Asynchronous-Logic QDI Quad-Rail Sense-Amplifier Half-Buffer Approach for NoC Router Design

Weng-Geng Ho, Kwen-Siong Chong, Kyaw Zwa Lwin Ne, Bah-Hwee Gwee, and Joseph S. Chang

Abstract—We propose a low area overhead and power-efficient asynchronous-logic quasi-delay-insensitive (QDI) sense-amplifier half-buffer (SAHB) approach with quad-rail (i.e., 1-of-4) data encoding. The proposed quad-rail SAHB approach is targeted for area- and energy-efficient asynchronous network-on-chip (NoC) router designs. There are three main features in the proposed quad-rail SAHB approach. First, the quad-rail SAHB is designed to use four wires for selecting four NoC router directions, hence reducing the number of transistors and area overhead. Second, the quad-rail SAHB switches only one out of four wires for 2-bit data propagation, hence reducing the number of transistor switches and dynamic power dissipation. Third, the quad-rail SAHB abides by QDI rules, hence the designed NoC router features high operational robustness toward process-voltage-temperature (PVT) variations. Based on the 65-nm CMOS process, we use the proposed quad-rail SAHB to implement and prototype an 18-bit ANoC router design. When benchmarked against the dual-rail counterpart, the proposed quad-rail SAHB ANoC router features 32% smaller area and dissipates 50% lower energy under the same excellent operational robustness toward PVT variations. When compared to the other reported ANoC routers, our proposed quad-rail SAHB ANoC router is one of the high operational robustness, smallest area, and most energy-efficient designs.

Index Terms—Asynchronous network-on-chip (ANoC) router, energy efficient, process–voltage–temperature (PVT) variations, quad-rail data encoding, quasi-delay insensitive (QDI).

I. INTRODUCTION

Multicore processing is a promising option to increase the computation speed in many applications nowadays. In the multicore processing, the use of network-on-chip (NoC) router to link all the processing cores for efficient data communication is widely adopted. To date, there are many NoC routers reported, including those based on fully synchronous logic (sync) and those in part based on asynchronous logic (async) [1]–[10]. The design of sync NoC routers, however, becomes increasingly more challenging due to the timing issues for operation correctness. The timing issues are further compounded in view of wider process–voltage–temperature (PVT) variations in deep-submicrometer fabrication processes.

Conversely, the design of async NoC (ANoC) routers [1]–[10] has increasingly become attractive in accommodating timing and data synchronization issues because their operations are essentially self-timed. In the ANoC router designs, the quasi-delay-insensitive (QDI) timing model with 1-of-n rails data encoding is well known to accommodate the PVT variations [11].

However, the ANoC router design using the async QDI handshake protocol has higher circuit complexity as compared to the sync NoC router design due to the following reasons. First, the async QDI handshake protocol requires input completion and output completion circuits to validate the input and output completeness for every pipeline stage [12]. Second, the use of well-established dual-rail data encoding increases the computation block complexity and the number of wires along the pipeline stage [11]. In short, the ANoC router suffers from high area overhead due to the increased circuit complexity and high dynamic power dissipation due to the increased number of switching in the data propagation.

In this brief, we propose a quad-rail sense-amplifier half-buffer (SAHB) approach with low area overhead, power-efficient, and high operational robustness toward PVT variations. Based on 65-nm CMOS process, we implement and prototype an ANoC router using the proposed quad-rail SAHB approach. When benchmarked against the dual-rail counterpart, the proposed quad-rail SAHB ANoC router features 32% smaller area and dissipates 50% lower energy under the same excellent operational robustness toward PVT variations. When compared to the other reported ANoC routers, our proposed quad-rail SAHB ANoC router is one of the smallest area, high operational robustness, and most energy-efficient designs.

This brief is organized as below. Section II elaborates our proposed quad-rail SAHB cell and the circuit and pipeline analysis. Section III elaborates our proposed ANoC router’s architecture. Section IV presents the chip implementation and measurement results. Finally, Section V draws the conclusion.

II. PROPOSED QUAD-RAIL SAHB

We leverage on our SAHB realization approach [12], and design the async cells in quad-rail encoding. For ANoC router design, the basic quad-rail cells are buffers (BUF), multiplexers (MUX), and deMUX (DEMUX). The SAHB quad-rail cells are designed at transistor level. For the sake of simplicity, we only elaborate the SAHB quad-rail 4-to-1 MUX cell; the other cells can be built based on the same concept.

Fig. 1(a) depicts the block diagram of the SAHB quad-rail 4-to-1 MUX cell (with generic interface signals), which comprises four quad-rail input data \(L_0, L_1, L_2, L_3\), a quad-rail input select data \(S\), a quad-rail output data \(R\), an acknowledgment signal from the succeeding pipeline (Rack), and acknowledgment signals to the preceding pipeline (Lack0, Lack1, Lack2, Lack3, and Lack\(k\)). The complimentary signals are bundled for simplifying the cell structure.

Fig. 1(b) depicts the pipeline diagram of the SAHB quad-rail 4-to-1 MUX, which comprises four functional blocks (FBs) plus two control logic cum sense amplifier (CLSA), input completion detection (ICD), output completion detection (OCD), and C-Muller tree (CT). The four FBs + two CLSAs block propagates \(L_0, L_1, L_2, L_3\), and \(S\) to \(R\) based on Rack. The ICD validates the input availability/nullity by generating input validity signal Val for each input data, the OCD validates the output availability/nullity, and finally CT generates Lack for each input data.

Fig. 1(c) depicts the cell diagram of the SAHB quad-rail 4-to-1 MUX, which comprises four FBs, two CLSAs, five ICDs, an OCD, and a CT. The FBs, as depicted in Fig. 1(d), evaluate/precharge \(L_0, L_1, L_2, L_3\), and \(S\) to \(R\), depending on Rack.
The CLSAs, as depicted in Fig. 1(e), control the sequence of async operation, and help amplifying/latching \( R \). The ICDs generate \( V_{L,0} \), \( V_{L,1} \), \( V_{L,2} \), \( V_{L,3} \), and \( V_{S} \) to validate the input availability/nullity. The OCD and CT generate \( \text{Lack}_0 \), \( \text{Lack}_1 \), \( \text{Lack}_2 \), \( \text{Lack}_3 \), and \( \text{Lack}_S \) to acknowledge the preceding stage.

In the quad-rail SAHB 4-to-1 MUX cell, there are two supply voltages, namely, \( V_{DDA} \) for FBs and \( V_{DD} \) for remaining blocks (CLSA, ICD, OCD, and CT). For proper operation, \( V_{DD} \geq V_{DDA} \). \( V_{DDA} \) is fixed as subthreshold voltage (\( \sim 0.3 \) V) to fully minimize the switching power without compromising the evaluation/precharge speed since the output is amplified through CLSA. On the other hand, \( V_{DD} \) is adjustable from nominal voltage (1.2 V) for speed-critical applications to near-threshold voltage (\( \sim 0.6 \) V) and even subthreshold voltage (\( \sim 0.3 \) V) for low-power low-speed applications.

To analyze and evaluate our proposed quad-rail SAHB cells, we compare their transistor count and number of transitions per cycle against the reported dual-rail SAHB cells [12]. Since the quad-rail cells propagate 2-bit data in the pipeline, we normalize the readings to single bit for fair comparison with dual-rail counterparts (propagate 1-bit data). The comparison is made for four basic NoC functions: 1-to-1 buffering, 1-to-2 demultiplexing, 2-to-1 multiplexing, and 4-to-1 multiplexing. For sake of simplicity, we only analyze 4-to-1 multiplexing function; the other functions can be analyzed similarly.

First, from the circuit-level perspective, the transistor count per bit, \( TC(4\text{-to-1 MUX}) \) with the transistor count of the cell’s components such as FB, CLSA, ICD, OCD, CT, and Inverter (Inv) can be calculated, as expressed in the following equation:

\[
TC(4\text{-to-1 MUX}) = \frac{1}{2} \left[ 4 \cdot TC(FB) + 2 \cdot TC(CLSA) + 5 \cdot TC(ICD) + TC(OCD) + TC(CT) + 4 \cdot TC(\text{Inv}) \right] 
\] (1)

Second, from the pipeline-level perspective, we depict the three-stage marked graph behavior for our proposed quad-rail SAHB cell approach in Fig. 2. This graph illustrates evaluate operation (\( F^{e} \), \( \text{Inv}^{e} \), \( \text{ICD}^{e} \), \( \text{OCD}^{e} \), and \( \text{CT}^{e} \) ) and precharge operation (\( F^{p} \), \( \text{Inv}^{p} \), \( \text{ICD}^{p} \), \( \text{OCD}^{p} \), and \( \text{CT}^{p} \)) to propagate the data within \( i \)th, \((i + 1)\)th, and \((i + 2)\)th pipeline stages in cycle basis.

From the three-stage marked graph behavior, the number of transitions per bit for one cycle, \( T(4\text{-to-1 MUX}) \) with delay, \( t \) of the cell’s components can be calculated, as expressed in the following equation:

\[
T(4\text{-to-1 MUX}) = \frac{1}{2} \left[ t(F^{p}) + t(\text{Inv}^{p}) + t(F^{e}) + t(F^{p+1}) + t(\text{Inv}^{p+1}) + t(F^{e+1}) + t(\text{Inv}^{p+2}) + t(F^{e+2}) + t(\text{OCD}^{e}) + t(\text{CT}^{e}) + t(F^{p+2}) + t(\text{OCD}^{p}) + t(\text{CT}^{p}) \right] 
\] (2)
TABLE I

<table>
<thead>
<tr>
<th>2-bit SAHB ANoC’s Functions</th>
<th>Logic Gate Combination</th>
<th>Transistor-Count/bit, TC</th>
<th>No. of Transitions/(cycle×bit), T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-to-1 Buffering</td>
<td>Quad-Rail</td>
<td>1 QR BUF</td>
<td>33 (1×)</td>
</tr>
<tr>
<td></td>
<td>Dual-Rail</td>
<td>2 DR BUF</td>
<td>34 (1.03×)</td>
</tr>
<tr>
<td>1-to-2 De-Multiplexing</td>
<td>Quad-Rail</td>
<td>1 QR 1-to-2 DEMUX</td>
<td>87 (1×)</td>
</tr>
<tr>
<td></td>
<td>Dual-Rail</td>
<td>2 DR 1-to-2 DEMUX</td>
<td>90 (1.03×)</td>
</tr>
<tr>
<td>2-to-1 Multiplexing</td>
<td>Quad-Rail</td>
<td>1 QR 2-to-1 MUX</td>
<td>73 (1×)</td>
</tr>
<tr>
<td></td>
<td>Dual-Rail</td>
<td>2 DR 2-to-1 MUX</td>
<td>80 (1.1×)</td>
</tr>
<tr>
<td>4-to-1 Multiplexing</td>
<td>Quad-Rail</td>
<td>1 QR 4-to-1 MUX</td>
<td>111 (1×)</td>
</tr>
<tr>
<td></td>
<td>Dual-Rail</td>
<td>6 DR 2-to-1 MUX</td>
<td>240 (2.16×)</td>
</tr>
<tr>
<td>Average</td>
<td>1×</td>
<td></td>
<td>1,33×</td>
</tr>
</tbody>
</table>

Fig. 3. Realization of 2-bit SAHB 4-to-1 multiplexing function. (a) Quad rail. (b) Dual rail.

where \( t(F^p) = t(F^p) \). 2 trans., \( t(\text{Inv}^p) = t(\text{Inv}^p) \) = 1 tran., \( t(\text{OCD}^p) = t(\text{OCD}^p) \) = 1 tran., and \( t(\text{CT}^p) = t(\text{CT}^p) \) = 2 trans.

IV. MEASUREMENT RESULTS

The proposed quad-rail SAHB ANoC router is implemented by the full-custom approach based on the 65-nm standard-threshold-voltage CMOS process \((V_t \sim 0.5 \text{ V})\) and fabricated through STMicroelectronics’s circuits multiprojects solutions. Fig. 6 depicts the microphotograph of the proposed quad-rail SAHB ANoC router and the test structure.

IV. Measurement Results

Four interfaces are used to transfer data to/from four directions [i.e., north (N), east (E), south (S), and west (W)] between two neighboring ANoC routers/IOs, and the last interface to transfer data between the ANoC router, and its respective processing core.

As seen, the area of direction switch (4-to-1 multiplexing function) is smaller area than the dual-rail counterpart. This is due to the simpler implementation in quad-rail SAHB library cells than the dual-rail implementation (which requires on average 1.33× more transistors, in Table I).

Fig. 7(a) and (b) depicts the area breakdown of the implemented quad-rail and dual-rail SAHB ANoC routers, respectively. The quad-rail SAHB ANoC router occupies 0.105 mm², 1.47× smaller area than the dual-rail counterpart. This is due to the simpler implementation in quad-rail SAHB library cells than the dual-rail implementation (which requires on average 1.33× more transistors, in Table I). As seen, the area of direction switch (4-to-1 multiplexing function) reduces significantly from 53% in dual-rail design to 38% in quad-rail design; this mainly reduces the overall area.

To demonstrate the operational robustness toward PVT variations, we compare the throughput and energy dissipation of the proposed quad-rail SAHB ANoC router against three related QDI counterparts, i.e., dual-rail SAHB [12], quad-rail weak-conditioned HB (WCHB) [6], and quad-rail precharged HB (PCHB) [7]. These comparisons are evaluated under various variations of supply voltages (0.3, 0.6, 0.9, and 1.2 V), operating temperatures...
Fig. 5. ANoC router interface. (a) IP. (b) OP.

TABLE II

<table>
<thead>
<tr>
<th>ANoC Routers</th>
<th>Word-length</th>
<th>Handshake Protocol</th>
<th>Data-encoding</th>
<th>Cell Design Approach</th>
<th>Robustness to PVT Variations</th>
<th>CMOS</th>
<th>( V_{DD} )</th>
<th>Normalized Area (mm²) *</th>
<th>Normalized Latency (ns) *</th>
<th>Normalized Throughput (MHz) *</th>
<th>Normalized Energy Disp. per-bit (fJ/bit) *</th>
</tr>
</thead>
<tbody>
<tr>
<td>QNoC [1]</td>
<td>10-bit</td>
<td>Bundled-data</td>
<td>Single-rail</td>
<td>Standard</td>
<td>Low</td>
<td>350nm</td>
<td>1.2V</td>
<td>0.016</td>
<td>1.86</td>
<td>404</td>
<td>-</td>
</tr>
<tr>
<td>MANGO [2]</td>
<td>33-bit</td>
<td>Bundled-data</td>
<td>Single-rail</td>
<td>Standard</td>
<td>Low</td>
<td>130nm</td>
<td>1.2V</td>
<td>0.009</td>
<td>4.10</td>
<td>1300</td>
<td>-</td>
</tr>
<tr>
<td>SCAFFI [3]</td>
<td>8-bit</td>
<td>Bundled-data</td>
<td>Single-rail</td>
<td>Standard</td>
<td>Low</td>
<td>180nm</td>
<td>1.8V</td>
<td>0.043</td>
<td>-</td>
<td>252</td>
<td>2468</td>
</tr>
<tr>
<td>HORAK [8]</td>
<td>32-bit</td>
<td>Bundled-data</td>
<td>Single-rail</td>
<td>Mousetrap</td>
<td>Low</td>
<td>65nm</td>
<td>1.2V</td>
<td>0.008</td>
<td>3.02</td>
<td>746</td>
<td>262</td>
</tr>
<tr>
<td>ARGO [10]</td>
<td>32-bit</td>
<td>Bundled-data</td>
<td>Single-rail</td>
<td>Mousetrap</td>
<td>Low</td>
<td>65nm</td>
<td>1.2V</td>
<td>0.008</td>
<td>3.02</td>
<td>746</td>
<td>262</td>
</tr>
<tr>
<td>ONIZAWA [9]</td>
<td>34-bit</td>
<td>Timed</td>
<td>Dual-rail</td>
<td>LEDR</td>
<td>High</td>
<td>130nm</td>
<td>1.2V</td>
<td>0.023</td>
<td>1.37</td>
<td>1052</td>
<td>247</td>
</tr>
<tr>
<td>PAUST [4]</td>
<td>32-bit</td>
<td>QDI</td>
<td>Dual-rail</td>
<td>WCHB</td>
<td>High</td>
<td>130nm</td>
<td>1.2V</td>
<td>0.113</td>
<td>3.00</td>
<td>320</td>
<td>635</td>
</tr>
<tr>
<td>ALPIN [6]</td>
<td>32-bit</td>
<td>QDI</td>
<td>Quad-rail</td>
<td>WCHB</td>
<td>High</td>
<td>65nm</td>
<td>1.2V</td>
<td>0.170</td>
<td>2.30</td>
<td>550</td>
<td>938</td>
</tr>
<tr>
<td>ALHUSSEN [7]</td>
<td>66-bit</td>
<td>QDI</td>
<td>Dual-rail</td>
<td>PCHB</td>
<td>High</td>
<td>65nm</td>
<td>1.0V</td>
<td>0.025</td>
<td>10.40</td>
<td>100</td>
<td>2824</td>
</tr>
</tbody>
</table>

Proposed SAHB 18-bit QDI Quad-rail SAHB

*The readings are normalized with respect to the proposed SAHB ANoC Router @65nm CMOS, 1.2V.

Fig. 6. Microphotograph of the proposed SAHB ANoC router.

(−40 °C, 27 °C, and 100 °C), and process threshold voltages (low threshold voltage (LVT), standard threshold voltage (SVT), and high threshold voltage (HVT)), as depicted in Fig. 8(a)–(c), respectively.

Now, we elaborate the overall comparison in these variations. When compared to the dual-rail SAHB, the proposed design features ~50% lower energy dissipation due to the adoption of quad-rail data encoding which reduces 1.33× transistor counts and 3× transitions/cycle. The tradeoff is slightly low throughput partly due to a more complex completion detection. When compared to the quad-rail WCHB, the proposed design features ~15% higher throughput and ~13% lower energy dissipation. When compared to the quad-rail PCHB, the proposed design features ~27% lower energy dissipation. These improved attributes are achieved because the proposed SAHB approach embodies FBs that operate in subthreshold region (\( V_{DDA} = 0.3 \) V) to dissipate significantly low power (<1 μW), and CLSAs that employ sense amplifiers to enhance the speed, resulting in high energy efficiency.

Fig. 9 depicts the latency and throughput of the proposed SAHB ANoC router by varying \( V_{DD} \) from 1.2 to 0.3 V; the results are normalized with respect to the readings at 1.2 V. As seen, \( V_{DD} \)
is adjustable from nominal voltage, 1.2 V (with 5.7-ns latency, 258-MHz throughput) for speed-critical applications, to near-threshold voltage, 0.6 V (with 2.7× latency, 0.35× throughput), and subthreshold voltage, 0.3 V (with 38× latency, 0.03× throughput) for low-power low-speed applications.

For completeness, Table II tabulates the comparison of various ANoC routers, where the results, i.e., area, latency, throughput, and energy dissipation per bit of the reported ANoC routers [1]–[10] are normalized with respect to the proposed SAHB ANoC router at 65-nm CMOS, 1.2 V based on technology scaling. In terms of robustness toward PVT variations, the ANoC routers based on the QDI handshake protocol feature high operational robustness as compared to those based on bundled data and timed handshake protocols. This is due to the absence of timing issues in QDI handshake protocol.

Furthermore, their process technology, design architecture, and implementation are different, hence a direct comparison of various ANoC routers is somewhat contentious. For example, the reported ARGO [10] and ONIZAWA [9] appear to dissipate lower normalized energy per bit than our SAHB design, but they are not robust toward PVT variations. In addition, the reported ARGO needs to be associated by an external controller for complete operations, hence dissipating additional energy overhead. When compared within all QDI ANoC routers, our SAHB designs are, respectively, 35%, 56%, and 85% more energy efficient than the reported FAUST [4], ALPIN [6], and ALHUSSIEN [7].

V. CONCLUSION

We have proposed an async QDI quad-rail SAHB approach with emphases on low area overhead and power efficient, for ANoC router design. Based on the 65-nm CMOS process, we have implemented and prototyped an 18-bit quad-rail SAHB ANoC router. When benchmarked against the dual-rail counterpart, the proposed quad-rail SAHB ANoC router features 32% smaller area and dissipates 50% lower energy under the same excellent operational robustness toward PVT variations. When compared to the other reported ANoC routers, our proposed quad-rail SAHB ANoC router is one of the highest robustness, smallest area, and most energy-efficient designs.

REFERENCES