IEEE CSC & ESAS SUPERCONDUCTIVITY NEWS FORUM (global edition), No. 49, March 2021. Invited presentation Wk1EOr1B-01 given at the virtual ASC 2020, October 27, 2020.

Wk1EOr1B-01

ASC 2020 Virtual Conference: October 24 – November 7, 2020

A 4-bit RISC-Dataflow AQFP MANA Microprocessor: Architecture, Design Challenges, and Demonstration

Christopher L. Ayala*

Ro Saito, Tomoyuki Tanaka, Tomohiro Tamura, Naoki Takeuchi, and Nobuyuki Yoshikawa

Yokohama National University, Yokohama, Kanagawa, Japan *email: ayala-christopher-pz@ynu.ac.jp • chris.ayala@ieee.org





Outline

- Motivation and background
- MANA processor
 - Design goals
 - Microarchitecture and ISA
- Tested breakout chips
- MANA prototype chip
- Outlook
- Summary

Motivation

Trend of rising electricity demand of information and communications technology (ICT).

Approaching 10% of the total electric power worldwide in 2020.

Facebook Data Center, Lulea, Sweden



Performance: 27-51 PFLOP/s Power 84 MW avg* (120 MW max)

D.S. Holmes, ISS 2013, Tokyo, Japan.

http://worldstopdatacenters.com/renewable-energy-output-rankings/



N. Jones, Nature, vol. 561, no. 7722, pp. 163–166, Sep. 2018.

Worst-case scenario: ICT could use as much as 50% of global electricity by 2030.

A. S. G. Andrae and T. Edler, Challenges, vol. 6, no. 1, pp. 117–157, Jun. 2015.

AQFP logic for computing

- 4
- Adiabatic quantum-flux-parametron (AQFP) logic
 - **Extremely small bit energy** $<< I_c \Phi_0$
 - Very small switching energy due to adiabatic operation
 - 1.4 zJ at 4.2 K in experiment [2]
 - High gain
 - 10-50x gain from µA's of input current
 - High robustness
 - Clock speeds on par with state-of-theart CMOS logic (5-10GHz)



N. Takeuchi et al., Supercond. Sci. Technol. 26, 035010 (2013).
 N. Takeuchi et al., *Appl. Phys. Lett.*, vol. 114, no. 4, p. 042602, Jan. 2019.

After cooling overhead [2], ~80x more efficient than 7nm FinFET with V_{DD} = 0.8V [3]

[2] D.S. Holmes et al., IEEE TAS, 23, no.3, (2013)[3] A. Stillmaker et al., Integration. 58, pp. 74-81 (2017)

AQFP logic a promising candidate for energy-efficient computing.

Adiabatic quantum-flux-parametron (AQFP)

5



Potential energy of the AQFP



a switching event.

Operation is based on conventional QFP gates [1]. Switching energy can be reduced below $I_c \Phi_0$ by using AC excitation currents, I_v .

[1] M. Hosoya et al., IEEE Trans. Appl. Supercond. 1, 77-89 (1991).

Adiabatic quantum-flux-parametron (AQFP)

6



 $+I_{in} \rightarrow SFQ$ stored in left loop, logic '1'. $-I_{in} \rightarrow SFQ$ stored in right loop, logic '0'. Potential energy of the AQFP



Potential energy changes adiabatically during a switching event.

Operation is based on conventional QFP gates [1]. Switching energy can be reduced below $I_c \Phi_0$ by using AC excitation currents, I_x .

[1] M. Hosoya et al., IEEE Trans. Appl. Supercond. 1, 77-89 (1991).

Adiabatic quantum-flux-parametron (AQFP)

7



 $+I_{in} \rightarrow SFQ$ stored in left loop, logic '1'. $-I_{in} \rightarrow SFQ$ stored in right loop, logic '0'. Potential energy of the AQFP



Potential energy changes adiabatically during a switching event.

Operation is based on conventional QFP gates [1]. Switching energy can be reduced below $I_c \Phi_0$ by using AC excitation currents, I_x .

[1] M. Hosoya et al., IEEE Trans. Appl. Supercond. 1, 77-89 (1991).

Data propagation in AQFP logic

8



[1] N. Takeuchi et al., Appl. Phys. Lett., vol. 114, no. 4, p. 042602, Jan. 2019.

Cell library: minimalist design



Any combinational logic gates can be designed by arraying the four building blocks.

N. Takeuchi et al., J. Appl. Phys., vol. 117, no. 17, p. 173912, May 2015.

Cell library: minimalist design

10

 $L_{in} = 1.13 \text{ pH}$ $L_{x} = 5.67 \text{ pH}$ $L_{d} = 6.16 \text{ pH}$ $L_{1}, L_{2} = 1.53 \text{ pH}$ $L_{q} = 7.88 \text{ pH}$ $L_{out} = 31.9 \text{ pH}$ $k_{d1}, k_{d2} = -0.154$ $k_{x1}, k_{x2} = -0.209$ kout = -0.515 $J_{1}, J_{2} = 50 \mu \text{A}$

Excitation/clock lines are 50Ω microstriplines

Interconnect are shielded striplines





4-layer Nb/AIO_x/Nb 10 kA/cm² high-speed standard process (HSTP) by AIST, Tsukuba, Japan

N. Takeuchi *et al., Supercond. Sci. Technol.*, vol. 30, no. 3, p. 035002, Mar. 2017. C. L. Ayala *et al., Supercond. Sci. Technol.*, vol. 33, no. 5, p. 054006, Mar. 2020. IEEE CSC & ESAS SUPERCONDUCTIVITY NEWS FORUM (global edition), No. 49, March 2021. Invited presentation Wk1EOr1B-01 given at the virtual ASC 2020, October 27, 2020.

Overall AQFP design flow



Perspective from ASC 2016



ASC 2016, Denver, Colorado USA - [1EOr2B-02] - September 05, 2016

MANA microarchitecture

13



MANA – Monolithic Adiabatic iNtegration Architecture

- **Goal:** Demonstrate AQFP can do both logic and memory
- RISC-like datapath + dataflow-like control
- In-order, single-issue
- 4-bit data word size
- 16-bit instr. word
- Program branching
- 21,460 JJs in 1 x 1 cm² chip
- 15 fJ/op at RT @ 5 GHz
- 4-phase 5 GHz clock
- Latency: 108 clock phases or 27 cycles (5.4 ns @ 5 GHz)

5.596 JJs 8 cycles (32 phases)

8,142 JJs 8 cycles (32 phases)

2.238 JJs 9 cycles (36 phases)

Ctrl buffer, routing, write-back (WB) 5.484 JJs 17 cycles (68 phases) overlapped 2 cycles (8 phases) write-back

MANA instruction set architecture

1	

Instruction Word									Vord				Description						
15	14	13	12	11	10	9	8	7 6 5 4 3 2 1 0			1	0	S=Stall bit, NPC= OPCODE=operation code,						
S	NF	PC 0		OF	CO	DE		RA			RB				RA = address for operand A, RB = address for operand B				
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NOP, no operation, used for stalling			
0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	HALT, end program			
1	0	0	0	0	0	0	1			JI	MP /	۱DD	R			JMP, absolute jump to [JMP ADDR]			
1	IBI N	NPC	0	0	0	1	0			JI	MP /	۱DD	R			BNEG, branch to [JMP ADDR] if NEG flag is set			
1	IBI N	NPC	1	0	0	1	0			JI	MP /	۱DD	R			BNNEG, branch to [JMP ADDR] if NEG flag is NOT set			
1	IBI N	VPC	0	0	0	1	1			JI	MP /	۱DD	R			BEQ, branch to [JMP ADDR] if EQ flag is set			
1	IBI N	NPC	1	0	0	1	1			JI	MP /	۱DD	R			BNEQ, branch to [JMP ADDR] if EQ flag is NOT set			
S	IBI N	NPC	0	0	1	0	0		IM	IM			R	В		ANDI, bitwise AND with immediate: R[RB]=R[RB]&&IMM			
S	IBI N	VPC	0	0	1	0	1		IM	М			R	В		LI, load immediate: R [RB] = [IMM]			
S	IBI N	NPC	0	0	1	1	0	0	0	AN	ΛT		R	В		<pre>SRL, shift right logical: R[RB]=R[RB]>>[AMT]</pre>			
S	IBI N	NPC	0	0	1	1	0	0	1	AN	ΛT		R	В		SRA, shift right arithmetic: R[RB]=R[RB]>>>[AMT]			
S	IBI N	NPC	0	0	1	1	1	0	0	٨N	ΛT		R	В		SLL, shift left logic: R[RB]=R[RB]<<[AMT]			
S	IBI N	NPC	0	0	1	1	1	0	1	AN	ΛT		R	В		SLA, shift left arithmetic (same as SLL) : R[RB]=R[RB]<<< [AMT]			
S	IBI N	NPC	0	1	0	0	0		R	A			R	В		ADD, addition: R [RB] = R [RA] + R [RB]			
S	IBI N	NPC	1	1	0	0	0		R	A		RB				ADDNW, add with no write: R[0]=R[RA]+R[RB]			
S	IBI N	VPC	0	1	0	0	1		R	A			R	В		SUB, subtract: R[RB]=R[RA]+(~R[RB]+1)			
S	IBI N	NPC	1	1	0	0	1		R	A			R	В		SUBNW, subtract with no write: R[0]=R[RA]+(~R[RB]+1)			
S	IBI N	NPC	0	1	0	1	0		R	A			R	В		XOR, bitwise XOR R: [RB]=R[RA]⊕R[RB]			
S	IBI N	VPC	0	1	0	1	1		R	A			R	В		XNOR, bitwise XNOR: $R[RB] = (R[RA] \oplus R[RB])$			
S	IBI N	VPC	0	1	1	0	0		R	A			RB			AND, bitwise AND: R[RB]=R[RA]&&R[RB]			
S	IBI N	NPC	1	1	1	0	0		R	A		RB			ANDNB, bitwise AND not B: R[RB]=R[RA]&&~R[RB]				
S	IBI N	VPC	0	1	1	0	1		R	A	RB			OR, bitwise OR: R[RB]=R[RA] R[RB]					
S	IBI N	NPC	1	1	1	0	1		RA RB			ORNB, bitwise OR not B: R[RB]=R[RA] ~R[RB]							
S	IBI N	NPC	0	1	1	1	0	MADD		DDR	DR			LFM, load from memory: R[14,R15]=MEM[MADDR][Hi,Lo]					
S	IBI N	NPC	0	1	1	1	1	MAD			DDR				WTM, write to memory: MEM[MADDR]=R[R14]:R[R15]				



Architecturally 2-stage pipeline

- Stage 1: determine stall based on prefetched stall bits (1 cycle latency)
- Stage 2: fetch instruction from IB, decode, execute, write back (107 cycles total)
- Allows peak IPC of 1

\backslash	Instruction	S	NF	ъС	Opcode				RA				RB				
Instr. 1	add \$4, \$3	1	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1
Instr. 2	add \$3, \$6	0	1	0	0	1	0	0	0	0	0	1	1	0	1	1	0
Instr. 3	xor \$5, \$7	0	1	1	0	1	0	1	0	0	1	0	1	0	1	1	1

Instr. 2 depends on Instr. 1's **\$3**. Compiler sets S-field of Instr. 1 to '1' thus Instr. 2 must wait.

- S-field stall-bit: compile-time hazard detection + hardware stall
- Stall-bit tells next instruction to wait
- Propagates with its instruction
- Returns with processed data as ACK signal to notify next instruction can be issued

Breakout: MANA instruction decoder-controller

15



- Input:
 - 12-bit instruction word
 - Processor flags (ZERO, NEG)
 - Debug input
- Generates 46 control signals + debug
- □ Logic synthesis and GA-based place-and-route
- Updates for MANA integration
 - 16-bit instruction word
 - Addition of CARRY flag
 - NPC (next PC) changed to request next block of 4 instructions
 - Was used to avoid long PC calculation every cycle
 - Now IB is stall logic controlled parallel shift register
 - All jumps/branches will force a stall
- 2.0 mm x 2.6 mm
- □ 2664 JJs
- □ Latency:
 - **7** cycles for datapath signals (1400 ps @ 5 GHz)
 - □ 1 cycle for stall logic (200 ps @ 5 GHz)

C. L. Ayala et al., ISEC (2019), Riverside, CA, USA.

Breakout: MANA register file (16 x 4-bit)

16



M. Nozoe et al., ASC 2018, Seattle, USA.

[1] C. L. Ayala et al., Supercond. Sci. Technol., vol. 33, no. 5, p. 054006, Mar. 2020.

IEEE CSC & ESAS SUPERCONDUCTIVITY NEWS FORUM (global edition), No. 49, March 2021. Invited presentation Wk1EOr1B-01 given at the virtual ASC 2020, October 27, 2020.

Breakout: MANA register file (16 x 4-bit)





Updates for MANA integration

- R0 remains zero-constant
- \square R1 \rightarrow normal register, no longer ones-constant
- R14, R15 modified to interface with serial I/O

MAJ DFF-based registers with additional circuitry for serial I/O (R14, R15)

Breakout: MANA EX high-speed chip

18



C. L. Ayala et al., IEEE Trans. Appl. Supercond., vol. 27, no. 4, pp. 1–7, Jun. 2017.
 C. L. Ayala et al., ISS 2017, Tokyo, Japan.

[3] N. Takeuchi et al., Appl. Phys. Lett., vol. 110, no. 20, p. 202601, May 2017.

- MANA EX high-speed chip
 - Nb/AIO_x/Nb 10 kA/cm² technology
 - ALU-shifter datapath + control signal buffering
 - ALU: MAJ-based Kogge-Stone adder [1] with in-place logic operators
 - **Shifter:** Synthesized logic-arithmetic data shifter [2]
 - **7** mm x 7 mm chip
 - 2.1 mm x 3.5 mm core
 - 2076 JJs
 - Latency: 9 cycles (1800 ps @ 5 GHz)
- High-speed considerations
 - Roundtrip of meandering clock < 5 mm
 - High-speed voltage drivers [3]
 - High-speed He-immersion chip probe
 - AC clock sources
 - BERT: Single bit data generator, single bit output checker



Breakout: MANA EX high-speed chip



Output interface uses unipolar return-to-zero encoding.

19

Functionally exhaustive low-speed test (100 kHz)

- Tested logical operators, addition/subtraction (random and carry propagate), and shifter operations.
- □ All tests passed.

Breakout: MANA EX high-speed chip



20

21





MANA prototype chip

- Nb/AIO_x/Nb 10 kA/cm² technology
- All stages integrated together by hand
- 1 cm x 1 cm
- Unoptimized clock network
- Wire-bonded
- 21,460 JJs
- Latency: 27 cycles (5.4 ns @ 5 GHz)

Experiment

- Low-speed testing
- 4x16-bit instruction blocks manually loaded to IB of IDI serially
- 4-bit debug output tapped from WB data

22

Smoke test: set and read registers



Smoke test passes at 100 kHz.

23

Test program 1: add/sub + branch



27-cycle stall due to data/ctrl hazards

Test program 1 successfully passes.

24

Test program 2: shift/sub + branch



Test program 2 successfully passes.

RF R/W, ALU execution, branching, and hardware stalling successfully demonstrated.

25

Excitation margins of chips



Statistics of measured chips

Wafer	Chip 1	Chip 2	Chip 3	l _c a	Working
MANA-W1	100 kHz⁵	X	100 kHz	110.3%	5 / 6 chips
MANA-W2	100 kHz	100 kHz	100 kHz	107.5%	
EX-W1	2.3 GHz	2.5 GHz	2.1 GHz	96.0%	7 / 12 chips
EX-W2	2.1 GHz	X	100 kHz	90.8%	
EX-W3	1.5 GHz	U	1.2 GHz	87.8%	
EX-W4	X	U	X	91.1%	

'U' denotes unstable or partial operation.

'X' denotes no meaningful output.

^a Measured I_c over designed I_c

^b Note that MANA chips were tested only up to 100 kHz.

Comparison with other demonstrated adiabatic work

	[1]	[2]	[3]	This work	<u>This work</u>
Circuit	16 b CLA	8 b DLX processor	16 b MIPS processor	4 b MANA processor	4 b EX (ALU- shifter)
Status	Tested	Tested	Layout in progress	Tested	Tested
Technology	0.8 µm CMOS	0.18µm NMOS	90nm CMOS	AQFP Nb/AlO _x /Nb	AQFP Nb/AIO _x /Nb
Clk. Rate	4 MHz (tested)	880 kHz (tested)	0.5 GHz (simulated)	100kHz (tested)	2.5 GHz (tested)
Supply	2.5 V dc	1.8 V dc	1 V dc	1 mA ac	1 mA ac
Energy/op	4 pJ	8.5 pJ	3 fJª	15 fJ ^b	1.6 fJ ^b

[1] J. Lim et al., IEEE JSSC, vol. 34, no. 6, pp. 898–903, Jun. 1999.

[2] S. Kim et al., in Proc. of Comp. Frontiers - CF '05, 2005.

[3] R. Celis-Cordova et al., in IEEE ICRC, Nov. 2019.

^a Authors simulated 3 b shift register. AQFP equivalent is 4.2 aJ with cooling. ^b Already includes cooling overhead coefficient of 1000x.

- Wide AC1/AC2 excitation margins
 - 5 dB / 4.6 dB for MANA at 100 kHz
 - 2.6 dB / 2.4 dB for EX chip at 2.5 GHz
- Measured tests repeatable across multiple chips and wafers.
- Superior speed and energy when compared with other demonstrated adiabatic work.
- More work to be done to have a clear competitive edge over bleeding edge FinFET.

Outlook

How do we move forward
with AQFP logic?

Comparison of AQFP INV and FinFET INV FO3/FO4

	AQFP	5 nm FinFET	7 nm FinFET
Power Supply	2 x 1 mA AC + 1 mA DC	0.45 V ~ 0.65 V	0.45 V ~ 1 V
Delay (ps)	~10 [3]	?? ~ 8.3	0.667 ~ 40
Switching Energy (fJ)	~0.0014 @ 5 GHz	0.106 ~ 0.291	0.111 ~ 1.317

Includes 1000x cooling overhead for AQFP

Area efficiency

Cell-level	
------------	--

- Advanced process such as MIT LL SFQ5ee [1]
- Directly coupled QFP (DQFP) [2]
- Novel compact memory

Design methodology

 Physical rows with multiple excitation phases available Delay line clocking [3]

Latency / clock distribution

- Power divider clocking [4]
- Clock domain crossing synchronizers

Flux trapping

Moat embedded interconnects [5]

Advanced EDA tools

- □ Flux trapping analysis [6]
- □ Flexible chip-level integration tools [7]
- More mature tools [7]

[1] Y. He et al., Supercond. Sci. Technol., vol. 33, no. 3, p. 035010, Feb. 2020.

[2] N. Takeuchi *et al.*, *Supercond. Sci. Technol.*, vol. 33, no. 6, p. 065002, May 2020.
[3] N. Takeuchi *et al.*, *Appl. Phys. Lett.*, vol. 115, no. 7, p. 072601, Aug. 2019.

[4] Y. He et al., Appl. Phys. Lett., vol. 116, no. 18, p. 182602, May 2020.

[5] C. J. Fourie *et al.*, *IEEE Trans. on Appl. Supercond.*, vol. 30, no. 6, pp. 1–9, Sep. 2020.
[6] K. Jackman *et al.*, *IEEE Trans. on Appl. Supercond.*, vol. 27, no. 4, pp. 1–5, Jun. 2017.
[7] IARPA SuperTools research program

FinFET data sources:

- E. Sicard, Introducing 7-nm FinFET technology in Microwind. 2017.
- A. Stillmaker and B. Baas, "Scaling equations for the accurate prediction of CMOS device performance from 180nm to 7nm," Integration, vol. 58, pp. 74–81, Jun. 2017.
- N. Collaert, "Device architectures for the 5nm technology node and beyond," presented at the SEMICON Taiwan, 2016.
- S. Sinha et al., "Design benchmarking to 7nm with FinFET predictive technology models," in Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design - ISLPED '12, Redondo Beach, California, USA, 2012, p. 15.

Summary

- 27
- AQFP Logic
 - Superconductor logic using Josephson junctions operating adiabatically
 - Energy dissipation: 1.4 aJ/op at 5GHz (includes cooling)
- MANA processor
 - 4-bit prototype design to show AQFP logic can do both processing and storage
 - Nb/AIO_x/Nb 10 kA/cm² superconductor process
 - Key processor operations demonstrated: R/W, ALU execution, stalling, program branching at 100 kHz
 - Standalone EX stage operated up to 2.5 GHz
 - First demonstration of adiabatic computing using superconductor logic
- Promising technology platform for next generation data centers and supercomputers
- Challenges still exist particularly in improving area efficiency and latency at large-scale

Stage	Description	Total JJs	Latency (cycles)	fJ/op ^a
IDI RFX EX WB	Instruction buffer, <u>D</u> ecode, <u>I</u> ssue <u>R</u> egister <u>F</u> ile with e <u>X</u> ternal I/O interface <u>EX</u> ecution stage (ALU-shifter) Write Back ctrl buffering routing	5596 8142 2238 5484	8 8 9 2	3.917 5.699 1.567 3.839
VVD	<u>Minte Dack, eth, builening, routing</u>	0-0-	2	0.000
	Total:	21460	27	15.022

Stage-by-stage summary of MANA

Includes 1000x cooling overhead

IEEE CSC & ESAS SUPERCONDUCTIVITY NEWS FORUM (global edition), No. 49, March 2021. Invited presentation Wk1EOr1B-01 given at the virtual ASC 2020, October 27, 2020.



Thank You

This work was supported by the Grant-in-Aid for Scientific Research (S) No. 19H05614 and the Grant-in-Aid for Early Career Scientists No. 18K13801 from the Japan Society for the Promotion of Science (JSPS).

This work was also supported by the VLSI Design and Education Center (VDEC) of the University of Tokyo in collaboration with Cadence Design Systems, Inc.

The circuits were fabricated in the Clean Room for Analog-digital superconductiVITY (CRAVITY) of the National Institute of Advanced Industrial Science and Technology (AIST) using the high-speed standard process (HSTP).