

A Charge Pump Based Receiver Circuit for Voltage Scaled Interconnect

Aatmesh Shrivastava
University of Virginia
Charlottesville,
VA, USA
as4xz@virginia.edu

John Lach
University of Virginia
Charlottesville,
VA, USA
jlach@virginia.edu

Benton H. Calhoun
University of Virginia
Charlottesville,
VA, USA
bcalhoun@Virginia.edu

ABSTRACT

This paper presents a charge-pump based low swing interconnect receiver circuit. The interconnect circuit is single ended and supports swings of 300mV or lower. A charge pump front end at the receiver boosts the arriving signal before restoring it to the full logic level, improving the performance of the interconnect. For a 10mm long interconnect wire in a 45nm CMOS process, the proposed scheme provides 3X energy reduction at constant speed and 3.5X delay improvement at constant energy relative to prior art. We deploy the interconnect scheme as the data bus between the L1-L2 caches of a 4-core Alpha processor. Over a set of Splash benchmarks, the proposed architecture reduces total energy consumption by 70% while maintaining the same performance.

1. INTRODUCTION

Studies have shown that 50% of total chip power is dissipated in interconnect wires and circuits in a modern microprocessor [2]. This number is close to 90% for reconfigurable architectures like FPGAs [3]. Interconnect power is going to become even a bigger concern for exascale computing [1] where 10 billion transistors are expected to be present in one square centimeter of chip area. Over the past decade, voltage scaling has been employed to reduce the power of interconnects [4-10]. In a voltage scaled interconnect, the interconnect wire is driven at a much lower voltage than the logic. A receiver circuit converts the low swing signal on the interconnect back to the full swing logic level. Various architectures for the interconnect driver and receiver have been proposed in the literature. These can broadly be categorized as single ended, differential, and capacitive interconnects. Table 1 shows an approximate energy-delay comparison of interconnect circuits reported in the literature. The basic and differential interconnect show the best performance, but they have higher

Table 1: Energy and Delay of existing interconnects

Schemes	Speed (GHz)	Swing (V)	Normalized Energy
Basic	>1	1	1
Single-ended [4,5,7]	<0.25	0.6	0.6
Differential [8-10]	>1	0.05	0.8
Capacitive [6]	<0.25	0.05	0.2

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'12, July 30– August 1, 2012, Redondo Beach, CA, USA.
Copyright 2012 ACM 978-1-4503-1249-3/12/07...\$10.00.

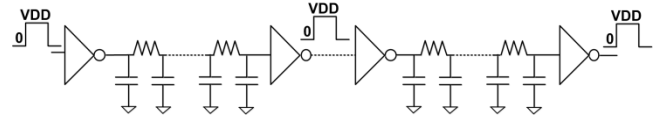


Figure 1. Basic interconnect circuit

energy consumption. The basic interconnect does not employ voltage scaling, and therefore consumes higher energy, while differential interconnects [8-10] use a differential amplifier and two wires per interconnect signal, which increases the energy consumption. The single ended interconnects [4-5,7] show reduced energy consumption, but their performance is poor because the lower input swing reduces drive current in the receiver. The capacitive interconnect schemes (e.g., [6]) have driver and receiver circuits capacitively coupled to the wire using series capacitors. The charge distribution between the wire and capacitor reduces the swing, which saves power. The receiver circuit in this scheme is either differential or single ended. The best reported work here [6] claims a bandwidth of 250MHz, much lower than the desired on-chip signal rate of a GHz or higher.

One of the primary reasons for the lower performance of single ended interconnects is the lower voltage swing at the receiver input. In this paper, we employ a charge pump to increase the swing at the receiver's input. The receiver sees three times the interconnect swing voltage at its input. This saves power without impacting performance.

2. PREVALANT INTERCONNECTS

Figure 1 shows the most basic type of interconnect architecture also known as CMOS interconnect. It does not employ voltage scaling. The metal interconnect between two points inside a chip can be approximated as a distributed π -RC network as shown Figure 1. Its delay increases quadratically with length due to Elmore delay. Repeaters are inserted at regular intervals to obtain the optimal delay point. We simulated a π -RC model of a 10mm long wire in 45nm CMOS with repeaters. By controlling the number of repeaters in the path, either a minimum delay or a minimum energy point can be achieved. However the overall energy consumption of this interconnect architecture is high. Voltage scaling has been employed to reduce this power in [4-10].

2.1 Voltage-scaled Interconnect Architectures

Figure 2 shows the architecture of a voltage scaled interconnect.

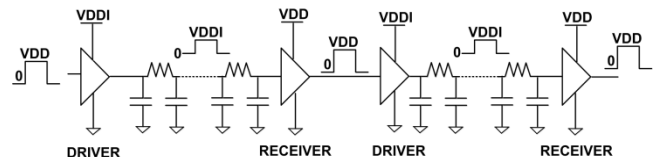


Figure 2. Voltage-scaled interconnect circuit

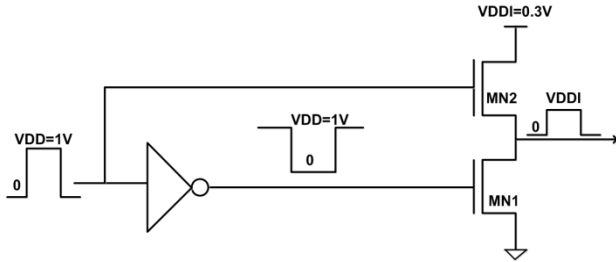


Figure 3. Asymmetric source follower driver [4]

It consists of an interconnect driver that converts the signal at the logic level (V_{DD}) to the lower interconnect voltage (V_{DDI}) and a receiver that converts the signal back to the logic level. The driver-receiver pairs are called repeaters as shown in Figure 2.

Interconnect Driver: Figure 3 shows an asymmetric source-follower driver [4], which is commonly used for low voltage interconnects. It uses NMOS devices to drive both *high* and *low* voltages. When *high* is to be passed, MN1 is off and MN2 acts as a source follower to pass V_{DD1} to the interconnect. The delay of transmitting a *high* through the NMOS is not increased because MN2 is over-driven. When a *low* is to be passed, MN2 is off, and MN1 turns on as in a regular inverter. This architecture is useful if the interconnect voltage is scaled down to very low voltages in the range from about 0.1V to 0.5V.

Interconnect Receiver: Receiver design plays an important role in the overall performance of interconnects. Figure 4 shows the energy-delay points of different interconnects reported in the literature for a 10mm long wire. We scaled the data accordingly to compare with the 45nm CMOS process with nominal operating voltage at 1V. The single ended interconnect implementations use a source follower receiver [4][5][7]. The performance of these receivers is poor because of the source degeneration. Also the interconnect signal should be at least $2 \cdot V_T$ for the receiver to operate correctly. This puts a limit on the amount of power that we can save using voltage scaling. These interconnect circuits fall on the right side of Figure 4. Usually, basic interconnect or differential interconnect [8-10] is used for higher performance. These schemes form the left side of Figure 4. They can have higher performance than the basic interconnect, but that costs additional energy. We find that existing solutions do not improve power and performance at the same time. There is a need to bring down the interconnect power while maintaining the performance [1]. The proposed interconnect circuit addresses this issue and gives a high performance yet lower power interconnect.

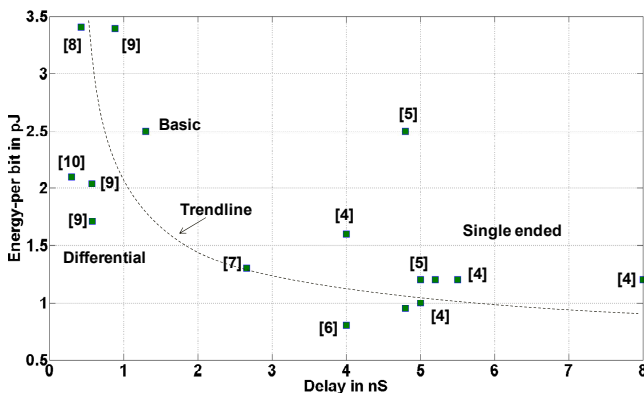


Figure 4. Energy-Delay points of interconnects in literature

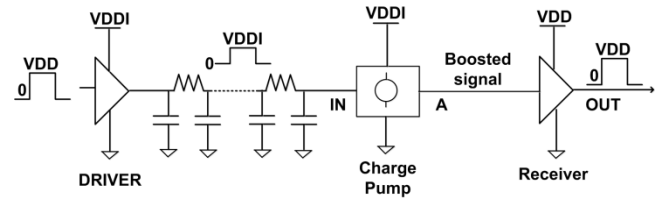


Figure 5. Proposed interconnect architecture

3. PROPOSED INTERCONNECT CIRCUIT

Figure 5 shows the block diagram of the proposed interconnect scheme. It can operate at an interconnect swing of 300mV or lower. We use an asymmetric source follower driver [4], which is well suited for this case. The receiver is novel and uses a charge pump, which boosts the signal and improves its performance. We will discuss each of these components in turn.

3.1 Charge-Pump

Figure 6 shows the charge pump. It has two capacitors (C_{CH} and C_{CL}) connected in series with the interconnect wire, IN. Nets A and C are dynamic nodes. Net A is controlled by NMOS MN4 and MN5. These transistors turn on for a very small duration of time to set the voltage at A to 0.3V or 0V. This voltage is then dynamically held at A. Net A is precharged to 0.3V before IN makes a high transition. As IN goes *high*, A gets charged to 0.6V. Similarly, A is precharged to 0 when IN is at 0.3V. Net A goes to -0.3V as IN makes a transition to zero. Therefore the overall swing at A is boosted to 0.9V. Similarly, C has a PMOS connected to ground and an NMOS to V_{DD} . It can swing from V_T to $V_{DD}-V_T$. It makes a transition from V_T to $V_T+0.3V$ when IN goes *high* and $V_{DD}-V_T$ to $V_{DD}-V_T-0.3V$ when IN goes *low* through the capacitive coupling of C_{CH} .

The series capacitors are implemented using NMOS in an Nwell. A $1.5\mu\text{m}$ by $1\mu\text{m}$ capacitor gives roughly 35fF of series capacitance, which was good enough to get the desired swing at A and C. The control and precharging of A and C is implemented using feedback discussed in Section 3.2. The proposed interconnect receiver circuit uses the charge pump and a pulse generator circuit in a feedback manner to properly set the voltage at A. This assists in receiving the incoming low swing signal and converting it to full swing. In the following section, we will explain the functioning and the circuit diagram of the receiver.

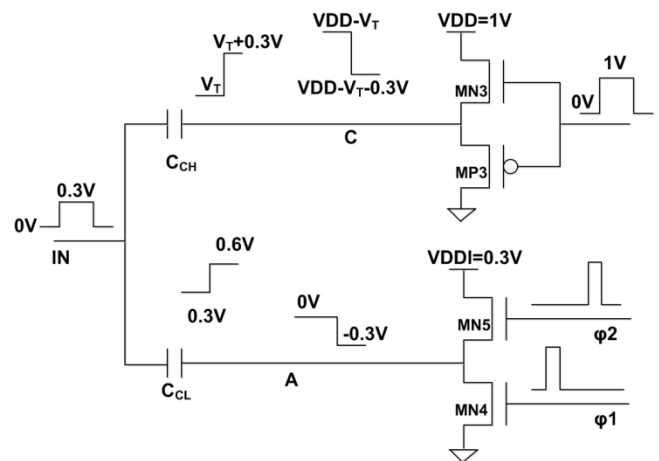


Figure 6. Proposed charge-pump circuit

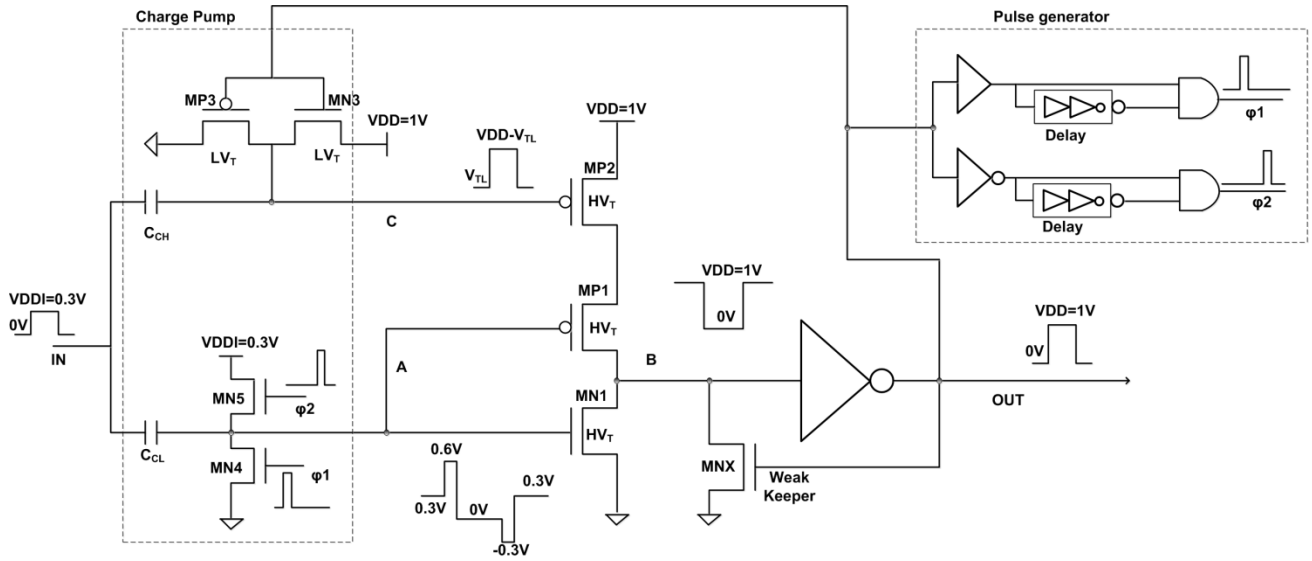


Figure 7. Proposed interconnect receiver circuit with charge-pump based pre-amplification

3.2 Proposed Interconnect Receiver

Figure 7 shows the full circuit diagram of the proposed receiver. It consists of a cascoded PMOS inverter (MP1, MP2, MN1) using HV_T transistors, a regular inverter, and a positive feedback NMOS transistor (MNX) along with the charge pump made of LV_T transistors and pulse generator circuit. Figure 8 shows the associated timing diagram of the receiver. We will explain the functioning of the circuit with the aid of this idealized timing diagram.

Transmitting HIGH: Consider the case at time=0, when $IN=0$.

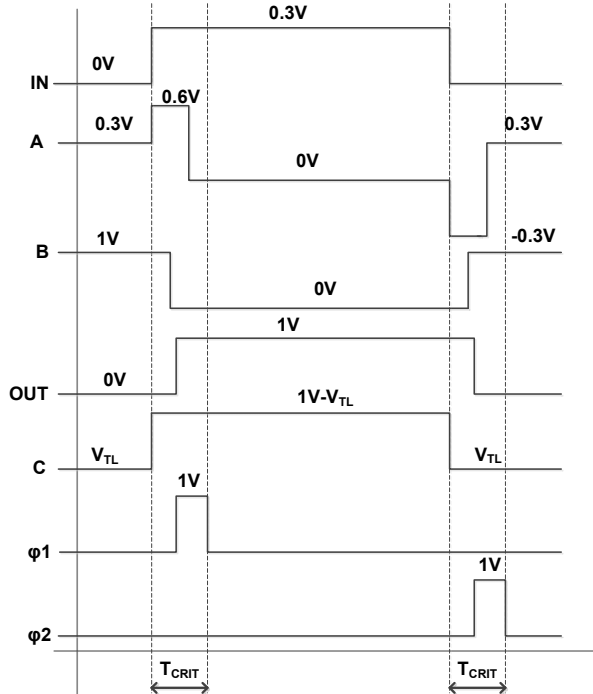


Figure 8 Timing diagram of the receiver

At this point, OUT should equal 0. Assuming $OUT=0$ and therefore $B=1$, so Net C should be equal to V_{TL} . Also, we assume that Net A is charged to $V_{DDI}=0.3V$. Since $A=0.3V$ and $C=V_{TL}$, MP1 and MP2 are on while MN1 is close to off, which means that our assumption for $B=1$ is consistent with this circuit. Now consider the case when IN goes high to 0.3V from 0V. Net A gets charged to 0.6V through the capacitive coupling of C_{CL} . Net C gets charged through capacitor C_{CH} to $V_{TL}+0.3V$. This causes the current drive of transistor MN1 to increase, while the drive of MP1 and MP2 decreases. The change in drive strength causes Net B to discharge to ground. As B goes to ground, OUT gets charged to 1V. Also, as OUT goes to 1V, MNX turns on and keeps B at ground. MNX is a weak keeper transistor and holds the state at net Net B in the absence of any other signal driving it. At this point, we have propagated a low to high transition through the receiver.

As OUT goes high, two more transitions take place. Net C will get charged to $V_{DD}-V_{TL}$ through MN3 in the charge pump, which turns off MP2. Also, the pulse generator circuit produces a small pulse at ϕ_1 , which pulls A to ground. Once ϕ_1 goes to ground, Net A remains charged to 0, but it becomes high impedance. The state at B is maintained by MNX through the positive feedback.

The delay from IN going *high* to ϕ_1 going *low* is called the critical delay, T_{CRIT} . It limits the maximum operating frequency of the circuit since the receiver is not ready to receive a low before ϕ_1 goes low.

Transmitting LOW: At this point, we have $OUT=1V$, $C=1V-V_{TL}$, $A=0$, and $B=0$. Now consider the case when IN goes to 0 from 0.3V. Net A will go to -0.3V through capacitor C_{CL} , and Net C will go to $V_{DD}-V_{TL}-0.3V$ through capacitor C_{CH} . This will turn on MP2 and increase the drive strength of MP1. The increased drive will overcome MNX (MNX is a weak keeper) and put a 1V at B, making OUT to go to 0V as shown in Figure 9. As OUT goes to 0V, C will be pulled to V_{TL} through MP3. At this point, the pulse generator circuit produces a pulse at ϕ_2 that pulls A to 0.3V making the receiver ready to receive a *high*.

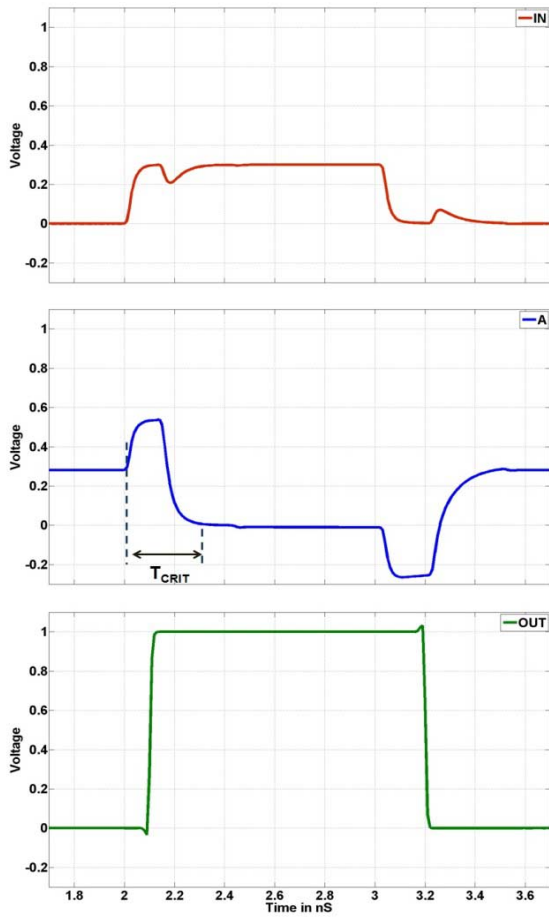


Figure 9. Simulation results of the receiver

Simulation: The interconnect circuit was designed in a 45nm CMOS process. We used HV_T and LV_T transistors and used NMOS in Nwell capacitors. The wire used is a distributed π -RC network that models the wire in our 45nm CMOS process. We selected a 10mm long wire and introduced the proposed interconnect driver and receiver circuits at regular intervals. The driver circuit reduces the swing from the logic level (1V) to the interconnect level (0.3V) and launches it to the long interconnect wire. The receiver circuit receives this signal and converts it back to the logic level. Figure 9 shows a typical simulation result in 45nm CMOS for key signals. Typical delay from A to OUT is close to 80ps. The receiver circuit gave the desired performance at

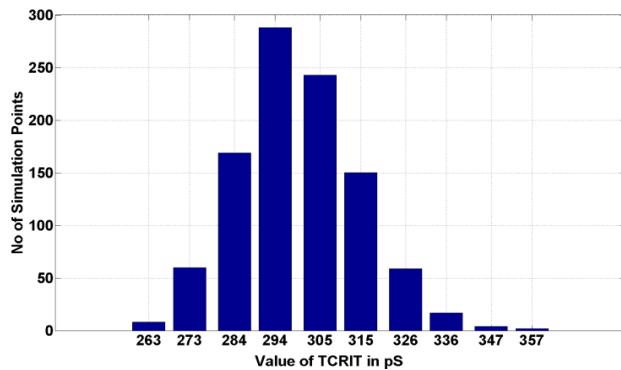


Figure 10. Histogram of T_{CRIT} from Monte-Carlo simulation

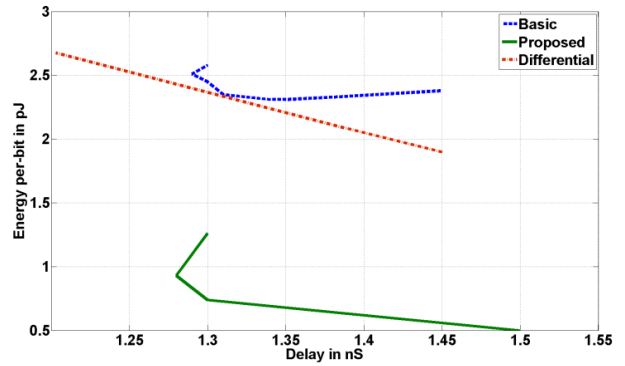


Figure 11. Energy-Delay curve of the proposed interconnect

fast, slow, typical, and skewed corners. Net A does not reach the ideal voltages shown in Figure 9 because of charge sharing, but it is good enough to obtain the desired performance. The small kink in the waveform of IN is caused by charge coupling from Net A. The T_{CRIT} of the circuit is 300ps, setting the maximum operating frequency at ~ 3 GHz. Figure 10 shows the histogram of T_{CRIT} obtained from 1000 local mismatch Monte-Carlo simulations of the circuit. The charge pumping technique can ideally make Net A of Figure 7 swing up to $(0.6+0.3V)$ 0.9V. Increased swing at A comes from the charge-pumps and not by increasing the interconnect swing, as was done in [4-5]. The lower voltage swing on the wire reduces energy dissipation, while the elevated swing at the receiver level ensures good performance.

Figure 11 shows the energy delay curve of the interconnect scheme obtained by varying the number of repeaters. We also used the differential interconnect of [8] and basic interconnect for comparison. Other single ended interconnects in literature cannot be used for this experiment owing to their high delay. We chose the same 10mm long wire discussed earlier. The proposed circuit has much lower energy at the same performance points as that of differential or the basic interconnect.

Initial conditions and leakage: The proposed circuit needs to be initialized properly before we can start transmitting data. This is because of the dynamic nature of the circuit. For example, Net A should be at 0.3V before the receiver can receive a *high*. At the very beginning this cannot be ensured by the feedback through the pulse generator. A RESET signal is used to initialize the voltage of the dynamic nets. Similarly, if the interconnect is left idle for very long, leakage can affect the dynamic nets. These nets are refreshed at a regular interval of time to ensure fidelity. We needed to refresh the circuit if it is idle for more than 100 μ s using the same RESET signal. Figure 12 shows the initialization

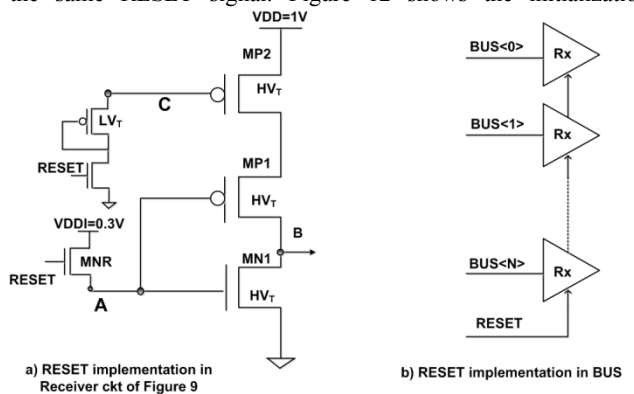


Figure 12. Initialization scheme for the receiver circuit

scheme using RESET. When RESET is held high, Net A is forced to 0.3V while Net C is taken to $\sim V_{TL}$, making B go high, which sets OUT to 0 (cf. Figure 7). This makes the circuit ready to receive a high. The RESET signal is low frequency, and only one signal is needed per bus. Therefore it does not contribute a significant area or power overhead.

Static current: The interconnect circuit consumes a small amount of static current that is present because pull-up and pull-down paths are not fully disabled. Figure 13a shows the first stage of the receiver circuit. When B is high, MP2 is connected to V_{TL} while MN1 is connected to 0.3V. MN1 is not completely cut off, which causes static current. However since the high V_T (V_{TH}) is greater than 0.3V, the static current is very small. Similarly, when B is low, MP2 is at $V_{DD}-V_{TL}$ which will result in static current. However $V_{TH} > V_{TL} + (-0.2V)$ ensures that leakage current is small.

Figure 13b shows the Monte-Carlo simulation result of the leakage in the receiver circuit. The simulation was performed at 30°C. The maximum leakage is less than 1μA, and average leakage is around 100nA. A basic interconnect receiver made of inverters has leakage in the range of $\sim 1nA$. The LHOS receiver of [5] will have leakage of $\sim 200nA$, while the HOA [5] will have leakage in the range of $\sim 1nA$. Static current is also present in differential interconnect receivers [8-10] in the form of bias current, which ranges from a few 100μA to a few mA. The leakage current in our receiver is an overhead that increases power consumption if the interconnect is not switching. However, as switching activity on the interconnect increases, this power will become insignificant. At 1GHz, for a 10mm long wire the switching energy of the proposed interconnect is 0.8pJ/bit, while leakage is 0.1fJ/bit. Energy benefits can be realized at switching activity of 0.03% and above. Later in the paper we show energy benefits in a real system. Also, if the interconnect is idle for a long time, then it can easily be power gated to save this power.

Voltage sensitivity: The interconnect circuit is sensitive to the variation in V_{DDI} . An increase in V_{DDI} increases the leakage in the first stage of the receiver as explained in previous section. We simulated the circuit with $V_{DDI}=0.35V$ and Figure 14a shows the Monte-Carlo simulation results across process. The average leakage increases to 316nA and maximum leakage goes to 1.5μA. However, this is still a small overhead when compared to the overall energy savings. In the other case when V_{DDI} goes lower, the drive to MN1 goes low, causing the receiver to lose performance. We simulated the circuit at $V_{DDI}=0.25V$ and measured the propagation delay from IN to OUT (Figure 7).

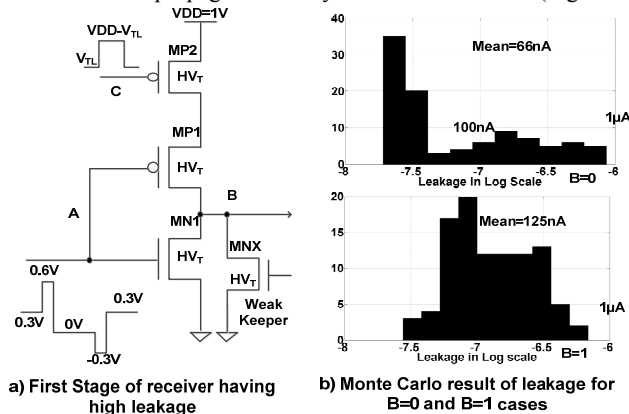


Figure 13. Static current consumption in the Receiver

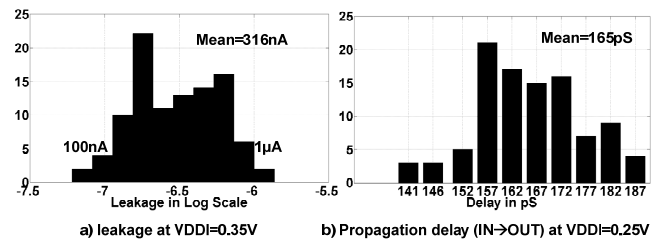


Figure 14. Leakage and delay with varying V_{DDI}

Figure 14b shows the Monte-Carlo simulation result. The receiver performance drops and average delay goes to 165ps. However, this is not a significant drop in performance because the overall delay of the interconnect will be dominated by wires and is close to 1ns for 10mm long wires. These simulations show that the proposed receiver circuit performs well for V_{DDI} varying from 0.25V to 0.35V (30% variation in V_{DDI}).

Parasitic and negative voltage on net A: The voltage seen at Nets A and C depends on the value of the series capacitance and input capacitance seen at those nets. Parasitics can increase the capacitance, reducing the swing seen at A and C. Increasing the value of series capacitances will make sure that the swing is not attenuated at these nets because of the parasitics. Note that this will only increase the area and will not affect the power. We simulated the circuit with 5fF of additional parasitic cap at both the nets A and C. All the simulation results in this paper include an additional 5fF of parasitic load on A and C.

Another concern can be the negative voltage of -0.3V at Net A, which can turn on the body diode of MN4 in Figure 7. However, the cut-in voltage for the body diode ranges from 0.5V-0.7V, and a voltage -0.3V will not result in a significant reliability issue. The duration of this negative voltage is small too.

Noise Performance: The interconnect circuit has better noise performance than alternative receivers. To understand the noise performance, let us consider the cases when the receiver circuit receives a *high* and a *low*. The receiver is designed to receive a *high* when Net A makes a transition from 0.3V to 0.6V. Therefore V_{IH} of the receiver can be anywhere between 0.3V to 0.6V; suppose for example that it is at 0.45V. Similarly, V_{IL} can be between 0 to -0.3V; suppose for example that it is at -0.15V. Therefore, the total hysteresis of the receiver can be $(V_{IH}-V_{IL})$ 0.6V. The worst case hysteresis is 0.3V. The receiver can tolerate a noise of 0.3V on A or equivalent 0.1V on IN. The high hysteresis in the receiver is produced because of the feedback path and charge-pumping technique. Most differential and single ended receivers in the literature do not have any hysteresis. Some receivers [4] [5] have hysteresis of 50mV or lower.

The interconnect circuit consumes approximately three times lower energy than the prevalent interconnect circuits at the same performance points. The use of a charge pump circuit in the receiver enables this energy benefit without any significant performance penalty.

4. RESULTS

Figure 15 compares the proposed interconnect with existing architectures. [4-5,7-8] have results based on simulation while [6][9-10] have silicon results. The differential interconnects [6, 8-10] use a single supply, and the interconnect swing is restricted by IR drop in the diffamp. [4-5][7] are single ended interconnects. In

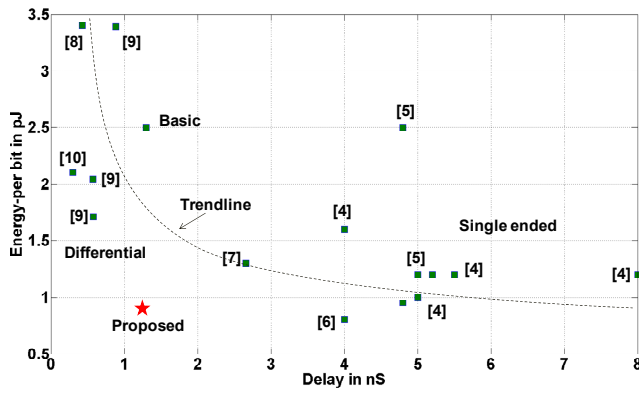


Figure 15. Proposed work in comparison with prior art

[4], authors present multiple circuits with different input swing on the wires. They use one logic supply and one or more interconnect supply voltages for their circuits. [5] and [7] present circuits with only one supply, and the interconnect wire swings from V_{DD} to $V_{DD}-V_T$, resulting in higher swing and hence higher power.

The proposed interconnect circuit uses a dedicated interconnect supply along with a supply voltage for the logic. The circuit has the best energy number and achieves very high performance. Table 2 compares the proposed circuit with existing architectures

Table 2: Energy, Delay and area of interconnect

Schemes	B/W (GHz)	Swing (V)	Norm. Energy	Area of 1 repeater
Basic	>1	1	1	2X
S-E [4,5,7]	<0.25	0.6	0.6	15-24X
Diff [8-10]	>1	0.05	0.8	100-250X
Cap[6]	<0.25	0.05	0.2	NA
This Work	>1	0.3	0.3	22X

We used the proposed interconnect to design the data bus connecting the L1 and L2 caches of a 4 core Alpha processor. Each core has a local L1 cache, while L2 is shared among all cores. The data bus between L1 and L2 will form a long interconnect, which makes the case for our experiment. We simulated the Alpha using m5 [11] and a spice model for the interconnect circuits implementing the data bus inside the Alpha.

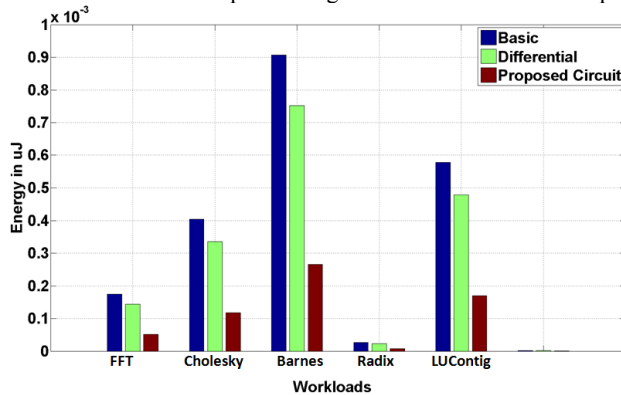


Figure 16. Interconnect energy dissipated in the data bus while simulating Splash workloads on an Alpha processor

We ran different Splash workloads to see the actual energy consumption in the interconnect during operation of a real processor. The energy consumption includes leakage as well as switching energy for the given interconnect circuits. The waveforms on the data bus were fed to the spice model of the interconnect circuit. This was done for differential, basic, and the proposed interconnect schemes set at the same delay constraint. Figure 16 shows that the proposed architecture saves up to 70% energy.

5. CONCLUSIONS

A new low power interconnect circuit was proposed and demonstrated. The proposed interconnect uses a charge-pumping technique to achieve high performance at 3X less energy than alternatives at comparable speeds. Simulations of a four core Alpha processor running Splash workloads show up to 70% energy savings at constant performance over alternative interconnect implementations.

6. REFERENCES

- [1] P. Kogge, K. Bergman, S Borka, et. al, "ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems" *DARPA/IPTO*, September 2008
- [2] D. Liu and C. Svensson, "Power Consumption Estimation in CMOS VLSI chips" *IEEE Journal of Solid-State Circuits*, Vol-29 No-6, June 1994.
- [3] E. Kusse and J.M. Rabaey, "Low-Energy Embedded FPGA Structures" *IEEE International Symposium on Low Power Electronics Design*, August 1998 .
- [4] H. Zhang, V. George and J.M. Rabaey, "Low-Swing On-Chip Signalling Techniques: Effectiveness and Robustness" *IEEE Transactions on Very Large Scale Integration (VLSI)*, Vol-8 No-3, June 2000
- [5] J.C.G. Montesdeoca, J.A. et. al, "CMOS Driver Receiver Pair for Low Swing Signalling for Low Energy On-chip Interconnects" *IEEE Transactions on Very Large Scale Integration (VLSI)*, Vol-17 No-2, February 2009.
- [6] R. Ho, I. Ono, F. Liu, et. al, "High Speed and Low Energy capacitively driven wires" *IEEE International Solid State Circuits Conference*, February 2007.
- [7] M. Ferretti and P.A. Beere "Low Swing Signaling Using a Dynamic Diode-Connected Driver" *European Solid-State Circuits Conference*, September 2001.
- [8] A. Narshimha, M. Kasotiya and R. Sridhar "A Low-Swing Differential signaling Scheme for on-chip Global Interconnects" *International Conference on VLSI Design*, January 2005.
- [9] N. Tzartzanis, W.W. Walker "Differential Current Mode Sensing for Efficient On-Chip global Signaling" *IEEE Journal of Solid State Circuits*, Vol-40 No-11, November 2005.
- [10] H. Ito, M. Kimura, K. Miyashita, et. al, "A Bidirectional and Multidrop Transmission Line Interconnect for Multipoint to Multipoint On-Chip Communication" *IEEE Journal of Solid State Circuits*, Vol-43 No-4, April 2008.
- [11] Binkert, N.L. , Dreslinski, R.G. , Hsu, L.R. , Lim, K.T. , Saidi, A.G. , Reinhardt, S.K. , "The M5 Simulator: Modeling Networked Systems" *IEEE Micro*, July 2006