MICRO-MODEM – RELIABILITY SOLUTION FOR NOC COMMUNICATIONS

Arkadiy Morgenshtein, Evgeny Bolotin, Israel Cidon, Avinoam Kolodny, Ran Ginosar

Electrical Engineering Department, Technion, Haifa, Israel, e-mail: arkadiy@tx.technion.ac.il

ABSTRACT

A new concept of Micro-Modem interface for reliable communications in Networks on Chip (NoC) is presented. The Micro-Modem addresses the major problems of sub-micron interconnect and contains techniques for reliability improvement. The architecture, data flow and components of the Micro-Modem are presented. Various techniques and processes are analyzed for compact on-chip implementation. Design and application considerations are discussed.

I. INTRODUCTION

The development of deep sub-micron technologies in recent years allowed the integration of multiple systems on a single chip, defining the System-on-Chip (SoC) concept. On-chip packet-switched networks were proposed as interconnect solution for SoC. The typical structure of NoC is presented in Figure 1 showing the concept of modules connection via a mesh-type network of routers. NoC allows design modularity and high level of abstraction in architectural modeling of the systems and should satisfy certain Quality of Service requirements. This is currently performed by defining various QoS levels and applying communication protocols for management and scheduling of data flow.

However, the physical layer of the NoC hierarchy is based on sub-micron technologies which, together with the advantages of compact implementation, manifest a set of challenges related to reliability of on-chip communications [1]. In this paper the Micro-Modem, a solution for reliability in NoC communications, is proposed and described. The presented interface addresses the major issues of reliability, such as immunity for noise, interferences and parameter uncertainty.

In Section II the technological challenges of sub-micron communications are listed and analyzed. Section III presents the proposed architecture and the components of the Micro-Modem. Discussion and future research directions are given in Section IV.

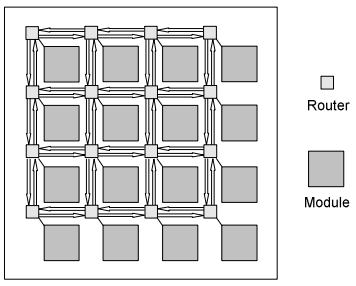

Figure 1. *Typical Network on Chip structure.*

II. CHALLENGES OF SUBMICRON ON-CHIP COMMUNICATIONS

Signal Integrity and Parameter Uncertainty

The noise and delay uncertainty in modern SoC designs are increasing significantly due to the increased coupling capacitances in deep submicron dimensions [2]. High-performance Network on Chip architectures are considered with multiple QoS levels and dedicated signal paths [3]. While the interconnect resources for packet transportation in various QoS levels are physically adjacent [3], the variation in activity factors, frequency and size of the transported data packets are contributing to the delay uncertainty and enhanced possible effect of the interferences due to the coupling capacitances.

The signal integrity is also affected by the random fluctuations of device properties [4] resulting in variations in transistor current drive capabilities and propagation delays. The fluctuations in threshold voltage and supply voltage in different locations on the chip can cause malfunction of circuits that depend on matched values of device properties and process parameters. It becomes even more important in high-end custom designs, utilizing “aggressive” techniques [5] as Dynamic Logic (sensitive to delay uncertainty) and Pass-Transistor Logic (sensitive to voltage fluctuations).

On-die temperature fluctuations are another important factor causing VLSI parameters uncertainty. Enhanced variations in operational activity and power consumption in different chip areas are very common in large, heterogeneous SoC, especially when it is running a variety of different applications. Thermal variations are also influencing the device properties as well as wire delays, noise and IR losses. In addition, the contribution of leakage to power consumption is increased in deep submicron technologies and is almost equal to the dynamic power [6]. This will cause an additional uncertainty in energy ranges – less active systems with multiple transistors might consume more static power than highly active systems consuming dynamic power.

As can be seen, the modern chip design can no more assume a micro-space with well pre-defined demands and dependencies between the logic blocks and interconnects. The SoC with variety of system types, the NoC with variety of QoS levels and the specifics of deep submicron fabrication technologies are introducing data traffic to a much more complex and aggressive environment.

Interfacing Communications in Network on Chip

The structure of Network on Chip (Figure 1) can be seen as assembly of routers and resources (IP cores and arbitrary logic

macro-cells) [7]. The communication concept of NoC is based on packet-type data transport between different source-destination pairs. The communication can be through the routers, or directly between the modules in case of high-bandwidth regular data transport. Thus, all the blocks involved in the network require a simple interface to the interconnect for uniform communication.

The complex implementations that are proposed for NoC are focusing on IP cores [9], but because of the implementation and operation complexity, it is not suitable for arbitrary logic. Other network channels propose simpler design which is, however, limited to 2-slot TDM schemes [7]. The variety of the systems, data formats and performance requirements dictates the need for simple but scalable circuit, which could be easily adapted for operation in different frequencies, parallelism rates and communication protocols.

Motivation

The assembly of the technological tendencies, requirements and limitations in the sub-micron System-on-Chip communications using Networks-on-Chip triggers a need for a new concept of reliable data transportation. The aggressive and complex environment influencing the signal integrity requires application of techniques for noise and interference immunity as well as error identification and correction. The controlled parallelism of the data link allows area efficiency and interconnection uniformity between the systems, but requires an interface that would be capable of translation of parallel data in different frequencies and formats into a different parallelism rate or into a sequence of serial signals. The interface should apply modulation techniques for successful transmission and receiving of the data via a limited communication medium. All these features have to be combined in a single robust, but simple circuit with enhanced scalability and reuse capability.

Basing on analysis of the list of demands and weighting the possible solutions we propose a novel concept of interfacing in on-chip communications – the Micro-Modem.

III. MICRO-MODEM CONCEPT

The proposed concept is based on implementation of a simple circuit which will be applied as a communication interface between the corresponding modules and routers in SoC/NoC (see example in Figure 2). This unit, is termed Micro-Modem and contains various blocks responsible for transmission and receiving of digital data via the interconnect.

As can be seen from the previous Section, the conditions and reliability demands of sub-micron SoC/NoC design closely resemble the conditions and requirements of RF digital communications [10]. The noisy environment with enhanced rate of parameter uncertainty, external and inter-symbol interferences, common communication medium, variety of systems and protocols – all these are common properties of Network-on-Chip and of wireless RF communications.

Thus, some of the techniques of RF communications can be adopted and adapted in the Micro-Modem. Due to the specifics of on-chip communications, some of the RF methods (like the variety of modulation techniques) could not be applied, or will get a new interpretation, while other techniques will be added to suit the on-chip world.

Far relatives of the Micro-Modem in on-chip design are the implementations of Transmitter (TX) and Receiver (RX) on-chip units which are currently researched and developed for high-speed on-board communications between chips [11][12]. These implementations are focusing on serial data transmission due to the limited number of available connectors and links, but are highly area consuming. Normally, a single TX/RX unit is placed on a chip for on-board inter-chip communication.

In case of the Micro-Modem the demand for compact implementation dictates a need for specific design concept. This will allow combining multiple Micro-Modem units on chip while minimizing the area and power overhead. Thus, the existing techniques-candidates for adoption in the new concept should be carefully examined through the prism of low-area and low-power demands.

Micro-Modem architecture

The architecture of the Micro-Modem can be observed from the example of the transmitter in Figure 3. Each Modem can perform as RX or TX (or contain both opposite paths). The data accepted from the router or the module has to be transferred to the interconnect for high-speed transportation. The Modem implements the Physical Layer and Data Link Layer of the OSI model, while the higher layers are performed in the Router and end nodes.

The parallel data from the source is stored in the input buffer which can be shared by the Modem and the source node. Afterwards, the packet is transformed into frame by addition of header and tail, as shown in Figure 4. Additional bit series is added to it to obtain error detection/repairing abilities; synchronization series and start/stop signals are attached for solving the delay uncertainty. The data undergoes additional processing to provide the Inter Signal Interference (ISI) and noise immunity. Then the frame is given the desired parallelism rate, or serialized and modulation is performed to allow data transport from multiple bit sources through a limited number of wires or a single wire. Link interface supplies the conditioning and buffering of the signal as it enters the wire.

At the wire level various optimizations are performed to assure the minimal delay, full swing and low ISI of the signal (e.g. repeater insertion). The combination of high-performance low-power circuitry with low area consumption, digital data modulation techniques, and wire optimizations should ensure the high-speed data transport based on a simple and compact interface.

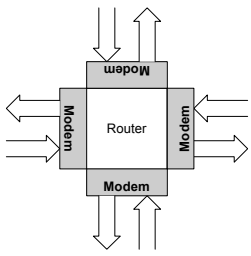

Figure 2. *Router architecture with modems.*

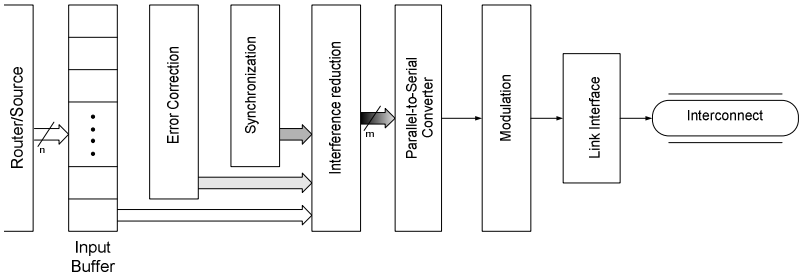

Figure 3. *Transmitter architecture in the micro-modem concept*

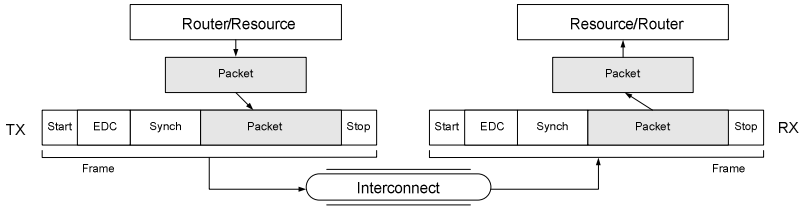

Figure 4. *Data transportation scheme*

Components of the Micro-Modem

Most of the components of the Micro-Modem are digital, due to the low area consumption, noise sensitivity and implementation complexity, as compared to analog circuits. In cases where the concern of area consumption is secondary to performance specifics, analog techniques can be also applied and will be mentioned here.

Input buffer – stores the packet for further parallel processing. As a basic unit on which most of the Modem processes are applied, it requires high-performance compact memory cells with high driving ability.

Error correction – computes the error detection code (EDC) that depends on the contents of the packet and is transmitted as additional bit series in order to detect corruption of the data in the Receiver. A variety of codes can be implemented in this stage – parity checks, cyclic redundancy codes (CRC), Humming code for error correction, etc. In general, each code can be described by a code rate Rc according to:

$$R_c = N_{in} / N_{out} \tag{1}$$

where N_{in} is the number of input bits and N_{out} is the number of resulting output bits (packet and EDC). The number of errors K that can be detected by (N_{in}, N_{out}) code is:

$$K = \left(N_{out} - N_{in} \right) / 2 \tag{2}$$

The EDC can be used both for enhancement of noise and interference immunity of the data, and for increased restoration ability of data losses due to voltage drops. Thus, error correction in the Micro-Modem is effective for overcoming low signal integrity due to ISI, cross-talk and noise; voltage drop in resistive wires; low-swing effects in low-area circuit designs like PTL. The EDC block has to be implemented using robust full-swing techniques as CMOS.

Synchronization – is responsible for delay uncertainty elimination. Several techniques can be applied in this stage: skew reduction between clock and data can be performed by generating *start/stop* signals; for asynchronous and GALS structures *ack/req* signals can be formed; clock recovery can be performed using 8x10 encoding. Sequential numbers can be generated for ordering and detection of lost packets. These signals together with the EDC create the header and the tail which are added to the original *n*-bit packet and assemble the *m*-bit frame.

Interference reduction – minimizes the effects of external and inter-signal interferences. Signal shaping can be performed for minimization of sequent serial signals interference by using Raised Cosine Roll-off Filter [10]. Interleaving of the data can be performed to enhance the immunity to burst noise and to improve the error correction ability of the EDC. The interleaver can be ether fixed for constant pattern, or programmable for the case of external control (according to noise characteristics). For enhanced Bit-Error Ratio the interleaving can be applied on several sequential packets of data. This, however, is costly in terms of area and latency due to the need in additional buffers for data storage.

Serializer – performs the conversion of the *m*-bit data from parallel to serial form or to different rate of parallelism. For serial data transportation it has to be high-speed to compensate the loss of parallelism. This creates a challenging trade-off between scaling of the transistors and the demand for compact and low-power implementation. The serializer in Figure 5 is based on a switch array as part of *m*-bit Time Division Multiplexing (TDM). It can be controlled by pipeline [13] initiated by system clock pulse for asynchronous protocols, or by synchronous multiplexer controlled by a fast clock. The advantage of the asynchronous implementation is in high-speed operation without a need for ultra-fast (*m*-times) clock generation. The serializer can be designed for various lane widths scenarios (*k*-bit output), or as a generic unit with lane width controller applied to the multiplexer and switch array.

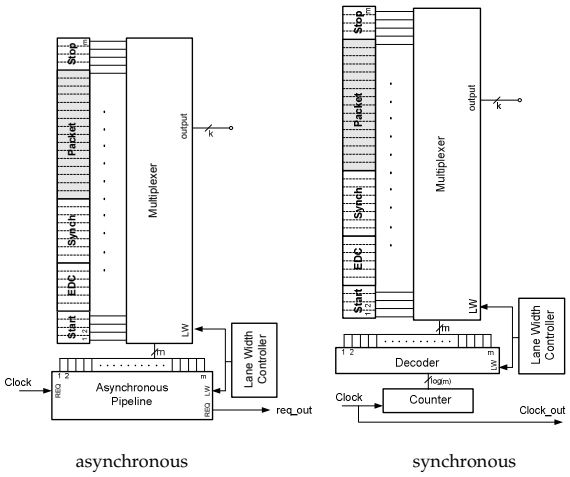

Figure 5. *Serializer structure.*

Modulation – adopts some of the standard techniques suitable for on-chip implementation. Simultaneous data transport in a single wire can be considered, using FDM. This raises two alternatives – current signaling, allowing unlimited number of channels, but requiring high-sensitive analog circuits; or voltage signaling using simpler multi-level digital circuits but resulting in limited number of channels. A widely used technique in on-board serial interfaces is the Low Voltage Differential Signaling (LVDS), which allows high noise and ISI immunity and high frequency, by using sensitive but complex and high-area analog and mixed-signal interfaces.

Link interface – performs conditioning and buffering of the signal as it enters the wire. The cascaded buffers and fan-out drivers in this stage are responsible for swing restoration and timing optimization of the signal.

IV. DISCUSSION AND FUTURE RESEARCH

The solution of reliability problems in SoC/NoC can be achieved by application of the Micro-Modem as link interface of the systems. While selecting the suitable techniques within the variety of existing wireless solutions, the designer should consider the specific area and power limitations of on-chip architecture. When multiple Micro-Modem interfaces have to be integrated in the network, the preference should be given to compact, low-power digital circuits and techniques. For single high-performance link interfaces, like in inter-chip links, more complex and big implementations can be adopted basing on some mixed-signal techniques.

It should be noted that the type of Modem and the way of its application are influencing the additional latency due to the signal processing in the TX/RX. Figure 6 shows how the Modem application influenced by noise characteristics and inter-system distance. For low noise activity and short data link, the Modems can be applied only at the end points of the communicating systems. In case of high noise activity along the link and long distance between the systems, Modems should be applied at each router point on the data path. This results in enhanced latency due to high number of interfaces and should be considered during the design.

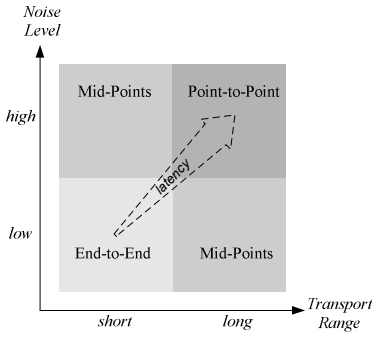

Figure 6. *Micro-Modem application types.*

The optimal application of Modems in NoC, as well as the selection of reliability techniques is related to the complete picture of on-chip noise sources, system activities, layout and routing details. Currently the combination of these parameters as part of reliability analysis is not performed during SoC design and should be a part of future research in the field. Design techniques for low-power compact Micro-Modem should be researched and developed to make it an integrated part of reliable SoC/NoC designs in sub-micron technologies.

REFERENCES

[1] ITRS, System-Level Design Challenges, pp. 18-19, 2003.

[2] J. Zhang and E.G. Friedman, "Crosstalk Noise Model for Shielded Interconnects in VLSI-based Circuits", *IEEE Int. SOC Conference*, pp.243-244, 2003.

[3] E. Bolotin, I. Cidon, R. Ginosar and A. Kolodny, "QNoC: QoS architecture and design process for Network on Chip", *Special issue on Networks on Chip, The J. of Systems Architecture*, December, 2003.

[4] H.P. Wong, D.J. Frank, P.M. Solomon, C.H.J. Wann, and J.J. Welser, "Nanoscale CMOS", Proc. of the IEEE, vol. 87, no. 4, April, 1999.

[5] R. Zimmermann and W. Fichtner, "Low-Power Logic Styles: CMOS Versus Pass-Transistor Logic", IEEE J. of Solid State Circuits, vol. 32, no. 7, p. 1079-90, 1997.

[6] G.E. Moore, "No Exponential Is Forever: But "Forever" Can Be Delayed!", Proc. of the ISSCC, pp. 20-23, vol.1, 2003.

[7] A. Joshi, J. Davis, "A 2-Slot Time-Division Multiplexing (TDM) Interconnect Network for Gigascale Integration (GSI)", *SLIP Conference*, pp. 64-68, 2004.

[8] P. Gupta, A.B. Kahng, Y. Kim, D. Sylvester, "Investigation of Performance Metrics for Interconnect Stack Architectures", *SLIP Conference*, pp. 23-29, 2004.

[9] P. Bhojwani and R. Mahapatra, "Interfacing Cores with On-chip Packet-Switched Networks", Proc. of the VLSI Design, pp. 382-387, January, 2003.

[10] A. Bruce Carlson, "Communication Systems – An Introduction to Signals and noise in Electrical Communication", McGraw-Hill, pp. 1-5, 1986.

[11] C.K. Yang and M.A. Horowitz, "A 0.8-µm CMOS 2.5 Gb/s Oversampling Receiver and Transmitter for Serial Links", *IEEE J. of Solid-State Circuits*, vol. 31, no.12, pp. 2015-2023, 1996.

[12] D. Friedman, m. Meghelli, B. Parker, J. Yang, H. Ainspan, and M. Soyuer, "A Single-Chip 12.5Gbaud Tranciever for Serial Data Communication", *IEEE Symposium on VLSI Circuits*, pp. 145-148, June, 2001.

[13] D.E. Muller and W.S. Bartky, "A Theory of Asynchronous Circuits," Proc. Int. Symp. on the Theory of Switching, Harvard University Press, pp. 204-243, April, 1959.