Latch Modeling for Statistical Timing Analysis

Sean X. Shi  Anand Ramalingam  Daifeng Wang  David Z. Pan
Department of ECE, University of Texas, Austin TX 78712
{xshi,anandram,wang,dpan}@ece.utexas.edu

Abstract—Latch based circuits are widely adopted in high performance circuits. But there is a lack of accurate latch models for doing timing analysis. In this paper, we propose a new latch delay model in the context of SSTA based on a new perspective of latch timing. The proposed latch model also takes into account the external timing variations such as data slew. The new latch model is integrated into SSTA by considering the timing analysis of both the combinational logic network and the clock distribution network simultaneously. The experimental results show that ignoring accurate latch modeling may lead to large errors (e.g., 50% at PDF peak).

I. INTRODUCTION

Process variations pose the biggest challenge to technology scaling into nanometer regime by being a major performance limiter. Statistical Static Timing Analysis (SSTA) has been proposed to perform full-chip analysis of timing under process variations and has been the subject of intense research recently [1-7].

In SSTA, the gate delays in the cell library are modeled as a first order approximation [4] or second order approximation [5] of process variations. Based on these models, statistical timing analysis and optimization can be applied to the combinational logic [6]. To attain more accuracy, SSTA is done considering the clock distribution network [7]. By these approaches one can predict both the data signal's statistical distribution at the end of each combinational logic chain and the clock distribution at each clock network terminal. However, so far there is no work accurate enough to combine the signal distribution from both networks and predict final signal distribution of the whole system. The major reason is because there are no accurate delay models for the sequential logic such as Flip-flop and latch. Flip-flop and latch are the most commonly used sequential elements whose purpose is synchronizing data signals. These elements will add some delay to timing and thus decrease the system performance.

In this paper, we concentrate on modeling latch accurately. This is because an edge-triggered flip-flop functionally is a back-to-back latch pair and also structurally made up of two latches [8]. Hence flip-flop models can be derived from accurate latch models.

A latch is a three-terminal element, having two inputs, data (D) and clock (clk/C) and one output (Q). The data must be stable tsetup before the falling edge of the clock (called the setup time) and thold after the falling edge of the clock (called the hold time) for the data to be correctly stored in the latch. For timing requirements, level sensitive latches are widely used in high performance ICs where timing analysis is more critical and challenging [9-11]. In the approaches presented in the literature, the latch delay model is deterministic; they ignore the impact of the input data signal and clock signal being statistical quantities. However, when a path is timing critical, the data would arrive very close to the falling edge of clock, and the mean value of \( t_{DC} \) (data-to-clock delay) might be close to the latch’s setup time with very limited or negative slack left leading to the increase in the delay of data D to output Q (tDQ). Moreover, with different slew distributions of data and clock, the tDQ to tDC function will be different. To keep things simple, traditional circuit design and timing analysis [12] have a constant setup time. But this simplification leads to less accurate statistical timing analysis and lesser flexibility in optimization [13].

In this paper, we propose a new latch delay model for statistical timing analysis. Our latch model captures the impact of delay and slew variations of both input data and clock on latch delay. Based on this new latch delay model, one can combine the timing analysis of data signal network with clock distribution network to do SSTA in an accurate way.

The main contributions of this paper include: a) a new latch timing model considering both logic and clock signal variations; b) integrating the proposed latch model into SSTA. Our experimental results show that ignoring latch modeling may lead to large errors (e.g., 50% at PDF peak).

The rest of this paper is organized as follows: in Section II, general timing diagram and structure of transparent latch are reviewed, with traditional latch delay model. A new point of view for latch working mode based on a 3-D analysis is proposed in Section III. Section IV presents our new latch delay model taking into account variations such as data slew, clock slew among others Statistical timing analysis for latch is discussed in Section V, followed by experimental results in Section VI and conclusion is drawn in the last section.

II. LATCH PRELIMINARIES

A. Timing diagram of latch

The timing diagram of latch is shown in Figure 1. Both setup and hold times of a latch are measured relative to the trailing edge of the clock. The data signal must be a constant in the timing window between the setup and hold time. This ensures that the data is sampled and latched correctly. In addition to setup and hold times, two more delay quantities \( t_{CQ} \) and \( t_{DQ} \) need to be defined. This is because of the following two scenarios: 1) Data is stable but the latch is closed due to the clock being low, and 2) Data stabilizing while the latch is open. In critical path analysis, when we assume that the data signals arrive quite close to the setup time while latch is open, \( t_{DQ} \) is the key delay to be analyzed. In this paper, we focus on modeling \( t_{DQ} \) accurately.

![Figure 1. Timing diagram of latch. The situation with the latch is different from flip-flop. Both setup and hold time of latch is measured relative to the trailing edge of the clock. The longest path “a1” must arrive at next latch “L2” before setup time and the shortest path “a2” must reach next latch “L3” after hold time.](image)

B. Structure of transparent latch

One of the most widely used latch structures is shown in Figure 2(a). In the semicustom datapath application, where the noise of the input signal can be well controlled, this latch structure is preferable for it is fast and compact [14]. With an additional inverter before the input data, the latch structure (Figure 2(b)) becomes robust and is widely...
used in standard cell applications [15]. Such a latch is recommended for all but the most performance-critical or area-critical design.

![Figure 2: Latch structures. (a) is one of the most widely used latch structures due to its speed and compactness. This paper focuses on this structure. (b) is widely used in standard cell applications with one additional inverter before the input in structure (a). The additional inverter makes (b) more robust compared to (a) at the cost of area and performance.](image)

In this paper, we focus on modeling the latch structure in Figure 2(a) but our modeling is generic enough to be applied to the latch structure in Figure 2(b) too.

The latch in Figure 2(a) can be decomposed into 3 parts: the transmission gate, output inverter, and the storage part. In next section, we will show that traditional latch modeling focuses on the feedback mechanism of the storage part and models it as two inverters.

C. Traditional timing model of latch

As shown in Figure 3, the traditional way of modeling latch focuses on the storage part of the latch [16], which is modeled as self-feedback system of two inverters as shown in Figure 3(a). Figure 3(b) shows the butterfly curve that results when the transfer function of the two inverters are superimposed. This feedback system has two stable states (point A & B) and one metastable state (point C) as shown in Figure 3(c).

\[ t_{DQ} = \tau_a \left[ \ln \Delta V - \ln a(0) \right], \]

where \( t_{DQ} \) is the delay from input D to output Q, and \( a(0) \) is a small signal offset from the original metastable point. \( \Delta V \) is some predefined constant voltage point to predict D-to-Q \( t_{DQ} \) delay.

![Figure 3: Traditional timing model of latch. (a) the storage part of a latch; (b) butterfly curves of the static transfer characteristics; (c) an analogy of a ball on a hill with one metastable state at the top of the hill and two stable states in the foothills.](image)

An additional assumption is that \( a(0) \) is proportional to \( (t_{DC}-t_m) \), where the input signal is a ramp that passes through the metastable state point at \( t_m \). Thus, the \( t_{DQ} \) delay can be modeled as log-linear function:

\[ t_{DQ} = a - b \cdot \ln(t_{DC} + c). \]

D. Limitation of traditional model

To better understand the traditional model of the latch, several HSPICE simulations were run to get the delays of latch around setup time. We used PTM [17] for 65nm in our simulation and fitted the resulting data using Eq. (2) and the result is shown in Figure 4.

![Figure 4: Limitation of the traditional latch model. Traditional model is only accurate when \( t_{DC} \) delay is much smaller than the setup time. However under statistical timing of critical paths, \( t_{DC} \) delay might be close to or bigger than the setup time.](image)

In Figure 4, the fanout of the latch is four, slew of the clock signal is 40ps and the slew of input data D is 80ps. Black dots are HSPICE simulation results and the red line is the curve fitted based on the traditional delay model in Eq. (2). Blue dash line is the input D-to-C \( t_{DC} \) delay distribution that has positive slack as the mean value of \( t_{DC} \) delay bigger than setup time. The setup time is defined according to \( t_{DC} \) when \( t_{DQ} \) is 10% bigger than its minimum value.

From the figure, we can see that when \( t_{DC} \) delay is around or bigger than setup time, the function Eq. (2) is quite inaccurate. The fitting is good only when \( t_{DC} \) delay is much smaller than setup time. For statistical timing analysis of critical longest paths, as the mean of \( t_{DC} \) delay is close to setup time and high percentage of \( t_{DC} \) delay distribution will be around the setup time of the latch, delay model of latch in Eq. (2) has difficulty to meet accuracy requirement of latches' statistical timing analysis.

Moreover, the model in Eq. (2) does not consider the impact of input data slew, clock slew or fanout. In fact, input data slew, clock slew and fanout, all of them could change the delay curves between \( t_{DQ} \) and \( t_{DC} \).

III. A NEW 3D VIEW OF LATCH TIMING

A. State transform in a latch storage part

If the two inverters in the storage part of the latch are the same and driving strength of the PMOS and NMOS in each inverter are also identical, the potential of the storage part can be drawn as Figure 5.

In Figure 5(a), the 3D potential figure is drawn while X and Y axis are \( V_1 \) and \( V_2 \) respectively. The 2D projection is shown in Figure 5(b). There are 5 special state points:

A: \( (V_1=0, V_2=Vdd) \), stable;
B: \( (V_1=Vdd, V_2=0) \), stable;
C: \( (V_1=V_2=Vdd/2) \), metastable;
D: \( (V_1=V_2=0) \), unstable with highest potential;
E: \( (V_1=V_2=Vdd) \), unstable with highest potential.

\( D' \) and \( E' \) are the \( D \) & \( E \)'s projection on 2D plane.

When the state of the storage part is at point A or B, the state is stable (the system is at its lowest potential at A and B). Point C is the only one metastable state in the system.
Traditional latch model in Eq. (1) only covers the state transfer from one stable state through metastable point and to another stable point, which is the dash dot line A-C-B in Figure 5 (white in 3D part of Figure 5 (a) and black on the 2D projection in both Figure 5 (a),(b)).

![Figure 5](image)

**Figure 5.** A new view of latch state transfer. (a) Potential of various states. (b) Projection onto a two dimensional space. Traditional latch delay function models the state transfer along A-C-B, where A and B are two stable states and C is the only metastable point. However, it is possible that the storage part of latch driven by a transmission gate goes directly from A to a point F (far away from C) and then goes from F to B.

On the projection plane of square A-D-B-E in Figure 5 (b), there are more state points than the points on line A-C-B. The colored solid lines show the equipotential lines. The dash lines show that the state moving tracks if there is no external signal input. For example, if the state is at D $(V_f=V_2=0)$ or E $(V_f=V_2=V_{DD})$, it will directly go to metastable point C along the black line D-C or E-C with red arrows, and then through C go to stable states of A or B. During this process D-C or E-C, if there is any noise, the state transfer will follow the grey dash lines in Figure 5 (b) and go to stable points A or B directly.

From the above analysis, one can infer that the simplification in traditional latch model leads to incorrect modeling of the state transformation process. This also explains why curve fitting (Eq. (2)) has difficulty in fitting the simulation results around setup time.

**IV. THE NEW LATCH MODEL FOR EXTERNAL VARIATIONS**

**A. Difficulty of latch modeling**

As discussed in previous section, the latch state transfer from one stable state to another stable state can be divided into two parts, A-F: close to linear driven by input data signal, and F-B: close to logarithmic which is self-feedback process of storage part in latch.

However, it is very difficult to develop an analytical function for latch modeling. SRAM which has a storage part like a latch has been modeled as dynamic system and an analytical function has been proposed to predict the critical time of noise [18]. However, the input signal's current waveform is quite complicated and can not be modeled as square wave.

Also the inverters in the practical latch are skewed since PMOS and NMOS have different driving strengths. As only some special functions can be solved in dynamic system [19], the above difficulties make the effort to derive an analytical function for latch modeling very hard.

Thus in this paper, instead of deriving an analytical model based on physics we develop a semi-empirical function for latch modeling. The proposed function covers all of the impacts including not only $t_{DC}$ delay but also input data slew, clock slew and fanout.

![Figure 6](image)

**Figure 6.** Voltage curves of each node in latch. $t_{DD}$ delay is made up of 2 parts: 1) from D$_{1/2}$ to F, which is driven by input data signal; 2) from F to Q$_{1/2}$, which is a self-feedback process.
B. Three regions of $t_{DQ} - t_{DC}$

We divide $t_{DQ}(t_{DC})$ into three regions as shown in Figure 7.

![Figure 7. Three regions of latch delay curve: constant region (red line/round dots), linear region (blue line/triangle dots), and exponential decay region (black line/square dots).](image)

1) Constant region (red line/round dots). In this region the latch is absolutely transparent and $t_{DQ}$ delay is a constant. During this process, clock is on, and the latch through $X$ to $Q$ is driven by input data signal.

2) Linear region (blue line/triangle dots). With the decreasing of $t_{DC}$ delay, the transmission gate is open for quite long period, and the input data signal drives the storage part from stable state (such as $A$) to some middle point $F$ which is quite close to another stable state (such as $B$). In this process, the part of $A$ direct to $F$ dominates the $t_{DQ}$ delay.

3) Exponential decay region (black line/square dots). In this region, the process from $F$ to $B$ is dominant in the total $t_{DQ}$ delay.

C. Latch modeling function

The proposed latch model is divided into two parts: when $t_{DC}$ is big enough, $t_{DQ}$ is constant; after $t_{DC}$ gets smaller, the model is made up of two components: linear part and exponential decay part, given by

$$ t_{DQ} = \begin{cases} t_{DQ0} & t_{DC} \geq t_{DQ0} \\ a \cdot \exp(-b \cdot t_{DC}) + c \cdot t_{DC} + d & t_{DC} < t_{DQ0} \end{cases} $$

(3)

where

$$ t_{DQ0} = a \cdot \exp(-b \cdot t_{DQ0}) + c \cdot t_{DQ0} + d. $$

If the variations of data slew, clock slew and fanout are within a small range or large approximation is acceptable during the statistical timing analysis, Eq. (3) can be simplified to an exponential decay function such as:

$$ t_{DQ} = \begin{cases} t_{DQ0} & t_{DC} \geq t_{DQ0} \\ a_1 \cdot \exp(-b_1 \cdot t_{DC}) + d_1 & t_{DC} < t_{DQ0} \end{cases} $$

(4)

Or even,

$$ t_{DQ} = a_2 \cdot \exp(-b_2 \cdot t_{DC}) + d_2. $$

(5)

However, over wide ranges of fanout, clock slew and data slew, our simulation results show that among Eq. (3), Eq. (4) and Eq. (5), only Eq. (3) can fit $t_{DQ}$ over a wide range of input data slew and clock slew very well as coefficient of multiple determination can be maintained always over 0.99. To some approximation, model Eq. (4) or Eq. (5) might be acceptable.

D. Multi-dimensional spline

After the latch delay model is proposed under specific fanout, clock slew and data slew, the fitting parameters in Eq. (3) under specific condition can be extracted and some table can be built up. The delay in the middle of nodes on the table has to be estimated.

In this paper, we have several parameters such as fanout, clock slew, delay slew. The interpolation problem is formulated as follows. Let $f$ denote fanout, $cs$ the clock slew and $ds$ the input data slew. We represent them as a three dimensional vector: $\vec{w} = (f, cs, ds)$. Therefore, the multi-dimensional cubic spline interpolation is considered here. The $t_{DQ}$ delay ($y$) is a function of $\vec{W}$ and $t_{DC}$ delay ($x$), given by:

$$ y = f(\vec{w}, x) = a(\vec{w}) \exp\left[-b(\vec{w}) \cdot x + c(\vec{w}) \cdot x + d(\vec{w})\right], $$

(6)

where coefficients $a, b, c$ and $d$ are all functions of $\vec{W}$.

V. LATCH MODELING IN STATISTICAL TIMING ANALYSIS FRAMEWORK

There are have been several works [9-11] which propose algorithms for statistical static timing analysis (SSTA) of latch based circuits. The accuracy of any proposed algorithm for SSTA can be compared with the Monte-Carlo (MC) simulations of the circuit. However, in these statistical algorithms and MC simulations, the basic latch delay model used was developed under deterministic timing analysis. In existing timing analysis, under certain fanout, both setup time and delay are fixed at its mean value.

And the final Q delay distribution should be:

$$ F_Q(t_Q) = \int_{-\infty}^{t_Q} \frac{1}{\sqrt{2\pi} \sigma_d} \exp\left[-\frac{(t_d - \mu_d)^2}{2\sigma_d^2}\right] \left[1 - \text{erf}\left(-\frac{t_d - \mu_d - \ln(t_f - t_d - d_z)/\sigma_d}{\sqrt{2}}\right)\right] dt_d $$

(11)

$$ p_Q(t_Q) = \frac{dF_Q(t_Q)}{dt_Q}; \quad \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt. $$

Obviously, such a distribution in Eq. (11) is different from the normal distribution in Eq. (7). The experimental results in the following section would show the above difference.

VI. EXPERIMENTAL RESULTS

Over a very wide range (fanout: 1-16; clock slew: 5-100ps, data slew: 5-100ps), our proposed latch delay model Eq. (3) can fit the
HSPICE simulation results with very high accuracy (coefficient of correlation is greater than 0.99). Therefore, in the following discussions and simulations, our proposed model will be regarded as golden model. We use a typical circuit, e.g., benchmark s27 [20] is used for post-latch SSTA. All other circuits have similar results.

A. The impact of clock slew and data slew

As discussed earlier, not only $t_{DC}$ delay but input data slew, clock slew and fanout also impacts the $t_{DQ}$ delay. Figure 8 and Figure 9 show the simulation results of $t_{DQ}$ delay variations caused by above external variations.

![Figure 8. Minimum delay dependency on clock slew and input data slew. Three-dimensional plot is shown in (a): the black square dots are latch’s minimum delays at different clock slews and data slews when fanout is 4; the blue round points are projection on plane of minimum delays and data slews; the red diamond points are projection on the plane of minimum delays and clock slews.](image)

Figure 8 shows that minimum $t_{DQ}$ delays (among different $t_{DC}$ delays) depend on clock slew and data slew. The fanout of the latch is fixed at 4. The black square dots in (a) are latch's minimum delays at different clock slew and data slew; the blue round points are projection on plane of minimum delays and data slews; the red diamond points are projection on the plane of minimum delays and clock slews. From the figure, we can observe that under different clock slews and data slews, the $t_{DQ}$ delays vary over 20ps. As the overall minimum $t_{DQ}$ delay is less than 20ps, such variation range is about 100%.

Red diamond points in Figure 8 (b) are projection of black square points in Figure 8 (a) on the plane of minimum delays and clock slews. From Figure 8 (b), even under the same clock slew, the input data slew can cause about 10ps $t_{DQ}$ delay variations.

Moreover our simulations show that external variations, such as data slew, clock slew, fanout, have big impact on $t_{DQ}$ delay. Hence if these factors are ignored, they lead to inaccurate yields from the statistical timing analysis of a circuit.

B. Statistical timing based on MC simulation

![Figure 9. Delay and slew distribution of a critical path in benchmark s27. This was obtained using Monte-Carlo SSTA [21].](image)

For benchmark s27 [20], after gate sizing, Monte Carlo (MC) simulation of gate length and threshold variations is done on a critical path made up of "AND2 -> INV1 -> NOR2 -> INV -> NAND2 -> NOR2 -> NOR2". The delay and slope results are shown in Figure 9.

The mean of delay is 266.3ps with standard deviation of 24.3ps (9.1% of mean). The mean of slew is 65.4ps while the standard deviation is 41ps (6.3% of mean). The above results were obtained from 10,000 MC simulations. The standard deviation of slew is much smaller than that of delay. One intuitive explanation is that a path delay is a simple addition of gate delays while the output slew gets regenerated at every gate in the path. Thus slew gets corrected at the output of every gate and the variation is reduced as the logic depth increases. An implied result is that the delay and slew might not be highly correlated which was verified from our MC simulations. We found that the correlation between delay and slew was 0.79 for the path in the s27 benchmark mentioned earlier.

In Figure 9, the black lines represent the normal distribution fitting of delay and slew. Compared to slew, delay distribution is closer to a normal distribution. However, as an approximation, it may be acceptable to use normal distribution for timing analysis.

In this part, the MC simulation results is directly sent to latch as external variations on data input terminal. The variations of clock delay and slew are omitted.

Figure 10 shows the simulation results and compares the Q delay distribution difference between our proposed model and traditional model presented in Eq. (7)) [9-11].

![Figure 10. Q delay distribution based on MC simulation results. The red lines are traditional output delay distribution of latch while the black lines are calculated according to our accurate latch model. The variations of clock delay and slew are omitted. (a)-(d) are different in clock frequency and fanout.](image)

As the red lines are calculated from in Eq. (7) without proposed latch delay model, they are marked as "w/o model". The black lines are the results based on proposed model. In this part, we did not use the normal distribution approximation of the data; we used the data delay and slew data from the MC simulations.

The results in Figure 10(a) is when the latch's fanout is set to 2, and the setup time of this latch is 33.4ps and the minimum $t_{DQ}$ delay is 33.6ps. The clock delay is set to 300ps and clock slew is set to 30ps. From Figure 10 (b) to (d) the fanout is set to 4, the setup time of this latch is 26.5ps, minimum $t_{DQ}$ delay is 39.9ps, clock slew is fixed at 60ps, and the clock delays are 300ps, 280ps, and 320ps respectively.

From Figure 10(a),(b), we can see that the PDF and CDF of output Q
delay distributions are quite different. For example, in Figure 10(a) the two PDFs have 20% difference at the peak. In some range, the CDF calculated based on method in previous SSTA papers is quite close to CDF based on our proposed accurate model. However, even within this range, the PDFs of two methods are still quite different from each other. These errors propagate across the gates when one does statistical timing analysis of a circuit. Figure 10 (c), (b) and (d) set the clock to be 280ps, 300ps, and 320ps, respectively. From another point of view, this means the slacks are increasing, and the paths become less timing critical. However for the critical paths, the traditional model becomes less accurate and the proposed latch delay model is necessary.

C. Discussion based on normal distribution approximation

As shown in Figure 9, the data delay and slew distributions are close to normal distributions. So normal distribution approximation is used to see the impact of correlation between delay and slew on latch delay. The original mean and standard deviation of delay and slew are used to approximate the normal distribution. The clock delay is approximated as a normal distribution with mean 300ps and standard deviation 30ps. The clock slew is approximated as a normal distribution with mean 60ps and standard deviation 8ps. The simulation results are shown in Figure 11.

Figure 11. Q delay distributions based on normal distribution approximation. The red lines are traditional output delay distribution of latch while the black lines are calculated according to our accurate latch model. The variations of clock delay and slew are considered. (a)-(c) are different in clock frequency and fanout. (d) compares PDFs of latch output based on models of different accuracy levels.

In Figure 11(a), data delays and slewes are generated independently and no clock variations are considered. In Figure 11(b), there is no clock variation and the correlation between data delay and sleew is set to 0.79 which is the same number obtained from MC simulation results. In Figure 11(c), the clock variations are involved with a correlation of 0.79 between delays and sleews. Finally in Figure 11(d), method in previous latch SSTA papers (black line) and condition in Figure 11(a) to Figure 11(c) (the purple, red and blue line, respectively) based on the proposed model are compared in the PDF curves. We can observe the following from the figures: As the left side and peak of purple line is larger than that of red line, the correlation between data delays and sleews is helpful to reduce latch delays. However, when clock variation is taken into account, the latch delay becomes worse and about 50% error at peak is observed in previous SSTA approaches when compared with our proposed accurate latch delay model.

VII. CONCLUSION

In this paper, we have studied the latch modeling for statistical timing analysis. Based on a new perspective of latch timing an accurate latch delay model is developed which can capture the impact of external variations of delay and slew from input data and clock. The proposed latch delay model is verified by simulations over a wide range of external variations and applied to statistical timing analysis. Compared with existing SSTA works for latch based circuits, our proposed model shows greater accuracy and it is essential to accurate statistical timing analysis of both the combinational logic network and the clock distribution network simultaneously.

ACKNOWLEDGEMENTS

This work is partially supported by NSF, SRC, IBM Faculty Award, Fujitsu, Qualcomm, Sun, Intel equipment donation.

REFERENCES