An Energy-Proportional, High-Speed Serial Link

Alex Huang, Amirmohammad Nazari, Robert Soulé, and Rajit Manohar
Computer Systems Lab
Yale University
New Haven, CT 06520
alex.huang@aya.yale.edu, {amirmohammad.nazari,robert.soule,rajit.manohar}@yale.edu

***Abstract*—Several recent projects have proposed approaches to traffic engineering that schedule workloads to maximize the idle time of network equipment in order to reduce overall power consumption. However, today's switches and network interface cards (NICs) consume significant power even while idle. This power inefficiency both increases operational costs and has a tremendous environmental impact. The key obstacle in realizing an energy proportional switch is physical link performance, i.e., creating a high-speed SERDES (serialization/deserialization) link.**

Asynchronous circuit designs, which do not rely on a global clock, make any digital computation energy proportional by their nature. However, asynchronous SERDES design is a relatively unexplored topic. In this paper, we improve on a previous SERDES design to incorporate continuous-time equalization hardware and circuit-level optimizations for the digital processing that takes the data rate to 20 Gbps in a 65nm technology. Through circuit simulation, we demonstrate that the SERDES link is energy proportional. Overall, this is an important step towards designing energy-proportional network switches and NICs.

***Index Terms*—Energy-proportional networking, High-speed serial link, Asynchronous circuits**

I. Introduction

Reducing the power consumption of large computing platforms is a critical goal for architects and operators, motivated by the need to reduce operational costs [1], [2] or meet sustainability goals [3], [4], [5]. A common approach towards reducing power consumption is to schedule workloads such that a set of servers are unused, and can therefore be placed into low-power modes [2], [3], [4].

There have been several proposals to apply similar approaches with network devices [6], [7], [8]. However, these proposals all suffer from a fundamental issue: unlike servers, network devices are not truly energy-proportional, such that energy consumption does not scale linearly with usage. The absence of energy-proportional network devices is a serious concern because the power consumed by network devices while idle is significant. El Zhar et al. [6] report that the idle power of a switch accounts for roughly 2/3rd of the overall power consumption at peak traffic load. Moreover, the energy consumed by computer networks as a whole is considerable. Malmodin et al. estimate that the electricity usage of data networks as a whole is as large, or larger than that of data centers [9].

By using asynchronous circuits, which do not rely on a global clock, it is possible to make the dynamic energy component of any digital computation energy-proportional. In fact, commercial products have already demonstrated the feasibility of implementing the core logic for a high-throughput switch using asynchronous circuits [10], [11], [12]. The key outstanding issue to making the entire switch energy-proportional is physical link performance—in particular, creating a high-speed SERDES (serialization/deserialization) link.

A traditional high-speed SERDES circuit includes a number of constantly operating elements, such as clock recovery logic and associated analog/mixed-signal signal processing equalization. The conventional wisdom is that this logic takes a long time to transition between idle and active states; hence, "waking them up" would add too much latency for packet processing.

In this paper, we present an alternate approach to SERDES design inspired by the asynchronous digital circuit design philosophy. We build on a previous design that showed that a clockless SERDES could achieve a 3 Gbps data rate. Our contributions are twofold: (i) the incorporation of continuous-time equalization hardware and circuit-level optimizations for the digital processing that takes the data rate from 3 Gbps to 20 Gbps in a 65nm technology; and (ii) detailed circuit simulation results with transmission line models that demonstrate the SERDES link is energy-proportional.

The rest of this paper is organized as follows. We first present background on traditional, clocked SERDES and prior work on asynchronous SERDES designs (§II). Next, we present our optimized, asynchronous SERDES (§III), and evaluate its power usage, data rate, and signal integrity (§IV). Finally, we discuss related work (§V) and make concluding remarks (§VI).

II. SERDES Primer

Before presenting our design, we first present background on clocked serial links and prior work on asynchronous designs.

Clocked SERDES. The goal of the SERDES is to convert parallel digital data operating at a standard on-chip clock frequency to/from a serial analog format operating at a much higher frequency that is suitable for the transmission line medium corresponding to the chip-to-chip link. In the analog format, the bits are separated by a period that determines the peak bit rate of the link. The clock is embedded within the data, rather than relying on a separate signal. The receiver uses a clock and data recovery circuit to reconstruct both the clock and data bit simultaneously [13]. A typical SERDES implementation is shown in Figure 1. High data rate is

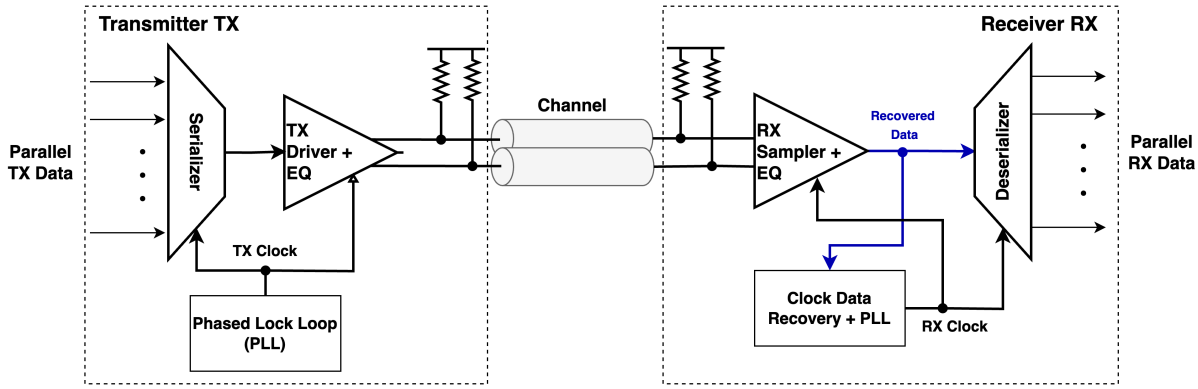


Figure 1. Clocked serial link block diagram showing major components on both the transmitter and receiver side of the link.

achieved by utilizing a multi-phase clock, where a new bit is transmitted at each phase boundary—effectively multiplying the data rate relative to the base clock frequency that is used for the parallel input.

The clock data recovery circuit is always active due to the fact that there is a high startup latency for the receiver reference clock to be established. Clock data recovery in commercial products can take hundreds of nanoseconds [14] and is often longer. This latency makes it prohibitive to shut down and then restart the link upon data arrival.

Asynchronous SERDES. Teifel and Manohar [15] explored an alternative, clockless approach to SERDES design. Rather than using a multi-phase clock to determine bit boundaries, their design used an asynchronous circulating token, where the token indicates permission to transmit a bit; hence, token arrival times determine bit boundaries. The approach exploits the fact that the token latency is significantly smaller than the frequency of token circulation, achieving the same effect as a multi-phase clock but without an explicit clock.

To avoid a clock and data recovery circuit at the receiver, Teifel and Manohar [15]'s design used a three-wire protocol first proposed by Røine [16]. Rather than using two-wire differential signaling, the protocol guarantees that the same wire is never used for two successive bits. This permits the receiver to use a state machine to determine when the next bit arrives without an explicit clock.

One of the motivations behind this work is the observation that eliminating the clock and data recovery circuit could be a path toward achieving an energy-proportional serial link, at the cost of introducing an additional wire. We believe that this is a desirable trade-off because the wire/board overheads of a high-speed link go beyond just the two wires. For example, adjacent links must be shielded, and dedicated on-chip analog power and ground pins are often needed in the package per lane, so an additional wire is much less than a 50% overhead that one might think at first glance for two reasons: (i) a normal SERDES footprint on a chip/package requires additional adjacent power/ground bumps; and (ii) the approach cuts the total maximum frequency per wire used by a factor of two because a wire cannot transition more often than once for every two bits.

While the Teifel and Manohar design demonstrated the feasibility of a clockless approach, it also suffered from several limitations. First, it was only able to achieve a data rate of 3 Gbps—which was respectable at the time but far away from a modern SERDES. Second, it did not incorporate any signal processing necessary to achieve modern serial link speeds. Third, it did not demonstrate (or in fact achieve) energy-proportionality, which is our primary motivation.

Our Contribution. In this paper, we build on these prior efforts to develop a SERDES with two key properties: (i) it is energy-proportional, with idle power being less than 2% of the overall power requirement; and (ii) it achieves a data rate of over 20 Gbps per lane in a 65nm technology, which is not that far from the highest reported links in that process. We also show how the notion of an eye-diagram—which depends on a fixed operating frequency—can be adapted for an asynchronous variable bit-rate SERDES, and how continuous-time signal processing techniques can replace the standard fixed-frequency signal processing functions used in a clocked SERDES.

III. System Design

In this section, we describe our asynchronous SERDES design. Our design improves upon Teifel and Manohar [15] by increasing the throughput and also addressing the accompanying signal integrity considerations.

As is common in asynchronous design, we adopt a token ring architecture. Token rings are used in asynchronous designs to implement mutual exclusion, i.e., when elements of the ring can access a shared resource. In our case, the shared resource is the transmission line.

A token ring consists of n concurrent processing elements connected in a ring, such that the output of process i is connected to the input of process $(i+1) \mod n$. The input/output of the elements is a token, which is passed around the ring. Possession of the token allows the element to access the shared resource. It is important to note that the token can stop at any element, allowing for an idle state. The token can then start

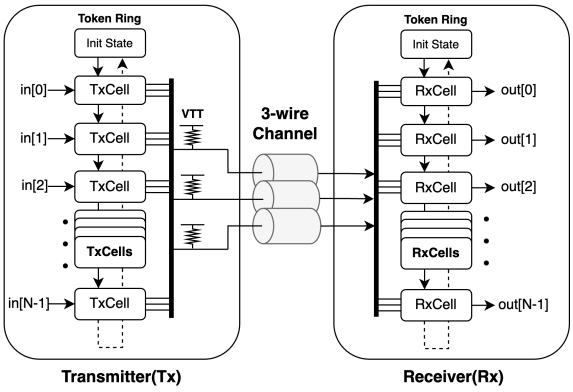


Figure 2. Teifel and Manohar's Design [15]

itself at anytime with no startup latency. This feature is what enables energy-proportionality.

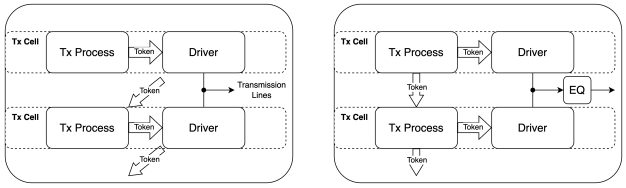
The sequential passing of the token serves as both a clock generator and a mechanism used to select the bit on the parallel bus that is to be transferred to/from the serial I/O link. The value of the token encodes the current state in a state machine, where the machine changes state based on the input bit and the previous state stored in the token. We use Røine [16]'s three-wire protocol and represent state transitions in the protocol by a short pulse on its respective wire. The signaling protocol ensures that two successive pulses are never sent on the same wire.

A. Token Ring Improvements

To improve the data throughput compared to the prior design, we focus on circuit-level improvements to the token-ring architecture. The key determinant of throughput in the token-ring based scheme is the propagation delay of the token from one stage to the next, i.e., the latency of token transfer between neighboring stages. Below, we discuss the changes incorporated into the transmitter and receiver that permit 20 Gbps transmission.

Transmitter Improvements. Each token ring stage in the transmitter is divided into two parts: (i) the main process which handles next state calculation and forwarding the next state to the next cell; and (ii) the driver which forwards the next state to the shared bus driver that drives the actual I/O link. The main process is implemented using digital asynchronous circuits, and the driver uses a combination of asynchronous digital and analog circuits.

The token propagation delay in Teifel and Manohar's design [15] was four gate delays. The circuit used to implement the TX process is shown in Figure 4. It corresponds to a slightly modified, pre-charge, half-buffer, asynchronous circuit template [17], [18]. The one-of-three encoded output token value ($T.0$, $T.1$, $T.2$) is computed from the input token ($U.0$, $U.1$, $U.2$) and the data bit ($L.0$, $L.1$). The forward latency for this process is two gate delays from input token arrival to output token generation. This token was passed to a driver process as shown in Figure 3(a), which uses the computed



(a) Previous Design (b) Optimized Design

Figure 3. Transmitter design, showing the change in information flow and token passing between the previous design and the proposed optimized design.

token value and passes it back to the TX process in the next stage using another pre-charge logic stage, adding two more gate delays to token propagation latency. This makes the process logic extremely robust to delay variations/uncertainties, isolating timing constraints to the analog driver module only.

As shown in Figure 3(b), we reduce the token propagation delay to two gate delays by simply sending the token to the driver in parallel with the next token ring stage. While this sacrifices some timing robustness in the asynchronous logic, it is a reasonable compromise because the actual driver logic already makes a number of internal timing assumptions in terms of signal pulse width, etc. and hence the entire circuit has to be carefully analyzed for correct operation regardless of the timing robustness of token propagation. Along with gate delay reduction, we implement aggressive transistors sizing tactics, including sizing down NMOS transistors to further decrease latency of tokens.

In Teifel and Manohar's design, data pulses were sent on the transmission line by swinging the voltage from ground to Vdd fully. However, this voltage swing takes a significant amount of time and power that can be avoided by using a smaller voltage swing. In our design, we use current mode logic to send pulses using a 200mV swing, allowing increased throughput.

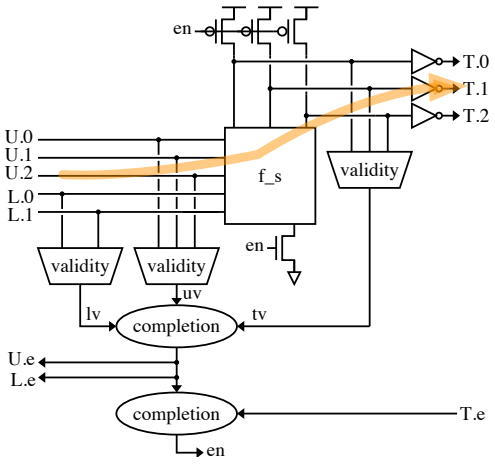


Figure 4. Circuit of the TX process from [15]. The token traversal path is highlighted, showing the two gate delays for token propagation. The function f_s computes the next state of the three wire protocol from the current state and data bit.

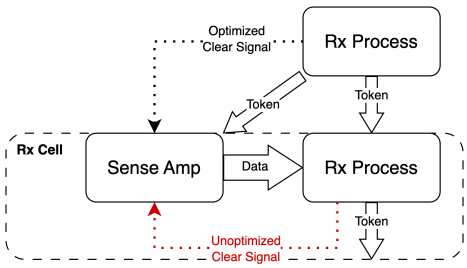


Figure 5. Receiver design, showing both the unoptimized and optimized signal path for the clear signal used by the sense amplifier.

Receiver Improvements. The receiver design is very similar to the transmitter architecture. It is also split into two parts: (i) a sense amplifier that captures sent data, and (ii) the main process that calculates the next state and final data bit. Additionally, the sense amplifier output is cleared before the next data bit is computed.

In Teifel and Manohar's receiver design [15], the token propagation delay was five gate delays. The previous approach passed the token to the sense amplifier, which computed the received data bit (two gate delays). The data bit is then used to compute the next token value (two gate delays). Each computation requires an inverter due to the complementary nature of CMOS transistors. Finally, the sense amplifier clear input took an additional gate delay, totaling five gate delays.

In our optimized design, shown in Figure 5, the sense amplifier clear input is pre-computed from the neighboring receiver cell, reducing the propagation delay by one gate. The logical computation was also redesigned to avoid inverters, resulting in another two gate delay reduction. This three gate delay reduction, combined with aggressive transistor sizing for speed, increases the throughput to 20 Gbps.

The optimized design also uses a single 3-input differential amplifier to turn analog pulses values into digital values. Relying on a differential signal improves the signal integrity and frequency response needed for a 20 Gbps channel.

B. Continuous-Time Equalization

With increased throughput, we must incorporate equalization measures to ensure signal integrity. As is well-known, transmission lines on printed circuit boards significantly attenuate high-frequency signal content. This means that signals with fast rise/fall times are distorted as they pass through the chip-to-chip link. Chip designers must consider the physical properties of an off-chip channel including the skin effect, dielectrics, parasitics, and other transmission line effects to ensure signal integrity and low bit-error rates. In the case of synchronous implementations, signal processing techniques like equalization and pre-emphasis are used that operate with the assumption of a fixed transmission frequency [13]. In the state-of-the-art synchronous links, the basic tool to correct transmission line distortions is to incorporate filters (in the signal processing sense) on both the transmitter and receiver end.

In the case of asynchronous circuits, digital equalization has been relatively unexplored since it is not amenable to standard frequency-domain analysis at a fixed operating frequency. One of the new aspects of our driver is that we incorporate a continuous-time feed-forward equalizer to improve analog signal integrity of the link.

Equalization is the process of correcting for the distortion introduced by the transmission line medium, so that the effective signal (with corrections) at the receiver is as close as possible to the ideal undistorted version sent by the transmitter. In our transmitter, we utilize a finite impulse response (FIR) filter. A FIR filter can shape the signal by performing a finite digital convolution over our signal as a weighted sum. The weighted sum looks at previous symbols sent down the line and accordingly will account for pre or post-symbol interference depending on its tuning [31].

In our design, we use a 4-tap FIR filter for this purpose. Our design differs from a traditional FIR filter that uses clocked registers operating at a fixed frequency. Instead, our design operates in continuous-time where the delays between the filter taps are set by the token propagation time.

C. Toward A State-of-the-Art 112.5 Gbps SERDES

The SERDES presented in this work operates at 20 Gbps, a notable improvement over the earlier 3 Gbps design. However, this data rate is not a fundamental limitation of the architecture. Our throughput limitation primarily comes from the maximum speed of logical computation from the receiver (two gate delays). Our current design is also limited by a two-level signal, where a wire has two states (idle and transmitting a pulse). Table I shows the data rates and technology nodes of representative SERDES papers from the literature, demonstrating that our design achieves state-of-the-art performance in 65 nm technology. It is important to note, faster links typically use PAM4 signaling, which, instead of sending a binary high or low signal, uses a four-level voltage scheme to encode two bits per symbol, effectively doubling the data rate within the same signaling bandwidth. These links are also typically limited by signal integrity and complex clock architectures that have very tight jitter timing requirements needed for clock-based precise temporal sampling.

In terms of a future, potential design path, using a smaller manufacturing technology node such as 16 nm or 7 nm would allow us to significantly increase throughput by the reduction in local gate delay compared to 65 nm. (For example, the gate delay in 22 nm is estimated to be 3.86× lower than in 65 nm [32].) To further improve data rate, we also plan to incorporate multi-voltage signalling similar to PAM4 that can result in an additional doubling of data rate. These design changes not only follow the well-understood path in the literature illustrated in Table I, but may also present fewer implementation challenges in an asynchronous architecture, as the circuits are more tolerant to jitter and have reduced equalization requirements (further discussed in §IV-B). Another advantage of three wiring signaling is that since a pulse cannot be sent on the same wire twice, the effective

Table I
SUMMARY OF SERDES DESIGNS, INCLUDING NODE, DATA RATE, AND EQUALIZATION.

Paper	Tech Node	Data Rate	Features	Equalization
Beukema *et al.* [19]	130nm	6.4 Gbps	4-tap FFE	5-tap DFE
Gondi *et al.* [20]	130nm	10 Gbps	–	–
Liu *et al.*[21]	65nm	12.5 Gbps	CML Driver	–
Philpott *et al.*[22]	65nm	20 Gbps	NRZ, 4-tap FFE	–
Menolfi *et al.* [23]	65nm	16 Gbps	4-tap FFE	–
This Paper	65nm	20 Gbps	Async, CT-FIR	4-tap CT-FFE
Tao *et al.* [24]	28nm	50 Gbps	PAM4, 4-tap FFE	–
Bassi *et al.* [25]	28nm	45 Gbps	PAM4, 4-tap FFE	–
Kwon *et al.* [26]	28nm	40 Gbps	PAM4	2-tap
Celik *et al.* [27]	28nm	32.5 Gbps	PAM4, 4-tap FFE	–
Xu *et al.* [28]	28nm	64 Gbps	PAM4, 4-tap FFE	–
Zhang *et al.* [29]	28nm	32.5 Gbps	NRZ, CTLE	2-tap DFE
Groen *et al.* [30]	7nm	112 Gbps	PAM4, FinFET	–

needed bandwidth on each wire is half of the total throughput, resulting in better signal integrity compared to a traditional two-wire transmission line.

Thus, once we have physically validated our current design through chip prototyping, we plan to take these natural next steps toward reaching or even exceeding speeds representative of the latest SERDES designs (112.5 Gbps).

D. Putting it All Together

Our new asynchronous SERDES chip uses asynchronous circuit design for power proportionality. The token ring design is able to effectively pass a token around to do calculations and generate pulses on the wire. The new design has been optimized to maximize throughput and ensures signal integrity across a transmission line. The new design also holds the same power proportionality due to the token's ability to stop at any moment, which causes an idle state where no transistors are being actuated and thus no dynamic power is used. Overall, the design uses significantly less power when idle and has no additional latency when switching between power states while still maintaining a substantial data rate.

IV. EVALUATION

In this section, we describe our methodology for implementing our proposed designed circuit, and evaluate the results of our simulations.

A. Implementation

Tooling. We designed our circuit using the Asynchronous Circuit Toolkit (ACT) [33], manually designing individual gates to optimize the energy and throughput of the digital circuits. We use the Sandia National Laboratories' Xyce [34] analog circuit simulator to develop an accurate electrical model for the communication link. The architecture is built using TSMC 65nm CMOS technology and supplied by a 1.2V power supply. The transmission line is biased with a voltage at .6V and a driver with 200mV of swing. Estimated parasitics are included in our simulations of all circuit components.

Line and Package Model. The transmission line used in our simulations is modeled off a FR4 PCB stripline of 0.2 meters. We use an RLC model with electrical properties set to the maximum frequency primarily using the 'transline' feature in Xyce [34]. It is important to note that even when the serializer is run at a frequency less than the maximum, no equalization change is necessary due to its continuous-time design. We also model QFN32 [35] package pins on the transmitter and receiver.

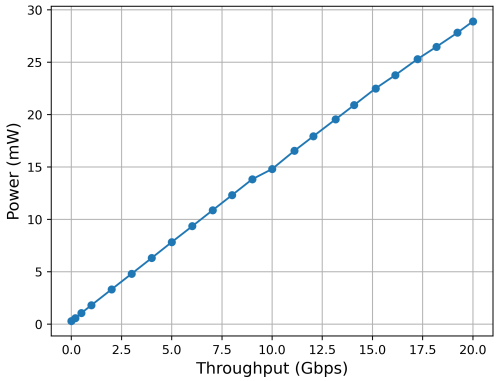
B. Results

In evaluating our design, we focused on the following metrics: (i) power consumption; (ii) peak attainable throughput; and (iii) signal integrity. We conducted a set of measurements in simulation. In all simulations, we used a token ring with a 16-bit word input. The input data to the simulation was generated by a pseudorandom bit sequence using the PRBS31 polynomial, typically used for SERDES testing. This input bit-sequence is broken into 16-bit digital words provided to the token ring. The throughput can be controlled by setting the delay between every bit sent.

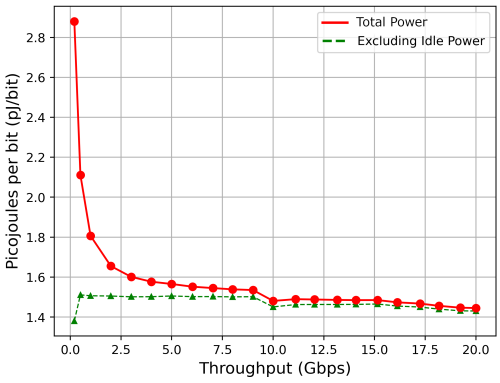
Energy Proportionality. For the first experiment, we measured power usage as we increased data throughput in the link from 200 Mbps to 20 Gbps. The results appear in Figure 6(a). We can see that power usage increases linearly as we increase the throughput, demonstrating that our design is indeed power proportional. Focusing on the lower data rates, Figure 6(b), shows that the energy usage per bit is sub-1.6pJ/bit from 2.5 Gbps, and remains roughly the same as we go to higher data rates. However, the energy more than doubles as we go to very low data rates—rates that would normally not require a SERDES at all.

Idle Power Usage. Idle power becomes significant in the sub-2.5 Gbps regime. In a simulation with no data transmitted, we found that idle operation only consumes 0.3 mW, representing $1/50^{\text{th}}$ the power consumption at 20 Gbps.

Maximum Throughput. The peak throughput is achieved when the environment is always ready to supply the next data

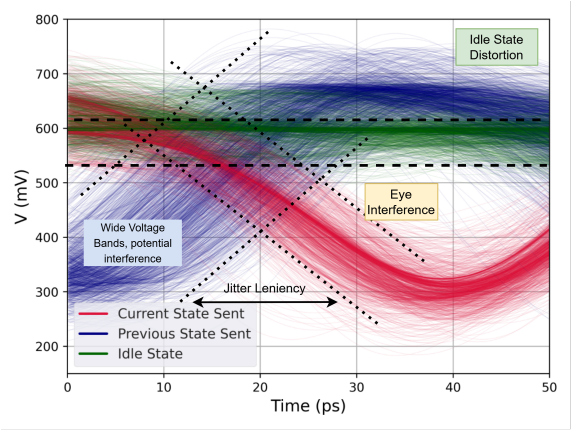


(a) Simulated Power Graph

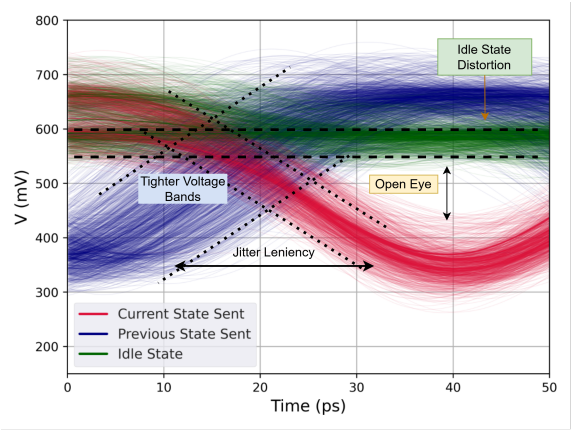


(b) Power consumption per bit

Figure 6. Energy Proportionality. The top figure shows the simulated power of the link as a function of the data rate. The bottom figure shows the energy per bit as a function the data rate, both including and excluding idle power. Below 2.5 Gbps, the idle power is a major contributor to the energy per bit.



(a) No Equalization Treye Diagram



(b) Equalized with 4-taps

Figure 7. Signal Integrity. Simulation results with and without equalization are shown. Note that the traditional notion of "eye opening" is not as relevant here because the receiver waits for signal separation between the idle wire (green) and current state wire (red).

word and when the receiver can operate at a data rate that is at least as fast as the transmitter. From simulations, the peak throughput of the link was 20 Gbps.

Signal Integrity. In traditional SERDES design, eye diagrams are used to demonstrate the integrity of the signal. An eye diagram consists of superimposed time-domain waveforms of the signal voltage of the two wires of traditional links. Since transmitters operate at a fixed frequency, and the receiver picks a fixed time offset to sample the signal, the "eye opening" at this offset can be used to determine if the receiver can reliability distinguish between a 0 or 1 [13].

However, our 3-wire protocol operates differently. In particular, the receiver waits until the wires are sufficiently separated in voltage before generating a data output bit. This means we don't have a fixed time offset for sampling the received signal, and hence the traditional metrics for eye opening no longer strictly apply. Furthermore, since a single wire cannot change twice in a row, the reset time for this wire is significantly relaxed compared to the bit rate of the link. Due to this property, our protocol can handle up to $T/2$ of jitter where T is the period of the operating frequency, significantly more lenient than the traditional sampling architecture.

Instead, we introduce the notion of a "Treye" (Tri) diagram to visualize the received signal. The diagrams shown in Figures 7(a) and 7(b) group the state being transmitted in red, the previous state sent in blue, and the idle state in green. It is important to note that each color does not necessarily correspond to an individual wire as the wire changes its role throughout the simulation lifespan. Evaluating the diagram requires ensuring that the idle state stays above the activation threshold, and that the previous state has sufficiently settled during the pulse of the transmitted state. The temporal alignment of the waveforms shown is based on the arrival of the state token at the bit location in the token ring.

In Figure 7(b), we can clearly see the subtle yet important effects of equalization. The set of superimposed waveforms have a much tighter spread, which means that there is significantly lower noise in all signals along with fewer reflections. The idle state line also has significantly fewer reflections and stays consistently near the voltage threshold of 0.6V,

reducing the chance of token transmission errors. The eye has significantly less interference and has opened up, also reducing the chance of errors.

V. Related Work

Below, we discuss closely related work on clocked and asynchronous SERDES design.

Clocked SERDES Designs. Dally's book [13] provides a foundational reference on clocked SERDES design. There has been extensive prior work on low-power, clocked SERDES designs [36], [37], [38], [24], [39]. All such approaches are similar in that they modify the clock and data recovery (CDR), MUX, or other blocks to reduce energy usage. None of these designs are energy proportional.

Sawaby et al. [40] design a 10Gb/s transceiver with 65nm technology that works with 9.57mW power. Continuous clocking in the CDR, CTLE, DFE, and other blocks leads to higher dynamic power consumption. Additionally, they utilize a CDR, which adds complexity to the design and will consume higher power at scale. In contrast, our design is data-driven and only consumes power during actual data transmission events. Moreover, by using the token ring architecture, we can easily increase the number of parallel links on either side of the transceiver with a small overhead.

Zhang et al. [29] proposed an energy-efficient design where only 20mW power is used at 32Gb/s rate. They use an aggressive 2-tap equalization technique to compensate for channel loss at high frequency.

Clockless SERDES. There is a paucity of work on clockless SERDES designs. This paper builds upon the design by Teifel and Manohar [15]. Both designs use an asynchronous transceiver with a three-wire signaling technique. However, their design used 180nm technology and could achieve a 77mW power usage at 3Gb/s data rate. Our work improves the data rate by more than 600%, leading to a 20 Gbps rate at 65nm technology, adds FFE channel equalization, and dramatically decreases power usage down to 28mW at full throughput.

Yoneda and Imai [41] also try to reduce SERDES idle power usage. However, their design keeps different clocks at the transmitter and receiver rather than connecting fully asynchronous circuits. They use a novel voltage encoder that simplifies clock recovery such that it can be done using a voltage comparison. They report 21mW of power consumption during active periods and almost 3mW during inactive periods. In contrast, our proposal removes the CDR, uses handshakes instead of a clock, maintains a higher bit rate, and has much lower idle power.

Asynchronous schemes for on-chip serial communication have been reported [42], [43] with data rates in excess of 50 Gbps for on-chip communication. This work differs in that it examines chip-to-chip communication where high data rates must be achieved in a signalling environment that is very lossy, requiring active compensation using equalization and pre-emphasis.

Recent work has also reported on an asynchronous bit-serial LVDS link [44]. They achieve a data rate of 1.5 Gbps in a 180nm technology with a power consumption of 22.87mW, idle power in the 0.1μW range, and use four wires for the link. They also use a token-ring based approach following [15]. Our approach targets a much higher peak data rate, which in turn also requires a higher power budget when the link is idle.

VI. Conclusion

Reducing power consumption is critical to lower operational costs and decrease the carbon footprint of data centers. Unfortunately, today's network equipment, including switches and network interface cards (NICs), consume significant power even while idle. The key remaining obstacle in developing energy-proportional network equipment is the SERDES.

In this paper, we present an optimized SERDES design inspired by the asynchronous digital circuit design philosophy. In particular, we improve on the state-of-the-art by incorporating continuous-time equalization into the transmitter and circuit-level optimizations for digital processing that takes the data rate from 3 Gbps to 20 Gbps in a 65nm technology. Our simulation demonstrates that the SERDES link is energy-proportional, and we plan to continue this work with a test chip to validate our design. Overall, this marks an important step towards designing energy-proportional switches and NICs.

Acknowledgments

This research was supported in part by Yale Planetary Solutions and the Three Cairns Climate Impact Innovation Fund, and in part by NSF awards #2420367 and #2504588. The authors would like to thank members of the asynchronous VLSI group for discussions on layout and design automation flows, and Arun Natarajan for advice on high-speed analog circuit design, packaging, and testing.

References

[1] L. A. Barroso, J. Clidaras, and U. Hlzle, *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. 2nd ed., 2013.

[2] J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, and R. P. Doyle, "Managing energy and server resources in hosting centers," in *Proceedings of the Eighteenth ACM Symposium on Operating Systems Principles*, SOSP '01, p. 103–116, 2001.

[3] A. A. Chien, R. Wolski, and F. Yang, "The zero-carbon cloud: High-value, dispatchable demand for renewable power generators," *The Electricity Journal*, vol. 28, no. 8, pp. 110–118, 2015.

[4] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez, "Greening the internet with nano data centers," in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '09, p. 37–48, 2009.

[5] M. Gupta and S. Singh, "Greening of the internet," in *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '03, p. 19–26, 2003.

[6] S. El-Zahr, P. Gunning, and N. Zilberman, "Exploring the benefits of carbon-aware routing," *Proceedings of the ACM on Networking*, vol. 1, no. CoNEXT '03, pp. 1–24, 2023.

[7] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, "Reducing network energy consumption via sleeping and rate-adaptation," in *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, NSDI'08, p. 323–336, 2008.

[8] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown, "Elastictree: saving energy in data center networks," in *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*, NSDI'10, p. 17, 2010.

[9] J. Malmodin and D. Lundén, "The energy and carbon footprint of the global ict and e&m sectors 2010–2015," *Sustainability*, vol. 10, no. 9, 2018.

[10] Intel, *Ethernet Switch FM5000/ FM6000 Datasheet*, 7 2017. [Online, accessed March 11, 2025].

[11] A. Lines, "Nexus: an asynchronous crossbar interconnect for synchronous system-on-chip designs," in *Proceedings of the 11th Symposium on High Performance Interconnects*, pp. 2–9, 2003.

[12] M. Davies, A. Lines, J. Dama, A. Gravel, R. Southworth, G. Dimou, and P. Beerel, "A 72-port 10g ethernet switch/router using quasi-delay-insensitive asynchronous design," in *20th IEEE International Symposium on Asynchronous Circuits and Systems*, pp. 103–104, 2014.

[13] W. J. Dally and J. W. Poulton, *Digital Systems Engineering*. 2008.

[14] K. Clark, H. Ballani, P. Bayvel, D. Cletheroe, T. Gerard, I. Haller, K. Jozwik, K. Shi, B. Thomsen, P. Watts, H. Williams, G. Zervas, P. Costa, and Z. Liu, "Sub-nanosecond clock and data recovery in an optically-switched data centre network," in *2018 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2018.

[15] J. Teifel and R. Manohar, "A high-speed clockless serial link transceiver," in *Ninth International Symposium on Asynchronous Circuits and Systems (ASYNC), 2003. Proceedings.*, pp. 151–161, IEEE, 2003.

[16] P. Røine, "A system for asynchronous high-speed chip to chip communication," in *Proceedings Second International Symposium on Advanced Research in Asynchronous Circuits and Systems (ASYNC)*, pp. 2–10, 1996.

[17] A. Lines, "Pipelined asynchronous circuits," Master's thesis, California Institute of Technology, Nov 1995.

[18] D. Fang and R. Manohar, "Non-uniform access asynchronous register files," in *IEEE International Symposium on Asynchronous Circuits and Systems*, pp. 78–85, IEEE, 2004.

[19] T. Beukema, M. Sorna, K. Selander, S. Zier, B. Ji, P. Murfet, J. Mason, W. Rhee, H. Ainspan, B. Parker, and M. Beakes, "A 6.4-gb/s cmos serdes core with feed-forward and decision-feedback equalization," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2633–2645, 2005.

[20] S. Gondi and B. Razavi, "Equalization and clock and data recovery techniques for 10-gb/s cmos serial-link receivers," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 9, pp. 1999–2011, 2007.

[21] B. Liu, Z. Wang, T. Zhang, and L. Zhang, "A 12.5gbps novel serdes transmitter for jesd204b physical layer," in *2021 5th International Conference on Communication and Information Systems (ICCIS)*, pp. 6–12, 2021.

[22] R. A. Philpott, J. S. Humble, R. A. Kertis, K. E. Fritz, B. K. Gilbert, and E. S. Daniel, "A 20gb/s serdes transmitter with adjustable source impedance and 4-tap feed-forward equalization in 65nm bulk cmos," in *2008 IEEE Custom Integrated Circuits Conference*, pp. 623–626, 2008.

[23] C. Menolfi, T. Toifl, P. Buchmann, M. Kossel, T. Morf, J. Weiss, and M. Schmatz, "A 16gb/s source-series terminated transmitter in 65nm cmos soi," pp. 446 – 614, 03 2007.

[24] B. Tao, C. Zhang, Y. Ren, and W. Qin, "A low-power 50 gb/s pam4 high-speed serdes transmitter," in *2024 9th International Conference on Integrated Circuits and Microsystems (ICICM)*, pp. 201–207, IEEE, 2024.

[25] M. Bassi, F. Radice, M. Bruccoleri, S. Erba, and A. Mazzanti, "A high-swing 45 gb/s hybrid voltage and current-mode pam-4 transmitter in 28 nm cmos fdsoi," *IEEE Journal of Solid-State Circuits*, vol. PP, pp. 1–14, 09 2016.

[26] D.-H. Kwon, M. Kim, K. Sung Geun, and W.-Y. Choi, "A low-power 40-gb/s pre-emphasis pam-4 transmitter with toggling serializers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, pp. 430–434, 03 2020.

[27] F. Celik, A. Akkaya, A. Tajalli, and Y. Leblebici, "A 32-gb/s pam-4 sst transmitter with four-tap ffe using high-impedance driver in 28-nm fdsoi," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. PP, pp. 1–9, 03 2021.

[28] H. Xu, X. Zheng, Z. Wang, C. Cai, W. Zhen, G. Yuan, Q. Hao, X. Guo, and Z. Jin, "A high-output-swing 64-gb/s pam-4 transmitter with a 4-tap hybrid ffe in 28-nm cmos," *IEICE Electronics Express*, vol. 21, 03 2024.

[29] X. Zhang, C. Zhang, Y. Song, and R. Song, "32gb/s low-power nrz signal serdes receiver," in *2024 9th International Conference on Integrated Circuits and Microsystems (ICICM)*, pp. 177–181, IEEE, 2024.

[30] E. Groen, C. Boecker, M. Hossain, R. Vu, S. Vamvakos, H. Lin, S. Li, M. Van Ierssel, P. Choudhary, N. Wang, M. Shibata, M. H. Taghavi, N. Nguyen, and S. Desai, "6.3 a 10-to-112gb/s dsp-dac-based transmitter with 1.2vppd output swing in 7nm finfet," in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, pp. 120–122, 2020.

[31] S. Palermo, "Slides from ecen 720: High-speed links circuits and systems." https://people.engr.tamu.edu/spalermo/ecen720.html, 2024. [Online, accessed March 11, 2025].

[32] A. Stillmaker, Z. Xiao, and B. Baas, "Toward more accurate scaling estimates of cmos circuits from 180 nm to 22 nm," Tech. Rep. ECE-VCL-2011-4, VLSI Computation Lab, ECE Department, University of California, Davis, 2011.

[33] S. Ataei, W. Hua, Y. Yang, R. Manohar, Y.-S. Lu, J. He, S. Maleki, and K. Pingali, "An open-source eda flow for asynchronous logic," *IEEE Design & Test*, vol. 38, no. 2, pp. 27–37, 2021.

[34] J. Verley, E. R. Keiter, and H. K. Thornquist, "Xyce: Open source simulation for large-scale circuits.," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2018.

[35] W. Xie, G. Wen, and H. Jin, "Extraction of parasitic parameters for qfn32," in *2009 IEEE Youth Conference on Information, Computing and Telecommunication (YC-ICT)*, pp. 224–226, 2009.

[36] H. Okuhara, A. Elnaqib, M. Dazzi, P. Palestri, S. Benatti, L. Benini, and D. Rossi, "A fully integrated 5-mw, 0.8-gbps energy-efficient chip-to-chip data link for ultralow-power iot end-nodes in 65-nm cmos," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 29, no. 10, pp. 1800–1811, 2021.

[37] S. Sen, U. Upadhyaya, K. R. Kondreddy, A. Goyal, S. Goyal, and S. Gupta, "A low jitter digital loop cdr based 8–16 gbps serdes in 65 nm cmos technology," in *2021 34th International Conference on VLSI Design and 2021 20th International Conference on Embedded Systems (VLSID)*, pp. 216–221, IEEE, 2021.

[38] C. Zhang, Y. Xu, Y. Wang, *et al.*, "Design of 40gb/s pam4 signal serdes low power cmos 64: 4 mux," in *2023 8th International Conference on Integrated Circuits and Microsystems (ICICM)*, pp. 457–462, IEEE, 2023.

[39] H. Kang, Y. Choi, J. Sim, J. Choi, Y. Kwon, S. Park, S. Kim, C. Sim, and C. Kim, "A 13-gb/s single-ended nrz receiver with 1-sample per 2-ui using data edge sampling for memory interfaces," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 71, no. 7, pp. 3328–3332, 2024.

[40] A. M. Sawaby, A. M. Elshorbge, O. T. Abdelhalim, M. A. Farghaly, M. S. Taha, Y. H. Yehia, S. El-Sawy, M. S. Fouad, and H. Mostafa, "A 10 gb/s serdes transceiver," in *2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 389–393, IEEE, 2021.

[41] T. Yoneda and M. Imai, "A new encoding mechanism for low power inter-chip serial communication in asynchronous circuits," in *2015 33rd IEEE International Conference on Computer Design (ICCD)*, pp. 395–398, IEEE, 2015.

[42] R. Dobkin, R. Ginosar, and A. Kolodny, "Fast asynchronous shift register for bit-serial communication," in *12th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC'06)*, pp. 10 pp.–127, 2006.

[43] R. R. Dobkin, Y. Perelman, T. Liran, R. Ginosar, and A. Kolodny, "High rate wave-pipelined asynchronous on-chip bit-serial data link," in *13th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC'07)*, pp. 3–14, 2007.

[44] N. Qiao and G. Indiveri, "A clock-less ultra-low power bit-serial lvds link for address-event multi-chip systems," in *2018 24th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, pp. 93–101, 2018.