

Analysis and Implementation of Charge Recycling for Deep Sub-micron Buses

Paul P. Sotiriadis

Department of EECS
Massachusetts Inst. of Technology
pps@mit.edu

Theodoros Konstantakopoulos

Department of EECS
Massachusetts Inst. of Technology
tkonsta@mit.edu

Anantha Chandrakasan

Department of EECS
Massachusetts Inst. of Technology
anantha@mit.edu

ABSTRACT : Charge recycling has been proposed as a strategy to reduce the power dissipation in data buses. Previous work in this area was based on simplified bus models that ignored the coupling between the lines. Here we propose a new Charge Recycling Technique (CRT) appropriate for sub-micron technologies. CRT is analyzed mathematically using a bus energy model that captures the energy loss due to strong line to line capacitive coupling. In theory CRT can result to energy reduction of a factor of 2. It becomes even more energy efficient when combined with Bus Invert coding (Stan '97, [6]). A circuit has been designed and simulated with all parasitic elements extracted from the layout. Taking into account the circuit energy overhead the net result in energy saving can be up to 32%.

1. INTRODUCTION

Over the past several years, significant emphasis has been placed on reducing the energy dissipation associated with on chip communication. Numerous schemes have been presented for reducing energy associated with driving wires including low swing signaling [1],[2],[3], charge re-cycling [4],[5] and data coding [6],[7],[8]. In this paper we introduce a new practical Charge Recycling Technique (CRT) appropriate for sub-micron technology buses. Its performance is verified by both mathematical analysis and circuit implementation.

The technique, based on charge recycling between the lines, consists of two steps. During the first step, charge redistribution takes place between the lines whose logical values are changing during the transition. All other lines remain connected to their drivers. During the second step, all lines are driven to the voltages corresponding to their new logical values. A similar technique was presented in [4] and [5] but for the case where there is no coupling between the lines. This difference is very essential. In sub-micron technology the strong capacitive coupling between the lines must be taken into account since it dramatically changes the energy consumption during bus transitions (with or without the CRT). For this purpose we use a sub-micron bus energy equivalent model presented in [7].

In this paper we also present a driver to implement CRT. Its operation is demonstrated on a 4-line and an 8-line bus using HSPICE. The driver works at 100MHz and results to a net energy saving of up to 32%. The circuit is directly expandable to larger buses. Large buses of 32 or 64 lines and FPGA interconnect networks are expected to have higher net energy savings with CRT (Rabaey [11]). Standard .18 μ CMOS technology has been used.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISLPED'01, August 6-7, 2001, Huntington Beach, California, USA.

Copyright 2001 ACM 1-58113-371-5/01/0008...\$5.00.

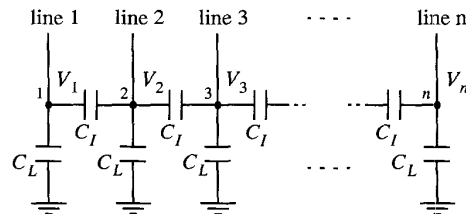


Figure 1: Sub-micron energy-equivalent Bus Model

2. SUB-MICRON BUS MODEL

The sub-micron bus energy model we use to evaluate the different energies is shown in Figure 1. It has been proven in [7] that this model has identical energy behavior to its distributed version. C_L is the capacitance between each line and the ground and C_i is the capacitance between adjacent lines. (The capacitance between non-adjacent lines is very weak and can be ignored). We define the technology dependent parameter $\lambda = C_i/C_L$. For .18 μ technology, λ is about 5. Also, λ tends to increase with technology scaling.

To simplify the theoretical analysis we set $V_{dd} = 1$. Then all energies calculated under this assumption must be multiplied by the factor V_{dd}^2 to give the real energy value.

3. BUS DRIVER MODEL

The drivers are modeled as in Figure 2, [9]. The resistors $R_i^P(t)$ and $R_i^N(t)$ correspond to the PMOS and NMOS transistors of the drivers. Their values can be almost arbitrary functions of time. The switches s_i and \bar{s}_i are complementary and their status corresponds to the desirable values of the lines. The parasitic capacitances of the drivers outputs can be lumped into C_L .

4. CHARGE RECYCLING

In this section we present the two steps of CRT. Suppose the bus has n lines and let T be the clock cycle period. New data is trans-

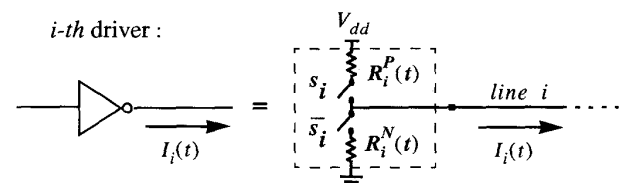


Figure 2: Bus Driver Model

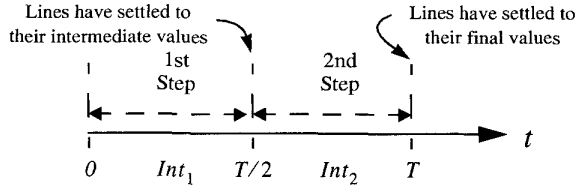


Figure 3: Timing of CRT

mitted through the bus every T seconds. The time interval $[0, T]$ is divided into the subintervals Int_1 and Int_2 . The two steps of CRT are timely related as in Figure 3.

Suppose that during the clock cycle the bus transitions from its current values, $x = [x_1, x_2, \dots, x_n]^T$, to its new values, $y = [y_1, y_2, \dots, y_n]^T$ (x_i, y_i correspond to line i). Normalizing by $V_{dd} = 1$, then all x_i and y_i belong to $\{0, 1\}$. The voltages of the lines as functions of time $t \in [0, T]$ are denoted by $V = [V_1, V_2, \dots, V_n]^T$. At $t = 0$ and $t = T$ it is $V(0) = x$ and $V(T) = y$ respectively. The CRT is presented in Figure 4 with the modified driving circuit. We agree that switch w_i has value 0 if node i is connected to the output of driver i , and value 1 if node i is connected to the common node q . During Int_1 the lines that change logical values (during the transition $x \rightarrow y$) are connected to node q and not to their drivers. The lines retaining their logical values remain connected to their drivers. (This is a major difference to the strategy in [5]. If there is coupling between the lines it makes a difference if the non-changing lines remain connected to their drivers or not during the charge redistribution). During Int_2 all lines are connected to their drivers.

4.1 First Step (Int 1)

For every line $i = 1, \dots, n$ we set $d_i = x_i \oplus y_i$. We also use the vector $d = x \oplus y$ where $d = [d_1, d_2, \dots, d_n]^T$ and the diagonal

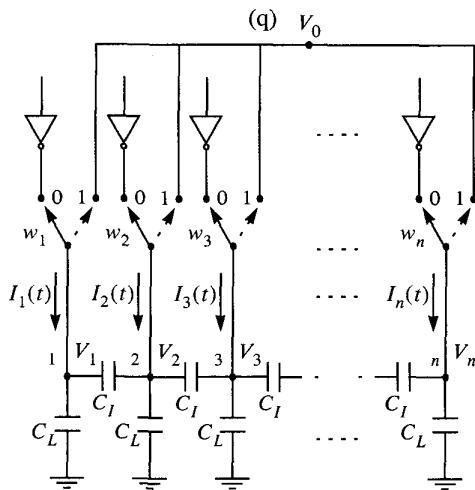


Figure 4: CRT - Network connections

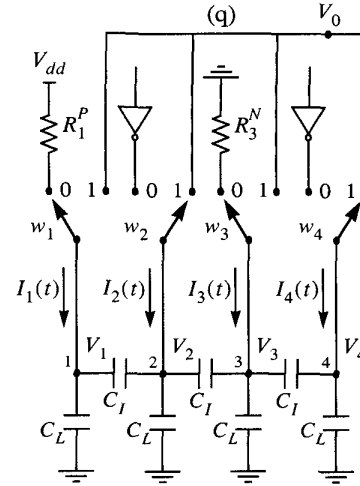


Figure 5: Example ($n=4$)

matrix,

$$D = \text{diag}(d_1, d_2, \dots, d_n) \quad (1)$$

During the transition, line i changes value if and only if $d_i = 1$. According to CRT, during the time interval $Int_1 = (0, T/2]$ the lines with changing values are connected to node q and therefore the i -th switch must have the value $w_i = d_i = x_i \oplus y_i$. The network is configured respectively. For example let $n = 4$, $x = [1, 0, 0, 1]^T$ and $y = [1, 1, 0, 0]^T$. Then $d = [0, 1, 0, 1]^T$ and during Int_1 the network is configured as in Figure 5.

During Int_1 the dynamics of the network satisfies the set of differential equations (see Figure 4),

$$\begin{aligned} I_1 &= C_L \cdot \dot{V}_1 + C_I \cdot (\dot{V}_1 - \dot{V}_2) \\ I_k &= C_I \cdot (\dot{V}_k - \dot{V}_{k-1}) + C_L \cdot \dot{V}_1 + C_I \cdot (\dot{V}_k - \dot{V}_{k+1}), \quad 1 < k < n \\ I_n &= C_L \cdot \dot{V}_n + C_I \cdot (\dot{V}_n - \dot{V}_{n-1}) \end{aligned} \quad (2)$$

If we define the currents vector $I = [I_1, I_2, \dots, I_n]^T$ and the $n \times n$ capacitance conductance matrix [7] of the network in Figure 4,

$$C_T = \begin{bmatrix} 1 + \lambda & -\lambda & 0 & \dots & 0 \\ -\lambda & 1 + 2\lambda & -\lambda & & 0 \\ 0 & -\lambda & & & \\ \vdots & \vdots & \vdots & 1 + 2\lambda & -\lambda \\ 0 & 0 & \dots & -\lambda & 1 + \lambda \end{bmatrix} \cdot C_L \quad (3)$$

then, equations (2) can be written in the compact form,

$$C_T \cdot \dot{V} = I \quad (4)$$

Now note that for $i = 1, \dots, n$ the current $I_i(t)$ is drawn from the driver if $d_i = 0$ and from the node q if $d_i = 1$. The charge con-

servation at node q implies that $\sum_{i: d_i=1} I_i(t) = 0$, or in vector form,

$$d^T \cdot I = 0 \quad (5)$$

Replacing (4) into (5) we get $d^T \cdot C_T \cdot \dot{V} = 0$ which integrated over the time interval $Int_1 = (0, T/2]$ of step 1 gives,

$$d^T \cdot C_T \cdot (V(\frac{T}{2}) - V(0)) = 0 \quad (6)$$

$V(0) = x = [x_1, x_2, \dots, x_n]^T$ are the initial conditions of the lines and $V(\frac{T}{2}) = [V_1(\frac{T}{2}), V_2(\frac{T}{2}), \dots, V_n(\frac{T}{2})]^T$ are the intermediate ones. Here we assume that the time length $T/2$ is sufficient for the voltages of the network to settle. This assumption is reasonable for the current technology and is always used in charge redistribution (and adiabatic) techniques. So for $i = 1, \dots, n$ the voltage $V_i(\frac{T}{2})$ is either x_i if $d_i = 0$ or $z \equiv V_q(\frac{T}{2})$ if $d_i = 1$. The value z is of course the same for all lines that change logical value. Algebraically we have that $V_i(\frac{T}{2}) = (1 - d_i) \cdot x_i + d_i \cdot z$ or in vector form that,

$$V(\frac{T}{2}) = (I - D) \cdot x + z \cdot d \quad (7)$$

where I is the $n \times n$ identity matrix. The matrix D and the vector d are as defined before. From (6), (7) and $V(0) = x$ we have,

$$(d^T \cdot C_T \cdot d) \cdot z = d^T \cdot C_T \cdot D \cdot x \quad (8)$$

Now note that matrix C_T is positive definite. This implies that the quantity $(d^T \cdot C_T \cdot d)$ is positive if and only if $(d \neq 0)$ or equivalently, if and only if at least one line change value during the transition $x \rightarrow y$. If there is no change during the transition then the energy dissipated during the transition is zero. For now we assume that at least one line changes value. Then from (8) we get,

$$z = V_0(\frac{T}{2}) = \frac{d^T \cdot C_T \cdot D \cdot x}{d^T \cdot C_T \cdot d} \quad (9)$$

4.2 Energy Dissipation on Step 1

Here we evaluate the energy that is **drawn from** V_{dd} on the first step of CRT. The current $I_{V_{dd}}(t)$ drawn from V_{dd} during (Int_1) is the sum of the currents drawn by the lines that do *not* change logical value during the transition *and* remain connected to V_{dd} through their drivers, i.e the lines $i = 1, \dots, n$ for which $x_i = y_i = 1$. So,

$I_{V_{dd}}(t) = \sum_{i: x_i=1} I_i(t) = \sum_{i=1}^n x_i \cdot (1 - d_i) \cdot I_i(t)$ which can be written in matrix form as,

$$I_{V_{dd}}(t) = x^T \cdot (I - D) \cdot I(t) \quad (10)$$

(Symbol I is used for both the current vector and the identity matrix. It should be clear what I represents each time). Using equation (4) and (10) we have that,

$$I_{V_{dd}}(t) = x^T \cdot (I - D) \cdot C_T \cdot \dot{V} \quad (11)$$

Because of the normalization $V_{dd} = 1$ the energy drawn from V_{dd} during step 1 is $E_1 = \int_0^{T/2} I_{V_{dd}}(t) dt$. By replacing (11) in the

integral we have $E_1 = x^T \cdot (I - D) \cdot C_T \cdot (V(\frac{T}{2}) - V(0))$. Finally we use (7) to get,

$$E_1 = x^T \cdot (I - D) \cdot C_T \cdot (z \cdot d - D \cdot x) \quad (12)$$

And by replacing z from (9) into (12) we have,

$$E_1(x, y) = x^T \cdot (I - D) \cdot C_T \cdot \left\{ \left(\frac{d^T \cdot C_T \cdot D \cdot x}{d^T \cdot C_T \cdot d} \right) \cdot d - D \cdot x \right\} \quad (13)$$

if $d \neq 0$ and $E_1 = 0$ if $d = 0$.

4.3 Second Step (Int 2)

During the second step of the CRT, the time interval $Int_2 = (T/2, T]$, every line is connected to its driver (with the new value y_i). So for all $i = 1, \dots, n$ it is $w_i = 0$. For the example with $n = 4$, $x = [1, 0, 0, 1]^T$ and $y = [1, 1, 0, 0]^T$, the network is configured as in Figure 6. Equation (4) holds during the second step as well.

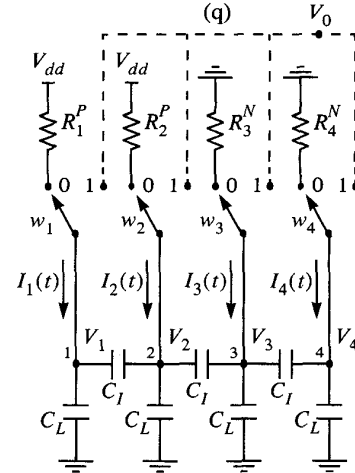


Figure 6: Step 2: Example (n=4)

4.4 Energy Dissipation on Step 2

During Int_2 the current drawn from V_{dd} equals the sum of the currents $I_i(t)$ of the lines connected to V_{dd} (through their drivers). So $I_{V_{dd}}(t) = \sum_{i: y_i=1} I_i(t) = \sum_{i=1}^n y_i \cdot I_i(t)$ or in vector form,

$$I_{V_{dd}}(t) = y^T \cdot I(t) \quad (14)$$

Replacing (4) into (14) we get,

$$I_{V_{dd}}(t) = y^T \cdot C_T \cdot V \quad (15)$$

The energy drawn from V_{dd} on step 2 is $E_2 = \int_{T/2}^T I_{V_{dd}}(t) dt$.

Replacement of (15) into the integral gives,

$$E_2 = y^T \cdot C_T \cdot (V(T) - V(\frac{T}{2})) \quad (16)$$

Finally, $V(T) = y$, (7) and (9) imply,

$$E_2(x, y) = y^T \cdot C_T \cdot \left\{ y - (I - D) \cdot x - \left(\frac{d^T \cdot C_T \cdot D \cdot x}{d^T \cdot C_T \cdot d} \right) \cdot d \right\} \quad (17)$$

5. ENERGY PROPERTIES OF CRT

The total energy $E(x, y)$ drawn from V_{dd} during the transition $x \rightarrow y$ is of course $E(x, y) = E_1(x, y) + E_2(x, y)$. Using the identity $(I - D) \cdot x = (I - D) \cdot y$ and expressions (13) and (17) we get,

$$E(x, y) = y^T \cdot C_T \cdot (y - x) + y^T \cdot D \cdot C_T \cdot (D \cdot x - z \cdot d) \quad (18)$$

where z is given by (9). The first term of the right part of (18) equals the energy drawn from V_{dd} by the bus during the transition $x \rightarrow y$ when no charge recycling is applied [7]. The other terms correspond to the energy difference (savings) due to CRT.

For a better intuition on how CRT influences the bus energy transition patterns, table 1 presents the case of a three line bus $n = 3$ when $\lambda = 5$. Five is a representative value of λ for the case of .18 μ technologies (with minimal distance between the wires). For simplicity we set $C_L = V_{dd} = 1$.

(x_1, x_2, x_3)	(y_1, y_2, y_3)							
	000	001	010	011	100	101	110	111
000	0	6	11	7	6	12	7	3
001	0	0	10	6	3	6	6.3	2
010	0	5.1	0	1	5.1	11	1	2
011	0	5	5	0	5.3	10	3	1
100	0	3	10	6.3	0	6	6	2
101	0	0	10	5.1	0	0	5.1	1
110	0	5.3	5	3	5	10	0	1
111	0	5	10	5	5	10	5	0

Table 1: Transition energies with and without the CRT

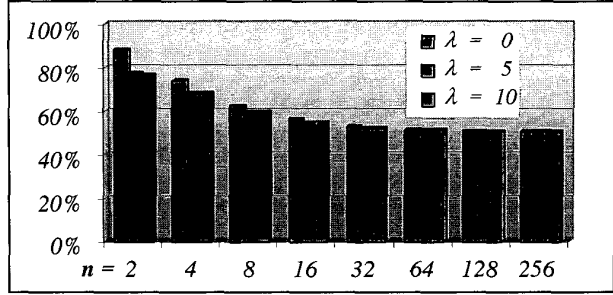


Figure 7: Energy with CRT / Energy without CRT

For each transition $(x_1, x_2, x_3) \rightarrow (y_1, y_2, y_3)$ the shadowed value (below) is the energy cost *without* CRT, equal to $y^T \cdot C_T \cdot (y - x)$. The numbers on the white background (above) are the energies *with* CRT, i.e. to the values given by (18). The energy *with* CRT is always smaller. Also, the highest percentage of energy reduction occurs in the most expensive transitions 010 \rightarrow 101 and 101 \rightarrow 010 where adjacent lines transit in the opposite direction and the interline capacitances are charged by $2 \times V_{dd}$.

6. ENERGY REDUCTION

The result for the transition energy, formula (18), allows us to estimate numerically the expected energy drawn by the bus when the CRT is used. We do this for the case of uniformly distributed i.i.d. data. In Figure 7 we see the expected energy using CRT as a percentage of the expected energy without CRT for the cases of $n = 2, 4, 8, 16, 32, 64, 128, 256$ and $\lambda = 0, 5, 10$. The figure suggests that for the number of lines $n = 32, 64, 128, 256$ the energy drawn from V_{dd} can be reduced to one half using CRT. Also, the results are independent of the capacitance to ground C_L and they slightly improve when λ increases. In general λ tends to increase with technology scaling.

7. CRT AND BUS-INVERT

In the previous sections we showed how CRT reduces energy consumption. In Figure 8 we present an architecture where CRT is combined with Bus-Invert coding [6].

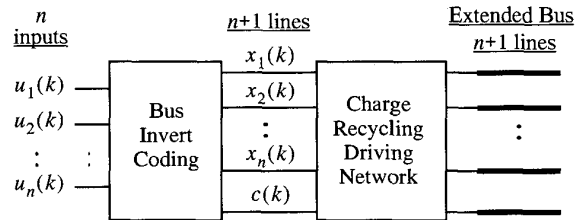


Figure 8: Combination of CRT with Bus Invert coding

The Bus Invert coding works in the following way. Let $u(k) = [u_1(k), u_2(k), \dots, u_n(k)]^T$ be the new input vector and $x(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$ be the new vector of the values of the

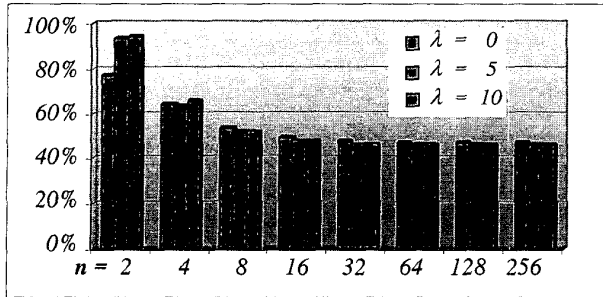


Figure 9: Energy with CRT and Bus Invert / Energy without them

lines. If the vector $u(k) \oplus x(k-1)$ contains more than $n/2$ ones then we set $x(k) = \overline{u(k)}$ and $c(k) = 1$, otherwise we set $x(k) = u(k)$ and $c(k) = 0$. The combined performance of CRT and Bus Invert is shown in Figure 9. We see a small improvement compared to the results of Figure 7. For buses with 16 lines or more the energy saving is more than 50%.

8. A CIRCUIT FOR CRT DRIVERS

To verify CRT we designed a circuit that implements the conceptual network of Figure 4. Our circuit implementation consisted of the bus and the CRT drivers of the lines. The CRT driver detects the transition of the line and connects it either to the common node (q) or to its regular driver (chain of inverters). The proposed CRT driver was designed and laid out in $.18\mu$ technology and its schematic is shown in Figure 10. Using this driver we tested CRT for a 4-line and an 8-line bus. The layout of both the CRT and the standard drivers for the two cases are shown in Figure 11. The CRT driver operates as follows. The switches w_1, w_2, \dots in Figure 4 are realized here by the pair of transmission gates.

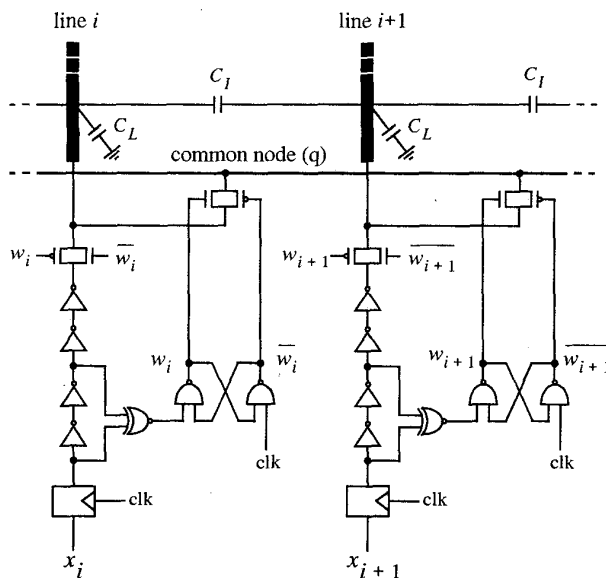


Figure 10: Efficient CRT - Driver

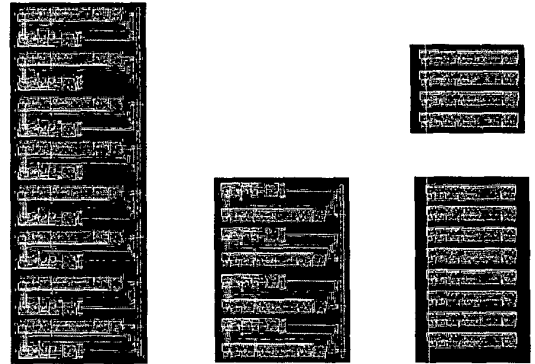


Figure 11: Layout of the CRT drivers

The charge recycling phase begins when CLK becomes 1. A negative spike appears at the output of the XNOR gate if the input x_i changes value. This sets the latch and connects the line to the common node q through the transmission gate. The charge recycling phase ends when CLK becomes 0. This resets the latch, isolates the line from the common node (q) and connects it to the buffer chain. If the input x_i does not make a transition, the latch remains reset during the whole clock cycle and the line remains connected to the buffer chain.

The same circuit can be used unchanged for buses with arbitrary number of lines.

9. SIMULATION AND RESULTS

CRT drivers of Figure 10 were used to drive the lines of a four and an eight line bus, $n = 4, n = 8$. A netlist was extracted from the layout of the drivers for the simulation with HSPICE. The lines were modeled as in Figure 1 and for the capacitor C_L we used the values $50fF, 100fF, 150fF$ and $200fF$. Note that these values could represent not only the line capacitors but all the loads as well. This is particularly the case of reconfigurable interconnect networks (e.g. in FPGAs) where long buses are loaded by the parasitic capacitances of several mosfets resulting to total capacitive loads of the size of a few picofarads [11]. The clock frequency in the simulations was set to $100MHz$ and the buses were fed with uniformly distributed i.i.d. sequences of data. In

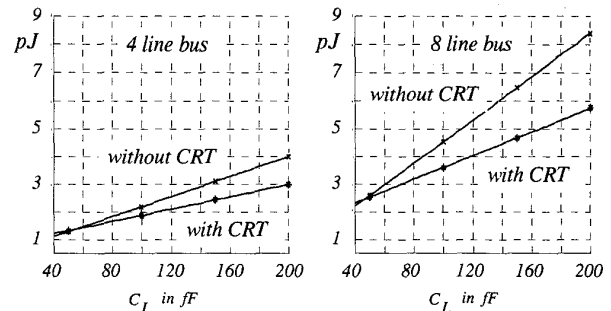


Figure 12: Average energy per cycle of a 4 and 8 line buses with and without CRT

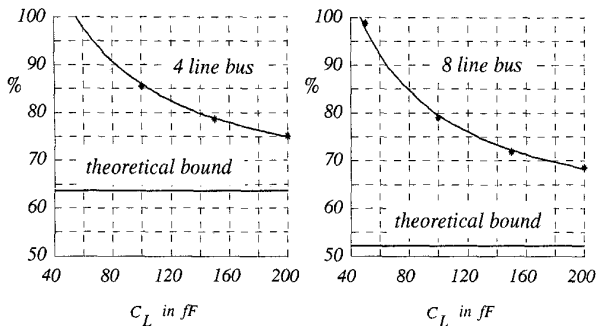


Figure 13: Normalized energy using CRT (HSPICE simulation)

Figure 12 we see the average energy per cycle of the four line bus (left) and the eight line bus (right). The curves in the graphs showing higher energy consumption correspond to the standard buses. The curves showing lower consumption correspond to buses with CRT drivers. In Figure 13 we see the average energy using CRT as a percentage of the average energy without CRT for the 4-line and 8-line buses. Again, the ratios are parametrized to C_L . The flat lines correspond to the minimum possible ratios resulting from the theoretical analysis and shown in Figure 7.

As it should be expected, for higher capacitive loads we get higher percentages of energy saving. This is because the average energy per cycle of the additional circuitry of the drivers is relatively independent of the loads. For larger loads this additional energy becomes less significant.

Finally, it is interesting to look at the waveforms of the individual lines during the two steps of the CRT. Figure 14 shows the waveforms of the line voltages of the 4-line bus. In this particular case, one line experiences a $1 \rightarrow 0$ transition and the rest three lines make a $0 \rightarrow 1$ transition. Since all lines transit they are all connected first to the common node q . The final voltage at node q during the charge redistribution period is of course z (equation (9)) and correspond to the converging point of the waveforms at time $T/2 = 5ns$.

It is interesting to note that for an individual transition the maximum energy saving with CRT occurs when all lines transition and

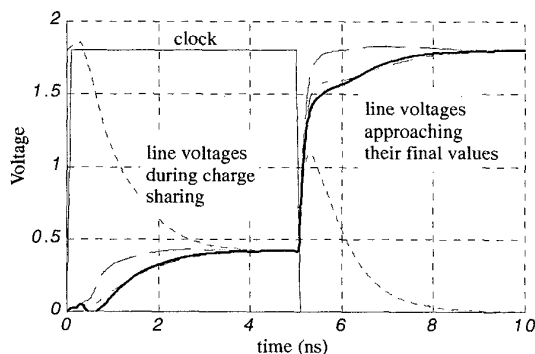


Figure 14: Line voltage waveforms during the two steps of CRT

adjacent lines transition in opposite directions. This generalizes to the fact that CRT performs very well when the sequences of the transitions of adjacent lines are negatively correlated.

10. CONCLUSIONS

A Charge Recycling Technique (CRT) for sub-micron buses has been proposed and analyzed. Closed form results for the transition energy have been given and used for the theoretical evaluation of the energy reduction with CRT. Reduction of the average transition energy by a factor of more than 2 can result in theory by the application of both CRT and Bus Invert coding. A line driver has been designed to implement CRT. Using it in an 8-line bus we have demonstrated net energy savings of up to 32%. Larger buses of 32 or 64 lines are expected to have higher energy savings.

Acknowledgements

The authors acknowledge support from the MARCO Focus Research Center on Interconnect which is funded at the Massachusetts Institute of Technology, through a subcontract from the Georgia Institute of Technology. Paul Sotiriadis is partially supported by the Alexander S. Onassis Public Benefit Foundation, the Greek Section of Scholarships and Research.

References

- [1] H. Zhang, J. Rabaey, "Low swing interconnect interface circuits," *IEEE/ACM International Symposium on Low Power Electronics and Design*, pp. 161-166, August 1998.
- [2] Y. Nakagome, K. Itoh, M. Isoda, K. Takeuchi, M. Aoki, "Sub-1-V swing internal bus architecture for future low-power ULSI's," *IEEE Journal of Solid-State Circuits*, pp. 414-419, April 1993.
- [3] H. Yamauchi, H. Akamatsu, T. Fujita, "An asymptotically zero power charge-recycling bus architecture for battery-operated ultrahigh data rate ULSI's," *Journal of Solid-State Circuits, IEEE, Vol. 30 Issue: 4*, pp. 423-431, April 1995.
- [4] K.Y. Khoo, A. Willson, Jr., "Charge recovery on a databus," *IEEE/ACM International Symposium on Low Power Electronics and Design*, pp. 185-189, 1995.
- [5] B. Bishop, M. J. Irwin, "Databus charge recovery: Practical consideration," *International Symposium on Low Power Electronics and Design*, pp. 85-87, August 1999.
- [6] M. Stan, W. Burleson, "Low-power encodings for global communication in cmos VLSI," *IEEE Transactions on VLSI Systems*, pp. 49-58, Vol. 5, No. 4, Dec. 1997.
- [7] P. Sotiriadis, A. Wang, A. Chandrakasan, "Transition Pattern Coding: An approach to reduce Energy in Interconnect", *ESSCIRC 2000*.
- [8] S. Ramprasad, N. Shanbhag, I. Hajj, "A coding framework for low-power address and data busses," *IEEE Transactions on VLSI Systems*, pp. 212-221, Vol. 7, No. 2, June 1999.
- [9] J.M. Rabaey, *Digital Integrated circuits*. Prentice Hall 1996.
- [10] H.B. Bakoglou, *Circuits, Interconnections, and Packaging in VLSI*. Addison-Wesley Pub. Co., 1990
- [11] E. Kusse, J. Rabaey, "Low-Energy Embedded FPGA Structures", *ISLPED '98*. August '98 Monterey USA.