

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

SINGULAR COMPUTING LLC,
Patent Owner.

Case No. IPR2023-00396
Patent No. 11,169,775

**PETITION FOR INTER PARTES REVIEW
UNDER 35 U.S.C. §§ 311-319 AND 37 C.F.R. § 42.1 et seq**

TABLE OF CONTENTS

I.	MANDATORY NOTICES	XVI
A.	Real Party-In-Interest	xvi
B.	Related Matters.....	xvi
1.	United States Patent & Trademark Office	xvi
2.	United States Patent Trial and Appeal Board	xvii
3.	U.S. District Court for the District of Massachusetts	xvii
C.	Counsel and Service Information - § 42.8(b)(3) and (4).....	xvii
II.	GROUND S AND STANDING	1
III.	'775 PATENT.....	1
A.	Specification	2
B.	Challenged Claims.....	3
C.	Person of Ordinary Skill in the Art (“POSA”)	4
D.	Prosecution History	4
E.	Priority Dates	4
IV.	CLAIM INTERPRETATION	5
V.	<u>GROUND 1</u> : CLAIMS 1-2, 5 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF KAN	5
A.	Schmidt-2000 (EX1007)	5
1.	Modification to 8-bit Architecture	7
2.	Implementing the 8-bit KPROC ISA	11
B.	Integrating a MIMD-Type Unit.....	11
1.	Kan (EX1015)	12
2.	Schmidt-Kan Combination.....	14
3.	Implementing the Schmidt-Kan Coprocessor	16
4.	Reasonable Expectations of Success.....	17
C.	Claim 1	20
1.	[1PRE]: “A computing system, comprising:”	20
2.	[1A]: “a host computer;”	20

3. [1B]: “a computing chip comprising:”	20
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;”	21
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;”	23
6. [1E]: “an input-output unit connected to the first” and “second edge processing element[s];”	26
7. [1F]: Each Processing Element of Limitation [1C] Has a “memory local to” It.....	29
8. [1H]: Each Processing Element of Limitation [1C] Comprises an “arithmetic unit”	30
9. [1I]: “a host connection at least partially connecting the input- output unit with the host computer;”	32
10. [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;”	33
11. [1G], [1K]: “a fifth arithmetic unit;” “wherein the fifth arithmetic unit comprises a corresponding multiplier circuit adapted to receive as inputs to the corresponding multiplier circuit two floating point values each of width at least 32 bits;”	36
12. [1L]: Each Multiplier Circuit of [1J] “comprises a [first/second/third/fourth] plurality of transistors and has no other transistors”	36

13. [1M]: “the multiplier circuit corresponding to the fifth arithmetic unit comprises a fifth plurality of transistors; and,”	37
14. [1N]: the “fifth plurality of transistors exceeds in number each of the” first through fourth “plurality of transistors”	37
D. Claim 2: “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.”	38
E. Claim 5: “the width of each [first/second] binary mantissa is no more than 10 bits plus a [first/second] sign bit.”	41
VI. <u>GROUND 2A: CLAIMS 3-4 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF KAN AND SCHMIDT-1999</u>	41
A. Implementing an On-chip Controller	41
B. Implementing an On-chip Cache.....	43
C. Reasonable Expectation of Success	45
D. Claim 3: “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.”	46
E. Claim 4: “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.”	47
VII. <u>GROUND 2B: CLAIMS 7-14 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF KAN, SCHMIDT-1999, AND POSA’S BACKGROUND KNOWLEDGE OF SEMICONDUCTOR PROCESS ADVANCES</u>	48
A. The 5000-PE Limitation Carries No Patentable Weight.....	48
B. Obviousness of Process Scaling the KPROC ISA	49
1. Motivation	49

2. Reasonable Expectation of Success	51
a. Moore’s Law	51
b. Process Node Scaling.....	53
c. Implementing an 8-bit N-KPROC ISA.....	56
C. Claim 7	60
1. [7PRE]: “A computing system, comprising:”	60
2. [7A]: “a host computer;”	61
3. [7B]: “a computing chip comprising:”	61
4. [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;”	61
5. [7D]: “an input-output unit connected to each of the first subset of the plurality of first processing elements;”	63
6. [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;”	63
7. [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;”	64
8. [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;”	65
9. [7H], [7L]: “a plurality of second processing elements; and a plurality of second arithmetic units, wherein each of the plurality of second processing elements has positioned therein	

at least one of the plurality of second arithmetic units;” “wherein the plurality of second arithmetic units 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;”	66
10. [7I]: “a host connection at least partially connecting the input- output unit with the host computer”	67
11. [7J]: “wherein the plurality of first arithmetic units each comprises a first corresponding multiplier circuit adapted to receive” the Same Inputs As Recited in [1J].....	67
12. [7K]: “wherein the first multiplier circuits corresponding to the plurality of first arithmetic units each comprises a first respective plurality of transistors and has no other transistors;”	67
13. [7M]: “wherein the second multiplier circuits corresponding to the plurality of second arithmetic units each comprises a second respective plurality of transistors;”	68
14. [7N]: “wherein each of the second respective pluralities of transistors of the second multiplier circuits corresponding to the plurality of second arithmetic units exceeds in number each of the first respective pluralities of transistors of the first multiplier circuits corresponding to the plurality of first arithmetic units.”	68
D. Claim 8	68
1. [8A]: “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”	68
2. [8B]: “wherein the plurality of first processing elements exceeds in number, by at least 100, the plurality of second processing elements.”	69
E. Claim 9: “the plurality of memory units each comprises a corresponding register”	69
F. Claim 10	69
G. Claim 11: “an aggregate maximum bandwidth of the plurality of processing element connections exceeds a maximum bandwidth of the host connection.”	69

H. Claims 12-14.....	71
VIII. <u>GROUND 3: CLAIMS 16, 23 ARE UNPATENTABLE OVER</u> SCHMIDT-2000 IN VIEW OF SCHIMMLER-1996.....	71
A. Add-On Board Implementation of Schmidt N-KPROC-chips.....	72
B. Claim 16	73
1. [16PRE]: “A circuit board connected to a host computer, the circuit board comprising:”	73
2. [16A]: “a plurality of processing element arrays, each comprising:”	73
3. [16B]-[16H].....	73
C. Claim 23	74
IX. <u>GROUND 4: CLAIMS 17-22 ARE UNPATENTABLE OVER</u> SCHMIDT-2000 IN VIEW OF SCHIMMLER-1996, SCHMIDT-1999, AND KAN	74
A. Claims 17-19.....	75
B. Claims 20-22.....	75
X. PUBLIC ACCESSIBILITY OF CITED PRINTED PUBLICATIONS	76
XI. SECONDARY CONSIDERATIONS	80
XII. NO BASIS EXISTS FOR DISCRETIONARY DENIAL	81
XIII. CONCLUSION.....	82
XIV. APPENDIX: CLAIM LISTING	83

TABLE OF AUTHORITIES

CASES

<i>In re Blaser</i> , 556 F.2d 534 (CCPA 1977).....	5
<i>In re Harza</i> , 274 F.2d 669 (C.C.P.A. 1960).....	
<i>In re Wertheim</i> , 646 F.2d 527 (C.C.P.A. 1981).....	5
<i>Lockwood v. Am. Airlines</i> , 107 F.3d 1565 (Fed. Cir. 1997)	5
<i>Microsoft Corp. v. IPA Techs. Inc.</i> , IPR2019-00810, Paper 44 (Oct. 15, 2020).....	
<i>Purdue Pharma L.P. v. Faulding Inc.</i> , 230 F.3d 1320 (Fed. Cir. 2000).....	5
<i>Sony Corp. v. Creative Tech.</i> , IPR2016-01407, Paper 13 (Dec. 6, 2016)	

STATUTES

35 U.S.C. § 282(b)	5
35 U.S.C. § 314(a)	

REGULATIONS

APPENDIX LISTING OF EXHIBITS

Exhibit	Description
1001	U.S. Patent No. 11,169,775
1002	Prosecution History of U.S. Patent No. 11,169,775
1003	Declaration of Paul Franzon, PhD.
1004	Curriculum Vitae of Paul Franzon, PhD.
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,” <i>Parallel and Distributed Computing Practices</i> 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 <i>Parallel Algorithms and Applications</i> , 19:2-3, 79-95 (2004)
1009	Schmidt et al., “A Parallel Accelerator Architecture for Multimedia Video Compression,” in <i>Euro-Par’99, Lecture Notes in Computer Science</i> 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,” <i>High Performance Computing and Networking, Lecture Notes in Computer Science</i> 797, Springer Verlag, 1994, pp. 487–488 (“Schimmler-1994”)
1011	Schimmler et al., “The Instruction Systolic Array in Image Processing Applications,” <i>Vision Systems: Sensors, Sensor systems, and Components</i> , 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,” <i>Proceedings of the IEEE</i> , Vol. 86, No. 1 (January 1998)

Exhibit	Description
1017	IEEE SSCS Newsletter 36-37 (reprinting Moore, “Progress in Digital Integrated Electronics,” International Electron Devices Meeting, IEEE, 1975, pp. 11-13).
1018	<i>Microprocessor Report</i> , “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,” <i>Parallel Computing</i> 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,” <i>Proceedings of IPPS/SPDP’98</i> (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)
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) (ITRS)
1030	International Technology Roadmap for Semiconductor, 2001 Edition (“ITRS-2001”)
1031	Declaration of Nathaniel E Frank-White (November 30, 2022) (Schmidt)

Exhibit	Description
1032	Declaration of Nathaniel E Frank-White (November 30, 2022) (Yeh)
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 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)

Exhibit	Description
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 0.171μm ² SRAM cell size in a 291Mb array,” 2008 IEEE International Electron Devices Meeting, 2008
1046	RESERVED – NOT USED
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
1056	RESERVED – NOT USED
1057	RESERVED – NOT USED
1058	U.S. Patent Appl. Publ. No. 2003/0033509A1
1059	U.S. Patent Appl. Publ. No. 2008/0288823A1
1060	RESERVED – NOT USED
1061	RESERVED – NOT USED
1062	RESERVED – NOT USED
1063	U.S. Patent Appl. Publ. No. 2009/0148068A1
1064	RESERVED – NOT USED
1065	RESERVED – NOT USED
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

Exhibit	Description
1070	U.S. Patent Appl. Publ. No. 2005/0237327A1
1071	RESERVED – NOT USED
1072	RESERVED – NOT USED
1073	RESERVED – NOT USED
1074	RESERVED – NOT USED
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”)
1079	RESERVED – NOT USED
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
1085	RESERVED – NOT USED
1086	RESERVED – NOT USED
1087	RESERVED – NOT USED
1088	VSI Alliance™: Virtual Component Interface Standard Version 2 (OCB 2 2.0) (April 2001)
1089	“Port,” in Microsoft Computer Dictionary, p. 412 (5th. ed. 2002)
1090	RESERVED – NOT USED
1091	RESERVED – NOT USED
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 μm CMOS platform with Cu/low-k interconnects for system on chip applications,” 2001 Symposium on

Exhibit	Description
	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 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”)
1099	RESERVED – NOT USED
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,
1101	RESERVED – NOT USED
1102	RESERVED – NOT USED
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

Exhibit	Description
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)
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 μm^2 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 <i>Computer</i> , vol. 38, no. 8, pp. 85-88, Aug. 2005
1121	RESERVED – NOT USED
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 <i>Multicore Processors and Systems. Integrated Circuits and Systems</i> (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 <i>Proceedings Fifth IEEE International Workshop on Computer Architectures for Machine Perception</i> , 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

Exhibit	Description
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
1132	RESERVED – NOT USED
1133	Declaration of Gordon MacPherson regarding EX1100
1134	Declaration of Gordon MacPherson regarding EX1110
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
1139	Amended Complaint for Patent Infringement, <i>Singular Computing LLC v. Google LLC</i> , 1:21-cv-12110 (D. Mass. Mar. 10, 2022)

I. MANDATORY NOTICES

A. Real Party-In-Interest

Petitioner is the Real Party-in-Interest.

B. Related Matters

1. United States Patent & Trademark Office

The application from which U.S. Patent No. 11,169,775 (“the ’775 patent”) issued is a Continuation of U.S. Patent Application No. 16/571,871 (U.S. Patent No. 10,664,236), which is a Continuation of U.S. Patent Application No. 16/175,131 (U.S. Patent No. 10,416,961), which is a Continuation of U.S. Patent Application No. 15/784,359 (U.S. Patent No. 10,120,648), which is a Continuation of U.S. Patent Application No. 14/976,852 (U.S. Patent No. 9,792,088), which is a Continuation of U.S. Patent Application No. 13/849,606 (U.S. Patent No. 9,218,156), which is a Continuation of U.S. Patent Application No. 13/399,884 (U.S. Patent No. 8,407,273), which is a Continuation of U.S. Patent Application No. 12/816,201 (U.S. Patent No. 8,150,902), which claims priority to U.S. Provisional Application No. 61/218,691.

The following U.S. patent applications also claim the benefit of the priority of the filing date of U.S. Provisional Application No. 61/218,691:

16/675,693; 16/882,694; 17/029,780; 17/367,051; 17/367,071; 17/367,097.

2. United States Patent Trial and Appeal Board

Concurrently with the present Petition, Petitioner is filing a second petition for *inter partes* review of the '775 patent. Petitioner is also filing petitions for *inter partes* review of related U.S. Patent No. 10,754,616.

3. U.S. District Court for the District of Massachusetts

Patent Owner asserted the '775 patent in *Singular Computing LLC v. Google LLC*, Case No. 1:21-cv-12110 (DMA). Patent Owner has represented to Petitioner that it will soon voluntarily dismiss that case without prejudice.

C. Counsel and Service Information - § 42.8(b)(3) and (4)

Lead Counsel	Elisabeth H. Hunt, Reg. No. 67,336
Backup Counsel	Marc S. Johannes, Reg. No. 64,978 Gregory S. Nieberg, Reg. No. 57,063
Service Information	<u>E-mail</u> : EHunt-PTAB@wolfgreenfield.com MJohannes-PTAB@wolfgreenfield.com GNieberg-PTAB@wolfgreenfield.com <u>Post and hand delivery</u> : Wolf, Greenfield & Sacks, P.C. 600 Atlantic Avenue Boston, MA 02210-2206 <u>Telephone</u> : 617-646-8000 <u>Facsimile</u> : 617-646-8646

A power of attorney is submitted with the Petition. Counsel for Petitioner consents to service of all documents via electronic mail.

II. GROUNDS AND STANDING

Petitioner requests *inter partes* review and cancellation of claims 1-5, 7-14, and 16-23 (the “challenged claims”) of the ’775 patent (EX1001).

The challenged claims would have been obvious under §103 per the following grounds. Each reference below is prior art under pre-AIA §102(b).

Ground Number and Reference(s)		Claims
1	Schmidt-2000 (EX1007), Kan (EX1015)	1-2, 5
2	Schmidt-2000, Kan, Schmidt-1999 (EX1009)	3-4, 7-14
3	Schmidt-2000, Schimmler-1996 (EX1011)	16, 23
4	Schmidt-2000, Schimmler-1996, Schmidt-1999, Kan	17-22

Petitioner certifies that the ’775 patent is available for IPR and Petitioner is not barred or estopped from requesting IPR of the challenged claims. 37 C.F.R. §42.104(a).

III. ’775 PATENT

The ’775 patent descends from a long line of continuation applications sharing a specification that states, “embodiments of the present invention are directed to computer processors or other devices which use low precision high dynamic range (LPHDR) processing elements to perform computations (such as arithmetic operations).” EX1001, 5:58-62. Patent Owner (“Singular”) sued Petitioner (“Google”) on three of those continuations in 2019; Google filed IPRs; and the Board invalidated those patents’ claims to a device comprising at least one

LPHDR execution unit. IPR2021-00155; IPR2021-00165; IPR2021-00179. After Singular had already sued Google, Singular filed the '775 patent's application, cobbling together disparate, cherry-picked claim limitations to support a second concurrent lawsuit against Google. But the '775 patent's claims still recite no more than obvious variations of what had long been known in the art.

A. Specification

The '775 patent's specification has one isolated statement that “[*if*] [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.”

EX1001, 5:63-6:3.¹ Having thus criticized this digital floating-point option, the specification then leaves it behind and spends the rest of its 30 columns describing logarithmic and analog embodiments. EX1003 (“Franzon”), ¶¶41-42.

Yet, because Google (and much of the world) uses floating-point arithmetic, Singular's litigation-inspired '775 patent claims now recite multiplier circuits using floating-point values of particular bitwidths. 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

¹ All emphases herein are added unless otherwise indicated.

(col. 28). None of these cherry-picked add-ons make Singular’s purported invention remotely patentable.

B. Challenged Claims

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. Franzon, ¶¶37-40, 49. 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*.”

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 decades old. Franzon, ¶¶50-51. And optimizing the bitwidth of floating-point multiplier circuits to reduce size and power consumption was well-known before Singular’s claimed priority date. Franzon, ¶¶52, 67.

Claims 7 and 16 recite that there are “*no less than 5000*” PEs in the array. Claim 16 further recites that a circuit board, connected to a host computer, comprises a plurality of such PE arrays. Each of the challenged claims would have been obvious over the prior art, as the grounds herein demonstrate.

C. Person of Ordinary Skill in the Art (“POSA”)

A POSA would have had a bachelor’s degree in Electrical Engineering, Computer Engineering, Applied Mathematics, or the equivalent, and two years of academic or industry experience in computer architecture. More education could substitute for experience, and vice versa. Franzon, ¶¶35-36.

D. Prosecution History

The examiner allowed the challenged claims, and the claims in all other applications in the priority chain, without substantively discussing any prior art. EX1002, 14.

E. Priority Dates

The ’775 patent claims priority to a 2009 provisional application that lacks written description support at least for claims 7-24 which recite “*no less than 5000*” PEs. Franzon, ¶53.

The 2009 provisional application 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.” EX1055, 8. Franzon, ¶¶45-46.

That does not remotely “constitute[] a full, clear, concise and exact description” of an array of PEs numbering “*no less than 5000*,” which would be required to satisfy the written description requirement under pre-AIA 35 U.S.C.

§112, ¶1. *In re Wertheim*, 646 F.2d 527, 538–39 (C.C.P.A. 1981); *Purdue Pharma L.P. v. Faulding Inc.*, 230 F.3d 1320, 1323 (Fed. Cir. 2000). Franzon, ¶46.

Although it would have been **obvious** to integrate as many PEs as semiconductor processing permitted (as discussed *infra* §VII.B), obviousness does not satisfy the written description requirement. *Lockwood v. Am. Airlines*, 107 F.3d 1565, 1571-72 (Fed. Cir. 1997); *In re Blaser*, 556 F.2d 534, 538 (C.C.P.A. 1977). Franzon, ¶47.

Thus, claims 7-24 are not entitled to the provisional application's 2009 filing date. The next-earliest application to which the '775 patent claims priority was filed June 15, 2010. Franzon, ¶48.

IV. CLAIM INTERPRETATION

Claim terms are construed herein using the standard used in civil actions under 35 U.S.C. §282(b), and have been given their ordinary and customary meaning as understood by a POSA in accordance with the specification and prosecution history. 37 C.F.R. §42.100(b).

V. GROUND 1: CLAIMS 1-2, 5 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF KAN

A. Schmidt-2000 (EX1007)

Schmidt-2000 is §102(b) prior art because it was publicly accessible in 2000. *See infra* §X.

Schmidt-2000 discloses “a parallel computer with 1024 floating-point processors on a single chip,” called the “KPROC-chip” (for KiloPROCessor). Schmidt-2000, 163. A “host computer” uses the KPROC-chip as a “coprocessor” to speed up computationally intensive tasks. Schmidt-2000, 163, 168-175. The KPROC-chip’s 1024 processors are arranged in a 32x32 “[i]nstruction systolic array” (ISA) in which each “identical processor[]” is “connected to its four direct neighbors.” Schmidt-2000, 163-165, 178. Fig. 2.1 (annotated below left) illustrates part of the ISA. Franzon, ¶56.

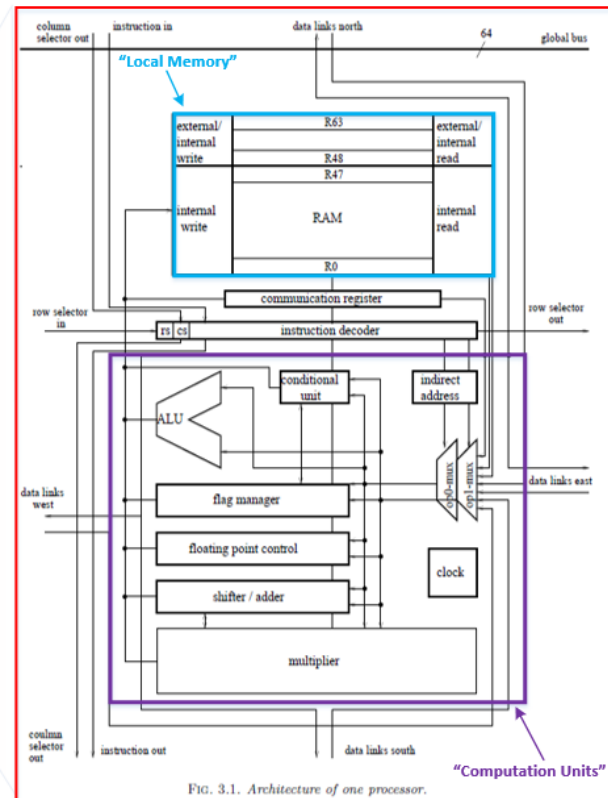
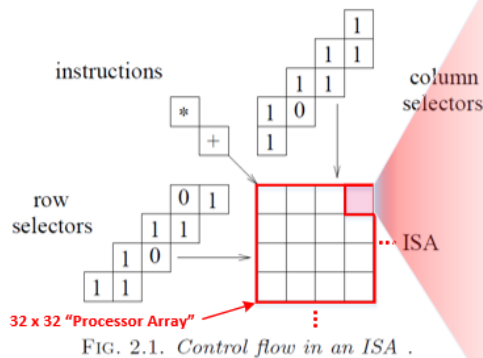


Fig. 3.1 (annotated above right) shows that each processor has 64 general-purpose registers of “local memory” (blue) implemented as random access

memory (RAM) that are independently addressable for read/write operations.

Schmidt-2000, 163-166; Franzon, ¶57. Each processor further comprises “computation units” (in purple box above) for performing computations including floating-point multiplications. Schmidt-2000, 166-167; Franzon, ¶58.

1. Modification to 8-bit Architecture

Schmidt-2000’s KPROC-chip used a 16-bit architecture in which each processor has 16-bit registers and computing units that operate on 16-bit operands one bit per clock cycle. Schmidt-2000, 163-164. POSAs would have had motivation and reasonable expectation of success in modifying this to an 8-bit architecture to provide a high-performance coprocessor optimized for image-processing applications. Franzon, ¶¶59-60.

Each of Schmidt-2000’s processors have 16-bit computing units, including a 16-bit integer multiplier and a 16-bit integer adder. Schmidt-2000, 166-167; Franzon, ¶60. Operands of more than 16 bits are partitioned into multiple 16-bit operands such that computations are performed iteratively, 16 bits at a time. Schmidt-2000, 166; Franzon, ¶60. Schmidt-2000’s floating-point unit uses the 16-bit integer multiplier to multiply mantissas and the 16-bit integer adder to add exponents such that higher-bitwidth multiplication (i.e., more than 16-bits) are likewise performed iteratively, 16 bits at time. In this way, Schmidt-2000’s processors can perform floating-point multiplication of higher-precision numbers

(e.g., 32-bit, 64-bit, etc.) by partitioning them into 16-bit operands and performing a sequence of 16-bit sub-operations. Schmidt-2000, 166-167; Franzon, ¶60.

Schmidt-2000 explains that the 16-bit architecture was chosen as a “compromise between the requirements of different application areas,” recognizing that the optimal bitwidth choice would be different for systems intended for different applications—e.g., an 8-bit architecture would be preferable for image processing; a 64-bit architecture for numeric applications. Schmidt-2000, 164-165 (“The wordlength...was chosen to 16 bits as a compromise between the requirements of different application areas like image processing (favourable 8 bits) and numeric (64 bits).”); Franzon, ¶61.

POSAs would have understood that because longer-bitwidth operands can be multiplied via iteration, the selection of the processor architecture’s native bitwidth—including the bitwidth of the processor’s registers and computing units that determines the bitwidth of the inputs received by the multiplier after longer-bitwidth operands are partitioned—was merely a design choice to optimize the trade-off between computation speed and the chip area and power needed for each processor, for whichever applications the chip is designed to primarily perform. Franzon, ¶62.

POSAs would have understood that Schmidt-2000’s 16-bit architecture is larger (e.g., approximately 1.7 times as large in chip area) than an 8-bit architecture

and would consume significantly more power than an 8-bit architecture. Franzon, ¶¶63-64, ¶¶190-195. Therefore, POSAs would have recognized the desirability of implementing Schmidt-2000's KPROC-chip with 8-bit processors to achieve a KPROC-chip optimized to primarily perform image processing applications in which a substantial percentage of the operands are contemplated to have fewer than 16 bits and for which Schmidt-2000 teaches 8 bits is "favourable"— and which rely on the types of computations for which Schmidt-2000 says its ISA is well-suited (matrix multiplication, convolution, FFT, motion estimation, histogram and filtered back-projection for tomographic image reconstruction, etc.)—with the capability of performing higher-bitwidth operations via iterative computations on 8-bit operands. Schmidt-2000, 164-165, 169-176. Franzon, ¶65.

In optimizing the KPROC-chip for such applications, POSAs would have understood that the speed-up of the disclosed 16-bit architecture for multiplying higher-bitwidth operands may not merit its significantly larger area and power consumption relative to an 8-bit architecture for implementations intended to primarily perform applications where operations are skewed towards 8-bit operands (like image processing and many digital signal processing and multimedia applications) but that may less frequently need higher-bitwidth computations as well. Franzon, ¶66.

Reduced-bitwidth floating-point formats' suitability for a broad range of applications was known, as were benefits of reducing power and conserving chip area resulting from their use. EX1034, 278-280; EX1125, 172-174; Franzon, ¶67. Thus, an 8-bit architecture would also reduce power consumption for many of the floating-point operations in image processing while retaining the higher-bitwidth floating-point capabilities where needed. Franzon, ¶67.

That POSAs would have understood Schmidt-2000's teachings, coupled with the POSA's background knowledge, to suggest implementing the KPROC-chip with an 8-bit architecture for an image-processing-optimized implementation is further corroborated by prior-art disclosure that Schmidt-2000's authors chose an 8-bit architecture when implementing a video accelerator using a similar ISA as described in Schmidt-2000, corroborating the POSA's knowledge that the choice of architecture bitwidth was a matter of design choice for a given type of application. EX1009 ("Schmidt-1999"), 953-956; Franzon, ¶68.

POSAs would thus have been motivated to implement an 8-bit KROC-chip to reduce the size of each processor (allowing for even more processors on the chip), reduce power consumption of both integer and floating-point operations, and increase performance for image processing applications, while retaining the ability to multiply higher-bitwidth operands via iteration when needed. Franzon, ¶¶69-70.

2. Implementing the 8-bit KPROC ISA

POSAs would have known how to implement an 8-bit KPROC ISA with a reasonable expectation of success. Franzon, ¶71. For example, 8-bit general-purpose and communication registers could be used in place of Schmidt-2000's disclosed 16-bit registers, and the computing units' native hardware reduced to 8-bit widths, using only ordinary skill. Franzon, ¶71.

Because Schmidt-2000's processors perform operations using a "parallel-serial" design (Schmidt-2000, 166), POSAs would have understood the bit-serial connections between registers and computing units would require little or no change, and the control circuitry would be modified in known ways to execute 8-clock-cycle instructions. Schmidt-2000, 164-166; Franzon, ¶72. As discussed above, Schmidt-2000's authors demonstrated a reasonable expectation of success in an 8-bit architecture for a video accelerator by 1999. Schmidt-1999, 953-956; Franzon, ¶72.

B. Integrating a MIMD-Type Unit

Schmidt-2000 teaches that an instruction systolic array, like the KPROC ISA, provides "high performance hardware for *specific* computation intensive applications" and is typically used by a sequential host as a coprocessor that "solves *only* the computationally intensive tasks within a global application." Schmidt-2000, 163. Indeed, single-instruction multiple-data (SIMD) architectures

(of which ISAs are one type) were known to provide high performance for a limited class of computations—large-scale repetition of the same simple computations on different data. Kan, 2:14-17; EX1019, 24; EX1020, 226; Franzon, ¶73.

However, POSAs also knew that many application areas such as image processing and multimedia applications also include other levels of parallelism or types of processing, such as task-level parallelism, irregular computations or process flow, or complex data dependency, for which SIMD-type architectures are not well suited. Kan, 1:50-2:49; Franzon, ¶74. Thus, for many image processing applications, acceleration is limited to portions of the global application when relying solely on SIMD-type architectures to increase performance. Franzon, ¶74.

1. Kan (EX1015)

Kan recognized that SIMD and MIMD (multiple-instruction multiple-data) architectures had respective advantages and disadvantages that made one well-suited to some application areas and not others (and vice versa). Kan, 1:50-68. Moreover, SIMD and MIMD architectures had respective advantages/disadvantages for different types of processing within a single application field. Kan, 2:7-49; Franzon, ¶75.

Kan discloses a parallel processing system combining both SIMD-type and MIMD-type units to take advantage of each architecture, while each compensates

for the other's drawbacks, thereby providing an improved parallel-processing system optimized for different types of processing within a given application field, such as image processing. Kan, 2:45-56, FIG. 1 (annotated below); Franzon, ¶76.

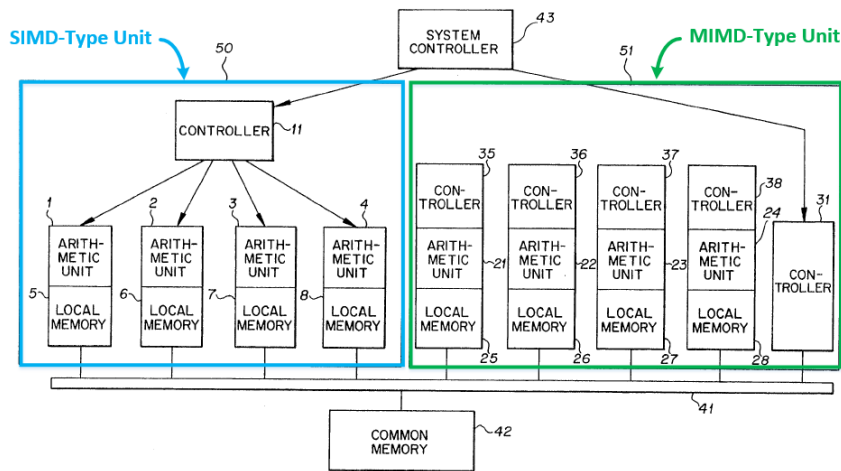


FIG. 1

“SIMD type” parallel data processing unit 50 and “MIMD type” parallel data processing unit 51 are both coupled to a system controller 43 that allocates processing of programs to each parallel-processing unit depending on which unit is more suitable for that type of processing. Kan, 4:10-32; Franzon, ¶77. System controller 43 can utilize SIMD-type unit 50 and MIMD-type unit 51 to execute separate programs operating on different data, execute portions of a program that operates on the same data (including data computed by the other unit) and/or execute the same program independently and compare the results. Kan, 4:24-5:4, 7:1-8:18; Franzon, ¶77.

2. Schmidt-Kan Combination

Given the known limitations of accelerating processing using a single type of parallel-processing architecture, POSAs would have been motivated to integrate aspects of Kan's hybrid parallel data processing system on the same chip as the 8-bit KPROC ISA to provide a more flexible single-chip coprocessor capable of accelerating a wider array of computations and more types of processing within a given application area, such as image processing for which the 8-bit KPROC ISA is optimized (*supra* §V.A.1). Kan, 1:24-3:55, 9:31-10:16; Franzon, ¶78.

In particular, POSAs would have been motivated to integrate a MIMD-type unit like Kan's on the 8-bit KPROC-chip to provide a hybrid single-chip coprocessor ("Schmidt-Kan coprocessor") that benefits from acceleration of not only large-scale parallelism via the KPROC array, but also processing tasks for which the KPROC array is not well suited but for which Kan's MIMD-type unit is optimized, such as course-grain parallelism, task parallelism (e.g., repetition of routines with substantial sequential or irregular computations), or complex data processing (e.g., complex data flow, frequent branch conditions or data dependency), etc., that are present in image processing applications. Kan, 1:50-2:55, 4:22-32; Franzon, ¶¶79-80 (also citing Schmidt-2000, 163, discussing goal of increasing chip's processing power).

Schmidt-2000's KPROC was the "first implementation of a parallel computer with 1024 floating-point processors on a single chip" using a 250-nm process and that the KPROC processors were specifically designed with the "the goal of integration of 1024 processors on a single chip." Schmidt-2000, 163-164, 166. Thus, Schmidt-2000 recognized the desirability of integrating as much processing power on a single chip as the available semiconductor processes would allow. Franzon, ¶80.

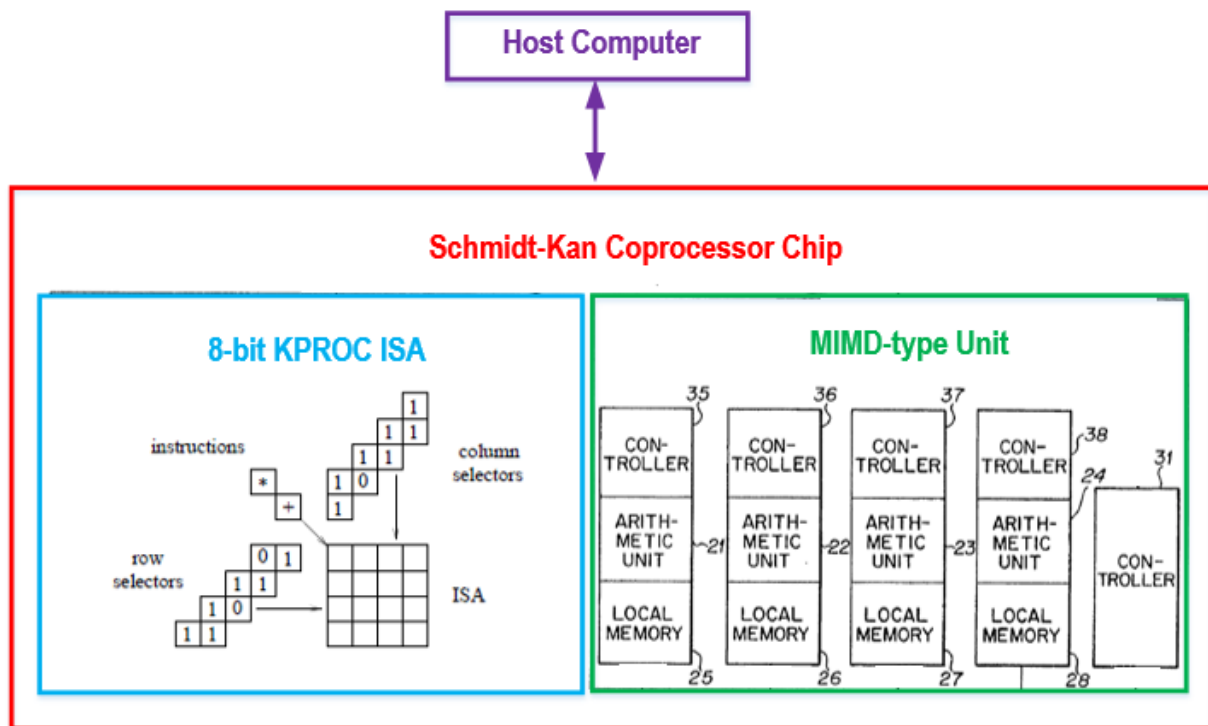
Indeed, Schmidt-2000 cites Schimmler-1996 (EX1011), authored by Schmidt-2000's co-author Schimmler, where a 1024-processor array was obtained by assembling sixteen 64-processor chips on a circuit board, corroborating the desirability of increased levels of single-chip integration as advances in semiconductor technology permitted. Schimmler-1996, 136-139; *see also* EX1010 ("Schimmler-1994"), 487 (also describing an add-on board implementation where "[t]here are 16 ISA chips on the board, each comprising 64 processors."); Franzon, ¶81 (citing same).

By extension, POSA would have likewise been motivated to implement the Schmidt-Kan coprocessor as a single-chip solution to take advantage of the known benefits of integrating desired functionality on a single-chip. Franzon, ¶82 (describing increasing single-chip functionality as driving the semiconductor

industry to continually develop smaller semiconductor components to increase single-chip functionality).

3. Implementing the Schmidt-Kan Coprocessor

An obvious implementation of a Schmidt-Kan coprocessor integrates a “MIMD type” parallel data processing system like Kan’s (annotated **green** in Kan’s FIG. 1 above) on the 8-bit KPROC-chip discussed *supra* §§V.A.1-V.A.2, in association with the 8-bit KPROC ISA as shown in the figure below (using Kan’s **MIMD-type unit** from Kan’s FIG. 1). Franzon, ¶83.



Kan’s system controller 43—which comprises “a data processor 90, a memory 91,” etc., that processes “user programs...stored in the memory 91,” transfers data “to the common memory 42,” and selects “either one of the parallel

data processing units 50 and 51” to execute user programs—operates as a host computer. Kan, 6:13-23; Franzon, ¶84; *see also* EX1001, 8:29-30, FIG. 1 (“The Host 102 is responsible for overall control of the computing system 100.”).

Thus, POSAs understood that Schmidt-2000’s “host” configured to execute programs that utilize the KPROC ISA to accelerate specialized tasks would likewise be configured to control execution of user programs on the Schmidt-Kan coprocessor in the manner of Kan’s system controller. Franzon, ¶85. Kan’s SIMD and MIMD units connect to shared bus 41 to access common memory 42, and POSAs likewise had motivation to connect the Schmidt-Kan coprocessor’s MIMD unit to Schmidt-2000’s “global bus” (Schmidt-2000, 165, Fig. 3.1) to access the host’s external memory, like the KPROC ISA performs I/O as discussed in §V.C.6 *infra* in connection with the claimed “input-output unit” in [1E]. Kan, 5:39-6:12; Franzon, ¶85.

4. Reasonable Expectations of Success

POSAs would have reasonably expected success implementing the above-described Schmidt-Kan coprocessor even on a single chip of the same size as Schmidt-2000’s KPROC-chip because it merely involves the ordinary skill of using the available semiconductor process technology in 2009. Franzon, ¶86.

Schmidt-2000 disclosed 1024 floating-point processors on a 1.2-cm² (120-mm²) chip using a 250-nm process nearly a decade earlier. Schmidt-2000, 163-

164; Franzon, ¶87. As discussed *infra* §§VII.B.2.b, semiconductor processing technology advanced through numerous successively smaller process nodes (180-nm, 130-nm, 90-nm, 65-nm, 45-nm, 40-nm, 32-nm nodes) that POSAs knew were available in 2009. These process nodes defined metrics that the semiconductor industry designed in developing next-generation products. Franzon, ¶87.

As demonstrated *infra* §§VII.B.2.b based on transistor densities that were achieved by commercial products at process nodes available in 2009, consistent with Moore's Law, POSAs understood that more than a factor-of-50 increase in transistor density was achieved between the 250-nm node used to implement Schmidt-2000's KPROC-chip and process nodes available in 2009. Franzon, ¶88.

As demonstrated *infra* §§VII.B.2.c, POSAs would have reasonably expected success in integrating a MIMD-type unit like Kan's on the same chip as the 8-bit KPROC ISA using ordinary skill in process technology available in semiconductor foundries by 2009, even if that ISA were scaled up to include well over 5000 8-bit Schmidt-2000 processors. Thus, POSAs would also have reasonably expected success in implementing the above-described Schmidt-Kan coprocessor (which includes a single 1024-processor KPROC ISA) on the same size chip as Schmidt-2000's KPROC-chip. Franzon, ¶89.

For example, based on the transistor-density scale factors for the 65-nm, 45-nm and 40-nm nodes discussed *infra* §VII.B.2.b and on the areas of Kan's MIMD-

type components discussed *infra* §VII.B.2.c, the area occupied by the KPROC ISA at 65-nm, 45-nm and 40-nm nodes (as well as how each area compares as a percentage of the 102.4-mm² area occupied by Schmidt-2000's KPROC ISA at 250-nm (Schmidt-2000, 164)), and the area of the above-described Schmidt-Kan coprocessor (with integrated MIMD-type unit) at each of those nodes, are shown in the table below. Franzon, ¶90.

Process Node	Scale Factor	<u>KPROC ISA</u> (% <u>KPROC ISA</u>)	Schmidt-Kan Coprocessor
250-nm	1	102.4 mm ²	
65-nm	24	4.27 mm ² (4%)	153.8 mm ²
45-nm	31	3.3 mm ² (3%)	118 mm ²
40-nm	52	1.9 mm ² (2%)	70.9 mm ²

As the table above shows, the Schmidt-Kan coprocessor would have been implementable as a single-chip solution at the 40-nm node without increasing the size of the KPROC-chip. Franzon, ¶91.

Additionally, Schmidt-2000 disclosed that the “chip area should not exceed 1.2 cm²,” but proposed adding an on-chip cache to the KPROC-chip which would have resulted in a larger chip design, demonstrating that the 1.2 cm² (120 mm²) chip was a design choice and not a limitation. Schmidt-2000, 164, 176; Franzon, ¶92. POSAs knew that significantly larger chips were routinely used before June 2009 (*see, e.g.*, die sizes used by example commercial products discussed *infra*

§VII.B.2.b), and would easily accommodate the above-described Schmidt-Kan coprocessor implementations. Franzon, ¶92.

C. Claim 1

The obvious Schmidt-Kan coprocessor implementation discussed *supra* §V.B.3 meets all limitations of claim 1.

1. [1PRE]: “A computing system, comprising:”

Schmidt-2000’s “parallel computer with 1024 floating-point processors on a single chip”—the “KPROC-chip”—is utilized by a “host computer” as a “coprocessor.” Schmidt-2000, 163. The host computer together with the Schmidt-Kan coprocessor implementation of this chip described *supra* §§V.B.2-V.B.3 meet [1PRE]’s “*computing system*.” Franzon, ¶¶93-94.

2. [1A]: “a host computer;”

Schmidt-2000’s “host computer” meets [1A]. Schmidt-2000, 163; Franzon, ¶95.

3. [1B]: “a computing chip comprising:”

Schmidt-2000’s “KPROC is an ISA with 1024 processors *integrated on a single chip*” to implement “a parallel *computer*.” Schmidt-2000, 163. The Schmidt-Kan coprocessor described *supra* §§V.B.2-V.B.3 that integrates a MIMD-type unit on the same chip meets [1B]’s “*computing chip*.” Franzon, ¶96.

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;”

The '775 patent “call[s] each unit[] which pairs memory with arithmetic[] a Processing Element.” EX1001, 16:65-67, 8:5-11 (“processing elements’...should be understood more generally as any kind of execution unit”).

Each Schmidt-2000 processor executes instructions using its computing units (including adder, multiplier, and floating-point unit) that operate on operands stored in addressable general-purpose registers and a communication register—i.e., each Schmidt-2000 processor is an execution unit that pairs memory with arithmetic, and thus is a claimed “*processing element*.” Schmidt-2000, 164-167; Franzon, ¶¶97-98.

In Schmidt-2000’s “processor array” (Schmidt-2000, 164), each processor is connected to adjacent processors in a “mesh connected processor grid,” meeting the claimed “*processing element array*.” Schmidt-2000, 163-164, 176 (Table 6.1 caption); Franzon, ¶99. A portion of this “KPROC array” (Schmidt-2000, 169) is illustrated in Fig. 2.1, reproduced below with **red** outline indicating processors positioned at an edge of the array and **blue** outline indicating processors positioned in the array’s interior (i.e., not on an edge).

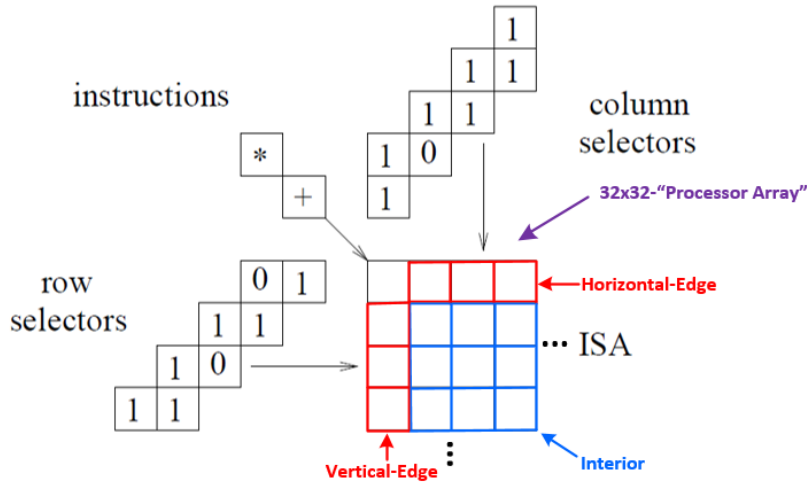


FIG. 2.1. Control flow in an ISA .

Any two of the red-outlined processors in the top row or left-most column (or bottom row or right-most column not shown in Fig. 2.1) are first and second “edge processing element[s] positioned at a first edge of the processing element array,” and any two processors in the blue-outlined interior that are adjacent to the two edge processors are first and second “interior processing elements.” Franzon, ¶100.

Example first and second edge PEs are shaded **red** below, and example first and second interior PEs are shaded **blue**, showing that they are “positioned at [first and second] location[s] in the interior of the processing element array.” Franzon, ¶100.

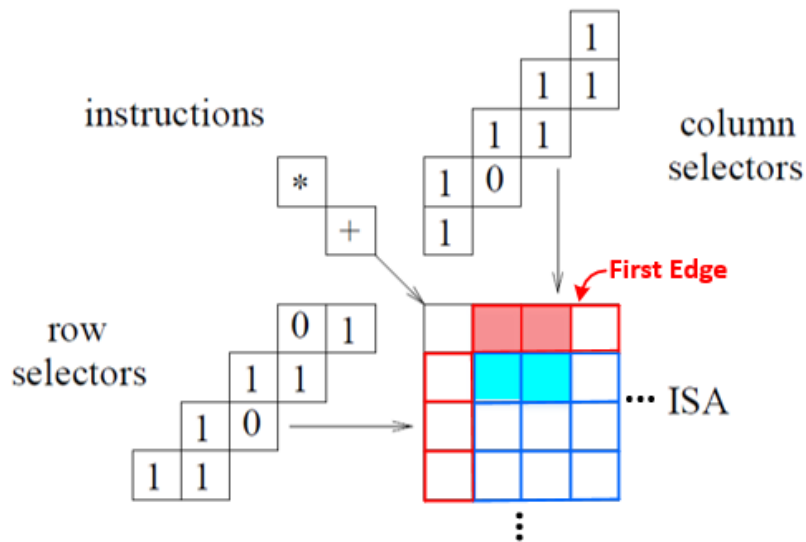


FIG. 2.1. Control flow in an ISA .

As discussed *supra* §§V.A.1-V.A.2 the Schmidt-Kan coprocessor’s 8-bit KPROC ISA uses this same mesh-connected “processor array” meeting [1C].

Franzon, ¶101.

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;”**

In Schmidt-2000’s “mesh connected processor grid,” “[c]ommunication among adjacent processors occurs via designated communication registers.” Schmidt-2000, 163-164 (“Every processor...has a designated communication register that can also be read by the four neighbour processors.”). Each processor’s communication register is connected to that of each of its four neighbors via a pair of wires for exchanging data to and from the adjacent processor. Schmidt-2000, 163 (“each processor is connected to its four direct neighbours by two

unidirectional data wires”), 164 (“Two adjacent processors can exchange data by writing on their own communication register and afterwards reading the communication register of the other....”); Franzon, ¶¶102-103.

The four pairs of wires connecting a processor’s communication register to the four adjacent neighbor processors, shown as north, south, east and west data links (highlighted **purple** in Schmidt-2000’s Fig. 3.1 below), provide PE connections. Franzon, ¶104.

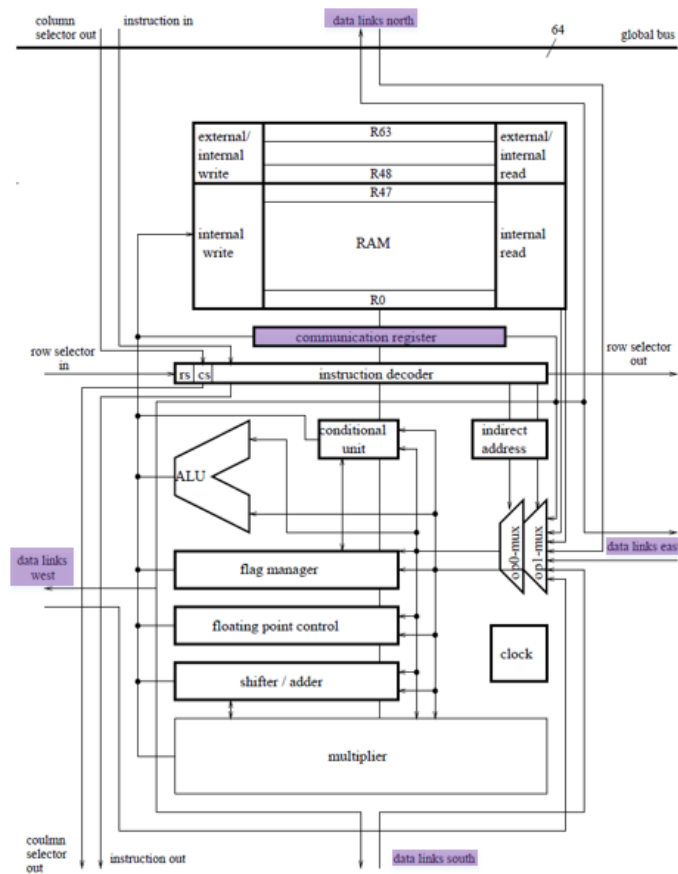


FIG. 3.1. Architecture of one processor.

As annotated in Schmidt-2000’s Figs. 2.1 and 3.1 juxtaposed and annotated below, the first edge PE has “data links south” meeting the claimed “*first*

processing element connection connecting the first edge processing element with the first interior processing element.” The second edge PE has the same “data links south” meeting the claimed “second processing element connection connecting the second edge processing element with the second interior processing element.” Franzon, ¶¶105-106.

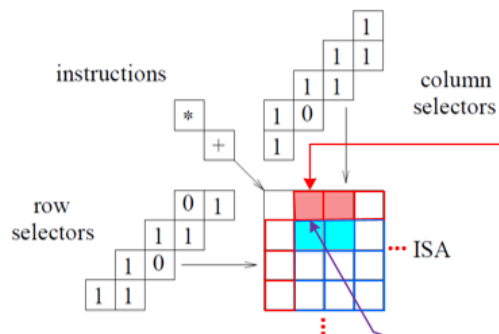


FIG. 2.1. Control flow in an ISA .

First Processing Element Connection

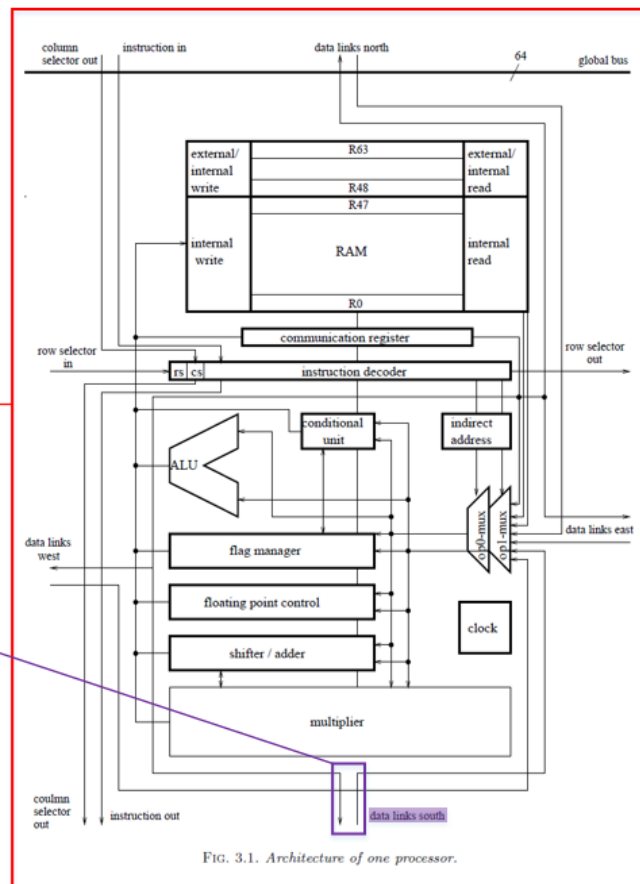
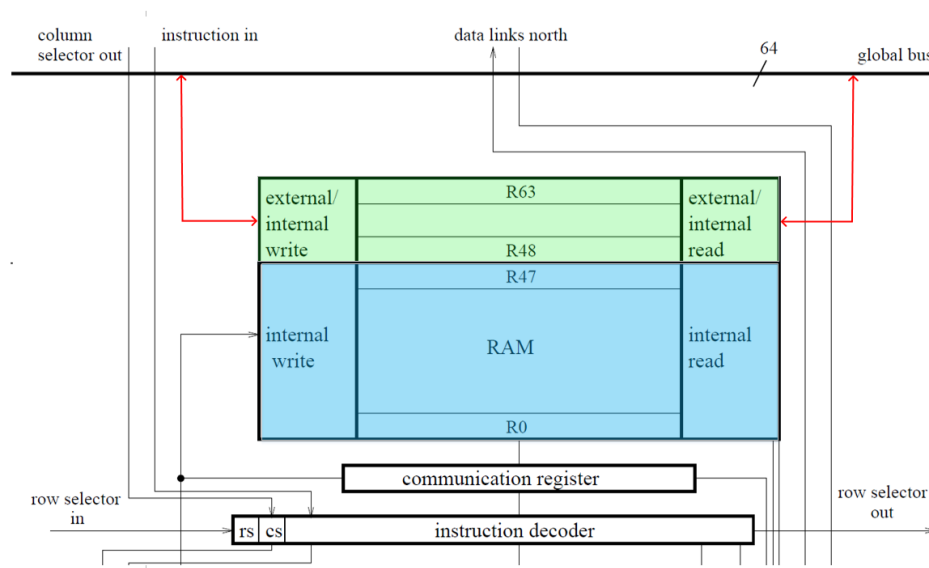


FIG. 3.1. Architecture of one processor.

The Schmidt-Kan coprocessor’s 8-bit KPROC ISA has the same data links south, meeting [1D]. Franzon, ¶106.

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

Schmidt-2000’s “array allows off-chip communication...between host computer and the KPROC-chip.” Schmidt-2000, 163. To implement this communication that “allows...[r]andom access data exchange with the external memory,” “every processor has two memory banks, from which one is connected to the outside, while the other serves as local storage.” Schmidt-2000, 168; Franzon, ¶107. Schmidt-2000’s Figure 3.1 shows the processor’s addressable “64 general-purpose registers” of “local memory” (Schmidt-2000, 164), including 48 registers (R0-R47 shaded in blue below) that form the memory bank serving as local storage (having only “internal write” and “internal read” access), and 16 registers (R48-R63 shaded in green) that form the memory bank connected to the outside for off-chip communication (having “external/internal write” and “external/internal read” access). Franzon, ¶108.



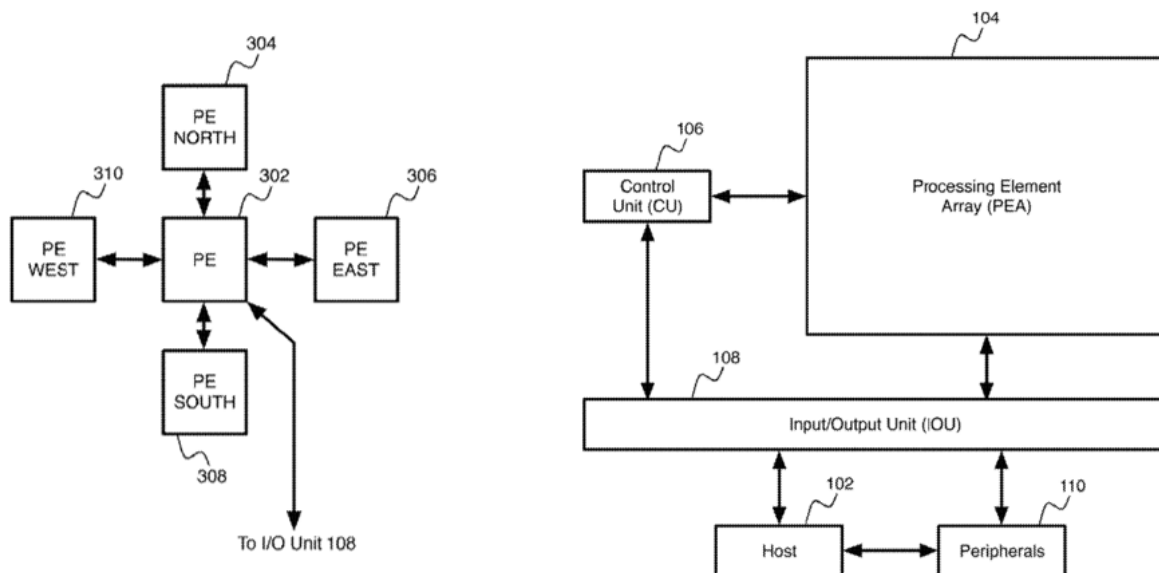
POSAs understood that registers R48-R63 are “connected to the outside” for “data exchange with the external memory” (Schmidt-2000, 168) of the host computer by input-output connections that allow the “off-chip communication...between host computer and the KPROC-chip” (Schmidt-2000, 163). Franzon, ¶109. Those input-output connections were not necessary for Schmidt-2000 to illustrate in Fig. 3.1, because POSAs understood that R48-R63’s “external” read/write is what “connect[s] to the outside” by connecting to the off-chip host computer’s “external memory” via the 64-bit “global bus” illustrated in Fig. 3.1 above, as illustrated by the red arrows annotated above. Franzon, ¶109.

Schmidt-2000 offers no other structure to allow the disclosed “[r]andom access data exchange with[] external memory” and connect registers R48-R63 “to the outside” as disclosed, and POSAs understood the obvious way to implement the functionality Schmidt-2000 describes was to couple R48-R63 to the global data bus. Franzon, ¶¶110-113 (citing as corroborating the POSAs understanding Schimmler-1996 (EX1011), which is bibliography reference [14] in Schmidt-2000, shares a co-author with Schmidt-2000, and shows the obvious use of the PC bus to exchange data between the off-chip host and the on-chip ISA).

Thus, registers R48-R63 of each processor and the global bus implement “off-chip communication...between host computer and the KPROC-chip” (Schmidt-2000, 163) and therefore (each individually or both together) meet the

claimed “*input-output unit.*” Each set of registers R48-R63 has “external/internal” read and write access (Schmidt-2000, Fig. 3.1) to communicate externally with the host and internally to the processor. Schmidt-2000’s input-output unit is thus “*connected to the first edge processing element and the second edge processing element,*” via those processors’ respective registers R48-R63. Franzon, ¶114.

The ’775 patent illustrates a similar configuration in Fig. 3 (below left) where the PE “includes data connection[] to the IOU 108.” EX1001, 9:55-56.



The ’775 patent does not describe what the input/output unit comprises, but merely illustrates it as a box in Figure 1 (above right) and discloses that “to get data into and out of the [optional] CU 106 and PEA 104, the I/O unit 108 may interface the CU 106 and PEA 104 with the Host 102” and that “I/O could be routed through” an “I/O unit 108.” EX1001, 9:5-7, 19-29. Schmidt-2000’s above-described input-output unit interfaces with the KPROC array and the host

computer to get data into and out of the KPROC array and meets [1E] for at least the same reasons the '775 patent's I/O unit 108 would. Franzon, ¶¶115-116.

The Schmidt-Kan coprocessor does not alter how I/O is performed and therefore meets [1E] (whether implemented with or without an on-chip cache as discussed *infra* §VI.B). Franzon, ¶117.

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

Schmidt-2000 discloses that “[e]very processor has read and write access to *its own memory*.” Schmidt-2000, 164. Each processor’s “local memory” was limited to 64 addressable general-purpose registers plus one communication register.” *Id.* The sixty-four addressable registers of local memory are shaded blue in Fig. 3.1 below. Franzon, ¶118; §V.A *supra*.

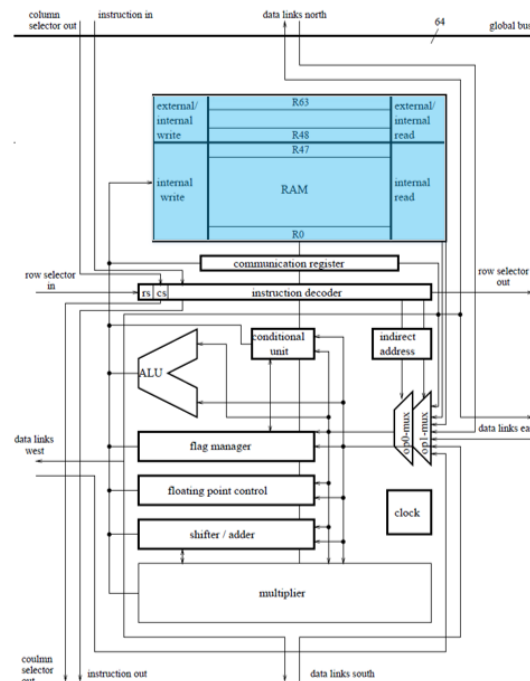


FIG. 3.1. Architecture of one processor.

[1F] is met by all 64 local memory registers R0-R63, or alternatively by the 48 registers R0-R47 having “internal” read/write access. Franzon, ¶119. The Schmidt-Kan coprocessor’s 8-bit KPROC ISA uses this same memory structure, with 8-bit (rather than 16-bit) registers, and meets [1F] for the same reasons. Franzon, ¶119.

8. [1H]: Each Processing Element of Limitation [1C] Comprises an “arithmetic unit”

Each KPROC processor includes “computation units” (within the yellow-shaded box in annotated Fig. 3.1 below) that perform arithmetic operations. Schmidt-2000, 166-167; Franzon, ¶120.

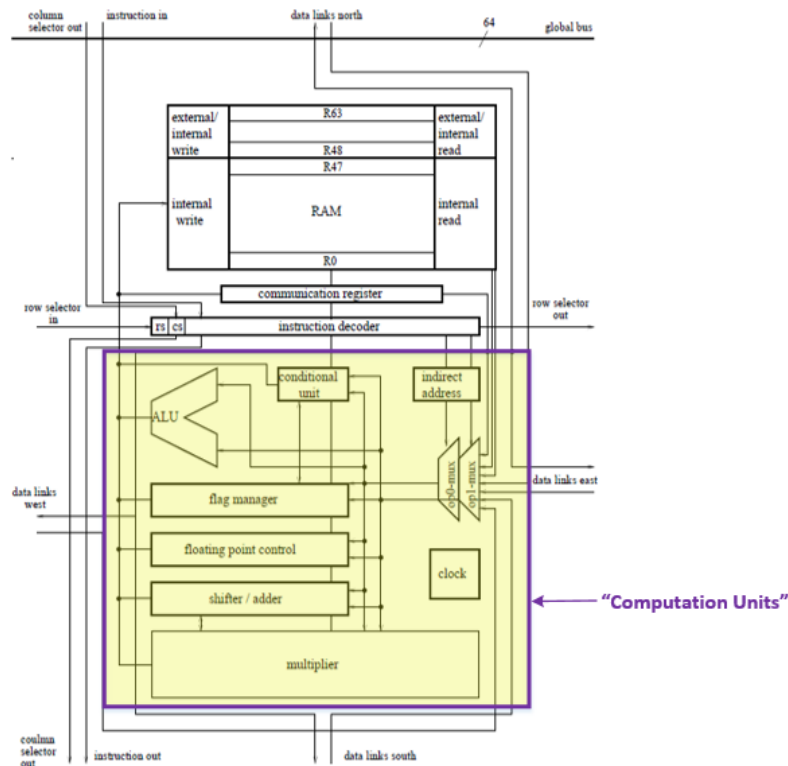
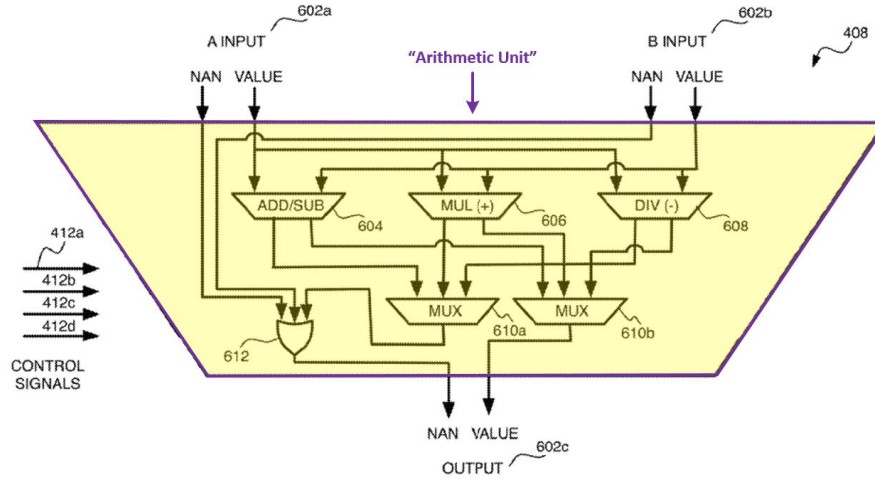


FIG. 3.1. Architecture of one processor.

Each processor's computation include a "unit for integer arithmetic and logical operations" including "a full adder" and "logic gates" implementing a "standard ALU" (Schmidt-2000, 166), plus an integer multiplier and a "floating-point unit" (Schmidt-2000, 167) that utilizes the integer multiplier and integer with the "floating point control" and "shift/adder" to perform floating-point operations. Schmidt-2000, 166-167; Franzon, ¶¶121-122.

These computation units meet the claimed "*arithmetic unit*" for each of [1C]'s PEs, consistent with the '775 patent's description of arithmetic unit 408 as including multiple different computing units ("ADD/SUB," "MUL," etc.). EX1001, 12:53-13:5, FIG. 6 (annotated below); Franzon, ¶¶123-124.



Each processor of the Schmidt-Kan coprocessor's 8-bit KPROC ISA includes this same set of computation units but with an 8-bit architecture, meeting [1H]. Franzon, ¶125.

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

“[O]ff-chip communication” (Schmidt-2000, 163) with Schmidt-2000’s host computer occurs via the KPROC-chip’s I/O pins, which POSAs understood provide the KPROC-chip’s “connect[ion] to the outside” (Schmidt-2000, 168) of the chip (including to the host computer). Schmidt-2000, 163 (describing bandwidth of “off-chip communication...between host computer and the KPROC-chip”), 176 (“I/O bandwidth” is limited by the chip’s “pin limitation”); Franzon, ¶126.

The ’775 patent’s specification does not use the term “host connection” and describes only that “the I/O Unit[] may interface...with the Host 102” (9:5-7). Schmidt-2000’s I/O pins likewise interface the host computer with the KPROC-chip by allowing “off-chip communication...between host computer and the KPROC-chip.” Schmidt-2000, 163; Franzon, ¶127. As discussed *supra* §V.C.6 ([1E]), registers R48-63 that provide external read/write access are part of the claimed “*input-output unit*.” The I/O pins providing the host computer’s interface to the Schmidt-Kan coprocessor chip are connected to the host computer and the chip’s input-output unit to allow off-chip data exchange. Schmidt-2000, 163, 168, 176; Franzon, ¶127. The I/O pins therefore “*connect[] the input-output unit with the host computer*,” meeting [1I]. *Id.*

10. [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;”

As discussed *supra* §V.A and §V.C.8 ([1H]), each KPROC processor includes “computation units” (in the yellow-shaded box annotated in Fig. 3.1 below) meeting the claimed “*arithmetic unit*.” A “floating-point unit” performs floating-point multiplication by using the integer adder to add the exponents and the integer multiplier to multiply the mantissas of two floating-point numbers. Schmidt-2000, 166-167 (“Floating point multiplication is performed in the usual way: Exponents are added and mantissas are multiplied using the integer units discussed above.”); Franzon, ¶128.

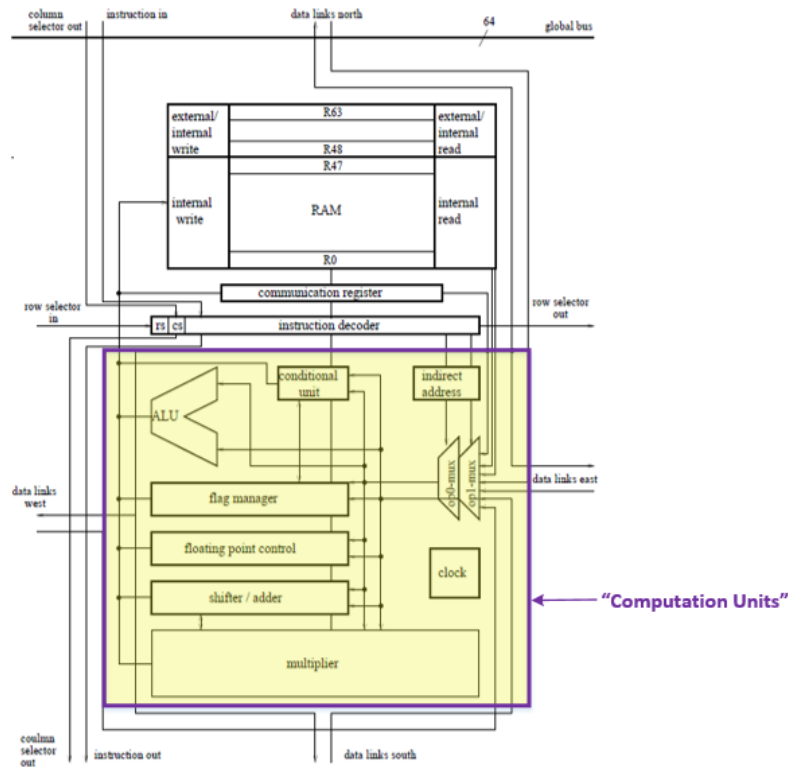


FIG. 3.1. Architecture of one processor.

The integer adder, integer multiplier, “floating point control” and “shift/add” units that are used to perform floating-point multiplication thus meet the claimed “multiplier circuit” and are adapted to receive first and second “floating point value[s]” as first and second inputs, as claimed. Schmidt-2000, 166-167; Franzon, ¶129.

Schmidt-2000’s processors (when not modified to an 8-bit architecture) have a 16-bit architecture with 16-bit registers providing 16-bit operands to computing units that operate on 16-bit operands on each instruction cycle. Schmidt-2000, 163-164, 166; Franzon, ¶130. To perform computations on operands of more than 16-bits, the operands are partitioned into multiple 16-bit words for storage in the

16-bit registers, and the computing units receive and each operate on 16-bit operands at a time.

As discussed above, the processor multiplies two floating-point numbers by adding their exponents in the 16-bit adder and multiplying their mantissas in the 16-bit multiplier. If the mantissas have more than 16 bits, the processor multiplies them by first loading 16 bits of each mantissa into the multiplier, multiplying those 16-bit mantissas, and then subsequently loading the remaining mantissa bits to perform “the same procedure...more than one times,...add[ing] the intermediate results” to arrive at the full product. Schmidt-2000, 166 (description of “Multiplier”); Franzon, ¶131.

In the 16-bit architecture, Schmidt-2000’s multiplier circuit in each processor is thus adapted to receive as first and second inputs two floating-point values, each having an exponent of up to 16 bits (added by the 16-bit adder) and a mantissa of up to 16 bits (the first mantissa partition multiplied by the 16-bit multiplier). Franzon, ¶131. Subsequently, the multiplier circuit can receive any further mantissa bits as inputs for multiplication. *Id.*

Schmidt-2000 discloses capabilities to perform “single-precision” floating-point operations, which involve 8-bit bit exponents. Schmidt, 167; Franzon, ¶132. Thus, the floating-point unit’s multiplier circuit is adapted to receive inputs with 8 exponent bits. *Id.*

As discussed *supra* §§V.A.1-V.A.2, the Schmidt-Kan coprocessor modifies Schmidt-2000 to use an 8-bit architecture. Thus, the floating-point multiplier circuit of each PE’s arithmetic unit is adapted to receive as first and second inputs two floating-point values, each having an 8-bit exponent (added by the 8-bit adder) and a mantissa of up to 8 bits (the first mantissa partition multiplied by the 8-bit multiplier), meeting [1J]. Franzon, ¶133.

11. [1G], [1K]: **“a fifth arithmetic unit;” “wherein the fifth arithmetic unit comprises a corresponding multiplier circuit adapted to receive as inputs to the corresponding multiplier circuit two floating point values each of width at least 32 bits;”**

The Schmidt-Kan coprocessor discussed *supra* §V.B.2 includes a MIMD-type unit having eight PEs, any one of which meets [1G]’s “*fifth arithmetic unit.*” Franzon, ¶134. Each comprises a “32-bit arithmetic unit 21” (Kan, 6:52-56) for “floating point calculations” (Kan, 9:41-44) on “32-bit real number data” (Kan, 9:46-59)—i.e., a floating-point multiplier circuit adapted to receive as inputs two floating-point values each of width at least 32 bits, meeting [1K]. Franzon, ¶134.

12. [1L]: **Each Multiplier Circuit of [1J] “comprises a [first/second/third/fourth] plurality of transistors and has no other transistors”**

Schmidt-2000’s 1024 processors were “integrated on one die using a 0.25-micron digital CMOS process” (Schmidt-2000, 163)—i.e., the 250-nm semiconductor process node for fabricating transistors. Franzon, ¶135. By the

2000s, floating-point multiplier circuits were conventionally implemented using transistors. Franzon, ¶135 (citing corroborating EX1100, 769).

13. [1M]: “the multiplier circuit corresponding to the fifth arithmetic unit comprises a fifth plurality of transistors; and,”

Just as the “first” through “fourth” multiplier circuits comprise respective pluralities of transistors (§V.C.12 above), POSAs knew the floating-point multiplier circuit in the “*fifth arithmetic unit*” (MIMD PEs—§V.C.11 above) comprises a “*fifth plurality of transistors.*” Franzon, ¶136.

14. [1N]: the “fifth plurality of transistors exceeds in number each of the” first through fourth “plurality of transistors”

The ’775 patent does not disclose the absolute or relative number of transistors for any multiplier circuit. It only says that “precision is costly”—e.g., a “double precision floating point multiplier may require on the order of a million transistors,” whereas fewer transistors are “required to perform a low precision multiplication.” EX1001, 7:8-13; Franzon, ¶137.

In the Schmidt-Kan coprocessor’s ISA array of “identical processors” (Schmidt-2000, 163), each 8-bit Schmidt-2000 processor has about the same number of transistors. Franzon, ¶138. As discussed *supra* §§V.A.1-V.A.2, §V.C.10, these processors’ 8-bit adder and multiplier units, used to perform floating-point multiplication, operate on 8-bit operands. The “fifth” multiplier

circuit (in a MIMD PE) has transistors needed to receive and multiply two 32-bit-wide floating-point values. §V.C.11 *supra*. Franzon, ¶138.

POSAs understood that a multiplier circuit adapted to receive and multiply two 32-bit-wide floating-point values conventionally has more transistors than a similarly architected multiplier circuit adapted to receive and multiply two lower-bitwidth floating-point values. Franzon, ¶139 (citing corroborating EX1076, 5:21-26 (“the size of the multiplier is proportional to the number of bits multiplied”)). In Schmidt-2000’s architecture, the multipliers are even smaller. Schmidt-2000, 164-166.

Thus, in the Schmidt-Kan coprocessor, the “fifth” plurality of transistors exceeds in number each of the “first” through “fourth” plurality of transistors, as claimed. Franzon, ¶140.

D. Claim 2: “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.”

POSAs would have readily understood that image pattern matching, such as image database search, is precisely the type of “computationally intensive task[]” in image processing for which the Schmidt-Kan coprocessor’s KPROC array is ideally suited due to the task’s highly parallelizable nature—involving large-scale repetitions of the same computations on different data. Schmidt-2000, 163, 169;

Franzon, ¶141 (citing as corroborating evidence of POSA's background knowledge EX1077 ("Kravec"), 7:6-12; EX1078 ("Caid"), 3:15-24, 13:28-52).

For example, Kravec corroborates the POSA's knowledge that many pattern matching problems involve comparing an input pattern to numerous reference patterns (e.g., an input fingerprint to a library of fingerprints, an input image to a database of images) to find the closest match (1:55-60, 7:6-12) and that sequential matching of patterns (e.g., on a conventional CPU) cannot achieve the desired high performance for these computations (3:58-60). Franzon, ¶142.

POSAs knew that in image matching, the pattern to be matched is typically a vector derived from an image's salient features or characteristics patterns (see Caid, 5:6-8:4; 4:33-56; 13:1-15), and image matching is typically performed by computing a distance between two vectors, such as a dot-product (Caid, Abstract, 2:19-21, 3:15-24, 4:33-56; 13:28-38), Euclidean distance (see Kravec, 10:35-55), Manhattan distance (*id.*), etc. Franzon, ¶143.

POSAs would have understood that such distance computations (involving repeated multiplication and addition operations) are the type of computations the KPROC array is designed for. Franzon, ¶144. The matrix multiplication and convolution operations Schmidt-2000 describes are composed of the same repeated multiplication and addition operations involved in computing the distance between an input pattern and massive numbers of reference patterns. Schmidt-2000, 169-

170; Franzon, ¶144. Thus, a POSA seeking to accelerate image database search and retrieval operations (e.g., as described in background art such as Kravec and Caid) would have implemented the distance computations in the manner Schmidt-2000 describes for matrix multiplication and convolution (Schmidt-2000, 169-170). Franzon, ¶¶144-145. In the Schmidt-Kan coprocessor, the host computer is programmed via processing a “user program” (Kan, 6:13-23) and allocating the distance-computation processing to the KPROC ISA—i.e., the host computer (like Kan’s system controller) provides the instruction stream corresponding to distance computations (via the on-chip controller) to the KPROC ISA for execution (*id.*). Franzon, ¶146.

This obvious use of the Schmidt-Kan coprocessor to perform known image-matching applications renders obvious claim 2. It programs the host computer “*to provide instructions to said computing chip that, when executed, cause said processing element array to perform an operation*” (e.g., desired distance computations between query vectors obtained from an input image and respective feature vectors of images stored in a database) “*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*” (e.g., the distance computation output by the processor array that is used to identify a similar image(s) in the database). Franzon, ¶147.

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

As discussed *supra* §V.C.10, the floating-point multiplier circuit of each 8-bit KPROC processor of the Schmidt-Kan coprocessor is configured to receive first and second inputs having up to 8-bit mantissas (i.e., mantissas of less than 10 bits plus a sign bit). Franzon, ¶148.

VI. GROUND 2A: CLAIMS 3-4 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF KAN AND SCHMIDT-1999

Schmidt-2000 discloses that instructions executed on the KPROC ISA are “input in the upper left corner of the processor array” as are row/column selectors (Schmidt-2000, 164), but does not describe the component that issues these instructions and selectors to the ISA. The obvious and conventional way to issue instructions to a parallel array was via a controller as in Schmidt-1999 (EX1009) (§X *infra*). Schmidt-1999, 955-956; Franzon, ¶149.

A. Implementing an On-chip Controller

In Schmidt-1999, Schmidt-2000’s authors disclosed an ISA implementation similar to Schmidt-2000’s for a video accelerator. Schmidt-1999, 951-955, Fig.1 (below left), Fig. 3 (below right); Franzon, ¶150; *see supra* §V.A.1.

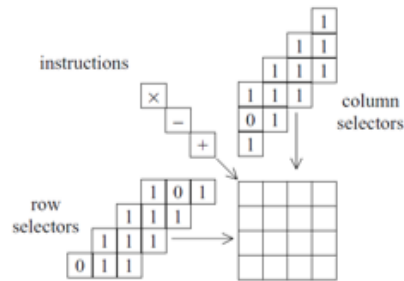


Fig. 1: Control flow in an ISA

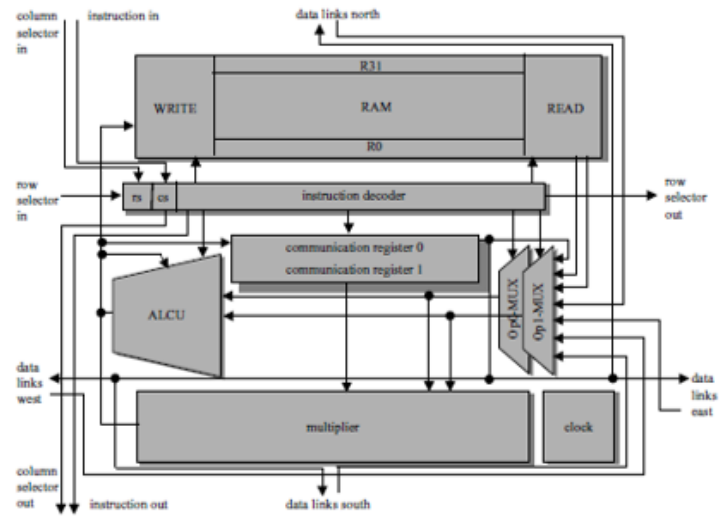


Fig. 3: Architecture of the array processor

Schmidt-1999's Fig. 4 below shows an on-chip controller that “supplies the processor array with instructions and selectors.” Schmidt-1999, 955-956. POSAs thus had motivation and reasonable expectation of success to use an on-chip controller to issue instructions and selectors to the 8-bit KPROC ISA. Franzon, ¶151.

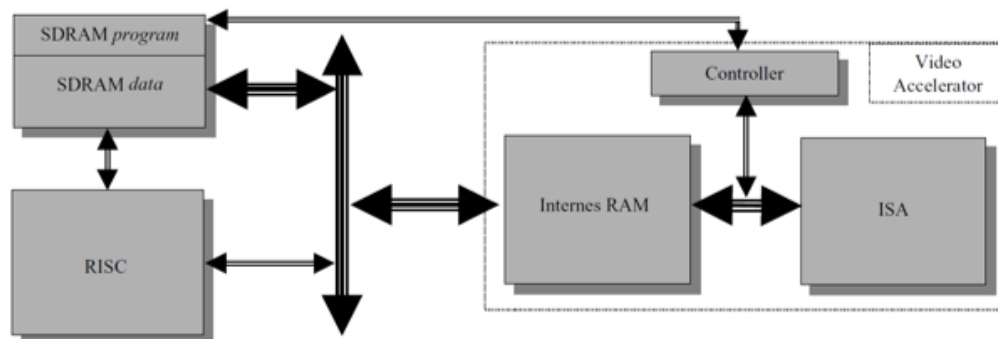


Fig. 4: Video accelerator architecture and data paths to the environment.

B. Implementing an On-chip Cache

To reduce the number of off-chip input/output operations, Schmidt-2000 proposed a “memory hierarchy” using a cache “integrated onto the KPROC chip.” Schmidt-2000, 176; Franzon, ¶152. Schmidt-1999 similarly adopted a “cascaded memory” organization “implemented on-chip” using a switched memory bank configuration like Schmidt-2000’s, and disclosed coupling the first memory bank to an on-chip RAM via a “fast data channel.” Schmidt-1999, 954-955, Fig. 4 above; Franzon, ¶152.

In view of these teachings by the same authors on very similar architectures, POSAs would have had motivation and reasonable expectation of success in implementing Ground 1’s Schmidt-Kan coprocessor described *supra* §§V.A-V.B to have an on-chip cache (as Schmidt-2000 proposed) and an on-chip controller (as Schmidt-1999 taught) to perform the disclosed issuing of instructions and selectors to the 8-bit KPROC ISA described *supra* §§V.A.1-V.A.2 and to control the disclosed data transfer between the on-chip cache and registers R48-R63 of the 8-bit KPROC ISA’s processors (*see* Schmidt-2000 Fig. 3.1 and §V.C.6 *supra* discussing “*input-output unit*”), as Schmidt-1999 taught. Schmidt-1999, 955 (“data transfer between ISA [via the “interface registers”] and internal RAM” is “controlled by [the] on-chip controller” via the “fast data channel”); Franzon, ¶153.

Schmidt-2000 discloses that the on-chip cache “must be global, i.e. every processor can access the cache for data exchange.” Schmidt-2000, 176. Thus, the obvious way to implement the on-chip cache would have been to couple the on-chip cache and the “global bus” (Schmidt-2000, 165, Fig. 3.1) to make the on-chip cache accessible by every processor (as Schmidt-2000 says it “must” be), and allow the cache to access and store data received from external memory (like Schmidt-1999’s on-chip RAM communicates “bidirectionally with the external SDRAM”—Schmidt-1999, 955). Schmidt-2000, 176; Franzon, ¶154; §V.C.6 *supra*. In this manner, the processors can access data from the cache and avoid repeating off-chip communication via the “limited I/O bandwidth” of the I/O pins, thereby “gain[ing] performance by reuse of data,” as disclosed in Schmidt-2000. Schmidt-2000, 176; Franzon, ¶154.

Schmidt-2000’s KPROC array exchanges data with external memory by writing data from external memory to a processor’s registers R48-R63 and reading data from those registers to external memory using a “special instruction.” Schmidt-2000, 168; Franzon, ¶155. A conventional and obvious way POSAs knew to implement the “special instruction” was to use conventional load and store instructions, performed by the on-chip controller. Franzon, ¶155.

For a load instruction (to load data operands into a processor’s registers), the on-chip controller would transfer the data needed by each processor from the cache

to registers R48-R63 (or if the data is not already in the cache, to first retrieve the data from the external memory and store it in the cache) and instruct the processors to switch the content of the memory banks so that the data on which each processor is to operate is loaded into the processor's internal registers R0-47. Schmidt-2000, 168; Franzon, ¶156.

For a store instruction (to read output results from the processors), the on-chip controller would first instruct each processor to switch memory banks and then read the contents of the external registers and store that data in the cache. Schmidt-2000, 168; Franzon, ¶157. In this way, the on-chip cache also can operate as an I/O buffer for data exchange with the host computer, like Schmidt-1999's on-chip RAM operates as an I/O buffer. Schmidt-1999, 955-957; Franzon, ¶157.

C. Reasonable Expectation of Success

As discussed *infra* §VII.B.2, POSAs would have reasonably expected success in single-chip N-KPROC Schmidt-Kan coprocessor implementations that include on-chip controller and cache with the 8-bit KPROC ISA, **plus** scale up the KPROC ISA to include well over 5000 8-bit Schmidt-2000 processors. Thus, POSAs would also have reasonably expected success in implementing the above-described Schmidt-Kan coprocessor chip (which includes a single 1024-processor KPROC ISA) with on-chip controller and cache, on the same size chip as Schmidt-2000's KPROC-chip. Franzon, ¶158.

Ground 2's Schmidt-Kan coprocessor with on-chip controller and cache meets claim 1's limitations for the same reasons as Ground 1's Schmidt-Kan coprocessor demonstrated *supra* §V.C. Further, with an on-chip cache, the global data bus coupled to registers R48-63 of each processor (together or individually) are an “*input-output unit*” as discussed *supra* §V.C.6. Alternatively, those components together with the on-chip cache also are “*an input-output unit*”—they provide an interface with the host computer to “get data into and out of” (EX1001, 9:5-7) the processor array. Franzon, ¶¶159-160 (citing Schimmler-1996, 138, corroborating POSA's knowledge of on-chip controller/RAM I/O and instruction-control architecture for “cascaded memory” like Schmidt-2000's and Schmidt-1999's).

D. Claim 3: “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.”

The on-chip controller discussed *supra* §§VI.A-VI.B meets claim 3's “*control unit*” because it is connected to the KPROC array, adapted to decode load/store instructions received in the instruction stream from the host computer via the input-output unit (including global bus) to load operands into or read results from each processor's registers, and adapted to thereafter issue the special

instruction (i.e., control signals that tell the processor how to operate) to the corresponding processor (claimed “*processing element*”) in the array to cause it to read/write data between the two register files as instructed. Franzon, ¶161.

Furthermore, for a load instruction, the subsequent instructions and row/column selectors issued to the array that instruct the processor to operate on that data are also “control signals” that cause the array to operate in accordance with the load/store instructions. Franzon, ¶161.

E. Claim 4: “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.”

A known and obvious way to implement the on-chip controller discussed *supra* §VI.A §§VI.A-VI.B was to include an on-chip instruction queue to buffer the instruction stream provided by the host computer via the input-output unit, so that the on-chip controller can synchronize execution of the instructions with the ISA’s “global clock” and can operate asynchronously with the host computer. Schmidt-2000, 163 (“The array is synchronised by a global clock.”); Franzon, ¶162 (citing Schimmler-1996, 136-139, corroborating known implementation of instruction queue for the on-chip controller). This on-chip instruction queue, from which the on-chip controller (“*control unit*”) is adapted to retrieve the instructions, meets claim 4’s “*instruction memory*.” Franzon, ¶162.

VII. GROUND 2B: CLAIMS 7-14 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF KAN, SCHMIDT-1999, AND POSA'S BACKGROUND KNOWLEDGE OF SEMICONDUCTOR PROCESS ADVANCES

Independent claim 7 recites a similar computing system as claim 1, but recites “*no less than 5000*” PEs in the claimed array. As discussed below, this limitation carries no patentable weight as it is merely a duplication of parts. *In re Harza*, 274 F.2d 669, 671 (C.C.P.A. 1960).

To the extent the “*no less than 5000*” limitation is given patentable weight, POSAs would have had motivation and reasonable expectation of success to scale-up the number of 8-bit KPROC processors on Ground 2’s Schmidt-Kan coprocessor to the levels process technology permitted, which POSAs knew allowed well over 5000 8-bit KPROC processors to be integrated on the coprocessor chip, even by the 2009 earliest date to which the ’775 patent claims priority.²

A. The 5000-PE Limitation Carries No Patentable Weight

The number of PEs recited in claim 7 carries no patentable weight because “[i]t is well settled that the mere duplication of parts has no patentable significance unless a new and unexpected result is produced.” *Harza*, 274 F.2d at 671. The

² As discussed *supra* §III.E, in fact claims 7-24 are not entitled to priority earlier than June 2010.

'775 patent only references “five thousand” elements once (EX1001, 28:46-51), among a generalized, indiscriminate listing of many different possible numbers of elements, all but one of which are significantly lower than 5000 (*id.*, 28:13-51) and fewer than Schmidt-2000’s 1024 processors, and none of which are limited to being on a single chip. EX1001, 28:13-29:4; Franzon, ¶163.

These embodiments are otherwise indistinguishable: the '775 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. Franzon, ¶163. Thus, claim 7’s “*no less than 5000*” limitation carries no patentable significance. *Harza*, 274 F.2d at 671; *Supercell Oy v. Gree, Inc.*, PGR2018-00008, Paper 42, 47 (Jan. 2, 2019) (“[I]t is axiomatic that duplication has no patentable significance.”) (citing *Harza*).

B. Obviousness of Process Scaling the KPROC ISA

1. Motivation

Schmidt-2000’s KPROC was the “first implementation of a parallel computer with 1024 floating-point processors on a single chip.” Schmidt-2000, 163. Schmidt-2000 discloses “building large arrays of many KPROC chips or speeding up small machines with a single KPROC as a coprocessor,” thus recognizing the performance benefit of using larger numbers of processors to

accelerate computationally intensive tasks beyond what could be integrated on a single chip in the year 2000. *Id.*; Franzon, ¶164.

Thus, Schmidt-2000 reflects the POSA's understanding of the computational-performance benefit of using very large numbers of processors, and the POSA's motivation to increase the number of processors integrated on a single chip as permitted by advances in semiconductor technology. Franzon, ¶165.

Indeed, Schimmler-1996 (discussed *supra* §§VI.A-VI.B) obtained a 1024-processor array by assembling sixteen 64-processor chips on a circuit board. Schimmler-1996, 136-139; *see also* EX1010 ("Schimmler-1994"), 487 ("[t]here are 16 ISA chips on the board, each comprising 64 processors."); Franzon, ¶165.

By natural extension of these teachings, POSAs would have likewise been motivated to integrate multiple KPROC arrays (forming an "N-KPROC ISA") on the single chip of Ground 1's or Ground 2A's Schmidt-Kan coprocessor (referred to herein as the "N-KPROC Schmidt-Kan coprocessor") as later advances in semiconductor processing allowed higher levels of within-chip integration. For example, scaling up the KPROC ISA to a 72x72 array ("5-KPROC" including 5,184 8-bit processors), or a 96x96 array ("9-KPROC" including 9,216 8-bit processors) implemented on the Schmidt-Kan coprocessor chip—are two examples of obvious and desirable implementations to "build[] large arrays" (Schmidt-2000,

163) on a single chip, in which POSAs would have reasonably expected success using semiconductor processes available before June 2009. Franzon, ¶166.

2. Reasonable Expectation of Success

POSAs would have reasonably expected success in implementing an N-KPROC Schmidt-Kan coprocessor (with or without an on-chip controller and cache) using semiconductor processes available in 2009, given that Schmidt-2000 disclosed 1024 floating-point processors on a 1.2-cm² (120-mm²) chip using a 250-nm CMOS process nearly a decade earlier. Schmidt-2000, 163-164; Franzon, ¶167.

a. Moore's Law

POSAs knew that advances in semiconductor process technology allowed for increasing levels of single-chip integration that kept pace with (and often exceeded) the prediction known as “Moore’s Law” that a chip’s integration capacity increases exponentially, doubling approximately every two years. Franzon, ¶¶168-169 (citing EX1026, 1, 86; EX1016, 83-84; EX1017, 36-37, EX1027, 1). The ’775 patent itself acknowledges that Moore’s Law “has proved valid for over 30 years and looks likely to continue in some form.” EX1001, 1:28-30, 6:30-33; Franzon, ¶169.

By application of the well-corroborated Moore’s Law, POSAs would have understood the increase in the number of Schmidt-processors implementable on a

single chip (of the same size as Schmidt-2000 disclosed) to be on the order of thirty-two (i.e., $32 \times 1024 = 32,768$ processors) by 2009, and would have understood that 8192 Schmidt-processors (well over 5000) was achievable at least by 2006, simply based on the time elapsed following Schmidt-2000's disclosure date in 2000. Franzon, ¶170.

Furthermore, the 250-nm process used by Schmidt-2000 was a legacy process node even by 2000, and the 1024-processor KPROC-chip Schmidt-2000 disclosed would have been achievable earlier, in the 1997-1998 timeframe when 250-nm was available and commercialized. Franzon, ¶171 (citing EX1022, 1; EX1023, 1; EX1018, 1; EX1030). Applying Moore's Law starting from that timeframe, POSAs would have understood that over a factor-of-50 increase—indeed, up to a factor-of-64 increase—in the number of Schmidt-processors (i.e., over 51,200 processors) was achievable by 2009, and that a factor-of-eight increase (i.e., 8192 processors) was achievable at least by 2004. Franzon, ¶172.

This background knowledge of the POSA based on Moore's Law (like the law itself) is corroborated by actual transistor-density increases achieved by commercial products fabricated at successively smaller process nodes available in the years up to June 2009, discussed below. Franzon, ¶172.

b. Process Node Scaling

Moore's Law has accurately predicted semiconductor process technology advancement over decades in part thanks to the ITRS, a well-known collaboration of leading companies and research institutions that established performance targets for the semiconductor industry with the stated goal of keeping pace with Moore's Law. Franzon, ¶173 (citing EX1026, 1-3, 87; EX1022, 1; EX1021). These performance targets defined metrics called "process nodes" that the semiconductor industry designed in developing next-generation products, meeting (and often exceeding) those targets. Franzon, ¶¶174-176 (citing EX1021-EX1028).

From the 250-nm process used by Schmidt-2000, semiconductor processing technology advanced through numerous successively smaller process nodes (180-nm, 130-nm, 90-nm, 65-nm, 45-nm, 40-nm, 32-nm nodes) that POSAs knew were available in 2009. As shown below, POSAs would have understood that (at least) the 90-nm node (available and commercialized by 2004) and all smaller nodes would have allowed more than 5000 Schmidt-2000 processors (i.e., a factor-of-five increase) to be integrated on the KPROC-chip without even changing the chip's size. Franzon, ¶177.

POSAs understood that commercially demonstrated process nodes would have allowed more than 50 times as many Schmidt-2000 processors to be integrated on the same size chip (e.g., using a 40-nm process node) according to

the transistor-density (i.e., the number of transistors integrated on a chip divided by the area of the chip) increases achieved by commercial products released before June 2009. Franzon, ¶178.

250-nm: Intel’s Celeron processor (“Mendocino”), released in 1998, integrated 19 million transistors on a 154-mm² die using a 250-nm process—a transistor density of 123,377 transistors/mm². EX1018, 2-3, Fig. 3; Franzon, ¶179.

90-nm: Intel’s 90-nm Pentium 4 processor (“Prescott”), released in 2004, integrated 112 million transistors on a 125-mm² die—a transistor density of 1.1 million transistors/mm², which is 9 times larger than in 250-nm Mendocino. EX1040, 1, Table 1; Franzon, ¶180.

65-nm: The 65-nm process node was available before June 2009 (EX1039, 1)—i.e., available in 2005 with commercial products released in 2006. Franzon, ¶181. Intel’s 65-nm Itanium processor integrated 2.05 billion transistors on a 698.75 mm² die—a transistor density of 2.93 million transistors/mm², which is 24 times larger than in 250-nm Mendocino. EX1039, 18; Franzon, ¶181.

45-nm: Intel’s 45-nm Penryn product, released in 2007, integrated 410 million transistors on a 107-mm² die—a transistor density of 3.8 million transistors/mm², which is 31 times larger than in 250-nm Mendocino. EX1042, 1; EX1052, 1; Franzon, ¶182.

40-nm: The 40-nm node was also available before June 2009 and achieved a “2.35 times density improvement” over the 65-nm node (EX1038, 3, 71; EX1043, 1; EX1137, 1). Franzon, ¶183. Coupled with the above-discussed factor-of-24 increase between 250-nm and 65-nm, this results in a factor-of-56 increase between 250-nm and 45-nm. Franzon, ¶184. This 50+ scale factor is further corroborated by a later 2009 report that AMD’s 40-nm Radeon HD 5870 had integrated 2.15 billion transistors on a 334-mm² die—a transistor density of 6.44 million transistors/mm², which is 52 times larger than in 250-nm Mendocino. EX1041, 1-2; EX1053, 1-2; Franzon, ¶183.

POSAs understood that the above-discussed transistor-density increases between process nodes provide an approximation of the factor increase between those nodes in the number of Schmidt-2000 processors implementable on a single chip without increasing the chip size. Franzon, ¶186. The table below shows these transistor-scale factors for the above-discussed process nodes that were available before June 2009, and the number of Schmidt-2000 processors on the KPROC-chip that would have resulted from each scaling factor, corroborating the above-discussed Moore’s Law estimation of over a factor-of-50 increase before 2009. Franzon, ¶186.

Process Node	Transistor-Density Scale Factor	#Schmidt-2000 Processors
250-nm	1	1024
90-nm	9	9256
65-nm	24	24,576
45-nm	31	31,744
40-nm	52	53,248

Additionally, at 32-nm in 2009, a further factor-of-1.5 increase over 40-nm was achievable. Franzon, ¶185; EX1038, 1; EX1045, 1; EX1094, 67.

Thus, POSAs would have understood from commercial products' transistor-density increases between Schmidt-2000's 250-nm process node and the 40-nm process node available before June 2009 that there would have been a reasonable expectation of success in implementing well over 5000 Schmidt-2000 processors on the KPROC-chip using ordinary skill and process technology available in semiconductor foundries by June 2009. Franzon, ¶187.

c. Implementing an 8-bit N-KPROC ISA

Examples of obvious and desirable 8-bit N-KPROC-ISA implementations to increase computational performance (*see supra* §VII.B.1) include scaling-up the KPROC ISA to include a 72x72 (5-KPROC), a 96x96 (9-KPROC) or a 128x128 (16-KPROC) array. Based on the above-discussed scale factors, the table below shows chip areas occupied by each of the 5-KPROC, 9-KPROC and 16-KPROC arrays at 65-nm, 45-nm and 40-nm nodes, as well as how each area compares with

the 102.4-mm² area occupied by Schmidt-2000's original array in the year 2000, as a percentage (in parentheses) of the original area. Franzon, ¶¶188-189.

Process Node	Scale Factor	Per-Processor Area	5-KPROC Area (72x72) (% KPROC-ISA)	9-KPROC Area (96x96) (% KPROC-ISA)	16-KPROC Area (128x128) (% KPROC-ISA)
250-nm	1	.1mm ²			
65-nm	24	.0042mm ²	21.3mm ² (21%)	38.4mm ² (38%)	68.3mm ² (67%)
45-nm	31	.0032mm ²	16.5mm ² (16%)	29.7mm ² (29%)	52.9mm ² (52%)
40-nm	52	.0019mm ²	9.8mm ² (10%)	17.7mm ² (17%)	31.5mm ² (31%)

The above areas for the N-KPROC ISA do not account for the reduction in processor size of using an 8-bit architecture, which POSAs knew would have reduced the per-processor size by an additional factor of 0.6 and reduced the chip's power consumption. Franzon, ¶¶190-194 (computing scale factor from sizes in Schmidt-1999, 955, Table 2), ¶195 (computing power consumption of 8-bit architecture). Thus, each of these obvious implementations that substantially increase the KPROC-chip's computational performance would have reduced power consumption and occupied only a fraction of the original KPROC-chip's area, and POSAs would have reasonably expected success in implementing them. Franzon, ¶¶196-197.

The sizes of 5-, 9-, and 16-KPROC ISAs for those N-KPROC Schmidt-Kan coprocessor implementations that include on-chip controller and cache (discussed *supra* §§VI.A-VI.B) are shown below. Franzon, ¶¶198-203 (citing Schmidt-1999, 955, Table 2 for size of controller and cache).

Process Node	5-KPROC-ISA	9-KPROC-ISA (Cache + Controller)	16-KPROC-ISA (Cache + Controller)
65-nm	28.7mm ²	51.6mm ²	91.7mm ²
45-nm	22.2mm ²	39.9mm ²	71mm ²
40-nm	13.2mm ²	23.8mm ²	42.3mm ²

Kan does not disclose the area of the MIMD-type unit's "eight processing elements 109" (Kan, 6:52-54), but POSAs knew a wide variety of processors were suitable for use as PEs 109, and that eight such processors could have been integrated on the N-KPROC Schmidt-Kan coprocessor chip using available processing nodes before June 2009. Franzon, ¶204.

For example, Yeh-ISCA07 (EX1012) and Yeh-Micro07 (EX1013) disclose using GPU shader-processors in a parallel-processor array that POSAs would have understood to be suitable for Kan's MIMD-type unit (Kan, 6:52-68) because each processor independently executes its own instruction sequence on data stored in its local memory (i.e., the array performs multiple-instruction multiple-data (MIMD) parallel processing). Yeh-ISCA07, 233-235, 238-241; Yeh-Micro07, 396, 400-403; Franzon, ¶¶204-207.

The table below illustrates the N-KPROC Schmidt-Kan coprocessor's area using 5-KPROC, 9-KPROC and 16-KPROC ISAs and an 8-processor MIMD-type unit using the largest areas disclosed for the Yeh processors (Yeh-ISCA07, 241; Yeh-Micro07, 401) plus a 1MB local memory that Kan discloses each MIMD PE

includes (using memory sizes in Schmidt-1999, 955, Table 2). Franzon, ¶¶208-210.

Process Node	Schmidt-Kan Coprocessor (5-KPROC-ISA)	Schmidt-Kan Coprocessor (9-KPROC-ISA)	Schmidt-Kan Coprocessor (16-KPROC-ISA)
65-nm	178.2mm ²	229.8mm ²	321.5mm ²
45-nm	137.8mm ²	155.6mm ²	186.7mm ²
40-nm	82.2mm ²	92.7mm ²	111.3mm ²

As shown in the table above, the 5-KPROC and 9-KPROC implementations (highlighted yellow) would have been available as single-chip solutions without increasing the KPROC-chip's size and without even accounting for the processor-size reduction of using an 8-bit architecture (or for process scaling of Schmidt-2000's pad frame and clock generation). Franzon, ¶¶211.

The table below shows sizes of the N-KPROC Schmidt-Kan coprocessor using **8-bit** 5-KPROC, 9-KPROC and 16-KPROC ISAs computed using the factor-of-0.6 scaling discussed above (see also §§V.A.1). Franzon, ¶¶192-194. As shown by yellow highlight, an 8-bit 16-KPROC ISA also would have fit on the chip without increasing its size. Franzon, ¶¶211-213.

Process Node	Schmidt-Kan Coprocessor (8-bit 5-KPROC-ISA)	Schmidt-Kan Coprocessor (8-bit 9-KPROC-ISA)	Schmidt-Kan Coprocessor (8-bit 16-KPROC-ISA)
65 nm	169.7mm ²	185.8mm ²	214mm ²
45 nm	131.2mm ²	143.7mm ²	165.5mm ²
40 nm	78.2mm ²	85.7mm ²	98.65mm ²

Additionally, a significant portion of the areas listed above is contributed by the large 1MB local memory assumed for each MIMD PE 109, and POSAs would have understood that many more single-chip N-KPROC Schmidt-Kan coprocessors could have been implemented without increasing the chip size by reducing the MIMD local memory size. Franzon, ¶213.

Furthermore, Schmidt-2000 disclosed that the “chip area should not exceed 1.2 cm²,” but proposed adding an on-chip cache which would result in a larger chip design, demonstrating that the 1.2 cm² (120 mm²) chip was a design choice and not a limitation. Schmidt-2000, 164, 176; Franzon, ¶¶197, 214. POSAs knew that significantly larger chips were routinely used before June 2009 (*see, e.g.*, commercial die sizes discussed *supra* §VII.B.2.b), and would easily accommodate any of the above-described example N-KPROC Schmidt-Kan coprocessor implementations. Franzon, ¶¶214-216.

C. Claim 7

The N-KPROC Schmidt-Kan coprocessor system discussed *supra* §VII.B meets all limitations of claim 7.

1. [7PRE]: “A computing system, comprising:”

Schmidt-2000’s “host computer” (Schmidt-2000, 163) together with the N-KPROC Schmidt-Kan coprocessor chip described *supra* §VII.B as a coprocessor meet [7PRE]’s “*computing system*.” Franzon, ¶218.

2. [7A]: “a host computer;”

Schmidt-2000’s “host computer” meets [7A]. Schmidt-2000, 163; Franzon, ¶219.

3. [7B]: “a computing chip comprising:”

The N-KPROC Schmidt-Kan coprocessor (*supra* §VII.B) is implemented on the same chip Schmidt-2000 discloses, meeting [7B]’s “*computing chip*” as discussed *supra* §V.C.3 ([1B]). Schmidt-2000, 163; Franzon, ¶220.

4. [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;”

As discussed *supra* §VII.B, the N-KPROC Schmidt-Kan coprocessor uses Schmidt-2000’s mesh-connected “processor array” (but with even more processors and meets the claimed “*processing element array*” for the same reasons discussed *supra* §V.C.4 ([1C]). Franzon, ¶221.

As discussed *supra* §VII.A, [7C]’s “no less than 5000” recitation confers no patentable weight. But even if given weight, the limitation is met because as discussed *supra* §VII.B, a 5-KPROC Schmidt-Kan coprocessor (5,184 SIMD processors), 9-KPROC Schmidt-Kan coprocessor (9,216 processors), and 16-KPROC Schmidt-Kan coprocessor (16,384 processors) (and numerous other

implementations) would each have been obvious, desirable scaled-up implementations by June 2009, and each implements well over 5000 “*first processing elements*” on a single chip. Franzon, ¶222.

As discussed *supra* §V.C.4 ([1C]), any processor in the top row or left-most column (or bottom row or right-most column) outlined **red** below is a PE positioned at an “*edge of the processing element array*,” and any of the processors outlined **blue** is “*positioned in the interior of the processing element array*.” Franzon, ¶223.

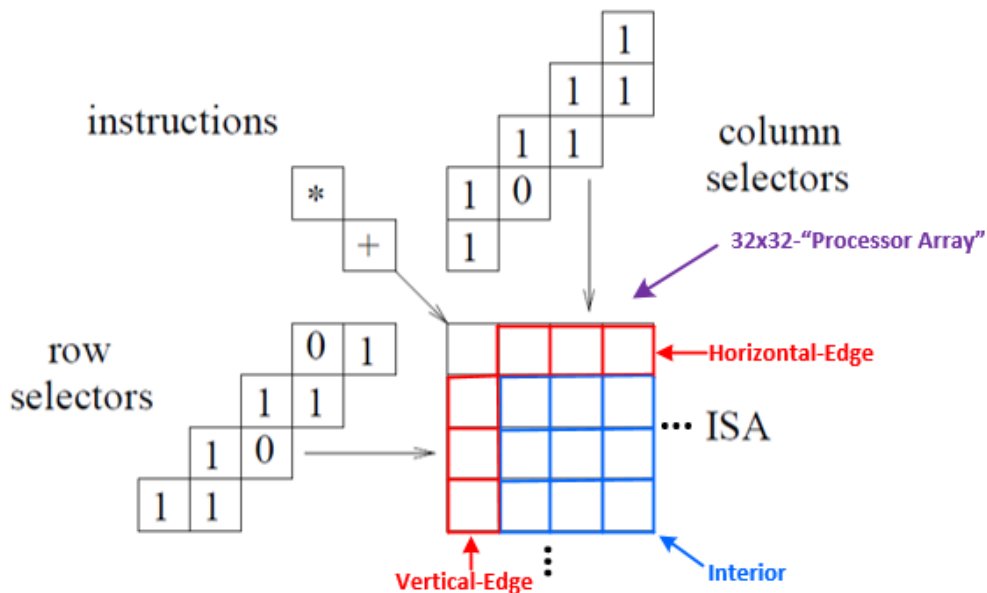


FIG. 2.1. Control flow in an ISA .

As an example meeting [7C], the processors in the top row (or alternatively in the left-most column) outlined **red** above meet “*a first subset*” of the first PEs, each of which is “*positioned at a first edge of the processing element array*,” and

processors highlighted **blue** meet “a second subset,” each of which is “*positioned in the interior of the processing element array.*” Franzon, ¶224.

5. **[7D]: “an input-output unit connected to each of the first subset of the plurality of first processing elements;”**

The N-KPROC Schmidt-Kan coprocessor uses the KPROC-chip’s input/output unit, meeting [7D] as demonstrated *supra* §V.C.6 ([1E]). Franzon, ¶225.

6. **[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;”**

As discussed *supra* §V.C.5 ([1D]), the four pairs of wires connecting the communication register of each Schmidt-2000 processor to its four adjacent neighbors (“data links” highlighted **red** in Fig. 3.1 below) are claimed “*processing element connections.*” Schmidt-2000, 163-164; Franzon, ¶226.

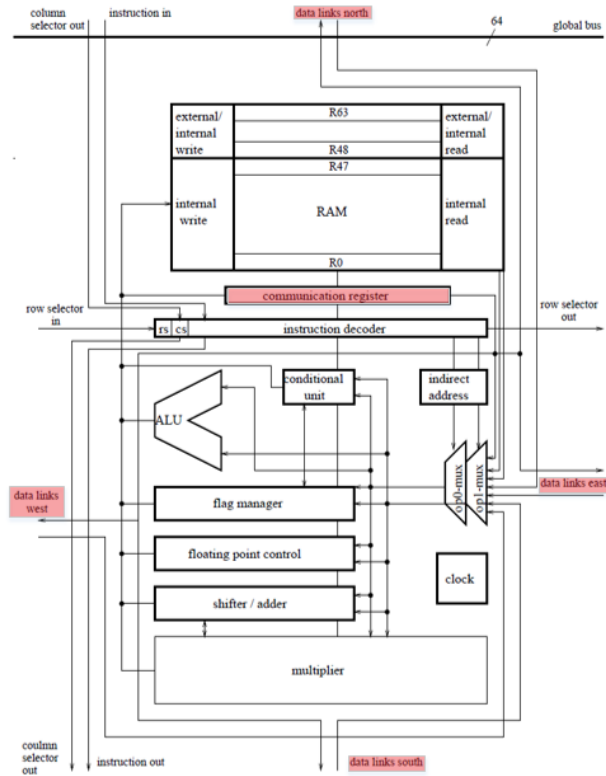


FIG. 3.1. Architecture of one processor.

These data links in the N-KPROC Schmidt-Kan coprocessor meet [7E].

Franzon, ¶227.

7. [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;”

As discussed *supra* §V.C.7 ([1F]), every Schmidt-2000 processor has its “own” 64 general-purpose registers of “local memory” shaded blue below.

Schmidt-2000, 164. The N-KPROC Schmidt-Kan coprocessor uses this same memory structure, but with 8-bit registers (*supra* §§V.A.1-V.A.2, §VII.B); therefore each Schmidt-2000 processor’s 64 registers (or alternatively, just

registers R0-R47) is a “memory unit” local to and associated with the respective processor. Franzon, ¶228.

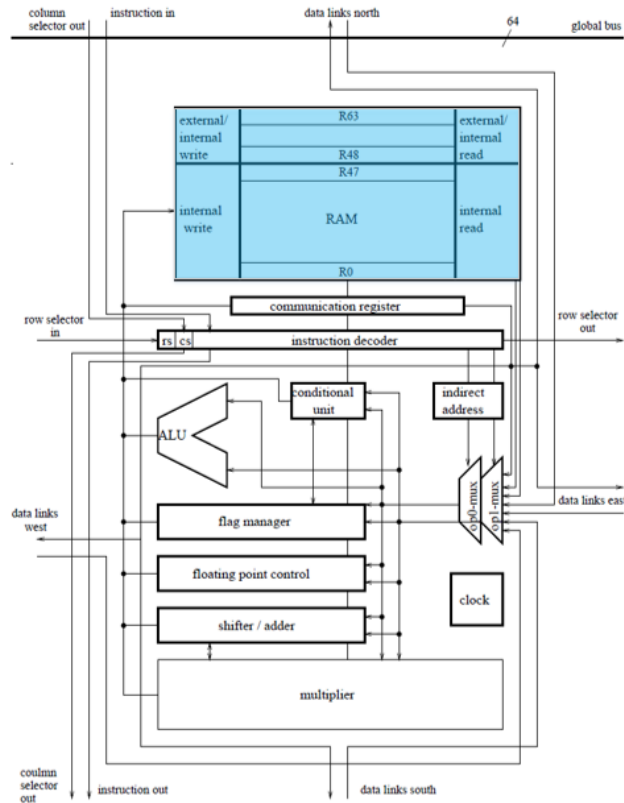


FIG. 3.1. Architecture of one processor.

8. [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;”

Each Schmidt-2000 processor’s “computation units” (in yellow-shaded box below) meet “*at least one...first arithmetic unit*” as demonstrated *supra* §V.C.8 ([1H]). Franzon, ¶229.

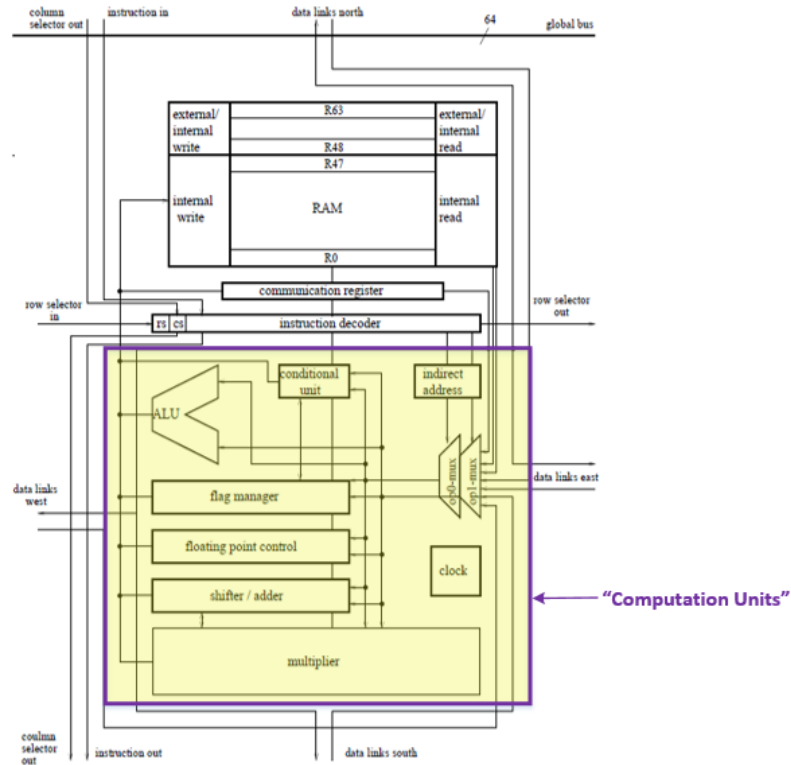


FIG. 3.1. Architecture of one processor.

In the N-KPROC Schmidt-Kan coprocessor, each Schmidt-2000 processor includes these computation units (but with an 8-bit architecture), meeting [7G].

Franzon, ¶230.

9. [7H], [7L]: “a plurality of second processing elements; and a plurality of second arithmetic units, wherein each of the plurality of second processing elements has positioned therein at least one of the plurality of second arithmetic units;” “wherein the plurality of second arithmetic units 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;”

As discussed *supra* §V.B, the N-KPROC Schmidt-Kan coprocessor includes a MIMD-type unit having eight PEs (“second processing elements”). Each

comprises a “32-bit arithmetic unit” (Kan, 6:52-56) for “floating point calculations” (Kan, 9:41-44) on “32-bit real number data” (Kan, 9:46-59)—i.e., an arithmetic unit comprising a multiplier circuit adapted to receive as inputs two floating-point values each of width at least 32 bits, to be multiplied, meeting [7H] and [7L]. Franzon, ¶231.

10. [7I]: “a host connection at least partially connecting the input-output unit with the host computer”

The N-KPROC Schmidt-Kan coprocessor has Schmidt-2000’s host connection that meets [7I] as demonstrated *supra* §V.C.9 ([1I]). Franzon, ¶232.

11. [7J]: “wherein the plurality of first arithmetic units each comprises a first corresponding multiplier circuit adapted to receive” the Same Inputs As Recited in [1J]

As discussed *supra* §V.C.10, the N-KPROC Schmidt-Kan coprocessor uses the 8-bit architecture described *supra* §V.A.2 in its KPROC ISA (with more processors), and therefore meets [7J] as demonstrated *supra* §V.C.10 ([1J]). Franzon, ¶233.

12. [7K]: “wherein the first multiplier circuits corresponding to the plurality of first arithmetic units each comprises a first respective plurality of transistors and has no other transistors;”

The 8-bit Schmidt-2000 processors’ multiplier circuits meet [7K] as discussed *supra* §V.C.12 ([1L]). Franzon, ¶234.

13. [7M]: “wherein the second multiplier circuits corresponding to the plurality of second arithmetic units each comprises a second respective plurality of transistors;”

The MIMD unit's 32-bit multiplier circuits meet [7M] as discussed *supra* §V.C.13 ([1M]). Franzon, ¶235.

14. [7N]: “wherein each of the second respective pluralities of transistors of the second multiplier circuits corresponding to the plurality of second arithmetic units exceeds in number each of the first respective pluralities of transistors of the first multiplier circuits corresponding to the plurality of first arithmetic units.”

Each of the MIMD unit's 32-bit multiplier circuits has more transistors than each of the 8-bit Schmidt-2000 processors' multiplier circuits, as discussed *supra* §V.C.14 ([1N]). Franzon, ¶236.

D. Claim 8

1. [8A]: “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”

As discussed *supra* §V.C.10 ([1J]), §VII.C.11 ([7J]), the N-KPROC Schmidt-Kan coprocessor's 8-bit N-KPROC array is composed of processors having an 8-bit architecture and, as such, not adapted to receive 32-bit floating-point values as inputs. The eight MIMD PEs (“*second processing elements*”) are the chip's only PEs comprising multiplier circuits that are so adapted. Franzon, ¶237.

2. [8B]: “wherein the plurality of first processing elements exceeds in number, by at least 100, the plurality of second processing elements.”

As discussed *supra* §VII.B, each of the 5-KPROC, 9-KPROC and 16-KPROC arrays includes thousands of 8-bit Schmidt-2000 processors, which “*exceeds in number, by at least 100,*” the eight MIMD PEs. Franzon, ¶238.

E. Claim 9: “the plurality of memory units each comprises a corresponding register”

As discussed *supra* §VII.C.7, each Schmidt-2000 processor’s “local memory” comprises addressable “general purpose registers.” Schmidt-2000, 164; Franzon, ¶239. The N-KPROC Schmidt-Kan coprocessor includes this register file but with 8-bit registers. Each register file comprises multiple registers, any one of which is “*a corresponding register.*” *Id.*

F. Claim 10

The N-KPROC Schmidt-Kan coprocessor meets claim 10’s additional limitations for the same reasons Ground 1’s or Ground 2A’s Schmidt-Kan coprocessor meets claim 2’s identical limitations (*supra* §V.D). Franzon, ¶240.

G. Claim 11: “an aggregate maximum bandwidth of the plurality of processing element connections exceeds a maximum bandwidth of the host connection.”

Schmidt-2000’s KPROC array “allows off-chip communication with a data rate of 800 Mbytes per second between host computer and the KPROC-chip” over the KPROC-chip’s pins (Schmidt-2000, 163)—i.e., the “*host connection*”

(§VII.C.10 above)—and this “I/O bandwidth” is “limited” relative to the KPROC array’s “remarkable processing power” that achieves “up to 24 giga instructions per second (with respect to 16 bit operations).” Schmidt-2000, 176; Franzon, ¶241.

Schmidt-2000 discloses that each processor can read from an adjacent processor’s 16-bit (two bytes) communication register via the “*processing element connection*” (§VII.C.6 above) and write to its own communication register on each instruction cycle. Schmidt-2000, 164; Franzon, ¶242. Therefore, the 1024-processor KPROC array can, in aggregate, transfer 2 Kbytes of information between the array’s communication registers each cycle (i.e., over the plurality of PE connections) at a rate of “24 giga instructions per second” (Schmidt-2000, 176)—a bandwidth of 48 terabytes/second, substantially more than the host connection’s 800 Mbytes/second. Franzon, ¶242.

In an 8-bit architecture, twice as many instructions can be performed per second, but the communication registers are half the size, so the aggregate bandwidth of the PE connections is the same as in Schmidt-2000’s 16-bit implementation. Franzon, ¶243. As an N-KPROC array (with more processors) would only increase the aggregate bandwidth of the PE connections by increasing the number of communication registers that can be read/written on each instruction

cycle, the N-KPROC Schmidt-Kan coprocessor's 8-bit N-KPROC array meets claim 11. Franzon, ¶243.

H. Claims 12-14

The N-KPROC Schmidt-Kan coprocessor meets claims 12-13 for the same reasons Ground 2's Schmidt-Kan coprocessor meets claims 3-4 (*supra* §§VI.D-VI.E), and meets claim 14 for the same reasons Ground 1's Schmidt-Kan coprocessor meets claim 5 (*supra* §V.E). Franzon, ¶¶244-245.

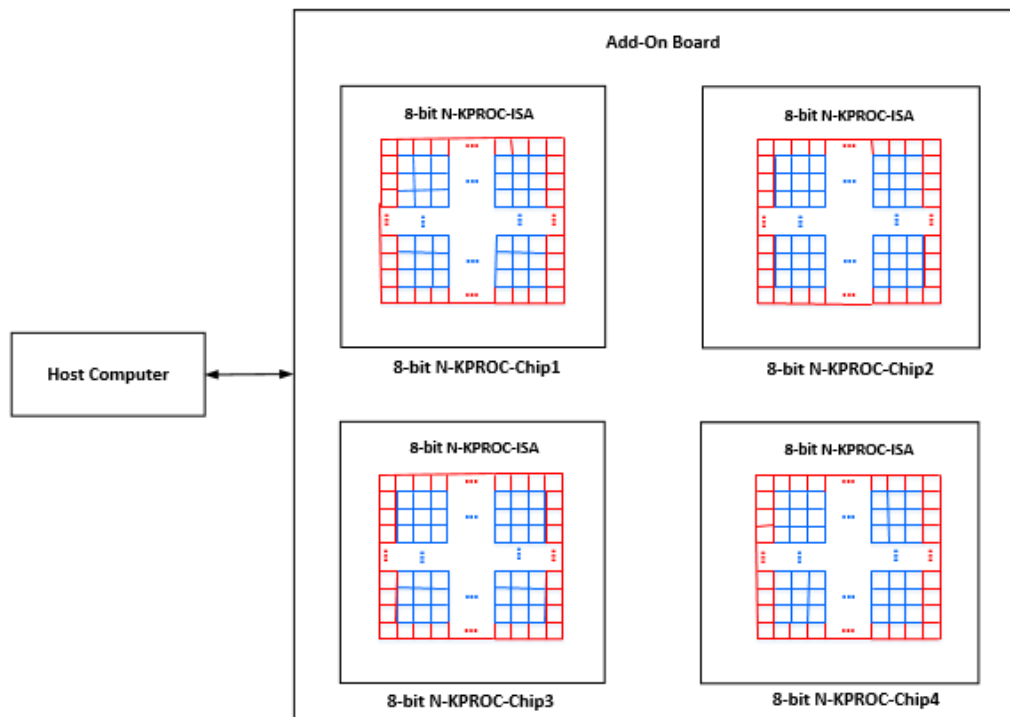
VIII. GROUND 3: CLAIMS 16, 23 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF SCHIMMLER-1996

As discussed *supra* §VII.B, POSAs would have been motivated to increase the number of processors integrated on a single chip as advances in semiconductor technology permitted increasing levels of integration. Franzon, ¶246. Schmidt-2000 discloses that the KPROC architecture's "modular organization" allows for "building large arrays of many KPROC chips." Schmidt-2000, 163. POSAs understood the conventional, obvious way to build larger arrays of processors than could be integrated on a single chip using available semiconductor technology was to tile and interconnect multiple chips on a circuit board—similar to the add-on board solution proposed in Schimmler-1996 (EX1011, cited in Schmidt-2000 and authored by Schmidt-2000's co-author Schimmler) to obtain a 1024-processor array by assembling sixteen 64-processor chips on a circuit board in 1996. *Id.* Schimmler-1996, 136-139; Franzon, ¶246.

It would have been obvious to similarly build large arrays from multiple 8-bit N-KPROC-chips (i.e., 8-bit KPROC-chips scaled to the numbers of 8-bit Schmidt-2000 processors discussed *supra* §VII.B (with or without integration of a MIMD-type unit and/or on-chip controller and cache on each chip) for use by the host computer as a coprocessor of increased computational performance as Schmidt-2000 disclosed, by arranging and interconnecting the chips on a circuit board in the conventional way Schimmler-1996 disclosed. Franzon, ¶247.

A. Add-On Board Implementation of Schmidt N-KPROC-chips

Any number of 8-bit N-KPROC-chips may be used for a Schmidt Add-On Board. An example having four 8-bit N-KPROC-chips is illustrated below. Franzon, ¶248.



B. Claim 16

1. [16PRE]: “A circuit board connected to a host computer, the circuit board comprising:”

The Add-On Board discussed *supra* §VIII.A is connected as a coprocessor to Schmidt-2000’s “host computer” (Schmidt-2000, 163), meeting [16PRE]. Franzon, ¶¶248-249.

2. [16A]: “a plurality of processing element arrays, each comprising:”

The Add-On Board discussed *supra* §VIII.A comprises a plurality of 8-bit N-KPROC-chips (e.g., 5-KPROC-chips, 9-KPROC-chips, or 16-KPROC-chips), each having an ISA processing element array. Franzon, ¶250.

3. [16B]-[16H]

Any one of the 8-bit N-KPROC-chips in the Add-On Board discussed *supra* §VIII.A, with its 8-bit N-KPROC array, meets [16B]-[16H] for the same reasons the N-KPROC Schmidt-Kan coprocessor having the same 8-bit N-KPROC array meets the corresponding limitations in claim 7, as shown below:

Claim 16 Limitation:	Corresponding Claim 7 Limitation:	Addressed in Above Sections:
[16B]	[7C]	§VII.C.4, §V.C.4
[16C]	[7E]	§VII.C.6, §V.C.5
[16D]	[7F]	§VII.C.7, §V.C.7
[16E]	[7G]	§VII.C.8, §V.C.8
[16F]	[7J]	§VII.C.11, §V.C.10
[16G]	[7D]	§VII.C.5, §V.C.6
[16H]	[7I]	§VII.C.10, §V.C.9

Franzon, ¶251.

C. Claim 23

Each 8-bit N-KPROC-chip on the Add-On Board meets claim 23's additional limitations for the same reasons Ground 1's Schmidt-Kan coprocessor meets claim 5's identical limitations (*supra* §V.E). Franzon, ¶252.

IX. GROUND 4: CLAIMS 17-22 ARE UNPATENTABLE OVER SCHMIDT-2000 IN VIEW OF SCHIMMLER-1996, SCHMIDT-1999, AND KAN

For the reasons discussed *supra* §V.B and §VII.B, it would have been obvious to integrate a MIMD-type unit like Kan's and on-chip controller and cache like Schmidt-1999's on each 8-bit N-KPROC-chip on the Add-On Board discussed

supra §VIII.A, resulting in multiple N-KPROC Schmidt-Kan coprocessor chips on a circuit board.

A. Claims 17-19

Claims 17-19 add to claim 16 the same limitations that claim 8 adds to claim 7, plus that the total count of first PEs exceeds the second PEs in number by:

- “at least 20 more than three times” (claim 17)
- “at least 1000 more than five times” (claim 18)
- “at least 100 more” (claim 19)

The limitations identical to claim 8 are met by each N-KPROC Schmidt-Kan coprocessor on the Add-On Board as discussed *supra* §VII.D (claim 8). The additional limitations above are met because, as discussed *supra* §VII.B.2.c, each N-KPROC Schmidt-Kan coprocessor includes over 5000 first PEs and only eight second PEs. Franzon, ¶¶254-255.

B. Claims 20-22

This grounds implementation discussed *supra* §IX with N-KPROC Schmidt-Kan coprocessors on a board meets claims 20-22 for the same reasons each N-KPROC Schmidt-Kan coprocessor meets the identical limitations in claims 10 and 12-13, discussed *supra* §VII.F (claim 10), §VII.H (claims 12-13), respectively. Franzon, ¶256.

X. PUBLIC ACCESSIBILITY OF CITED PRINTED PUBLICATIONS

This section addresses public accessibility of the printed publications cited in the Petition’s grounds. Franzon, ¶¶54-55. Each exhibit below dated prior to June 2008 is §102(b) prior art. Each exhibit dated prior to June 19, 2009 is §102(a) prior art to all challenged claims and §102(b) prior art to at least claims 7-24 whose priority date is no earlier than June 19, 2010 (*supra* §III.E). Each exhibit dated prior to June 2010 is §102(a) prior art to claims 7-24.

EX1007 (“Schmidt-2000”) (at 163) bears “© 2000 Nova Science Publishers, Inc.” and serial number “ISSN 1097-2803,” and its cover material indicates it was published in the June 2000 issue of *Parallel and Distributed Computing Practices* by “well-known, reputable publisher” Nova Science Publishers, Inc. EX1005 (“Bolin”), ¶¶27-38. EX1007 was indexed, catalogued, and publicly accessible in various libraries, and publicly orderable from the publisher’s website in the early 2000s. Bolin, ¶38; Franzon, Appendix §II.A.1; EX1031, ¶¶2-7, pp. 32-39, 44-58; *Hulu v. Sound View Innovations*, IPR2018-01039, Paper 29, 19 (Dec. 20, 2019) (precedential) (finding reasonable likelihood of printed publication based on copyright date, printing date, and ISBN date from “established publisher” of “well-known book series”).

EX1009 (“Schmidt-1999”) (at 3) bears “© Springer-Verlag Berlin Heidelberg 1999” and serial numbers “ISSN 0302-9743,” “ISBN 3-540-66443-2,”

and indicates it was published in 1999 by well-known, reputable publisher Springer. Bolin, ¶¶46-53. EX1009 was indexed, catalogued, and publicly accessible in various libraries, and the publisher's website ("SpringerLink") indicates it was publicly available in 1999. Bolin, ¶¶51-53. Franzon, Appendix §II.A.2.

EX1010 ("Schimmler-1994") (at 3) bears a 1994 copyright, serial numbers, and other indicia of publication by Springer, and was indexed, catalogued, and publicly accessible in various libraries, both in print and online, by 2005. Bolin, ¶¶54-60; Franzon, Appendix §II.A.3.

EX1011 ("Schimmler-1996") (at 1, 3) bears 1996 library stamps, copyright, ISBN serial number and Library of Congress Catalog number, and indicates it was published by well-known, reputable publisher SPIE. Bolin, ¶¶61-66. EX1011 was indexed, catalogued, and publicly accessible in various libraries; and SPIE's Digital Library lists a publication of August 1996. Bolin, ¶¶64-66; Franzon, Appendix §II.A.4.

EX1018 bears "August 24, 1998" and was published by Microprocessor Report, "a well-known, reputable newsletter in the microprocessor industry." Franzon, Appendix ¶97. It was also publicly accessible from Microprocessor Report's website by 2004. Franzon, Appendix §II.C.6.

EX1019 bears “© 2002 Elsevier Science B.V.” and was publicly accessible from www.sciencedirect.com in 2002. Franzon, Appendix §II.C.7.

EX1021-EX1027, EX1030 were published by the ITRS, “widely distributed to POSAs throughout the world” (Franzon, Appendix §II.B; EX1033), and publicly accessible from the ITRS’s website, before June 2008 (EX1021-EX1026, EX1030) or by April 2010 (EX1027). Franzon, Appendix §II.B (citing Internet Archive declaration EX1029).

The exhibits below bear copyrights on the pages cited, indicating they were published before June 2008 by the ACM and/or IEEE, which are “well-known, reputable compiler[s] and publisher[s] of scientific and technical publications.” *Ericsson v. Intellectual Ventures I*, IPR2014-00527, Paper 41, 10–11 (May 18, 2015); *Microsoft Corp. v. IPA Techs. Inc.*, IPR2019-00810, Paper 44, 159 (Oct. 15, 2020); *id.*, Paper 12, 81.

- EX1012, 232;
- EX1013, 394 & 3 (library stamp);
- EX1016, 2;
- EX1017, 36;
- EX1020, 226;
- EX1034, 273;

- EX1039, 18 & 1, 3 (library stamp);
- EX1094, 67;
- EX1100, 759;
- EX1125, 170.

See also Franzon, Appendix §II.C.1, §II.C.2, §II.C.4, §II.C.5, §II.C.8, §II.C.43, §II.C.10, §II.C.20, §II.C.23, §II.C.40; Bolin, ¶¶67-80, ¶¶101-105, ¶¶113-122 (discussing EX1012-EX1013, EX1039, EX1125).

Further, testimony from IEEE representatives demonstrates public accessibility of the following publications as of the corresponding years below:

- EX1013 (2007)—*see* EX1127, ¶¶6-12;
- EX1034 (2000)—*see* EX1035, ¶¶6-12;
- EX1039 (December 2008)—*see* EX1128, ¶¶6-12;
- EX1045 (December 2008)—*see* EX1129, ¶¶6-12;
- EX1094 (June 19, 2008)—*see* EX1130, ¶¶6-12;
- EX1100 (2006)—*see* EX1133, ¶¶6-12;
- EX1125 (2000)—*see* EX1136, ¶¶6-12.

See also Franzon, Appendix §II.C.16 (discussing EX1045).

EX1040 (at 5) bears “© 2005 Dell Inc.” and was additionally publicly accessible from Dell’s website in 2007. Franzon, Appendix §II.C.11; EX1047, 60-65; EX1048, 7-16.

EX1041 was publicly accessible from guru3d.com in October 2009. Franzon, Appendix §II.C.12; EX1047, 9-14; EX1048, 69-71, 116-127.

EX1042, EX1052-EX1053 were publicly accessible from anandtech.com in 2007, February 2009, and June 6, 2009, respectively. Franzon, Appendix §II.C.13, §II.C.17-§II.C.18; EX1047, 15-52; EX1048, 52-74.

EX1038, 72 bears “© 2008 Altera Corporation” and “June 2008” and was additionally publicly accessible from Altera’s website by December 2008. Franzon, Appendix II.C.9; EX1047, 81-153; EX1048, 107-112.

EX1043 bears “Issued on: 2008/11/17” and was publicly accessible from TSMC’s website by that date. Franzon, Appendix §II.C.14.

EX1137 bears “Issued on: 2008/11/17” and was publicly accessible from TSMC’s website by March 2009. Franzon, Appendix §II.C.42.

XI. SECONDARY CONSIDERATIONS

In IPRs on related patents, Singular alleged evidence of secondary considerations, which the Board found was not probative of nonobviousness. Singular “d[id] not show skepticism as to whether or how a problem could be solved;” its “alleged industry praise [wa]s too ambiguous and high-level;” it “d[id]

not establish a difference between the alleged [unexpected] results and the closest prior art;” its alleged commercial success evidence was “too speculative;” and it failed to demonstrate nexus with accused products. IPR2021-00155, Paper 62, 38-51. The same failings would apply if Singular raised the same evidence and arguments for the ’775 patent.

If Singular seeks in later filings to raise secondary considerations evidence/arguments directed to the ’775 patent, the Board should not consider such evidence/arguments where “Petitioner has not had an opportunity to respond.” *Sony Corp. v. Creative Tech.*, IPR2016-01407, Paper 13, 24 (Dec. 6, 2016).

XII. NO BASIS EXISTS FOR DISCRETIONARY DENIAL

No prior proceeding has asserted invalidity of the ’775 patent’s claims,³ and Singular represented to Google that it will soon voluntarily dismiss (without prejudice) the only litigation asserting the ’775 patent, which is in its infancy. Schmidt-2000, Schmidt-1999, Kan, and Schimmler-1996 were not previously presented to the Office, nor are they cumulative. Therefore, there is no basis for discretionary denial under §314(a) or §325(d).

³ Petitioner’s concurrent IPR petition is addressed in a separate Ranking paper filed herewith.

XIII. CONCLUSION

The Board should institute review and cancel claims 1-5, 7-14, and 16-23.

Dated: December 22, 2022

Respectfully submitted,
Google LLC

/Elisabeth Hunt/
Elisabeth H. Hunt, Reg. No. 67,336
WOLF, GREENFIELD & SACKS, P.C

VIII.B.2

XIV. APPENDIX: 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 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;
[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; a fourth memory local to the second interior processing element; and,
[1G] a fifth arithmetic unit;
[1H] 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,
[1I] a host connection at least partially connecting the input-output unit with the host computer;
[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;
[1K] wherein the fifth arithmetic unit comprises a corresponding multiplier circuit adapted to receive as inputs to the corresponding multiplier circuit two floating point values each of width at least 32 bits;
[1L] wherein the multiplier circuit corresponding to the first arithmetic unit comprises a first plurality of transistors and has no other transistors, the multiplier circuit corresponding to the second arithmetic unit comprises a second plurality of transistors and has no other transistors, the multiplier circuit corresponding to the third arithmetic unit comprises a third plurality of transistors and has no other transistors, the multiplier circuit corresponding to the fourth arithmetic unit comprises a fourth plurality of transistors and has no other transistors, and
[1M] the multiplier circuit corresponding to the fifth arithmetic unit comprises a fifth plurality of transistors; and,
[1N] wherein the fifth plurality of transistors exceeds in number each of the first plurality of transistors, the second plurality of transistors, the third plurality of transistors, and the fourth plurality of transistors.
Claim 2
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 3
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.
Claim 4
The computing system of claim 3, 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 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;
[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;
[7H] a plurality of second processing elements; and a plurality of second arithmetic units, wherein each of the plurality of second processing elements has positioned therein at least one of the plurality of second arithmetic units; and
[7I] a host connection at least partially connecting the input-output unit with the host computer;

[7J] wherein the plurality of first 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;
[7K] wherein the first multiplier circuits corresponding to the plurality of first arithmetic units each comprises a first respective plurality of transistors and has no other transistors;
[7L] wherein the plurality of second arithmetic units 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;
[7M] wherein the second multiplier circuits corresponding to the plurality of second arithmetic units each comprises a second respective plurality of transistors;
[7N] wherein each of the second respective pluralities of transistors of the second multiplier circuits corresponding to the plurality of second arithmetic units exceeds in number each of the first respective pluralities of transistors of the first multiplier circuits corresponding to the plurality of first arithmetic units.
Claim 8
[8PRE] The computing system of claim 7, wherein,
[8A] 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
[8B] wherein the plurality of first processing elements exceeds in number, by at least 100, 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 circuit board connected to a host computer, the circuit board comprising:
[16A] a plurality of processing element arrays, each comprising:
[16B] a plurality of first processing elements, wherein the plurality of first processing elements of said processing element array is no less than 5000 in number, wherein each of a first subset of the plurality of first processing elements of said processing element array is positioned at a first edge of said processing element array, and wherein each of a second subset of the plurality of first processing elements of said processing element array is positioned in the interior of said processing element array;
[16C] a plurality of processing element connections, each processing element connection connecting one of the plurality of first processing elements of said processing element array with another of the plurality of first processing

elements of said processing element array, wherein each of the plurality of first processing elements of said processing element array is connected to at least one other of the plurality of first processing elements of said processing element array by at least one of the plurality of processing element connections of said processing element array;
[16D] a plurality of memory units, wherein each of the plurality of first processing elements of said processing element array is associated with a corresponding one of the plurality of memory units of said processing element array, and wherein each of the plurality of memory units of said processing element array is local to its associated one of the plurality of first processing elements of said processing element array; and,
[16E] a plurality of arithmetic units, wherein each of the plurality of first processing elements of said processing element array has positioned therein at least one of the plurality of arithmetic units of said processing element array;
[16F] wherein the plurality of arithmetic units of said processing element array 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,
[16G] an input-output unit connected to at least one of the plurality of processing element arrays; and,
[16H] a host connection at least partially connecting the input-output unit with the host computer.
Claim 17
[17A] The circuit board of claim 16, wherein the circuit board 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;
[17B] wherein other than the second processing elements, the circuit board 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,
[17C] wherein a total count of the pluralities of first processing elements of the plurality of processing element arrays exceeds in number, by at least 20 more than three times, the plurality of second processing elements.

Claim 18
[18A] The circuit board of claim 16, wherein the circuit board 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;
[18B] wherein other than the second processing elements the circuit board 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,
[18C] wherein a total count of the pluralities of first processing elements of the plurality of processing element arrays exceeds in number, by at least 1000 more than five times, the plurality of second processing elements.
Claim 19
[19A] The circuit board of claim 16, wherein the circuit board 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;
[19B] wherein other than the second processing elements the circuit board 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,
[19C] wherein a total count of the pluralities of first processing elements of the plurality of processing element arrays exceeds in number, by at least 100 more, the plurality of second processing elements.
Claim 20
The circuit board of claim 18, wherein said circuit board is adapted to receive instructions from the host computer that, when executed, cause at least one of said processing element arrays 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 21
The circuit board of claim 18, wherein the circuit board further comprises a control unit connected to at least one of the plurality of processing element arrays, 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 at least one of the plurality of processing

element arrays to cause the at least one of the plurality of processing element arrays to operate according to the at least one instruction.
Claim 22
[22A] The circuit board of claim 21, wherein the circuit board further comprises an instruction memory connected to the control unit, the instruction memory comprising circuitry adapted to store the at least one instruction received from the host computer via the input-output unit, and
[22B] wherein the control unit further comprises circuitry adapted to retrieve the at least one instruction from the instruction memory.
Claim 23
The circuit board 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 24
The circuit board of claim 23, 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.

CERTIFICATE OF SERVICE UNDER 37 C.F.R. § 42.6 (E)(4)

I certify that on December 22, 2022, I will cause a copy of the foregoing document, including any exhibits or appendices filed therewith, to be served via Priority Express Mail at the following correspondence address of record for the patent:

Blueshift IP LLC
1 Broadway, 14th Floor
Cambridge, MA 02142

Date: December 22, 2022

/MacAulay Rush/
MacAulay Rush
Paralegal
WOLF, GREENFIELD & SACKS, P.C.

CERTIFICATE OF WORD COUNT

Pursuant to 37 C.F.R. § 42.24, the undersigned certifies that the foregoing Petition for *Inter Partes* Review contains 13,947 words excluding a table of contents, a table of authorities, Mandatory Notices under § 42.8, a certificate of service or word count, or appendix of exhibits or claim listing. Petitioner has relied on the word count feature of the word processing system used to create this paper in making this certification, manually adding to that the number of words present in annotations to figures.

Date: December 22, 2022

/MacAulay Rush/

MacAulay Rush

Paralegal

WOLF, GREENFIELD & SACKS, P.C.