Optimal bus coding for OR-chained buses

Kon-Woo Kwon*

Abstract This letter presents a novel bus coding technique suitable for a bus with OR chaining that requires data parking at the end of data transmission. Based on the fact that the data parking state is known a priori, the proposed technique encodes the last sequential data value by jointly considering two switching activities, one that occurred due to a change in valid data values and the other that will be subsequently occurred by data parking. Compared to the conventional bus coding scheme that picks a codeword that minimizes only the former, the proposed technique can achieve >3% additional saving of switching activity in average. Evaluation results based on 22-nm CMOS technology show that the proposed scheme, even with its higher encoding energy, can obtain the lower total energy for >0.59 mm bus length in comparison with the conventional data bus invert technique.

key words: bus coding, data bus invert, OR chain, switching activity

Classification: Integrated circuits (memory, logic, analog, RF, sensor)

1. Introduction

Data movement in modern processors requires a large amount of energy owing to charging and discharging the capacitance of data bus [1, 2]. Studies [3, 4, 5] show that moving data may consume >10x higher energy than computing on a processor. As CMOS technology scales down, the energy gap between data movement and computation is expected to increase further [1, 5, 6, 13, 26, 30]. Hence, reducing data movement energy has a practical importance for energy efficiency of future processors.

Bus coding [7, 8, 9, 10, 11, 12, 27, 28, 29] is a well-known technique that can decrease data movement energy by encoding data value into a codeword before transmitting over a bus. If the codeword is formed to lower the number of voltage level transitions on wires, data movement energy can be improved due to the decreased switching activity. The most commonly used bus coding technique is data bus invert (DBI) coding [9, 15, 16, 17], which encodes a group of multiple data bits by pairing with a single extra control bit. If DBI coding is applied at 8-bit granularity of data, which is the most relevant case in practice [17, 18, 19], 8-bit data \( u(t) \) is converted to 9-bit codeword \( v \) that will be sent over a bus. As shown in Fig. 1, if the number of bit transitions from the previous bus value \( v(t-1) \) to the current data value \( u(t) \) is greater than four, the control bit is set to 1 and the data bits are inverted. Otherwise, the control bit is set to 0 and the data bits are encoded as equal to the original 8-bit data. With the assumption of uniformly random distribution for data bits, DBI coding was proven to be optimal in the sense that given one-extra-control-bit overhead no other bus coding can achieve the higher reduction in the number of bit transitions between the previous bus value and the current codeword [9].

In this letter, we identify that the conventional DBI coding can actually increase the switching activity when applied for a bus with OR chaining [20, 21, 22, 23, 24]. We also propose a new coding technique optimized for OR-chained buses. OR chaining has been considered as a hardware-efficient solution that can collect multiple modules without requiring multiplexers as shown in Fig. 2. Lack of multiplexers can reduce the amount of hardware and further provide an easier bus control solution because control signals for multiplexers are not present [20, 21, 23]. The OR-chained bus, however, requires data parking, i.e., driving all zero values, at the end of data transmission to facilitate the subsequent operation [20]. It is noteworthy that data parking can change the voltage level on wires, and thus, cause the additional switching activity. In the case of the encoding by conventional DBI, this extra switching activity is ignored, motivating a need for designing the bus encoder optimized for OR-chained bus. The key idea in this work is to encode...
data value based in part on a determination of whether a next bus value would transition from a codeword to a parked state. Based on the fact that the parked state is known a priori, the proposed technique jointly considers the previous bus value and the parked state for improving the total number of bit transitions occurred in the bus.

2. Conventional DBI over OR-chained bus

Consider data \( u(t) \) that is encoded to a codeword \( v(t) \), either \((u(t), 0)\) or \((u(t)', 1)\), before transmission. If a transmission of \( v(t) \) is followed by data parking, two different switching activities, termed \( \alpha \) and \( \beta \), can occur as shown in Fig. 3(a). The former, \( \alpha \), occurs due to a change between codeword values \( v(t-1) \) and \( v(t) \), while \( \beta \) occurs due to signal transitions by data parking. The conventional DBI is designed to pick \( v(t) \) that only minimizes \( \alpha \), hence, can actually increase the average switching activity if the resultant \( \beta \) is too high, which in turn results in energy consumption larger than that of the un-coded baseline. We define such case a defective encoding. An example of the defective encoding is illustrated in Fig. 3(b) and (c) where the DBI encoding of \( u(t) \) adversely increases the number of bit transitions from \( 7 (=6 \text{ bit toggles} + 1 \text{ bit toggle}) \) to \( 11 (=3 + 8) \).

In order to evaluate the probability of defective encodings by DBI, we have measured the number of defective encodings for a sequence of million random data transmitted in two bursts. The application of conventional 8-bit DBI scheme leads to 14.4% probability of defective encoding. The DBI coding at the different data granularity of 4, 12, and 16 bits also results in >10% probability of defective encoding as summarized in Table I.

| DBI data width (bits) | Probability of defective encoding (%) |
|-----------------------|--------------------------------------|
| 4                     | 11.7                                 |
| 8                     | 14.4                                 |
| 12                    | 15.9                                 |
| 16                    | 16.8                                 |

3. Proposed bus coding

In order to eliminate an occurrence of defective encoding over OR-chained bus, we propose a scheme that jointly considers the number of \( \alpha \)-associated bit transitions \( (n_\alpha) \) and the number of \( \beta \)-associated bit transitions \( (n_\beta) \). Note, although \( \beta \) occurs later than the encoding of \( u(t) \), \( n_\beta \) can be pre-computed since the value of parked state is known a priori. For example, in case of a codeword \( v(t) \) chosen to be \((u(t), 0)\), the corresponding \( n_\alpha \) and \( n_\beta \) can be computed as

\[
\begin{align*}
  n_\alpha &= w[v(t-1) \oplus (u(t), 0)], \\
  n_\beta &= w[u(t)],
\end{align*}
\]

where \( w[v] \) refers to a Hamming weight of a vector \( v \). The proposed scheme calculates the sum of \( n_\alpha \) and \( n_\beta \) if \( u(t) \) is the last sequential data followed by data parking (See Fig. 4 that illustrates the proposed coding of eight-bit data). If the sum of \( n_\alpha \) and \( n_\beta \) is greater than eight, \( u(t) \) is encoded in an inverted form, limiting the maximum total number of bit transitions by \( n_\alpha \) and \( n_\beta \) to a number of bus wires (= 9). Hence, this encoding never results in a defective encoding. The control bit for encoding can be generated as in Fig. 5.
4. Results

Fig. 6 shows the reduction in bit transitions achieved when the conventional and proposed bus coding schemes are applied for a sequence of random data transmitted in two bursts over OR-chained bus. First, it is observed that both schemes perform better for smaller data bit width. Hence, in order to lower the switching activity, it can be one approach to break the bus into multiple groups with each being encoded separately. Note, however, such approach leads to the higher wire overhead because each coding group needs one extra control bit. Second, in comparison with the conventional DBI, the proposed scheme achieves higher reduction of bit transitions due to its improved encoding algorithm. It is noteworthy that the proposed scheme at 8-bit granularity performs better even when compared to the DBI at 4-bit granularity. Hence, applying the proposed scheme may be the better approach to achieve the lower switching activity than partitioning an 8-bit bus into two 4-bit sub-buses with each being encoded by DBI.

To quantify the potential of the proposed scheme with respect to the total energy balance, we derive the break-even length \( L_{BE} \) at which the proposed scheme and the conventional DBI achieve the same energy efficiency. From the length longer than \( L_{BE} \) the proposed scheme obtains the lower total energy balance compared to the DBI. The value of \( L_{BE} \) can be estimated as

\[
L_{BE} = \frac{E_{\text{codec}_\text{pro}} - E_{\text{codec}_\text{dbi}}}{0.5 \cdot C_{\text{wire}} \cdot V_{dd}^2 (\alpha_{\text{dbi}} - \alpha_{\text{pro}})},
\]

where \( E_{\text{codec}_\text{pro}} \) (\( E_{\text{codec}_\text{dbi}} \)) is the per-bit energy dissipated by encoder and decoder of the proposed (conventional DBI) scheme, \( C_{\text{wire}} \) the wire capacitance per length in \( \text{mm} \) unit, \( V_{dd} \) the operating voltage, and \( \alpha_{\text{pro}} \) (\( \alpha_{\text{dbi}} \)) the average switching activity by the proposed (conventional DBI) scheme. Based on aforementioned results, the operating voltage of 0.89V, and the wire capacitance of 276.37 \( fF/\text{mm} \) from the predictive technology model in [25], \( L_{BE} \) is calculated as 0.59\( \text{mm} \). That is, the proposed scheme can achieve the lower total energy dissipation for >0.59\( \text{mm} \) bus length in comparison with the conventional DBI technique.

5. Conclusion

This letter discussed the bus coding technique suitable for a bus with OR chaining. We first identified that the conventional DBI coding can adversely increase the switching activity due to occurrences of defective encoding. We also proposed and implemented a new coding technique optimized for an OR-chained bus. Based on the fact that the parked state value is known a priori, the proposed scheme jointly considers the previous bus value and the parked state for an improved encoding, resulting in >3% additional saving of average switching activity. The simulation results show that the proposed scheme, even with its higher encoding energy, can obtain the lower total energy for >0.59\( \text{mm} \) bus length compared to the conventional DBI.

Acknowledgments

This work was supported by the Hongik University new faculty research support fund.

References
[1] Shekhar Borkar and Andrew A Chien: “The future of microprocessors,” Communications of the ACM, 54(5):67–77, 2011 (DOI: 10.1145/1941478.1941507).

[2] Bill Dally: “Power, programmability, and granularity: The challenges of exascale computing,” 2011 IEEE International Test Conference, pages 12–12. IEEE Computer Society, 2011 (DOI: 10.1109/ITP.2011.420).

[3] William J Dally: “Circuit challenges for future computing systems,” 2011 Symposium on VLSI Circuits-Digest of Technical Papers, pages 100–103. IEEE, 2011.

[4] Gokcen Kestor, Roberto Gioiosa, Darren J Kerbyson, and Adolfy Hoisie: “Quantifying the energy cost of data movement in scientific applications,” 2013 IEEE international symposium on workload characterization (IISWC), pages 56–65. IEEE, 2013 (DOI: 10.1109/IISWC.2013.6704670).

[5] Robert Lucas, et al.: “Doe advanced scientific computing advisory subcommittee (ascac) report: ten exascale research challenges,” Technical report, USDOE Office of Science (SC)(United States), 2014.

[6] Ron Ho, Kenneth W Mai, and Mark A Horowitz: “The future of wires,” Proceedings of the IEEE, 89(4):490–504, 2001 (DOI: 10.1109/5.920580).

[7] Sparsh Mittal and Subhrajit Nag: “A survey of encoding techniques for reducing data-movement energy,” Journal of Systems Architecture, 97:373–396, 2019 (DOI: 10.1016/j.sysarc.2018.11.001).

[8] Youngsoo Shin, Kiyoung Choi, and Young-Hoon Chang: “Narrow bus encoding for low-power dsp systems,” IEEE transactions on very large scale integration (VLSI) systems, 9(5):656–660, 2001 (DOI: 10.1109/92.953499).

[9] Mircea R Stan and Wayne P Burleson: “Bus-invert coding for low-power i/o,” IEEE Transactions on very large scale integration (VLSI) systems, 3(1):49–58, 1995 (DOI: 10.1109/92.365453).

[10] Mircea R Stan and Wayne P Burleson: “Low-power encodings for global communication in cmos vlsi,” IEEE transactions on very large scale integration (VLSI) systems, 5(4):444–455, 1997 (DOI: 10.1109/92.645071).

[11] Leiou Wang, Donghui Wang, and Chengpeng Hao: “Low power address bus encoding using loop prediction,” IEICE Electronics Express, 11(12):20140379–20140379, 2014 (DOI: 10.1587/elex.11.20140379).

[12] Shibbo Wang and Engin Ipek: “Reducing data movement energy via online data clustering and encoding,” 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), pages 1–13. IEEE, 2016 (DOI: 10.1109/MICRO.2016.7783735).

[13] Hadi Esmaeilzadeh, Emily Blem, Renee St Amant, Karthikeyan Sankaranigam, and Doug Burger: “Dark silicon and the end of multicore scaling,” 2011 38th Annual international symposium on computer architecture (ISCA), pages 365–376. IEEE, 2011 (DOI: 10.1145/2000064.2000108).

[14] V Huard, F Cacho, and X Federspiel: “Technology scaling and reliability challenges in the multicore era,” 2013 IEEE International Reliability Physics Symposium (IRPS), pages 3A–5. IEEE, 2013 (DOI: 10.1109/IRPS.2013.6531975).

[15] Youngsoo Shin, Soo-Ik Chae, and Kiyoung Choi: “Partial bus-invert coding for power optimization of application-specific systems,” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 9(2):377–383, 2001 (DOI: 10.1109/92.924059).

[16] Timothy M Hollis: “Data bus inversion in high-speed memory applications,” IEEE Transactions on Circuits and Systems II: Express Briefs, 56(4):300–304, 2009 (DOI: 10.1109/TCSII.2009.2015395).

[17] Seung-Jun Bae, et al.: “An 80 nm 4 gb/s/spin 32 bit 512 mb gddr4 graphics dram with low power and low noise data bus inversion,” IEEE journal of solid-state circuits, 43(1):121–131, 2008 (DOI: 10.1109/JSSC.2007.908002).

[18] Kyomin Sohn, et al.: “A 1.2 v 30 nm 3.2 gb/s/pin 4 gb ddr4 sram with dual-error detection and pvt-tolerant data-fetch scheme,” IEEE journal of solid-state circuits, 48(1):168–177, 2012 (DOI: 10.1109/JSSC.2012.6176868).

[19] D. D. R Micron. 2014. SDRAM, 4Gb: x4, x8, x16 DDR4 SDRAM features, white paper. Micron Technology, Inc (2014).

[20] William James Dally and Brian Patrick Towles: Principles and practices of interconnection networks (Elsevier, 2004).

[21] Sudeep Pasricha and Nikil Dutt: On-chip communication architectures: system on chip interconnect (Morgan Kaufmann, 2010).

[22] Nan Jiang, George Michelogiannakis, Daniel Becker, Brian Towles, and William J Dally: Booksim 2.0 user’s guide (Stanford University, 2010).

[23] Erno Salminen, Tero Kangas, Timo D Hämäläinen, Jouni Riikimäki, Vesa Lahтинen, and Kimmo Kuusilinna: “Hibi communication network for system-on-chip,” Journal of VLSI signal processing systems for signal, image and video technology, 43(2-3):185–205, 2006 (DOI: 10.1007/s11265-006-7270-6).

[24] Kon-Woo Kwon, Dinesh Somasekhar, and Sang Phill Park: US Patent App. 14/569,985 (2016).

[25] Predictive Technology Model. (2012) [Online]. Available: http://ptm.asu.edu

[26] Wei Zhao and Yu Cao: “Predictive Technology Model for Nano-CMOS Design Exploration,” ACM Journal on Emerging Technologies in Computing Systems (JETC), 3(1):1–es, 2007 (DOI: 10.1109/NANONET.2006.346227).

[27] Donghyuk Lee, Mike O’Connor, and Niladrich Chatterjee: “Reducing Data Transfer Energy by Exploiting Similarity within a Data Transaction,” 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA), pages 40–51. IEEE, 2018 (DOI: 10.1109/HPCA.2018.00014).

[28] Yuanwei Song and Engin Ipek: “More is less: Improving the energy efficiency of data movement via opportunistic use of sparse codes,” Proceedings of the 48th International Symposium on Microarchitecture, pages 242–254, 2015 (DOI: 10.1145/2830772.2830806).

[29] Somrita Ghosh, Prasun Ghosal, Nabanita Das, Saraju P Mohanty, and Oghenekaro Okoiah: “Data correlation aware serial encoding for low switching power on-chip communication,” 2014 IEEE Computer Society Annual Symposium on VLSI, pages 124–129. IEEE, 2014 (DOI: 10.1109/ISVLSI.2014.48).

[30] Dhinakaran Pandian and Carole-Jean Wu: “Quantifying the energy cost of data movement for emerging smart phone workloads on mobile platforms,” 2014 IEEE International Symposium on Workload Characterization (IISWC), pages 171–180. IEEE, 2014 (DOI: 10.1109/IISWCI.2014.6983056).

IEICE Electronics Express, Vol.VV, No.NN, 1–4