Efficient High-speed Current-mode Links for Network-on-Chip Performance Optimization

Hamed S. Kia and Cristinel Ababei
Department of Electrical and Computer Engineering
North Dakota State University, Fargo ND, USA
Email: \{hamed.sajadikia, cristinel.ababei\}@nds.edu

Abstract—We propose a novel cost-effective long-range NoC interconnect design based on current-mode signaling. The proposed CMOS based long-range link reduces the communication delay of long wires significantly without using traditional pipelined repeaters, making it a cost efficient alternative to wireless and optical interconnects. Spice simulations are performed to verify the performance of the proposed current-mode transceiver. The performance of the proposed interconnect is analyzed and compared with wireless, optical and existing CMOS based interconnects. Also a modified NoC simulator is utilized to demonstrate the benefit of current-mode long-range links.

I. INTRODUCTION

Continuous technology downscaling enabled increased device densities and improved device performance. However, downsampling also lead to an increase of the resistance and capacitance per unit length for interconnects, which in turn resulted into a shift of the circuit delay bottleneck from logic gates to wires. This shift is further accelerated by the trend of global interconnects to become longer due to the increase in maximum die size. Early solutions to this global delay problem included pipelining of the interconnects via repeater insertion. However, repeater insertion is not a scalable or sustainable solution anymore. One of the most promising recent solutions to address the global delay problem and the need for higher communication concurrency in today’s complex systems-on-chip (SoCs) including multiprocessor systems-on-chip (MPSoCs) is the network-on-chip (NoC) [1], [2]. NoCs replace the traditional global on-chip buses and provide a generic on-chip interconnection network realized by specialized routers or switches connected via short-range links. While the delay of these short-range links continue to represent a design challenge, it is the latency between remote network routers that is of even higher concern. This problem can be addressed via insertion of long-range links that can be implemented via traditional interconnects with repeaters. Since the idea of long-range link insertion has been introduced in the context of NoCs [3], several different types of implementations have been proposed. These implementations can be basically classified into three main classes: wireless, optical, and CMOS wires. Optical interconnects require converting signals between optical and electrical domains, an increased area, and have to deal with fabrication technology incompatibility. Wireless interconnects have issues related to extra area and power consumption of the transceivers. Yet another recent alternative solution is current-mode signaling. Current-mode links are fully CMOS and can achieve nearly speed-of-light speed. Their main drawback is the area penalty needed to implement transceivers that implement voltage-to-current and current-to-voltage converters.

As mentioned above, one of the alternatives to the traditional metal interconnects is wireless interconnects. The network architecture is augmented with wireless nodes spread across the chip. These nodes form a wireless network superimposed with the regular mesh network, which can be employed to transmit data via wireless paths with smaller number of hops compared to the regular mesh network. For example, RF interconnect shortcuts have been used to design novel NoC reconfigurable architectures in [4]. A simultaneous tri-band on chip RF interconnect for NoCs was reported in [5], where two RF bands in mm-wave frequencies were modulated using amplitude-shift keying.

Optical or photonic interconnects integrate one or several high bandwidth optical links that span the whole NoC area to provide high speed links between remote network routers. Because of the potentially significant performance benefits that optical communications can offer a lot of research concentrated on this area. For example, an optical NoC (ONoC) was designed and investigated in [9]. A hybrid micro-architecture for NoCs that combines a broadband photonic circuit-switched network with an electronic packet-switched control network is described in [10]. [11] combines photonic rings on multiple photonic layers within 3D mesh NoC architectures. Other recent studies on optical interconnects and their use within NoC based systems as well as network simulators include [12]–[21].

Most of the traditional CMOS interconnects work in voltage-mode. While they are cheap to fabricate, they suffer from the rapid increase of their delay with length. This problem is typically addressed via interconnect pipelining with repeaters or buffers, which translate into area penalty. An alternative solution is offered by several earlier proposals for interconnects whose operation is in current-mode. Current-mode links operate with currents as the main information carriers while voltage variation is limited to narrow ranges. As a result, current-mode links can achieve nearly speed-of-light latencies. In addition, current-mode transceivers are entirely CMOS thereby cost-effective and easy to fabricate. Their main disadvantage compared to voltage-mode links is that they require transceivers that may result into larger area penalty compared to that of repeaters of the voltage-mode.
links. One of the earlier current-mode link proposals for on-chip interconnects is [22]. A new transceiver design based on differential current-mode sensing to reduce both delay and energy dissipation compared to the standard inverter repeater approach is proposed in [23]. Authors in [24] use a sharp current-pulse data transmission to modulate transmitter energy to higher frequencies, where the effect of wire inductance is maximized, allowing the on-chip wires to function as transmission lines with considerably reduced dispersion. [25] reports an asynchronous serial link for on-chip communication as an alternative to standard bit-parallel links. The serial interconnect uses a differential level encoded dual-rail two-phase asynchronous protocol, and avoids per-bit handshake and eliminating per-bit synchronization. A high-performance long-range NoC link design based on multilevel current-mode signaling and delay insensitive two-phase 1-of-4 encoding is studied in [26]. This design achieves higher throughput and lower latency compared to voltage-mode interconnects.

In this paper, we focus on CMOS based current-mode links that represent a compromise between the high performance of the wireless and optical interconnects and the reduced area and cost of the voltage-mode interconnects. We propose a new current-mode link with the main objective of minimizing the area of the transceiver without sacrificing performance. The remainder of this paper is organized as follows. In the next section, we introduce the proposed current-mode link design. Section III reports simulation results. In Section IV, we study the use of the proposed current-mode link as long-range links in NoCs. We conclude in Section V by summarizing our main contribution.

II. PROPOSED CURRENT-MODE LINK

In this section, we introduce the current-mode link design that represents the main contribution of this paper. Because the main application of the proposed current-mode link is the implementation of long-range links in NoCs, our primary goal is to develop a cost effective solution without compromising the link performance. Cost effective current-mode links are desirable to minimize the area increase of network routers compared to the traditional router architectures. Earlier efforts on current-mode links used differential solutions [24] that were motivated by noise or architectural constraints. The overall current-mode link architecture is shown in Fig.1.

The transmitter is implemented using a current-mode differential driver [27] formed basically by the two PMOS transistors M1 and M2, which are controlled by the outputs Q and Q̅ of the flip-flop that in turn is driven by the pulses of the input voltage Vin. During normal operation, when Vin is logic 1, the transistor M1 turns off while transistor M2 turns on to direct the current I0 to resistor R2 and the actual link. On the other hand, when Vin is logic 0, M2 turns off while M1 turns on to drive current I0 to resistor R1 and away from the link.

The receiver is formed by four transistors M3, M4, M5 and M6, a Schmitt trigger, and two inverters as shown in Fig.1. The key point of the receiver architecture is that the voltage swing of Vr is limited to a small range through the action of the Schmitt trigger and the feedback created via transistors M5 and M6. The two inverters regenerate the output from the Schmitt trigger to provide the output voltage Vout and to drive the inverter formed by M5 and M6. The small voltage swing of Vr is essential to achieving a fast generation of the output voltage Vout. The range for the variation of Vr is centered at a reference voltage Vref determined by M3 and M4. Vref is selected such that it is in the middle of the voltage range given by the low and high threshold voltages of the Schmitt trigger, VTTL and VTH (see also Fig.2).

In the following subsections, we describe the operation of the proposed current-mode link when either a logic 1 or a logic 0 is transmitted.

A. Operation when logic 1 is transmitted

When the input voltage Vin is logic 1, the differential pair M1 and M2 directs all current I0 to resistor R2 and the actual link. R2 is selected such that most of the current flows into the link. In our analysis, we model the physical link as a distributed RC transmission line whose technology parameters can be estimated via technology prediction models.

Once the current I0 is directed to the link, the equivalent capacitance of the link and its load starts charging. In other words, the net equivalent capacitance of the node Vs is getting charged and Vr starts to increase until it reaches the upper threshold voltage VTH of the Schmitt trigger. At this moment, the output of the Schmitt trigger goes high, which in turn results into the output voltage Vout to become logic 1. Once Vout switches to logic 1, the feedback formed by the transistors M5 and M6 helps to stop the increase of Vs by pulling (via M5) some of the current that charges the node Vs. This operation is illustrated for example in the fifth clock period or cycle shown in Fig.2.

B. Operation when logic 0 is transmitted

On the other hand, when the input voltage Vin is logic 0, the differential pair M1 and M2 directs all current I0 to resistor R1. In this case, no current drives the physical link. As a result, the voltage Vs starts to decrease until it reaches the lower threshold voltage VTL of the Schmitt trigger, which triggers and changes its output to low. This in turn results into the output voltage Vout to become logic 0. Once this happens, the feedback formed by the transistors M5 and M6 helps to stop the decrease of Vs by pushing (via M6) current to sustain the node Vs. This operation is illustrated for example in the seventh clock period shown in Fig.2.

III. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

To test the performance of the proposed current-mode link we implemented the transceiver shown in Fig.1 using 0.18μm CMOS technology parameters. The 0.18μm technology was chosen intentionally because we wanted to compare our proposed interconnect with previous work. The actual physical link is modeled as a distributed RC transmission line.
whose technology parameters are estimated via the technology prediction model from [28]. We utilize Cadence Spectre [29] for our simulations. In what follows we report delay and power consumption simulation results.

The latency is defined as the difference between the time the input is applied to the transceiver and the time the output signal is received. In other words, link latency includes the delay of the transmitter, the link, and the receiver from Fig. 1. The variation of the overall transceiver latency or delay versus the link length is shown in Fig. 3 for a wire width of $0.36\mu m$ and the link length varied between $1mm$ and $6mm$. Simulations are performed under a clock period of $T_{clock} = 0.5ns$ with 1% edge rise and fall times.

The variation of the overall transceiver power consumption versus the link length is shown in Fig. 4. Simulations are performed for a $V_{DD} = 1.8V$.

**A. Performance sensitivity to power supply and variation with technology**

The impact of the power supply changes on the performance of the proposed link is illustrated in Fig.5 and Fig.6. As expected, increasing the power supply leads to better performance but increased power consumption, while any decrease in power supply reduces the power consumption but increases the link latency.

Fig.7 shows simulations results of the link latency for different CMOS technologies. In our simulations we use technology parameters reported by the predictive technology model [28]. We notice that link delay increases linearly as a result of that in more recent technologies wire resistances and capacitances increase.

**B. Qualitative comparison with related work**

In this section, we attempt to compare the proposed current-mode link to previous works. It is difficult to conduct a fair comparison because to quantify the advantages and disad-
Fig. 4. Power consumption versus link length.

Fig. 5. Illustration of sensitivity of the measured latency results to power supply changes for 6mm long link.

Fig. 6. Illustration of sensitivity of the measured power consumption results to power supply changes for 6mm long link.

Vantages of a given long-range link approach, one has to consider similar constraints. For example, if one wants to compare traditional voltage-mode links versus optical links in terms of delay and bandwidth, the comparison should be done assuming similar area and/or fabrication cost budgets. In addition, some interconnect solutions may offer features that cannot be achieved by others. For example, wireless links can transmit to multiple receivers or they can be easily reconfigured to form new routing paths. Therefore, here, we present a qualitative comparison of the proposed current-mode link with previous works.

Table I compares the proposed current-mode link with several other reported CMOS based links. The comparison is done for a link width of 0.36μm and a length of 6mm. We note that the proposed current-mode link outperforms or offers comparable performance to that of previously proposed links. The only previous work that reports a shorter delay is [24] however for a link length of 3mm and wire width of 4μm. If we use the same wire dimensions, then the proposed current-mode transceiver’s performance would be 30ps.

Table II compares the proposed current-mode link with some of the wireless and optical interconnects that were recently proposed also in the context of NoC design. We note that the proposed link is implemented completely in CMOS technology. Therefore, it requires less area and is cheaper to implement compared to the wireless and optical counterparts.

IV. APPLICATIONS

In this section we discuss how current-mode links are utilized to improve the performance of NoC based systems.

The concept of long-range link insertion within the context of NoCs has been proposed in [3] and it consists of the superposition of a few long-range links and a standard mesh network. Inserting application-specific long-range links enables a higher network throughput and a lower average
To investigate the above claim, we conduct simulation experiments on several 4 × 4 networks. In our simulations we use the smallnoc algorithm [3] to find the location of the long-range links to insert in each of the investigated networks. Information on the location of the long-range links is then utilized as part of the input given to the wormsim cycle-accurate NoC simulator, which is able to simulate regular networks as well as networks modified with long-range links. Both tools can be downloaded from [30]. Note that the smallnoc tool also calculates the content of all routing tables to guarantee deadlock free routing after the long-range link insertion. We have modified the wormsim simulator to be able to simulate also the proposed current-mode links.

In the first simulation, we study hotspot and transpose synthetic traffic workloads. For hotspot traffic, three nodes (5, 11, and 12) are selected randomly to act as hotspot nodes. Each node in the network sends packets to all other nodes. The hotspot nodes are setup such that they receive 5% more packets on average compared to the rest of the nodes. For transpose traffic, each node communicates only with the symmetric node with respect to the diagonal of the network. In the second simulation, we evaluate the performance of the telecom benchmark, which is provided as part of the wormsim tool and which was also studied in [3]. In all our simulations we use the following settings for the wormsim simulator: packet size of one flit, input and output buffer size of two packets, and one virtual channel per input port. The remaining settings are kept to their default values. Similar to [3], we run the smallnoc tool with a constraint of $S = 12$, where $S$ is the total length of all long-range links measured in terms of standard links or hops. The solutions achieved by the smallnoc tool are shown in Fig.9.

The variation of the average packet latency as a function of the packet injection rate is shown in Fig.10 for the hotspot, transpose, and telecom traffics. As expected, when long-range links are implemented as current-mode links the average packet latency improves across all values of the packet injection rate.

V. CONCLUSION

We have introduced a novel high-speed cost-effective current-mode interconnect design. Simulations demonstrate that the proposed current-mode transceiver can offer a latency of 126 ps at a power consumption of 3.35 mW for an actual wire length of 6 mm and wire width of 0.36 μm in 0.18 μm CMOS technology. These results improve on the state-of-the-art current-mode links and represents a better compromise between the benefits of high speed of wireless and optical interconnects and of reduced area and cost of traditional repeater-pipelined voltage-mode interconnects. Simulations with a modified NoC simulator demonstrate that current-mode long-range links can improve the average flit latency.

REFERENCES

Fig. 10. Average packet latency versus traffic injection rate

(a) Average packet latency versus traffic injection rate for hotspot traffic.
(b) Average packet latency versus traffic injection rate for transpose traffic.
(c) Average packet latency versus traffic injection rate for telecom benchmark.