Design of an Energy-Efficient Ternary Current-Mode Intra-Chip Communication Link for an Asynchronous Network-on-Chip

Akira MOCHIZUKI[^a], Member, Hirokatsu SHIRAHAMA[^b], Nonmember, Yuma WATANABE[^c], Student Member, and Takahiro HANYU[^d], Member

SUMMARY An energy-efficient intra-chip communication link circuit with ternary current signaling is proposed for an asynchronous Network-on-Chip. The data signal encoded by an asynchronous three-state protocol is represented by a small-voltage-swing three-level intermediate signal, which results in the reduction of transition delay and achieving energy-efficient data transfer. The three-level voltage is generated by using a combination of dynamically controlled current sources with feedback loop mechanism. Moreover, the proposed circuit contains a power-saving scheme where the dynamically controlled transistors also are utilized. By cutting off the current paths when the data transfer on the communication link is inactive, the power dissipation can be greatly reduced. It is demonstrated that the average data-transfer speed is about 1.5 times faster than that of a binary CMOS implementation using a 130nm CMOS technology at the supply voltage of 1.2V.

key words: asynchronous communication link, network-on-chip, multiple-valued logic, current-mode

1. Introduction

The growing performance in network-on-chip (NoC) is getting accelerated by the development of nanometer-scale VLSI fabrication technologies. Asynchronous NoC architecture has potential advantages such as low-power consumption maintaining total performance [1]–[6]. An asynchronous routing is suitable for a many-core processing system which consists of processing elements (PEs) and communication links. When a large network system is built, the energy efficiency of the I/O interface circuits and communication links is required with maintaining the performance in the asynchronous NoC platform.

In this paper, a ternary current-mode I/O interface circuit at a driver is proposed for an intra-chip asynchronous communication link. Since a timing-information signal instead of clock is required at the asynchronous data transfer, the four-phase dual-rail [11] encoding is utilized. This encoding needs a 1-bit data and a dummy data (called Spacer) for distinguished between present data and previous one on a continuous data stream, so that two wires are required. Using a ternary encoding, the two-bit information is superimposed on only one wire. Hence, the wire complex problem can be solved. A combination of dynamically controlled current sources with the feedback loop is also utilized in the proposed circuit. There are four controllable current sources on a one-level voltage generator. There are three features in this structure. The first one is to keep the voltage at the middle level of supply voltage on the wire because the voltage swing on the wire is kept small. The second one is to accelerate the transition time to the threshold level. The last one is to slower the transition over the threshold level for realizing the energy efficiency.

Using the proposed ternary current-mode I/O driver at the asynchronous router, the performances of 2x2 and 3x3 mesh-structure NoC platforms are estimated. The performance estimation of the NoC platform utilizes the high-speed high-accurate simulator with SystemC and the asynchronous routing model [7]–[9] which are our previous work. In case of the 3x3 full-mesh topology at the packet injection rate of 20 (packets/μsec) and under a 130nm CMOS technology at the supply voltage of 1.2V, the average data-transfer speed is about 1.5 times faster than that of a binary CMOS implementation, where the energy-delay product (EDP) of the proposed I/O interface circuit is about 75 percent of the CMOS one.

2. Asynchronous Encoding and Its NoC Structure

2.1 Four-Phase Dual-Rail Protocol

Figure 1 shows a block diagram of a typical model for asynchronous systems. There is no global clock and a handshake mechanism which has Request, Data and Acknowledgment. In this paper, a quasi delay insensitive (QDI) logic-circuit style with a four-phase dual-rail protocol [11] is used as this hardware implementation. Since this protocol is represented by three states which have logic values “0” and “1” as binary data, and an extra phase state “Spacer”, two wires as binary signaling are required. Table 1 shows the assignment of dual-rail signals (x, x’). Figure 2 shows the timing chart of data stream and its signal transition. The acknowledgement signal (Ack) is sent back to the previous module to notice receiving data.

The module is waiting for changing to the next data when the Ack signal is alternate transition. In Fig. 2 (a), the sequence of two-time data transfer is described and to send 1-bit data requires four phases, “send Data” and “wait Ack” at Data Phase, and “send Spacer” and “wait Ack” at Spacer...
Phase. In Fig. 2 (b), there are three important issues. The first one is that Data or Spacer is not sent to the following module until the Ack signal is alternatively changed. The second one is that the Ack signal is not changed until one of the signals x or x’. These dependencies make sure that the data transfer without clock. The last one is that the dual-rail signals are based on Return-to-Zero protocol, so that only one signal, x or x’, is alternated when the phase is changed.

The purpose in this paper is to decrease the number of link wires using ternary signaling, which result in the reduction of wiring complexity and wiring delay.

2.2 Asynchronous NoC Structure and Its Router

Figure 3 shows a block diagram of a typical mesh-topology NoC structure. The processing element (PE) contains a CPU and/or the instruction and the data memories. The asynchronous router (AR) is connected to four direction’s other ARs between communication link wires and one PE. The data transfer on the AR and on the links is performed asynchronously.

Figure 4 (a) and (b) show the mesh-type topological structures of 2x2 and 3x3 PEs, respectively, for example. In this paper, the performance of the NoC platforms is simulated using above two types. To simplify the simulation, the data-transfer to the outside is ignored.
connects to the inputs of OU0, OU1, OU2, and OU4. IU0 and OU1 are connected to the PE. The other units are connected neighbor AR through the communication links.

Figure 6(a) and (b) shows block diagrams of the IU and the OU, respectively. The asynchronous routing function such as controllers and latches and their delay table are based on the previous work [8]. This paper focuses design of the I/O interface circuit at the AR using current-mode circuit techniques. On the Data line, a three-level voltage driver and a three-level detector are utilized.

This communication link wire is reduced to a wire per 1-bit using representation of three-level voltage signal. On the Ack line, a two-level voltage driver and a two-level detector are utilized. This communication link wire is represented by two-level voltage signal, not a CMOS full-swing signal, which results in reduction the delay of the Ack signal.

3. Proposed I/O Interface Circuit

3.1 Ternary I/O Interface Circuit

Figure 7 shows the block diagram of the three-valued I/O interface circuit. On the driver, there are two voltage generators for producing the ternary level signal and three kinds of inverters for feedback control are required. On the detector, there are three kinds of inverters which are as same as those of the driver. The pair of binary voltage signals \((in, in')\) is the four-phase dual-rail coding data whose voltage swing is from GND to VDD. The pair of voltage outputs \((out, out')\) from the detector becomes also the binary full-swing signal. The communication link is single wire and its voltage becomes the middle levels of VDD. These voltage levels are generated by two voltage generators. To generate the three-level signal, the proposed circuit contains two 1-level generators whose corresponding logic level is 0 or 1. Using a liner summation technique, a combination of two generators’ condition enables to produce three-level signal.

Table 2 shows the three-valued level assignment of 1-bit four-phase dual-rail coding on one link wire where the input pair signal \((in, in')\) is as same as the output pair \((out, out')\). Each three-valued level corresponds to around middle voltage of supply voltage VDD. In fact, the voltage-level assignment is shown in Table 3 by using a 130nm CMOS technology at the supply voltage of 1.2V.

In Fig. 7, two kinds of inverters compare the voltage level of the link wire with its own threshold voltage Vth. For example, the \(V_{0.5} \) indicates the middle voltage corresponding to logic values between 0 and 1, so that the output of the inverter becomes high if the input logic level is 0. In contrast, the output becomes low if the input is 1 or 2. The point of view of voltage level, the \(V_{0.5} \) is equivalent to \(5/12\text{VDD} \) which is calculated from Table 3. The result of the small-swing comparison is represented by full-swing binary voltage since the inverter acts as an amplifier. Two inverters with no Vth indication are the same as a standard CMOS gate. The shifted-threshold inverters are realized by determined the beta ratio where the gate-channel widths of the NMOS and the PMOS are mainly changed from a standard CMOS inverter cell. It is easy to design the inverters.
since their threshold voltages are around 1/2VDD. Actually, the threshold voltages of \( V_{0.5} \) and \( V_{1.5} \) become 5/12VDD and 7/12VDD, respectively.

Figure 8 shows the circuit diagram of the voltage generator. The voltage generator produces logic values 0 or 1 whose voltage is 1/3VDD or 1/2VDD, respectively. Logic value 2 whose voltage is 2/3VDD can be produced when both ones generate logic level 1. To perform the voltage summation and to keep the middle voltage level can be done because the voltage level is generated by the current balance of four current sources, M1, M2, M3 and M4. Two voltage levels are determined by the current levels of M1/M4 and M2/M3, so that each size of a MOS transistor is carefully set in order to generate the target voltage.

The detail behavior of the MOS transistors is provided as follows:

- To save power, the external input signal \( \text{actv} \) controls the current paths of four current sources. The generator becomes in operation mode when the \( \text{actv} \) is high. In contrast, it becomes in sleep mode when the \( \text{actv} \) is low. Hence, all the current paths are cut off. This mode is used for no data-transfer on the link.
- The output wire becomes floating state and the voltage is not determined when the generator is in the sleep mode. In case of the sleep mode, the detector becomes unstable and has steady current path of each inverter. To suppress this condition, the output wire is connected to GND by turning ON of MS when the sleep mode is activated.
- MP and MN act as an inverter which tern ON/OFF of M1 and OFF/ON of M2, respectively, when its switch state depends on the voltage level of an external input signal \( \text{in1} \), low/high.
- M1 and M2 are controllable current sources. The current levels of M1 and M2 are determined by the voltage level on the output wire \( \text{out} \) which connects the communication link. In other words, the currents of M1 and M2 are controlled by the self-feedback loop.
- The lower the output voltage becomes, the larger the current of M1 flows. On the other hand, the higher the output voltage becomes, the larger the current of M2 flows.
- M3 and M4 are switched current sources through their current mirrors controlled by the other input signal \( \text{in2} \). The input current of the PMOS current mirror are controlled by MPS. Similarly, the input current of the NMOS current mirror are controlled by MNS. The state ON/OFF of MPS and MNS is determined by the input signal level of \( \text{in2} \).
- As the basic operation, M3 is ON and M4 is OFF when the voltage level of the output wire becomes low. On the other hand, M3 is OFF and M4 is ON when the voltage level of the output wire becomes high.
- On the stable conditions of four current sources, if M1 and M4 are ON, their current levels generate higher voltage level at the output wire. If M2 and M3 are ON, lower voltage level is generated.

3.2 Binary I/O Interface Circuit

Figure 9 shows the block diagram of the binary I/O interface circuit. This circuit utilizes for generating the Ack signal. To improve the handshaking speed at the asynchronous control, all the current-mode circuits are required instead of CMOS implementation. There is one voltage generator and a threshold-shifted inverter for feedback control in the driver. The detector is structured by only one inverter with a \( V_{0.5} \) threshold. The logic level on the communication link becomes 0 or 1. Consequently, the voltage level becomes 1/3VDD or 1/2VDD corresponding with the logic level. To achieve energy efficiency, the binary generator has also the sleep mode and the input \( \text{actv} \).

3.3 Dynamically Controlled Current Sources with Feedback Loop

Figure 10 shows the timing chart of the voltage transition model. Assume that the voltage generator be operated in case of switching the logic level from 1 to 0 and from 0 to 1. There are six modes, from (1) to (6). At the mode (1) and (4), the generator is at stable condition which is before and after transition. At the modes (2) and (5), the generator is to accelerate the voltage transition for reduction the delay until achieving the threshold voltage \( V_{0.5} \). At the modes (3) and (6), the generator is to slow down the voltage transition.
for realizing the energy efficiency. It is enough to accelerate the voltage transient till the threshold voltage, because the transition voltage is already over the threshold voltage. This acceleration consumes more power, so that the slow transient is performed for lowering energy after detecting the change the logic value.

The detail behavior of the four current sources as shown in Fig. 11 is explained as follows:

- In Fig. 11 (a), at the mode (1), the generator is on the stable condition before transition and produces the higher voltage level $V_1$ determined by the current levels of M1 and M4. M2 and M3 are OFF.
- When the input signal $in1$ is changed, the inverter transistors MP/MN are switched from ON/OFF to OFF/ON at the mode (2). The current sources M1/M2 is immediately switched to OFF/ON. The voltage level of the output wire out is quickly going down, because two current sources, M2 and M4, discharges the output wire connected the communication link.
- Moreover, large current is flowing thorough M2 since the current driving capability of M2 is high by the self-feedback connection between the gate and the output wire. In other word, the voltage level of the output wire connected to the gate of M2 is still high, so that the ON current is getting higher.
- Hence, since the voltage transition is accelerated by the M2 and M4 at the mode (2), the detector can quickly sense the change of the logic level.
- In Fig. 11 (b), when the voltage level of the output wire is already achieved to the threshold voltage $V_{0.5}$ at the mode (3), the detector also senses and switches the input signal $in2$ which result in turning M3/M4 into ON/OFF.
- At the same time, the current driving capability of M2 is going down due to raising the output voltage level, so that the transition is slowing.
- In Fig. 11 (c), when the output voltage level becomes $V_0$ at the mode (4), the current level of M2 is stable and set to the determined current to keep $V_0$ with M3.

In Fig. 11 (d), (e) and (f), the behaviors are shown when the logic value is changed from 0 to 1.

- In Fig. 11 (d), at the mode (4), the generator is produces the lower voltage level $V_0$ determined by the current levels of M2 and M3. M1 and M4 are OFF.
- When the input signal $in1$ is changed, the inverter transistors MP/MN are switched from OFF/ON to ON/OFF at the mode (5). The current sources M1/M2 is immediately switched to ON/OFF. The voltage level of the output wire out is quickly going up, because M2 and M4 charge the output wire.
- The large current is flowing thorough M1 since the current driving capability of M1 is high by the self-feedback. The voltage level of the output wire connected to the gate of M1 is still low, so that the ON current is getting higher.

- Then, since the voltage transition is accelerated by the M1 and M3 at the mode (5), the detector can quickly sense the change of the logic level.
- In Fig. 11 (e), when the output voltage level is already achieved to the threshold voltage $V_{0.5}$ at the mode (6), the detector also senses and switches the input signal
in2 which result in turning M3/M4 into OFF/ON.

- At the same time, the current driving capability of M1 is going down, so that the transition is slowing.
- In Fig. 11 (f), when the output voltage level becomes \( V_1 \) again at the mode (1), the current level of M1 is stable and set to the determined current to keep \( V_1 \) with M4.

To realize the energy efficiency, the voltage transition when the wasted energy is consumed can be slowing to control the current level dynamically by using the feedback loop.

The I/O interface circuits are designed for the variation tolerant. Actually, the transistor size of one of the current sources is about 70 times larger than that of the logic gate cell, since the output-voltage variation is designed within one percent for the process variation. The larger transistors are not only obtained the large current level, but also reduce the variation influences using the large channel length \( L \).

4. Evaluation

We have evaluated the proposed I/O circuit which is designed by using a 130nm CMOS technology at the supply voltage of 1.2V. Even if the proposed circuit is based on current-mode style, there is no unit current. The current sources are utilized for determined the voltage level, and are dynamically changed the current value.

4.1 HSPICE Simulation

To evaluate the performance of the proposed intra-chip I/O interface circuit, the communication link model is utilized as shown in Fig. 12. The length of the link wire is 1mm. Figure 13 shows the waveforms of the input/output signals and the communication link signal. On the input and the output waveforms, the solid line and the broken line indicate \( x \) and \( x' \), respectively, at the four-phase dual-rail coding (\( x, x' \)). The outputs are obtained correctly. The expected link signal is also obtained in spite of the RC parasitic influence.

Figure 14 shows the waveforms of input/output and the communication link signals corresponding to Fig. 13 at the worst PVT condition, where the worst condition of the process parameters are used, the supply voltage and signal voltages are 10 percent reduction of 1.1V, and the temperature is \(-40 \) degree Celsius. In the worst process condition, the both switching speeds of n-channel and p-channel transistors are the slowest/slowest at the temperature of \(-40 \) degree Celsius rather than 125 degree Celsius. Even if the circuit condition is under the worst, the normal operation of the communication link is done. Although each signal levels become low, the threshold level of the inverters at the driver and the detector is relatively decreasing, so that the detection of the three-valued level can be performed. Moreover, the normal detection can be done in this condition, even if the voltage swing becomes low because of higher gain of each inverter.

4.2 Performance Comparison

Table 4 shows the performance comparison among the proposed I/O interface circuit, a conventional current-mode interfaces such as current-mode logic (CML) [12], and a conventional CMOS implementation. Assume that a packet consist of three Data flits and three Spacer flits. The Data contains a header, a body, and a tail flits. The length of a flit is 12 bit which has 2-bit control and 10-bit payload. The 2-bit control indicates the header, the body, the tail or the Spacer. The 10-bit payloads of the body and the tail are real data. The header payload contains five 2-bit routing data which indicates one of four direction without own one to send the packet to next AR. Five routing data are enough to send the packet at all routing ways of the 3x3-mesh NoC. The EDP of the proposed interface circuit is about 75 percent of the CMOS implementation. Although the area of
Table 4  Performance comparisons.

<table>
<thead>
<tr>
<th></th>
<th>CMOS</th>
<th>CML</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Binary</td>
</tr>
<tr>
<td>Delay (ns)</td>
<td>0.85</td>
<td>0.26</td>
<td>0.34</td>
</tr>
<tr>
<td>Throughput (Gbps)</td>
<td>0.87</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Number of wires</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Throughput per wire (Gbps)</td>
<td>0.44</td>
<td>0.45</td>
<td>0.80</td>
</tr>
<tr>
<td>Power (500MHz, mW)</td>
<td>0.38</td>
<td>14</td>
<td>1.4</td>
</tr>
<tr>
<td>Static state (nW)</td>
<td>2.7</td>
<td>14x10^6</td>
<td>42</td>
</tr>
<tr>
<td>Energy per packet (pJ)</td>
<td>9.1</td>
<td>94</td>
<td>5.4</td>
</tr>
<tr>
<td>EDP per packet (pJ·ns)</td>
<td>7.8</td>
<td>24</td>
<td>5.9</td>
</tr>
<tr>
<td>Area (Driver+Detector) per wire (um^2)</td>
<td>94</td>
<td>613</td>
<td>380</td>
</tr>
</tbody>
</table>

I/O interfaces is limited to 13 percent in comparison with that of an AR, where the interfaces are five directions and each direction has 12-bit width. Figure 15 shows the average latency for transferring a packet between arbitrary PEs at the 2x2-mesh type NoC. Figure 16 shows also the average latency at the 3x3-mesh type NoC. The injection rate means the average number of packets per time. One PE injects a packet into the link network. The injection timing and the destination PE are randomly decided [10].

The transition times of entering the sleep mode and the wakeup from the sleep are 350ps and 480ps, respectively. The wakeup time is enough to start the transfer data from the AR.

When the NoC structure is a 2x2-mesh style and the injection rate is 25 (packets/µsec), the average speed is 1.43 times faster than that of a binary CMOS implementation. When the structure is a 3x3-mesh style and the injection rate is 20 (packets/µsec), the average speed is 1.52 times faster. In above cases, the EDP of the proposed circuit is about 75 percent of the CMOS realization. These performance estimations utilize the simulator with SystemC and the asynchronous routing model [7]–[9] using a 130nm CMOS technology at the supply voltage of 1.2V. These result indicates that the energy saving is achieved where the average speed of the packet transfer is normalized.

5. Conclusion

An energy-efficient ternary current-mode interface circuit has been proposed for the intra-chip asynchronous communication link. In this proposed circuit, four-phase dual-rail coded three-valued levels are superimposed on one wire. The combination of the ternary coding and the dynamically controlled current with feedback loop mechanism enhances energy-efficient asynchronous data communication under the normalized the transfer speed. Actually, the average packet-transfer speed of the proposed design method is faster than that of the conventional one.

As the future prospects, using advanced process technologies such as 65nm, 45nm and more, the data transfer speed of the proposed asynchronous interface circuit at the communication link archives the present binary current-mode transmission rate which is up to the 10Gbps.

Acknowledgments

This work is supported by Core Research for Evolutional Science and Technology (CREST) of Japan Science and Technology Agency (JST). This work is also supported partially by VLSI Design and Education Center (VDEC), Tokyo University in collaboration with Synopsys, Inc. and Cadence Design Systems, Inc.

References


Akira Mochizuki received the B.E., the Master of Information Science and the D.E. degrees from Tohoku University, Sendai, Japan, in 1993, 1995 and 2006, respectively. From 1995 to 2002, he joined NEC, Japan. From 2002 to 2007, he joined Renesas, Japan. In both the semiconductor companies, he was engaged in the design and development of the embedded CPU IPs. From 2002 to 2007, he was a Research Associate for researching high-performance integrated circuits in Tohoku University. He is currently an Assistant Professor in CSIS, Tohoku University. His main interests and activities are a design of high-speed low-power VLSI and its application. Dr. Mochizuki is a member of the IEEE.

Hirokatsu Shirahama received the B.E., M.E. and D.E. degrees in Electronic Engineering from Tohoku University, Sendai, Japan, in 1984, 1986 and 1989, respectively. He is currently a Professor in the Research Institute of Electrical Communication, Tohoku University. His general research interests include nonvolatile logic circuits and their applications to ultra-low-power and/or PVT-variation-free VLSI processors, and multiple-valued current-mode circuit and its application to power-aware asynchronous Network-on-Chip systems. He received the Sakai Memorial Award from the Information Processing Society of Japan in 2000, the Judge’s Special Award at the 9th LSI Design of the Year from the Semiconductor Industry News of Japan in 2002, the Special Feature Award at the University LSI Design Contest from ASP-DAC in 2007, the APEX Paper Award of Japan Society of Applied Physics in 2009, the Excellent Paper Award of IEICE, Japan, in 2010, Ichikawa Academic Award in 2010, the Best Paper Award of IEEE ISVLSI 2010, and the Paper Award of SSDM 2012. Dr. Hanyu is a Senior Member of the IEEE.

Yuma Watanabe received the B.E. degree from Tohoku University, Japan, in 2012. He is currently working towards the M.E degree at Tohoku University. His main interests and activities are in the asynchronous Network-on-Chips and data-transfer links.

Takahiro Hanyu received the B.E., M.E. and D.E. degrees in Electronic Engineering from Tohoku University, Sendai, Japan, in 1984, 1986 and 1989, respectively. He is currently a Professor in the Research Institute of Electrical Communication, Tohoku University. His general research interests include nonvolatile logic circuits and their applications to ultra-low-power and/or PVT-variation-free VLSI processors, and multiple-valued current-mode circuit and its application to power-aware asynchronous Network-on-Chip systems. He received the Sakai Memorial Award from the Information Processing Society of Japan in 2000, the Judge’s Special Award at the 9th LSI Design of the Year from the Semiconductor Industry News of Japan in 2002, the Special Feature Award at the University LSI Design Contest from ASP-DAC in 2007, the APEX Paper Award of Japan Society of Applied Physics in 2009, the Excellent Paper Award of IEICE, Japan, in 2010, Ichikawa Academic Award in 2010, the Best Paper Award of IEEE ISVLSI 2010, and the Paper Award of SSDM 2012. Dr. Hanyu is a Senior Member of the IEEE.