IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Parallel vs. Serial On-Chip Communication

Rostislav (Reuven) Dobkin, Arkadiy Morgenshtein, Avinoam Kolodny, Ran Ginosar

CCIT Report #674
December 2007

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #674 December 2007

Parallel vs. Serial On-Chip Communication

Rostislav (Reuven) Dobkin, Arkadiy Morgenshtein, Avinoam Kolodny, Ran Ginosar
VLSI Systems Research Center, Electrical Engineering Department
Technion – Israel Institute of Technology
Haifa, Israel
[rostikd@tx.technion.ac.il]

***Abstract*—Synchronous parallel links are widely used in modern VLSI designs for on-chip inter-module communication. Long range parallel links occupy large area and incur high capacitive load, high leakage power and cross-coupling noise. The problems exacerbate for applications having low utilization of the links or suffer from congestion of the interconnect. While standard synchronous serial links are unattractive due to limited bit-rate, novel high performance serial links may change the balance. In this paper we show that novel serial links provide better performance than parallel links for long range communications, beyond several millimeters. We analyze the technology dependence of link performance. An example for 65 nm technology is presented, and compare wave-pipelined and register-pipelined parallel links to a high performance serial link in terms of bit-rate, power, area and latency.**

Index Terms—Serial Link, Parallel Link, Asynchronous Circuits, Dual-Rail, Long-Range Interconnect

I. Introduction

Transistor size scaling drastically improves on-chip clock rates, practically doubling the performance every five years [1]. While local interconnect follows transistor scaling, global lines do not, challenging long range on-chip data communications in terms of latency, throughput and power [1]. In addition, as Systems-on-Chip (SoC) integrate an ever growing number of modules, on-chip inter-modular communications become congested and the modules must turn to serial interfaces, similar to the trend from parallel to serial inter-chip interconnects.

Long-range bit-parallel data links provide high data rates at the cost of large chip area, routing difficulty, noise and power. In addition, such links are often utilized only a small portion of the time, but dissipate leakage power at all times. Leakage is incurred at the line drivers and also at the repeaters, which are often necessary for long interconnects [2][3]. Parallel link performance is bounded by available clock rate and by clock skew, delay uncertainty due to process variations, cross-talk noise, and layout geometries.

Bit-serial communications offer an alternative to bit-parallel interconnects, mitigating the issues of area, routability, and leakage power, since there are fewer wires, fewer line drivers, and fewer repeaters. However, to provide the same throughput as an N-bit parallel interconnect, the serial link must operate N times faster. Simple synchronous serial links that employ the system clock are incapable of providing the required throughput. Recently proposed novel wide-bandwidth serial link circuits [4]—[14], which operate faster than the system clock, may deliver the required bandwidth.

Synchronous serial links are typically employed for off-chip communications, where pin-out limitations call for a minimal number of wires per link. Source-synchronous protocols are often used for these applications [15]—[20]. A common timing mechanism for serial interconnects injects a clock into the data stream at the transmitting side and recovers the clock at the receiver. Such clock-data recovery (CDR) circuits often require a power-hungry PLL, which may also take a long while to converge on the proper clock frequency and phase at the beginning of each transmission. If the receiver and transmitter operate in different clock domains, the transaction must also be synchronized at both ends, incurring additional delay and power. Alternatively, an asynchronous data link employs handshake instead of clocks. Traditional asynchronous protocols are relatively slow due to the need to acknowledge transitions [14][21]. In [22] asynchronous protocols share data lines, but their performance depends on wire delays.

High-speed serial schemes, having data cycle of a few gate delays (down to single gate-delay cycle), have been recently proposed [4]—[14]. These fast schemes exploit wave-pipelining, low-swing differential signaling, fast clock generators and asynchronous protocols. In addition, these schemes require channel optimization to support wide-bandwidth data transmission over the link wires. A wave-front train serialization scheme was presented in [11]. The serializer is based on a chain of MUXes (similar to [23]). The link is single-ended and employs wave-pipelining. The link data cycle is approximately $7 \cdot d_4$ (3Gbps@180nm), where d_4 is an inverter FO4 delay. Wave-pipelined multiplexed (WPM) routing

technique was presented in [12][13]. WPM routing employs source synchronous communication and its performance is limited by the clock skew and delay variations. Employing low-voltage differential pairs for on-chip serial interconnect was discussed in [9][10], where data was sampled at the receiver without any attention to synchronization issues. A three level voltage swing was presented in [24], requiring non-standard amplifiers.

Circuits that had originally been designed for off-chip communications [15][20] were adopted for on-chip serial link in [8]. An output-multiplexed transmitter is connected to a multiplexed receiver, requiring clock calibration at the receiver side. Both transmitter and receiver use multi-phase DLL circuits. The link employs low-swing differential signaling and transfers eight-bit words. The output-multiplexed architecture delivers better performance than input-multiplexing (down to $2 \cdot d_4$ data cycle), but at the expense of much higher output capacitance (that grows linearly with the word-width). A fabricated chip demonstrated an operational 3mm link.

In this paper we consider a different novel architecture [4][5] achieving a data cycle of a single gate delay (d_4) and throughput that is independent of the word width. This novel link is studied in comparison with parallel links that provide the same throughput (a preliminary comparison was presented in [25]). Such comparative analysis is of great importance for predicting system-level interconnect performance and is the main subject of this paper. We analyze the various costs of serial versus parallel links. Keeping bandwidth the same, we compare area, power and latency. We show that long range serial links outperform parallel links. The rest of the paper is structured as follows. In Section II we define the parallel and serial links under study. Section III provides analytical models for bit-rate, area, power and latency, and comparison results are presented in Section IV.

II. High Bit-Rate Parallel and Serial Communication

In this section we define the parallel and serial links under study and explain why these specific architectures were selected.

A *Parallel Link* comprises at least N wires that can carry N bit simultaneously. Either no data or N bits traverse the link together and pass through any given cross section of the link at the same time. The data rate is $F_{PAR} \cdot N$, where F_{PAR} is the rate at which the words are presented at the input of the link.

A *Serial Link* is one or more wires that are able to carry single-bit words. The bit is presented to the link at the transmitter side and is sampled subsequently at the receiver side. Either no data or a single bit traverse the link at any given cross section of the link. The bandwidth of the serial link is F_{SER}, where F_{SER} is the rate at which one-bit words are presented at the input of the link.

The serial link is a special case of the parallel link for N=1. In this study, $8 \le N \le 128$.

Different implementations of parallel and serial links exist. Some are used more than others in actual chips. In this paper we study only a few representative architectures, as defined below.

A. Parallel Links

We consider two types of parallel links, *register-pipelined* and *wave-pipelined*.

The widely used "*register-pipelined*" parallel link is fully synchronous where the interconnect is considered as combinational logic between two registers. When the interconnect delay exceeds the clock cycle, the link is pipelined to yield the required bit-rate (Figure 1). The single clock is either generated globally or sent with the data from the transmitter (*source synchronous* communication). Interconnect delay is usually optimized by means of buffering (repeaters).

A primary drawback of the register-pipelined parallel link is the high cost of pipelining that is incurred when a high bit-rate is desired over a long range. Wave-pipelining [26]—[29] exploits buffers and wire delays instead of flip-flops. In a source-synchronous *wave-pipelined* parallel link (Figure 2), the bit rate is limited by the relative skew of the link wires rather than by the clock cycle. Multiple N-bit words can transverse the link simultaneously. The data is presented to the bus on each rising edge of CLKT and is sampled by a receiver register on each falling edge of CLKR. Wave-pipelining may improve the bit-rate relative to register-pipelined links [27] .

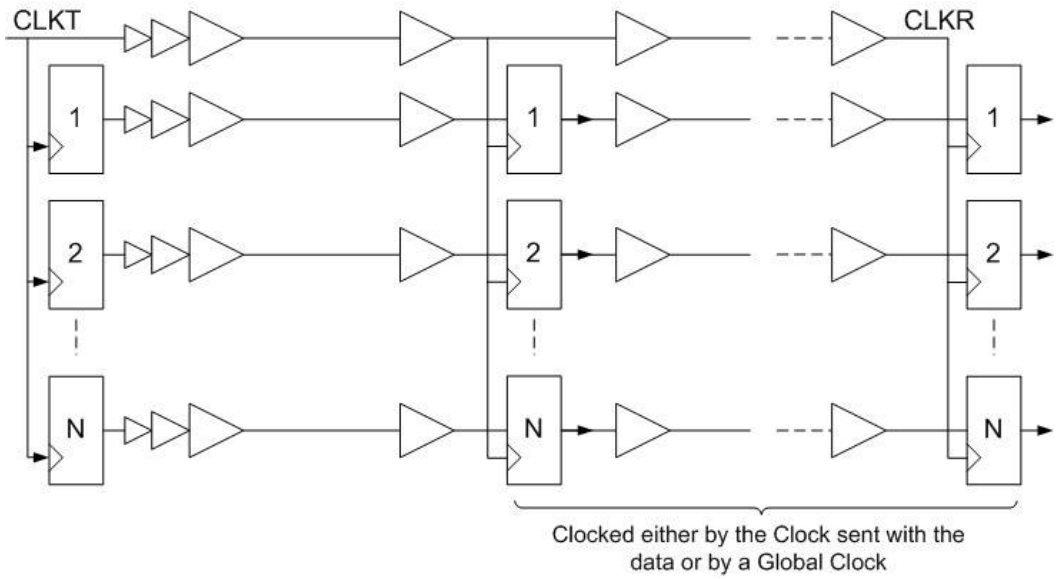

Figure 1: Register-pipelined parallel link (no wave-pipelining)

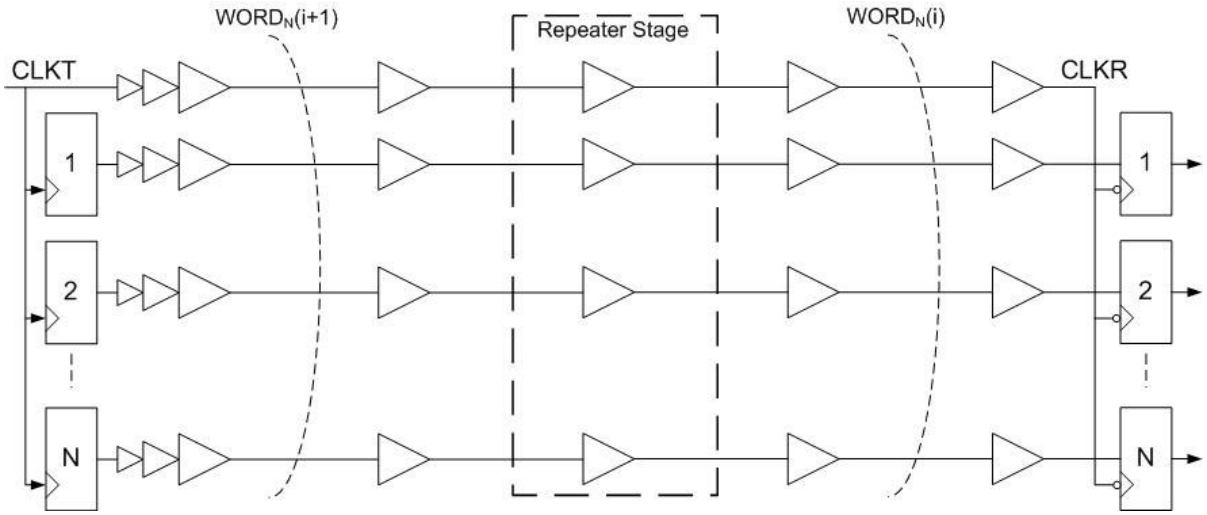

Figure 2: Wave-pipelined parallel link

Several enhancements and variations may be applied to the basic architectures of Figure 1 and Figure 2 to mitigate crosstalk and reduce power. Examples include shielding, interleaved bi-directional lines, asynchronous signaling, data encoding, staggered repeaters, special worst-case transition patterns handling [30] and power-saving techniques [2]. While most variations result in minor performance differences, shielding may significantly affects performance. We consider the two extremes of shielding (in terms of achieved bit rates and required area): unshielded and fully-shielded wires (Figure 3).

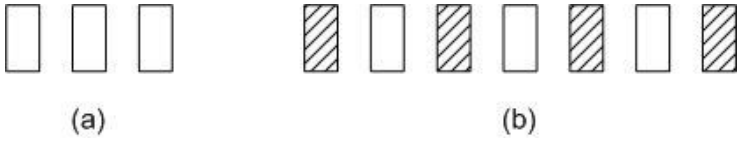

Figure 3: Differently shielded links: (a) unshielded, (b) fully-shielded

B. A High Performance Serial Link

Standard on-chip serial links are unattractive due to their inferior bit-rates relative to the parallel link. With the same clock as the parallel link, the bit-rate of a standard synchronous serial link is limited to *N* times lower than the parallel link.

A novel high bit-rate asynchronous serial link was presented in [4]—[7]. The link (Figure 4) employs low-latency synchronizers at the source and sink [31], two-phase NRZ Level Encoded Dual Rail (LEDR) data/strobe (DS) encoding and an asynchronous handshake protocol (allowing non-uniform delay intervals between successive bits) [32]—[34], serializer and de-serializer and line drivers and receivers. Acknowledgment is returned only once per word, rather than bit by bit, enabling multiple bits in a wave-pipelined manner over the serial channel. The wires (D and S) employ (fully shielded) wave-guides, enabling multiple traveling signals. On a well-designed wave-guide long wires may carry a number of successive bits simultaneously.

The minimal data cycle of the serial link is bounded by one d_4 gate delay [4][5] due to the digital logic forming the serializer and de-serializer circuits, which consist of fast shift-registers that can deliver and consume one bit every d_4 [4]. The *N*-bit shift-register consists of *N*-1 Transition-Latch (XL) stages [5]. The serial link channel consists of either two or four lines for single-ended or differential signaling, respectively. Although differential signaling is preferred for lower power and higher rates, in this paper we analyze only the single-ended case since all other circuits that are compared below are single-ended. For 65nm technology, the typical d_4 gate delay is 15ps and the typical link data rate is hence 67 Gbps.

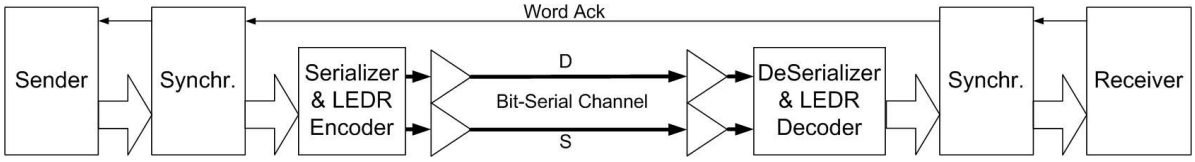

Figure 4: Serial communication scheme

III. Analytical Models

This section presents an analytical study of the performance and cost functions of the parallel and serial links.

A. *Bit-Rates*

1) *Bit-Rate of the Parallel Link*

The following factors bound the bit-rate B of the synchronous parallel link:

a. *Fastest available clock*. The shortest clock cycle that can be generated using a ring oscillator is typically bounded by $8 \cdot d_4$ [20], resulting in 8 GHz for 65nm technology [4]. However, most SoC modules operate at slower clock rates, e.g. $11d_4$ for modern fast processors [35] and $100\text{-}400 \cdot d_4$ for standard SoC/ASIC (a digital IC based on standard cells and designed using standard EDA tools) [36].

b. *Synchronization Latency*. Data synchronization is required at the receiver side of the link. Synchronization latency depends on the relationship between transmitter and receiver clocks and on synchronizer architecture. The worst case relates to mutually-asynchronous clocks, when synchronization may take several cycles. Faster synchronization is possible when using high performance synchronizers [37][38].

c. *Clock uncertainty*. This is typically added to the design critical path. In source synchronous communication, clock uncertainty extends the minimal data cycle on the link, as explained below.

d. *Delay and Delay Uncertainty of the link*. In register-pipelined links (Figure 1), both global clock cycle and delay uncertainty bound the link performance. In wave-pipelined links (Figure 2), the data rate is not bounded any more by the clock cycle, but only by delay uncertainty. A long parallel wave-pipelined link may carry multiple words simultaneously when its delay is longer than the transmitter clock cycle. The delay uncertainty of the link results from the following factors:

   i. The skew and jitter of the clock.

   ii. Repeater delay variations. Uncertainty grows monotonically with the number of repeaters [29].

   iii. Wire delay variations, mostly due to variations in metal thickness that affect resistance [39][40].

   iv. Via variations [39].

   v. Cross-Coupling. Unknown bit-patterns sent through the parallel link may result in cross-talk noise that affects the delay of the victim lines in an unpredictable way. Links should be optimized for worst-case

switching patterns that cause worst cross-coupling noise. Cross-coupling is usually mitigated by means of shielding and spacing [3][30].

vi. Geometry. Wide busses may encounter routing limitations, resulting in different geometries for different link lines (even in the same metal layer). This, of course, changes the worst-case link delay. In a multi-layer link structure the link delay is bounded by its slowest (lowest) metal layer (this paper analyzes only single layer interconnects).

Seeking to achieve maximal bit-rate, we first analyze the delay uncertainty of the wave-pipelined parallel link and then extend the results to register-pipelined links. The following two worst cases bound the minimal clock cycle of the link:

a. Latest data clocking: the latest signal should arrive early enough to be clocked by the sampling register at the receiver (namely the signal should arrive before CLKR in Figure 2).
b. Earliest data clocking: The first arrival of the next signal should not interfere with sampling of the previous word.

We adopt the notation of [27] and draw the delay uncertainty for source-synchronous communication in Figure 5. The clock cycle is restricted as follows:

$$T_{CLK} > 2 \cdot (\delta_{MAX} - \delta_{MIN}) + 4 \cdot \Delta_{CLK} + T_{SU} + T_H \qquad (1)$$

where δ_{MAX} and δ_{MIN} are the max and min data delays (which are also the clock uncertainty in source-synchronous communication), Δ_{CLK} is the one side clock skew, and T_{SU}, T_H are the setup and hold times of a flip-flop. Below we explore the dependency of T_{CLK} on other parameters: the link width *N* and the link length *L*.

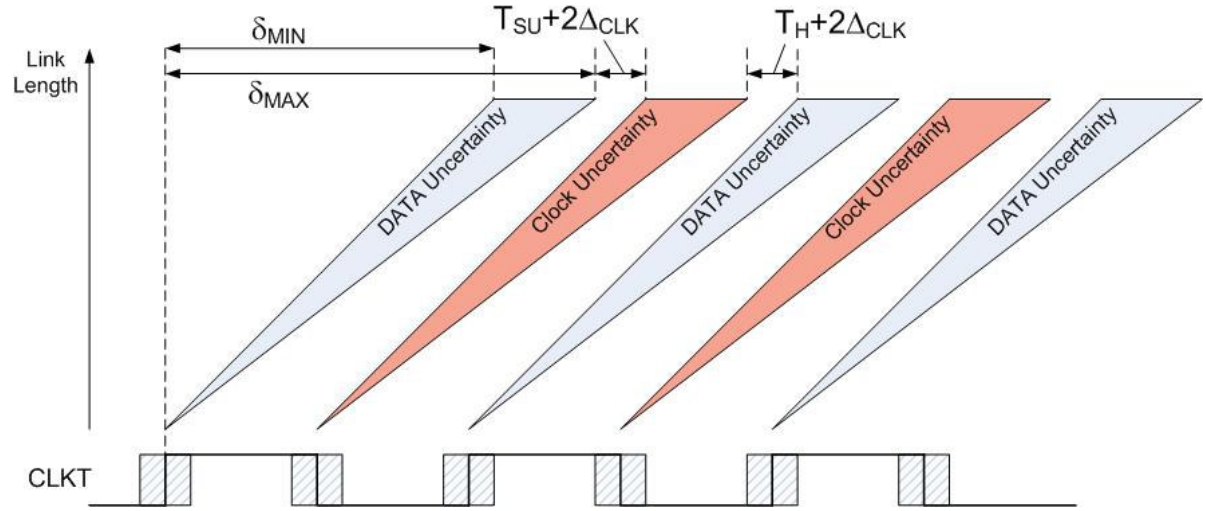

Figure 5: Parallel link minimal clock cycle is limited by clock jitter and skew and by link-length-dependent delay differences among the parallel wires due to variations and cross-talk

There are two types of in-die variations: random variations of closely placed devices and "systematic" variations, which typically depend on location on the die [41][42]. When a single line is considered, its delay may vary significantly (up to a factor of four) relative to the nominal delay, due to systematic variations in repeaters that are placed far from one another, and due to interconnect variations [39]. But when multiple lines are involved, such as in a parallel link, the effect of their relative skew (δ_{MAX}–δ_{MIN}) on T_{CLK} should also be analyzed. The effect of repeater transistor variation on that skew is low, thanks to the correlation between δ_{MAX} and δ_{MIN}, as follows. Repeaters that belong to the same stage (see Figure 2) are highly correlated in terms of systematic variations since they are placed close together. In addition, since repeaters are typically large [43], their random variations are averaged out. The length of the link also affects the skew: many repeaters result in smaller relative skew, because systematic inter-stage variations are averaged along the link. To conclude, even though the delay uncertainty of a single wire can be very high, the relative skew among the lines of a parallel link due to process variations is small and can be neglected.

The worst case delay of a single-wire with repeaters can be expressed as follows:

$$D^{Worst} = v_{SI} \cdot K \cdot d_{RPTR} + v_{INT} \cdot K \cdot d_{INT}(L) \qquad (2)$$

where v_{SI} is transistor variation coefficient (up to 30%), K is the number of repeaters, d_{RPTR} is the nominal delay of single repeater, v_{INT} is the interconnect variation coefficient (up to ×3 [39]), L is the wire length, $d_{INT} \approx 0.5 \cdot R_{INT} \cdot C_{INT} \cdot (L/K)^2$ [44][45], R_{INT} and C_{INT} are wire resistance and capacitance per length computed according to [43] and [46]. Note that this approximation of d_{INT} leads to an optimistic estimate of the total delay of the parallel link, and consequently to an optimistic estimate of the performance of the parallel link. When cross-coupling is also taken into account the wire delay is multiplied by η as follows:

$$D^{Worst}_{+cross-talk} = v_{SI} \cdot K \cdot d_{RPTR} + \eta \cdot v_{INT} \cdot K \cdot d_{INT}(L) \qquad (3)$$

The worst case skew Φ between two lines in the parallel link happens when one line is victimized by the worst possible aggression, while the other one experiences no aggression at all. Since the relative skew due to process variation can be neglected, $\Phi \approx \delta_{MAX} - \delta_{MIN}$:

$$\delta_{MAX} - \delta_{MIN} \approx \Phi(L) = D^{Worst}_{+cross-talk} - D^{Best}_{+cross-talk} = (\eta^{Worst} - \eta^{Best}) \cdot (v_{INT} \cdot K \cdot d_{INT}(L)) \qquad (4)$$

Values of η differences, based on PTM [46] and [30], are listed in Table 1. Actual shielding as in Figure 3 yield factors somewhat larger than zero, but we use zero for fully shielded links for the sake of simplicity.

Table 1: Coupling Factors Residual, η^{WC}—η^{BC}

Shielding	η^{Worst}—η^{Best}
Not-Shielded	1.9
Fully-Shielded	0

Combining Eq. (1) and (4) we get:

$$T_{CLK} > 2 \cdot \Phi(L) + 4 \cdot \Delta_{CLK} + T_{SU} + T_H \qquad (5)$$

The parallel link clock frequency is demonstrated for 65nm technology in Figure 6 as a function of length. This is based on the following assumptions: Δ_{CLK} is 10% of the clock cycle, $T_{SU}+T_H$=50 ps (about $3d_4$), and $v^{WC}_{INT} = 3$ [39]. The minimal clock cycle is $8 \cdot d_4$, as discussed above. Note that for the fully-shielded link and for very short distances of the unshielded link, the rate is bounded by clock cycle rather than by delay uncertainty. Since in typical SoC the clock cycle is substantially longer than $8 \cdot d_4$, the maximal link rate is smaller. This is expressed as follows:

$$T^{PAR}_{CLK} = \max\{2 \cdot \Phi(L) + 4 \cdot \Delta_{CLK} + T_{SU} + T_H, \quad T_{SYSTEM-CLOCK}\} \qquad (6)$$

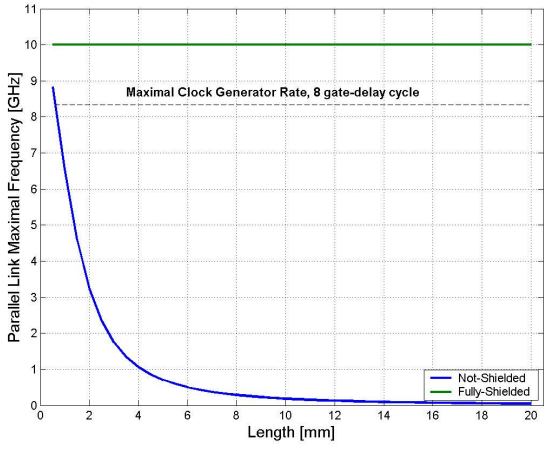

Figure 6: Wave-pipelined parallel link maximal frequency

In the register-pipelined parallel link the clock rate is equal to the system clock rate, and delay uncertainty affects the distance between successive pipeline stages.

2) Serial Link Bit-Rate

Serial links differ from parallel links in two ways: the serial link consists of only two wires (Figure 4), and the coupling

factor over the serial link is always known [4]. The skew due to in-die variations over a serial link is much smaller than for even the narrowest of parallel links (eight-bit parallel) and therefore the skew is neglected. In addition, thanks to the fact that in the serial link only one of the two lines changes per every new bit, the skew is not affected by cross-coupling, and as a result the link delay is the same for all symbols. The minimal data cycle of the link is d_4 (a new bit is sent every gate delay), resulting in maximal bit-rate of:

$$B_{SER} = \frac{1}{d_4} \qquad (7)$$

E.g., for 65nm, B_{SER}=67Gbps.

3) Interconnect Characteristics

Note that the parallel and serial lines operate at very different rates. While the parallel link operates at the "RC" region, the serial link operates at the "RLC" region [47]. Repeater insertion is treated differently for these two domains [48]. Moreover, the cost and latency of the serial link can be further improved when interconnects without repeaters are considered [8]. These considerations are applied in the following.

B. Area

1) Wave-Pipelined Parallel Link Area

In general, the total area requirement consists of silicon (drivers and repeaters) and interconnect (wires, shields and spacing):

$$A_{PAR} = (A^{PAR}_{DRIVERS} + A^{PAR}_{REPEATERS}) + (A_{WIRES} + A_{SHIELDS}) \qquad (8)$$

More precisely the area requirement is expressed as follows:

$$A_{PAR} = (N+1)\cdot\left[A^{PAR}_{DRIVER} + k^{PAR}_{RPTR}\cdot L\cdot A_{RPTR} + (k^{PAR}_{RPTR}\cdot L+1)\cdot(A_{WIRE} + s\cdot A_{SHIELD})\right] \qquad (9)$$

where:

$N+1$	N data bits and one clock line,
A^{PAR}_{DRIVE}	area of the parallel wire driver
k^{PAR}_{RPTR}	number of repeaters per unit length of a wire,
A_{RPTR}	area of the parallel wire repeater,
$k^{PAR}_{RPTR}\cdot L+1$	number of wire segments.
A_{WIRE}	wire segment area, including spacing,
s	shielding coefficient (s=0 for no shielding and s=1 for full shielding),
A_{SHIELD}	shield segment area, including spacing,

We assume that $A_{SHIELD}\approx A_{WIRE}$,:

$$A_{PAR} = (N+1)\cdot\left[A^{PAR}_{DRIVER} + k^{PAR}_{RPTR}\cdot L\cdot A^{PAR}_{RPTR} + (k^{PAR}_{RPTR}\cdot L+1)\cdot(s+1)\cdot A^{PAR}_{WIRE}\right] \qquad (10)$$

Driver and repeater areas are computed according to Eq. (11) and (12), respectively, where A_{INV} is minimal inverter size and $\{h\}$ are sizing factors optimized for minimal delay [43][49][50]. The drivers assume cascaded buffers as in Figure 1.

$$A^{PAR}_{DRIVER} = \sum_{i=1}^{K_{CAS}} h^i_{CSCD}\cdot A_{INV} = \delta\cdot A_{INV} \qquad (11)$$

$$A^{PAR}_{RPTR} = h^{PAR}_{RPTR}\cdot A_{INV} \qquad (12)$$

In Section IV we consider the active silicon and the interconnect separately, since they differ significantly in area, scaling and leakage issues. Let us split Eq. (10) into Eq. (13) and (14) for active and interconnect areas.

$$A^{PAR}_{ACTIVE} = (N+1)\cdot\left[\delta + k^{PAR}_{RPTR}\cdot L\cdot h^{PAR}_{RPTR}\right]\cdot A_{INV} \qquad (13)$$

$$A^{PAR}_{INT} = (N+1)\cdot(k^{PAR}_{RPTR}\cdot L+1)\cdot(s+1)\cdot A^{PAR}_{WIRE} \qquad (14)$$

2) Register-pipelined Parallel Link Area

The register-pipelined parallel link contains pipeline stages as well as repeaters in between them. The number of stages M_{FF} depends on the clock cycle, on the link length and on the delay uncertainty of the link. Note that the pipeline stages act as repeaters, in addition to re-synchronizing the signal. The active area is obtained as follows (A_{FF} is area of a single flip-flop):

$$A_{ACTIVE}^{PAR} = (N+1)\cdot\left[A_{DRIVER}^{PAR} + (k_{RPTR}^{PAR}\cdot L - M_{FF})\cdot A_{RPTR}^{PAR} + M_{FF}\cdot A_{FF} \right] \quad (15)$$

M_{FF} is computed as follows:

$$M_{FF} = \frac{L}{T_{CLK}^{PAR}\cdot V_P} \quad (16)$$

where V_P is the propagation velocity of the voltage front, L is the link length and T is computed according to Eq. (6). We assume that the flip-flop is larger than the repeater by a factor θ:

$$A_{FF} = \theta\cdot A_{RPTR}^{PAR} = \theta\cdot h_{RPTR}^{PAR}\cdot A_{INV} \quad (17)$$

Thus:

$$A_{ACTIVE}^{PAR} = (N+1)\cdot\left[\delta + (k_{RPTR}^{PAR}\cdot L - M_{FF})\cdot h_{RPTR}^{PAR} + M_{FF}\cdot\theta\cdot h_{RPTR}^{PAR} \right]\cdot A_{INV} \quad (18)$$

The interconnect area is similar to that of the wave-pipelined parallel link (Eq. (14)).

3) Serial Link Area

In addition to the components of Eq. (8), the area of the serial link contains also the SERDES (shields are neglected for serial links):

$$A_{SER} = (A_{SERDES} + A_{DRIVERS}^{SER} + A_{REPEATERS}^{SER}) + A_{WIRES} \quad (19)$$

Given transistor count and ratios for XL stages (transition latches) of the *N*-bit shift-register [5], the SERDES area normalized to A_{INV} is:

$$A_{SERDES} = \kappa\cdot N\cdot A_{INV} = 240\cdot N\cdot A_{INV} \quad (20)$$

According to [5], $\kappa = 240$. The serial link also contains a LEDR encoder of negligible area [4]. Substituting expression (20) into (19) and noting that the link consists of two wires, we get:

$$A_{SER} = \kappa\cdot N\cdot A_{INV} + 2\cdot[A_{DRIVER}^{SER} + k_{RPTR}^{SER}\cdot L\cdot A_{RPTR}^{SER} + (k_{RPTR}^{SER}\cdot L + 1)\cdot A_{WIRE}^{SER}] \quad (21)$$

The number and size of the repeaters on the serial link are smaller than for parallel links by factors λ,γ thanks to RLC characteristics of the serial link [48]:

$$h_{RPTR}^{SER} = \sqrt{\frac{R_0\cdot C_{INT}}{R_{INT}\cdot C_0}}\cdot\lambda = h_{RPTR}^{PAR}\cdot\lambda \quad (22)$$

$$k_{RPTR}^{SER} = \sqrt{\frac{0.4\cdot R_{INT}\cdot C_{INT}}{0.7\cdot R_0\cdot C_0}}\cdot\gamma = k_{RPTR}^{PAR}\cdot\gamma \quad (23)$$

The trade-off between latency and power can be optimized further [51], but this is not considered here. A_{RPTR}^{SER} is computed according to Eq. (22):

$$A_{RPTR}^{SER} = h_{RPTR}^{SER}\cdot A_{INV} = \lambda\cdot h_{RPTR}^{PAR}\cdot A_{INV} \quad (24)$$

Like the repeaters, the driver is also smaller than in the parallel link:

$$A_{DRIVER}^{SER} = \chi\cdot A_{INV} \quad (25)$$

Thus, the active and interconnect areas are:

$$A_{ACTIVE}^{SER} = \left[\kappa\cdot N + 2(\chi + k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR}\cdot\lambda\cdot\gamma)\right]\cdot A_{INV} \quad (26)$$

$$A_{INT}^{SER} = 2\cdot(k_{RPTR}^{PAR}\cdot L\cdot\gamma + 1)\cdot A_{WIRE}^{SER} \quad (27)$$

4) Area Ratio Expressions

Given the area expressions for each case, we can derive the parallel-to-serial area ratios. For example, the area ratio of the

active portions of the serial and wave-pipelined links (Eq. (13) and (26)), which is also indicative of leakage power, is:

$$\rho_{ACTIVE\_AREA} = \frac{A_{ACTIVE}^{PAR}}{A_{ACTIVE}^{SER}} = \frac{(N+1)\cdot\left[\delta + k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR}\right]}{\kappa\cdot N + 2\cdot(\chi + k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR}\cdot\lambda\cdot\gamma)} \quad (28)$$

We look for the link length L_{AREA} above which the serial link takes less area than the parallel link ($\rho_{ACTIVE\_AREA}\geq 1$). In Eq. (28) driver's components δ and χ can be neglected relative to the repeaters:

$$\begin{aligned} &\frac{(N+1)\cdot k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR}}{\kappa\cdot N + 2\cdot k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR}\cdot\lambda\cdot\gamma} \geq 1 \\ &L_{AREA} \geq \frac{\kappa\cdot N}{k_{RPTR}^{PAR}\cdot h_{RPTR}^{PAR}\cdot(N+1-2\cdot\lambda\cdot\gamma)} \end{aligned} \quad (29)$$

Eq. (29) has three main parameters k_{REP}^{PAR}, h_{REP}^{PAR} and N ($\lambda\cdot\gamma$ is negligible). Both k_{REP}^{PAR} and h_{REP}^{PAR} depend on the technology node. According to data from [1][46], the number of repeaters per unit length k_{REP}^{PAR} grows roughly linearly as feature size shrinks (the extrapolation in Figure 7 beyond 65nm is speculative), while the repeater relative size h_{REP}^{PAR} stays almost the same. Therefore, as technology advances, the serial approach becomes preferable for shorter ranges. For example, Figure 8 shows L_{AREA} for the fully-shielded parallel link (N=8, the bit-rates are the same for the parallel and serial links, and the extrapolation beyond 65nm is speculative).

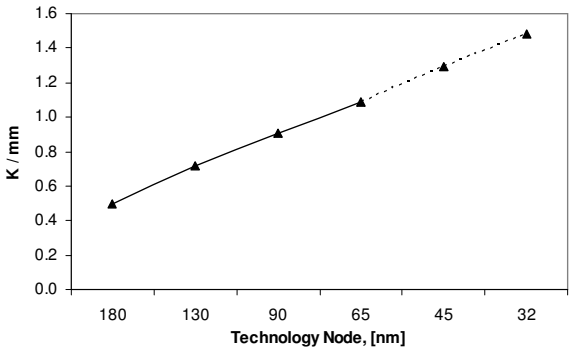

Figure 7: Number of repeaters (per millimeter) vs. technology node

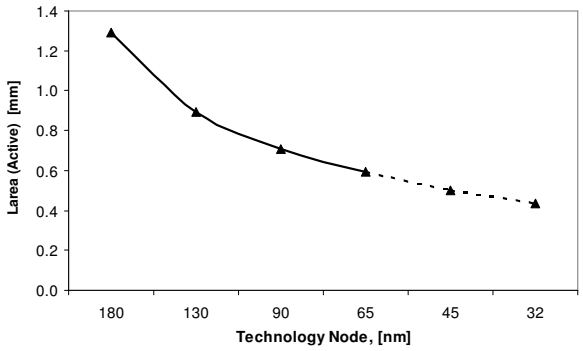

Figure 8: Minimal link length above which the active area and leakage of the serial link are smaller than in parallel link

The ratio of interconnect areas (practically the total link area) is expressed as follows (Eq. (14)):

$$\rho_{INT\_AREA} = \frac{A_{INT}^{PAR}}{A_{INT}^{SER}} = \frac{(N+1)\cdot(s+1)}{2} \quad (30)$$

Again, in fully shielded case where N≥8 it is clear from Eq. (30) that the serial link always consumes less interconnect area.

C. Power

1) Wave-Pipelined Parallel Link Power

The total link power comprises dynamic and standby power,

$$P_{TOT} = u\cdot P_{DYN} + P_{STANDBY} \quad (31)$$

The utilization u can be very low, typically less than 30%. Assume that average data patterns would incur $N/2$ transitions per word, plus two clock transitions:

$$P_{DYN}^{PAR} = (N/2+2)\cdot C_{TOT}\cdot V_{DD}^2\cdot F_{PAR} = (N/2+2)\cdot (k_{RPTR}^{PAR}\cdot L\cdot C_{RPTR}^{PAR} + C_{DRIVER}^{PAR} + (k_{RPTR}^{PAR}\cdot L+1)\cdot C_{WIRE}^{PAR})\cdot V_{DD}^2\cdot F_{PAR} \quad (32)$$

C_{TOT} comprises driver, repeaters and wire capacitances, and F_{PAR} is the link clock frequency. Normalizing to minimal inverter capacitance,

$$C_{RPTR}^{PAR} = h_{RPTR}^{PAR}\cdot C_{INV} \quad (33)$$

$$C_{DRIVER}^{PAR} = \delta\cdot C_{INV} \quad (34)$$

$$C_{WIRE}^{PAR} = \beta\cdot C_{RPTR}^{PAR} = \beta\cdot h_{RPTR}^{PAR}\cdot C_{INV} \quad (35)$$

where β is the ratio between the capacitances of wire segment and repeater (β≈1). Eq. (32) is now rewritten as

$$\begin{aligned} P_{DYN}^{PAR} &= (N/2+2)\cdot (k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR} + \delta + (k_{RPTR}^{PAR}\cdot L+1)\cdot \beta\cdot h_{RPTR}^{PAR})\cdot C_{INV}\cdot V_{DD}^2\cdot F_{PAR} \\ &= (N/2+2)\cdot (k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR}\cdot (1+\beta) + \delta + \beta\cdot h_{RPTR}^{PAR})\cdot C_{INV}\cdot V_{DD}^2\cdot F_{PAR} \end{aligned} \quad (36)$$

When the clock is gated (neglecting the power of clock gating circuits) the standby power is the leakage power:

$$P_{STANDBY} = P_{LEAK} = A_{ACTIVE}\cdot V_{DD}\cdot I_{OFF} = (N+1)\cdot (k_{RPTR}^{PAR}\cdot L\cdot A_{RPTR} + A_{DRIVER}^{PAR})\cdot V_{DD}\cdot I_{OFF} \quad (37)$$

where I_{OFF} is the off-current per device area [1][52].

2) Register-pipelined Parallel Link

The dynamic power expression for the pipelined wire case is as follows[1]:

$$P_{DYN}^{PAR} = (N/2+2)\cdot C_{TOT}\cdot V_{DD}^2\cdot F_{PAR} = \left[(N/2+2)\cdot ((k_{RPTR}^{PAR}\cdot L - M_{FF})\cdot C_{RPTR}^{PAR} + C_{DRIVER}^{PAR} + (k_{RPTR}^{PAR}\cdot L+1)\cdot C_{WIRE}^{PAR}) + (N/2)\cdot M_{FF}\cdot C_{FF}\right]\cdot V_{DD}^2\cdot F_{PAR} \quad (38)$$

Where

$$C_{FF} = \theta\cdot C_{RPTR}^{PAR} = \theta\cdot h_{RPTR}^{PAR}\cdot C_{INV} \quad (39)$$

Normalizing to C_{INV} we get:

$$P_{DYN}^{PAR} = \left[(\frac{N}{2}+2)\cdot (k_{RPTR}^{PAR}\cdot L\cdot h_{RPTR}^{PAR}(1+\beta) + \delta + h_{RPTR}^{PAR}(\beta - M_{FF})) + (N/2)\cdot M_{FF}\cdot \theta\cdot h_{RPTR}^{PAR}\right]\cdot C_{INV}\cdot V_{DD}^2\cdot F_{PAR} \quad (40)^2$$

In the register-pipelined case there is an additional leakage power component due to pipeline stage logic (here again gated-clock is assumed):

$$P_{LEAK} = (N+1)\cdot (A_{DRIVER}^{PAR} + (k_{RPTR}^{PAR}\cdot L - M_{FF})\cdot A_{RPTR} + M_{FF}\cdot A_{FF})\cdot V_{DD}\cdot I_{OFF} \quad (41)$$

3) Serial Link

Total power for serial link is computed according to Eq. (31), noting that the serial link always incurs one transition per bit [4]. The dynamic power of the serial link is:

$$P_{DYN}^{SER} = P_{DYN}^{SERDES} + P_{DYN}^{CHANNEL} = 2\cdot a_{SR}\cdot C_{SR}\cdot V_{DD}^2\cdot F_{PAR} + (k_{RPTR}^{SER}\cdot L\cdot C_{RPTR}^{SER} + C_{DRV}^{SER} + (k_{RPTR}^{SER}\cdot L+1)\cdot C_{WIRE}^{SER})\cdot V_{DD}^2\cdot B_{SER} \quad (42)$$

The factor of two relates to the two shift-registers in the transmitter and in the receiver. C_{SR} is a single shift-register capacitance and a_{SR} is the activity factor, accounting also for toggling inside the shift-register. A more detailed $a_{SR}\cdot C_{SR}$ product is:

[1] The clock input capacitance of flip-flops is ignored.

[2] The capacitance of the clock tree is ignored.

$$a_{SR} \cdot C_{SR} = \frac{1}{2} \cdot S_F \cdot \left[80 \cdot (N / S_F)^2 + 20 \cdot 0.5 \cdot 2 \cdot \frac{(N / (2 \cdot S_F)) \cdot ((N / (2 \cdot S_F)) + 1)}{2} \right] \cdot C_{INV} \quad (43)$$

The factor of 1/2 is for the equivalent frequency per each sub shift-register [5], and there are S_F sub-registers, each comprising the equivalent of 80 minimal inverters in the control section and 20 in each of the N/S_F XL stages in the data path. The activity factor of the data path is 0.5, and there are two data paths in each XL. Finally, $(N/(2 \cdot S_F)) \cdot ((N/(2 \cdot S_F))+1)/2$ transitions are required each word in each sub-register [5]. As evident from Eq. (43), the dynamic power of the serial link depends quadratically on N, and therefore for long words the link is split into parallel registers [4]. The split factor S_F defines the split extent, while the minimal size of sub-register is eight bits.

As previously, we define:

$$C_{RPTR}^{SER} = h_{RPTR}^{SER} \cdot C_{INV} = \lambda \cdot h_{RPTR}^{PAR} \cdot C_{INV} \quad (44)$$

$$C_{DRV}^{SER} = \chi \cdot C_{INV} \quad (45)$$

$$C_{WIRE}^{SER} = C_{WIRE}^{PAR} \quad (46)$$

Normalizing for C_{INV} we obtain:

$$\begin{aligned} & P_{DYN}^{SER} = P_{DYN}^{SERDES} + P_{DYN}^{CHANNEL} = \\ & (42 \cdot N^2 + 5 \cdot N) \cdot C_{INV} \cdot V_{DD}^2 \cdot F_{PAR} + \\ & (k_{RPTR}^{PAR} \cdot h_{RPTR}^{PAR} \cdot L \cdot \gamma \cdot \lambda + \chi + (k_{RPTR}^{PAR} \cdot L + 1) \cdot \beta \cdot h_{RPTR}^{PAR}) \cdot C_{INV} \cdot V_{DD}^2 \cdot B_{SER} \\ & = (42 \cdot N^2 + 5 \cdot N) \cdot C_{INV} \cdot V_{DD}^2 \cdot F_{PAR} + \\ & (k_{RPTR}^{PAR} \cdot h_{RPTR}^{PAR} \cdot L \cdot (\gamma \cdot \lambda + \beta) + \chi + \beta \cdot h_{RPTR}^{PAR}) \cdot C_{INV} \cdot V_{DD}^2 \cdot B_{SER} \end{aligned} \quad (47)$$

Leakage power is dissipated by the serdes, line driver and repeaters:

$$P_{LEAK}^{SER} = \left[A_{SERDES} + 2 \cdot (A_{DRIVER}^{SER} + k_{RPTR}^{SER} \cdot L \cdot A_{RPTR}) \right] \cdot V_{DD} \cdot I_{OFF} \quad (48)$$

4) Power Ratio Expressions

We look for the link length L_{POWER} above which the serial link dissipates less power than the parallel link. To compare same bit rates, $N=B_{SER}/F_{PAR}$.

$$\begin{aligned} & \rho_{POWER} = \frac{P_{DYN}^{PAR}}{P_{DYN}^{SER}} = \\ & = \frac{(N/2+2) \cdot (k_{RPTR}^{PAR} \cdot L \cdot h_{RPTR}^{PAR} \cdot (1+\beta) + \delta + \beta \cdot h_{RPTR}^{PAR}) \cdot C_{INV} \cdot V_{DD}^2 \cdot F_{PAR}}{\left[(42 \cdot N^2 + 5 \cdot N)/N + (k_{RPTR}^{PAR} \cdot h_{RPTR}^{PAR} \cdot L \cdot (\gamma \cdot \lambda + \beta) + \chi + \beta \cdot h_{RPTR}^{PAR}) \right] \cdot C_{INV} \cdot V_{DD}^2 \cdot B_{SER}} \\ & = \frac{(N/2+2) \cdot (k_{RPTR}^{PAR} \cdot h_{RPTR}^{PAR} \cdot (1+\beta) + \delta + \beta \cdot h_{RPTR}^{PAR})}{\left[(42 \cdot N + 5) + (k_{RPTR}^{PAR} \cdot h_{RPTR}^{PAR} \cdot L \cdot (\gamma \cdot \lambda + \beta) + \chi + \beta \cdot h_{RPTR}^{PAR}) \right] \cdot N} \end{aligned} \quad (49)$$

Solving for $\rho_{POWER} \geq 1$ and neglecting δ and χ as above,

$$\frac{(N/2+2) \cdot (k_{RPTR}^{PAR} \cdot L \cdot h_{RPTR}^{PAR} \cdot (1+\beta) + \beta \cdot h_{RPTR}^{PAR})}{\left[(42 \cdot N + 5) + (k_{RPTR}^{PAR} \cdot h_{RPTR}^{PAR} \cdot L \cdot (\gamma \cdot \lambda + \beta) + \beta \cdot h_{RPTR}^{PAR}) \right] \cdot N} \geq 1 \quad (50)$$

$$L_{POWER} \geq \frac{42 \cdot N^2 + (5 + \beta \cdot h_{RPTR}^{PAR}/2) \cdot N - 2 \cdot \beta \cdot h_{RPTR}^{PAR}}{k_{RPTR}^{PAR} \cdot h_{RPTR}^{PAR} \cdot \left[(1+\beta) \cdot (N/2+2) - (\gamma \cdot \lambda + \beta) \cdot N \right]}$$

Eq. (50) depends on k_{REP}^{PAR}, h_{REP}^{PAR} and N, similarly to Eq. (29), and likewise the length threshold becomes smaller with the shrinking feature size (Figure 9, extrapolated speculatively). The results are slightly better when exact computations are performed as shown in later sections.

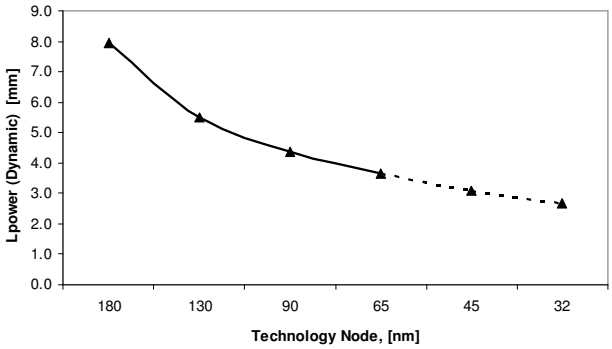

Figure 9: Minimal link length above which less dynamic power is dissipated by the serial link than the parallel link.

D. Relative Latency Overhead

As above, the inherent wire delay is $d_{INT} \approx 0.5 \cdot R_{INT} \cdot C_{INT} \cdot (L)^2$. We consider the additional latency incurred by the various links.

1) Wave-Pipelined Parallel Link

The relative latency overhead is

$$\Lambda_{WP}^{PAR} = (\eta^{Worst} - 1) \cdot v_{INT} \cdot d_{INT} \quad (51)$$

2) Register-pipelined Parallel Link

For register-pipelined parallel link the overhead consists also of additional pipeline stage delays:

$$\Lambda_{RP}^{PAR} = (\eta^{Worst} - 1) \cdot v_{INT} \cdot d_{INT} + M_{FF} \cdot (d_{FF} - d_{RPTR}) \quad (52)$$

3) Serial Link

The serial link latency consists of the time preserved for serialization and the flight time over the channel (d_{INT}). Then, the overhead is the serialization time: $N \cdot d_4$.

Delay uncertainty affects also the serial link resulting in skew in between the D and S lines. However, the delay uncertainty was found to be much smaller since in-die variations for two closely placed wires are much smaller than for a wider link. In addition, the number of repeaters was also smaller thanks to working in RLC region. Hence we neglect the delay uncertainty due to in-die variations.

The coupling noise is also small in the serial structure. Since in the serial link there are no concurrent transitions, the same pattern is sent for each bit [30], resulting always in the same delay over the channel. Special layout mitigates the crosstalk further enabling differential encoding over the channel [4]. In addition, since density restrictions are less strict for serial channels, wider spacing can be employed for further cross-talk mitigation. Hence, we assume that for a serial line the cross-coupling noise can also be neglected.

IV. Comparative Analysis for 65nm

In this section we compare the area, power and latency of serial and parallel links that deliver bandwidth B_{SER}, the bit rate of the serial link (Eq. (7)). Figure 10 and Figure 13 show the parallel link widths that are required to achieve that bit rate in wave-pipelined and register-pipelined parallel links respectively.

Note that for ranges above 6mm, the unshielded wave-pipelined parallel link requires hundreds and thousands of lines in order to provide the required bit-rate. The same is true for register-pipelined links operating at low rates (clock cycle > $130 \cdot d_4$). Wide links over 128 lines are impractical and are marked by dotted lines in the analysis. Note that fully shielded links which double the number of wires may be limited to 64 bit lines. Figure 11 and Figure 14 compare active area of the links.

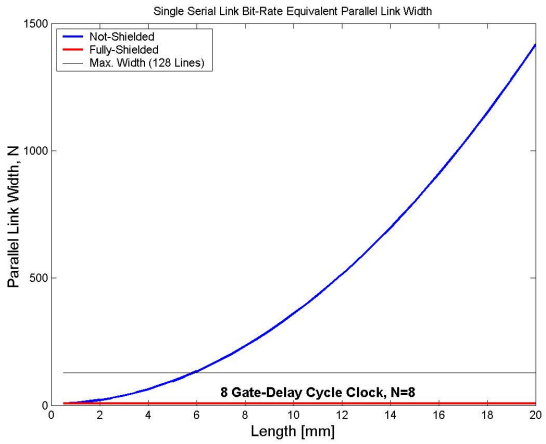

Figure 10: *Wave-pipelined* link width required to deliver B_{SER} bit rate

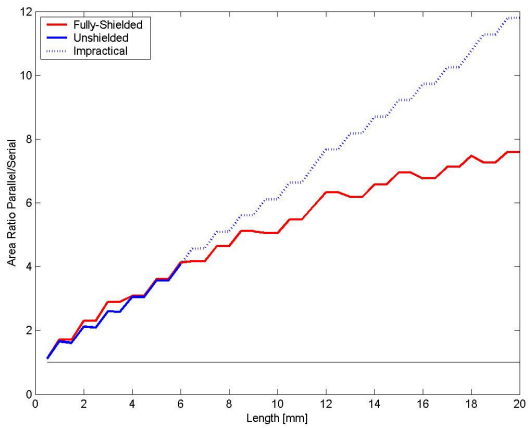

Figure 11: Ratio of active area and leakage power (wave-pipelined/serial)

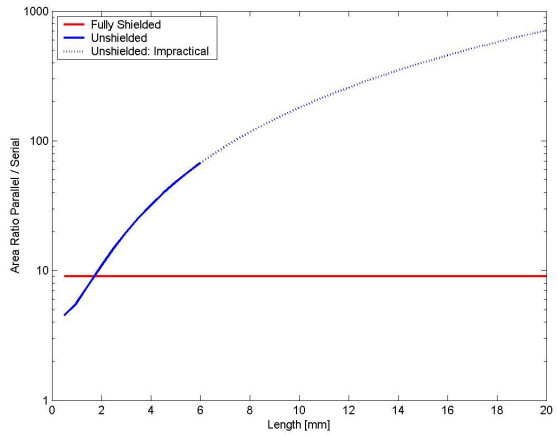

Figure 12: Ratio of interconnect Area (wave-pipelined/serial)

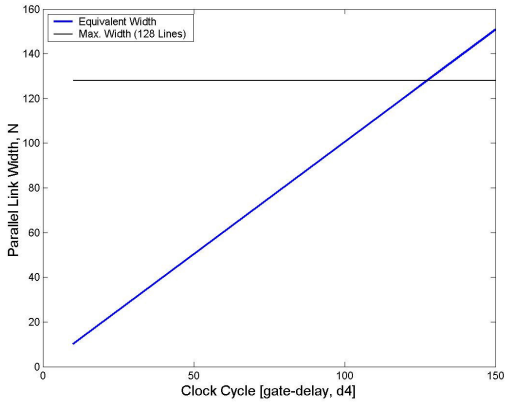

Figure 13: *Register-pipelined* link width required to deliver B_{SER} bit rate (same for unshielded and fully-shielded, bounded by system clock)

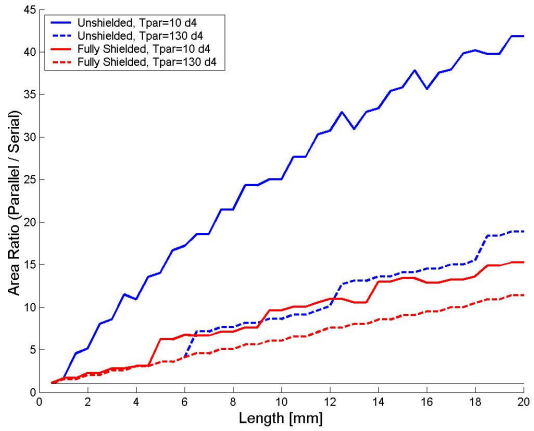

Figure 14: Ratio of active area and leakage power (register-pipelined/serial)

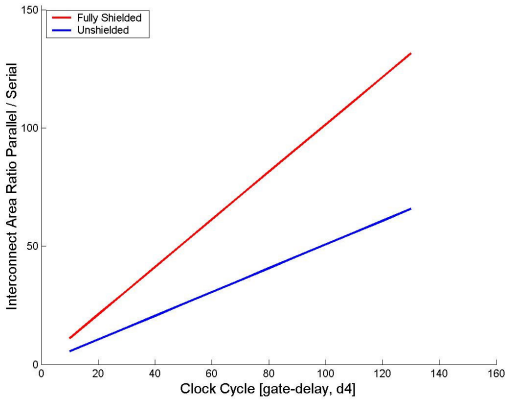

Figure 15: Ratio of interconnect area (register-pipelined/serial)

As expected there is a clear improvement in interconnect and total area requirement (Figure 12—Figure 17).

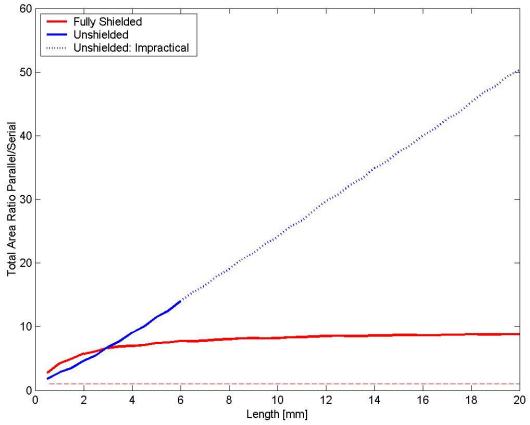

Figure 16: Total area ratio (wave-pipelined)

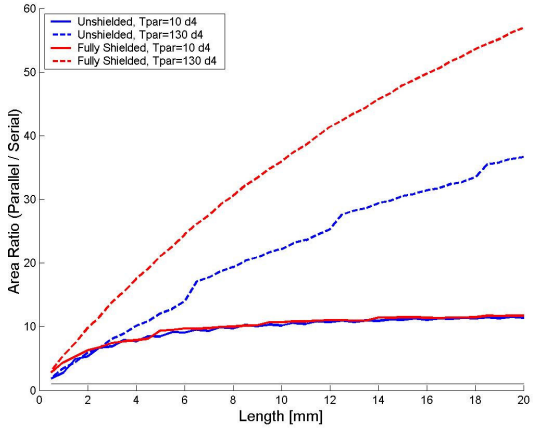

Figure 17: Total area ratio (register-pipelined)

The leakage power ratios are the same as active area (Figure 11 and Figure 14), as expected from Eq. (37), (41) and (48). The serial link dissipates less dynamic power than the fully-shielded wave-pipelined parallel link (Figure 18) at ranges above 2mm, and the unshielded wave-pipeline link dissipates less power at shorter lengths. Similarly, the serial link dissipates less power than the pipelined wire link (Figure 19), except for the fully shielded, slow and very wide parallel link (which requires significant area). Note that the dynamic power of the serial link consists of both channel power and the power dissipated by the SERDES registers. In all links, dynamic power is significantly higher than leakage, as is evident in Figure 20 and Figure 21 (one exception is the dotted segment in Figure 20, where leakage is proportional to the impractically large area). This observation is independent of utilization levels.

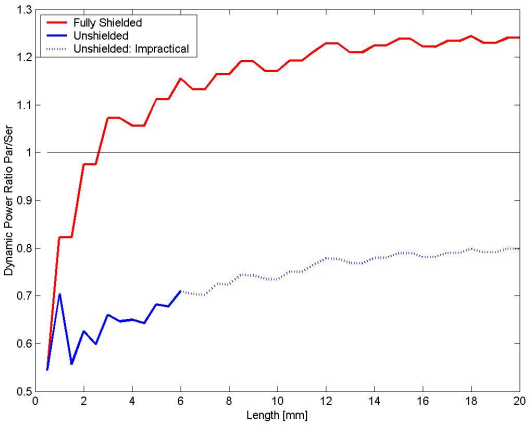

Figure 18: Ratio of dynamic power (wave-pipelined/serial)

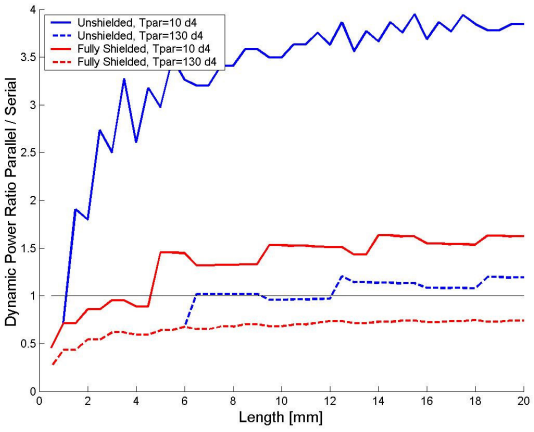

Figure 19: Ratio of dynamic power (register-pipelined/serial)

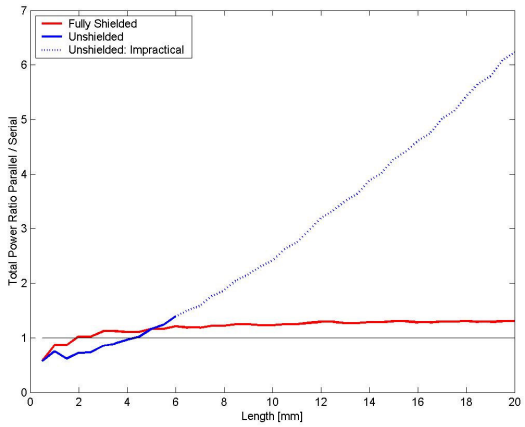

Figure 20: Total power ratio, 20% utilization (wave-pipelined parallel link)

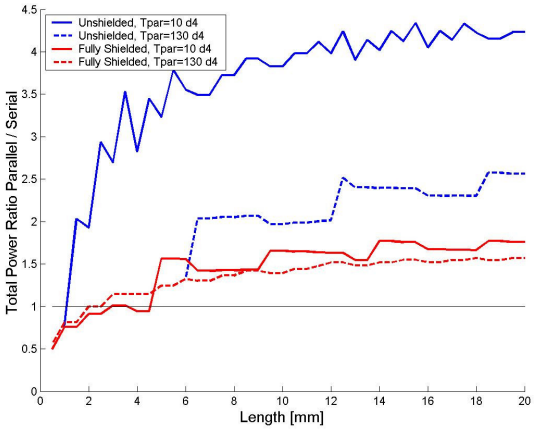

Figure 21: Total power ratio, 20% utilization (register-pipelined parallel link)

Latency overhead is presented in Figure 22 and Figure 23. The serial link incurs higher latency than unshielded wave-pipelined parallel link due to long SERDES shift-registers. In Figure 23, the register-pipelined links incur higher latencies at longer wires.

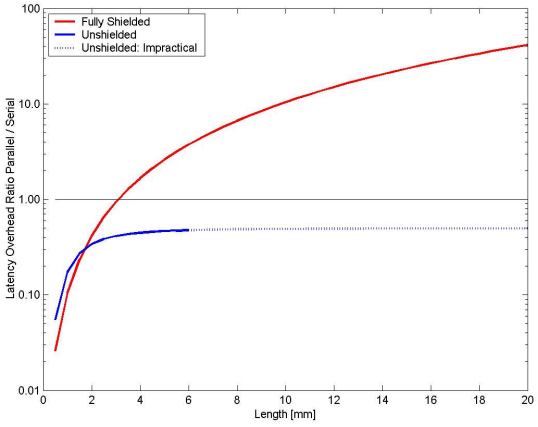

Figure 22: Latency overhead ratio (wave-pipelined)

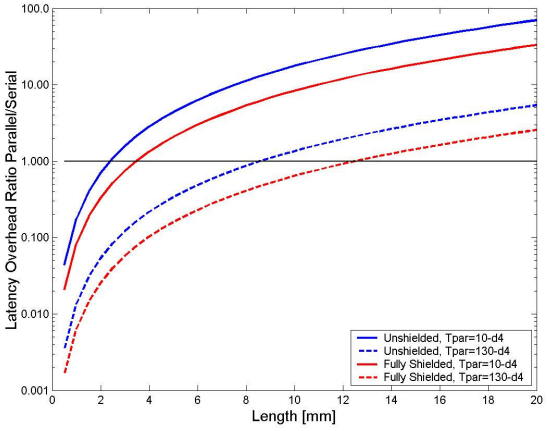

Figure 23: Latency overhead ratio (register-pipelined)

Table 2 summarizes the decision thresholds and costs for serial link employment. In the table, we specify minimal ranges for which a serial link is preferred over parallel. In some cases, the serial link is never better and for the specified minimal length incurs some penalty, which is also specified in the table. When there is no penalty, but only an improvement, only the minimal length is specified.

Table 2: 65nm example – minimal length above which the serial link is preferred (various criteria)

	Wave-Pipeline vs. Serial		Register-pipelined vs. Serial			
Shielding	Fully Shielded	Unshielded	Fully Shielded		Unshielded	
Length of parallel link	unlimited	Up to 6mm	unlimited		unlimited	
Clock cycle of parallel link	$8d_4$	$8d_4$	$10d_4$	$130d_4$	$10d_4$	$130d_4$
To minimize the following:	choose a serial link for links longer than:					
area	Always	Always	Always		Always	
power	2 mm	4mm	3mm	3mm	1mm	3mm
latency	2 mm	Never*	4mm	12mm	2mm	9mm

* The serial link incurs 2-10× latency overhead penalty for 0-6 mm link.

The table provides length thresholds above which we should prefer the serial link, depending on whether area, power or latency, or their combinations, are minimized.

V. CONCLUSIONS

Novel serial links outperform standard parallel links when long range communication is considered. This advantage scales with technology, making the serial links more attractive for shorter links in future technologies. Future large SoCs should employ serial links to mitigate the cost of communication in terms of area, congestion, power and latency. We have provided a detailed analysis of the serial link with an example for 65nm technology. In the example we compared the serial link versus two typical parallel links.

REFERENCES

[1] International Technology Roadmap for Semiconductors (ITRS), 2005.
[2] A. Morgenshtein, I. Cidon, A. Kolodny, R. Ginosar, "Low-Leakage Repeaters for NoC Interconnects", ISCAS, 600-603, 2005.
[3] R. Weerasekera, D. Pamunuwa, L. Zheng, H. Tenhunen, "Minimal-Power, Delay-Balanced Smart Repeaters for Interconnects in the Nanometer Regime," SLIP, 113-120, 2006.
[4] R. Dobkin, Y. Perelman, T. Liran, R. Ginosar, A. Kolodny, "High Rate Wave-Pipelined Asynchronous On-Chip Bit-Serial Data Link," ASYNC, 3-14, 2007.
[5] R. Dobkin, R. Ginosar, A. Kolodny, "Fast Asynchronous Shift Register for Bit-Serial Communication," ASYNC, 117-126, 2006.
[6] R. Dobkin, R. Ginosar, A. Kolodny, "High-Speed Serial Interconnect for NoC," NoC Workshop, DATE, 2006.
[7] R. Dobkin, I. Cidon, R. Ginosar, A. Kolodny, A. Morgenshtein, "Fast Asynchronous Bit-Serial Interconnects for Network-on-Chip," CCIT TR529, EE Pub No. 1480, EE Dept., Technion, 2005.
[8] A.P. Jose, G. Patounakis, K.L. Shepard, "Pulsed Current-Mode Signaling for Nearly Speed-of-Light Intrachip Communication," JSSC, 41(4), 772–780, 2006.
[9] I. Saastamoinen, T. Suutari, J. Isoaho, J. Nurmi, "Interconnect IP for gigascale system-on-chip," ECCTD, 116-120, 2001.
[10] T. Suutari, J. Isoaho, H. Tenhunen, "High-speed Serial Communication with Error Correction Using 0.25µm CMOS Technology," ISCAS, IV-618-621, 2001.
[11] S.J. Lee, K. Kim, H. Kim, N. Cho, H.J. Yoo, "Adaptive Network-on-Chip with Wave-front Train Serialization Scheme," Proc. VLSI Circuits, 104-107, 2005.
[12] A.J. Joshi, J.A. Davis, "Wave-pipelined multiplexed (WPM) routing for gigascale integration (GSI)," TVLSI 13(8): 889-910, 2005.
[13] A.J. Joshi, G.G. Lopez, J.A. Davis, "Design and Optimization of On-Chip Interconnects Using Wave-Pipelined Multiplexed Routing," TVLSI 9(15): 990-1002, 2007.
[14] J. Teifel, R. Manohar, "A High-Speed Clockless Serial Link Transceiver," ASYNC, 151-161, 2003.
[15] C.K.K. Yang, "Design of High-Speed Serial Links in CMOS", PhD Thesis, Stanford University, 1998.
[16] S. Sidiropoulos, "High Performance Inter-Chip Signaling," Tech. Rep. CSL-TR-98-760, Stanford Univ., 1998.
[17] W.F. Ellersick, "Data Converters for High Speed CMOS Links," PhD Thesis, Stanford Univ., 2001.
[18] H.O. Johansson, J. Yuan and C. Svensson, "A 4 Gsamples/s Line-Receiver in 0.8 um CMOS," Proc. Int. Symp. VLSI Circuits, pp. 116-117, 1996.
[19] C. Svensson and J. Yuan, "High Speed CMOS Chip to Chip Commication Circuit," ISCAS, 2228-2231, 1991.
[20] M.J.E. Lee, "An Efficient I/O and Clock Recovery for TERABIT Integrated Circuits Design," PhD Thesis, Stanford Univ., 2001.
[21] W. Bainbridge, S. Furber, "Delay Insensitive System-on-Chip Interconnect using 1-of-4 encoding", ASYNC, 118-126, 2001.
[22] R. Ho, J. Gainsley, R. Drost, "Long Wires and Asynchronous Control," ASYNC, 240-249, 2004.
[23] G. Lakshminarayanan, B. Venkataramani, "Optimization Techniques for FPGA-Based Wave-Pipelined DSP Blocks," IEEE Trans. VLSI, 13(7), 2005.
[24] C. Svensson, J. Yuan, "A 3-Level Asynchronous protocol for a Differential Two-Wire Communication Link," JSSC, 29(9), 1994.
[25] A. Morgenshtein, I. Cidon, A. Kolodny, R. Ginosar, "Comparative Analysis of Serial vs. Parallel Links in Networks on Chip," SoC, 185-188, 2004.

[26] J. Xu, W. Wolf, "Wave Pipelining for Application-Specific Networks-on-Chips," Proc. Int. Conf. Compilers, Architecture, and Synthesis for Embedded System, 198-201, 2002.
[27] J. Xu, W. Wolf, "Wave-Pipelined On-chip Interconnect Structure for Networks-on-Chips," HOTI, 10-14, 2003.
[28] W. P. Burleson, M. Ciesielski, F. Klass,W. Liu, "Wave-Pipelining: A Tutorial and Research Survey," TVLSI, 6(3):464-474, 1998.
[29] B.D. Winters, M.R. Greenstreet, "A Negative-Overhead, Self-Timed Pipeline," ASYNC, 37-46, 2002.
[30] L. Li, N. Vijaykrishman, M. Kandemir, M.J. Irwin, "A Crosstalk Aware Interconnect with Variable Cycle Transmission," DATE, 102-107, 2004.
[31] R. Dobkin, R. Ginosar, C. P. Sotiriou, "High Rate Data Synchronization in GALS SoCs," TVLSI, 14(10):1063-1074, 2006.
[32] M.T. Dean, T. Williams et al. "Efficient Self-Timing with Level-Encoded 2-Phase Dual-Rail (LEDR)," ARVLSI, 55-70, 1991.
[33] D.H. Linder, J.C. Harden, "Phased Logic: Supporting the Synchronous Design Paradigm with Delay-Insensitive Circuitry," IEEE Trans. Computers 45(9):1031-1044, 1996.
[34] DS-DE, IEEE1355-1955, http://grouper.ieee.org/groups/1355/index.html.
[35] D.Pham et al, "The Design and Implementation of a First-Generation CELL Processor," ISSCC, 184-592, 2005.
[36] International Technology Roadmap for Semiconductors (ITRS), 2003.
[37] R. Ginosar, "Fourteen Ways to Fool Your Synchronizer," ASYNC, 89-96, 2003.
[38] R. Dobkin, R. Ginosar, "Zero phase latency synchronizers using four and two phase protocols," TR, EE Dept, Technion, 2007, www.ee.technion.ac.il/~ran/papers/zerolatency.pdf.
[39] L. Scheffer, "An Overview of On-chip Interconnect Variation," SLIP, 27-28, 2006.
[40] R. O. Topaloglu, A. B. Kahng, "Generation of Design Guarantees for Interconnect Matching," SLIP, 29-34, 2006.
[41] H. Chang, H. Qian, S. S. Sapatnekar, "The Certainty of Uncertainty: Randomness in Nanometer Design," PATMOS, 36-47, 2004.
[42] K.A. Bowman, S.G. Duvall, J.D. Meindl, "Impact of Die-to-Die and Within-Die Parameter Fluctuations on the Maximum Clock Frequency Distribution for Gigascale Integration," JSSC, 37(2):183-189, 2002.
[43] H.B. Bakoglu, "Circuits, Interconnections and Packaging for VLSI", Adison-Wesley, 194-219, 1990.
[44] W. C. Elmore, "The transient response of damped linear networks with particular regard to wideband amplifiers," J. Applied Physics, 19(1), 1948.
[45] M. Moreinis, A. Morgenshtein, I. Wagner, A. Kolodny, "Logic gates as Repeaters (LGR) for Timing Optimization," TVLSI, 14(11):1276-1281, 2006.
[46] Predictive Technology Model (PTM), http://www.eas.asu.edu/~ptm.
[47] C. Svensson, "Electrical Interconnects Revitalized," TVLSI, 10(6):777-788, 2002.
[48] Y.I. Ismail, E.G. Friedman, "Repeater Insertion in RLC Lines for Minimum Propagation Delay," IEEE Int. Symp. Circuits and Systems, 404-407, 1999.
[49] E.G. Friedman, G. Chen, "Low-Power Repeaters Driving RC and RLC Interconnects With Delay and Bandwidth Constraints," IEEE Trans. VLSI, 14(2), pp. 161-172, 2006.
[50] Y. Cao, C. Hu, A.B. Kahng, S. Muddu, D. Stroobandt, and D. Sylvester, "Effects of global interconnect optimizations on performance estimation of deep submicron designs," Proc. Int. Conference on Computer-Aided Design, pp. 56-61, 2000.
[51] Y. Cao, C. Hu, A.B. Kahng, S. Muddu, D. Stroobandt, D. Sylvester, "Effects of Global Interconnect Optimizations on Performance Estimation of Deep Submicron Designs," IEEE/ACM CAD, 56-61, 2000.
[52] R. Venkatesan, J.A. Davis, K.A. Bowman, J.D. Meindl, "Minimum Power and Area N-Tier Multilevel Interconnect Architectures Using Optimal Repeater Insertion," ISLPED, 167-172, 2000.