On Potential Design Impacts of Electromigration Awareness

Andrew B. Kahng[†‡], Siddhartha Nath[†] and Tajana S. Rosing[†]
[†]CSE and [‡]ECE Departments, UC San Diego, La Jolla, CA 92093
{abk, sinath, tajana}@ucsd.edu

***Abstract*—Reliability issues significantly limit performance improvements from Moore's-Law scaling. At 45nm and below, electromigration (EM) is a serious reliability issue which affects global and local interconnects in a chip and limits performance scaling. Traditional IC implementation flows meet a 10-year lifetime requirement by overdesigning and sacrificing performance. At the same time, it is well-known among circuit designers that Black's Equation [2] suggests that lifetime can be traded for performance. In our work, we carefully study the impacts of EM-awareness on IC implementation outcomes, and show that circuit performance does not trade off so smoothly with mean time to failure (MTTF) as suggested by Black's Equation. We conduct two basic studies: EM lifetime versus performance with fixed resource budget, and EM lifetime versus resource with fixed performance. Using design examples implemented in two process nodes, we show that performance scaling achieved by reducing the EM lifetime requirement depends on the *EM slack* in the circuit, which in turn depends on factors such as timing constraints, length of critical paths and the mix of cell sizes. Depending on these factors, the performance gain can range from 10% to 80% when the lifetime requirement is reduced from 10 years to one year. We show that at a fixed performance requirement, power and area resources are affected by the timing slack and can either decrease by 3% or increase by 7.8% when the MTTF requirement is reduced. We also study how conventional EM fixes using per net Non-Default Rule (NDR) routing, downsizing of drivers, and fanout reduction affect performance at reduced lifetime requirements. Our study indicates, e.g., that NDR routing can increase performance by up to 5% but at the cost of 2% increase in area at a reduced 7-year lifetime requirement.**

I. Introduction

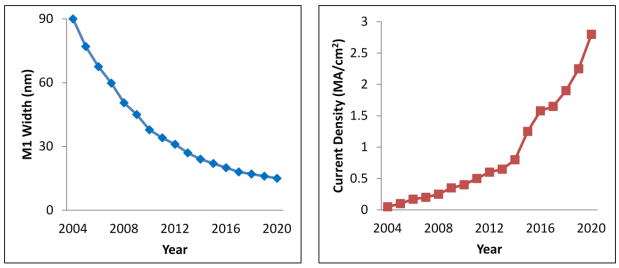
At deep-submicron process nodes, electromigration (EM) poses significant reliability challenges which prevent further scaling of chip performance [18]. Electromigration causes shorts and voids in metal interconnects, leading to failures of the interconnects and decreased mean time to failure (MTTF) of the chip.[1] Particularly for copper metallization at 45nm and below, EM becomes a strong limiter of current density scaling [28]. Black's Equation [2] models median lifetime with respect to EM failures as

$$t_{50} = \frac{A^*}{J^n} \cdot e^{\frac{E_a}{kT}} \quad (1)$$

where t_{50} is the median lifetime, A^* is a geometry-dependent constant, J is the current density, n is a model parameter for current density, E_a is the activation energy of metal ions, T is the temperature of the interconnect, and k is Boltzmann constant. The median lifetime (t_{50}) is approximated as the MTTF of the interconnect. In our work, which focuses on (AC) signal EM, we use the root mean square (RMS) current density.

Besides the long-standing issue of DC electromigration, which affects power delivery networks, AC EM has become a serious concern for interconnects that carry clock and logic signals. AC EM limits become tighter, and EM design violations worsen, with technology scaling since (1) on-current of drivers increases with smaller channel lengths, (2) widths of interconnects decrease, (3) faster switching of transistors increases frequency of operation, and (4) capacitive loads increase due to scaling non-idealities. Together, these factors cause a significant increase in current density, leading to vulnerabilities in the design. Figures 1(a) and (b) show the exponential increase of current density, along with the exponential decrease of wire width, with successive technology nodes.

Figure *INTC*9 from the Interconnect Chapter of the 2011 *International Technology Roadmap for Semiconductors* (ITRS) [31] indicates that maximum current density limits (from EM reliability considerations) become the barrier to further frequency scaling from 2018 onwards. This motivates a closer look at the potential benefits of improved analysis, and/or relaxation, of EM reliability limits. Today, most circuits are still designed to meet 10-year lifetime requirements [13]; however, recent trends suggest that some types of high-volume products are replaced by users after much smaller lifetimes. For example, in the United States people replace their cell phones roughly every two years [34] and laptops every three to five years [35]. Servers are replaced by enterprises every three to seven years [36]. From Black's Equation (1), MTTF can be traded for an increase in the RMS current, I_{rms}, which can be further used to increase the maximum operating frequency (F_{max}) of a design. Therefore, one may seek to design products with reduced MTTF requirements and use the resulting *EM slack* (see discussion in Section III) to improve F_{max}. (Alternatively, EM slack may be used to save chip area and power resources while keeping the same product performance.)



(a) (b)

Fig. 1. (a) Wire width scaling, and (b) current density scaling [28], [31].

However, the MTTF vs. performance relation is not as straightforward as suggested by Black's Equation. Different designs have different performance scaling for the same MTTF tradeoff, and the scaling depends on several other factors such as EM slack, upper bounds on area and peak temperature. In our work, we perform two basic studies.
Experiment 1: We study MTTF vs. F_{max} tradeoffs at fixed budgets for area, power and temperature.
Experiment 2: We study MTTF vs. area and power tradeoffs at a fixed performance requirement.

Experiment 1 leads to three key observations. (1) We show that F_{max} scaling across different circuits is dependent on *EM slack* which is determined by timing constraints, mixes of cell sizes, and lengths of critical paths in the circuit. (2) We identify the dominating constraints (area, EM or temperature) that can limit frequency scaling as lifetime requirements range from 10 years to one year. (3) We empirically determine the design guideline that area and temperature limit F_{max} scaling only when the lifetime requirement is less than four years. Experiment 2 leads to two key observations. (4) Where timing constraints are loose, resource usage can reduce as MTTF reduces; on the other hand, (5) where timing constraints are tight, resource usage will increase as MTTF reduces (hence, EM-awareness will give no benefit). We also compare how conventional fixes for EM and signal integrity (SI) issues affect performance. Our results show, e.g., that up to a lifetime of seven years, Non-Default Rules (NDRs) created by comprehending I_{rms} violations on each net are effective in improving performance by up to 5% at an area cost of 2%. Further, downsizing drivers and reducing fanouts are less effective levers to improve performance (the ratio of increase in performance to increase in area is 1.5 whereas it is 2.5 for NDR routing).

Our main contributions are as follows.

1) Our study shows that EM slack is needed to maximize F_{max} when lifetime requirement is reduced. In general, 20% EM slack allows increase of F_{max} by 20% and 83% EM slack allows increase of F_{max} by 80% when the lifetime requirement is reduced to one year.
2) Our studies of MTTF vs. area and power at a fixed F_{max} show

[1]In the constant failure rate region of the reliability bathtub curve, reliability is modeled using the exponential distribution. In the exponential distribution, the median is ln(2) times the mean. Our use of the MTTF term throughout this paper refers to the mean measure; however, using the median measure will not change the trends in our results.

978-1-4673-3030-5/13/$31.00 ©2013 IEEE

that tightness vs. looseness of timing constraints determine area and power trends at reduced MTTF.

3) Our results give front-end designers and architects insight into lifetime tradeoffs, particularly opportunities to scale F_{max} at no area cost. Physical design engineers can obtain new understanding of performance vs. area tradeoffs at different lifetimes, when NDR routing, driver downsizing and fanout reductions are the available levers for EM fixes.

The rest of our paper is organized as follows. Section II discusses related work, and Section III gives added background on electromigration and implications of Black's Equation. Section IV describes our analytical model, problem formulation and description of our optimization flow. In Section V we discuss validation and results of our optimization flow. The paper concludes with directions for future work in Section VI.

II. Related Work

Previous works on electromigration can be broadly classified as (1) models to accurately estimate EM MTTF, (2) impact of technology scaling and architectural techniques to mitigate EM failures, and (3) synthesis and physical design techniques to reduce current density violations and meet lifetime of the chip.

The EM modeling literature begins with Black's proposal [2] of an analytical model for EM median time to failure (t_{50}). Liew et al. [9] extend Black's Equation to propose DC and AC lifetime models by accounting for EM healing during vacancy relaxation time. Clement [3] uses analytical models for EM failure to describe situations when void growth or stress build-up are the dominant factors. Lloyd [10] proposes void growth and nucleation models to explain that EM failure is also governed by the time it takes for a void to grow to its failure point. These models are essentially parametric modifications of Black's Equation. Orio et al. [5] study the well-known EM failure models to identify gaps in models and implementation. Shatzkes et al. [15] empirically demonstrate that the current density exponent in Black's Equation should be $n = 2$. Wu et al. [19] and Lu et al. [12] propose EM lifetime models that comprehend joule heating and thermal gradient effects on interconnects.

In the architecture literature, Srinivasan et al. [17] consider requirements to change conventional microprocessor architectures such that they comprehend reliability issues of the underlying hardware. Romanescu et al. [14] propose *core cannibalization* to improve lifetime performance by sharing resources between cores when there are hard faults. Their architecture brings implementation challenges due to significant changes in hardware design (e.g., multiplexing of every resource in every pipeline stage). Lu et al. [11] propose to improve performance – in the sense of reclaiming excess design margin – by accounting for temperature gradients in reliability models.

Synthesis and physical design mechanisms include work by Dasgupta et al. [4], who propose a synthesis methodology for high-reliability netlists by taking into account EM degradation arising from switching activity correlations on a microprocessor data transfer bus. Jerke et al. [7] propose a design rule methodology to check current density violations. They use a quasi-3D model to hierarchically calculate current density and perform thermal simulations to check violations in wires and vias. Lienig et al. [8] use post-route area adjustment of critical net structures to size wires and vias and thereby fix current density violations. Banerjee et al. [1] study the extent of EM failure under AC bipolar stress and joule heating in signal nets. Hunter [6] demonstrates that joule heating will limit the peak current densities in circuits and will pose a bigger reliability issue on top of EM failures due to current density violations. Finally, recent whitepapers from Synopsys [26] and Cadence [21], [22] propose ways to fix EM and other signal integrity violations after routing to meet lifetime requirements, but do not quantify impacts on design metrics such as area, congestion and power.

In all of these previous works on EM, the goal has been to meet lifetime bounds by accurately estimating the degradation, then applying fixes using physical design techniques. Our work takes a very different perspective: we directly examine the potential design impacts of trading off EM lifetime bounds, whether to increase the performance of the design, or to reduce area and power given fixed performance and thermal budget.

III. Implications of Black's Equation

We now give a brief background on EM, the design and runtime parameters that affect EM, and implications of Black's Equation. As noted above, EM causes voids and shorts in IC interconnects through momentum transfer between metal ions and conducting electrons; this reduces the reliability of IC interconnects. EM lifetime is affected by *design parameters* such as wire width, fanout, driver size, and operating voltage – since all of these affect current density. *Runtime parameters* that affect EM are switching activity (α), frequency, and temperature. Figure 2 shows the relation to EM MTTF of these factors, with positive (negative) correlations to MTTF shown as red (blue) directed edges.

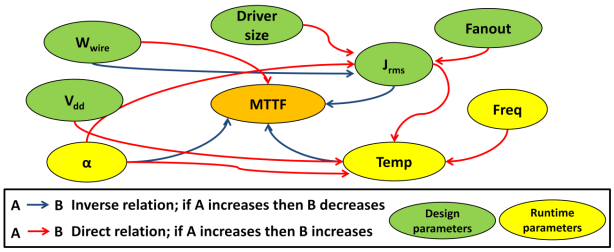


Fig. 2. Design and runtime factors affecting EM MTTF.

EM slack is a function of the RMS current ($I_{rms-net}$) and RMS current limit ($I_{rms-limit}$) according to the technology Library Exchange Format (LEF) file [32]. Under AC stress the EM slack is given by

$$I_{rms-net} = I_{rms-limit} \cdot \sqrt{\frac{MTTF_{def}}{MTTF_{red}}} \quad (2)$$

where $MTTF_{def}$ is the default lifetime requirement (typically 10 years) and $MTTF_{red}$ is the reduced lifetime requirement for the design. EM-awareness can improve performance if there is positive EM slack, i.e., $I_{rms-net} - I_{rms-limit} > 0$. Therefore, for any circuit, Equation (2) determines the theoretically available performance improvement when MTTF is reduced. From Black's Equation, $I_{rms-net}$ is given by

$$I_{rms-net} = \left(\frac{1}{MTTF} \cdot A^*(W \cdot H)^n e^{\frac{E_a}{kT}}\right)^{\frac{1}{n}} = K_1 \cdot \left(\frac{1}{MTTF}\right)^{\frac{1}{n}} \quad (3)$$

$$K_1 = \left(A^*(W \cdot H)^n e^{\frac{E_a}{kT}}\right)^{\frac{1}{n}} \quad (4)$$

where W and H respectively denote the width and height of a wire segment of a net. With all other parameters remaining constant, Equation (3) shows that $I_{rms-net}$ scales inversely with MTTF.

In an actual circuit, $I_{rms-net}$ of a net is given by

$$I_{rms-net} = C \cdot V_{dd} \cdot \sqrt{\alpha \cdot F_{max}\left(\frac{1}{t_r} + \frac{1}{t_f}\right)} \quad (5)$$

where C is the sum of load and wire capacitances, V_{dd} is the operating voltage, t_r and t_f are respectively the rise and fall times of the driver, and α is the switching activity on the net with $0 \leq \alpha \leq 1$.

By combining Equations (5) and (3), we may express F_{max} as

$$F_{max} = \left(\frac{K_1}{K_2}\right)^2 \cdot \frac{1}{C^2 V_{dd}^2} \cdot \frac{1}{\alpha} \cdot \left(\frac{1}{MTTF}\right)^{\frac{2}{n}} \quad (6)$$

$$K_2 = \sqrt{\frac{1}{t_r} + \frac{1}{t_f}} \quad (7)$$

Notice that Equation (6) is quite useful: it shows how performance and reliability interact when the synthesis, place and route (SP&R) degrees of freedom are only gate sizing (affects K_2 and C) and wire width/spacing sizing (affects K_1 and C). Conventional methods for EM fixes use these degrees of freedom. In our work, we use these degrees of freedom along with EM slack as given by Equation (2) to study potential circuit performance at reduced MTTF requirements.

IV. Design of Experiments

We now describe our experimental flows. Experiment 1 requires us to find the highest possible maximum frequency of a design, subject to upper bounds on area, peak temperature and number of critical EM

violations, for a given reduced MTTF requirement. Formally, we

$$\begin{aligned} &\text{maximize } F_{max} \\ &\text{subject to} \\ &\quad T_{peak} \leq T_{UB} \\ &\quad UTIL_{eff} \leq UTIL_{UB} \\ &\quad \sum \left(\frac{I_{rms-net}}{I_{rms-limit}} > 1.1 \right) \leq VIONETS_{UB} \end{aligned}$$

where F_{max} is the maximum operating frequency (which we want to maximize), T_{peak} is the peak temperature of the design, T_{UB} is the upper bound on the peak temperature of the design, $UTIL_{eff}$ is the post-P&R effective utilization of the design, $UTIL_{UB}$ is the upper bound on the effective utilization (which is the upper bound on the area), and $VIONETS_{UB}$ is the maximum number of wire segments that violate $I_{rms-limit}$ by more than 10% (= our upper bound on critical EM violations).

Alg. 1 Find F_{max}

Require: $(T_{peak} \leq T_{UB}) \wedge (UTIL_{eff} \leq UTIL_{UB})$
$\wedge \left(\sum \left(\frac{I_{rms-net}}{I_{rms-limit}} > 1.1 \right) \leq VIONETS_{UB}\right)$
$F_{max} \leftarrow$ *initial guess*
while All constraints are satisfied **do**
  $F_{val} \leftarrow F_{max}$
  $F_{max} \leftarrow F_{max} + f_{step}$
end while
if At least one constraint is violated **then**
  $F_{inval} \leftarrow F_{max}$
  while $F_{inval} - F_{max} \leq f_{\delta}$ **do**
    Search for F_{max} between F_{val} and F_{inval} using binary search
  end while
end if

We perform this optimization using Algorithm 1, which seeks a (global optimum) maximum value for F_{max}. We begin with an initial guess for F_{max} and greedily increment this by f_{step} as long as all constraints are satisfied. The last valid frequency that satisfies all the constraints is stored in F_{val}. If a violation occurs, the violating frequency is stored in F_{inval}. As long as $F_{inval} - F_{max} \leq f_{\delta}$,[2] binary search is used to find the F_{max} which is bounded by F_{inval} and F_{val}. The greatest frequency that meets all the constraints within this bound is the F_{max}, since frequency keeps increasing by f_{step} in each loop until a violation is met. When a violation occurs, binary search efficiently determines the final F_{max}. The final F_{max} differs from the frequency that violates at least one constraint by f_{δ} and hence is the maximum achievable F_{max} for the design. We use Algorithm 1 in the flow described in Section IV-B below.

Experiment 2 requires us to fix timing constraints for a circuit, sweep MTTF from 10 years down to one year, and observe changes in area and power. We keep designs free from EM violations by construction, by limiting cell sizes as shown in Table II (see discussion in Section V).

A. LEF Current Density Characterization

In both experiments, we characterize the current density limits for each metal layer in the technology LEF [32] for a given reduced MTTF requirement ($MTTF_{red}$), switching activity at the primary inputs (PIs) (α_{PI}), and peak temperature (T_{UB}). We assume that at T_{UB}, the design is free from electromechanical stress and that T_{UB} includes joule heating effect due to AC stress on the signal interconnects. The new peak and RMS current density limits of each metal layer in the LEF, $J_{peak-new}$ and $J_{rms-new}$ respectively, are calculated from the default $J_{peak-def}$ and $J_{rms-def}$ using Black's Equation:

$$J_{peak-new} = \left(\frac{MTTF_{def}}{MTTF_{red}} \cdot J^{n}_{peak-def} \cdot e^{\frac{E_a}{k}\left[\frac{1}{T_{UB}} - \frac{1}{T_{def}}\right]} \right)^{\frac{1}{n}} \quad (8)$$

$$J_{rms-new} = \left(\frac{MTTF_{def}}{MTTF_{red}} \cdot J^{n}_{rms-def} \cdot \left(\frac{\alpha_{PI}}{\alpha_{def}} \right)^{\frac{n}{2}} \cdot e^{\frac{E_a}{k}\left[\frac{1}{T_{UB}} - \frac{1}{T_{def}}\right]} \right)^{\frac{1}{n}} \quad (9)$$

where $MTTF_{def}$ is the default lifetime defined by the foundry, α_{def} is the default switching activity of metal layer, T_{def} is the default temperature at which the LEF is characterized by the foundry, n is the model parameter for current density, E_a is the activation energy of the interconnect metal or metal alloy, and k is Boltzmann constant.

[2] f_{δ} is the minimum tolerable frequency difference with F_{inval}.

B. F_{max} Flow

Figure 3 shows our flow for Experiment 1. The flow is described in detail by the following steps.

1) Start with an initial timing constraint (SDC [30] file).
2) Synthesize a design from its RTL description, cell library with cell mixes for different lifetime requirements and SDC.
3) Characterize current density limits in each metal layer in the technology LEFs based on T_{UB}, switching activity, and reduced MTTF requirements ($MTTF_{red}$) using Black's Equation as described in Section IV-A.
4) Place and route the post-synthesis netlist using the newly characterized technology LEF from the previous step.
5) Perform post-route extraction, timing and signal integrity (SI) analysis. Fix EM violations using P&R tool commands.
6) Calculate peak temperature of the design (T_{peak}) from chip and core areas, ambient temperature (T_{amb}), and power using *Hotspot* [16] calibrated to a 45nm Qualcomm SoC package.
7) Check that the number of wire segments violating the I_{rms} limit by more than 10% is less than $VIONETS_{UB}$, slack is met, $T_{peak} \leq T_{UB}$ and $UTIL_{eff} \leq UTIL_{UB}$.
8) If all constraints are met, then increase frequency (decrease clock period in SDC) using binary search. If any constraint is violated, decrease frequency (increase clock period) using binary search.
9) If the next frequency has already given a valid solution, then exit, else repeat the flow from Step 2 using a frequency obtained from the previous step.

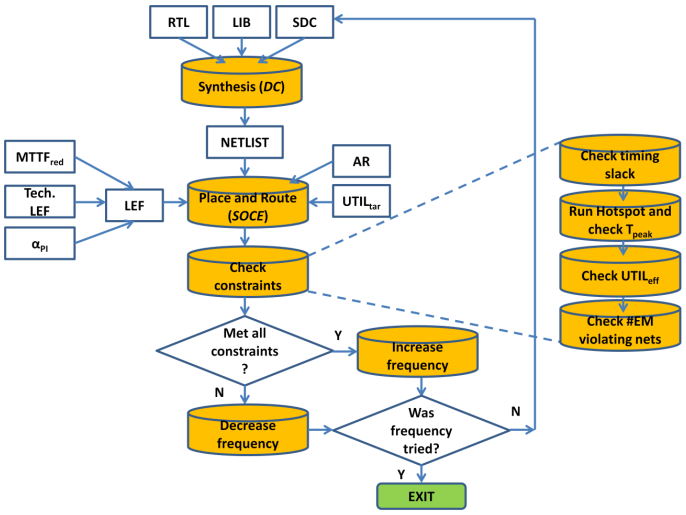


Fig. 3. F_{max} determination flow.

C. Post-Route EM Fix Flow

In both of our experiments, we fix critical EM violations after routing to meet required lifetimes. We use a conventional physical design (PD) methodology that relies on NDRs in engineering change order (ECO) routing to widen wires and spacings, along with fanout reduction and downsizing of drivers [26], [21], [7]. We also study how such standard methods can affect performance and area of a design. We implement these techniques using *Cadence SOC Encounter vEDI10.1 (SOCE)* [25], [20]. Figure 4 shows our flow to create NDRs per net; the following steps describe implementation in *SOCE*.

1) Group EM critical nets by I_{rms} (which is a function of the switching activity on the net) and create NDRs for wire-width and spacing depending on the extent of I_{rms} violation.[3]
2) Execute command *verifyACLimit* in *SOCE* to find the extent of current density violations vis-a-vis LEF in each metal layer, as described in Section IV-A. Apply NDRs to grouped nets

[3] To avoid creating too many NDRs, nets with I_{rms} violation in the range +10% to -10% are grouped together.

using the command *setAttribute -nets netname -non_default_rule rulename*. If any of these nets are "FIXED" (such as clock nets), then convert them to SPECIAL nets for re-routing.

3) Select all violating nets for ECO routing using the command *selectNet netname*.
4) Perform ECO routing of these nets using the commands *setNanoRouterMode -routeWithECO true; setNanoRouterMode -routeSelectedNetsOnly true; setNanoRouterMode -routeStrictlyHonorNonDefaultRule true; globalDetailRoute.* Then, verify AC limit violations after re-route. If there are AC limit violations, decrease fanout of these nets using the command *set_max_fanout fanout* and perform ECO routing.
5) If violations remain after the second ECO route, swap large drivers with smaller drivers and redo timing analysis. If no timing violations, then accept this new frequency and exit, else reduce frequency and run the flow from Section IV-B. If frequency $\leq F_{max}$ obtained from the flow in Section IV-B, then exit.

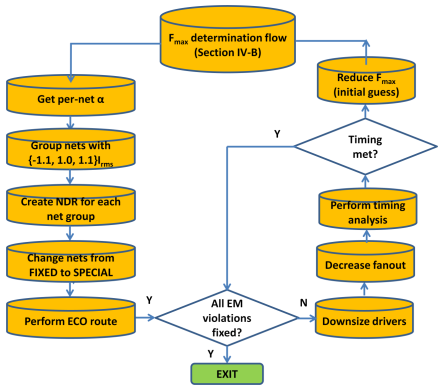


Fig. 4. Per-net NDRs flow to fix EM I_{rms} violations.

V. Validation and Results

Table I describes the implementation parameters used in our experiments. We synthesize the design using *Synopsys Design Compiler vD-2010.03-SP1-64* [29] to obtain a post-synthesis netlist. We set floorplan aspect ratio to 1.0 and target utilization to 60%, then place and route the post-synthesis netlist using *Cadence SOC Encounter vEDI10.1* [25]. We reject any solution that has an effective utilization of more than 66% because we want to ensure that there is enough whitespace to fix any remaining EM violations by widening wires and re-routing the design using NDRs. To eliminate tool noise, we execute each SP&R run five times by perturbing the timing constraints by $\pm 0.5\%$ and $\pm 1\%$. Our fully automated flow is implemented in bash and TCL scripts. The default MTTF for our study is 10 years [13], and the reduced MTTFs from nine to one years. Although we use three different switching activity factors, ($\alpha = 0.05, 0.2, 0.5$) in our experiments, we report results for $\alpha = 0.05$ since this is a typical switching activity at primary inputs for practical workloads. We use six designs from [33] (JPEG encoder, AES cipher, Wishbone DMA controller, PCI bridge, 5×5 network-on-chip router and OpenSparcT1 fpu) and two standard-cell libraries, TSMC45GS and TSMC65GPLUS. We present results of JPEG, AES and DMA because they are representative of all our observations.

As we alluded above, a standard "correct by construction" EM reliability methodology used by industry PD teams is to determine allowed cell strengths for SP&R based on the EM lifetime requirement. Because EM fixes post-P&R – e.g., ECO routing with new NDRs per violating net – are time-consuming and disruptive, the typical methodology goal is to choose cell strengths that will lead to at most several hundred EM violations at the conclusion of the automated SP&R flow. We have performed analyses of our 65nm and 45nm cell libraries to determine cell strengths that result in at most 25 EM violations in automated block implementation. Our flows apply the results of this analysis, which are shown in Table II.

A. Experiment 1: MTTF vs. F_{max}

We validate the Section IV-B flow against the analytical model for EM slack in Equation (2) using three designs – JPEG, AES and DMA. Figures 5(a) and (b) respectively show the percentage of increase in

TABLE I
Implementation Parameters

Parameter	Definition	Value
T_{UB}	Peak temperature upper bound	105^oC
$UTIL_{UB}$	Effective utilization upper bound	66%
$VIONETS_{UB}$	# of EM violating net segments	25
$MTTF_{def}$	Default MTTF	10 years
$MTTF_{red}$	Reduced MTTF	{9, 8, 7, ..., 2, 1} years
α_{PI}	Switching activity at PIs	{0.05, 0.2, 0.5}
T_{amb}	Ambient temperature	45^oC
T_{def}	Default temperature of LEF	85^oC
E_a	Activation energy of interconnect	0.7eV
k	Boltzmann constant	8.617×10^{-5}eV/K
n	Current density model parameter	2
f_{step}	Frequency change step	200MHz
f_δ	Minimum tolerable frequency delta	25MHz
$UTIL_{tar}$	Target utilization	60%
AR	Floorplan aspect ratio	1.0

TABLE II
Cell strengths required to limit the number of EM violations to less than 25 after SP&R.

MTTF (years)	Cell Strength
{10, 9, 8}	all cells up to X12
{7, 6, 5}	all cells up to X16
{4, 3}	all cells up to X24
{2, 1}	all cells

the maximum frequency of these designs at 45nm and 65nm, with $\alpha = 0.05$. We observe that, unlike what is suggested by Black's Equation, frequency does not increase continuously for all types of circuits at both process nodes. As expected, frequency scaling is greater at 45nm because EM slack is larger, cell areas are smaller and transistors are faster than at 65nm. Since 45nm and 65nm trends are similar, we use the 45nm results as the backdrop for our observations. Figures 6(a) and (b) show the EM slack and timing slacks available in the circuits at each lifetime requirement; the positive slack indicates that none of the circuits is limited by tight timing constraints and that timing margin is available for further frequency scaling. Lifetime reduction can increase F_{max} proportional to the amount of EM slack available in the critical paths of the circuits. The key observations from this experiment are:

Observation 1: Positive EM slack, given by the extent of I_{rms} violation at each lifetime, determines the potential frequency scaling benefit of reduced MTTF. The JPEG design has frequency scaling benefit from reduced MTTF all the way to one year lifetime because I_{rms} violations offer sufficient EM slack that can be used to scale frequency. For example, at a two-year lifetime the violations are over 2x in the clock nets, so frequency increases by 18% when MTTF is reduced to one year. Figure 7(a) shows the percentage of EM violations remaining in the critical paths. For the constraints in our experiment, all violations are resolved at one-year lifetime. For the DMA design, F_{max} increases by 60% when MTTF reduces from 10 to four years. Below four years, the violations are less than 20% and the flow cannot make use of the theoretical EM slack of 41%[4] from Equation (2).

Observation 2: The EM slack in each circuit is determined by the mix of cells. According to Table II, EM slack increases as MTTF reduces because we increase the size of cells. Figures 8(a) and (b) show that cell size and frequency determine the total power. Large cells are more frequently used as MTTF reduces, which makes the ratio of leakage to total power nearly constant. EM slack is also a function of length of critical paths. At MTTF less than six years, JPEG and DMA have short critical paths (20 stages on average), so the switching activity on each net more closely tracks that of the PIs. This leads to more I_{rms} violations, which can be used as EM slack. AES, on the other hand, has longer critical paths (30 stages on average) below seven years MTTF, which makes switching activity on nets and wire capacitances (as well as net lengths) small; thus, the extent of I_{rms} violations is small. We see that AES has at most 27% performance increase despite having slack in timing and utilization as shown in Figures 6(b) and 7(b) respectively.

Observation 3: For MTTF requirements below three years, area plays an important role in F_{max} scaling. Figure 7(b) shows that the DMA design cannot scale frequency further when MTTF drops below seven years, because area upper bounds are violated, i.e., total cell area

[4]In Figure 6(a), EM slack is expressed as a percentage of $I_{rms-limit}$.

exceeds available layout area. However, until MTTF is decreased to four years, these designs can increase frequency by different amounts at no area cost: frequency increases for AES, JPEG and DMA are 20%, 70% and 50%, respectively. The potential for increased frequency without area cost may be motivation for front-end designers and architects to incorporate EM-awareness in their designs. Figures 8(a) and (b) respectively show that the percentage increase in total power grows as F_{max} increases, and that leakage power remains almost constant relative to total power.

Figures 9(a) and (b) show the dominating constraints for F_{max} scaling at different lifetimes for the 45nm JPEG and DMA designs. For both designs, EM is the dominating constraint until a five-year MTTF. At one-year lifetime, area constraints prevent further F_{max} scaling. Figures 9(c) and (d) show the dominating constraints for these designs at 65nm. Again, EM is a dominant constraint until a six-year MTTF, but for lifetimes less than four years peak temperature becomes a dominating constraint as larger cells dissipate more power, and more instances are used to meet F_{max} constraints at 65nm.

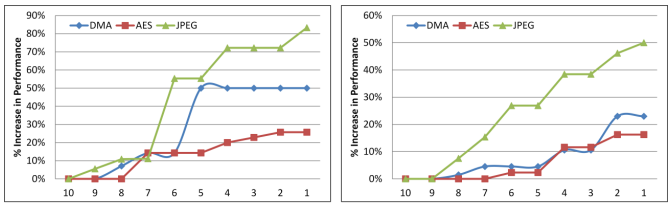


Fig. 5. Percentage increase in F_{max} at (a) 45nm and (b) 65nm due to reduced MTTF requirement.

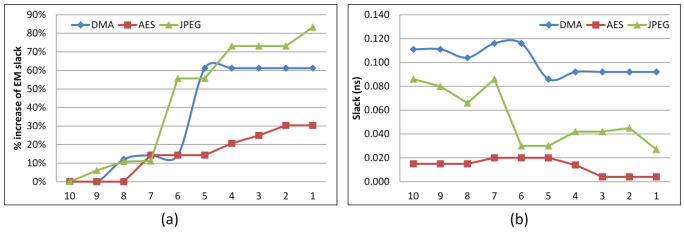


Fig. 6. (a) Percentage increase of EM slack and (b) timing slack in circuits at 45nm due to reduced MTTF requirement.

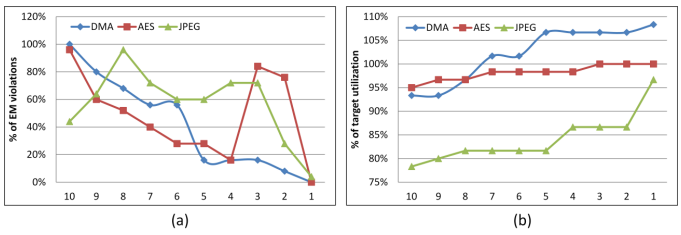


Fig. 7. (a) Percentage of critical paths with EM violations and (b) total cell area relative to available area at 45nm as MTTF requirement is reduced.

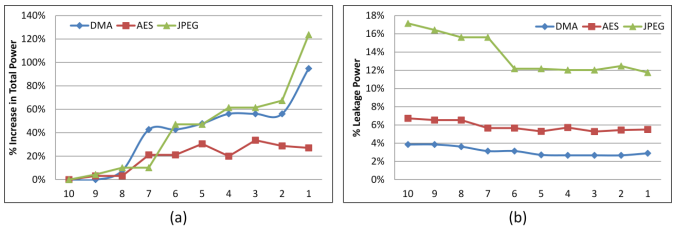


Fig. 8. Percentage of (a) increase in total power and (b) leakage power relative to total power as MTTF requirement is reduced.

B. Experiment 2: MTTF vs. Area and Power

Our Experiment 2 results show how timing constraints determine available scaling of area and power. We implement AES (1100MHz), JPEG (850MHz) and DMA (2000MHz) designs at 45nm using the cell strengths at each MTTF in Table II. The constraints for JPEG are

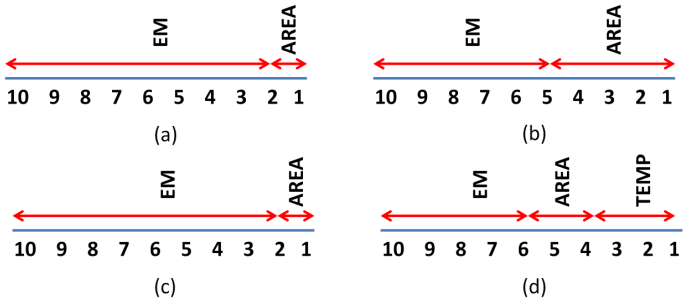


Fig. 9. Dominating constraints on F_{max} scaling at different MTTF limits: (a) JPEG 45nm, (b) DMA 45nm, (c) JPEG 65nm, (d) DMA 65nm.

loose, with 93ps of slack after SP&R; DMA and AES have 2ps and 1.6ps of post-P&R slack, respectively.

Observation 4: If a design has large positive timing slack at 10 years, then area and power *decrease* when MTTF is reduced. Figures 10(a) and (b) show the area and power reduction for JPEG. The positive slack at 10-year lifetime allows the design to replace small-sized buffers and inverters with large drivers when MTTF is reduced, thereby reducing the number of instances. The number of instances is also reduced with MTTF because EM violations are reduced, and post-route fixes do not need to add extra buffers when reducing fanouts or downsizing drivers. Collectively, these factors reduce area by 3.7% and power by 2.5%.

Observation 5: If a design has little or no positive timing slack at 10-year lifetime, then area and power *increase* when MTTF is reduced. For AES (and DMA), the constraint is very tight at 10 years. As MTTF reduces, smaller drivers are replaced by larger drivers to meet timing, but the number of instances does not decrease. Although the number of EM violations reduces with MTTF which, in turn, reduces the resources needed to fix these violations after routing, the number of instances does not reduce significantly. As MTTF reduces, these factors increase area by 2.02% and power by 7.08% as shown in Figures 11(a) and (b). Hence, EM-awareness does not benefit circuits with tight timing constraints.

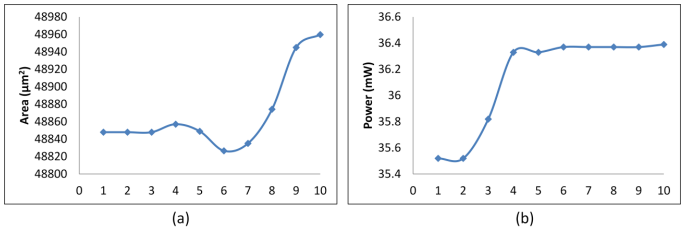


Fig. 10. JPEG (a) area and (b) power at 45nm.

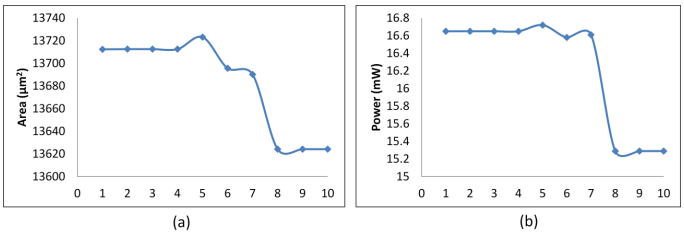


Fig. 11. AES (a) area and (b) power at 45nm.

C. Performance and Area Impact of Conventional EM-Fix Methods

We additionally study how conventional EM and SI fixing methods affect performance and area at reduced lifetime requirements. We sweep MTTF from 10 years down to one year and apply per-net NDRs, driver downsizing and fanout reduction fixes as described in Section IV-C to the three 45nm testcases. Our analyses show that NDRs are more flexible in repairing EM violations.

Observation 6: Using NDRs to fix EM violations helps to increase F_{max} only until MTTF reaches seven years. Figure 12 shows that the increase in F_{max} is less than 5% for all the designs. For JPEG, there is no increase although area increases by 2%; this is because the extensive I_{rms} violations cannot be repaired by the conventional fixing methods. AES, however, gains about 4.65% in frequency by paying 0.4% area, since its smaller EM slack permits repair using conventional methods.

Observation 7: When drivers are downsized, numbers of buffers and inverters increase to meet timing; when fanout is reduced, (too many) cloned buffers are added to meet clock skew. Figures 13(a) and (b) show how area and performance change by changing fanout and driver sizes, respectively, at MTTF of seven years for the AES design. A performance increase of 3% comes at the cost of a 1.86% increase in area. Compared to downsizing and fanout reduction, NDRs appear to improve F_{max} with less increase in area.

From these observations, we conclude that conventional EM fixing methods have limited capability to scale performance. Below seven-year MTTF, these methods are only useful for fixing EM at the cost of increased area, and cannot create sufficient EM slack to increase F_{max}. Physical design engineers may be able to reduce the time taken for these conventional fixes by reducing MTTF.

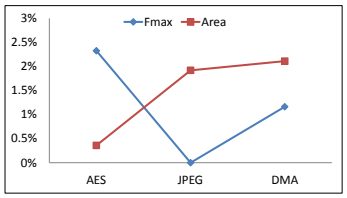


Fig. 12. Percentage increase in F_{max} and area after NDR re-route at 45nm.

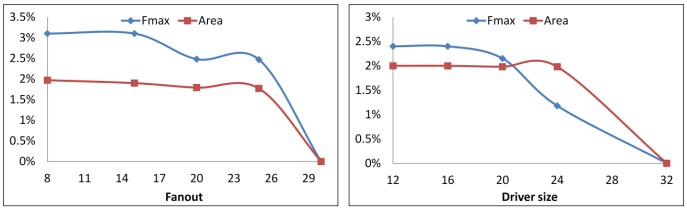


Fig. 13. Percentage increase in F_{max} and area of AES by varying (a) fanout and (b) driver size.

VI. Conclusions and Future Work

Electromigration is a significant reliability concern, and limits performance and current density scaling in leading-edge technology nodes. In this work, we study the potential impacts of improved EM-awareness in designs, through two basic experiments that seek understanding of MTTF versus F_{max} tradeoffs at fixed resource budgets, and MTTF versus area and power tradeoffs at fixed performance requirement. We use a fully automated flow based on commercial SP&R tools to find the maximum F_{max} for a given block implementation under EM reliability, peak temperature and area bounds; using this flow, we extract and validate trends using three designs across two technology nodes.

A key observation from our first experiment is that the available performance scaling from MTTF reduction is dependent on *EM slack*, which is affected by circuit characteristics such as length of critical paths, timing constraints and cell mixes in critical paths. We also show the dominating constraints (area, temperature or EM) that can limit F_{max} scaling at different lifetimes for different types of circuits. Our results further shed light on the available range of performance scaling, e.g., by reducing lifetime from 10 years to one year it is possible to increase F_{max} by 80% given the same amount of EM slack. Results of our second experiment indicate that area and power will decrease when MTTF is reduced only in (parts of) designs where timing constraints are loose. Additional studies shed light on the performance and area impact of conventional EM fix methods at a reduced MTTF requirement of seven years (with EM lifetimes below seven years, such methods offer no performance benefit in the technologies that we have studied). Specifically, performance can be increased by less than 5% with 2% increase in area when using NDRs to fix EM violations, down to a lifetime bound of seven years. Such results may give physical designers insight into when to trade off turnaround time spent on conventional EM fixes for increase in performance.

Our ongoing work studies the potential performance, area and power impacts when EM reliability requirements are relaxed for designs that have multiple operating modes – in particular, frequency overdrive modes that strongly stress interconnect reliability. Our end goal is to understand the combined impact of EM and other back end of line reliability mechanisms (time-dependent dielectric breakdown, stress-induced voiding, etc.) on interconnect lifetime and resilient design mechanisms at the architecture and system levels.

Acknowledgments

Research supported in part by funding from NSF, SRC, MARCO/DARPA, Qualcomm and Samsung.

References

[1] K. Banerjee and A. Mehrotra, "Coupled analysis of electromigration reliability and performance in ULSI signal nets" *Proc. ICCAD*, 2001, pp. 158-164.
[2] J. R. Black, "Electromigration failure modes in aluminium metallization for semiconductor devices", *IEEE Letters* 57(9) (1969), pp. 1587-1594.
[3] J. J. Clement, "Electromigration modeling for integrated circuit interconnect reliability analysis", *IEEE TDMR* 33(1) (2001), pp. 33-42.
[4] A. Dasgupta and R. Karri, "Electromigration reliability enhancement via bus activity distribution", *Proc. DAC*, 1996, pp. 353-356.
[5] R. L. de Orio, H. Ceric and S. Selberherr, "Physically based models for electromigration: From Black's equation to modern TCAD models", *Microelectronics Reliability* 50 (2010), pp. 775-789.
[6] W. Hunter, "The implications of self-consistent current density design guidelines comprehending electromigration and Joule heating for interconnect technology evolution", *Proc. IEDM*, 1995, pp. 483-486.
[7] G. Jerke and J. Lienig, "Hierarchical current-density verification in arbitrarily shaped metallization patterns of analog circuits", *IEEE TCAD* 23(1) (2004), pp. 80-90.
[8] J. Lienig and G. Jerke, "Current-driven wire planning for electromigration avoidance in analog circuits", *Proc. ASPDAC*, 2003, pp. 783-788.
[9] B.-K. Liew, N. W. Cheung and C. Hu, "Electromigration interconnect lifetime under AC and pulse DC stress", *Proc. IRPS*, 1989, pp. 215-219.
[10] J. R. Lloyd, "Black's law revisited: Nucleation and growth in electromigration failure", *Microelectronics Reliability* 47 (2007), pp. 1468-1472.
[11] Z. Lu, J. Lach, M. Stan and K. Skadron, "Banking chip lifetime: Opportunities and implementation", *Proc. 1st Workshop on HPCRI*, 2005.
[12] Z. Lu, W. Huang, M. R. Stan, K. Skadron and J. Lach, "Interconnect reliability prediction for reliability-aware systems", *IEEE TVLSI* 15(2), 2007, pp. 159-172.
[13] C. A. Papameletis, "Development of design methodologies and CAD tools for system-level evaluation of interconnect reliability issues in SoC designs", *Diploma Thesis*, National Technical University of Athens, 2010.
[14] B. F. Romanescu and D. J. Sorin, "Core cannibalization architecture: Improving lifetime chip performance for multicore processors in the presence of hard faults", *Proc. PACT*, 2008, pp. 43-51.
[15] M. Shatzkes and J. R. Lloyd, "A model for conductor failure considering diffusion concurrently with electromigration resulting in a current exponent of 2", *Journal of Applied Physics* 59(11) (1986), pp. 3890-3893.
[16] K. Skadron, M. R. Stan, W. Huang, S. Veluswamy, K. Sankaranarayanan and D. Tarjan, "Temperature-aware microarchitecture", *Proc. ISCA*, 2003, pp. 2-13.
[17] J. Srinivasan, S. V. Adve, P. Bose and J. A. Rivers, "The case for lifetime reliability-aware microprocessors", *Proc. ISCA*, 2004, pp. 276-287.
[18] K.-C. Wu, M.-C. Lee, D. Marculescu and S.-C. Chang, "Mitigating lifetime underestimation: A system-level approach considering temperature variations and correlations between failure times", *Proc. DATE*, 2012, pp. 1269-1274.
[19] W. Wu, J. S. Yuan, S. H. Kang and A. S. Oates, "Electromigration subjected to Joule heating under pulsed DC stress", *Solid-State Electronics* 45 (2001), pp. 2051-2056.
[20] "Clock routing rules." *http://www.cadence.com/Community/blogs/di/archive/2011/05/10/five-minute-tutorial-setting-up-clock-routing-rules.aspx*
[21] "Signal and design integrity whitepaper." *http://http://w2.cadence.com/whitepapers*
[22] "Virtuoso Ultrasim full-chip simulator netlist-based electromigration voltage drop (EMIR) flow." *http://http://w2.cadence.com/whitepapers*
[23] "28 nm IC design: The devil is in the details.' *http://www.cadence.com/Community/blogs/di/archive/2011/03/14/28-nm-ic-design-the-devil-is-in-the-details.aspx*
[24] "Eliminating routing congestion issues with logic synthesis." *http://www.cadence.com/rl/resources/white_papers/routing_congestion_wp.pdf*
[25] "Cadence SOC Encounter user guide." *http://www.cadence.com/products/di/first_encounter/pages/default.aspx*
[26] "Signal electro-migration analysis and fixing research in IC Compiler." *http://www.synopsys.com.cn/information/snug/2011/signal-electro-migration-analysis-and-fixing-research-in-ic-compiler-2*
[27] "Advanced design closure techniques using IC Compiler and Zroute on high performance designs at 32nm and below." *https://www.synopsys.com/news/pubs/snug/austin2011/pb2_han_paper.pdf*
[28] "Understand and avoid electromigration (EM) & IR-drop in custom IP blocks." *http://www.synopsys.com/Tools/Verification/CapsuleModule/CustomSim-RA-wp.pdf*
[29] "Synopsys Design Compiler user guide." *http://www.synopsys.com/Tools/Implementation/RTLSynthesis/DCUltra/pages/default.aspx*
[30] "SDC user's guide." *http://www.actel.com/documents/SDC_AN.pdf*
[31] "ITRS 2011 edition reports." *http://www.itrs.net/Links/2011ITRS/Home2011.htm*
[32] "LEF DEF reference." *http://www.si2.org/openeda.si2.org/projects/lefdef*
[33] "OpenCores." *http://www.opencores.org*
[34] "Americans replace their cell phones every 2 years, Finns – every six, a study claims." *http://www.phonearena.com/news/Americans-replace-their-cell-phones-every- 2-years-Finns–every-six-a-study-claims_id20255*
[35] "When should you buy a new laptop?" *http://www.pcworld.com/article/219826/when_should_you_buy_a_new_laptop.html*
[36] "How often should you plan to replace your servers?" *http://www.xennix.com/how-often-should-you-plan-to-replace-your-servers*