Optimizing the Architecture of SFQ-RDP (Single Flux Quantum-Reconfigurable Datapath)

Mehdipour, Farhad
Graduate School of Information Science and Electrical Engineering, Department of Informatics, Kyushu University

Honda, Hiroaki
Institute of Systems, Information Technologies and Nanotechnologies

Kataoka, Hiroshi
Graduate School of Information Science and Electrical Engineering, Department of Informatics, Kyushu University

Inoue, Koji
Graduate School of Information Science and Electrical Engineering, Department of Informatics, Kyushu University

他

Optimizing the Architecture of SFQ-RDP (Single Flux Quantum- Reconfigurable Datapath)

Farhad Mehdipour, Hiroaki Honda, Hiroshi Kataoka, Koji Inoue, and Kazuaki Murakami

Abstract—A large-scale reconfigurable data-path (LSRDP) processor based on single-flux quantum circuits is designed to overcome the issues originating from the CMOS technology. The LSRDP micro-architecture design procedure and its outcome will be presented in this paper.

I. INTRODUCTION

FOR providing high computational power to individual researchers in various scientific areas a desk-side tera-flop scale computer has been introduced [5] which consists of a CMOS general purpose processor, a memory and a single-flux quantum (SFQ)- based Reconfigurable Large-Scale Data-Path processor (SFQ-LSRDP) as an accelerator Fig. 1). A SFQ circuit is based on superconductor technology featuring low-power consumption, high speed and smaller switching energy. It is also suitable for pipeline processing and there is no additional cost for latch implementation. LSRDP constructed based on SFQ circuits is a pipelined architecture comprising a two-dimensional array of processing elements (PEs) such that one PE can be connected through Operand Routing Networks (ORNs) to one or more PEs in the next row. Since the LSRDP is aimed to target various scientific applications, it should be featured with dynamically reconfigurable PEs and ORNs. Each PE is supposed to include a functional unit (FU) and a transfer unit (TU) for transferring data to inconsequent rows. Moreover, ORNs are implemented using SFQ cross-bar switches [1].

II. LSRDP DESIGN PROCEDURE

A. Tool Chain

Fig. 2 shows the proposed tool chain for the LSRDP design and compilation phases. In the first stage, a hw/sw partitioning is performed on the input application manually. In this stage, critical segments of the code are isolated and the corresponding DFGs are generated. DFGs are mapped onto the LSRDP through placing DFG nodes on the PEs, routing interconnections among them as well as positioning input/output nodes on the proper ports.

Placement and routing procedures should be repeated until a valid map satisfying the LSRDP constraints is generated. Configurations' bit-stream corresponding to each one of DFGs
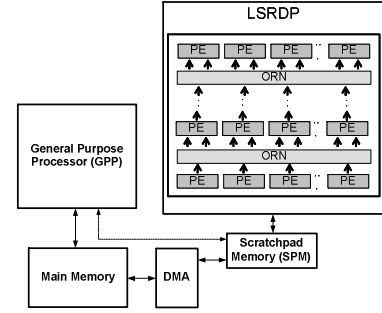
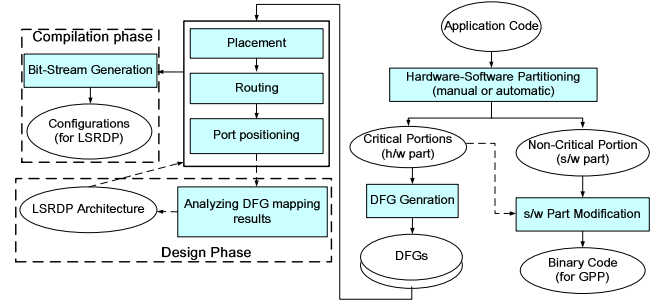Fig. 1.   Overall architecture of the SFQ-LSRDP computer

Fig. 2.   A detailed outline of the proposed hw/sw compilation and design flow

can be generated after completion of the mapping stage. An executable code including non-critical segments of the application code and a piece of code for interfacing the LSRDP ought to be produced. A part of compiler tools can be customized for utilizing in the LSRDP design phase as shown in Fig. 2.

B. DFG Extraction

DFGs are manually extracted by the designer from the applications. Four applications are attempted as benchmark scientific applications including: one-dimensional heat (referred as Heat) and vibration equations (Vibration), two-dimensional Poisson equation (Poisson) [4], and recursion calculation part of electron repulsion integral (ERI [3]) as a quantum chemistry application. All calculations consist of ADD, SUB, and MUL operations. As basic blocks of applications might be small and mapping small DFGs is inefficient therefore, larger DFGs are generated through expanding small basic DFGs.

C. Design Stages

Different methods can be used for determining LSRDP's detailed architectural specifications. One approach is to utilizing quantitative analysis of DFGs and extracting their

properties. In this approach, various characteristics of DFGs are quantitatively obtained and utilized in the design procedure of LSRDP architecture. The design flow is an iterative procedure of gathering statistics and analysis of results, therefore, the designer should decide the priority of design parameters in the first step. Then, for determining each design parameter, DFGs should be mapped onto the LSRDP and outcome should be analyzed. There is no limitation in the initial architecture and the mapping process is performed without forcing any constraint. In the next stage, the results of mapping should be analyzed by the designer to decide an appropriate value for the intended parameter. This interactive process is repeated until fixing entire specifications of the architecture.

D. DFG Classification

DFGs generated from various applications have different qualities in terms of size, number of inputs/outputs, connection complexity and etc. For each application at least one DFG is generated. Moreover, for some DFGs different versions are introduced to be able to meet different architectural specifications. Among various implementations of a DFG, usually a larger DFG is preferred due to higher achievable speedup. Three factors including the available PEs in the LSRDP and the number of input/output ports are considered as the main criteria for classifying DFGs. Four classes including Small (S), Medium (M), Large (L) and X-Large (XL) are identified. TABLE I shows the values determined for above parameters and the list of DFGs in each class.

TABLE I.   **Various LSRDP configurations and a list of classifieds DFGs**

	# of PEs	# of Inputs	# of Outputs	# of DFGs
LSRDP-S	128	19	12	Heat (3), Poisson (1), Vibration (2), ERI (4)
LSRDP-M	512	19	12	Heat (1), Poisson (1), Vibration (1), ERI (4)
LSRDP-L	1024	38	24	Heat (2), Poisson (1), Vibration (2), ERI (5)
LSRDP-XL	> 1024	64	52	Heat (1), Poisson (1), Vibration (2), ERI (5)

E. Placement and Routing

During mapping process, firstly, DFG nodes are placed on appropriate positions (PEs) in the LSRDP. Here, the mail goal is to minimizing maximum horizontal connection length that directly impacts the ORN sizes. Routing process is the next stage that determines connections between the PEs in the LSRDP by means of ORNs and TUs.

F. I/O Positioning

The last step of mapping process is to locating input/ouput nodes of DFG on the appropriate port positions around LSRDP. In the LSRDP it is assumed that input and ports are located in the top and bottom borders, respectively. Between the first/last LSRDP rows and input/output ports, ORNs are utilized as routing resources as well.

G. Connection-length minimization

Connection length is calculated based on the distance of source and destination PEs of a net in horizontal direction. Maximum ORN size is decided based on maximum connection length (MCL) value among whole DFGs. Since the number of ORNs strongly impacts the LSRDP various costs, reducing ORN size is an important challenge. The number of inputs required for an ORN would be 2 x (2 x MCL+1). It is doubled

because each PE has two outputs. In addition, three outputs are assigned for each ORN corresponding to three inputs of PEs.

III. LSRDP GENERAL ARCHITECTURE AND SPECIFICATIONS

The design procedure introduced above is employed for the LSRDP and the outcome is as TABLE II.

TABLE II.   **Specifications of the designed LSRDP**

Functional units		ADD/SUB, MUL		
Layout		Checker (ADD/SUB and MUL in every other PE)		
Operations		64-bit floating point		
Processing structure		Pipelined		
PE structure		FU, T, FU+T, T+T		
LSRDP Type		**Small**	**Medium**	**Large**
No. of inp/out ports		19/12	19/12	38/24
Width/Height		16/16	32/16	58/25
Conf. bit-stream size	Imm. Regs	13*13*64	32*16*64	58*25*64
	ORNs	13*BSS(ORN)	32* BSS(ORN)	58*BSS(ORN)
	PEs	13*13*2	32*16*2	58*25*2
ORN	inputs, outputs	22 , 3	26 , 3	26 , 3
	# in a row	13	32	58
	Structure	Cross-bar switch		
	Conn. Type	One-directional		
Internal memory		Type	Immediate registers	
		Size and count	64-bit registers, One reg. for each PE	
		Commu'n mechanism	Serial	
External memory		No. of memory modules	16	
		Date trans. rate	1800Mbps/pin	
		Overall data trans. rate	24 GB/s	
		Mem. to LSRDP bus width	64 bit	
		Channels per module	Two	
Reconf. mechanism		Bit serial configuration through a serial chain		

IV. CONCLUSION

A high-performance computer comprising an accelerator implemented by superconducting circuits was introduced which is suitable for executing massive computational-intensive scientific applications. A quantitative design procedure was followed and a number of data flow graphs extracted from scientific application were subjected to the design procedure. LSRDP architectures with various dimensions have been designed and evaluated. Evidences from experiments demonstrate that the high-performance computer equipped with SFQ-LSRDP is promising for achieving considerable performances.

REFERENCES

[1] A. Fujimaki, S. Iwasaki, K. Takagi, R. Kasagi, I. Kataeva, H. Akaike, M. Tanaka, N. Takagi, N. Yoshikawa, K. Murakami, Demonstration of an SFQ-Based Accelerator Prototype for a High-Performance Computer, Applied Superconductivity Conference (ASC 2008), 2EZ01, Chicago, 2008.

[2] J. Makino, K. Hiraki and M. Inaba, GRAPE-DR: 2-Pflops massively-parallel computer with 512-core, 512-Gflops processor chips for scientific computing, SuperComputing (SC'07), 2007.

[3] S. Obara, Efficient recursive computation of molecular integrals over Cartesian Gaussian Functions, J. Chem. Phys., Vol.84, pp.3963, 1986.

[4] W.H. Press, B.P. Flannery, S.A. Teukolsky, and T.W. Vetterling, Numerical Recipes in C, Cambridge University Press, 1988.

[5] 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. on Elec., E91-C(3):350-355, 2008.

[6] W. Wulf and S. McKee "Hitting the Memory Wall: Implications of the Obvious," ACM SIGArch Computer Architecture News, 23 (1):20-24, 1995.