

An Efficient Clocking Scheme for On-Chip Communications

Mahdi Nazm Bojnordi, Nariman Moezzi Madani, Mehdi Semsarzade, and Ali Afzali-Kusha
Nanoelectronics Center of Excellence
School of Electrical and Computer Engineering
University of Tehran, Tehran, Iran
{m.bojnordi, n.moezi, m.semsar}@ece.ut.ac.ir, afzali@ut.ac.ir

Abstract— In Globally Asynchronous Locally Synchronous (GALS) architectures, System on Chip (SoC) deals with the risk of synchronization failure for the locally generated clocks. Also, this is an issue for Networks on Chips (NoCs). This paper proposes an efficient synchronizer for on-chip communication. The proposed module can be used in NoC switches satisfying the requirements for designing GALS SoCs. Proposed architecture is based on mesochronous clocking scheme and can endure frequency drift between two clock domains. Efficiency of the proposed architecture is evaluated by implementing in 0.18 μm CMOS technology. In comparison with a similar architecture, based on Delay Locked Loop (DLL), the new architecture saves around 74% of power consumption. Such power reduction is achieved while the number of transistors is decreased about 29% rather than the DLL-based architecture.

Keywords—mesochronous, NoC, clocking scheme.

I. INTRODUCTION

NoC is a new paradigm in designing efficient SoCs which exploits multi-billion transistor capacity composed of hundreds of interconnected IP blocks, such as CPU and DSP cores, application specific hardware and large memories[1]. It also has a profound impact on simplifying on-chip clock distribution networks. In contrast to SoC buses, NoCs are more scalable and power efficient leading to NoC-based multi-processor systems which are characterized by high structural complexity and functional diversity [2].

Concept of NoC is borrowed from wide area domain off-chip networks. NoCs are being used in current SoCs to manage all on-chip communications. This is a new solution to tackle IP cores inter-communication bottleneck in SoCs. Basic blocks in each NoC are switches and IP core interfaces. NoC's infrastructure is defined based on its basic block clocking scheme. Traditionally, two general clocking schemes have been widely considered in SoC designs: *Synchronous* and *Asynchronous*. Globally Synchronization (GS) -all SoC's blocks are synchronized using a single clock- is commonly used due to its simple design process and many available CAD tools. This scheme, however, has some drawbacks as it is mentioned here. The most important one is that the maximum clock frequency is inversely proportional to the chip physical size [3]. GS is not easily scaleable and efficient for the large SoCs. Furthermore, using GS method in synchronization of

large SoCs needs, H-trees for clock distribution on the chip consuming a large amount of energy [3]. The second types of clocking scheme are based on fully asynchronous approaches which are potentially power- optimized and efficient. The design of fully asynchronous circuits, however, is more complex especially when the size of the system increases. In addition, there are not many CAD tools available for asynchronous designs.

As a new alternative, the possibility of using a mixed system based on the GALS scheme has been considered [4]. In fact for the reasons mentioned above, GS should be avoided in SoC designs where different clock domains may exist. Therefore, there is a special need for synchronizing circuits which avoids reading the data during the meta-stability. The goals may be achieved through using a GALS scheme. In this approach, the inter-block communications are performed asynchronously while a synchronous clocking scheme is used for intra-block communications/ computations. Current researches focus on this methodology because the GALS systems offer a solution for SoC/NoC implementers seeking good performance and low power consumption [5]. Note that as an advantage for the GALS architecture, different blocks can have their own clock-frequencies allowing some parts to operate at as low clock frequency as possible. In addition, when some blocks of the system are idle, their clocks may be removed to minimize the power consumption.

To design a GALS architecture, the internal communications within the on-chip network may be implemented in two different ways which are partly/fully asynchronous and mesochronous solutions [6]. A mesochronous system can be viewed as a fully synchronous system where each block may be considered as a pipeline stage whose clocks may have some delay with respect to each other or are not the same [7]. In the mesochronous scheme, two different clock signals with the same frequency but with different phases may be used in the sender and the receiver modules. Another possibility is that the clock frequencies are not the same for the two blocks due to the fact that the clock signals may be generated by different clock sources [7] and, hence, the designer should consider the possible drift between the clock signals of different cores.

In the literature, some methods for asynchronous on-chip communication have been proposed (see, e.g. [8][9]). An approach to implement the mesochronous clocking scheme