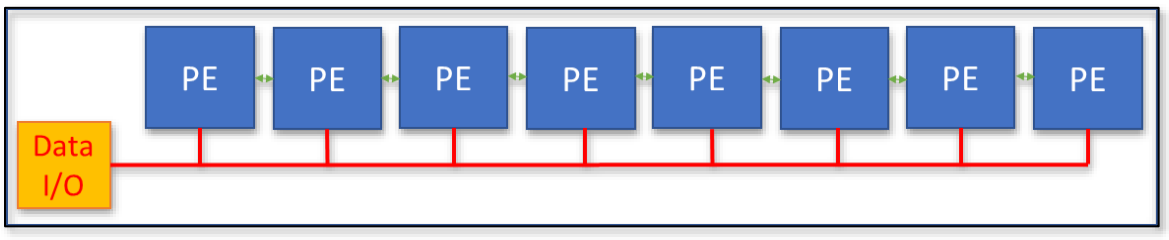
POSAs had reasons to manufacture a physical chip using a two-dimensional PE layout. For evidence corroborating see EX1081, at 8:46-47, 21:5-20, which I discuss further below.

100. **First**, POSAs knew that a two-dimensional layout reduces the overall length of wiring in the chip, like Stuttard's "channels" that transfer data between PEs and the host. Stuttard, 17:20-23, 23:26-32; §VII.B.6 below. These data channels are connected to every PE in the device. Stuttard, 17:20-23 (disclosing that each PE has "One or more I/O channels" and that "Data is transferred in to and out of, the PE's memory using I/O instructions"), 23:26-32 (disclosing that "the SIO channels are used for continuously streaming packets into, and out of, PE memory" and that the "PIO channels are used by the PEs to send requests to another subsystem that handles lookups"), FIG. 7 (illustrating that the "PIO Engine" and "SIO Engine" are connected to every PE via a "PIO Node" and "SIO Node," respectively). By physically arranging PEs in a two-dimensional array, the overall length of each channel is reduced, which beneficially reduces (i) signal delays, (ii) resistances (and thus heat, which is proportional to resistance) and (iii) and undesirable "parasitic" capacitances. For evidence corroborating POSAs knowledge in this regard EX1066 at [0007] taught that "the length of interconnections between cells impacts the performance. For instance, longer lines have **increased resistance, increasing heat dissipation** in the integrated circuit.

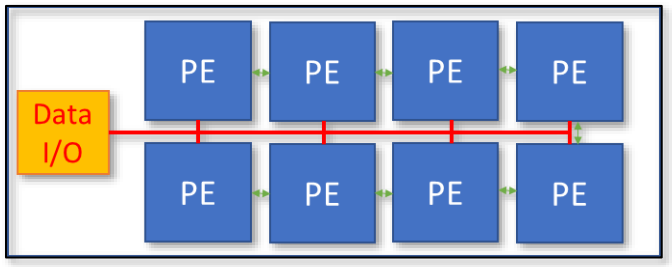
Longer lines **also increase the capacitance** of the overall design." POSAs would understand that the increased capacitances from longer interconnections lines is a "parasitic" capacitance because it is (generally) not a useful capacitance and thus undesirable. POSAs would understand that increased wire resistance and capacitance associated with longer lines would increase the power consumption and wiring delays.

101. To illustrate the benefits of a 2D dimensional array of PEs over a 1D array, the first demonstrative that I created below shows 8 PEs arranged in a one-dimensional line; the second demonstrative shows the same 8 PEs arranged in a two-dimensional array. The red line in each figure represents a channel connected to each PE (e.g., Stuttard's PIO channel); the green lines represent a potential "swazzle" path connection between nearest neighbor PEs. As demonstrated below, both layouts allow for a linearly connected swazzle path (i.e., only a connection to one or two neighbors) but the overall length of the channel in the 2D array can be approximately half as long as the length of the channel in the 1D line.

Linear Layout



Two-dimensional Layout



I also note that while the above two-dimensional layout illustrates that a 1D swazzle path can be maintained in a two-dimensional physical layout, a POSA would have understood that a 2D "swazzle" path – i.e., an interconnection between North-East-West-South (so-called "NEWS" connections) would have also been a conventional, and advantageous interconnection between PEs physically arranged in two-dimensions, as I discuss below with corroboration from Stuttard's applicant.

102. **Second**, POSAs knew it was desirable and conventional for computing chips to have aspect ratios (width-to-length ratios) closer to 1, because very high or very low aspect ratios cause "mechanical stress[es]" and "adversely change the electrical properties of the IC circuitry," as taught by EX1067 at [0007]:

> ***Aspect ratio is a critical parameter*** in the design of most monolithic ICs, including, by way of example, power integrated circuit devices. An IC fabricated on a semiconductor die having ***a very large or very small aspect ratio often suffers from mechanical stress caused by the molding compound used to package the die***. This ***stress can adversely change the electrical properties of the IC circuitry***. For minimum stress a ***semiconductor die should have an aspect ratio that is close to 1.0, i.e., a length that is substantially equal to its width***.

Though EX1067 refers to "power integrated circuit devices," a POSA would understand that its teachings about how the aspect ratio of a computing chip impacts the mechanical stress and may adversely change electrical properties would be generally applicable to all "monolithic ICs," such as a computing chip based on Stuttard's teachings.

103. **Third**, other prior art related to Stuttard's general architecture further corroborates my opinion that a two-dimensional interconnect topology (and corresponding two-dimensional physical topology) would have been a conventional alternative to a linear topology. For example, EX1108 – co-authored by two of Stuttard's inventors – described "a simple extension to the C [programming] language" for use in "[m]assively parallel SIMD array architectures" including "MTAP" architectures like Stuttard's, and corroborates the POSA's knowledge that with this architecture, "as the number of PE cores increases **other (than linear)** interconnect topologies could be introduced." EX1108, 35. Thus, EX1108 corroborates that non-linear (e.g., 2D) interconnect topologies would have been a well-known design option for an MTAP architecture like Stuttard's.

104. Similarly, EX1081—assigned to Stuttard's Applicant "ClearSpeed Technology"—teaches a SIMD device similar to Stuttard's SIMD Processor for use with image processing having "a processing element (PE) array" with "PE

interconnections" that move data "along the PE array (hereinafter swazzled)" (EX1081, 5:51-62) and discloses the array could have its "PE interconnection" arranged in "a 1D ring/line array **or a 2D grid array**." EX1081, 5:62-65, 8:40-47 (noting "[t]wo interconnection networks are generally used: a 1D array/ring and **a 2D grid**"). EX1081 further taught that "a 2D array with a two dimensional bi-directional (up, down, left, right) nearest-neighbor configuration, often known as a mesh" was a "**common** configuration" and sometimes "**preferable** to exploit the 2D connectivity." EX1081, 21:10-22. Thus, EX1108 and EX1081 demonstrate that using a 2D-array in Stuttard-Shirazi would have been using a known technique (2D-mesh) to improve similar devices (parallel processor having an array of PEs) in the same way (by facilitating applications that benefited from full-2D connectivity between PEs).

105. Applying the above-discussed principles to Stuttard-Shirazi would have resulted in a two-dimensional layout of poly PEs, an exemplary design of which is depicted below (right, which I adapted from Stuttard's Figure 7, left).

Stuttard, FIG. 7

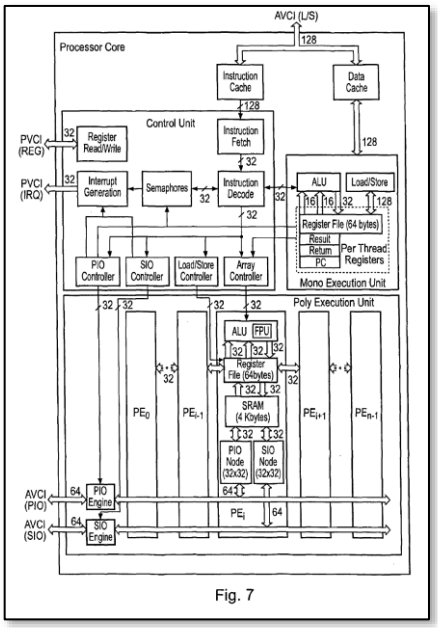
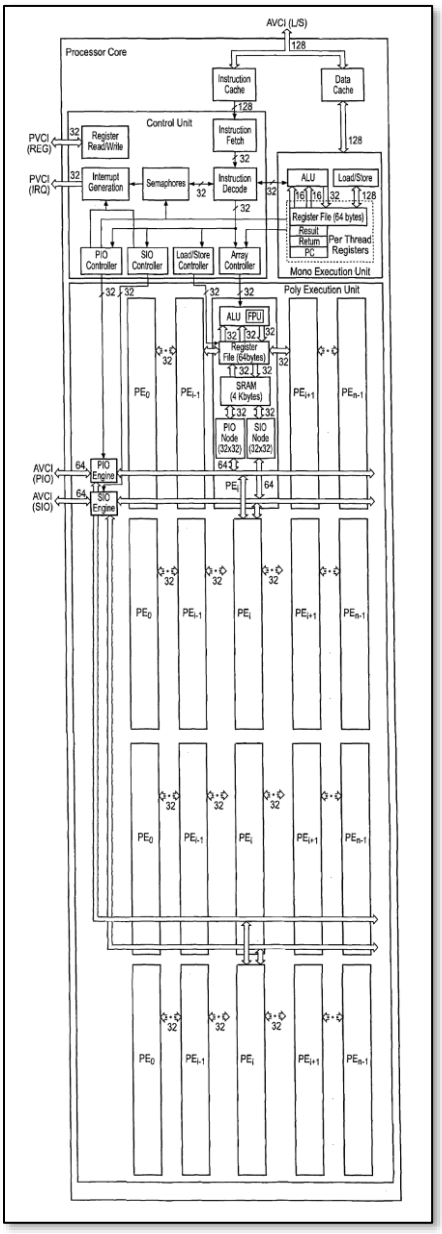


Fig. 7

Stuttard-Shirazi



B. Claim 1

1. [1PRE]: "A computing system, comprising:"

106. The Stuttard-Shirazi processor (§VII.A.3 above) would have been for "use[] in a number of **computing** applications" (Stuttard, 1:3-5) by "integration" into "a variety of **systems**" (Stuttard, 9:15-16). Data and instructions in such a

system would have been provided by a host computer. This is corroborated by Stuttard's disclosure at 1:21-24, which notes as background to the invention, that with SIMD processors (like Stuttard's) the "*host* processor…will send data to be processed to the SIMD array with an indication of the function to be performed, typically a pointer to the start of a sequence of SIMD array ***instructions*** or micro***instructions***."); e.g., Stuttard, Title ("Unified **SIMD Processor**"); §VII.C.2 below (describing, for limitation [3B], how the control unit on the Stuttard-Shirazi chip would have received instructions from the host computer). Thus, the Stuttard-Shirazi processor and host computer are integrated in a computing system, as claimed. Stuttard, 5:12-14 ("The SIMD part [poly unit] and the scalar part [mono unit] preferably have a common memory area **for instructions** and scalar data."), 13:21-25 ("Instruction set: ***The processor has a fairly standard RISC-like instruction set***. ***Most instructions can operate on mono or poly operands*** and are executed by the appropriate part of the execution unit.").

2. [1A]: "a host computer;"

107. The Stuttard-Shirazi processor (Section VI.A.3 above) communicates with a "host system" that includes a "host processor" that runs an "application or operating system." Stuttard, 21:13-31. Stuttard at 1:11-2:3 further discloses as background information for SIMD processors (like Stuttard's) that the SIMD PEs "are fed with instructions and data by a dedicated controller" that is "controlled by

a general purpose processor which also runs the bulk of the application software." *See also* Stuttard, 13:19-20 ("The other allows the MTAP processor to generate interrupts ***to the host system***."), 20:5-8 (describing interface for "transfer[ring] data to and from" "external memory…or a host processor"), 22:4-17 (describing an exemplary MTAP processor built as a "real device" that has an "***external port[]***" at each end of a bus that allows the device "to connect to…a host interface."); *see also* EX1089 (Microsoft Computer Dictionary), 2 ("Copyright © 2002 by Microsoft Corporation"), 3 (first definition of port: "***An interface through which data is transferred between a computer and other devices…or a direct connection to another computer***. The port appears to the CPU as one or more memory addresses that it can use to send or receive data. Specialized hardware, such as in an add-on circuit board, places data from the device in the memory addresses and sends data from the memory addresses to the device. Ports may also be dedicated solely to input or to output.").

108. POSAs would have understood this "host" to be a host computer, as claimed. Stuttard, 1:11-24 (teaching as background information that in "SIMD processor architectures," like Stuttard's, the "***host* processor**…will send data to be processed to the SIMD array with an indication of the function to be performed"); The '616 patent at 3:18-20 broadly states that a "computer[]" can be a "typical microprocessor" and at 29:27-36 states that a "computer includ[es] a processor and

other components (such as memory coupled to the processor by a data path)." The '616 patent at 8:12-9:18 also discloses that the "host is responsible for overall control." Consistent with Stuttard's teaching at (e.g., 1:11-24), POSAs understood Stuttard's host to at least include a host processor and memory and, as I discussed above, the host is responsible for overall control of the Stuttard-Shirazi processor, which meets the '616 patent's descriptions of a "host."

3. [1B]: "a computing chip comprising:"

109. The Stuttard-Shirazi processor would have been fabricated as a computing chip, following Stuttard's teachings that "illustrate the use of an MTAP processor in a real device" manufactured as an "evaluation **chip**." Stuttard, 22:3-7. Stuttard similarly notes (at 7:3-4) that "Figure 8 is a block diagram of a specific implementation of the MTAP architecture as applied to ***an evaluation chip***." Likewise, Stuttard at 22:12-19 (describes the evaluation chip's "ClearConnect bus" connecting "from one **chip** to another" and "on-**chip** SRAM."

4. [1C]: "a processing element array comprising" "first" and "second" "edge processing element[s] positioned at a first edge of the processing element array" and "first" and "second" "interior processing element[s] positioned at [first and second] location[s] in the interior of the processing element array"

110. Stuttard's "processing elements" (PEs) are in a SIMD "array." Stuttard, 3:7-9 ("***The processing elements in the SIMD array*** and the mono execution unit have a common instruction set."), 3:30-34 ("Having a full range of

addressing modes and completely separate memory blocks in each PE make it simple to support poly data structures and poly pointers ***on the SIMD array***…."), 5:29-30 ("Each processing element ***of the SIMD array*** may comprise one or more function units, a register file, said local memory, and I/O means."), 10:12-19 ("A data processor as claimed in any of claims 1, 3 or 5, ***wherein the execution unit comprises a SIMD part*** comprising an array of processing elements (PEs) adapted to operate on said parallel data and a scalar part adapted to operate on said scalar data."), claim 6 ("A data processor as claimed in any of claims 1, 3 or 5, wherein ***the execution unit comprises a SIMD part*** comprising an array of processing elements (PEs) adapted to operate on said parallel data and a scalar part adapted to operate on said scalar data."); *see* Section VII.B.10 below. As discussed (Section VII.A.3.b above) and illustrated below right, POSAs had reason to make the array two-dimensional in the Stuttard-Shirazi processor:

Stuttard, FIG. 7

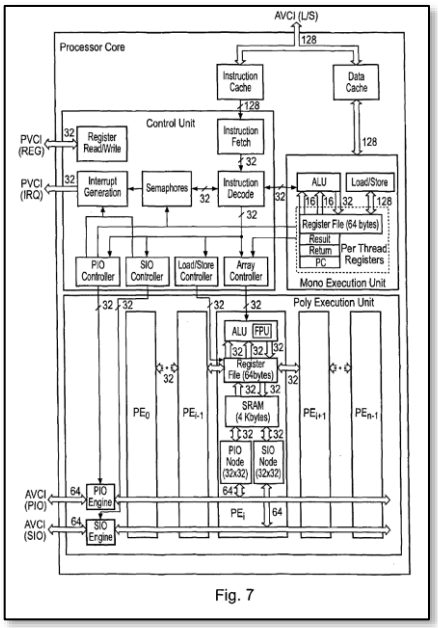
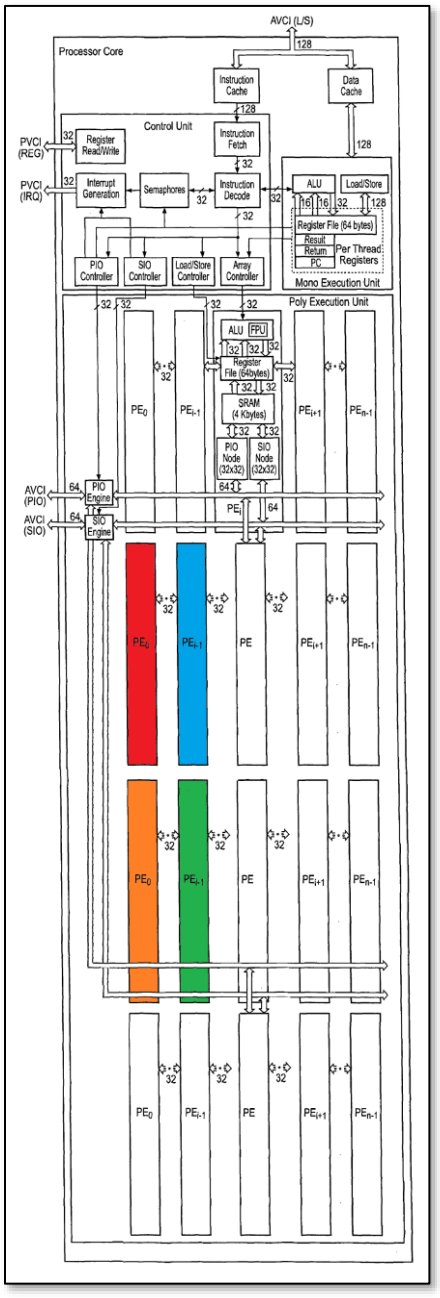


Fig. 7

Stuttard-Shirazi

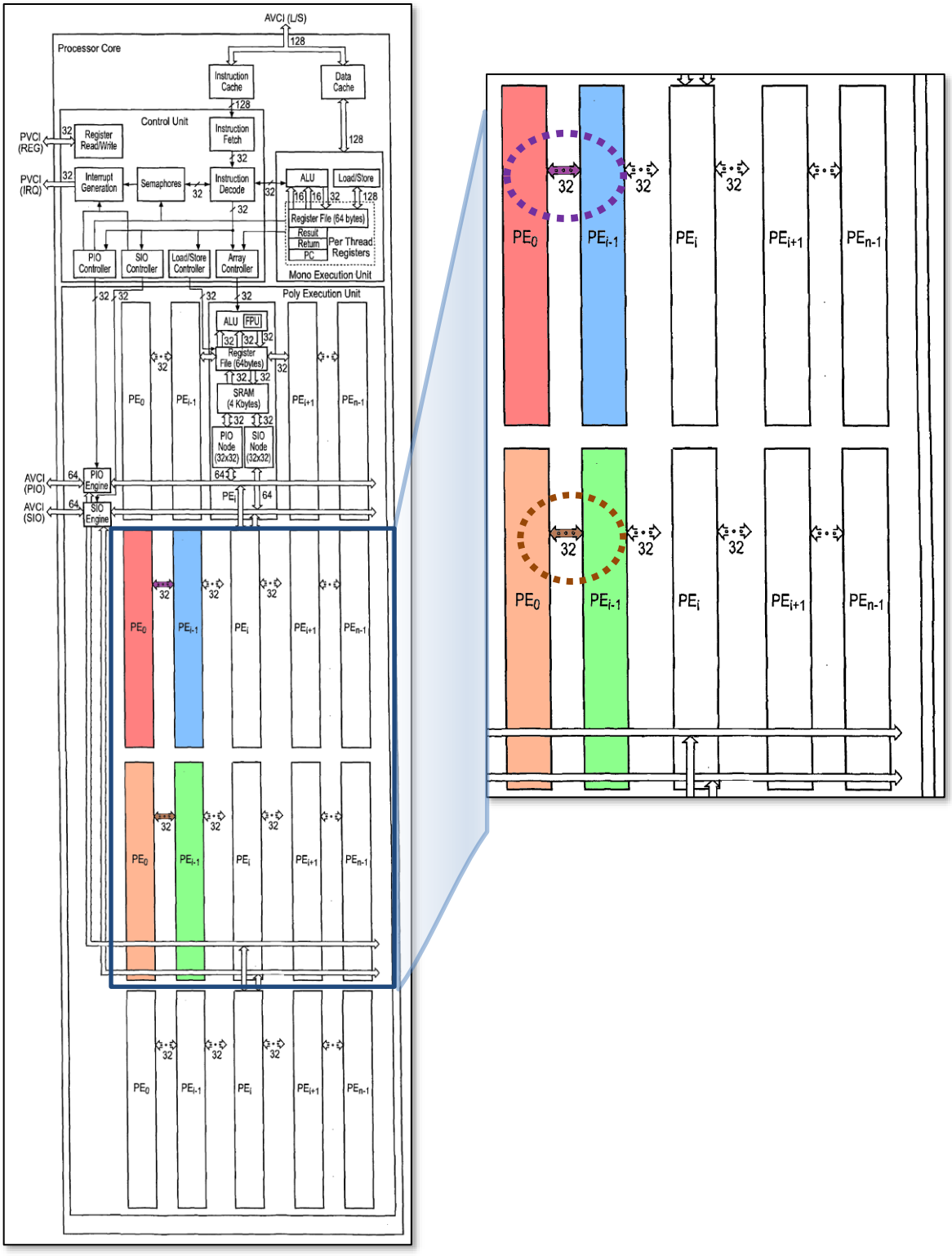


111. As annotated above (right), this obvious implementation of Stuttard-Shirazi has a first edge PE (red) and second edge PE (orange), both positioned at a first edge of the PE array (i.e., the left-hand edge of the 2D array common to these two PEs) and a first interior PE (blue) at a first location in the interior of the PE

array (i.e., PEs that are not along an outer edge of the array) and a second interior PE (green) located at a second location in the interior of the PE array, as claimed.

5. [1D]: "a first processing element connection connecting the first edge processing element with the first interior processing element; a second processing element connection connecting the second edge processing element with the second interior processing element;"

112. Stuttard's PEs in the Stuttard-Shirazi processor "communicate with one another via what is known as the swazzle path that **connects** the register file of **each PE with** the register files of **its left and right neighbours**." Stuttard, 20:31-33. These swazzle connections include a first PE connection (circled purple below) connecting the first edge PE (red) with the first interior PE (blue) and a second PE connection (circled brown) connecting the second edge PE (orange) to the second interior PE (green), as claimed. These same connections would have also existed in the alternative implementation of Stuttard-Shirazi I discussed above (§VII.A.3.b), where each PE's register is also connected to North and South neighbors.

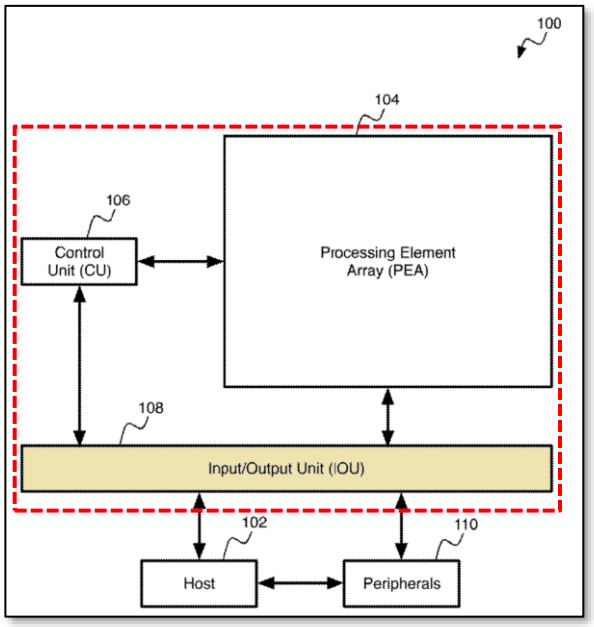

AVCI (L/S)
Processor Core
Instruction Cache
Data Cache
Control Unit
Instruction Fetch
Register Read/Write
Interrupt Generation
Semaphores
Instruction Decode
ALU
Load/Store
Register File (64 bytes)
Result
Return
PC
Per Thread Registers
PIO Controller
SIO Controller
Load/Store Controller
Array Controller
Mono Execution Unit
Poly Execution Unit
FPU
Register File (64bytes)
SRAM (4 Kbytes)
PIO Node (32x32)
SIO Node (32x32)
PIO Engine
SIO Engine
PVCI (REG)
PVCI (IRQ)
AVCI (PIO)
AVCI (SIO)
PE0
PEi-1
PEi
PEi+1
PEn-1

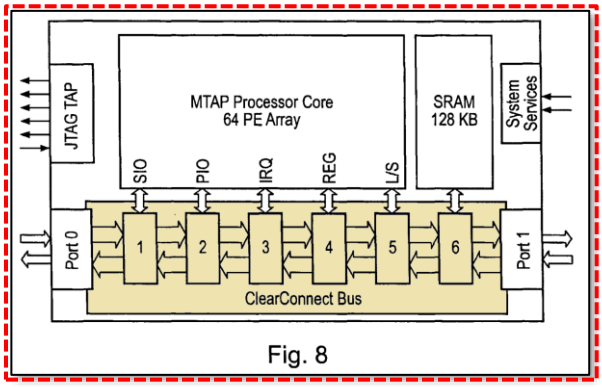
6. [1E]: "an input-output unit connected to the first" and "second edge processing element[s];"

113. The '616 patent's specification mentions an "I/O unit (IOU) 108" but does not limit it to being composed on any particular components. EX1001, 8:14-15 ("Sometimes present are a control unit (CU) 106, an I/O unit (IOU),…"). The specification merely states that "I/O could be routed…through the separate I/O unit," which "interfac[es] the CU [control unit] 106 and PEA [processing element array] 104 with the Host 102" to "get data into and out of the CU 106 and PEA 104." EX1001, 9:6-13, 9:24-28.

114. The '616 Patent's Figure 1 (annotated below left) illustrates the Input/Output Unit (IOU) connecting the on-chip Control Unit (CU) and Processing Element Array (PEA) to the off-chip Host and Peripherals. '616 Patent, 8:12-16 ("I/O unit (IOU) 108"). Stuttard's similar Figure 8 (annotated below right), illustrates "a block diagram…of the MTAP architecture as applied to an evaluation chip" (Stuttard, 7:3-4) which includes an on-chip "ClearConnect bus" with "two external ports at each end" that, like the '616 patent's I/O Unit, connect the chip to "peripherals…or a host interface." Stuttard, 22:3-17.

’616 Patent, FIG. 1 **Stuttard, FIG. 8**





115. As an aside, I note that Stuttard at 22:9 references “Figure 9,” but a POSA would recognize this is a typographical error meant to reference “Figure 8.” This is because Stuttard has only eight figures, and the description of the “evaluation chip EVI” at Stuttard, 22:3-19 corresponds with what is illustrated in Figure 8, which Stuttard at 7:3-4 says illustrates an “evaluation chip.” See, e.g., Stuttard, 22:3-19 (describing a figure that illustrates “five ports” of the MTAP processor (i.e., SIO, PIO, IRQ, REG, and L/S shown in Figure 8) and “one port of the embedded memory block (i.e., the port number 6 to “SRAM 128 KB” in Figure 8) and “two external ports at each end of the bus (i.e., “Port 0” and “Port 1” in Figure 8). I reviewed Stuttard’s prosecution history and it shows that other references to “Figure 9” in the application were corrected to refer to “Figure 8,”

but the error that appears at 22:9 appears to have been overlooked and left uncorrected (i.e., the specification continued to refer to a figure intended to illustrate "power efficiency" that does not exist). *See* EX1095 (Stuttard's PCT prosecution history), 32 (original application referring to "Figure 1" through "Figure 9" and describing "Figure 6" as "a graphical representation of the results of running an example program using the present invention" – i.e., what Stuttard's FIG. 6 illustrates – and describing "Figure 7" as "a graph indicating the power efficiency of the MTAP architecture" – i.e., Stuttard's description of FIG. 6, but is not what Stuttard's FIG. 6 illustrates); EX1095, 37 (lines 18-21) (stating "Figure 7 shows how performance and power scale with the number of PEs…," which is how Stuttard (incorrectly) describes FIG. 6 (*see* Stuttard, 12:18-19)); EX1095, 51-62 (only including eight figures; figure originally identified as "Fig. 7" (EX1095, 59) clearly showing "a block schematic diagram depicting an exemplary MTAP processor" ***not*** "a graph indicating the power efficiency of the MTAP processor"). In other words, what appears to have happened is that the applicant's failed to include a Figure 7 with "a graph indicating the power efficiency of the MTAP architecture" and rather than delete the description of Figure 7 in the Brief Description of the Drawings, the applications deleted the original brief description of Figure 6 (EX1095, 32), rather than the original brief description of Figure 7 (EX1095, 32) and shifted the original brief descriptions of Figures 7-9 down to

Figure 6-8, failing to recognize that they were now misdescribing Figure 6 as the figure they failed to include in the application. The applicants also failed to correct the reference to "Figure 9" on page and replace it with "Figure 8." In any event, for all the reasons discussed above, POSA's would have recognized that (1) Stuttard's reference at 22:9 to "Figure 9" is a typographical error meant to reference "Figure 8"; (2) that Stuttard's statement at 12:18 that "Figure 6 shows how performance and power scale with the number of PEs" is a typographical error that refers to a figure not included in the application; and (3) that Stuttard's brief description of "Figure 6" at 6:30-34 is a typographical error that refers to the same figure that was not included in the application and that what Stuttard's Figure 6 actually illustrates is "a graphical representation of the results of running an example program using the present invention," stated in Stuttard's original PCT application (EX1095, 32).

116. As I illustrate below with annotations, the MTAP Processor Core in Stuttard's Figure 8 contains a control unit and a PE Array, as illustrated, for example in Stuttard's Figure 7, which has a "Control Unit" (red) and array of PEs in the "Poly Execution Unit" (green).

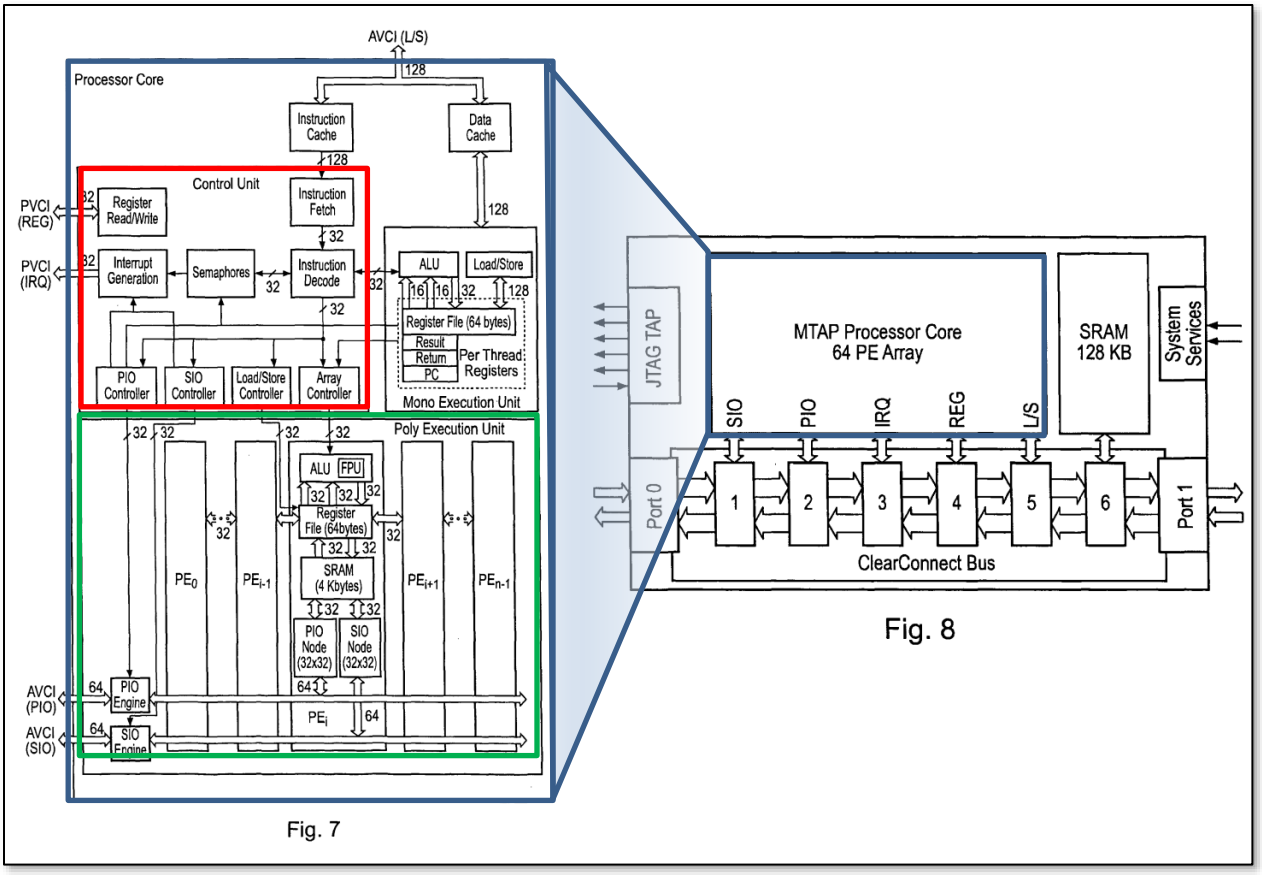


Fig. 7

Fig. 8

Thus, Stuttard teaches that in a real-world chip the ClearConnect bus connects a control unit and processing elements to off-chip peripherals and a host system (via the bus's "Port 0" and/or "Port 1"), just like the '616 Patent's IOU 108 in the '616 Patent's Figure 1 connects a Control Unit (CU) 106 and PEA 104 to a Host 102 and Peripherals 110, as illustrated a few paragraphs above.

117. As Stuttard's Figure 8 (above right) illustrates, the exemplary ClearConnect bus includes nodes 1-5 that connect to "five ports of the MTAP processor [core]." Stuttard, 22:9-12; *see also* EX1049, 19:22 ("The nodes are the connection points where data leaves or enters the bus."); Stuttard, 8:7-8 (disclosing

details for ClearConnect are in "copending patent application GB 0321186.9, which published as EX1049); *see* EX1051, codes (21), (87) (identifying EX1049 as the international publication of Stuttard's PCT application). In Figure 7 (above right), the ClearConnect nodes have corresponding virtual component interfaces (VCIs) to the processor core. Stuttard, 13:6-20, 21:20-22. Though different VCIs are shown in various locations in Figure 7's "conceptual view" (Stuttard, 13:1-5), POSAs understood that they all physically connect to the same ClearConnect Bus. This is corroborated, for example, by EX1082, which discloses (at 1:13-21) that "[o]n-chip buses are used to interconnect 'virtual' components" which "perform distinct functions ***but are not implemented as physically separate devices***." Instead, POSAs understood all of the "virtual" interfaces map to the same physical ClearConnect bus on the chip.

118. Stuttard discloses that VCIs are "defined by the VSI Alliance." Stuttard, 13:6-9. The VSI Alliance teaches that a VCI provides "a point-to-point connection between **two units**." EX1088, 5. *See* Appendix (regarding my opinions on public accessibility of EX1088 and evidence that VSI standards were available from VSI Alliance's website by 2002, as reflected by Internet Archive captures). As Stuttard Figure 8 illustrates, the two units connected by Stuttard's VCIs are the MTAP processor core and ClearConnect bus.

119. Stuttard's Figure 7 illustrates three AVCIs (Advanced VCIs) for the processor core: "Programmed I/O (PIO)" and "Streaming I/O (SIO)" (on the left-side of Figure 7) and "L/S" (load/store) (on the top-right of Figure 7). Stuttard, 19:12-20 (disclosing "two I/O mechanisms provided for transferring date between PE memory and **devices outside the MTAP core**": "*Programmed I/O* (PIO)" and "*Streaming I/O* (SIO)"), 13:6-16 (disclosing "a single *Advanced* VCI (AVCI) interface which is used for mono **loads and stores**, and for instruction fetching" (i.e., the "AVCI (L/S)" in Figure 7)). The PIO and SIO AVCIs are for PEs to exchange data with off-chip devices (e.g., the host). For example, the PIO is "for accessing memory external to the MTAP core," like accessing "variable sized data **by the PEs**." Stuttard, 10:20-23. Similarly, the SIO "allows chunks of contiguous data to be streamed **directly into the memory within PEs**." Stuttard, 10:23-25.

120. Stuttard's Figure 7 shows that these interfaces connect the ClearConnect bus unit to respective PIO/SIO "engines," which connect to every PE in the poly-execution unit array via the "channels" discussed above in this section. Stuttard, 19:22-31 ("The PIO and SIO controllers decode I/O instruction…The [PIO and SIO] I/O engines are basically DMA engines which manage the actual data transfer. There is a Controller and Engine for each I/O channel. A single Controller can manage several I/O Engines."), FIG. 7 ("PIO Node" and "SIO Node" exists in every poly PE). Exemplary connections from the

channels to a processing element in the Stuttard-Shirazi chip are circled below. Similar connections exist to every PE in the processor core. Stuttard, 20:1-4 ("There is an ***I/O node in each PE***" that is "activated" by an "I/O Engine"). Thus, the PIO/SIO AVCIs are input-output interfaces that connect to every PE in the processor core (via their engines and channels).

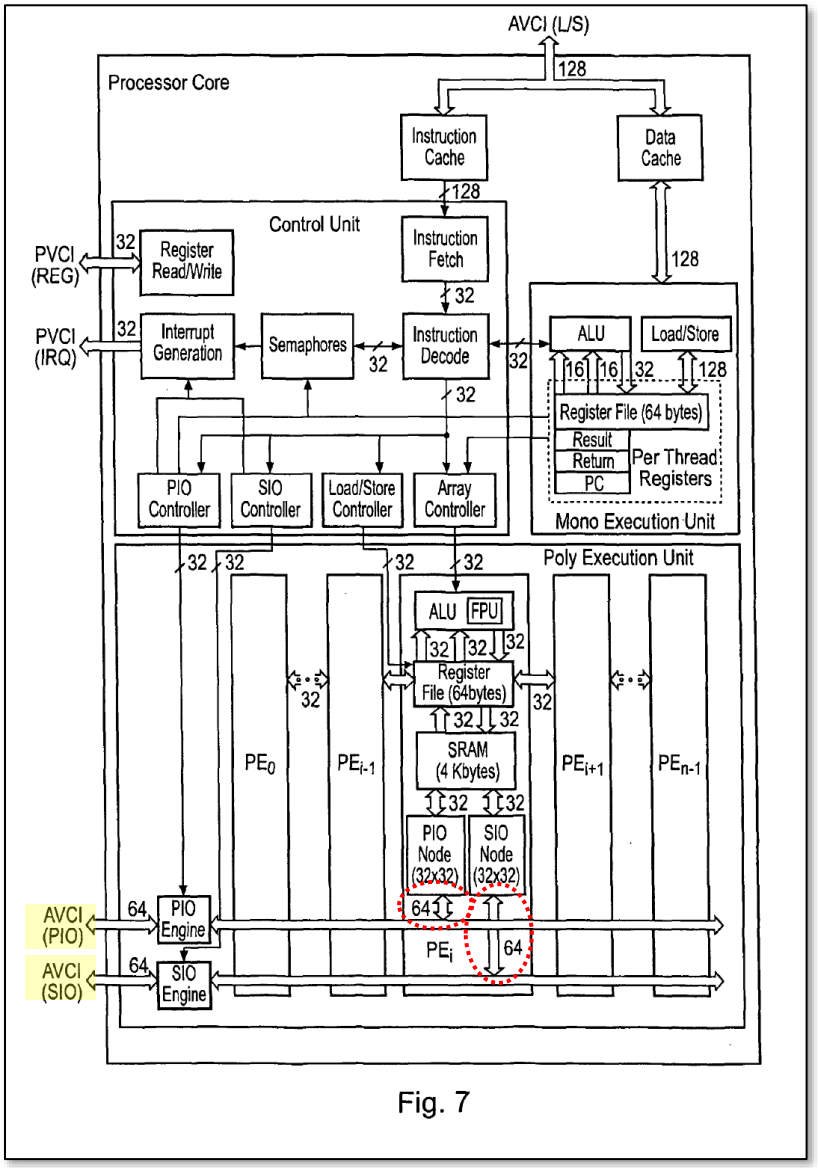


Fig. 7

121. The AVCI (L/S) (illustrated at the top of FIG. 7, though it can be physically located in other locations) is another input-output interface to the MTAP processor core: inputs via the L/S include instructions fetched from the host; outputs via the L/S include data from the mono unit. Stuttard, 13:6-13 ("There are three categories of interface: *Mono data & instructions*: This is a single Advanced VCI (AVCI) interface which is used for mono loads and stores, and for instruction fetching" (i.e., instructions for the entire processor, not just the mono unit)); 15:17-25 ("Load and store instructions are used to transfer data between memory and registers…. Poly loads and stores transfer data between the PE register file and PE memory. Data is transferred between the PEs and external memory using the l/0 functions described later."); 9:17-21, FIG. 5. As Figure 7 illustrates, the AVCI (L/S) connects every PE in the poly unit to the host: e.g., instructions from the host pass through the L/S port into the control unit (via the "Instruction Cache"), then to the PIO/SIO engines (via the PIO and SIO Controllers), which send the instructions via the channels connected to every PE in the array (as discussed above). Stuttard, 14:29-34 (disclosing that "Instructions…are provided in several versions" that vary in "sizes," "data types," etc.), 15:17-25, 17:20-23 ("**Load and store instructions** [i.e., L/S] move data between a PE's register file and memory, while the ALU operates on data in the register file. Data is transferred in to, and out of, the PE's memory using I/0 instructions.").

122. Thus, the PIO, SIO, and L/S AVCIs and their respective ports on the ClearConnect bus provide connections between every PE in the poly array and the ClearConnect bus. As discussed above, the ClearConnect connects the MTAP processor to, for example, a host interface. Stuttard, 22:3-17. The ClearConnect bus (or alternatively, the bus plus the I/O engines and channels), therefore, is an input-output unit connected to the first and second edge PEs as claimed (because it is connected to all the poly PEs, including the first and second edge PEs discussed above §VII.B.4).

7. [1F]: Each Processing Element of Limitation [1C] Has a "memory local to" It

123. Limitation [1F] generally recites memory local to each of limitation [1C]'s four PEs that I discussed in Section VII.B.4 above. Stuttard's Figure 4 (which I have annotated below) illustrates that "[e]ach PE [] has its own independent **local memory**" and an addressable "register file," each of which is memory local to the PE, as claimed. Stuttard, 7:14-16, 5:29-30, 10:1-4, 10:14-19, 17:12-23, FIGS. 3-4, 7. POSAs also understood a register file is an array of registers that is ***addressable*** as a unit of memory. For evidence corroborating that understanding, see EX1058 at [0003] teaches that a "'register file' is an array of registers that may be ***addressed as a unit***" and EX1075 at 2:14-25, which teaches that a "register file is ***a memory array***" and that there is an "***address word*** for the register file" (i.e., the register file is addressable as a unit, as taught in EX1058).

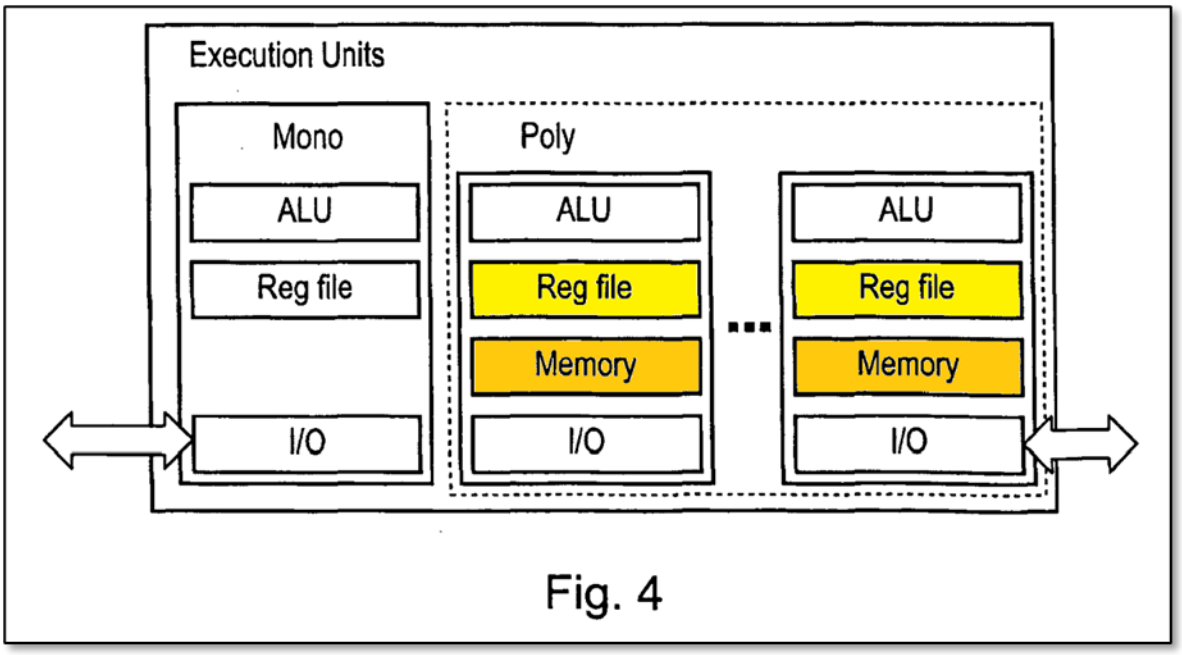


Fig. 4

8. [1G]: Each Processing Element of Limitation [1C] Comprises an "arithmetic unit"

124. Limitation [1G] recites that each of limitation [1C]'s PEs – i.e., the reduced-bitwidth PEs in the Stuttard-Shirazi poly unit – comprises a respective arithmetic unit. As Stuttard's Figure 4 (annotated below) illustrates, "**[e]ach PE**" comprises "an ALU" (**arithmetic** logic **unit)**. Stuttard, 10:1-4, 10:14-19 (disclosing that "data [is] being processed" by each PE), 5:28-30 ("***Each processing element*** of the SIMD array may comprise one or more ***function units***, a register file, said local memory, and I/O means."), 7:14-16 ("***each PE has an ALU***, a register file and memory"), 10:1-4 ("***Each PE consists of an ALU***, registers, memory and I/O."), 17:12-23 ("***Each PE*** in the poly execution unit

consists of: • An 8-bit ALU with optional extensions such as a multiply-accumulate (MAC) unit…"), Claim 9-10, FIGS. 3-4, 7.

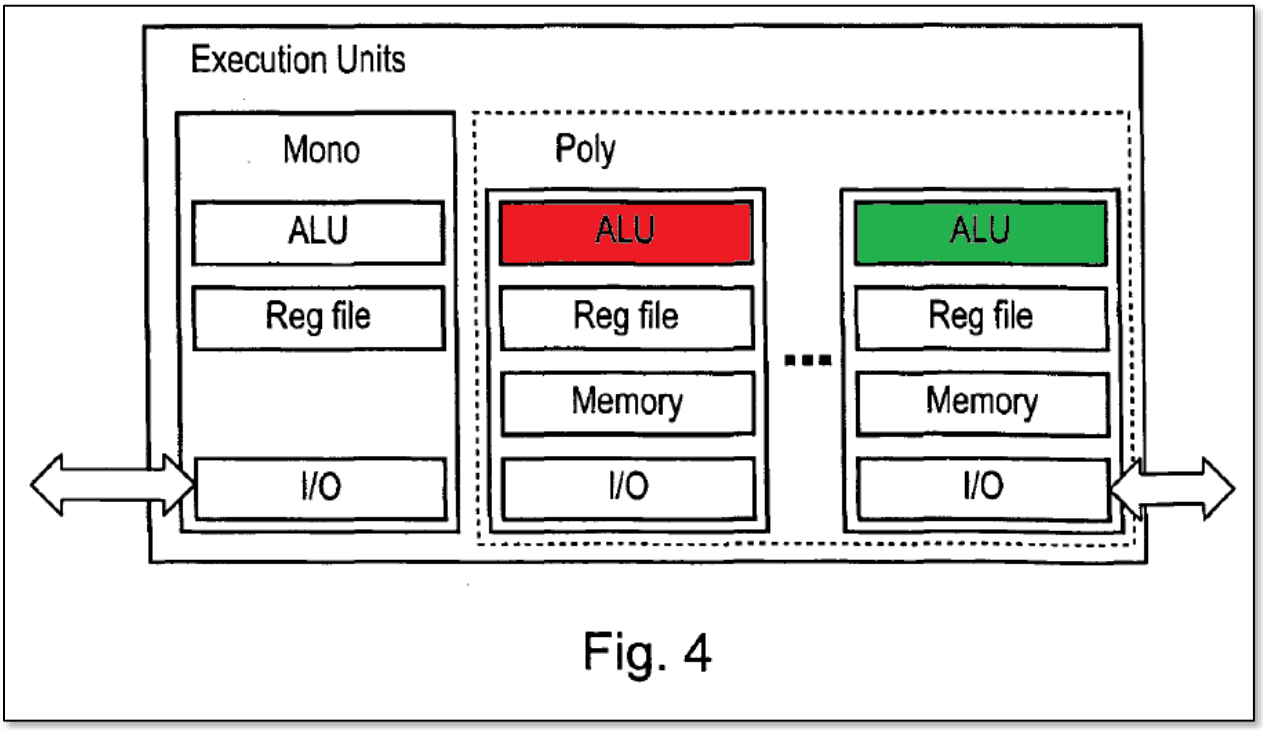


Fig. 4

Thus, limitation [1C]'s two edge and two interior PEs (which I discussed in Section VII.B.4 above) each comprise an ALU, meeting the claimed first through fourth arithmetic units.

9. [1H]: "a host connection at least partially connecting the input-output unit with the host computer;"

125. As I discussed above in Section VII.B.6, POSA would have understood the input-output unit is the ClearConnect bus or, alternatively, the bus plus I/O engines and channels. As I also discussed in that section, the ClearConnect bus has several on-chip nodes (e.g., 1←→SIO, 2←→PIO) that are

connected to the processor core and "external ports" (Ports 0 and 1 annotated below) that are "connect[ed] to an FPGA [field programmable gate array] to provide other functions, such as…a **host interface**." Stuttard, 22:8-19, 21:13-15 (processor has "interfaces to the host system…").

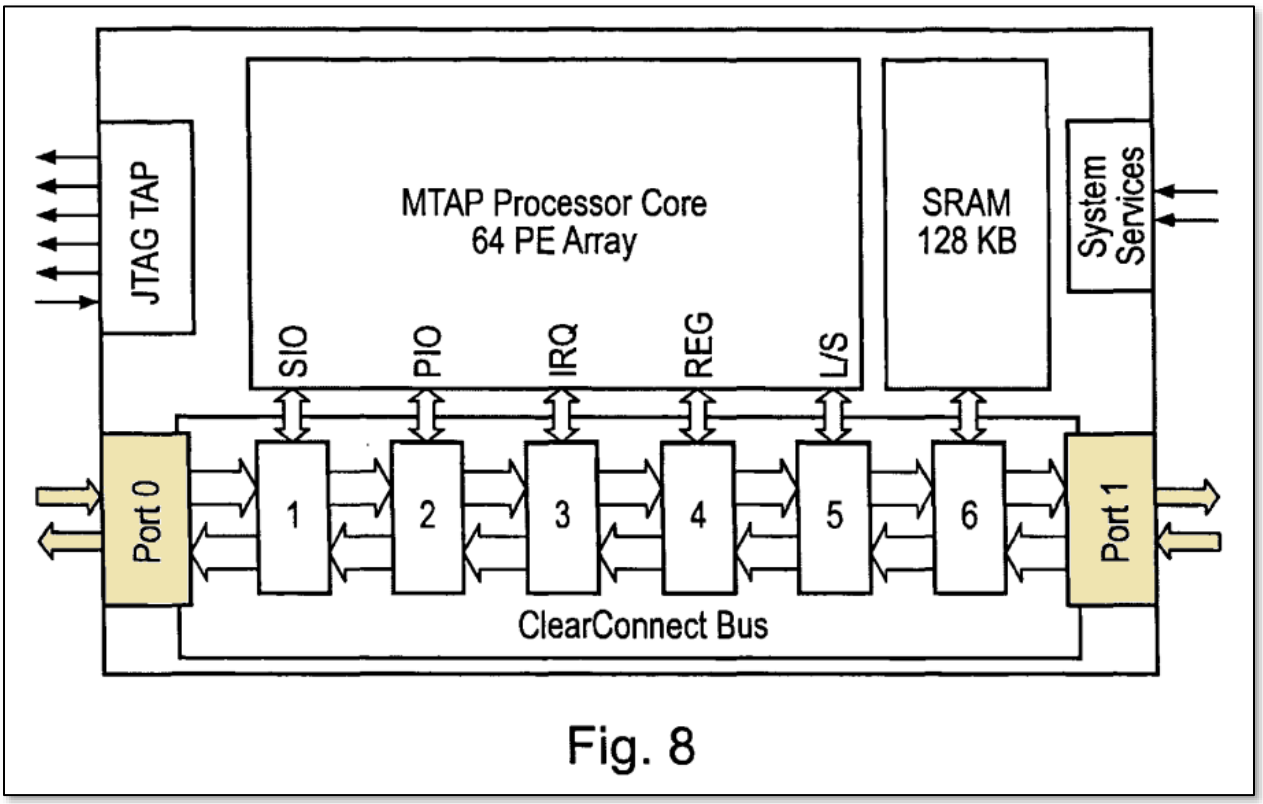


Fig. 8

126. The PIO node ("2" in FIG. 8) is used "to transfer data to and from the outside world," including to/from "a host processor." Stuttard, 20:5-9 ("PIO is closely coupled to program execution and is the normal way for the processor to ***transfer data to and from the outside world*** (e.g. external memory, hardware accelerators ***or a host processor***)."), 20:22-30 ("SIO is used for streaming high-bandwidth data directly to and from PE memory. This is less flexible but very

efficient ***for transferring blocks of data in and out of the system***."), 17:22-23 ("Data is transferred between the PEs ***and external memory*** using the I/O functions described later."), 19:12-17 ("There are two I/O mechanisms provided for transferring data ***between PE memory and devices outside the MTAP core***."). Similarly, the L/S (load/store) node ("5" in FIG. 8) provides an "external interface[]" for "instruction fetching." Stuttard, 13:6-13.

127. POSAs also understood it was conventional for a host system that provides data (as in Stuttard) to also provide instructions for processing that data. For example, Stuttard notes as background information that SIMD co-processors, of which Stuttard is an example (*see, e.g.*, Stuttard, Title, Abstract), conventionally receive data and instructions from a "general purpose [host] processor." Stuttard, 1:16-24.

128. All of the above communications with the host occur via the ClearConnect bus's external ports (i.e., one or both of "Port 0" or "Port 1" in FIG. 8). *See* Stuttard, 22:8-19 ("The two ***external ports*** at each end of the bus allow the ClearConnect bus to be connected from one chip to another…. The bus ports can also be used to connect to… peripherals, an external memory or a host interface.") *compare* '616 Patent, FIG. 1 (similarly illustrating the IUO 108 connecting the Control Unit 106 and Processing Element Array 104 to off-chip Host 102 and Peripherals 110).

129. Thus, the external ports connecting the on-chip bus (input/output unit) to the host system (providing the "host interface" that Stuttard describes—Stuttard, 22:8-19, 21:13-15) are one or more host connections at least partially connecting the input-output unit (including the bus) with the host computer, as claimed.

10. [1I]: "wherein the first, second, third and fourth arithmetic units each comprises a corresponding multiplier circuit adapted to receive as a first input to the corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits;"

130. As I discussed above in Section VII.B.8, each PE in the poly unit includes an ALU (a claimed arithmetic unit). As shown in Stuttard's Figure 7 (which I have annotated below), each ALU has an FPU (floating-point unit), which comprises circuits for performing floating-point operations, including floating-point multiplication. For example, Stuttard (at 12:7-15) discloses the performance of an exemplary MTAP core where each "PE [has] an FPU [floating-point unit] extension," which achieves 51,200 "MFLOPS" (**M**ega, i.e., million, **FL**oating point **O**perations **P**er **S**econd"), 13:21-28 (disclosing instructions for "floating point adds, subtracts, **multipliers**, divides" for "both the mono and **poly execution units**"), 17:24-33 (ALUs can include "extensions" "to accelerate functions for floating point").

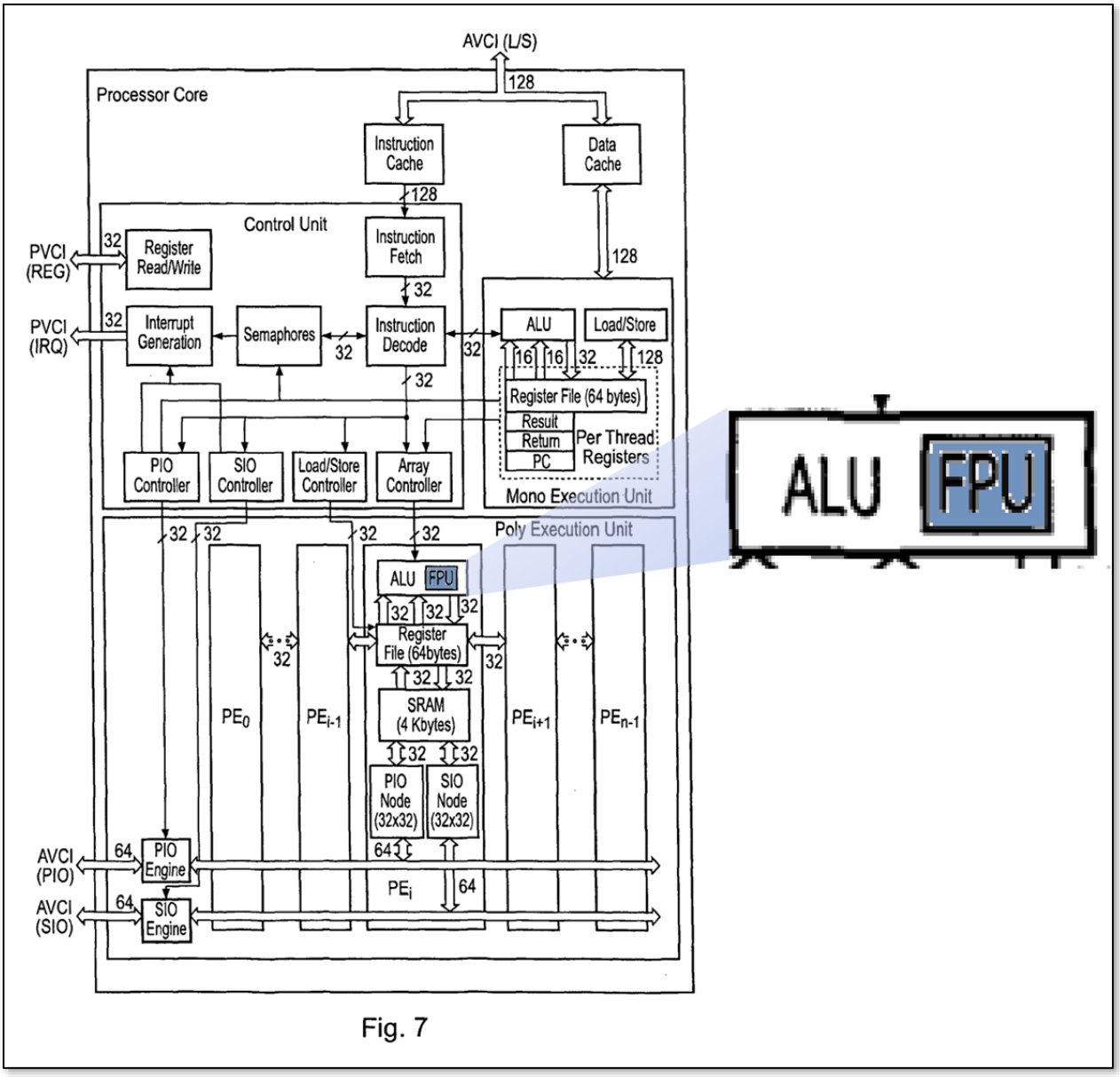


Fig. 7

131. Stuttard further discloses that the mono unit (discussed further in Section VIII.C.1 below) "include[s] a multiplier" for floating-point operations. Stuttard, 16:13-20. POSAs understood that, like the FPU in the mono-execution unit, each FPU in the poly-execution would have also included a multiplier circuit for performing the floating-point multiplication that Stuttard describes both the

mono unit and the poly PEs perform. A multiplier circuit would have also been the conventional way to perform multiplication in an FPU. For evidence corroborating that, by the 2000s, floating point operations were conventional performed by multiplier circuits, EX1109 discloses (at 1:17-36) that in the prior art "[a] ***floating point type multiplier circuit*** is constructed of a fixed point multiplier for exclusive use of mantissa arithmetic and an adder for exclusive use of exponent arithmetic."

132. In the Stuttard-Shirazi processor, the multiplier circuit in each poly PE is adapted to receive as inputs two of Shirazi's reduced-bitwidth floating-point values, as I discussed above Section VII.A.3.a above. In one implementation using Shirazi's 18-bit format (FIG. 2 annotated below), each input to the multiplier circuit would be adapted to receive a 7-bit-wide exponent, a 10-bit-wide mantissa, and one sign bit; in another implementation using Shirazi's 16-bit format (FIG. 3 annotated below), each input to the multiplier circuit would be adapted to receive a 6-bit-wide exponent, a 9-bit-wide mantissa, and one sign bit; as illustrated in Shirazi's Figures 2-3 (which I have annotated below):

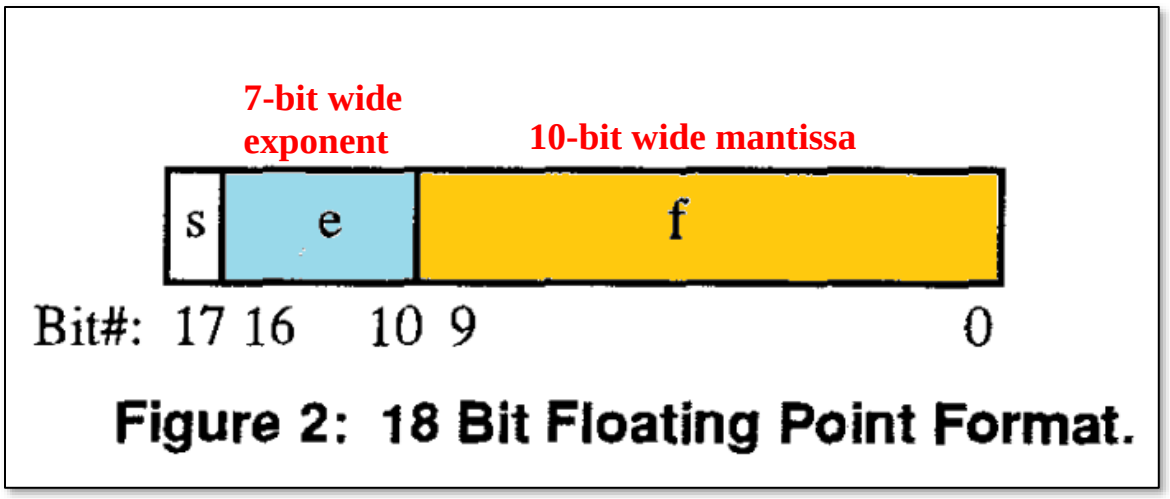


Figure 2: 18 Bit Floating Point Format.

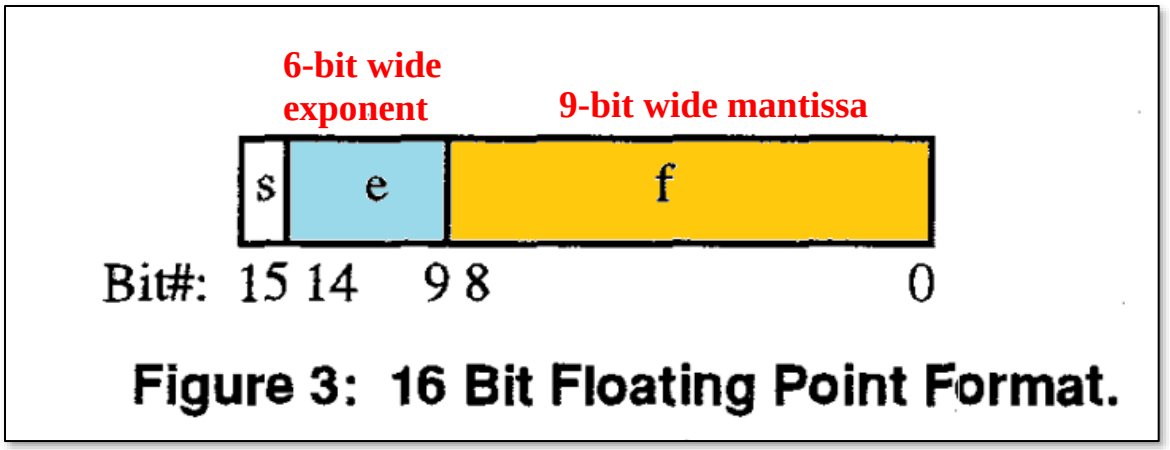


Figure 3: 16 Bit Floating Point Format.

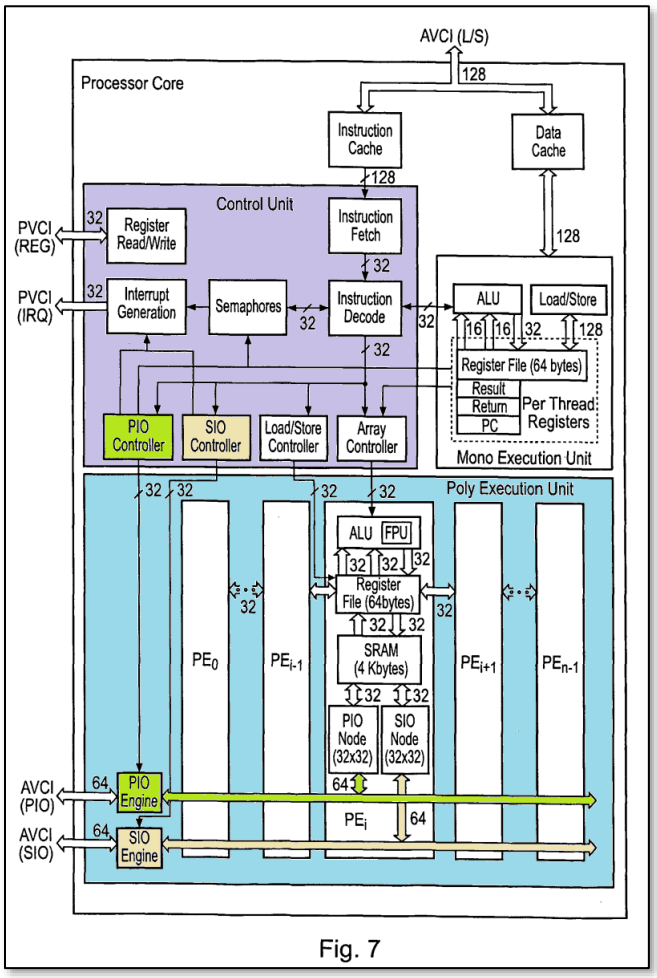
Thus, each multiplier circuit in the poly unit (including the first through fourth arithmetic units that I discussed above in this section) in the Stuttard-Shirazi processor is adapted to receive as inputs two values having respective first and second binary exponents of width at least 6 bits and respective first and second binary mantissas of width no more than 11 bits, as claimed.

133. Stuttard-Shirazi thus renders obvious claim 1 of the ’616 patent for the reasons discussed above.

C. Claim 2

1. [2A]: "the computing chip further comprises a control unit connected to the processing element array,"

134. As Stuttard's Figure 7 (which I have annotated below) illustrates, the Stuttard-Shirazi processor core comprises a control unit (purple). Stuttard, 4:20-5:28, 14:8-10, 19:22-27, 21:20-22, Claims 1-4, FIG. 3. The control unit is connected to the poly PE array (blue), as claimed, via, for example, the PIO (green) and SIO (yellow) controllers, engines, and channels. Stuttard, 14:8-26, 19:22-27 ("PIO and SIO controllers…coordinate with the rest of the control unit").



2. [2B]: "the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and"

135. The "control unit **fetches instruction** from memory [and] **decodes them**." Stuttard, 14:9-10, 9:17-18 ("The control unit fetches, decodes and dispatches instructions to the execution units. The processor executes a fairly standard, three operand instruction set."). The PIO and SIO controllers in the control unit similarly "**decode I/O instructions**" for the PE array. Stuttard, 19:25-26. POSAs would have understood these decoding operations to be performed by the circuitry in these units that is adapted to perform the decoding. Thus, the control unit comprises circuitry adapted to decode at least one instruction, as claimed.

136. POSAs would have understood that instructions decoded by the control unit are received from the host computer via the AVCI (L/S) interface to the ClearConnect bus (i.e., the input-output unit, Section VII.B.6 above). Stuttard, 13:11-13 ("single *Advanced* VCI (ACI) interface which is used…for instruction fetching"). Stuttard also teaches as background information that with SIMD arrays, like Stuttard-Shirazi's PE array, the "host processor" will "send data to be processed…with an indication of the function to be performed, typically a pointer to the start of a sequence of SIMD **array instructions** or micro**instructions**" (i.e., the host computer provides the instructions to be decoded). Stuttard, 1:21-24. The

electrical pathway for instructions received via the AVCI (L/S) into the Control Unit's "Instruction Decode" block is shown in Stuttard's Figure 7 (which I have annotated below). For example, instructions retrieved via the AVCI (L/S) are loaded to the "Instruction Cache," and when the control unit is ready to dispatch instructions, it will fetch the instructions from the cache, using the "Instruction Fetch" and then decode the instructions in the "Instruction Decode." Stuttard, 9:6-33, 13:6-20, 14:23-26 ("By *pre-fetching* data in the 1/0 thread, the programmer (or the compiler) can ensure the data is available when it is required by the execution units and the processor can run without stalling.").

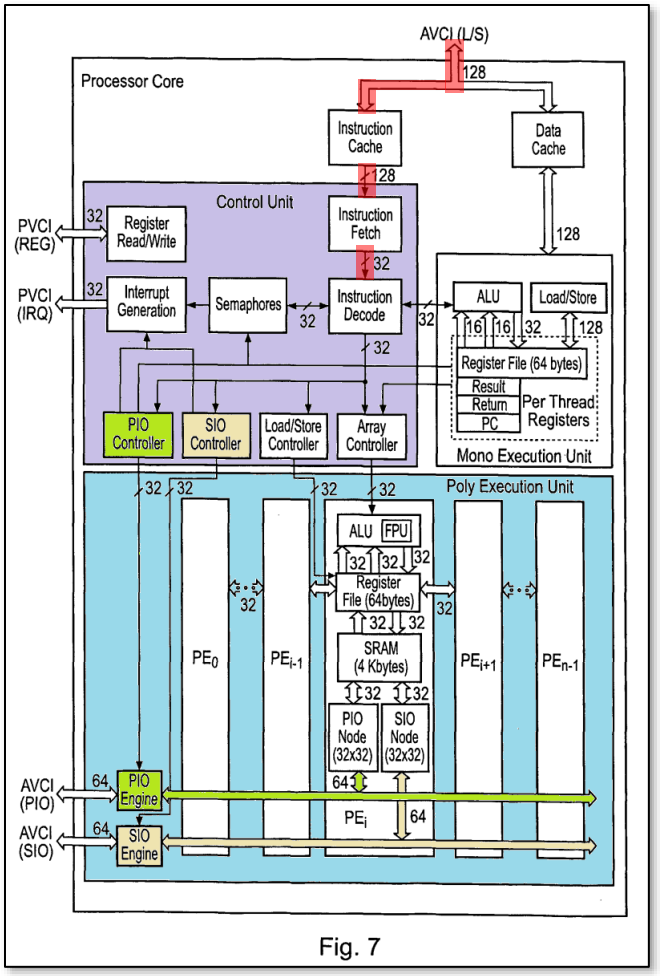


Fig. 7

3. [2C]: "to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction."

137. After the control unit "decodes the[]" instruction(s), it "despatches them to the appropriate functional unit," including the "poly execution unit." Stuttard, 14:8-10 ("The control unit fetches instructions from memory, decodes them and despatches them to the appropriate functional unit: this may be the mono or poly execution unit or one of the I/O controllers."); *see also* Stuttard, 9:17-18. POSAs would understand the "despatche[d]" instructions comprise signals that

cause the PEs in the poly unit to operate according to these instruction(s), as claimed (e.g., signals causing PEs to perform instructions for "floating point adds, subtracts, multipliers divides"). Stuttard, 13:21-14:26. The instructions "despatche[d]" by the control unit thus include control signals, as claimed.

138. Stuttard-Shirazi thus renders obvious claim 2 of the '616 patent for the reasons discussed above.

D. Claim 3: "the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory."

139. As Stuttard's Figure 7 (which I have annotated below) illustrates, the Stuttard-Shirazi processor "include[s] instruction and data caches to minimize the latency of memory accesses." Stuttard, 9:6-11, 9:24-28, 22:20-27.

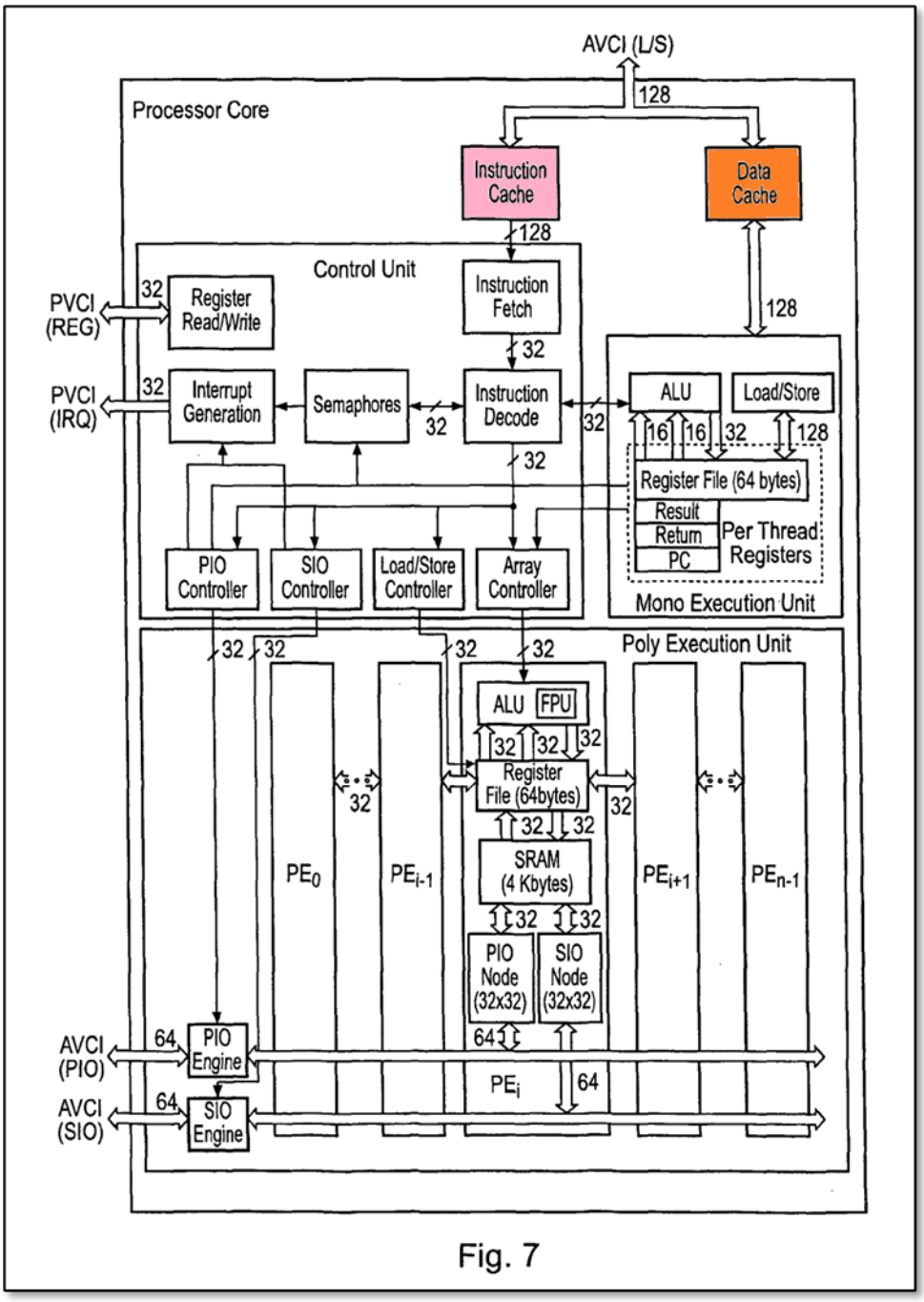


Fig. 7

140. Stuttard's FIG. 3 (which I have annotated below) similarly illustrates a "Cache" that feeds an "Instruction Fetch / Decode" block.

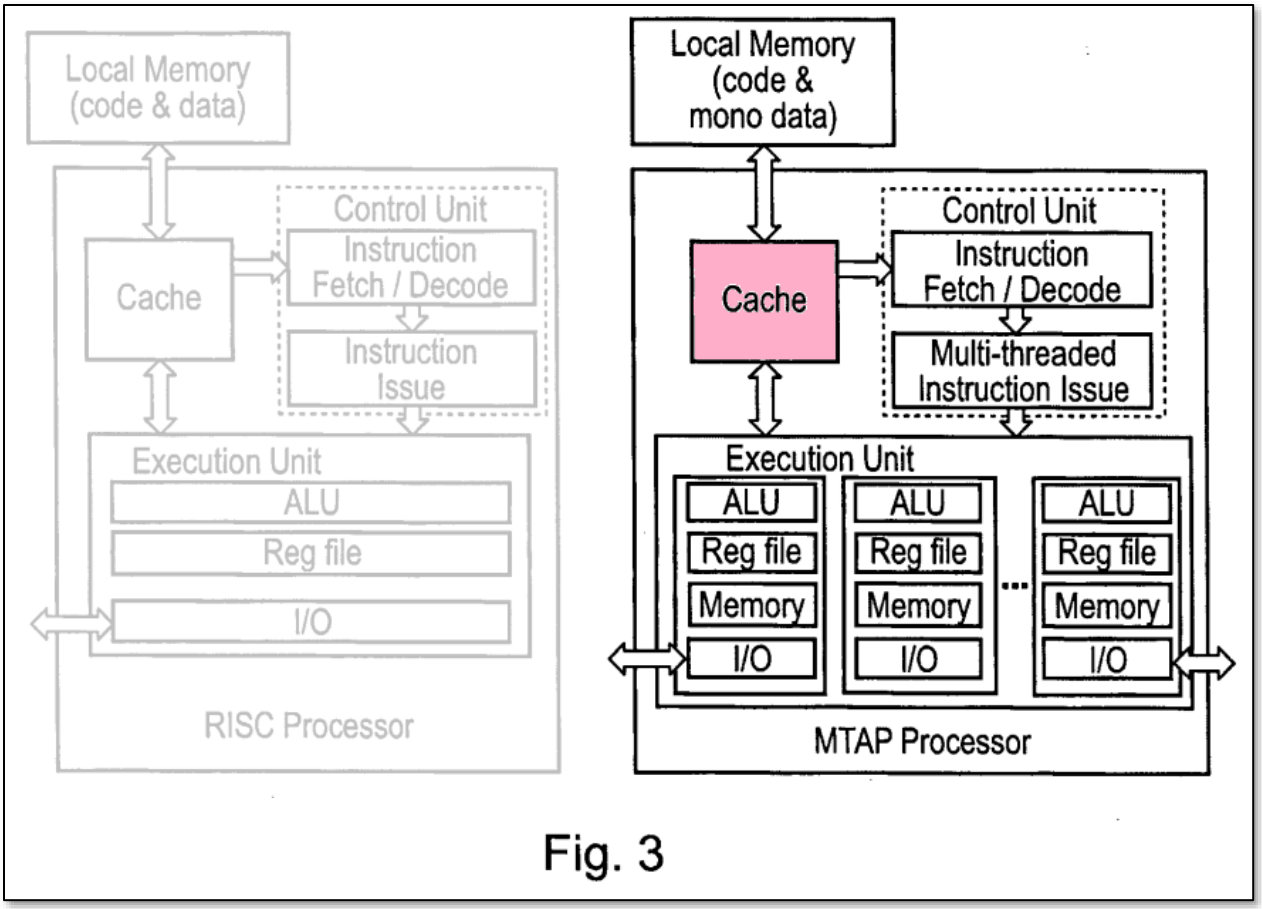


Fig. 3

141. POSAs would understand Stuttard teaches latency is minimized when the on-chip control unit retrieves instructions from the instruction cache, rather than from the off-chip host computer, because on-chip communication to an instruction cache is generally faster than off-chip communication to the host system. For example, Stuttard at 22:16-19 teaches that "**on-chip** SRAM…**provides fast access** to code and data."

142. Thus, POSAs would understand that the "instruction cache" is memory adapted to store the at least one instruction discussed above (Section

VII.C.2) that is received from the host computer via the input-output unit (e.g., ClearConnect bus), as claimed.

143. Stuttard-Shirazi thus renders obvious claim 3 of the '616 patent for the reasons discussed above.

E. Claim 4: "said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image."

144. As I discussed above in Section VII.B.9, the Stuttard-Shirazi host computer is programmed to provide instructions that, when executed, cause the Stuttard-Shirazi processor to perform data-processing operations. Stuttard's processor architecture is suitable for accelerating "a number of computing applications." Stuttard, 1:1-6, 8:26-28, 23:17-24:21 (describing two exemplary applications: "Network processing" and "Bio-informatics"). One application POSAs had reasons to accelerate using the Stuttard-Shirazi processor is an image search-and-identification application.

145. Stuttard's processor is for "applications [that] are characterized by very high data rates." Stuttard, 8:24-9:5. POSAs knew that image processing applications like image-identification often involved large amounts of data and generally required very high data rates and were advantageously performed with parallel processing elements, like the PEs in Stuttard's SIMD processor. For

evidence corroborating that POSAs understood image-identification applications to be well-suited to SIMD processor's like Stuttard's, EX1063 at [0023]-[0024] teaches that "image classification systems" require "raw processing power" and "[m]any operations in image processing are fundamentally Single Instruction Multi Data (SIMD)." EX1063 (at [0080]) similarly teaches that "the primate visual cortex [e.g., monkeys but also humans] operates in a retinotopic fashion, which is an inherently SIMD type of processing…." Similarly, EX1111 (at 1-2) teaches that "Content Based Image Retrieval (CBIR) is a technique that allows one to find out images of a database that are (at least) partly similar to a given reference image" and "often require[s] a low response time" (i.e., fast processing speed and high data rates) and, as a result, "a parallel machine [like Stuttard's] is the most natural choice" for "CBIR systems." *See* Appendix (discussing public accessibility of EX1111). As additional corroboration, EX1103 (at 1101) teaches that "pattern classification" "is time devouring" and benefits from parallel processors like GPUs (which are conventionally implemented as SIMD processors). *See also* Stuttard, Title ("Unified ***SIMD Processor***"), 3:7-9 ("The processing elements in the ***SIMD array*** …."), 4:25-27 ("[T]he execution unit comprises ***a SIMD part comprising an array of processing elements adapted to operate on said parallel data***…."). Thus, POSAs knew that Stuttard-Shirazi's massively parallel PE array would have been particularly well-adapted to image-identification applications.

146. As further corroboration that Stuttard's processor was well-suited for image processing applications, including image retrieval and pattern matching applications, a contemporaneous patent application by Suttard's Applicant (EX1050) taught using parallel processors like Stuttard's (8:19-25) to perform "image comparison" applications, like "Optical Character Recognition (OCR)" and "Identify Friend or Foe (IFF) systems," with "rapid **comparison of an image with a 'target**'" using sum of square of difference (SSD) computations (18:1-8). *See also id.*, 3:30-4:6 (noting SSD requires multiplications). Similarly, EX1107 discloses a similar SIMD processor (at 822-823) and notes (at 825-826) that there are "many application areas for this type of computing, including: "Image Processing" and "Search Engine" and "Supercomputing Applications"). Other well-known image-identification applications included fingerprint identification and image-database search and retrieval. For corroborating evidence in that regard, see, for example, EX1077 at 1:55-60 ("In pattern recognition, it is often desired to compare elements of an IP to many RPs. For example,…a finger print…to a library of RP (all scan results on file)") and 7:6-12 (teaching that "possible applications that can benefit from the capabilities of PPDE [parallel pattern matching engine] 100's high performance pattern matching are: … database reach image processing…." Additional corroboration can be found in EX1078, which teaches at 3:15-24 an image retrieval application by querying

images or "image portions" and teaches at 13:28-52 searching for images performed computing "the dot product" (a multiplication) that is repeatedly performed, and thus would have been well-suited to Stuttard-Shirazi's PE array in the poly unit.

147. These applications conventionally required identifying at least one image, from a plurality of images (e.g., fingerprint database, or other set of images being searched), that is similar to one input image (e.g., a fingerprint of a person to be identified, or a photo of a person or object of which other images are being searched for), as claimed. For evidence corroborating see EX1077 (at 1:55-60, 2:19-21, 3:58-60, 7:6-12, 10:35-53) and EX1078 (at Abstract, 1:35-38, 2:19-21, 3:15-24, 4:33-56, 5:6-8:4, 13:1-15, 13:28-52), which I discussed above. Similarly, EX1104 at 81-82 teaches that the known benefit of "face recognition" applications achieving "performance increase[s] through the use of massively parallel data processing," e.g., GPUs with "SIMD" "architecture." As I discussed above, Stuttard has a SIMD architecture and thus EX1104 (and the other prior art I discussed above) corroborate that POSAs had reason to use Stuttard's processor for various image-retrieval applications, like face recognition, fingerpint analysis, and CBIR applications.

148. Moreover, POSAs understood the additional benefits would have been provided by Stuttard-Shirazi's reduced-bitwidth PEs because they would have

produced accurate results for image-processing applications, while using less memory and processing power compared to a computing chip with an equal number of full-precision PEs. POSAs understood that the human visual system is not capable of detecting very small differences in color or brightness, so using fewer precision bits can still produce visually accurate results without a noticeable loss of quality. *See, e.g.*, Tong, 273 (teaching that "[i]t has long been known that many… applications can get by with less precision," such as "image processing"); EX1125, 171-174 (teaching that determining image moments is useful in pattern-recognition applications and that "[e]xperimental results show" that using floating point numbers with "8 bits for the exponent (*e*) and 10 bits used for the fractional part (*f*) of the mantissa" were sufficient); *see also* §VII.A.3.a above. Moreover, as I discuss in Section VIII.A.1 below, a POSA understood that, all other things being equal, they could fit more reduced-bitwidth PEs on a computing chip compared to using full-precision PEs. POSAs would have understood increasing the number of PEs on a chip would have been particularly useful for real-time image processing applications where speed is a critical factor. *See, e.g.*, EX1110, 607 (teaching that "time is critical in on-line CBIR systems as the response time requires [sic] to be low for good interactivity") (citing support); EX1111, 1 ("most applications CBIR…often require a low response time" and have to search large databases (e.g., "a 30,000 image database")).

149. Accordingly, POSAs would have readily understood the suitability and desirability of using Stuttard-Shirazi's massively parallel PE array to identify an image, from a plurality of images to be searched, that is similar to an input image.

150. As I have also demonstrated above with corroboration, using a massively parallel processor like Stuttard-Shirazi to perform image-matching applications would have also been applying a known image-matching technique (identifying an image from a plurality of images) to improve a similar massively parallel-processor chip like Stuttard-Shirazi in the same way (by supporting image-matching applications).

151. In my opinion, POSAs would have reasonably expected success using the Stuttard-Shirazi processor to perform these image-identification applications because, as noted above, Stuttard's processor is for "very high data rate[]" applications (8:24-28) using a "highly parallel execution unit" that could "be **easily customized** for the target application" (9:9-14). For example, Stuttard also disclosed "a simple extension to standard C" (a conventional programming language) "to identify data which is to be processed in parallel," thereby facilitating the design and use of the chip for various applications, including the image processing applications discussed above. Stuttard, 11:11-14.

152. When performing such image-identification applications, POSAs understood that the host computer would have provided instructions to be performed by the PEs in the Stuttard-Shirazi processor to accelerate the image-identification application, thereby meeting claim 2. For example, Stuttard at 1:21-24 notes as background information that with SIMD processors (like Stuttard's) "th[e] *host* processor … will send data to be processed to the SIMD array with" instructions to be executed. Stuttard at 14:9-10 similarly discloses that "[t]he control unit fetches instructions from memory, decodes them and despatches them to the appropriate functional unit." *See also* Stuttard, 9:17-18. POSAs understood that the instructions that are fetched from memory were "loaded into memory" via the "interfaces to the host system" that "initiatiz[e] and boot[]" the processor to "start code running on the MTAP process." Stuttard, 21:13-25. For example, instructions for the processor are "fetch[ed]" from the host system via an "external interface[]" like the ACVI (L/S) that I discussed in Sections VII.B.6 and VII.B.9 above. *See* Stuttard, 13:6-13. Stuttard at 2:34-3:6 similarly also teaches that the "processing elements in the SIMD array and the mono execution unit have a common instruction set" and (at 6:3-26) that data to be processed by the poly unit PEs are retrieved over the I/O interfaces that, as I discussed above in §VII.B.6 [1E], are used to exchange data with the host computer. Stuttard at 7:5-10 teaches that "[t]he design of this processor architecture was driven by the desire to make it

practical and efficient to program," which POSAs understood would be programmed via the host system. *See also* Stuttard, 10:27-11:16 (describing programming the MPTA processor from "a programmer's perspective"). *See also* Section VII.C.2 below (Ground 2A, Claim 3: describing control signals conveying instructions from the host to the processing elements).

153. Stuttard-Shirazi thus renders obvious claim 4 of the '616 patent for the reasons discussed above.

F. Claim 5: "the width of each [first/second] binary mantissa is no more than 10 bits plus a [first/second] sign bit"

154. As I discussed above in Section VII.B.10, in Stuttard-Shirazi, each PE in the poly unit is adapted to multiply two reduced-bitwidth floating-point values. Based on Shirazi's teachings, each PE's multiplier circuit is adapted to receive two floating-point values comprising 10-bit-wide or 9-bit-wide mantissas plus one sign bit, as illustrated in Shirazi's Figures 2-3 (which I have annotated below):

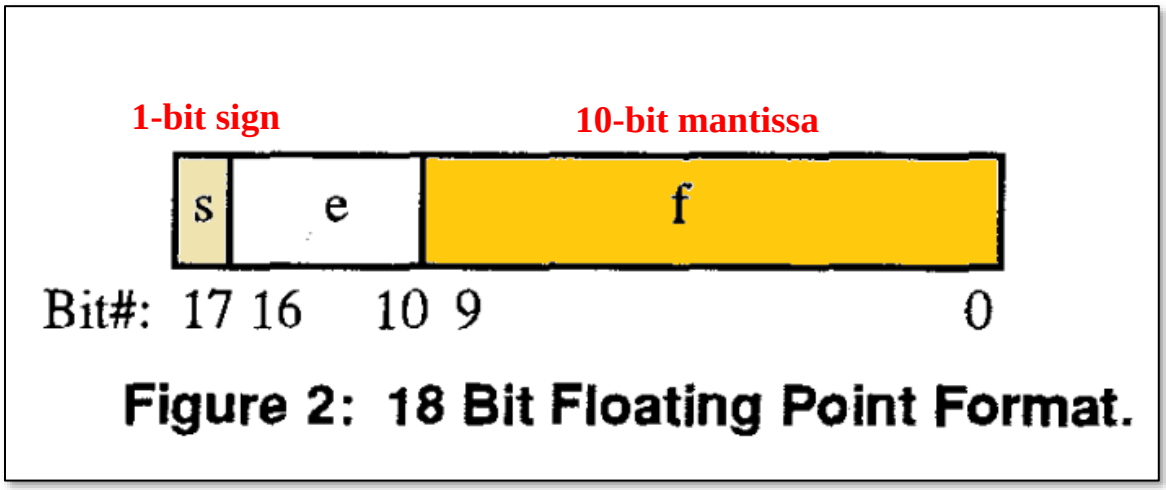


Figure 2: 18 Bit Floating Point Format.

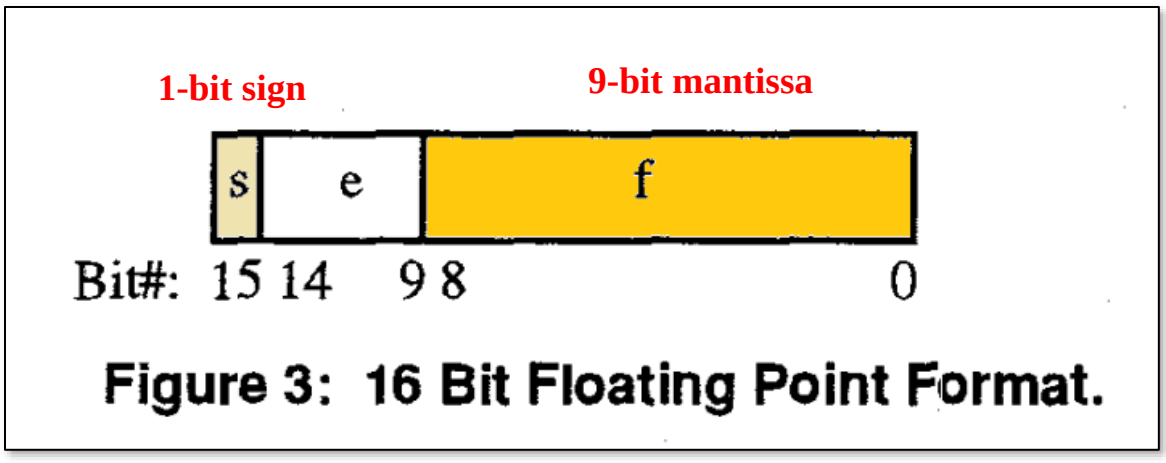


Figure 3: 16 Bit Floating Point Format.

Shirazi, 156.

155. Thus, in Stuttard-Shirazi each of the first and second binary mantissas are no more than 10 bits plus a sign bit, as claimed. Accordingly, Stuttard-Shirazi thus renders obvious claim 5 of the '616 patent for the reasons discussed above.

VIII. GROUND 1B: CLAIMS 7-14 ARE OBVIOUS OVER STUTTARD IN COMBINATION WITH SHIRAZI

A. Scaling the Stuttard-Shirazi Processor

1. 5,000+ Processing Elements

156. Limitation [7C] (which I address in Section VIII.B.2 below) recites a "plurality of first processing elements [that] is no less than 5000 in number." For the reasons I explain below, POSAs had reasons to scale the Stuttard-Shirazi processor in a manner that would have resulted in over 5000 reduced-bitwidth PEs on one computing chip.

157. **First**, Stuttard taught that one "poly execution unit may consist of…***thousands of PEs***" within an MTAP processor core. Stuttard, 10:12. Stuttard

at 12:15-16 similarly disclosed that each processor "core can be scaled…to 256 or even ***thousands of PEs.***" Similarly, EX1108, which was co-authored by two of Stuttard's inventors (Neil Hickey and David Stuttard), similarly taught that the "*poly* execution unit [in an MTAP processor core]…may consist of…***thousands of identical processor element (PE) cores***." EX1108, 32-33 (emphasis on "poly" original); *see also* EX1108, 2 (identifying authors); Stuttard, code (75); Appendix (discussing public accessibility of EX1108).

158. Moreover, Stuttard's processor architecture was expressly intended for "high performance, high data rate processing." Stuttard, 8:24-25. The "processing power ***can scale*** with increasing function demands." Stuttard, 9:2-3. "The poly execution unit," in particular, "get[s] its performance largely from ***the number of PEs brought to bear*** on the problem." Stuttard, 7:19-20. As discussed (Section VII.C above), Stuttard's array of PEs in its poly unit is well-suited to image-retrieval applications. POSAs would have known, for example, as EX1110 (at 607-608) corroborates, that in content-based image retrieval (CBIR) systems the "response time requires to be low for good interactivity," and so these systems "could be implemented using" SIMD processors (like Stuttard's). As another example, EX1113 (at 543) similarly disclosed that "reduc[ing] the retrieval time…is very important to the massive CBIR systems." *See* Appendix (discussing public accessibility of EX1110, EX1113). Thus, POSAs would have recognized

the benefit in, and market forces driving, increasing the number of PEs on a chip to be as high as reasonably achievable (*see* Section VIII.A.3 below discussing POSAs' reasonable expectation of success in achieving 5000+ PEs), to increase the system's processing performance for applications like CBIR (and other highly parallel applications), at lower power and/or lower cost compared to multi-chip systems having equivalent processing performance. The below is a list of additional evidence corroborating my opinion that POSAs had reason to and would have been aware of the market forces driving an increase in the number of processing elements in a parallel processing chip like Stuttard-Shirazi:

- EX1114, 89 (Stuttard's Applicant noting "the current unstoppable trend to put more and more cores on a single die," which would apply to both Stuttard's MTAP core and the individual PE "cores" in the poly unit);

- EX1013, 394 ("[T]he ***more cores that can be packed into a single piece of silicon***, the more parallelism that can be exploited."); 403 ("an ***increase in the number of cores*** will mean ***more exploitation of parallelism***");

- EX1014, 9 ("[R]eduction in core area [from lower-precision PEs] can be ***leveraged to increase the number of cores in the same silicon area***."), 1 ("reduc[ing] the size of floating point units (FPUs)[] allow[s] more of them to occupy the same silicon area"; "***increased number of FPUs can significantly improve the performance of future physics accelerator*s**");

- EX1084, 1:25-37 ("[T]hroughput of a SIMD parallel processor is determined by the number of PEs multiplied by the throughput of each individual PE…. Generally, ***the more PEs per chip, the less costly the implementation*** for a given level of performance.");

- EX1106, 4 ("Parallel execution in multi-core designs will then allow us to ***take advantage of these greater transistor densities to provide greater performance***…. : Many simple cores can be built within the same area as a small number of large complex cores."), *id.* ("[W]e ***must increase the*** efficiency and ***number of cores on a chip***." (again, this would be true of both the MTAP core overall, which includes the mono unit, and the number of PEs in the PE array));

- EX1116, 1 ("Advances in semiconductor technology have ***driven the hardware industry to integrate an increasing number of cores on-chip***")).

159. Thus, in my opinion, POSAs had reasons to include 5000+ PEs in a Stuttard-Shirazi computing chip. *See also* Appendix (my opinions regarding public accessibility evidence of the above-cited references).

160. **Second**, I understand from counsel that the mere duplication of parts has no patentable significance unless a new and unexpected result is produced. The '616 patent only references 5000 PEs once, among a generalized, boilerplate description of "certain embodiments," with all but one reciting significantly fewer than 5000 PEs, and none of which are limited to being on a single chip (rather than

multiple chips). *See* '616 patent, 28:46-51. These embodiments are otherwise indistinguishable: the patent does not allege that any particular number of PEs is critical to the alleged invention, allowed for qualitatively different functionality, or produced new or unexpected results. *See* '616 patent, 28:13-29:4. Thus, I understand that claim 7's limitation reciting 5000+ PEs (Section VI.B.2 below) should carry no patentable significance, reinforcing that it would have been an obvious number of PEs to use in Stuttard-Shirazi for the reasons I've explained above.

161. **Third**, I understand that it is not inventive to discover the optimum or workable ranges by routine experimentation when the optimized variable is a result-effective variable. Stuttard expressly taught that the number of PEs is a result-effective variable, optimized by routine experimentation: "**choosing the number of PEs** and the clock frequency" provides the designer with "a great deal of **flexibility to tune** performance, bandwidth, cost and power for the needs of a given application." Stuttard, 12:18-21; *see also* 9:13-14 (Stuttard's processor is "**easily customized** for the target application"). EX1084 at 1:25-36 similarly teaches that "throughput of a SIMD parallel processor ***is determined by the number of PEs*** …," which corroborates my opinion that selecting the number of PEs in a Stuttard-Shirazi chip would have been routine optimization of a result-

effective variable, such that arriving at 5000+ PEs would have been obvious to POSAs.

162. Thus, selecting the number of PEs in a Stuttard-Shirazi chip would have been routine optimization of a result-effective variable, such that arriving at 5000+ PEs would have been obvious. I will refer to this implementation of Stuttard-Shirazi – with 5000+ reduced-bitwidth PEs in the PE array – as the "scaled Stuttard-Shirazi processor."

2. Obvious Ways to Implement the Scaled Stuttard-Shirazi Processor

163. For the reasons I discussed above (§VIII.A.1) POSAs had reasons to implement the scaled Stuttard-Shirazi processor to include over 5000 reduced-bitwidth PEs. POSAs knew this could be done in several alternative ways that I discuss below.

a. First Obvious Implementation of the Scaled Stuttard-Shirazi Processor

164. Stuttard teaches scaling a single MTAP processor core's poly unit to "thousands of PEs." Stuttard, 12:15-16. "Thousands," at a minimum, includes at least 2000 PEs in one processor core (and, in my opinions, POSAs would have understood "thousands" to indicate more than 2000, such as many as could reasonably fit on the chip). Thus, in one implementation, given that Stuttard teaches scaling a single MTAP processor core's poly unit to "thousands of PEs"

(Stuttard, 12:15-16), POSAs would have been motivated to scale the poly unit in a single MTAP core to include as many "thousands of PEs" as could reasonably be achieved using available process technology, for the reasons I discussed above in Section VIII.A.1. I demonstrate in Section VIII.A.3 below demonstrates that well over 5000 PEs could be fabricated on a single chip before the '616 patent's earliest claimed priority date.

b. Second Obvious Implementation of the Scaled Stuttard-Shirazi Processor

165. Another conventional way POSAs would have achieved the known goal of increasing the number of PEs on the chip (Section VIII.A.1 above) would have been to replicate an existing processor core design (with its corresponding ratio of poly to mono execution units) multiple times on the chip. This would have been a conventional technique in the art to scale up the capabilities of a computing chip, to leverage functional blocks from existing/prior generations of products and thereby save cost and time by avoiding unnecessary redesign. For evidence corroborating my opinion that duplicating an existing processor core was a known, beneficially technique for scaling up the number of cores and/or processing elements on a chip, EX1068, at [0002]-[0007] explains as background knowledge that "[m]odern microprocessor systems derive significant performance enhancements from the ***implementation of multiple, identical cores on a single integrated circuit*** (IC)" by making "***identical replications of the core***…on a single

die" which results in "***lower design costs***." POSAs would have understood that duplicating the design of existing cores lowers design costs because the existing core design can be leverage and integrated with duplicates using ordinary semiconductor layout skill, rather than have to design (and test) a completely new core design. Similarly, EX1069 at [0004]-[0005] explains as background knowledge that "[c]hip multithreading (CMT) processors and chips may include a number of cores" that is achieved by repeating the "same [core] design copied over multiple times on a die." As additional corroboration, EX1106 at 4 (a White Paper by Intel) taught that "***standard, repeated tiles***…into a common infrastructure should ***allow more reuse of designs between generations of processors***." That is, by maintaining a common architecture for a single core (or "tile") and duplicating that design on a chip, POSAs would have understood they could continue to fit more and more cores on a single chip as semiconductor fabrication technology continued to advance with ever-shrinking transistor sizes, thereby lowering design costs, as I explained above. Thus, duplicating Stuttard-Shirazi's MTAP core with "thousands" of PEs, to beneficially achieve even more PEs on the chip, would have also been using a known technique to improve similar devices (adding PEs to MTAP-based processors) in the same way (by duplicating the pre-designed MTAP core as many times as desired).

166. POSAs would have also recognized the benefit in having multiple mono units on a chip to provide several full-precision FPUs on the chip, in addition to thousands of reduced-bitwidth FPUs. For example, Stuttard discloses an exemplary arithmetic operation where each PE in the poly unit multiplies a single value stored in the mono unit. Stuttard, 3:12-14. Having multiple mono units would have allowed subsets of the chip's PEs to multiply different values stored in different mono units over a shorter time period, compared to a chip having a single mono unit. Having a small number of PEs in the mono units with high floating point precision and others with lower levels of precision on the same computing chip would also provide flexibility in terms of the types of calculations that can be performed. Different applications may require different levels of precision, and having a mix of processing elements with different precision levels would allow the processor to switch between them as needed. This would allow the processor to perform a wider range of calculations more efficiently, as it would not be limited to a single level of precision. Moreover, having a mix of processing elements with different precision levels on the same chip would be more efficient from a hardware standpoint. In many cases, the high-precision processing elements would only be needed for a small number of calculations, while the lower-precision elements would be used for the majority of calculations. By having both types of processing elements on the same chip, the processor would be able to use the

lower-precision elements for most calculations, saving space and reducing the overall cost of the processor. In short, a POSA would want some processing elements with high floating point precision and some with lower precision levels on the same chip because it would provide flexibility and efficiency in terms of the types of calculations that can be performed. That is, POSAs understood having several full-precision FPUs on the chip along with many reduced-bitwidth PEs would "provid[e]…granularity options," allowing the chip to handle a wider range of tasks and data types. EX1124, 73-76 (teaching "researchers have proposed heterogenous multicore (also known as asymmetric multicore) processors as a means to improve power efficiency or performance" by having different sized processing elements on a die; "expect[ing] this trend to continue").

167. For additional evidence corroborating my opinion that POSAs understood the desirability of having a small number of higher precision PEs in the mono units and a much larger number of reduced-bitwidth PEs in the poly unit, see corroboration EX1106 (at 6), which describes a multi-core architecture that had a "mix of element types matched to the significant workloads in the segment it is to server" and noting that "the number of each and relative number of each type could be varied." EX1116 (at 1) similarly recognized that "higher energy efficiency" is achieved by [using] (asymmetric (or heterogenous) architectures," meaning PEs of different complexity and precision. Similarly, EX1117 (at 746-749) disclosed "[a]

many-core system on a die" that "may have a few large cores to deliver higher single-thread performance, but will predominantly have large number of small cores."

168. Stuttard also discloses an exemplary MTAP processor, with a mono unit having "[s]upport for 8 threads." Stuttard, 22:20-30. Stuttard further teaches that "[t]he number of threads is a configurable option" and that "threads are prioritized" so the "processor will run the highest priority thread," which can "pre-empt a lower priority thread." Stuttard, 14:18-22. EX1119 at 1:23-25 notes that a "'thread' refers generally to an instance of execution of a particular program using particular input data." And EX1122 at [0013] taught that "[t]here is ***generally a desire to have as many processing cores as possible each concurrently processing as many threads as possible*** in order to obtain the greatest processing power and efficiency from the processor." Thus, another reason POSA would have wanted to increase the number of mono units on the chip would be to support parallel processing of additional, highest-priority threads on the chip that could not be efficiently parallelized (i.e., computations better suited for the mono units).

169. As additional corroboration I note that EX1012 at 234, for example, describes different phases of physics simulations for video game applications, some of which are "not easily mapped to parallel architectures" while others "exhibit[] massive fine-grain (FG) parallelism." Thus, EX1012 corroborates my

opinion that POSAs understood that some applications would likely have a mixture of processing that can be easily parallelized (and thus well-suited for the Stuttard-Shirazi poly PEs) and processing that could not be easily parallelized (and thus better-suited for the Stuttard-Shirazi mono units). *See also* §V.D.2 (describing known benefits of having parallel processors with a large number of efficient cores, like Stuttard-Shirazi's reduced-bitwidth poly PEs, and a smaller number of (but more than one) more-complex processing elements, like Stuttard-Shirazi's mono unit).

170. Finally, as I discussed above (§VIII.A.1) I understand that it is not inventive to discover the optimum or workable ranges by routine experimentation when the optimized variable is a result-effective variable. Stuttard expressly taught that the number of PEs is a result-effective variable, optimized by routine experimentation: "***choosing the number of PEs*** and the clock frequency" provides the designer with "a great deal of flexibility to tune performance, bandwidth, cost and power for the needs of a given application." Stuttard, 12:18-21; *see also* 9:13-14 (Stuttard's processor is "easily customized for the target application"). While the majority of the PEs are in Stuttard's poly unit, the mono unit itself has a PE. *See, e.g.*, Stuttard, 3:8-9 ("The processing elements in the SIMD array ***and the mono execution unit*** have a common instruction set.").

171. Thus, for the same reason selecting the number of PEs in a Stuttard-Shirazi chip would have been routine optimization of a result-effective variable, selecting the number of mono units (and mono PEs) would have also been routine optimization of a result-effective variable, such that arriving at two (or more) mono units on a chip would have been obvious to a POSA. The '616 patent's specification also does not allege that any particular number of 32-bit floating-point multipliers is critical to the alleged invention, allowed for qualitatively different functionality, or produced new or unexpected results. Thus, I understand from counsel that the number of 32-bit floating point multipliers recited in the claims should carry no patentable weight for the same reason the number of PEs on the chip should carry no patentable weight, reinforcing the obviousness of having more than one 32-bit floating point multiplier on the chip.

172. As I noted above, Stuttard teaches scaling a single MTAP processor core's poly unit to "thousands of PEs." Stuttard, 12:15-16. "Thousands," at a minimum, includes at least 2000 PEs in one processor core (and, in my opinions, POSAs would have understood "thousands" to indicate more than 2000, such as many as could reasonably fit on the chip). If a POSA following this teaching scaled a Stuttard-Shirazi processor core to, for example, 2048 PEs (*e.g.*, a 64x32 PE array—just barely "thousands" of PEs and numbers that are multiplies of 2), then duplicating this 2048 PE / one mono unit core three times on a chip would

have provided an array of 6144 reduced-bitwidth PEs and 3 full-precision mono units (each having a 32-bit floating-point multiplier circuit, as I discussed in §VII.B.10 above for the mono unit in Stuttard-Shirazi and will discuss in §VIII.C.1 below for mono units in Scaled Stuttard-Shirazi).

c. Third Obvious Implementation of the Scaled Stuttard-Shirazi Processor

173. As another alternative example implementation of the Scaled Stuttard-Shirazi processor, I note again that an exemplary MTAP processor core discussed in Stuttard had one mono execution unit and 256 poly-unit PEs. Stuttard, 10:12-19, 12:15-16. Duplicating this core's architecture 20 times on a chip would have provided 5120 reduced-bitwidth PEs and 20 mono execution units. POSAs would have understood that the ratio of the number of reduced-bitwidth poly PEs to 32-bit mono units would have been a well-understood design choice, based on, for example, whether the chip were being designed for particular applications, balancing the expectation of computations that would and would not be easily parallelized. *See*, *e.g.*, EX1013, 401-404, specifically 403 (teaching that "[p]hysics requires varying amount of numerical accuracy across different phases and scenarios" and that, at times, "a full precision FPU in a physics processor architecture" "is required" and so POSAs understood that there would have been benefit to having multiple higher-precision processing elements on a chip for applications like this that require varying degrees of numerical accuracy). *See*

§V.A.2 above (describing various applications that were known to benefit from a mixture of many efficient PEs on a chip and a smaller plurality of more-complex PEs on a chip).

3. Reasonable expectation of success

174. In my opinion, POSAs would have reasonably expected success in scaling the Stuttard-Shirazi processor discussed in Section VIII.A.1 above in any of the obvious implementations I discussed in Section VIII.A.2 above.

175. Again, I note that Stuttard's processor architecture is expressly designed for "easy scalability" (1:5-6) by being "highly modular" to "allow[] many details of a specific implementation to be ***easily customized*** for the target application" (9:13-14). *See also* Stuttard, 8:29-32 ("The MPAP processor…has a ***clear scalability*** path for future requirements."). Stuttard teaches that a "designer has a great deal of flexibility" in "choosing the number of PEs." Stuttard, 12:15-21. Stuttard described simulating "***the number of PEs***, size of memory, and other options" in software "before the system architecture is fully defined" and put into "silicon implementation" (i.e., before fabricated as a real chip) (Stuttard, 24:28-32), which POSAs understood facilitated designing the number of PEs on the chip.

176. As an aside, I note that Stuttard at 12:18 states that "Figure 6 shows how performance and power scale with the number of PEs." See also Stuttard, 6:34, 10:5-6. However, as I explained above (§VII.B.6), this is a typographical

error because the referenced figure was never included in the Stuttard application. *See generally* EX1095, 32 (specification from Stuttard's original PCT application at 4:15-26 (describing nine figures)), *id.*, 51-62 (only eight figures included in application)). POSAs would have understood that what published as Stuttard's Figure 6 illustrates the "graphical representation of the results of running an example program" (EX1095, 32 (described in the original specification at 4:21-22)) that is described in Stuttard at 11:18-12:6 ("calculat[ing] 64 values of a sine function across the PE array"), and Stuttard has no figure showing how performance and power scale with the number of PEs.

177. There are several ways by which I can demonstrate that a POSA would have reasonably expected success in the Scaled Stuttard-Shirazi chip discussed above (§VIII.A.1).

178. In 2008, Intel disclosed a microprocessor chip based on its 65-nm process node having over 2 billion transistors. EX1039, 18; *see* Appendix (opinions regarding public accessibility of EX1039). By 2008, TSMC, another semiconductor fabrication company, had disclosed a 40-nm process node. *See* EX1038, 3 (noting Altera had introduced commercial "40-nm devices" (e.g., "FPGAs and ASICs") using TSMC's 40-nm node); EX1043 (2008 TSMC press release noting 40-nm process had moved to "volume production"). Appendix (opinions regarding public accessibility of EX1038 and EX1043,

respectively). TSMC's 40-nm node process had a "***2.35 times density improvement*** compared to devices built on 65nm process technology." EX1038, 71.

179. Thus, even if I only consider the above, which was known to POSAs as of June 19, 2009 (the filing date of the '619 Provisional, *see* §V.C above), by mid-2009 POSAs would have reasonably expected to be able to make one integrated circuit on a chip with over 4 billion transistors using a 40nm node. This is because POSAs would have understood that they could have (at least) scaled 65-nm Intel's 2 billion transistor chip to TSMC's 40-nm process node and achieved a density improvement of approximately 2.35 (i.e., 2 billion transistors times a 2.35 scaling factor equals more 4 billion).

180. By mid-2009, however, POSAs would have also been aware that they could have achieved even more transistors on a single chip using the denser 32nm and 22nm process nodes, which had been demonstrated in the prior art.

181. For example, EX1045 is a paper published in February 2009 by members of Intel. EX1045, 1. *See* Appendix (opinions regarding public accessibility of EX1045). That paper demonstrated a 32nm chip (an "SRAM test vehicle") that achieved an SRAM "0.171μm^2 cell size." EX1038 (at 71) noted that TSMC's 40-nm process had an "SRAM cell size at 0.242 μm^2." Thus, the 32-nm node process disclosed in EX1045 would have allowed for an additional 1.4x

scaling over TSMC's 40nm SRAM cell size (0.242 μm^2 / 0.171μm^2). As another example, EX1094 is a paper published in 2008 by members of IBM. *See* Appendix (opinions regarding public accessibility of EX1094). That papers demonstrated a "32nm" process achieving "functional high-density SRAM with a cell size of 0.157μm^2. EX1094, 67. As noted above, TSMC's 40-nm process node had an "SRAM cell size at 0.242 μm^2." EX1038, 71. Thus, the 32-nm process node disclosed in EX1094 would have allowed for an additional 1.5x scaling over TSMC's 40-nm process node based on their respective SRAM cell sizes (0.242 μm^2 / 0.157 μm^2).

182. Thus a 32-nm process node would have resulted in over 6 billion transistors on a chip (i.e., by scaling Intel's 2-billion plus transistor chip built in 65-nm technology to a 32-nm process node technologies, which had a 2.32 scaling factor going from 65-nm to 40-nm technology times the additional 1.4-1.5 scaling factor going from 40-nm to 32-nm technology).

183. As another example, EX1115 is another paper published by members of IBM (in February 2009). EX1115, 1. *See* Appendix (opinions regarding public accessibility of EX1115). That paper demonstrated "22 nm node technology[] compatible with fully functional 0.1 μm^2 6T-SRAM cell" (also noting they achieved a 0.09 μm^2 cell"). As noted above, TSMC's 40-nm process node had an "SRAM cell size at 0.242 μm^2." EX1038, 71. Thus, the 22-nm process node

disclosed in EX1115 would have allowed for an additional 2.42x scaling over TSMC's 40-nm process node based on their respective SRAM cell sizes (0.242 μm^2 / 0.1 μm^2). Accordingly, this 22-nm process node would have resulted in nearly 10 billion transistors on a chip (i.e., by scaling Intel's 2-billion plus transistor chip built in 65-nm technology to a 32-nm process node technologies, which had a 2.32 scaling factor going from 65-nm to 40-nm technology times the additional 2.42 scaling factor going from 40-nm to 22-nm technology).

184. I also note that in September 2009, Intel demonstrated 22nm node working "test chip" with "SRAM memory and logic circuits" having "2.9 billion transistors" at "approximately **double the density of the previous 32nm** generation." EX1126, 1. *See* Appendix (opinions regarding public accessibility of EX1126). I understand that EX1126 is prior art to (at least) claims 7-14 because those claims do not have written description support in the '619 Provisional, for the reasons I discussed in §V.E above. Because EX1126 disclosed a 22-nm node process with "double the density of the previous 32nm generation" (EX1126, 1) a POSA would have understood that by September 2009 the 22-nm node process described in EX1126 would have achieved an additional ~3x scaling over TSMC's 40-nm process node based on their respective SRAM cell sizes (because the 32-nm processes I discussed above achieves ~1.4x – 1.5x scaling over TSMC's 40-nm process node).

185. Thus, POSAs would have reasonably expected success in using the 32-nm and 22-nm processes I described above to achieve between six and 12 billion transistors on chip (i.e., by scaling Intel's 2-billion plus transistor chip built in 65-nm technology to the 32-nm and 22-nm process node technologies having the scaling values I explained above).

186. In the early 2000s, Stuttard's Applicant demonstrated to POSAs that a 64-poly-PE MTAP core could be fabricated on a chip using a "0.13-micron" (130-nm) process node, and that this core would require "40 million" transistors. EX1118, 1 (describing the CS301 processor); *see also* EX1107, 822 (showing the CS301 had an MTAP architecture similar to Stuttard's)). As discussed above, POSAs knew before June 2009 that over 4 billion transistors could be fabricated on a single chip using the 40-nm process node (and even more could be fabricated using a 32-nm or 22-nm process node). Four billion transistors is 100 times as many transistors as the 40 million transistors used in the above-discussed 64-poly-PE MTAP core that was disclosed in EX1118. Transistors are the basic building blocks of digital circuitry and the number of transistors that could fit on a chip was a well-known, and well-publicized figure of merit for a processing technology. Thus, POSAs would have expected that the 100 times as many transistors that would have been achievable with a 40-nm process node (compared to the 65-nm node) would have allowed for fabricating approximately 100 times as many of a

given component (e.g., processor cores) on the chip. Thus, POSAs would have reasonably expected success in implementing up to (at least) one hundred 64-poly-PE MTAP cores on a single chip (because it would have required approximately 4-billion transistors, which was reasonable for the reasons I discussed above). That chip would have included 6400 poly PEs and 100 mono execution units, both of which exceeds any of the Scaled Stuttard-Shirazi processors I described above in §VIII.A.2 above.

187. EX1118 disclosed that the CS301 (the 64-PE MTAP processor described above) had "an array of 64 PE's, each of which consisted of two floating-point execution units." EX1107 (at 822-823) disclosed that the CS301 supported "32-bit" "Floating Point (IEEE)." Thus, a POSA's expectation of success in the Scaled Stuttard-Shirazi processor would have been bolstered by their understanding that Stuttard-Shirazi's reduced-bitwidth floating-point units would have likely required less, but certainly no more, chip area than the 32-bit floating-point units within the types of poly PEs described in EX1118 and EX1107 (at 822-23).

188. As additional corroboration regarding my opinion that POSAs reasonably expected success in the Scaled Stuttard-Shirazi processor, EX1106 at 4 teaches that "[m]any simple cores can be built within the same area as a small number of large complex cores." Similarly, EX1014 (at 9) taught that "reduction

in core area [from lower-precision PEs] can be leveraged to increase the number of cores in the same silicon area") and taught (at 1) that "reduc[ing] the size of floating point units (FPUs)[] allow[s] more of them to occupy the same silicon area." Thus, EX1106 and EX1014 further corroborate that by reducing the precision of the poly PEs (as in Stuttard-Shirazi), POSAs would have been expected to fit even more of those PEs on a chip compared to the IEEE 32-bit PEs in the CS301 (all other things, like chip size and process node, equal).

189. As an additional demonstration that POSAs would have reasonably expected success in the Scaled Stuttard-Shirazi processor, I note again that Stuttard's Applicant demonstrated a 64-PE chip using a 130nm process node. *See, e.g.*, EX1118 (noting the CS301 used a "0.13-micron process[]"). EX1124 (at 74) notes that the die size for that 64-PE chip was "72" "mm2" and that the die size for Intel's 2-billion transistor Itanium chip (EX1039) was "699" mm2. Thus, simply increasing the size of the chip from 72mm2 to 699mm2, POSAs would have understood that they could have achieved approximately 9.7 times as many PEs and mono-units on a single chip in the 65-nm process node (i.e., 699 mm2 / 72mm2).

190. EX1097 is a 2001 paper by members of IBM. EX1097, 101. *See* Appendix (opinions regarding public accessibility of EX1097). That paper that disclosed a "0.13µm" (130nm) SRAM cell of "2.28µm^2." EX1097, 101. As noted

above, TSMC's 40-nm process node had an "SRAM cell size at 0.242 μm^2." EX1038, 71. Thus, TSMC's 40-nm process node would have allowed for a scaling factor of ~9.4 over the 130-nm process node based on their respective SRAM cell sizes (2.28 μm^2 / 0.242 μm^2).

191. The 9.7 scaling factor I discussed above based on comparing the chip size of the CS301 to the chip size of Intel's Itanium processor is independent of the 9.4 scaling factor I just described in moving from the CS301's 130-nm process node to a 40-nm process node. That is, POSAs would have understood that these two scaling factors could be multiplied together to get a combined scaling factor of approximately 91 times, which is consistent with the 100 times scaling factor I discussed above based on transistor counts. Even assuming a scaling factor of 91 meant that the 64-PE MTAP processor described in the prior art could have been scaled to 5824 PEs and 91 mono units, which has significantly more mono units than any of the Scaled Stuttard-Shirazi processors I described in §§VIII.A.2 above and also does not account for the likely area savings that would have been achieved from using Stuttard-Shirzai's reduced-bitwidth PEs compared to the 32-bit PEs in the 64-PE MTAP processor described in EX1118 and EX1107 (at 822-23). (And, for the reasons I discussed above, higher scaling factors would have been achievable using the 32-nm and 22-nm processes that were described in the prior art that I discussed above.)

192. Thus, by (at least) mid-2009, POSAs would have reasonably expected success in fabricating the scaled Stuttard-Shirazi device to include well over 5000 poly PEs (having reduced-bitwidth FPUs) on one chip (the first obvious implementation of the scaled Stuttard-Shirazi chip, described in §VIII.A.2.a above) and a mono execution unit (with an IEEE-standard compatible 32-bit FPU). Additionally, by (at least) mid-2009, POSAs would have also reasonably expected success in fabricating the scaled Stuttard-Shirazi processor as one chip on which a 2048-PE Stuttard-Shirazi processor core was duplicated 3 or more times (the second obvious implementation of the scaled Stuttard-Shirazai chip, described in §VIII.A.2.b above)—or alternatively on which a 256-PE Stuttard-Shirazi processor core was duplicated 20 or more times (the third obvious implementation of the scaled Stuttard-Shirazi chip, described in §VIII.A.2.c above)—resulting in a chip having at least 6144 (or 5120) reduced-bitwidth PEs and a plurality of mono execution units (with an IEEE-standard compatible 32-bit FPU).

193. My opinion that POSAs would have reasonably expected success in the Scaled Stuttard-Shirazi processor is reinforced by the fact that the ’665 patent itself identifies no challenges or specific fabrication techniques required to achieve 5000+ PEs on a chip. Thus, the ’616 patent’s silence in this regard reinforces that only ordinary skill would have been required to integrate such numbers of PEs

(and, if desired for the expected applications, multiple 32-bit multipliers) on one computing chip.

B. Claim 7

1. [7PRE] "A computing system, comprising:" [7A] "a host computer;" [7B] "a computing chip comprising:"

194. As I show in the table below, limitations [7PRE]-[7B] are identical to limitations [1PRE]-[1B], respectively.

Claim 7 limitation	Claim 1 limitation
[7PRE] A computing system, comprising:	**[1PRE]** A computing system, comprising:
[7A] a host computer;	**[1A]** a host computer;
[7B] a computing chip comprising:	**[1B]** a computing chip comprising:

195. Therefore, the Scaled Stuttard-Shirazi system (Section VI.A above) meets limitations [7PRE], [7A], and [7B] for the same reasons I discussed in Sections VII.B.1-VII.B.3 above with respect to limitations [1PRE], [1A], and [1B], respectively, because Stuttard-Shirazi's computing system meets the limitations identically recited in limitations [1PRE]-[1B] and limitations [7PRE]-[7B].

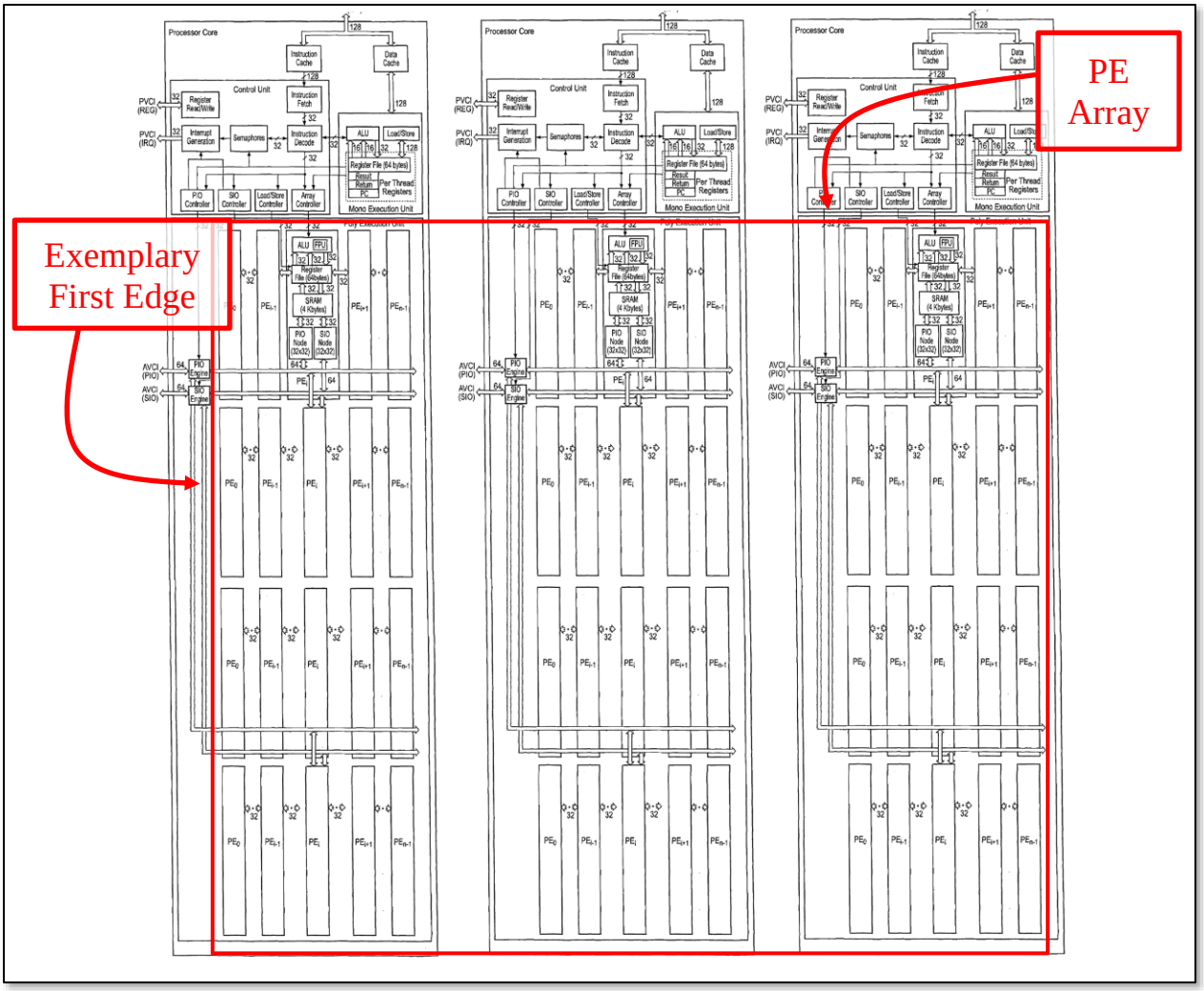
2. [7C]: "a processing element array comprising a plurality of first processing elements, wherein the plurality of first processing elements is no less than 5000 in number, wherein each of a first subset of the plurality of first processing elements is positioned at a first edge of the processing element array, and wherein each of a second subset of the plurality of first processing elements is positioned in the interior of the processing element array;"

196. Even if [7C]'s "no less than 5000 in number" recitation is given patentable weight, the limitation is met because the Scaled Stuttard-Shirazi chip that I described in §VIII.A above because that chip comprises an array with, e.g., 6144 (or alternatively 5120) reduced-bitwidth PEs, which is not less than 5000 in number, as claimed. *See also* §VII.A.3.a above (describing the reduced-bitwidth PEs in the Stuttard-Shirazi chip).

197. For the same reasons I described above in §VII.B.4 (concerning [1C] reciting first and second edge PEs and first second interior PEs), the Scaled Stuttard-Shirazi processor comprises a first subset of the 6144 (or 5120) reduced-bitwidth PEs at a first edge of the PE array (i.e., the same edge identified that I identified in §VII.B.5 above) and a second subset of the PEs positioned in the interior of the PE array (i.e., the PEs not positioned at an edge), as claimed.

198. Below is a schematic illustration I prepared of one exemplary implementation of the Scaled Stuttard-Shirazi processor (having three mono-unit PEs and 3x2048 = 6144 poly-unit PEs). I have outlined the PE array in red and identified an exemplary "first edge" on its left-hand side. In the exemplary

implementation below, the entire Stuttard-Shirazi MTAP core has been duplicated three times. However, POSAs would have also understood that alternative designs would have been consistent with the Scaled Stuttard-Shirazi implementations I described in §VII.B.5. For example, rather than duplicate each MTAP core control unit three times on the chip, POSAs would have understood that a single control unit could have been used to decode instructions and issue control signals to the PE array. Similarly, rather than duplicate the PIO Controller, SIO Controller, Load/Store Controller, and Array Controller and PIO and SIO engines three times on the chip, only one of each circuitry could have been included on the chip (with the PIO and SIO channels extending across the entire PE Array). Indeed, Stuttard at 19:22-31 discloses "[t]here is a Controller and Engine for each I/O channel" and "[a] single Controller can ***manage several I/O Engines***." Moreover, even in implementations where each MTAP core has a control unit, it would have been obvious design choice to include logic on the chip that would have allowed any one of the MTAP core control units to issue instructions to all PEs in the entire PE array on the chip. Thus, Stuttard corroborates my opinion that the number of MTAP cores (and microarchitecture of each MTAP core) was a design choice that required only ordinary skill to implement.



3. [7D]: "an input-output unit connected to each of the first subset of the plurality of first processing elements;"

199. For the same reason I explained above (§VII.B.6) that Stuttard-Shirazi's ClearConnect bus (or alternatively, the bus plus the I/O engines and channels) is an input-output unit connected to its first and second edge PEs, in the Scaled Stuttard-Shirazi chip, the ClearConnect bus (or alternatively, the bus plus the I/O engines and channels) is an input-output unit connected to each MTAP core on the chip, including each of the PEs in the first subset of the plurality of first PEs

on the chip that I described in Section VIII.B.2 above, as claimed *See* Stuttard, 22:8-15, FIG. 8.

4. [7E]: "a plurality of processing element connections, each processing element connection connecting one of the plurality of first processing elements with another of the plurality of first processing elements, wherein each of the plurality of first processing elements is connected to at least one other of the plurality of first processing elements by at least one of the plurality of processing element connections;"

200. As I discussed above in Section VIII.B.2, the plurality of first processing elements are PEs in the Stuttard-Shirazi poly units. Stuttard teaches that each PE in a poly unit can "communicate with one another via" a path "that **connects** the register file of each PE with the register files of its left and right neighbors." Stuttard, 20:31-21:2. Stuttard's Figure 7 annotated below shows connections between neighboring PEs in one poly execution unit as double-sided arrows (which I have circled below), which are a plurality of PE connections, as claimed. The Scaled Stuttard-Shirazi processor has the same connections in each of its poly units. As illustrated below, at least one PE connection exists between each PE and one of its neighboring PEs, which meets limitation [7E].

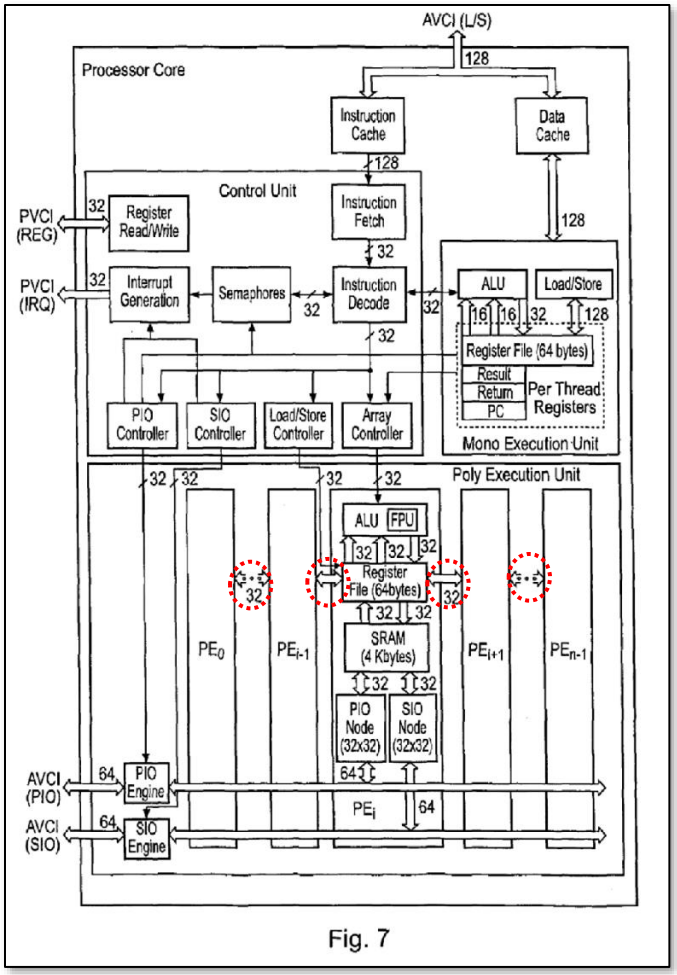

AVCI (L/S)
128
Processor Core
Instruction Cache
Data Cache
128
128
Control Unit
Instruction Fetch
32
PVCI (REG)
32
Register Read/Write
PVCI (IRQ)
32
Interrupt Generation
Semaphores
32
Instruction Decode
32
32
ALU
Load/Store
16
16
32
128
Register File (64 bytes)
Result
Return
PC
Per Thread Registers
Mono Execution Unit
PIO Controller
SIO Controller
Load/Store Controller
Array Controller
Poly Execution Unit
32
32
32
32
ALU
FPU
32
32
32
Register File (64bytes)
32
32
32
SRAM (4 Kbytes)
32
32
PIO Node (32x32)
SIO Node (32x32)
PE0
PEi-1
PEi+1
PEn-1
64
64
AVCI (PIO)
PIO Engine
PEi
64
64
AVCI (SIO)
SIO Engine

Fig. 7

5. [7F]: "a plurality of memory units, wherein each of the plurality of first processing elements is associated with a corresponding one of the plurality of memory units, and wherein each of the plurality of memory units is local to its associated one of the plurality of first processing elements;"

201. As I discussed above in Section VII.B.7 and illustrated in Stuttard's Figure 4 (annotated below), "[e]ach PE [] has its own independent **local memory**" and register files (which are addressable as memory for the reasons I discussed in that section). Thus, the local memory and addressable register file are each a memory unit local to its associated PE, including each of the first PEs, as claimed. Stuttard, 10:14-19, 5:29-30, 7:14-16, 10:1-4, 17:12-23, FIGS. 3-4, 7.

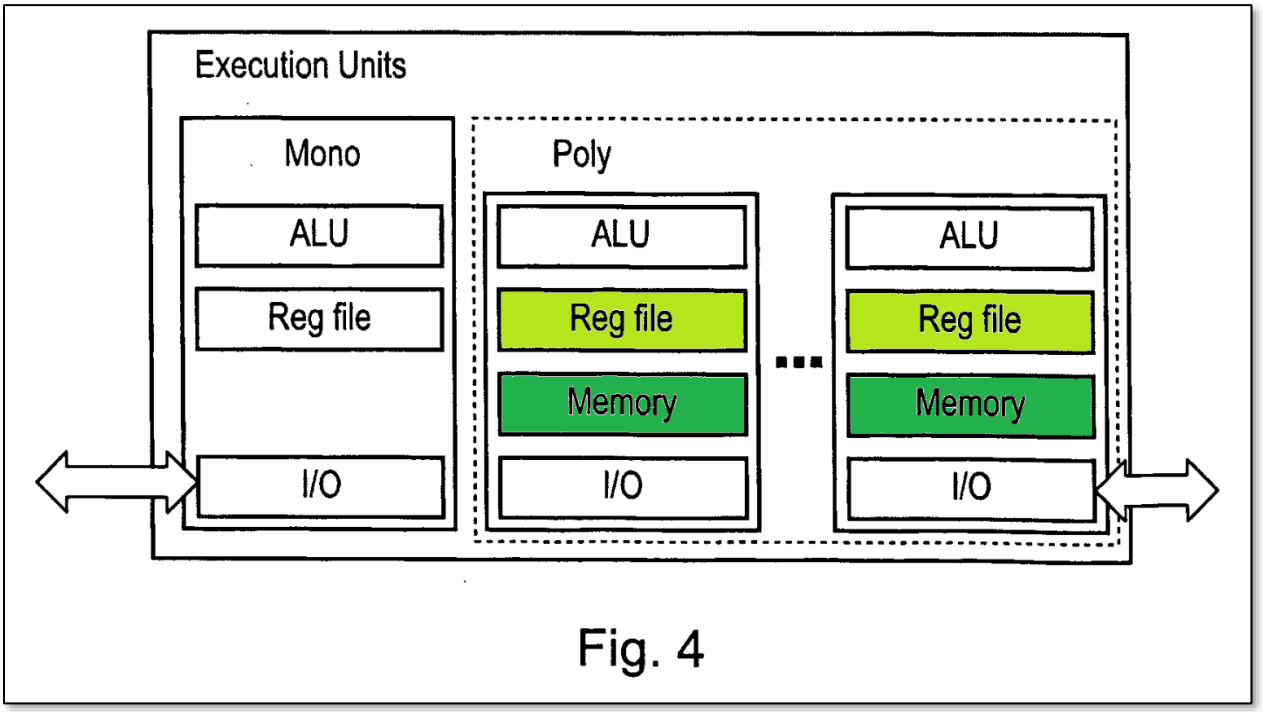


Fig. 4

6. [7G]: "a plurality of first arithmetic units, wherein each of the plurality of first processing elements has positioned therein at least one of the plurality of first arithmetic units;"

202. For the same reasons I discussed above in Section VII.B.8 (for limitation [1G]), each poly PE in the Scaled Stuttard-Shirazi processor has positioned therein a first arithmetic unit, as claimed, which I have annotated in blue below from Stuttard's Figure 4. Stuttard, 5:29-30, 7:14-16, 10:1-4, 10:14-19, 17:12-23, FIG. 3.

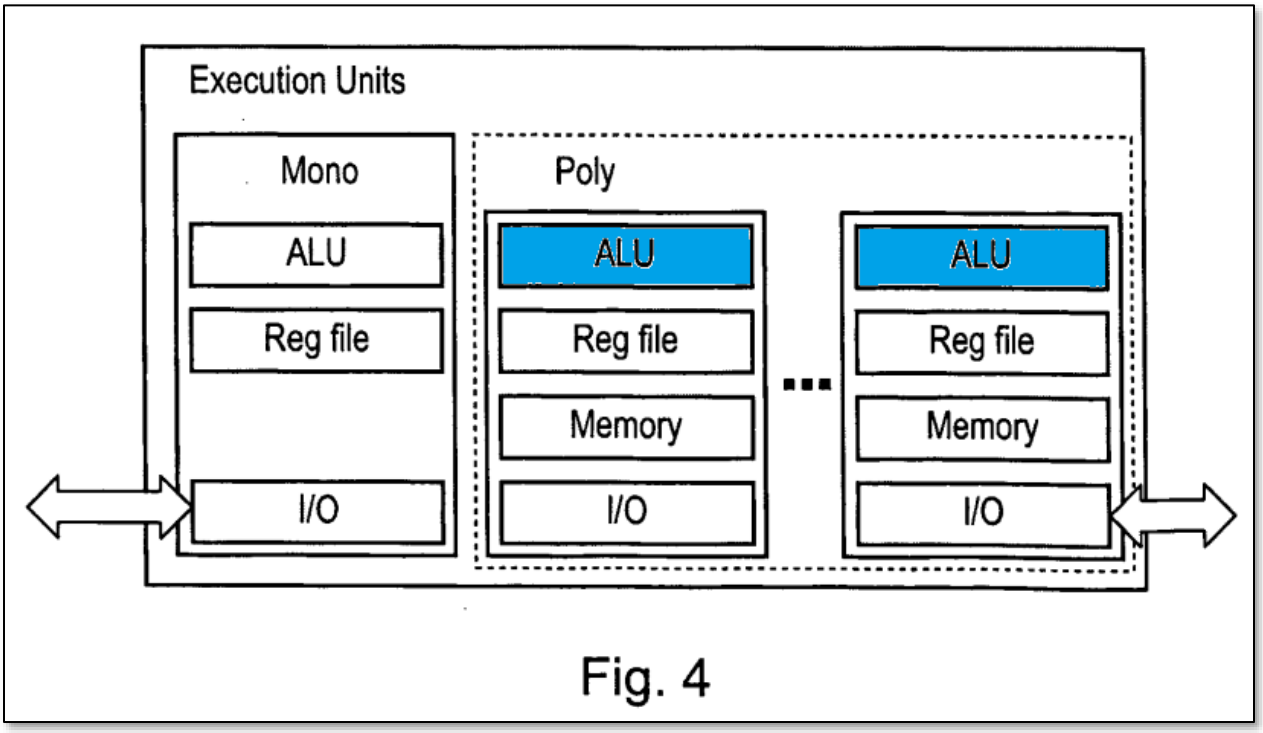


Fig. 4

7. [7H]: "a host connection at least partially connecting the input-output unit with the host computer;"

203. The Scaled Stuttard-Shirazi processor has the same host connection that Stuttard-Shirazi has that I discussed above in Section VII.B.9 ([1H]), which at

least partially connects the input-output unit – the ClearConnect bus or, alternatively, the bus plus the I/O engines and channels – with the host computer, as claimed, for the reasons I discussed in that section.

204. Thus, Scaled Stuttard-Shirazi processor meets [7H].

8. [7I]: "wherein the plurality of arithmetic units each comprises a first corresponding multiplier circuit adapted to receive as a first input to the first corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the first corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits;"

205. The table below shows limitation [7I] side-by-side with limitation [1I]:

Claim 7 limitation	**Claim 1 limitation**
[7I] "wherein the plurality of arithmetic units each comprises a first corresponding multiplier circuit adapted to receive as a first input to the first corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the first corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits;"	**[1J]** "wherein the first, second, third and fourth arithmetic units each comprises a corresponding multiplier circuit adapted to receive as a first input to the corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits"

206. In comparing the two limitations, a POSA would understand that Scaled Stuttard-Shirazi meets limitation [7I] for the same reasons Stuttard-Shirazi meets limitation [1I], for the reasons I discussed in Section VII.B.10 above.

207. Thus Scaled Stuttard-Shirazi renders obvious claim 7 of the '616 patent for the reasons discussed above.

C. Claim 8

1. [8A]: "the computing chip further comprises a plurality of second processing elements, wherein the plurality of second processing elements each comprises a second corresponding multiplier circuit adapted to receive as inputs to the second corresponding multiplier circuit two floating point values each of width at least 32 bits;"

208. In Stuttard, "one PE [processing element] forms the mono execution unit." Stuttard, 10:5-9, 16:13-24, 3:15-18 ("[M]ono execution unit has a 32-bit ALU."). In the second and third obvious implementations of the scaled Stuttard-Shirazi processor that I described in Sections VIII.A.2.b-VIII.A.2.c above there are a plurality (e.g., 3, or alternatively 20) of mono execution units, and thus there is a "second plurality of processing elements" (in the mono units), as claimed.

209. As Stuttard's Figure 4 (which I have annotated below) illustrates, each mono unit has an ALU (**arithmetic** logic **unit**). Stuttard, 7:14-24 ("…the mono execution unit is likely to require a wider and more complex ***ALU***" than the ALUs in the poly PE array), 3:15-29 ("the mono execution unit has a 32-bit ALU"), 16:13-20 ("The mono execution unit is a 16-bit processing element consisting of: •

A 16-bit ALU with optional extensions, … The mono ALU ***extensions include a multiplier***, a barrel shifter and a normalizer, which is important ***for accelerating software implementations of floating point operations***."), FIGS. 3, 4, 7.

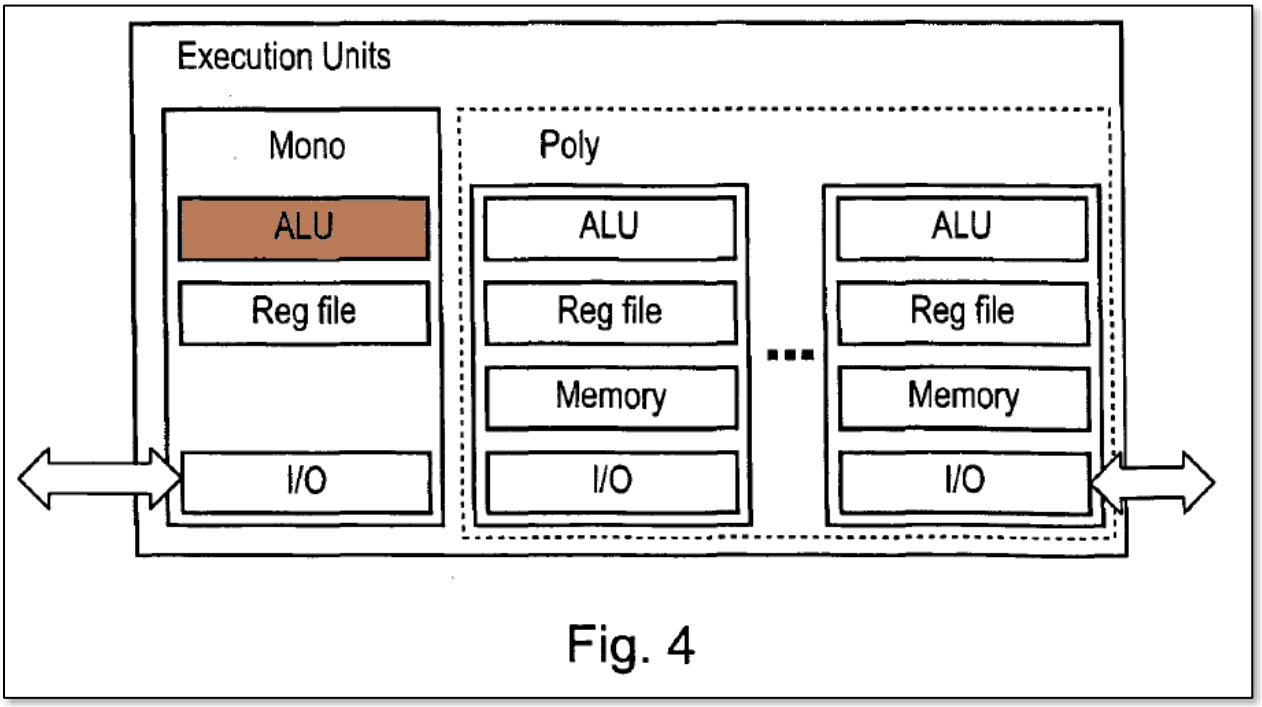


Fig. 4

210. Like the ALU in each poly PE, the ALU in the mono unit also performs "floating point…**multiplies**" (Stuttard, 12:26-28) using floating point "extensions" that "include a multiplier" (Stuttard, 16:13-20). In one embodiment, Stuttard's mono execution unit's ALU is "a 32-bit ALU" (Stuttard, 3:15-18); thus a POSA would have understood that Stuttard teaches providing a 32-bit floating-point multiplier circuit in each mono unit, each of which receive as inputs to the multiplier circuit two floating point values each of width at least 32 bits, as claimed.

211. POSAs also had reason to include in each mono execution unit at least one multiplier circuit capable of multiplying two IEEE standard (full-precision) 32-bit-wide floating-point values, for all the reasons I discussed above in §VIII.A.2.b and discuss in this VIII.C.1. For example, Stuttard taught that "[i]t is not necessary for the [poly] PEs to be identical to the mono execution unit in every detail: e.g. …the **width and complexity of the ALU are likely to be different in practice**." Stuttard, 7:14-24 ("mono execution unit is likely to require a wider and more complex ALU"). This would have suggested to POSAs including in Stuttard-Shirazi's mono unit an FPU that supported higher bit-widths than the reduced-bitwidth (reduced number of bits and, therefore, complexity of) FPUs in the poly unit. POSAs had reason to use the IEEE's 32-bit full-precision values in each mono execution unit because POSAs would have wanted the Stuttard-Shirazi chip to support data and instructions from a conventional, IEEE-standard compliant host. For evidence corroborating my opinion in that regard, EX1012 (at 232-233) for example, teaches that "existing designs" for GPUs (which have SIMD architectures like Stuttard) "combine some number of large coarse-granularity (CG) cores [like Stuttard-Shirazi's mono unit] with a large number of fine-granularity (CG) cores [like Stuttard-Shirazi's PEs in the poly unit]." *See also* EX1034, 274 ***("[M]ost desktop microprocessors support the IEEE 754 standard***" which "use[es] 1 sign bit, 8 bits of exponents, and 23 bits of mantissa").

212. Thus, POSAs would have recognized the benefit in having the floating-point multiplier circuit in each mono execution unit on the chip be adapted to receive as inputs two IEEE-standard 32-bit-wide floating-point values, which meets limitation [8A]. POSAs would have reasonably expected success because, as noted above, Stuttard taught that the ALU in the mono unit could be "wider and more complex" than the ALUs in the poly unit and could be "a 32-bit ALU." Stuttard, 7:14-24 ("***It is not necessary for the PEs to be identical to the mono execution unit*** in every detail: e.g. the number of registers and ***the width and complexity of the ALU are likely to be different in practice***."), 3:15-18 (". For example, the ***poly processing elements could have an 8-bit ALU, while the mono execution unit has a 32-bit ALU***."). *See also* §VIII.A.3 (opinions regarding reasonable expectation of success in scaling up the number PEs and mono units on a Stuttard-Shirazi chip).

2. [8B]: "wherein, other than the plurality of second processing elements, the computing chip has no other processing element that comprises a multiplier circuit adapted to receive as inputs to the multiplier circuit two floating point values each of width at least 32 bits; and"

213. In the second and third obvious implementations scaled Stuttard-Shirazi (Section VIII.A.2.b-VIII.A.2.c above), the only PEs on the chip that comprise a multiplier circuit adapted to receive as inputs two floating-point values each of width at least 32 bits are the multiplier circuits in the plurality of mono

units, which are the plurality of second PEs for the reasons I discussed in Section VIII.C.1 above. The only other multiplier circuits on the chip are the multipliers in the poly units, which are the plurality of first PEs, for the reasons I discussed in Sections VII.B.10, VIII.B.2 above. That plurality of first PEs are only adapted to receive Shirazi's floating point values of 16- or 18-bit width, for the reasons I discussed above in Sections VIII.B.8 and VII.B.11. Thus, scaled Stuttard-Shirazi meets limitation [8B].

3. [8C]: "wherein the plurality of first processing elements is greater in number, by at least 100, than the plurality of second processing elements."

214. In Scaled Stuttard-Shirazi (Section VIII.A above), the plurality of first PEs (in the poly units) number 6144 (or alternatively 5120) PEs, for the reasons I described in Section VIII.B.2 above. Either of those numbers of PEs is greater in number, by at least 100, the three (or alternatively 20) PEs in the plurality of second PEs (which are the PEs in the mono units that I discussed in Section VIII.C.1 above), as claimed.

215. Thus Scaled Stuttard-Shirazi renders obvious claim 8 of the '616 patent for the reasons discussed above.

D. Claim 9: "the plurality of memory units each comprises a corresponding register."

216. As Stuttard's FIG. 4 (which I have annotated below) illustrates, each PE in a poly unit comprises a "register file" ("Reg file"), which is "an array of

bytes which can be addressed as registers" (Stuttard, 15:9-16), as claimed. *See* Section VII.B.7 above (explaining why register files are addressable "memory units"). Stuttard, 5:29-30 ("Each processing element of the SIMD array may comprise…a register file…."), 7:14-24, 10:1-4, 10:18-19, 15:21-25, 17:21-23, 20:31-21:12. For evidence corroborating that understanding, see EX1058 at [0003] teaches that a "'register file' is an array of registers that may be addressed as a unit" and EX1075 at 2:14-25, which teaches that a "register file is a memory array" and that there is an "address word for the register file" (i.e., the register file is addressable as a unit, as taught in EX1058).

217. Thus Scaled Stuttard-Shirazi renders obvious claim 9 of the '616 patent for the reasons discussed above.

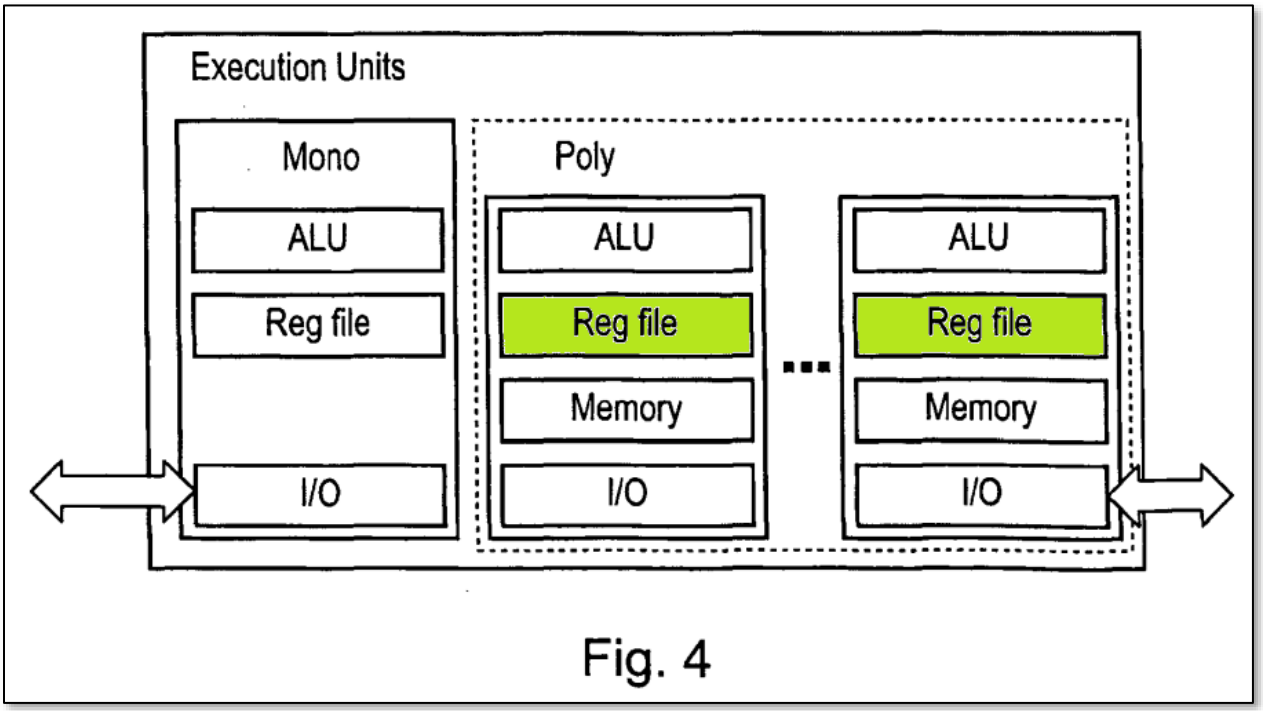


Fig. 4

E. Claim 10: "said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image."

218. The table below shows claim 10 side-by-side with claim 4:

Claim 10	Claim 4
[10] "The computing system of claim 8, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image."	**[4]** "The computing system of claim 1, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image."

219. In comparing the two limitations, a POSA would understand that Scaled Stuttard-Shirazi meets claim 10 for the same reason Stuttard-Shirazi meets claim 4, for the reasons I discussed in Section VII.C above. Thus Scaled Stuttard-Shirazi renders obvious claim 10 of the '616 patent for the reasons discussed above.

F. Claim 11: "an aggregate maximum bandwidth of the plurality of processing element connections exceeds a maximum bandwidth of the host connection."

220. Stuttard teaches that a 64-PE array, running at 400 MHz, has an Inter-PE Bandwidth of 50 Gigabytes/sec (i.e., aggregate bandwidth of 3,200

Gigabytes/sec). Stuttard, 10:15-19. Scaled Stuttard-Shirazi has at least eighty times as many parallel PEs and thus 80 times the aggregate bandwidth as un-scaled Stuttard (i.e., 200+ Terabytes/sec).

221. As I discussed above in Sections VII.B.9, VIII.B.7, Scaled Stuttard-Shirazi's host connection is provided by "external ports" connecting the input-output unit's bus to the host system. Stuttard discloses that more details for the bus unit can be found in "copending patent application GB 0321186.9." Stuttard, 8:7-8. GB 0321186.9 (which is EX1051) was published as WO2002/065700 (which is EX1049). *See* EX1051, codes (21) ("Application No: 0321186.9"), (87) ("International Publication Data: WO2002/065700"). EX1049 (at 39:8-21) disclosed that a ClearConnect bus for a processor like Stuttard's had two datapaths with "raw bandwidth [of] 51.2 Gbits/sec" and another with "raw bandwidth [of] 102.4 Gbits/sec," which are orders of magnitude below Scaled Stuttard-Shirazi's 200+ Terabytes/sec aggregate bandwidth for the poly PEs discussed above.

222. A POSA would also understand that the maximum bandwidth of the host connection would have been no wider than (and likely significantly narrower than) the maximum bandwidth of the ClearConnect bus itself, and thus the aggregate maximum bandwidth of the poly PEs in Scaled Stuttard-Shirazi would have exceeded the maximum bandwidth of the host connection, as claimed. For corroboration in this regard, EX1020 (at 85-86) notes the "relatively low

bandwidth available between the CPU and GPU" in accelerator systems with "SIMD[] capabilities." POSAs would have understood that it would have been conventional that the SIMD Scaled Stuttard-Shirazi processor would have also had a "relatively low bandwidth" to the host system, compared to the aggregate bandwidth of the PE connections.

223. Thus, Scaled Stuttard-Shirazi renders obvious claim 11 of the '616 patent for the reasons discussed above.

G. Claim 12: "the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction."

224. The table below shows claim 12 side-by-side with claim 2:

Claim 12	Claim 2
[12] "The computing system of claim 11, wherein the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction."	**[2]** "The computing system of claim 1, wherein the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction."

225. In comparing the two limitations, a POSA would understand that Scaled Stuttard-Shirazi meets claim 12 for the same reasons Stuttard-Shirazi meets claim 2, for the reasons I discussed in Section VII.C above. For example, in the first obvious implementation I discussed in §VIII.A.2.a, there is one control unit in the mono unit. And in the second and third obvious implementations I discussed in §§VIII.A.2.b-VIII.A.2.c, which duplicate entire MTAP cores on a chip, the control unit in any one of the MTAP cores is a "control unit" that meets claim 12, because it would decode instructions received from the host computer via the input-output unit (e.g., ClearConnect bus) and send control signals to the PE array (either in the same or different cores), for the same reasons the control unit I discussed above in §VII.C in Stuttard-Shirazi sends control signals to the PE array. Thus, Scaled Stuttard-Shirazi renders obvious claim 12 of the '616 patent for the reasons discussed above.

H. Claim 13: "the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory."

226. The table below shows claim 13 side-by-side with claim 4:

Claim 13	Claim 4
[13] "The computing system of claim 12, wherein the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory."	**[4]** "The computing system of claim 1, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image."

227. In comparing the two limitations, a POSA would understand that Scaled Stuttard-Shirazi meets claim 13 for the same reasons Stuttard-Shirazi meets claim 4, for the reasons I discussed in Section VII.E. *See also* §VIII.G above (describing how any of the MTAP core control units in the obvious implementations of Scaled Stuttard-Shirazi is a "control unit," as claimed). Thus, Scaled Stuttard-Shirazi renders obvious claim 13 of the '616 patent for the reasons discussed above.

I. Claim 14: "the width of each [first/second] binary mantissa is no more than 10 bits plus a [first/second] sign bit."

228. The table below shows claim 14 side-by-side with claim 5:

Claim 14	Claim 5
[14] "The computing system of claim 7, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, and wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit."	**[5]** "The computing system of claim 1, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, and wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit."

229. In comparing the two limitations, a POSA would understand that Scaled Stuttard-Shirazi meets claim 14 for the same reasons Stuttard-Shirazi meets claim 5, for the reasons I discussed in Section VII.F above. Thus, Scaled Stuttard-Shirazi renders obvious claim 14 of the '616 patent for the reasons discussed above.

IX. GROUND 1C: CLAIMS 16-22, 24 ARE OBVIOUS OVER STUTTARD IN COMBINATION WITH SHIRAZI

A. Two Connected Stuttard-Shirazi Chips on a Circuit Board

230. As I discuss below, a POSA would have had reasons to mount two Stuttard-Shirazi computing chips (§VII.A.3 above) on a circuit board connected to the host computer. These reasons would have applied whether or not each Stuttard-Shirazi chip is scaled to include over 5000 PEs in their respective PE arrays (which, I note, is not a limitation recited in any of claims 16-25).

231. **First**, Stuttard taught mounting its MTAP chips to a circuit board. Specifically, Stuttard's "evaluation chip EVI" is a "simple embodiment of" Stuttard's invention that used the ClearConnect "bus to interface ***various ports*** of the processor… ***to external pins***." Stuttard, 22:4-7. POSAs understood that it was conventional for "external pins" to form an electrical connection between I/O ports of a computing chip and a circuit board. For evidence corroborating my opinion in that regard, EX1096 at 1:16-28 teaches that in "[c]onventional integrated circuit devices" the "integrated circuit chip is [] packaged" "to enable th[e] chip to be incorporated in a larger electronic system in a ***convenient conventional way, such as by mounting the packages in a printed circuit board***." EX1096 further teaches there that a conventional chip package – a so-called "dual inline package (DIP)" – "may have two or more rows of ***external pins***, with ***connections made*** within the package between pads formed on the integrated circuit chip and the external pins." POSAs would have understood that it this and alternative "packages" for chips, it was conventional for the external pins to be connected to the chip's pads and to a circuit board (which, in turn, connected could be used to connect the chip to other devices, like Stuttard's host computer). *See also* EX1089 (Microsoft Computer Dictionary), 2 ("Copyright © 2002 by Microsoft Corporation"), 3 (first definition of port: "An interface through which data is transferred between a computer and other devices…or a direct connection to another computer. The port appears to the

CPU as one or more memory addresses that it can use to send or receive data. ***Specialized hardware, such as in an add-on circuit board, places data from the device in the memory addresses and sends data from the memory addresses to the device***. Ports may also be dedicated solely to input or to output.").

232. Stuttard's Figure 8 (which I have annotated below) shows that the evaluation chip had a "JTAG TAP," which POSAs understood is an IEEE standard for testing "interconnections…after installation on a **printed circuit board**." EX1059, [0001]-[0002]. Thus, Stuttard taught connecting an MTAP chip to a circuit board.

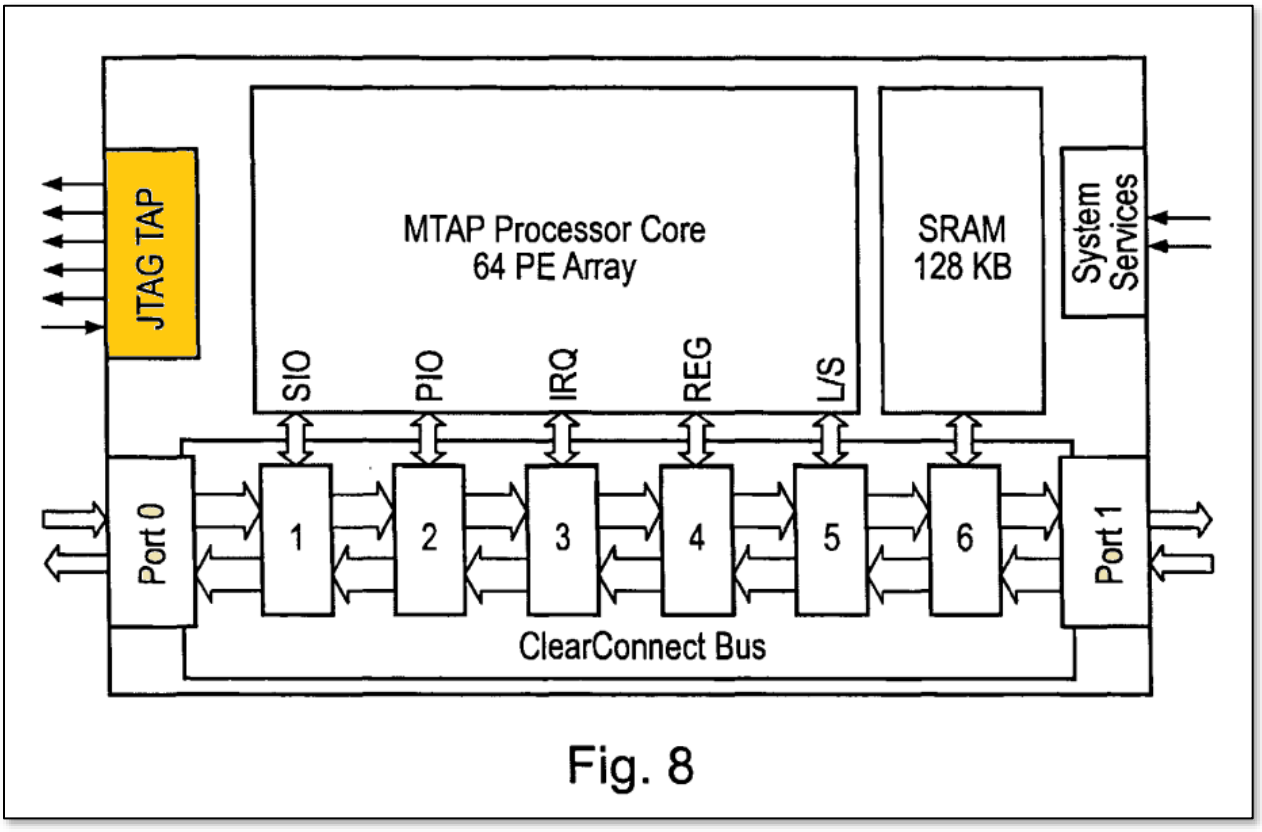


Fig. 8

233. **Second**, Stuttard taught that its system bus allowed a "connect[ion] from **one chip to another**," which would have conventionally been implemented by mounting the multiple chips on a common circuit board. Stuttard, 22:12-15. For evidence corroborating that it was conventional to mount two similar chips to a common circuit board, EX1108 (at 34) noted the disclosure in the prior art of the "ClearSpeed Advance board having ***two*** CSX600 ***chips***"). Similarly, EX1107, (at 822-23), disclosed in FIG. 2 (annotated below) a "board [that] includes 2 SIMD chips" with a double-sided arrow indicating the connection between the two chip via the "ClearConnect Bus" (e.g., via a port like one of the external ports 0 and 1 shown in Figure 8 above) and to an FPGA, which, in turn, has double-sided arrows pointing off the board, which POSAs would have understood indicated the connection between the board (and its chips) to another device, for example, a slot on a motherboard with the host computer and processor. *See also* EX1070,[0001]-[0003] (discussed in the paragraph below).

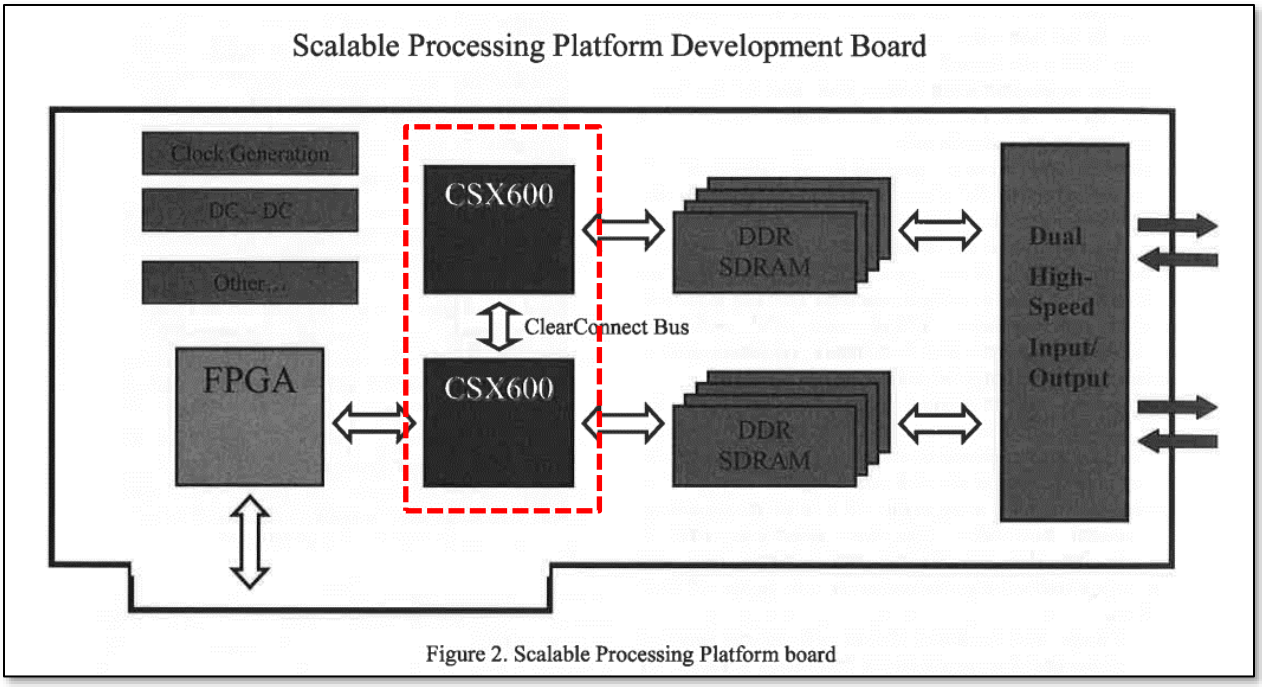


Figure 2. Scalable Processing Platform board

234. **Third**, it was background knowledge for POSAs that incorporating multiple SIMD processors, like Stuttard-Shirazi, on a circuit board connected to a host computer would have been using a known technique (circuit boards) to improve similar devices (coprocessors) in the same way (facilitating incorporation of the coprocessor into a system with a host computer). For evidence corroborating that this was a known technique (in addition to EX1107 and EX1108 above), I note that EX1070 teaches at [0001]-[0003] a system having a "general purpose central processing unit (CPU) and [a] graphics processing subsystem" wherein the "graphics processing subsystem is ***located on a separate circuit board*** that is connected with the motherboard" (having the CPU) and that the "graphics processing subsystem" can be "part of a chip" or "as one or more ***separate***

graphics and/or memory ***chips***." Thus, EX1070 further corroborates that it was a known technique to have multiple chips on a circuit board in communication with a host processor (like a general purpose CPU). POSAs would have understand that such a system would have provided the benefit of, for example, increasing the processing power of the combined two-chip system compared to a single chip system.

235. In this combination implementation, two Stuttard-Shirazi chips are mounted to a common circuit board connected to the host computer using a conventional connection between a host computer and circuit board (e.g., a PCI slot), with all other aspects of each Stuttard-Shirazi chip remaining unchanged.

236. A POSA would have had a reasonable expectation of success in this implementation because, as discussed above, it was conventional to incorporate multiple computing chips, like Stuttard-Shirazi, onto circuit boards connected to a host computer.

1. [16PRE], [16A]-[16H], [16M]

237. As illustrated by the table below, limitations [16PRE] and [16A]-[16H] recite identical limitations as corresponding limitations [1PRE] and [1A]-[1H], which also identifies where I discussed the corresponding limitations in claim 1 above. As also illustrated by the table below, limitation [16M] is identical to [1I] as to the "first" through "fourth" arithmetic units.

Claim 16 Limitation	Claim 1 Limitation
[16PRE] A computing system, comprising:	**[1PRE]** A computing system, comprising: Discussed in §VII.B.1 above
[16A] a host computer;	**[1A]** a host computer; Discussed in §VII.B.2 above
[16B] a computing chip comprising:	**[1B]** a computing chip comprising: Discussed in §VII.B.3 above
[16C] a processing element array comprising a first edge processing element positioned at a first edge of the processing element array, a second edge processing element positioned at the first edge of the processing element array, a first interior processing element positioned at a first location in the interior of the processing element array, and a second interior processing element positioned at a second location in the interior of the processing element array;	**[1C]** a processing element array comprising a first edge processing element positioned at a first edge of the processing element array, a second edge processing element positioned at the first edge of the processing element array, a first interior processing element positioned at a first location in an interior of the processing element array, and a second interior processing element positioned at a second location in the interior of the processing element array; Discussed in §VII.B.4 above
[16D] a first processing element connection connecting the first edge processing element with the first interior processing element; a second processing element connection connecting the second edge processing element with the second interior processing element;	**[1D]** a first processing element connection connecting the first edge processing element with the first interior processing element; a second processing element connection connecting the second edge processing element with the second interior processing element;

	Discussed in §VII.B.5 above
[16E] an input-output unit connected to the first edge processing element and the second edge processing element;	**[1E]** an input-output unit connected to the first edge processing element and the second edge processing element; Discussed in §VII.B.6 above
[16F] a first memory local to the first edge processing element; a second memory local to the second edge processing element; a third memory local to the first interior processing element; and a fourth memory local to the second interior processing element;	**[1F]** a first memory local to the first edge processing element; a second memory local to the second edge processing element; a third memory local to the first interior processing element; and, a fourth memory local to the second interior processing element; Discussed in §VII.B.7 above
[16G] wherein the first edge processing element comprises a first arithmetic unit; wherein the second edge processing element comprises a second arithmetic unit; wherein the first interior processing element comprises a third arithmetic unit; and wherein the second interior processing element comprises a fourth arithmetic unit;	**[1G]** wherein the first edge processing element comprises a first arithmetic unit; wherein the second edge processing element comprises a second arithmetic unit; wherein the first interior processing element comprises a third arithmetic unit; and wherein the second interior processing element comprises a fourth arithmetic unit; and, Discussed in §VII.B.8 above
[16H] a host connection at least partially connecting the input-output unit with the host computer;	**[1H]** a host connection at least partially connecting the input-output unit with the host computer; Discussed in §VII.B.9 above
[16M] wherein the first, second, third, fourth, and fifth arithmetic units each	**[1I]** wherein the first, second, third and fourth arithmetic units each comprises

comprises a first corresponding multiplier circuit adapted to receive as a first input to the first corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the first corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits; and	a corresponding multiplier circuit adapted to receive as a first input to the corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits Discussed in §VII.B.10 above

238. In the implementation that I discussed in §IX.A above that includes two Scaled Stuttard-Shirazi chips on one board connected to Stuttard's "host," the host meets limitation [16A] for the reason discussed in §VII.B.2 above ([1A]). Either one of the Stuttard-Shirazi chips on the board meets [16B] for the reasons I discussed in §VII.B.3 above ([1B]). Stuttard's "host" and the board with two Stuttard-Shirazi chips together meet the "computing system" of [16PRE], for at least the same reason one Stuttard-Shirazi chip and host computer are a "computing chip" (as I discussed above in §VII.B.1 ([1PRE]).

239. The poly-unit PE array in either one of the Stuttard-Shirazi chips on the board meets [16C]-[16D] for the same reasons I discussed above in §§VII.B.4-VII.B.5 ([1C]-[1D]).

240. The ClearConnect bus (or alternatively the bus plus I/O engines and channels) on either Stuttard-Shirazi chip is an "input-output unit" connected to the first and second edge PEs, which meets [16E] for the reasons that I discussed in §VII.B.6 ([1E]). And each Stuttard-Shirazi chip has a claimed "host connection" meeting [16H] for the reasons that I discussed in §VII.B.9 ([1H]) (e.g., either or both of the chips would have had a connection to the host computer via one of their external bus ports).

241. Each reduced-bitwidth poly PE on each Stuttard-Shirazi chip has a "local memory" meeting [16F] and "arithmetic unit" meeting [16G] and [16M] (as to the "first" through "fourth" arithmetic units), for the reasons that I discussed above §§VII.B.7-VII.B.8 ([1F]-[1G]) and in §VII.B.10 ([1I]). I discuss the fifth arithmetic unit further in §IX.A.2 below.

2. [16I], [16K]-[16M]: "an external edge processing element external to the computing chip" comprising "a fifth arithmetic unit" comprising a "multiplier circuit adapted to receive" as inputs a first and second floating point value each having a "binary mantissa of width no more than 11 bits" and a "binary exponent of width at least 6 bits"; and "a fifth memory local to the external edge processing element"

242. With two connected Stuttard-Shirazi chips on one board (as I described above in §IX.A), any edge poly PE on the second Stuttard-Shirazi chip (e.g., any PE at the top, bottom, left-most, or right-most edge of the chip's PE

array) meets [16I] because the edge poly PEs on the second chip are all external to the first chip (the claimed "computing chip").

243. This external edge poly PE would have comprised the same architecture as each other reduced-bitwidth poly PE that exists in the chips on the same circuit board (e.g., all reduced-bitwidth poly PEs on the chips would have had the same architecture for their respective local memory, ALU, and reduced-bitwidth floating-point multiplier circuit). Thus, the external edge PE meets [16K]-[16M] for the same reasons I demonstrated above in §§VII.B.7-VII.B.8 ([1F]-[1G]) and above in §VII.B.10 ([1I]) for claim 1's first-through-fourth PEs.

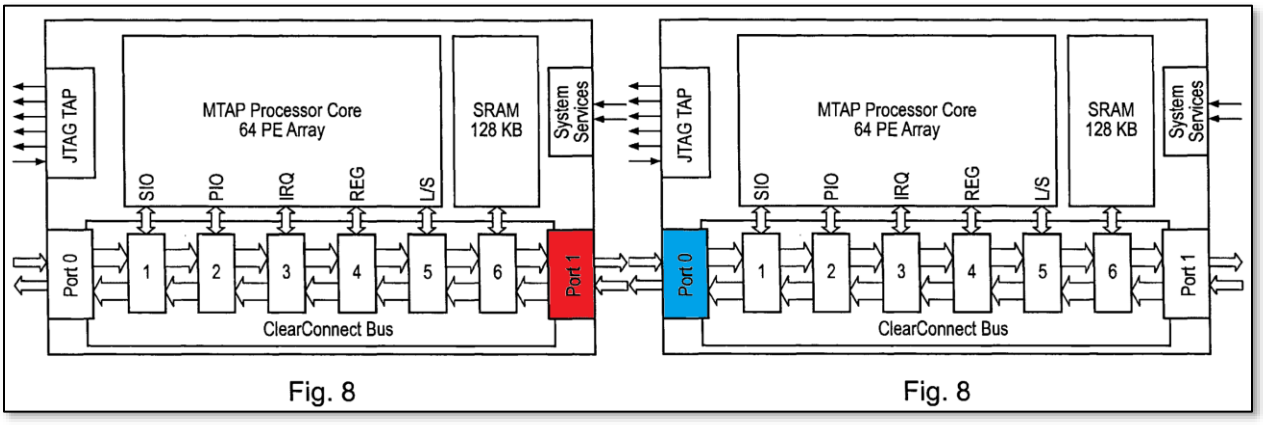
3. [16J] "an external input-output unit external to the computing chip;"

244. With two connected Stuttard-Shirazi chips on a board (*see* §IX.A above), the second chip on the board would have had an input-output unit – its local ClearConnect bus or, alternatively, the bus plus the I/O engines and channels – for the same reasons the first chip has an input-output unit, as I discussed above in §VII.B.6 ([1E]). The second chip's input-output unit is "an external input-output unit external to" the first chip, and thus meets [16J].

4. [16N] "wherein the first edge processing element is connected with the external edge processing element through the input-output unit and the external input-output unit."

245. As Stuttard discloses for an implementation of a real-world chip, Stuttard's chip has "two external ports at each end of the bus allow the ClearConnect bus to be connected ***from one chip to another***." Stuttard, 22:12-15.

246. When placing two Stuttard-Shirazi chips on the same board board (*see* §IX.A above), POSAs would understand that one external port of one Stuttard-Shirazi chip (e.g., "Port 1," which I have annotated red below) would be connected to an external port of the second Stuttard-Shirazi chip on the board (e.g., "Port 0," which I have annotated blue below). (POSAs would have understood that at least one of the other "external" ports below would have been available for connection to the host, e.g., Port 0 on the bottom left.)



247. As I discussed in §VII.B.6 above, each Stuttard-Shirazi chip's bus (or alternatively, the bus plus I/O engines and channels) is an input-output unit connected to every poly PE.

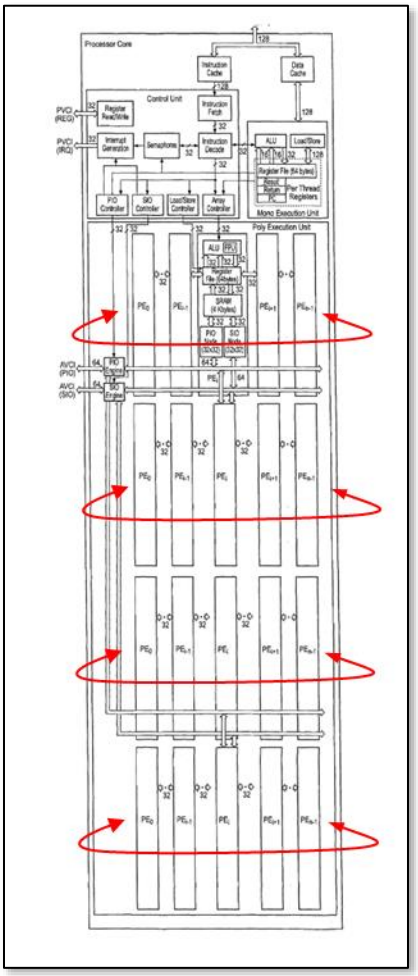
248. Thus, in the implementation having two Stuttard-Shirazi chips on a board that I discussed in § IX.A above, every poly PE in the first Stuttard-Shirazi chip (including the first edge PE, discussed §VII.B.4 above) is connected with every poly PE in the second Stuttard-Shirazi chip (including the external edge PE, discussed §IX.A.2 above) through the first chip's "input-output unit" and the second chip's "external input-output unit," thereby meeting limitation [16N].

249. Thus, the implementation that I discussed in §IX.A above that includes two Scaled Stuttard-Shirazi chips on one board renders obvious claim 16 of the '616 patent for the reasons discussed above.

B. Claim 17: "the first edge processing element is connected with the external edge processing element using a torus interconnect."

250. As the '616 patent itself discloses, a torus interconnect was "well studied" connection scheme that connects "the outside edges" of an array of processing elements and the desirability of such an interconnect to allow communication between the first and last processors in a row were known. EX1001, 10:22-29. In Stuttard-Shirazi's design, POSAs would have achieved a horizontal torus interconnect by, for example, swazzle-connecting a row's first and

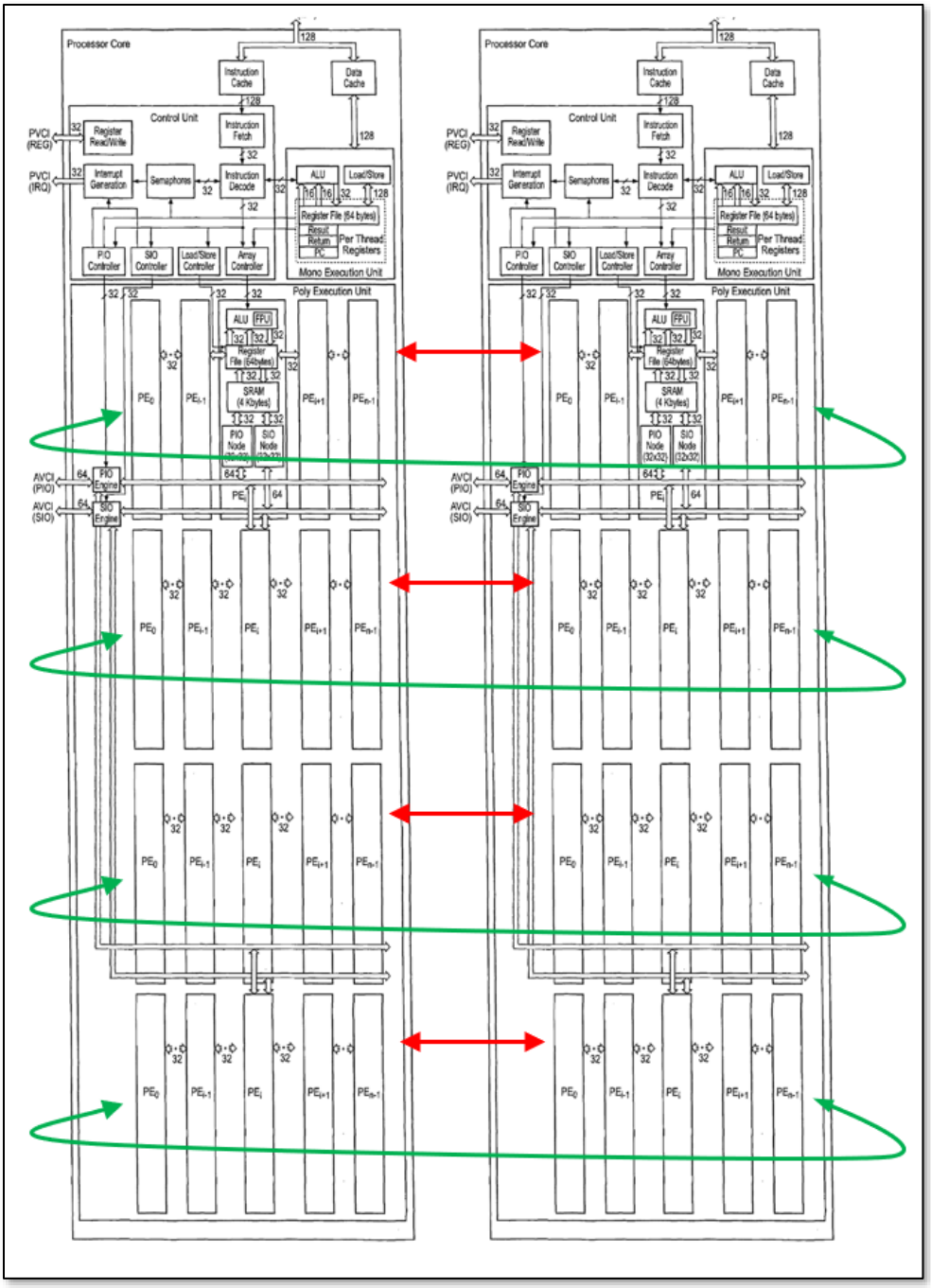
last processors, as I have illustrated below based on the exemplary Stuttard-Shirazi chip that I discussed in §VII.A.3.b above.



251. POSAs had motivation to implement these torus connections to achieve their known benefits, including reducing the number of instructions needed to propagate data between ends of rows. POSAs would have understood that the obvious and desirable *intra*-chip torus interconnection illustrated above would also apply to *inter*-chip connections illustrated below. Thus, POSAs had reason to apply *inter*-chip torus connections to the implementation that I discussed in §IX.A

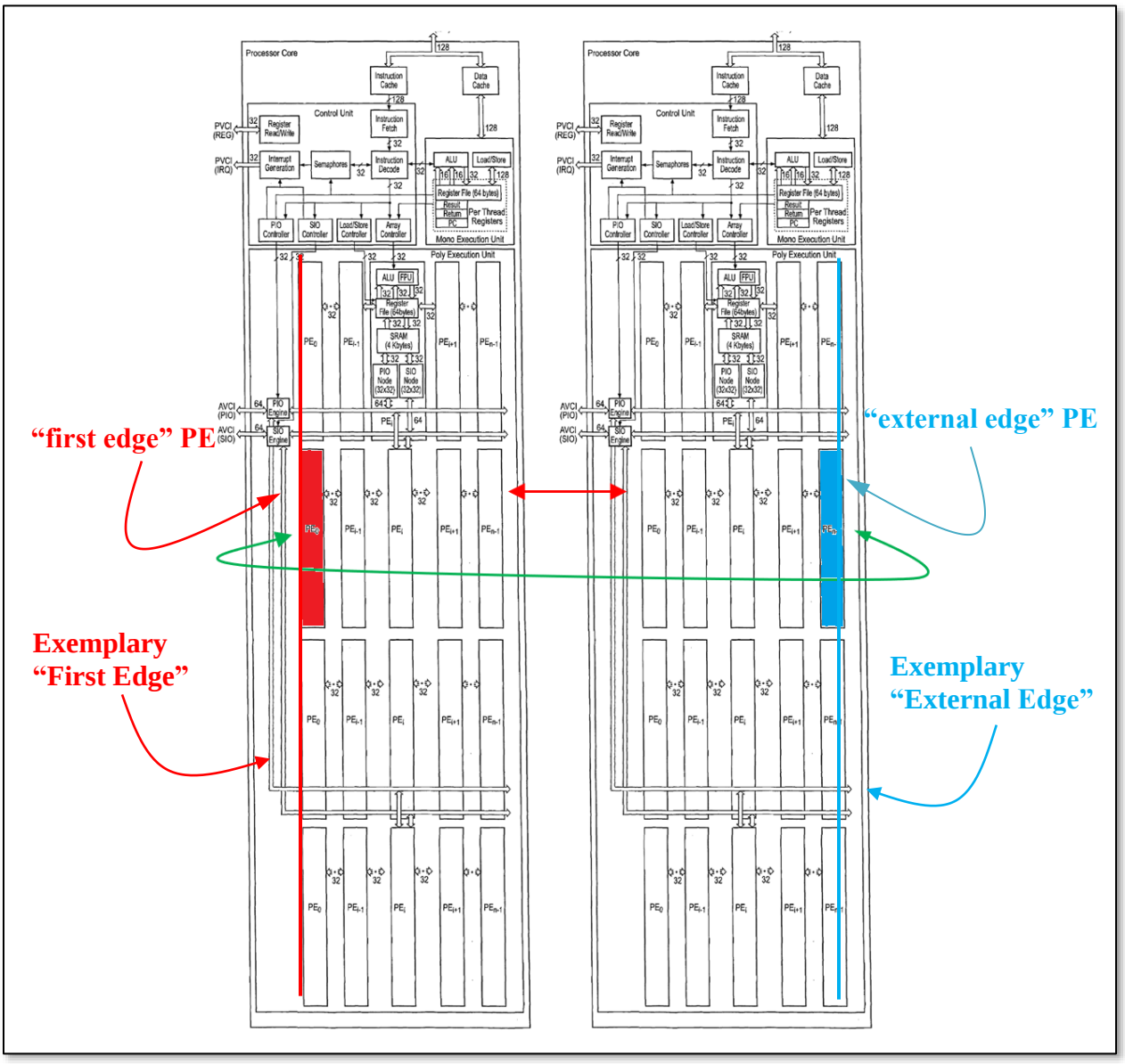
above having two Stuttard-Shirazi chips on a board, for the same well-known benefits of torus interconnections that the ’616 patent acknowledges were known in the prior art.

252. In the demonstrative below, I have illustrated an exemplary implementation of *inter*-chip torus connections for two Stuttard-Shirazi chips on one circuit board (*see* §IX.A). The green torus connections connect each left-edge PE in the left Stuttard-Shirazi chip to a right-edge PE in the right Stuttard-Shirazi chip; and the red connections complete the torus by connecting each right-edge PE in the left Stuttard-Shirazi chip to a left-edge PE in the right Stuttard-Shirazi chip.



253. As I discussed above in §VII.B.4 ([1C]) and §VIII.B.2 ([7C]), any Stuttard-Shirazi poly PE on an edge and adjacent poly PE in the interior of the poly array meet the claimed edge/interior processing elements, and any Stuttard-Shirazi

edge processor on the other chip on the board meets the claimed "external edge processing element." In the demonstrative below, I have annotated an exemplary "first" edge PE (red) and exemplary "external" edge PE (blue) and their torus interconnect (green) which connects these two PEs on the circuit board.



254. Thus, the obvious and conventional use of an inter-chip "torus interconnect" that I described above connects the "first edge processing element" and the "external edge processing element," as claimed. Accordingly, the implementation that I discussed in §IX.A above that includes two Scaled Stuttard-Shirazi chips on one board renders obvious claim 17 of the '616 patent for the reasons discussed above.

C. Claim 18

1. [18A]: "the processing element array further comprises a plurality of first additional processing elements and a plurality of first additional arithmetic units, each of the plurality of first additional processing elements having positioned therein at least one of the plurality of first additional arithmetic units;"

255. Limitation [18A] recites: "wherein the processing element array further comprises a plurality of first additional processing elements and a plurality of first additional arithmetic units, each of the plurality of first additional processing elements having positioned therein at least one of the plurality of first additional arithmetic units." In the implementation that I discussed in §IX.A above that includes two Scaled Stuttard-Shirazi chips on one board, each chip's poly PE array includes many more PEs than just the four PEs that meet limitation [16C], for the reasons I discussed above in §IX.A.1, §VII.B.4. Thus, those PEs meet limitation [18A]'s "first additional processing elements." Each PE has positioned

therein an ALU, thereby meeting [18A]'s "first additional arithmetic units," for the reasons I discussed above in §VII.B.8.

2. [18B]: "wherein the plurality of first additional arithmetic units each comprises a second corresponding multiplier circuit adapted to receive" As Third and Fourth Inputs Floating-Point Values of the Same Bitwidths As Recited in [1I]

256. Limitation 18[B] recites: "wherein the plurality of first additional arithmetic units each comprises a second corresponding multiplier circuit adapted to receive as a third input to the second corresponding multiplier circuit a third floating point value having a third binary mantissa of width no more than 11 bits and a third binary exponent of width at least 6 bits, and to receive as a fourth input to the second corresponding multiplier circuit a fourth floating point value having a fourth binary mantissa of width no more than 11 bits and a fourth binary exponent of width at least 6 bits." For the reasons I discussed above in §VII.B.10 ([1I]), each poly PE's ALU in each Stuttard-Shirazi chip comprises a multiplier circuit adapted to receive the inputs recited in [18B] (and [1I]).

3. [18C]: "wherein the computing chip further comprises a plurality of second additional processing elements" Having the Same Limitations As Limitation [8A]'s "second processing elements"

257. Limitation [18C] recites: "wherein the computing chip further comprises a plurality of second additional processing elements, wherein the plurality of second additional processing elements each comprises a third

corresponding multiplier circuit adapted to receive as inputs to the third corresponding multiplier circuit two floating point values each of width at least 32 bits." As I discussed above in §§VIII.A.2.b-VIII.A.2.c, POSAs had motivation to implement, and a reasonable expectation of success in implementing, each Stuttard-Shirazi chip on the board discussed in §IX.A above to include a duplicated MTAP architecture having a plurality of mono-unit, each with a 32-bit PE (e.g., an IEEE-standard compatible PE). The mono-unit PEs in either of those obvious implementations of the Stuttard-Shirazi chip meet limitations [18C]'s "plurality of second additional processing elements" for the same reasons as [8A]'s "second processing elements" (as I discussed above in §VIII.C.1).

258. Thus, obvious implementations of two Scaled Stuttard-Shirazi chips on one board that I discussed in §IX.A above meet limitation 18[C] for the reasons discussed above.

4. [18D]-[18E]

259. As illustrated in the table below, limitations [18D]-[18E] recite identical limitations for the first and second "additional processing elements" as [8B]-[8C] recite for the first and second "processing elements."

Claim 18 Limitation	Claim 8 Limitation
[18D] wherein other than the second additional processing elements the computing chip has no other processing element that comprises a multiplier circuit adapted to receive as inputs to the multiplier circuit two floating point values each of width at least 32 bits; and,	**[8B]** wherein, other than the plurality of second processing elements, the computing chip has no other processing element that comprises a multiplier circuit adapted to receive as inputs to the multiplier circuit two floating point values each of width at least 32 bits; and: Discussed above in §VIII.C.2 above
[18E] wherein the plurality of first additional processing elements is greater in number, by at least 100, than the plurality of second additional processing elements.	**[8C]** wherein the plurality of first processing elements is greater in number, by at least 100, than the plurality of second processing elements. Discussed above in §VIII.C.3

260. The poly-unit PEs ("first" PEs) and mono-unit PEs ("second" PEs) in either Stuttard-Shirazi chip in the obvious implementations discussed above in §IX.C.3 ([18C]) meet limitations [18D]-[18E] for the same reasons that I discussed above in §§VIII.C.2-VIII.C.3 ([8B]-[8C]). Accordingly, obvious implementations

of two Scaled Stuttard-Shirazi on a circuit board render obvious claim 18 of the '616 patent for the reasons discussed above.

D. Claims 19-22

261. As illustrated in the table below, except for differences in claim dependencies, claims 19-22 recite the identical limitations as claims 10, 12, 13, and 14, respectively.

'616 Patent Claim	**'616 Patent Claim**
[19] "The computing system of claim 18, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image.	**[10]** "The computing system of claim 8, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image. Discussed above in §VIII.E; *see also* §VII.E (claim 4)
[20] "The computing system of claim 18, wherein the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate	**[12]** "The computing system of claim 11, wherein the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate

according to the at least one instruction.	according to the at least one instruction. Discussed above in §VIII.G; *see also* §VII.C (claim 2)
[21] "The computing system of claim 20, wherein the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory.	**[13]** "The computing system of claim 12, wherein the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory. Discussed above in §VIII.H; *see also* §VII.D (claim 3)
[22] "The computing system of claim 16, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, and wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit.	**[14]** "The computing system of claim 7, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, and wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit. Discussed above in §VIII.I; *see also* §VII.F (claim 5)

262. Thus, the obvious implementations that I discussed above in §§VIII.A.2.b-VIII.A.2.c and IX.C.3 including two Stuttard-Shirazi chips (each having a plurality of 32-bit multipliers in their mono-units) meet the limitations of claims 19-22 for the same reasons one Scaled Stuttard-Shirazi chip meets the

corresponding limitations in claims 10 and 12-14, for the reasons discussed in the sections of my report identified in the table above. *See also* §VIII.G above (describing how any MTAP core control unit in the obvious implementations of Scaled Stuttard-Shirazi is a "control unit" as claimed). Accordingly, obvious implementations of two Scaled Stuttard-Shirazi on a circuit board render obvious claims 19-22 of the '616 patent for the reasons discussed above.

E. Claim 24: "The computing system of claim 18, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit, wherein the width of each third binary mantissa is no more than 10 bits plus a third sign bit, and wherein the width of each fourth binary mantissa is no more than 10 bits plus a fourth sign bit."

263. Claim 24 recites that the width of each binary mantissa received by claim 16's "first" and claim 18's "second" multiplier circuits is no more than 10 bits plus a sign bit. As I discussed above in §§IX.A.1, IX.C.2, the claimed "first" and "second" multiplier circuits are met by each Stuttard-Shirazi chip's poly-PE multipliers. Those multipliers meet claim 24's bitwidth limitations for the same reasons that I discussed above in §VII.F above (claim 5) and §VIII.I (claim 14).

264. Thus, obvious implementations of two Scaled Stuttard-Shirazi on a circuit board render obvious claim 24 of the '616 patent for the reasons discussed above.

X. CONCLUSION

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

I declare under penalty of perjury that the foregoing is true and correct.

Dated: 12/22/2022



XI. CLAIM LISTING

The following claim listing assigns element labels (e.g., [1PRE], [1A], etc.) to certain claims for convenience of reference.

Claim 1
[1PRE] A computing system, comprising:
[1A] a host computer;
[1B] a computing chip comprising:
[1C] a processing element array comprising a first edge processing element positioned at a first edge of the processing element array, a second edge processing element positioned at the first edge of the processing element array, a first interior processing element positioned at a first location in an interior of the processing element array, and a second interior processing element positioned at a second location in the interior of the processing element array;
[1D] a first processing element connection connecting the first edge processing element with the first interior processing element; a second processing element connection connecting the second edge processing element with the second interior processing element;
[1E] an input-output unit connected to the first edge processing element and the second edge processing element;
[1F] a first memory local to the first edge processing element; a second memory local to the second edge processing element; a third memory local to the first interior processing element; and, a fourth memory local to the second interior processing element;
[1G] wherein the first edge processing element comprises a first arithmetic unit; wherein the second edge processing element comprises a second arithmetic unit; wherein the first interior processing element comprises a third arithmetic unit; and wherein the second interior processing element comprises a fourth arithmetic unit; and,
[1H] a host connection at least partially connecting the input-output unit with the host computer;
[1I] wherein the first, second, third and fourth arithmetic units each comprises a corresponding multiplier circuit adapted to receive as a first input to the corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits.

Claim 2
[2A] The computing system of claim 1, wherein the computing chip further comprises a control unit connected to the processing element array,
[2B] the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and
[2C] to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction.
Claim 3
The computing system of claim 2, wherein the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory.
Claim 4
The computing system of claim 1, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image.
Claim 5
The computing system of claim 1, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, and wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit.
Claim 6
The computing system of claim 5, wherein the width of each first binary exponent is at least 5 bits plus a third sign bit and wherein the width of each second binary exponent is at least 5 bits plus a fourth sign bit.
Claim 7
[7PRE] A computing system, comprising:
[7A] a host computer;
[7B] a computing chip comprising:
[7C] a processing element array comprising a plurality of first processing elements, wherein the plurality of first processing elements is no less than 5000 in number, wherein each of a first subset of the plurality of first processing elements is positioned at a first edge of the processing element array, and wherein each of a second subset of the plurality of first processing elements is positioned in the interior of the processing element array;
[7D] an input-output unit connected to each of the first subset of the plurality of first processing elements;

[7E] a plurality of processing element connections, each processing element connection connecting one of the plurality of first processing elements with another of the plurality of first processing elements, wherein each of the plurality of first processing elements is connected to at least one other of the plurality of first processing elements by at least one of the plurality of processing element connections;
[7F] a plurality of memory units, wherein each of the plurality of first processing elements is associated with a corresponding one of the plurality of memory units, and wherein each of the plurality of memory units is local to its associated one of the plurality of first processing elements; and,
[7G] a plurality of arithmetic units, wherein each of the plurality of first processing elements has positioned therein at least one of the plurality of arithmetic units; and,
[7H] a host connection at least partially connecting the input-output unit with the host computer;
[7I] wherein the plurality of arithmetic units each comprises a first corresponding multiplier circuit adapted to receive as a first input to the first corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the first corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits.
Claim 8
[8PRE] The computing system of claim 7,
[8A] wherein the computing chip further comprises a plurality of second processing elements, wherein the plurality of second processing elements each comprises a second corresponding multiplier circuit adapted to receive as inputs to the second corresponding multiplier circuit two floating point values each of width at least 32 bits;
[8B] wherein, other than the plurality of second processing elements, the computing chip has no other processing element that comprises a multiplier circuit adapted to receive as inputs to the multiplier circuit two floating point values each of width at least 32 bits; and
[8C] wherein the plurality of first processing elements is greater in number, by at least 100, than the plurality of second processing elements.
Claim 9
The computing system of claim 7, wherein the plurality of memory units each comprises a corresponding register.

Claim 10
The computing system of claim 8, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image.
Claim 11
The computing system of claim 8, wherein an aggregate maximum bandwidth of the plurality of processing element connections exceeds a maximum bandwidth of the host connection.
Claim 12
The computing system of claim 11, wherein the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction.
Claim 13
The computing system of claim 12, wherein the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory.
Claim 14
The computing system of claim 7, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, and wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit.
Claim 15
The computing system of claim 14, wherein the width of each first binary exponent is at least 5 bits plus a third sign bit and wherein the width of each second binary exponent is at least 5 bits plus a fourth sign bit.
Claim 16
[16PRE] A computing system, comprising:
[16A] a host computer;
[16B] a computing chip comprising:

[16C] a processing element array comprising a first edge processing element positioned at a first edge of the processing element array, a second edge processing element positioned at the first edge of the processing element array, a first interior processing element positioned at a first location in the interior of the processing element array, and a second interior processing element positioned at a second location in the interior of the processing element array;
[16D] a first processing element connection connecting the first edge processing element with the first interior processing element; a second processing element connection connecting the second edge processing element with the second interior processing element;
[16E] an input-output unit connected to the first edge processing element and the second edge processing element;
[16F] a first memory local to the first edge processing element; a second memory local to the second edge processing element; a third memory local to the first interior processing element; and a fourth memory local to the second interior processing element;
[16G] wherein the first edge processing element comprises a first arithmetic unit; wherein the second edge processing element comprises a second arithmetic unit; wherein the first interior processing element comprises a third arithmetic unit; and wherein the second interior processing element comprises a fourth arithmetic unit;
[16H] a host connection at least partially connecting the input-output unit with the host computer;
[16I] an external edge processing element external to the computing chip;
[16J] an external input-output unit external to the computing chip; and,
[16K] a fifth memory local to the external edge processing element;
[16L] wherein the external edge processing element comprises a fifth arithmetic unit;
[16M] wherein the first, second, third, fourth, and fifth arithmetic units each comprises a first corresponding multiplier circuit adapted to receive as a first input to the first corresponding multiplier circuit a first floating point value having a first binary mantissa of width no more than 11 bits and a first binary exponent of width at least 6 bits, and to receive as a second input to the first corresponding multiplier circuit a second floating point value having a second binary mantissa of width no more than 11 bits and a second binary exponent of width at least 6 bits; and
[16N] wherein the first edge processing element is connected with the external edge processing element through the input-output unit and the external input-output unit.

Claim 17
The computing system of claim 16, wherein the first edge processing element is connected with the external edge processing element using a torus interconnect.
Claim 18
[18PRE] The computing system of claim 16,
[18A] wherein the processing element array further comprises a plurality of first additional processing elements and a plurality of first additional arithmetic units, each of the plurality of first additional processing elements having positioned therein at least one of the plurality of first additional arithmetic units;
[18B] wherein the plurality of first additional arithmetic units each comprises a second corresponding multiplier circuit adapted to receive as a third input to the second corresponding multiplier circuit a third floating point value having a third binary mantissa of width no more than 11 bits and a third binary exponent of width at least 6 bits, and to receive as a fourth input to the second corresponding multiplier circuit a fourth floating point value having a fourth binary mantissa of width no more than 11 bits and a fourth binary exponent of width at least 6 bits;
[18C] wherein the computing chip further comprises a plurality of second additional processing elements, wherein the plurality of second additional processing elements each comprises a third corresponding multiplier circuit adapted to receive as inputs to the third corresponding multiplier circuit two floating point values each of width at least 32 bits;
[18D] wherein other than the second additional processing elements the computing chip has no other processing element that comprises a multiplier circuit adapted to receive as inputs to the multiplier circuit two floating point values each of width at least 32 bits; and,
[18E] wherein the plurality of first additional processing elements is greater in number, by at least 100, than the plurality of second additional processing elements.
Claim 19
The computing system of claim 18, wherein said host computer is programmed to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation whose output is used to identify at least one image, from a plurality of images to be searched, that is similar to at least one input image.

Claim 20
The computing system of claim 18, wherein the computing chip further comprises a control unit connected to the processing element array, the control unit comprising circuitry adapted to decode at least one instruction received from the host computer via the input-output unit, and to send at least one control signal to the processing element array to cause the processing element array to operate according to the at least one instruction.
Claim 21
The computing system of claim 20, wherein the computing chip further comprises an instruction memory adapted to store the at least one instruction received from the host computer via the input-output unit, wherein the control unit is further adapted to retrieve the at least one instruction from the instruction memory.
Claim 22
The computing system of claim 16, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, and wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit.
Claim 23
The computing system of claim 22, wherein the width of each first binary exponent is at least 5 bits plus a third sign bit and wherein the width of each second binary exponent is at least 5 bits plus a fourth sign bit.
Claim 24
The computing system of claim 18, wherein the width of each first binary mantissa is no more than 10 bits plus a first sign bit, wherein the width of each second binary mantissa is no more than 10 bits plus a second sign bit, wherein the width of each third binary mantissa is no more than 10 bits plus a third sign bit, and wherein the width of each fourth binary mantissa is no more than 10 bits plus a fourth sign bit.
Claim 25
The computing system of claim 24, wherein the width of each first binary exponent is at least 5 bits plus a fifth sign bit, wherein the width of each second binary exponent is at least 5 bits plus a sixth sign bit, wherein the width of each third binary exponent is at least 5 bits plus a seventh sign bit, and wherein the width of each fourth binary exponent is at least 5 bits plus an eighth sign bit.

APPENDIX

APPENDIX TABLE OF CONTENTS

I. OVERVIEW OF SEVERAL RESOURCES....3

A. Internet Archive....3

B. IEEE and IEEE Xplore....5

C. ACM and ACM Digital Library....6

II. OPINIONS REGARDING PUBLIC ACCESSIBILITY....8

A. Public Accessibility of the Schmidt Articles (EX1007, EX1009-EX1011)....8

1. Public Accessibility of Schmidt-2000 (EX1007)....8

2. Public Accessibility of Schmidt-1999 (EX1009)....15

3. Public Accessibility of Schimmler-1994 (EX1010)....16

4. Public Accessibility of Schimmler-1996 (EX1011)....18

B. Public Accessibility of NTRS/ITRS Reports (EX1021-EX1027, EX1030)....19

1. Public Accessibility of NTRS 1997 (EX1021: "NTRS-1997")....24

2. Public Accessibility of ITRS 1999 (EX1022: "ITRS-1999")....26

3. Public Accessibility of ITRS 2000 Update (EX1023: "ITRS-2000")....29

4. Public Accessibility of ITRS 2001 (EX1030: "ITRS-2001")....31

5. Public Accessibility of ITRS 2002 Update (EX1024: "ITRS-2002")....33

6. Public Accessibility of ITRS 2004 Update (EX1025: "ITRS-2004")....36

7. Public Accessibility of ITRS 2007 (EX1026: "ITRS-2007")....38

8. Public Accessibility of ITRS 2009 (EX1027: "ITRS-2009")....40

C. Public Accessibility of Additional Prior Art....42

1. Public Accessibility of EX1012 ("Yeh-ISCA07")....42

2. Public Accessibility of EX1013 ("Yeh-Micro07")....48

3. Public Accessibility of EX1014 ("Yeh Fool Me Twice")....50

4. Public Accessibility of EX1016 ("Moore 1965")....52

5. Public Accessibility of EX1017 (“Moore 1975”) 53
6. Public Accessibility of EX1018 (“Mendocino”) 54
7. Public Accessibility of EX1019 (“Sim”) 56
8. Public Accessibility of EX1020 (“Taveniku”).....58
9. Public Accessibility of EX1038 (“Altera Product Catalog”).....59
10. Public Accessibility of EX1039 (“Intel-Itanium”).....62
11. Public Accessibility of EX1040 (“Janakiraman”) 64
12. Public Accessibility of EX1041 (“Radeon-5700”) 66
13. Public Accessibility of EX1042 (“Penryn”) 69
14. Public Accessibility of EX1043 (“TSMC 40nm Press Release 1”) 73
15. Public Accessibility of EX1044 (“CPU Cheatsheet”) 74
16. Public Accessibility of EX1045 (“Natarajan”) 76
17. Public Accessibility of EX1052 (“Intel’s 32nm Update”).....77
18. Public Accessibility of EX1053 (“AMD Evergreen”).....80
19. Public Accessibility of EX1088 (“VSI Standard”) 81
20. Public Accessibility of EX1094 (“Chen”) 84
21. Public Accessibility of EX1097 (“Schiml”) 86
22. Public Accessibility of EX1098 (“Sudha”).....87
23. Public Accessibility of EX1100 (“Oh”).....88
24. Public Accessibility of EX1103 (“Prabhu”) 89
25. Public Accessibility of EX1104 (“Poli”) 90
26. Public Accessibility of EX1105 (“Ligon”) 91
27. Public Accessibility of EX1106 (“Held”).....93
28. Public Accessibility of EX1107 (“Reddaway”).....95
29. Public Accessibility of EX1108 (“Lokhmotov”).....96
30. Public Accessibility of EX1110 (“Shahbahrami”).....98
31. Public Accessibility of EX1111 (“Noumsi”).....99
32. Public Accessibility of EX1113 (“Pengdong”).....101

33. Public Accessibility of EX1114 (“Bradley”) 102
34. Public Accessibility of EX1115 (“Haran”) 104
35. Public Accessibility of EX1116 (“Li”) 105
36. Public Accessibility of EX1117 (“Borkar”).....106
37. Public Accessibility of EX1118 (“Hannibal”) 107
38. Public Accessibility of EX1120 (“Manocha”).....110
39. Public Accessibility of EX1124 (“Burger”).....111
40. Public Accessibility of EX1125 (“Roma”) 112
41. Public Accessibility of EX1126 (“Intel 22nm”) 113
42. Public Accessibility of EX1137 (“TSMC 40nm Press Release 2”) 117
43. Public Accessibility of EX1034 (“Tong”) 118
44. Public Accessibility of EX1036 (“Shirazi”) 118
45. Public Accessibility of EX1092 (“Aty”).....119
46. Public Accessibility of EX1138 (“Thornton”).....119

1. I understand from counsel that a printed publication qualifies as publicly accessible as of the date it was disseminated or otherwise made available such that any member of the public interested in the subject matter could locate it through the exercise of reasonable diligence.

2. While I understand that the determination of public accessibility under the foregoing standard rests on a case-by-case analysis of the facts particular to an individual publication, I also understand that a printed publication is rendered "publicly accessible" if it is cataloged and indexed by a library such that a member of the public interested in the subject matter a could locate it exercising reasonable diligence (i.e., I understand that cataloging and indexing by a library in a manner that permits a member of the public interested in the subject matter to locate the publication is sufficient, though there are other ways that a printed publication may qualify as publicly accessible, such as being indexed and catalogued in an online database). One manner of sufficient indexing is indexing according to subject matter category. I understand that the cataloging and indexing by a single library of a single instance of a particular printed publication is sufficient, even if the single library is in a foreign country. I also understand that the cataloging and indexing of information that would guide a person interested in the relevant subject matter to the printed publication, such as the cataloging and indexing of an abstract

for the printed publication, is sufficient to render the printed publication publicly accessible.

3. I understand that a document may also be publicly accessible if it is catalogued and indexed in an online digital database such that a person interested in the relevant subject matter could locate the document in the database exercising reasonable diligence. For example, as Dr. Bolin discussed in paragraphs 17-18 of her declaration, the ACM Digital Library and IEEE Xplore are examples of digital databases that are well-regarded among the academic community and considered to be accurate and trustworthy sources of information regarding the scientific and academic publications catalogued in those databases. As discuss in §§I.B-I.C below, I have personally used these databases and, in my opinion, POSAs would have been aware of these databases and would have used them to search for and retrieve publications in the art of computing and related fields at the time of the alleged invention.

4. I understand that routine business practices, such as general library cataloging and indexing practices, can be used to establish an approximate date on which a printed publication became publicly accessible. I further understand that the date on which digital databases, like IEEE Xplore and ACM Digital Library, indicate a document was published and/or added to the database can be used to establish a date on which the document became publicly accessible.

5. I agree with Dr. Bolin that POSAs would have been engaged in research, learning through study and practice in their field, and possibly through formal instruction, the bibliographic resources relevant to their research. *See* Bolin, ¶25. I also agree with Dr. Bolin that in the late-2000s POSAs would have had access to a vast array of long-established print and online resources providing indexing information, abstracts, and full-text services for publications in the field of computing. Bolin, ¶25.

6. I agree with Dr. Bolin that POSAs would have been able to locate the material discussed in Dr. Bolin's declaration as well as any additional documents I discuss in the sections below on their own or with the assistance of a research librarian, with relative ease using the tools and resources described in Dr. Bolin's declaration and in the sections below. *See* Bolin, ¶26.

I. OVERVIEW OF SEVERAL RESOURCES

A. Internet Archive

7. Exhibits 1029, 1031-1032, 1047-1048, and 1123 were provided by Nathaniel E Frank-White, the Records Request Processor at the Internet Archive, regarding various websites and files that were captured by the Internet Archive at the indicated dates and made browsable by the Internet Archive's Wayback Machine. The Internet Archive's Wayback Machine is well known to researchers and the general public as a service that provides archives of webpages as they

existed on dates in the past. I am personally familiar with the Internet Archive and its Wayback Machine from my own use of the service in my work and research. It is well known that the Internet Archive collects and archives captures of webpages using web crawlers, which are software programs that surf the Web, visiting websites and automatically storing copies of the webpages and other files represented by the URLs that the crawler visits. Crawlers are also used by Internet search engines (e.g., Google, Bing, etc.) to access webpages to be provided as search results, as is also well known to researchers and the general public. When the Internet Archive's crawlers visit a URL on a particular date, they store a copy of that webpage or file at the URL as it existed on that date. Users of the Wayback Machine can later access that archived capture of the URL to view and interact with the webpage or retrieve a file as it existed on that past date.

8. It is well known to users of the Internet Archive's Wayback Machine, and I am personally aware from my own use of the Wayback Machine, that the date on which an archived webpage was captured by the Internet Archive's crawlers can be observed in the URL that the Internet Archive assigns to the archived capture (what the Internet Archive calls an "extended URL"). When an archived capture of a webpage is viewed through the Wayback Machine on the Internet Archive's website (http://web.archive.org), the URL of the archived capture is http://web.archive.org/web, followed by numbers indicating the year,

month, day, and time at which the webpage was captured by the Internet Archive's crawlers, followed by the original URL of the archived webpage. For example, a Wayback Machine URL of http://web.archive.org/web/20111231171057/http://www.uspto.gov:80/ is the Internet Archive's archived capture of the USPTO's homepage (http://www.uspto.gov:80/) captured on December 31, 2011 (2011/12/31), at 17:10:57 (5:10 p.m. and 57 seconds). This general practice of the Internet Archive for assigning URLs indicating the date and time of capture of an archived URL is well known to the Wayback Machine's users (including myself), and is also confirmed by the Frank-White declarations (*e.g.*, EX1029, ¶5).

B. IEEE and IEEE Xplore

9. The Institute of Electrical and Electronics Engineers (IEEE) "is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity."[1] According to the IEEE official homepage, there are more than 409,000 IEEE members in more than 160 countries. The IEEE's website includes a digital database (currently called IEEE Xplore) for searching and retrieving copies of publications of the IEEE. The IEEE Xplore digital library currently has more than 5 million documents and more than 15

[1] https://www.ieee.org/about/at-a-glance.html (last visited Dec. 19, 2022)

million downloads each month. I have personally used the IEEE Xplore digital database (and the IEEE's predecessor digital database) to retrieve copies of articles or presentations published by the IEEE and in my opinion POSA at the time of the alleged invention would have been aware of IEEE publications and could have retrieved an article or presentation published by the IEEE by exercising reasonable diligence (e.g., by obtaining a copy of an article from a library, by being a member of or subscriber of the IEEE publication, or by navigating to the IEEE's digital database).[2] *See also* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines").

C. ACM and ACM Digital Library

10. The Association for Computing Machinery (ACM) is "the world's largest educational and scientific computing society, delivers resources that advance computing as a science and a profession."[3] According to the ACM's

[2] *See, e.g.*, August 9, 2008 Wayback Machine archive of IEEE Xplore, at https://web.archive.org/web/20080809051147/http://ieeexplore.ieee.org:80/Xplore/guesthome.jsp

[3] https://www.acm.org/about-acm (last visited Dec. 19, 2022)

official homepage, there are "more than 100,000 members" of the ACM.[4] The ACM's website links to a digital database (currently called the ACM Digital Library), which is "the world's most comprehensive database of full-text articles and bibliographic literature covering computing and information technology. This renowned repository includes the complete collection of ACM publications plus an extended bibliographic database of core works in computing from scholarly publishers."[5] I have personally used the ACM Digital Library to retrieve copies of articles or presentations published in journals in the field of computing and information technology and in my opinion POSA at the time of the alleged invention would have been aware of the ACM and the ACM Digital Library and could have retrieved an article or presentation published by the ACM or publishers indexed by the ACM Digital Library by exercising reasonable diligence (e.g., by obtaining a copy of an article from a library, by being a member of or subscriber of

[4] https://www.acm.org/about-acm/about-the-acm-organization (last visited Dec. 19, 2022)

[5] https://www.acm.org/publications/digital-library (last visited Dec. 19, 2022); *see also* https://dl.acm.org/ (last visited Dec. 19, 2022)

the publication, or by navigating to the ACM Digital Library).[6] *See also* Bolin, ¶17 (identifying ACM Digital Library as "a well-known and widely used resource for researchers, educators, and professionals in the field of computer science and related disciplines").

II. OPINIONS REGARDING PUBLIC ACCESSIBILITY

A. Public Accessibility of the Schmidt Articles (EX1007, EX1009-EX1011)

1. Public Accessibility of Schmidt-2000 (EX1007)

11. EX1007 (which I refer to herein as "Schmidt-2000") is titled "KPROC – An Instruction Systolic Architecture for Parallel Prefix Applications." Based on the evidence I discuss below, it is my opinion that Schmidt-2000 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence.

12. According to the face of EX1007, Schmidt was published in the June 2000 issue of *Parallel and Distributed Computing Practices* (PDCP) by Nova Science Publishers, Inc. (Huntington, NY 2000). Page 163 of Schmidt-2000

[6] *See, e.g.*, June 24, 2006 Wayback Machine archive of the ACM Digital Library, at https://web.archive.org/web/20060624020808/http://portal.acm.org:80/dl.cfm

similarly bears the text "© 2000 Nova Science Publishers, Inc." indicating it was copyrighted and thus publicly accessible by the year 2000. According to Dr. Bolin, the PDCP journal has a creation date of December 9, 1997 in the OCLC MARC database and the University of Hong Kong has the PDCP journal in its collection and online catalog. *See* Bolin, ¶¶28-30; *see also id.* ¶¶7-15 (discussing MARC records). EX1007 is the copy of Schmidt-2000 that Dr. Bolin obtained from the University of Hong Kong Library. Bolin, ¶¶28-29. Based on this evidence, Dr. Bolin concluded that Schmidt-2000 was indexed, searchable and publicly accessible no later than early 2001. Bolin, ¶¶31-38. Based on the above evidence, I too conclude that Schmidt-2000 would have also been publicly accessible to a POSA exercising reasonable diligence at least by early 2001.

13. As I discuss below, a copy of Schmidt-2000 also appears to have been publicly accessible from a Schmidt-2000 author's websites more than one year before the earliest possible priority date of the Asserted Patent. I understand from counsel that a POSA may use other publicly accessible information as "research aids" as a roadmap to locating additional prior art in the same or related field. Various publications publicly available before the earliest possible priority date of the Asserted Patent would have directed a POSA to Schmidt-2000. For example, one of Schmidt-2000's authors – Dr. Bertil Schmidt – co-authored several articles published by Springer-Verlag and the IEEE with subject matter relating to the field

of the Asserted Patent, which would have directed a POSA exercising reasonable diligence to Schmidt-2000. *See also* §I.B above (discussing a POSA's knowledge of the IEEE as a reputable compiler and publisher of scientific and technical books and journals); Bolin, ¶51 (identifying Springer-Verlag as "a well-known and respected publisher (founded in 1842) of academic journals and books"). Several of Dr. Schmidt's IEEE- and Springer-published articles indicated that he was employed at the Nanyang Technological University (NTU) in Singapore at that time of their publication. *See, e.g.*, https://ieeexplore.ieee.org/document/1016589/authors#authors; https://ieeexplore.ieee.org/document/1556805/authors#authors; https://ieeexplore.ieee.org/document/1524318/authors#authors. POSAs in the field of the challenged patent would have had reason to navigate to the website of an author of an article in a field relevant to the challenged patent, to find additional information about publications or, in some cases, copies of publications relating to that author's teaching, research and research groups, projects and other academics and graduate students in the same and related academic department.

14. Page 33 of EX1031 is the Internet Archive's archived capture of NTU's homepage, http://www.ntu.edu.sg/publicportal/, as it existed on November 6, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061106005508/http://www.ntu.edu.sg/publicportal.

See EX1031, ¶¶5-6, p. 32. At the top of the NTU's homepage, there is a dropdown box for "Quick Links," one of which is to an "A-Z Website Directory." Selecting "A-Z Website Directory" from the dropdown box navigates to NTU's "Website Directory Service" webpage at https://wis.ntu.edu.sg/webexe88/owa/wdr_search.index_search. Page 35 of EX1031 is the Internet Archive's archived capture of that webpage as it existed on October 6, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061006171248/https://wis.ntu.edu.sg/webexe88/owa/wdr_search.index_search. *See* EX1031, p. 34.

15. As shown on Page 35 of EX1031, NTU's Website Directory included hyperlinks to a list of NTU Staff, reachable by clicking on the letter that matched the first letter of the Staff's name. Clicking on "B" in the Staff row navigates to https://wis.ntu.edu.sg/pls/webexe88/wdr_search.index_search?P_category=SF&P_index=B, which provides a list of "Staff websites [that] begin with letter B." Page 37 of EX1031 is the Internet Archive's archived capture of that webpage as it existed on October 6, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061006173813/https://wis.ntu.edu.sg/pls/webexe88/wdr_search.index_search?P_category=SF&P_index=B. *See* EX1031, p. 36.

16. As shown on Page 37 of EX1031, there was a hyperlink to the website of "Bertil Schmidt." Clicking on "Bertil Schmidt" navigates to

http://www.ntu.edu.sg/home/asbschmidt/, which was Dr. Schmidt's homepage hosted on NTU's website. Page 39 of EX1031 is the Internet Archive's archived capture of that webpage as it existed on October 9, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061009222645/http://www.ntu.edu.sg/home/asbschmidt/. *See* EX1031, p. 38.

17. As shown on Page 39 of EX1031, Dr. Schmidt's homepage included a link at the top of homepage to "Publications." Clicking on "Publications" navigates to a webpage listing various Books, Journal Papers, and Conferences Papers authored by Dr. Schmidt. Pages 56-58 of EX1031 contain the Internet Archive's archived capture of that webpage as it existed on September 24, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20060924015515fw_/http://www.ntu.edu.sg/home/asbschmidt/publications.htm. *See* EX1031, p. 55.

18. As shown on Page 56 of EX1031, item number "17" is identified as "B. Schmidt, M. Schimmler, KPROC - An Instruction Systolic Architecture for Parallel Prefix Applications, Parallel and Distributed Computing Practices, Vol. 3, No. 2, 2000," which is the title and journal listed on the face of Schmidt-2000. Clicking on the hyperlink for item number "17" navigates to http://www.ntu.edu.sg/home/asbschmidt/paper/PDCP00.htm, which is a webpage

that includes a "Title," "Abstract," and "Reference" citation that matches what is shown on the face of Schmidt-2000 (EX1007). Page 45 of EX1031 is the Internet Archive's archived capture of that webpage as it existed on October 26, 2004, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20041026062341fw_/http://www.ntu.edu.sg/home/asbschmidt/paper/PDCP00.htm. *See* EX1031, p. 44.

19. As shown on Page 45 of EX1031, there was a hyperlink to "File download" a pdf at the Internet Archive's extended URL https://web.archive.org/web/20041026062341fw_/http:/www.ntu.edu.sg/home/asbschmidt/paper/KPROC2.pdf. This extended URL suggests that this webpage provided a link to download a PDF copy of Schmidt-2000 that the Internet Archive had archived on October 26, 2004. This extended URL, however, does not presently link to a PDF copy of Schmidt-2000 and instead navigates to a recent archive of NTU's website, which indicates that the Internet Archive's extended hyperlink to the Schmidt-2000 PDF is no longer valid. However, as discussed below, several other articles listed on the archived version of Dr. Schmidt's webpage of "Journal Articles" are still archived by the Internet Archive from August 2007, which indicates to me that Schmidt-2000 likely would have also been available to download from Dr. Schmidt's website, likely by October 26,

2004, or at least by August 2007, consistent with the Internet Archive's archived copies of other Schmidt-authored articles from Dr. Schmidt's website.

20. But even if Dr. Schmidt's hyperlink to download Schmidt-2000 had never been active, a POSA exercising reasonable diligence would have been able to obtain a copy of Schmidt-2000, or information about how to obtain a copy of Schmidt-2000, by contacting Dr. Schmidt directly. As shown on Page 39 of EX1031, Dr. Schmidt's homepage (archived on October 9, 2006) included an "Address," "Tel[ephone]" number, "Fax" number, and "E-mail" for contacting Dr. Schmidt. Thus, Dr. Schmidt's website provided a POSA several methods by which to contact Dr. Schmidt to ask for a copy of the Schmidt-2000 article, which I know from personal experience would have been a common courtesy of professors and researchers in this field. If Dr. Schmidt had been unable (or unwilling) to provide a copy of the Schmidt-2000 article, it reasonable to assume that Dr. Schmidt would have at least provided information about where to obtain a copy of Schmidt-2000 from the journal that published Schmidt (the PDCP) and/or the PDCP's publisher (Nova Science).

21. Based on the above evidence, I too conclude that Schmidt-2000 would have also been publicly accessible, via Dr. Schmidt's website, to a POSA exercising reasonable diligence by October 26, 2004, or at least by August 2007.

22. As I noted above in this section and discussed in §I.A, the Internet Archive maintains active archived links to copies of another paper co-authored by Dr. Schmidt, which indicates to me that the Schmidt-2000 article also would have been available for download at around the same time, consistent with the "File download" hyperlink I discussed above.

2. Public Accessibility of Schmidt-1999 (EX1009)

23. EX1009 (which I refer to herein as "Schmidt-1999") is titled "A Parallel Accelerator Architecture for Multimedia Video Compression." Based on the evidence I discuss below, it is my opinion that Schmidt-1999 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence.

24. According to the face of EX1009, Schmidt-1999 was published in connection with the August/September 1999 5th International Euro-Par Conference in Toulouse, France by the publisher Springer-Verlag. Page 3 of Schmidt-1999 similarly bears the text "© Springer-Verlag Berlin Heidelberg 1999," indicating it was copyrighted and thus publicly accessible by the year 1999. According to Dr. Bolin, the publisher Springer-Verlag "was and is a well-known and respected publisher (founded in 1842) of academic journals and books …

[and] provides an online search tool ('SpringerLink')[] for providing researchers with access to millions of academic publications." Bolin, ¶16.

25. According to Dr. Bolin, the proceedings of the Euro-Par'99 conference have a creation date of August 6, 1999 in the OCLC MARC database and the University of Wisconsin at Madison has a copy of Schmidt-1999. Bolin, ¶¶13-15; *see also* Bolin, ¶¶7-15 (discussing MARC records). EX1009 is the copy of Schmidt-1999 that Dr. Bolin obtained from the University of Wisconsin. Bolin, ¶12. Attachment 3c to Dr. Bolin's declaration is a copy of Schmidt-1999 that Dr. Bolin obtained from the SpringerLink online database, which indicated a "first online" date for the article of January 1, 1999 and an online publication date of August 6, 1999. Bolin, ¶17. Based on this evidence, Dr. Bolin concluded that Schmidt-1999 was indexed, searchable and publicly accessible no later than early the end of 1999. Bolin, ¶18. Based on the above evidence, I too conclude that Schmidt-1999 would have been publicly accessible to a POSA exercising reasonable diligence at least by the end of 1999.

3. Public Accessibility of Schimmler-1994 (EX1010)

26. EX1010 (which I refer to herein as "Schimmler-1994") is titled "The Instruction Systolic Array – Implementation of a Low-Cost Parallel Architecture as Add-On Board for Personal Computers." Based on the evidence I discuss below, it is my opinion that Schimmler-1994 is prior art to the challenged patent because it

was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence.

27. According to the face of EX1010, Schimmler-1994 was published in connection with the April 1994 High-Performance Computing and Networking (HPCN) conference in Munich, Germany by the publisher Springer-Verlag. Page 3 of Schimmler-1994 similarly bears the text "© Springer-Verlag Berlin Heidelberg 1994," indicating it was copyrighted and thus publicly accessible by the year 1994. According to Dr. Bolin, the publisher Springer-Verlag "was and is a well-known and respected publisher (founded in 1842) of academic journals and books … [and] provides an online search tool ('SpringerLink')[] for providing researchers with access to millions of academic publications." Bolin, ¶16.

28. According to Dr. Bolin, the proceedings of the 1994 HPCN conference have a creation date of June 30, 1994 in the OCLC MARC database and the University of Wisconsin at Madison has a copy of Schimmler-1994. Bolin, ¶¶21-23; *see also* Bolin, ¶¶7-15 (discussing MARC records). EX1010 is the copy of Schimmler-1994 that Dr. Bolin obtained from the University of Wisconsin. Bolin, ¶20. According to Dr. Bolin, the Springer Link online database indicates that Schimmler-1994 has a "first online" date for the article of January 1, 2005 and an online publication date of May 26, 2005. Bolin, ¶25. Based on this evidence,

Dr. Bolin concluded that Schimmler-1994 was indexed, searchable and publicly accessible by May 26, 2005. Bolin, ¶18.

29. Based on the above evidence, I too conclude that Schimmler-1994 would have been publicly accessible to a POSA exercising reasonable diligence at least by May 26, 2005.

4. Public Accessibility of Schimmler-1996 (EX1011)

30. EX1011 (which I refer to herein as "Schimmler-1996") is titled "The Instruction Systolic Array in Image Processing Applications." Based on the evidence I discuss below, it is my opinion that Schimmler-1996 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence.

31. According to the face of EX1011, Schimmler-1996 was published in connection with the June 10-12, 1996 Vision Systems: Sensors, Sensor Systems, and Components conference in Besancon, France by the publisher SPIE–The International Society for Optical Engineering, which is a well-known international technical society and respected publisher of engineering and scientific journals and proceedings. Bolin, ¶30. Page 3 of Schimmler-1996 similarly bears the text "Copyright © 1996, The Society of Photo-Optical Instrumentation Engineers," indicating it was copyrighted and thus publicly accessible by the year 1996.

32. According to Dr. Bolin, the proceedings of the 1996 Vision Systems: Sensors, Sensor Systems, and Components conference have a creation date of May 16, 1996 in the OCLC MARC database and the University of Wisconsin at Madison has a copy of Schimmler-1996. Bolin, ¶¶28-30; *see also* Bolin, ¶¶7-15 (discussing MARC records). EX1011 is the copy of Schimmler-1996 that Dr. Bolin obtained from the University of Wisconsin. Bolin, ¶27. According to Dr. Bolin, the SPIE Digital Library indicates that Schimmler-1996 has a publication date of August 21, 1996. Bolin, ¶31. Based on this evidence, Dr. Bolin concluded that Schimmler-1996 was published at least by September 17, 1996. Bolin, ¶31.

33. Based on the above evidence, I too conclude that Schimmler-1996 would have been publicly accessible to a POSA exercising reasonable diligence at least by September 17, 1996.

B. Public Accessibility of NTRS/ITRS Reports (EX1021-EX1027, EX1030)

34. The International Technology Roadmap for Semiconductors (ITRS) is a document that outlines the expected future development of the semiconductor industry and incorporates the latest information and insights from the semiconductor industry when the roadmap is being updated. The ITRS was established to determine research needs for the semiconductor industry. Several times a year, teams of semiconductor professionals get together and establish capabilities and forecast research problems that needed addressing. Before the

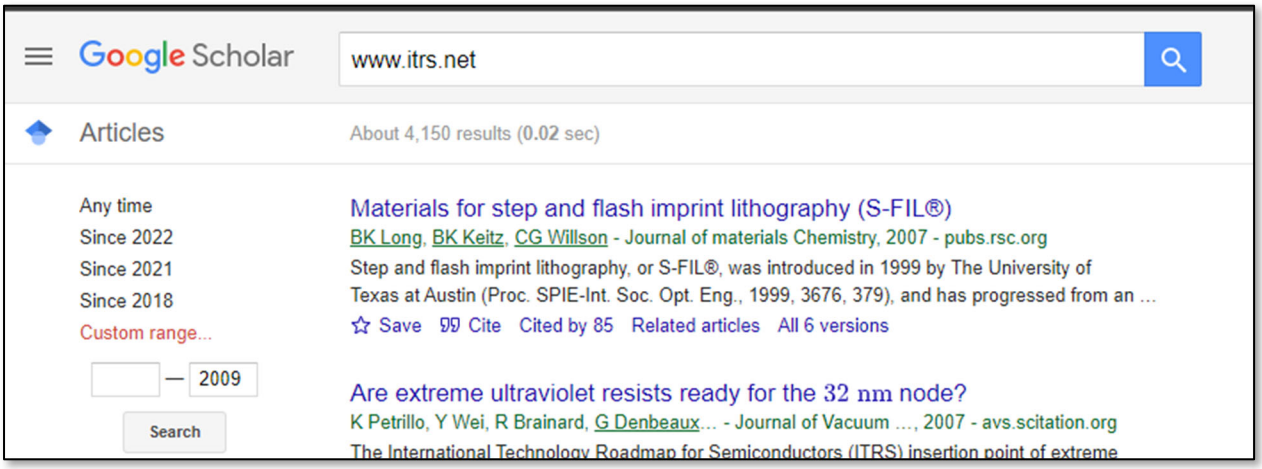
ITRS there was the National Technology Roadmap for Semiconductors (NTRS), which was initiated by the Semiconductor Industry Association (SIA), an association for the semiconductor industry, whose members include and have included leading companies in the semiconductor industry, such as Broadcom, IBM, Intel, Micron, Qualcomm, ARM, Samsung, TSMC, and others. In 1998, the NTRS evolved into the ITRS after the SIA invited semiconductor industry associations from around the world to contribute to and sponsor the roadmap moving forward. *See* EX1029, p. 94 (Foreward).

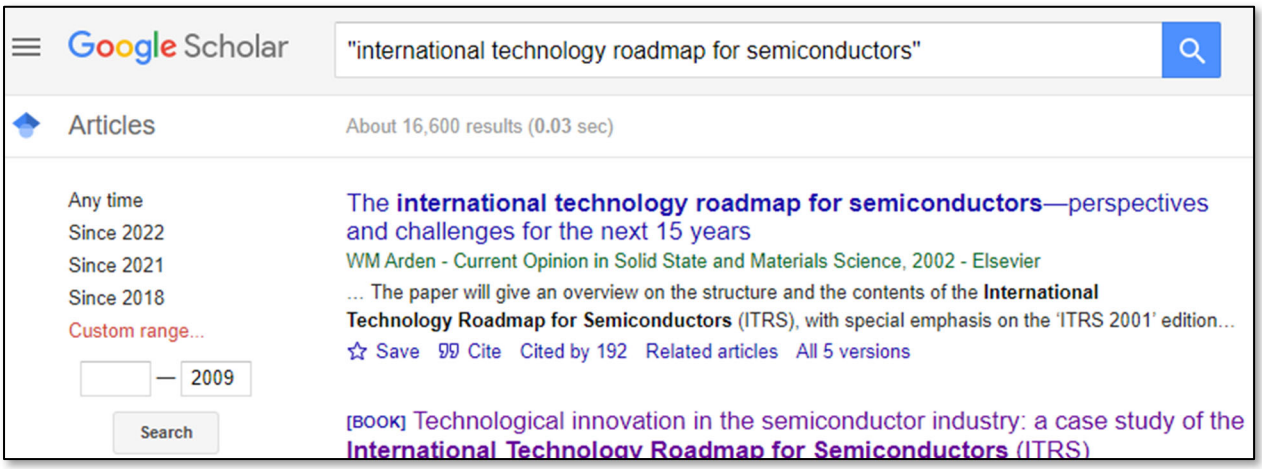
35. The ITRS roadmap has been updated every two years by a consortium of semiconductor industry organizations from around the world, like the SIA, European Semiconductor Industry Association, Korean Semiconductor Industry Association, Taiwan Semiconductor Industry Association, and Japan Electronics and Information Technology Industries Association. Though the ITRS roadmap has been published every two years, there have often been yearly updates. Each roadmap and update provides a long-term view of the key technologies and trends that are expected to drive the growth of the semiconductor industry, including areas such as device scaling, materials and processes, manufacturing, and design and architecture. The ITRS is an important tool for companies and researchers in the semiconductor industry, as it helps them to plan and prioritize their technology development efforts.

36. I personally served on the ITRS for several years in the mid-2010s, and was aware of and relied on the roadmaps throughout the 2000s. In my opinion, POSAs would have been aware of the NTRS and ITRS roadmaps that were published in the late-1990s through the 2000s and into the 2010s. The roadmaps would have been relied upon by POSAs at the time of the alleged invention for reliable data about both logic chip and memory chip development, which POSAs would have used to estimate the chip area and transistor density for the design of semiconductor devices like computing chips, for example. The roadmaps were published and widely distributed to POSAs throughout the world at no charge, and, as I discuss below, were available for download directly from the ITRS website, which POSAs would have been aware of and been able to locate with reasonable diligence.

37. For example, as shown below, a recent search of Google Scholar returned 4,150 articles from 2009 and earlier that included the ITRS's website – www.itrs.net – and 16,600 articles that included the text "international technology roadmap for semiconductors," corroborating my opinion that POSAs would have

been well-aware of the ITRS website and roadmaps.[7] *See also* Bolin, ¶6 (identifying Google Scholar as a reliable repository of scholarly and research data).





[7] https://scholar.google.com/schhp?hl=en&as_sdt=0,22 (search for "www.itrs.net"; from filter tool select "Custom range…"; enter "2009" into right field for date range; click "Search"); *id.*, (same search pathway for "international technology roadmap for semiconductors").

The ITRS roadmap is also referenced and/or cited in numerous U.S. patent application publications and issued patents from before the time of the alleged invention. Several specifically identify the ITRS's website as a source for retrieving copies of the roadmaps. *See, e.g.*, U.S. Patent Appl. Publ. No. 2003/0219157A1, [0184] ("Data of the chip areas and the technologies is referred from ITRS 2000[ ]Update Road Map (Reference: 'ITRS 2000 Update: The International Technology Roadmap for Semiconductors 2000 Update', **URL** **http://public.itrs.net/(2000)**)."); U.S. Patent Appl. Publ. No. US2003/0040130A1 [0041] ("As used herein, the term "'technology node'" is in accordance with the definition for Technology node provided in The International Technology Roadmap for Semiconductors (2000 Up**date), published by the Semiconductor Industry Association (SIA), San Jose Calif.; http://public.itrs.net/**.").

38. Further corroborating my opinion that the roadmaps would have been widely distributed to POSAs, ITRS chairman Paolo Gargini, wrote in a 2000 IEEE publication, that "**It is the purpose of** the 1999 ITRS, of the 2000 ITRS Update, and of future **ITRS documents**, **to provide a set of up-to-date reference documents** of requirements, potential solutions, and their timing for the semiconductor industry. **These documents have already provided a common reference for international forums that have initiated discussions on cooperation among the leading silicon semiconductor manufacturers** and the

leading suppliers of equipment, materials, and software, **as well as researchers from university and government laboratories**." EX1033, 4. Thus, it is my opinion that at the time of the alleged invention, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

39. A "Technology Node" in the roadmap "was defined as the achievement of an approximate 0.7x reduction per node." EX1023 (ITRS 2000), 5. The node name generally tracks the DRAM minimum metal half-pitch or polysilicon half pitch. EX1023 (ITRS 2000), 31. The ITRS roadmaps typically also included information about the area required to fabricate a 4-transistor logic gate. *See, e.g.*, EX1028 (ITRS 2010 Executive Summary), 67. A POSA would understand that many useful gates require four transistors, such as NAND and NOR gates, and so the area required to fabricate a conventional 4-transistor logic gate would have been a useful metric for estimating the transistor density achievable with a particular technology node.

1. Public Accessibility of NTRS 1997 (EX1021: "NTRS-1997")

40. The first page of EX1021 (which I refer to herein as "NTRS-1997") is titled "The National Technology Roadmap for Semiconductors Technology Needs 1997 Edition," by the Semiconductor Industry Association (SIA). As I discussed

in §II.B above, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

41. Page 3 of EX1021 bears a copyright "© 1997 Semiconductor Industry Association," which indicates it was published in 1997. As discussed in §§II.B.2-II.B.7 below, the Internet Archive has various archived captures of the ITRS's website from before the year 2009. *See, e.g.*, EX1029, 6, 10. As I will discuss (§II.B.3 below), the Internet Archive captured the ITRS's "ITRS Edition Reports and Ordering" webpage as it existed on April 1, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080401032429/http://www.itrs.net:80/reports.html. *See also* EX1029, ¶¶5-7, p. 11. Page 12 of EX1029 is the Internet Archive's archived capture of that webpage. As shown at the bottom of that webpage, the ITRS disclosed that "[a] limited number of hard copy books are available for the 2005 ITRS **and previous roadmaps** (1994, **1997,** 1999, 2001, 2003)." As I discussed above (§II.B), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled artisan would have

known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located NTRS-1997 by searching the Web for keywords of interest such as "NTRS" or "National Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I discussed above, NTRS-1997 was published and publicly accessible and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) April 1, 2008.

2. Public Accessibility of ITRS 1999 (EX1022: "ITRS-1999")

42. The first page of EX1022 (which I refer to herein as "ITRS-1999") is titled "International Technology Roadmap for Semiconductors 1999 Edition," "Overall Roadmap Technology Characteristics & Glossary." As I discussed in §II.B above, POSAs would have been aware of and had reason to be familiar with

the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

43. The Internet Archive has an archived capture of the ITRS's website, http://public.itrs.net/, as it existed on February 1, 2001, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20010201085500/http://public.itrs.net/. As shown on the left side of that webpage is the text "Reports," which navigates to, http://public.itrs.net/Reports.htm. The Internet Archive has an archived capture of that webpage as it existed on February 6, 2001, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20010206133820/http://public.itrs.net/Reports.htm. As shown on that page, there is a hyperlink to "1999 Edition," which navigates to http://public.itrs.net/files/1999_SIA_Roadmap/Home.htm. The Internet Archive has an archived capture of that webpage as it existed on February 6, 2001, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20010206135413/http://public.itrs.net/files/1999_SIA_Roadmap/Home.htm. As shown on that page, there is a hyperlink to "Overall Technology Roadmap Characteristics and Glossary" (ORTC), which navigates to a PDF of the 1999 roadmap's ORTC, as it existed on June 4, 2001, as reflected in the Internet Archive's extended URL:

https://web.archive.org/web/20010604100132/http://public.itrs.net/files/1999_SIA_Roadmap/ORTC.pdf. That PDF is EX1022.

44. As I discussed above (§I above), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled artisan would have known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located ITRS-1999 by searching the Web for keywords of interest such as "ITRS" or "International Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I discussed above, ITRS-1999 (EX1022) was published and publicly accessible and

could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) February 6, 2001.

3. Public Accessibility of ITRS 2000 Update (EX1023: "ITRS-2000")

45. The first page of EX1023 (which I refer to herein as "ITRS-2000") is titled "International Technology Roadmap for Semiconductors 2000 Update: Overall Roadmap Technology Characteristics." As I discussed in §II.B above, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

46. Page 10 of EX1029 is the Internet Archive's archived capture of the ITRS's homepage, http://www.itrs.net/home.html/, as it existed on April 1, 2008. This is reflected in the Wayback Machine URL on page 9 of EX1029: https://web.archive.org/web/20080401032723/http://www.itrs.net/home.html.

47. As shown on the left side of Page 10 of EX1029, the ITRS home page included a link to "ITRS Edition Reports and Ordering," which navigates to http://www.itrs.net/reports.html. Page 12 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 1, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080401032429/http://www.itrs.net:80/reports.html. *See also* EX1029, ¶¶5-7, p. 11.

48. As shown on the middle of Page 12 of EX1029, the ITRS home page included a link to the "2000 Update," which navigates to http://www.itrs.net/Links/2000UpdateFinal/2kUdFinal.html. Page 14 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 20, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080420120224/http://www.itrs.net/Links/2000UpdateFinal/2kUdFinal.html. *See also* EX1029, ¶¶5-7, p. 13. EX1023 combines the portions of the 2000 Update retrievable from the hyperlinks respectively labeled (1) "Overall Roadmap Technology Characteristics (ORTC)" and (2) "PIDS" (i.e., Process Integration, Devices and Structures), as they existed on July 13, 2007, as reflected in their respective Internet Archive's extended URLs: *e.g.*, https://web.archive.org/web/20070713160806/http://www.itrs.net/Links/2000UpdateFinal/ORTC2000final.pdf. *See* EX1029, ¶7, pp. 29-92 (Folders 1-2).

49. As I discussed above (§I above), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled artisan would have known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular

roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located ITRS-2000 by searching the Web for keywords of interest such as "ITRS" or "International Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I discussed above, ITRS-2000 was published and publicly accessible and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) April 20, 2008.

4. Public Accessibility of ITRS 2001 (EX1030: "ITRS-2001")

50. The first page of EX1030 (which I refer to herein as "ITRS-2001") is titled "International Technology Roadmap for Semiconductors 2001 Edition: Executive Summary." As I discussed in §II.Babove, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

51. As discussed (§II.B.3 above), the Internet Archive captured the ITRS's "ITRS Edition Reports and Ordering" webpage as it existed on April 1, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080401032429/http://www.itrs.net:80/reports.html. *See also* EX1029, ¶¶5-7, p. 11. Page 12 of EX1029 is the Internet Archive's archived capture of that webpage.

52. As shown on the middle of Page 12 of EX1029, the ITRS home page included a link to the "2001 Edition," which navigates to http://www.itrs.net/Links/2001ITRS/Home.htm. The Internet Archive archived a capture of that webpage as it existed on December 22, 2007, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20071222013746/http://www.itrs.net/Links/2001ITRS/Home.htm. *See also* EX1029, ¶¶5-7. As shown on that page, there is a hyperlink to "Executive Summary," which navigates to a PDF of the 2001 roadmap's Executive Summary, as it existed on July 11, 2007, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20070711213510/http://www.itrs.net/Links/2001ITRS/ExecSum.pdf. That PDF is EX1030.

53. As I discussed above (§II.B), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the

ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled artisan would have known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located ITRS-2001 by searching the Web for keywords of interest such as "ITRS" or "International Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I discussed above, ITRS-2001 was published and publicly accessible and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) July 11, 2007.

5. Public Accessibility of ITRS 2002 Update (EX1024: "ITRS-2002")

54. The first page of EX1024 (which I refer to herein as "ITRS-2002") is the "Foreward" to the "International Technology Roadmap for Semiconductors:

2002 Update." As I discussed in §II.Babove, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

55. Page 10 of EX1029 is the Internet Archive's archived capture of the ITRS's homepage, http://www.itrs.net/home.html/, as it existed on April 1, 2008. This is reflected in the Internet Archive's extended URL on page 9 of EX1029: https://web.archive.org/web/20080401032723/http://www.itrs.net/home.html.

56. As shown on the left side of Page 10 of EX1029, the ITRS home page included a link to "ITRS Edition Reports and Ordering," which navigates to http://www.itrs.net/reports.html. Page 12 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 1, 2008, as reflected in the Internet Archive's extended URL. *See also* EX1029, ¶¶5-7, p. 11.

57. As shown on the middle of Page 12 of EX1029, the ITRS home page included a link to the "2002 Update," which navigates to http://www.itrs.net/Links/2002Update/Home.htm. Page 16 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on January 26, 2008, as reflected in the Internet Archive's extended URL. *See also* EX1029, ¶¶5-7, p. 15. EX1024 combines the portions of the 2002 Update retrievable from the hyperlinks respectively labeled (1) "Foreward," (2) "Table of Contents,"

(3) "Overall Roadmap Technology Characteristics (ORTC)," and (4) "Process Integration, Devices & Structures (includes Emerging Research Devices)," as they existed on July 12, 2007, as reflected in their respective Internet Archive's extended URLs. *See* EX1029, ¶7, pp. 94-154 (Folders 3-7)

58. As I discussed above (§I above), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled artisan would have known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located ITRS-2002 by searching the Web for keywords of interest such as "ITRS" or "International Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I

discussed above, ITRS-2002 was published and publicly accessible and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) April 20, 2008.

6. Public Accessibility of ITRS 2004 Update (EX1025: "ITRS-2004")

59. The first page of EX1025 (which I refer to herein as "ITRS-2004") is titled "International Technology Roadmap for Semiconductors 2004 Update: Overview and Summaries." As I discussed in §II.B above, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

60. Page 6 of EX1029 is the Internet Archive's archived capture of the ITRS's homepage, http://public.itrs.net/, as it existed on March 21, 2005. This is reflected in the Wayback Machine extended URL on page 5 of EX1029.

61. As shown on Page 6 of EX1029, the ITRS homepage included the following text – "The 2004 Update is now available here" – which ended in a hyperlink ("here") that navigates to http://www.itrs.net/Common/2004Update/2004Update.htm. Page 8 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 3, 2005, as reflected in the Internet Archive's extended URL. *See also* EX1029, ¶¶5-7.

62. As shown on Page 8 of EX1029, the ITRS "2004 Update Tables" included hyperlinks to download the 2004 Update to the 2003 roadmap. Page 8 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 3, 2005, as reflected in the Internet Archive's extended URL. *See* EX1029, ¶¶5-6, p. 7. EX1025 combines the portions of the 2004 Update retrievable from the hyperlinks respectively labeled (1) Overview and Summaries, (2) Overall Roadmap Technology Characteristics (ORTC), (3) Design, and (4) Process Integration, Devices & Structures, as they existed on May 4, 2005, as reflected in their respective Internet Archive's extended URLs. *See* EX1029, ¶7, pp. 156-320 (Folders 8-11).

63. As I discussed above (§II.B), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled artisan would have known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler

software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located ITRS-2004 by searching the Web for keywords of interest such as "ITRS" or "International Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I discussed above, ITRS-2004 was published and publicly accessible and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) May 4, 2005.

7. Public Accessibility of ITRS 2007 (EX1026: "ITRS-2007")

64. The first page of EX1026 (which I refer to herein as "ITRS-2007") is titled "International Technology Roadmap for Semiconductors 2007 Edition: Executive Summary." As I discussed in §II.B above, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

65. Page 10 of EX1029 is the Internet Archive's archived capture of the ITRS's homepage, http://www.itrs.net/home.html/, as it existed on April 1, 2008. This is reflected in the Wayback Machine extended URL on page 9 of EX1029.

66. As shown on the left side of Page 10 of EX1029, the ITRS home page included a link to "ITRS Edition Reports and Ordering," which navigates to http://www.itrs.net/reports.html. Page 12 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 1, 2008, as reflected in the Internet Archive's extended URL. *See also* EX1029, ¶¶5-7, p.11.

67. As shown on the middle of Page 12 of EX1029, the ITRS home page included a link to the "2007 Edition," which navigates to http://www.itrs.net/Links/2007ITRS/Home2007.htm. Page 18 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 22, 2008, as reflected in the Internet Archive's extended URL. *See also* EX1029, ¶¶5-7, p. 17. EX1026 combines the portions of the 2007 Roadmap retrievable from the hyperlinks respectively labeled (1) "Executive Summary" (as it existed on November 23, 2009) and (2) "Process Integration, Devices & Structures" (as it existed on March 6, 2009), as reflected in their respective Internet Archive's extended URLs. *See* EX1029, ¶7, pp. 322-487 (Folders 12-13).

68. As I discussed above (§II.B), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled

artisan would have known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.Aabove), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located ITRS-2007 by searching the Web for keywords of interest such as "ITRS" or "International Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I discussed above, ITRS-2007 was published and publicly accessible and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) April 22, 2008.

8. Public Accessibility of ITRS 2009 (EX1027: "ITRS-2009")

69. The first page of EX1027 (which I refer to herein as "ITRS-2009") is titled "International Technology Roadmap for Semiconductors 2009 Edition: Executive Summary." As I discussed in §II.B above, POSAs would have been aware of and had reason to be familiar with the NTRS/ITRS roadmaps and would

have been able to obtain any roadmap by exercising reasonable diligence, such as by navigating to the ITRS's website.

70. Page 20 of EX1029 is the Internet Archive's archived capture of the ITRS's homepage, http://www.itrs.net/home.html, as it existed on April 13, 2010. This is reflected in the Wayback Machine URL on page 19 of EX1029.

71. As shown on the middle of Page 20 of EX1029, the ITRS home page included a link to "The NEW **ITRS 2009 Edition** is now released!," which navigates to http://www.itrs.net/Links/2009ITRS/Home2009.htm. Page 22 of EX1029 is the Internet Archive's archived capture of that webpage as it existed on April 7, 2010, as reflected in the Internet Archive's extended URL. *See also* EX1029, ¶¶5-7, p. 21. EX1027 combines the portions of the 2009 Roadmap retrievable from the hyperlinks respectively labeled (1) "Executive Summary" and (2) "Process Integration, Devices & Structures," as they existed on March 31, 2010, as reflected in their respective Internet Archive's extended URLs. *See* EX1029, ¶7, pp. 489-611 (Folders 14-15)

72. As I discussed above (§II.B), I have personal knowledge that POSAs with a background or interest in the semiconductor industry, were familiar with the ITRS by the 2009-2010 timeframe. Such POSAs would have been interested in the ITRS roadmaps, which were accessible via the ITRS's website as demonstrated by the above-discussed archived webpage captures. In other words, a skilled

artisan would have known to navigate to the ITRS homepage and browsing the ITRS website would have known to navigate to the page of a particular roadmap or roadmap update, to retrieve documents on that page regarding a particular roadmap or roadmap update. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located ITRS-2009 by searching the Web for keywords of interest such as "ITRS" or "International Technology Roadmap for Semiconductor," which would be terms of interest to a POSA interested in the semiconductor industry and/or the roadmaps and updates. Based on the evidence I discussed above, ITRS-2009 was published and publicly accessible and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence by (at least) April 7, 2010.

C. Public Accessibility of Additional Prior Art

1. Public Accessibility of EX1012 ("Yeh-ISCA07")

73. EX1012 (which I refer to herein as "Yeh-ISCA07") is titled "ParallAX: An Architecture for Real-Time Physics." Based on the evidence I discuss below, it is my opinion that Yeh-ISCA07 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date

for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence.

74. According to the face of EX1012, Yeh-ISCA07 was "[p]ublished by the Association for Computing Machinery" (ACM) in the May 2007 issue of Computer Architecture News, in connection with the Proceedings of the 34th International Symposium on Computer Architecture (ISCA'07). *See also* §I.C above (discussing a POSA's knowledge of the ACM as a reputable compiler and publisher of scientific and technical books and journals). I am personally familiar with ISCA conferences, which are and have been since before the alleged invention date one of the premier conferences on computer architecture. Page 232 of EX1012 similarly bears the text "Copyright 2007 ACM," indicating it was copyrighted and thus publicly accessible in the year 2007. According to Dr. Bolin, the ACM Digital Library database indicates Yeh-ISCA07 was published on June 9, 2007. Based on this evidence, Dr. Bolin concluded that Yeh-ISCA07 was indexed, searchable and publicly accessible no later than June 2007. Bolin, ¶¶68-73. Based on the above evidence, I too conclude that Yeh-ISCA07 would have been publicly accessible to a POSA exercising reasonable diligence by June 2007.

75. As I discuss below, a copy of Yeh-ISCA07 was also publicly accessible from one of the author's websites more than one year before the earliest possible priority date of the Asserted Patent. I understand from counsel that a

POSA may use other publicly accessible prior art as "research aids" that provide a roadmap to locating additional prior art in the same or related field. Before 2008, I was personally aware of Dr. Glenn Reinman – a Yeh-ISCA07 co-author – as a well-known, reputable researcher in the field of parallel computing architectures. In my opinion, a POSA would have also been aware of Dr. Reinman and his research group at UCLA and thus would have had reason to go UCLA's homepage and navigate to the homepages of professors and graduate students associated with Dr. Reinman's research. *See* Yeh-ISCA07, p. 232 (identifying three of the four authors as being associated with the "Computer Science Department, UCLA" with "@cs.ucla.edu" email addresses).

76. Page 5 of EX1032 is the Internet Archive's archived capture of UCLA's homepage, http://www.ucla.edu/, as it existed on February 1, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080201003048/http://www.ucla.edu/. *See* EX1032, ¶¶5-6, p. 4. At the top of UCLA's homepage, there is a hyperlink to "ACADEMICS," which navigates to UCLA's "ACADEMICS" webpage at http://www.ucla.edu/academics.html. Page 7 of EX1032 is the Internet Archive's archived capture of that webpage as it existed on February 8, 2008, as reflected in the Internet Archive's extended URL:

https://web.archive.org/web/20080208230223/http://www.ucla.edu/academics.html . *See* EX1032, ¶¶5-6, p. 6.

77. As shown on Page 7 of EX1032, UCLA's Academics webpage included hyperlinks to various "Professional Schools," including the "Henry Samueli School of Engineering and Applied Science," which navigates to http://www.engineer.ucla.edu/, which is the landing page for UCLA's School of Engineering. Page 9 of EX1032 is the Internet Archive's archived capture of that webpage as it existed on February 6, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080206081155/http://www.engineer.ucla.edu/. *See* EX1032, ¶¶5-6, p. 8.

78. As shown on Page 9 of EX1032, UCLA's Engineering school webpage included a hyperlink "DEPARTMENTS," which navigates to www.engineer.ucla.edu/departments/index.html. Page 11 of EX1032 is the Internet Archive's archived capture of that webpage as it existed on February 11, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080211124057/http://www.engineer.ucla.edu/departments/index.html. *See* EX1032, ¶¶5-6, p. 10.

79. As shown on Page 11 of EX1032, the "Departments in the School" included a hyperlink to the "Computer Science" department, which navigates to

http://www.cs.ucla.edu/. Page 13 of EX1032 is the Internet Archive's archived capture of that webpage as it existed on February 8, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080208153715/http://www.cs.ucla.edu/. *See* EX1032, ¶¶5-6, p. 12.

80. As shown on Page 13 of EX1032, the Computer Science Department included a hyperlink to "people" in the department, which navigates to http://www.cs.ucla.edu/csd/people/index.html. Page 15 of EX1032 is the Internet Archive's archived capture of that webpage as it existed on February 6, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080206005807/http://www.cs.ucla.edu/csd/people/index.html. *See* EX1032, ¶¶5-6, p. 14.

81. As shown on Page 15 of EX1032, the "CSD People Pages" included a hyperlink to "Faculty" in the department, which navigates to http://www.cs.ucla.edu/csd/people/faculty.html. Pages 18-25 of EX1032 is the Internet Archive's archived capture of that webpage as it existed on February 14, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080214085121/http://www.cs.ucla.edu/csd/people/faculty.html. *See* EX1032, ¶¶5-6, p. 17.

82. As shown on Page 20 of EX1032, the list of "Faculty" included a hyperlink for "Petros Faloutsos" – one of Yeh-ISCA07's authors – which navigates to http://www.cs.ucla.edu/~pfal/. Page 17 of EX1032 is the Internet Archive's archived capture of Dr. Faloutsos's ebpage hosted at UCLA.edu as it existed on February 20, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080220222723/http://www.cs.ucla.edu/~pfal/. *See* EX1032, ¶¶5-6, p. 26.

83. As shown on Page 27 of EX1032, Dr. Faloutsos' webpage included a hyperlink to "Publications," which navigates to http://www.cs.ucla.edu/~pfal/papers.html. Page 29 of EX1032 is the Internet Archive's archived capture of that webpage as it existed on February 22, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080222132652/http://www.cs.ucla.edu/~pfal/papers.html. *See* EX1032, ¶¶5-6, p. 28.

84. As shown on Page 29 of EX1032, item number "8" under "*Refereed Conferences*" is identified as "'ParallAX: An Architecture for Real-Time Physics', Tom Yeh, Petros Faloutsos, Sanjay Patel, Glenn Reinman, in 34th Annual International Symposium on Computer Architecture, pp. 232-243, June 2007," which is the title and conference identified on the face of Yeh-ISCA07. At the end of item number "8" is a hyperlink to a "pdf." Clicking on that hyperlink

downloads a PDF of Yeh-ISCA07 (EX1012) that is identical to Yeh-ISCA07 at EX1012. Pages 60-71 of EX1032 which is substantively identical to EX1012, is the Internet Archive's archived capture of the PDF as it existed on February 21, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080221202312/http://www.cs.ucla.edu/~pfal/papers/ISCA-07-ParallAX.pdf. *See* EX1032, ¶¶5-6, p. 59.

85. Based on the above evidence, I too conclude that Yeh-ISCA07 would have also been publicly accessible, via Dr. Faloutsos' website, to a POSA exercising reasonable diligence by February 21, 2008.

2. Public Accessibility of EX1013 ("Yeh-Micro07")

86. EX1013 is titled "The Art of Deception: Adaptive Precision Reduction for Area Efficient Physics Acceleration." Based on the evidence I discuss below, it is my opinion that Yeh-Micro07 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence.

87. According to the face of EX1013, Yeh-Micro07 was published by the "IEEE Computer Society" in connection with "The 40th Annual IEEE/ACM International Symposium on Microarchitecture" on December 1-5, 2007 in Chicago, Illinois, USA. *See also* §I.B above (discussing a POSA's knowledge of

the IEEE as a reputable compiler and publisher of scientific and technical books and journals). Page 2 of Yeh-Micro07 similarly bears the text "© 2007 IEEE," indicating it was copyrighted and thus publicly accessible in the year 2007. According to Dr. Bolin, the IEEE Xplore database indicates Yeh-Micro07 was added December 26, 2007. Based on this evidence, Dr. Bolin concluded that Yeh-Micro07 was indexed, searchable and publicly accessible no later than December 2007. Bolin, ¶¶75-80. Testimony from an IEEE representative demonstrates public accessibility of EX1013 no later than December 5, 2007. *See* EX1127, ¶¶6-11. Based on the above evidence, I too conclude that Yeh-Micro07 would have been publicly accessible to a POSA exercising reasonable diligence in December 2007.

88. My opinion that Yeh-Micro07 would have been publicly accessible to a POSA exercising ordinary is further corroborated by the fact that a copy of Yeh-Micro07 was available from Dr. Faloutsos' website by at least early 2008, as I discuss below. As I discussed (§II.C.1 above), a POSA exercising reasonable diligence had reason to and could have navigated to Dr. Faloutsos' homepage hosted on the UCLA website, as demonstrated by the Internet Archives archived captures on Pages 4-31 of EX1032.

89. As shown on Page 29 of EX1032, item number "2" under Dr. Faloutsos' list of "*Refereed Conferences*" is identified as "'The Art of Deception:

Adaptive Precision Reduction for Area Efficient Physics Acceleration', Tom Yeh, Petros Faloutsos, Sanjay Patel, Milos Ercegovac, and Glenn Reinman, The 40th Annual International Symposium on Microarchitecture (MICRO), Dec 2007, to appear," which is the title and conference identified on the face of Yeh-Micro07. At the end of item number "2" is a hyperlink to a "pdf." Clicking on that hyperlink downloads a PDF of Yeh-Micro07 (EX1013) that is identical to Yeh-Micro07 at EX1013. Pages 46-58 of EX1032, which is substantively identical to EX1013, is the Internet Archive's archived capture of the PDF as it existed on February 21, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080222132652/http://www.cs.ucla.edu/~pfal/papers/MICRO-07-art.pdf.

90. Based on the above evidence, I too conclude that Yeh-Micro07 would have also been publicly accessible, via Dr. Faloutsos' website, to a POSA exercising reasonable diligence by February 21, 2008.

3. Public Accessibility of EX1014 ("Yeh Fool Me Twice")

91. EX1014 is an article titled "Fool Me Twice: Exploring and Exploiting Error Tolerance in Physics-Based Animation." Based on the evidence I discuss below, it is my opinion that EX1014 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged

patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence.

92. As discussed (§II.C.2 above), Yeh-Micro07 (EX1013) was published by the IEEE in 2007 and was also publicly accessible from Dr. Faloutsos' website by February 21, 2008. Reference [34] in Yeh-Micro07 is identified as an article by, among others, Dr. Faloutsos titled "Fool me twice: Exploring and exploiting error tolerance in physics-based animation," which is the title of EX1014. *See* Yeh-Micro07, p. 406. Yeh-Micro07's reference also indicates that a copy of the article was accessible from a webpage hosted at www.cs.ucla.edu (i.e., UCLA's Computer Science Department). Thus, a POSA interested in Yeh-Micro07 – published in the IEEE – and exercising reasonable diligence would have navigated to the website for UCLA's computer science department to retrieve a copy of EX1014 from the URL indicated in Yeh-Micro07 and, if not there, via the homepage of one of the other author's identified in Yeh-Micro07.

93. A copy of EX1014 was available from Dr. Faloutsos' website by at least early 2008. As I discussed (§II.C.1 above), a POSA exercising reasonable diligence had reason to and could have navigated to Dr. Faloutsos' homepage hosted on the UCLA website, as demonstrated by the Internet Archives archived captures on Pages 4-31 of EX1032. As shown on Page 29 of EX1032, item number "1" under "*Refereed Journals*" is identified as "'Fool Me Twice:

Exploring and Exploiting Error Tolerance in Physics-Based Animation', Tom Yeh, Sanjay Patel, Petros Faloutsos, Glenn Reinman, ACM Transactions on Graphics, accepted with revisions, 2007," which is the title and journal identified on the face of EX1014. At the end of item number "1" is a hyperlink to a "pdf." Clicking on that hyperlink downloads a PDF of the journal paper that is identical to EX1014. Pages 34-44 of EX1032, which has been excerpted as EX1014, is the Internet Archive's archived capture of the PDF as it existed on February 21, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080222132652/http://www.cs.ucla.edu/~pfal/papers/TOG-08-Fool.pdf. Based on the above evidence, I too conclude that EX1014 was publicly accessible, via Dr. Faloutsos' website, to a POSA exercising reasonable diligence by February 21, 2008.

4. Public Accessibility of EX1016 ("Moore 1965")

94. EX1016 is titled "Cramming More Components onto Integrated Circuits." As seen on the face of the document, EX1016 was published in the January 1998 edition of the Proceedings of the IEEE. *See also* EX1016, p. 3 (indicating this as Vol. 86, No. 1 of the proceedings of the IEEE "published monthly"). As I discussed above (§I.B), the IEEE is a well-known, reputable and widely used resources for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines. As shown on

Page 5 of EX1016, the article "Cramming More Components onto Integrated Circuits" by Gordon Moore (of Moore's Law fame), was originally published on April 19, 1965 in the journal "Electronics." IEEE Xplore indicates that Vol. 86, Issue 1 of the Proceedings of the IEEE, which republished Moore 1965, had a "Date of Publication" of "January 1998."[8] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines").

95. Based on the evidence above, it is my opinion that EX1017 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by January 1998.

5. Public Accessibility of EX1017 ("Moore 1975")

96. EX1017 is titled "Progress in Digital Integrated Electronics." As seen on the face of the document, EX1017 was printed in the September 2006 edition of the IEEE SSCS (Solid-State Circuits Society) Newsletter. EX1017 bears a "Copyright 1975 IEEE" based on the article's original publication in the Technical

[8] https://ieeexplore.ieee.org/document/658762

Digest, International Electron Devices Meeting. IEEE Xplore indicates that Vol. 11, Issue 3 of the IEEE Solid-State Circuits Society Newsletter, which republished Moore 1975, had a "Date of Publication" of "September 2006."[9] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the evidence above, it is my opinion that EX1017 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by September 2006.

6. Public Accessibility of EX1018 ("Mendocino")

97. EX1018 is titled "Mendocino Improves Celeron." According to the face of the document, EX1018 was published in Volume 12, No. 11 of Microprocessor Report ("MPR") on August 24, 1998. I am personally familiar with MPR as a well-known, reputable newsletter in the microprocessor industry and in my opinion POSAs would have also been aware of MPR at the time of the alleged invention and would have considered MPR to be reputable source of information in the microprocessor industry. In the 2000s, MPR was widely

[9] https://ieeexplore.ieee.org/document/4804410

distributed to POSAs interested in the microprocessor industry on or around the dates indicated on the face of an article published by Microprocessor Report.

98. The Internet Archive's Wayback Machine has several archives of the MPR's website from the 2000s. For example, the Internet Archive has an archive of that website as it existed on April 2, 2001, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20010402030508/http://mpronline.com/. The MPR's main webpage had a search box in the upper lefthand, which would have allowed a user to keyword search the entire contents of the MPR website, including all articles published by MPR. As shown at the bottom of the April 2, 2001 archive of the MPR's website, there is a link to "Subscribe," which navigates to the "About Microprocessor Report" page as it existed on February 5, 2004, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20040205074721/http://www.mpronline.com/mpr_public/about.html. As shown on that page, any interested member of the public could subscribe to MPR to "access current articles and **an eight-year archive** at www.MPRonline.com."

99. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I

above), such that a POSA interested in the semiconductor industry or interested in the roadmaps, could have located EX1018 by searching the Web for keywords of interest such as "Microprocessor Report" or "Mendocino," which would be terms of interest to a POSA interested in the semiconductor industry generally and/or information about Intel's Mendocino processor specifically.

100. Based on the evidence I discussed above, it is my opinion that EX1018 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence by (at least) the 1999-2001 timeframe, but certainly no later than the 2004 timeframe from when archives of articles in Microprocessor Report could be obtained from MPRonline.com, as I discussed above.

7. Public Accessibility of EX1019 ("Sim")

101. EX1019 (which I refer to herein as "Sim") is titled "MIMD–SIMD hybrid system—towards a new low cost parallel system." As indicated on the face of EX1019, Sim was published in Volume 29 of the journal "Parallel Computing" in 2003. Parallel Computing "is an international journal presenting the practical use of parallel computer systems, including high performance architecture, system

software, programming systems and tools, and applications."[10] I am personally aware of Parallel Computing as a widely distributed, reputable journal in the field of computing and thus, in my opinion, journal that POSAs interested in the field of computing in the 2000 timeframe had reason to review articles published in Parallel Computing.

102. The face of EX1019 bears "© 2002 Elsevier Science B.V.," indicating it was published by Elsevier Science in 2002. Elsevier is "a Dutch academic publishing company specializing in scientific, technical, and medical content. Its products include journals such as The Lancet, Cell, the ScienceDirect collection of electronic journals."[11] "Elsevier published more than 600,000 articles annually in over 2,700 journals; as of 2018 its archives contained over 17 million documents and 40,000 e-books, with over one billion annual downloads."[12] Sim was also "Available online" from www.sciencedirect.com on December 4, 2002, which also indicates Sim was published on pages 21-36 of Issue 1, Volume 29 of Parallel Computing.[13] ScienceDirect, which "launched in March 1997," is "a website which

[10] https://www.sciencedirect.com/journal/parallel-computing

[11] https://en.wikipedia.org/wiki/Elsevier (last visited Dec. 19, 2022)

[12] *Id.*

[13] https://www.sciencedirect.com/science/article/abs/pii/S0167819102001825

provides access to a large bibliographic database of scientific and medical publications of the Dutch publisher Elsevier. It hosts over 18 million pieces of content from more than 4,000 academic journals and 30,000 e-books of this publisher."[14] I was personally aware of www.sciencedirect.com in the 2000 timeframe as a reputable digital database for publications in the field of computing and, in my opinion, POSAs in the 2000 timeframe would have been aware of www.sciencedirect.com as indexed, searchable, and publicly accessible source of publications in the field of computing.

103. Based on the evidence above and my understanding of the Parallel Computing journal and Science Direct website, it is my opinion that Sim is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence by December 4, 2002 (from ScienceDirect.com), but not later than 2003 (from Volume 29, Issue 1 of the journal "Parallel Computing").

8. Public Accessibility of EX1020 ("Taveniku")

104. EX1020 (which I refer to herein as "Taveniku") is titled "The VEGA Moderately Parallel MIMD, Moderately Parallel SIMD, Architecture for High

[14] https://en.wikipedia.org/wiki/ScienceDirect (last visited Dec. 19, 2022)

Performance Array Signal Processing." Page 226 of Taveniku bears a copyright "© 1998 IEEE," which indicates it was published in 1998 by the IEEE. IEEE Xplore indicates that Taveniku was published by the IEEE in connection with a conference held on March 30, 1998 – April 3, 1998, and had a "Date Added to IEEE Xplore" of August 6, 2002 and had a "Print ISBN" number and "Print ISSN" number.[15] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines").

105. Based on the above evidence and my understanding of the IEEE and IEEE Xplore, it is my opinion that EX1020 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by August 2002.

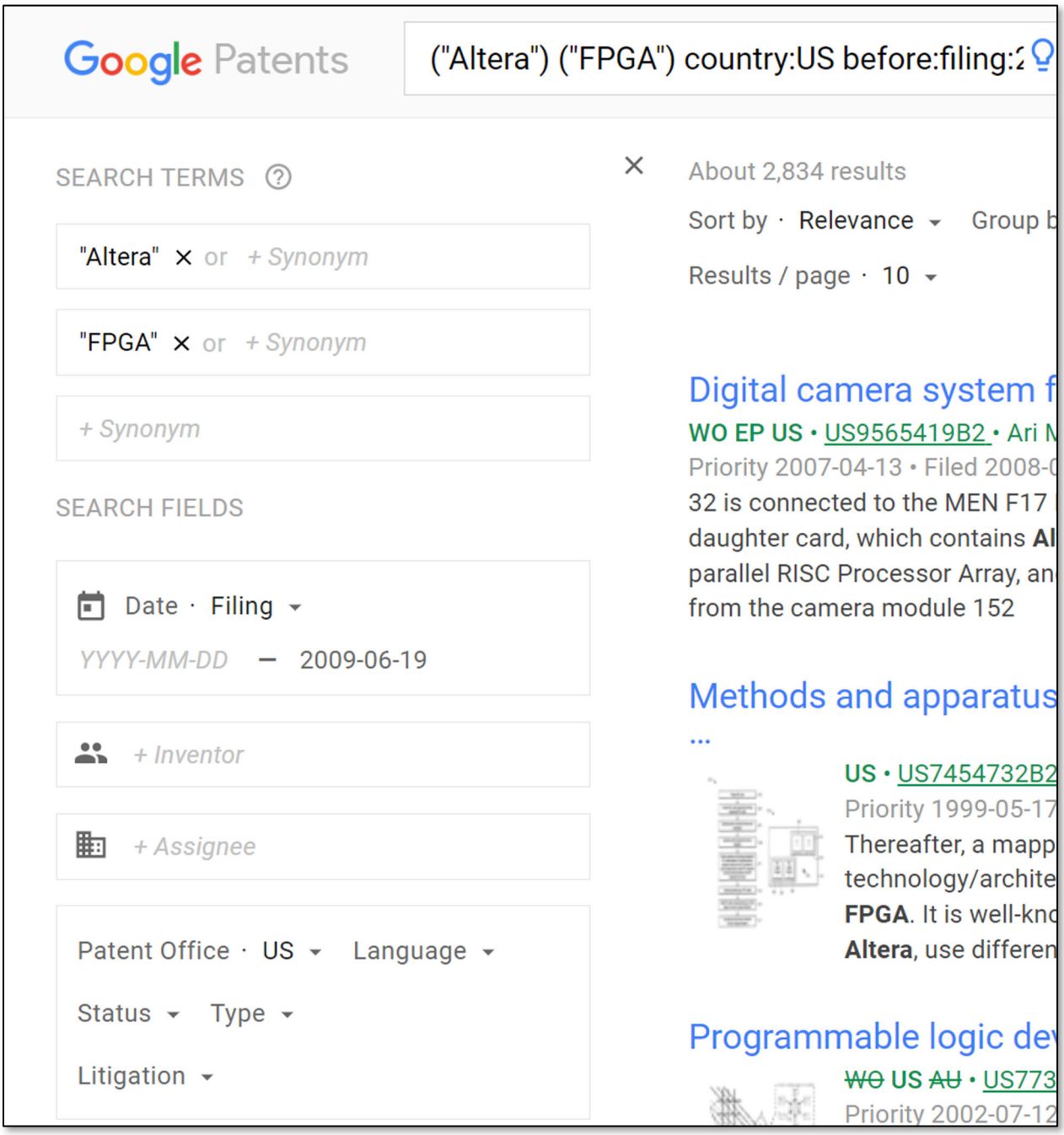
9. Public Accessibility of EX1038 ("Altera Product Catalog")

106. EX1038 is titled "Altera Product Catalog" from Altera, which "was a manufacturer of programmable logic devices (PLDs) headquartered in San Jose, California" that "was founded in 1983 and acquired by Intel in 2015."[16] As

[15] https://ieeexplore.ieee.org/document/669915

[16] https://en.wikipedia.org/wiki/Altera (last visited Dec. 14, 2002)

corroborated by Intel's acquisition of Altera, Altera was one of the leading manufacturers of PLDs at the time of the alleged invention. For example, a recent search of Google Patents for US patents or patent applications filed before June 19, 2009 with the words "Altera" and "FPGA" returned 2,834 results:



Thus, POSAs interested in the field of computing exercising reasonable diligence had reason to navigate to Altera's website to review the information and documentation available there, such as white papers and press releases, available from Altera's website regarding Altera's products, research, and development.

107. Page 72 of EX1038 bears a "Copyright © 2008 Altera Corporation," which indicates it was published in 2008. The very bottom of page 72 states "June 2008," which is consistent with a 2008 Copyright date.

108. Page 112 of EX1048 is the Internet Archive's archived capture of Altera's homepage, http://www.altera.com/, as it existed on June 2, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080602092253/http://www.altera.com/. *See* EX1048, ¶¶5-6, p. 111. At the top of the Altera's homepage there is the text "Products," which navigates to Altera's products webpage at http://www.altera.com/products/prd-index.html. Page 110 of EX1048 is the Internet Archive's archived capture of that webpage as it existed on June 2, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080602092611/http://www.altera.com/products/prd-index.html. *See* EX1048, p. 109.

109. As shown on Page 110 of EX1048, there is the text "View All" next to the heading for Altera's "Devices," which navigates to Altera Devices webpage

at http://www.altera.com/products/devices/dev-index.jsp. Page 108 of EX1048 is the Internet Archive's archived capture of that webpage as it existed on May 29, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080529214527/http://www.altera.com/products/devices/dev-index.jsp. *See* EX1048, p. 107.

110. Near the bottom right of that page is a hyperlink with the text "***Product Catalog** (PDF)*," which I right clicked on to save the document at the link, which downloaded a PDF that was at the URL http://www.altera.com/literature/sg/product-catalog.pdf. Pages 82-153 of EX1047 (which is EX1038) is the Internet Archive's capture of that PDF from December 20, 2008, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20081220145857/http://www.altera.com:80/literature/sg/product-catalog.pdf. *See* EX1047, p. 81.

10. Public Accessibility of EX1039 ("Intel-Itanium")

111. EX1039 (which I refer to herein as "Intel-Itanium") is titled "A 65 nm 2-Billion Transistor Quad-Core Itanium Processor." As indicated on the face of EX1039, Intel-Itanium was published by the IEEE in the IEEE Journal of Solid-State Circuits (Volume: 44, Issue: 1, January 2009). I have reviewed EX1039, which appears to be a library scan of a physical copy of the issue of the IEEE Journal of Solid-State Circuits that contained Stackhouse, which is listed in the

TOC on page 3 of EX1039. Page 2 of EX1039 bears a "Copyright © 2009 by The Institute of Electrical and Electronics Engineers, Inc." and an "ISSN" number. Pages 1 and 3 of EX1039 bear a stamp from the "Linda Hall Library." Page 5 of Intel-Itanium bears a Copyright "© 2008 IEEE," which indicates it was published in 2008. IEEE Xplore indicates that Intel-Itanium had a "Date of Publication" of December 30, 2008.[17] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Testimony from an IEEE representative demonstrates public accessibility of EX1039 no later than December 31, 2008. *See* EX1128, ¶¶6-11.

112. Based on the above evidence and my review of the indicia appearing on the Intel-Itanium document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Intel-Itanium is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by January 2009. *See also* Bolin, ¶¶101-105 (concluding that attachment 12b to her declaration, an IEEE conference paper about the Itanium Processor by several authors indicated on the

[17] https://ieeexplore.ieee.org/document/4735536

face of EX1039, "was publicly accessible at least as of January 9, 2009 and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence, on their own or with the assistance of a research librarian").

11. Public Accessibility of EX1040 ("Janakiraman")

113. EX1040 (which I refer to herein as "Janakiraman") is titled "Performance Analysis of Intel® Pentium® M and Pentium 4 Processors." As indicated on the face of EX1040, Janakiraman is a white paper by Dell System Performance Analysis and bears the date "July 2005." Page 5 of EX1040 bears a copyright "© 2005 Dell Inc.," indicating it was published in 2005. Dell was and still is a large developer and seller of computing devices, like personal computers and servers. In the mid-2000s, Dell had an active research and development group that produced various white papers in the field of computing that POSAs exercising reasonable diligence would have been aware of to learn about developments in the industry. *See, e.g.*, Page 12 of EX1048 (showing the Internet Archive's archived capture of the "Innovation @ Dell" webpage under the Dell.com domain as it existed on February 3, 2007, as reflected in the Internet Archives extended URL: https://web.archive.org/web/20070203155947/http://www.dell.com/content/topics/global.aspx/innovation/en/index?c=us&l=en&s=corp).

114. Page 8 of EX1048 is the Internet Archive's archived capture of a web page under the Dell.com domain summarizing Janakiraman, http://www.dell.com/content/topics/global.aspx/vectors/en/2005_p4_pentiumm?c=us&l=en&s=corp, as it existed on September 6, 2007, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20070906020846/http:/www.dell.com/content/topics/global.aspx/vectors/en/2005_p4_pentiumm?c=us&l=en&s=corp. *See* EX1048, ¶¶5-6, p. 7. *See also* EX1048, pp. 9-16 (showing the Internet Archive's captures of Dell.com's homepage and webpages from which POSAs could navigate to EX1048, as they existed in February 2007, as reflected in the Internet Archive's extended URLs). On the right of the Dell.com webpage summarizing Janakiraman there is a hyperlink to "Download PDF," that the Internet Archived captured on February 5, 2007, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20070205153105/http://www.dell.com/downloads/global/vectors/2005_p4_pentiumm.pdf. That PDF is EX1048. *See* EX1047, ¶¶5-6, pp. 60-65. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry, could have located EX1048 by searching the Web for keywords of interest such as "Pentium 4," die size," and

”transistor process,” which are terms that appear in EX1048 and would be terms of interest to a POSA interested in the computing industry.

115. Based on the above evidence and my review of the indicia appearing on EX1048, it is my opinion that EX1048 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by February 2007.

12. Public Accessibility of EX1041 (“Radeon-5700”)

116. EX1041 (which I refer to herein as the “Radeon-5700 article”) is an article from guru3d.com titled “Radeon HD Series 5700 features.” According to the face of EX1041, EX1041 was published on October 21, 2009 on the website guru3d.com, a website which has “well over 15 million pageviews monthly” and “is one of World’s leading Interactive Entertainment and PC Hardware fan sites.”[18] The site “provides up-to-date news on latest developments in the interactive entertainment industry and in-depth reviews and articles targeted at the PC enthusiast and gamer” and “covers the hottest hardware such as video cards, [and] mainboards.”[19] An article that published on December 1, 2008 in the “ACM

[18] https://www.guru3d.com/content-page/contact-us.html

[19] *Id.*

Transactions on Graphics" regarding algorithm to exploit "the fine-grained parallelism of GPUs" cited an article from guru3d.com as a reference, demonstrating that guru3d.com was a reliable source of information in the computing industry that POSAs would have been aware of by the mid-2000s timeframe.[20] *See also* §I.C above (discussing ACM and ACM Digital Library).

117. Pages 120-127 of EX1048 is the Internet Archive's archived capture of The Guru of 3D's homepage, https://www.guru3d.com, as it existed on October 26, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20091026163855/https://www.guru3d.com/. *See* EX1048, ¶¶5-6, p. 119. At the right of the guru3d.com homepage, there is the text "Quick article links," which includes hyperlinked text that says "Radeon HD 5750 review (Crossfire)", which navigates to guru3d.com's review of the ATI Radeon HD 5750 GPU webpage at http://www.guru3d.com/article/radeon-hd-5750-review-crossfirex/. Pages 117-118 of EX1048 is the Internet Archive's archived capture of that webpage as it existed on October 25, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20091025050805/http://www.guru3d.com/article/rade

[20] https://dl.acm.org/doi/abs/10.1145/1409060.1409079 (citing a product review hosted on Guru3D.com for a device in the computing industry)

on-hd-5750-review-crossfirex/. *See* EX1048, p. 116. As shown near the top of that page there is a drop-down box to navigate to different portions of the review. Selecting the text "Radeon HD Series 5700 features" from the drop down list navigates to the webpage at http://www.guru3d.com/article/radeon-hd-5750-review-crossfirex/3. Pages 13-14 of EX1047 (which are EX1041) is the Internet Archive's archived capture of that webpage as it existed on September 26, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090926023317/http://www.guru3d.com/article/radeon-hd-5870-review-test/3. *See* EX1047, ¶¶5-6, p. 12.

118. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry and data about commercial GPUs, could have located EX1041 by searching the Web for keywords of interest such as "processing elements" and "GPU," "process," "transistors," and "die size," which are terms that appear in EX1041 and would be terms of interest to a POSA interested in the computing industry.

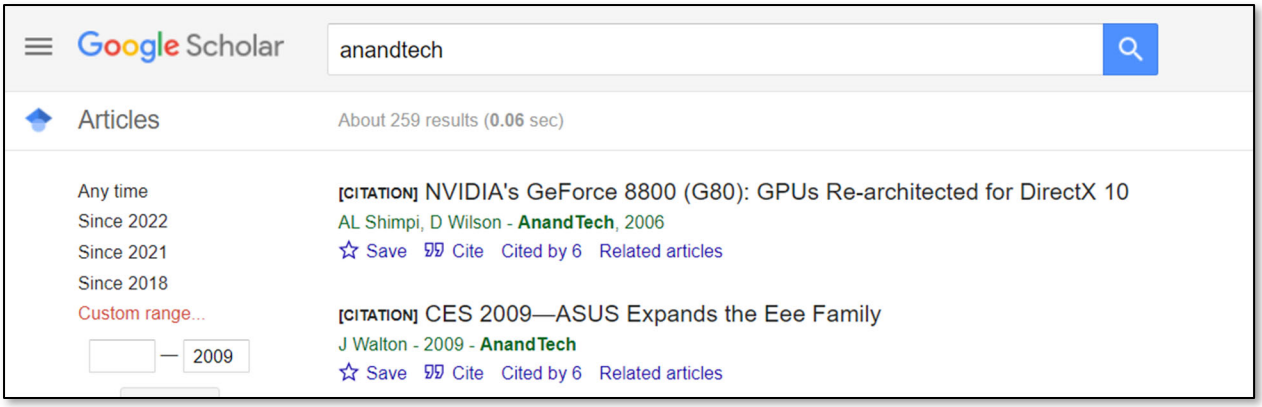
119. Based on the above evidence and my review of the indicia appearing on EX1041, it is my opinion that EX1041 was publicly accessible before the June 15, 2010 filing date for U.S. Appl. No. 12/816,201 (challenged patent, code (63))

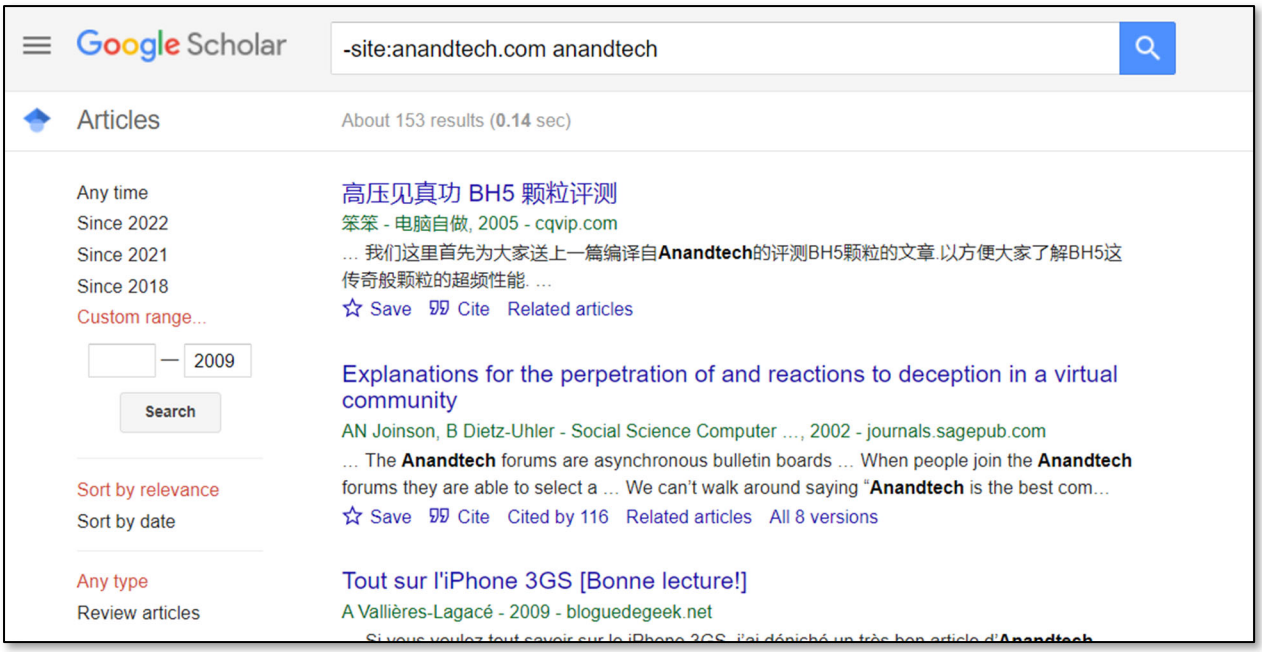
by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by October 26, 2009.

13. Public Accessibility of EX1042 ("Penryn")

120. EX1042 (which I refer to as "Penryn") is an article from AnandTech.com titled "Intel Core 2 Extreme QX9650 - Penryn Ticks Ahead." The face of EX1042 is dated "Oct[ober] 29, 2007," indicating it was publicly accessible from AnandTech.com by that date. AnandTech is website that reviews computing products like GPUs and microprocessors that began in 1997 (though under different domain).[21] A recent search of Google Scholar returned 259 articles from 2009 and earlier that included "anandtech" and 153 articles hosted on sites that were not under the anandtech.com domain name, demonstrating that AnandTech was a well-known website for those interested in technology, such as the technology of computing devices. *See also* Bolin, ¶6 (identifying Google Scholar as a reliable repository of scholarly and research data).

[21] https://www.anandtech.com/home/about/





121. Articles from AnandTech have also been cited in various publications from the ACM and IEEE, demonstrating that AnandTech was a reliable source of information in the computing industry that POSAs would have been aware of by

the mid-2000s timeframe.[22] *See also* U.S. Patent Publication Application 2008/0094427A1 (filed October 28, 2005), at [0055] (citing article from anandtech.com regarding a graphics control).

122. Pages 70-71 of EX1048 is the Internet Archive's archived capture of AnandTech's homepage, https://www.anandtech.com, as it existed on October 29, 2007, as reflected in the Internet Archive's extended URL:

---

[22] *See, e.g.*, https://ieeexplore.ieee.org/abstract/document/1412184/references#references (articled added to IEEE Xplore on April 4, 2005; reference no. 15 from anandtech.com); https://ieeexplore.ieee.org/abstract/document/4523358/references#references (articled published by the IEEE on May 16, 2008; reference no. 15 from anandtech.com). *See, e.g.*, https://ieeexplore.ieee.org/abstract/document/1412184/references#references (articled added to IEEE Xplore on April 4, 2005; reference no. 15 from anandtech.com); https://ieeexplore.ieee.org/abstract/document/4523358/references#references (articled published by the IEEE on May 16, 2008; reference no. 15 from anandtech.com).

https://web.archive.org/web/20071029173916/https://www.anandtech.com/. *See* EX1048, ¶¶5-6, p. 69. In the middle of AnandTech.com's homepage there is hyperlinked text "Intel Core 2 Extreme QX9650 – Penryn Ticks Ahead," which navigates to AnandTech's review of Intel's Penryn chip at http://www.anandtech.com/cpuchipsets/intel/showdoc.aspx?i=3137. Pages 30-40 of EX1047 (which is EX1042) is the Internet Archive's archived capture of that webpage as it existed on October 31, 2007, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20080602092611/http://www.altera.com/products/prd-index.html. https://web.archive.org/web/20071031095146/http://www.anandtech.com/printarticle.aspx?i=3137. *See* EX1047, p. 29.

123. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry and data about commercial GPUs and microprocessors, could have located EX1042 by searching the Web for keywords of interest such as "Intel" and "45nm," "65nm," "Intel," "AMD," and "transistors," which are terms that appear in EX1042 and would be terms of interest to a POSA interested in the computing industry.

124. Based on the above evidence and my review of the indicia appearing on EX1041, it is my opinion that EX1041 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by October 2007.

14. Public Accessibility of EX1043 ("TSMC 40nm Press Release 1")

125. EX1043 is a press release by TSMC titled "TSMC Ramps Most Advanced Available Process Technology to Volume Production" and appears to be a press release "Issued by: TSMC." EX1043, 1. The face of EX1043 includes the date "November 17, 2008," which indicates it was published by TSMC on November 17, 2008. I retrieved EX1043 from TSMC's website at the URL https://pr.tsmc.com/english/news/1526. Because TSMC is one of the leading semiconductor manufacturers in the world and a publicly traded company, in my opinion the date identified on the face of a TSMC press release is a reliable indication of when the press release would have been publicly available from TSMC's website. Based on the above evidence and my review of the indicia appearing on the face of EX1043, it is my opinion that EX1043 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or

otherwise made available to POSAs exercising reasonable diligence at least by November 17, 2008.

15. Public Accessibility of EX1044 ("CPU Cheatsheet")

126. EX1044 (which I refer to as "CPU Cheatsheet") is an article from AnandTech.com titled "Updated CPU Cheatsheet - Seven Years of Covert CPU Operations." The face of EX1044 is dated "Aug[ust] 28, 2004," indicating it was publicly accessible from AnandTech.com by that date. As I discussed above (§II.C.13) Anandtech.com is a source of information for commercial GPUs and microprocessors that POSAs interested in the field of computing would have been aware of considered a reliable source of information regarding commercial chips.

127. Pages 73-74 of EX1048 is the Internet Archive's archived capture of AnandTech's homepage, https://www.anandtech.com, as it existed on August 29, 2004, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20040829013031/http://anandtech.com/. *See* EX1048, ¶¶5-6, p. 72. Near the top of that webpage there is hyperlinked text "Updated CPU Cheatsheet - Seven Years of Covert CPU Operations," which navigates to http://www.anandtech.com/cpuchipsets/showdoc.aspx?i=2178. Pages 51-52 of EX1047 is the Internet Archive's archived capture of that page as it existe3d on September 9, 2004, as reflected in the Internet Archives extended URL:

https://web.archive.org/web/20040909115936/http://www.anandtech.com/cpuchipsets/showdoc.aspx?i=2178.

128. Towards the bottom of page there is the text "Print this article." Clicking on that navigates to a webpage containing the entire "Updated CPU Cheatsheet - Seven Years of Covert CPU Operations" article. Pages 42-49 of EX1047 (which is EX1044) is the Internet Archive's archived capture of the entire article as it existed on January 22, 2005, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20050122063415/http://www.anandtech.com/printarticle.aspx?i=2178. *See* EX1047, p. 41.

129. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry and data about commercial GPUs and microprocessors, could have located EX1042 by searching the Web for keywords of interest such as "Intel," "AMD," "die size," and "transistor count," which are terms that appear in EX1044 and would be terms of interest to a POSA interested in the computing industry.

130. Based on the above evidence and my review of the indicia appearing on EX1044, it is my opinion that EX1044 is prior art to the challenged patent

because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by January 2005.

16. Public Accessibility of EX1045 ("Natarajan")

131. EX1045 (which I refer to herein as "Natarajan") is titled "A 32nm logic technology featuring 2nd-generation high-k + metal-gate transistors, enhanced channel strain and 0.171μm2 SRAM cell size in a 291Mb array." IEEE Xplore indicates that Natarajan was "Published in: 2008 IEEE International Electron Devices Meetings" by the IEEE in connection with a conference held on December 15-17, 2008, and had a "Date Added to IEEE Xplore" of February 27, 2009 and had an "ISBN" number and "Print ISSN" number.[23] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Testimony from an IEEE representative demonstrates public accessibility of EX1045 no later than December 17, 2008. *See* EX1129, ¶¶6-11. Based on the above evidence and my understanding of the IEEE and IEEE Xplore, it is my opinion that Natarajan is prior art to the challenged patent because it was publicly accessible before the

[23] https://ieeexplore.ieee.org/document/4796777

earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence by December 17, 2008, or at least no later than February 27, 2009.

17. Public Accessibility of EX1052 ("Intel's 32nm Update")

132. EX1052 is an article from AnandTech.com titled "Intel's 32nm Update: The Follow-on to Core i7 and More." The face of EX1052 is dated "Feb[ruary] 11, 2009," indicating it was publicly accessible from AnandTech.com by that date. As I discussed above (§II.C.13) Anandtech.com is a source of information for commercial GPUs and microprocessors that POSAs interested in the field of computing would have been aware of considered a reliable source of information regarding commercial chips.

133. Pages 63-64 of EX1048 is the Internet Archive's archived capture of AnandTech's homepage, https://www.anandtech.com, as it existed on June 3, 2009 as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090603130817/https://www.anandtech.com/. *See* EX1048, ¶¶5-6, p. 62. Near the top of that page there is a tab with the text "CPU/Chipset." Clicking on that text navigates to AnandTech's "CPU & Chipset Section" at http://www.anandtech.com/cpuchipsets/. The Internet Archive captured that page as it existed on May 31, 2009, as reflected in the Internet Archive's extended URL:

https://web.archive.org/web/20090531115149/http://www.anandtech.com/cpuchipsets/. *See* EX1048, ¶¶5-6; *see also* EX1048, pp. 56-59 (Internet Archives capture of that same webpage as it existed on June 17, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090617064007/http://www.anandtech.com/cpuchipsets/default.aspx?manu=77&topicid=101).

134. Towards the middle of the CPU & Chipset Section webpage there is hyperlinked text "Intel's 32nm Update: The Follow-on to Core i7 and More," which navigates to http://www.anandtech.com/cpuchipsets/showdoc.aspx?i=3513. Pages 53-54 of EX1048 is the Internet Archive's archived capture of that page as it existed on June 1, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090601070325/http://www.anandtech.com/cpuchipsets/showdoc.aspx?i=3513. See EX1048, ¶¶5-6, p. 52.

135. Towards the bottom of page there is the text "Print this article." Clicking on that navigates to a webpage containing the entire "Intel's 32nm Update: The Follow-on to Core i7 and More" article. Pages 21-28 of EX1047 (which is EX1052) is the Internet Archive's archived capture of the entire article as it existed on December 31, 2009, as reflected in the Internet Archive's extended URL:

https://web.archive.org/web/20091231063002/http://anandtech.com/printarticle.aspx?i=3513. *See* EX1047, ¶¶5-6, p. 20.

136. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry and data about commercial GPUs and microprocessors, could have located EX1052 by searching the Web for keywords of interest such as "Intel," "32nm process," "CPUs," amd "dual-core," which are terms that appear in EX1052 and would be terms of interest to a POSA interested in the computing industry.

137. Based on the above evidence and my review of the indicia appearing on EX1052, it is my opinion that EX1052 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by June 17, 2009 (as reflected in the Internet Archive's extended URL for the main page of the "Intel's 32nm Update: The Follow-on to Core i7 and More," as discussed above). At the very latest, EX1052 was publicly accessible by December 31, 2009 (as reflected in the Internet Archive's extended URL for printing the entire article, as discussed above).

18. Public Accessibility of EX1053 ("AMD Evergreen")

138. EX1053 (which I refer to herein as "AMD Evergreen") is an article from AnandTech.com titled "AMD Evergreen: 40nm DirectX 11 GPU Silicon Arrives." The face of EX1053 is dated "June 3, 2009," indicating it was publicly accessible from AnandTech.com by that date. As I discussed above (§II.C.13) Anandtech.com is a source of information for commercial GPUs and microprocessors that POSAs interested in the field of computing would have been aware of considered a reliable source of information regarding commercial chips.

139. Pages 63-64 of EX1048 is the Internet Archive's archived capture of AnandTech's homepage, https://www.anandtech.com, as it existed on June 3, 2009 as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090603130817/https://www.anandtech.com/. *See* EX1048, ¶¶5-6, p. 62. Near the top of that page is hyperlinked text that reads "AMD Evergreen: 40nm DirectX 11 GPU Silicon Arrives," which navigates to http://www.anandtech.com/video/showdoc.aspx?i=3573. Pages 16-19 of EX1047 (which is EX1053) is the Internet Archive's archived capture of that page as it existed on June 6, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090606053120/http://www.anandtech.com/video/showdoc.aspx?i=3573. *See* EX1047, ¶¶5-6, p. 15.

140. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry and data about commercial GPUs and microprocessors, could have located EX1053 by searching the Web for keywords of interest such as "AMD," "GPU," "transistors," and "die" which are terms that appear in EX1053 and would be terms of interest to a POSA interested in the computing industry.

141. Based on the above evidence and my review of the indicia appearing on EX1053, it is my opinion that EX1053 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by June 6, 2009.

19. Public Accessibility of EX1088 ("VSI Standard")

142. EX1088 (which I refer to herein as the "VSI Standard") is titled "VSI Alliance™ Virtual Component Interface Standard Version 2 (OCB 2 2.0) On-Chip Bus Development Working Group April 2001." Stuttard (EX1006) refers to the "Virtual Component Interface Standard by the VSI Alliance" (Stuttard, 13:7-9), which indicates that the VSI Alliance's Virtual Component Interface Standard was publicly accessible to POSAs interested in its subject matter. Thus, POSAs

interested in Stuttard's teachings would have had reason to obtain standard documents published by the VSI Alliance, such as by navigating to the VSI Alliance's website as it was (and still is) conventional for standards bodies and alliances to provide copies of their standards and publications for members and other interested persons on their websites. As shown below, a recent search of Google Scholar returned 465 articles from 2009 and earlier that included the VSI Alliance's website – www.vsi.org –including IEEE publications, corroborating my opinion that POSAs would have been well-aware of the VSI Alliance's website and standards. *See also* Bolin, ¶6 (identifying Google Scholar as a reliable repository of scholarly and research data).



143. Pages 105-106 of EX1048 is the Internet Archive's archived capture of the VSI Alliance's homepage, https://vsi.org/, as it existed on January 22, 2002, as reflected in the Internet Archive's extended URL:

https://web.archive.org/web/20020122030423/https://vsi.org. *See* EX1048, ¶¶5-6, p. 104. At the left of the VSI's homepage, there is the text "members list," which navigates to the VSI Alliance's Member List webpage at http://www.vsi.org/members.htm. Pages 7-8 of EX1047 is the Internet Archive's archived capture of that webpage as it existed on February 10, 2002, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20020210130046/http://www.vsi.org/members.htm. *See* EX1047, p. 6. As shown on Pages 7-8 of EX1047, the VSI Alliance's Members included many of the largest and well-known corporations in the semiconductor and related industries, like Analog Devices, ARM, Hewlett-Packard, IBM, Motorola, Samsung, and TSMC.

144. At the top of the VSI's homepage (EX1048, 105), there is the text "VSIA Specifications," which navigates to the VSI Alliance's Technical Documents & Specifications webpage at http://www.vsi.org/library/specs/summary.htm. Pages 101-103 of EX1048 is the Internet Archive's archived capture of that webpage as it existed on February 15, 2002, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20020215225401/http://www.vsi.org/library/specs/summary.htm. *See* EX1048, p. 100.

145. As shown on Pages 101-103 of EX1048, the VSI Alliance's Technical Documents & Specifications webpage included links to retrieve various publications of the VSI Alliance, one of which include the "Virtual Component Interface Standard Version 2," which is the name of the document on the face of EX1088. The VSI's Alliance's Technical Documents & Specifications webpage also notes that this is the OCB 2 2.0 – Revision Date: April 2001, consistent with the face of EX1088 which includes the text "OCT 2 2.0" and the date "April 2001." As shown on Page 103 of EX1048, the Alliance's Technical Documents & Specifications webpage included hyperlinks for a "Member Download" and a "Non Member Buy a Single Copy here." Based on the above evidence, I conclude that EX1088 would have been publicly accessible to a POSA exercising reasonable diligence at least by the year 2002.

20. Public Accessibility of EX1094 ("Chen")

146. EX1094 (which I refer to herein as "Chen") is titled "A cost effective 32nm high-K/ metal gate CMOS technology for low power applications with single-metal/gate-first process." EX1094 appears to be a scan of a print copy of the cover matter of the journal in which the Chen article was published along with the table of contents for the publication and a copy of the Chen article. The face of EX1094 indicates that it was published by the IEEE in connection with the 2008 Symposium on VLSI Technology held on June 17-19, 2008 and provides an

"ISBN 13" number. Page 3 of EX1094 includes what appears to be a stamp from the SUNY Institute of Technology at Utica/Rome – Library. Page 4 of EX1094 includes a "Copyright © 2008 by The Institute of Electrical and Electronics Engineers, Inc.," which indicates it was published by the IEEE in 2008. Page 67 of the Chen article similarly includes a copyright "©2008 IEEE." IEEE Xplore indicates that Chen was "Published in: 2008 Symposium on VLSI" by the IEEE in connection with a conference held on June 17-19, 2008, and had a "Date Added to IEEE Xplore" of August 5, 2008 and had an "ISBN" number and "Print ISSN" number.[24] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Testimony from an IEEE representative demonstrates public accessibility of EX1094 no later than June 19, 2008. *See* EX1130, ¶¶6-11. Based on the above evidence and my understanding of the IEEE and IEEE Xplore, it is my opinion that Chen is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence by June 19, 2008, or at least by August 5, 2008.

[24] https://ieeexplore.ieee.org/document/4588573

21. Public Accessibility of EX1097 ("Schiml")

147. EX1097 (which I refer to herein as "Schiml") is titled "A 0.13µm CMOS platform with Cu/low-k interconnects for system on chip applications." As indicated on the face of EX1097, Schiml was published in connection with the 2001 Symposium on VLSI Technology Digest of Technical Papers. IEEE Xplore indicates that Schiml was published by the IEEE in connection with a June 12-14, 2001 conference and had a "Date Added to IEEE Xplore" of August 7, 2002 and had a "Print ISBN" number.[25] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Testimony from an IEEE representative demonstrates public accessibility of EX1097 no later than June 14, 2001. *See* EX1131, ¶¶6-11 Based on the above evidence and my review of the indicia appearing on the Schiml document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Schiml is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence by June 14, 2001, or at least by August 2002.

[25] https://ieeexplore.ieee.org/document/934969

22. Public Accessibility of EX1098 ("Sudha")

148. EX1098 (which I refer to herein as "Sudha") is titled "An Efficient Digital Architecture for Principal Component Neural Network and its FPGA Implementation." As indicated on the face of EX1098, Sudha was published in Volume 53, No. 5, September-October 2007 issue of the IETE Journal of Research. I have reviewed EX1098, the first page of which appears to be a library scan of the front page of a physical copy of the issue of the IETE Journal of Research in which the Sudha article was published. I am aware from my past work experience that the IETE, or Institute of Electronics and Telecommunication Engineers, is an Indian organization that is often considered to be the IEEE of India. Sudha, p. 1. As the official website for the IETE states, the IETE is "is India's leading recognized professional society devoted to the advancement of Science and Technology of Electronics, Telecommunication & IT" and has "more than 1,25,000 members."[26] The cover page of Sudha bears the label "Volume 53, No. 5, September-October 2007" and an "ISSN" number of 0377-2063. In addition, the cover page of Sudha bears a stamp from the "Linda Hall Library" dated December 28, 2007. Page 425 of Sudha bears a "Copyright © 2007 by the IETE," which indicates that the document was published in the year 2007. Based

[26] https://www.iete.org/aboutiete.html

on my review of the indicia appearing on the Sudha document and my understanding of the IETE, it is my opinion that Sudha is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by the end of 2007.

23. Public Accessibility of EX1100 ("Oh")

149. EX1100 (which I refer to herein as "Oh") is titled "A fully pipelined single-precision floating-point unit in the synergistic processor element of a CELL processor." As indicated on the face of EX1100, Oh was published by the IEEE in Volume 41, Number 4 of the IEEE Journal of Solid-State Circuits (April 2006). I have reviewed Oh, which appears to be a library scan of a physical copy of the issue of the journal that contained Oh. The cover page of Oh bears an "ISSN" number of 0018-9200 and a stamp from the "Linda Hall Library" dated April 10, 2006. Page 759 of Oh bears a Copyright "© 2006 IEEE," which indicates it was published in 2006. IEEE Xplore indicates that Oh had a "Date of Publication" of March 27, 2006.[27] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of

[27] https://ieeexplore.ieee.org/document/1610620

electrical engineering, computer science, and related disciplines"). Testimony from an IEEE representative demonstrates public accessibility of EX1100 no later than March 31, 2006. *See* EX1133, ¶¶6-11. Based on the above evidence and my review of the indicia appearing on the Oh document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Oh is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by the end of 2006. *See also* Bolin, ¶¶91-95 (concluding that the Oh article "was publicly accessible by July 2005 and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence, on their own or with the assistance of a research librarian").

24. Public Accessibility of EX1103 ("Prabhu")

150. EX1103 (which I refer to herein as "Prabhu") is titled "SOMGPU: An Unsupervised Pattern Classifier on Graphical Processing Unit." As indicated on the face of EX1103, Prabhu was published by the IEEE in connection with the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). Page 1101 of Prabhu bears a Copyright "© 2008 IEEE," which indicates it was published by the IEEE in 2008. IEEE Xplore

indicates that Prabhu had a "Date Added to IEEE *Xplore*" of September 23, 2008.[28] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Prabhu document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Prabhu is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by September 23, 2008.

25. Public Accessibility of EX1104 ("Poli")

151. EX1104 (which I refer to herein as "Poli") is titled "Processing Neocognitron of Face Recognition on High Performance Environment Based on GPU with CUDA Architecture." As indicated on the face of EX1104, Poli was published by the IEEE in connection with the 2008 20th International Symposium on Computer Architecture and High Performance Computing. I have reviewed Poli, which appears to be a library scan of a physical copy of the issue of the journal that contained Poli. Page 2 of EX1104 bears a "Copyright © 2008 by The

[28] https://ieeexplore.ieee.org/abstract/document/4630920

Institute of Electrical and Electronics Engineers, Inc." an "ISBN" number and an "ISSN" number. Page 4 of EX1104 bears a stamp from the "Linda Hall Library." Page 7 of Poli bears a Copyright "© 2008 IEEE," which indicates it was published in 2008. IEEE Xplore indicates that Poli had a "Date Added to IEEE *Xplore*" of November 21, 2008.[29] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Poli document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Poli is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by November 21, 2008.

26. Public Accessibility of EX1105 ("Ligon")

152. EX1105 (which I refer to herein as "Ligon") is titled "A re-evaluation of the practicality of floating-point operations on FPGAs." As indicated on the face of EX1105, Ligon was published by the IEEE in connection with the IEEE Symposium on FPGAs for Custom Computing Machines (Cat. No.98TB100251)

[29] https://ieeexplore.ieee.org/document/4685731

April 17, 1998. I have reviewed Ligon, which appears to be a library scan of a physical copy of the issue of the journal that contained Ligon. Page 3 of EX1105 bears a "Copyright © 2008 by The Institute of Electrical and Electronics Engineers, Inc." an "ISBN" number and an "ISSN" number. Page 3 of EX1105 also bears a stamp from the "University of Wisconsin-Madison." Page 206 of Ligon bears a Copyright "© 1998 IEEE," which indicates it was published in 1998. IEEE Xplore indicates that Ligon had a "Date Added to IEEE *Xplore*" of August 6, 2002.[30] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Ligon document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Ligon is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by August 6, 2002.

[30] https://ieeexplore.ieee.org/document/707898

27. Public Accessibility of EX1106 ("Held")

153. EX1106 (which I refer to herein as "Held") is titled "From a Few Cores to Many: A Tera-scale Computing Research Overview." As shown on the face of EX1106, Held is a "White Paper" from Intel, one of the leading semiconductor manufacturers and designers of microprocessors in the world. POSAs interested in the field of computing exercising reasonable diligence had reason to navigate to Intel's website to review the information and documentation available there, such as white papers and press releases, available from Intel's website regarding Intel's products, research, and development.

154. Page 82 of EX1048 is the Internet Archive's archived capture of Intel's homepage, http://www.intel.com/, as it existed on December 6, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061206101550/http://www.intel.com/. *See* EX1048, ¶¶5-6, p. 81. At the top of Intel's homepage, there is the text "Technology & Research," which navigates to http://www.intel.com/technology/index.htm?iid=homepage+hdr_nav1_technology. Page 80 of EX1048 is the Internet Archive's archived capture of that webpage as it existed on December 6, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061206013632/http://www.intel.com/technology/index.htm?iid=homepage+hdr_nav1_technology. *See* EX1048, p. 79.

155. On the left of that webpage, there is the text "Intel® multi-core," which navigates to http://www.intel.com/multi-core/index.htm. Page 78 of EX1048 is the Internet Archive's archived capture of that webpage as it existed on December 2, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061202183758/http://www.intel.com/multi-core/index.htm. *See* EX1048, p. 77.

156. On the bottom right of that webpage, there is the text "Tera-scale Research," which navigates to http://www.intel.com/technology/techresearch/terascale/index.htm. Page 76 of EX1048 is the Internet Archive's archived capture of that webpage as it existed on November 28, 2006, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20061128200422/http://www.intel.com/technology/techresearch/terascale/index.htm. *See* EX1048, p. 75.

157. On the bottom of that webpage there is the text "A whitepaper detailing our tera-scale computing research vision," which provides a hyperlink to a PDF that the Internet Archived captured on July 3, 2007, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20070703180231/http://download.intel.com/research/platform/terascale/terascale_overview_paper.pdf. That PDF is EX1106. *See* EX1047, ¶¶5-6, pp. 66-78.

158. Based on the above evidence, it is my opinion that EX1106 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least in the year 2007.

28. Public Accessibility of EX1107 ("Reddaway")

159. EX1107 (which I refer to herein as "Reddaway") is titled "Ultra-high performance, low-power, data parallel radar implementations." As indicated on the face of EX1107, Reddaway was published by the IEEE in connection with the IEEE International Radar Conference. I have reviewed Reddaway, which appears to be a library scan of a physical copy of the issue of the journal that contained Reddaway. Page 4 of EX1107 bears a "Copyright © 2005 by The Institute of Electrical and Electronics Engineers, Inc." an "ISBN" number and an "ISSN" number. Page 5 of EX1107 bears a stamp from the "Library of Congress Copyright Office" dated July 29, 2005. Page 822 of Reddaway bears a Copyright "© 2005 IEEE," which indicates it was published in 2005. IEEE Xplore indicates that Reddaway had a "Date Added to IEEE Xplore" of June 6, 2005.[31] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for

[31] https://ieeexplore.ieee.org/document/1435940

researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Reddaway document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Reddaway is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least in the year 2005.

29. Public Accessibility of EX1108 ("Lokhmotov")

160. EX1108 (which I refer to herein as "Lokhmotov") is titled "Revisiting SIMD Programming." As indicated on the face of EX1108, Lokhmotov was published by the publisher Springer in connection with the 20th International Workshop of Languages and Compilers for Parallel Computing (LPCP) in Urbana, IL, USA, October 11-13, 2007. According to Dr. Bolin, the publisher Springer-Verlag "was and is a well-known and respected publisher (founded in 1842) of academic journals and books … [and] provides an online search tool ('SpringerLink')[] for providing researchers with access to millions of academic publications." Bolin, ¶16. According to Dr. Bolin, the book containing the Lokhmotov article has a creation date of August 12, 2008 in the OCLC MARC database. *See* Bolin, ¶119; *id.* ¶¶7-15 (discussing MARC records). As shown in

Attachment 15b to Dr. Bolin's declaration, the Springer Link database indicates a copyright of "© 2008 Springer-Verlag Berlin Heidelberg," consistent with the face of EX1108. Bolin, ¶120. Attachment 15c to Dr. Bolin's declaration includes an electronic copy of the Lokhmotov article, which is identical to EX1108, that Dr. Bolin retrieved from the Springer Link database along with cover matter for the book that contained the Lokhmotov article. Bolin, ¶121. Based on this evidence and the other evidence Dr. Bolin reviewed, Dr. Bolin concluded that the Lokhmotov article was indexed, searchable and publicly accessible by August 12, 2008. Bolin, ¶¶119-123.

161. I have also reviewed Lokhmotov, which appears to be a library scan of a physical copy of the issue of the journal that contained Lokhmotov. Page 2 of EX1107 bears a Copyright "© Springer-Verlag Berlin Heidelberg 2008" and an "ISBN-10" number, "ISBN-13" number, and an "ISSN" number. Page 2 of EX1108 also bears a stamp from the "Cornell University Libraries." Page 6 of Lokhmotov bears a Copyright "© Springer-Verlag Berlin Heidelberg 2008," which indicates it was published in 2008. The ACM Digital Library indicates that the Lokhmotov article was published October 1, 2007.[32] *See* Bolin, ¶17 (identifying

[32] https://dl.acm.org/doi/10.1007/978-3-540-85261-2_3; *see also* https://link.springer.com/chapter/10.1007/978-3-540-85261-2_3#Abs1

the ACM Digital Library as "a well-known and widely used resource for researchers, educators, and professionals in the field of computer science and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Lokhmotov document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Lokhmotov is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by August 12, 2008.

30. Public Accessibility of EX1110 ("Shahbahrami")

162. EX1110 (which I refer to herein as "Shahbahrami") is titled "Tradeoff analysis of integer multiplier circuits implemented in FPGAs." As indicated on the face of EX1110, Shahbahrami was published by the IEEE in connection with the 2008 Tenth IEEE International Symposium on Multimedia. I have reviewed Shahbahrami, which appears to be an electronic copy from the issue of the journal that contained Shahbahrami. Page 4 of EX1110 bears a "Copyright © 2008 by The Institute of Electrical and Electronics Engineers, Inc.," which indicates it was

---

(SpringerLink database to retrieve digital copy of the Lokhmotov article); Bolin, ¶16 (identifying Springer-Verlag as a "well-known and respected published (founded in 1842) of academic journals and books").

published by the IEEE in 2008, and an "ISBN" number. Page 607 of Shahbahrami bears a Copyright "© 2008 IEEE," which indicates it was published in 2008. IEEE Xplore indicates that Shahbahrami had a "Date Added to IEEE Xplore" of January 9, 2009.[33] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Testimony from an IEEE representative demonstrates public accessibility of EX1110 no later than December 17, 2008. *See* EX1134, ¶¶6-11. Based on the above evidence and my review of the indicia appearing on the Shahbahrami document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Shahbahrami is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence by December 17, 2008, or at least by January 9, 2009.

31. Public Accessibility of EX1111 ("Noumsi")

163. EX1111 (which I refer to herein as "Noumsi") is titled "Acceleration of a content-based image-retrieval application on the RDISK cluster." I have reviewed EX1112, which appears to be a scan of the face and table of contents of

[33] https://ieeexplore.ieee.org/document/4741235

the journal of a bound collection of abstracts of papers published in connection with the 20th IEEE International Parallel & Distributed Processing Symposium. Page 3 of EX1112 indicates that this publication included "Abstracts and CD-ROM," of the articles listed therein, which suggests that the CD-ROM that accompanied this document included electronic copies of the articles listed in the table of contents starting on 7 of EX1112. Page 3 of EX1112 bears a stamp from the "Library of Congress Copyright Office," dated June 2, 2006. Page 4 of EX1112 bears a "Copyright © 2006 by The Institute of Electrical and Electronics Engineers, Inc.," an "ISBN" number and an "ISSN" number. Page 13 of EX1112 identifies an article titled "Acceleration of a content-based image-retrieval application on the RDISK cluster," by the authors "A. Noumsi, S. Derrien, and P. Quinton," which is the title and authors listed on the face of the Noumsi article in EX1111. Page 38 of EX1112 includes a copy of the abstract of the Noumsi article that matches the abstract on the face of EX1111. IEEE Xplore indicates that Noumsi had a "Date Added to IEEE Xplore" of June 26, 2006.[34] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my

[34] https://ieeexplore.ieee.org/document/1639346

review of the indicia appearing on the Noumsi document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Noumsi is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least in the year 2006.

32. Public Accessibility of EX1113 ("Pengdong")

164. EX1113 (which I refer to herein as "Pengdong") is titled "Performance Comparison between Color and Spatial Segmentation for Image Retrieval and Its Parallel System Implementation." As indicated on the face of EX1113, Pengdong was published by the IEEE in connection with the 2008 International Symposium on Computer Science and Computational Technology. I have reviewed Pengdong, which appears to be a library scan of a physical copy of the issue of the journal that contained Pengdong. Page 2 of EX1113 bears a "Copyright © 2008 by The Institute of Electrical and Electronics Engineers, Inc." and an "ISBN-13" number. Page 4 of EX1113 bears a stamp from the "Linda Hall Library." Page14 of Pengdong bears a Copyright "© 2008 IEEE," which indicates it was published in 2008. IEEE Xplore indicates that Pengdong had a "Date Added

to IEEE *Xplore*" of December 30, 2008.[35] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Pengdong document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Pengdong is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by December 30, 2008.

33. Public Accessibility of EX1114 ("Bradley")

165. EX1114 (which I refer to herein as "Bradley") is titled "Exploiting Loop-Level Parallelism for SIMD Arrays Using *OpenMP*." Based on the evidence I discuss below, it is my opinion that Bradley is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by November 11, 2008.

166. The face of EX114 bears a copyright "© Springer-Verlag Berlin Heidelberg 2008," which indicates it was published in 2008. According to Dr.

[35] https://ieeexplore.ieee.org/document/4731486

Bolin, the publisher Springer-Verlag "was and is a well-known and respected publisher (founded in 1842) of academic journals and books … [and] provides an online search tool ('SpringerLink')[] for providing researchers with access to millions of academic publications." Bolin, ¶16. According to Dr. Bolin, the book containing the Bradley article has a creation date of November 11, 2008 in the OCLC MARC database, which was published in connection with the Proceedings of the 3rd International Workshop on OpenMP, IWOMP 2007, Beijing China, June 3-7, 2007. *See* Bolin, ¶1078; *id.* ¶¶7-15 (discussing MARC records). EX1114 is the copy of the Bradley article that Dr. Bolin obtained from the Springer Link database. Bolin, ¶108. As shown in Attachment 13b to Dr. Bolin's declaration, the Springer Link database indicates a copyright of "© 2008," consistent with the face of EX1114. Attachment 13c to Dr. Bolin's declaration, which is EX1114, is an electronic copy of the Bradley article that Dr. Bolin retrieved from the Springer Link database. Bolin, ¶109. Based on this evidence, Dr. Bolin concluded that the Bradley article was indexed, searchable and publicly accessible by at least November 11, 2008. Bolin, ¶¶107-112. Based on the above evidence, I too conclude that Schmidt-2000 would have also been publicly accessible to a POSA exercising reasonable diligence at least by November 11, 2008.

34. Public Accessibility of EX1115 ("Haran")

167. EX1115 (which I refer to herein as "Haran") is titled "22 nm technology compatible fully functional 0.1 μm^2 6T-SRAM cell." I have reviewed EX1115, which appears to be a digital copy of an article retrieved from IEEE Xplore (see footer of page 1). IEEE Xplore indicates that Haran was published by the IEEE and "Published in: 2008 IEEE International Electron Devices Meeting," for a conference held on December 15-17, 2008, and had a "Date Added to IEEE Xplore" of February 27, 2009, and had a "Print ISBN" number and "Print ISSN" number.[36] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Testimony from an IEEE representative demonstrates public accessibility of EX1115 no later than December 17, 2008. *See* EX1135, ¶¶6-11. Based on the above evidence and my understanding of the IEEE and IEEE Xplore, it is my opinion that EX1115 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence by December 17, 2008, or at least by February 27, 2009.

[36] https://ieeexplore.ieee.org/document/4796769

35. Public Accessibility of EX1116 ("Li")

168. EX1116 (which I refer to herein as "Li") is titled "Operating System Support for Shared-ISA Asymmetric Multi-core Architectures." The Semantic Scholar database indicates that Li was published in 2008.[37] The ACM Digital Library indicates that Li was cited in a paper published on April 21, 2009 that concerned multicore processors, which indicates that Li was publicly accessible to POSAs in the art exercising reasonable diligence at least by the April 2009.[38] *See* Bolin, ¶17 (identifying ACM Digital Library as a "well-known and widely used resource for researchers, educators, and professionals in the field of computer science and related disciplines"). Based on the above evidence and my understanding of the ACM and ACM Digital Library, it is my opinion that Li is

[37] https://www.semanticscholar.org/paper/Operating-System-Support-for-Shared-ISA-Asymmetric-Li-Brett/c835d52e0c8ac9c861f1a5ca215dbdc7d48aeff9; https://www.semanticscholar.org/about ("Semantic Scholar provides free, AI-driven search and discovery tools, and open resources for the global research community. We index over 200 million academic papers sourced from publisher partnerships, data providers, and web crawls.").

[38] https://dl.acm.org/doi/10.1145/1531793.1531801 (identifying Li as cited Reference 22).

prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by April 21, 2009.

36. Public Accessibility of EX1117 ("Borkar")

169. EX1117 (which I refer to herein as "Borkar") is titled "Thousand Core Chips—A Technology Perspective." According to Dr. Bolin, the ACM Digital Library indicates that the Borkar article was published on June 4, 2007 in connection with a 2007 conference. Bolin, ¶97; §I.C above (discussing the ACM and ACM Digital Library). Attachment 11b to Dr. Bolin's declaration, which is EX1117, is the copy of Borkar that Dr. Bolin retrieved from the ACM Digital Library. Bolin, ¶98. As Dr. Bolin noted, the first page of EX1117 includes a "Copyright 2007 ACM." Based on this evidence, Dr. Bolin concluded that the Borkar article was indexed, searchable and publicly accessible no later than early June 2007, which is consistent with evidence I have reviewed. Bolin, ¶100 (concluding that the Borkar article "was publicly accessible by June 2007 and could have been located by any member of the public interested in the subject matter by exercising reasonable diligence, on their own or with the assistance of a research librarian"). Based on the above evidence, I too conclude that Borkar

would have also been publicly accessible to a POSA exercising reasonable diligence by June 2007. *See also* Bolin, ¶¶96-100.

37. Public Accessibility of EX1118 ("Hannibal")

170. EX1118 (which I refer to herein as "Hannibal") is an article from Arstechnica.com titled "ClearSpeed kicks it up a notch." Based on the evidence I discuss below, it is my opinion that EX1118 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by October 13, 2004.

171. According to the face of EX1118, EX1118 was published on October 11, 2004 on the website arstechnica.com. According to Arstechnica's website, Ars Technica "was founded in 1998" as a "a publication devoted to technology" and "[i]n 2001, Ars launched a digital subscription service" and in 2008 it was acquired by "the parent company of Condé Nast," a well-known and reputable publisher of magazines, such as Wired and The New Yorker.[39] For example, as shown below, a recent search of Google Scholar returned 1,880 articles from 2009 and earlier that included "arstechnica" and 770 articles hosted on sites that were not under the

[39] https://arstechnica.com/about-us/; https://www.condenast.com/brands-and-markets/brands

arstechnica.com domain name, demonstrating that Ars Technica was a well-known website for those interested in technology, such as the technology of computing devices.[40,41] *See also* Bolin, ¶6 (identifying Google Scholar as a reliable repository of scholarly and research data).

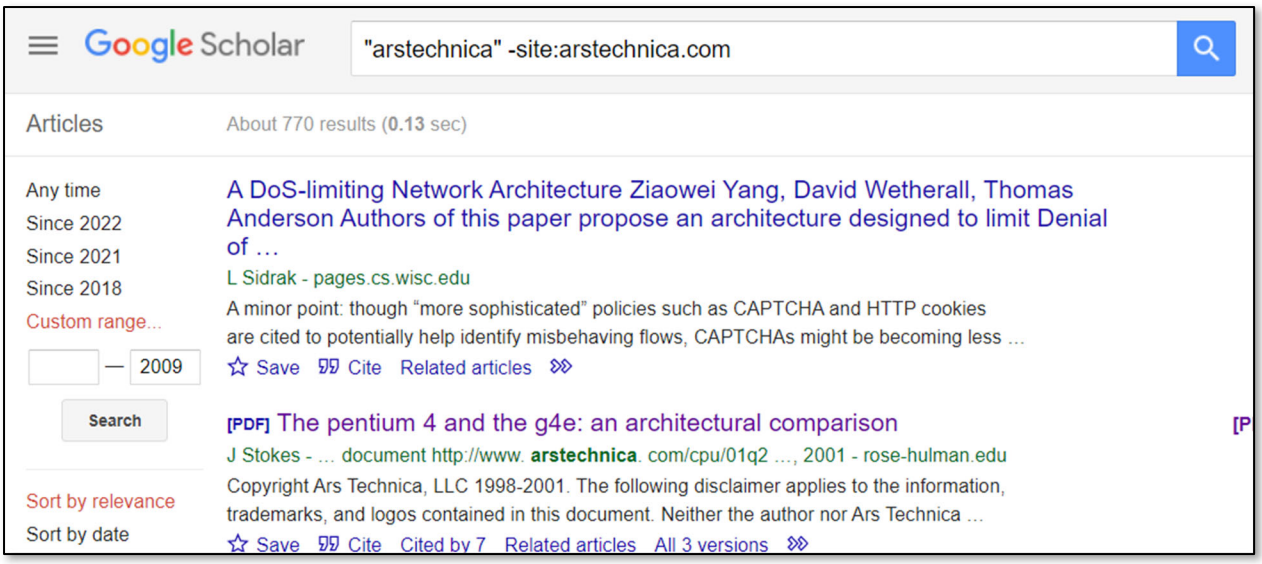


[40] https://scholar.google.com/scholar?q=%22...

arstechnica%22&hl=en&as_sdt=0%2C33&as_ylo=&as_yhi=2009

[41] https://scholar.google.com/scholar?hl=en&as_sdt=0%2C33&as_yhi=2009...

&q=%22arstechnica%22+-site%3Aarstechnica.com&btnG=



172. Page 8 of EX1123 is the Internet Archive's archived capture of Arstechnica.com's homepage, http://arstechnica.com/index.ars, as it existed on October 13, 2004, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20041013014107/http://arstechnica.com/index.ars. *See* EX1123, ¶¶5-6, p. 7. On the left, towards the bottom of Arstechnica's homepage, there is a hyperlink to the "Full Story" for an article titled "ClearSpeed kicks it up a notch," which navigates to the article in EX1118. Page 5-6 of EX1123, which is EX1118, is the Internet Archive's archived capture of that webpage as it existed on October 13, 2004, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20041013015215/http://arstechnica.com/news.ars/post/20041011-4293.html. *See* EX1123, p. 4. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the

same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry, could have located EX1118 by searching the Web for keywords of interest such as "processing elements" and "FPU" and "floating-point," which are terms that appear in EX1118 and would be terms of interest to a POSA interested in the computing industry. Based on the above evidence and my review of the indicia appearing on EX1118, it is my opinion that EX1118 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by October 13, 2004.

38. Public Accessibility of EX1120 ("Manocha")

173. EX1120 (which I refer to herein as "Manocha") is titled "A cost effective 32nm high-K/ metal gate CMOS technology for low power applications with single-metal/gate-first process." As indicated on the face of EX1120, Manocha was published by the IEEE in Volume 38, Issue 8 of the magazine Computer. I have reviewed Manocha, which appears to be a library scan of a physical copy of the issue of the issue of Computer that contained Manocha. Page 1 of EX1120 bears a library stamp date August 8, 2005. Page 2 of EX1120 bears a "Copyright © 2005 by The Institute of Electrical and Electronics Engineers, Inc."

and page 5 bears "ISSN" number. Page 85 of Manocha states that it was "Published by the IEEE Computer Society" and the date "August 2005." IEEE Xplore indicates that Manocha had a "Date of Publication" of August 8, 2005.[42] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Manocha document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Manocha is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by August 8, 2005.

39. Public Accessibility of EX1124 ("Burger")

174. EX1124 (which I refer to herein as "Burger") is titled "Composable Multicore Chips." As indicated on page 2 of EX1124, Burger was published in a book titled "Multicore Processors and Systems" by the publisher Springer. According to Dr. Bolin, the publisher Springer-Verlag "was and is a well-known and respected publisher (founded in 1842) of academic journals and books …

[42] https://ieeexplore.ieee.org/document/1492276

[and] provides an online search tool ('SpringerLink')[] for providing researchers with access to millions of academic publications." Bolin, ¶16. According to Dr. Bolin, the book containing the Burger article has a creation date of June 18, 2009 in the OCLC MARC database. *See* Bolin, ¶82; *id.* ¶¶7-15 (discussing MARC records). As shown in Attachment 8c to Dr. Bolin's declaration, the Springer Link database indicates the book containing the Burger article was "First Online: 01 January 2009" and had a copyright of "© 2009 Springer-Verlag US," consistent with page 3 of EX1124. Bolin, ¶83. Attachment 8d to Dr. Bolin's declaration, which is EX1124 includes an electronic copy of the Burger article that Dr. Bolin retrieved from the Springer Link database along with cover matter for the book that contained the Burger article. Bolin, ¶184. Based on this evidence and the other evidence Dr. Bolin reviewed, Dr. Bolin concluded that the Burger article was indexed, searchable and publicly accessible by January 1, 2009, which is consistent with evidence I have reviewed. Bolin, ¶¶81-85. Based on the above evidence, I too conclude that Burger would have also been publicly accessible to a POSA exercising reasonable diligence by January 1, 2009.

40. Public Accessibility of EX1125 ("Roma")

175. EX1125 (which I refer to herein as "Roma") is titled "In the Development and Evaluation of Specialized Processors for Computing High-Order 2-D Image Moments in Real-Time." As indicated on page 170 of EX1125, Roma

was published by the IEEE and bears a copyright "© 2000 IEEE," which indicates it was published in the year 2000. As shown in attachment 14b to Dr. Bolin's declaration, IEEE Xplore indicates that the Roma article was added to IEEE Xplore on August 6, 2002. Bolin, ¶114. Testimony from an IEEE representative demonstrates public accessibility of EX1125 no later than September 13, 2000. *See* EX1136, ¶¶6-11. Based on this evidence and the other evidence Dr. Bolin reviewed, Dr. Bolin concluded that the Roma article was indexed, searchable and publicly accessible by August 6, 2002, but at least no later than March 1, 2004, which is consistent with evidence I have reviewed. Bolin, ¶¶113-117. Based on the above evidence, I conclude that Roma would have been publicly accessible to a POSA exercising reasonable diligence by 2004.

41. Public Accessibility of EX1126 ("Intel 22nm")

176. As shown on its face, EX1126 is a "News Fact Sheet" from Intel titled "Intel Developer Forum 22nm News Facts" dated September 22, 2009. EX1126 states that "Intel Corporation is holding its Intel Developer Forum in San Francisco from Sept. 22-24" and "Paul Otellini, Intel president and CEO, **today** [i.e., September 22, 2009] displayed the world's first working chips built on 22nm process technology." In my opinion, POSAs interested in the field of computing would have been familiar with the technological developments announced by Intel at its developer forums and would have been the type of people that attended

Intel's developer forums, which were "a biannual gathering of technologists to discuss Intel products and products based on Intel products."[43] POSAs that did not personally attend the Intel developer's forum had reason to navigate to Intel's website to find information about the technologies discussed at the forum, such as through Intel press releases and white papers posted to and/or retrievable from Intel's website.

177. Page 98 of EX1048 is the Internet Archive's archived capture of Intel's homepage, http://www.intel.com/index.htm/, as it existed on October 19, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20091019172712/http://www.intel.com/index.htm/#/en_US_01. *See* EX1048, ¶¶5-6, p. 97. In the middle of Intel's homepage, there is the text "About Intel" which opens a drop-down box that includes the text "Press," which navigates to http://www.intel.com/pressroom/index.htm?iid=gg_about+intel_pressroom. Page 96 of EX1048 is the Internet Archive's archived capture of that webpage as it existed March 6, 2010, as reflected in the Internet Archive's extended URL:

[43] Wikipedia, Intel Developer Forum, https://en.wikipedia.org/wiki/Intel_Developer_Forum (last visited Dec. 20, 2022)

https://web.archive.org/web/20100306035503/http://www.intel.com/pressroom/index.htm?iid=gg_about+intel_pressroom. *See* EX1048, ¶¶5-6, p. 95.

178. At the left of Intel's press room webpage there is the text "Innovation@Intel," which navigates to http://www.intel.com/pressroom/innovation/innovation.htm?iid=pr1_ln_innovation. Pages 90-94 of EX1048 is the Internet Archive's archived capture of that webpage as it existed February 27, 2010, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20100227023739/http://www.intel.com/pressroom/innovation/innovation.htm?iid=pr1_ln_innovation. *See* EX1048, ¶¶5-6, p. 89.

179. On the left of that webpage there is the text "Press Kits," which navigates to http://lz1.intel.com/appzone/pressroom/PressReleases.asp?ReleaseType=VPK. Pages 87-88 of EX1048 is the Internet Archive's archived capture of that webpage as it existed March 9, 2010, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20100309202006/http://lz1.intel.com/appzone/pressroom/PressReleases.asp?ReleaseType=VPK. *See* EX1048, ¶¶5-6, p. 86.

180. Under the heading "Services" of that webpage there is a hyperlink with the text "PRESS KIT - Intel Developer Forum (IDF) San Francisco 2009 which navigates to

http://www.intel.com/pressroom/kits/events/idffall_2009/index.htm. Pages 84-85 of EX1048 is the Internet Archive's archived capture of that webpage as it existed September 23, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090923235718/http://www.intel.com/pressroom/kits/events/idffall_2009/index.htm. *See* EX1048, ¶¶5-6, p. 83.

181. Towards the bottom left of that page under the heading "Press Materials" there is a hyperlink with the text "Intel Developer Forum 22nm News Facts (PDF 70KB)," which navigates to a PDF of a document that the Internet Archive captured on October 7, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20091007000925/http://download.intel.com/pressroom/kits/events/idffall_2009/pdfs/22nm_factsheet.pdf. That PDF, which is on page 80 of EX1047, is EX1126. *See also* EX1047, ¶¶5-6, p. 79.

182. Based on the above evidence, it is my opinion that EX1126 was publicly accessible before the June 15, 2010 filing date for U.S. Appl. No. 12/816,201 (challenged patent, code (63)) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by October 7, 2009.

42. Public Accessibility of EX1137 ("TSMC 40nm Press Release 2")

183. EX1137 is a press release by TSMC titled "TSMC Ramps 40nm Volume Production to Promote Innovation as Foundries Assume a Larger Role for $300 Billion Industry." The face of EX1137, states that it was "Issued by: TSMC" and "Issue on: 2008/11/17," which indicates it was published by TSMC on November 17, 2008. I retrieved EX1137 from the Internet Archive's Wayback Machine at the following extended URL, which indicates EX1137, is a copy of the press release as it existed on March 18, 2009, as reflected in the Internet Archive's extended URL: https://web.archive.org/web/20090318021909/http:/www.tsmc.com/tsmcdotcom/PRListingNewsArchivesAction.do?action=detail&newsid=3222&language=E. *See, e.g.*, EX1029, ¶5. Additionally, Internet search engines (e.g., Google, Yahoo, etc.) would have had access to the same webpages using the same type of crawler software that the Internet Archive used to capture the archived pages (*see* §I.A above), such that a POSA interested in the computing industry, could have located EX1137 by searching the Web for keywords of interest such as "TSMC," "semiconductor foundry," and "40nm process," which are terms that appear in EX1137 and would be terms of interest to a POSA interested in the computing industry.

184. Based on the above evidence and my review of the indicia appearing on the face of EX1137, it is my opinion that EX1137 is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by March 18, 2009.

43. Public Accessibility of EX1034 ("Tong")

185. EX1034 (which I refer to herein as "Tong") is titled "Reducing Power by Optimizing the Necessary Precision/Range of Floating-Point Arithmetic." As indicated on page 273 of EX1034, Tong was published by the IEEE and bears a copyright "© 2000 IEEE," which indicates it was published in the year 2000. Testimony from an IEEE representative demonstrates public accessibility of EX1034 no later than the last date of June 2000. *See* EX1035, ¶¶6-11; *see* EX1035, Exhibit A (showing IEEE Xplore information for this exhibit). Based on the above evidence, I conclude that Tong would have been publicly accessible to a POSA exercising reasonable diligence in 2000.

44. Public Accessibility of EX1036 ("Shirazi")

186. EX1036 (which I refer to herein as "Shirazi") is titled "Quantitative Analysis of Floating Point Arithmetic on FPGA Based Custom Computing Machines." As indicated on page 155 of EX1036, Shirazi was published by the IEEE and bears a copyright "© 1995 IEEE," which indicates it was published in

the year 1995. Testimony from an IEEE representative demonstrates public accessibility of EX1036 no later than April 21, 1995. *See* EX1037, ¶¶6-11; *see* EX1037, Exhibit A (showing IEEE Xplore information for this exhibit). Based on the above evidence, I conclude that Shirazi would have been publicly accessible to a POSA exercising reasonable diligence in 1995.

45. Public Accessibility of EX1092 ("Aty")

187. EX1092 (which I refer to herein as "Aty") is titled "High-Speed, Area-Efficient FPGA-Based Floating-Point Multiplier." As indicated on page 274 of EX1092, Aty was published in connection with the ICM 2003 conference on December 9-11 in Cairo, Egypt, indicating it was publicly available in the year 2003. Testimony from an IEEE representative demonstrates public accessibility of EX1092 no later than December 11, 2003. *See* EX1093, ¶¶6-11; *see* EX1093, Exhibit A (showing IEEE Xplore information for this exhibit). Based on the above evidence, I conclude that Aty would have been publicly accessible to a POSA exercising reasonable diligence in 2003.

46. Public Accessibility of EX1138 ("Thornton")

188. EX1138 (which I refer to herein as "Thornton") is titled "Tradeoff analysis of integer multiplier circuits implemented in FPGAs." As indicated on the face of EX1138, Thornton was published by the IEEE in connection with the 1999 IEEE Pacific Rim Conference on Communications, Computers and Signal

Processing (PACRIM 1999). I have reviewed Thornton, which appears to be a library scan of a physical copy of the issue of the journal that contained Thornton. Page 2 of EX1138 bears a "Copyright © 1999 by The Institute of Electrical and Electronics Engineers, Inc." an "ISBN" number a stamp from the University of Wisconsin - Madison. Page 301 of Thornton bears a Copyright "© 1999 IEEE," which indicates it was published in 1999. IEEE Xplore indicates that Thornton had a "Date Added to IEEE Xplore" of August 6, 2002.[44] *See* Bolin, ¶18 (identifying IEEE Xplore as "a well-known and widely used resource for researchers, educators, and professionals in the fields of electrical engineering, computer science, and related disciplines"). Based on the above evidence and my review of the indicia appearing on the Thornton document and my understanding of the IEEE and IEEE Xplore, it is my opinion that Thornton is prior art to the challenged patent because it was publicly accessible before the earliest possible priority date for the challenged patent (June 19, 2009) by being disseminated or otherwise made available to POSAs exercising reasonable diligence at least by August 6, 2002. *See also* Bolin, ¶¶86-90 (concluding that the Thornton article "was publicly accessible by August 2002 and could have been located by any

[44] https://ieeexplore.ieee.org/document/799536

member of the public interested in the subject matter by exercising reasonable diligence, on their own or with the assistance of a research librarian").