

MANA: A Monolithic Adiabatic iNtegration Architecture Microprocessor Using 1.4-zJ/op Unshunted Superconductor Josephson Junction Devices

Christopher L. Ayala¹, Senior Member, IEEE, Tomoyuki Tanaka², Ro Saito², Member, IEEE, Mai Nozoe², Member, IEEE, Naoki Takeuchi², Member, IEEE, and Nobuyuki Yoshikawa², Senior Member, IEEE

Abstract—We conducted the first successful demonstration of an adiabatic microprocessor based on unshunted Josephson junction (JJ) devices manufactured using a Nb/AIO_x/Nb superconductor IC fabrication process. It is a hybrid of RISC and dataflow architectures operating on 4-b data words. We demonstrate register file R/W access, ALU execution, hardware stalling, and program branching performed at 100 kHz under the cryogenic temperature of 4.2 K. We also successfully demonstrated a high-speed breakout chip of the microprocessor execution units up to 2.5 GHz. We use a logic primitive called the adiabatic quantum-flux-parametron (AQFP), which has a switching energy of 1.4 zJ per JJ when driven by a four-phase 5-GHz sinusoidal ac-clock at 4.2 K. These demonstrations show that AQFP logic is capable of both processing and memory operations and that we have a path toward practical adiabatic computing operating at high-clock rates while dissipating very little energy.

Index Terms—Adiabatic, adiabatic quantum-flux-parametron (AQFP), ALU, energy efficient, microprocessor, superconductor electronics.

I. INTRODUCTION

SUPERCONDUCTOR Josephson junction (JJ) devices have recently garnered substantial attention due to their use in qubits for quantum annealing and general quantum computing processors [1], [2]. These JJs are also very capable of performing conventional computation, operating as classical digital circuits and have the potential to be an energy-efficient foundation for the next generation of supercomputers and data centers [3]. With the power demand of information and

communications technology expected to rise to over 20% by 2030, half of which is attributed to data centers [4], [5], there is a need to consider other technological alternatives for the future development of our information and communications infrastructure, and this is an area where JJ-based superconductor logic may fulfil.

A number of JJ-based energy-efficient superconductor logic families exist, including energy-efficient rapid single flux quantum (ERSFQ) logic [6], energy-efficient SFQ (eSFQ) logic [7], reciprocal quantum logic (RQL) [8], LR-biased RSFQ logic [9], and low-voltage RSFQ (LV-RSFQ) logic [10]. But these are all non-adiabatic logic families in which their switching energy dissipation remains unchanged regardless of the clock rate they operate at.

Adiabatic quantum-flux-parametron (AQFP) logic is a superconductor logic family shown to operate with a switching energy of 1.4 zJ per JJ when driven by a four-phase 5-GHz ac power-clock at 4.2 K in experiments using unshunted JJ devices [11]. As expected from adiabatic circuits, the switching energy of the AQFP also decreases linearly as the power-clock frequency decreases, or more specifically, as the rise time of the power-clock increases [12]. With a cryogenic cooling overhead taken into account [3], the switching energy increases 1000× to 1.4 aJ at 5 GHz, which is still ~80× more efficient than a 7-nm multi-gate technology with a $V_{DD} = 0.8$ V [13]. Fig. 1(a) shows the schematic of the AQFP. The AQFP is composed of two components: 1) the two-junction dc superconducting quantum interference device (dc-SQUID) which corresponds to the J_1 - L_1 - L_2 - J_2 loop and 2) the output transformer which corresponds to the mutual inductance pair of L_q and L_{out} . An excitation clock line (L_x) and a dc-offset line (L_d) are coupled to the dc-SQUID via L_1 and L_2 . An input data current is provided through L_{in} . When an ac current (power-clock) ramps up through the clock excitation line (L_x), a single-flux-quantum (SFQ) is induced in the J_1 - L_1 - L_q loop if the input current through L_{in} was positive (logic ‘1’) or the J_2 - L_2 - L_q loop if it was negative (logic ‘0’). A large current is induced by the SFQ through L_q whose direction is also positive for logic ‘1’ or negative for logic ‘0.’ The current through L_q induces an output current through L_{out} via their mutual inductance. This is the active state of the AQFP. When the ac power-clock ramps down, the SFQ in either loop of the

Manuscript received August 26, 2020; revised October 24, 2020; accepted November 17, 2020. Date of publication December 15, 2020; date of current version March 26, 2021. This work was supported in part by the Grant-in-Aid for Scientific Research (S) under Grant N° 19H05614 and in part by the Grant-in-Aid for Early Career Scientists from the Japan Society for the Promotion of Science under Grant N° 18K13801. This article was approved by Associate Editor Yusuke Oike. (Corresponding author: Christopher L. Ayala.)

Christopher L. Ayala and Naoki Takeuchi are with the Institute of Advanced Sciences, Yokohama National University, Yokohama 240-8501, Japan (e-mail: chris.ayala@ieee.org; takeuchi-naoki-kx@ynu.ac.jp).

Tomoyuki Tanaka, Ro Saito, Mai Nozoe, and Nobuyuki Yoshikawa are with the Department of Electrical and Computer Engineering, Yokohama National University, Yokohama 240-8501, Japan (e-mail: tanaka-tomoyuki-wy@ynu.jp; saito-ro-mw@ynu.jp; nozoemai@gmail.com; yoshikawa-nobuyuki-gt@ynu.ac.jp).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSSC.2020.3041338>.

Digital Object Identifier 10.1109/JSSC.2020.3041338

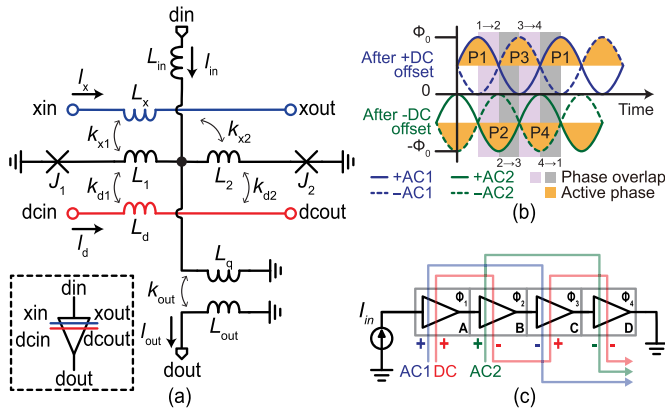


Fig. 1. AQFP logic primitive. (a) Inset: Shows the schematic of the AQFP with its corresponding symbol. (b) Shows how two ac sinusoidal power-clocks with a dc offset can create the four-phase clock to drive an AQFP circuit. (c) Shows how the power-clock lines and dc offset meander throughout a simple four-stage AQFP shift register.

AQFP dissipates, and the output current goes to zero. This is the reset or null state of the AQFP. This structure effectively behaves as a clocked buffer. A clocked inverter is created by inverting the coupling coefficient between L_q and L_{out} . A constant 1/0 cell is created by adding physical asymmetry to the L_1 - L_2 inductors of the dc-SQUID. By combining these minimal variations of the AQFP, majority-based Boolean logic cells can be created [14], [15].

Our AQFP circuits are clocked by a four-phase current-based ac power-clock. Two sinusoidal ac sources, AC1 and AC2, plus a dc offset create the four-phases [14], as shown in Fig. 1(b). The power-clock excitation lines are typically routed in a meandering fashion, as shown in Fig. 1(c), for a simple four-stage shift register composed of four AQFP buffers connected in series. The meandering directions create positive and negative versions of the ac power-clocks and the dc offset. AC1 and AC2 are in quadrature (90° out of phase) and when, respectively, combined with the positive and negative dc offset, all four phases of the clock can be generated. The nominal peak ac amplitude is approximately 0.9 mA and the nominal dc offset is 1.2 mA. These power-clock lines never physically “sink” into the AQFP but rather provide an excitation current through inductive coupling.

The simulation of a four-stage AQFP shift register is shown in Fig. 2 which was performed using the freely available superconductor analog circuit simulator Josephson Simulator (JSIM) [16]. The four power-clock phases are annotated on the ac signals as observed directly from the ac sources before they are shifted by the dc offsets in the AQFP circuits. Input data are provided in the form of a positive or negative current which the first-stage AQFP samples when the first phase of the power-clock arrives. Upon arrival of the power-clock, the AQFP generates a corresponding positive or negative output current, which, in turn, is sampled by the next AQFP in series. As the power-clock ramps down, so does the AQFP output and eventually, it goes into the reset or null state where the output current is zero. Then, on the next power-clock cycle, the process repeats. Because of the overlap between active clock phases, the AQFP can transfer data from one phase to the next, as shown in the simulation waveform. The overlaps between

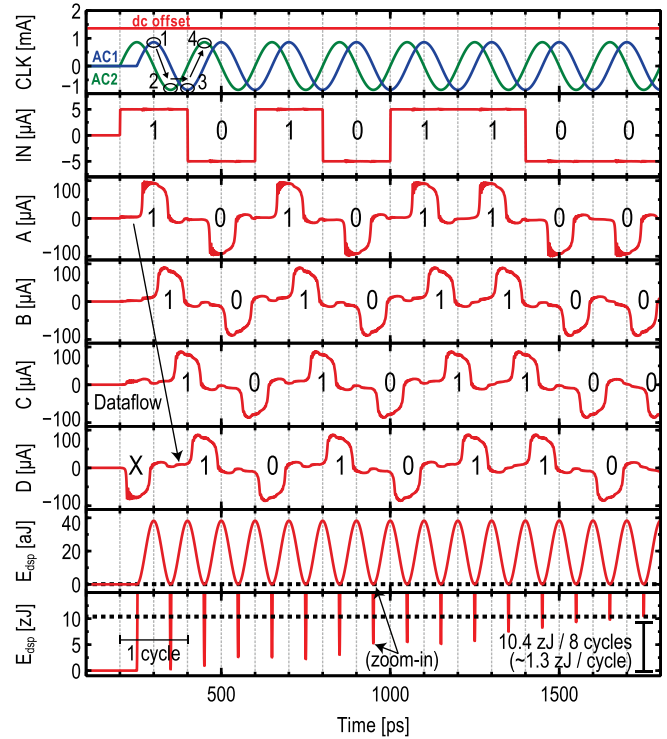


Fig. 2. Analog simulation of a four-stage AQFP shift register using JSIM. The bottom two traces show the switching energy profile of the first stage buffer. The “X” observed in the D trace is a random output event as the shift register has yet to be initialized with valid data.

active phases are visualized in Fig. 1(b). It is only during these overlaps between adjacent phases that data can transfer from the earlier phase of the two to the later phase of the two, as the earlier phase provides a signal output that the later phase can sample to produce its own output. In this clocking approach, we also cannot skip phases. For example, in Fig. 1(b), Phase 1 (P1) has a 180° phase difference with P3; therefore, there is no way for the AQFP circuits to properly sample outputs between these phases. Data also cannot be transferred from P1 to P4 because P4 is already active before P1 starts; thus, data can only be transferred from P4 to P1 (not vice versa). In short, data propagates from one phase to the next with each phase consisting of only one logic stage. Because we are using four-phase clocking, we can only propagate through four stages of logic in a single cycle. This severely limits how much logic we can perform in a single cycle, and it is what drove the design decisions discussed in Section II.

Also shown in the waveform is the total dissipated energy of the first buffer which exhibits a typical adiabatic switching profile where energy is dissipated and then recovered. The switching energy trace is obtained by performing the rolling integral shown in (1), where $V_x(t)$ is the voltage across L_x , and $I_x(t)$ is the current through L_x

$$E_{disp} = \int V_x(t)I_x(t)dt. \quad (1)$$

An example of an AQFP-based full-adder using majority logic gates is shown in Fig. 3 as a schematic (a) and physical layout (b). The majority gates are a mix-and-match of normal AQFP buffers, inverters, and constant cells tied together in parallel. Gates are arranged in logic rows where each row is

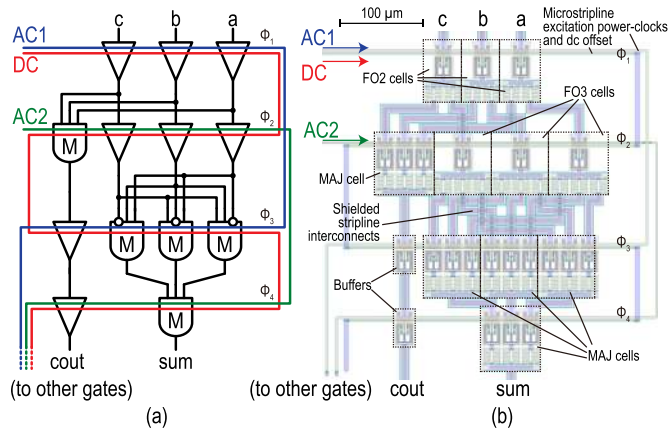


Fig. 3. Example of a gate-level AQFP full-adder based on majority logic. (a) Shows the logic schematic and clocking scheme. (b) Shows the physical layout.

clocked by one phase of the power-clock. Data propagate from one row to the next row, phase by phase. As previously mentioned, it is not possible to skip phases or propagate through multiple gates within the same phase. Furthermore, all connections are one-to-one, meaning we have no notion of tri-state buffers that can drive the same interconnect line, and gates do not have passive fan-out. Active fan-out cells up to FO3 are used to split data signals. The power-clock lines are made of superconductor microstriplines whereas the gate-to-gate interconnect are made of shielded striplines. Due to the parasitic inductances of the interconnect, the AQFP cells have a limited driving distance of up to only 1 mm [15], after which another buffer must be inserted as a repeater to re-amplify the signal.

As previously mentioned, there are a number of other JJ-based superconductor logic families (ERSFQ, eSFQ, RQL, LR-biased RSFQ, and LV-RSFQ). They all fall under the class of SFQ logic families where the data tokens of logic “1” or “0” are represented as the presence or absence of an SFQ voltage pulse, respectively [17]. AQFP logic differs from these SFQ logic families in that data is not transmitted as SFQ voltage pulses but as positive (logic “1”) or negative (logic “0”) current with a reset state (zero current). But the most important difference is that AQFP logic operates adiabatically which limits the clock rate to around 10 GHz in order to remain in the adiabatic regime. The SFQ logic families are non-adiabatic, which means they are capable of running at extremely fast clock rates as high as 770 GHz [18] at the cost of much higher switching energy. The switching energy of AQFP is so low that its modest clock rates put its energy-delay-product (EDP) at about 2–3 orders of magnitude lower than that of SFQ logic [19].

This work describes a proof-of-concept microprocessor as a first step toward demonstrating adiabatic data processing and storage using practical circuits. We call this microprocessor: MANA, Monolithic Adiabatic iNtegration Architecture. The efforts described herein build upon previous work, including the use of an established AQFP cell library based on unshunted JJs for a four-layer Nb/AlO_x/Nb superconductor IC process with a superconducting critical current density of 10 kA/cm² manufactured by the National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan [14]. The

fabrication process is referred to as the AIST high-speed standard process (HSTP). We also use a semi-custom AQFP design environment described in [20] which consists of a rudimentary logic synthesis framework [21], and a genetic-algorithm-based place and route tool [22]. Some of the components of MANA are variations of previously demonstrated circuits, including a 16 × 1 b register file (RF) in [23], 8–b Kogge–Stone adders in [11], [24], and a 4–b data shifter in [22]. Preliminary low-frequency demonstrations of actual MANA units have been reported for the ALU and shifter in [20].

The remaining sections of this article are organized as follows: Section II describes the architecture of MANA and how it was developed specifically for AQFP logic. Section III describes the implementation of MANA and the general challenges of designing the microprocessor. Section IV shows the experimental demonstrations of the MANA chip and a high-speed breakout chip of the execution components of MANA. Section V makes a brief comparison with other adiabatic circuits and microprocessors and discusses the challenges in moving the technology forward. Section VI estimates the performance of a refrigerated supercomputer implemented in AQFP logic. Section VII concludes this article.

II. ARCHITECTURE

We developed the microarchitecture of MANA as a demonstration vehicle to show that AQFP logic is capable of performing computation, including the processing of logic and the storage of data into AQFP-based memory structures, all within a single technology, single-logic family, and single chip. Optimizing raw performance in terms of throughput and instructions per cycle (IPC) was not our priority. It was specifically catered to AQFP logic to take into account the following.

- 1) *Four-Phase Clocking*: Data can only propagate through four-levels of logic per clock cycle; therefore, program control must be very simple to achieve a peak IPC of 1.
- 2) *Lack of System Integration Tools*: Integration of components at the chip-level is done by hand; therefore, a design with a large number of control signals broadcasted all over the chip is not feasible. Instead, the architecture should be simple so that data and control signals have a straightforward dataflow but still perform a meaningful computation.

To this end, we combined aspects of conventional RISC design [25] with some concepts taken from dataflow architectures [26], [27]. The microarchitecture diagram of MANA is shown in Fig. 4. Table I lists the core set of supported instructions. MANA operates on 4–b data words via 16–b instruction words whose formats are shown in the inset of Fig. 4. The functional organization of MANA consists of an instruction buffer-decode-issue (IDI) stage, an RF with eXternal I/O access stage (RFX), an EXecution stage (EX), and Writeback (WB) stage. The components of each stage are color coded in Fig. 4.

The IDI components are a 4 × 16 b instruction buffer (IB), 16–b instruction decoder (ID), an 8–b program counter (PC) and instruction fetch (IFETCH) controller. We do not include

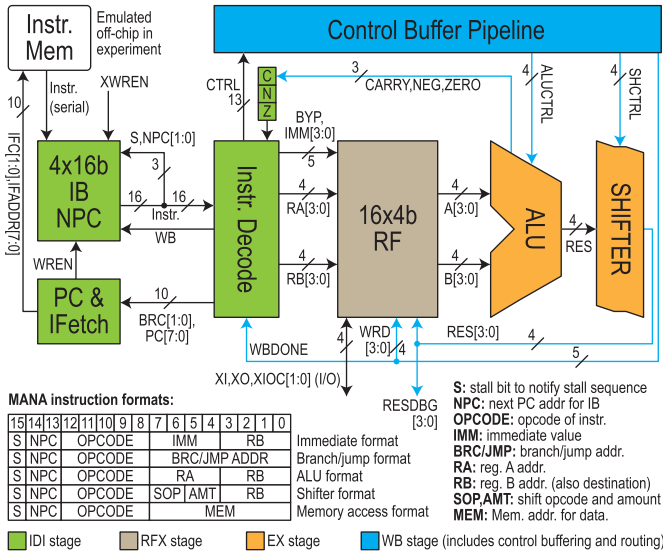


Fig. 4. Organization of MANA and its instruction formats.

a large instruction memory due to its design complexity; therefore, we only emulate its behavior in the experiment. The small IB is sufficient to perform key demonstrations of the implemented processor. MANA instructions have an NPC (Next PC) field which corresponds to which of the other three instructions in the IB will be executed next. If ID determines that the NPC field refers to the current instruction (e.g., the instruction in the 11_2 slot of IB has an NPC of 11_2), this notifies IFETCH to increment PC and request the next block of four instructions to be serially loaded into IB. A branch instruction whose branch address is outside the block of four instructions would also notify IFETCH. This eliminates the need to calculate the NPC within the block of four instructions. This concept was inspired by dataflow architectures, where instructions have an understanding of what would be executed next without the need of a PC. Furthermore, MANA instructions also have an S bit, also known as the stall bit. This S bit is encoded by the compiler to notify the IDI stage on whether the next instruction has a dependence on the current instruction. The simple detection for this bit makes it very easy for the IDI stage to decide whether it must stall the processor to resolve both data and control hazards in a single cycle. When an S bit is detected, it prevents the next instruction from being fetched from the IB, and it propagates through the pipeline with the current instruction in flight. It returns to the IDI stage during writeback as an acknowledgment signal to notify that the next instruction can finally be issued.

Thanks to the S bit and NPC field, program control decisions are made at the very start of the pipeline. Once those program control decisions are determined, the rest of the instruction is decoded by ID, and all necessary control signals are generated and buffered appropriately. This actually makes MANA a two-stage pipeline from a pure computer architecture point of view. The first stage checks on the last S bit status and performs either a hardware stall or enables the fetching of the next instruction. This first stage is very simple and can be completed within four stages of logic (one clock cycle in four-phase clocking) so that a peak IPC of 1 can be obtained. The second stage consists of everything else in the processor,

TABLE I
MANA CORE INSTRUCTION SET

Instr.	Format	Description
nop	N/A	No operation, used for stalling
halt	N/A	End program
jmp	Branch	Absolute jump to [JMP ADDR]
bneg	Branch	Branch to [BRC ADDR] if NEG flag is set
bnneg	Branch	Branch to [BRC ADDR] if NEG flag is not set
beq	Branch	Branch to [BRC ADDR] if EQ flag is set
bneq	Branch	Branch to [BRC ADDR] if EQ flag is not set
andi	Immediate	AND with immediate $R[RB]=R[RB]\&\&[IMM]$
li	Immediate	Load immediate $R[RB]=[IMM]$
sr1	Shift	Shift right logical: $R[RB]=R[RB] \gg [AMT]$
sra	Shift	Shift right arithmetic: $R[RB]=R[RB] \ggg [AMT]$
sll	Shift	Shift left logical: $R[RB]=R[RB] \ll [AMT]$
s1a	Shift	Shift left arithmetic (s1l): $R[RB]=R[RB] \lll [AMT]$
add	ALU	Add: $R[RB]=R[RA]+R[RB]$
addnw	ALU	Add with no write: $R[0]=R[RA]+R[RB]$
sub	ALU	Subtract: $R[RB]=R[RA]+(\sim R[RB]+1)$
subnw	ALU	Subtract with no write: $R[0]=R[RA]+(\sim R[RB]+1)$
xor	ALU	XOR: $R[RB]=R[RA] \oplus R[RB]$
xnor	ALU	XNOR: $R[RB]=\sim(R[RA] \oplus R[RB])$
and	ALU	AND: $R[RB]=R[RA] \& R[RB]$
andnb	ALU	AND not B: $R[RB]=R[RA] \& \sim R[RB]$
or	ALU	OR: $R[RB]=R[RA] \mid R[RB]$
ornb	ALU	OR not B: $R[RB]=R[RA] \mid \sim R[RB]$
lfm	Memory	Load from memory: $R[14, 15]=MEM[MADDR][Hi, Lo]$
wtm	Memory	Write to memory: $MEM[MADDR]=R[14]:R[15]$

including the enabled fetching of the instruction from IB. It has a total latency of 26 cycles. The gate-level clocking nature of AQFP logic allows the multi-cycle second stage of MANA to accommodate 26 instructions in-flight although MANA will only issue a maximum of four instructions successively due to the small four-instruction-word IB. To keep the architecture as simple as possible, we do not support other pipeline hazard resolution mechanisms, such as data forwarding. This kept the number of feedback loops small, and thus, kept manual timing closure feasible.

The RFX stage is a two-read/one-write 16×4 b RF with registers \$14 and \$15 to hold I/O data, and \$0 as a constant zero register. External I/O data are transferred in and out of \$14 and \$15 serially such that \$14 contains the higher nibble of the memory byte data and \$15 contains the lower nibble. Control flags are available to notify the processor that data in \$14 and \$15 are valid data from I/O and likewise, a similar set is also available to notify I/O devices that data \$14 and \$15 are valid from the processor.

The EX stage has a 4-b integer ALU and 4-b data shifter unit. The units are connected in series so data must propagate through both units regardless of the instruction.

The WB stage routes processed data back to the RFX stage to the appropriate destination register. This stage also includes the necessary buffering of control signals to be tapped OFF when needed as the instruction flows through the pipeline and clocked signal repeaters to transmit data across the long distances of the WB path.

III. IMPLEMENTATION

A. Design Environment

Before implementing MANA, we developed a semi-custom design environment [20] which features a four-phase AQFP cell library [14] in a Cadence design environment that has been

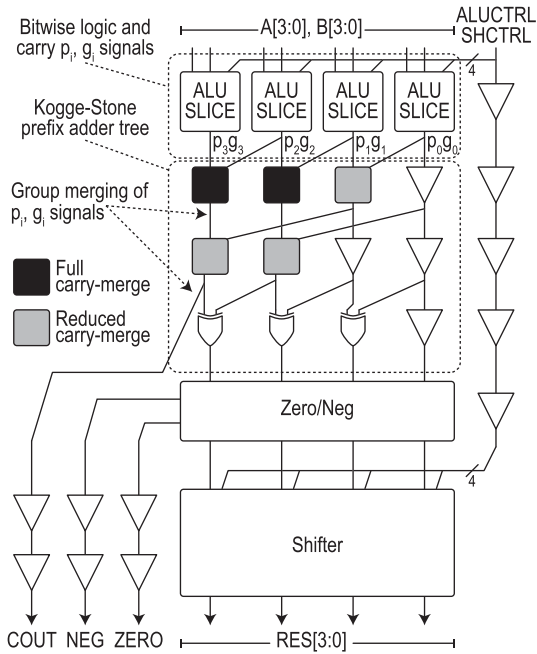


Fig. 5. Block diagram of the EX stage of MANA featuring a Kogge–Stone-based ALU and a synthesized data shifter.

augmented with scripts and external programs to do AQFP-based combinational logic synthesis [21] and component-level place and route using the genetic algorithm and left-edge channel routing [22], [28]. Place and route tools operate on adjacent logic rows and are aware of the limited driving ability of AQFP logic cells. Together, they attempt to make all interconnect in the routing channel to be less than 1 mm. If it is not possible, buffers are inserted as signal repeaters automatically. The design environment also includes gate-level models specified in SystemVerilog [29] with timing information extracted from freely available superconductor circuit simulators, such as JSIM and Josephson Simulator (JoSIM) [16], [30], [31]. The AQFP cells require accurate extraction of the 3-D inductors and their mutual coupling factors, all of which can be extracted through a tool called InductEx [32]. The post-extraction simulation results align well with the characterization experiments of fabricated cell tests of our library [14], [15].

B. Combinational Logic Design

We applied our combinational logic synthesis along with place and route on the ID unit of the IDI stage and the data shifter unit of the EX stage. The shifter was our first example in which we showed a full demonstration from a behavioral Verilog description to a working circuit in experiments [20], [22]. In this work, the ID unit was also synthesized using a subset of an RTL description of MANA. It is straightforward decoding of input signals, mainly the 16–b instruction word, into control signals and source/destination addresses.

We designed an integer ALU by hand using a majority logic-based parallel prefix carry look-ahead Kogge–Stone adder as its structural foundation composed of majority-based carry-merge blocks (black and gray cells) [24]. As shown in Fig. 5 via the ALU slices, we modified the first stage of the adder normally used to generate bitwise carry propagate p_i and

generate g_i signals so that bitwise logic operations can also be performed in-place. When performing addition, ALU slices will produce the appropriate p_i and g_i signals as expected in a Kogge–Stone adder. When performing subtraction, the ALU slices will generate the two’s complement of operand B by inverting B in each slice and adding “1” in the least significant bit by introducing a carry-in. In the case of a logic operation, we produce the bitwise logic result in each of the ALU slices and push the result out as a p_i signal. No bitwise g_i signals are produced in this case so the logic result propagates through the adder tree unchanged. We first experimentally validated the design of the ALU and data shifter each in separate breakout chips at low frequency in [20], but in this work, we combined them together in a high-speed test chip described later. Furthermore, we integrated architectural glue logic into the ALU, such as the generation of carry (C), negative (N), and zero (Z) flags.

C. Memory

Memory structures are needed for the IDI and RFX stages of MANA. We created an adiabatic latch by connecting AQFP logic gates in a short four-phase feedback loop similar to the academic latches composed of NAND or NOR gates [33]. This gate-level implementation of an AQFP latch has an enable input which is asserted on the same clock phase as the input data when we want to write into the latch.

Our logic synthesis and place and route framework currently do not support circuits with memory and feedback structures. Such structures had to be designed by hand. The 16×4 b RF unit of the RFX stage is the largest and most complicated component of the entire MANA processor. The design is based on a previously demonstrated 16×1 b RF unit [23]. For the development of MANA, efforts were made to establish a compact bit-slice design methodology by hand to systematically scale up the design to 4–b data words. A block diagram of RF is shown in Fig. 6 with the read and write decoders omitted. The D cells in the diagram are gate-implemented D latches serving as the memory cells whereas DS cells are modified D cells but with additional “serial ctrl” interfaces to enable separate serial read and write with external I/O. AQFP logic has limited fan-out with discrete active fan-out cells up to only FO3. We also cannot use shared lines, as CMOS does with tri-state buffer access of an output bus bit line. The “rd ctrl” components in Fig. 6 that control the enabling of a word line output are implemented as AND gates whose output is then merged into the OR-trees for each bit column inside the “bit merge tree” block. The output of the “bit merge tree” forms one bit of the 4–b data output. These one-to-one connections contribute to the large area of the RF unit ($4.7 \text{ mm} \times 6.6 \text{ mm}$).

Other memory structures in MANA, such as the IB and PC in the IDI stage, follow the same design concepts as the RF implementation but are actually much simpler. The IB has only four words, one read port, and is loaded serially. The PC is just a set of eight D latches updated by ID after all instructions in IB have finished execution.

D. Clock Distribution

As illustrated in the full-adder example in Fig. 3, we generally distribute the power-clock network of AQFP logic circuits

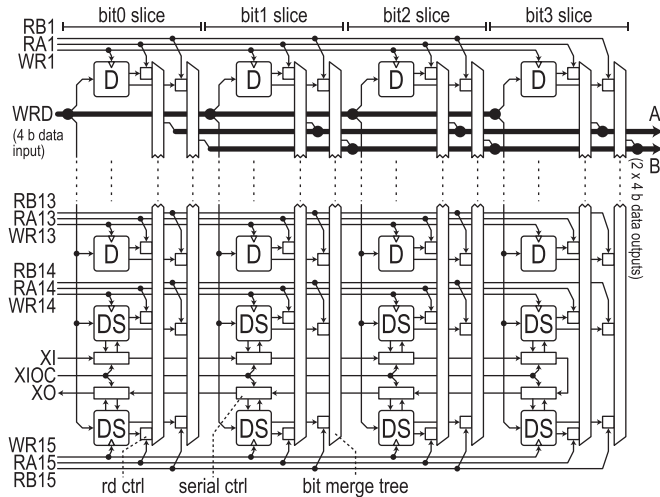


Fig. 6. Block diagram of the 16×4 b RF unit with integrated I/O serial interfaces. The two read decoders and one write decoder are not shown. A bit-slice methodology was used. Note that data in \$0 are always 0000₂.

in a meandering fashion from one logic row to the next. This is perfectly fine for small circuits, but if we increase the scale to that of the RF unit, we start to see the impact of clock skew because the microstriplines have an effective signal velocity. The AIST HSTP fabrication process used in this work has four Nb superconductor layers. The Nb layer and dielectric thicknesses are specified in [14]. We made a first-order extraction of the microstripline velocity to be $161 \mu\text{m/ps}$ or a delay of 6.21 ps/mm in [20]. Fig. 7(a) shows an illustration of round trip clock skew which is governed by the total length of the AC1 line traveling from the start of the phase 1 logic row to the end of the phase 3 logic row. Detailed timing analysis is beyond the scope of this article, but as a soft rule, we constrain this round trip length to be no more than 5 mm. This results in a maximum clock skew of about 30 ps, which we expect to be tolerable for AQFP circuits operating at 5 GHz given that we also cannot drive data interconnects longer than 1 mm [20].

But to apply this constraint on to large circuits, we need to divide the power-clock using microwave power dividers so that we can create multiple local power-clock networks in parallel with each one abiding by the round trip constraint, as shown in Fig. 7(b). Wilkinson power dividers [34] would be a good choice for AQFP circuit designs deemed ready for operating at high clock rates. They are lossless when their output ports have properly matched impedances, but the narrow bandwidth of these dividers does not allow us to verify the circuit at lower frequencies making experimentation more difficult. Because this is the first demonstration, we opted for a resistor-based power divider network which we can operate over a large range of frequencies. We applied it to the RF unit via a 1-to-4 power divider with a $50\text{-}\Omega$ resistor placed in series of every output of each 1-to-2 dividing stage. This implies that we need $4\times$ the nominal ac excitation current amplitude to operate the circuit ($4 \times 0.9 \text{ mA}$). The round trip length of each meandering local power-clock network is about 3 mm. Because the divider is resistor-based, additional power will be dissipated through the resistors. The equivalent resistance of

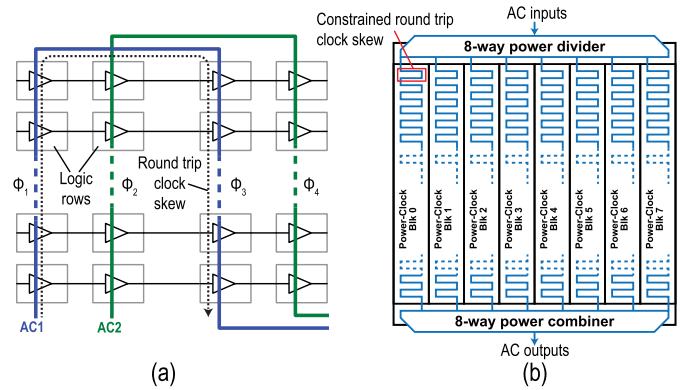


Fig. 7. (a) Illustration of the round trip clock skew that we must constrain to approximately 30 ps (or a length of about 5 mm) for high-speed design. (b) One approach to apply the round trip clock skew constraint at the chip-level using power dividers.

the entire 1-to-4 divider network is 37.5Ω . We apply an ac sinusoidal voltage with a peak amplitude of 135 mV so that ac power-clock currents are at the nominal peak amplitudes of 0.9 mA in each of the four branches of the 1-to-4 power divider. The resulting dissipated power from the divider is $P = (135 \text{ mV}/\sqrt{2})^2/37.5 \Omega = 243 \mu\text{W}$, which is about three orders of magnitude larger than the power dissipated from the AQFP circuits in MANA. When using a well-designed Wilkinson power divider, we expect very little to no power dissipation from the power-clock network on chip.

The area of the resistor-based power divider is almost negligible as they can be designed as a straightforward compact binary tree with resistors in series at each dividing stage. The Wilkinson power divider, on the other hand, requires a substantial area to accommodate microstripline meanders for achieving the desired impedances. A 1-to-4 Wilkinson power divider may take up an area of $1 \text{ mm} \times 2 \text{ mm}$ [35] and an additional copy of the divider is needed to recombine the divided power-clocks, as shown in Fig. 7(b). A fabrication process with more superconductor layers can help alleviate this challenge.

E. Component Integration

While our design environment allows us to generate stand-alone combinational processor components, we do not have an automated way to integrate them together at the chip-level. This was done by hand. All processor components were designed independently without any real consideration of how they will be connected to each other. Each component was designed with the assumption that its first stage of logic is clocked by the first phase of the ac power-clock with the corresponding ac lines physically coming from the top-left of each block. When connecting components together, manual buffer insertion was necessary for two reasons: 1) to synchronize the output signals of the transmitting component such that they arrive at the starting phase of the receiving component and 2) to re-amplify signals traveling across long distances ($>1 \text{ mm}$).

In addition, the meandering local power-clock network of the individual components was designed without consideration of additional buffering of signals adjacent to them, such as

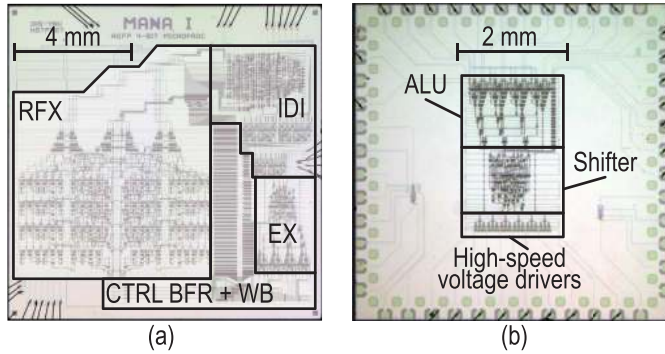


Fig. 8. Microphotographs of the prototype chips. (a) 1 cm \times 1 cm MANA chip. (b) 7 mm \times 7 mm EX chip.

TABLE II
BREAKDOWN OF MANA

Stage	Description	Total JJs	Latency (cycles)	fJ/op ¹
IDI	Instruction buffer, Decode, Issue	5596	8	7.834
RFX	Register File with eXternal I/O	8142	8	11.399
EX	EXecution stage (ALU-shifter)	2238	9	3.133
WB	Writeback, ctrl. buffering, routing	5484	2	7.678
Total:		21460	27	30.044

¹ At 5 GHz and includes cooling overhead coefficient of 1000 \times . It does not include dissipation from the resistor-based power divider which can be neglected in the case of a Wilkinson power divider.

control signals that are simply passing through for the next stage. It was necessary to stretch or even completely redesign (if stretching resulted in too large of a round trip clock skew) the local power-clock network to accommodate those extra control signals.

Efforts to improve and automate this aspect of the design flow through electronic design automation (EDA) tools are underway (see Section V).

F. Taped-Out Chips

We manufactured our chips using the AIST HSTP 10kA/cm² Nb/AIO_x/Nb superconductor IC fabrication process [14]. In this work, we taped out two designs. The first chip is a low-frequency design of the MANA processor. We underestimated the overall size of the design which resulted in having to use a larger 1 cm \times 1 cm chip. At the time of this writing, we, unfortunately, did not have a high-speed experimental probe that can support this chip size; therefore, our experiments for the MANA chip are limited to 100 kHz using our low-frequency probes. We included a debug output port “RESDBG” which is tapped from the writeback data right after the EX stage so that we can observe the state of the microprocessor step-by-step. A microphotograph of the MANA chip is shown in Fig. 8(a). The active circuit core consists of 21460 JJs and has an area of 8.5 mm \times 9.5 mm. Table II shows a stage-by-stage breakdown of the MANA processor.

To demonstrate that at least part of the MANA processor is capable of high-speed operation, we produced a second design which is a breakout chip of the EX stage of MANA. It includes the ALU and data shifter connected together along with the

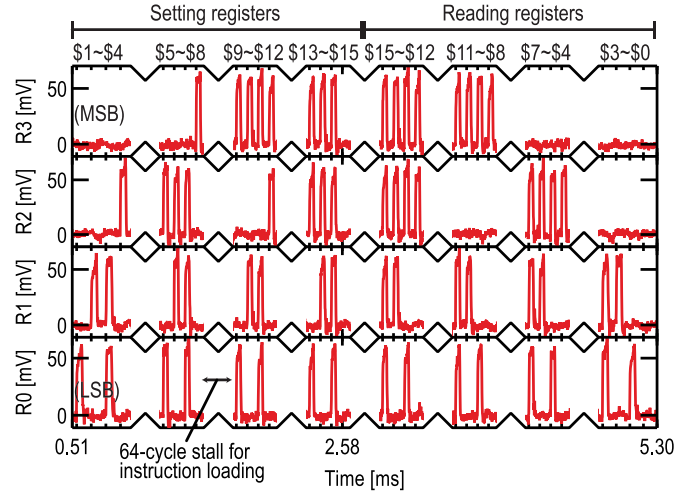


Fig. 9. Measured output trace of the initial smoke test for MANA where registers are incrementally set and readout. A 64-cycle stall is observed between bursts of three–four instructions to load the next three–four instructions.

buffering of their respective control signals, but we omitted the generation of the ALU flags. This EX chip was designed primarily for high-speed demonstration (> 1 GHz) so we made sure that the round trip clock length was less than 5 mm, and we interfaced the outputs of the EX stage with high-speed dc-SQUID stack voltage drivers, so we can observe the high-speed output [36]. The design is placed on a 7 mm \times 7 mm chip which is a compatible size for our high-speed experimental probe. The chip contains a frame of 48 pads to interface with the probe. A microphotograph of the EX chip is shown in Fig. 8(b). The active circuit core consists of 2076 JJs and has an area of 2 mm \times 3.5 mm.

Both designs were taped out, and we experimentally demonstrated their operation across multiple wafers and chips as detailed in Section IV.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

We measured each chip using an immersion probe lowered into a liquid helium Dewar to achieve a 4.2 K cryogenic temperature. We surrounded the chip carrier end of the probe with two-layers of Mu-metal shielding. A low-speed probe suitable for testing 1 cm \times 1 cm chips via wire-bonding was used for the MANA chip, whereas a high-speed probe that accepts 7 \times 7 mm chips via a 48-pin pad frame was used for the EX chip. We used a function generator to create two ac sinusoidal sources (AC1 and AC2) in quadrature (90° out of phase). A dc offset is coupled to both ac excitation lines and dc offset are 50- Ω terminated at room temperature. The nominal amplitude of AC1 and AC2 are both 0.9 mA. The nominal dc offset is 1.2 mA. Because the MANA chip uses a 1-to-4 power divider, we applied approximately 4 \times the nominal values of AC1, AC2, and the dc offset. We used a data pattern generator to create a +10 μ A logic “1” and a –10 μ A logic “0.” An on-chip dc-SQUID stack voltage driver interface [36] with additional room temperature amplification provides output readout. The signal produced from the output

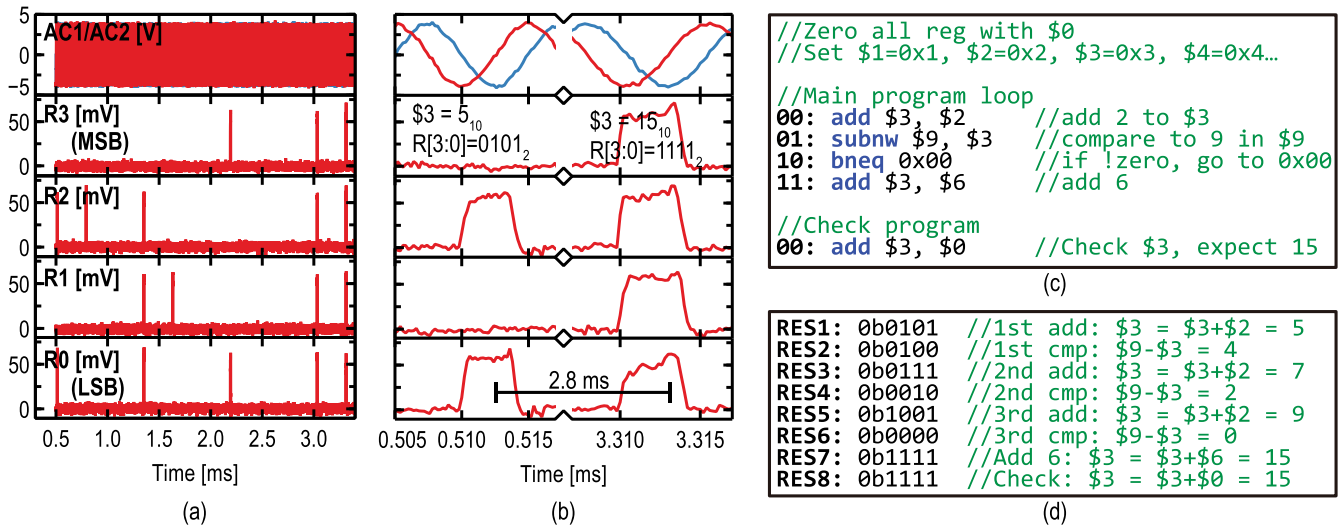


Fig. 10. Measured outputs from the RESDBG port for the first program demo of MANA using the add instruction. (a) Shows the entire measured waveform from 0.5 to 3.5 ms. (b) Shows a zoomed-in view of the first and last result. (c) Lists the actual test program. (d) Lists the expected output sequence.

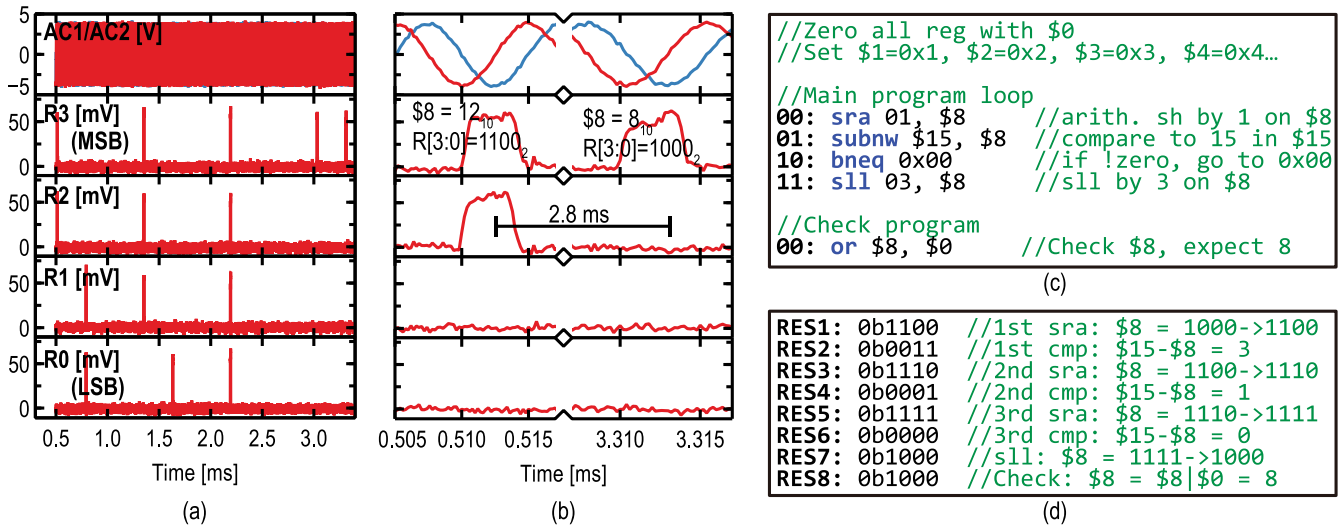


Fig. 11. Measured outputs from the RESDBG port for the second program demo of MANA using the sra and sll instructions. (a) Shows the entire measured waveform from 0.5 to 3.5 ms. (b) Shows a zoomed-in view of the first and last result. (c) Lists the actual test program. (d) Lists the expected output sequence.

interface is a unipolar return-to-zero (RTZ) signal in which a logic “1” is represented by a positive output voltage for a duration that is proportional to the experimental clock period before returning to zero. When the output is a logic “0,” no activity is observed on the output trace.

B. MANA Chip Test

During the MANA chip test, we operate the circuit only at low speed (100 kHz) and 16-b instruction words were serially loaded into the instruction buffer 1b/cycle. We first zeroed all registers in the RF by ANDing each register with \$0. Next, we used the load-immediate instruction to write each register with its register number (e.g., \$1 := 0001₂; \$2 := 0010₂; ... \$0 is always fixed to 0000₂) and then we read each register out to confirm their values starting from \$15 down to \$0 by observing the RESDBG port. This serves as a smoke test to check that at least the RF unit (the most complex component) is operating before proceeding to other tests. The setting and reading of the registers are shown in Fig. 9, where

R_3 is the most significant bit (MSB) of the output, and R_0 is the least significant bit (LSB) of the output. This smoke test also shows that we can achieve an IPC of 1 for up to four-instruction bursts, which is limited by the size of IB. A 64-cycle stall between bursts is used to serially load the next three–four instructions into the IB unit.

We demonstrated two short four-instruction programs for MANA. The first program loaded into the IB incremented \$3 by the value of 2 stored in \$2 until the contents of \$3 equals 9. When \$3 equals 9, we add the value of 6 stored in \$6 to \$3 so that it is now 15. We checked the contents of \$3 by adding it with \$0, and observed the resulting output on the RESDBG output port. The experimental waveform is shown in Fig. 10(a) with a zoomed-in view on the first and last results in Fig. 10(b). The test program and output sequence are shown in Fig. 10(c) and (d), respectively. Note that each instruction in the program has either a data or control hazard. To resolve these hazards, a 27-cycle hardware stall is performed by MANA after every instruction. This can be

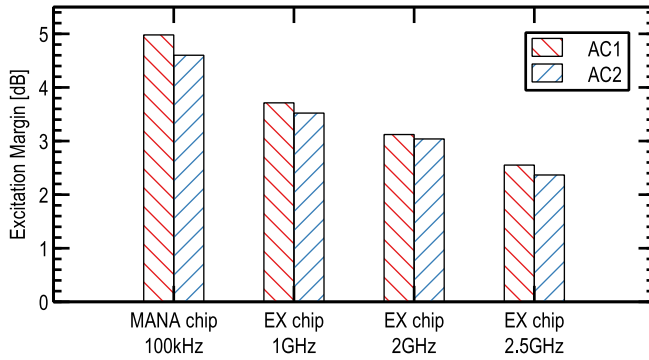


Fig. 12. Measured excitation power-clock margins for the MANA chip and EX chip averaged across fully functional chips capable of operating at the given clock rates.

seen by the large gaps of idle activity between the outputs of the RESDBG port shown in Fig. 10(a). Some gaps might appear even larger because the `bneg` instruction produces no observable output on RESDBG. In addition, the RESDBG outputs show all ALU calculations even though some are not necessarily written to the RF unit as is the case for the `subnw` (`compare`) instructions.

After resetting the registers back to the initial state, we loaded the second program, which tests a few shifter operations. It starts by performing shift right arithmetic (`sra`) by 1 on \$8 until it is 1111_2 . When it is 1111_2 , it performs a shift left logic (`sll`) by 3 so that \$8 should now hold the value of 8. Finally, we checked the contents of \$8 by ORing it with \$0. The corresponding waveforms, test program, and output sequence are all shown in Fig. 11.

Both programs demonstrate successful RF R/W access, execution of a few arithmetic/logic operations, hardware stalling, and program branching. Furthermore, we were able to replicate these results on a total of five out of six chips across two wafers that we have tested, as shown in Table III, with one chip failing to produce any output, possibly due to a poor connection between the immersion probe and chip. We also varied the amplitudes of the excitation clock amplitudes of AC1 and AC2 until we induced bit errors. The operating margins of AC1 and AC2 shown in Fig. 12 are reasonably wide at 5.0 and 4.6 dB, respectively, when averaged across all fully functional chips.

C. EX Chip Test

For the EX chip, we first confirmed all ALU and shifter operations at 100 kHz, as shown in Fig. 13. The test starts by setting operands A and B such that they cover all input combinations across the bits for the logic operators. We repeated the test with the bits reversed and swapped between A and B. Then, we test a few random operands for addition and subtraction, including a set of test vectors that show carry propagating from the LSB to carry out (R_4). Lastly, we fixed the B operand to 0111_2 and 1000_2 with the A operand fixed to 0000_2 to test the shift operations. All shift operations and shift amount values have been tested.

Next, we perform the high-speed test. Our setup is limited to a single data input from an Agilent N4906B BERT with

TABLE III
EXPERIMENTAL YIELD

Wafer	Chip 1	Chip 2	Chip 3	I_c^1	Working Chips
MANA-W1	100 kHz ²	X	100 kHz	110.3%	5 of 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 of 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%	

¹‘U’ denotes unstable or partial operation, and ‘X’ denotes no output or oscillating output.

² Measured I_c over designed I_c of a $1.1\ \mu\text{m}^2$ test JJ averaged from two process chips on the same wafer.

³ Note that MANA chips were tested only up to 100 kHz.

single output monitoring. Our high-speed test pattern consists of the critical carry propagation test where control signals are set for addition, operand $A_{3:0}$ is fixed to 1111_2 , $B_{3:1}$ is fixed to 000_2 , and B_0 is toggled at high-speed as a pseudorandom input from the BERT. The carry-out R_4 and sum $R_{3:0}$ outputs were observed on a high-speed oscilloscope (Fig. 14) with carry out R_4 also being monitored by the BERT. When B_0 is 1, R_4 is 1 and $R_{3:0}$ is 0000_2 . When B_0 is 0, R_4 is 0 and $R_{3:0}$ is 1111_2 .

For our best sample, we confirmed functional operation at 1, 2, and 2.5 GHz and measured excitation margins of AC1 and AC2 to be 2.6 and 2.4 dB, respectively, at 2.5 GHz, as seen in Fig. 12. Beyond 2.5 GHz, we noted bit errors, and upon closer investigation, it revealed that 101_2 patterns did not fully return to zero. We believe this is not a problem with AQFP circuit itself but perhaps with the voltage driver interface or our experimental setup. In total, we tested 12 EX chips (three chips across four different wafers) and reproduced correct functionality on seven chips, as shown in Table III. For the high-speed test, the maximum operating frequency ranged from 1.2 to 2.5 GHz across all working chips. Some chips had unstable or oscillating outputs which we attribute to the superconducting phenomenon of flux trapping [37]. Such unwanted flux may be inadvertently trapped near sensitive circuits which may modulate the data signal currents along the interconnect striplines. Nonetheless, the results are indeed promising, and this is in spite of the fact that the critical current (I_c) of the fabricated JJs, which is governed by the effective physical area of the junction, had as much as a 12% difference from their designed values.

V. OUTLOOK

The experimental results show that we can indeed perform computation using superconductor AQFP logic. A number of other adiabatic circuits have been reported, including an experimentally demonstrated 16-b carry look-ahead adder (CLA) [38], a demonstrated 8-b DLX-based microprocessor [39], and an in-progress physical design of a 16-b microprocessor [40]. We compare these circuits in Table IV. Our superconductor AQFP circuits surpass all of them in terms of clock rates by approximately three orders of magnitude compared with most adiabatic circuits, with a $5\times$ advantage when compared with [40]. These much higher clock rates are still possible while also exhibiting higher energy efficiency by two

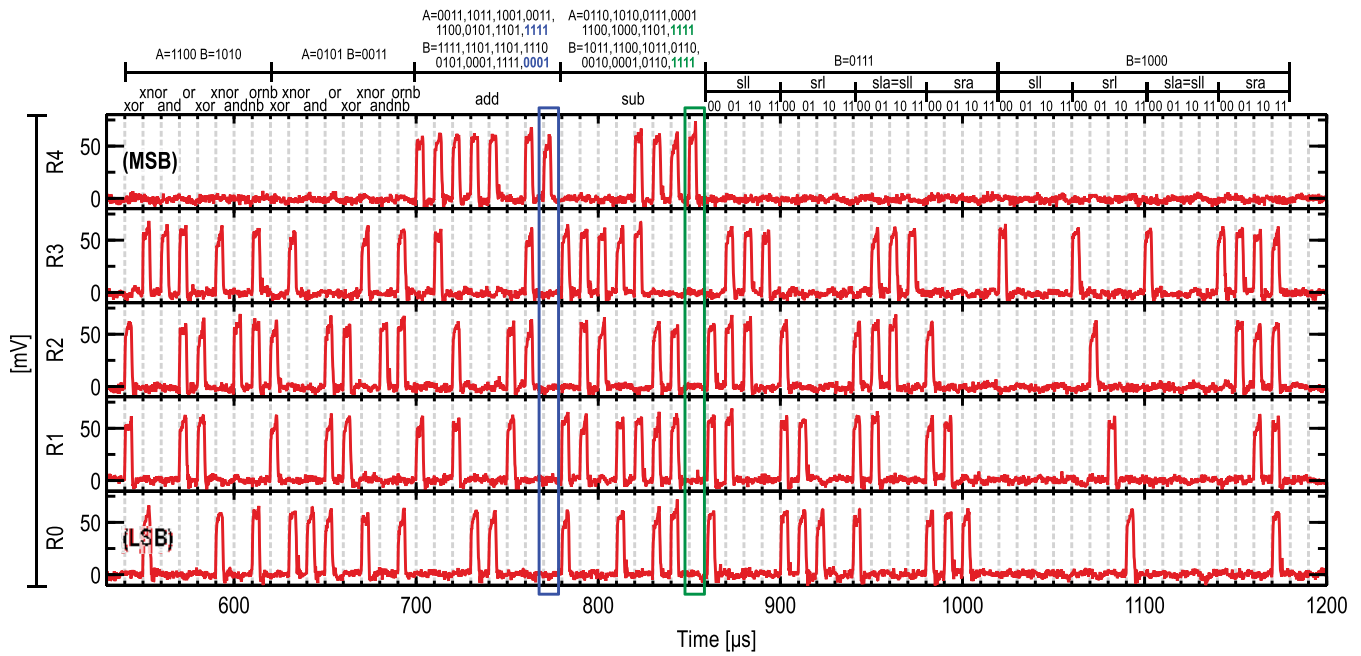


Fig. 13. Measured low-frequency waveform of the EX chip demonstrating logic operations, addition, subtraction, and all shifter operations.

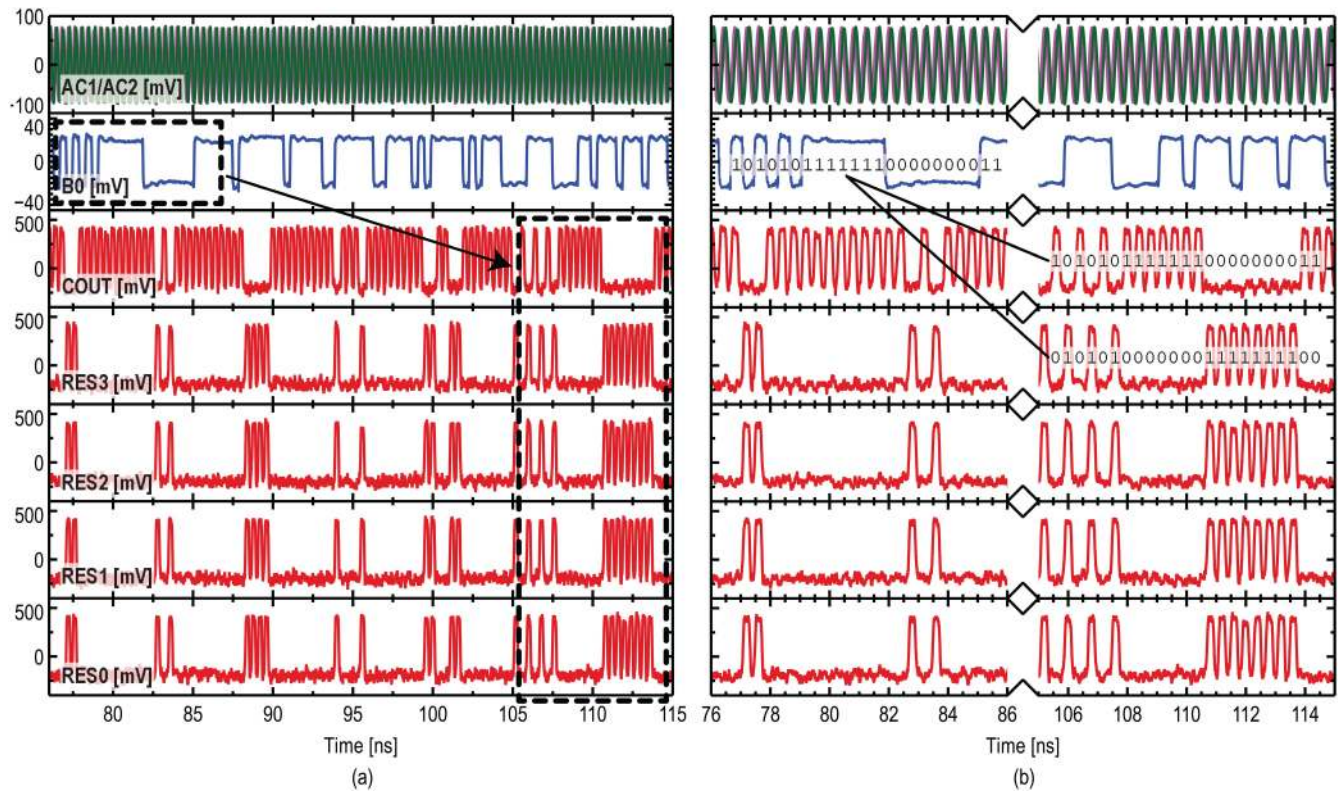


Fig. 14. Measured high-frequency waveform of the EX chip operating at 2.5 GHz, where B0 is the LSB of input operand B toggling at high-speed, and R4 (MSB, carry-out) through R0 (LSB) is the 5b output result. (a) Shows the full waveform and (b) zooms in on the outlined areas annotated in (a).

to three orders of magnitude despite the extra power needed to cool the circuits to cryogenic temperatures. However, the other adiabatic circuits all have a major practical advantage in that they are highly suitable for mobile applications. Superconductor digital circuits, in general, must showcase their advantages in the supercomputing, accelerator, and data center markets, where much more complex circuits are needed, in contrast to the simple architecture of MANA.

To move forward with this technology, improvement in the following areas are necessary: area efficiency, latency and clock distribution, flux trapping, and development of advanced EDA tools.

A. Area Efficiency

The chip photographs in Fig. 8 show that there is a lot of inefficient use of space. This is partly because of the design

TABLE IV
SUMMARY AND COMPARISON

	[38]	[39]	[40]	This work	This work
Circuit	16 b CLA	8 b DLX processor	16 b MIPS processor	4 b MANA processor	4 b ALU & shifter
Status	Tested	Tested	Layout in progress	Tested	Tested
Technology	0.8 μ m CMOS	0.18 μ m NMOS	90 nm CMOS	AQFP	AQFP
Clk. Rate	4 MHz (tested)	880 kHz (tested)	0.5 GHz (simulated)	100 kHz (tested)	2.5 GHz (tested)
Supply	2.5 V dc	1.8 V dc	1 V dc	4 mA ac	1 mA ac
Energy/op	4 pJ	8.5 pJ	3 fJ ¹	30 fJ ²	3.2 fJ ²

¹ Authors simulated 3 b shift register. AQFP equivalent is 8.4 aJ with cooling.

² Already includes cooling overhead coefficient of 1000 \times .

methodology of using only a single-clock phase per row of logic and the large logic cell size due to the output transformer of the AQFP. We have already shown promising work on using a more advanced eight-layer superconductor IC fabrication process provided by MIT Lincoln Laboratory, Lexington, MA, USA [41] to create a more compact AQFP cell library [42]. We can also potentially remove the large output transformer completely by using directly coupled QFP (DQFP) [43].

But being able to place any logic cell in any physical row regardless of which phase it is clocked would improve row utilization substantially. More investigation is needed to intelligently achieve this through more routing layers in the process or through careful design of the multi-phase clock network such that clock lines are sufficiently isolated from each other within the same row so that they do not interfere with each other or inadvertently excite a logic cell on multiple phases. Another area of investigation to help improve the area is a novel memory design that is compatible with AQFP logic instead of using gate-level implementations of latches.

B. Latency and Clock Distribution

The four-phase power-clocking scheme limits data propagation to only four stages of logic per cycle. This introduced a number of difficulties in trying to operate a microprocessor at the native clock frequency, as discussed in Section II. The limited amount of logic stages per cycle makes it extremely difficult or perhaps impossible to implement hazard resolution techniques to make the overall pipeline stall free. However, AQFP logic is actually an intrinsically fast logic family despite the adiabatic operation. If we consider other clocking schemes, such as delay line clocking [44] or N-phase power-dividing clocking [45], the latency can be reduced substantially by allowing more stages of logic to operate in a single-clock cycle. Such approaches will allow circuit designers and architects more flexibility to implement advanced pipelining techniques.

Because these novel clocking schemes still operate at the nominal 5-GHz clock rate used in four-phase clocking, the switching energy remains unchanged. But these clocking

schemes come with more complexity on how to distribute their networks at the chip-level and on how to close feedback loops which may require the use of non-adiabatic latches [46] as clock synchronizers.

C. Flux Trapping

As briefly touched upon in the experimental results, superconductor circuits are prone to be negatively affected by unwanted flux trapping [37]. One way to ameliorate this problem is to design moat structures to attract unwanted flux to areas relatively far from the sensitive circuits [47]. Our cell library does integrate some moat structures, but our interconnect currently does not. Development of a systematic strategy on moat design and placement would make AQFP logic and superconductor electronics in general, more resilient and practical.

D. Advanced EDA Tools

All investigations into the aforementioned challenges can be accelerated with the right set of CAD and EDA tools. Of particular importance are tools to rapidly prototype and evaluate 3-D inductor structures, accurate and fast analog simulation, flux trapping analysis, and flexible chip-level integration tools. With the IARPA SuperTools program in progress for developing EDA and TCAD for VLSI superconductor electronics [48], [49], the outlook looks positive.

VI. TOWARD PRACTICAL CRYOSYSTEMS

AQFP logic is based on superconductivity, which requires cryogenic cooling for the devices to transition into the superconducting state to operate. This implies that AQFP logic is impractical for mobile devices and is better suited for larger computing facilities, such as data centers and supercomputers, where cryocooling systems can be accommodated. Some estimations on the power budgets for large-scale superconductor-based computing systems already exist in [3]. To extend this estimation to AQFP logic, we considered the supercomputing system featuring two Linde LR280 helium reliquefier refrigeration systems from [3]. The details of this system are summarized in the following.

- 1) *Refrigeration System*: Linde LR280 (2 \times).
- 2) *Cost*: 6 million USD.
- 3) *Refrigeration Capacity*: 1020 W at 4.4 K.
- 4) *Refrigeration Efficiency*: 395 W/W_{4.4K}.
- 5) *Power for Cryoelectronics*: 0.4 MW at maximum capacity.
- 6) *Power for Non-Cryoelectronics*: 1.6 MW.
- 7) *Total Power of Computing System*: 2 MW.

Specifically, the above-mentioned system requires a room temperature power of 0.4 MW to cool 1020 W worth of cryoelectronics. An estimate of an additional 1.6 MW is for the non-refrigerated components, such as room temperature interfaces, power supplies, and storage [3]. Thus, the total power of 2 MW is estimated for such a system.

Using this system, we made some estimations based on the NVIDIA Ampere GA100 GPU [50], [51], and the Intel Teraflops Research Chip (codenamed Polaris) [52]. The GA100

TABLE V
COOLING CAPACITY OF AQFP CHIPS IN
LR280-COOLED SUPERCOMPUTER

Chip architecture	Estimated JJs	Clock [GHz]	Power ¹ at 4.4 K [mW]	Total chips in system ¹
NVIDIA GA100 (A)	54×10^9	1.41	15.2	6.698×10^4
NVIDIA GA100 (B)	216×10^9	1.41	60.9	1.675×10^4
Intel Polaris (A)	100×10^6	5.1	0.7	1.429×10^6
Intel Polaris (B)	400×10^6	5.1	2.9	3.571×10^5

(A) assumes JJ count equals transistor count of original design.

(B) assumes JJ count will be $4\times$ as many transistors.

¹ LR280-based supercomputer has a refrigeration capacity of 1020 W at 4.4 K. Physical capacity is not considered here.

operates with a boost clock rate of 1.41 GHz; therefore, we used the switching energy of the AQFP operating at 1.5 GHz which is approximately 0.2 zJ per JJ [42] to calculate its power at 4.4 K. The Intel Polaris chip operates at 5.1 GHz; therefore, we used the AQFP switching energy of 1.4 zJ per JJ obtained from experiments at 5 GHz [11] to calculate its power at 4.4 K.

Table V shows how many chips of the aforementioned architectures can be cooled within the refrigeration capacity of the LR280-cooled supercomputer. Note that we are only considering the cooling power budget and not the logistics of physically inserting chips into the cooling system and interfacing them to room temperature. To estimate the complexity of the AQFP-based implementations of the two architectures, we made two separate assumptions: (A) we assumed that the number of JJs is equivalent to the number of transistors reported in the design and (B) we assumed that the number of JJs is $4\times$ the number of transistors. (B) is considered the more conservative estimation as we expect more JJ devices to be used for data buffering (synchronization), repeaters along long interconnects, active fan-out, and memory when compared with the transistor-based implementation. On the other hand, AQFP logic is intrinsically a majority logic family, which can yield better quality-of-results in majority-logic-optimized synthesis when compared with conventional logic primitives [24], [53].

Noting approximately how many chips can be cooled in the LR280-based supercomputing system, we can estimate performance from the reported TFLOPS of each architecture for different floating-point (FP) operations [50]–[52]. The estimates are shown in Table VI. The goal of the U.S. Department of Energy (DoE) Exascale Computing Initiative, Washington, DC, USA, is to build a high-performance computing system that has a LINPACK performance of 1 EFLOPS within a power budget of 20 MW [54]. The scaled EFLOPS shown in Table VI shows the estimated exascale performance of the LR280-cooled supercomputing system when it is scaled from the original total power budget of 2 MW to the DoE power budget constraint of 20 MW. The results indicate that a conservative estimation (B) of an AQFP-based GA100 GPU would exceed the 1-EFLOPS goal for both single-precision (3.3 EFLOPS) and double-precision (1.6 EFLOPS). But the GA100 uses a lot of devices not just for raw computation but for working with graphics and textures as well. The Intel Polaris chip was specifically designed for terascale computing; therefore, it has a higher TFLOPS-to-device ratio when

TABLE VI
PERFORMANCE OF LR280-COOLED SUPERCOMPUTER
USING AQFP LOGIC

Chip architecture and FP operation	Chip TFLOPS	System ¹ PFLOPS	System ¹ GFLOP/J	Scaled ² EFLOPS
NVIDIA GA100				
Bfloat16 (A)	312	20898	10449	209.0
Single-precision (A)	19.5	1306	653	13.1
Double-precision (A)	9.7	650	325	6.5
Bfloat16 (B)	312	5225	2612	52.2
Single-precision (B)	19.5	327	163	3.3
Double-precision (B)	9.7	162	81	1.6
Intel Polaris				
Single-precision (A)	1.63	2329	1164	23.3
Single-precision (B)	1.63	582	291	5.8

(A) assumes JJ count equals transistor count of original design.

(B) assumes JJ count will be $4\times$ as many transistors.

¹ LR280-cooled supercomputer system with a total estimated power of 2 MW.

² Scaled to the DoE exascale power budget of 20 MW.

compared with the GA100. In the conservative case (B) of an AQFP-based Intel Polaris supercomputer, the performance is nearly $6\times$ as much as the 1-EFLOPS goal of the DoE. If we can sufficiently address the challenges mentioned in Section V, an expected performance of somewhere between our (A) estimations and (B) estimations should be possible.

Our implementation of MANA is nowhere near the complexity and performance of the aforementioned architectures, but its successful low-speed functional demonstration along with the high-speed demonstration of its EX stage indicate that AQFP logic can indeed perform practical adiabatic computation. The extremely low switching energy and relatively high-clock rates (2–5 GHz) of AQFP logic also enable it to interface closely to superconductor-based qubits in the form of a controller for the read-out of quantum states for quantum computing [55]. With further refinement in the design methodologies to improve area efficiency and latency and the emergence of superconductor EDA tools to help with flux trapping analysis, system-level integration, and clocking, more feature-rich performance-driven AQFP circuits may be feasible in the near future.

VII. CONCLUSION

We have designed and demonstrated an adiabatic microprocessor called MANA using unshunted superconductor JJ devices. We implemented this microprocessor using AQFP logic. The basic AQFP cell has a measured switching energy of only 1.4 zJ per JJ at 4.2 K. Even after considering cooling costs, the switching energy is approximately only 1.4 aJ per JJ. The microprocessor was manufactured using a four-layer $10\text{ kA}/\text{cm}^2$ Nb/AIO_x/Nb superconductor integrated circuit process. Our test programs show that our MANA chip is capable of performing all the basic operations of the processor, including R/W into the RF unit, ALU execution, hardware stalling, and program branching under a 100–kHz clock rate in experiments. In addition, we developed a high-speed breakout chip of the EX stage of MANA and showed correct operation up to 2.5 GHz. This validates that AQFP logic is not only energy efficient but also a very capable foundation for high-performance design. With these successful demonstrations, we see that there is a potential future in

applying this technology for the next generation of energy-efficient supercomputers and data centers.

ACKNOWLEDGMENT

The authors would like to thank Mr. Taiki Yamae for finding an error in their energy calculations.

Design and EDA tools were provided by VLSI Design and Education Center, The University of Tokyo, Tokyo, Japan, in collaboration with Cadence Design Systems, Inc., San Jose, CA, USA. The circuits were fabricated in the Clean Room for Analog-digital superconductivity, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, using the high-speed standard process.

REFERENCES

- [1] M. W. Johnson *et al.*, "Quantum annealing with manufactured spins," *Nature*, vol. 473, no. 7346, pp. 194–198, May 2011. [Online]. Available: <https://www.nature.com/articles/nature10012>
- [2] F. Arute *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019. [Online]. Available: <http://www.nature.com/articles/s41586-019-1666-5>
- [3] D. S. Holmes, A. L. Ripple, and M. A. Manheimer, "Energy-efficient superconducting computing—Power budgets and requirements," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, Jun. 2013, Art. no. 1701610.
- [4] N. Jones, "How to stop data centres from gobbling up the world's electricity," *Nature*, vol. 561, no. 7722, pp. 163–166, Sep. 2018.
- [5] A. Andrae and T. Edler, "On global electricity usage of communication technology: Trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, Apr. 2015.
- [6] D. E. Kirichenko, S. Sarwana, and A. F. Kirichenko, "Zero static power dissipation biasing of RSFQ circuits," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 776–779, Jun. 2011.
- [7] O. A. Mukhanov, "Energy-efficient single flux quantum technology," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 760–769, Jun. 2011.
- [8] Q. P. Herr, A. Y. Herr, O. T. Oberg, and A. G. Ioannidis, "Ultra-low-power superconductor logic," *J. Appl. Phys.*, vol. 109, no. 10, May 2011, Art. no. 103903.
- [9] N. Yoshikawa and Y. Kato, "Reduction of power consumption of RSFQ circuits by inductance-load biasing," *Superconductor Sci. Technol.*, vol. 12, no. 11, pp. 918–920, Nov. 1999, doi: [10.1088/0953-2048/12/11/367](https://doi.org/10.1088/0953-2048/12/11/367).
- [10] M. Tanaka, M. Ito, A. Kitayama, T. Kouketsu, and A. Fujimaki, "18-GHz, 4.0-aJ/bit operation of ultra-low-energy rapid single-flux-quantum shift registers," *Jpn. J. Appl. Phys.*, vol. 51, May 2012, Art. no. 053102, doi: [10.1143/JJAP.51.053102](https://doi.org/10.1143/JJAP.51.053102).
- [11] N. Takeuchi, T. Yamae, C. L. Ayala, H. Suzuki, and N. Yoshikawa, "An adiabatic superconductor 8-bit adder with 24kBT energy dissipation per junction," *Appl. Phys. Lett.*, vol. 114, no. 4, Jan. 2019, Art. no. 042602, doi: [10.1063/1.5080753](https://doi.org/10.1063/1.5080753).
- [12] N. Takeuchi *et al.*, "Margin and energy dissipation of adiabatic quantum-flux-parametron logic at finite temperature," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, Jun. 2013, Art. no. 1700304.
- [13] A. Stillmaker and B. Baas, "Scaling equations for the accurate prediction of CMOS device performance from 180 nm to 7 nm," *Integration*, vol. 58, pp. 74–81, Jun. 2017.
- [14] N. Takeuchi *et al.*, "Adiabatic quantum-flux-parametron cell library designed using a 10 kA cm⁻² niobium fabrication process," *Superconductor Sci. Technol.*, vol. 30, no. 3, Jan. 2017, Art. no. 035002.
- [15] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Adiabatic quantum-flux-parametron cell library adopting minimalist design," *J. Appl. Phys.*, vol. 117, no. 17, May 2015, Art. no. 173912.
- [16] E. S. Fang and T. V. Duzer, "A Josephson integrated circuit simulator (JSIM) for superconductive electronics application," in *Proc. Ext. Abstr. 2nd Int. Supercond. Electron. Conf.*, 1989, pp. 407–410. [Online]. Available: <https://github.com/coldlogix/jsim>
- [17] K. K. Likharev and V. K. Semenov, "RSFQ logic/memory family: A new Josephson-junction technology for sub-terahertz-clock-frequency digital systems," *IEEE Trans. Appl. Supercond.*, vol. 1, no. 1, pp. 3–28, Mar. 1991.
- [18] W. Chen, A. V. Rylyakov, V. Patel, J. E. Lukens, and K. K. Likharev, "Rapid single flux quantum T-flip flop operating up to 770 GHz," *IEEE Trans. Appl. Supercond.*, vol. 9, no. 2, pp. 3212–3215, Jun. 1999.
- [19] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Energy efficiency of adiabatic superconductor logic," *Superconductor Sci. Technol.*, vol. 28, no. 1, Nov. 2014, Art. no. 015003.
- [20] C. L. Ayala *et al.*, "A semi-custom design methodology and environment for implementing superconductor adiabatic quantum-flux-parametron microprocessors," *Superconductor Sci. Technol.*, vol. 33, no. 5, Mar. 2020, Art. no. 054006.
- [21] Q. Xu, C. L. Ayala, N. Takeuchi, Y. Murai, Y. Yamanashi, and N. Yoshikawa, "Synthesis flow for cell-based adiabatic quantum-flux-parametron structural circuit generation with HDL back-end verification," *IEEE Trans. Appl. Supercond.*, vol. 27, no. 4, pp. 1–5, Jun. 2017.
- [22] T. Tanaka, C. L. Ayala, Q. Xu, R. Saito, and N. Yoshikawa, "Fabrication of adiabatic quantum-flux-parametron integrated circuits using an automatic placement tool based on genetic algorithms," *IEEE Trans. Appl. Supercond.*, vol. 29, no. 5, pp. 1–6, Aug. 2019.
- [23] N. Tsuji, C. L. Ayala, N. Takeuchi, T. Ortlepp, Y. Yamanashi, and N. Yoshikawa, "Design and implementation of a 16-word by 1-bit register file using adiabatic quantum flux parametron logic," *IEEE Trans. Appl. Supercond.*, vol. 27, no. 4, pp. 1–4, Jun. 2017.
- [24] C. L. Ayala, N. Takeuchi, Y. Yamanashi, T. Ortlepp, and N. Yoshikawa, "Majority-logic-optimized parallel prefix carry look-ahead adder families using adiabatic quantum-flux-parametron logic," *IEEE Trans. Appl. Supercond.*, vol. 27, no. 4, pp. 1–7, Jun. 2017.
- [25] D. Patterson, *Computer Organization and Design: The Hardware/Software Interface*. Cambridge, MA, USA: Morgan Kaufmann, 2018.
- [26] J. R. Gurd, C. C. Kirkham, and I. Watson, "The Manchester prototype dataflow computer," *Commun. ACM*, vol. 28, no. 1, pp. 34–52, Jan. 1985.
- [27] M. Dorojevets, C. L. Ayala, and A. K. Kasperek, "Data-flow microarchitecture for wide datapath RSFQ processors: Design study," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 787–791, Jun. 2011.
- [28] Y. Murai, C. L. Ayala, N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Development and demonstration of routing and placement EDA tools for large-scale adiabatic quantum-flux-parametron circuits," *IEEE Trans. Appl. Supercond.*, vol. 27, no. 6, pp. 1–9, Sep. 2017.
- [29] Q. Xu, C. L. Ayala, N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "HDL-based modeling approach for digital simulation of adiabatic quantum flux parametron logic," *IEEE Trans. Appl. Supercond.*, vol. 26, no. 8, pp. 1–5, Dec. 2016.
- [30] J. A. Delpoit *et al.*, "JoSIM—Superconductor SPICE simulator," *IEEE Trans. Appl. Supercond.*, vol. 29, no. 5, pp. 1–5, Aug. 2019.
- [31] C. L. Ayala, O. Chen, and N. Yoshikawa, "AQFPTX: Adiabatic Quantum-Flux-Parametron timing eXtraction tool," in *Proc. IEEE Int. Superconductive Electron. Conf. (ISEC)*, Jul. 2019, pp. 1–3.
- [32] C. J. Fourie, O. Wetzstein, T. Ortlepp, and J. Kunert, "Three-dimensional multi-terminal superconductive integrated circuit inductance extraction," *Superconductor Sci. Technol.*, vol. 24, no. 12, Nov. 2011, Art. no. 125015.
- [33] N. Tsuji, N. Takeuchi, Y. Yamanashi, T. Ortlepp, and N. Yoshikawa, "Majority gate-based feedback latches for adiabatic quantum flux parametron logic," *IEICE Trans. Electron.*, vol. E99.C, no. 6, pp. 710–716, 2016.
- [34] E. J. Wilkinson, "An N-Way hybrid power divider," *IEEE Trans. Microw. Theory Techn.*, vol. MTT-8, no. 1, pp. 116–118, Jan. 1960.
- [35] M. Elsbury *et al.*, "Broadband lumped-element integrated N-way power dividers for voltage standards," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 8, pp. 2055–2063, Aug. 2009. [Online]. Available: <http://ieeexplore.ieee.org/document/5175347>
- [36] N. Takeuchi, H. Suzuki, and N. Yoshikawa, "Measurement of low bit-error-rates of adiabatic quantum-flux-parametron logic using a superconductor voltage driver," *Appl. Phys. Lett.*, vol. 110, no. 20, May 2017, Art. no. 202601.
- [37] Y. Polyakov, S. Narayana, and V. K. Semenov, "Flux trapping in superconducting circuits," *IEEE Trans. Appl. Supercond.*, vol. 17, no. 2, pp. 520–525, Jun. 2007.
- [38] J. Lim, D.-G. Kim, and S.-I. Chae, "A 16-bit carry-lookahead adder using reversible energy recovery logic for ultra-low-energy systems," *IEEE J. Solid-State Circuits*, vol. 34, no. 6, pp. 898–903, Jun. 1999.
- [39] S. Kim and S.-I. Chae, "Implementation of a simple 8-bit microprocessor with reversible energy recovery logic," in *Proc. 2nd Conf. Comput. Frontiers CF*. New York, NY, USA: ACM, 2005, pp. 421–426.

- [40] R. Celis-Cordova, A. O. Orlov, T. Lu, J. M. Kulick, and G. L. Snider, "Design of a 16-bit adiabatic microprocessor," in *Proc. IEEE Int. Conf. Rebooting Comput. (ICRC)*, Nov. 2019, pp. 1–4.
- [41] S. Tolpygo *et al.*, "Advanced fabrication processes for superconducting very large-scale integrated circuits," *IEEE Trans. Appl. Supercond.*, vol. 26, no. 3, Apr. 2016, Art. no. 1100110.
- [42] Y. He *et al.*, "A compact AQFP logic cell design using an 8-metal layer superconductor process," *Superconductor Sci. Technol.*, vol. 33, no. 3, Feb. 2020, Art. no. 035010.
- [43] N. Takeuchi, K. Arai, and N. Yoshikawa, "Directly coupled adiabatic superconductor logic," *Superconductor Sci. Technol.*, vol. 33, no. 6, May 2020, Art. no. 065002.
- [44] N. Takeuchi, M. Nozoe, Y. He, and N. Yoshikawa, "Low-latency adiabatic superconductor logic using delay-line clocking," *Appl. Phys. Lett.*, vol. 115, no. 7, Aug. 2019, Art. no. 072601.
- [45] Y. He, N. Takeuchi, and N. Yoshikawa, "Low-latency power-dividing clocking scheme for adiabatic quantum-flux-parametron logic," *Appl. Phys. Lett.*, vol. 116, no. 18, May 2020, Art. no. 182602.
- [46] N. Takeuchi, T. Ortlepp, Y. Yamanashi, and N. Yoshikawa, "Experimental demonstration of quantum-flux-latch-based circuits," *IEEE Trans. Appl. Supercond.*, vol. 25, no. 3, Jun. 2015, Art. no. 1300803.
- [47] V. K. Semenov and M. M. Khapaev, "How moats protect superconductor films from flux trapping," *IEEE Trans. Appl. Supercond.*, vol. 26, no. 3, pp. 1–10, Apr. 2016.
- [48] *IARPA SuperTools Program*. Accessed: 2017. [Online]. Available: <https://www.iarpa.gov/index.php/research-programs/supertools>
- [49] C. J. Fourie *et al.*, "ColdFlux superconducting EDA and TCAD tools project: Overview and progress," *IEEE Trans. Appl. Supercond.*, vol. 29, no. 5, pp. 1–7, Aug. 2019.
- [50] *NVIDIA A100 Tensor Core GPU Architecture: Whitepaper*. Accessed: Jun. 2020. [Online]. Available: <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/nvidia-ampere-architecture-whitepaper.pdf>
- [51] *NVIDIA A100 Tensor Core GPU: Datasheet*. Accessed: Jun. 2020. [Online]. Available: <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet.pdf>
- [52] S. Vangal *et al.*, "An 80-tile 1.28TFLOPS network-on-chip in 65 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 98–589.
- [53] L. Amaru, P.-E. Gaillardon, and G. De Micheli, "Majority-inverter graph: A new paradigm for logic optimization," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 5, pp. 806–819, May 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7293649/>
- [54] W. Harrod. *DOE Exascale Computing Initiative (ECI)*. Accessed: 2012. [Online]. Available: <https://science.osti.gov/-media/ascr/ascac/pdf/meetings/aug12/2012-ECI-ASCAC-v4.pdf>
- [55] N. Yoshikawa, "Superconducting digital electronics for controlling quantum computing systems," *IEICE Trans. Electron.*, vol. E102.C, no. 3, pp. 217–223, Mar. 2019.



Christopher L. Ayala (Senior Member, IEEE) received the B.Eng. and M.Sc. degrees in 2009 and the Ph.D. degree in 2012 from Stony Brook University, Stony Brook, NY, USA, all in electrical and computer engineering.

From 2013 to 2015, he was a Post-Doctoral Fellow with IBM Research, Zürich, Switzerland. Since 2015, he has been with the Institute of Advanced Sciences, Yokohama National University, Yokohama, Japan, where he is currently an Associate Professor. His research interests include emerging circuit technologies, superconductor logic, NEMS-MEMS, novel computer architectures, and electronic design automation (EDA).

Dr. Ayala is a member of the IEEE HKN, the Japan Society of Applied Physics, the Institute of Electronics, Information and Communication Engineers of Japan, and the Tau Beta Pi Engineering Honor Society.



Tomoyuki Tanaka received the B.S. and M.E. degrees in electrical and computer engineering from Yokohama National University, Yokohama, Japan, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree with a focus on research on ultra-low power computers using superconductor devices.

Mr. Tanaka is a member of the Institute of Electronics, Information and Communication Engineers.



Ro Saito (Member, IEEE) received the B.S. degree in electrical and computer engineering, and the M.S. degree in informational science from Yokohama National University, Yokohama, Japan, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree in information systems.

His research interests include adiabatic quantum-flux-parametron logic, computer architecture, electronic design automation (EDA), and logic synthesis.

Mr. Saito is a member of the Institute of Electronics, Information and Communication Engineers of Japan.



Mai Nozoe (Member, IEEE) received the B.S. and M.E. degrees in electrical and computer engineering from Yokohama National University, Yokohama, Japan, in 2017 and 2019, respectively.

She was involved in research on adiabatic quantum-flux-parametron logic, with an emphasis on the design of scalable register file arrays for energy-efficient computing. Since 2019, she has been with the Nomura Research Institute, Tokyo, Japan.

Ms. Nozoe is a member of the Japan Society of Applied Physics.



Naoki Takeuchi (Member, IEEE) received the B.S., M.E., and Ph.D. degrees from Yokohama National University, Yokohama, Japan, in 2008, 2010, and 2014, respectively, all in electrical and computer engineering.

He was an Electronic Circuit Designer with SONY, Tokyo, Japan, from 2010 to 2011, a Post-Doctoral Researcher with the National Institute of Information and Communications Technology, Tokyo, and the Japan Society for the Promotion of Science from 2014 to 2015, and a PRESTO Researcher with the Japan Science and Technology Agency, Kawaguchi, Japan, from 2015 to 2019. Since 2015, he has been with the Institute of Advanced Sciences, Yokohama National University, where he is currently an Associate Professor. His research interests include: superconductor logic, including rapid single flux quantum and adiabatic quantum-flux-parametron, cryogenic detectors, including SNSPD, and non-von Neumann computing, including reversible computing, quantum computing, and bio-inspired computing.

Dr. Takeuchi is a member of the Japan Society of Applied Physics and the Institute of Electronics, Information and Communication Engineers of Japan.



Nobuyuki Yoshikawa (Senior Member, IEEE) received the B.Eng., M.Eng., and Ph.D. degrees in electrical and computer engineering from Yokohama National University, Yokohama, Japan, in 1984, 1986, and 1989, respectively.

Since 1989, he has been with the Department of Electrical and Computer Engineering, Yokohama National University, where he is currently a Professor. His research interests include superconductive devices and their application in digital and analog circuits, quantum computing devices, and cryo-

CMOS devices.

Dr. Yoshikawa is the Chair of the Superconducting Electronics Committee of the Japan Society for the Promotion of Science and the Technical Committee on Metal and Ceramics of the Institute of Electrical Engineering of Japan.