Evaluation of GALS Methods in Scaled CMOS Technology: Moonrake Chip Experience

Miloš Krstić, IHP, Frankfurt (Oder), Germany
Xin Fan, IHP, Frankfurt (Oder), Germany
Eckhard Grass, IHP, Frankfurt (Oder), Germany
Luca Benini, University of Bologna, Bologna, Italy
M. R. Kakooee, University of Bologna, Bologna, Italy
Christoph Heer, Intel Mobile Communications, Neubiberg, Germany
Birgit Sanders, Intel Mobile Communications, Neubiberg, Germany
Alessandro Strano, University of Ferrara, Ferrara, Italy
Davide Bertozzi, University of Ferrara, Ferrara, Italy

ABSTRACT

In this paper the authors present the concept and evaluation results of a complex GALS ASIC demonstrator in 40 nm CMOS process. This chip, named Moonrake, compares synchronous and GALS synchronization technology in a homogeneous experimental setting: same baseline designs, same manufacturing process, same die. The chip validates GALS technology for both point-to-point and network-centric on-chip communications, demonstrating its potentials for different applications. The design analysis, measurement and test results confirm the potential of GALS approach for the scaled technologies, showing the significant benefits in respect to area, power, and EMI when it comes to the complex system implementation. Furthermore, 91% of the tests performed on the GALS network-on-chip test structures completed successfully, validating the timing robustness of new area and latency-efficient synchronization schemes and proving that the design flow for GALS synchronization technology can be implemented by means of mainstream industrial tools.

Keywords: Asynchronous, Globally-Asynchronous Locally-Synchronous (GALS), Network-on-Chip (NoC), Pausable Clocking, Synchronization

DOI: 10.4018/jertcs.2012100101
INTRODUCTION

Globally Asynchronous Locally Synchronous (GALS) technology has been proposed many years ago as an alternative to the traditional synchronous paradigm for chip synchronization (Krstic, 2006). Although significant potential was reported by the academia, the GALS methodology has never taken off in the industry. However, the growing challenges, imposed by the unrelenting pace of technology scaling to the nanoscale regime, urge for an efficient and safe system-level integration methodology. Consequently, we have targeted the implementation of a chip, named Moonrake, in the advanced 40 nm CMOS process, aiming at the assessment of GALS technology for nanoscale designs.

Our intention was to evaluate GALS vs. standard synchronous technology on the same die, by implementing synchronous and GALS counterparts of the same baseline designs, both in the point-to-point as well as in the network-on-chip (NoC) scenarios for on-chip communication. The two scenarios are very different, hence motivating the different choice of baseline designs for their analysis. In point-to-point communication, once an optimized GALS interface is selected, the focus is on the implications of redesigning an entire system around these links. In this direction, we took a state-of-the-art multi-million gate synchronous system, an OFDM baseband transmitter developed for a 60 GHz transceiver with a gigabit throughput as presented by Krstic in 2008, and re-implemented it with GALS methodology, using the optimized interfaces for pausible (stoppable) clocking as defined by Fan in 2009. One major goal was to explore Electromagnetic Interference (EMI) and switching noise properties of GALS designs and special algorithms and circuits for noise reduction based on the GALS methodology, initially analyzed by Fan in 2010. Within the chip, the switching noise (and correspondingly EMI) is caused by simultaneous switching activity of the digital circuits and it can lead to various problems including ground bounce, power integrity, IR drop, substrate noise etc.

For on-chip networking applications, the communication landscape is more heterogeneous since it results from the interconnection of domains with different synchronization assumptions. Therefore, our focus was on the provision of flexible and cost-effective interfaces for arbitrary composability. In this direction, the novel synchronization interfaces presented by Strano (2010) and Ludovici (2010), aiming at low-area/power/latency overhead while preserving timing robustness, were integrated into NoC test structures exposing (and comparing) a range of flexible GALS solutions.

The contributions of this paper are as follows:

• The GALS partitioning criteria for a state-of-the-art OFDM transmitter is presented, highlighting the optimized asynchronous link crossing scheme and the partitioning granularity and strategy at the system level.
• The design flow followed for different GALS systems is illustrated: from pausible clocking to mesochronous synchronization to mixed-timing systems. Compatibility with mainstream standard cell libraries and design toolflows is discussed.
• The feasibility of GALS NoCs linking sub-systems with heterogeneous timing assumptions by means of area/power/latency optimized interfaces while preserving timing margins has been demonstrated.
• Synchronous and GALS counterparts of the same baseline designs (the OFDM transmitter and a NoC sub-set), implemented in the same demonstrator chip, have been compared in terms of area, pointing out counterintuitive benefits of the GALS design style.
• Finally, the test and measurement results of Moonrake chip are presented and analyzed, with the focus on EMI and power measurements showing the benefits of GALS for complex system integration. Additionally, NoC test structures getting the clock from the external world provided an excellent result: frequencies from 25 to
Performance-Enhanced Caching Scheme for Web Clusters for Dynamic Content
www.igi-global.com/article/performance-enhanced-caching-scheme-web/58502?camid=4v1a