

A Reconfigurable Data-Path Accelerator Based on Single Flux Quantum Circuits

Hiroshi KATAOKA^{†a)}, Hiroaki HONDA[†], Farhad MEHDIPOUR[†], *Nonmembers*, Nobuyuki YOSHIKAWA^{††}, Akira FUJIMAKI^{†††}, Hiroyuki AKAIKE^{†††}, Naofumi TAKAGI^{††††}, and Kazuaki MURAKAMI[†], *Members*

SUMMARY The single flux quantum (SFQ) is expected to be a next-generation high-speed and low-power technology in the field of logic circuits. CMOS as the dominant technology for conventional processors cannot be replaced with SFQ technology due to the difficulty of implementing feedback loops and conditional branches using SFQ circuits. This paper investigates the applicability of a reconfigurable data-path (RDP) accelerator based on SFQ circuits. The authors introduce detailed specifications of the SFQ-RDP architecture and compare its performance and power/performance ratio with those of a graphics-processing unit (GPU). The results show at most 1600 times higher efficiency in terms of Flops/W (floating-point operations per second/Watt) for some high-performance computing application programs.

key words: *single flux quantum, reconfigurable data-path, accelerator*

1. Introduction

Superconducting single flux quantum (SFQ) circuit technology [1] is expected to be a next-generation circuit technology which enables ultra high-speed computation with ultra low-power consumption. The basic component of SFQ circuits is a superconducting loop with Josephson junctions (JJs). A single flux quantum pulse appears as a 1 mV extremely low width (in the range of pico-seconds) and carries information at very high speed (up to the light speed) through the circuit, while in CMOS circuits the speed is limited due to the time required for charging/discharging capacitors [1].

Further, SFQ circuits operate based on pulse logic and each gate is a clocked gate and has a function of a latch. Hence, latches are implemented without additional costs and this makes SFQ circuits suitable for pipeline processing on streaming data. One issue in implementing very large-scale high-performance circuits by means of semiconductor circuits is the heat radiation and difficulty in high-density packaging. Unlike CMOS circuits, SFQ circuits exhibit extremely low-power consumption and also very smaller area compared to the CMOS counterpart [2]. On the contrary,

SFQ circuits are not suitable for processing feedback loops and conditional branches.

Nowadays, memory wall problem is a well-known problem which comes from the gap between the operating speed of processors and that of memory. In other words, the memory bandwidth required for data transfer between general-purpose processors (GPPs) and the memory is not enough to fill this gap. In SFQ circuits, this problem might be more crucial, because implementing large-scale superconducting random-access memory is difficult while SFQ processing units can operate so fast. Most of the accesses to memory are due to reading/writing intermediate data to/from a finite number of registers, which associates with spill code. An appropriate architecture capable of reducing the impact of spill code can alleviate the impact of memory wall problem as well.

In order to exploit the benefits of SFQ circuits in providing considerably high speed and low power consumption compared to CMOS technology, a new system architecture, or SFQ-RDP system, has been introduced which consists of a CMOS GPP, a CMOS memory, and an SFQ-based reconfigurable data-path (RDP) accelerator (SFQ-RDP accelerator) [3]. Hundreds of processing elements (PEs) are arranged over a two-dimensional array so that the output of each PE can be forwarded to the inputs of one or more PEs through flexible operand-routing networks (ORNs).

Since the RDP accelerator as a main component of the target architecture is implemented by SFQ circuits, obviously it can resolve most of the above-mentioned power-consumption issues originated from CMOS technology. Further, due to pipelined architecture of the RDP, the cascaded PEs can generate final results without temporally memorizing intermediate data, therefore the number of load/store operations will be reduced and the memory-bandwidth pressure can be relieved. In addition, through providing the RDP with dynamic reconfiguration facility it would be possible to program processing elements and routing resources at run-time. Hence, the RDP is an adaptable architecture as well.

In this paper, we provide a complete, but condensed, survey on our work on the SFQ-RDP architecture and system, most of which are reported in [4]–[6], in the project “Low-power, High-performance, Reconfigurable Processor Using Single-Flux-Quantum Circuits” [3] in the Research Area “Technology Innovation and Integration for Information Systems with Ultra Low Power” supported by JST

Manuscript received July 16, 2013.

Manuscript revised November 1, 2013.

[†]The authors are with Kyushu University, Fukuoka-shi, 819-0395 Japan.

^{††}The author is with the Yokohama National University, Yokohama-shi, 240-8501 Japan.

^{†††}The authors are with Nagoya University, Nagoya-shi, 464-8601 Japan.

^{††††}The author is with Kyoto University, Kyoto-shi, 606-8501 Japan.

a) E-mail: kataoka@soc.ait.kyushu-u.ac.jp

DOI: 10.1587/transele.E97.C.141

CREST Program (FY 2006–2012). The design procedure of SFQ-RDP as well as optimization through design space exploration is exploited. In the next section, a general specification of RDP architectures is given. In Sect. 3, detailed design of SFQ-RDP accelerators is discussed, and the design procedure of SFQ-RDP accelerators as well as optimization through design space exploration is exploited. In Sect. 4, results of evaluation of performance and power/performance ratio for target application programs implemented on SFQ-RDP system as well as comparisons to graphics-processing unit (GPU) are reported. Finally, Sect. 5 concludes the paper.

2. Reconfigurable Data Path (RDP) Architecture

As mentioned in Sect. 1, it is difficult to implement feedback loops and conditional branches using SFQ circuits. Further, any practical on-chip and off-chip SFQ memories have never been developed yet. To overcome these disadvantages, we have proposed a reconfigurable data-path (RDP) architecture and its application program to SFQ circuits, or SFQ-RDP [3].

Generally, RDP is a pipelined architecture comprising a two-dimensional array of processing elements (PEs) and operand-routing networks (ORNs) such that one PE can be connected through ORNs to a number of PEs in the next row. The RDP is implemented by SFQ circuits as a substitute to CMOS circuits. Figure 1 displays the overall architecture of the proposed high-performance computer system consisting of a GPP, the RDP as accelerator, and memory elements.

Data flow graphs (DFGs) are pulled-out from critical segments of application programs, then configuration bitstreams, which are used for configuring the RDP, are generated by using a dedicated tool. During execution of an application program, configurations associated with the crit-

ical segments are loaded onto the RDP and executed in favor of achieving higher performance and lower power consumption. Since the cascaded PEs can generate a final result without temporally memorizing intermediate data, the number of memory load/store operations corresponding to spill code can be reduced. Therefore, memory bandwidth required to gain a high-performance computation might decrease as well. Furthermore, as a loop body mapped onto the PE array is executed in a pipeline fashion, RDP can provide a high computing throughput.

In addition to the basic properties of the RDP architecture, it is necessary to determine the following architectural specifications. These architectural specifications are determined through the design procedure discussed in Sect. 3.1.

2.1 Input/Output Ports

I/O ports are located on top and bottom boundaries of the RDP as displayed in Fig. 1. Limitation on the number of ports depends on available memory bandwidth, RDP operation frequency, width of data bus, and the number of memory read/write channels.

2.2 RDP Dimensions and Layout

Figure 1 shows the height and width of RDP as the number of rows and columns, respectively.

Layout of the RDP indicates the type of functional units (FUs) and their distribution (Fig. 2). Three layout types are examined for the RDP during the design procedure. In a normal layout (Layout I), each FU can implement operations including addition/subtraction (ADD/SUB) and multiplication (MUL). In Layout II, every other PE implements only one of the ADD/SUB and MUL operations. In the Layout III every other row of the RDP implements one of the operations. In other words, all PEs in the first row only implement MUL operations while in the second row only ADD/SUB operations are implemented.

2.3 PE Types

Three PE architectures (depicted in Fig. 3) are examined for

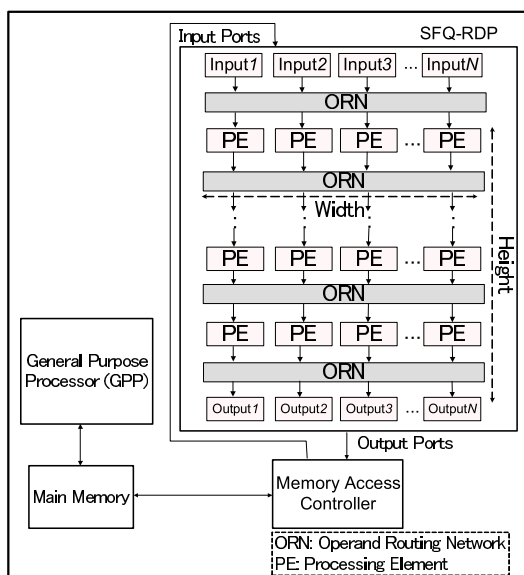


Fig. 1 Overall architecture of the SFQ-RDP system.

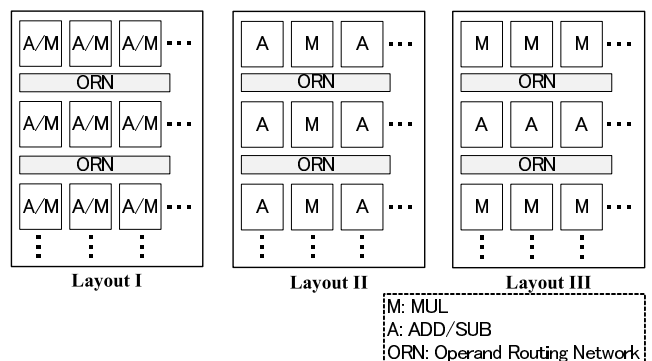


Fig. 2 Layout types.

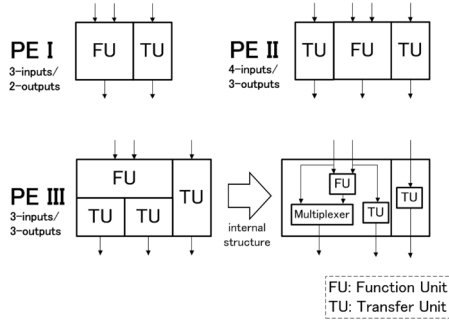


Fig. 3 Various PE architectures.

the RDP and the most suitable one is selected during the design procedure. The first PE architecture (PE I) includes an FU for implementing desired operations and a TU (transfer unit). As ORNs provide only routing resources between consecutive rows, TUs are utilized to connect two PEs not locating on consecutive rows. It is also possible to use an FU for implementing transfer operation. In addition, each PE has three inputs (two inputs for FU and one for TU) and two outputs (one from FU and another from TU). The second PE architecture (PE II) has one additional TU for increasing the flexibility of routing and it has 4-inputs/3-outputs. The third type of PE architecture (PE III) resembles the first one, though the difference is in extending capability of implementing two simultaneous TUs by the FU (totally three TUs). An additional multiplexer should be used inside the PE to choose between FU's output and the input. Each type of PE can implement various configurations depending on the states of the FU and TU (Fig. 4). Either FU or TU can be in used or unused states depending on the required functionality for the PE which is determined during the DFG mapping process. Also, each FU can be configured to operate as a TU. For example, in PE I, it is possible to use FU and TU together (Conf. I-1), or only use the FU while the TU is unused (Conf. I-2). Further, along with utilizing the TU, FU can be either configured as unused (Conf. I-3), or can be used as a TU, which consequently provides the possibility of implementing two TUs by a single PE (Conf. I-4). Similarly, PE II, and PE III can have 6, and 5 different configurations, respectively.

Each FU can implement basic 64-bit double-precision floating-point operations (e.g., ADD, SUB, and MUL). Control instructions (branches) and direct memory accesses via PEs are not supported. Further, 64-bit immediate registers are located in each PE in order to handle immediate operands (Fig. 5).

2.4 Operand Routing Network (ORN)

PEs of each row are connected to a number of PEs in the next row through ORNs as routing resources. Figure 6 shows the definition of connection length and the maximum connection length (MCL) on a piece of RDP architecture. It can be seen that the connection length of two PEs locating in two consecutive rows is the absolute value of horizontal

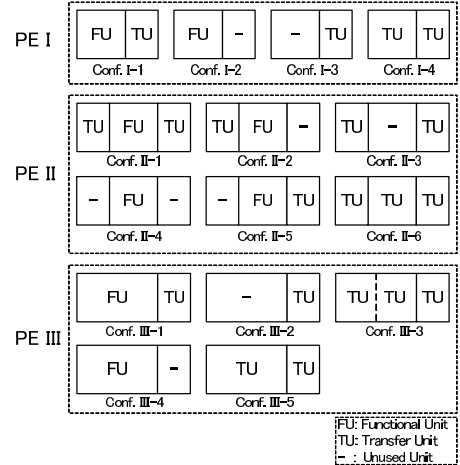


Fig. 4 Various configurations for the PEs.

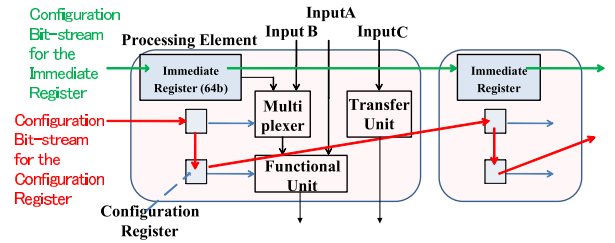


Fig. 5 PE's detailed structure including immediate register and reconfiguration facilities.

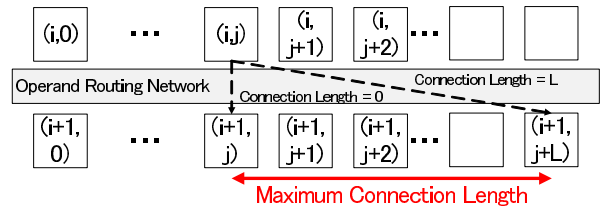


Fig. 6 Definition of maximum connection length (MCL).

distance between them. (e.g., in Fig. 6, for the two PEs located on (i, j) and $(i + 1, j + 2)$, the connection length is 2.)

Correspondingly, the MCL size is the maximum horizontal distance of two PEs located in two subsequent rows. ORNs should provide all outputs of a PE with totally $N_i \times (2 \times MCL + 1)$ connections to the PEs of consequent row, where N_i stands for the number of inputs for each PE. ORN's functionality is similar to a multiplexer, however ORNs are composed of cross-bar switches (CBs) as the basic elements of the RDP, CBs are also implemented by means of Josephson junctions (JJs) as the basic elements of the SFQ circuits [7]. A typical ORN structure locating between two rows of PEs is displayed in Fig. 7. T is the shift register which has 1-input/1-output. 2/1 CB is the crossbar with restricted functionality for 1-output. The crossbar-based ORN has a regular pipelined structure that does not limit the performance of the RDP and can be reconfigured

on the fly. It can also be easily re-designed and expanded for any given complexity by adding a necessary number of extra rows of crossbars [7]. A unified ORN is implemented between every two rows rather than multiple ORNs. Each PE can be connected to any of $2 \times MCL + 1$ PEs in the consecutive row via dedicated ORNs. Three factors including the PE architecture, MCL size, and the RDP width are basic factors which highly impact the ORN structure and size. In Fig. 7, PE I including three inputs and two outputs

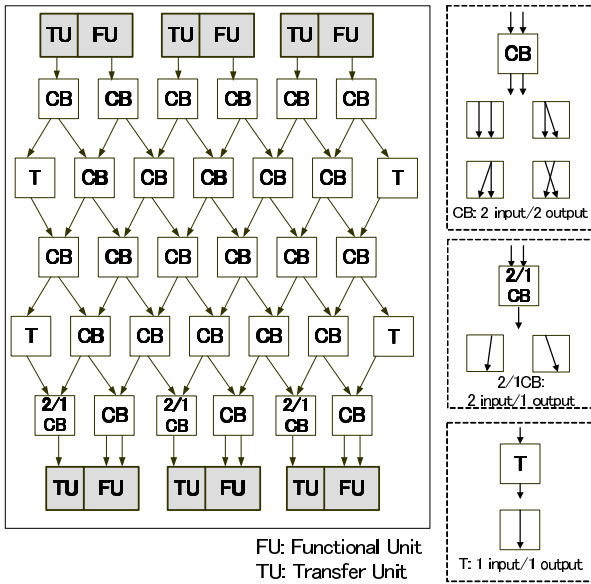


Fig. 7 ORN detailed architecture ($W = 3$ and $MCL = 1$).

is employed. Assuming W as the RDP width, the ORN will totally consist of $8 \times W \times MCL - 2 \times MCL + W$ CBs.

2.5 Reconfiguration Mechanism

RDP is an adaptable hardware that can be reconfigured at run-time using the configuration bit-streams which are statically generated within compilation phase for the DFGs (more details in Sect. 3.1). Figure 5 shows the architecture of a PE and how it can be reconfigured during the configuration phase. Besides programming immediate registers, multiplexers, and functional units inside PEs, ORNs should also be programmed using the configuration bits. In order to configure each component, the configuration bit-stream is serially transferred to configuration registers.

3. Design of SFQ-RDP Accelerator

3.1 Design Procedure

Compiler support for the RDP architecture is an essential requirement during the design procedure and hardware utilization phase. The extracted DFGs include a large number of operations which necessitates a sophisticated mapping tool to map DFGs onto the RDP and satisfy the architectural constraints as well. Figure 8 shows the tool chain which is similar to the conventional ones, however it has been customized for the RDP architecture [4]. The basic functionality of this flow is to generate configuration bit-streams and an executable code for the reconfigurable accelerator. Firstly, a hardware/software partitioning is performed on an

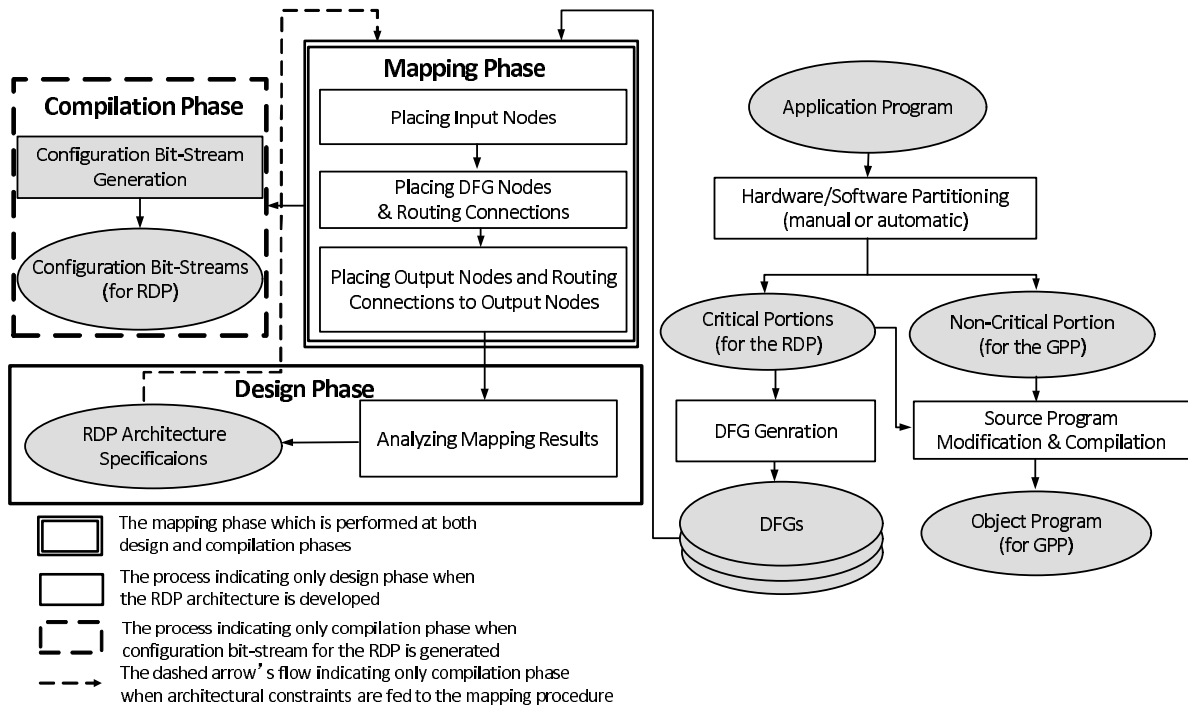


Fig. 8 The design and compilation flow for the RDP architecture.

Table 1 Results of design space exploration.

	PE I			PE II			PE III		
	MCL	W×H	Area (kJJ)	MCL	W×H	Area (kJJ)	MCL	W×H	Area (kJJ)
Layout I	8	24×11	29737	4	18×11	23567	5	22×12	28251
Layout II	7	26×12	21493	4	20×13	20444	4	22×14	19387
Layout III	12	24×18	36967	4	19×20	29589	7	19×18	25005

input application program. Critical segments of the program are isolated and the corresponding DFGs are generated. Due to the complexity of some attempted application programs. (e.g., quantum chemistry application program [8]) and the necessity of careful analysis and deep knowledge on the details of their implementations, the DFGs are extracted manually, however automatic generation is possible for some of application programs (e.g., finite difference equations [9]). The DFGs are mapped onto the RDP through placing DFG nodes on the PE array, routing interconnections as well as assigning input/output nodes to suitable I/O ports, while considering the RDP architectural constraints. Limitations originated from the basic architecture of the RDP and from SFQ circuits (e.g., sparse routing resources, difficulties of using memory elements and register file for intermediate data, etc.) necessitate particular placement and routing algorithms.

Within the mapping phase, one main objective is to minimize the MCL such that the interconnections could be successfully routed. We will show that the MCL should be minimized for reducing the ORN size. Placement and routing procedures ought to be iterated until a valid map satisfying the RDP constraints are generated. A configuration bit-stream associated with each of DFGs can be generated after completion of the mapping stage. Consequently, an executable code including noncritical segments of the program code and a piece of code for RDP architecture interfacing is generated. A part of compiler tools is customized to be utilized within the SFQ-RDP architecture design phase as shown in Fig. 8.

Determining the following essential architectural specifications of the SFQ-RDP is aimed, as mentioned in Sect. 2, during the design procedure:

- the total number of required PEs in the PE array,
- suitable layout and PE architecture,
- the number of PEs in each row (width),
- the number of PE rows (height),
- the number of input/output ports,
- optimized ORN architecture with minimum number of CBs,
- the total number of ORNs, and
- the size of ORNs including the numbers of inputs/outputs and CBs for each ORN which necessitates obtaining the MCL size as well.

During the design procedure, first all the DFGs extracted from target application programs are mapped onto various RDP architectures with different layouts (depicted in Fig. 2) and PE architectures (shown in Fig. 3) (mapping phase- Fig. 8). Then the results are analyzed to decide ap-

propriate architectural specifications for the RDP (design phase- Fig. 8). In the design phase, DFGs are mapped onto the RDP without forcing any constraint except those relating to the RDP basic properties, (e.g., unidirectional data flow over the PE rows, availability of routing resources only between subsequent rows, etc.), RDP layout and PE architecture. The total number of RDP resources including the number of PEs and ORNs as well as ORN size, hence overall area -in terms of the number of JJs- required for implementing each DFG can be calculated after mapping phase [7]. Consequently, specifications of the RDP architecture which can accommodate all the DFGs with minimum overall area can be determined (more details in Sect. 3.2).

3.2 SFQ-RDP Design Space Exploration and Result

Design of the RDP architecture entails a multitude of design parameters. Variety of design parameters indicates the high complexity of the design procedure and proves the requirements for a methodological approach. A major challenge within the design procedure is to find a right balance between the different quality requirements that a system has to meet. Design space exploration (DSE) is one approach to achieve it. DSE is the process of analyzing various implementation alternatives to identify an optimal solution.

Evidences from the experiments show that aiming the MCL reduction does not necessarily result in the overall area minimization. Obviously, the RDP area is the summation of PE array's and the ORN's areas. The structure and area of an ORN are influenced by the RDP width, the PE structure, and the MCL size. Therefore, ORN area is differently calculated for three PE types (more details can be found in [4]).

The SFQ-RDP design space is explored for various MCL values, three PE architectures, and three layouts to obtain the minimum RDP area. The SFQ-RDP total area for each design point is calculated and the design point giving the minimum overall area is selected among the evaluated ones.

Consequently, final specifications of the RDP can be derived from Table 1, in which the MCL size, width, and height are recognized as 4, 22, and 14, respectively. Also, the Layout II is the most suitable one and PE architecture III can provide reasonably sufficient routing resources and routing flexibility to keep the overall area the smallest.

4. Performance Evaluation

4.1 Performance Evaluation Method

In this section, we evaluate the SFQ-RDP system which

consists of a CMOS GPP, a CMOS memory, and the SFQ-RDP accelerator. Evaluation metrics are the performance and the power/performance ratio, which are compared with GPU, a well-known accelerator. Benchmarks are three sorts of finite difference method (FDM) application programs: (1) three-dimensional heat diffusion (3D-Heat), (2) two-dimensional heat diffusion (2D-Heat), and (3) two-dimensional finite difference time domain (2D-FDTD).

Performance evaluation of the SFQ-RDP system was accomplished, based on simple analytical model and simulation. Total execution time of a program T_{total} is obtained as shown in Eq. (1).

$$T_{total} = T_{gpp} + T_{rdp} + T_{oh} \quad (1)$$

T_{gpp} is the execution time obtained by executing application's object program on SimpleScalar [10] which is a cycle-accurate processor simulator. T_{rdp} is the execution time of the pipelined execution model which we developed in [5]. T_{oh} is an overhead time required for driving the SFQ-RDP accelerator. Architectural parameters used for the performance evaluation are shown in Table 2.

The GPUs used for comparisons are NVIDIA Tesla M2050 and GTX280 [11]. Performance evaluation results for the GPU are borrowed from [6], [12], [13]. In addition, memory bandwidths assumed for SFQ-RDP evaluation are 148.0 GB/s, 147.5 GB/s, and 158.0 GB/s for 3D-Heat, 2D-Heat, and 2D-FDTD, respectively [11]. These values are corresponding to the memory bandwidths available in the above GPU boards.

For the estimation of the power/performance ratio, the power consumption of SFQ-RDP accelerator, GPU processor, and GPU board are required. The power consumption of the SFQ-RDP accelerator is estimated as a sum of power

consumption of all the floating point units (FPU) and total ORNs. The power/performance ratio of SFQ-RDP's FPU with a single-precision adder is already reported as 4500 GFlops/W [14]. The power consumption of ORNs is estimated as 20.1% of that of total FPUs.

The power consumption of the GPU board is reported as 225 W and 236 W for NVIDIA Tesla M2050 and GTX280, respectively [11]. According to [15], the power consumption of DRAM memory module is around 30% of the entire GTX280 board, which corresponds to ~70 W. Assuming that the memory module consumes 70 W for both M2050 and GTX280, we can conclude that the GPUs themselves consume 155 W and 166 W, respectively.

Total power consumption of the SFQ-RDP system is estimated as the sum of the power consumed by memory, SFQ-RDP accelerator, and cryocooler which is a specific cooling system for the SFQ circuits. We assume that the power consumption of memory for SFQ-RDP accelerator and GPU are the same, and the cryocooler consumes 1000 times higher than that of the SFQ-RDP accelerator [16].

4.2 Performance Evaluation Results

The performance of 3D/2D-Heat and 2D-FDTD application programs on SFQ-RDP system and GPU are shown in Fig. 9. For 3D-Heat, 2D-Heat, and 2D-FDTD, the SFQ-RDP system is almost 18%, 20% and 25% less, respectively compared with the GPU. Although the performance of the SFQ-RDP system is lower than that of GPU, it is almost comparable to GPU.

The power/performance ratio results are shown in Table 3. In terms of the power/performance ratios, SFQ-RDP

Table 2 Configurations of GPP and SFQ-RDP processor.

GPP	Processor type	Out-of-order	
	Frequency	3.2 GHz	
	Instruction issue width	4instruction/clockcycle	
	Instruction decode width	4instruction/clockcycle	
	Level-1 data cache	size	64 KB
		line size	128B
		associativity	2way
		hit latency	2clockcycle
	Level-1 instruction cache	size	64 KB
		line size	64 B
		associativity	1way
		hit latency	1clockcycle
	Level-2 unified cache	size	4 MB
		line size	128 B
		associativity	4way
		hit latency	16clockcycle
Main memory	bus width	64 B	
	frequency	400 MHz	
	access latency	300 clockcycle	
RDP	SFQ-RDP frequency	80 GHz	
	Reconfiguration latency	30000 clockcycle ^a	
	Memory bandwidth	147.5,148.0,158.0 GB/s	
	No. of PEs in a row	22	
	No. of PE rows	14	

^a Based on SFQ-RDP frequency (80 GHz)

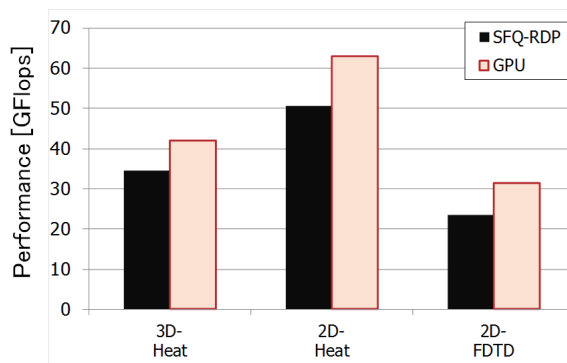


Fig. 9 Performance comparisons.

Table 3 Power/performance ratio comparisons.

Application program	Processor ¹		Board + Cryocooler ²	
	SFQ-RDP	GPU	SFQ-RDP	GPU
3D-Heat	430	0.27	0.23	0.19
2D-Heat	628	0.40	0.33	0.27
2D-FDTD	290	0.20	0.15	0.13

¹ Estimated with the power consumption of processor chip

² Estimated with power consumption of processor chip, memory, and cryocooler

outperforms GPU by 1600, 1600, and 1500 times for 3D-Heat, 2D-Heat, and 2D-FDTD, respectively. Therefore, the power/performance ratio of SFQ-RDP accelerator is extraordinarily higher than that of GPU. However, when the impact of power consumption of CMOS memory and cryocooler components is taken into account, advantage of the power/performance ratio decreases to around 1.2.

Current results show that SFQ-RDP system can be used as an efficient accelerator for FDM application programs in terms of power/performance ratio, which is important metric in high-performance computing systems. However, it is necessary to decrease total power consumption of cryocooler and memory by developing more efficient cryocooler and practical SFQ memory devices for practical usage.

5. Conclusions

We introduced a high-performance computing system, or SFQ-RDP system, which comprises a CMOS GPP, a CMOS memory, and an SFQ-RDP accelerator. We discussed the design procedure of SFQ-RDP accelerators, through which various PE architectures and layouts were examined and an optimized RDP architecture with minimum area in terms of the number of Josephson junctions was obtained.

To demonstrate the effectiveness of the SFQ-RDP system, it was compared with GPU using three finite difference method programs, in terms of performance and power/performance ratio. We obtained maximally 1600 times higher efficiency for 3D-Heat application program with respect to power/performance ratio. This ratio decreases to 1.2 when the power consumption of memory and cryocooler components is taken into account. Therefore, in terms of power/performance ratio, it is concluded that the SFQ-RDP accelerator can be used for practical scientific calculations especially for those based on finite difference methods. However, to achieve higher efficiency, decreasing total power consumption by developing practical SFQ memory and low-power cryocooler is inevitable.

Acknowledgement

This research was supported in part by the Research Area “Technology Innovation and Integration for Information Systems with Ultra Low Power” of JST (Japan Science and Technology Corporation)’s CREST Program (FY 2006–2012).

References

- [1] K. Likharev and V. Semonov, “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, March 1991.
- [2] O.A. Mukhanov, “Energy-efficient single flux quantum technology,” *IEEE Trans. Appl. Supercond.*, vol.21, pp.760–769, 2011.
- [3] N. Takagi, K. Murakami, A. Fujimaki, N. Yoshikawa, K. Inoue, and H. Honda, “Proposal of a desk-side supercomputer with reconfigurable data-paths using rapid single flux quantum circuits,” *IEICE Trans. Electron.*, vol.E91-C, no.3, pp.350–355, March 2008.
- [4] F. Mehdipour, H. Honda, K. Inoue, H. Kataoka, and K. Murakami, “A design scheme for a reconfigurable accelerator implemented by single-flux quantum circuits,” *J. Systems Architecture - Embedded Systems Design*, vol.57, no.1, pp.169–179, Jan. 2011.
- [5] H. Honda, F. Mehdipour, H. Kataoka, K. Inoue, and K. Murakami, “Performance evaluations of finite difference applications realized on a single flux quantum circuits-based reconfigurable accelerator,” *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2011 (APSIPA ASC 2011)*, Xi’an, China, Oct. 2011.
- [6] H. Honda, F. Mehdipour, H. Kataoka, K. Inoue, N. Yoshikawa, A. Fujimaki, H. Akaike, N. Takagi, and K. Murakami, “Enhancing performance and power efficiencies of scientific computing through an SFQ reconfigurable data-path processor,” *5th Superconducting SFQ VLSI Workshop 2012 (SSV2012)*, Nagoya, Japan, Dec. 2012.
- [7] I. Kataeva, A. Fujimaki, N. Yoshikawa, N. Takagi, K. Inoue, H. Honda, and K. Murakami, “An operand routing network for an SFQ reconfigurable data-paths processor,” *IEEE Trans. Appl. Supercond.*, vol.19, no.3, pp.665–669, June 2009.
- [8] S. Obara and A. Saika, “Efficient recursive computation of molecular integrals over cartesian Gaussian functions,” *J. Chem. Phys.*, vol.84, no.7, pp.3963–3974, April 1986.
- [9] W.H. Press, B.P. Flannery, S.A. Teukolsky, and T.W. Vetterling, *Numerical Recipes in C*, Cambridge University Press, 1988.
- [10] T. Austin, E. Larson, and D. Ernst, “SimpleScalar: An infrastructure for computer system modeling,” *Computer*, vol.35, no.2, pp.59–67, Feb. 2002.
- [11] NVIDIA, “Tesla M2050 and Tesla M2070/M2070Q Dual-slot Computing Processor Modules,” http://www.nvidia.com/docs/IO/43395/BD-05238-001_v03.pdf, Aug. 6. 2010.
- [12] T. Aoki and A. Nukada, *CUDA programming primer (Japanese)*, Kougakusya, Tokyo, 2009.
- [13] N. Takada, T. Shimobaba, N. Masuda, and T. Ito, “Speeding up of FDTD finite difference calculations by efficient use of GPU and shared memory,” (Japanese), *Proc. Forum of Information Science and Technology*, no.8, pp.457–462, Aug. 2009.
- [14] Y. Shimamura, Y. Yamanashi, N. Yoshikawa, A. Fujimaki, N. Takagi, and K. Takagi, “Design and implementation of SFQ floating-point multiplier and adder using 10 kA/cm² Nb process,” *Superconductivity Centennial Conference*, Den Haag, Netherlands, Sept. 2011.
- [15] S. Hong and H. Kim, “An integrated GPU power and performance model,” *Proc. 37th annual international symposium on Computer architecture*, pp.280–289, Saint-Malo, France, June 2010.
- [16] H. Ogihara, ed., *Survey of Cryogenic Engineering (Japanese)*, Tokyo Denki University, Tokyo, 1999.



Hiroshi Kataoka received the B.E. and M.E. degrees in engineering from Kyushu University, Japan in 2008 and 2010, respectively. Currently, he is a researcher of Research Institute for Information Technology in Kyushu University. He is the member of IPSJ (Information Processing Society of Japan).



Hiroaki Honda was born in Ishikawa prefecture, Japan in 1970. He received the Ph.D. degree in the graduate school of science of Hokkaido university, Japan in 2000. Currently, he is a research associate professor of Research Institute for Information Technology in Kyushu University. His research interests are quantum chemistry, high-performance computing and computer architecture. He is a member of the Chemical Society of Japan, Japan Society for Molecular Science, IPSJ and ACM.



Hiroyuki Akaike received the B.E., M.E., and Ph.D. degrees in electronic engineering from Nagoya University in 1990, 1992, and 1995, respectively. Since 1995, he has been with Nagoya University, where he is currently an associate professor. In 2003, he stayed in Superconductivity Research Laboratory, ISTE to be engaged in the development of Nb advanced fabrication process for a year. He has been involved in the research of superconducting devices and fabrication processes. He is a member of the Japan Society of Applied Physics.



Farhad Mehdipour received the B.Sc. degree from Sharif University of Technology in 1996, and the M.Sc. and Ph.D. degrees in Computer Systems Architectures from the Amirkabir University of Technology in 1999 and 2006, respectively. He was a visiting researcher at System LSI Lab, Kyushu University during November 2005 to June 2006. He joined the School of Information Science and Electrical Engineering at the Kyushu University as a researcher in December 2006. Since August 2010, he has

been an Associate Professor in E-JUST Center at Kyushu University. Dr. Farhad Mehdipour is a member of IEEE and his research interests include electronic design automation, high-performance and low-power micro-architectures, and cyber-physical systems.



Naofumi Takagi received the B.E., M.E., and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1981, 1983, and 1988, respectively. He joined Kyoto University as an instructor in 1984 and was promoted to an associate professor in 1991. He moved to Nagoya University, Nagoya, Japan, in 1994, and promoted to a professor in 1998. He returned to Kyoto University in 2010. His current interests include computer arithmetic, hardware algorithms, and logic design. He received Japan

IBM Science Award and Sakai Memorial Award of the Information Processing Society of Japan in 1995, and The Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology of Japan in 2005.



Nobuyuki Yoshikawa received the B.E., M.E., and Ph.D. degrees in electrical and computer engineering from Yokohama National University, 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. He is also interested in single-electron-tunneling devices,

quantum computing devices and cryo-CMOS devices. Prof. Yoshikawa is a member of the Institute of Electronics, Information and Communication Engineers of Japan, the Japan Society of Applied Physics, the Institute of Electrical Engineering of Japan, and the Institute of Electrical and Electronics Engineers.



Kazuaki Murakami received the B.E., M.E., and Ph.D. degrees in computer science and engineering from Kyoto University, Japan in 1982, 1984, and 1994, respectively. From 1984 to 1987, he worked for the Fujitsu Limited, where he was a Computer Architect of the mainframe computers. In 1987, he joined the Department of Information Systems of Kyushu University, Japan. He is currently a Professor of the Department of Advanced Information Technology, and also the Vice President of the Institute of Systems, Information Technologies and Nanotechnologies (ISIT).

He is a fellow of the IPSJ, and a member of the ACM, the IEEE, the IEEE Computer Society, the IEICE, and the JSIAM. His research interests include computer system architectures, high-performance computing (HPC), embedded technologies, cyber-physical systems (CPS), and so on.



Akira Fujimaki received the B.E., M.E., and Dr. Eng. degrees from Tohoku University, Sendai, Japan, in 1982, 1984, and 1987, respectively. He was a Visiting Assistant Research Engineer at the University of California, Berkeley, in 1987. Since 1988, he has been working on superconductor devices and circuits at the School of Engineering, Nagoya University, Nagoya, Japan, where he is currently a professor. His current research interests include single-flux-quantum circuits and their applications based on low- and high-temperature superconductors.

based on low- and high-temperature superconductors.