Abstract

Accurate current analysis is required in circuit designs to analyze electromigration failure rate, power consumption, voltage drop, among additional factors. An accurate time-domain current waveform simulator which is based on the charge-based current model is presented in this paper. The current waveform of a CMOS gate during a transition consists of three components: the capacitor-differential current which occurs when the input changes, the load-capacitance charging/discharging current and the short-circuit current which exist only when the output changes. These three components are characterized by triangular pulses with four parameters \((T_s, T_p, T_{e}, \Delta Q)\). These three time parameters could be reasonably obtained from the voltage waveforms acquired from a timing simulator. The area of the triangle is equal to the charge \(\Delta Q\). The author has calculated the charge associated to each component during a transition, then the total current waveform is obtained by summing these current components. This model has been embedded into the switch-level timing simulator to generate the current waveform. The simulated current waveform helps solve the VLSI reliability problems due to electromigration and excess voltage drops in the power buses. Comparing the results obtained using SPICE with results obtained using this simulator, a good agreement was found, especially on the time points at which current pulses occur. The high accuracy is primary due to accurate time points, which are from accurate timing simulator, and isosceles triangular current pulses, which approach to the real current waveforms.

Key words:

Time domain simulation, CMOS, Charge Based, and Timing.

1. Introduction

In large integrated circuits, such as VLSI/ULSI, current flow in power and ground buses leads to the problems of voltage drop and metal migration, which are the major reliability problems1-5. The problems are especially important in the widely used CMOS technology, where switching transients from different parts of a circuit can occur almost simultaneously and thus, may generate large noise spikes (in the form of voltage drops) in the power/ground buses. Unrestricted voltage drops in the power/ground buses may result in incorrect logic operation and degradation in switching speed6-7. When the voltage drop is big, then the \(V_{DD}\) voltage of the some gates inside the circuit will be far less than the pre-defined value, say 3.5V, then the driving capability of these gates will decrease. It means these gates need more time to charge or discharge the load capacitance, thus degrading the speed. At the same time, the output signal cannot reach the pre-defined \(V_{DD}\). If the voltage drop is too big, then the high state will be far less than the pre-defined \(V_{DD}\). If this signal is used to drive another gate, the gate will never change to low. Pattern-independent techniques8-16 have been used to estimate the current/power consumption while avoiding the problem that a great deal of vectors need simulating. These techniques, however, may not reflect the “real” current pattern since the waveform shape, compared with the power consumption, is much more input-dependent.

Pattern-dependent approaches17-23, on the other hand, have been proposed to provide the real current waveform associating...
with an input vector. Among the existing approaches, the current associated with a high-to-low/low-to-high state transition is approximated by an isosceles triangular current pulse\(^1\). This model can achieve great accuracy when the load capacitance of each gate is small. The current pulse is asymmetric when the load or the fan out is large, therefore this model will result in large errors if the current pulse is represented as an isosceles triangle. Also, due to asymmetry, the peak current obtained by this model will deviate from the real peak current, therefore the peak value of the total current supported by the power bus is not correct. Another method is proposed\(^1\) which uses a set of non-linear exponential function to match a current pulse. Although this method can give accurate results, but the problem is that the pre-characterization is necessary. Asymmetric triangular pulse\(^2\) has been the reasonable approach in recent years.

Regardless of the model used in previous works, simulation or estimation, only the current flowing during the rising and falling edges of the output signals of CMOS gates is considered. The author finds that large errors exist in some large circuits, even the short-circuit current is considered\(^2\). The primary reason is the assumption of neglecting the current that flows across the gate capacitance of CMOS gates. Currents flow across these gate capacitors when the input signals change, whether the gate changes state or not. Therefore, the author develops another current model which takes this current into account. In this model, the researcher decomposes the total current supported by the \(V_{dp}\) bus into three components; one is owing to the parasitic capacitance, another the load capacitance, and the other owing to simultaneous conduction of the P- and N-block of a CMOS gate. The author has calculated the individual charge for each of these three components. Since the current is the time derivative of the charge, the current waveforms of these three components can be obtained. After summing these three components up, the total current is obtained. This current model has been implemented into the timing simulator BTS as a post-processor to obtain the current waveform after the voltage waveforms are obtained.

2. Background and Previous Work

The author has implemented a timing simulator BTS\(^2\), which shows the time-domain voltage waveforms of CMOS circuits with less than 10% waveform differences, when compared with those computed by SPICE. The simulation speed approaches that of a typical switch-level logic simulator as it is based on an event-driven logic simulation algorithm. It reads in the SPICE circuit description file, partitions the circuit into blocks, and then tests whether the feedback paths exist or not by applying level assignment. If not, a waveform calculation algorithm is used to calculate the whole waveform of the output voltage of each block. Otherwise, the simulation process is based on an event-queue data structure. A state change, either from high to low or low to high, is viewed as an event. When an event occurs, the circuit must respond to this event, so the simulator simulates the circuit at the time points that the event and other new events occur. If no event occurs, the circuit is assumed to be idle and no change occurs inside the circuit. Therefore, the simulator only simulates the circuit at the time points that events occur. This technique can reduce extremely the computation time.

Each MOS transistor in switch-level timing simulators is modeled with a linear resistor between its drain and its source terminals, and with a grounded capacitor at its gate terminal. To embrace the capacitive effects found in MOSFETs, two grounded capacitors are connected to source and drain terminals. Thus, charges can be stored in these “internal” capacitors and these “internal” charges play a very important role in the state transition of a CMOS gate. The researchers use the series-parallel PN tree\(^2\) to represent a logic gate, in order to quickly estimate the time-domain state-transition waveforms based on a recursive RC calculation algorithm\(^2\). Figure 1(a) shows a 7-input CMOS compound gate whose function is \( Z = \{ A \cdot (B \cdot (C + D) + (E + F) \cdot G) \} \), and Figure 1(b) shows its equivalent series-parallel tree. Each vertex of the tree consists two items. The first item is the type of node, such as P for parallel, S for series, L for leaf, or O for output. A leaf node represents the corresponding MOSFET. P and S mean the left and right children of this vertex are connected in parallel and serially, respectively. Meanwhile, S corresponds to an actual node inside the gate. \( S_1 \) and \( S_6 \) in the tree represent the internal nodes 1 and 6, shown in Figure 1 (a). The second item is either the gate signal if the vertex is a leaf vertex or the high/low state (H/L) if the vertex represents an internal node. For example, if \( A = B = G = 1 \) and \( C = D = E = F = 0 \), then the output voltage is HIGH. Meanwhile, nodes 1, 2, 4, 5 and 6 are charged to HIGH, so the second items of \( S_1 \), \( S_3 \), \( S_4 \), \( S_5 \) and \( S_6 \) are H and the second item of \( S_2 \) is L, as shown in Figure 1(b).

![Figure 1. (a) A 7-input CMOS complex gate, (b) the equivalent PN tree of (a).](image)

The PN tree can represent the real net-list connections of the circuit; it also illustrates the positions of different MOS transistors inside the circuit. So all conducting paths can be obtained easily from the PN tree. Since there is a delay between the input signal and the output signal of a CMOS gate, one can define the “switching delay” as the difference between the time when the output signal begins to change and the time when the input signal begins to change. The switching delay consists of (1) the time needed to form the conducting path, and (2) the time needed to charge/discharge the internal capacitors connected to the con-
ducting paths before the output voltage begins to change. To improve the accuracy of the voltage waveform, one does need to calculate the RC time constant due to the internal charge. The circuit configuration from an internal node to the reference source can be represented as that shown in Figure 2 (a). For each $U_i$, $D_i$ and $P_i$, it could be open-circuit, short-circuit or one finite resistance. For example, Figures 2 (b) and (c) show the circuit configurations from internal node $S_5$ and $S_2$ to $V_{DD}$ and GND, respectively. This circuit diagram can be viewed as the mesh shown in Figure 2 (d) which can be constructed directly from the PN tree. Thus, the equivalent resistance from $S$ to Source can be obtained. More accurate timing information can be obtained at the expense of the CPU time.

3. Current Model

When evaluating the average power consumption of a CMOS gate, parasitic capacitors can be assumed to be connected to a common node with a constant voltage level (usually, the ground), that acts a sink. The overall energy drawn from power supply in a whole charging/discharging cycle does not depend on which node the capacitors are lumped to. This phenomena, however, the time-domain current profile. There are many parasitic capacitors existing inside a CMOS gate, e.g. $C_{gb}$, $C_{gd}$. In order to find a consistent model with that in SPICE, the author has used the Level 1 MOSFET model in SPICE to represent these parasitic capacitors. Figure 3 shows the MOSFET models in SPICE and in these simulators. Eight Capacitors, $C_{gs}$, $C_{gd}$, $C_{gb}$, $C_{gs0}$, $C_{gd0}$, $C_{gb0}$, $C_{sb}$ and $C_{db}$ are used to model the gate overlap capacitance and the source/drain junction capacitance. The junction capacitor is connected either to power or to ground routes, according to the bulk connection of the corresponding devices. There will be current flowing if the voltage across any two terminals of a MOS transistor changes. In this paper, the authors have assumed that current flows only when the inputs of the gate change. This means that leakage currents and the current due to the fluctuation in $V_{DD}$ and GND are neglected.

Figure 2. (a) The circuit configuration form an internal node $S$ to the reference source, (b) and (c) the equivalent circuit configurations from $S_5$ to $V_{DD}$ and from node 2 to GND, and (d) the mesh structure that can be used to calculate the total effective resistance from an internal node $S$ to the Source.

Figure 3. (a) SPICE Level 1 model, (b) the MOS model used in this program.
ter which state the other input A is. When the input \( V_{in} \) increases from 0V to 5V, the voltage of node A will decrease, but the output of the NOR gate remains unchanged. This input transition has the following effects: (1) discharging the capacitors \( C_{gs11}, C_{gs23}, \) and \( C_{gs32} \) (2) charging the capacitors \( C_{gs12}, C_{gs21}, C_{gs31} \) and load capacitance \( C_{load} \), and (3) activating temporary conducting paths between the supply and ground. Figure 5 (a) shows the voltage and current waveforms obtained using SPICE.

The supply current drawn by a CMOS gate corresponding to an input transition can always be viewed as consisting of three contributions:

- A gate capacitor differential current \( I_{d} \) that increases/decreases the charge of the gate capacitor,
- A charging current \( I_{c} \) that increases the total charge of internal and load capacitor, and
- A short-circuit current \( I_{s} \) that does not affect the charge status of the cell.

In the past, many researchers have focused their attention on the calculation of the charging/discharging current of CMOS gates, and few consider the short-circuit current. They assume that the current flows only during the rising and falling edges of the output transitions of CMOS gates. It means that only the currents \( I_{13} \) and \( I_{32} + I_{33} \) are considered in the above case. But owing to the change of the voltage drop across the gate capacitor \( C_{gs} \) of \( M_{1} \), the currents \( I_{2} \) and \( I_{31} \) flowing from \( V_{DD} \) to node A across the gate capacitors, and the current \( I_{11} \), flowing to \( V_{DD} \) across the gate capacitor of the inverter gate, could not be neglected. In the above case, this 2-input NOR gate will not change its state, therefore, so it is assumed that no current produced by this gate in previous models. But the current will still flow across the gate capacitance, so there will be large error if one would neglect this component.

### 3.1. Modeling Current Pulses

Similar to the current model used References, the current model is a charge-based model. Thus, the charge as a state variable was used. The reason is that the calculation of the quantity of the transferred charge during a transition is easier than that of the current because \( \Delta Q \) is equal to \( C \cdot \Delta V \), where \( C \) is the parasitic/load capacitor that the current charges/discharges, and \( \Delta V \) is the voltage variation across the capacitor. \( \Delta Q \) is also equal to the area of the current pulse, since \( \Delta Q = \int \Delta I \cdot dt \). The current pulse is extremely nonlinear and its variation is larger than the voltage waveform. The author had proposed an approach that approximates the current pulse with three exponential functions. Another easier method was proposed that uses an isosceles triangular current pulse to approximate the real current pulse, but the accuracy is not enough since the current pulse is not symmetric. The accuracy can be improved if the current waveform is represented as an asymmetric triangle pulse with four parameters \( (T_s, T_p, T_e, D_Q) \), as shown in Figure 5(b). \( \Delta Q \) is the total charge transferred to/from the \( V_{DD} \) bus and is equal to the area of this triangle. These three time parameters designate the time when...
the current waveform begins, reaches its peak value and stops changing, and their values can be determined from the voltage waveforms. The author has tried to calculate the charge transferred and determine the values of these three time parameters from the voltage waveforms, then each current waveform can be obtained. After summing each current waveform up, the total current is obtained.

3.2. Capacitor Differential Current

To simplify the calculation, one would consider only the currents flowing to/from VDD, though there are many parasitic capacitances existing inside the CMOS gate. This current can be calculated as \( \frac{dCV}{dt} \). When the input increases, the current flows into the VDD bus; when the input decreases, the current flows from the VDD bus. Since \( C_{gs} \) and \( C_{gd} \) are input-dependent and nonlinear, it is difficult to describe the current waveform with an analytical formula. The total charge transferred due to this current can be calculated as follows,

\[
Q = \int_{V_{IS} - V_e}^{V_{DS} - V_e} dC \cdot V = \int_{V_{IS} - V_e}^{V_{DS} - V_e} C(V) \cdot dV
\]

where \( V_e \) and \( V_s \) are the initial and final values of the input during this transition, respectively. Since \( C_{gs}, C_{gd} \) and \( C_{gb} \) are input-dependent and nonlinear, it is difficult to describe the current waveform with an analytical formula. One can use the approximation method\(^{28}\) to consider the input-dependent characteristic. A PMOS driven by a falling input for illustration was used, as shown in Figure 6, and consider two cases: (a) the source of this PMOS is connected to VDD, like M1 in Figure 4, and (b) the source is open, like M2 in Figure 4 if M1 is off.

As the input changes from High to Low, the PMOS transistor will be in one of the following three regions in sequence: off, saturation, and linear regions.

- **Off region**, where \( V_{DS} - V_T \leq |V_{tp}| \). In this case, PMOS is off, \( C_{ox} = C_{os} \) and \( C_{gs} = C_{go} = 0 \).
- **Saturation region**, where \( \left| V_{gs} \right| \leq V_{DD} - V_{in} \) and \( V_{d} - V_{in} < |V_{tp}| \). In this region, the channel is heavily inverted, \( C_{gs} = C_{gs} = 0 \) and \( C_{gs} = 4 C_{os} \).
- **Linear region**, where \( V_{d} - V_{in} \geq |V_{tp}| \). In this region, the depletion layer depth remains relatively constant, \( C_{gs} = 0 \) and \( C_{gs} = C_{gs} = \frac{1}{2} C_{os} \).

If \( V_d \) is initially high, the PMOS will not enter the saturation region. The value of the charge transferred depends on the total capacitance considered and can be calculated as follows,

\[
\Delta Q = \int_{0}^{V_{ds}} C_{ds} dV + \int_{0}^{V_{ds}} C_{ds} dV + \int_{0}^{V_{ds}} C_{ds} dV \quad \text{if} \quad V_{d}(0) = 0
\]

\[
\int_{0}^{V_{ds}} C_{ds} dV + \int_{0}^{V_{ds}} C_{ds} dV \quad \text{if} \quad V_{d}(0) = V_{DD}
\]
where

\[
\begin{align*}
\text{Case(a):} & \\
C_1 &= C_{p0} + C_{m0} + C_{n0} \parallel C_{o0} \\
C_2 &= C_{p0} + \frac{1}{2} C_{m0} + C_{n0} \parallel C_{o0} \\
C_3 &= C_{p0} + \frac{1}{2} C_{m0} + \left( C_{p0} + \frac{1}{2} C_{m0} \right) \parallel C_{o0} \\
\text{Case(b):} & \\
C_4 &= C_{p0} \parallel C_{o0} + C_{m0} + C_{n0} \parallel C_{o0} \\
C_5 &= (C_{p0} + \frac{1}{2} C_{m0}) \parallel C_{o0} + C_{n0} \parallel C_{o0} \\
C_6 &= \left( C_{p0} + \frac{1}{2} C_{m0} \right) + \left( C_{p0} + \frac{1}{2} C_{m0} \right) \parallel C_{o0}
\end{align*}
\]

From the above equations, one can find that the value of the charge transferred is dependent on the status of the inner structure of the CMOS gate and its initial state, so it cannot be calculated by using a simple equation, like \( \int C_{eff} \cdot dV \), with an effective capacitance \( C_{eff} \). Therefore, the value could be determined only during the simulation process to see how many parasitic capacitances that should be considered. The values of these capacitances can be obtained using the model parameters of SPICE, such as, CGDO, CGSO, and the physical parameters of this MOSFET, such as, L and W^29.

If \( V_s \) in the above equation is smaller than \( V_e \), then \( Q \) is negative and it means that the differential current flows into the \( V_{DD} \) bus. If \( V_s \) is larger than \( V_e \), then \( Q \) is positive and the currents flow from the \( V_{DD} \) bus. \( I_{p1} \), \( I_1 \), and \( I_{p2} \) shown in Figure 4 belong to this kind of current. Since this current is due to the change of the input, the values of \( T_{s1} \) and \( T_{e1} \) are set as the time points when the input signal begins and stops changing, respectively. The peak current occurs when the voltage slope of the input is at its maximum. But the voltage waveform is expressed as a linear segment with an exponential tail in this simulator^26, so \( T_{p1} \) is set as the time when the input is \( V_{DD}/2 \) approximately.

### 3.3. Capacitor Charging Current

Let us consider the generic CMOS gate structure shown in Figure 7, which is used in many literatures\(^{1,19}\). The output node capacitance is split into two lumped capacitors, \( C_{s} \) to \( V_{DD} \) and \( C_{n} \) to GND. On a low-to-high transition, the currents flowing through \( C_{s} \) and \( C_{p} \) at the output node are \( i_{p1} \) and \( i_{n2} \), respectively, as shown in Figure 7. The corresponding \( I_{n1} \) and \( I_{n2} \) for a high-to-low transition are also shown. Only \( i_{p1} \) and \( i_{n1} \) are considered\(^{13,15}\). The primary problem of this model is the way to split parasitic capacitors into \( C_{s} \) and \( C_{p} \). As one may know, not all parasitic capacitances are connected directly to \( V_{DD} \) or GND, as described in the last subsection, so the values of \( C_{s} \) and \( C_{n} \) are time-variant. Errors will occur if the output capacitance is split carelessly, especially when the fanout is large. Another problem is that \( i_{p1} \) is produced by the current gate while \( i_{n1} \) is not.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Fig7.png}
\caption{Current paths for a generic CMOS gate.}
\end{figure}

The primary character of an event-driven simulator is that only one CMOS gate is processed at a time. This means that the status of a CMOS gate is known only when it is processed. Precisely speaking, it is hard to know the value of \( C_{p} \) and \( C_{n} \), except that one would use very complex data structure and algorithm at an expense of CPU time. This method is to focus the attention on calculating how much the current produced inside the current CMOS gate, even though some of the current will return to \( V_{DD} \) by flowing across the capacitance \( C_{s} \) of the driven gate. Therefore, all the parasitic capacitances that are connected to the output node are lumped as one load capacitance, and use this load capacitance to calculate the charging current. The value obtained this way will be larger than that obtained by using previous models, since part of it, as differential current, will return to \( V_{DD} \) through other CMOS gates. But the over-estimated part will be compensated when one move the attention to these other gates.

Since the total charge transferred to the load capacitor during the transition is \( V_{DD} \cdot C_{load} \), the key point is how to calculate the load capacitance accurately. The load capacitance is also a function of the design parameters (\( L \), \( W \), and other parameters), and its value can be obtained by simply lumping all the parasitic capacitances that are connected to the output node.

\[
C_{load} = \sum C_{gd} + \sum C_{gs} + \sum C_{gb}
\]

The total charge transferred can be calculated as \( \Delta Q = V_{DD} \cdot C_{load} \). This current is labeled as \( I_1 \) shown in Figure 4. Since this current is due to the change of the output voltage, the values of \( T_{s2} \) and \( T_{e2} \) are set as the time when the output voltage begins and stops changing, respectively. The peak current occurs when the voltage slope of the output is at its maximum. Therefore, \( T_{s2} \) is the time when the output voltage is approximately \( V_{DD}/2 \).
The objective is to obtain the total current supported by the $V_{DD}$ bus, thus, the author is interested in the current on the low-to-high transition. Since $C_p$ and $C_n$ are combined into one effective load capacitance $C_{load}$ and the charging current is calculated based on this capacitance. Of course, this current will be larger than $i_p$ shown in Figure 7. This current is equal to the summation of $i_{p1} + i_{p2}$. But since $C_p$ is the gate capacitance at the next stage, there will be a capacitance-differential current flowing to $V_{DD}$, which is equal to $i_{p2}$. Therefore, the over-estimated current will be subtracted when one would process the next stage. On the other hand, the current will be under-estimated on the high-to-low transition owing to neglecting the discharging current. But, since $C_n$ is the gate capacitance at the next stage, there will be a capacitance-differential current flowing from $V_{DD}$. Therefore, the under-estimated current will be also complemented.

### 3.4. Short-Circuit Current

A static CMOS inverter does not dissipate power during the absence of transients on the inputs. Since the input could not change from one state to another state immediately, so, as the input signal is larger than the threshold voltage of N-MOSFET and less than the threshold voltage of the P-MOSFET, the N-MOSFET and P-MOSFET are turned-on at the same time, as shown in Figure 8. As a result, during a transient on the input, there will be a small time period in which both the N-MOSFET and the P-MOSFET conduct, causing a current to flow from $V_{DD}$ to ground. This current is called short-circuit current, since there exists a direct path from $V_{DD}$ to ground. This current flows as long as the input voltage is higher than a threshold voltage above $V_{DD}$ and lower than a threshold voltage below $V_{DD}$. Clearly, this short-circuit current strongly depends on the physical design parameters, so it is very difficult to describe the waveform of the short-circuit current precisely, and the analytical solution of this current can be obtained only when the gate is a CMOS inverter gate under some assumptions. Veenendrick has proposed an analytic equation to estimate average short-circuit power dissipation of an unloaded inverter,

$$P_s = \frac{\beta}{12} (V_{DD} - 2V_t)^2 \cdot \frac{\tau}{T}$$  \hspace{1cm} (3)

where $\beta$ is the gain factor, and $V_t$ is the threshold voltage of a MOS transistor. $T$ is the signal period and $\tau$ is the rising or falling time of the input signal. The total charge transferred due to the short-circuit current can be obtained as follows,

$$\Delta Q = P_s \cdot \frac{T}{2} = \frac{\beta}{6} (V_{DD} - 2V_t)^2$$  \hspace{1cm} (4)

Although more accurate results have been obtained, but they are too complex. Therefore, the author has used the above equation to calculate the charge due to the short-circuit current.

This current is labeled as $I_3$ shown in Figure 4. The short-circuit occurs when the NMOS (PMOS) turns on and stops when PMOS (NMOS) turns off during the charging (discharging) process, therefore, the values of $T_{p3}$ and $T_{n3}$ are set equal to these time points. $T_{p3}$ could be obtained through \( \frac{\Delta Q}{I_{VDD}} \) instead of $n$. In this case, SPICE level 1 equations for $I_{ns}$ are used. So $T_{p3}$ can be derived as follows,

$$T_{p3} = \frac{t_n}{t_n + 2 \cdot t_p} \left( 1 + K \cdot \frac{t_n}{V_{DD}} + K \cdot \frac{t_n}{t_{fp}} \right) t_p$$

The definitions of $K$, $t_n$, and $t_p$ are shown in Figure 8.

![Figure 8. The relationship between the input/output voltage waveforms and the short-circuit current waveform.](image)

### 4. Modeling the Event Overlapping Cases

The accuracy of the current simulation depends strongly on the voltage waveforms. So far, the author has constructed a refined model of the supply current drawn by a generic CMOS gate corresponding to a transition between two input patterns. The author has implicitly made the assumption that all switching inputs have the same arrival times. Unfortunately, in real circuits, the internal signals are in general slightly misaligned and may give rise to glitches and overlapping transitions. A glitch is a small pulse in the time domain, while the overlapping transition is a transition that change the slope in the midway. One can consider the relationship between contiguous events. According to the time point that the second event occurs and the signs of the slopes of event $V_1$ and $V_2$, there are 7 cases that can be classified, as shown in Figure 9. Assuming that the slope of event $V_1$ is positive,

- $V_1$ occurs after $V_1$ reaches its steady state.
- $V_2$ occurs before $V_1$ reaches ONn and the slope of $V_2$ is negative.
- $V_2$ occurs before $V_1$ reaches ONn and the slope of $V_2$ is positive.
- $V_2$ occurs before $V_1$ reaches ONp but after $V_1$ reaches ONn, and the slope of $V_2$ is negative.
- $V_2$ occurs before $V_1$ reaches ONp but after $V_1$ reaches ONn, and the slope of $V_2$ is positive.
- $V_2$ occurs after $V_1$ reaches ONp but before $V_1$ reaches its steady state, and the slope of $V_2$ is negative.
Time-Domain Current Waveform Simulation for CMOS VLSI Circuits Using Charge-Based Model

- $V_2$ occurs after $V_1$ reaches $ONp$ but before $V_1$ reaches its steady state, and the slope of $V_2$ is positive.

![Figure 9. The relationships between two contiguous events.](image)

Case 1 is the normal case. Event-overlapping, the Case 2 to case 6, occurs when two adjacent events are so close that the second event occurs before the first one finishes. It can be classified into two groups: (1) Superposition occurs when the signs of the two adjacent events are the same, and (2) Glitch occurs when the two adjacent events are opposite. Only one event in the Superposition case is effective. And the two events in Glitch case could be both effective or invalid according to that the amplitude of the glitch is greater than a threshold value or not. The shape of the current pulse would not be a perfect triangle in both cases. And no matter how large its amplitude is, a glitch may draw significant amounts of currents. Therefore, it should not be neglected when one considers the current waveform.

Though these phenomena of event-overlapping have a sizable effect on power consumption and current flows, they have never been modeled at gate-level for two reasons. First, they elude any pre-characterization attempt due to the intractable number of possible combinations of signal skews. Second, the corresponding current waveforms are no longer shaped as triangular pulses.

To handle glitches, not like the interpolation method used in, the author has a more reasonable and more accurate method. One can consider a 3-input CMOS NOR gate. In case 1, the time difference between $Vin1$ and $Vin2$ is 2.5ns, could reach its final state before the second event occurs that makes the decrease. The voltage and current waveforms are shown in Figure 10(a) and (b) as solid lines. In case 2, the time difference is 0.3ns, a glitch occurs on and its amplitude is. The voltage and current waveforms are shown in Figure 10(a) and (b) as dashed lines. One can see that the shape of the current pulse due to this glitch is extremely irregular and could not be modeled as a triangle. But, the author could find that the current waveform in case 2 seems to be the superposition of the two current pulses in case 1 if one would scale down the second pulse with a factor and moved it close to the first pulse. According to this observation, one uses the following method to calculate the current waveform corresponding to a glitch.

When calculating the current pulse corresponding to the first event, the author assumes that this event can reach its final state, in order for the whole current waveform corresponding to the first event would be obtained. But since the second event occurs before the first event finishes, so this current waveform is cut at the time that the second event occurs, as shown in Figure 10 (b). The current pulse corresponding to the second event is calculated as it is a whole-swing event, and then scale the current pulse corresponding to the second event according to the peak ratio, $V_{z}/V_{DD}$, where $V_z$ is the voltage swing of the glitch.

![Figure 10. (a) The voltage waveforms of a 3-input CMOS NOR gate simulated using SPICE. The dashed lines show the case 1 and the solid lines show the case 2. (b) The current waveforms.](image)

This is the basic concept used to model the event-overlapping cases. Next, the concept is applied in the current calculation. One can add another time parameter $T_{cut}$ to model the pulse cut in event-overlapping cases. So the current pulse is represented as an asymmetric triangle pulse with five parameters $(T_s, T_p, T_{cut}, \Delta Q)$. In normal cases, $T_{cut}$ is set equal to $T_s$. See Figure 10. Case 1 is the normal case. Case 2 and 4 are glitch cases, and case 3 and 5 are superposition cases. In all other cases, the current pulse due to event 1 is cut at the time the second event occurs.

- Differential current: In cases 2 and 4, the current pulse corresponding to the second event is calculated as it is a whole-swing event, and then scale the current pulse corresponding to the second event according to the peak ratio, $V_{z}/V_{DD}$, where $V_z$ is the voltage swing of the glitch. In cases 3 and 5, the current pulse due to the second event is calculated as followings,
The time-domain current waveform supported by the \( V_{DD} \) bus is the summation of the current waveforms of each block. The current waveform of each block consists of a series of triangular pulses, and is represented as a current source to the \( V_{DD} \) bus. Since the metal resistance is negligibly small within a sub-circuit, one can combine several sources in a sub-circuit into a single current source, which represents the power-bus current drawn by the sub-circuit. Let the addition of a triangle to the waveform be a basic unit of calculation. And let \( M \), \( N \), and \( K \) be the total event number of all blocks, the total event number of all primary inputs of this circuit, and the total event number of the primary outputs of this circuit, respectively. When one calculate the current waveforms, the number of the addition operation required is as follows,

\[
\frac{M - N}{2} \text{ calculating the capacitor differential current}
\]

\[
\frac{2}{K}M \text{ calculating the load capacitor current}
\]

\[
M - N \text{ calculating the short circuit current}
\]

Therefore, the total number of the addition is \( \frac{M - N}{2} (N + K) \). Because \( M \), in general, is larger than \( N + K \), the CPU-time overload of the current calculation to the timing simulation is strongly dependent on the event number, no matter how large the circuit.

The simulator has been tested extensively for basic modules such as counters, decoders, adders and ALUs. The CPU time comparisons are summarized in Table 1. The CPU time consists of the time that the timing simulator BTS used and that the current calculation used on a SUN SPARC station 1+. Since BTS considers the effects of the internal charges, the simulator is not as fast as other timing simulators, but the waveforms it derives are more accurate. It is very important that the accuracy of the current waveform calculation is strongly coupled to the accuracy of the timing information. In Table 1, one would list three more error values, DC, Max_I and Max_T. DC error is the relative error between the average current \( I_{\text{avg,BTS}} \) and \( I_{\text{avg,SPICE}} \). Max_I is the relative error between the value of the peak current \( I_{\text{peak,SPICE}} \) and \( I_{\text{peak,BTS}} \). Max_T is the absolute error about the time that the peak current occurs. Figure 12(a) and (b) show the voltage waveforms of the circuit shown in Figure 4. The dashed lines are obtained using SPICE, and the solid lines are obtained using BTS. Figure 13 shows the current comparisons of an encoder SN74147 with SPICE’s result. For the purpose of comparison, the current waveform is the summation of all the currents from individual sub-circuits.

Table 1. The comparisons between BTS and SPICE.

<table>
<thead>
<tr>
<th>Circuits (MOS No.)</th>
<th>DC</th>
<th>MAX_I</th>
<th>MAX_T</th>
<th>BTS</th>
<th>SPICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>74381(584)</td>
<td>0.50%</td>
<td>-8.52%</td>
<td>-8.21e-11</td>
<td>1.483</td>
<td>1478.6</td>
</tr>
<tr>
<td>7483(258)</td>
<td>-1.71%</td>
<td>-4.15%</td>
<td>-6.39e-13</td>
<td>1.383</td>
<td>282.37</td>
</tr>
<tr>
<td>74147(146)</td>
<td>1.78%</td>
<td>5.76%</td>
<td>7.64e-10</td>
<td>1.107</td>
<td>121.55</td>
</tr>
<tr>
<td>Inverter Chain(200)</td>
<td>2.2%</td>
<td>-7.43%</td>
<td>-4.24e-12</td>
<td>0.700</td>
<td>100.7</td>
</tr>
</tbody>
</table>

5. Implementation and Simulation Results

This current model has been embedded into this timing simulator BTS. This timing simulator uses a recursive technique to calculate the delays in the series-parallel MOS circuits\(^{27} \). The voltage waveform slope is approximated by a linear segment followed by an exponential tail\(^{28} \). The results are accurate since BTS uses an accurate delay calculation method that computes the switching delays and slopes in series-parallel RC networks with the considerations of the effects of the internal charges\(^{26} \). When the timing simulation finishes, the event list at the output of each gate can be obtained. Then, the current pulse associating with each event could be obtained using the current model mentioned above.
6. Conclusions

In this paper, the author has presented our accurate time-domain current waveform simulator. This simulator could be viewed as a post-processor of a timing simulator which is used to obtain the voltage waveforms. The primary contribution of this approach is to separate a current pulse into three components and calculate each component respectively. A charge-based current model is described which can be used to generate the time-domain transient current waveforms in the power bus lines. Methods to calculate the charge associated with each current component are discussed. The simulated waveforms in general differ by no more than 10% from those simulated by SPICE. Its speed is $10^2 \sim 10^3$ times faster than that of SPICE for circuits with hundreds of transistors, and the speed ratio is expected to be even more significant for larger-scale circuits. At the cost of a little speed ratio, the results are more accurate, especially the time at which the peak current occurs. The accurate current waveform will support solving the VLSI reliability problems due to electromigration and excess voltage drops in the power buses, since these two problems are primary resulted from the current.

References


