

29.3 A 48GHz 5.6mW Gate-Level-Pipelined Multiplier Using Single-Flux Quantum Logic

Ikki Nagaoka¹, Masamitsu Tanaka¹, Koji Inoue², Akira Fujimaki¹

¹Nagoya University, Nagoya, Japan

²Kyushu University, Fukuoka, Japan

A multiplier based on superconductor single-flux-quantum (SFQ) logic is demonstrated up to 48GHz with the measured power consumption of 5.6 mW. The multiplier performs 8×8 -bit signed multiplication every clock cycle. The design is based on a bit-parallel, gate-level-pipelined structure that exploits ultimately high-throughput performance of SFQ logic. The test chip fabricated using a 1.0- μm , 9-layer process consists of 20,251 Nb/AlOx/Nb Josephson junctions (JJs). The correctness of operation is verified by on-chip high-speed testing.

The SFQ logic [1] is a superconductor ultrafast digital circuit technology based on magnetic flux quantization and quantum interference in superconductor rings containing JJs. Binary information is represented by absence or presence of an SFQ in a superconductor ring, and JJs are used as switching elements. An impulse-shaped voltage is generated only when an SFQ travels across a JJ, which leads to ultralow power consumption ($\sim 10^{-19}$ J per switching event) in principle. A superconductor stripline wiring behaves like a lossless waveguide and propagates a mass-free SFQ signal at the speed of light for a long distance ($> 10\text{mm}$). Based on these features, several projects on the SFQ-based VLSI technology have been undertaken in the US and Japan toward high-performance computing in the post-Moore's era.

SFQ signals can be split or merged using a few JJs, and easily stored in large-inductance superconducting rings. Unlike voltage level logic, synchronization of an input signal with the reference signal (clock signal) is essential at a storage ring in each SFQ logic gate in order to distinguish between that the input signal has not arrived yet and that a logic value '0' has arrived. In other words, each storage ring intrinsically has latch or memory functionality, and all the SFQ logic gates are clocked gates except for pulse splitters and mergers. Such a unique feature expands an architectural design space of SFQ circuits that would never be reached with conventional CMOS logic circuits. Especially gate-level, ultradeep pipelining, with target clock frequencies in a range of several tens of GHz, is promising for high-performance circuits, because no additional pipeline registers are required and because operating frequencies are not limited by the power-wall problem but simply determined by the sum of setup and hold times in SFQ logic. Figure 29.3.1 summarizes the SFQ circuit elements used in this work [2], by which 50-to-100GHz operation is possible. However, it has been believed that gate-level pipelining is too difficult to be applied to complex SFQ circuits because of timing design. Several approaches including asynchronous design [3] and bit-serial processing [4] have been proposed to avoid the timing design complexity, even though these approaches result in lower clock frequency or poor processing performance.

In this work, the authors demonstrate an ultimately high-throughput multiplier based on a bit-parallel, gate-level-pipelined structure to show the maximum potential of SFQ logic. Figure 29.3.2 illustrates the block diagram of the designed multiplier. The total number of pipelining stages is 15, and the multiplier is divided into three component circuits, i.e., the partial product generator (PPG), partial product accumulator (PPA), and final stage adder (FSA). In PPG, all partial products (PPs) are generated in parallel by AND gates in one clock cycle. Then, the PPs are processed in PPA, which calculates based on carry-save addition using full adders and half adders. The final two PPs are added by a parallel prefix adder. For successful gate-level pipelining, several D flop-flops are inserted into paths to equalize the logic depth, and multiplication is performed every clock cycle. In layout design, clock signals to all the logic gates are carefully distributed to minimize timing jitter, and clock skews are intentionally added to conceal signal propagation delays, in order to achieve the highest clock frequency. The wire lengths are also precisely controlled, with an accuracy of $10\mu\text{m}$, because signal propagation at the speed of light is no longer negligible in SFQ circuit design.

Figures 29.3.3 and 29.3.4 display the micrograph of the chip and the cross-sectional device structure fabricated using a 1.0 μm , 9-layer process [5]. In addition to the designed multiplier, there are shift-registers (SRs) and a clock generator (CG) for on-chip testing. The CG generates a high-speed clock signal train around 50GHz, depending on the supplied voltage. In total, 20,251 JJs are integrated on a 6.03mm \times 5.22mm area.

The chip is tested in a liquid helium bath at 4.2K. The correctness of the multiplier operation is verified using many test vectors. One of the on-chip high-speed test results is shown in Fig. 29.3.5. The six waveforms at the top correspond to input signals, where an SFQ pulse is generated at each rising edge. The others are outputs, whose signal transitions represent the arrival of SFQ signals. In this test sequence, first, four sets of test vectors are written in the SRs at low frequency ($\sim 1\text{kHz}$) from room-temperature electronics (i). Then, the CG is triggered to generate high-speed clock signals, and multiplication is performed on the chip (ii). Finally, the results are read out from the SRs at low speed and verified (iii). Figure 29.3.6 shows operating voltage vs. clock frequency for the component circuits. The maximum clock frequency obtained in measurement is 48GHz with 5.6mW power.

Figure 29.3.7 shows the performance summary and comparison to prior works on demonstrated SFQ multipliers whose bit length is longer than four [3, 4]. Thanks to bit-parallel, gate-level pipelining, the measured throughput achieves 48 giga-operations per second (GOPS), and the resultant efficiency is almost 10×10^{12} operations per second per watt (10 TOPS/W). The authors believe that the successful demonstration of the gate-level pipelined multiplier is the starting point toward tera-FLOPS/W-class, ultimately throughput-oriented high-performance cryogenic computing using the SFQ VLSI technology.

Acknowledgements:

This work is supported by JSPS KAKENHI Grant Numbers JP16H02796, JP18H05211 and JP18H01498. The circuits are designed with the support by VDEC of the University of Tokyo, in collaboration with Cadence Design Systems, Inc. and fabricated in the CRAVITY of National Institute of Advanced Industrial Science and Technology.

References:

- [1] 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, pp. 3–28, Mar. 1991.
- [2] Y. Yamanashi, et al., "100 GHz demonstrations based on the single-flux-quantum cell library for the 10 kA/cm² Nb multi-layer process," *IEICE Trans. Electron.*, vol. E93-C, pp. 440–444, Apr. 2010.
- [3] M. Dorjjevets, A. K. Kasperek, N. Yoshikawa, and A. Fujimaki, "20-GHz 8×8 -bit parallel carry-save pipelined RSFQ multiplier," *IEEE Trans. Appl. Supercond.*, vol. 23, p. 1300104, June 2013.
- [4] X. Peng, et al., "High-speed demonstration of bit-serial floating-point adders and multipliers using single-flux-quantum circuits," *IEEE Trans. Appl. Supercond.*, vol. 25, p. 1301106, June 2015.
- [5] S. Nagasawa, et al., "Nb 9-layer fabrication process for superconducting large-scale SFQ circuits and its process evaluation," *IEICE Trans. Electron.*, vol. E97-C, pp. 132–140, Mar. 2014.

Cell	JJ count	Clocked Gate	Delay, ps	Setup, ps	Hold, ps
Splitter	3	No	4.3	n/a	n/a
Merger	7	No	8.2	n/a	n/a
D flip-flop	6	Yes	5.1	1.2	-0.9
NOT	10	Yes	9.6	1.2	5.0
AND	14	Yes	7.9	-1.4	2.0
OR	12	Yes	5.0	5.8	-0.2
Exclusive OR	11	Yes	6.5	3.7	4.1

Figure 29.3.1: SFQ library developed for 1.0- μ m, 9-layer process.

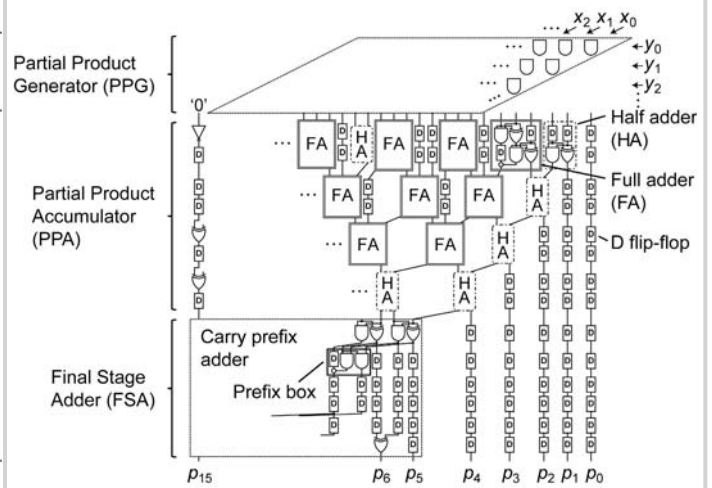


Figure 29.3.2: Schematic diagram of gate-level pipelined multiplier.

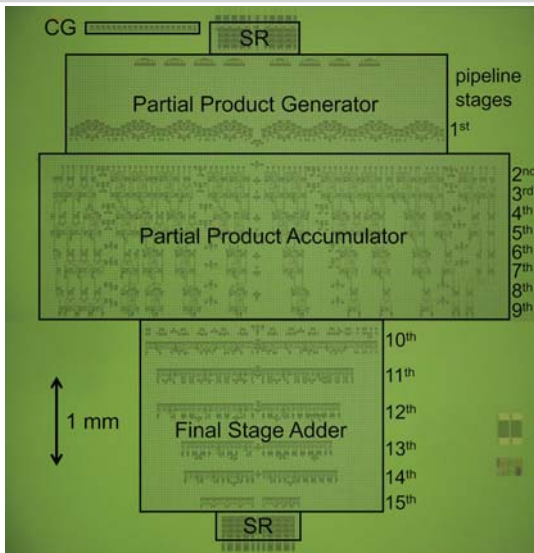


Figure 29.3.3: Die micrograph of multiplier with on-chip test circuitry.

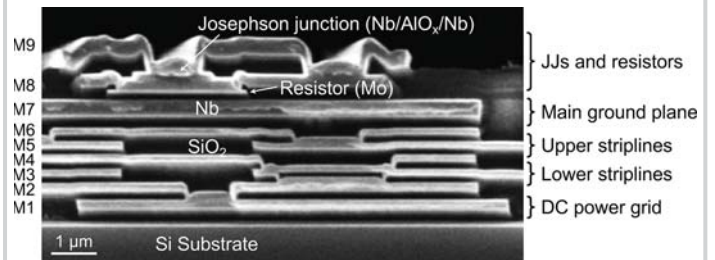


Figure 29.3.4: Cross-sectional scanning electron microscopy image of integrated device.

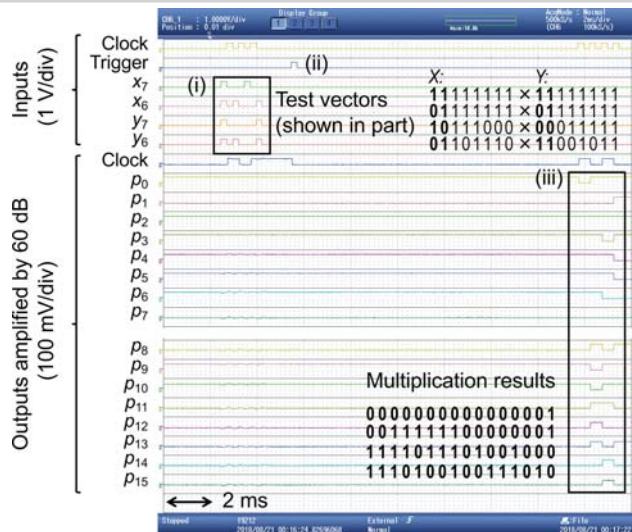


Figure 29.3.5: Oscilloscope waveforms of on-chip high-speed test (oscilloscope screens were captured twice and merged).

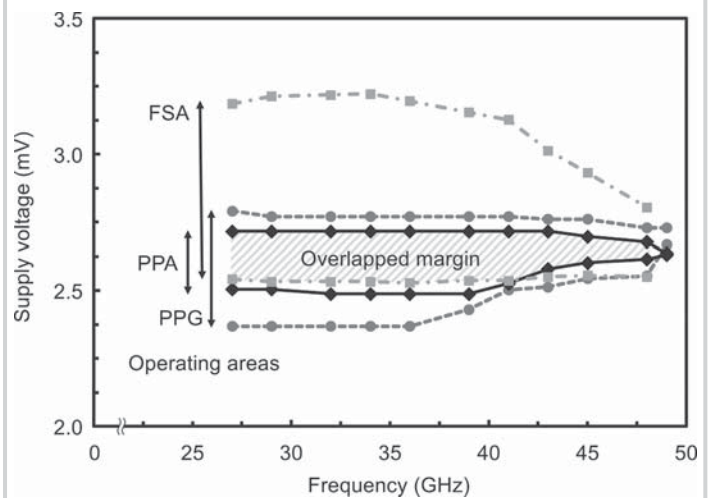


Figure 29.3.6: Frequency dependence of operating margin in supply voltage.

	This work	[3] (2013)	[4] (2015)
Process	1.0- μm , 9-layer, Nb/AIO _x /Nb JJs		
Bit processing	Parallel	Parallel	Serial
Design scheme	Gate-level pipeline	Asynchronous wave-pipeline	Pipelined systolic array
Operands	8-bit, signed integers	8-bit, unsigned integers	32-bit floating-point numbers
Bit width of product	16 bits	8 bits	32 bits
Clock Frequency, GHz	48	20 [†]	59
Throughput, GOPS	48	20 [†]	2.4
Power, mW	5.6	1.7 [†]	5.8
JJ count	17,488	5,948	18,766
Area, mm ²	5.07 × 5.22	2.50 × 1.45	3.77 × 6.66

[†] Simulation results

Figure 29.3.7: Experimentally-obtained performance summary and comparison to prior works.

--	--

--	--