

Repeated On-Chip Interconnect Analysis and Evaluation of Delay, Power, and Bandwidth Metrics under Different Design Goals

Ling Zhang¹, Hongyu Chen², Bo Yao³, Kevin Hamilton⁴, Chung-Kuan Cheng¹

¹University of California, San Diego, CA 92093, ²Synopsys Inc., CA 94043

³Mentorgraphics Corp., CA 95131, ⁴Qualcomm Inc., CA 92121

¹{lizhang, ckcheng}@cs.ucsd.edu, ²hongyu.chen@synopsys.com

³bo_yao@mentor.com, ⁴kevinham@qualcomm.com

Abstract

As semiconductor process technologies shrink, interconnect planning presents ever-greater challenges to designers. In this paper, we analyze, evaluate and compare various metrics with optimized wire configurations in the contexts of different design criteria: delay minimization, delay-power minimization and delay²-power minimization. We show how various design criteria influence interconnect performance and we have several observations: (1) the optimal inverter to wire capacitance ratio depends only on the technology and design goal, not on wire pitch, (2) at min-pitch, the width pitch ratios of wire for different objective functions are different: the ratio is 0.52 for minimizing delay, 0.31 for minimizing delay²-power product and 0.21 for minimizing delay-power product, (3) we derive the quantitative delay-energy trade-offs for the three objective functions: the delay-power product reduces power by 67% with a cost of 40% larger delay, while the delay²-power product reduces power by 50% with a cost of 10% larger delay, which implies that delay²-power product results a decent power saving with little cost on speed and (4) We derive the quantitative results of the impact of wire pitch on wire performance. Particularly at 70nm technology node, for bandwidth, the optimal pitch is at min-pitch, while for power, the optimal pitch is 2.35x the min-pitch, and for bandwidth over power, the optimal pitch is 1.76x min-pitch.

1. Introduction

Interconnect strategy, or interconnect planning has become a critical part of chip design. One well-known reason for this comes from the growing significance of wire delay relative to gate delay in the total delay equation. [4] noted that RC delay is dominated by the global interconnect component and that the benefits of new materials alone are insufficient to meet overall long-term performance requirements. Another important factor is the increasing relative power consumption of wires versus gates. In [8], the authors found that interconnect power alone accounted for half the total dynamic power of a 0.13 μ m micro-processor that was designed for power efficiency. As a result, interconnect power consumption has the potential to be a limiting factor in the realization of Moore's law.

A great challenge has been posed: how to design on-chip wires to meet increasing requirements for communication speed within specific power and area constraints. Facing such a multi-objective optimization problem, the determination of a reason-

able objective function is of primary importance.

Under the objective function of minimizing delay, the performance of repeated wire has been well studied. The closed form expressions of optimal inverter size and inverter interval are derived in [7] and [9] based on different transistor models. Meanwhile, there are many previous optimization works on wire sizing, wire spacing in the context of minimizing delay. We also include the analysis under the objective function of minimizing delay in our work.

For the energy-delay optimization, much work [3],[14], [16] has been done from the gate level to the architecture level optimization. [3],[14] are focusing on finding the energy-delay tradeoffs on devices via gate sizing, supply voltage and threshold voltage optimization, while [16] concentrates on evaluating the energy-delay tradeoffs both in circuit and architectural level by defining hardware intensity. So far from our knowledge, no quantitative result of interconnect optimization for minimizing delay-power has been shown in published works. In this paper, we showed the energy-delay tradeoffs on wire and its impact on all performance metrics.

Our work focuses on repeated local wires, which has been widely used in practice, and much recent research [2], [6],[9] is focusing on it. In [5], the authors found that local circuit connections comprise a dominant majority (90%) of on-chip wiring, and according to [8], the power dissipated by local wires is over 60% of total interconnect power.

With multiple layers of metal and multiple levels of design hierarchy, minimum wiring pitch is not always the best choice for every routing problem [12], [11]. In our work, the repeated on-chip interconnect configuration is revamped for multi-objective optimization. Our goal is to develop and present methods and guidelines to aid the designer in choosing the best interconnect strategy.

Our main contributions are as follows:

(1) We formulate various metrics to measure the quality of wire types and configurations.

(2) Although many works have been done on wire optimization for minimizing delay, we summarize the analytical expressions for comparison and completeness. We also obtain the closed form long-term trends of technology shrinkage according to the proposed metrics.

(3) We apply numerical experiments to verify all our analytical results, and demonstrate the optimal value of wire config-

urations, performance metrics and their relations for minimizing delay-power and delay²-power product. We consider short circuit current and leakage current in our buffer model, and adopt an accurate wire capacitance model, Elmore delay model and ITRS technology parameters.

We have several observations:

(1) The optimal inverter to wire capacitance ratio depends only on the design goal and technology, and remains constant with changing wire pitch. In [7], the author has pointed out that optimal wire configuration for minimizing delay requires a constant capacitance ratio of wire to the inverter. We expand this conclusion to minimizing delay-power and delay²-power product.

(2) Some previous works has found that to save power, thinner wire is preferred. In our work, we derived the closed form optimal width pitch ratio for minimizing delay, and found the optimal width pitch ratio for the other two design goals by numerical experiments. At min-pitch, the ratio is 0.52 for minimizing delay, 0.31 for minimizing delay²-power product and 0.21 for minimizing delay-power product.

(3) We derive the quantitative delay-energy trade-offs for the three objective functions: the delay-power product reduces power by 67% with a cost of 40% larger delay, while the delay²-power product reduces power by 50% with a cost of 10% larger delay, which implies that delay²-power product results a decent power saving with little cost on speed.

(4) We derive the quantitative results of the impact of wire pitch on wire performance. Particularly at 70nm technology node, for bandwidth, the optimal pitch is at min-pitch, while for power, the optimal pitch is 2.35x the min-pitch, and for bandwidth over power, the optimal pitch is 1.76x min-pitch.

2. Glossary

To clarify further expression, we define symbols and expressions used in this paper. Most of notations follow [7].

2.1. Variables

These variables can be controlled in physical layout and we want to understand their effect on wire performance.

- w is the width of a wire used for interconnect.
- $pitch$ is the wire pitch.
- s_{inv} is the scaled size of an inverter.
- l_{inv} is the distance between placed inverter instances.
- l is the total length of a wire.

2.2. Parameters

These two interconnect parameters play critical role in wire performance.

- r_w is the wire resistance per unit length.
- c_w is the wire capacitance per unit length.

2.3. Functions

These functions produce the values with which we gauge performance.

- d_{total} is the delay of a repeated wire.
- d_{stage} is the delay of a wire segment between two adjacent inverters.
- $p_{stage} = p_{wire} + p_{inv}$ which shows the sum of a wire segment power and the power of its driver is the power of a single stage.
- $delay_n$ is the wire-length-normalized delay.
- $power_n$ is the wire-length-normalized power.
- $bandwidth = 1/(delay_n \times pitch)$ is the amount of data that can be transferred per unit area per unit time.
- $bandwidth/power = 1/(delay_n \times pitch \times power_n)$ is the bandwidth for a given power budget.

Table 1. Technology data from ITRS[4]

year	1999	2001	2003	2006
Technology node(nm)	180	130	100	70
Metal 1 pitch(nm)	450	350	240	170
Metal 1 aspect ratio	1.8	1.6	1.7	1.7
Conductor effective resistivity (Ω -cm)	2.2	2.2	2.2	2.2
Voltage supply(volt)	1.8	1.5	1.2	1.1
Interlevel metal insulator dielectric constant	3.1	3.1	3.1	3.1
Leakage current(25°C)(uA/um)	0.001	0.01	0.03	0.05

Table 2. Hspice simulation result of output resistance of min sized CMOS inverter

Technode(nm)	180	130	100	70
r_0 (k Ω)	7.07	8.32	8.38	9.43

2.4. Technology data

We list technology related data besides those already listed in Table [4]. We performed Hspice simulation to obtain the output resistance of inverters, and leakage currents. (More details are in Section 3.2.1.)

- $w_0 = 2 \times \text{technode}$ is the min-sized NMOS gate width.
- $g = 1.34$ is the P/N ratio of transistor width.
- $f = 1$ is the ratio of diffusion capacitance to gate capacitance of a transistor.
- $c_g = 1.75 f F/um$ is the value of NMOS gate capacitance per micron [7].
- $c_{mos} = c_g w_0$ is the min-sized NMOS gate capacitance.
- r_0 is the output resistance of a min-sized inverter.
- ρ is the resistivity of copper.
- ϵ is the interlevel metal insulator dielectric constant.
- t is the wire thickness.
- h is the interlevel metal insulator thickness. We assume $h = t$.
- s is wiring spacing, which satisfies $s = pitch - w$.
- v_{dd} is the supply voltage.
- a and b are constants related to the transistor switching model. When switching occurs at half the voltage swing, $a = 0.4$ and $b = 0.7$, and those are the values we use.
- I_{leak} is the leakage current, which relies on technology, gate size and working temperature.
- η_{leak} is the ratio of leakage power and dynamic power. It is technology depended. (Details are in Section 3.2.1.)
- s_w is the switching factor of signal wires, we assume $s_w = 0.15$ [15].

3. Evaluation procedure and models

3.1. Evaluation procedure

Our evaluation process is as follows.

For a given objective function and process technology, a reasonable range of wire pitch is selected. We sweep the pitch in very small step (10nm) from the lower to upper bound. At each step, we find the optimal w , s_{inv} and l_{inv} by numerical search so that the given objective function is minimized. With this optimal wire configuration, metrics (defined below) at that pitch value are evaluated.

There are four design metrics that are our primary concerns: $delay_n$, $power_n$, $bandwidth$, $bandwidth/power$

The meanings of $delay_n$ and $power_n$ are straightforward, and the $bandwidth$ measures the speed and capacity of wire communication while $bandwidth/power$ shows the communication efficiency in terms of power. Note that for a given pitch,

Table 3. The value of η_{leak}

Technode(nm)	180	130	100	70
η_{leak}	1.4%	12%	35%	45%

with the definition of $bandwidth = 1/(delay_n \times pitch)$, the maximization of $bandwidth$ and $bandwidth/power$ is equivalent to the minimization of the objective functions $delay_n$ and $delay_n \times power_n$.

We define three objective functions as follows:

$$delay_n, delay_n \times power_n, delay_n^2 \times power_n$$

We explore performance in the context of each objective function for a range of wire configurations, and note how the design goal influences the perception of the "best" configuration.

We define procedures of minimizing $delay_n$, $delay_n \times power_n$ and $delay_n^2 \times power_n$ as min-d, min-dp and min-ddp. For min-d, we summarize the closed form of optimal s_{inv} and l_{inv} (shown in section 4).

3.2. Assumptions and models

To simplify our study, we make several assumptions:

- Driver and receiver inverters are of the same size. Inverters are inserted at equal intervals into wires.
- There is no scattering effect for wiring resistivity.
- There is no minimal wiring width constraint when searching for optimal w , s_{inv} and l_{inv} .

3.2.1 Elmore delay and power model

We use the Elmore model from [1] to derive the wire delay and power per bit, which are also derived by [7]. The wire delay can be written as:

$$d_{total} = \frac{l}{l_{inv}} d_{stage} = l \left(\frac{b(1+g)(1+f)r_0 c_{nmos}}{l_{inv}} + ar_w c_w l_{inv} + \frac{br_0 c_w}{s_{inv}} + b(1+g)r_w c_{nmos} s_{inv} \right) \quad (1)$$

The power for every wire segment and its driver can be written as $p_{stage} = p_{wire} + p_{inv}$, in which the power consumed by wire is $p_{wire} = c_w l_{inv} v_{dd}^2$, and the power consumed by driver includes dynamic power, short circuit power and leakage power. Dynamic power can be estimated by $p_{dyn} = (1+f)(1+g)c_{nmos} s_{inv} v_{dd}^2$. The work in [10] shows that the short circuit power is roughly 10% of the dynamic power regardless of technology scaling. Considering the exponential relations between the leakage current and temperature, we choose the temperature of 100°C to explore the importance of leakage effect. To estimate the leakage power, we performed Hspice simulations under 180nm technology to obtain the leakage current of a NMOS transistor with zero gate voltage, and supply voltage at drain. We find that at 100°C, leakage current becomes 30 times of the value at 25°C [4], and we assume this relation holds for all technologies we concern. By having that relation and clock rate from [1], we can define:

$$\eta_{leak} = p_{leak} / (s_w p_{dyn}) = I_{leak} / [s_w (1+f)(1+g)c_g v_{dd} f_{clock}]$$

and the computation results are listed in Table 3.

In summary, the total power for a single stage is expressed as

$$p_{stage} = (c_w l_{inv} + (1.1 + \eta_{leak})(1+f)(1+g)c_{nmos} s_{inv}) v_{dd}^2 \quad (2)$$

3.2.2 Models of wire capacitance

To determine the wire capacitance, c_w , we assume that there are power and ground planes above and below RC wires. We use the equations (12-15) from [13] for the numerical experiments, which have less than 5% error compared with results of RC extraction. For analytical derivation we apply a much simpler form:

$$c_s = \epsilon w / h, c_c = \epsilon t / (p - w) \quad (3)$$

In (3), c_s is ground capacitance, and c_c is coupling capacitance. Our experiments show that around min-pitch, 3 gives a result of around 30% less than that in [13], and c_c is around 3 times of c_s . The simpler form enables us to derive analytical estimations on wire performance at min-pitch area with tolerable error, as shown in Sections 4 and 5.

We consider the average coupling effect of a wire sandwiched by other two, which means adjacent wires don't switch to worsen or improve coupling. Then summing up the capacitance on both sides, we have the wiring capacitance:

$$c_w = 2c_c + 2c_s \quad (4)$$

Equation 5 determines the wire resistance r_w .

$$r_w = \rho / (wt) \quad (5)$$

3.2.3 Wire length normalized delay and power model

Instead of being interested in total delay in equation 1, we are interested in normalized delay, defined as $delay_n$ in equation (7), since it is independent from the wire length.

$$delay_n = d_{stage} / l_{inv} = b(1+g)(1+f)r_0 c_{nmos} / l_{inv} + ar_w c_w l_{inv} + br_0 c_w / s_{inv} + b(1+g)r_w c_{nmos} s_{inv} \quad (6)$$

Also, we normalize power of each stage by the inverter interval:

$$power_n = \frac{p_{stage}}{l_{inv}} = (c_w + \frac{(1.1 + \eta_{leak})(1+f)(1+g)c_{nmos} s_{inv}}{l_{inv}}) v_{dd}^2 \quad (7)$$

It can be seen that now both the delay and power metrics are independent of wire length. They only rely on physical parameters l_{inv} , s_{inv} , r_w , and c_w .

4. Analysis of min-d procedure

In this section, the well known expressions ([1][8]) of optimal wire configuration for min-d are given, and the trends of metrics over technology are summarized in Table 4. for later comparison.

To perform minimization, we take the derivative of $delay_n$ with respect to s_{inv} and l_{inv} , and let them equals to zero. We have the following results, which are first derived by [1]:

$$l_{inv_mind} = \sqrt{b(1+g)(1+f)r_0 c_{nmos} / (ar_w c_w)} \quad (8)$$

$$s_{inv_mind} = \sqrt{r_0 c_w / [(1+g)r_w c_{nmos}]} \quad (9)$$

Hence the minimal delay and power are derived by plugging in (8,9) back to definitions (6,7):

$$delay_{n_mind} = 2(\sqrt{a(1+f)} + \sqrt{b}) \sqrt{(1+g)br_0 c_{nmos} r_w c_w} \quad (10)$$

$$power_{n_mind} = (1 + (1.1 + \eta_{leak} \sqrt{(1+f)a/b})) c_w v_{dd}^2 \quad (11)$$

To obtain optimal w , we employ (10,11,3), and take the derivative of $delay_{n_mind}$ with respect to w , and let it equal to zero. We then have

$$w_{mind} = 0.5pitch \quad (12)$$

As technology scales down, if assumed that r_0 is roughly constant and apply (8-11), we can summarize the trends for variables and metrics over feature size for min-d in Table 4.

Substituting l_{inv_mind} and s_{inv_mind} with formula (8,9), we also derive the ratio of inverter capacitance to wire capacitance for min-d:

$$\begin{aligned} c_{gate_mind} / c_{wire_mind} &= (1+f)(1+g) \frac{c_{nmos} s_{inv_mind}}{c_w l_{inv_mind}} \\ &= \sqrt{a(1+f)/b} \end{aligned} \quad (13)$$

Which means $c_{gate_mind} / c_{wire_mind}$ is independent of technology and wire configuration. For min-dp and min-ddp, power is related with technology via η_{leak} , hence, this ratio varies with technology scaling. These trends are verified in Section 5.

Table 4. Relations of variables and metrics with technode around min-pitch for min-d

variables	trend	variables	trend
$w/pitch$	0.5	c_w	constant
c_{nmos}	w_0	l_{inv_mind}	$\sqrt{w_0^3}$
r_w	$1/w_0^2$	s_{inv_mind}	$\sqrt{w_0}$
metrics	trend	metrics	trend
$power_n$	$\eta_{leak} v_{dd}^2$	$delay_n$	$1/\sqrt{w_0}$
$bandwidth$	$1/\sqrt{w_0}$	$bandwidth/power$	$1/(\eta_{leak} v_{dd}^2 \sqrt{w_0})$

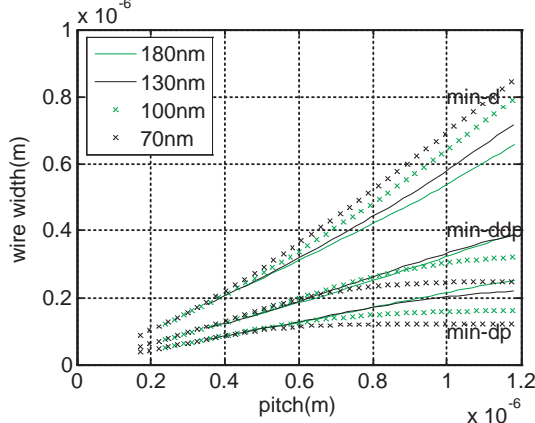


Figure 1. Optimal widths

5. Experimental Results

All experiments were performed using Matlab 7.0. We use Matlab function *fminsearch* to find the minimum value of given function. Technology nodes include 70nm, 100nm, 130nm, and 180nm. The pitch ranged from the minimum Metal 1 pitch of each technology to 1.2um. This range exceeds the practical range used for local wires in each technology.

Optimal wire configurations are discussed in Subsection 5.1, and metrics are evaluated in Subsection 5.2. We discuss the results roughly from three perspectives: (1) spreading from 180nm to 70nm technology nodes, values at minimum pitches are compared, (2) values at the same technology node with a range of pitches are compared (In both (1) and (2), we compare the analytical derivation for min-d procedure to the numerical results.) (3) values at 70nm technology are shown.

Some related results have been derived and scattered in many previous works, our contributions are summarizing and comparing them in a systematic way so that the relations between technology, design goals, wire configurations and performance can be explored thoroughly.

5.1. Optimized wire configuration

The optimal wire widths, inverter sizes and inverter distances under different objective functions and technology nodes are shown in Fig.1 to Fig.3.

We have several observations regarding the optimized wire configurations:

Wire width:

At minimum pitch, the $w/pitch$ ratio is 0.52 for min-d, which matches our prediction in section 4, and 0.31 for min-ddp and 0.21 for min-dp (Fig.1). It is reasonable that power consideration favors narrower wire.

Inverter size:

(1) Spreading through four different technology nodes for min-d at min-pitch, the trend of inverter size can be observed (Fig.2).

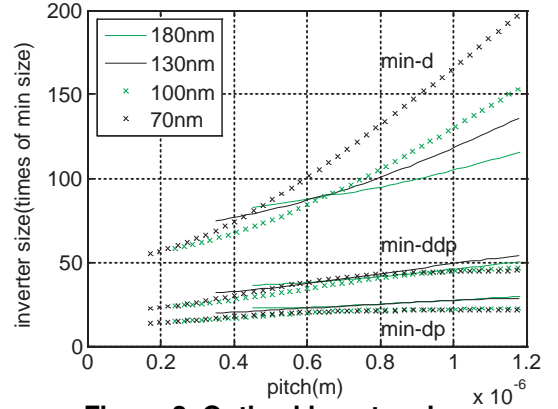


Figure 2. Optimal inverter sizes

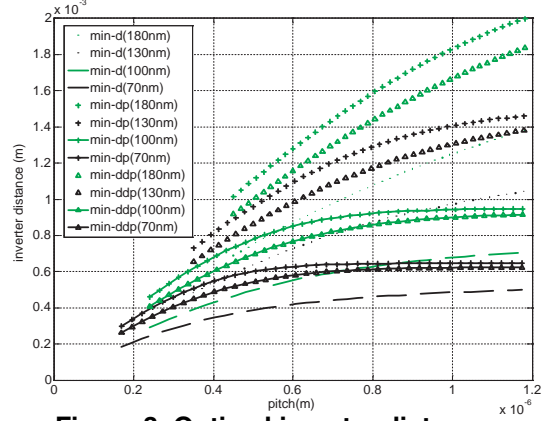


Figure 3. Optimal inverter distances

(2) For the same technology node with a range of pitches, we can derive analytical expression for min-d procedure by using (3,5,9,12) and the assumption $h = t$. Noting that r_0 and c_{nmos} are constant under the same technology, the result is $s_{inv_mind} \propto \sqrt{r_0 c_w / (r_w c_{nmos})} \sqrt{t^2 + 0.25p^2}$. This explains why the inverter size increases linearly near the min-pitch range in Fig.2 (The same approach is used in other variables and metrics evaluations.)

(3) For min-ddp and min-dp procedures, comparing values for different technology nodes at minimum pitch vs. values at the same technology node with a range of pitches, we can also observe similar trends but they are much less drastic.

(4) For 70nm technology node and at min-pitch, the inverter size of min-ddp is 41% of that of min-d, and the inverter size of min-dp is 25% of that of min-d.

Inverter distance:

(1) Spreading through four different technology nodes for min-d at min-pitch range, the trend of $w_0^{3/2}$ can be seen (Fig.3).

(2) For the same technology node with a range of pitches, inverter distance increases (Fig.3). The analytical expression can be derived from (3,5,8,12):

$$l_{inv_mind} \propto \sqrt{c_{nmos} r_0 / (r_w c_w)} \propto \sqrt{1 / (4/p^2 + 1/t^2)} \quad (14)$$

(3) For min-ddp and min-dp procedures, comparing values for different technology nodes at minimum pitch vs. values at the same technology node with a range of pitches, we can also observe similar trends.

(4) On average, the inverter distance of min-dp is 140% of that of min-d, and min-ddp is 130% of that of min-d.

Capacitance ratio c_{gate}/c_{wire} :

In our experiments, we found that for min-d, c_{gate}/c_{wire} is a constant: $c_{gate_mind}/c_{wire_mind} = 1.07$, which matches the

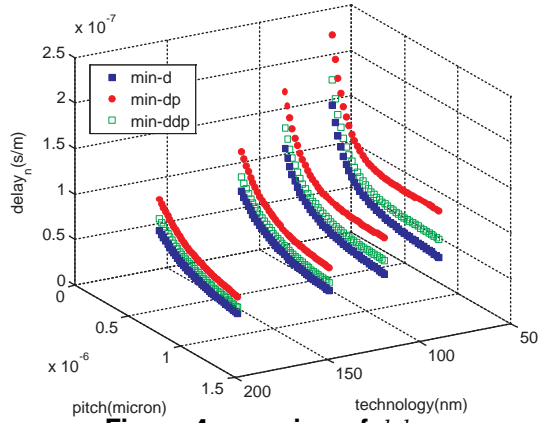


Figure 4. overview of $delay_n$

result in [7]. This means the optimum wire configuration results in a particular c_{gate}/c_{wire} value, regardless of the wire pitch and technology. In other words, the proper c_{gate}/c_{wire} value is essential to optimize min-d wire configuration of repeated RC wire. For min-dp and min-ddp, the c_{gate}/c_{wire} value depends on technology, but not on wire pitch. $c_{gate_mindp}/c_{wire_mindp}$ is $0.45\sim 0.50$, and $c_{gate_mindp}/c_{wire_mindp}$ is $0.27\sim 0.33$.

In summary, around min-pitch, if we formulate the cost of inverter as s_{inv}/l_{inv} , then according to the above comparison, the cost of min-dp is 15%~18% of min-d, and the cost of min-ddp is 29%~33% of min-d. Increasing pitch induces a linear increase in wire width, inverter size and inverter distance in min-d, but has much less effect on min-dp and min-ddp.

5.2. Metric evaluation

In this section, 4 3-D figures show the trends of delay, power, bandwidth, and bandwidth over power with respect to wire pitch and technology under different design goals.

Normalized delay evaluation:

(1) Spreading through 180nm to 70nm technology nodes at min-pitch, delay increases proportional to $1/w_0^{0.5}$ (Table 4, Fig.4). The min-d delay increases by around 100% from 180nm to 70nm technology node.

(2) For the same technology node with a range of pitches, the trend can be derived by adopting (3,5,10,12): $delay_{n_mind} \propto \sqrt{c_{nmos}r_w c_w} \propto \sqrt{1/(4p^2) + 1/t^2}$ which makes $delay_n$ decrease (Fig.4). As pitch further increases, delay becomes insensitive to pitch. Such a change of trend leads to the definition of a saturating pitch: the pitch at which the decrease rate of delay becomes smaller than a threshold rate. Increasing pitch can not effectively reduce $delay_n$ when the pitch is larger than the saturating pitch. Here, we pick the threshold rate as $0.03s/m^2$ which means $delay_n$ decreases 30ps/mm when pitch increases 1 micron.

(3) At saturating pitch, the delay of min-dp is about 140% of that of min-d, and the delay of min-ddp is about 115% of that of min-d.

(4) Saturating pitch scales with technology.

(5) In the 70nm technology node, saturating pitch is $0.6\sim 0.7\mu m$, and the delay of min-d is about 68ps/mm, which is about 1/20 the speed of light.

Normalized power evaluation:

(1) Spreading through 180nm to 70nm technology nodes at min-pitch, power decreases (Fig.5) due to lower supply voltage. Smaller technology can bring more than 60% power reduction

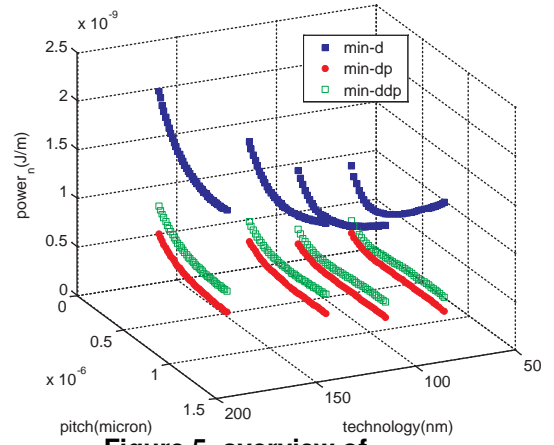


Figure 5. overview of $power_n$

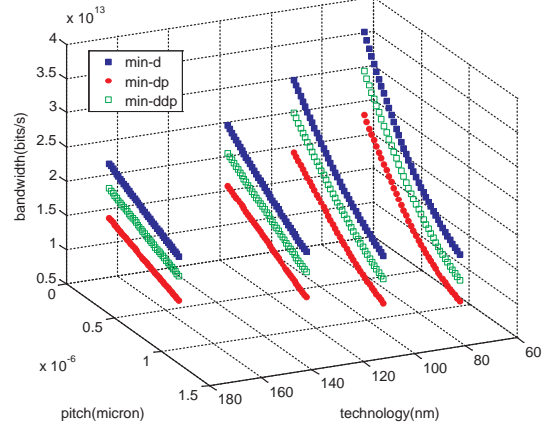


Figure 6. overview of $bandwidth$

at min-pitch (Fig.5), since v_{dd}^2 trend (Table 1) gives power the reduction of 63% for min-d.

(2) For the same technology node with a range of pitches, optimal pitch for min-d can be derived analytically using (3,5,7,12): $power_{n_mind} \propto \eta_{leak} c_w v_{dd}^2 \propto (2t/p + 0.5p/t) \eta_{leak} v_{dd}^2$. Power decreases since coupling capacitance c_c dominates ground capacitance c_s near min-pitch. At larger pitch area, with wire width increasing relatively faster, larger c_w enables wire to consume more power.

(3) For min-ddp and min-dp in the higher pitch range, the trend is almost constant (Fig.5). The reason is that neither wire capacitance nor inverter capacitance change much for these objective functions (Fig.1-3), which implies wire spacing is not quite effective for power saving.

(4) At optimal pitch, compared with min-d, min-ddp reduces power by 47%~60%, and min-dp by 67%. Further, optimal pitch scales with technology. The power at minimum pitch is around $1.3x\sim 1.5x$ of that at optimal pitch.

(5) At the 70nm technology node, the optimal pitch is around $0.4\mu m$, and the power is about $0.75pJ/mm$ for min-d, $0.23pJ/mm$ for min-dp, and $0.33pJ/mm$ for min-ddp at optimal pitch. At min-pitch, the power is about $1.3pJ/mm$ for min-d, $0.33pJ/mm$ for min-d, and $0.45pJ/mm$ for min-ddp.

Bandwidth evaluation:

(1) Spreading through 180nm to 70nm technology nodes at min-pitch, the trend of bandwidth follows $1/w_0^{0.5}$ (Table 4) which is verified in Fig.6.

(2) For the same technology node with a range of pitches, the trend is clear: $bandwidth$ almost inversely proportional to pitch because of the definition of $bandwidth$ is proportional to

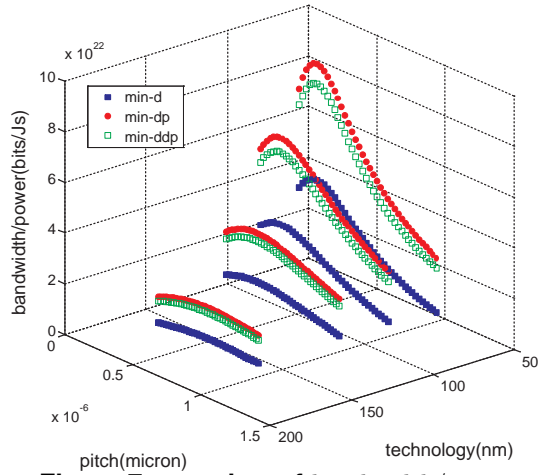


Figure 7. overview of $bandwidth/power$

$1/pitch$.

(3) The min-d procedure results in the best $bandwidth$ (Fig.6). The reduction of $bandwidth$ for min-ddp is 15%, and for min-dp is 33%.

(4) Min-d enjoys the highest $bandwidth$ of 37.5bits/ps at the minimum pitch for the 70nm technology.

Bandwidth over power evaluation:

(1) Spreading through 180nm to 70nm technology nodes at min-pitch, Fig.7 verifies the $1/(v_{dd}^2 w_0^{0.5})$ trend of $bandwidth/power$. Since both v_{dd} and w_0 shrink, smaller technologies show increasing sensitivity of $bandwidth/power$ to pitch, and the largest values for this metric.

(2) For the same technology node with a range of pitches, the trend can be inferred easily from the optimal pitch of $power_n$ and the decrease of $bandwidth$.

(3) Optimal pitch scales with technology, and min-dp has the greatest value for this metric (Fig.7).

(4) At the optimal pitch of 0.3um for 70nm technology, min-dp's $bandwidth/power$ is 0.088bits*m/(pspJ), which is 113% larger than min-d, and is around 9.4% larger than min-ddp. At min-pitch, the min-dp's $bandwidth/power$ is 0.075bits*m/(pspJ), which is 108% larger than min-d and 8.7% larger than min-ddp. On average, $bandwidth/power$ at optimal pitch is around 1.4x~1.7x of that at min-pitch.

6. Conclusions

In this paper, we studied the optimized wiring strategies for three objective functions, and evaluated their effects on four design metrics. In addition to numerical experiments, we summarized the analytical explanations for the min-d procedure, and the analytical results match well with the numerical experiments. Our observations are as follows:

(1) The inverter to wire capacitance ratio depends only on the objective function and technology, and remains constant when wire pitch changes.

(2) At min-pitch, the width pitch ratios of wire for different objective functions are different: the ratio is 0.52 for minimizing delay, 0.31 for minimizing $delay^2$ -power product and 0.21 for minimizing delay-power product.

(3) Among the commonly used objective functions studied, Min-ddp shows a better trade-off between delay and power compared with min-d. It reduces $power_n$ by 50%, increases $bandwidth/power$ by 60~100%, while the cost is 10% increase

in $delay_n$ and 15% reduction in $bandwidth$. In contrast, min-dp reduces $power_n$ by 67% and increases $bandwidth/power$ by more than 100%, but the costs on $delay_n$ and $bandwidth$ are over 40%.

(4) Each metric has its own optimal pitch region, and the region scales down with technology. At 70nm technology node, for $bandwidth$, the optimal pitch is at min-pitch, while for $power_n$, it is 2.35x min-pitch (0.4um), for $bandwidth/power$, it is 1.76x min-pitch (0.3um), and for $delay_n$, it is larger than 0.6um.

Repeated RC wire is still used most widely in current chip design as local interconnect. Analysis and numerical evaluation in this work give favorable pitch values for different metrics and depict how different design goals choose different trade-offs between delay vs. power: to choose different c_{gate}/c_{wire} . The $delay^2$ -power product acquires much power saving with relative low cost in wire speed.

References

- [1] V. Agarwal, M. S.Hrishikesh, S. W. Keckler, and D. Burger. Clock rate versus ipc: The end of the road for conventional microarchitectures. 2000.
- [2] K. Banerjee and A. Mehrotra. Accurate analysis of on-chip inductance effects and implications for optimal repeater insertion and technology scaling. 2001.
- [3] R. Brodersen, M. A. Horowitz, D. Markovic, B. Nikolic, and V. Stojanovic. Methods for true power minimization. In *Int. Conf. Computer-Aided Design Dig. Tech. Papers*, 2002.
- [4] I. R. Committee. International Technology Roadmap for Semiconductors.
- [5] A. Deutsch, P. Coteus, G. Kopcsay, H. Smith, C. Surovic, B. Krauter, D. Edelstein, and P. Restle. On-chip wiring design challenges for gigahertz operation. *IEEE Proceedings*, pages 529–555, April 2001.
- [6] Y. Ismail, E.Friedman, and J. Neves. Effects of inductance on the propagation delay and repeater insertion in vlsi circuits. *IEEE Trans on VLSI System*, 8:195–206, 2000.
- [7] P. Kapur, G. Chandra, and K. C. Saraswat. Power estimation in global interconnects and its reduction using a novel repeater optimization methodology. In *DAC*, 2002.
- [8] N. Magen, A. Kolodny, U. Weiser, and N. Shamir. Interconnect power dissipation in a microprocessor. In *SLIP*, 2004.
- [9] A. Nalamalpu and W. Bursleson. Repeater insertion in deep sub-micron cmos: ramp-based analytical model and placement sensitivity analysis. *IEEE International Symposium on Circuits and Systems*, pages 766–769, 2000.
- [10] K. Nose and T. Sakurai. Analysis of future trend of short-circuit power. *IEEE Trans. Computer-Aided Design*, 19:1023–1030, 2000.
- [11] N. Rohrer, C. Lichtenau, P. Sandon, P. Kartschoke, E. Cohen, M. Canada, T. Pfluger, M. Ringler, R. Hilgendorf, S. Geissler, and J. Zimmerman. A 64-bit microprocessor in 130nm and 90nm technologies with power management features. *IEEE J. of Solid State Circuits*, 2005.
- [12] G. Sai-Halasz. Performance trends in high-end processors. *IEEE Proceedings*, 1995.
- [13] S. Sim, S.Krishnan, D. Petranovic, and N. Arora. A unified rlc model for high-speed on-chip interconnects. *IEEE Trans. Electron Devices*, 2003.
- [14] V. Stojanovic, D. Markovic, B. Nikolic, M. Horowitz, and R. Brodersen. Energy-delay tradeoffs in combinational logic using gate sizing and supply voltage optimization. In *Proc. ESSCIRC*, 2002.
- [15] D. Sylvester and K. Keutzer. Getting to the bottom of deep submicron. In *Proc. ICCAD*, 1998.
- [16] V. Zyuban and P. N. Strenski. Balancing hardware intensity in microprocessor pipelines. *IBM J. Res. and Dev.*, 2003.