

# Body Bias Optimization for Variable Pipelined CGRA

Takuya Kojima, Naoki Ando, Hayate Okuhara, Ng. Anh Vu Doan and Hideharu Amano  
Dept. of Information and Computer Science, Keio University, Yokohama, Japan  
Email: wasmii@am.ics.keio.ac.jp

---

**Abstract**—Variable Pipeline Cool Mega Array (VPCMA) is a low power Coarse Grained Reconfigurable Architecture (CGRA) based on the concept of CMA (Cool Mega Array). It implements a pipeline structure that can be configured depending on performance requirements, and the silicon on thin buried oxide (SOTB) technology that allows to control its body bias voltage to balance performance and leakage power. In this paper, we propose a methodology to optimize exactly with an Integer Linear Program the VPCMA body bias while considering simultaneously its variable pipeline structure. For the studied applications, we evaluate that it is possible to achieve an average reduction of energy consumption of 19.3% and 11.8% when compared to respectively the zero bias (without body bias control) and the uniform (control of the whole PE array) cases, while respecting performance constraints. Besides, with appropriate body bias control, it is possible to extend the possible performance, hence enabling broader trade-off analyzes between consumption and performance. These promising results show that applying an adequate optimization technique for the body bias control while simultaneously considering pipeline structures can not only enable further power reduction than previous methods, but also allow more trade-off analysis possibilities.

## I. INTRODUCTION

Recent advanced IoTs (Internet of Things) and wearable computing require a relatively high performance with extremely low energy consumption. CGRA (Coarse-Grained Reconfigurable Architecture) is a candidate of accelerators for such devices thanks to its high degree of performance per limited energy budget. The principle of CGRAs consists of an array of small processing elements (PEs) which can execute simple computational operations, and distributed memory modules connected together with an interconnection network. Highly efficient computing can be performed by changing the type of operations and their interconnection.

VPCMA (Variable Pipeline Cool Mega Array) [1] is a low power CGRA based on the concept of CMA (Cool Mega Array) [2]. It provides a large PE array without dynamic reconfiguration and a tiny microcontroller with banked data memory. The pipeline structure in the PE array can be configured so as to fit target algorithms and required performance.

This work has been done within the “Ultra-Low Voltage Device Project” of LEAP funded and supported by METI and NEDO. It has also been supported by the VLSI Design and Education Center (VDEC) from the University of Tokyo in collaboration with Cadence Design Systems, Inc. The stay of Ng. Anh Vu Doan in Keio University has been supported by the Erasmus Mundus EASED programme (Grant 2012-5538/004-001) coordinated by CentraleSupélec.

Also, VPCMA uses the Silicon on Thin Buried Oxide (SOTB) technology, a type of fully depleted silicon on insulator (FD-SOI). So a balance between performance and leakage power can be kept by controlling the body bias voltages.

Although the basic trade-off of changing the pipeline structure of VPCMA has been discussed in [1], body bias control has not been applied. Here, we propose a bi-objective optimization method of both energy and performance considering simultaneously the body bias voltages, the pipeline structure, and the target application. At first sight, the problem may seem complex, and one could consider to apply multi-objective metaheuristics such as genetic algorithms to tackle it. However, while these methods have successfully been used for various similar cases, they do not always provide optimal solutions and we propose in this work a model and analysis of this problem that allow to solve it quickly by using an ILP (Integer Linear Program) model, with guarantee of optimality. All optimization results are based on parameters from an existing developed design, and the results can be directly applied to a real chip now under evaluation.

The rest of the paper is organized as follows. Section II introduces VPCMA, SOTB process technology, and fundamental body bias control for VPCMA. Then, an optimization method is proposed in Section III with preliminary evaluation for building an ILP. The optimization results are presented in Section IV. After discussion comparing with related works in Section V, we conclude with a brief summary in Section VI.

## II. VARIABLE PIPELINE COOL MEGA ARRAY (VPCMA)

### A. The architecture of VPCMA

The VPCMA is classified into Straight Forward CGRAs (SF-CGRAs), a class of simple CGRAs. They consist of a pipelined array of processing elements (PEs), memory modules and networks for transferring data between them. Data are read out from the memory modules, transferred to the input of pipelined array through a permutation network, and the results are written back to the memory modules with another permutation network.

The VPCMA architecture is a simple SF-CGRA that focuses on reducing any energy usage other than that required for computation. The PE array is built with a simply pipelined combinatorial circuits to eliminate the power needed to distribute a clock to each PE. As shown in Fig. 1, the VPCMA

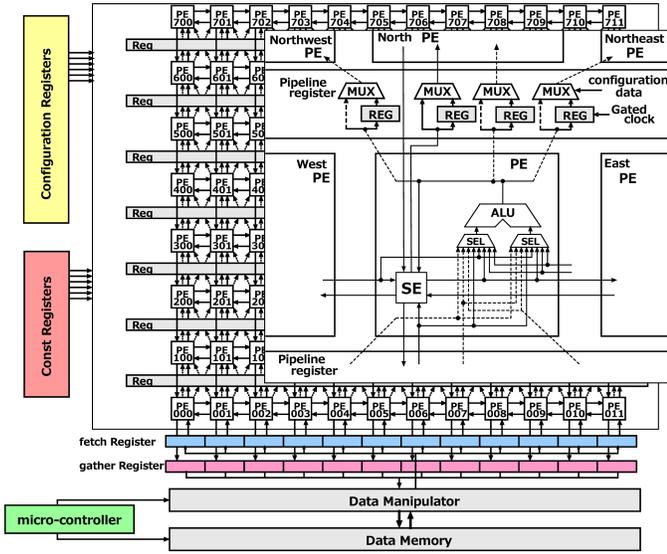


Fig. 1: Diagram of VPCMA with details of PE and Pipeline Registers

consists of a large PE array with pipeline registers, a micro-controller and banked data memory. The pipeline registers are placed between every row of the PE array. As they are all independently switchable, the VPCMA can freely change its pipeline structure. The implementation of the registers is also illustrated. This structure enables the implementation of various application programs without a power-hungry dynamic reconfiguration in the PE array.

### B. SOTB

SOTB is classified as an FD-SOI technology in which transistors are formed on thin buried oxide (BOX) layer. It has been developed so that the delay and leakage power consumption can be optimized by controlling the bias voltage to the body,  $V_{BN}$  and  $V_{BP}$  respectively given to NMOS and PMOS transistors. Three possibilities of biasing are available. First, there is the zero-bias, where  $V_{BN}$  and  $V_{BP}$  are equal to the source voltage ( $V_S$ ), which means that the transistor works with its normal voltage  $V_{TH}$ . Second, we have the reverse-bias ( $V_{BN} < V_S$  and  $V_{BP} > V_S$ ) where  $V_{TH}$  is increased, and the leakage current is exponentially reduced, while the delay is increased. Finally, with the forward-bias ( $V_{BN} > V_S$  and  $V_{BP} < V_S$ ),  $V_{TH}$  is decreased and the leakage current is increased, while the operational speed is enhanced.

### C. Row-level body bias control for VPCMA

In the original paper on VPCMA [1], body bias control was not considered and only zero bias was used to study the benefit of a pipeline structure. In this work, we propose a row-level body bias domain for the PE array, as shown in Fig. 2, to balance the delay time of each pipeline stage and have more flexible choices on the bias voltages. By using a row-level body bias control, we can apply a reverse (forward) body bias to every stage whose delay is shorter (longer) than the largest (shortest) one until they become (nearly) equal.

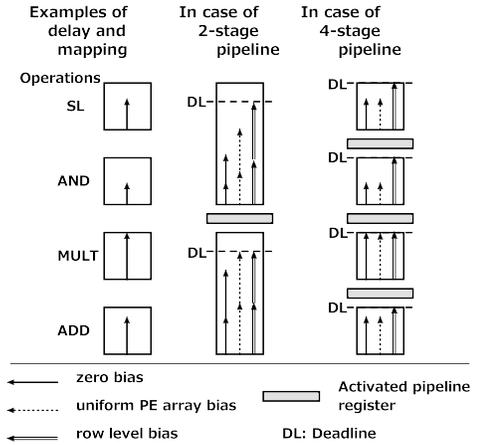


Fig. 2: Row-level body bias control with pipeline registers (2 and 4 stages)

## III. PROBLEM DEFINITION AND PROPOSED METHOD

Assuming that the pipeline structure and the body bias voltages are controlled simultaneously, there are several possibilities of trade-off as shown in Table I. More advanced analyses are therefore required to assess the trade-off possibilities between performance and power consumption which both depends differently on the pipeline registers configuration and the body bias control. Therefore, we propose in this paper to optimize the choices on the body bias control while simultaneously considering the pipeline structure.

TABLE I: Trade-off between performance and power

	Number of pipelined stage	
	large	small
Performance	high	low
Dynamic power of register and clock tree	increases	decreases
Dynamic power of the glitches	decreases	increases
	Body bias voltage	
	forward bias	reverse bias
Performance	low	high
Static power	decreases	increases

On basis of the aforementioned trade-off information, we can define the problem as the following bi-objective optimization problem: given an application, how to optimize the power consumption and the performance of the VPCMA with choices on simultaneously the body bias voltages and the pipeline structure.

The equations required to model this problem can be formulated as follows:

$$V_{ij} = \begin{cases} 1 & \text{if the } i\text{-th PE row is set with } V_{BN_j} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$preg_k = \begin{cases} 1 & \text{if the } k\text{-th pipeline register is used} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$P_{dyn} = f_{req} \times (E_{comb}(preg_k) + \sum_{k=0}^6 (E_{reg} + E_{clk})preg_k) \quad (3)$$

$$P_{stat} = \sum_{i=0}^7 \sum_{j=0}^{12} P_{leak,row,j}(V_{ij}) + P_{leak,reg} + P_{leak,clk} \quad (4)$$

$$D_l = \sum_{i=0}^7 \sum_{j=0}^{12} D_{PE,j} V_{ij} \quad (5)$$

where:

- $V_{ij}$  represents the body bias assignment, with  $i = \{0, 1, \dots, 7\}$  since the VPCMA possesses eight rows, and  $j = \{0, 1, \dots, 12\}$  since there are 13 possible voltages as it will be explained in the following
- $VBN_j$  is the  $j$ -th available body bias voltage
- $preg_k$  represents the configuration of the  $k$ -th pipeline register, with  $k = \{0, 1, \dots, 6\}$  since the VPCMA implements 7 registers
- $P_{dyn}$  and  $P_{stat}$  are respectively the dynamic and static power of the PE array (considering body bias control and pipeline structure)
- $E_{comb}$ ,  $E_{reg}$ , and  $E_{clk}$  are the energy consumption of respectively the combinatorial circuits, a pipeline register, and clock tree
- $P_{leak,row,j}$ ,  $P_{leak,reg}$ , and  $P_{leak,clk}$  are the leak power of respectively a row on  $VBN_j$ , a pipeline register, and the clock tree
- $D_l$  and  $D_{PE,j}$  are the delay time of respectively the  $l$ -th datapath and a PE supplied with  $VBN_j$ ;  $D_l$  is therefore calculated as the sum of the delays caused by the PEs located in the  $l$ -th datapath.

In this work, the optimization problem is to minimize the sum of  $P_{dyn}$  and  $P_{stat}$ . The parameters in the above model such as  $P_{leak,row,j}$  or  $P_{comb}(preg_k)$  are obtained by several simulations for the four applications listed in Table II. The design used in the simulations are based on a real VPCMA chip.

TABLE II: Simulated applications

Application	Description
gray	24 bit (RGB) gray scale
sepia	8 bit sepia filter
af	24 bit (RGB) alpha blender
sf	24 bit (RGB) sepia filter

The size of the solution space is  $2^7 \times 13^8$ . Indeed, the VPCMA can configure  $2^7 = 128$  patterns of pipeline structure since for each of the seven registers, it is possible to choose to use it or not. For the row level body bias, each of the eight rows in the PE array can select among thirteen possible voltages (denoted  $VBN_0, \dots, VBN_{12}$ ), so there are  $13^8$  possibilities. As a test, for one pipeline structure, it takes 3 hours to elicit and simulate all these possibilities on a 1.6GHz dual-core Intel Core i5 with 8GB of DDR3 RAM.

Given the size of the solution space and the complex formulation of some equations (e.g.  $P_{dyn}$ ), techniques such as metaheuristics could be applied. However, a close examination of the problem shows that it is possible to formulate this problem as an ILP (Integer Linear Problem) which, unlike metaheuristics, gives a guarantee of optimality. Indeed, when the pipeline structure is fixed, that is,  $preg_i$  is fixed,  $P_{dyn}$  is constant. Therefore, with the remaining equations being linear, it is possible to formulate this problem as only 128 ILPs (one for each pipeline structure). Moreover, its bi-objective nature can be simplified by considering the performance as a constraint that needs to be reached. Since the design focus of the VPCMA is low power, the problem can be re-formulated as follows: given an application and a fixed pipeline structure, how to optimize the power consumption of the VPCMA while reaching required performance with choices on the body bias voltages.

#### A. ILP model

The ILP can then be formulated as follows:

$$\min P_{stat,rows} = \sum_{i=0}^7 \sum_{j=0}^{12} P_{leak,row,j}(V_{ij}) \quad (6)$$

subject to

$$\sum_{i=0}^7 V_{ij} = 1 \quad \forall j = \{0, 1, \dots, 12\} \quad (7)$$

$$D_l \leq D_{req}, \quad \forall \text{datapath } l \quad (8)$$

$$V_{ij} = \{0, 1\}, \quad \forall i = \{0, 1, \dots, 7\}, \quad \forall j = \{0, 1, \dots, 12\} \quad (9)$$

where the constraint (7) ensures that the row level body bias is respected (same body bias for the PEs on the same row) and (8) expresses that the required performance  $D_{req}$  is reached. It is worth noting that  $P_{leak,reg}$  and  $P_{leak,clk}$  are constant (not controlled by body bias) and therefore do not have to be included in the objective function.

## IV. EVALUATION

To analyze the possibilities of the proposed method, we perform the power optimization for several different performance requirements and for each application described in Section III. To evaluate the energy reduction achieved by the proposed method, we simulate other policies of body bias control as comparison basis:

- control for the whole PE array (uniform)
- no body bias control (zero bias)

As shown in Fig. 3, using the body bias control allows to reach higher achievable performance. For instance, without body bias control (zero bias), the performance cannot exceed  $3.12 \times 10^9$ . However, both the uniform control and the proposed method allow higher performance values. Also, unlike the uniform control, the proposed method can keep a steady increase of the power even at high performance, since

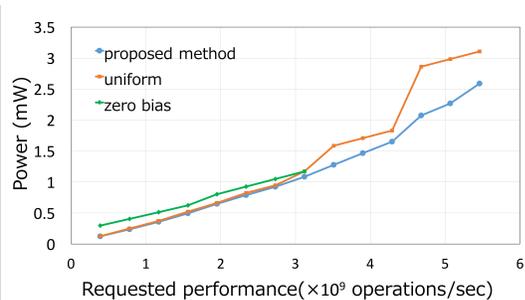


Fig. 3: Comparisons between each methods ( $VDD = 0.55$  V)

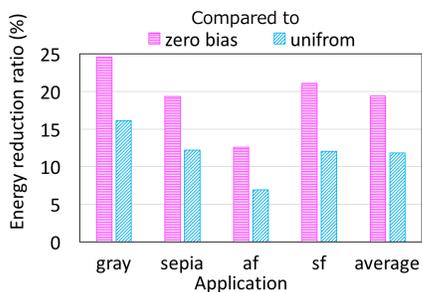


Fig. 4: Energy reduction ratio for each application ( $VDD = 0.55$  V)

forward bias has to be applied only to the row which causes a bottleneck in the critical path.

To compare the energy between different methods, the average energy of all performances is calculated for each application and for each methods. Fig. 4 illustrates the reduction ratio of the energy between the proposed method and the other two policies. With the proposed method, it is possible to achieve an energy consumption of 24.5% and 16.1% lower than respectively the zero bias and the uniform cases (best reduction with “gray” application). In average, the consumption is 19.3% and 11.8% lower than respectively the zero bias and the uniform cases.

In terms of algorithmic performance, it is worth noting that the proposed method gives a guarantee of optimality and is indeed faster than an explicit elicitation. Compared to the previously-mentioned 3 hours to simulate all the possibilities for a fixed pipeline structure, the ILP takes around 4 minutes in the worst simulated case.

## V. RELATED WORKS

Variable pipeline structure is widely used to select various trade-off between the performance and power. It was applied to a CPU [3], H.264 decoder [4] and routers [5], [6]. Some of them control the power supply voltage when the pipeline structure is changed but a body bias control has not been applied.

Variable body bias control technique has been applied to a dynamically reconfigurable processor [7] and the CMA [8]. However, the former focuses on finding the optimal body bias domain size at the design stage whereas the latter is also

searching for the optimal size of body biasing, but targeting instead groups of PE array with combinatorial circuits, and a genetic algorithm which cannot give guarantee of optimality was used. They did not consider the pipeline processing and so the optimization only focused on body biasing. Since the goal of this paper is multi-objective optimization of both the power and the performance considering simultaneously body bias control and pipeline structure, the optimization methods and results are completely different.

## VI. CONCLUSION

In this paper, we have proposed a methodology based on ILP to optimize simultaneously the power consumption and the performance of a variable pipelined CGRA, the VPCMA, while considering both body biasing and pipeline structure. The simulation results demonstrated that the proposed method allows to reach lower consumption than previous work while meeting required performance. Moreover, the range of possible performance can be stretched with appropriate body biasing and pipeline structure, hence enabling broader trade-off analyzes between consumption and performance.

As future works, although all the parameters used for the simulations are based on an existing developed design, tests on a real chip (now under evaluation) have yet to be carried out. Besides, it is worth noting that the optimization is currently performed considering a fixed application mapping on the PEs. Since the body bias control and the pipeline structure both depend on the mapping, a change on the latter (for instance, a more compact mapping) may alter the optimality of previously-found bias voltages and pipeline registers configuration. An application mapping tool considering both body bias control and pipeline structure would allow even further optimization and analyzes.

## REFERENCES

- [1] N. Ando, K. Masuyama, H. Okuhara, and H. Amano, “Variable pipeline structure for coarse grained reconfigurable array cma,” in *2016 International Conference on Field-Programmable Technology*, 2016, pp. 231–238.
- [2] N. Ozaki, Y. Yoshihiro, Y. Saito, D. Ikebuchi, M. Kimura, H. Amano, H. Nakamura, K. Usami, M. Namiki, and M. Kondo, “Cool megarray: A highly energy efficient reconfigurable accelerator,” in *Field-Programmable Technology (FPT), 2011 International Conference on*. IEEE, 2011, pp. 1–8.
- [3] T. Shimada, et al., “A novel low-power processor with variable pipeline control,” in *Proc. of IEEE International Symposium on VLSI-DAT*, 2008, pp. 263–266.
- [4] Chanhoo Lee, and Seohoon Yang, “Design of an H.264 decoder with variable pipeline and smart bus arbiter,” in *Proc. of ASP-DAC2012*, Jan. 2012, pp. 407–412.
- [5] H. Matsutani and Y. Hirata and M. Koibuchi and K. Usami and H. Nakamura and H. Amano, “A multi-Vdd dynamic variable pipeline on-chip router for CMPs,” in *Proc. of ASP-DAC2012*, Jan. 2012, pp. 407–412.
- [6] C. Y. Lee and N. K. Jha, “Variable-Pipeline-Stage Router,” in *IEEE Trans. on VLSI system*, vol. 21, no. 9, Jan. 2013, pp. 1669–1682.
- [7] J.M. Kuehn, H. Amano, O. Bringmann, W. Rosenstiel, “Leveraging FDSOI through Body Bias Domain Partitioning and Bias Search,” in *Proc. of 53rd Design Automation Conference*, Jul. 2016.
- [8] Y. Matsushita, H. Okuhara, K. Masuyama, Y. Fujita, R. Kawano, and H. Amano, “Body bias grain size exploration for a coarse grained reconfigurable accelerator,” in *2016 26th International Conference on Field Programmable Logic and Applications (FPL)*, Aug 2016, pp. 1–4.