Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

Deep Sub-Micron Bus Invert Coding

Tina Lindkvist, Jacob Löfvenberg, Oscar Gustafsson

Dept. of EE, Linköpings Universitet, SE-581 83 Linköping, Sweden
Email: {tina, jacob, oscarg}@isy.liu.se

ABSTRACT

In this paper we present a simplified model for deep sub-micron, on-chip, parallel data buses. Using this model a coding technique similar to Bus Invert Coding is presented, but with a better performance in the proposed model. The coding technique can be realized using low-complexity encoding and decoding circuitry, and with a complexity that scales linearly with the bus width. Simulation results show that the energy dissipation decreases with approximately 20% for buses with up to 16 wires.

1. INTRODUCTION

1.1. Background

The continuing decrease in the minimum feature size in modern CMOS circuits and the corresponding increase in chip density and operating frequency have made power consumption a major concern in ULSI design. Chip area and throughput may no longer be primary system limiting factors except in very high-volume integrated circuits (tens of millions circuits per year) and in general-purpose computing.

1.2. Restrictions

In this paper our concern is not high-speed, but instead very low power. This means that the buses we consider are energy optimized, even if this means that we have to sacrifice throughput. As a result, problems with cross-talk and inductive couplings will not be discussed.

One important factor in power consumption, and one that needs addressing, is leakage. This problem is however not a topic in this paper. We will instead focus on the energy that is dissipated due to parasitic capacitances between nodes in the circuit.

1.3. Bus Model

In on-chip, parallel buses the energy dissipation stems from parasitic capacitances between wires, which we call inter-wire capacitances, and capacitances between wires and other metal layers (or the substrate) that have to be charged and discharged as the bus state changes.

In Fig. 1 a cross section of the metal layers in a 180nm process is shown, and different capacitances for metal layer 4 is shown. A more detailed figure would have shown also the capacitances between non-adjacent wires. For the sake of clarity we have chosen not to do so here and since they are much smaller than the capacitances between adjacent wires we will disregard them in this paper.

The capacitance C_d can be split into the parts C_{d3}, C_{d2}, C_{d1} and C_{dGND}, for the capacitances to the different layers below metal layer 4. In the same way C_u can be split into three different capacitances. Of these the capacitive coupling to the adjacent layers will be the greater.

We assume signals in different layers to be independent, implying that the energy dissipation due to such capacitive couplings depend only on the frequency of state changes on the bus wires under consideration. For every bus wire we can lump together all the capacitances to nodes in other layers and view them as a single capacitance, connecting the bus wire with a single node with non-changing charge. The value of this charge does not affect power dissipation, so we will assume it to be ground and call it the wire-to-ground capacitance.

The fringe capacitance, C_f, is in general greater than the wire-to-ground capacitance and less than the inter-wire capacitance, but in order to simplify the model C_f is often taken to be C_i or zero (see [4]).

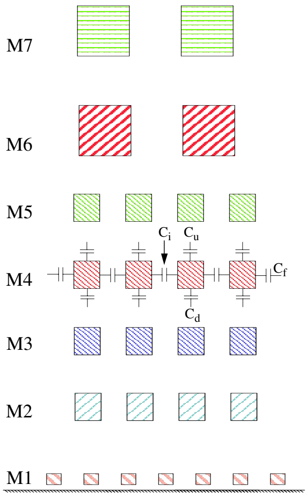

Fig. 1. A cross section of the metal layers in a 180nm process with seven metal layers.

The relation between the different capacitances is interesting. In older models the inter-wire and fringe capacitances were disregarded and only the wire-to-ground capacitances were taken into account. Such assumptions motivated the use of Gray codes for address bus coding, and Bus Invert coding for data buses. However, as processes shrink the ratio of the inter-wire capacitances to the wire-to-ground capacitances grows, and in modern processes the inter-wire capacitances (between adjacent wires) can no longer be disregarded. In Fig. 2 the inter-wire to wire-to-ground capacitance ratio is shown. The numbers for the figure is taken from Table 1 in [7], which in turn is based on [3].

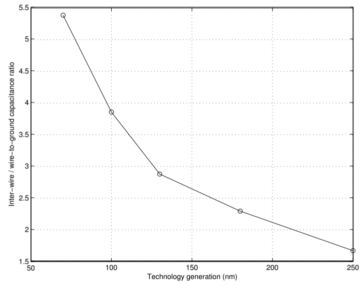

Fig. 2. Ratio between inter-wire and wire-to-ground capacitances for different technology generations.

As is seen in Fig. 2 the inter-wire capacitance is much greater than the wire-to-ground capacitance for modern processes, and the trend is that the ratio grows. If this trend continues the inter-wire capacitances will soon be dominating. This motivates us to use a simple model where the wire-to-ground capacitances are disregarded. For the sake of simplicity, we also ignore the fringe capacitance. For small processes this is a reasonable simplification, especially if the bus under consideration is wide enough that the energy dissipated due to the fringe capacitance is small compared to what is dissipated in the rest of the bus. These simplifications lead us to a model of a one layer parallel bus as shown in Fig. 3.

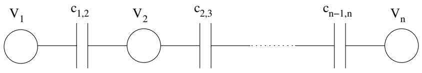

Fig. 3. Simplified model of a one layer, parallel bus. V_i are node potentials and $c_{i,j}$ are inter-wire capacitances.

We will later show that the coding system that is derived using this simplified model works well also in a more realistic setting, with non-neglected wire-to-ground capacitances.

Given an initial and a final state of the n-wire bus, with $V^i = (V_1^i, \dots V_n^i)$ and $V^f = (V_1^f, \dots V_n^f)$ representing the bus wire potentials, we can express the energy dissipated during the transition as

$$E(V^i, V^f) = (1/2)(V^f - V^i)C(V^f - V^i)^T, \quad (1)$$

where $C =$

$$\begin{bmatrix} c_{1,2} & -c_{1,2} & 0 & 0 & \dots 0 \\ -c_{1,2} & c_{1,2} + c_{2,3} & -c_{2,3} & 0 & \dots 0 \\ 0 & -c_{2,3} & c_{2,3} + c_{3,4} & -c_{2,3} & \dots 0 \\ & & & & \vdots \\ 0 & 0 & \dots 0 & -c_{n-1,n} & c_{n-1,n} \end{bmatrix}$$

is the capacitance conductance matrix [4], where we have set the ground capacitances $c_{i,i}$ to zero. At any time the bus wire potentials will define a vector of values. We will consider the system only when the wires have settled, so in our model the potentials will be either 0 or V_{dd}. We will represent these two states with the binary values 0 and 1.

As a result of the chosen model, the energy dissipated when the bus changes state can be expressed as the sum of the energies dissipated for each pair of adjacent wires during the state transition.

We assume all $c_{i,i+1}$ to be equal, and denote this capacitance $\tilde{C}$. In Table 1 the energy dissipation when switching two wires from one state to another is shown in multiples of $V_{dd}^2\tilde{C}/2$, what we call the normalized energy dissipation.

Table 1. Energy dissipation in multiples of $V_{dd}^2\tilde{C}/2$ for a pair of wires that changes state.

after/before	00	01	10	11
00	0	1	1	0
01	1	0	4	1
10	1	4	0	1
11	0	1	1	0

2. DEEP SUB-MICRON BUS INVERT CODING

We will realize the Bus Invert coding of Stan and Burleson [6], but with an energy measure adapted to the model of the deep sub-micron bus in Fig. 3. Both the proposed coding and the Bus Invert coding can be seen as special cases of the coding technique by Sotiriadis and Chandrakasan [5]. Another relevant reference is Zhang, Lach, Skadron and Stan [8], which is also a generalization of the Bus Invert coding, but not as general as that in [5]. Our contribution is a coding scheme that can be realized with low complexity, while at the same time being suited for modern, deep sub-micron buses.

The idea of Bus Invert coding is to send each data word either unchanged or inverted, together with a flag that is 0 for unchanged and 1 for inverted data (the INV flag). The data words are inverted if this yields a lower energy dissipation. In the original Bus Invert coding scheme the energy cost was taken as the Hamming distance between words. Fig. 4 below shows an overview of a Bus Invert encoder.

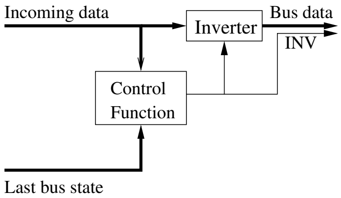

Fig. 4. Encoder overview.

The control function chooses to output either the unchanged, incoming data, together with a logical zero on the INV line, or the inverse of the incoming data, together with a logical one on the bus invert line. The construction of the control function is shown in Fig. 5 below.

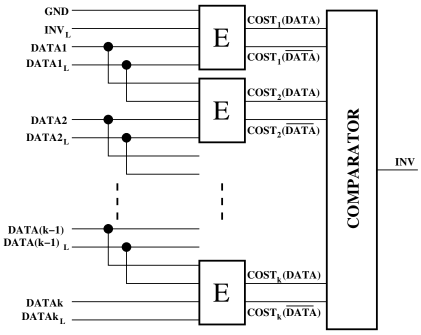

Fig. 5. Control function overview. GND is the value of the next, unraised, invert flag, $\mathrm{INV_L}$ is the invert flag of the last state, DATAi is the ith data bit of the next (uninverted) state and $\mathrm{DATAi_L}$ is the ith data bit of the last state.

This structure is similar to the one used in Bus Invert coding, and also more generally in [5]. We will however use another energy cost function E, as defined in Table 2 below.

Table 2. Pairwise energy cost function E. D is the data inputs.

D$(i-1)$	D$(i-1)_L$	Di	Di_L	Cost(data)	Cost($\overline{\text{data}}$)
1	0	0	1	1	0
0	1	1	0	1	0
1	0	1	0	0	1
0	1	0	1	0	1
everything else				0	0

The comparator compares the number of ones on its two sets of input lines, and outputs zero if there are more ones on the Cost(data) than on the Cost($\overline{\text{data}}$) lines, and one otherwise.

The energy comparison takes into account only the most expensive transitions, namely those from 01 to 10 and those from 10 to 01, that is, those with cost 4 in Table 1. However, when comparing the transmission costs for the unchanged and the inverted words, this comparison will always yield the same result as if the correct energy cost function had been used. The reason for this is, as can be seen by inspecting Table 1, that all other transition costs are unchanged by inversion of the next bus state, thereby not affecting the outcome of any energy cost comparison. It is this property that makes it possible to use the very simple energy cost function E in Table 2, an realization of which will be described in Section 3.1.

The construction in Fig. 5 is similar to that in [2], with the main difference being that in the latter a threshold comparison is done, that is, the energy cost is compared to a fix value. In our solution we compare the costs of the two alternatives, always choosing the cheaper and thereby yielding a better result. Such a comparison is also done by the construction in [8]. Compared to that solution our coding scheme dissipates more energy, but has coding and decoding circuits with a much lower complexity (the block corresponding to our E-block contains twelve full-adders, four half-adders, three 2-input gates and one multiplexer with four 2-bit inputs). The solution in [8] also needs two extra bus wires instead of one.

3. REALIZATION ISSUES

It is important that the overhead caused by the coding scheme does not cancel the gain obtained by the coding. Hence, a low power realization of the coding and decoding is required. The power savings due to the coding will depend on the bus length, while the coding circuitry is independent of bus length. Therefore, there will be a critical bus length, when the bus coding introduces a decrease in the total power consumption.

3.1. Energy Cost Function E

The energy cost function E can be realized using 2.5 XNOR gates(in average), two three-input NOR gates and one inverter. This solution is shown in Fig. 6 below. Note that the middle XNOR gate can be shared with one of the adjacent E blocks, so it suffices with one such gate per two E blocks.

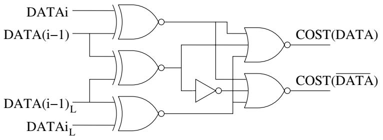

Fig. 6. Possible E function realization.

Another realization was made using the program Design Compiler from Synopsis together with the function description in Table 2, yielding a solution using eight standard cells and three inverters. All of the costs are taken per pair of adjacent wires of the bus.

3.2. Comparator

The function of comparing which of the inputs COST(DATA) and COST($\overline{\text{DATA}}$) has the most ones can be realized in a number of ways. In Fig. 7 a circuit based on fall time comparison is shown. The input to the fall time detector which has the most corresponding inputs set to one will discharge faster than the other one. The fall time detector can be realized using a sense amplifier-like circuit.

An alternative is to use an arithmetic realization with counters and a comparator. Furthermore, a delay based realization can be used [1]. Finally, in [6] an analog approach using only resistors and a voltage comparator was proposed for a very similar problem.

In determining which comparator to use the required speed and number of information bits must be taken into account. This is considered as future work. However, most coding approaches require one or more circuits of this type. Hence, our approach is comparable in this area.

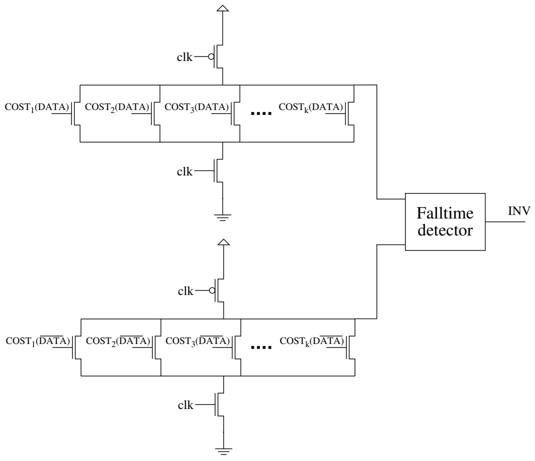
clk
COST$_1$(DATA)
COST$_2$(DATA)
COST$_3$(DATA)
....
COST$_k$(DATA)
clk
Falltime detector
INV
clk
COST$_1$($\overline{\text{DATA}}$)
COST$_2$($\overline{\text{DATA}}$)
COST$_3$($\overline{\text{DATA}}$)
....
COST$_k$($\overline{\text{DATA}}$)
clk

Fig. 7. An fall time based realization of a comparator.

It should be noted that analog constructions does not necessarily yield an accurate result every time. However, this does not result in corrupt data, since what happens is only that sometimes the data will be inverted when it should not be inverted (or vice versa), thereby resulting in a slightly increased power dissipation.

4. SIMULATION RESULTS

For reasons of comparison we have simulated the energy dissipation defined in section 1.3 using random data taken from a uniform and memoryless distribution. In Table 3 below is given the results in our model for uncoded data, and the relative gain for Bus Invert coding and Deep Sub-micron (DSM) Bus Invert coding. The energy cost is given in multiples of $V_{dd}^2\tilde{C}/2$ per information bit, and the reduction in energy dissipation is given compared to this. All simulations were done with one million pseudo-random information words.

Table 3. Energy dissipation comparison.

Info bits	Uncoded	Bus Inv	DSM Bus Inv
4	0.750	16.5%	22.9%
8	0.875	16.9%	22.7%
12	0.917	15.6%	20.4%
16	0.938	14.3%	18.7%

To verify that the DSM Bus Invert works well also for scenarios where the wire-to-ground capacitance is not zero, as we have assumed in our model, we have simulated the energy dissipation for non-zero wire-to-ground capacitances.

The next emerging technology is 65nm. Looking at Fig. 2 we find that for 65nm the ratio between inter-wire and wire-to-ground capacitances is approximately five, and we use this ratio in the simulations. For our simulations we assumed the fringe capacitance to be zero. The simulations where done with one million pseudo-random information words.

Table 4. Energy dissipation comparison.

Info bits	Uncoded	Bus Inv	DSM Bus Inv
4	0.850	17.3%	21.5%
8	0.975	17.0%	21.5%
12	1.017	15.6%	19.5%
16	1.038	14.4%	17.8%

As can be seen in the table the decrease in energy dissipation when wire-to-ground capacitance is used is similar to the decrease when the simplified model was used. We see this as an argument for the relevance of the simplified model.

5. CONCLUSION

By introducing redundancy it is possible to reduce the energy dissipation in on-chip, parallel data buses. We present a construction that introduces very little redundancy, while at the same time reducing the energy dissipation per information bit approximately 20% for small buses. The proposed coding technique requires little in terms of encoding circuitry, and it has a complexity that scales linearly with the bus width. Using simulations we have verified that the coding technique constructed based on the simplified model in Fig. 3 also works in more realistic settings.

REFERENCES

[1] M. Fujino and V. G. Moshnyaga, “Dynamic operand transformation for low-power multiplier-accumulator design,” *IEEE International Symposium on Circuits and Systems*, pp. 345–348, May 2003.

[2] K. Kim, K. Baek, N. Shanbhag, C. L. Liu, and S. Kang, “Coupling-driven signal encoding scheme for low-power interface design,” *IEEE/ACM International Conference on CAD*, pp. 318–321, November 2000.

[3] National Technology Roadmap for Semiconductors, Semiconductor Industry Association, 1997.

[4] P. P. Sotiriadis, *Interconnect modeling and optimization in deep sub-micron technologies*, Thesis (Massachusetts Institute of Technology), May 2002.

[5] P. P. Sotiriadis and A. Chandrakasan, “Low power bus coding techniques considering inter-wire capacitances,” *The 2000 IEEE Custom Integrated Circuits Conference*, pp. 507–510, May 2000.

[6] M. R. Stan and W. P. Burleson, “Bus-invert coding for low-power I/O”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 3 No. 1, pp. 49–58, March 1995.

[7] D. Sylvester, O. S. Nakagawa and C. Hu, “An analytical crosstalk model with application to ULSI interconnect scaling”, *SRC Technical Conference*, Las Vegas, NV, September 1998.

[8] Y. Zhang, J. Lach, K. Skadron, and M. R. Stan , “Odd/even bus invert with two-phase transfer for buses with coupling,” *International Symposium on Low Power Electronics and Design 2002*, pp. 80–83, August 2002.