Power-Centric Design of High-Speed I/Os

Hamid Hatamkhani
UCLA
9 Marisol,
Newport Coast, CA 92657
1 (949) 246-6129
hhatam@icsl.ucla.edu

Frank Lambrecht
Rambus Inc.
4440 El Camino Real
Los Altos, CA 94022
1 (650) 776-7905
frankl@rambus.com

Vladimir Stojanovic
MIT
MIT 38-260
77 Massachusetts Ave
Cambridge, MA 02139
1 (617) 324-4913
vlada@mit.edu

Chih-Kong Ken Yang
UCLA
56-147A EIV
420 Westwood Plaza
Los Angeles, CA 90095
1 (310) 206-3665
yang@icsl.ucla.edu

ABSTRACT
With increasing aggregate off-chip bandwidths exceeding terabits/second (Tb/s), the power dissipation is a serious design consideration. Additionally, design of I/O links is constrained by a complex set of specifications such as voltage levels, voltage noise, signal deterministic jitter, random jitter, slew rate, BER etc. These specifications lead to complex tradeoffs for both circuits and circuit architecture in order to minimize power. This paper presents a design framework that enables the analysis of tradeoffs in the design of an I/O transmitter. The design framework includes BER analysis with a channel model coupled with logic sizing optimization that is constrained by the desired signaling specification.

Categories and Subject Descriptors
B.7.1,B.7.2 [**Integrated Circuits**]: Types and Design Styles – *advanced technologies, input/output circuits.* Design Aids – *simulation.*

General Terms
Algorithms, Performance, Design, Standardization.

Keywords: I/O, Serial Link, Power Minimization, Convex Optimization, Channel Model.

1. INTRODUCTION
Power dissipation has emerged as a focal point of integrated circuit design mostly due to the increasing density of features, performance, and device leakage. The power problem permeates all areas including the design of the I/O sub-system. A brief look at the ITRS roadmap [1] shows that one can anticipate an exponential scaling of off-chip data rates and packaging pins. Fig. 1 shows the aggregate bandwidth normalized to the current 90-nm process technology.[1] To illustrate the power scaling trend, Fig. 2 shows the energy cost per bit of transmitted data (energy/bit) of published I/O links. While the trend line shows an exponential decrease in power with device scaling, the scaling factor is lower than the scaling of the aggregate bandwidth indicating that the power cost of I/Os is increasing exponentially. Currently, I/O power can contribute >5% of the total power (i.e. >5W), for high-performance ICs [2].

Logic design methodologies have adapted their flow to incorporate more and more power analysis and optimization. Recent analysis [4] has shown that power dissipation can be quite sensitive to a slight relaxation of performance requirement. I/O links are in need of similar improvements on tradeoff analysis and power minimization. While many low-power circuit techniques have been introduced [5], a systematic optimization and analysis method is still needed. The paper begins with a brief introduction to the components of an I/O system and with specifications such as voltage levels, timing jitter, and bit error rate (BER). The power cost of I/Os are due to driving a signal through a lossy 50-Ω medium and the switching and amplification cost of the electronics. The next section presents a custom logic optimizer that is applied to the transmitter using I/O specifications as constraints to minimize energy cost. The paper next relates the energy cost with the BER using a statistical channel/receiver oracle that outputs the BER based on the I/O channel characteristics and noise conditions. Finally, examples of analysis that can be performed using this framework are discussed.

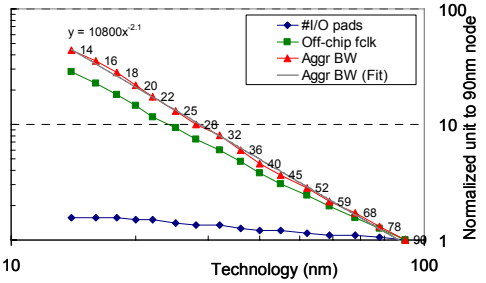

Fig. 1 Scaling of pad count, I/O rate, and aggregate bandwidth (normalized to 90-nm technology)

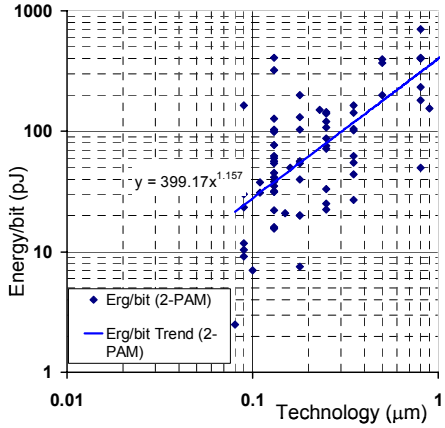

Fig. 2 Scatter plot of published energy/bit of I/Os

[1] The current aggregate data rate for the 90-nm node can be >0.5Tb/s for a high performance design [3].

DAC 2006, July 24–28, 2006, San Francisco, California, USA.

2. I/O SUB-SYSTEM

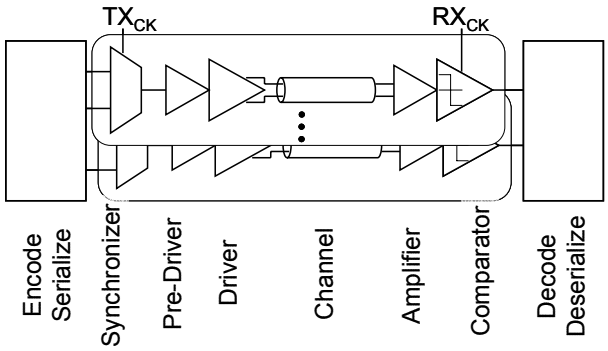

Fig. 3 Architecture of an I/O sub-system

Fig. 3 illustrates an I/O sub-system. The transmitter encodes[2] and serializes the on-chip data. Data stream is converted into well-conditioned multiple off-chip signals using a pre-driver and driver for each pin or pair of pins. The signals are distorted as they traverse lossy transmission medium ("channel"). A receiver amplifies and conditions each input signal. The signal waveform is sampled in time and quantized in voltage to result in digital values. The values are deserialized and retimed to accommodate the synchronization of the core logic. To ensure clean timing information, the synchronization at the transmitter uses a clean clock, TX_{CK}, commonly generated from a phase-locked loop (PLL). The data timing is also recovered at the receiver using a clock, RX_{CK}, that is synchronized to the transmitted signal using a PLL.

2.1 Signal Conditioning Circuits

A number of signal conditioning circuits are needed at both the receiver and the transmitter. First, since the signaling voltages are not the on-chip digital levels, the output driver must produce the correct voltage levels. Fig. 4 shows two common architectures: a high common-mode driver (HCM) that typically steers current and the transistors are kept in saturation, and a low common-mode driver (LCM) that typically uses the transistors as push-pull resistive switches to a low signaling supply. The driver devices and load resistances are carefully sized to maintain the desired output voltage and impedance. The driver and receiver have appropriate impedances to minimize signal reflections from any impedance mismatch of the transmission medium.

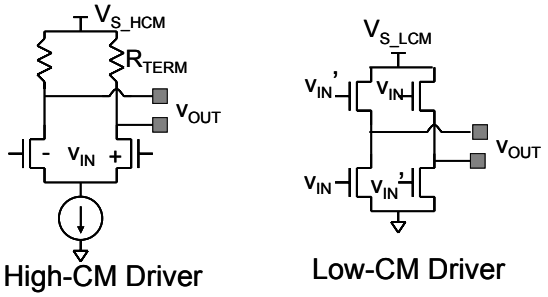

Fig. 4 Two common types of output drivers.

A second common signal conditioning circuit is a noise filter. At the transmitter, energy that is outside of the signal-bandwidth can interfere with other signals. Hence, sharp transitions containing high-frequencies are filtered by constraining the output slew-rate. At the receiver, to eliminate out-of-band noise, a low-pass filter is used.

A third circuit, one of increasing importance, compensates for the frequency characteristics of the channel. At high frequencies, even very short channels experience substantial high-frequency attenuation. Fig. 5 illustrates the frequency response of an 8" FR4 channel where the signal is attenuated by 8dB at 5GHz and 13dB at 7.5GHz for a data rate of 10Gb/s and 15Gb/s respectively. The compensation involves distorting the signal that is transmitted, or equalizing the received signal with a filter at the receiver. Fig. 6-(left) illustrates the basic concept of pre-emphasis that uses a compensation FIR filter at the transmitter. Fig. 6-(right) shows the pulse response of the pre-emphasis both at the input of the channel and at the output. Without the compensation, the pulse response broadens and exceeds the bit-time. Inter-symbol interference (ISI) is the effect of the interference due to consecutive pulses.

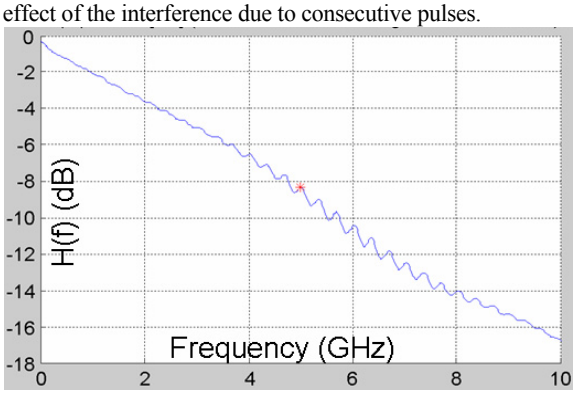

Fig. 5 Frequency characteristics of the transmission medium (8" of FR4 + flip-chip connection).

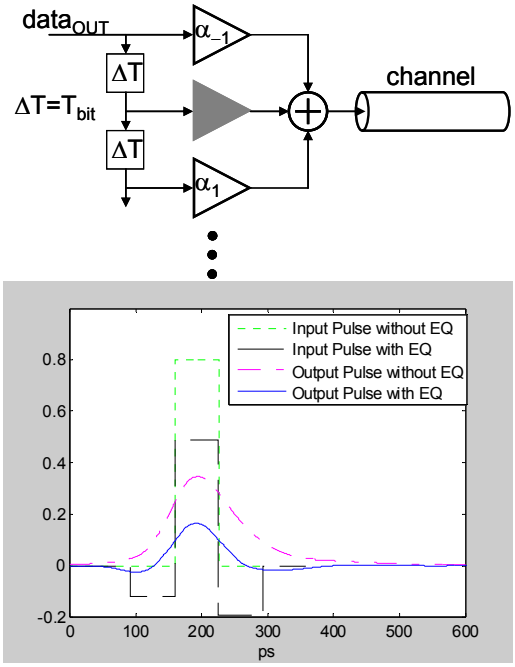

Fig. 6 (top) Architecture of transmitter pre-emphasis with 3 taps (1 pre- and 1 post-cursor tap). (bottom) Pulse response w/ and w/o pre-emphasis before and after the channel.

2.2 I/O Specifications

I/O links have commonly been specified by the data rate (or bit time, Tbit) and a data eye (eye mask) at the output of the transmitter

[2] Some encoding such as transition encoding may be used to embed phase information, further reduce power, or improve BER.

and/or the input of the receiver. The eye mask is used to define the amount of allowable timing noise and voltage noise. An example of an eye mask is shown in Fig. 7. Timing noise is expressed as "total jitter" (TJ). The TJ comprises of a deterministic component (DJ) and a random component (RJ).

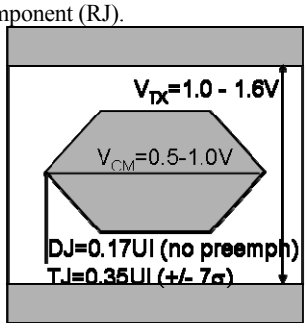

Fig. 7 Example of an eye-mask at the transmitter (Infiniband)

As the lossy channel increases its impact on the signal, while the timing and voltage noise are still important for performance, the shape of the eye-mask becomes less meaningful with significant amounts equalization. A metric that represents the robustness of an I/O link is the bit error rate, BER. Some links require extremely low error rates and hence can only be characterized by the amount of voltage margin allowable at the receiver before errors start to occur.

3. ENERGY COST OF I/O LINKS

The power needed for an I/O is the sum of the power cost of signaling through the low-impedance medium, and the power cost of the electronic circuits driving and receiving the signal. Power can be converted to the energy-per-bit for a given data rate.

The signaling power depends on the voltage levels and voltage swings in a specification. Fig. 8 shows the power cost as a function of swing for a signal with high common-mode that swings from an I/O supply voltage (Vdd). Power dissipation can be related to the voltage margin at the receiver (for BER=10^{-15}) by plotting the margin as a function of the voltage swing on the same figure for the 8" channel at 15Gb/s. The equivalent BER for the corresponding voltage margin is shown in the plot for one of the data sets.

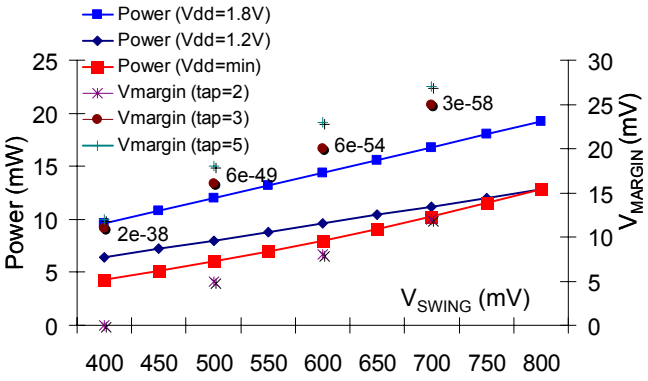

Fig. 8 Power cost of signaling with and HCM driver and the corresponding voltage margin at the receiver with varying equalization FIR.

The power cost of the electronics can be considerable and depends on the data rate and the processing technology. Similar to logic optimization, design considerations such as circuit architecture and sizing can lead substantial power tradeoffs. The remainder of this paper uses the transmitter to illustrate the design tradeoffs.

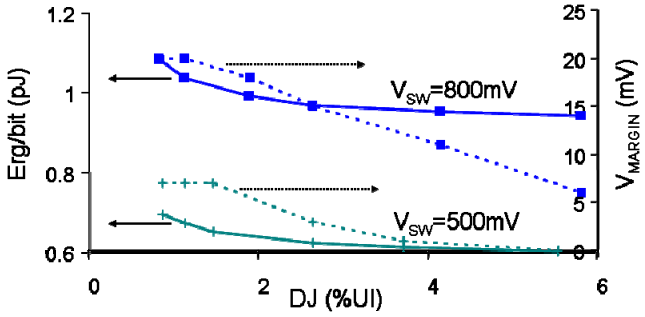

Fig. 9 Energy/bit versus the % DJ for signaling with 800mV and 500mV. The corresponding receiver voltage margin is on the 2nd y-axis.

An example of such a tradeoff is shown in Fig. 9 where substantial amount of power can be saved by relaxing the amount of deterministic jitter at the output of the transmitter. Fig. 9 also shows the impact of the DJ specification on the receiver voltage margin hence relating the power tradeoff with BER.

As shown in Fig. 10, the deterministic jitter at the transmitter is due to the data dependent delay of the pre-driver gates in the transmitter after the final stage of synchronization (i.e. a multiplexer or a flipflop). The delay variation is a function of the logic family, the circuit implementation, and most importantly the fanout of the logic (equivalently, the bandwidth of the logic gate). Sizing[3] can be optimized for each gate in a manner similar to the approach described in [4] except that the circuits are tailored for signal-integrity considerations (i.e. controlling the voltage swing, output impedance, and slew-rate, etc), and the optimization is constrained by the requirements of the I/O link [5].

The DJ is modeled as a difference in delay. The first-order model is expressed in (2), where τ is time constant of the system. For several stages of logic, we approximate deterministic jitter of a stage denoted by i, with an expanded model in (3). The A, B, C and D parameters are approximated by mean-square estimation and are fitted to simulated data from various logic blocks in the transmitter (i.e. static CMOS and pass-transistor logic gates). The delays (t_{di} and t_{di-1}) are modeled to within 15% error using an α-power model and the Elmore delay formula [11]. The model can also be adapted to analyze the impact of supply noise since supply variation can be modeled correctly in the α-power model.

$$DJ = -(\tau / T_{bit}) \ln(1 - e^{-T_{bit}/\tau}) \quad (2)$$

$$\begin{cases} DJ_i = DJ_{i-1} - (At_{di} \ln(1 - e^{-X}) + Ct_{di-1} \ln(1 - e^{-Y}))/T_{bit} \\ X = \dfrac{C(1 - DJ_{i-1})T_{bit}}{t_{di}}, \; Y = \dfrac{D(1 - DJ_{i-1})T_{bit}}{t_{di-1}} \end{cases} \quad (3)$$

Sizing is optimized using the delay model that satisfies final output eye constraints. The sizes are used to calculate the total power. The optimization result from our model can be compared with Spice simulations and shows good agreement with error $< \pm 10\%$.

[3] Bias and supply voltage optimization may also be needed depending on the circuit structure.

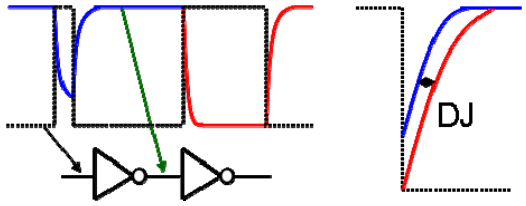

Fig. 10 Illustration of DJ caused by data-dependent delay variation of logic gates due to large fanout.

4. STATISTICAL LINK MODEL

In order to relate the BER to the voltage swing at the receiver, we use Rambus' LinkLab simulation and modeling environment as our stochastic channel oracle. This tool breaks the stochastic link modeling into three basic components: 1) interference distribution calculation, 2) effective receiver referenced voltage noise distribution from transmitter and receiver jitter, and 3) CDR dither jitter distribution caused by signal shape, interference and jitter [12][13]. Total received signal error can be described by Eq. (1), where x_k^{ISI}, x_k^{jitTx}, x_k^{jitRx} represent interference, and receiver voltage noise from transmitter and receiver jitter, respectively.[4]

$$x_k = x_k^{ISI} + x_k^{jitTX} + x_k^{jitRX} \tag{1}$$

$$x_k^{ISI} = \sum_{n=-sbS}^{sbE} b_{k-n} p_n$$

$$x_k^{jitTx} = \sum_{n=-sbS}^{sbE} b_{k-n}\left(h_{n-1}\varepsilon_{k-n+1}^{TX} - h_n \varepsilon_{k-n}^{TX}\right)$$

$$x_k^{jitRx} = \varepsilon_k^{RX} \sum_{n=-sbS}^{sbE} b_{k-n}\left(h_n - h_{n-1}\right)$$

Assuming random data, interference distribution is calculated as a convolution of data probability mass functions (pmfs), which are binomial, and each scaled in voltage axis by channel pulse response samples p_n. The resulting interference distribution calculated for each phase offset is shown in Fig. 11.

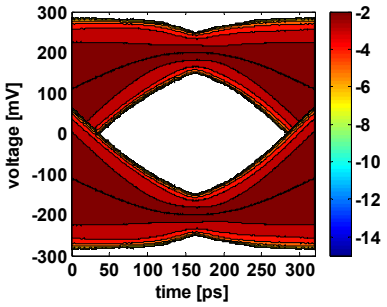

Fig. 11 Statistical cumulative distribution of the received signal vs. receiver phase. Probabilities <10^{-13} are shown as white.

Transmitter and receiver jitter are mapped to effective voltage noise at the receiver using the perturbation method on the convolution of jitter pulses, Fig. 12, with the channel impulse response. This approximation of the convolution integral relies on the large ratio of channel impulse response width and jitter pulse duration (Fig. 12) [12][13].

[4] Data pulses are labeled as b_k, edge jitter values as ε_k, channel pulse response samples as p_k, and channel impulse response samples as h_k. Indexes sbS and sbE mark the beginning and end of the channel pulse/impulse response.

This method allows us to separately analyze the impact of different types of jitter. Gaussian random jitter is most straightforward to model as the effective voltage noise variables x_k^{jitTx} and x_k^{jitRx} will be also Gaussian and will accurately capture any correlated data and jitter values. In addition to the Gaussian jitter, we can also model the impact of arbitrary jitter distributions, under the assumption that jitter samples are independent. An example is the jitter as a result of DJ caused by the transmitter. Fig. 13 shows an example of the distribution. In the case of arbitrary jitter distribution, we can calculate the x_k^{jitTx} and x_k^{jitRx} distributions as convolutions of jitter sample pmfs, scaled by $b_k h_{n-k}$ values (note that this procedure is identical to the interference distribution calculation). In our previous work [12], we show that independent (uncorrelated Gaussian) jitter assumption results in the largest rms values for x_k^{jitTx}, hence this assumption makes the oracle a bit conservative which is good. We will use this extension of the tool to capture the effects of random data jitter and supply noise-induced jitter on the link performance.

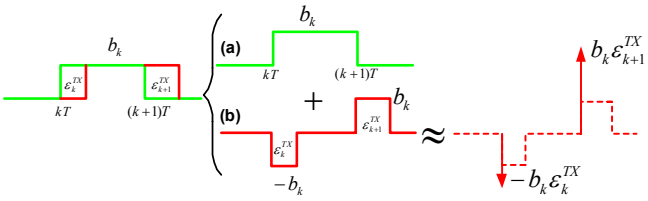

Fig. 12 Jittered pulse decomposition. A symbol transmitted with jitter is converted to a symbol with no jitter (a), plus a noise term where the noise-symbol's widths (b) are equal to ε_k^{TX} and ε_{k+1}^{TX}.

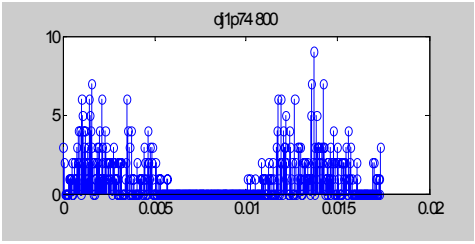

Fig. 13 Probability density of transition edges due to DJ at the output of the transmitter. X-axis is the UI (normalized T_{bit}).

Having calculated these distributions for the total receiver referenced noise (consisting of receiver thermal voltage noise and effective voltage noise from transmitter and receiver jitter) we can use the interference distribution as conditional distribution for the total receiver referenced noise. Receiver offset and overdrive requirements are also included in decision probability calculation as decision threshold shifts and meta-stable decision boundary.

Decision probabilities on data and edge samples differ since edge samples are only meaningful on valid transitions, which fix the values of the neighboring bits. The ISI and receiver noise distributions have to be calculated with that in mind (since the calculations rest on the independence assumption). Different edge possibilities for the two fixed transition bits are calculated combinatorially and then used to condition the ISI and noise distributions originating from other symbols in the sequence (for which the independence assumption still holds). Decision probabilities on data and edge samples drive the bang-bang clock and data recovery loop, where they are filtered into up/down/hold transition probabilities in the Markov-chain representation of the loop state machine as shown in Figure 14.

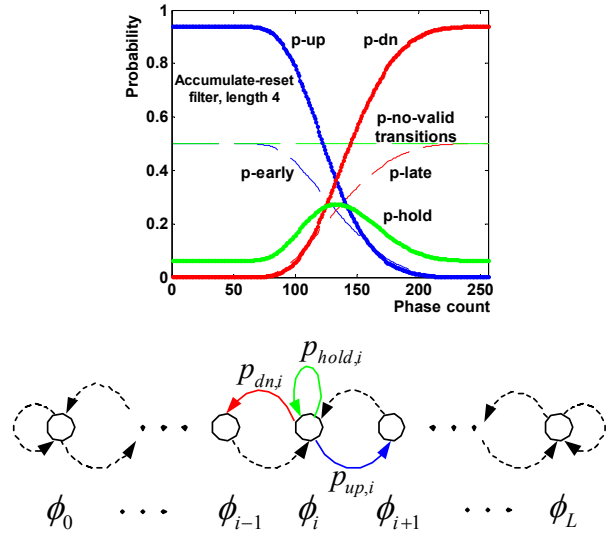

Fig. 14 (top) The raw input probabilities (p-early, p-late, p-no-valid transition) are converted by a filter to state transition probabilities (p-up, p-dn, p-hold) for each possible phase position; (bottom) First-order Markov chain phase-state model. Each state represents a different phase position, and the arcs are the probability of transition, given that position.

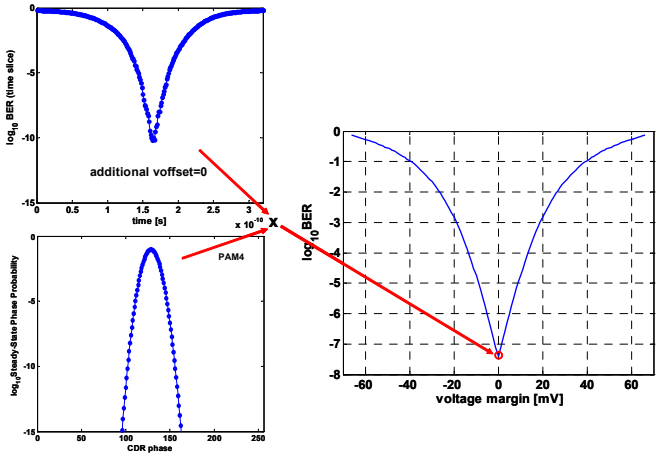

Fig. 15 Computing the BER by combining the time conditioned BER (upper-left: bathtub curve) and CDR phase probabilities (lower left).

Converging these phase-state transition probabilities to steady state yields CDR dither jitter histogram. The final bit error rate is obtained when this dither jitter histogram is used to condition the bit-error rate calculations from interference and random noise, at each possible CDR phase offset. The flow of probability conditioning is illustrated in Fig. 15.

Since the conditional BERs are computed with receiver thresholds set at nominal levels with receiver sensitivity of 10 mV, the computed BER has no additional voltage margin. If in turn, we compute the conditional BERs assuming some additional voltage offset, for example 50mV, then the computed aggregate BER would have an additional voltage margin of ±50mV depending on the sign of the applied offset.

5. ANALYSIS EXAMPLES

By combining the sizing optimization with LinkLab, better power tradeoff analysis can be performed. The results of analysis shown in Fig. 9 illustrate that the receiver voltage margin degrades roughly linearly with increasing DJ for large DJ percentages. However, interestingly, decreasing DJ unnecessarily (from 2% to 1%) can increase power by 30%, while the impact on voltage margin is nearly negligible. This result indicates that the designer's inclination to eliminate DJ at the transmitter output is an over-design that can substantially penalize power dissipation.

An architectural analysis using this methodology can evaluate the energy cost of compensating for the channel. For instance, a transmitter pre-emphasis design that uses an FIR filter (i.e. shown in Fig. 6) requires additional logic per I/O pin for each additional tap. As shown in Fig. 16, the improvement in voltage margin for 2 additional post-taps is 2mV while the energy cost is 40fJ/bit. Fig. 8 implies the default method of improving the voltage margin, by simply increasing the voltage swing. The receive voltage margin increases proportionally with the transmitted swing as well as the transmitter power. The energy cost per mV of receiver voltage margin can be estimated from the figure to be roughly 20fJ/bit/mV at 15Gb/s. So the added complexity from 3 taps to 5 taps is no better in terms of energy than simply increasing the voltage swing. However, eliminating the precursor tap reduces the margin by 10-mV and can be recovered only with a substantial increase in voltage swing and a 200-fJ/bit cost in energy.

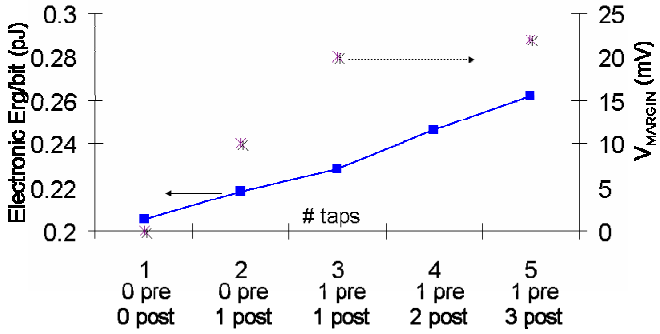

Fig. 16 The energy cost of adding a multi-tap FIR pre-emphasis filter at the transmitter. The 2nd axis shows the corresponding receiver voltage margin.

A second architectural analysis can evaluate the impact of changing the position of the synchronizing multiplexer in a transmitter, as shown in Fig. 17. By moving the clocked multiplexer closer to the output, the amount of accumulated DJ decreases at the cost of increasing in power. As shown in Fig. 18, the energy cost for our 90-nm process technology is 40fJ for the same DJ. Interestingly, while the DJ statistical distributions of the two multiplexer positions are different, the resulting impact on the receiver voltage margin is roughly the same. The relationship is still linear and only a slight change in the sensitivity of the voltage margin to DJ (difference in the slope of the curve) is observed.

A third architectural analysis can compare the energy cost of different driver architectures. The two common types of output drivers are shown in Fig. 19. The signaling energy of an LCM driver can be substantially lower than an HCM driver. For instance, for 500mV of differential swing, an HCM driver can dissipate between 6-8mW depending on the supply voltage, while an LCM driver dissipates only 1.25mW. However, an LCM driver requires twice as many devices and larger devices in order to switch the voltage with low impedance. Hence, for large DJ values where on-

chip fanouts are large, an LCM design maintains small energy cost. However, the large DJ degrades the voltage margin to 0 and results in high BER. For DJ below 3% UI, the energy cost of an LCM design increases substantially and exceeds that of an HCM driver. Similar to the previous analysis, even though the DJ jitter distribution is considerably different for an HCM and an LCM driver, the relationship between DJ magnitude and voltage margin is quite similar.

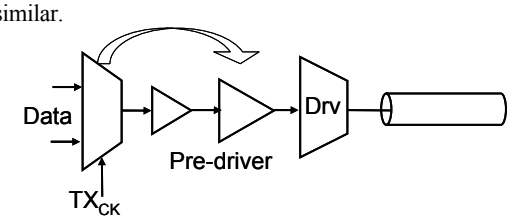

Fig. 17 Possible synchronizing/multiplexing positions.

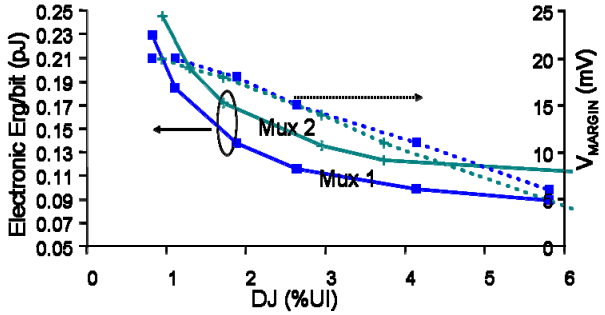

Fig. 18 Energy costs of different multiplexer positions as a function of DJ. The 2nd y-axis is the voltage margin.

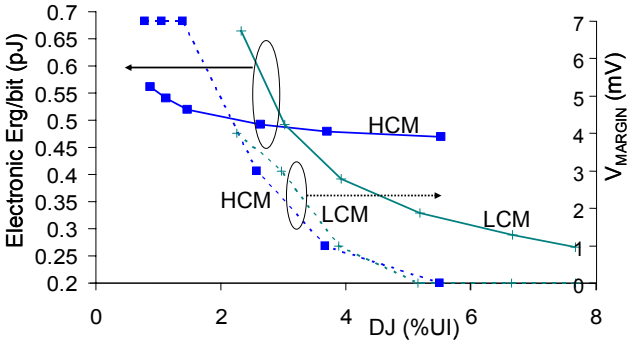

Fig. 19 Comparison of the energy/bit (with associated receiver voltage margin) for LCM (green) and HCM (blue) architectures.

6. CONCLUSION

As power becomes increasingly important in each portion of an IC, additional methodology is needed to evaluate the tradeoff between power and performance of an I/O sub-system. Similar to logic design, the power performance of an I/O link depends on the process technology and the sizing of the circuits. In addition, there is a strong dependence on the channel characteristics and the circuit architecture for better signal integrity. In order to properly evaluate power tradeoffs, an analytical framework is presented that includes size optimization and modeling of delay variation coupled with a stochastic channel oracle that determines the receiver voltage margin for a given channel and input signal. This methodology ties transistor-level design to I/O system specification such as error rate. The paper illustrates the potential benefit of such a framework through several examples of power tradeoffs for a 15-Gb/s I/O link through an 8" FR4 channel.

7. ACKNOWLEDGMENTS

The authors would like to acknowledge the funding from UC Micro. Special thanks to Rambus Inc. for the use and adaptation of LinkLab, I. Stojanovic, and P.L. Yang for their support.

8. REFERENCES

[1] International Technology Roadmap Service, http://public.itrs.net

[2] J. Hart, et.al., "Implementation of a Fourth-Generation 1.8GHz Dual-Core SPARC V9 Microprocessor", JSSC 01/06, pp210

[3] D.C. Pham, et.al., "Overview of the Architecture, Circuit Design, and Physical Implementation of a First-Generation Cell Processor", JSSC 01/06, pp178

[4] D. Markovic, et.al., "Methods for True Energy-Performance Optimization", JSSC 08/04 pp1282-93

[5] K.L. Wong, et.al., "A sub-30mW 3.6Gbps Transceiver", JSSC, 04/04, pp 602-12

[6] J. Kim, M. Horowitz, "Adaptive Supply Serial Links With Sub-1-V Operation and Per-Pin Clock Recovery," JSSC 11/02, pp 1403-13

[7] Y. Moon, et.al., "A Quad 6Gb/s Multi-rate CMOS Transceiver with TX Rise/Fall-Time Control", ISSCC 02/06 pp 84-5

[8] B. Casper, et.al., "A 20Gb/s Forward Clock Transceiver in 90nm CMOS", ISSCC 02/06, pp 90-1

[9] A. Emami-Neyastak, et. al., "A Low-Power Receiver with Switched-Capacitor Summation DFE", VLSI Circuit Symposium 06/06, to appear

[10] K.J. Wong, et.al., "A 5-mW 6-Gb/s Quarter-Rate Sampling Receiver with a 2-Tap DFE Using Soft Decisions", VLSI Circuit Symposium 06/06, to appear

[11] T. Sakurai, A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas" JSSC 07/92.

[12] V. Stojanovic, M. Horowitz, "Modeling and Analysis of High-Speed Links", CICC '03, pp.589-94

[13] V. Stojanovic, et.al. "Optimal linear precoding with theoretical and practical data rates in high-speed serial-link backplane communication", ICC'04, pp.2799-2806