A Fast and Retargetable Framework for Logic-IP-internal Electromigration Assessment Comprehending Advanced Waveform Effects

Palkesh Jain, *Member IEEE,* Jordi Cortadella, *Fellow, IEEE* and Sachin S. Sapatnekar, *Fellow IEEE*

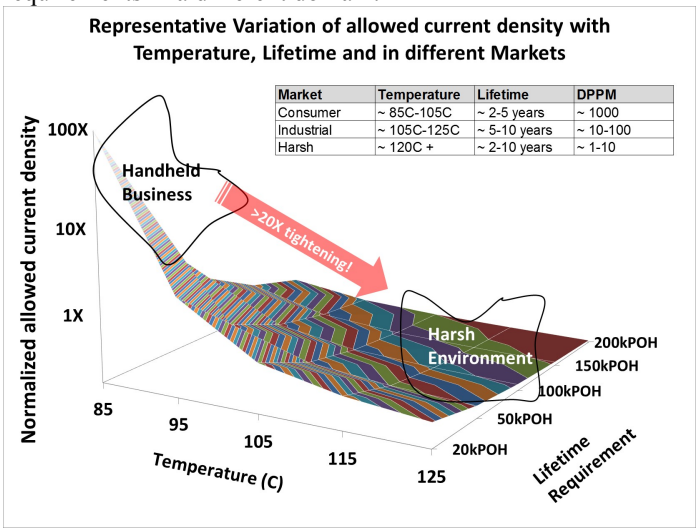
Abstract— **A new methodology for SoC-level logic-IP-internal EM verification is presented in this work, which significantly improves accuracy by comprehending the impact of the parasitic RC loading and voltage-dependent pin capacitance in the library model. It additionally provides an on-the-fly retargeting capability for reliability constraints by allowing arbitrary specifications of lifetimes, temperatures, voltages and failure rates, as well as interoperability of the IPs across foundries. The characterization part of the methodology is expedited through intelligent IP-response modeling. The ultimate benefit of the proposed approach is demonstrated on a 28nm design by providing an on-the-fly specification of retargeted reliability constraints. The results show a high correlation with SPICE and were obtained with an order of magnitude reduction in the verification runtime.**

***Index Terms*—reliability, electromigration, signal probability, retargeting, pin capacitance**

I. Introduction

ELECTROMIGRATION (EM) is a major product aging mechanism revolving around the containment of the average and RMS current densities in interconnects. This, in turn, requires *cell-external analysis* for signals and power nets connecting to the cells and *cell-internal analysis* for wires within a logic-IP (standard cells) or mixed signal IP block. Recently, a great deal of innovation and improvement has been seen on the verification and design strategies for cell-external signal and power grid EM [1-4]. However, there has not been adequate focus on the robust design and reuse of the standard cells. Ensuring EM reliability for standard cells and IPs in a design implies that the exact context at which the IP is used must be bounded to guarantee its robustness in the design. This context could be stated in terms of design limits (loads, slews, frequencies, supply voltage), or reliability (temperature, lifetime, or a failure rate specification tied to current density limits). Without rigorous assessments, a set of IPs designed for a particular *reliability condition* (*e.g.*, 1.2V, 105C, 100k power-on hours (POH), 0.1% cumulative failure and 10C Joule heating (JH) limit) cannot be guaranteed to be EM-safe at another condition (*e.g.*, 1.0V, 115C, 200kPOH, 0.01% cumulative failure and 15C JH limit).

Nevertheless, tradeoffs on these constraints are increasingly in demand in industry due to accelerated inroads of semiconductor houses into newer businesses with different reliability demands [5]. For example, industrial designs demand more stringent operating conditions than traditional computing applications [6, 7]. From an EM standpoint, meeting these specifications is challenging, as seen from Fig. 1, which highlights the representative current density per μm^2 across various temperature and lifetime specifications. As can be seen, amongst the various environments, the current carrying capability becomes over 20x more stringent. Not only amongst different application markets, even for the same SoC itself, different complex IPs (*e.g.*, CPU core or a DSP) can have different reliability requirements, based on their ON times and temperature specifications. The challenges increase when such reliability requirements could be only made available *on-the-fly*: that is, either during the final SoC verification or even after the SoC tape-out; in which cases, the original reliability targets for the IP, characterized for one application domain, may not match with the reliability requirements in a different domain.



Manuscript received June 19, 2015; revised October 12, 2015; accepted November 27, 2015.

P. Jain was with Texas Instruments India Pvt, Ltd., Bangalore 560093, India. He is now with Qualcomm Technologies Inc., Bangalore 560066, India (e-mail: palkesh@qti.qualcomm.com).

J. Cortadella is with the Computer Science Department, Universitat Politècnica de Catalunya, Barcelona 08034, Spain (e-mail: jordi.cortadella@upc.edu).

S. S. Sapatnekar is with the University of Minnesota, Minneapolis, MN 55455 USA (e-mail: sachin@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2015.2505504

Fig. 1: Typical current density limits as a function of temperature and lifetime, showing >20X differences between various environments.

One way to meet such diverse specifications is to approach the design in a bottom-up manner with a fresh logic-IP portfolio that meets targeted domain-specific reliability specifications. However, this is very expensive, and economic and design effort considerations often dictate that the product integration over all application domains be based on the same IP portfolio. This implies that the logic-IPs require a disciplined utilization procedure, making it important to assess their exact usage boundaries at arbitrary conditions.

A starting point towards this is to ensure that the cell is EM-safe at a specific load and frequency by selecting wire widths so that EM constraints are met. However, this only implies that a lower load and lower frequency can be considered EM-safe. The cell may (or may not) be EM-safe at a lower load and higher frequency, or a higher load and lower frequency, or a higher load and higher frequency, and this can only be uncovered through costly detailed analysis.

As an improvement, some industrial implementation tools [8, 9] use a precharacterized table that models the tradeoffs in various design/reliability parameters. Fig. 2a shows a representation of one such table, where x-axis represents an operating constraint of the cell (load here, but this could be slew, supply voltage, or any reliability constraint) and the y-axis represents the alternative constraint (frequency here), at a baseline reliability condition.

The intuition behind such a table (frequency versus load; f-L) is simple: the current flow in the IP increases with the operating load, and hence the frequency should be lowered to meet the reliability specification. This model can be used at the chip level to determine the safe frequency (f_{safe}) of an instance for any design/reliability parameter, and then make corresponding design fixes. Needless to say, most of the EM-critical cells are the ones that operate at higher loads, frequencies or slews.

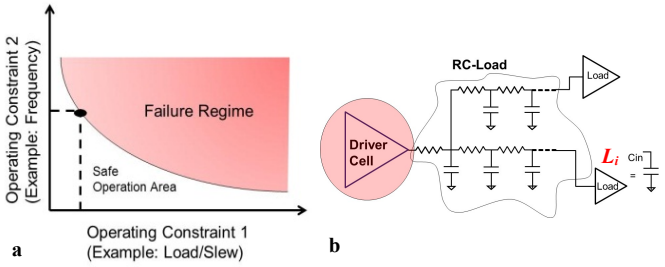


Fig. 2a: Traditional approach for EM verification using the safe operating region concept. Fig. 2b: Schematic highlighting the EM-critical cell, driving an RC load network (vis-à-vis safe frequency obtained for pure C load)

However, such a specification is also simplistic and with advancing technology and convoluted circuit effects, this model is inadequate in accurately predicting EM safety for several reasons. First, the frequency in Fig. 2a refers to the output switching frequency: for multi-input cells, the failure rate depends on the switching frequency at each input. This corresponds to a multidimensional space that is computationally expensive to characterize. Second, EM constraints are often specified in terms of average current density thresholds or RMS [10]. However, having multiple relationships between operating parameters is infeasible. Lastly, while the traditional f-L model is characterized at purely capacitive loads, in reality, cells drive RC loads (Fig. 2b), and fast prediction of cell-internal EM safety under RC loads is an open problem.

Our goal is to address four limitations (L1–L4) associated with chip-level cell-internal EM analysis:

- **L1** – Inability to incorporate the impact of arbitrary switching rates on inputs pins and effects such as clock gating: We overcome this by discretely characterizing the individual current density components (switching or leakage). Additionally, our frequency constraints are self-consistent, which simultaneously address the average and RMS current density criteria, based on formulations proposed by Hunter [10].
- **L2** – Inability to comprehend RC loads (Fig. 2b) and to model voltage-dependent pin capacitance (C_{in}): We apply intelligent moment-matching-based techniques as in [3], and propose a novel formulation for C_{in} estimation.
- **L3** – Inability to retarget reliability specifications on-the-fly for different reliability conditions: We develop the concept of equivalent stress and present closed-form formulae.
- **L4** – Nonscalability of cell characterization data for an entire library due to prohibitive simulation runtimes, with ~600 simulations per cell: We perform these simulations efficiently using intelligent response modelling.

The core methodology of our work naturally enables model retargeting by separating the current density computation part from the verification, as against the tight coupling in the model of Fig. 2a, where the f-L curve must be characterized at each reliability condition. In our approach, the reliability conditions need to be specified in-situ: only at the design verification stage. Moreover, our model can take the operating frequency (f_{op}) of an instance as an input, or it can provide the maximum safe operating frequency as an output.

II. EM Modeling – Basic Framework Under Purely Capacitive Loads

A. Electromigration Basics

In this section, we review the key parameters affecting EM. In our terminology, we refer to metal segments of the IP as *resistors*. These resistors are obtained by parasitic extraction, which retains key information such as the width, length, and the metal-level for every resistor in the netlist.

Since EM is a statistical process, the time to failure for metal segments stressed in similar conditions also varies [11]. Industrial markets demand low failure rates (*e.g.*, 100 defective parts per million (DPPM) over the chip lifetime). Chip reliability engineers translate this chip-level specification to specific fail fraction (*FF*) targets, in units of failures-in-time (FITs), on individual resistors.

The classic Black's equation [11] relates the mean time to failure (t_{50}, time to failure for half of the population) to the average current density J across the interconnect cross-section and the wire temperature T as:

$$t_{50} = A\, J^{-n}\, e^{Q/k_B T} \qquad (1)$$

Here, Q is the activation energy, k_B is Boltzmann's constant, n is the current exponent (typically between 1 to 2), and A is a fitting parameter.

Black's equation predicts the time to failure, and in practice, it is predominantly used to determine the average current density thresholds to meet a target *FF*. It has been demonstrated that *FF* follows a lognormal dependency on the time to failure (t_f, also known as stress time) [11]. The lognormal-transformation parameter (z), relates to the time-to-failure as follows, where σ is the standard deviation of the distribution, which is process-dependent:

$$z = \frac{\ln(t_f) - \ln(t_{50})}{\sigma} \; ; \; FF = \int_{-\infty}^{z} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \qquad (2)$$

The transformation variable z helps in directly representing the cumulative failure rate with a normal cumulative distribution function [34]. For example, at stress time (t_f) = t_{50}, z and *FF* consistently evaluate to 0 and 0.5 respectively.

In signal wires, currents flow in both directions, leading to a limited damage recovery, which can be incorporated by an empirically estimated recovery factor, ξ, used for adjusting the computed average current density in Black's equation [12], as:

$$J_{avg} = J_{avg}^{+} - \xi J_{avg}^{-} \qquad (3)$$

Here, J_{avg}^{+} and J_{avg}^{-} indicate the average current density during current conduction in the positive and negative directions, respectively. Additionally, the wire heating (ΔT) has an inherent dependence on the RMS current density, J_{RMS}, as:

$$\Delta T = c J_{RMS}^2 \qquad (4)$$

Eq. (4), with c as a fitting parameter, follows directly from heat conduction principles. Typically, the limit on the maximum temperature rise due to Joule heating is a design constraint, and this places automatically limits on RMS current densities.

Pioneering work by Hunter [10] combined the two effects of average EM fails and RMS-induced Joule heating in a self-consistent manner through the concept of duty cycles, making it possible to simultaneously check both conditions. Thus, given the constraints of stress temperature, lifetime and Joule heating limit, we can arrive at the EM thresholds that should be met by all metal segments in the IP. Once we have the EM thresholds in place, we can embark on the EM verification process across various resistors in the IP.

It must however be noted that fundamentally, EM is induced by divergence of atomic flux – which is typically highest at sites such as vias, contacts, or even points where the leads merge. Further, it has been reported in literature that even if the incoming atomic flux (signified by high current density) is high at such sites, the site itself may not fail due to it maintaining a low divergence; while a simple, individual-lead based Black's equation continues to predict failure for such a structure. This inefficiency has been recently revisited by various researchers resulting into evolution of alternate paradigms in EM checking [31-33]. Fundamentally, such alternative methods rely on computing some form of atomic flux divergence at EM-probable sites and subsequently comparing them against set thresholds. One such method, as reported in [31] is vector via-node based method, wherein the physical and directional interactions amongst various leads is incorporated to perform the reliability verification. Notably, however, the fundamental inputs required to perform these calculations still remain the individual current density in every single interconnect of the circuit, along with additional information like the circuit topology.

Consequently, we note that even for alternative EM checking methodologies, the discrete current density in individual interconnects still is the vital input – which is discussed in greater details in Section III.

B. Traditional Approach for Modeling EM Reliability

We begin by revisiting the traditional approach, as outlined in Fig. 2a. Given the physical design of the IP, EM verification requires a model that provides a tradeoff amongst various operating conditions such that within the bounds of those tradeoffs, the IP remains EM-safe. The generation of this model requires an iterative search: for example, in Fig. 2a, at a fixed loading and reliability condition (say, 50fF, 1.0V, 105C, 100kPOH), an iterative search over the frequency space is required to determine the maximum f_{safe}, where all resistors within the IP are EM-safe. This is computationally expensive since each iteration involves a SPICE-simulation-based verification. A typical optimized procedure requires ten binary search iterations at each loading condition. For a single input cell, whose operating load/slew space is covered through an 8×8 matrix in the liberty file, the number of required iterations are about 64×10 = 640 for fixed values of other parameters (supply voltage and reliability specifications). To support operation at multiple supply voltages, as well as IP reuse across application domains, this number must be multiplied by the number of use cases, resulting in a formidable characterization overhead.

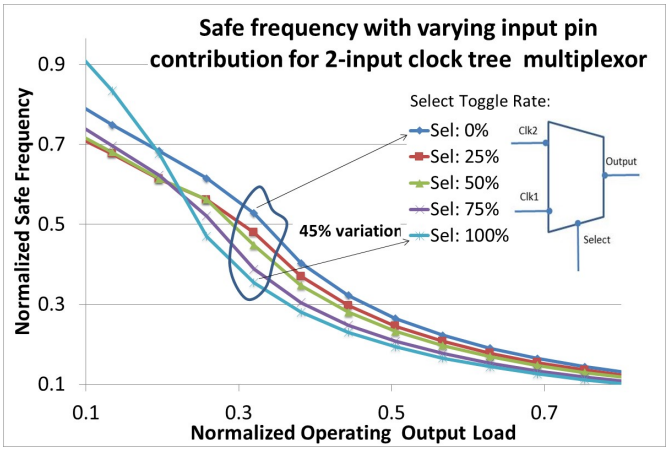


Fig. 3 f_{safe} plot for a 2-input clock-multiplexor cell. Both input clocks switch at 100%, while the select pin chooses one of them, with varying likelihoods.

While this may even be tractable for single-input cells, for multiple-input cells, this characterization becomes challenging, not just from a computational point of view, but also from the fundamental modeling (**L1**) viewpoint. To illustrate this idea, consider the example of a two input clock-tree mux IP block that is used to alternate amongst clocks for downstream propagation. The user may examine typical workloads and use cases and provide an EM analysis tool with information about the switching rates of the input pins of a block. In this experiment, both the input pins (Clk1, Clk2) switch at 100%, but the select pin is toggled to allow passing of first and second clocks in varying amounts (going from 0% to 100% in steps of 25%).

The f-L plots for the five cases are shown in Fig. 3, and show a variation of up to 45% in f_{safe} estimates, depending on how often Clk1 or Clk2 is selected over the lifetime, but the

traditional model will choose the pessimistic f_{safe} over all cases. We also see that while at very low loads, select = 100% (meaning Clk2 being transmitted) results in least EM stress, it switches over as the load increases. This switch results from the interplay of various currents (short-circuit and the output switching current) in the circuit. Such an asymmetric response can only be captured by the traditional model by individually generating and storing the f-L data for various input excitations, which is computationally expensive. Further, effects like clock gating are not straightforward to handle in the traditional model.

Another significant drawback (**L2**) with the traditional model is the fact that it has been generated using a lumped C load, while real applications involve RC loading. Due to resistive shielding effects, a direct application of traditional model to assess reliability of instances that drive RC loads turns out to be severely pessimistic. Finally, we also note that the traditional model is locked to a particular reliability specification (supply voltage, temperature, lifetime and failure rate target), and is incapable in allowing a tradeoff on these (**L3**), unless, the f-L data is regenerated along these vectors, which becomes computationally unaffordable for the entire library (**L4**). With the above background and detailed understanding of the traditional model (including generation, usage and associated limitations), we now look at building the proposed model, which can address the various limitations.

III. Addressing L1 – Incorporating Arbitrary Switching and Clock Gating in Frequency Estimation

A. Library Level Current Density Characterization

In order to build the model which can help predict the reliability of an IP for arbitrary switching scenarios, we begin in an *ab initio* manner by trying to classify the current flow in the IP as either leakage or switching current. We observe that for a combinational IP with m inputs, 2^m distinct static states (various combinations of input pins at logic 1 or 0) are possible. Each of these states can have different leakage flow. Additionally, based on the IP functionality, there could be several paths (later referred to as *arcs*) from an input pin resulting in an output transition. Every such output transition, causes a switching current flow in the IP-internal resistors (belonging to the resistor-set $\Re$).

Thus, first step in our approach is to discretely characterize the current flow: *average and RMS,* both through every resistor R in the IP (resistor-set $\Re$), in every legal state (for leakage current) or arc through the cell (for switching current). Such a characterization will be used to compute the eventual effective current density through any resistor of the cell as a weighted summation of the current densities in unique scenarios, coupled with the information of arc switching rates and probabilities of legal state occurrences.

The salient feature of our characterization is that it remains independent of the reliability condition, which is actually an input during chip-level verification. As the leakage current density in the cell depend only on the static states of inputs, we can easily obtain the current density through R by cycling through all possible input states in SPICE (note that average and RMS remain the same due to DC nature of the waveform). On the other hand, switching current densities are tied to a particular input-pin to output-pin combination (also referred to as a timing *arc*), through a fixed cell-internal path, with other inputs in non-controlling states enabling the transition. For example, for a three-input AOI gate (Y = !(A + BC)), the output Y can fall because of a rise on A in three different states of BC, namely, 00, 01 and 10. Hence, for this particular A→Y arc, the current density must be computed through R for these three states of BC. We can leverage the simulation framework of industrial timing characterization systems [9], to obtain information about all such arcs and states through the cell. For a particular arc i and associated non-controlling state k, we denote the time duration over which this current density is calculated as s_{ik}. A similar convention is followed by $J_{avg,R_{ik}}$ and $J_{rms,R_{ik}}$ to define the average and RMS current densities through R. As we leverage the timing characterization framework, we do not recompute s_{ik}, but reuse it from the timing analysis step [8, 13]. Moreover, s_{ik} is typically greater than the delay itself, and therefore accurately captures the tail effects.

B. Effective Current Density Estimation for a Chip-Level Instance

After characterizing the leakage and switching current densities for various arcs and states, we now present the calculations for the average and RMS densities in the circuit.

B.1 Effective Leakage Current Density Through a Resistor Across All States

For an m-input gate, let the leakage current density through resistor R for a state k (of 2^m states) in the positive [negative] direction be denoted by L_{+R_k} [L_{-R_k}]. Then, the average effective leakage current density ($L_{avg,R}$) covering all the states and incorporating recovery (eq. (3)) would be:

$$L_{avg,R} = \sum_{k=1}^{l} P_k^+ \, L_{+R_k} - \xi\left(\sum_{i=1}^{2^m-l} P_k^- L_{-R_i}\right) \qquad (5a)$$

Here l is the number of states with positive current density, and P_q^+ [P_q^-] is, the probability of occurrence of state q in which the current flows in the positive [negative] direction. These probabilities are a function of the duty cycle at the inputs of the gate.

The RMS effective leakage current density is given by

$$L_{rms,R}^2 = \sum_{k=1}^{l} P_k^+ \, L_{+R_k}^2 + \sum_{i=1}^{2^m-l} P_k^- L_{-R_k}^2 \qquad (5b)$$

B.2 Effective Switching Current Density Through a Resistor Across All Switching Arcs

In similar spirit, the effective-average-switching current density through R ($J_{avg_{sw},R}$) is given by:

$$J_{avg_{sw},R} = \sum_{i=1}^{all\ arcs}\left(\sum_{k=1}^{all\ states} P_{ik} J_{avg,R_{ik}} \frac{s_{ik}}{T_{clk}}\right) \qquad (6)$$

Here, P_{ik} and T_{clk} are the design-level parameters – the switching probability of the particular arc, and the switching period respectively. The scaling factor, s_{ik}/T_{clk}, translates the characterized current density ($J_{avg,R_{ik}}$), which was averaged during the characterization over the switching duration s_{ik}, to the entire clock period. This scaling factor accounts for the fact that the current is inactive during the remainder of the clock period. Similar calculations for RMS current density ($J_{rms_{sw},R}$) yield:

$$J_{rms_{sw},R} = \sqrt{\sum_{i=1}^{all\ arcs}\left(\sum_{k=1}^{all\ states} P_{ik} J_{rms,R_{ik}}^2 \frac{s_{ik}}{T_{clk}}\right)} \qquad (7)$$

B.3 Effective Average and RMS Current Densities

After computing the effective switching and leakage current densities independently, we must now compute the effective average and RMS current densities. In a normal design flow, the chip level probabilistic activity propagation tools already provide the effective switching rate ($f_{ik} = \frac{P_{ik}}{T_{clk}}$) for any given arc i and associated non-controlling state k, along with the state probabilities (P_q^+ in eq. (5a)) for all gates of the design.

Since equations (5), (6) and (7) discretely describe the leakage and switching current densities, we can sum them to derive the effective average current density ($J_{avg,R}$) and add them in an RMS manner to derive the effective RMS current density ($J_{rms,R}$) for any resistor R in the cell.

We compute the average and RMS current densities by consolidating eqs. (5)–(7) as

$$J_{avg,R} = J_{avg_{sw},R} + L_{avg,R} \quad (8)$$

$$J_{rms,R}^2 = J_{rms,sw,R}^2 + L_{rms,R}^2 \quad (9)$$

It must be mentioned here that RMS formulations work under the assumption that the different current density (leakage and switching) are non-overlapping. This strictly is not true; however, we find that this assumption leads to very marginal errors. Next, we look at incorporating clock gating in the formulations.

B.4 Incorporating Clock Gating

Clock gating is a widely-used technique for reducing the dynamic clock power by disabling the clock signal to the idle parts of the circuit – thereby also directly affecting the reliability of the signals in the gated domain [28]. In order to assess the reliability impact of clock gating, we notice that as a phenomenon, clock gating can occur in an arbitrary way over the lifetime of the chip. For instance, the clock could be gated for a fixed number of cycles, after every specific period of activity, in a repeated manner. Such uniform gating is akin to a direct reduction in the operating frequency and can be readily approximated by specifying the activity-rate-adjusted frequency in above eqs. (8), (9).

However, the cases when the clock gating is non-uniform, or is uniform only in the intervals, are nontrivial and require equivalent reliability-lifetime calculations. The key determinant in such calculations is the thermal time constant of Joule heating in interconnect (typically in several microseconds for copper [30]), which signifies the duration after which the interconnect responds to the RMS current in the form of a temperature rise. Hence, if the time interval between successive clock gating events is larger than the thermal time constant, then, the full current (without activity correction), should be ideally used for RMS and average density estimations, for the appropriate durations.

We will defer treatment for non-uniform clock gating to Section V.B (subsequent to incorporation of arbitrary reliability specifications), and focus the formulations now only for the uniform case. This makes the solution similar to setting a pin specific activity rate on the cell. Hence, if a 1GHz clock tree element remains gated-high for 25% of the lifetime, we would note the corrected f_{ik} as 750MHz in eq. (8), (9), and state probability as 0.375 (assuming 50% duty cycle for clock). The computation procedure can thus be captured as:

Algorithm 1 Current density computation through every resistor

Input: SPICE setup (with all resistors), timing characterization setup
Output: $J_{avg,R_{ik}}$ and $J_{rms,R_{ik}}$

for each library cell; **for** every load/slew in the 8×8 matrix
  simulate for every legal input state combination (k)
    for each resistor R of the cell
      store average leakage density $L_{R,k}$ (eq. (5a))
    end
    for every legal switching scenario (arc i)
      for each resistor R, store $J_{avg,R_{ik}}$ and $J_{rms,R_{ik}}$
    end ## every arc
  end ## every state
end cell characterization
for each instance in the design
  estimate f_{ik}, P_q^+, s_{ik} for all input pins, arcs and states
  for each resistor R of the instance at chip level
    query-and-add $J_{avg,R_{ik}}$, $J_{rms,R_{ik}}$ and $L_{R,k}$ as in eq. (8) (9)
    store $J_{avg,R}$, $J_{rms,R}$ at given condition (f_{ik}, P_q^+, s_{ik})
  end ## for every resistor of the instance
end ## for every cell

C. Instance Safe Frequency Estimation at Chip Level

Once we have estimated the current densities in the cell, the EM checking procedure can subsequently be approached in two manners, as noted in Section II earlier:

- **Predict the safety of the cell** (pass or fail), given a full set of operating conditions of the cell.
- **Calculate a set of safe operating parameters** for the cell under a partial set of operating conditions. For example, if the frequency, slew and supply voltage are given, the safe load may be computed.

The first is rather trivially obtained from the above discussion, since eqs. (8), (9) and Algorithm 1 lend themselves readily to allow substitution of the exact operating conditions, and subsequent verification of current densities (through all resistors) against the foundry EM thresholds.

In real designs, however, the actual operating frequency of the instance could be arbitrary, and we must work the problem backwards by recommending a maximum f_{safe} based on other parameters. *In contrast to the f–L data of Fig. 2a obtained by iterated binary-search SPICE simulations, our approach here provides closed-form solutions for f_{safe}.*

It must be noted that potentially, every resistor in the cell could have unique frequency dependence, and therefore, the maximum f_{safe} procedure must find the minimum safe frequency over all resistors in the instance.

Let $J_{avg,th}(T,t)$ and $J_{rms,th}(\Delta T)$ represent the current density limits for average and RMS current densities respectively, as a function of stress temperature, stress time, and maximum heating constraint. Further, note that in eqs. (5)–(7), the dependence on the frequency $f = 1/T_{clk}$ appears only in the expressions for the average and RMS switching current densities. By setting the left-hand sides of eqs. (8) and (9) to be no larger than the threshold densities and combining them with eqs. (5)–(7), we can constrain the RMS or average-limited frequencies ($f_{max,AVG,R}$ and $f_{max,RMS,R}$, respectively) for each intra-cell resistor R in following manner:

$$f_{max,AVG,R} = \frac{J_{avg,th}(T,t) - L_{avg,R}}{\sum_{i=1}^{all\ arcs}\left(\sum_{k=1}^{all\ states} P_{ik} J_{avg,R_{ik}} s_{ik}\right)} \quad (10)$$

$$f_{max,RMS,R} = \frac{J_{rms,th}^2(\Delta T) - L_{rms,R}^2}{\sum_{i=1}^{all\ arcs}\left(\sum_{k=1}^{all\ states} P_{ik} J_{rms,R_{ik}}^2 s_{ik}\right)} \quad (11)$$

Since all parameters on the right-hand sides of the above equations are known for each resistor in each instance, we can now apply the self-consistent formulations [10] to estimate the safe parameter (frequency) of the resistor. The entire process has to be approached iteratively, as shown in Algorithm 2, to determine the safe operating frequency for an instance, which can be then used as a design constraint. The safe frequency for a resistor is the lower of the two values in eqs. (10) and (11) and the safe frequency f_{safe} for a cell instance is the smallest safe frequency over all resistors in the instance.

Algorithm 2 Self-consistent safe frequency estimation of the instance

Input: T_{clk}, $J_{avg,R}$, $J_{rms,R}$ from Algorithm 1 and the average and RMS Electromigration thresholds from foundry
Output: f_{safe} for every instance

for every instance of the design; set high f_{safe}
  for every resistor R in $\Re$
    start with a low estimate of $f_{max,RMS,R}$
    estimate ΔT for this RMS current (from eq. (4))
    while (($\Delta T < \Delta T_{limit}$) **and** ($J_{avg,R} < J_{avg,th}(T + \Delta T)$) **do**
      estimate $f_{max,AVG,R}$, using $J_{avg,th}(T + \Delta T)$ (eq. (11))
      $f_{max,RMS,R} = f_{max,AVG,R}$
      estimate ΔT
    end while
    ## found self consistent frequency for R
    $f_{safe,R}$ = **min** ($f_{max,AVG,R}$, $f_{max,RMS,R}$)
    if ($f_{safe,R} < f_{safe}$) $f_{safe} = f_{safe,R}$
  end ## for every resistor in $\Re$
  return f_{safe}
end ## for full design

To evaluate this procedure, we revisit the two-input clock tree mux from the earlier discussion around Fig. 3. Fig. 4 provides the f_{safe} plot for this case, for a fixed operating condition and output load, showing the results of binary-search-based SPICE simulation, our approach, and the traditional method that chooses f_{safe} pessimistically over all switching conditions. We see that the proposed model fits the SPICE behavior very well and can model the arbitrary switching rates on different pins, as against the large pessimism in the traditional approach.

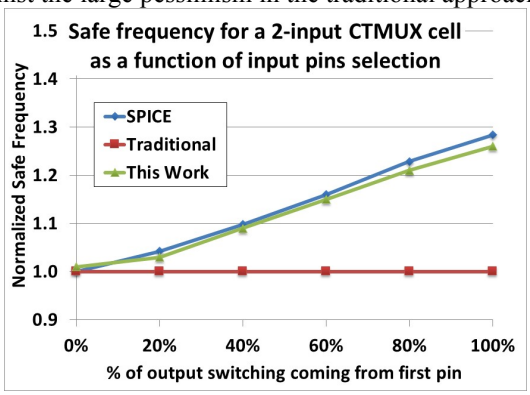


Fig. 4. Evaluation of f_{safe} for the circuit in Fig. 3, at a selected load point: 0.45x. The f_{safe} varies based on the extent of switching coming from the first or second pin. The proposed model completely captures the behavior, but the traditional is excessively pessimistic.

While the results shared in Fig. 4 were from a single cell, consolidated results from the entire 28nm design library will be shared later in Section VII. It must also be noted that thus far, we have demonstrated Black's equation (eq. (1)) based EM verification. However, as our methodology aptly decouples the current density computation and the verification part, it easily lends itself to other EM verification schemes, such as the via-node based scheme (Section IIA).

IV. Addressing L2: Modeling the Impact of Arbitrary RC Loading

The model developed so far is capable of covering following parameters: lumped capacitive load (C load), slews, multi-input gates, arbitrary switching rate and clock gating. This is directly relevant at the chip level, when the IPs are used at arbitrary frequencies and under clock gating. Next, we look at incorporating RC load into the assessment.

A. Overview of Prior Work

In the last section (Sec. III), we used the lumped load as one of the metrics for EM reliability. To a great extent, the C load model in itself can be used for accurate estimation of average current density, and is largely independent of resistive effects [3], provided the rail-to-rail swings for the output net.

On the other hand, RMS current densities additionally depend on the duration of transfer, and are thereby directly impacted by the resistive effects of RC loads on the cell [14-19]. The effect of RC load (Fig. 2b) for signal EM reliability was addressed earlier in [3]. It was further established that resistive shielding cannot be accounted using the traditional C_{eff} approach, derived from timing constraints. Hence, a current-criterion-based moment matching was devised to come up with a C_{eff}, by performing the RC tree traversal along with the basic timing information.

B. Prior Work: Limitations

We notice that there are at least two limitations of the prior work associated both with the traditional model (of Fig. 2a) and the model proposed in the Sec. III.

Firstly, RC loads affect the current flow in all segments: cell-external as well as cell-internal. While the cell-external problem was solved in [3], the cell-internal piece of the problem has remained unsolved. In fact, it was proposed to simulate the entire active network with the actual distributed load (at transistor level) through SPICE. As we will later see (Section VII), the number of such simulations required for a block/SoC could run in thousands, becoming a major computational and logistical overhead.

Secondly, we notice that not just the effective capacitance, but also the lumped capacitance depends on the current waveform shape — making the load capacitance itself as voltage-dependent. This can be explained by the fact that the pin capacitance has an inherent voltage dependence [20]. Hence, even though there is no explicit dependence of the average current on the network resistance [3], it implicitly exists because of the dependence on $C_{in}(V)$. Therefore, assuming a fixed value of C_{in} for performing current density calculations on a net becomes very pessimistic. We now attempt to solve both problems.

C. Proposed Solution: RC Loading and C_{in} Modeling

We begin by observing that the basic challenge for cell-internal EM arises from the fact that the characterization of the fundamental current densities (through Algorithm 1) must be performed using the single C load values, but the data must be applied to instances that drive RC loading. Hence, we require a good proxy of the RC load, which can be used to query the characterized data. Extending the concepts developed in [3], if we use C_{eff} to query only the RMS component of the current density from the precharacterized data, an accurate match can be achieved. Indeed, we do see a reduction in error (compared with SPICE) in this manner: as we will later see in Fig. 6, the mean error of about 2X in RMS estimation reduces to about 20% with the C_{eff} incorporation. However, there still are outliers and upon detailed investigation, a majority of them are attributable to C_{in} modeling of the load pins.

Next, as an improvement, we also compared the current densities derived from the case, when the load cells were modeled as a C1/C2 combination, where C1 represents the pin capacitance from 0-50% swing of the voltage, and C2 from 50-70% [21]. However, we notice that at an individual load pin level itself, this approach does not yield high accuracy due to ignorance of the tail effects [3, 22].

Hence, we propose calculating an effective C_{in} ($C_{in,eff}$) from the multi-piece $C_{in}(V)$ table (typically 8 points). Since C_{in} is a function of the voltage waveform, which in turn is a function of C_{in}, the entire computation must be carried out in an iterative manner. Accordingly, in the k-th iteration, we make use of the starting current waveform (as incident on the load cell L_i of Fig. 2b). Such a current waveform ($I_{L_{i,k}}(t)$), is obtained through a single $C_{in,k}$ and uses a double exponential model with estimated parameters – $A_{0,k}$, $T_{a,k}$ and $T_{b,k}$. The estimation of these parameters is performed by RC-tree traversal and moment matching technique with assumption on the waveform shape at the driver (a mixture of ramp/exponential) [3]. The current waveform is modeled as:

$$I_{L_{i,k}}(t) = A_{0,k}\left(e^{-t/T_{a,k}} - e^{-t/T_{b,k}}\right) \quad (12)$$

Subsequently, the voltage waveform $V_{L_{i,k}}(t)$, as seen on the load pin, can be generated as an area under the curve of this current waveform, using a constant $C_{in,k}$.

$$V_{L_{i,k}}(t) = \frac{1}{C_{in,k}}\int_0^t I_{L_{i,k}}(t')dt' \quad (13)$$

This voltage waveform can then be used along with the varying C_{in}: $C_{in}(V)$ table, to reconstruct a new current waveform $I'_{L_{i,k}}(t)$ as:

$$I'_{L_{i,k}}(t) = C_{in}(V)\frac{d}{dt}(V_{L_{i,k}}(t)) \quad (14)$$

Note that only an update in the current waveform at the load pin is required, since we are interested in the current specifically at this point. Assuming the duration of this current waveform as d (approximated by the 0-100% slew at the load pin obtained through STA), its RMS is given by:

$$RMS\ for\ I'_{L_{i,k}}(t) = \sqrt{\frac{1}{d}\int_0^d I'^2_{L_{i,k}}(t)\,dt} \quad (15)$$

Note that for the next iteration, we require an updated value for C_{in}. Hence, we make use of the RMS current through $I'_{L_{i,k}}(t)$, to derive a single effective C_{in}, assuming an equivalent triangular current waveform (with d being the delay at the load pin). For such a triangular waveform, the RMS current expression is standard: $\sqrt{\frac{4}{3}}\frac{CV}{d}$, where C is the equivalent load. In order to obtain an equivalent pin capacitance which can match the RMS current of $I'_{L_{i,k}}(t)$, we equate eq. (15) to the RMS current of triangular waveform, to get below capacitance (to be used for next iteration) as:

$$C_{in,k+1} = \frac{d}{V}\sqrt{\frac{3}{4}}\int_0^d I'^2_{L_{i,k}}(t)dt \quad (16)$$

We expect convergence in 2-3 iterations, though, for our work, we have made only a single update to starting C_{in}. In similar way, the average current case can be approached. While this means that we must ideally compute two separate capacitances: namely $C_{in,eff,RMS}$ and $C_{in,eff,AVG}$, our experiments indicate acceptable errors for the average case, and hence we do iterative computation only for RMS matching.

In summary, we accurately incorporate the impact of the voltage-dependent input pin capacitance as well as the impact of parasitic RC loading on the cell-internal currents, by:

- Making an initial estimate of the current at the driving points and iterating with the load's voltage-dependent pin capacitance to arrive at the final current flow.
- Estimating the effective capacitance (C_{eff}), which matches the final current flow in the network.
- Using this C_{eff} to query the precharacterized cell-internal RMS current density database and C load for querying AVG cell-internal current densities.

Note that in absence of this method, we would have used C load to query the cell-internal AVG as well as RMS current densities, which is very pessimistic. Note also that the formulations for incorporating voltage-dependent pin capacitance automatically improve the accuracy of cell-external current densities as well.

D. RC Loading and C_{in} Model Validation: Results

We perform validation at multiple levels of the RC and C_{in} modeling approaches. First, we validate the C_{in} approach at the load-level circuit, followed by the validation of the combined RC loading and C_{in} modeling in the driver-load pair case (Fig. 2b), followed finally by the results from several driver instances (driving unique RC loads).

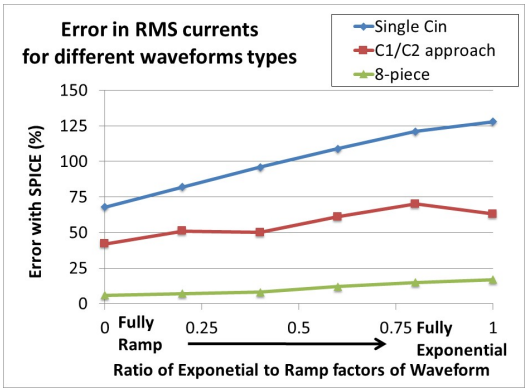


Fig. 5. Error in the RMS estimates (versus SPICE) for various Cin modeling approaches and waveform types (x-axis; going from fully ramp to fully exponential)

We begin by showing the load-level comparison first, for effective C_{in} estimation (versus SPICE) for different C_{in}

models (Fig. 5). This comparison is at the load circuit level, where we apply a voltage waveform at the load pin and the load cell is modeled as: a) single C_{in}, b) a two-piece voltage dependent capacitor in SPICE [23], and c) a single $C_{in,eff}$ (obtained from eq. (12)–(16)).

Moreover, as discussed earlier, since C_{in} is a function of the starting input voltage waveform, we have computed the errors for different types of input voltage waveforms (the x-axis represents waveforms going from fully ramp to fully exponential), whose shape is controlled with the coefficient a in below equation, with T_r being the rise time:

$$V_{in}(t) = \begin{cases} a\left(1 - e^{-t/T_r}\right) + (1-a)\frac{t}{T_r}, & t \le T_r \\ 1 - ae^{-t/T_r}, & t > T_r \end{cases} \quad (17)$$

Hence, setting a to zero in above equation, results in a fully saturated ramp input waveform, whereas setting a to unity makes it complete exponential. Such a formulation is a good representation of the various input waveforms which can be incident on the load pin.

As we can see from Fig. 5, the traditional approach of single C_{in} leads to almost 2X error as compared to SPICE. The error reduces using the C1/C2 model, but it still remains unacceptable and the effective capacitance computation approach from an eight-piece piecewise-linear table fits the SPICE results in a better solution. We can also see that because of increased tail effects in the exponential input voltage waveform, the errors are higher for all models for a completely exponential case.

Fig. 6 shows the maximum error from several instances (which drive different RC loads; plotted on the x-axis). While the left y-axis shows the errors, the *C_{eff}/C-load* ratio, an indicative of the extent of resistive load the instance is driving, is plotted on the right y-axis. The exact set of instances and their driving RC load information is obtained from a 28nm production design. We show the comparison of: the traditional case (using lumped load for current density querying), C_{eff} model alone and the combined C_{eff} + C_{in} model.

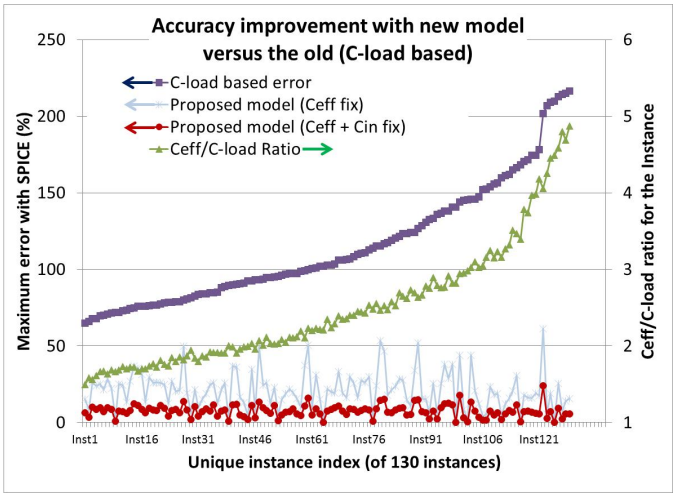


Fig. 6. Maximum error in RMS current density estimation across several instances driving different kinds of RC loading (indicated by the C_{eff}/C-load ratio) at the design level.

Overall, amongst all cases, we find about 2X mean error in RMS current density estimation with the usage of lumped load, which drops to about 21% mean with the usage of C_{eff} model, and further down to about 7% mean error with the combined usage of C_{eff} and C_{in} model. We also see that for instances driving severely resistive loads (indicated by the ratio of C_{eff} to C load), the original error with C load usage is very high, with outliers that cross 50% error.

The final algorithm for estimating the accurate cell-internal currents for arbitrary loading is below:

Algorithm 3 Accurate EM verification considering RC loads

Input: SPEF, Capacitance(Voltage) for all cells
Output: Pass/Fail results for instances after EM verification

```
for every instance in the design
    obtain the timing information (load, slew and the SPEF)
    for every load cell, L_i, of the instance, in k^th iteration
        set C_in,k, compute I_Li,k(t), using eq. (12)
        use I_Li,k(t) to construct I'_Li,k(t) using C_in(V) eq. (12-16)
        recalculate C_in,eff and iterate till acceptable accuracy
    end
    compute network C_eff by using C_in,eff for every load [3]
    for every resistor R of the instance
        use C_load to query average, C_eff to query RMS densities
        verify current densities against thresholds;
    end; flag pass if all resistors pass
end for every instance
```

Thus, we have examined the impact of RC loading on the EM reliability, and demonstrated significant improvement in accuracy with the proposed method.

V. Addressing L3 – On-The-Fly Retargeting of Reliability for Arbitrary Specifications

The formulations of Sec. III and IV were dependent on the library data, characterized at one set of operating conditions, and the foundry EM thresholds at a specified reliability condition. However, as described in Section I, there is an increasing need for on-the-fly reliability retargeting, at design verification stage, as the IP library is used under different reliability conditions. As noted earlier, meeting this goal is impractical under the traditional methodology, as it requires a new characterizations of the entire IP library (Fig. 2a) at each new condition.

The core methodology of this work enables the ability to perform this retargeting efficiently, since the current density computation part is separated out from the verification part (whereas these are tightly coupled in the traditional approach). We begin with the fundamental relation between EM lifetime and the lognormal variable. From eqs. (1), (2), taking logarithm, we obtain:

$$\sigma z = \ln(t_f) - \ln(A) + n\ln(J) - \frac{Q}{k_B T} \quad (18)$$

Now, if we have two different sets of stresses, denoted by subscripts a and b, each is described by the same fitting parameter A, but other terms in eq. (18) may differ. Naturally, their reliability is related as follows (by substituting the parameter A):

$$\sigma z_{cond,b} = \sigma z_{cond,a} + ln\left(\frac{t_b J_b^n}{t_a J_a^n}\right) - \frac{Q}{k_B}\left(\frac{1}{T_b+\Delta T_b} - \frac{1}{T_a+\Delta T_a}\right) \quad (19)$$

Here, the variables t, J, T, and ΔT represent the stress time, current densities, stress temperature and Joule heating,

respectively, while the subscripts a and b refer to the two different conditions.

This equation is a powerful representation of the scaling factors that can either be used to assess a) required tradeoffs in new reliability conditions to meet the same fail fraction levels or b) the actual fail fractions at the new reliability conditions. For example, we can directly use above equation to find the equivalent stress time (t_b) that causes the same reliability loss as benchmark condition, but with increased current densities. In order to do so, we must set $z_{cond,b} = z_{cond,a}$, since the reliability loss has to be equated and obtain the equivalent lifetime t_b as a function of (t_a, J_a, and J_b). Obviously, if $J_b > J_a$, t_b will be estimated to be lower than t_a.

We now look at the application of the retargeting concepts, based on eq. (19), to some of the case studies, followed by application to non-uniform clock gating. Unlike uniform clock gating, which was previously treated with generic activity reductions in section IIIB, non-uniform clock gating requires a more accurate sliding window based analysis, wherein, every frame potentially becomes a new reliability condition.

A. Case Studies Incorporating Reliability Retargeting

Case I: Variations in Temperature If the use temperature and/or POH specification are different from the original conditions, then it is straightforward to address this by using eq. (19) to determine new current density thresholds, and then updating f_{safe} in eq. (11). Such a modification only affects the average, and not the RMS reliability.

A second situation is the common industry scenario when the stress profile is provided by the user as a temperature profile, as the series $\{(J_1,T_1,t_1), (J_2,T_2,t_2), \ldots , (J_m,T_m,t_m)\}$, *i.e.*, from time t_{k-1} to time t_k, it experiences current stress J_k at temperature T_k. If the baseline stress is characterized for J_0 at temperature T_0, then can relate the k^{th} stress vector to the baseline stress at (J_0,T_0) with an equivalent stress time $t_{k,0}$. In other words, the stress at temperature T_k is transposed to an equivalent stress time at temperature T_0. Consequently, our stress retargeting scheme will map the entire stress to $(J_0,T_0,t_{eq,0})$, where $t_{eq,0} = \sum_{k=1}^{m} t_{k,0}$.

Case II: Variation in Operating Voltage If the eventual use voltage of the library is different from the characterization voltage, current scaling must be performed. Such a scaling is straightforward in our framework, since the leakage and switching related components are separately stored, as described in eqs. (6), (7). Based on our experiments, we see that a linear scaling works very well for voltage scaling, while an exponential model is required for leakage. Note that this scaling must be performed for every discrete component of the current densities for every resistor in the circuit (eq. (6), (7)).

A second situation (arising due to power management scenarios like dynamic voltage frequency scaling (DVFS) [24]), is when the voltage is represented as a series: $\{(V_1,t_1) \ldots ((V_m,t_m)\}$. In such a case, we can follow the scaling procedure to obtain a series of currents, which can then be dealt in the same way as the earlier case.

Case III: Variation in Failure Rate Specification The $J_{avg,th}(T,t)$ in eq. (11) is really a function of the fail fraction FF, which in turn is a function of z (eq. (2)), wherein, z_{target} is the inverse function of FF_{target}. Hence, if the FF specified by the end user changes from, say, 0.1% to say 0.01% cumulative, it can be readily translated to z, translated to a current density limit using eq. (19), and then used in eq. (11) for verification.

Fig. 7 shows a representation of such a retargeting using the proposed model from a representative cell. For exposition ease, we represent our model at a fixed slew, as in Fig. 2a.

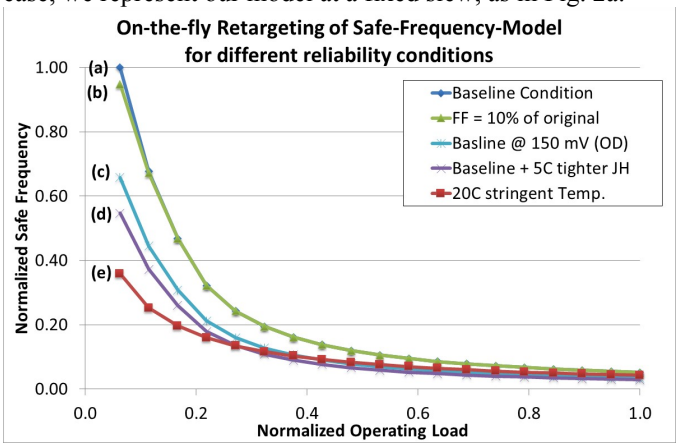


Fig. 7 Demonstrating on-the-fly retargeting of the basic frequency-load curve (Fig. 2a) with changes in the constraining criteria (at a fixed slew point)

Curve (a) represents the reliability at the baseline condition. If the FF requirement of the design changes and drops to 10% of the original, the curve slides down to (b) due to reduction in EM capability at tighter FF requirement. The drop is not drastic as this specific IP is RMS-current-limited, rather than being limited by the average current density. Similarly, if the use voltage has a 150mV overdrive over the characterized value, the reliability is represented by curve (c), which shows degraded reliability due to increased current flow. Similar behavior is seen in curve (d) if the Joule heating (RMS current density specification) is tightened by 5C. Finally, if the temperature requirement becomes 20C higher, design closure becomes more challenging with the reliability being now represented by the curve (e) – almost 3X tightening.

The case study in Fig. 7 is handled very naturally in our approach. We reiterate that handling them in the traditional approach require a complete recharacterization of the f_{safe} model at various conditions.

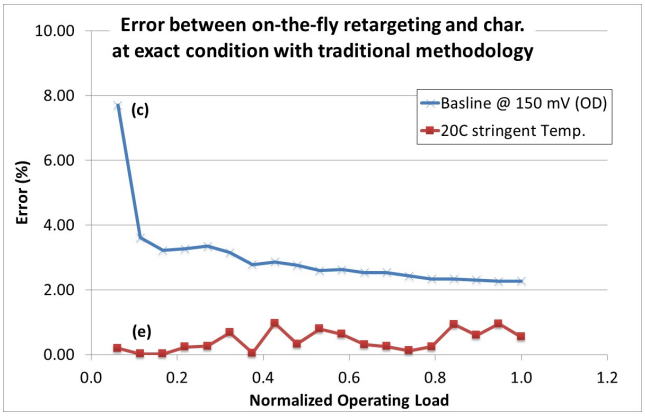


Fig. 8 Validation of retargeting methodology versus SPICE for two conditions, (c) and (e), of Fig. 7.

Next, to validate the retargeting methodology, we directly compare the curves of Fig. 7 to the curves of traditional methodology (obtained by the actual characterization at the

exact condition). As earlier, we present results from a single representative cell.

We show the percentage error for two conditions, (c) and (e) in Fig. 8. For (e), where the temperature specification is altered, the required retargeting only affects the verification part (as the current density limits are scaled), which incurs little error. For (c), the retargeting is due to 150mV overdrive, where we use a more approximate current-scaling model. The error here, although high, is acceptable, considering the fact that it is in a lower load regime (usually a low-current, EM-safe zone).

B. Incorporating Non-uniform Clock Gating

The case studies of previous subsection were helpful in outlining a general thought process on approaching the problem, when the eventual use scenario is different from the baseline one. We now consider the extension of those principles to the problem of clock gating, which was previously (Section IIIB.3), analyzed under uniform assumption. In order to do so, a key input required is an activity profile of the design over several clock cycles, as shown in Fig. 9 [25]. We noted previously that the thermal time constant is a key determinant in addressing the non-uniform clock gating [26, 27], and any change in current profile (of a larger duration), should be handled individually, and cannot be combined as a time-weighted summation.

Hence, we follow a sliding window approach (with the duration as the thermal time constant), wherein the complete clock activity profile is scanned in a step-by-step manner. For every single time window scanned, we compute the *effective* activity rate, which can then be used to compute the cell-internal current densities through every resistor, arcs and states of the cell. Eventually, for a resistor R, in the i-th arc and k-th state, we can represent the current densities as: $\{(J_{avg,R_{ik},0}, T_0), \ldots . (J_{avg,R_{ik},m}, T_m) \ldots\}$, where the index m refers to the index of the sliding window. Clearly, every window can have a unique activity rate. For instance, in Fig. 9, the sampling windows S_a, S_b, S_c and S_d correspond to a 75%, 50%, 100% and 67% activity rate respectively. This current stress can then be collapsed into a single equivalent stress, based on the concepts developed earlier and using eq. (19).

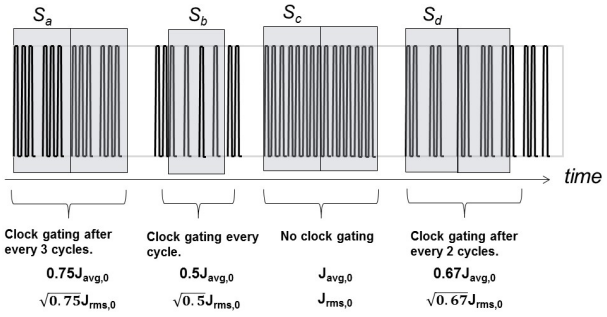


Fig. 9. Representative clock activity profile for a large duration. Different sampling windows show different activity rates (and corresponding J_{avg}, J_{rms}).

Indeed, for a variety of examples considering clock gating, we can notice a significant difference in the reliability. Fig. 10 shows the normalized reliability (of the clock tree element), based on the extent and the nature of the clock gating.

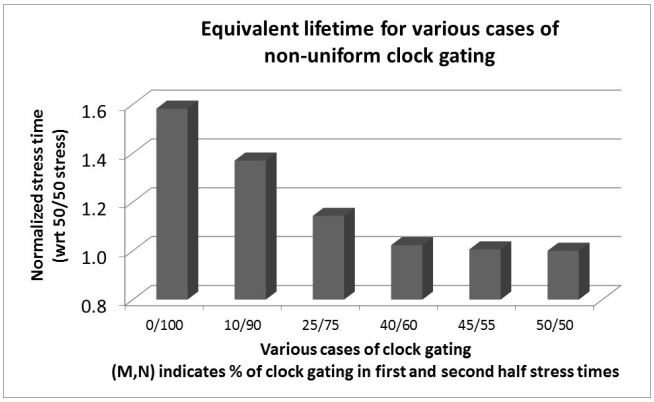


Fig. 10. Variation in reliability based on the extent of uniform clock gating in first half and second half of the stress time.

For this experiment, we considered a single clock tree element, which underwent different kinds of clock gating, though all amounting to a net 50% duration of gating in the chip lifetime. For example, the third bar in the Fig 10 corresponds to a case where the clock remains 25% *uniformly* gated (meaning gated every one in four cycles) the first half of stress time, followed by 75% gated in the other half; and so on, for the other cases. After reliability computations based on eq. (19), we plot the equivalent stress times for all cases, considering the (50%,50%) case as the baseline.

As we can see, for the same clock gating duration amongst all cases (50%), the worst case reliability occurs for the case in which the full-throttle events are clustered together – thereby meaning maximizing the average current, as well as JH together. On the other hand, if the clock gating is completely uniform, the JH is lowered, causing the least reliability loss. For the sake of completeness, we also note that for the same case, a free running clock corresponds to an equivalent lifetime of about 3X as compared to the (50%,50%) case. Thus, we can capture the algorithm to incorporate the exact clock gating impact in following way:

Algorithm 4 Incorporating non-uniform clock gating

Input: Clock gating profile, characterized cell and timing info.
Output: $J_{avg,R}$ and $J_{rms,R}$

1. **for each** clock tree instance in the design
2. **obtain** timing information: free-running frequency *f*, slew *s*
3. **for every** $window_m$ of clock profile
4. **compute** activity rate for the $window_m$ and reuse *f*, *s*
5. **for each** resistor *R* of the instance
6. **query-and-add** various current components (8), (9)
7. **find** equivalent time for *R*'s $window_m$ stress at baseline condition (free running); $t_{eq,R,m,0}$; **add** to $t_{eq,R}$
8. **end;**
9. **end;** all windows
10. **for each** resistor *R* of the instance
11. **use** $t_{eq,R}$ to find resistor pass/fail
12. **end;** report pass for all *R* passing
13. **end**

It must be mentioned that such a profiling data maybe hard to come by in real designs. Therefore, in absence of information, it is recommended to either assume no clock gating, or assume clock gating in the non-uniform manner.

VI. Addressing L4: Accelerated Data Generation Using Cell Response Modeling

Having looked at the various determinants of cell-EM reliability and ways to incorporate them in our model, we now look at expediting the characterization process. As discussed in Section IIIC, for the traditional methodology, the safe frequency estimation requires 640 SPICE simulations per cell. Indeed, complete data generation for a production 28nm library, consisting of a few thousand cells, can run into days of effort. Such high runtimes for just a single baseline reliability condition make the process of EM characterization prohibitive under the traditional model. Although the efficiencies suggested earlier in this work can greatly reduce this overhead, it is still essential to use the baseline operating condition and characterize the current densities using eqs. (6), (7): a process that can be very compute-intensive when carried out for all load/slew conditions. Hence, we must optimize the characterization process and in this work, we use response modeling approach.

In the retargeting discussion of Sec. V, we noticed that the traditional methodology is inflexible, as it **commingles the processes of current computation and EM verification**. For the same reason, it also does not lend itself for application of response modeling. The challenge here is twofold:

- From the circuit point of view, operating parameters such as load and slew non-uniformly affect the individual RMS and average resistor **current densities** in various arcs and states.
- At the same time, the reliability specifications like lifetime and fail-fraction requirements non-uniformly influence the average and RMS **current-density thresholds**.

Both of above eventually cause the average and RMS-limited frequencies (discussed in the self-consistent estimation in Algorithm 2) to be asymmetrically impacted, thereby making the traditional *frequency-level* abstraction as non-scalable across load/slews (illustrated for example in Fig. 11) and reliability specifications (discussed earlier in Fig. 7).

On the other hand, a key feature of our approach is to keep characterization and verification disjoint, which presents an opportunity for model building during the characterization phase and accelerate the data generation process.

As noted in Section IV earlier, average current flow through the resistor is purely a function of the total charge transferred (lumped load), while the RMS current density also has an inverse relationship with slew [3]. Based on these observations, we attempt to model the current density through any given resistor in the IP as a polynomial function of output loads/inputs slews:

$$J_{R_j} = a_0 + b_1 L + b_2 L^2 + c_1 s + c_2 s^2 + d_1 Ls + d_2 L^2 s^2 \quad (20)$$

Here, a_i, b_i and c_i are fitted coefficients, and L and s are the loads and slews, respectively. We identify seven critical points (in the 8x8 load/slew matrix) that help shape up the polynomial model: the four corners, (1,1), (1,8), (8,1), (8,8), and a few internal points (2,4), (4,4) and (4,2), where the indices represent the index of the load and slew, respectively, in the table. The parameter fitting is then performed, based on eq. (20), providing a model to predict the current densities at any arbitrary load/slew point. Note that the response modeling must be performed for every current density component (of eqs. (6)–(9)) of the resistor R. The number of models corresponds to the total number of unique arcs and states of the cell. For example, for a single input clock-tree inverter, we would require a total of four simulations: two to cover the arcs (input rise to output fall, and vice versa), and two to cover the static states (input high and input low).

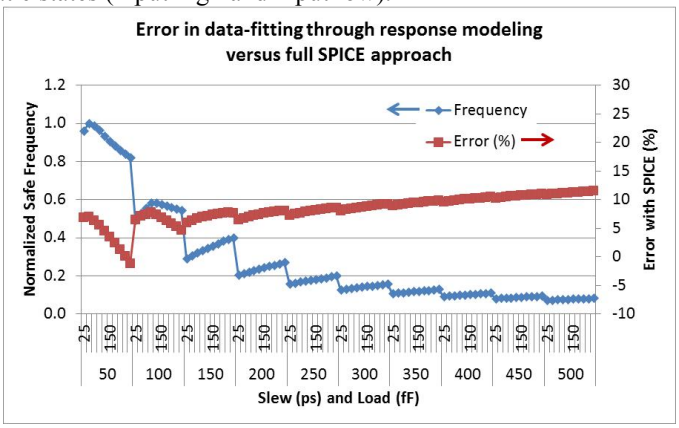


Fig. 11. Comparison of the response modeling approach (eq. (14)) with full SPICE (red). f_{safe} obtained through response modeling (blue)

We now examine the validation of the response model for a representative IP cell in Fig. 11; the results from the entire library will be presented in an end-to-end manner in the next section.

For various load/slew points on the x-axis, we first develop the characterization data based on full SPICE simulations (using eqs. (6), (7)). Subsequently, the model from eq. (20), is built using the simulation data from seven sampled points, and later evaluated at each of the 64 load/slew points. The normalized f_{safe} is plotted on the left y-axis, and the error between model and SPICE, on the right y-axis.

The non-monotonic behaviour of f_{safe} with load/slew can be readily observed from this plot. Such non monotonicity arises from the fact that at different load/slew indices, the metal segments which limits the EM performance of the cell varies. For example, at a fixed load condition (say 100fF), a lower input slew (~ 25 ps) would mean large RMS current in the output signal resistors, while a smaller short-circuit current in the power-ground resistors. On the other hand, a higher input slew (~ 200ps) means vice versa. Thus, for sharp input slews, the output signal resistors may often limit the cell reliability (due to RMS constraint), while, at the sluggish slews, the cell-internal power-ground resistors may be limiting (due to the average constraint). Such an interplay finally leads to a non-monotonic f_{safe} behaviour of the cell with load/slews.

Our methodology, however, works only at the current density level, and hence, remains unaffected by the reliability constraints which bring in the non-monotonicity. Using the representation from eq. (20) (for every resistor, per arc state), we can readily obtain the current densities at the chosen load/slew condition, and can subsequently, use those to compute the safe frequency of the cell by using Algorithm 2, which additionally requires the reliability condition. Consequently, we can cover all the load/slew points to get the safe frequency plot, and as we can see, the response modeling approach works reasonably well in predicting the current densities and f_{safe}, with an acceptable marginal error. Next, the runtime impact for a single cell is summarized in Table 1.

Methodology	Simulations required per cell	Overall compute
Traditional (Fig. 2a)	~640×*n* (*n*: number of design/reliability conditions)	~10 mins. per cell
Proposed (eq. 20)	28 (7 load-slew×4 unique sims)	~ 50 sec

Table 1. Runtime comparisons with proposed and traditional methods, for a single cell

As we see, the characterization runtime for our approach drops down significantly as compared to the traditional methodology, which takes ~10 minutes to generate the safe frequency data for all load/slew points, whereas the proposed methodology (response modeling) is completed in about a minute. It must be mentioned here that the number of simulations required in the traditional methodology grows linearly with the number of design/reliability conditions required (as discussed in Section V). For example, every change in the voltage, stress temperature or lifetime requires a new characterization. On the other hand, the proposed methodology comprehends the design/reliability conditions on-the-fly, using the same database (Section V), hence further keeping down on the number of simulations required. Thus, the methodology in this section directly addresses the limitation L4 and significantly speeds up characterization.

VII. Production Design Analysis

We now examine the final application of the proposed methodology in an industrial scenario, discussing the setup and the workflow. A 28nm high performance block (2mmx2mm; ~600K instances, >10M transistors), operating at 1GHz clock frequency is taken, which is part of a large industrial SoC. The entire flow is outlined in Fig. 12 for the proposed method.

The new method, in essence, is a three-step process: (a) IP characterization at a baseline reliability condition, (b) determining the reliability constraints for this design and (c) integration into the timing/implementation tool. Note that the true retargeting flexibility of the proposed approach comes in form of (b), which is a runtime-level input to verification that is completely detached from (a). The flow of (c) uses a standard industrial design methodology.

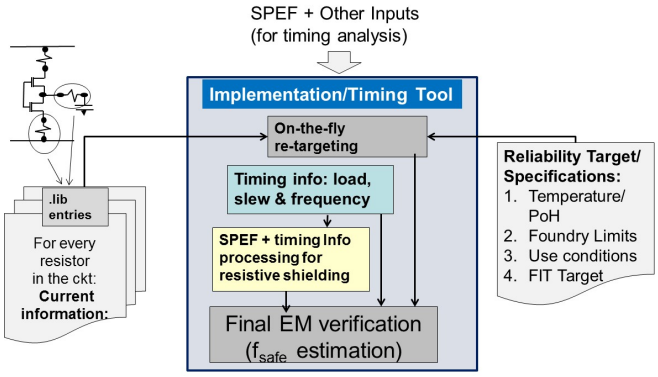


Fig. 12. Overall methodology and data-flow diagram for the proposed method

A. Library Characterization

The entire library of a few thousand cells was characterized in two ways: (a) a full SPICE-based approach, where the traditional f_{safe} table was generated at a baseline condition, and (b) the methodology proposed in this work. Parallelized and multithreaded SPICE simulations (using Cadence Spectre) were used. The runtime for (a) was about 800 CPU hours of raw simulation, excluding extraction, whereas the methodology in (b) completes in about 80 CPU hrs. For (b), the production characterization framework for timing was used to arrive at the various arcs and logical states for switching and leakage current characterization.

B. Final Reliability Verification

The final application of the library-generated data was performed in the timing tool (Encounter Timing System), through a custom developed scriptware, which reads in both the characterization data types. The timing analysis of the design was performed at the baseline condition, to arrive at the slews and probabilistic switching rates through all the input pins. In the traditional approach, the scriptware steps through the timing information of every instance in the design and compares the queried f_{safe} (from the traditional model) to the operating frequency. Note that since this approach suffers from the problems discussed earlier (specifically, L1 and L2), a final full SPICE simulation (with the RC loading of the driver instance) is required after the initial results from the frequency comparisons. A total of about 600K instances were analyzed in this way, and finally, the instances with the frequency ratio > 1 (around 4500), were simulated further. The excitations for the SPICE simulations were a simple 1010 transition (at operating frequency), since all the instances were single input clock tree inverters, buffers and gater cells (only eight unique cells). The final set of violations after the full SPICE simulations came down to 426.

On the other hand, in the new approach, the scriptware additionally implements Algorithms 1-3, and based on the reliability specifications (lifetime/temperature/voltage/fail fraction), the equations are updated on-the-fly for the final frequency comparison of every instance.

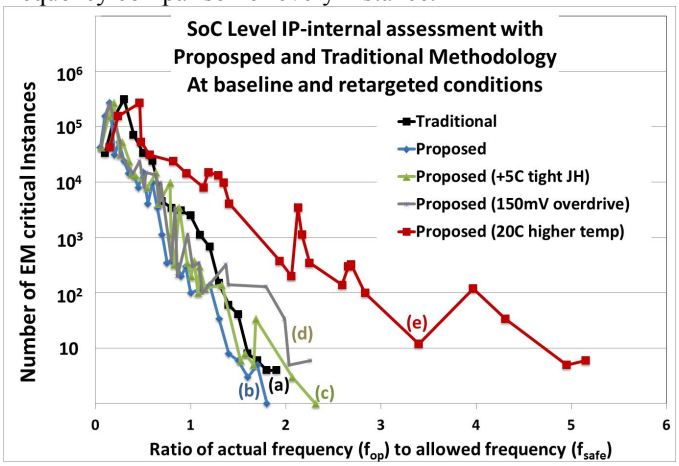


Fig. 13. Distribution plot for a 28nm block (>600K instances), highlighting the number of EM-critical instances and violations (with f_{op}/f_{safe} ratio > 1) for a), b) baseline reliability analysis with traditional and proposed methods; c), d), e): retargeted reliability condition analysis with proposed methodology.

Finally, we plot the population distribution of frequency ratios in Fig. 13. We consider five cases: a), b) corresponding to analysis with traditional and proposed methods at baseline reliability conditions respectively and c), d), e) corresponding to the analysis at *retargeted* reliability conditions of a tighter JH limit, an overdrive case and a high temperature requirement, respectively.

For every method, we plot the ratio of f_{op} to f_{safe}, which signifies the EM criticality for that instance. Hence, an instance with f_{op} greater than f_{safe} (red region in the plot) is deemed as EM failure and must be acted upon for fixing (either by load reduction or replacement). The y-axis shows the distribution of number of instances in design with a particular f_{op}/f_{safe} ratio. We document the total number of violations from various analyses in Table 2.

As we can see from Fig. 13 and Table 2, the proposed approach reports a total of 442 violations, 421 of which overlaps with the traditional methodology (+ SPICE). The remaining: false (21) and escaped (5) violations from the new approach were found to be relatively less critical, with frequency ratios in the range of 1.14 to 0.9. Thus, the new approach agrees well with SPICE.

Analysis Type	Reliability Condition	Violations
Traditional (SPICE)	Baseline	426
Proposed Methodology	Baseline	442 (421)
	Baseline + 5C tighter JH	1297
	Baseline + 150mV overdrive	1093
	Baseline + 20C higher temperature requirement	56945

Table 2. Overall comparison of traditional versus proposed methodology. Traditional method was run only at baseline condition due to runtime issues, whereas the proposed method could run at various reliability conditions.

Next, we demonstrate the final retargetability of the proposed approach is evident by the curves c), d) and e) in Fig. 13, analyzed at retargeted reliability conditions. Run c), corresponding to an additionally tight constraint of 5C lower JH, results in almost 3X increased violations, due to tighter RMS limits. Run d), which is at overdrive conditions results in a similar violation profile. However, run e), which corresponds to a 20C higher stress temperature run results in a plethora of violations. Such a run is a close proxy to a direct application of IPs meant for handheld businesses to harsher environments!

Finally, based on the stage of the chip-design execution, design community has multiple ways to act upon this EM verification feedback. Although a detailed solution to developing EM fixes is beyond the scope of this paper, we provide some pointers in the rest of this paragraph. In many cases the harshness of reliability criterion softens due to a lower lifetime requirement – for instance, in infotainment category chips [29]. Alternatively, an avoidance strategy can be followed upfront, wherein, based on the logic, high drive-strength cells are used to drive large fanout points. However, this requires careful consideration since unwarranted improvement in drive-strength is associated with sharp output-slew reduction resulting in increased RMS currents. On the other hand, a forceful lowering of drive-strength for instances with timing slack causes slew degradation resulting in increased short-circuit currents. A better approach may be through fanout-load or activity reduction, which predictably reduces the current flow.

VIII. Conclusion

In summary, an accurate and retargetable methodology for IP-internal EM verification was presented in this work. Generic switching rates for various pins of the IP are comprehended, including aspects of clock gating. Significantly high accuracy, with respect to SPICE, was achieved by incorporating the impact of arbitrary parasitic loading, and, an intelligent way of coming up with the effective pin capacitance of load cells. The methodology was shown to be highly flexible, in terms of allowing on-the-fly retargeting for the reliability. Finally, the complete data generation process at library level is expedited by application of cell response modeling. Results on a 28nm production setup were shared, to demonstrate significant relaxation in terms of violations, along with close correlation to SPICE. We shared various cases of runtime-level reliability retargeting, by specifying varying reliability conditions for the production block verification. The methodology presented in this work is most suitable in a third-party-IP context. The need is only underlined further with the increasing porting of designs from one business segment to a different one, which requires on-the-fly assessment of the reliability of all the components.

References

[1] J. Lienig, "Electromigration and its impact on physical design in future technologies," in *Proc. of the ACM International Symposium on Physical Design,* 2013, pp. 33-40.

[2] X. Huang, T. Yu, V Sukharev, and S. X-D. Tan, "Physics-based Electromigration Assessment for Power Grid Networks," in *Proc. of the ACM/IEEE Design Automation Conference*, 2014, pp. 1-6.

[3] P. Jain and A. Jain, "Accurate current estimation for interconnect reliability analysis," *IEEE Transactions on VLSI Systems*, vol. 20, no. 9, pp. 1634-1644, Sept. 2012.

[4] "ITRS Interconnect Summary, 2013": http://www.itrs.net

[5] C. Yuan, D. Tipple, and J. Warner, "Optimizing standard cell design for quality," in *Proc. of the SPIE Design-Process-Technology Co-optimization for Manufacturability 9053*, 2014.

[6] "JEDEC reliability standards, JC14," http://www.jedec.org

[7] "AEC Q100 specification, AEC-Q003 Rev-A": www.aecouncil.com/AECDocuments

[8] "Encounter Design Implementation tool user manual, 2014,": http://www.cadence.com/products/di/edi_system/pages/default.aspx

[9] "Magma-Talus Implementation tool user manual, 2014": http://www.synopsys.com

[10] W. Hunter, "Self-consistent solutions for allowed interconnect current density. II. Application to design guidelines," *IEEE Transactions on Electron Devices*, vol. 44, no. 2, pp. 310-316, 1997.

[11] J. Black, "Electromigration failure modes in aluminum metallization for semiconductor devices," *Proc. of the IEEE*, vol. 57, no. 9, pp. 1587-94, Sept. 1969.

[12] K.-D. Lee, "Electromigration recovery and short lead effect under bipolar-and unipolar-pulse current," in *Proc. of the IEEE International Reliability Physics Symposium*, 2012, pp. 6B-3.

[13] "Altos tool user-manual, 2012," http://www.cadence.com

[14] S. S. Sapatnekar, *Timing*, Springer US, New York, NY, USA, 2004.

[15] S. P. McCormick, "Modeling and Simulation of VLSI Interconnections with Moments." Ph.D. Thesis, MIT, Cambridge, MA, USA, 1989.

[16] J. Qian, S. Pullela, and L. Pillage, "Modeling the effective capacitance for RC interconnect of CMOS gates," *IEEE Transactions on Computer Aided Design*, vol. 13, no. 12, pp. 1526-35, Dec. 1994.

[17] C. Wang and D. Markovic, "Delay estimation and sizing of CMOS logic using logical effort with slope correction," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 56, no. 8, pp. 634-638, Aug. 2009.

[18] J. Croix and D. F. Wong, "Blade and Razor: cell and interconnect delay analysis using current-based models," in *Proc. of the ACM/IEEE Design Automation Conference*, 2003, pp. 386-389.
[19] S. Gupta and S. S. Sapatnekar, "Compact current source models for timing analysis under temperature and body bias variations," *IEEE Transactions on VLSI Systems,* vol. 20, pp. 2104-2117, 2012.
[20] D. Sinha and S. Abbaspour "Method of employing slew dependent pin capacitances to capture interconnect parasitics during timing abstraction of VLSI circuits," U.S. Patent 8103997, 2012.
[21] B. Mullen, "CCS-Timing: Composite Current Source Delay Modeling", in *Proc. of DAC* 2005.
[22] D. D. Ling *et al.,* "A moment-based effective characterization waveform for timing analysis," in *Proc. of the ACM/IEEE Design Automation Conference*, 2009, pp. 19-24.
[23] "HSPICE User Manual, 2015," http://www.synopsys.com
[24] S. Herbert and D. Marculescu, "Analysis of dynamic voltage/frequency scaling in chip-multiprocessors," in *Proc. of the ACM International Symposium on Low Power Electronics and Design*, 2007, pp. 38-43.
[25] "PTPX power estimation tool, 2014": http://www.synopsys.com
[26] J. Choi *et al.* "Thermal-aware task scheduling at the system software level," in *Proc. of the ACM International Symposium on Low Power Electronics and Design*, 2007, pp. 213-218.
[27] Y. Zhan, S. V. Kumar, and S. S. Sapatnekar, "Thermally-Aware Design," *Foundations and Trends in Electronic Design Automation*, vol. 2, no. 3, pp. 255-370, Oct. 2008.
[28] A. Todri and M.-S. Malgorzata, "A study of reliability issues in clock distribution networks," in *Proc. of the IEEE International Conference on Computer Design*, 2008, pp. 101-106.
[29] "Automotive Processors Overview, Aug. 2015" http://www.ti.com/lsds/ti/processors/dsp/automotive_processors/overview.page
[30] K. Banerjee and A. Mehrotra, "Coupled analysis of electromigration reliability and performance in ULSI signal nets," in *Proc. of the ACM International Conference on Computer Aided Design*, Nov. 2001, pp. 158-164.
[31] Y.-J. Park, P. Jain, and S. Krishnan, "New electromigration validation: Via Node Vector Method", in *Proc. of the IEEE International Reliability Physics Symposium*, 2010, pp. 698-704.
[32] S. Alam *et al.* "Circuit level reliability analysis of Cu interconnects," in *Proc. of the International Symposium on Quality Electronic Design,* 2004.
[33] Z. Guan, *et al.* "Atomic flux divergence based current conversion scheme for signal line electromigration reliability assessment," in *Proc. of Interconnect Technology Conference,* 2014.
[34] K.-D. Lee, "Electromigration Critical Length Effect and Early Failures in Cu/oxide and Cu/low k Interconnects", Ph.D. Thesis, University of Texas at Austin, Austin, TX, USA, 2003.

Palkesh Jain (M'04) Palkesh Jain graduated from the Indian Institute of Technology Bombay, in 2004 with Bachelors and Masters in Electrical Engineering. He joined the ASIC group at Texas Instruments India, where he defined and developed, several of the GHz enabling reliability methodologies. Subsequently, he joined the Yield and Product Engineering team at Qualcomm India in 2014, where he is involved with system level power and thermal management methodologies. He holds 15 US patents (granted/ pending) and is also a part time doctoral student at the Universitat Politècnica de Catalunya.

Jordi Cortadella (S'87-M'89-F'15) is Professor of the Computer Science Department at the Universitat Politècnica de Catalunya. He is a Fellow of the IEEE and member of the Academia Europaea. His research interests include formal methods and computer-aided design of VLSI systems with special emphasis on asynchronous circuits, concurrent systems and logic synthesis. Prof. Cortadella has served on the technical committees of several international conferences in the field of Design Automation and Concurrent Systems and is associate editor of the IEEE Transactions on CAD of Integrated Circuits and Systems. He received best paper awards at the Int. Symp. on Advanced Research in Asynchronous Circuits and Systems (2004), the Design Automation Conference (2004) and the Int. Conf. on Application of Concurrency to System Design (2009).

Sachin Sapatnekar (S'86, M'93, F'03) received the B. Tech. degree from the Indian Institute of Technology, Bombay, the M.S. degree from Syracuse University, and the Ph.D. degree from the University of Illinois. He taught at Iowa State University from 1992 to 1997 and has been at the University of Minnesota since 1997, where he holds the Distinguished McKnight University Professorship and the Robert and Marjorie Henle Chair. He has received six conference Best Paper awards, a Best Poster Award, an ICCAD 10-year Retrospective Most Influential Paper Award, the SRC Technical Excellence award and the SIA University Researcher Award.