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

Roshan Weerasekera, Dinesh Pamunuwa*, Li-Rong Zheng and Hannu Tenhunen
Department of Electronic, Computer and Software Systems,
KTH Information and Communication Technology,
ELECTRUM 229, 164 40 Kista, Sweden.
*Centre for Microsystems Engineering, Faculty of Applied Sciences,
Lancaster University, Lancaster LA1 4YR, United Kingdom.
{roshan,lirong,hannu}@imit.kth.se, d.pamunuwa@lancaster.ac.uk

ABSTRACT

In this paper we propose a smart repeater that consumes less energy and is suitable for driving global interconnections in nanometre technologies. When there is coupling between interconnects, the effective capacitance of a given wire is a function not only of the physical geometry, but also the relative switching pattern described by the bits on the wire in question (the victim) and the adjacent wires (aggressors). The drive strength of a traditional repeater is static, resulting in 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 *SMART* repeater, the drive strength is dynamically altered depending on the relative bit pattern, by partitioning it into a *Main Driver* and *Assistant Driver*. For a higher effective load capacitance both drivers switch, while for a lower effective capacitance the assistant driver is quiet. 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 shown that the potential average saving in energy can be as much 15% with a 18% jitter reduction over a traditional repeater for typical global wire lengths in nanometre technologies.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Types and Design Styles—*VLSI (Very Large Scale Integration)*

General Terms

Design,Performance

Keywords

Delay-Balanced, Interconnects, Minimal-Power, Repeaters.

SLIP'06, March 4–5, 2006, Munich, Germany.

1. INTRODUCTION

Shrinking of the minimum feature size used in fabrication of ICs has resulted in exponential growth of performance and functionality over the past four decades. The integration of millions of devices on a single die however poses many difficult engineering challenges, most notably in power management and on-chip communication. As chip complexity and area grow, despite the best efforts to exploit locality with innovative architectural solutions, the average distance across which a bit has to be transferred has increased, and interconnection delay is a key bottleneck in modern digital design. Scaling of wires and tighter integration has also resulted in signal integrity problems which only add to the interconnection woes; cross-talk between signal lines results in signal corruption and variable delay, depending on the respective switching patterns.

A key technique in reducing propagation delay and signal degradation 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 [13, 21, 14]. Further, over 90% of this power is concentrated in only 10% of the interconnects; i.e. those which are classed as global and run for a significant fraction of the die length.

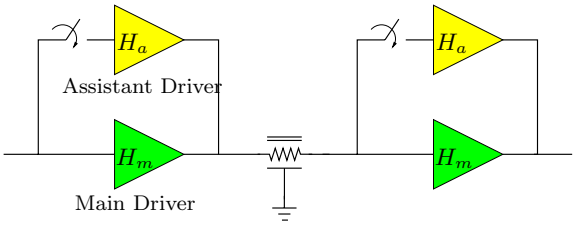


Figure 1: Basic schematic of the proposed driver scheme

In this paper we propose a repeater that consumes less energy, and is suitable for exactly these kinds of global interconnections. It exploits the fact that in a parallel wire structure, the effective capacitance of a given wire is dynamic; i.e. it is a function of not only the physical geometry, but also 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*. For a higher effective load capacitance both drivers switch, while for a lower effective capacitance the assistant driver is quiet [27]. 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 shown that the potential average saving in energy can be as as much 15% over a traditional repeater for typical global wire lengths in nanometre technologies.

The ramifications of the dynamically changing load in coupled interconnects have received a fair amount of attenion in the literature. A comprehensive analysis of design considerations for repeater insertion in a bus structure with heavy coupling was presented in [19]. A scheme proposed in [9] 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) [8], Charge Recycling Technique (CRT) [24], Boosters [15], the TAGS receiver [11], the Aggressor-Aware Repeater [10], and the Capacitor Coupled Trigger and Accelerator combination [7]. Some of these use skewed inverters to trade-off noise margin for speed [8, 15, 11], while others consume more energy [10] and occupy a larger area [8, 15, 11] to produce a faster response.

The work done in [6] and [17] also seek to reduce the delay by avoiding simultaneous switching similar to [9], but they accomplish this by introducing static delays in the repeaters rather than by physical offsets in the placement. They report an overall reduction in the delay for the worst-case pattern of upto 20%, but this scheme dissipates more power for transitions in the same direction, due to additional charging and discharging of the coupling capacitance. [4] and [16] report average energy savings of upto 25% by introducing a delay dependant on the relative transition pattern between two adjacent wires, but this additional delay introduces a timing penalty. The worst-case pattern for the delay is also the worst case pattern for the energy, and hence any energy saving is at the cost of an increase in the cycle time, which may not always be possible.

Another approach is to use Error Control or Transition Coding Techniques [26, 22, 25, 12, 23, 1, 20] to overcome the effects of inter-symbol interference. The relatively complex codec circuitry causes additional delay and consumes more power, rendering the coding ineffective in many cases [3]. Even otherwise, these schemes mostly address the problem of reducing transitions on a given wire, which is less important than reducing the relative switching activity between lines, given that the aspect ratio of on-chip interconnect emphasises the coupling capacitance over the self capacitance.

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 [2], resulting in easy amalgamation in CAD flows at different levels of hierarchy from initial signal planning to detailed place and route [18].

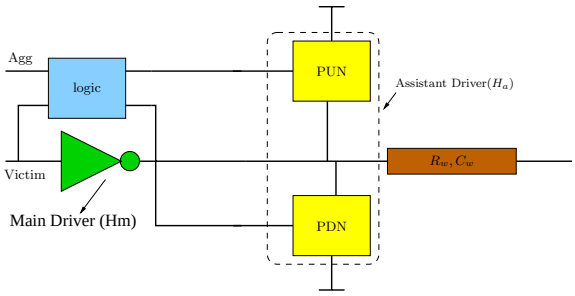


Figure 2: Schematic of proposed SMART repeater, shown here only for two coupled nets

One of the main advantages 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. We present a design methodology similar to that for traditional single-wire inverting repeaters, including an RC equivalent circuit and closed-form expressions for the first-order approximation to the delay. Therefore this repeater can be very easily modeled in tasks such as delay calculation, signal integrity analysis and timing driven optimisation in any CAD flow for physical design.

The rest of this document is structured as follows. Section 2 describes the repeater circuit and analyses the energy saving resulting from it. The following section presents the circuit abstraction and design methodology. In the next section we present simulation results that validate the delay analysis and show the average saving in energy for typical wire lengths in a 180nm technology. We end with a discussion.

2. CONCEPT OF THE SMART DRIVER

The effective interconnect capacitance varies with the transitions of neighboring lines and can be written as $C_s + \lambda C_c$, where C_s is the self capacitance of the wire, λ is the switch factor and C_c is the inter-wire capacitance. In this work we use different switch factors for delay and power estimation (given in Table 1) based on the experimental validation in [5] which proposes power-based switch factors that are slightly different from the delay-based ones. The variation of the effective capacitance with the relative switching pattern introduces a spread in the arrival time at the far end of the wire. To demonstrate this 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. 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); *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.

To ensure error-free operation, timing constraints have to be satisfied for the switching pattern that causes the worst-case delay, which are the $\uparrow\downarrow$ *and* $\downarrow\uparrow$ combinations. Since the effective load is highest for these patterns, the size

Group	Case	Switching Event on		Switch Factor		Energy Dissipation for wire i ($\times\frac{1}{2}V_{dd}^2$) with	
		wire i	wire j	Delay-Based(λ^d)	Power-Based(λ^p)	Traditional driver	Smart driver
1	1	$\downarrow$	$\downarrow$	0	0.25	$C_{w\_trad} + 0.25C_c$	$C_{w\_smart} + 0.25C_c$
	2	$\uparrow$	$\uparrow$	0	0.25	$C_{w\_trad} + 0.25C_c$	$C_{w\_smart} + 0.25C_c$
2	3	0	0	n.a.	n.a.	0	0
	4	0	1	n.a.	n.a.	0	0
	5	1	0	n.a.	n.a.	0	0
	6	1	1	n.a.	n.a.	0	0
3	7	0	$\uparrow$	1	1	0	0
	8	$\uparrow$	0	1	1	$C_{w\_trad} + C_c$	$C_{w\_smart} + C_c$
	9	0	$\downarrow$	1	1	0	0
	10	$\downarrow$	0	1	1	$C_{w\_trad} + C_c$	$C_{w\_smart} + C_c$
4	11	1	$\uparrow$	1	0	0	0
	12	$\uparrow$	1	1	0	C_{w_trad}	C_{w_smart}
	13	1	$\downarrow$	1	0	0	0
	14	$\downarrow$	1	1	0	C_{w_trad}	C_{w_smart}
5	15	$\uparrow$	$\downarrow$	2	1.75	$C_{w\_trad} + 1.75C_c$	$C_{w\_trad} + 1.75C_c$
	16	$\downarrow$	$\uparrow$	2	1.75	$C_{w\_trad} + 1.75C_c$	$C_{w\_trad} + 1.75C_c$

Table 1: Switching Activities on the lines and the variation of effective capacitance.

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 [19]. 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. A basic schematic of the proposed *SMART* repeater is shown in Figure 2. If the switching pattern belongs to Groups 1, 3, or 4, a single inverter (the Main driver) drives the interconnect. When a switching pattern in Group 5 occurs, another inverter (the Assistant) also drives the line, increasing the total drive strength appropriately. By disconnecting the Assistant driver when it is not needed, part of the parasitic capacitance is disconnected for the majority of the switching patterns, leading to a saving in the average energy consumption.

2.1 Energy Saving with the SMART Driver

Each time a wire is driven from 0 to V_{DD}, an energy amounting to $C_{eff}V_{dd}^2$ is drawn from the power supply. Half of this is stored in the load capacitance while the rest is dissipated in the pull-up network of the driver. During a V_{DD} to 0 transition, the energy stored in the capacitance is dissipated in the pull-down network of the driver. The energy dissipation when two wires are coupled together is given in Table 1 for all possible switching patterns. In accordance with common terminology, the size of a traditional inverting repeater is defined in terms of multiples of a minimum sized repeater as H_t. Since the driving portions of the SMART driver are two inverters, they can be characterized in a similar fashion as H_m and H_a, which denote the sizes of the Main and Assistant drivers respectively. The total static capacitive load of the traditional driver, C_{w_trad}, can be defined as $C_s + H_t(C_{dmin} + C_{gmin})$ (i.e. the sum of its own parasitic drain capacitance, the self capacitance of the wire, and the gate capacitance of the target load (a repeater for the purpose of this analysis) at the end of the wire. Here C_{gmin} and C_{dmin} are the gate capacitance and the drain diffusion capacitance of a minimum sized inverter. Similarly, C_{w_smart} can be described as $C_s + H_tC_{dmin} + H_mC_{gmin}$.

The energy dissipation per cycle depends on whether or not switching transitions occur, and on the relative switching pattern as given in Table 1. A switching transfer is a probabilistic event, and in general there can be temporal and spatial correlations between transitions on the same line and also on different lines. The average energy dissipation for wire i with traditional repeater insertion is

$$\begin{aligned} E_{avg}^{trad}(i) &= \frac{V_{DD}^2}{2}\left[p_{s,s}\left(C_{w\_trad} + 0.25C_c\right) + p_{e,1}C_{w\_trad}\right. \\ &+ \left. p_{e,0}\left(C_{w\_trad} + C_c\right) + p_{o,o}\left(C_{w\_trad} + 1.75C_c\right)\right] \end{aligned} \quad (1)$$

where $p_{x,y}$ is the probability that wires i, j switch as defined below:
(s, s) - both wires switch in same direction;
$(e, 0)$ - wire i switches up or down while wire j is quiet at 0;
$(e, 0)$ - wire i switches up or down while wire j is quiet at 1;
(o, o) - both wires switch in different directions.

This can be simplified to:

$$\begin{aligned} E_{avg}^{trad}(i) &= \frac{V_{DD}^2}{2}\left[(p_{s,s} + p_{e,0} + p_{e,1} + p_{o,o})C_{w\_trad}\right. \\ &+ \left. (0.25p_{s,s} + p_{e,0} + 1.75p_{o,o})C_c\right] \end{aligned} \quad (2)$$

As enumerated in Table 1, the capacitive load C_w for the SMART driver in a bus structure is C_{w_smart} for Group 1-4 and C_{w_trad} for Group 5. Hence the energy dissipation is:

$$\begin{aligned} E_{avg}^{smart}(i) &= \frac{V_{DD}^2}{2}\left[(p_{s,s} + p_{e,0} + p_{e,1})C_{w\_smart}\right. \\ &+ \left. p_{o,o}C_{w\_trad} + (0.25p_{s,s} + p_{e,0} + 1.75p_{o,o})C_c\right] \end{aligned} \quad (3)$$

From (3) and (4) the average total energy saving for a single segment of a repeater-inserted net is found to be:

$$\Delta E = 0.5(p_{s,s} + p_{e,0} + p_{e,1})(C_{w\_trad} - C_{w\_smart})V_{DD}^2 \quad (4)$$

Substituting $C_{w\_trad} = C_s + H_t(C_{gmin} + C_{dmin})$ and $C_{w\_smart} =$

$C_s + H_t C_{dmin} + H_m C_{gmin}$, this simplifies to:

$$\Delta E = \frac{H_a}{2}(p_{s,s} + p_{e,0} + p_{e,1}) C_{gmin} V_{DD}^2 \quad (5)$$

If all switching events are random uniformly distributed events with no correlation between neighbouring lines, $p_{s,s} = p_{e,1} = p_{e,0} = \frac{2}{16}$ and (5) reduces to

$$\Delta E_{avg} = \frac{3}{16}(H_t - H_m) C_{gmin} V_{dd}^2 \quad (6)$$

This can be verified by averaging the energy disspation given in Table 1:

$$E_{avg\_trad}(i) = \frac{V_{DD}^2}{32}(8C_{w\_trad} + 6C_c) \quad (7)$$

$$E_{avg\_smart}(i) = \frac{V_{DD}^2}{32}(2C_{w\_trad} + 6C_{w\_smart} + 6C_c) \quad (8)$$

It can be easily seen the energy saving given by (7)-(8) is the same as (6).

This energy analysis assumes that the energy consumed by the selection logic of the SMART driver is negligible. This is true for repeaters driving global wires, and the smaller the technology, the more accurate is the assumption.

To obtain the same performance as a traditional driver for the worst-case, the main and assistant drivers are designed such that $H_t = H_m + H_a$. In essence (5) affirms that the larger the Assistant driver, the larger the energy saving, as this results in a lower average load.

2.2 Jitter Reduction with the SMART Driver

In Figure 3 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*). Our aim, with the SMART driver, is to delay the response for the best-case without affecting the worst-case, so that the variation in delay becomes as small as possible. In other words, the concept is to make the response slower with minimum effective capacitance, as the cycle delay has to be set to the worst-case delay anyway.

3. DESIGN METHODOLOGY

3.1 Delay Modeling

The delay analysis for repeater insertion is carried out by characterizing a minimum sized inverter by an output resistance R_{gmin}, in addition to the input gate capacitance C_{gmin} and output drain-diffusion capacitance C_{dmin} already defined. This linearization allows us to use superposition to find the total delay, and is sufficiently accurate for the intended design tasks of global signal planning and incremental physical optimization. The 50% delay of the circuit can be approximated by the Elmore delay, with a factor of 0.4 being used to account for the distributed nature of the wires. The delay of the pattern sensing circuitry has been neglected.

3.1.1 Delay Analysis with both Drivers Switching

With the linearisation of the driver, the equivalent circuit for one repeater segment can be shown to be the circuit in Figure 4. Hence the 50% delay for the wire can be expressed

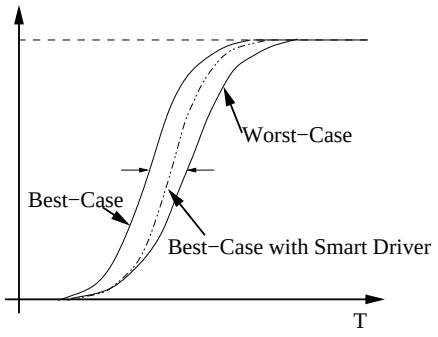


Figure 3: Method of Jitter reduction using SMART driver

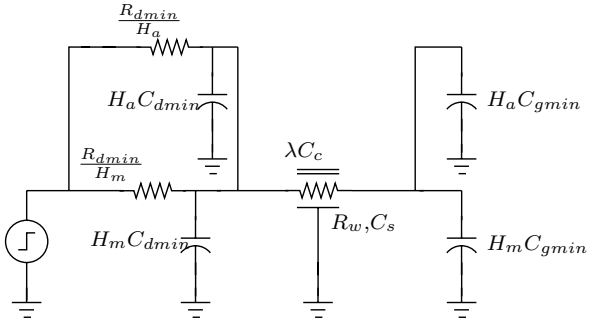


Figure 4: Equivalent Circuit for the case when both drivers are switching

as

$$T_{MA} = k\left\{0.7R_d\left(C_d + \frac{C_w}{k} + C_g\right) + 0.7\frac{R_w}{k}C_g + 0.4\frac{R_w}{k}\frac{C_w}{k}\right\} \quad (9)$$

Where k is the number of repeaters, and the parasitics are $R_d = \left(\frac{R_{dmin}}{H_m} \parallel \frac{R_{dmin}}{H_a}\right) = \frac{R_{dmin}}{H_m + H_a}$, $C_g = C_{gmin}(H_m + H_a)$, $C_d = C_{dmin}(H_m + H_a)$ and $C_w = C_s + \lambda C_c$. Here H_m and H_a are the sizes of the Main and Assistant drivers respectively, and λ is the switching factor. Since the Assistant driver switches only when adjacent lines switch in opposite directions, λ=2. To simplify the delay equation, the following time constants are defined: $t_{Dout} = R_{dmin}C_{dmin}, t_{DWs} = R_{dmin}C_s, t_{DWc} = R_{dmin}C_c, t_{Din} = R_{dmin}C_{gmin}, t_{WD} = R_w C_{gmin}, t_{Ws} = R_w C_s$ and $t_{Wc} = R_w C_c$. This results in:

$$T_{MA} = 0.7k(t_{Dout} + t_{Din}) + \frac{0.7(t_{DWs} + 2t_{DWc})}{(H_m + H_a)} + 0.7t_{WD}(H_m + H_a) + 0.4\frac{(t_{Ws} + 2t_{Wc})}{k} \quad (10)$$

3.1.2 Delay Analysis with the Assistant Quiet

When the Assistant driver is quiet while the Main driver is switching, the gate capacitance of the Assistant will not add to the load, as it is disconnected by a switch in which the input capacitance is negligible compared to the Assistant driver's input capacitance (see Figure 5). However the parasitic drain-diffusion capacitance will always add to the load.

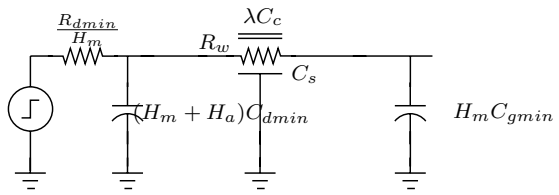


Figure 5: Equivalent Circuit for the case when the Main driver is switching

The delay expression is now

$$T_M = k\left\{0.7\frac{R_{dmin}}{H_m}\left[C_d + \frac{C_s + \lambda C_c}{k} + H_m C_{gmin}\right] + 0.7\frac{R_w}{k}H_m C_{gmin} + 0.4\frac{R_w(C_s + \lambda C_c)}{k}\right\}$$

Replacing the time constants defined earlier, this is reduced to:

$$T_M = 0.7k\left[t_{Dout}\left(1 + \frac{H_a}{H_m}\right) + t_{Din}\right] + 0.7H_m t_{WD} + \frac{0.7(t_{DWs} + \lambda t_{DWc})}{H_m} + 0.4\frac{t_{Ws} + \lambda t_{Wc}}{k} \quad (11)$$

where $\lambda \in \{0, 1\}$
Equations (10) and (11) are the two principal delay equations of the SMART driver for its two states of Main and Assistant drivers switching, and Main driver switching while the Assistant driver is quiet. The accuracy of these delay expressions was checked against simulated values, and the results are presented in Table 2. Here R_{dmin}, C_{gmin} and C_{dmin} are chosen as $9k\Omega$ and $C_{dmin} = C_{gmin} = 1.8fF$ as being representative values for a repeater in a 180 nm technology. The average error of under 10% is low enough to allow this first order delay model to be used for optimisation tasks early in the design flow.

3.2 Optimum Buffer Sizes

The delay equations (10) and (11) predict a global minimum for the delay for *optimal* k,H_m and H_a values. The variation of T_{MA} and T_M with H_m and H_a is shown in Figures 6 and 7. T_{MA} is a convex function of H_m and H_a, and hence also of $(H_m + H_a)$. T_M is a convex function of H_m, while it has a linear dependence on H_a for a given value of H_m. This is a consequence of the fact that the Assistant driver contributes a parasitic capacitance to the load while not contributing any drive strength for the switching combinations represented by T_M.

We obtain the optimal H_m, H_a and k values by deriving an expression for $(H_m + H_a)$ and k by minimising t_{MA}. Then H_m is obtained by minimising t_{MA}, which then allows us to also solve for H_a[1]. The delay when the assistant driver is switching, t_{MA}, is minimised by setting the partial derivatives of t_{MA} with respect to k and $(H_m + H_a)$ to zero, which results in:

[1]A value for the optimal k can also be obtained by setting the partial derivative of T_M with respect to k to zero, but the difference is not significant.

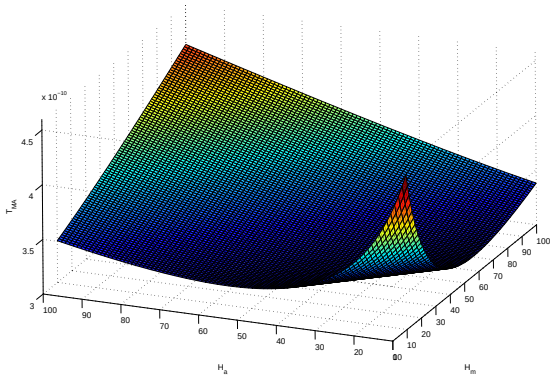


Figure 6: The Variation of T_{MA} with H_m and H_a

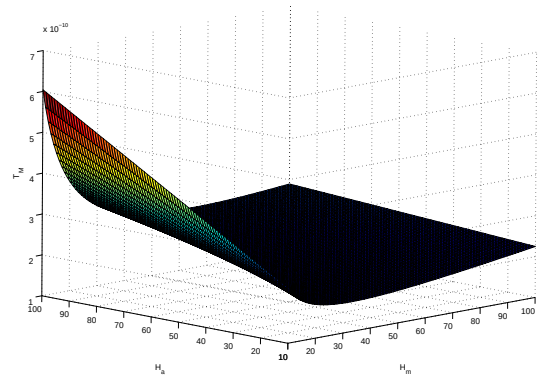


Figure 7: The Variation of T_M with H_m and H_a

$$H_{m\_opt} + H_{a\_opt} = \sqrt{\frac{t_{DWs} + 2t_{DWc}}{t_{WD}}} \quad (12)$$

$$k_{\_opt} = \sqrt{\frac{0.4(t_{Ws} + 2t_{Wc})}{0.7(t_{Dout} + t_{Din})}} \quad (13)$$

Since (10) is a convex function of total drive strength $(H_m + H_a)$, and also of k, (12) gives the optimum value for $(H_m + H_a)$ and (13) the optimal number of sections, $k_{\_opt}$, to minimise delay. Now the optimal value for H_m is obtained by setting the partial derivative of T_M with respect to H_m to zero.

$$\frac{\partial T_M}{\partial H_m} = 0 \Rightarrow H_m^2 = \frac{kH_a t_{Dout} + t_{DWs} + \lambda t_{DWc}}{t_{WD}} \quad (14)$$

By substituting for H_m in (14) from (12) the following expression is obtained:

$$H_a^2 - \left(2H_t + \frac{kt_{Dout}}{t_{WD}}\right)H_a + \left(H_t^2 - \frac{t_{DWs} + \lambda t_{DWc}}{t_{WD}}\right) = 0 \quad (15)$$

where $H_t = H_{m\_opt} + H_{a\_opt} = \sqrt{\frac{t_{DWs} + 2t_{DWc}}{t_{WD}}}$
The two possible solutions to the above quadratic equation

R_w /(Ohms)	C_s /(fF)	C_c /(fF)	$T_{MA}/(ps)$			$T_M/(ps)$		
			Model	Simulated	% Error	Model	Simulated	% Error
400	30.00	70.00	106.28	113.394	6.27	75.306	86.186	12.62
400	30.00	100.00	112.47	119.865	6.17	75.306	86.128	12.57
400	30.00	130.00	118.66	125.953	5.79	75.306	86.426	12.87
400	50.00	70.00	110.41	118.021	6.45	79.96	90.669	11.81
400	50.00	100.00	116.59	124.187	6.12	79.96	90.96	12.09
400	50.00	130.00	122.78	130.333	5.80	79.96	90.895	12.03
400	70.00	70.00	114.53	122.033	6.15	84.615	95.033	10.96
400	70.00	100.00	120.72	128.065	5.74	84.615	95.196	11.11
400	70.00	130.00	126.91	134.099	5.36	84.615	95.381	11.29
900	30.00	70.00	206.75	212.233	2.58	132.51	142.467	6.99
900	30.00	100.00	218.93	224.466	2.47	132.51	142.518	7.02
900	30.00	130.00	231.12	236.577	2.31	132.51	142.445	6.97
900	50.00	70.00	214.87	220.57	2.58	141.17	150.895	6.44
900	50.00	100.00	227.06	232.635	2.40	141.17	150.382	6.13
900	50.00	130.00	239.25	244.967	2.33	141.17	150.86	6.42
900	70.00	70.00	223	228.248	2.30	149.82	159.309	5.96
900	70.00	100.00	235.18	241.478	2.61	149.82	159.528	6.09
900	70.00	130.00	247.37	253.616	2.46	149.82	159.477	6.06

Table 2: Comparison of the delay and delay predicted by the model

are

$$H_a = H_t + \frac{kt_{Dout}}{2t_{WD}} \pm \sqrt{\frac{\mu}{4} + \frac{t_{DWs} + \lambda t_{DWc}}{t_{WD}}}$$

Where $\mu = \frac{4H_t k t_{Dout}}{t_{WD}} + \left(\frac{kt_{Dout}}{t_{WD}}\right)^2$
Since $H_a < H_t$, the meaningful value for H_a is:

$$H_a = H_t + \frac{kt_{Dout}}{2t_{WD}} - \sqrt{\frac{\mu}{4} + \frac{t_{DWs} + \lambda t_{DWc}}{t_{WD}}} \quad (16)$$

and therefore

$$H_m = \sqrt{\frac{\mu}{4} + \frac{t_{DWs} + \lambda t_{DWc}}{t_{WD}}} - \frac{kt_{Dout}}{2t_{WD}} \quad (17)$$

Here λ can be either 0 or 1. It can be seen that when there is no coupling (i.e. $\lambda = 0$) (15) gives $H_a = 0$ and consequently $H_m = H_t$ which is the logical result.

3.3 Delay Balancing with the SMART Driver

As explained earlier, the SMART driver saves energy by reducing the capacitive load for certain switching combinations, which in turn is achieved by switching off part of the driver. This means that the driver is essentially slower for the switching combinations that give rise to a lower capacitive load, and hence reduces jitter - the variation between the best-case and worst-case delays. This is a secondary benefit of the driver, and here we present a design methodology for sizing the SMART driver to minimize jitter.

Since the Assistant driver switches only for the worst-case switching pattern defined by Group 5 in Table 1, the size of the Assistant driver, H_a, can be used to tune the delays for the other switching combinations defined by Groups 1 and 3-4. The expressions in (11) and (10) represent the delay for all these switching combinations. For clarity of explanation, say T_1, T_2 and T_3 are the wire delays for Groups 1, 3-4, and 5 respectively. Hence $T_1 = T_M|_{\lambda=0}$, $T_2 = T_M|_{\lambda=1}$ and $T_3 = T_{MA}$. Now increasing H_a increases T_M (see Figure 7), and hence H_a can be sized so that either $T_1 = T_3$ or $T_2 = T_3$ ($T_1 = T_2 = T_3$ is not possible because the relative delay variation between T_1 and T_2 is not a funtion of H_a).

The delay variation can be quantified as

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

By setting $\Delta T = 0$, delay balancing can be achieved. Substituting for T_{MA} and T_M from (11) and (10) and using the relation $H_{mDB} = H_t - H_{aDB}$ the following quadratic for H_{aDB} can be obtained.

$$AH_{aDB}^2 + BH_{aDB} + C = 0 \quad (18)$$

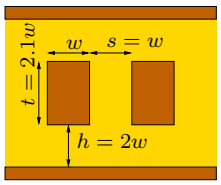
where

$$\begin{aligned} A &= 0.7t_{WD} \\ B &= 0.7\left[kt_{Dout} - t_{WD}H_t \right. \\ &+ \left. \frac{t_{DWs} + 2t_{DWc}}{H_t} + \frac{(2-\lambda)t_{Wc}}{k}\right] \\ C &= -(2-\lambda)\left[0.7t_{DWc} + \frac{0.4t_{WC}}{k}\right] \end{aligned}$$

Now sizing H_{aDB} to equalise T_1 and T_3 results in T_2 being larger than T_3, which may not always be possible due to constraints on T_3, the worst-case delay. However equalising T_2 and T_3 does not result in any such adverse effect. For the former strategy $\lambda = 0$ while for the latter strategy $\lambda = 1$ in (18). Here H_t and k can be calculated according to the strategy adopted. For example for optimal repeater insertion, H_t can be calculated from Equation (12) and k from Equation (13).

4. DISCUSSION ON DELAY AND ENERGY CONSUMPTION

Having identified the delay expressions and possible ways of optimizing the buffers, both the main driver and assistant, calculations were carried out in order to verify the strategy proposed in the paper. We designed the assistant driver considering two cases: the first case is optimization



$R_w = 39\ \Omega/mm$
$C_s = 36\ fF/mm$
$C_c = 98.6\ fF/mm$
$Length = 10mm$

Figure 8: Wire Geometrical Parameters

of delay for Group 1 ($\lambda = 0$), and the second case is delay optimization for Group 3-4 ($\lambda = 1$). These two cases has termed as *Design Strategy One* and *Design Strategy Two*. Apart from that the drivers can be sized according to the delay balancing technique too. The calculations carried out considering minimum sized global wires (Refer Figure 8 with $w = 525nm$) in 180 nm technology node and the results are shown in the Table 3. Switching probabilities to evaluate the energy saving (ΔE) were found using two uniformly distributed pseudo random bit sequences(PRBS) with probabilities $p_{s,s} = 0.1111$, $p_{e,0} = 0.1311$,$p_{e,1} = 0.1542$ and $p_{o,o} = 0.1001$. For all the strategies the maximum delay is 352 ps, and ΔE ($= \frac{E_{trad} - E_{smart}}{E_{trad}}$) and ΔT ($= \frac{\Delta T_{trad} - \Delta T_{smart}}{\Delta T_{trad}}$) values are represented as a percentage as compared to the traditional method.

Case	k	H_t	H_m	H_a	ΔT	ΔE
Strategy One	4	173	105	68	3.1%	10.9%
Strategy Two	4	173	142	31	-1.6%	5.0%
Delay Balanced	4	173	75	97	18.7%	15.6%
Traditional	4	173	N.A.	N.A.	-	-

Table 3: H_m and H_a for Two Optimization Strategies

The optimum driver sizes given in Table 3 are used to evaluate the propagation delays for the switching patterns related to Group 1, Groups 3 and 4, and Group 5. The propagation delays are shown in the Figure 9. Switching patterns in the Group 5 shows the maximum (worst-case) delay whereas that for Group 1 is the minimum while Group 3-4 is in between.

When the delay balancing technique is used the delay for Group 5 and Group 3-4 becomes equal and the delay for Group 1 increases a little reducing the delay variation, which is shown in the Figure 10. The reduction in jitter is 18% with the delay balanced sizing of SMART repeater.

5. CONCLUSIONS

With the ever decreasing feature size, the delay and energy dissipation of on-chip global level wires are posing great challenges in the nanometer regime. Usually repeaters are used to reduce their delay trading off some active area and power. Due to the static nature of the traditional repeater there is an excessive power loss in bus wire structures. We addressed this unnecessary power dissipation problem by dynamically changing the size of a repeater: The SMART repeater concept. Required worst-case delay based driver has been distributed among two drivers namely main and assistant. What we pay for all these power and jitter reduction is just some extra area for a logic circuit and the assistant driver. However, compared to a bus wire structure, the area for the extra gates are considerably small. With the same

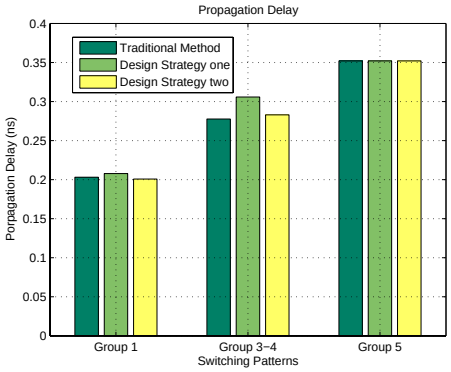


Figure 9: Propagation Delay

maximum delay as with traditional repeaters, SMART repeaters presented in this work shows a 15% energy reduction with a 18% jitter reduction.

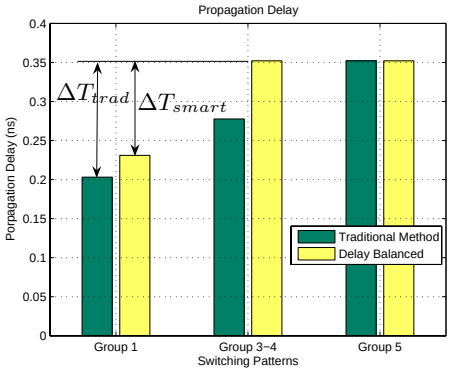


Figure 10: Propagation Delay using with balancing the delay

This work contained a detailed Elmore delay model for the SMART repeater and optimum sizes for Assistant and the Main drivers. Also, the delay balancing technique is also presented. In section 4 probabilistic average energy model is described for clear analysis of energy saving in SMART repeater concept. Unlike the other interconnect repeater concepts other than the original concept by Backoglu [2],this SMART repeater concept can easily be implemented in an interconnect planning CAD tool. However, a detailed circuit level implementation and practical design issues of the SMART driver, and it's usefulness for different applications will be discussed elsewhere.

Acknowledgment

The authors would like to acknowledge the financial support of SIDA under the auspices of the research capacity building project at the Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka.

6. REFERENCES

[1] M. Anders, N. Rai, R. K. Krishnamurthy, and S. Borkar. A transition-encoded dynamic bus technique for high-performance interconnects. *IEEE*

Journal of Solid-State Circuits, 38(05):709–714, May 2003.

[2] H. B. Backoglu. *Circuits, Interconnections and Packaging for VLSI*. Addison-Wesley, 1990.

[3] D. M. Claudia Kretzschmer, André K. Nieuwland. "why transition coding for power minimization of on-chip buses does not work". In *DATE '04: Proceedings of the conference on Design, automation and test in Europe*, pages 512–517, February 2004.

[4] M. Ghoneima and Y. I. Ismail. Utilizing the effect of relative delay on energy dissipation in low-power on-chip buses. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(12):1348–1359, December 2004.

[5] P. Gupta and A. B. Kahng. Quantifying error in dynamic power estimation of cmos circuits. In *IEEE Intl. Symp. on Quality Electronic Design*, pages 273–278, March 2003.

[6] K. Hirose and H. Yasuura. A bus delay reduction technique considering crosstalk. In *DATE '00: Proceedings of the conference on Design, automation and test in Europe*, pages 441–445, 2000.

[7] H.-Y. Huang and S.-L. Chen. Interconnect accelerating techniques for sub-100-nm gigascale systems. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(11):1192–1200, November 2004.

[8] T. Iima, M. Mizuno, T. Horiuchi, and M. Yamashina. Capacitance coupling immune, transient sensitive accelerator for resistive interconnect signals of subquarter micron ulsi. *IEEE Journal of Solid-State Circuits*, 31(4):531–536, April 1996.

[9] A. B. Kahng, S. Muddu, E. Sarto, and R. Sharma. Interconnect tuning strategies for high-performance ics. In *DATE '98: Proceedings of the conference on Design, automation and test in Europe*, pages 471–478, 1998.

[10] A. Katoch, S. Jain, and M. Meijer. Aggressor aware repeater circuits for improving on-chip bus performance and robustness. *European Solid-State Circuits, 2003. ESSCIRC '03. Conference on*, pages 261–264, Sep 2003.

[11] H. Kaul and D. Sylvester. Low-power on-chip communication based on transition-aware global signaling(tags). *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(05):464–476, May 2004.

[12] K.-W. Kim, K.-H. Baek, N. Shanbhag, C. L. Liu, and S.-M. Kang. Coupling-driven signal encoding scheme for low-power interface-design. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 318–321, November 2000.

[13] D. Liu and C. Svensson. Power consumption estimation in cmos vlsi chips. *IEEE Journal of Solid-State Circuits*, 29(6):663–670, 1994.

[14] N. Magen, A. Kolodny, U. Weiser, and N. Shamir. Interconnect-power dissipation in a microprocessor. In *SLIP '04: Proceedings of the 2004 international workshop on System level interconnect prediction*, pages 7–13, 2004.

[15] A. Nalamalpu, S. Sirinivasan, and W. P. Burleson. Boosters for driving long on chip interconnects-design issues, interconnect synthesis, and comparison with repeaters. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 21(1):50–62, January 2002.

[16] A. K. Nieuwland, A. Katoch, and M. Meijer. Reducing cross-talk induced power consumption and delay. In *PATMOS 2004 Proceedings, Lecture Notes in Computer Science*, volume LNCS 3254, pages 179–188, September 2004.

[17] K. Nose and T. Sakurai. Two schemes to reduce interconnect delay in bi-directional and uni-directional buses. In *VLSI Circuits, 2001. Digest of Technical Papers. 2001 Symposium on*, pages 193–194, 2001.

[18] D. Pamunuwa, S. Elassaad, and H. Tenhunen. Modeling delay and noise in arbitrarily coupled rc trees. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 24(11):1725–1739, November 2005.

[19] D. Pamunuwa, L.-R. Zheng, and H. Tenhunen. Maximizing throughput over parallel wire structures in the deep submicrometer regime. *IEEE transactions on Very Large Scale Integration (VLSI) Systems*, 11(2):224–243, April 2003.

[20] K. N. Patel and I. L. Markov. Error-correction and crosstalk avoidance in dsm busses. In *SLIP '03: Proceedings of the 2003 international workshop on System-level interconnect prediction*, pages 09–014, 2003.

[21] T. Sakurai. Perspectives on power-aware electronics. In *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC.*, volume 1, pages 20–26, 2003.

[22] Y. Shin, S.-I. Chae, and K. Choi. Partial bus-invert coding for power optimization of system level bus. In *ISLPED '98: Proceedings of the 1998 international symposium on Low power electronics and design*, pages 127–129, 1998.

[23] Y. Shin and T. Sakurai. Coupling-driven bus design for low-power application-specific systems. In *DAC '01: Proceedings of the 38th conference on Design automation*, pages 750–753, 2001.

[24] P. Sotiriadis, T. Konstantakopoulos, and A. Chandrakasan. Analysis and implementation of charge recycling for deep sub-micron buses. In *ISLPED '01: Proceedings of the 2001 international symposium on Low power electronics and design*, pages 364–369, 2001.

[25] P. P. Sotiriadis and A. Chandrakasan. Bus energy minimization by transition pattern coding (tpc) in deep sub-micron technologies. In *ICCAD '00: Proceedings of the 2000 IEEE/ACM international conference on Computer-aided design*, pages 322–328, 2000.

[26] M. R. Stan and W. P. Burleson. Bus-invert coding for low-power i/o. *IEEE Transactions on Very Large Scale Integration Systems*, 3(1):49–58, 1995.

[27] R. Weerasekera, L.-R. Zheng, D. Pamunuwa, and H. Tenhunen. Switching sensitive interconnect driver to combat dynamic delay in on-chip buses. In *PATMOS 2005 Proceedings, Lecture Notes in Computer Science*, volume LNCS 3728, pages 277–285, September 2005.