Prediction and Comparison of High-Performance On-Chip Global Interconnection

Yulei Zhang, *Student Member, IEEE*, Xiang Hu, *Student Member, IEEE*, Alina Deutsch, *Fellow, IEEE*, A. Ege Engin, *Member, IEEE*, James F. Buckwalter, *Member, IEEE*, and Chung-Kuan Cheng, *Fellow, IEEE*

Abstract—As process technology scales, numerous interconnect schemes have been proposed to mitigate the performance degradation caused by the scaling of on-chip global wires. In this paper, we review current on-chip global interconnect structures and develop simple models to analyze their architecture-level performance. We propose a general framework to design and optimize a new category of global interconnect based on on-chip transmission line (T-line) technology. We perform a group of experiments using six different global interconnection structures to discover their differences in terms of latency, energy per bit, throughput, area, and signal integrity over several technology nodes. Our results show that T-line structures have the potential to outperform conventional repeated RC wires at future technology nodes to achieve higher performance while using less power and improving the reliability of wire communication. Our results also show that on-chip equalization is helpful to improve throughput, signal integrity, and power efficiency.

Index Terms—On-chip global interconnect, passive equalization, performance prediction, transmission line.

I. INTRODUCTION

AS semiconductor technology advances in the ultra deep sub-micrometer (UDSM) era, on-chip global interconnect has become an ever-greater barrier to acheiving the performance requirements of increasingly large system-on-chip (SoC) designs. Shrinking of wire geometries results in greater per-unit-length resistance. Even with a shrinking dielectric constant, an increasing RC delay per unit wire length is observed as technology scales. Meanwhile, the average length of global wires, determined by chip size, remains fixed as technology scales due to increasingly compelex SoC designs. According to the ITRS roadmap [1], the RC delay of 1-mm-long, minimum pitch global wire will be 542 ps at the 45 nm node, while the 10 level fan-out of 4 (FO4) delay of minimum sized inverter will be 145 ps at the same node. A substantial performance gap is growing between global interconnect and logic gates.

Global wires also consume a significant portion of the total power in digital systems. In [2], Magen *et al.* found that interconnect power accounted for half the total dynamic power in a 0.13-μm microprocessor designed for power efficiency. Further, nearly one-third of the total power was dissipated in global wires, comprising global clocks and signals. The widely-used repeated RC wire structure for global interconnect requires significant power overhead because it uses strong repeaters to drive relatively short wire segments [3]. As shown in [4], to minimize total latency, the optimal repeated structure has equal amounts of wire and gate capacitance, which means that half the total dynamic power is dissipated in repeaters.

To break the "interconnect wall" caused by the scaling of global wires, many approaches have been proposed to hasten on-chip global communication. The repeater insertion method has been widely adopted [5]. By breaking the long wire into segments and adding buffers, the repeater insertion method reduces total wire delay at the cost of additional power overhead. To further reduce latency and energy per bit, transmission-line (T-line) effects of on-chip wires have been utilized by adopting fat top-layer wires driven by low impedance transmitters [6]. However, the inter-symbol interference (ISI) due to the resistive loss severely limits the bandwidth of such T-line schemes.[1] To counter ISI and increase throughput density, equalization techniques have been employed [7]. Different approaches have been proposed using passive [8], [9] or active components [7], [10] to build equalized T-line structures for high-throughput on-chip global communication.

In this work, six global interconnection structures are explored and their performance compared across multiple technology nodes. Extending a previously published conference paper [11], we add the following features: 1) as a means to improve the throughput of repeated RC wires, pipelined RC wire is analyzed and compared with other global interconnect structures; 2) chip areas consumed by different global interconnect structures are modeled and discussed; and 3) wire length is added as a new variable in performance models to capture and study the critical length of different interconnect schemes in terms of specific performance metrics.

The rest of this paper is organized as follows. In Section II, the various global interconnect structures are introduced in

[1]In this work, bandwidth of interconnect is defined as the highest signal frequency that the whole interconnection system can support in order to meet specific voltage swing requirement (e.g., full-swing for repeated RC wire and minimum detectable voltage for T-line schemes) at the receiver side.
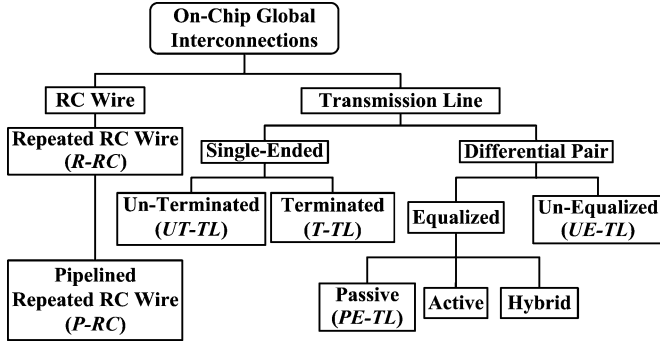
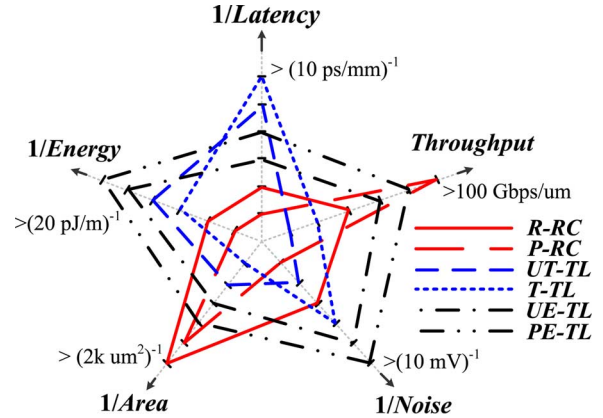Fig. 1.   Organization of on-chip global interconnect structures.

Fig. 2.   Multidimensional design tradeoffs of different global interconnect structures. For each interconnect structure, the more area the pentagon covers, the better overall performance can be achieved.

detail, and approximate analysis is performed to model performance metrics for different structures. Section III discusses the design methodology for on-chip T-line interconnect and a framework to optimize such schemes is proposed. Performance prediction results of different interconnect schemes are shown in Section IV. In Section V, we discuss the signal integrity of T-line interconnects, focusing on crosstalk effects. Finally, Section VI concludes, highlighting several general observations on performance scaling trends of global interconnect.

II. ON-CHIP GLOBAL INTERCONNECT

This section begins with an overview of the design considerations of different on-chip global interconnect structures. Six chosen interconnect schemes are then detailed, including the design and modeling of wires and transceivers. Finally, architecture-level performance metrics of the various structures are analyzed and corresponding scaling trends are discussed.

A. Overview

On-chip global interconnect schemes can be divided into categories based on the operating region of wires, the signaling method, and other factors, as shown in Fig. 1. The widely-used repeated RC wire approach (referred as $R-RC$ in this paper) belongs to the first category, which uses RC-mode dominant wires. To improve the bandwidth of repeated RC wires, the R-RC structure may be pipelined by breaking optimized R-RC wire into segments and inserting flip-flops. This pipelined RC wire strategy is subsequently denoted $P-RC$. The other main category utilizing T-line effects of on-chip wires is comprised of two configurations, namely single-ended and differential pair, based on their respective signaling methods. For the single-ended configuration, capacitive or resistive loading (unterminated or terminated T-line, referred as $UT-TL$ or $T-TL$) can be used at the wire end depending upon the throughput requirement [12]. For the differential pair configuration, conventional design mainly focuses on the optimization of T-line transceivers without adopting any equalization (referred as $UE-TL$). Passive networks [8], [9] are used in some recent research to equalize on-chip T-line (referred as $PE-TL$) whereas other on-chip equalization implementations using active circuits or even hybrid structures could be potential future research directions.

Multidimensional design tradeoffs, which are normally related to the latency, energy dissipation, throughput, area/cost, and reliability (noise), should be considered while designing an on-chip global interconnect scheme. For the six different structures mentioned above, we use a 45-nm CMOS process as an example to illustrate the tradeoff relations along multiple performance dimensions in Fig. 2. By observing this figure, designers can easily identify complex design tradeoffs and make determinations based on given specific applications. It can be seen that, by using 45 nm CMOS, RC wire has advantages in throughput density (using P-RC) and area/cost (for both R-RC and P-RC) because of their small wire dimensions. On the other hand, single-ended T-lines (UT-TL and T-TL) could be used for low-latency application by utilizing wave propagation. In terms of low energy and noise, differential T-lines (UE-TL and PE-TL) should be better candidates due to their larger wire input impedance, low-power transceiver circuitry, and differential configuration. In order to identify these complex tradeoff relations at the early-stage and also from the architecture-level, we have developed simple performance models to help designers to do approximate but trend-following estimation, which will be discussed later in Section II-D.

B. Interconnect Schemes

We show the detailed structure for each global interconnection scheme mentioned in Section II-A and briefly introduce the features of these schemes as follows.

For repeated RC wires (R-RC), the long wire is divided by repeaters into several RC segments. The strength of repeater (size of inverter) and length of wire segments could be optimized according to different design objectives for a given wire geometry. To further improve the bandwidth, P-RC is proposed as shown in Fig. 3. Assuming the R-RC wire between two flip-flops is **already optimized** based on one specific objective (minimum latency in this study), the only variable for P-RC wire optimization is the number of flip-flops inserted (a.k.a. pipelining depth). By utilizing pipeline, bandwidth of the R-RC wire is improved with overhead of energy and latency, therefore, the best pipelining depth can be decided in terms of the lowest
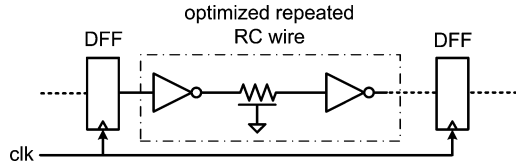
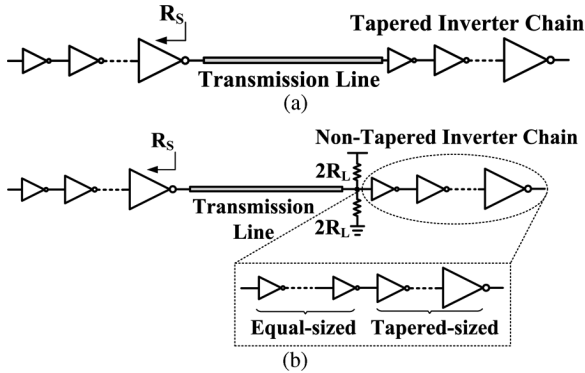Fig. 3. One stage of pipelined repeated RC wire (P-RC structure).

Fig. 4. Single-ended T-line schemes for on-chip global interconnect. (a) Unterminated T-line scheme. (b) Terminated T-line scheme.
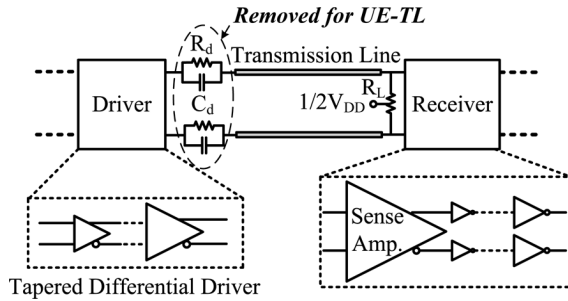

Fig. 5. Differential T-line schemes for on-chip global interconnect.

TABLE I
DESIGN PARAMETERS FOR GLOBAL R-RC AND P-RC WIRES BASED ON ITRS ROADMAP 2007 AND SPICE SIMULATION

Tech Node/nm	90	65	45	32	22
supply voltage/(V)	1.2	1.1	1.0	0.9	0.8
dielectric constant ϵ_r	3.1	2.9	2.6	2.4	2.1
ρ_{Cu}/($\mu\Omega \cdot cm$)	2.53	2.73	3.10	3.52	3.93
min global pitch/(nm)	300	210	135	96	75
aspect ratio (A/R)	2.2	2.3	2.4	2.5	2.6
min-inv FO4 delay*/(ps)	11.9	8.4	4.6	2.8	1.6
flip-flop T^*_{c-q}/(ps)	34	24	13	8	4
flip-flop T^*_{setup}/(ps)	22	15	10	6	3
flip-flop C^*_{DFF}/(fF)	34.3	23.8	17.0	9.0	4.3

*Data are obtained by simulation using a predictive model [14].

For the differential T-line schemes (UE-TL and PE-TL, shown in Fig. 5), the tapered differential drivers[2] could be used to provide the low driver impedance, whereas at the receiver side, a sense-amplifier (SA), based on the work [13], is adopted to amplify the attenuated signal and the following inverter chain further increases the slew rate to improve signal quality. Circuit design of T-line receiver has been discussed in the previous work [9]. In this work, we improve the design there to facilitate the SA bandwidth by using more sophisticated transistor-sizing strategy, which improves the bandwidth by about 2× compared with design results shown in [9]. Since the receiver is designed and optimized for each given technology individually, noise and sensitivity performance (capability of recovering 50 mV voltage difference) of receiver is guaranteed even for smaller technology node by automated transistor-sizing. As a result, real performance change of receiver circuit with technology scaling is modeled and considered during the estimation of whole T-line structures in this work. For the equalization approach, the passive-equalized scheme adopts a parallel RC network at the driver side to flatten the overall frequency response by utilizing the high-pass characteristic. In this scheme, driver impedance, resistance and capacitance value in the passive equalizer, and terminated resistance are optimized with the constraint that enough eye-opening should be observed at the wire-end in order to be safely captured by the receiver.

C. Global Wire Modeling and Implementation

We model on-chip global wires using different approaches based on operating frequency and wire geometry.

Global RC wires in scheme R-RC are normally represented by distributed Π model composed of wire resistance and capacitance. Following [4], 2-D closed-from equations in [15] are utilized to calculate wire capacitance. The wire geometry and other design parameters for R-RC structure are listed in Table I, based on the predictions of the 2007 ITRS roadmap [1]. For P-RC structure, flip-flop parameters including the clock-to-q time, setup time, and effective capacitance are derived by SPICE simulation using a predictive device model, as listed in the last three rows of Table I.

For other T-line schemes, we adopt single-ended or differential strip-line configurations to model global wires, as shown

energy over bandwidth ratio (conceptually similar to the energy-delay product, refer to Appendix A for the mathematical derivation). In practice, there is an upper-bound for the maximum pipelining depth, so in the following experiments, P-RC is optimized based on **the lowest energy/bandwidth with the maximum pipelining depth constraint**.

For the single-ended T-line schemes (UT-TL and T-TL) which are shown in Fig. 4, tapered or non-tapered inverter chain is adopted as the driver and receiver, depending on different termination scenarios. Compared with unterminated scheme [as shown in Fig. 4(a)], resistive termination improves the bandwidth by alleviating the ISI, but lowers the swing of output signal and burns extra power on the termination. As a result, a non-tapered inverter chain [as shown in Fig. 4(b)] is devised to amplify the received signal and recover it back to digital level. In this kind of single-ended schemes, driver impedance (and terminated resistance, if any), first inverter size, and number of stages in the inverter chain are the variables to be optimized during the design.

---

[2]Drivers could be CML or other types, but in following optimization and experiments, differential T-line drivers are assumed to be voltage sources with output resistance R_s to simplify the analysis and optimization.
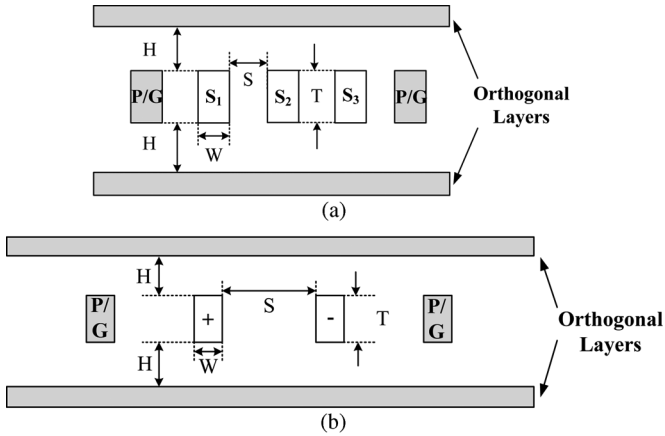
Fig. 6. Wire configurations for on-chip T-line schemes. (a) Cross section of single-ended stripline. (b) Cross section of differential stripline.

TABLE II
PARAMETERS OF ON-CHIP GLOBAL T-LINE USED IN UT-TL/T-TL SCHEMES

W (μm)	S (μm)	H (μm)	T (μm)	R (Ω/mm)	Z_0 (Ω)	f_{LC}^* (GHz)
1.6	1.6	3.2	2.4	6.0	40	1.79

* f_{LC} indicates the corner frequency of RC and LC region.

in Fig. 6. For single-ended scenarios, we insert power/ground (P/G) lines every three wires (shown in Fig. 6(a), following the typical wiring and power arrangement for global wide data bus [16]) to provide current return paths in order to form well-controlled on-chip T-line structures. The adjacent orthogonal layers could be replaced by the ground planes if performing 2-D capacitance extraction. Considering orthogonal loading, the capacitance obtained by a 2-D extraction is slightly overestimated compared with the 3-D value [17], but still acceptable with the assumption that on-chip loading density and lateral wire-to-wire coupling are very high.[3] The dimensions of this single-ended T-line structure are listed in Table II, following the settings in [12]. Fat and unscaled wires implemented on the top-layer are utilized to reduce the resistive loss in this scenario, first proposed in [18], to alleviate the increasing RC delay of scaled on-chip wires. With the improvement of device speed, the transmission-line effect does kick in and cannot be neglected while modeling such fat global wires; on the other hand, it has been verified by previous research works [6], [12], [17] that on-chip bus configuration comprising a low-impedance driver and uninterrupted fat wire outperforms repeated wire structures in high-performance applications (e.g., high-end processors [16]) due to the T-line effect. As a result, we adopt the configuration comprising a low-impedance driver and uninterrupted fat wire and assume the utilized wire geometry shown in Table II maintains as technology scales. As shown in the last column of Table II, LC-mode behavior is dominant for this wire geometry, which speeds up the signal transmission through wave propagation.

[3]For the narrower lines traveling in the same layer, this is a reasonable approximation. As shown in [6], the impact of orthogonal layers on the capacitance extraction depends on the wire geometries. When t/w ratio is large, coupling capacitance changes a little with different orthogonal loading scenarios, also alleviating the internal reflection due to non-uniform capacitance distribution of on-chip T-line in the practical cases.

TABLE III
DESIGN PARAMETERS FOR UE-TL/PE-TL SCHEMES (WIRE LENGTH = 5 mm)

Tech Node/nm	90	65	45	32	22
SA bandwidth/(GHz)	12.5	20.0	40.0	66.7	125
UE-TL wire width/(μm)	0.350	0.350	0.400	0.448	0.504
PE-TL wire width/(μm)	0.300	0.315	0.375	0.400	0.444

$H = T = S = 2W$ for T-lines used in UE-TL and PE-TL.

We also devise a similar coplanar configuration for differential T-lines as shown in Fig. 6(b). Here, only one pair of wires is placed in a P/G bay in order to reduce the crosstalk noise. Wire dimensions of such configuration are determined by the resistive loss at given signal frequency, which is changing with the technology. Considering the differential T-line schemes discussed in this work, the overall signal bandwidth is limited by the SA in transceiver, as listed in Table III at each technology node. We derive the minimum wire widths of differential T-lines that satisfy the eye-opening constraint by binary search and SPICE simulations, as shown in the third and forth row of Table III.[4] By comparison, it can be seen that equalization helps to improve the data density by supporting narrower wires at the same bit rate.

The modeling and simulation of on-chip T-lines generally incorporates two steps. First, we extract the frequency-dependent RLGC parameters for the given wire structure using field solvers. For on-chip wires, since dielectric loss can be ignored and wire capacitance is basically frequency independent, $R(f)L(f)C$ extraction is generally performed [6]. The frequency-dependent impedance extraction requires a group of P/G wires located in a signal layer and sub-adjacent layers (parallel to the signal layer) to serve as return paths in order to capture the wide-band characteristic of wire inductance [17]. As a result, parameter extraction can generate the tabular model or other kinds of SPICE-compatible macro-model. In the second step, SPICE simulations can be performed to study transient characteristics and signal integrity. In this work, we evaluate performance metrics of all the T-line schemes based on the tabular model generated by 2-D-$R(f)L(f)C$ extraction. As a more practical modeling approach, a stable compact circuit model [19] synthesized from 2-D-$R(f)L(f)C$ extraction is used to study the signal integrity of global T-line schemes, as discussed in Section V.

The previously discussed T-line structures can be implemented using CMOS process. For single-ended schemes, T-lines are implemented on the top-layers of copper stacking with well designed power/ground arrangement to control the T-line effect, as shown in the bus design for high-end processors [16]. Differential T-lines also have been implemented recently for global clock distribution [20]. Similar configuration can be borrowed here to implement global differential data bus [21]. There is no any further specific layout style required for such T-line configuration, however, to improve the signal integrity (e.g., crosstalk), twisted structure might be used for the real chip designs [20], [22].

[4]The wire width values in this table are different from the previous work [11] because of the improved wire width optimization subroutine in this work. It is shown that narrower T-lines can be utilized to satisfy the eye-opening constraint, resulting in higher throughput density in the following results and also affecting other metrics.

TABLE IV
MODELING PERFORMANCE METRICS (NORMALIZED DELAY, NORMALIZED ENERGY, NORMALIZED THROUGHPUT, AREA) OF SIX GLOBAL INTERCONNECTION
STRUCTURES USING TECHNOLOGY-DEFINED PARAMETERS

Structures	R-RC/P-RC	UT-TL/T-TL	UE-TL/PE-TL
Normalized Delay	$K_1\sqrt{\epsilon_r/\tau} + K_2\tau/L$	$K_1\sqrt{\epsilon_r} + K_2\tau/L$	
Normalized Energy	$K_1\epsilon_r V_{DD}^2 + K_2\tau V_{DD}^2/L$	$K_1 + K_2 V_{DD}^2/L$	$K_1 V_{DD}^2 + K_2 V_{DD}^2/L$
Normalized Throughput	$1/(K_1\tau^2 + K_2\sqrt{\epsilon_r\tau}L)$	$K_1 1/\tau + K_2 1/L$	$1/(K_1 L\sqrt{\tau} + K_2\sqrt{L})$
Area	$K_1\tau L + K_2\tau^2$	$K_1 L + K_2\tau^2$	$K_1 L^{3/2} + K_2 L^2/\sqrt{\tau}$

D. Performance Analysis

An approximate high-level analysis is performed here to reveal how architecture-level performance metrics of different global interconnects behave with technology scaling under the **min-d** (minimum delay) objective.[5] As a result, we derive simple linear models, which can be used for designers to approximately estimate the performance of different interconnect structures at the early stage. In the following analysis, basic technology-determined parameters, including supply voltage V_{DD}, dielectric constant ϵ_r, min-sized inverter FO4 delay τ, as well as the total wire length L, are chosen to be the variables to build such models.

1) Latency: For the latency evaluation, we define the **normalized delay** as

$$\mathrm{delay}_n = \frac{\text{propagation delay}}{\text{wire length}} \qquad (1)$$

where the propagation delay includes both the wire delay and the gate delay (repeaters and flip-flops in RC wires or transceivers in other T-line schemes).

For R-RC structure, it can be shown that [4]

$$\mathrm{delay}_n^{R-RC} \propto \sqrt{r_0 c_{\mathrm{nmos}} r_w c_w} \propto \sqrt{\epsilon_r/\tau} \qquad (2)$$

with the assumption that the output resistance of a min-sized inverter r_0 is roughly constant across different technologies, and FO4 delay τ reduces with the same scaling factor as the feature size. For P-RC structure, additional delay is introduced by the inserted flip-flops and is linearly proportional to the pipelining depth and FO4 delay τ. For long global RC wires (≥ 5 mm) and advanced technologies (beyond 45 nm), the experimental results show that the maximum pipelining depth is chosen to reduce the overall energy over bandwidth ratio. Therefore, in our simple models, latency overhead of P-RC wire is assumed to be a linear function of FO4 delay τ only.[6]

For other T-line schemes, total latency can be expressed as the sum of wire delay and transceiver delay. For LC-mode dominant T-lines, normalized wire delay is proportional to $\sqrt{\epsilon_r}$. The transceiver delay could be simply represented by the FO4 delay τ linearly.

Considering the total wire length in our delay models, the final results are shown in the second row of Table IV, where coefficients K_1 and K_2 reflect the different process technologies. It can be seen that if $\sqrt{\epsilon_r/\tau}$ item is dominant in the normalized

delay expression for the R-RC and P-RC structures, the trend is an increase in the normalized delay of RC wires as technology scales. The same table shows the opposite trend for the T-line schemes, where normalized delay decreases with the reduction of dielectric constant and scaling of transistors.

The most significant sources of error in our proposed delay model come from the simple modeling of transistor gate capacitance and the approximation of the P-RC scheme in the short wire-length range. As technology scales, the gate capacitance per unit width actually reduces instead of being constant,[7] resulting in the decreasing c_{nmos} in (2), which partly cancels out the RC-wire slowing caused by the wire scaling. Modeling error of the P-RC scheme in the short wire-length range ($L \leq 3$ mm) is due to the neglect of optimal pipelining depth changing in the delay model. As discussed in Appendix A, the optimal pipelining depth in terms of lowest energy-bandwidth ratio reduces as the wire length decreases, and is proportional to the wire length. Therefore, for the cases with short wire length, the τ/L item in the delay model of P-RC structure approaches τ, causing relatively larger errors. This type of error source also applies to energy and throughput modeling of P-RC structure, as shown below.

Using the proposed delay models, the average percent error for RC wires is less than 15% and maximum percent error is limited by 30%. For other T-line schemes, maximum and average error are less than 13% and 5%, respectively.

2) Energy per Bit: The **normalized energy per bit** is used to evaluate the energy dissipation of global interconnect, which is defined as follows:

$$\mathrm{energy}_n = \frac{\text{energy per bit}}{\text{wire length}} = \frac{\text{power}}{\text{bit rate} \times \text{wire length}}. \qquad (3)$$

The bit rate for RC wires is assumed to be the inverse of propagation delay over the total wire length for R-RC structure (not pipelined), or the inverse of delay between two flip-flops for P-RC structure (pipelined).

As discussed in [4], the normalized energy per bit for R-RC structure satisfies that

$$\mathrm{energy}_n^{R-RC} \propto c_w V_{\mathrm{DD}}^2 \propto \epsilon_r V_{\mathrm{DD}}^2. \qquad (4)$$

For the P-RC structure, additional energy consumed by inserted flip-flops is represented by $C_{FF} V_{\mathrm{DD}}^2$, which is approximately

---

[5]P-RC scheme is optimized based on min-energy/bandwidth w/maximum pipelining depth constraint, as discussed in Section II-B.

[6]This assumption also holds for the following analysis of P-RC wire. Regarding the performance analysis of ideal P-RC structure without maximum pipelining depth constraint, refer to Appendix A.

[7]CMOS gate capacitance per unit width (C_g) equals to $\epsilon_{ox} L/t_{ox}$, where L is the channel length, and t_{ox} indicates the oxide thickness. For long channel devices, L/t_{ox} is roughly constant due to the same scaling factor along different dimensions of transistor, whereas for short channel devices, oxide thickness cannot scale as fast as transistor width and length due to leakage and process considerations. Therefore, gate capacitance per unit width decreases as technology scales. In our study, C_g is 1.5 fF/μm at 90 nm node and 0.8 fF/μm at 22 nm node.

proportional to τV_{DD}^2. Similar to the delay modeling shown above, linear models are built by combining energy consumed on wire and gate together, shown in the third row of Table IV.

For T-line schemes, we consider the power dissipation on the wire and transceiver individually. The power consumed on T-lines is basically proportional to V_{DD}^2 if assuming wire input impedance remains constant across technologies. Transceiver dynamic power is linearly proportional to fCV_{DD}^2, where f is the clock frequency and C represents the total gate capacitance of the transceiver. Combining these two factors together

$$\text{energy}_n^{TL} \propto T_C V_{DD}^2/L \propto \tau V_{DD}^2/L \qquad (5)$$

with the assumption that cycle time T_C and gate capacitance C scale by the same rate as τ. Linear models based on the analysis are shown in Table IV. As shown in (4), compared with RC wires, if $\epsilon_r V_{DD}^2$ item is dominant in the total normalized energy expression, T-lines will consume less energy as technology scales since τ shrinks more rapidly than ϵ_r does.

For energy modeling, the largest errors in R-RC/P-RC modeling come from the same sources as discussed in previous subsection. Errors in the modeling of T-lines may relate to the neglect of wire input impedance variations across different technologies.

As a result, the maximum and average percent errors for RC wires are less than 34% and 12%, respectively, whereas for T-line schemes, those values become 27% and 13%.

3) Throughput: The **normalized throughput** (or throughput density) is defined as

$$\text{throughput}_n = \frac{\text{bit rate}}{\text{wire pitch}} \qquad (6)$$

which is adopted to compare the amount of data can be transmitted for a given cross area in a given time interval.

From (2) and assuming that wire pitch also scales down as FO4 delay τ, it is observed that, for an R-RC structure

$$\text{throughput}_n^{R-RC} \propto 1/(\sqrt{\epsilon_r}\tau L). \qquad (7)$$

Regarding a P-RC structure, normalized throughput can be derived from the normalized delay expression. The flip-flop delay should account for the τ^2 item in the denominator of normalized throughput expression. The general throughput model for RC wires is summarized in the third row of Table IV.

Unlike the RC-dominated structures, the bit rate of T-line schemes is usually limited by the bandwidth of transceivers (except for UT-TL, which is determined by the wire itself). As a result, assuming the transceiver bandwidth is inversely proportional to τ, the bit rate of T-TL/UE-TL/PE-TL structures is inversely proportional to τ. For UT-TL, the bit rate is constant as technology scales, but is approximately inversely proportional to wire length L. Therefore, a general model of bit rate of T-lines can be represented by

$$\text{Bit Rate} = C_1 1/\tau + C_2 1/L \qquad (8)$$

where C_1, C_2 are fitting coefficients. Considering wire pitch, for single-ended T-lines, wire pitch does not change with technology and wire length. However, for differential T-lines, larger wire pitch is required as technology scales and wire becomes

longer. An approximate relation is shown below based on simple modeling of wire resistance considering dc and ac components separately,[8]

$$\text{Wire Pitch} = C_1\sqrt{L} + C_2 L/\sqrt{\tau} \qquad (9)$$

where C_1, C_2 are fitting coefficients. In deriving this equation, we neglect the minor change of supply voltage V_{DD} as technology scales, and assume the most important frequency component of T-line skin effect for each technology is also proportional to $1/\tau$. Combining (8) and (9), we derive the throughput models for each T-line scheme in Table IV.

For most transceiver-limited T-line structures, even when considering the increasing wire pitch as technology scales, the throughput density will still exceed that of an R-RC structure due to the rapid improvement of transceiver bandwidth.

The proposed model may have a larger error for T-TL scheme at the most advanced technology node (22 nm) as wire length increases (≥ 7 mm) because the bandwidth of the overall structure becomes wire-limited and does not improve as technology scales.

For throughput models, maximum and average errors of RC wires are 18% and 9%, respectively. For T-lines, fitting errors become larger, which are 32% for maximum scenario and 11% for average scenario.

4) Area: Chip area consumed by different interconnect structures comprises two parts: wire area and circuit area. Wire area is the wire pitch multiplied by the total wire length L. The pitch scaling trend has been discussed in the previous section. Circuit area will reduce quadratically as technology scales, approximately proportionally to τ^2. Based on the analysis in previous subsections, the area model for each interconnection scheme is shown in the forth row of Table IV.

Typically, wire area dominates total chip area. As a result, RC wires consume less area compared to T-line structures, and RC wire area decreases more quickly as technology scales compared to T-line structures. The area of differential T-lines actually increases as technology scales due to increasing wire pitch.

Area model of RC wires may show a substantial error due to the reasons cited in the subsection of delay modeling. The maximum and average errors are 34% and 11% for RC wires. For T-line schemes, due to the dominance of wire area, fitting errors are smaller compared with RC wires. The maximum and average errors are 13% and 4%, respectively.

III. DESIGN METHODOLOGY

In this section, methodologies to design and optimize the six global interconnection structures are discussed. As previously mentioned, for the R-RC scheme, we adopt the optimization framework in [4], which is based on analytical formulae and numerical experiments to study the performance metrics under different design goals across multiple technology nodes. In terms of the P-RC structure, we develop a simple MATLAB flow to optimize the pipeline depth based on the lowest energy/bandwidth ratio, with the assumption that R-RC wire between flip-flops are

---

[8]In (9), the $\sqrt{L}$ value comes from the dc component of wire resistance, whereas $L/\sqrt{\tau}$ item comes from the ac component of wire resistance, caused by the skin effect.
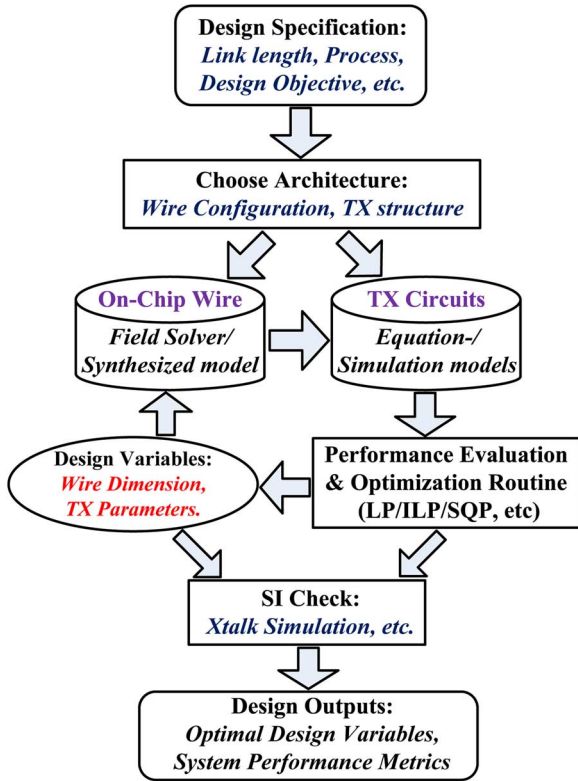
Fig. 7. Design framework for on-chip global T-line structures (UT-TL/T-TL/UE-TL/PE-TL).

optimized and the maximum pipeline depth is given. A detailed flow description is omitted here for the sake of brevity. For more information regarding performance analysis of ideal pipelined repeated RC wires (without maximum pipelining depth limit), the basis of our pipelined RC wire flow, please refer to the Appendix A.

This section focuses on the design of on-chip T-line schemes. Here, we propose a general framework by modeling on-chip T-line and transceiver circuitry separately and utilizing well-behaved optimization routines to generate the optimal design for given design specification, as illustrated in Fig. 7. We will introduce the application of this design framework on the single-ended T-line schemes (including UT-TL/T-TL) and differential T-line schemes (including UE-TL/PE-TL), respectively, in the following.

A. Single-Ended T-Line Schemes

The methodology to optimize single-ended T-line structures is proposed and discussed in [12]. Here, we summarize this methodology according to the general design framework shown in Fig. 7.

Corresponding to the proposed framework, on-chip wire is modeled using the frequency dependent tabular model generated by the field solver, and the characteristic of the transceiver circuit is obtained by SPICE simulation. Also, we use SPICE to evaluate the performance metrics of whole structure. Since wire dimensions are well defined, design variables of interest relate only to the transceiver circuit (the inverter chain), including the first inverter size S_1 and number of stages N. For the terminated

scheme (T-TL), termination resistance R_L also needs to be optimized for achieving high throughput. The optimization routine for this kind of scheme is comprised of two phases, namely determining the optimal clock rate and choosing the optimal variables in terms of the given objective. Finally, signal integrity is studied by SPICE simulation and the framework outputs the optimal design variables and corresponding performance metrics.

B. Differential T-Line Schemes

For differential T-line schemes, the adopted methodology is based on the constrained nonlinear programming formulation [8] and sequential quadratic programming (SQP) approach [23]. We discuss the details of this flow corresponding to the framework in Fig. 7 as follows.

The flow begins with modeling of wires and transceivers using different means. For on-chip wires, 2-D-$R(f)L(f)C$ tabular model is still utilized. For the transceiver circuit, though, we adopt a closed-form equation-based model, which is generated by fitting SPICE simulation data.[9] To evaluate the performance metrics of the whole structure, we combine the models of wire and transceiver together, and then utilize the approach in [24] to estimate the wire-end eye-opening. The optimization routine for differential schemes initially tries to find the smallest wire dimension that satisfies the eye-opening constraint by using binary search (which generates the data in Table III), and then calls the SQP subroutine to optimize the design variables for the given design objective. The design variables include driver impedance R_s, passive equalizer parameters R_d, C_d (for PE-TL structure), and termination resistance R_l. The key element in formulating the differential schemes is the eye-opening constraint. In this work, we choose the method used in [8] to consider this constraint by adding an exponential item to the cost function. After optimization, we check the signal integrity and finally output the system performance metrics.

IV. PREDICTION AND COMPARISON OF PERFORMANCE METRICS

Applying the design methodologies discussed in Section III, we perform the experiments in this section to study the performance metrics of our six different global interconnect structures under the **min-d** design objective across technology nodes, from 90 nm down to 22 nm.

A. Experimental Settings

For parameter extraction of on-chip lossy T-lines, we use the 2-D field solver CZ2D of the EIP tool suite from IBM [25] to build T-line structures shown in Fig. 6, and extract the frequency-dependent tabular model for SPICE simulation. For circuit design and modeling, we adopt a predictive transistor model [14], which is a Synopsys level3 MOSFET model with the parameters tuned following the ITRS roadmap.

---

[9]The details of these models can be referred to [8]. In this work, we adopt the same closed-form equations but recalculate the coefficients based on the newer receiver circuit design generated. As shown in [8], these equation-based models can achieve less than 2% and 5% relative error for the delay and power fitting, respectively.
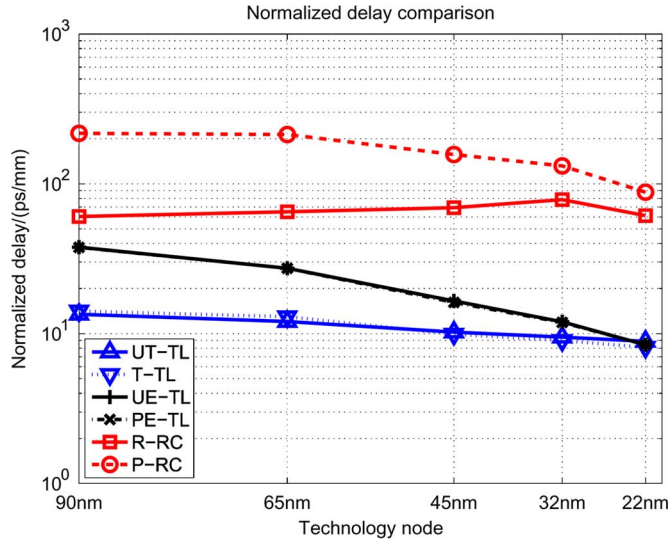
Fig. 8. Normalized delay of different global interconnection structures under min-d objective.

Fig. 9. Normalized energy per bit of different global interconnection structures under min-d objective.

For system simulation and optimization, HSPICE is used to simulate the transient response of wires, evaluate the performance of circuit and the entire interconnection structure. Linear and nonlinear regression methods and SQP routine implemented in MATLAB are adopted to build circuit models and perform optimization.

In our study, we set the maximum pipeline depth to 20, and choose 5 mm as the wire length (which represents typical critical length for on-chip global interconnect) to evaluate and compare the performance metrics of different structures. Power dissipation is estimated using a PRBS pattern with the activity factor around 0.23. We also extend each experiment to different wire lengths (0.5, 1, 3, 7, 9 mm) to study the wire length crossing points for some representative structure pairs in terms of different performance metrics, as shown in Section IV-D.

B. Latency

A comparison of the normalized delays of various global interconnect structures under the min-d objective is shown in Fig. 8. The trends of normalized latency with technology scaling verify our previous analysis in Section II-D. Normalized delay of R-RC structure increases due to the dominant effect of $\sqrt{\epsilon_r/\tau}$ on the total latency, whereas latency of P-RC decreases as the flip-flops dominate the total delay. Therefore, the latency penalty of pipelining RC wires is alleviated as technology scales.

On the other hand, due to the opposite scaling trend, all T-line structures outperform R-RC in terms of latency beyond the 90 nm node. The single-ended T-lines achieve lowest delay across all the five technology nodes. At 22 nm node, all the T-line structures show a similar delay around 8 ps/mm, whereas this number is 60 ps/mm for R-RC and around 90 ps/mm for P-RC. Therefore, a delay reduction of at least 87% could be obtained by replacing global RC wires with T-line structures in this scenario.
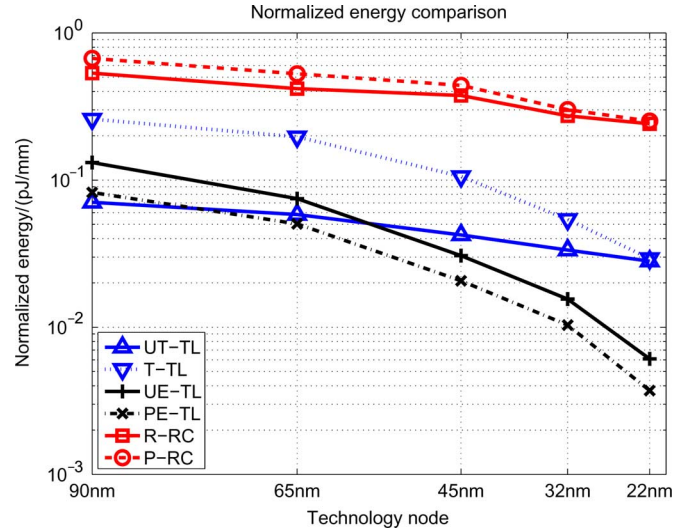
C. Other Metrics

Under the min-d objective, every interconnection scheme shows a decreasing trend in energy dissipation as technology scales (see Fig. 9), verifying our previous analysis in Section II-D. RC wires consume the largest energy among all six interconnection structures. Pipelining, under the optimization criterion of min-d, increases the energy of R-RC further due to the additional energy consumed by flip-flops, but this overhead decreases as technology scales because of the scaling of flip-flop capacitance. On the other hand, T-line structures consume less energy at each technology node. Beyond the 65 nm node, differential T-lines (UE-TL/PE-TL) consume the least energy due to power efficient SA-based receivers and the higher bit rate achieved by reducing signal swing at the wire-end. Further, the energy per bit could be reduced by nearly 40% using a passive equalizer. At the 22 nm node, differential T-line schemes will reduce the energy per bit by two orders of magnitude compared with RC wires.

The throughput density of different schemes under the min-d objective is shown in Fig. 10. As discussed in Section II-D, this metric is improved for all the schemes as technology scales (except for UT-TL, which throughput density is constant, limited by the wire itself). P-RC achieves the highest throughput density across all the technologies by increasing the R-RC bandwidth using the smallest wire pitch. For T-line schemes, differential T-lines have larger throughput density compared with single-ended ones because of the higher achievable bit rate by utilizing SA-based receiver. Furthermore, the introduction of passive equalization makes the utilization of narrower wires possible, increasing the density even further. Beyond 45 nm node, differential T-lines will finally outperform R-RC in terms of throughput density.

The chip areas consumed by the various interconnect structures are compared in Fig. 11. According to the analysis performed in Section II-D, assuming wire area is dominant in total area consumption, RC wire area will decrease exponentially as technology scales, whereas area of other T-line schemes will
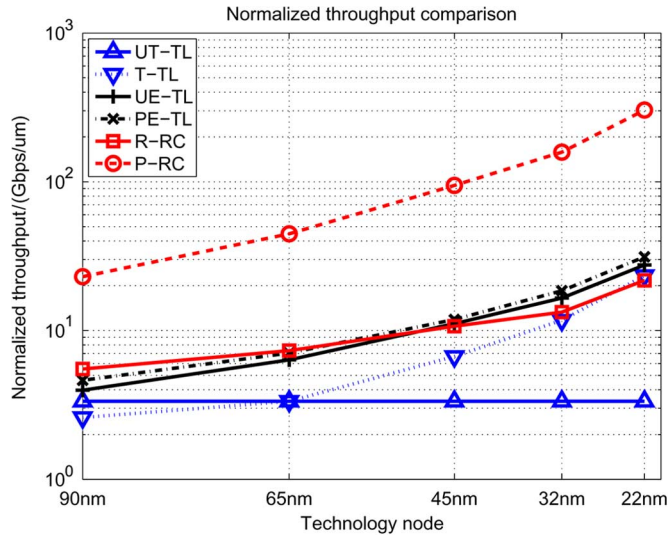
Fig. 10. Throughput density of different global interconnection structures under min-d objective.
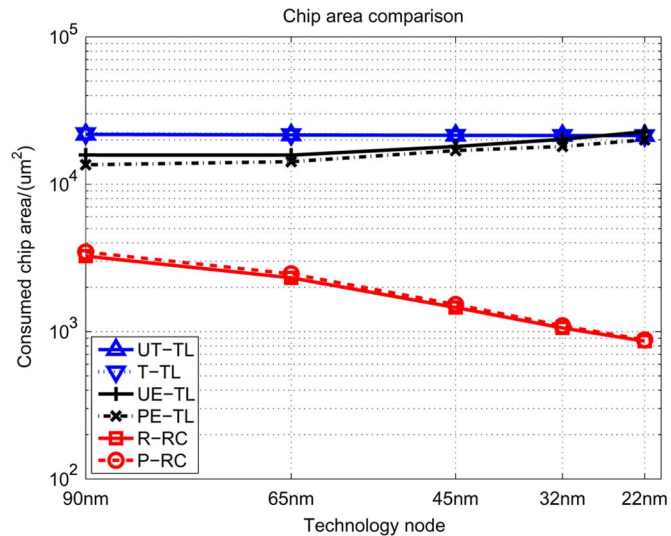

Fig. 11. Chip areas consumed by different global interconnection structures under min-d objective.

remain the same (single-ended T-lines) or even increase (differential T-lines), as shown in Fig. 11.

D. Critical Length

A critical length study is also performed by running the optimization flow in several different wire lengths, from 0.5 to 9 mm. The results are summarized in Fig. 12. In this figure, a dashed line and dotted line located on the upper and lower sides indicate the upper-bound and lower-bound of wire length for on-chip global interconnects, corresponding to 10 and 0.5 mm, respectively. We chose eight representative interconnect structure pairs, and show the scaling trend of their critical lengths in terms of four different performance metrics. As an illustration, for "Delay:PE-TL versus R-RC" case, which corresponds to the solid line with upper triangle marker, the critical length at 90 nm
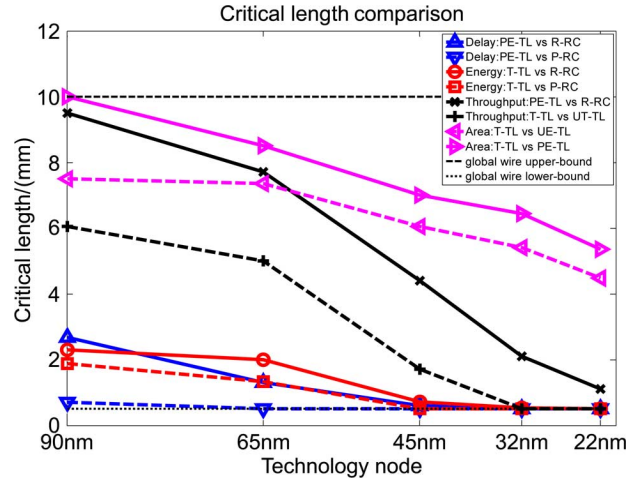

Fig. 12. Critical length of several chosen interconnect structure pairs in terms of different performance metrics under min-d objective.

node is about 2.5 mm, which means that when the wire length is larger than 2.5 mm at this node, PE-TL will outperform R-RC in terms of normalized delay. Based on this illustration to understand Fig. 12, we make the following general observations.

1) As technology scales (beyond the 45 nm node), T-line schemes will outperform RC wires in terms of normalized delay and energy within the entire length range of on-chip global wires.

2) In terms of throughput, at the 22 nm node, PE-TL will outperform R-RC while wire length is larger than 1 mm, and UT-TL will be replaced by T-TL within the entire length range.

3) Single-ended T-lines will consume less chip area compared with differential counterparts for longer wire lengths. At the 22 nm node, T-TL occupies less area than PE-TL and UE-TL when the wire length is longer than 5.4 and 4.5 mm, respectively.

V. SIGNAL INTEGRITY

In this section, we discuss the signal integrity issues of different interconnection structures, with the focus on the T-line schemes. Basically, we will study signal integrity by simulating the maximum crosstalk noise at the wire-end of quiet lines and the eye-height with and without crosstalk effects. For the maximum noise simulation, based on the previous work [26] and SPICE simulations, the worst case switching patterns of single-ended and differential T-lines are given in Fig. 13. In terms of eye-height simulation, HSPICE transient simulations for 500 cycle times are performed using one or several different PRBS input patterns. All the experimental results are summarized as follows.

A. Single-Ended T-Lines

Single-Ended structures tend to be more sensitive to noise. For the unterminated scheme (UT-TL), simulation shows that the maximum peak noise will be 380 mV at 45 nm node (1
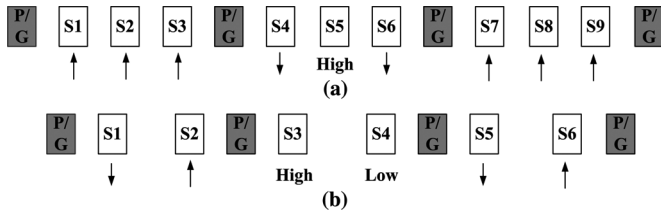
Fig. 13. Wire configurations and worst case switching patterns of T-line structures for testing crosstalk effects: (a) single-ended; (b) differential.

TABLE V
CROSSTALK EFFECTS ON THE T-TL STRUCTURE

Tech Node/nm	90	65	45	32	22
Cycle Time/(ps)	90	70	40	25	15
Max Xtalk Noise/(mV)	212	188	170	137	78
Eye Height w/o Xtalk/(mV)	1002	851	769	574	405
Eye Height w/ Xtalk/(mV)	750	740	706	539	383
Supply Voltage/V	1.2	1.1	1.0	0.9	0.8

TABLE VI
MAXIMUM CROSSTALK PEAK NOISE (mV) OF DIFFERENTIAL T-LINE STRUCTURES

Tech Node/nm	90	65	45	32	22
UE-TL	7	8	10	11	13
PE-TL	2	6	8	10	13

V supply voltage),[10] and this situation could be more severe as the technology scales since the supply voltage drops. Therefore, considering the crosstalk, full-swing signals cannot be guaranteed at the wire-end, which makes this conventional on-chip bus structure less reliable at advanced technology nodes in spite of its high-performance.

In comparison, T-TL provides improved noise performance as well as higher bandwidth. Since the cycle time of this structure changes as technology scales, we perform the simulation at different nodes and summarize results in Table V. The peak crosstalk noise reduces with technology scaling due to the reduced termination resistance and supply voltage (can be derived based on the formula presented in [28]). At the 45 nm node, the noise is only 170 mV, less than half that of UT-TL. Eye-heights also reduce because of the increasing bit rate. However, an eye around 380 mV could still be achieved at the 22 nm node even with the impact of crosstalk noise.

B. Differential T-Lines

Differential T-lines enjoy greater immunity to crosstalk due to the termination resistance and the impact of common-mode noise rejection [29]. Similar crosstalk peak noise simulations are performed using the switching pattern described in Fig. 13(b) for UE-TL and PE-TL structures, and the results are listed in Table VI. The table shows that peak noise is far lower in differential T-lines than that of single-ended T-lines. Even with the higher inductive coupling as the bit rate increases, the peak noise in the differential T-line is only around the 10 mV range.

---

[10]Here we follow the crosstalk simulation method in [27] and focus on the far-end noise (FEN). [27] also provides a more comprehensive study on the frequency-dependent crosstalk effects of on-chip single-ended unterminated data bus.
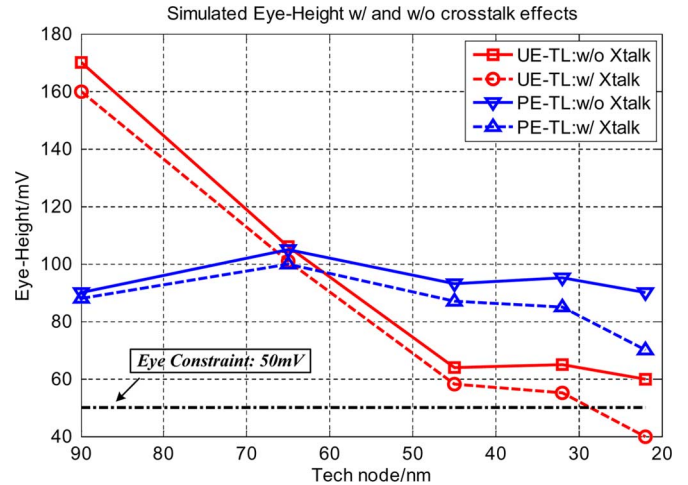

Fig. 14. Influence of crosstalk effects on the eye-height of UE-TL and PE-TL structures.

Eye-heights with and without crosstalk effects for two differential T-line structures are simulated and illustrated in Fig. 14. For UE-TL structure, the optimal eye-height reduces as technology scales due to the increased bit rate. Considering the crosstalk, it will be harder for this scheme to meet the 50 mV eye constraint at advanced technology nodes (see the violation at 22 nm node). In comparison, by using passive equalization, PE-TL can achieve larger than a 70 mV eye across all technologies even in the presence of crosstalk. Therefore, equalization improves signal integrity by boosting the eye-heights at higher bit rates.

VI. DISCUSSION AND CONCLUSION

A. Discussions

In this work, **latency** was chosen to be the design objective for different interconnect schemes specifically designed for global wires (e.g., wide bus) in conventional high-end processors. To meet the increasing demand for computing capacity as process technology scales, throughput-centric interconnect design has become a hot research topic. New computing architectures have appeared, such as multicores and networks-on-chips (NoCs) [30]. New design metrics have also been proposed to balance throughput, energy, and chip area during the interconnect planning stage for different applications, as shown in [31]. Conventional repeater insertion with min-d optimization cannot satisfy the increasing bandwidth requirement for global interconnect. To enhance signal bandwidth, pipelining and other similar concepts (e.g., wave-pipelining [32]) are utilized to compensate for this performance gap. Since the purpose of our paper is to explore the potential of different interconnect options in **high-performance** applications, we did not include much optimization freedom during the preliminary study of the P-RC scheme.[11] By adopting voltage scaling, buffer and wire sizing, the performance of P-RC scheme can be more fully studied, and the energy gap between RC wire and T-lines could perhaps be

---

[11]Actually, the optimization here for P-RC is only one of possible choices, with only one design variable (pipelining depth N) tuned. The complete solution space could be a continuum ranging from extremely high throughput (high energy cost) to the limit of no pipelining (equivalent to R-RC).

reduced. This would be an interesting future research topic, exploring a different direction.

Some research has been done recently regarding chip-level CMOS implementation of novel global interconnect (e.g., uninterrupted RC wire, equalized on-chip interconnect, etc.) [10], [22]. For the state-of-the-art equalized on-chip interconnect design using a 90 nm process [10], measured throughput density is similar to our prediction for differential T-line scheme (about 2 Gb/s/μm), and energy per bit is about 1/2 of the passive equalized T-line scheme (about 700 fJ/b). Another possible option for global interconnect is low-swing signaling on RC wire, according to a recent study [33]. Although the energy dissipated using such a scheme can be very low (similar or even lower than T-line schemes based on 0.13-μm simulation results), its latency is very large (2–4× of repeated RC wires). Therefore, we did not include reduced-swing RC signaling in this study.

B. Conclusion

In this paper, we compare six different global interconnect structures in terms of latency, energy per bit, throughput, chip area, and signal integrity, across technology nodes ranging from 90 nm down to 22 nm. A set of simple linear models is provided to link the architecture-friendly performance metrics of these interconnect structures with technology-defined parameters, and is verified by experimental results. A general design framework is introduced to optimize and evaluate the performance metrics of on-chip T-line interconnects.

Several observations based on the performance trends observed with technology scaling are summarized as follows: 1) T-line structures have the potential to replace RC wires at future technology nodes due to improved delay, energy per bit, throughput density (compared with R-RC), and reliability (crosstalk noise), but such schemes consume greater chip area; 2) differential T-lines are better for high-throughput, low-power, and low-noise application compared with single-ended counterparts; and 3) equalization approaches (such as passive equalization) can be utilized for on-chip global interconnects to improve throughput density and reduce energy dissipation.

APPENDIX A
PERFORMANCE ANALYSIS OF IDEAL PIPELINED REPEATED RC WIRES

We analyze the performance metrics of pipelined repeated RC wire without maximum pipelining depth limit in the following, and define some parameters shown in Table VII. Using the above defined parameters, and assuming the energy and delay is evenly distributed within each pipelining stage the same as in previous long R-RC wires without flip-flop insertion, formulae for performance estimation can be derived as follows:

$$\text{Latency} = N T_{FF} + t_{RC} L \qquad (10)$$

$$\text{Power} = \frac{N C_{FF} V_{\text{DD}}^2 + e_{RC} L}{T_{FF} + t_{RC}(L/N)} \qquad (11)$$

$$\text{Bandwidth} = \frac{1}{T_{FF} + t_{RC}(L/N)} \qquad (12)$$

$$\text{Energy} = N C_{FF} V_{\text{DD}}^2 + e_{RC} L. \qquad (13)$$

TABLE VII
DESIGN PARAMETERS USED IN PERFORMANCE ANALYSIS OF P-RC STRUCTURE

symbol	description
N	Pipelining depth (# of flip-flops inserted)
L	Total wire length
T_{FF}	Total latency of flip-flop (sum of T_{c-q} and T_{setup})
C_{FF}	Effective capacitance of flip-flop (obtained by power simulation)
t_{RC}	Normalized delay of optimized R-RC wire
e_{RC}	Normalized energy of optimized R-RC wire

To derive the optimal pipelining depth N, we take the derivative of **Energy/Bandwidth**, and let it equal zero. The optimal N is shown to be

$$N_{\text{opt}} = \sqrt{\frac{e_{RC} t_{RC}}{C_{FF} V_{\text{DD}}^2 T_{FF}} L} \qquad (14)$$

which is proportional to the wire length and shows an increasing trend with technology scaling. If there is no limit on the upper-bound of pipelining depth (ideal P-RC case), the performance metrics of P-RC in terms of **min-Energy/Bandwidth** can be obtained by plugging in (14) back to (10)–(13)

$$\text{Latency} = \left(\sqrt{\frac{e_{RC} t_{RC} T_{FF}}{C_{FF} V_{\text{DD}}^2}} + t_{RC} \right) L \qquad (15)$$

$$\text{Power} = \left(\frac{e_{RC}}{T_{FF}} \right) L \qquad (16)$$

$$\text{Bandwidth} = \frac{1}{T_{FF} + t_{RC}\sqrt{\frac{C_{FF} V_{\text{DD}}^2 T_{FF}}{e_{RC} t_{RC}}}} \qquad (17)$$

$$\text{Energy} = \left(\sqrt{\frac{e_{RC} t_{RC} C_{FF} V_{\text{DD}}^2}{T_{FF}}} + e_{RC} \right) L. \qquad (18)$$

It can be seen that, for ideal P-RC structure, most metrics (latency, power, energy) are linearly proportional to L, except for the bandwidth, which is independent of the wire length L. Also, the bandwidth increases nearly exponentially as the technology scales, similar to the trend of transistor performance scaling.

REFERENCES

[1] International Technology Roadmap for Semiconductors, Semiconductor Industry Association 2007. [Online]. Available: http://www.itrs.net

[2] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," in *Proc. Int. Workshop Syst. Level Interconnect Prediction*, Paris, France, Feb. 2004, pp. 7–13.

[3] D. Sylvester and K. Keutzer, "A global wiring paradigm for deep submicron design," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 19, no. 2, pp. 242–252, Feb. 2000.

[4] L. Zhang, H. Chen, B. Yao, K. Hamilton, and C.-K. Cheng, "Repeated on-chip interconnect analysis and evaluation of delay, power, and bandwidth metrics under different design goals," in *Proc. IEEE Int. Symp. Quality Electron. Des.*, San Jose, CA, Mar. 2007, pp. 251–256.

[5] H. B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*. Boston, MA: Addison-Wesley, 1990.

[6] A. Deutsch, P. W. Coteus, G. V. Kopcsay, H. H. Smith, C. W. Surovic, B. L. Krauter, D. C. Edelstein, and P. L. Restle, "On-chip wiring design challenges for Gigahertz operation," *Proc. IEEE*, vol. 89, no. 4, pp. 529–555, Apr. 2001.

[7] B. Kim and V. Stojanovic, "Equalized interconnects for on-chip networks: Modeling and optimization framework," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, San Jose, CA, Nov. 2007, pp. 552–559.

[8] L. Zhang, Y. Zhang, A. Tsuchiya, M. Hashimoto, E. S. Kuh, and C.-K. Cheng, "High performance on-chip differential signaling using passive compensation for global communication," in *Proc. Asia South Pac. Des. Autom. Conf.*, Yokohama, Japan, Jan. 2009, pp. 385–390.

[9] Y. Zhang, L. Zhang, A. Tsuchiya, M. Hashimoto, and C.-K. Cheng, "On-chip high performance signaling using passive compensation," in *Proc. IEEE Int. Conf. Comput. Des.*, Lake Tahoe, CA, Oct. 2008, pp. 182–187.

[10] B. Kim and V. Stojanovic, "A 4 Gb/s/ch 356 fJ/b 10 mm equalized on-chip interconnect with nonlinear charge-injecting transmit filter and transimpedance receiver in 90 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2009, pp. 66–68.

[11] Y. Zhang, X. Hu, A. Deutsch, A. E. Engin, J. F. Buckwalter, and C.-K. Cheng, "Prediction of high-performance on-chip global interconnection," in *Proc. Int. Workshop Syst. Level Interconnect Prediction*, San Francisco, CA, Jul. 2009, pp. 61–68.

[12] Y. Zhang, L. Zhang, A. Deutsch, G. A. Katopis, D. M. Dreps, J. F. Buckwalter, E. S. Kuh, and C.-K. Cheng, "Design methodology of high performance on-chip global interconnect using terminated transmission-line," in *Proc. IEEE Int. Symp. Quality Electron. Des.*, San Jose, CA, Mar. 2009, pp. 451–458.

[13] D. Schinkel, E. Mensink, E. Klumperink, E. Tuiji, and B. Nauta, "A double-tail latch-type voltage sense amplifier with 18 ps setup+hold time," in *Proc. IEEE Int. Solid-State Circuits Conf.*, San Francisco, CA, Feb. 2007, pp. 314–316.

[14] S. Uemura, A. Tsuchiya, and H. Onodera, "A predictive transistor model based on ITRS roadmap," in *Proc. General Conf. IEICE*, Mar. 2006, p. 81.

[15] S. Sim, S. Krishnan, D. Petranovic, and N. Arora, "A unified RLC model for high-speed on-chip interconnects," *IEEE Trans. Electron Devices*, vol. 50, no. 6, pp. 1501–1510, Jun. 2003.

[16] H. Smith, A. Deutsch, S. Mehrotra, D. Widiger, M. Bowen, A. Dansky, G. V. Kopcsay, and B. Krauter, "R(f)L(f)C coupled noise evaluation of an S/390 microprocessor chip," in *Proc. IEEE Custom Integr. Circuits Conf.*, San Diego, CA, May 2001, pp. 237–240.

[17] I. M. Elfadel, A. Deutsch, H. H. Smith, B. J. Rubin, and G. V. Kopcsay, "A multiconductor transmission line methodology for global on-chip interconnect modeling and analysis," *IEEE Trans. Adv. Packag.*, vol. 27, no. 1, pp. 71–78, Feb. 2004.

[18] G. A. Sai-Halasz, "Performance trends in high-end processors," *Proc. IEEE*, vol. 83, no. 1, pp. 20–36, Jan. 1995.

[19] G. V. Kopcsay, B. Krauter, D. Widiger, A. Deutsch, B. J. Rubin, and H. H. Smith, "A comprehensive 2-D inductance modeling approach for VLSI interconnects: Frequency-dependent extraction and compact circuit model synthesis," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 10, no. 6, pp. 695–711, Dec. 2002.

[20] J. Wood, T. C. Edwards, and S. Lipa, "Rotary traveling-wave oscillator arrays: A new clock technology," *IEEE J. Solid-State Circuits*, vol. 36, no. 11, pp. 1654–1665, Nov. 2001.

[21] L. Zhang, J. Wilson, R. Bashirullah, L. Luo, J. Xu, and P. Franzon, "Driver pre-emphasis techniques for on-chip global buses," in *Proc. IEEE Int. Symp. Low Power Electron. Des.*, Aug. 2005, pp. 186–191.

[22] D. Schinkel, E. Mensink, E. A. Klumperink, E. van Tuijl, and B. Nauta, "A 3-Gb/s/ch transceiver for 10-mm uninterrupted RC-limited global on-chip interconnects," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 297–306, Jan. 2006.

[23] M. C. Biggs, "Constrained minimization using recursive quadratic programming: Some alternative subproblem formulations," in *Towards Global Optimization*, L. C. W. Dixon and G. P. Szego, Eds. Amsterdam, The Netherlands: North-Holland, 1975, pp. 341–349.

[24] R. Shi, W. Yu, Y. Zhu, E. S. Kuh, and C.-K. Cheng, "Efficient and accurate eye diagram prediction for high speed signaling," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, San Jose, CA, Nov. 2008, pp. 655–661.

[25] IBM, "IBM electromagnetic field solver suite of tools," 2006. [Online]. Available: http://www.alphaworks.ibm.com/tech/eip

[26] I. M. Elfadel, A. Deutsch, G. V. Kopcsay, B. J. Rubin, and H. H. Smith, "A CAD methodology and tool for the characterization of wide on-chip buses," *IEEE Trans. Adv. Packag.*, vol. 28, no. 1, pp. 63–70, Feb. 2005.

[27] A. Deutsch, H. Smith, C. Surovic, G. Kopcsay, D. Webber, P. Coteus, G. Katopis, W. Becker, A. Dansky, G. Sai-Halasz, and P. Restle, "Frequency-dependent crosstalk simulation for on-chip interconnections," *IEEE Trans. Adv. Packag.*, vol. 22, no. 3, pp. 292–308, Aug. 1999.

[28] R. Venkatesan, J. A. Davis, and J. D. Meindl, "Compact distributed RLC interconnect models-part IV: Unified models for time delay, crosstalk, and repeater insertion," *IEEE Trans. Electron Devices*, vol. 50, no. 4, pp. 1094–1102, Apr. 2003.

[29] Y. Massoud, J. Kawa, D. MacMillen, and J. White, "Modeling and analysis of differential signaling for minimizing inductive cross-talk," in *Proc. IEEE/ACM Design Autom. Conf.*, Las Vegas, NV, Jun. 2001, pp. 804–809.

[30] , A. Jantsch and H. Tenhunen, Eds., *Networks on Chip*. Norwell, MA: Kluwer, 2003.

[31] V. V. Deodhar and J. A. Davis, "Optimal voltage scaling, repeater insertion, and wire sizing for wave-pipelined global interconnects," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 4, pp. 1023–1030, May 2008.

[32] J. Xu and W. Wolf, "A wave-pipelined on-chip interconnect structure for networks-on-chips," in *Proc. IEEE Symp. High Perform. Interconnects*, Aug. 2003, pp. 10–14.

[33] J. C. Montesdeoca, J. A. Montiel-Nelson, and S. Nooshabadi, "CMOS driver-receiver pair for low-swing signaling for low energy on-chip interconnects," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 2, pp. 311–316, Feb. 2009.

Yulei Zhang (S'08) received the B.E. degree in electrical engineering from Tsinghua University, Beijing, China, in 2007, and the M.S. degree in electrical and computer engineering from University of California-San Diego (UCSD), La Jolla, in 2009, where he is currently pursuing the Ph.D. degree from the Department of Electrical and Computer Engineering.

Since Fall of 2009, he was an intern with Bluetooth IC Design Group, Broadcom Corporation, San Diego, CA. His research interests include design and optimization of high-speed, low-power on-chip/off-chip interconnects and low-power clock distribution network design.

Xiang Hu (S'08) received the B.E. and M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2005 and 2007, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering from the University of California, San Diego.

His research interests include the areas of analysis and optimization of power distribution networks, and circuit simulation. He is also working as an intern at Qualcomm Inc. on the simulation and analysis of power distribution networks for through-silicon-stacked ICs.

Alina Deutsch (F'99) received the B.S. and M.S. degrees in electrical engineering from Columbia University, NY, and Syracuse University, Syracuse, NY, in 1971 and 1976, respectively.

She worked with IBM since 1971 and retired after 38 years, in 2009. She worked in several areas, including testing of semiconductor and magnetic bubble memory devices. She designed unique lossy transmission line configurations, developed unique high-frequency high impedance coaxial probes, and a novel short-pulse measurement technique for characterization of resistive transmission lines that is now an IPC industry standard. She was a Research Staff Member that worked on the design, analysis, and measurement of packaging and VLSI chip interconnections for future digital processor and communication applications. Her work involved the three dimensional modelling, signal integrity and noise simulation, and testing of a large range of package lossy transmission lines from printed-circuit boards, cables, connectors, to thin-film wiring on multi-chip modules and on-chip wiring. She was also the manager of the Interconnect and Packaging Analysis

Project that developed advanced electromagnetic field-solver codes. She was the author of 45 papers published in refereed technical journals, has given numerous invited and tutorial talks, and holds 16 patents.

Ms. Deutsch was a recipient of Outstanding Technical Achievement, Research Division, and S/390 Division Team Awards from IBM in 1990, 1993, 1996, 1999, 2000, 2001, 2002, 2003, 2005, 2006, and 2009. She co-chaired for four years the IEEE Topical Meeting on Electrical Performance of Electronic Packaging, she was technical program co-chair for the IMAPS Next Generation IC and Package Design Workshop for three years, co-chaired the CPMT Society Future Directions in IC and Package Design workshop for six years, served as Guest Editor of the IEEE TRANSACTIONS ON ADVANCED PACKAGING for five years, served as Associate Editor of IEEE TRANSACTIONS ON COMPONENTS AND PACKAGING TECHNOLOGIES for seven years, is member of Tau Beta Pi, Etta Kappa Nu, served as elected member of the IEEE Components, Packaging, and Manufacturing Technology Society Board of Governors for 2000-2002, and was Vice-Chair of the CPMT Society TC-EDMS Technical Committee for seven years.

James F. Buckwalter (S'01–M'06) received the B.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology (Caltech), Pasadena, in 1999 and 2006, respectively, and the M.S. degree in electrical engineering from University of California at Santa Barbara, in 2001.

He was a Research Scientist with Telcordia Technologies from 1999 to 2000. He worked with the IBM T. J. Watson Research Center, Yorktown Heights, NY, during the summer of 2004. In 2006, he joined Luxtera, Carlsbad, CA, where he developed high-speed circuits for optical interconnects. In July 2006, he joined the Faculty of the University of California-San Diego, where he is an Assistant Professor of electrical engineering.

Dr. Buckwalter was a recipient of the Analog Devices Outstanding Student Designer Award in 2003, an IBM Ph.D. fellowship in 2004, and a DARPA Young Faculty Award in 2007.

A. Ege Engin (M'05) received the B.S. and M.S. degrees in electrical engineering from Middle East Technical University, Ankara, Turkey, and from University of Paderborn, Germany, in 1998 and 2001, respectively, and the Ph.D degree with *summa cum laude* from the University of Hannover, Germany, in 2004.

He worked as a Research Engineer with the Fraunhofer-Institute for Reliability and Microintegration, Berlin, Germany. From 2006 to 2008, he was an Assistant Research Director of the Microsystems Packaging Research Center, Georgia Institute of Technology. He is currently an Assistant Professor with the Electrical and Computer Engineering Department, San Diego State University. He has over 70 publications in journals and conferences in the areas of signal and power integrity modeling and simulation, one patent, and three patent applications. He is the coauthor of the book *Power Integrity Modeling and Design for Semiconductors and Systems* (Prentice-Hall, 2007).

Dr. Engin was a recipient of the Semiconductor Research Corporation Inventor Recognition Award in 2009. He has coauthored publications that received the Outstanding Poster Paper Award in the Electronic Components and Technology Conference (ECTC) 2006, Best Paper Award Finalist in the Board-Level Design Category at DesignCon 2007, and Best Paper of the Session Award in IMAPS Advanced Technology Workshop on RF and Microwave Packaging 2009.

Chung-Kuan Cheng (S'82–M'84–SM'95–F'00) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, and the Ph.D. degree in electrical engineering and computer sciences from University of California, Berkeley, in 1984.

From 1984 to 1986, he was a Senior CAD Engineer with Advanced Micro Devices Inc. In 1986, he joined the University of California, San Diego, where he is a Professor in the Computer Science and Engineering Department, an Adjunct Professor in the Electrical and Computer Engineering Department. He served as a Chief Scientist at Mentor Graphics in 1999. He was appointed as an Honorary Guest Professor of Tsinghua University 2002-2008. His research interests include medical modeling and analysis, network optimization and design automation on microelectronic circuits.

Dr. Cheng was an Associate Editor of IEEE TRANSACTIONS ON COMPUTER AIDED DESIGN for 1994-2003. He was a recipient of the Best Paper Awards, IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN in 1997, and in 2002, the NCR Excellence in Teaching Award, School of Engineering, UCSD, 1991, and IBM Faculty Awards in 2004, 2006, and 2007.