Reconfigurable Self-timed Regenerators for Wide-Range Voltage Scaled Interconnect

Jingcheng Wang, Nathaniel Pinckney, David Blaauw, and Dennis Sylvester
University of Michigan, Ann Arbor, MI, USA
{jiwang, npfet, blaauw, dmcs}@umich.edu

***Abstract* — A reconfigurable self-timed regenerator based global interconnect scheme enables graceful degradation of performance and power in wide range dynamic voltage/frequency scaled systems. A test chip demonstrates up to 40% and 25% better performance scaling than a traditional repeater based interconnect at 1V and 0.5V, respectively, in 45nm SOI CMOS.**

Keywords — DVFS, Interconnect, Near Threshold, Regenerator, Wide-Range Voltage Scaling

I. Introduction

Near-threshold (NT) operation has been shown to provide a reasonable balance between energy efficiency and performance demands for a wide range of applications [1-2], particularly in the mobile space. However, even with the recent focus on energy efficiency, high single-thread performance demands still dictate nominal voltage operation at times. Wide-range dynamic voltage and frequency scaling (DVFS) enables operation across the energy/performance design space, but requires underlying circuits to scale across voltage in a robust and predictable manner. Without this, the ability to adapt to dynamic runtime constraints will be limited.

Recent work has shown how to optimize logic [3-4] and memory [5] across both near-threshold and full voltage regimes. However, little work has addressed interconnect optimization across this wide voltage range. Unlike logic delay, which changes dramatically with supply voltage, interconnect RC delay is insensitive to voltage scaling. This leads to different optimization approaches in comparison to logic and memory. As designs are limited by their critical path, interconnections that are poorly optimized for certain voltage modes cause the entire design to suffer.

Optimal repeater insertion for a long interconnect differs significantly at full and near-threshold (NT) voltages. The optimal repeater count N_{opt} and size w_{opt} are given by the well-known equations [6] in Fig. 1. As supply voltage reduces, the effective repeater driver resistance R_d increases relative to the interconnect resistance r_w, which remains constant. Wire capacitance c_w and gate capacitance C_g also remain constant as voltage scales. Therefore, $N_{opt} \propto 1/\sqrt{R_d(1+\gamma)}$ and $w_{opt} \propto \sqrt{R_d}$, such that at low VDD fewer, yet larger, repeaters are optimal.

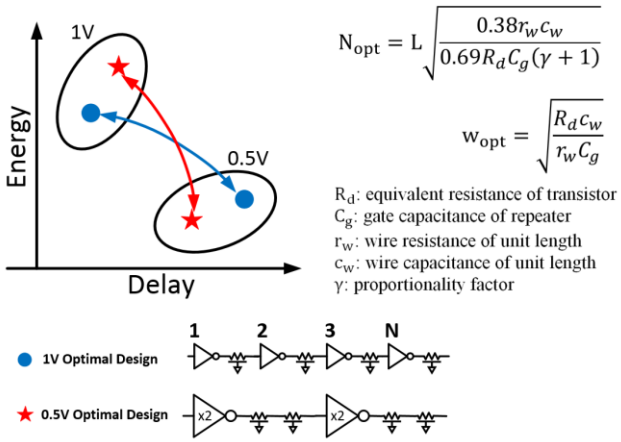

Fig. 1. Differing optimal repeater designs for high and low supply voltages lead to sub-optimality in wide-range voltage scaled systems.

For the 45nm SOI technology used in this work, nominal voltage is 1V while 0.5V can be considered near threshold, hence we consider this range during optimization. In this technology, R_d increases by roughly 4× from 1V to 0.5V, therefore an optimized interconnect at 0.5V uses half as many repeaters of twice the size as an interconnect optimized for 1V. Operating repeated interconnects at a voltage they were not targeted for leads to large sub-optimality in energy and delay, shown conceptually in Fig. 1.

On-chip interconnect has been studied in-depth by the circuit community with many specialized designs, such as low-swing transceivers, being proposed to save energy and increase throughput. However, within circuit blocks, long wires are repeated with inverters and buffers by commercial place and route tools. While specialized transceivers are desirable for well-defined interconnections spanning long distances, we propose using *regenerators* for shorter, within-block, wired interconnects in voltage scaled systems when simplicity, low overhead, and ease of integration into a design is valued over absolute performance and energy improvements. This proposed technique does not replace specialized interconnect techniques, but instead is meant to replace repeaters for general purpose use.

II. PROPOSED APPROACH

The poor voltage scalability of repeater-based interconnect currently forces the designer to choose between a design that is optimal at either full or NT voltages, but not both. Furthermore, the interconnect delay does not track the fanout-of-4 (FO4) inverter delay, characteristic of how digital circuits scale with voltage, and hence the interconnect will become performance-limiting for the entire design during either full or NT operation if traditional design methodologies are followed. SPECTRE simulations of industrial wire and device models provided by a 45nm foundry are shown in Table I with results matching the analytical predictions of Figure 1. The baseline repeaters were inverters in this simulation. As expected, NT favored fewer, larger repeaters as compared to nominal voltage.

TABLE I.
SIMULATED OPTIMAL REPEATER DESIGN

VDD (V)	Optimal Delay (ps/mm)	Optimal Size w_{opt} (µm)	Optimal # N_{opt}
0.5 (NT)	1680	12.6	35
1.0 (Nom.)	740	6.3	49

*Interconnect configuration: 10mm length with minimum width and spacing.

An obvious approach to overcome the N_{opt} discrepancy between VDD and NT operation is to selectively disable repeaters along an interconnect. However, this only shifts the problem from drivability of the repeater to drivability of the bypass devices, amounting to a zero sum game. For instance, if transmission gates are used to bypass repeaters then they suffer similar R_d degradation to that of the repeater, unless driven by a separate nominal voltage supply, which incurs considerable level shifting and power delivery overheads.

We propose using single-ended regenerators based on [7-8] which, unlike conventional repeaters, are single-ended gates attached along a wire. Instead of discrete input and output pins, regenerators rely on detection circuits to sense partial transitions along the wire, triggering a temporary regenerative drive of the wire until it has fully transitioned to a new value. Regenerators have the unique property of not partitioning a long interconnect into separate wire segments. If a regenerator is enabled, it acts as a repeater passively monitoring the interconnect and then actively driving it to transition. Disabling the regenerator in effect extends the repeated distance, as the inactive regenerator does not change the characteristics of the wire other than added parasitic capacitance. Using regenerators addresses the scalability of the number of inserted repeaters, but to address repeater size we also add regenerators in parallel and selectively enable them.

Fig. 2 shows a circuit schematic for our proposed regenerator, named Reconfigurable Self-Timed Regenerator (RSTR), which is based on [6] but with extensions for reconfiguration. The new reconfigurable components are highlighted in red.

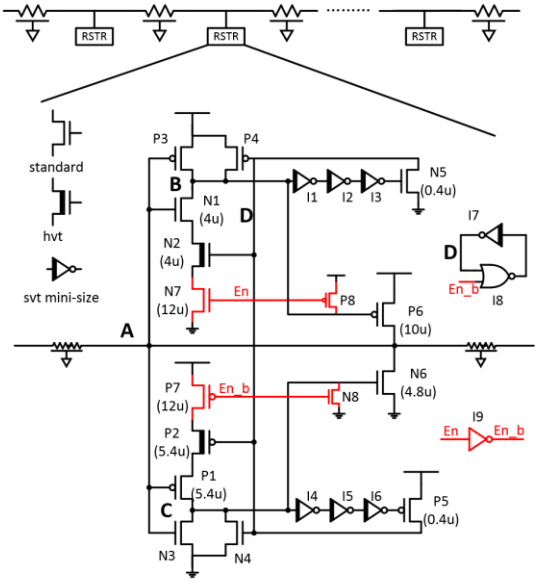

Fig. 2. RSTR schematic with transistor sizing. Transistors with unlabeled sizes are minimum width (152nm). Enable signal and header/footer transistors provide reconfigurability.

The circuit operates by early detection of a transition along the interconnect wire at point A. The transition is then aided by turning on either the PMOS or NMOS driving transistor, *P6* and *N6*, to supply additional current in driving the wire. To avoid global control signals a self-timed delay chain (*I1-3* and *I4-6*) turns off the driving transistors and awaits the next transition. The regenerator is enabled through the *En* signal that, when asserted, activates *N1-2* and *P3-4* forming a NAND structure to sense the low-to-high transition and turn on driver *P6*, while remaining insensitive to high-to-low transitions. Similarly, high-to-low transitions are detected by a NOR (*P1-2*, *N3-4*) that controls *N6*. To allow for this hysteresis, *I7* and *I8* form a latch to store the previous value on the wire. Lastly, *N7* and *P7* in the NAND/NOR detection circuits disable the sensing of transitions while *P8* and *N8* disable the output drivers.

Because of the internal delay chain, RSTR controls its own pulse width, namely the duration of the pull-up/pull-down time, hence careful delay selection is needed to ensure that the wire transitions substantially before the RSTR resets itself across a range of Vdd. Also the delay should not be so long that it interferes with the next signal transition. The delay chain consists of three SVT minimum-sized stacked inverters; simulation across design corners and process variation ensures all these requirements are met.

Fig. 3 shows the energy-delay curve for repeaters and

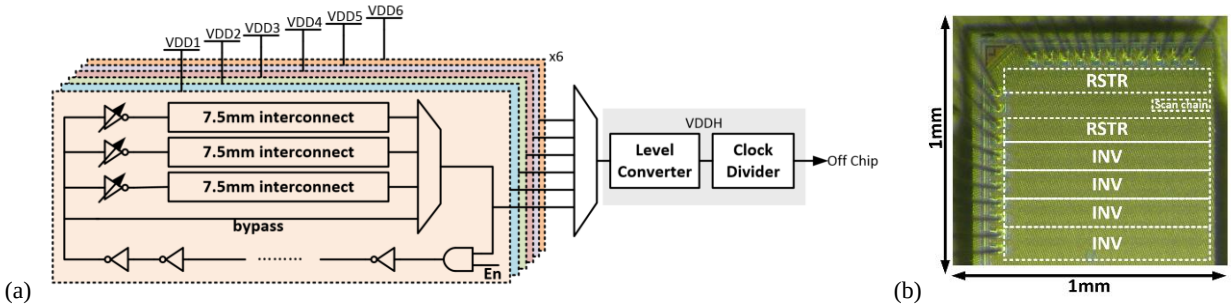

Fig. 4. (a) Reported delays are measured based on the frequency of a ring oscillator structure. Each interconnect design is in a separate voltage domain to measure energy. Each interconnect under test has adjacent neighbors with 140nm spacing (1× min.). (b) Die photo of 45nm SOI test chip. The 7.5mm interconnect is folded ten times.

RSTR at 0.5V and 1V, simulated with the industrial 45nm SOI CMOS models. The driven interconnect is a 7.5mm intermediate (2× thickness) wire with 140nm spacing (1× min.) and 280nm width (2× min.), chosen to represent a reasonably long within-block interconnect. At both voltages, the size and number of repeaters are swept to find the optimal energy/delay points, marked as the Pareto frontier curve in Fig. 3. On the 1V frontier, we chose "INV #23" to represent the 1V-optimized design containing N_{opt}=23 inverter repeaters, each with size w_{opt} = 12µm PMOS and 6µm NMOS. On the 0.5V frontier, design "INV #9" is selected with N_{opt}=9 inverters (w_{opt} = 24µm PMOS, 12µm NMOS).

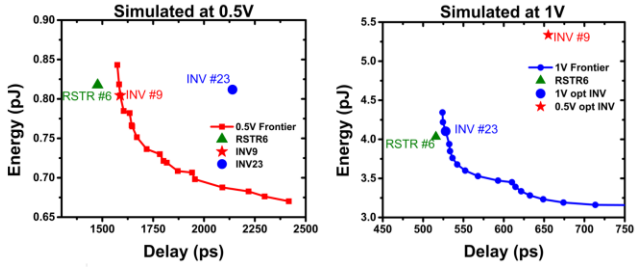

Fig. 3. Simulated energy versus delay curves for RSTR. Optimal inverter and RSTR designs are chosen from the frontier curves at each voltage.

The RSTR design space is similarly swept and we observed that some configurations on its 1V frontier also appeared on the 0.5V frontier. One such design "RSTR #6" uses N_{opt}=6 RSTR along the 7.5mm wire with device sizes given in Fig. 2. This RSTR design is labeled on both plots of Fig. 3 for comparison. Unlike traditional repeated interconnects, RSTR can achieve better performance and energy characteristics over a wide voltage range, such as 0.5V to 1V as demonstrated in this simulation.

Despite the simplicity of the proposed RSTR scheme (the regenerator topology adds only small overhead beyond the design in [7]) it provides the following important benefits over traditional repeated interconnects:

1) RSTR remains optimal (in energy/delay space) across the full VDD range.

2) RSTR reconfigurability provides a new knob for adaptive designs to compensate for variability at NT operation. This is achieved by selectively turning on/off RSTRs along a wire to trade performance for power (e.g., 24% performance loss for 40% lower energy).

3) RSTR is faster than an optimal repeater design at both full and NT supply while maintaining energy efficiency.

4) RSTR does not partition the wire, allowing for bi-directionality.

III. MEASUREMENTS AND RESULTS

A test chip was fabricated in 45nm SOI CMOS to evaluate the efficacy of RTSRs in silicon and validate simulation predictions. A total of four inverter repeater (INV) designs and two proposed RSTR designs were included on the test chip, which measured 1×1mm (Fig. 4). Fig. 5 shows the test harness; the interconnect matches the structure simulated above and is implemented as a bypassable delay chain within a ring oscillator. After level conversion and a clock divider, frequency is measured off chip both with and without interconnect to assess delay.

Fig. 6 shows measured results confirming the relatively poor voltage scalability of repeater-based designs. A 1V optimal design is 31% slower than the 0.5V optimal design when operating at 0.5V. Conversely, a 0.5V optimal design is 18% slower with 29% higher energy than a 1V optimal design when both operate at 1V. In contrast the RSTR design shows good voltage scalability. Specifically at 1V it is 28% faster than the 1V optimal INV design while consuming 5% less energy. At 0.5V, the "RSTR #6" energy and delay essentially match the 0.5V optimal INV design. In addition to being superior to INV-based designs, recall that "RSTR #6" appears along the Pareto optimal frontier at both supply voltages. This indicates that excellent performance can be obtained across voltage scaling, relative to other RSTR designs.

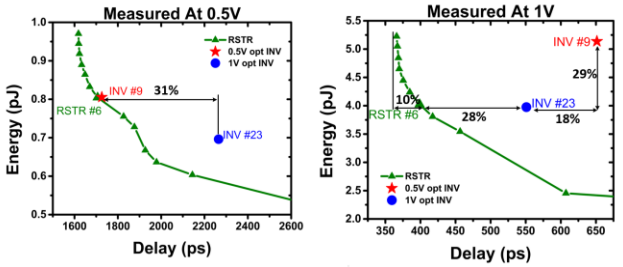

Figure 6. Measured energy versus delay curves showing RSTR and repeater performance. Green triangles represents different RSTR configurations (i.e., different number of RSTR enabled).

Green triangles in Fig. 6 represent RSTR energy-delay points with varying number of RSTR enabled, representing dynamic reconfiguration options depending on real-time energy-performance priorities. This allows the RSTR design to also operate at lower energy with faster delay than "INV #23" at 0.5V, if desired. Also, if interconnect was performance limiting for the design at full VDD (1V), turning on six additional RSTR along the wire (reconfiguring RSTR #6 into RSTR #12) offers 10% faster performance, potentially rebalancing the overall design. In NT mode (0.5V), regenerators can then be turned off to achieve a minimal energy of 0.6pJ in this example.

Fig. 7 shows measured delay scaling of repeater and RSTR designs across VDD, indicating the sub-optimality of using a single inverter-based repeater design in wide-range voltage scaling. RSTR is able to achieve better performance across the entire 0.5V to 1V range. Fig. 8 plots this measured data normalized to inverter FO4 delay across a range of voltages. Ideally an interconnect scales identical to circuit delay, which would be shown as a fixed line at 1.0 of FO4 in Fig. 8. Again, this supports the more graceful scaling of delay offered by an RSTR design over a conventional repeater-based approach.

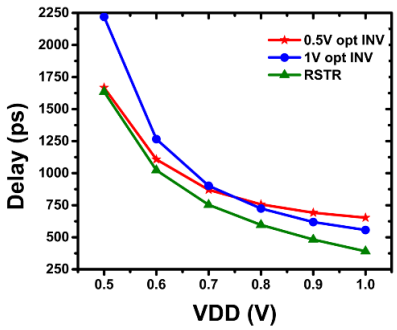

Figure 7. Measured delay dependency on supply voltage shows better scalability of RSTR.

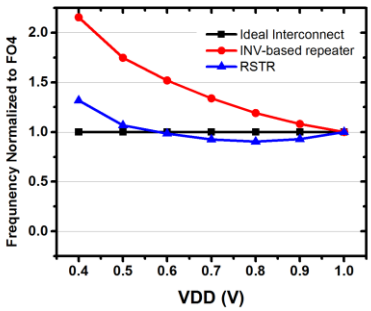

Figure 8. RSTR speed scales more similarly to digital logic than inverter-based repeated wires.

IV. Conclusion

Today's emerging mobile applications require high energy efficiency, which is often provided by scaling supply voltage across a wide range according to real-time workload variation. We present a reconfigurable, self-timed, regenerator-based interconnect scheme that remains optimal in terms of energy-delay efficiency at both full and near-threshold voltages. RSTR interconnect delay tracks FO4 logic delay more closely than repeated wires. In addition, RSTR offers higher speed and better energy efficiency overall compared to traditional repeater approaches.

References

[1] R.G. Dreslinski et al., "Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits", *Proc. Of IEEE*, Feb. 2010.

[2] S. Jain et al, "A 280mV-to-1.2V Wide-Operating-Range IA-32 Processor in 32nm CMOS", *IEEE International Solid-State Circuit Conference*, pp. 66-67, 2012.

[3] S. Hsu et al, "A 280 mV-to-1.1 V 256b Reconfigurable SIMD Vector Permutation Engine With 2-Dimensional Shuffle in 22 nm Tri-Gate CMOS", *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 118-127, 2013.

[4] G. Chen et al, "A 340mV-to-0.9V 20.2Tb/s Source-Synchronous Hybrid Packet/Circuit-Switched 16×16 Network-on-Chip in 22nm Tri-Gate CMOS" , *IEEE International Solid-State Circuit Conference*, pp. 276-277, 2014.

[5] M. Chen et al, "A 260mV L-shaped 7T SRAM with bit-line (BL) Swing expansion schemes based on boosted BL, asymmetric-VTH read-port, and offset cell VDD biasing techniques", *Symposium on VLSI Circuit*, pp. 112-113, 2012.

[6] N. Weste et al, "CMOS VLSI Design: A Circuits and Systems Perspective, Fourth Edition", pp. 232, 2010.

[7] P. Singh et al, "Self-Timed Regenerators for High-Speed and Low-Power On-Chip Global Interconnect", *IEEE Trans. VLSI Syst.*, vol. 16, no. 6, pp. 673-677, 2008.

[8] A. Nalamalpu et al, "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, 2002.