Delay-Insensitive On-Chip Communication Link using Low-Swing Simultaneous Bidirectional Signaling

Ethiopia Nigussie, Juha Plosila, and Jouni Isoaho
Communication Systems Laboratory
Department of Information Technology, University of Turku
Turku, Finland
{ethnig, juplos, jisoaho}@utu.fi

Abstract

In this paper we present the circuit implementation of a new asynchronous delay-insensitive on-chip link structure, where two modules placed on the opposite sides of the link can exchange data simultaneously. Unlike the conventional delay-insensitive dual-rail link which requires $2N + 1$ interconnects to transfer N-data bit, $N +1$ interconnects are required in this design. As two transceivers can access simultaneously the same physical interconnect the number of required interconnects halves compared to bidirectional transfer based on two separate unidirectional dual-rail links.This makes the link cost effective for future SoC. The transceiver circuits are designed using multiple-valued current-mode logic, linear summation is implemented by wiring without active devices simplifying the resulting circuitry. By using 110mV voltage swing the power consumption of the link is 8.32mW for 689ps propagation delay and 5mm interconnect length. Some of the potential application areas of this link are between locally clocked modules in GALS system, between routers of NoC nodes, and in adaptive and reconfigurable system where feedback information is crucial. The circuit is designed and simulated using Cadence Analog Spectre with 0.13um CMOS technology.

1. Introduction

System-on-Chip (SoC) designers face a major challenge in achieving the required functionality, performance and testability whilst minimizing design cost and time to market. The key to achieving this goal is a design methodology that allows component reuse. These design methodologies rely upon the use of a standardised interconnection interface for connecting component blocks together to form an on-chip system. This system-level interconnect encounters delay, power consumption, and signal integrity problems as feature sizes advances towards nanometer regime and multi-gigahertz frequencies.

Due to the challenge to overcome clock skew and switching noise in synchronous CMOS design, asynchronous circuit design becomes an increasingly practical alternative. Some potential advantages of asynchronous circuits over synchronous circuits are higher speed, lower power dissipation, and higher modularity [1-2]. Asynchronous circuits are also free from the problem of clock skew and have relatively low electromagnetic interference because of their distributed switching activities in time. The advantage of Multiple-Valued Current-Mode (MVCM) logic over conventional CMOS logic style include circuit simplicity, higher speed due to much smaller signal swing and frequency independent power dissipation, which results in lower power dissipation in multi-GHz frequencies. In addition, the steady current source of MVCM reduces the power supply fluctuations and noise. This property and the small voltage swing on the interconnect help to minimize crosstalk.

Asynchronous simultaneously bidirectional delay-insensitive link is presented in this paper. It uses MVCM signaling which combines the benefits of multiple-valued logic and current-mode operation. These make the link a good candidate for a future high speed, energy efficient and noise tolerant on-chip link.

2. Signaling protocol

The signal transmission system used in CMOS circuits can be broadly classified into two categories:

voltage-mode and current-mode signaling. In voltage-mode the voltage has to swing rail-to-rail over the entire length of the interconnect. This leads to large transient currents consuming more power, larger delay and it also generates power-supply noise [7]. The optimal repeater insertion technique [4]–[6] used in voltage-mode signaling, was developed to reduce the wire delay and improve performance of lengthy global interconnections. However, with the increase in number and density of interconnects with technology, the number of repeaters necessary would increase manifold, presenting significant overhead in terms of power and area. Hence, there is a need for a new interconnection scheme that reduces wire delay, its variation and increases noise immunity. Unlike voltage-mode signaling, current-mode signaling allows voltage-swing to be reduced without separate voltage references and isolates the received signal from power supply noise [7]. It also translates into increased bandwidth performance [8], decreased delay and power dissipation and higher noise immunity. For these reasons, current-mode signaling technique is a better alternative to voltage-mode technique, for the future multi-GHz, noise-prone SoC chips.

In this link current-mode signaling is designed using multiple-valued logic. This enables to encode the data together with request using different current levels. In addition this logic allows linear current summation and subtraction just by wiring without active devices. Therefore the combination of multiple-valued logic with current-mode signaling makes the link circuit simple, propagates the signal faster, dissipates less power and isolates the signal from power-supply noise.

Communication on the link and between modules and transceivers follows 2-phase signaling protocol. This signaling protocol is a good alternative to 4-phase signaling for long interconnects because each communication across the link requires only two communication actions as opposed to four. This saves energy and time by eliminating the return to zero phase of four-phase signaling.

In asynchronous data encoding, dual-rail encoding is preferable over single-rail (bundled-data) encoding due to its delay-insensitive property. This property makes the data transfer very robust, because the sender and receiver can communicate reliably regardless of delays in the wire, in a delay-insensitive manner. But at the same time it has wiring cost because it requires 2N wires to encode N-bit data. In the conventional single-rail encoding data signals use normal Boolean levels to encode data and separate request and acknowledgement

wires are required [3]. In this encoding data should be stable before the data validity signal is activated. If the data interconnect has larger delay than the request one, the communication can be erroneous due to the arrival of request before the stable data reaches to the receiver. This timing constraint makes the communication delay-sensitive. But in this link special single-rail encoding which has delay-insensitive property is designed. This is achieved by using different current values to encode data and request together on one interconnect. Since the communication protocol is two-phase signaling there are four possible combinations of request and data as shown in Figure 1. These four different combinations are represented using four different current values of encoder current output as shown in Figure 3.
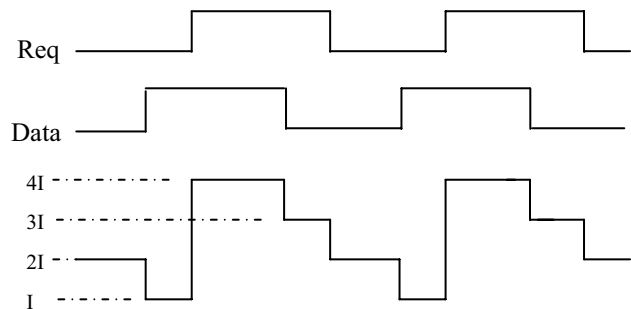

Figure 1. Encoding of Request and Data together using current-values

3. Link architecture

The proposed asynchronous link is capable of transferring data between two modules simultaneously in opposite directions. The block diagram of the link is shown in Figure 2, which has two transceivers placed opposite each other. The transceivers receive data from their own modules and output the data which is sent by the opposite module. The functions of transceiver blocks are explained as follows.

The encoder encodes the received data and request together into four different current levels. Then this encoded current is divided into two by the Current Subtractor unit and the pre-line. To restore to full current levels before it is fed to the line a current multiplier unit is used. The function of current Subtractor unit is to generate the current encoded by the opposite encoder. The decoder transforms the encoded current levels into its previously defined voltage levels after it takes output of Current Subtractor unit. In other words it decodes the current which is encoded by the other side of encoder into data and request. Since one acknowledgment signal is generated

per N-bit data transfer, only one Current Adder unit is required per N-bit data transfer. This current adder sums at least the output of three different Current Subtractor unit. For example, if there is 8-bit data transfer from both sides of modules. It takes first bit, last bit and 4th bit output of Current Subtractor and sum up together using the current adder. Then this sum sends to the acknowledgment line. The current adder which is placed in the other transceiver also performs the same thing. This means the output of the two current adders superimposed on each other on the acknowledgment interconnect. Then the current comparator compares the acknowledgments interconnect current with its own current adder output current. If the current of the acknowledgment interconnect is greater than the output current of the current adder then acknowledgment will be send to the module. This indicates the data sent by the module is received by the other module which is placed in the other side of the link. The current multiplier unit is designed from simple current mirror.

4. Design of link blocks

Both transceivers receive data and handshake signals from the module in voltage form then transform them into current representation and send to the interconnect. Then each transceiver changes data and handshake signals which are sent by the opposite module back to voltage form and make them ready for the module. The designs of link components are discussed below.
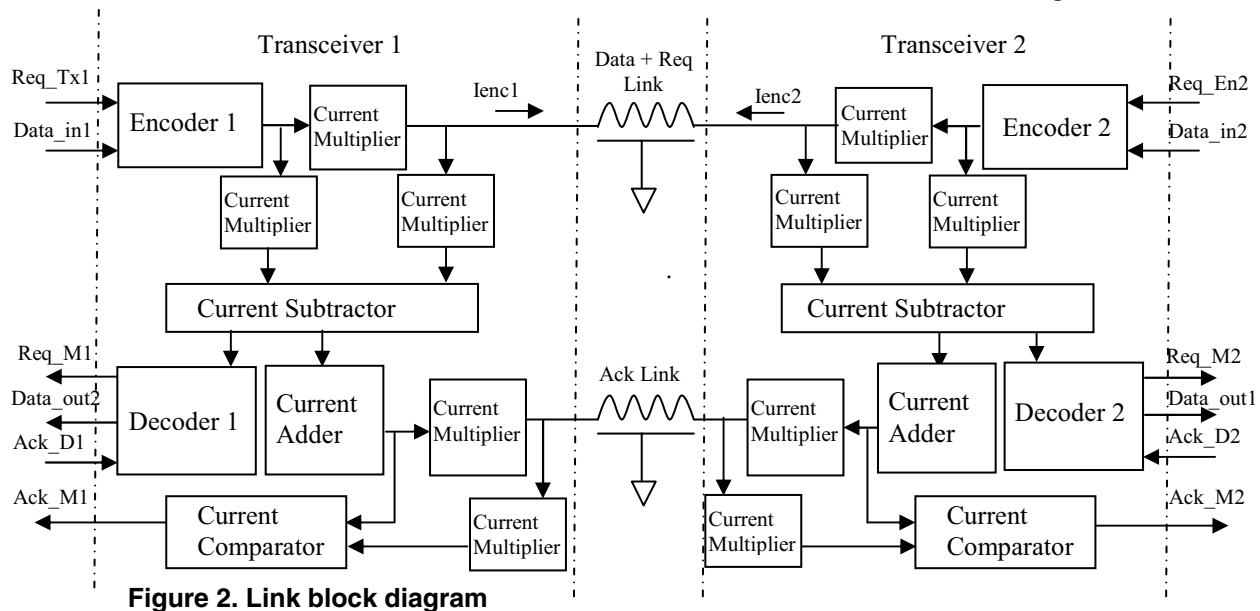
4.1. Encoder

The encoder bundles data and request together and change into four different current values because there are four possible combinations of data and request as shown in Figure 2. The encoder circuit is shown in Figure 3. Mn and Mp transistors serve as current source which generates constant current Is. Then Mp1, Mp2, Mp3 and Mp4 duplicate the current Is into four different values. The NMOS pass transistors are used to assign the current levels which are mapped to the combinations of data and request signals.
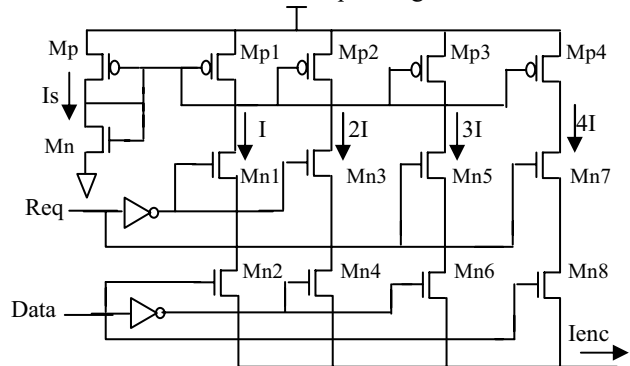

Figure 3. Encoder circuit

4.2. Current Subtractor

The Data + Req interconnect current is the sum of the two encoders current output, I_link = Ienc1 + Ienc2. The current subtractor shown in Figure 4 subtracts its side encoder current output from the

Figure 2. Link block diagram

interconnect current. As shown in Figure 4 the input I_link current divides between Mn2 and Mn3. By using simple current mirror the current which can pass through Mn2 becomes Ienc1. Since I_link = Ienc1 + Ienc2, current through Mn3 becomes Ienc2. Thus the output of the current subtractor is the current encoded by the encoder which is placed on the opposite transceiver.

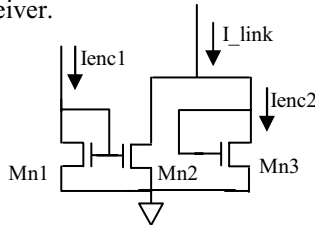**Figure 4. Current subtractor circuit**

4.3. Decoder

The decoder consists of current comparator, XNOR and NAND gates. The XNOR gate is designed using Differential Cascode Voltage Switch Logic (DCVSL), which eliminates the static power consumption. The current comparator compares the input current with four different threshold currents as shown in Figure 5a. Transistors Mn and Mp used as current source and other four PMOS transistors which duplicate the Ith current into four different threshold currents. Its input current Ienc2 comes from the current subtractor output, the current which is sent by the other side of the encoder. This input current has four different current levels as shown in Table II. When the input current is I only V1 goes down because M1 consumes the entire 0.5I threshold current to drive the current. When the input current is 2I, V1 and V2 go down because M1 and M2 consume their threshold currents to drive the 2I current. The same principle applies for all voltage values shown in Table I. The data output which is sent by the other side of the module is XNOR of V2 and V4. And the request output is NAND of V3 and V4 as shown in Figure 5b. For example take the case where Ienc2 is 3I, then Data_out2 = V2 XNOR V4 = 0 XNOR 1= 0 and Req_M1= V3 NAND V4 = 0 NAND 1 = 1. This result can be cross checked from Figure 2, the current representation of 3I means   Data = 0 and Req = 1.

Table 1. Decoding of encoded current to voltage

Ienc2	I	2I	3I	4I
V1	0	0	0	0
V2	1	0	0	0
V3	1	1	0	0
V4	1	1	1	0
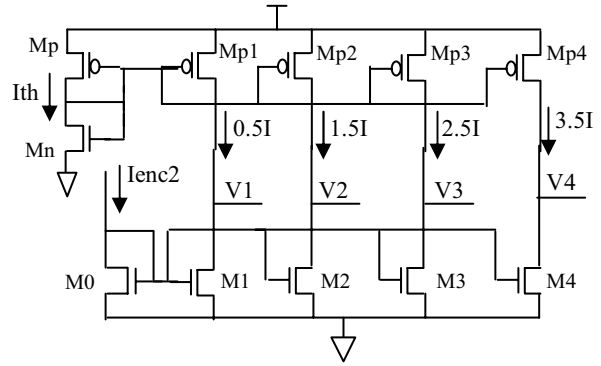
a) Current comparator
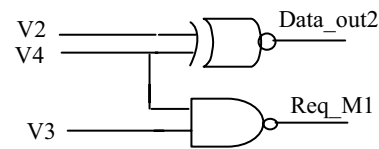
b) Binary logic gates

Figure 5. Decoder circuit

4.4. Current Adder and current Comparator

The purpose of these two blocks is to generate an acknowledgment signal. One acknowledgment signal is required for N-data bit transfer. Thus one current adder and current comparator is required for N-bit data transfer in each transceiver. Each of these current adders takes current output of three different Current Subtractor units, sum up together and send it to acknowledgment link. Then the current comparator compares the current on the acknowledgment link with its current adder output. When the acknowledgment link current is greater than the current adder output, the D-latch output, Ack_M1 becomes Req_Tx1. In other words acknowledgment signal follows request signal transition. The circuit which outputs acknowledgment signal is shown in Figure 6.
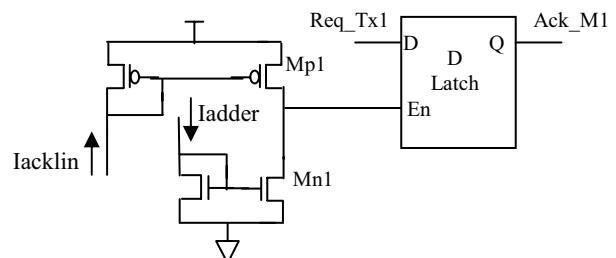

Figure 6. Circuit for generating acknowledgment signal to the module

5. Performance Comparison

The corresponding unidirectional link consists of only Encoder and Decoder. The number of transistors used for unidirectional and bidirectional link design is 44 and 112 respectively. The number of transistors used for bidirectional is greater than two times of the unidirectional, the remaining 24 transistors used in multiplying and subtracting the current. On-chip interconnects can be divided into three different types depending on their length [9]. Local interconnect less than 2mm, intermediate interconnect within 2 to 4mm and global interconnect greater than 4mm. The propagation of signal in unidirectional link is two times faster compared to bidirectional for local interconnects as shown in Figure 7. For intermediate interconnects both link has comparable signal propagation speed. But for global interconnects bidirectional link has faster signal propagation. In case of average power consumption of these links, bidirectional link consumes greater than four times of unidirectional as shown in Figure 8. Therefore the tradeoff for reducing the number of interconnects by half and becoming faster for global interconnects is power consumption. In terms of power supply noise both links isolate the signal from power supply noise.
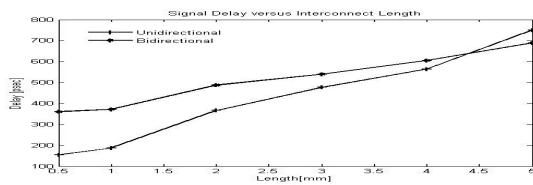

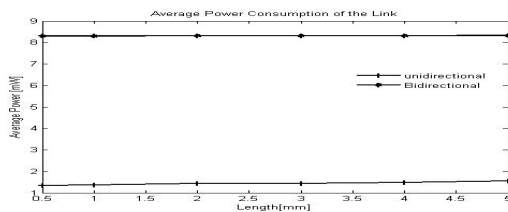**Figure 7. Signal Propagation Delay Versus Interconnect Length**

Figure 8. Average Power Consumption versus Interconnect length

5. Simulation of bidirectional link

The link between the two transceivers is designed using 10π RC model of interconnect. When the interconnect length varies from 0.5 to 5mm, the voltage swing on the link and average power dissipation varies

from 95 to 110mV and 8.30 to 8.32mW respectively. The effect of interconnect length variation on signal propagation delay is shown in Figure 8. Since the link is simultaneous bidirectional link, the sum of the two encoder current output on the link results in seven different current levels as shown in Table II. As expected the seven distinct current-levels on the link is shown in simulation waveforms Figure 9. Also waveforms of the two encoder's current output, the voltage-swing on the link and the two decoder's data and request outputs are shown.

Table 2. Sum of two encoders current output on the link

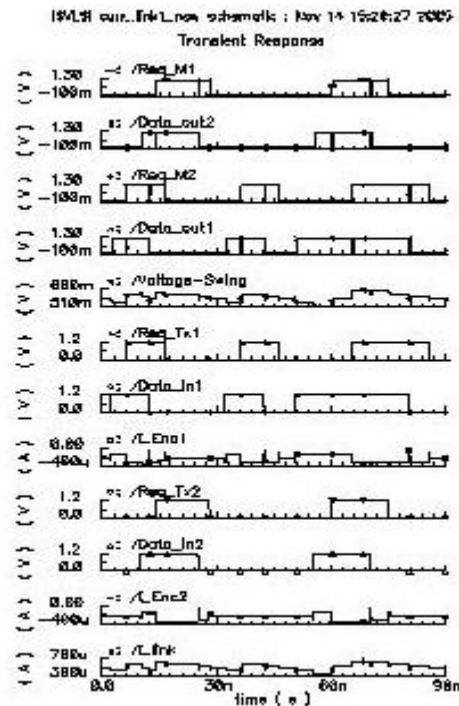| Ienc1 / Ienc2	I	2I	3I	4I
I	2I	3I	4I	5I
2I	3I	4I	5I	6I
3I	4I	5I	6I	7I
4I	5I	6I	7I	8I

Figure 9. Simulation waveforms

6. Application of the Link

This link can be used in Globally Asynchronous Locally Synchronous System (GALS). In SoC clock distribution and alignment has become an increasingly challenging problem, consuming an increasing portion

of resources such as wiring area, power, and design time. One of the solutions to this problem is to build SoC from several independently clocked subsystems which communicate each other through self-timed handshake signaling [3]. Such system enables flexible use of stoppable clocks providing automatic power down of idle system modules [10]. In addition to this, it makes easy to have a modular system. This simplifies the design process of a complex system enabling easy re-usage of different synchronous nodes. So this link can be used in GALS system between synchronously clocked local modules which has synchronous to asynchronous interface.

The other application area of this link is between two routers of Network-on-Chip (NoC). The design of larger more complex systems becomes an increasingly difficult task because of the many different issues related to the productivity, design reuse, technology and cost that have to be tackled simultaneously [11-12]. In this context Intellectual Property (IP) reuse becomes more and more important. However, building a system reusing existing IP blocks requires standardaised interfaces. One of a disciplined and scalable solution is offered by NoC [13]. Thus our link can be used between two routers which have one or more IP attached to it. Since the link is bidirectional, it is inherently reconfigurable in terms of information transfer direction. This means the router can route the information in opposite directions without any other effort.

It is also a good nominee for adaptive and reconfigurable link. Due to increasing number of noise sources and its level, technological parameter and environmental variations, communicating information reliably becomes difficult in future SoC. So to have reliable communication there should be some way of error detecting and correcting mechanism. To do this there should be feedback from the receiver side about the information transfer quality. This type of feedback communication can be done easily without any other effort using our bidirectional link.

7. Conclusions

In this work the realization of simultaneous bidirectional delay-insensitive asynchronous link using 2-phase handshake protocol and multiple-valued current mode scheme is presented. Unlike the conventional dual-rail delay-insensitive link which requires 2N +1 interconnects N/2 + 1 interconnect is required for N-data bit transfer since both transceivers

can send data simultaneously. The link circuit is composed of simple current mirrors, current comparators, and current subtractors. From circuit-level simulation results using CMOS 0.13um technology when the interconnect length varies from 0.5 to 5mm the power consumption of the link varies from 8.30 to 8.32mW. Also the voltage swing and signal propagation delay on the link varies from 95 to 110mV and from 360 to 689ps respectively. These results show that the effect of link length on power consumption and signal propagation delay is not as significant as the conventional voltage-mode link. Furthermore its signal propagation delay for 5mm length is 689psec which makes the link suitable for high speed data transfer on global on-chip interconnect. As performance comparison of unidirectional and bidirectional link indicates the cost of making the link simultaneously bidirectional is power consumption increase.

8. References

[1] M. Shams, et al, "Asynchronous Circuits", John Wiley's Encyclopedia of Electrical Engineering, pp. 716-725, 1999.

[2] Peter A. Beerel, "Asynchronous Circuits: An increasingly Practical Solution", Proc. ISQED, pp. 367-372, 18-21 March 2002.

[3] J. Sparso, and S. Furber, "Principles of Asynchronous Digital Design – A System Perspective", Kluwer Academic Publishers, Boston 2001.

[4] H. B. Bakoglu, Circuits, interconnections and Packaging for VLSI. Addison-Wesley, 1990.

[5] V. Adler, et al, "Repeater design to reduce delay and power in resistive interconnect," IEEE Trans on Circuits and Systems - II, vol. 45, no. 5, pp. 607–616, May 1998.

[6] D. Pamunuwa, H. Tenhunen, "Repeater insertion to minimise delay in coupled interconnects", VLSI Design, 3-7 Jan 2001, pp. 513-517.

[7] W. J. Dally and J. W. Poulton, Digital Systems Engineering. Cambridge University Press, 1998.

[8] R. Bashirullah, W. Liu, and R. K. Cavin, "current-mode signaling in deep submicrometer global interconnects", vol. 11, no. 3, pp. 406-417

[9] J. Nurmi, H. Tenhunen, J. Isoaho, and A. Jantsch, "Interconnect-Centric Design for Advanced SoC and NoC ", Kluwver Academic Publishers, 2004.

[10] J. Muttersbach, et al, "Practical Design of Globally-Asynchronous Locally-Synchronous Systems", Proc. ASYNC, April 2000.

[11] A. Jantsch, and H. Tenhunen, "Will Networks-on-Chip close the productivity gap?", in Networks on chip Kluwver Academic Publishers, 2003, pp. 3-18.

[12] L. Benin, and G. De Micheli, "Networks on Chips: A new SoC Paradign", Computer, 2002, 35, (1), pp. 70-78.

[13] Dally, W.J., and Towles, B.: 'Route packets, not wires: on-chip inteconnection networks'. Proc. 38th Annual Design Automation Conf. (ACM Press, 2001), pp. 684–689.

IEEE
COMPUTER
SOCIETY