# *Energy Efficient Synchronous - Asynchronous Circuit-Switched NoC*

Sandy A. Wasif[1], Salma Hesham[1,2], Diana Goehringer[3], Klaus Hofmann[4] and Mohamed A. Abd El Ghany[1,4]

[1]Electronics Department, German University in Cairo, Egypt

[2] Department of Electrical Engineering and Information Technology, Ruhr-University Bochum, Germany

[3]Adaptive Dynamic Systems Chair, TU Dresden, Germany

[4]Integrated Electronic Systems Lab, TU Darmstadt, Germany

E-mails: {sandy.abdelmalak,salma.hesham, mohamed.abdel-ghany}@guc.edu.eg, diana.goehringer@tu-dresden.de, klaus.hofmann@ies.tu-darmstadt.de

***Abstract*— The advancement in technology has led to the integration of multiple cores onto a single chip, calling for new means of on-chip communication substituting the simple physical wires. Networks-on-Chip (NoCs) were introduced as the emerging solution for a high performance and scalable communication infrastructure in the multi-core platforms. However, with the increasing chip complexity heading to thousand cores, NoCs are significantly contributing in the chip power consumption. This paper presents the energy efficient sync-async Circuit-switched NoC (CS-NoC) with a synchronous control sub-router and an asynchronous data transfer sub-router. CS was chosen to benefit from the fact that CS NoC presets the data-path. The proposed NoC is synthesized using Synopsys Design Compiler for 65nm technology. The obtained results for 65nm show a reduction of 80% in the dynamic power and 7% for the leakage power as compared to pure synchronous CS-NoCs.**

***Keywords—Circuit Switching; NoC; Power consumption;***

## I. INTRODUCTION

The decreasing transistor size has opened the door for integrating multiple cores into one chip which is known as Multi-processor system-on-chip (MPSOC). This enhancement in technology may lead to thousand cores on a single chip [1]. The communication between this large numbers of cores can no longer be sustained by physical interconnections as it will limit the performance. Researchers introduced the Network-on-chip (NoCs), a reconfigurable network that offers better performance [2]. However NoCs consume a large percentage of the total chip power [3]. This leads to the need of designing NoCs with lower power consumptions. Circuit switching NoCs can be designed to provide lower power consumption since the data path is preset [4]. Another way to further reduce power consumptions is to introduce asynchronous NoCs since they have lower power consumption than synchronous designs as no dynamic power is consumed when there is no transmission [5].

Several researchers tackled different design techniques to reduce the power consumption. The techniques implemented varied from Dynamic voltage/Frequency scaling (DVFS) [6] to power gating (PG) [3]. DVFS relies on adjusting the voltage and frequency of the system to optimize power consumption, while PG simply shuts off the blocks that are not currently used. DVFS primarily targets dynamic power while PG aims for leakage power. In [7], the dark silicon was utilized in CS NoCs to benefit from the separation of the control and data sub-router, it showed significant reduction in power. In [8], a two layered NoC is used; one layer operates at near threshold voltage to minimize power consumption at low communication load and the other operates at the nominal voltage only at high communication load.

Asynchronous designs eliminate the need for clock signal to synchronize the circuit operation. It employs various handshaking protocols, such as the request-acknowledge [9]. A request is sent to initiate communication, when the communication is terminated the acknowledgment signal is sent back. One of the common approaches to design NoCs is the globally asynchronous, locally synchronous (GALS), yet it needs many synchronization blocks from the source to the destination [10]. In [11], an asynchronous router was designed and compared to the synchronous one and it showed significant reduction in power consumption. However in their work the entire design is asynchronous; internally and externally as well. In [12], the synchronous and asynchronous routers were also compared and it concludes that asynchronous routers are better options for limited power budget while synchronous routers are more suitable for real-time applications.

In this work, we propose the sync-async CS NoC; it is a CS NoC with synchronous control and an asynchronous data transfer. CS was chosen due to the independence between the control and the data paths. This work introduced modifications on the traditional CS router by having an asynchronous data transfer between routers to benefit from the lower dynamic power consumption while maintaining a normal synchronous control sub-router. The proposed design also benefits from eliminating the common synchronous design problems such as clock skew. This design allows data to be transferred from the source to the destination with the source rate without any need for synchronization blocks. The proposed design provides lower area and power consumption as compared to the all synchronous CS NoC; however it has higher data arrival time.

This paper is organized as follows: the basic CS NoC architecture is described in section II. The proposed sync-async CS NoC architecture is presented in section III. The

comparisons and results are presented in section IV. Finally, section V discusses the conclusion and future work.

## II. CIRCUIT-SWITCHED NOC

### A. Basic Operation

For CS NoCs, a path is reserved from the source to destination for each transfer request. The sender requests to send data to a specific destination and the control part of the router sends a control flit to reserve a dedicated path for that transfer which is known as path setup. During data transmission, no one else can use the reserved ports, which means that the path established is used exclusively by the sender. After transmission, the control router sends a release flit to indicate that all the ports are no longer reserved and can be accessed by others which is known as path release [13]. This is arguably not the most efficient use of resources; but it offers a fixed data transfer rate. Also, it allows total independence between the control sub-router and the data transfer.

### B. CS Router Architecture

The router in CS NoC is mainly divided into two main parts; the data sub-router and the control sub-router as shown in Fig. 1. The two are separate and each part has its own clock signal. The data sub-router consists of a clocked cross bar to connect each output port to the corresponding input port according to the control signals. The control sub-router is responsible for reserving a path, generating the control signals to the data sub-router and releasing the path again after the end of transmission. The control also includes the routing logic and an arbiter.

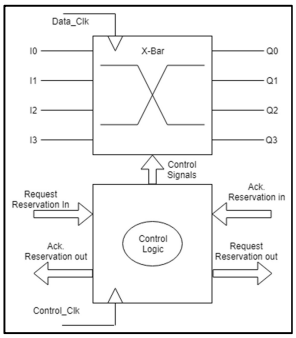


Fig. 1 Circuit-Switched Router Architecture

## III. PROPOSED NOC ARCHITECTURE

In this section the proposed CS NoC router architecture with synchronous control sub-router and asynchronous data transfer sub-router is presented in details.

### A. Overview of the proposed router architecture

The router is divided into two main blocks, the control sub-router which will remain synchronous as presented in the basic architecture and the date transfer sub-router which will be converted into asynchronous. This was achievable as CS routers offer independence between the control and data sections. The only connections between the two sections are the control signals from the control to the data. The idea to have asynchronous transfer is to maintain data transfer between the source and the destination with the source rate. The independence factor means that this transfer does not need any addition of synchronization blocks. The suggested design aims at having lower power consumption due to the reduction in dynamic power. The dynamic power equation is shown in (1).

$$P = ACV^2F \quad (1)$$

Where $A$ is the activity factor, $C$ is the switched capacitance, $V$ is the supply voltage and $F$ is the clock frequency. By converting to asynchronous data transfer, the high frequency clock is removed and its contribution to the dynamic power is eliminated. The other contributor to the dynamic power is the control sub-router; however, its clock has lower frequency which means much less dynamic power consumption.

### B. Proposed Router Architecture

The conversion from synchronous to asynchronous design was only implemented in the data sub-router, as it deals with much higher traffic and transitions than the control sub-router. The original cross-bar design simply consisted of five multiplexers (one multiplexer for each output port) and each multiplexer has five possible inputs and three select lines. The multiplexer design was combinational and there was a register at each multiplexer input to synchronize all the signals with the clock edge. To convert this synchronous to asynchronous design, the register is replaced by an asynchronous pipeline stage. The pipeline stage is designed using two different techniques: 2-phase and 4-phase single rail as shown in Fig. 2 respectively. The pipeline stages are based on the basic designs mentioned in [10]. The pipeline stages employ the handshaking protocol which is based on requests and acknowledgments. Another possible approach is to design a dual rail 4 phase pipeline stage; it merges the data signal and the request signal into one by creating two wires per bit. One wire will hold logic "1" and the other is logic "0". If both wires are holding logic "0", then this means that there is no data and it holds an empty state. The two wires cannot hold logic "1" at the same time. The dual rail pipeline stage is shown in Fig. 3 below. For asynchronous designs, the combinational logic must remain transparent to the handshaking protocol, for this reason the multiplexer was redesigned as shown in Fig. 4. The addition of the C-elements at each input ensures that the multiplexer will not perform its functionality and pass the request to the next stage unless the input request has arrived at its input. This restriction assures synchronization between all routers. The same goes for the acknowledgment, they will only be sent once the acknowledgment is received from the next stage. This means that each pipeline stage will not receive any new inputs or requests until knowing that the next pipeline stage has already received the data. These blocks were added to each router and the requests and acknowledgments were mapped between all the routers accordingly.

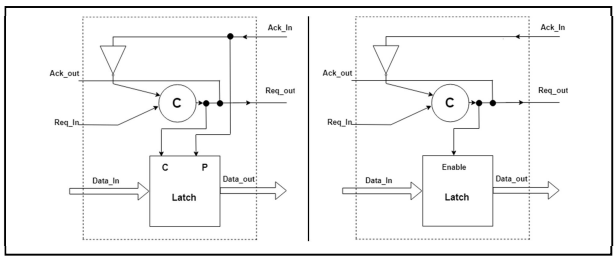


Fig. 2 Asynchronous Pipeline Stage as 2-phase (a) & 4-phase (b)

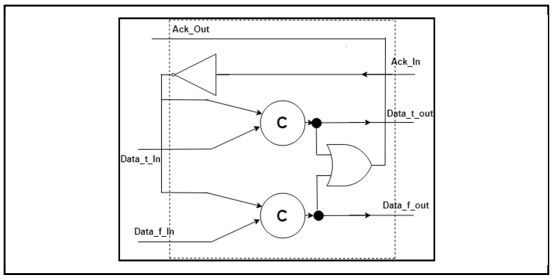


Fig. 3 Asynchronous Pipeline Stage for dual rail 4-phase

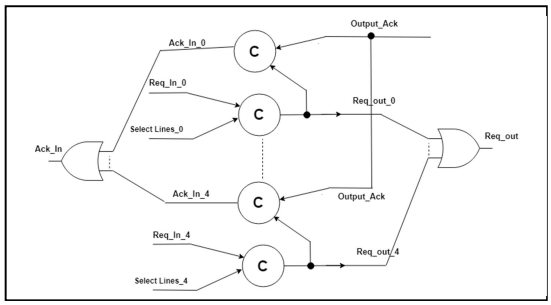


Fig. 4 Multiplexer implementation

### C. *Synchronous to Asynchronous interface*

The control signals in the multiplexer implementation (control signals) come from a synchronous design. Typically, there would be an interface between the two sections to avoid any glitches or failure in communication. However, CS means that the path is selected before the beginning of the transfer and dedicated to the data during transmission. It can only be released after the end of transmission. This means that the control signals are constant during the data transfer and no changes are expected in these lines. This leads to the conclusion that there is no need to add any more hardware in implementing an interface to synchronize between them. This is valid only for CS NoCs; however, for any other type such as packet switching where the data and the control are not independent, an interface must be implemented between them.

## IV. RESULTS

The suggested router design with asynchronous data sub-router was compared to the router with synchronous data sub-router in terms of Latency, power and area. The synchronous router used in comparison is based on the one presented in [7], but modified to include only a single synchronous data layer. Both routers were implemented using VHDL and synthesized using Synopsys Design Compiler for 65nm technology library. Matlab was used to generate random traffic for all NoCs to verify the dynamic power measurement under same traffic. The traffic produced was used to create a test-bench and a SAIF file was generated to capture the switching activity. The SAIF file along with the source files and constraints were injected to Synopsys Design Compiler. A timing analysis was conducted for the design and it showed a data arrival time of 0.44ns for the 2-phase asynchronous data sub-router. The clock period for the data in the synchronous design was set to a similar value to the asynchronous data arrival time for fair comparisons regarding dynamic power. The frequency for the control router clock was set to be 200MHz and for the synchronous router the frequency of the data clock was chosen to be 500MHz. The results are shown in the below figures, indicating much lower dynamic power consumption in both asynchronous designs as compared to the synchronous design. The leakage power and the area were both slightly reduced as well for the 4-phase design.

Fig. 5 presents 4x4 NoC area comparisons between the three techniques. The results show that the 2-phase design has the largest area among the three architectures due to the added complexity of the design. However this is a slight increase of 1.12% as compared to the synchronous. The 4-phase has the lowest area which is 3% lower than the synchronous design

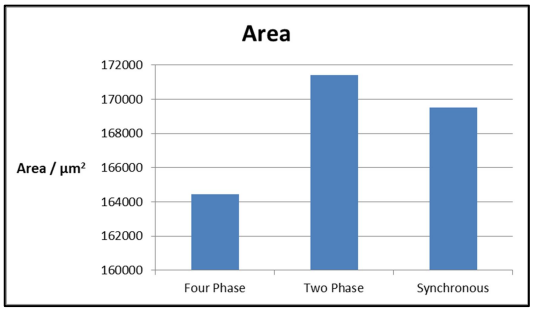


Fig. 5 4x4 NoC Area Comparisons under 65nm

Fig. 6 presents the dissipated leakage power comparisons between the three techniques for a 4x4 NoC. Both asynchronous techniques showed slightly lower power dissipation than the synchronous router. The proposed 2-phase router shows a reduction in leakage power by 7% while the 4-phase router shows a reduction in leakage power by 13%. The 4-phase has lower leakage power as it has a smaller area. The results also shows that the leakage power is negligible compared to the dynamic power (all the leakage power values presented are in the Nano Watt range).

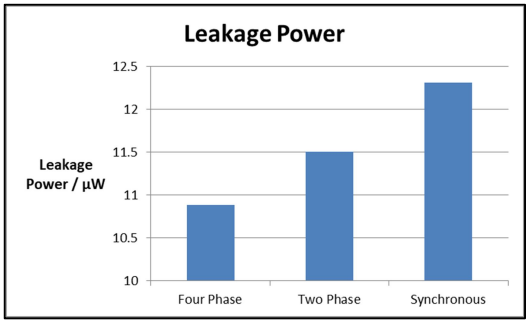


Fig. 6 4x4 NoC Leakage Power Comparisons under 65nm

Fig. 7 presents the dynamic power consumption comparisons between the three techniques for a 4x4 NoC. Both asynchronous techniques showed significantly lower power dissipation than the synchronous router. The proposed 2-phase and 4-phase NoC architecture shows a reduction in dynamic power by 80%. This large difference is obtained under fair conditions as all architecture have the same applied random traffic and the frequency for the synchronous router is chosen

to be matching the frequency of the applied traffic. The 4-phase has a slightly larger dynamic power as it has more transitions to achieve a single cycle when compared to the 2-phase.

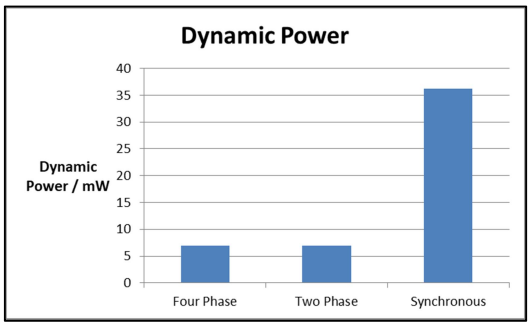


Fig. 7 4x4 NoC Dynamic Power Comparisons under 65nm

Fig. 8 presents the latency per cycle comparisons between the three techniques. It is a measure of the time taken to perform one complete data transfer through a single router. Both asynchronous techniques consume more time than the synchronous one due to the added hardware. The proposed 2-phase router has latency 70% larger than the synchronous one. The 4-phase router has latency 80% larger than the synchronous one. This shows that the improvement in power consumption leads to significant reduction in performance.

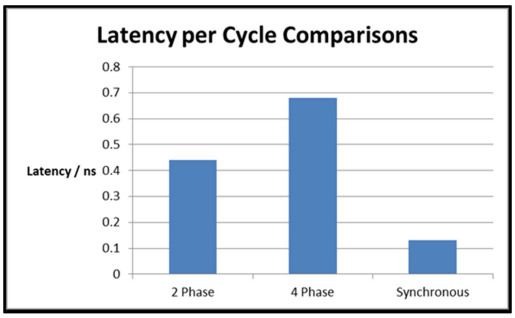


Fig. 8 Latency comparison under 65nm

The final comparison presented in Fig. 9 shows the comparison between the presented single rail schemes and the dual rail schemes. The dual rail scheme presents complete robustness against process variations which is very important with scaling down transistor sizing. However, it shows much larger area and dynamic power which makes it unrealistic and inefficient to use pure dual rail.

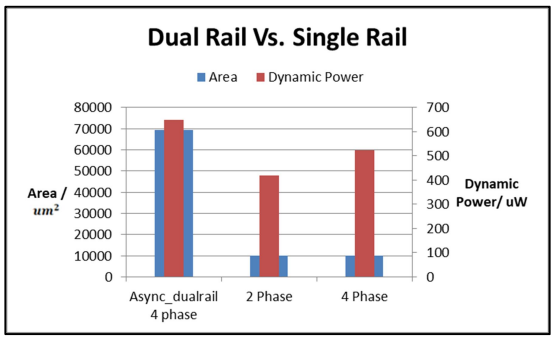


Fig. 9 Results comparison for Dual and Single under 65nm

## V. CONCLUSION

This paper presented an energy efficient sync-async CS NoC router with synchronous control sub-router and asynchronous data transfer sub-router. The aim was to introduce an energy efficient design to reduce power consumption. The proposed design was compared to the normal synchronous design in terms of latency, power and area. It showed more than 3% and 13% reduction in both area and leakage power respectively for 4-phase under 65nm. It also showed reduction in dynamic power by almost 80% under 65nm. The drawback is that it showed longer data arrival time for the asynchronous design by a factor of 70%. A final comparison was conducted to compare single rail and dual rail; it illustrated that single rail consumes much lower area and power. This proves that pure dual rail in not realistic to implement and approximations must be considered. In future work, the idea of a synchronous layer and an asynchronous layer alternating could be investigated. According to application, one layer is activated and the other is kept dark to exploit the concept of dark silicon for reduction in energy consumption. PG can also be introduced to shut down the inactive layer to further minimize the leakage power.

## VI. REFERENCES


[1] S. Borkar, "Thousand Core Chips: A Technology Perspective," In IEEE Design Automation Conference, pp. 746-749, June, 2007.

[2] L. Carloni, P. Pande and Y. Xie, "Networks-on-chip in emerging interconnect paradigms: Advantages and challenges," In IEEE International Symposium on Networks-on-Chip, pp. 93-102, May, 2009.

[3] J. Wu, D. Dong, X. Liao and L. Wang, "Energy-efficient NoC with multi-granularity power optimization," Journal of Supercomputing, vol. 73, no. 4, pp. 1654-1671, 2017

[4] Y. He, "Opportunistic circuit-switching for energy efficient on-chip networks," In IEEE International Conference on Very Large Scale Integration (VLSI-SoC), pp. 106, September, 2016.

[5] T. Bjerregaard, S. Mahadevan, "A survey of research and practices of network-on-chip," In: ACM Computing Surveys, June 2006.

[6] L. Cremona, W. Fornaciari, A. Marchese, M. Zanella and D. Zoni, "DENA: A DVFScapable hEterogeneous NoC Architecture," in ISVLSI, pp. 489-494, July, 2017.

[7] S. Hesham, D. Goehringer and M. El-Ghany, "Call-up for Circuit-Switched NoCs in the Dark-Silicon Era," In IEEE Nordic Circuits and Systems Conference (NORCAS), pp. 1-6, October, 2017.

[8] Rajmanikkam, J. Rajesh, k. Chakraborty and S. Roy, "Energy Efficient Network-on-Chip Architectures for Many-Core Near-Threshold Computing System," In *Journal of Low Power Electronics, pp. 115-128, 2019*

[9] J. Sparsø and S. Furber, "Asynchronous circuit design—A tutorial," in Principles of Asynchronous Circuit Design—A Systems Perspective, chs. 1–8, 2001.

[10] L. Weber, F. Moraes, L. Oliveira and E. carara, "Exploring Asynchronous End-to-End Communication Through a Synchronous NoC", In Symposium on Integrated Circuits and Systems Design, pp. 1-6, August, 2018.

[11] W. Jiang, D. Bertozzi, G. Mironadi, S. Nowick, W. Burleson and G. Sadowski, "An asynchronous NoC router in a 14nm FinFET library: comparison to an industrial synchronous counterpart," In Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017.

[12] M. Imai, T. Chu, K. Kise and T. Yoneda, "The synchronous vs. asynchronous NoC routers: an apple-to-apple comparison between synchronous and transition signaling asynchronous designs" In IEEE/ACM International Symposium on Networks-on-Chip, 2016.

[13] S. Liu, "New circuit switching techniques in on-chip networks," Doctoral dissertation, KTH Royal Institute of Technology,Sweden,2016