Transactions Briefs

Minimal-Power, Delay-Balanced SMART Repeaters for Global Interconnects in the Nanometer Regime

Roshan Weerasekera, Dinesh Pamunuwa, Li-Rong Zheng, and Hannu Tenhunen

Abstract—A SMART repeater is proposed for driving capacitively-coupled, global-length on-chip interconnects that alters its drive strength dynamically to match the relative bit pattern on the wires and thus the effective capacitive load. This is achieved by partitioning the driver into main and assistant drivers; for a higher effective load capacitance both drivers switch, while for a lower effective capacitance the assistant driver is quiet. In a UMC 0.18-μm technology the potential energy saving is around 10% and the reduction in jitter 20%, in comparison to a traditional repeater for typical global wire lengths. It is also shown that the average energy saving for nanometer technologies is in the range of 20% to 25%. The driver architecture exploits the fact that as feature sizes decrease, the capacitive load per transistor shrinks, whereas global wire loads remain relatively unchanged. Hence, the smaller the technology, the greater the potential saving.

Index Terms—Buffer, interconnects, nanometer design, on-chip signaling, repeaters.

I. INTRODUCTION

A key technique in reducing propagation delay and signal degradation in global on-chip interconnects is repeater insertion. Although very effective and simple, this has an adverse effect on power consumption and it has been estimated that over 50% of the power in a high performance microprocessor is dissipated by repeaters charging and discharging interconnects [1]. Further, [1] concludes that over 90% of this power is concentrated in only 10% of the interconnets; i.e., those which are classed as global and run for a significant fraction of the die length.

The repeater we propose in this paper exploits the fact that in a parallel wire structure, the effective capacitance of a given wire is dynamic; i.e., it is a function not only of the physical geometry, but also of the relative switching pattern described by the bits on the wire in question (the victim) and the adjacent wires (aggressors). With a traditional repeater, since the drive strength is static, the result is a spread of the propagation delay, with the repeater strength being essentially too much for every bit pattern *other* than the worst-case pattern. In the proposed repeater, the drive strength is dynamically altered depending on the relative bit pattern, by partitioning it into a *main driver* and *assistant driver* (see Fig. 1). For a higher effective load capacitance, both drivers switch, while for a lower effective capacitance the assistant driver is quiet [2]. By disconnecting part of the repeater when it is not needed, the total load capacitance to the previous stage is reduced, resulting in reduced energy consumption for those instances. It is experimentally
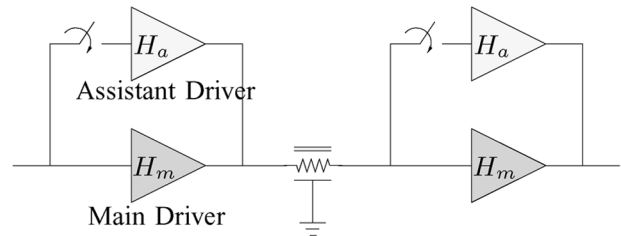
Fig. 1. Basic schematic of the proposed driver scheme.

shown that for a UMC 0.18-μm technology the potential average saving in energy can be as much as 10% over a traditional repeater for typical global wire lengths in nanometer technologies. Since this SMART repeater works by reducing the variation in the delay, an added benefit is that the jitter is reduced. In the same technology, the jitter reduction was as much as 20%.

The ramifications of the dynamically changing load in coupled interconnects have received a fair amount of attention in the literature. A comprehensive analysis of design considerations for repeater insertion in a bus structure with heavy coupling was presented in [3]. A scheme proposed in [4] staggers the repeaters so that opposing transitions only persist for the length of the offset between repeaters, and become best-case patterns for the remainder, resulting in a delay reduction. Many innovative alternatives to the traditional repeater have also been proposed, such as the transient sensitive accelerator (TSA) [5], charge recycling technique (CRT) [6], boosters [7], the TAGS receiver [8], the aggressor-aware repeater [9], and the capacitor-coupled trigger and accelerator combination [10]. Some of these use skewed inverters to tradeoff noise margin for speed [5], [7], [8], while others consume more energy [9] and occupy a larger area [5], [7], [8] to produce a faster response.

In general, not only do these alternatives to traditional repeaters require much effort in circuit design similar to library cell design, but they also lack a clear high-level abstraction; in contrast, performance metrics such as delay and energy consumption can easily be quantified in terms of a few critical design parameters for the traditional inverting repeater [11], resulting in easy amalgamation in computer-aided design (CAD) flows at different levels of hierarchy from initial signal planning to detailed place and route.

A secondary advantage of the repeater circuit proposed here is that the relatively minor increase in circuit complexity required to obtain the energy saving and delay equalization described above can be completely abstracted in the performance analysis.

II. IMPLEMENTATION OF THE SMART DRIVER

A. Concept

In order to demonstrate the variation of effective capacitance of wires, a pair of coupled lines is used as a constituent unit for a bus. For two simultaneously switching lines, sixteen possible switching combinations can be identified as given in Table I. These can be categorized into five different groups according to the effective capacitance as follows.

Group 1   Both switch in the same direction.
Group 2   Both lines are quiet (at 0 or 1).

TABLE I
SWITCHING ACTIVITIES ON THE LINES AND THE VARIATION
OF EFFECTIVE CAPACITANCE

| Group | Case | Switching Event On | | Effective Wire Capacitance | |
		wire i	wire j	Traditional driver	Smart driver
1	1	↓	↓	C_{wt}	C_{ws}
	2	↑	↑	C_{wt}	C_{ws}
2	3	0	0	0	0
	4	0	1	0	0
	5	1	0	0	0
	6	1	1	0	0
3	7	0	↑	0	0
	8	↑	0	$C_{wt} + \frac{0.9C_c}{k}$	$C_{ws} + \frac{0.9C_c}{k}$
	9	0	↓	0	0
	10	↓	0	$C_{wt} + \frac{0.9C_c}{k}$	$C_{wt} + \frac{0.9C_c}{k}$
4	11	1	↑	0	0
	12	↑	1	$C_{wt} + \frac{0.9C_c}{k}$	$C_{wt} + \frac{0.9C_c}{k}$
	13	1	↓	0	0
	14	↓	1	$C_{wt} + \frac{0.9C_c}{k}$	$C_{ws} + \frac{0.9C_c}{k}$
5	15	↑	↓	$C_{wt} + \frac{1.7C_c}{k}$	$C_{wt} + \frac{1.7C_c}{k}$
	16	↓	↑	$C_{wt} + \frac{1.7C_c}{k}$	$C_{wt} + \frac{1.7C_c}{k}$

Group 3 One line is switching while the other is quiet at 0.
Group 4 One line is switching while the other is quiet at 1.
Group 5 The lines switch in opposite directions.

In Table I, $C_{wt} = (C_s/k) + H_t(C_{d\min} + C_{g\min})$ and $C_{ws} = (C_s/k) + H_tC_{d\min} + H_mC_{g\min}$, where H_m and H_a denote the sizes of the main and assistant drivers, respectively; $H_t = H_m + H_a$; k is the number of wire segments; $C_{g\min}$ and $C_{d\min}$ are the gate capacitance and the drain diffusion capacitance of a minimum sized inverter; and C_s and C_c are the total wire-to-ground and wire-to-wire capacitances, respectively.

To ensure error-free operation, timing constraints have to be satisfied for the switching pattern that causes the worst-case delay, which are the ↑↓ and ↓↑ combinations. Since the effective load is highest for these patterns, the size of the buffer designed statically for the worst-case delay is much larger than would be necessary for the same timing requirements for other patterns [3]. Now this worst-case condition occurs only twice out of 16 possible input switching patterns, with a probability of 1/8 for simultaneously switching lines if the transitions are equally distributed as in a random bit stream. For the 14 other cases, the wire is driven faster, which just translates to slack which typically cannot be used, consuming energy unnecessarily. The driver proposed here changes its drive strength depending on the neighbour's switching direction by using some simple logic.

The other useful feature in the SMART driver is its ability to reduce jitter while saving energy. The SMART driver achieves this energy saving by delaying the response for the best-case without affecting the worst-case (*delay-balancing*), so that the variation in delay is as small as possible [2], [12]. This is illustrated in Fig. 2, where the curves with solid lines represent the output response of a conventional driver for minimum effective capacitance (*best-case*) and maximum effective capacitance (*worst-case*).

We propose this SMART driver circuit for a regular bus structure, and in that case, there will be at least two aggressors for a middle wire. This problem is addressed by adding an extra assistant as in our previous work [2] for each extra aggressor, and a unified analytic optimization algorithm is not derived, but an empirical sizing methodology can easily be obtained. In terms of the hardware, there is very little added complexity in the control logic. Similarly, irregularly spaced aggressors do not impose any special problems, because we can merely stagger the
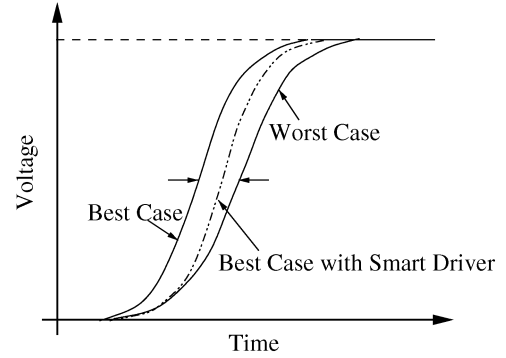

Fig. 2. Jitter reduction using SMART driver.

spacing of repeaters appropriately, so that long extra control signals are not necessary.

B. Circuit Realization

The implementation has been carried out in a UMC 0.18-μm CMOS technology, with a V_{DD} of 1.8 V. All simulations are carried out using Cadence Spectre.

In the implementation, a decision is made prior to the next transition about whether or not it constitutes a worst-case pattern. This decision is based on the relative logic values of the aggressor and the victim at the current time. Since the assistant driver needs to switch on for the worst-case patterns described in Group 5 in Table I, any time the present state has opposing logic values on the victim and aggressor, the assistant is turned on. This actually turns the assistant on for two other patterns which are not worst-case, namely patterns 10 and 12 in Table I, which reduces the energy saving from the theoretical maximum, but allows a robust and fairly simple circuit implementation.

The complete schematic is shown in Fig. 3. The transistors P_a and N_a form the assistant driver, whereas the inverter $I1$ is the main driver. Two transmission gates (TG_p and TG_n), drive the pull-up and pull-down networks of the assistant driver. The weak transistors P_k and N_k act as keepers ensuring that the assistant driver is turned off properly when the corresponding transmission gate is disabled.

The propagation delay of the selection logic T_{logic} is designed to be greater than or equal to the delay of the main driver (from node $Victim\_In$ to node $Victim\_Out$) so that node $Victim\_Out$ is able to change before Pa or Na change.

Figs. 4 and 5 show the simulation results at the far end of a 2.5-mm-long wire driven by a SMART repeater and a traditional repeater (inverter). The waveforms show the delay equalization for different switching patterns taking place.

Along with the delay performance, the noise resilience of the proposed driver is of paramount importance. In our implementation, we avoid the use of skewed inverters while using complementary logic with a switching threshold of $V_{DD}/2$ throughout the control circuitry. The only exception is the transmission gate pair, which are protected by keepers. Hence, a preliminary analysis reveals a relatively high noise margin for the circuit. Nevertheless, a more comprehensive noise analysis is earmarked for future work.

III. ENERGY MODEL OF THE SMART DRIVER

1) Dynamic Energy: If all switching events are random uniformly distributed events with no correlation between neighboring lines, the average energy dissipation per transition for wire i can be obtained by averaging out the dynamic energy consumption for each pattern. Then

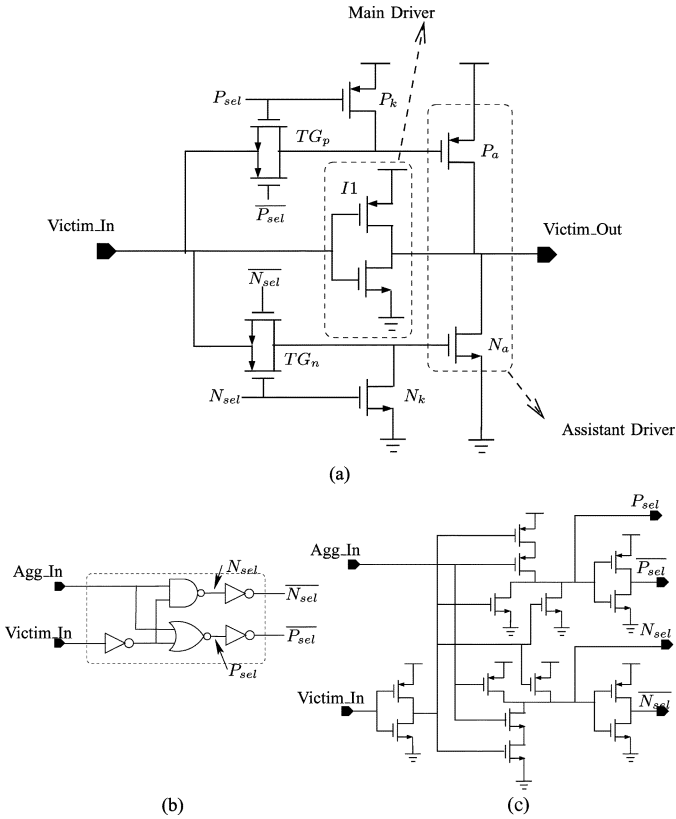| Feature size (nm)	180	130	90	65	45	32
L_{eff} (nm)	120	49	35	24.5	17.5	12.6
V_{dd} (V)	1.8	1.3	1.2	1.1	1.0	0.9
I_{dsat} $(\mu A/\mu m)$	554	1000	1100	1150	1200	1250
t_{ox} (nm)	4.2	1.6	1.4	1.2	1.1	1
V_{th} (V)	0.53	0.288	0.284	0.289	0.292	0.295
$I_{off}(nA/\mu m)$	20	30	50	70	100	150
$freq$ (GHz)	1.0	1.6	2.0	2.5	3.0	3.5
R_{dmin} $(k\Omega)$	8.27	14.15	16.62	20.82	25.40	30.48
C_{gmin} (fF)	2.31	0.43	0.25	0.14	0.077	0.043
C_{dmin} (fF)	2.00	0.49	0.33	0.22	0.15	0.10
Width$(w)(nm)$	525	335	205	145	102	70
Aspect Ratio (AR)	2.1	2.1	2.1	2.2	2.3	2.4
k_{ILD}	3.5	3.3	2.8	2.5	2.1	1.9
$r_w(\Omega/mm)$	38	93	249	475	919	1870
$c_s(fF/mm)$	36	34	29	26	22	20
$c_c(fF/mm)$	101	96	81	75	65	61

Fig. 3. Circuit schematic of the proposed SMART driver and the selector logic. (a) Driver circuit. (b) Gate-Level schematic of selector logic. (c) Transistor-level schematic of selector logic.
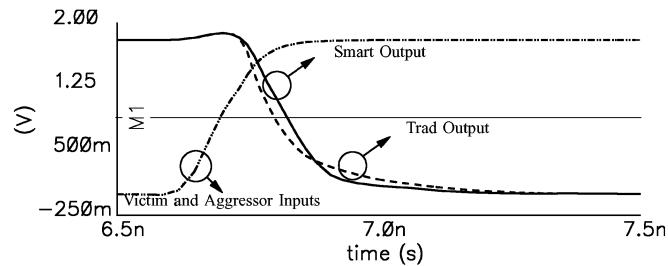

Fig. 4. Waveforms when the aggressor and victim switch in the same direction with only the main driver being active.
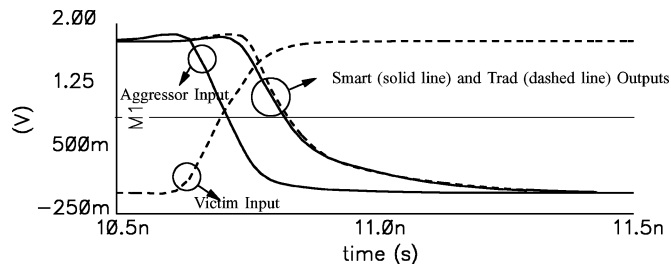

Fig. 5. Waveforms when the aggressor and victim switch in opposite directions with both drivers active.

the dynamic energy dissipation for a wire buffered with k traditional repeaters is

$$E_{\text{trad}}^{\text{dyn}} = \frac{V_{\text{DD}}^2}{32}\left(8C_{wt} + 7\frac{C_c}{k}\right) \tag{1}$$

and that for a wire buffered with k SMART repeaters is

$$E_{\text{smrt}}^{\text{dyn}} = \frac{V_{\text{DD}}^2}{32}\left(4C_{wt} + 4C_{ws} + 7\frac{C_c}{k}\right). \tag{2}$$

The dynamic energy consumption of the selection logic is found by estimating the total effective load capacitance including parasitic capacitances of all the gates and multiplying it by $V_{\text{DD}}^2/2$ and the activity factor.

2) Short Circuit Energy: Assuming the short-circuit current spike is a triangle with a peak I_{peak}, and a base t_{sc}, the short-circuit energy is shown in [13], [14] to be

$$E_{SC_{l\to h}} = \frac{1}{2}I_{\text{peak}}t_{sc}V_{\text{DD}} \tag{3}$$

where

$$I_{\text{peak}} = I_{\text{dsat}}\left(\frac{V_{gs} - V_t}{V_{\text{DD}} - V_t}\right)^{1.3} \tag{4}$$

$$t_{sc} = 1.1\left[R_d(C_d + C_g + C_w) + R_w C_g + 0.4R_w C_w\right] \tag{5}$$

with R_d being the device resistance averaged over the switching range during which the short-circuit current flows and V_{gs} is the source to gate voltage of the MOS transistor. It is assumed that the peak current occurs in the middle of the transition and hence, $V_{gs} \approx V_{\text{DD}}/2$.

In modeling the short circuit power consumed in the selector logic, the series connected pMOS/nMOS combination is represented by an equivalent single pMOS/nMOS device for the purpose of computing the driving resistance. This resistance is multiplied by the load capacitance to obtain t_{sc}, which is

$$t_{sc\_gate} \approx R_{\text{gout}}(C_{\text{dout}} + C_{\text{gin}}) \tag{6}$$

where R_{gout} is the equivalent output resistance of the gate, C_{dout} is the output capacitance, and C_{gin} is the fan-out capacitance.

3) Leakage Energy: The average leakage energy of a MOS transistor is given by

$$E_{\text{leakage}} = \frac{V_{\text{DD}}I_{\text{leakage}}}{f_{\text{clk}}} \tag{7}$$

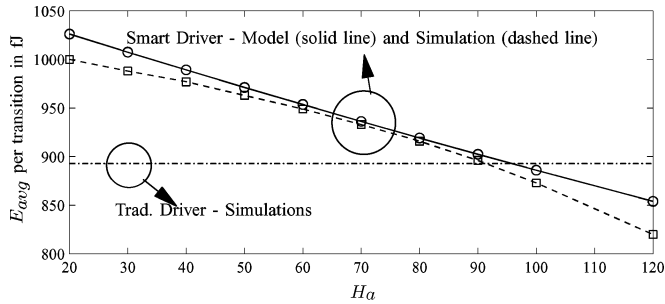$$= \frac{V_{\text{DD}}(I_{\text{offn}}W_n + I_{\text{offp}}W_p)}{2f_{\text{clk}}}. \tag{8}$$

Fig. 6. Variation of energy per transition with the size of assistant driver.

TABLE III
ENERGY DISSIPATION FOR EACH SWITCHING GROUP ($H_a = 104$)

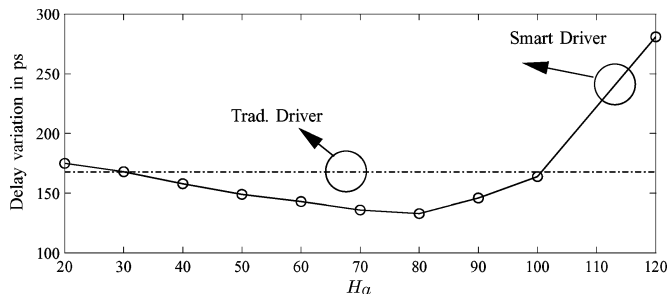|	Driver	Energy Dissipation (fJ)			E_{avg} (fJ)
		Group 1	Group 3/4	Group 5	
Model	Trad.	1507	1888	2195	939
	Smart	934	1274	2195	789
	Selector	77	77	77	77
	ΔE	33%	28%	-4%	7.8%
Simu	Trad.	1530	1735	1997	893
	Smart	994	1248	2065	753
	Selector	57	102	215	71
	ΔE	31%	22%	-14%	8%

Fig. 7. Delay variation with the size of assistant driver.

IV. ENERGY MODEL VALIDATION

The energy models derived in Section III are validated for the UMC 0.18-μm implementation by running simulations for a wire length of 10 mm. The relevant parameters for this technology node are given in column one of Table II. For the traditional inverting repeater, the parameters of repeater size ($H_t = 156$) and number of repeaters ($k = 5$) for minimizing delay are obtained from the well-known methodology described in [3]. Applying the general methodology described in the Appendix for optimizing the SMART driver, the parameters $H_m = 52$, $H_a = 104$, and $k = 5$ are obtained.

The simulations show that the energy models derived for the traditional and SMART repeaters are accurate to within 95% of their simulated values as evidenced in Fig. 6 and summarized in Table III. As predicted by the model, increasing the size of the assistant driver will increase the energy saving, although at the cost of increased delay, if the size is increased beyond the optimal (refer to Fig. 7).

It is evident from Table III that the energy loss introduced by the extra selection logic for switching patterns in Group 5, where both the assistant and main drivers switch, is more than offset by the energy saving for those patterns in Groups 1, 3, and 4, where the assistant

TABLE IV
MAXIMUM CROSSTALK ON A QUIET LINE

	Near End	Far End
Smart Repeater	0.113 V	0.213 V
Traditional Repeater	0.069 V	0.206 V

TABLE V
ENERGY SAVING FOR FUTURE GENERATIONS

Tech. node (nm)	130	90	65	45	32
k	14	24	36	54	84
H_t	325	268	277	278	282
H_a	202	162	163	158	154
E_{smrt}/E_{trad}	0.74	0.75	0.77	0.80	0.83

does not switch. On average, assuming equally likely occurrences of all patterns, the total energy saving is around 10%.

There is a slight increase in the peak crosstalk voltage with a smart driver compared to a traditional driver but the peak crosstalk at the far-end is about 12% of V_{DD}, which is in the normally acceptable range of 20% of V_{DD}. It can be seen from Table IV that the near-end cross-talk with the smart repeater is doubled compared to a traditional repeater, but the total percentage is 0.1% of V_{DD}.

V. IMPACT OF TECHNOLOGY SCALING

In this section, the potential of the SMART Driver to save energy in future technology nodes is investigated. As the feature size decreases, the short circuit energy increases fairly sharply, which adversely affects the energy saving due to the fact that the SMART driver has a few transistors in the selector logic. However, this is offset to some degree due to the relative decrease in area and the associated dynamic energy consumption of the selection logic in comparison to the driving inverters. Since global wires are scaled selectively, the wire parasitics remain approximately the same, or are worse, and the driving transistors see no reduction in size [15]. In contrast, the selection logic can be implemented with minimum sized transistors and the dynamic energy consumed becomes truly negligible. An analysis was carried out using ITRS predictions to derive the relevant technology parameters, as summarized in columns 2–6 in Table II. The predicted total average energy saving in driving global length wires is shown in Table V, highlighting the usefulness of the SMART driver right up to the 32-nm node.

VI. CONCLUSION

In this paper, we addressed the issue of reducing energy consumption by exploiting the switching-pattern-dependent delay of repeater-inserted global wires. The proposed circuit was implemented in an UMC 0.18-μm CMOS technology and tested for proof of concept. The average energy saving was shown to be around 10%, and the jitter reduction to be 20% for a data rate of 1 GB/s.

A comprehensive delay and energy analysis was presented, including a design methodology to obtain the optimal repeater configurations for minimizing delay while also minimizing jitter. Further, as processes scale, the selector latency shrinks, and higher data rates can be achieved. The total energy saving that can be achieved by the SMART driver in future nanometer technologies is found to be in the range of 20%–25%.

APPENDIX
DELAY-BALANCED DRIVER SIZING

The methodology for delay-balanced driver sizing for the SMART driver is described extensively in our previous work [12]. The delay analysis for SMART repeater insertion uses the characterization of a minimum-sized repeater in terms of an output resistance $R_{d\min}$, input gate capacitance $C_{g\min}$, and output drain-diffusion capacitance $C_{d\min}$. When both the main and assistant drivers are switching, $R_d = R_{d\min}/(H_m + H_a)$ and $C_g = C_{g\min}(H_m + H_a)$. An expression can be derived for the associated delay T_{MA}, by using the Elmore delay as in [13]. When the assistant driver is quite, the driver resistance changes to $R_d = R_{d\min}/H_m$, and the gate capacitance to $C_g = H_m C_{g\min}$. Thus, the delay expression (T_M) when the assistant is quiet is found. Since the assistant driver switches only for the worst-case switching pattern defined by Group 5 in Table I, the size of the assistant driver H_a can be used to tune the delays for the other switching combinations defined by Groups 1, 3, and 4. The delay variation can be quantified as

$$\Delta T = T_{MA} - T_M.$$

By setting $\Delta T = 0$, delay balancing can be achieved. Using the relation $H_{mDB} = H_t - H_{aDB}$, a quadratic equation for H_{aDB} can be obtained, the solution to which gives the delay balanced assistant driver size.

REFERENCES

[1] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," in *Proc. Int. Workshop Syst. Level Interconnect Prediction (SLIP)*, 2004, pp. 7–13.

[2] R. Weerasekera, L.-R. Zheng, D. Pamunuwa, and H. Tenhunen, "Switching sensitive interconnect driver to combat dynamic delay in on-chip buses," in *Proc. Lecture Notes Comput. Sci. (PATMOS)*, Sep. 2005, vol. 3728, pp. 277–285.

[3] D. Pamunuwa, L.-R. Zheng, and H. Tenhunen, "Maximizing throughput over parallel wire structures in the deep submicrometer regime," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 11, no. 2, pp. 224–243, Apr. 2003.

[4] A. B. Kahng, S. Muddu, E. Sarto, and R. Sharma, "Interconnect tuning strategies for high-performance ICS," in *Proc. Conf. Des., Autom. Test Eur. (DATE)*, 1998, pp. 471–478.

[5] T. Iima, M. Mizuno, T. Horiuchi, and M. Yamashina, "Capacitance coupling immune, transient sensitive accelerator for resistive interconnect signals of subquarter micron ULSI," *IEEE J. Solid-State Circuits*, vol. 31, no. 4, pp. 531–536, Apr. 1996.

[6] P. Sotiriadis, T. Konstantakopoulos, and A. Chandrakasan, "Analysis and implementation of charge recycling for deep sub-micron buses," in *Proc. Int. Symp. Low Power Electron. Des. (ISLPED)*, 2001, pp. 364–369.

[7] A. Nalamalpu, S. Sirinivasan, and W. P. Burleson, "Boosters for driving long on chip interconnects-design issues, interconnect synthesis, and comparison with repeaters," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 21, no. 1, pp. 50–62, Jan. 2002.

[8] H. Kaul and D. Sylvester, "Low-power on-chip communication based on transition-aware global signaling (TAGS)," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 5, pp. 464–476, May 2004.

[9] A. Katoch, S. Jain, and M. Meijer, "Aggressor aware repeater circuits for improving on-chip bus performance and robustness," in *Eur. Solid-State Circuits Conf. (EUSCIRC)*, Sep. 2003, pp. 261–264.

[10] H.-Y. Huang and S.-L. Chen, "Interconnect accelerating techniques for sub-100-nm gigascale systems," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 11, pp. 1192–1200, Nov. 2004.

[11] H. B. Backoglu, *Circuits, Interconnections and Packaging for VLSI*. New York: Addison-Wesley, 1990.

[12] R. Weerasekera, D. Pamunuwa, L.-R. Zheng, and H. Tenhunen, "Minimal-power, delay-balanced smart repeaters for interconnects in the nanometer regime," in *Proc. International Workshop Syst.-Level Interconnect Prediction (SLIP)*, 2006, pp. 113–120.

[13] T. Sakurai and A. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.

[14] D. Sylvester, W. Jiang, and K. Keutzer, "Berkeley advanced chip performance calculator," [Online]. Available: http://www.eccs.umich.edu/dennis/bacpac/

[15] Semiconductor Corp., "The international technology roadmap for semiconductors (ITRS)," 2003 [Online]. Available: http://www.itrs.net/Links/2003ITRS/Home2003.htm

[16] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm design exploration," in *Proc. 7th Int. Symp. Quality Electron. Des. (ISQED)*, 2006, pp. 585–590.

[17] L.-R. Zheng, D. Pamunuwa, and H. Tenhunen, "Accurate a priori signal integrirty estimation using a dynamic interconnect model for deep submicron VLSI design," in *Proc. Conf. Euro. Solid-State Circuits (ESSCIRC)*, 2000, pp. 324–327.