Reliability Issues in Deep Deep Submicron Technologies: Time-Dependent Variability and its Impact on Embedded System Design

Antonis Papanikolaou[1], Hua Wang[1,2], Miguel Miranda[1],
Francky Catthoor[1,2] and Wim Dehaene[2]

[1] IMEC vzw, Kapeldreef 75, 3001 Leuven, Belgium
{papaniko,wanghua,miranda,catthoor}@imec.be
[2] Katholieke Universiteit Leuven, ESAT Dept., Kasteelpark Arenberg 10, 3001 Leuven, Belgium
wim.dehaene@esat.kuleuven.be

Abstract. Technology scaling has traditionally offered advantages to embedded systems in terms of reduced energy consumption and die cost as well as increased performance, without requiring significant additional design effort. Scaling past the 45 nm technology node, however, brings a number of problems whose impact on system level design has not been evaluated yet. Random intra-die process variability, reliability degradation mechanisms and their combined impact on the system level parametric quality metrics are prominent issues that will need to be tackled in the next few years. Dealing with these new challenges will require a paradigm shift in the system level design phase.

1 Introduction

Embedded system design is especially demanding and challenging in terms of requirements that need to be satisfied, e.g. real-time processing, cost effectiveness, low energy consumption and reliable operation. These requirements have to be properly balanced until a financially viable global solution is found. Novel mobile multimedia and communication applications pose extremely severe requirements on the amount of storage, processing and functionality capabilities of the system. Near future embedded systems will have to combine interactive gaming with advanced 3D and video codecs together with leading edge wireless connectivity standards, like software defined radio front-ends and protocol stacks for cognitive radio. This will increase the platform requirements by at least a factor of 10. Meanwhile, battery capacity is only increasing by about 7% per year and users demand longer times between battery recharges. Optimizing any one of these requirements by compromising on another is a rather straightforward design task. However, in embedded system design the solution must obey the constraints in all four requirement axes.

Products containing some sort of embedded system implementation targeting safety critical applications (i.e. advanced braking systems and traction control of modern cars, biomedical devices, etc.) impose aggressive constraints on the design of embedded systems, especially in terms of meeting reliability and fail-safe operation targets during the guaranteed product lifetime. This translates onto very low field return targets during that time, since failures can lead to dire financial consequences or catastrophic results. On the other hand, systems that belong to the low end consumer electronics market are also subject to tight lifetime and reliability targets. They are usually deployed in very large volume, thus even a small percentage of failures can lead to a large amount of field returns that cost both financially and in consumer loyalty and in company image. For all these reasons fail-safe reliable operation throughout a guaranteed product lifetime becomes a strategically important property for the design of embedded systems.

Technology scaling has traditionally enabled improvements in three of the design quality metrics: increased processing performance, lower energy consumption per task and lower die cost. Reliability targets were also guaranteed at the technology level by using well controlled processes and well characterized materials. Unfortunately this ”happy scaling” scenario where technology and design could be kept decoupled is coming to an end [1]. New technologies become far less mature than earlier ones, e.g. the nanometer range feature sizes require the introduction of new materials and process steps that are not properly characterized by the time they start being used in commercial products, leading to potentially less reliable products. On the other hand, progressive degradation instead of abrupt failure of electrical characteristics of transistors and wires becomes reality as an intrinsic consequence of the smaller feature sizes and interfaces as well as increasing electric fields and operating temperatures (see [2] and its references). Effects considered as second-order in the past, become a clear threat now for the correct operation of the circuits and systems since they start affecting their parametric features (e.g., timing but also energy dissipation) while the functionality remains unaltered. Moreover, as we show in this work, the combined impact of manufacturing uncertainty (e.g. process variability) and reliability degradation results in time-dependent variability. The electrical characteristics of the transistors and the wires will vary statistically in a spatial and a temporal manner, directly translating into design uncertainty during fabrication and even during operation in the field, especially as a function of the application's functionality influence in the system as such. Unfortunately, current reliability models based on traditional worst case stress analysis are not sufficient to capture these more dynamic system level interactions, resulting in over-pessimistic implementations [2]. Research in fully integrated analysis models (from technology to full system) is urgently needed.

On the solution side, a number of conventional techniques already exist for dealing with uncertainty. However, most of them rely on the introduction of worst-case design slacks at the process technology, circuit and the system level in order to absorb the unpredictability of the transistor and interconnect performance and to provide implementations with predictable parametric features. But trade-offs are always involved in these decisions, which result in excessive energy consumption and/or cost leading to infeasible design choices. From the designers perspective reliability degradation mechanisms manifest themselves as time-dependent uncertainties in the parametric performance metrics of the devices. In the future sub 45 nm regime, these uncertainties would be way too high to be handled with existing worst-case design techniques without incurring significant penalties in terms of area/delay/energy. As a result, reliability becomes a great threat to the design of reliable complex digital systems-on-chip (SoC) implementations. We believe this will require the development of novel reliability models at all three levels, namely device, circuit and system level. They should be capable of capturing the impact of the application functionality on the system as well as new design paradigms for embedded system design in order to build reliable systems out of technology which will be largely unpredictable in nature. This problem cannot only be solved at the technology and circuit level anymore. A shift toward Technology-Aware Design solutions will be required to keep designing successful systems in future aggressively scaled technologies.

2 Reliability Degradation Mechanisms for Scaled Technology Nodes

Reliability has always been a concern in the technology development community. In the past decades however, technology scaling involved shrinking the feature sizes of transistors and wires as well as the supply voltage with minimal intervention on the materials used. The available reliability margins were quite large and guaranteeing a life time of ten years for each of the transistors in the design was a feasible target, even under worst-case assumptions on the operating conditions. Furthermore, the first transistor to break in the die has been assumed to render the entire die non-functional which is another worst-case assumption that reliability engineers have always made in order to guarantee life-time under all circumstances. Still these conditions were based on reasonable assumptions. But scaling toward Deep Deep Sub-Micron (DDSM) technology nodes is not business as usual. Along with feature miniaturization, process technologists have also introduced a number of novel materials and process steps in the leading edge manufacturing processes. Examples include the high-k materials used for the transistor gate insulation from the channel, the low-k materials for the implementation of the dielectrics in the metal stack, the re-introduction of copper for the implementation of interconnect wires a couple of technology nodes ago etc. Characterizing these materials and their interactions for reliability degradation mechanisms is an extremely complex task. Typically they are used in commercial processes before full understanding of the physical degradation mechanisms is available. At the same time, the supply voltage scaling has been saturating in order to keep enough headroom between the transistor threshold voltage and the supply voltage hence increasing the electric fields and stress conditions for these devices. Furthermore, effects that in the past have been considered second order are now becoming a clear threat for the parametric and functional operation of the circuits and systems in near future technologies. Examples include soft-breakdowns (SBD) in gate oxide of transistors (especially dramatic in high-k oxides) [2], Negative Bias Temperature Instability (NBTI) issues in the threshold voltage of the PMOS transistors [3], Electro-Migration (EM) problems in copper interconnects [4], breakdown of dielectrics in porous low-k materials [5], etc.

The net result is that it becomes increasingly difficult to guarantee the life time of transistors and wires for new technology nodes, as will be discussed in the remainder of this section. Apart from the reliability mechanisms, transistors and wires are also subject to manufacturing imperfections which lead to static manufacturing time variability. This is also aggravated by novel transistor architectures. The development of FinFETs is a good example. Variability due to random dopant fluctuations can be severely reduced by alleviating or reducing the need of dopant atoms in the channel. But implementing FinFETs in a stable and reliable process requires the controlled and precise manufacturing of very complex three-dimensional structures (fins), which leads to a significant increase in the variability contribution due to line edge roughness in all three dimensions.

NBTI effects [6, 3] in PMOS transistors and (Soft) gate oxide Break-Downs (SBD) in NMOS transistors [7] are becoming two of the most important sources of progressive degradation of electrical properties of devices in DDSM technologies. Thinner equivalent gate oxides, due to dimension scaling, and a deficient supply voltage scaling are leading to higher electrical fields in the oxide interfaces, hence in larger tunneling currents that degrade the electrical properties of the oxide, resulting in electric traps in the interfaces. These traps translate in both NBTI and SBD effects. NBTI appears as a progressive drift of the threshold voltage of the PMOS transistors over time, which can partially be recovered once the negative voltage stress between the gate and the drain/source becomes zero or positive. SBD appears when enough traps align in the gate dielectric. A conducting path is created resulting in ”micro” tunneling currents through the gate. After some time the path created will ”burn out” leading to an electrical short or Hard Break-Down (HBD) resulting in a catastrophic failure of the transistor. The transition from the initial conducting path to the HBD is not abrupt, the gate leakage current will start to progressively increase long before the HBD actually occurs (Fig. 1). Moreover, changes of the stress conditions due to the application usage of the platform, like activity, and the way this is translated into operating conditions of the devices and wires will also have a major impact on the actual dynamics of the degradation phenomena.

Similar effects are predicted for wires from the 45nm technology node on. Both electro-migration in the metal wires and reliability problems in the dielectrics between them are becoming serious concerns for guaranteeing correct and reliable operation during a specified product lifetime. The ever decreasing widths of the local wires combined with the slowing scaling of the supply voltage lead to an increase in current densities along technology nodes, which is accelerating electro-migration problems not only in aluminum but also in the more robust copper interconnects [4]. The problem is not alleviated by assuming a decreasing fan-out condition which would provide a temporary partial solution

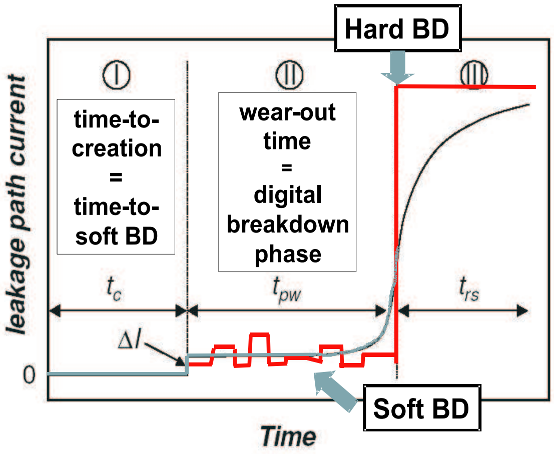

Fig. 1. Wear-out and breakdown model for normal (SiON) and high-k (HfO2) gate oxides [8]

to current densities control. For relatively long local and intermediate interconnects even though the current densities can increase due to large fan-outs, electro-migration is not a considerable problem. System-on-Chip level communication typically has more relaxed constraints on energy consumption per task and performance. Local interconnects, on the other hand, which are used to implement processing elements or local communication between processors and local memories/caches have all the fore-mentioned stringent constraints. Guaranteeing real-time performance and improving density in order to minimize area (die cost) leads to the minimization of the lateral dimensions of the wires [9]. These conditions significantly speed up the electro-migration mechanism in this context.

Similar to SBD effects in transistors, electro-migration is also translated to a progressive degradation of the associated resistance of the wire. The thinner the wire is, the earlier the degradation will start [4] (see Fig. 2). This is aggravated by asymmetries in the printed interconnect features, such as connections between wires and vias. Interfaces between different materials across the conducting path are especially susceptible to electro-migration problems. Also irregularities in the critical dimensions of the interconnects, due to Line Edge Roughness [10] as a consequence of sub-wavelength lithography, will make the whole metal structure far more vulnerable to electro-migration problems. This can lead to uncontrollable (location- and impact-wise), random hot-spots.

A similar case can also be drawn for breakdowns in the dielectrics in the interconnect stack, where the wire pitch is reducing in each new technology node. This leads to reduced thicknesses of the dielectrics between metal wires, while the supply voltage does not reduce at the same pace. As a result the main figure of merit for reliability, Mean Time To Failure (MTTF), drastically reduces [5]

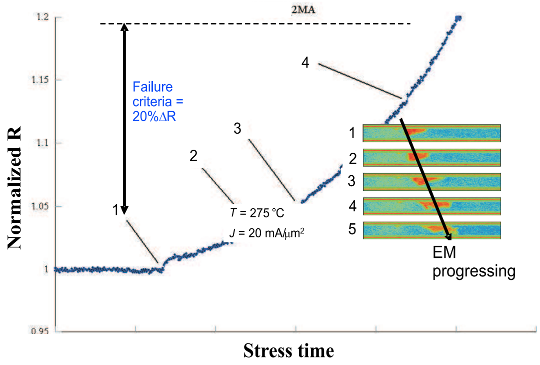

Fig. 2. EM signature in narrow lines (<120 nm line width) [4]

(see Fig. 3). The reason is the combination of the increasing electric fields in active wires due to the insufficient voltage scaling and the introduction of low-k dielectric materials for improving the RC delay of wires based on less electrically robust porous materials. Even when this failure phenomenon manifests itself as catastrophic without an explicit progressive degradation phase, the number of dielectric breaks over time and the time to first break becomes less predictable than earlier. Imperfections of the low-k dielectric material, like granularity of the material grains and/or air gaps, are dramatically increasing the uncertainty on the actual useful life-time of the product.

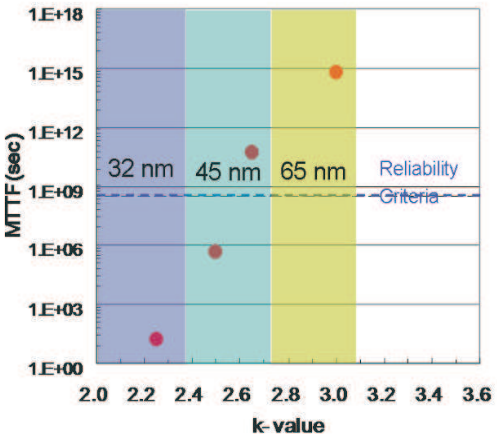

Fig. 3. Reliability targets and projected MTTF in advanced Cu-low-K materials [5]

3 The Impact of Reliability Degradation Mechanisms on the Circuit Level Performance Metrics

For a proper evaluation of the impact that the fore-mentioned reliability problems have in circuit and system design, it is not sufficient to have models representing the mechanism and effect of a particular reliability effect in a single device or interconnect. Not even considering possible interactions with other reliability phenomena is sufficient, e.g. studying the combined impact of NBTI and SBD effects in the behavior of an SRAM cell [11]. The real problems need to be evaluated in the context of the particular circuit where the device/interconnect subject to degradation is situated. The fact that a progressive degradation effect may manifest mildly when looking at each single transistor/wire separately does not provide any information about its impact on the circuit level performance metrics. For instance, oxide breaks manifest themselves as a slight increase in the total gate leakage [12] that may not have strong impact on the transistor current-voltage characteristics [8], since the drain current does not change significantly at the moment the soft oxide breakdown occurs. However, when looking to the interaction that the gate current increase may have with the circuit operation, although small, it can affect the parametric figures of the circuit by affecting the current of another device whose drain is connected to that gate. A typical example where small changes in the gate current of a single transistor can cause major problems at the circuit level are SRAM sense amplifiers or other circuits that work under a common mode rejection mode. Affecting the bias conditions of one of the transistors even slightly may have detrimental effects for the functionality of the circuit. Different types of circuits are much more robust toward breakdowns, for example ring oscillators can tolerate hard breakdowns on several of their transistors before they stop oscillating at the specified frequency [13]. This means that in order to evaluate the impact that reliability degradation mechanisms have on the circuit level performance metrics we need analysis and modeling tools that can take into account the context where the affected transistor/wire is operating in.

In the general case, the gate leakage current of FETs can either impede or favor the charging/discharging process of the output node of a gate leading to longer/shorter delays. In terms of equivalent SBD resistance, previous research has predicted that it is in the order of several hundred kilo-Ohm and above for sub-45nm technologies [14]. Furthermore, the extra leakage contributes directly to the increase of total energy consumption. A lower than nominal voltage swing can be observed at the gate of the output node, due to the soft oxide breakdown induced gate leakage. Such a voltage swing then slows down the downstream logic driven by the defective gate [15]. Delay degradation induced by such a defect has already been observed in simple logic NOR/NAND gates and small data-paths (full adder)[15, 16].

Apart from the standard logic gates, it has recently been shown that SBDs in the NMOS transistors of SRAM components can also bring shifts in their performance. The energy and delay of both sense amplifiers and individual SRAM cells are dramatically affected by having a single SBD in one of their transistors.

A variation of 36% in energy and 22% in delay is reported for the sense amp and a similar variation is reported in the SRAM cell parametric (energy/delay) operation [14]. The amount of drift is mainly due to the impact of soft oxide breakdown on the internal feedback loops of these sub-circuits. Similar to combinational logic, the infected feedback loop can also reduce/increase the delay of the actual component. Such drifts come from the second-order interactions of the gate leakage increase enabled via the circuit topology and a more significant variation in the circuit parametric figures is also expected when the soft gate oxide breakdown starts affecting the first order characteristics of the device. The actual behavior of the associated sub-circuit under SBD effect is more difficult to model than those of logic gates because of these feedback loops.

Moreover, these complex interactions exhibit a multiplicative effect when considered in combination with random intra-die process variability. The time dependent nature of the degradation effects and the uncertainty in the initial parametric figures due to variability lead to time dependent variability that is very difficult to predict and control by countermeasures that are only based on design time analysis and solutions. Given that the breakdown resistance value and location are random in nature [7], it is reasonable to expect a more dramatic impact of this combined effect on the energy and delay of the SRAM in the DDSM era. Figure 4 illustrates the increase in the uncertainty ranges of the sense amplifier performance metrics when a single soft break-down is considered in one of its transistors. The delay and energy consumption ranges increase by more than a factor of two. This additional uncertainty largely prohibits the designers to predict the run-time circuit behavior at design time. Thus it is impossible to steer circuit level optimizations, e.g. timing slacks, device sizing decisions, etc., that make the circuit robust enough to both effects combined.

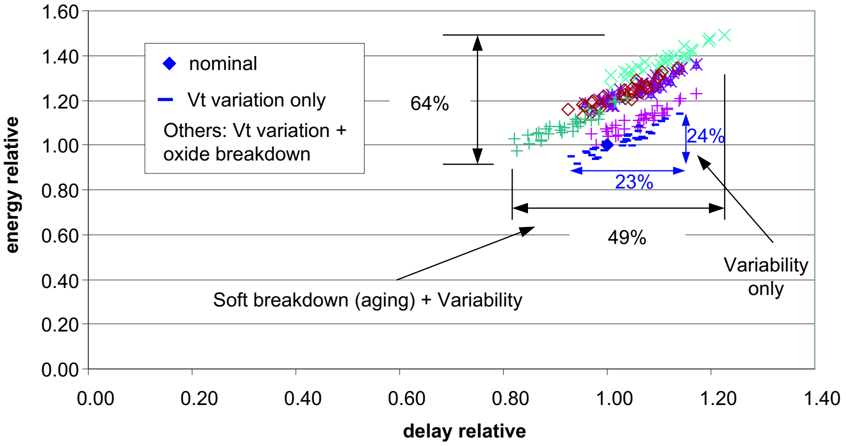

Fig. 4. Impact of variability and gate oxide breakdown in the energy consumption and delay of an SRAM sense amplifier when only one transistor suffers a soft breakdown

In the case of a complete SRAM cell matrix the conclusion is quite different. Only the access delay is greatly affected by SBDs, while the associated access energy is only marginally influenced. The matrix consists of a very large number of cells, where a few of them are accessed in parallel in every memory read or write operation. The matrix static energy consumption is an accumulation of the static energy of all the cells, so a change in the leakage current of one of the transistors in the matrix is unlikely to impact the total static energy consumption significantly. Dynamic energy consumption also exhibits the same trend. The impact on break downs on the dynamic energy consumption of the matrix is also rather small. In the case of delay, however, things are quite different. The break downs have a significant impact on the relative driving strengths of the transistors in the cross-coupled inverter pair, which leads to a significant impact on the delay of reading or writing the activated cells. Transistor level simulations have been carried out to evaluate the impact of soft break downs on the main performance metrics of the SRAM matrix. The difference in dynamic and static energy consumption of the matrix incurred by injecting soft break downs in four individual transistors is around 1%, so it is indeed negligible. The impact of these break downs on delay however are much larger. The standard deviation on the read delay is about 20% of the nominal, while it increases to 60% in the case of the write delay. In addition, the number of soft breakdowns present in the matrix also affects the variation range and distribution of the matrix delay. Such effects can be clearly observed in Fig. 5 which shows the cumulative density functions of the cell matrix delay in the case of one, two or three individual transistors suffering SBDs. The results are obtained via transistor-level simulations of the matrix assuming negligible process variability. The slopes of the cumulative functions indicates the degree of uncertainty, the "slower" the slope the larger the uncertainty and vice versa. Initially no break downs have occurred and the delay of the matrix is completely deterministic. For an increasing number of SBDs it is interesting to note that the delay variation range increases and leads to a more evenly distributed delay over the range. But the mean value of the delay also shifts for a different number of break downs. Moreover, in this case of an SRAM matrix, additional SBDs always increase the mean and the second moment of the delay distribution. The conclusion for this example is that both delay and spread deteriorate for each new SBD suffered. The mechanism behind this is simply due to the increasing interactions between SRAM sub-circuits that have suffered a SBD. For instance, the interaction between a defective SRAM cell and sense amp in the same column during the read operation not only increases the delay variation range, but also leads to a larger uncertainty in delay. Adding the impact of random process variability on delay on top of the fore-mentioned figures gives a perspective on the scale of the real problem. The circuit topology and context are extremely important in determining which circuit metrics will be influenced by degradation mechanisms and which will be unaffected.

Finally, the effect of the application running on the hardware and consequently the bit-level activity that defines the operating voltages of the devices and interconnect is essential to fully characterize the actual impact that the

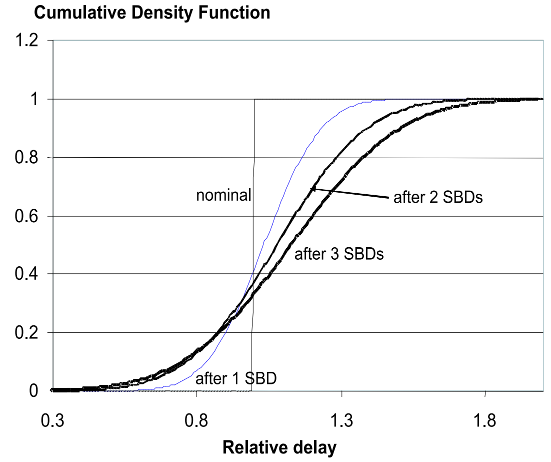

Fig. 5. Cumulative distribution of SRAM matrix delay variation under SBD

reliability effects will have in the time-dependent parametric variations of the system. Trying to characterize this impact at design time becomes extremely difficult, if not impossible in sub-45 nm technologies using existing commercial tools and design flows. Todays worst case analysis and system design paradigms are breaking down in the presence of the increasing dynamism which is present in the modern application in both the multimedia and wireless domains. The way reliability problems appear within the circuit is a rather random process and it depends on the actual operating conditions: time, temperature and stress voltages [7]. This is especially true for large circuit and systems featuring many transistors which can undergo significantly different stress conditions when executing dynamic applications. The actual location of the progressive defect and severity degree is hard to estimate at design time in this case. Moreover due to the varying nature of the stress induced by the application the defect generation rate also becomes very difficult to capture unless this is done at operation time (run-time). These facts simply indicate that innovation in circuit and system level design and analysis has to take place to counteract the impact that progressive parametric degradation mechanisms will have in the actual useful life-time of the system.

For the past decades variations have always existed on critical parameters during the design and operation of electronic systems. The most common such parameters are temperature, activity and other operating conditions. The circuits must always operate within the specified performance constraints for a given range of temperature and humidity conditions. In recent years, variations have also been observed in the electrical parameters, like capacitance, drive current etc., of the transistors and wires due to tolerances during the processing of the wafers. The conventional solution for dealing with these variations is to incorporate worst-case margins so that the circuit will always meet the target constraints under all possible specified conditions. The minimum and maximum value of each varying parameter is characterized and the combinations of these values for all the parameters form the corners of the parameter space which defines the working conditions of the design. Designers typically tune their designs to meet the performance constraints for all the corner-points, this technique is called corner-point analysis.

This technique is still widely used in the industry, but it suffers from a number of disadvantages. The corner points are usually very pessimistic; it is extremely unlikely that all the parameters will have their maximum or minimum values simultaneously. Thus, the design margins required to make the circuits operational under all corner conditions are excessive. Furthermore, the number of parameters affected by time-dependent variability becomes very large. This means that circuit designers will have to deal with parameter spaces of many dimensions and extremely large numbers of corner points. Finally, corner-point analysis techniques cannot handle the impact of intra-die time-dependent variability, which is spatially uncorrelated in nature [17], because the electrical parameters of each transistor would become an additional axis in the parameter space and the complexity would become unmanageable. So similar to the evolution at the system level, also here the worst case design paradigm is breaking down.

The most prominent alternative for corner-point analysis, which is already finding its way into the design flows of the major companies of the consumer electronics segment, is Static Statistical Timing Analysis (SSTA). Instead of just working with the value ranges of each electrical parameter, SSTA works with the statistical distribution of each of the parameters. Standard cell libraries are calibrated in order to correctly reflect the impact of variability on the transistor threshold voltage, beta and other electrical parameters on the delay of the standard cells. Then the delay of the complete circuit is estimated by statistically adding the delays of the critical path standard cells. This opens an entirely new perspective to circuit designers. Instead of blindly trying to achieve functional and parametric compliance in all corner points, they can evaluate the sensitivity of the design margins on the timing yield of the circuit. Thus, designers can trade-off the magnitude of the required design margins against the parametric yield of the circuit in a qualitative manner. Accepting some parametric yield loss can significantly limit the required margins, which is beneficial for energy consumption and area.

Mani et al. [18] have quantified the impact of corner-point analysis and statistical analysis on the power consumption, performance and yield of small logic circuits comprising a few hundred gates for the 130 nm technology node. They have assumed a limited impact of variability on the performance characteristics of the gate, a 25% delay variation in terms of $3\sigma/\mu$, which was reasonable for the 130 nm technology node. In their paper they demonstrate that in order to achieve a yield of 3σ (99.73%) using statistical timing analysis, squeezing the last 5.5% out of the circuit delay to meet the performance constraint incurs a power overhead of about 65% even for a small circuit. The overheads that corner-point analysis incurs, on the other hand, are about 30% larger on average. This illustrates one of the walls that circuit designers have to face due to the increased variability. The larger spreads of the delays due to variability lead to a need to excessively over-design the circuit, so that the nominal or average delay becomes much faster than the target. This headroom between the average and the target delay is there to absorb the spread due to variability. But faster circuits consume more energy, so an implicit energy consumption vs. timing yield trade-off exists for a given performance specification. Furthermore, fundamental limits exist for the maximum speed of circuits. Increasing the transistor sizes, for instance, fails when self-loading exceeds the output load. Further increases in transistor sizes lead to degrade energy consumption and delay.

In the meantime, variability on the electrical characteristics of devices and wires and hence of the circuits is growing in magnitude as technology scales. Moreover it is becoming randomly time-dependent as illustrated in the previous section and verified by the results in Fig. 5 due to the more progressive degradation of the key electrical parameters of devices and wires. As a result the uncertainty region collecting the actual electrical properties of the devices/wire will move randomly in space as time progress. This leads to a new global region of uncertainty resulting from the collection of the local variability "clouds" (see Fig. 6) which becomes far bigger than the corresponding one right after manufacturing. In the conceptual view in Fig. 5, t0 represents manufacturing time and t1,t2 represent moments in time during the product normal operation in the field. The 1 sigma, 2 sigma and 3 sigma contours correspond to iso-yield boundaries. It is clear from the above discussion, that the various degradation mechanisms will force the initial uncertainty cloud to shift in different directions as well as increase in magnitude. The region of uncertainty that is relevant for the designer is not just the initial (t0) cloud, but rather the aggregate area of all the clouds, because the design may be situated in any of these points during its life time. If the total cloud becomes too large, the possibility exists that it will be impossible to design a circuit for a given combination of performance and power budget constraints.

4 Impact of Time-Dependent Variability and Progressive Degradation in System Design

This increase in uncertainty has a very significant impact on system design as well. It effectively means that the system architect should build a system out of components that have unpredictable performance and quality metrics (that cannot even be fully bounded at design time anymore) as well as limited reliability guarantees. Conventional system design optimization techniques include trading-off energy consumption for performance at design-time, where most options are available at the component level. For instance, if a system has to meet a given clock frequency target, memories from a high-speed memory library might be used instead of slower low-power memories to guarantee sufficient timing slack. Typically components that are significantly faster than the given requirement are used in order to guarantee the parametric system target is met with

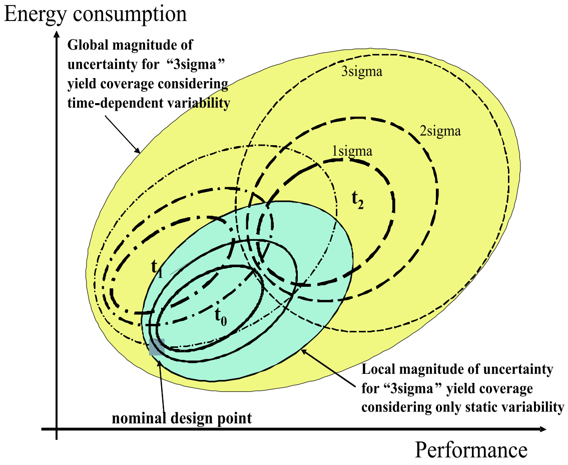
Energy consumption
Global magnitude of uncertainty for "3sigma" yield coverage considering time-dependent variability
3sigma
2sigma
1sigma
t2
t1
t0
Local magnitude of uncertainty for "3sigma" yield coverage considering only static variability
nominal design point
Performance

Fig. 6. Evolution of the uncertainty region of the system-level energy consumption and performance

reasonable yield. This is a worst-case margin that is usually added by system designers on top of the worst-case circuit tuning already performed by circuit designers. However, the large performance and energy consumption uncertainty at the component level combined with the requirement for very high yield forces designers at all levels to take increasingly larger safety margins. Stacking all these margins leads to systems that are nominally much faster than required and hence, much more energy hungry and potentially costly as well. It becomes clear that using margins is an acceptable solution only if we can give up on one of the major embedded system requirements (real-time performance, low energy consumption, low cost, high yield). Design margins trade-off energy consumption for performance, redundancy trades off cost for yield, parallelism trades off cost for performance and so on. No solution exists, however, that can optimize all these cost metrics simultaneously.

Furthermore, it is not yet known whether the corner points for each of the varying parameters will be fully characterizable, because they will depend on the detailed operating conditions on each device, like activity and stress conditions on the transistors and wires. Furthermore, these operating conditions heavily depend on the applications that are running on the system and the way they use the system resources. This means that the corner points and the distributions of each parameter, which guide the corner point and the SSTA analysis and optimization techniques respectively, will not be available anymore at design time. The only reasonable way out in the current design flows is to add second order design margins, namely on the place of the corner points to tackle the uncertainty due to time-dependent variability. Putting the fore-mentioned results of the SSTA technique in perspective of this unpredictability of the magnitude of the growing time-dependent variability, we conclude that design-time tuning of the circuit will be impossible for the target constraints of real-time performance, low energy and low cost.

5 Inadequacy of State of the Art Solutions

Even though both variability and reliability mechanisms affect the quality metrics of the same transistors and wires, the communities working on processing and reliability aspects at the technology level are different and usually disjoint. Plenty of literature exists in the process technology community about the sources and impact of variability in devices. At the technology level though little can be done to reduce the magnitude of random intra-die process variability. Random dopant fluctuations, for instance, are an unavoidable side-effect of the shrinking dimensions due to the limited amount of dopant atoms in the channel region of the transistor. Thus this type of variability has to be dealt with by the design community. The reliability community, on the other hand, generally focuses on the impact of the physical breakdown and degradation mechanisms on individual transistors and interconnects in typically small circuits and test structures which are not fully representative of the design reality. The main assumption there is the classical way of reliability lifetime prediction, which is based on extensive accelerated testing and extrapolations toward real operating conditions, design sizes and time scales. But the reliability community typically fails to also take into account the impact of random variability, since few test structures are used and statistics on manufacturing imperfections cannot be extracted with sufficient confidence.

Circuit and system designers have always been confronted with process variability and reliability degradation issues especially in the analog domain. A variety of alternative solutions has been developed in the previous years to deal with them. Good examples of such solutions include the one-time post fabrication tuning and binning technique, adaptive body bias, statistical static timing analysis, asynchronous design styles, architectural error detection and correction techniques and redundancy mechanisms, among others.

Binning has been the most popular technique used in general purpose microprocessors to deal with fabrication process induced inter-die variations. Instead of clocking every chip (of the same design) at the same frequency, the capable frequency of a chip is decided after fabrication with the help of at-speed testing. In parallel, chip-level supply voltage (V_{dd}) and body-bias voltage (V_{bb}) can be adjusted so as to increase the percentage of chips that can meet the design target frequency [19]. As frequency, V_{dd}, and V_{bb} are coarse chip-level controls, this method is not suitable to deal with stochastic intra-die variability, which requires some of control parameter that operates at a much finer granularity level.

Prevailing worst-case design methodologies use best-case and worst-case process corners to predict the impact of intra-die variability and enable potential optimizations. But they also fail in handling the complete problem in a generic manner. Static timing analysis (STA) which computes the critical path delays and hence clock period uses a single worst-case gate delay, which is the result of the most pessimistic corner for delay. As corners move farther and farther apart due to the increasing random intra-die variability component, STA based design incurs significant overheads (in terms of area/delay/energy depending upon the specific design objectives) which could jeopardize the scaling benefits. Statistical STA (SSTA) exploits the fact that device parameters and hence gate delays are stochastically distributed. As a result the path delay is much smaller than the sum of worst-case delays due to the averaging effect [20–22] of adding statistical distributions. SSTA calculates path delay distributions and hence the clock period distribution, which allows trade-offs between parametric timing yield and performance. Use of SSTA also improves the efficacy of circuit optimizations, such as circuit sizing under intra-die uncertainty [18]. But it suffers from a major drawback: it can only handle sequential or combinational logic circuits comprising standard cells, which is usually only a small part of current embedded system designs.

Razor [23] is a micro-architectural error technique based on dynamic detection and correction of circuit timing errors. The key idea of Razor is to tune the supply voltage by monitoring the error rate during circuit operation, thereby eliminating the need for voltage margins. A Razor flip-flop is introduced that double-samples pipeline stage values, once with a fast clock and again with a time-borrowing delayed clock. A meta-stability-tolerant comparator then validates the latch values sampled with the fast clock. In the event of a timing error, a modified pipeline misspeculation recovery mechanism restores the correct program state. This solution can guarantee correct I/O functional behavior of the processor pipeline. But it works on the principle of error detection and correction, so the timing at the application level cannot be guaranteed because the number of faulty cycles cannot be a priori known. So this is not directly portable to real-time embedded systems.

Asynchronous design styles produce circuit implementations that are inherently very robust toward local performance uncertainties [24]. Functionality in terms of correct input/output behavior of the circuit can be easily guaranteed, since no synchronization boundaries exist to create timing violations. Their major drawback is that their actual performance is completely unpredictable, thus mapping real-time applications on asynchronous circuits is very difficult.

Redundancy has been a popular technique to tackle reliability concerns in the past. Historically designers have been treating reliability degradation mechanisms as a pure functional concern and hence built reliability support by exploiting one (or some combination) of three forms of redundancy: information, hardware or time [25]. Use of information redundancy, such as parity or error correction codes (ECC), allows detection and/or correction of certain classes of bit errors. Systems achieve hardware redundancy by carrying out the same computation on multiple, independent hardware units at the same time and corroborating the redundant results to expose errors. Systems with triple (or higher) redundancy can obtain a correct answer through a majority voting scheme. Time redundancy techniques are based on redundant computation in time, they repeat the same operation multiple times on the same hardware. They mostly target to counteract soft errors, but they cannot handle catastrophic failures in a circuit. All forms of redundancy, however, come with a large associated overhead. Time redundancy incurs a significant delay penalty, which is not acceptable in the domain of real-time performance embedded systems. Hardware redundancy, on the other hand, incurs significant area overheads and does not provide adequate solutions. Time-dependent variability influences both the performance characteristics of processing elements and memories as well as those of communication networks. Thus communication becomes the weak link of the system. Existing redundancy solutions rely on perfect communication between the various degrading blocks in order to find an optimal assignment of tasks to system resources. Moreover, the new degradation mechanisms incur parametric drifts in all the utilized system components, thus they will all degrade uniformly. This makes it impossible to detect which redundant component has a "defect". Finally, existing testing fault models are not appropriate for dealing with the parametric degradations, because they have been developed for catastrophic defects that impact one or a few of the redundant layers [26]. In the case of parametric time-dependent variability all the layers will be affected, thus conventional redundancy solutions cannot be applied. In conclusion, redundancy techniques are only suited to partly deal with functional reliability issues, not with parametric ones.

All the fore-mentioned techniques, however, were developed to tackle the manifestation of variation and degradation mechanisms of past technology nodes. Post-fabrication tuning and binning techniques, for instance, are very successful at recovering dies that suffer from systematic variations, like die-to-die and wafer-to-wafer variations etc. Coarse-grain redundancy mechanisms based on majority voting can easily overcome malfunctions in limited parts of the design, due to failures related to sudden break downs of parts of the design. But the nature of the currently prominent process variability and reliability degradation effects has changed significantly by scaling feature dimensions into the DDSM regime. Systematic process variations are being overshadowed by random spatially-uncorrelated intra-die variability. Binning and adaptive body bias techniques cannot tackle the impact of variability on the quality metrics of the design, because they operate at a very coarse-grain level thus failing to deal with the spatial dynamics of variability. Reliability degradation mechanisms, on the other hand, are shifting from effects causing abrupt failures which are catastrophic for the circuit operation to gradual and graceful degradations of the circuit performance and energy consumption during normal operation. Redundancy mechanisms fail to provide adequate solutions for these new effects, since all the redundant components of the design will also degrade along with the original ones if they are used in parallel, thus providing negligible improvements in the product life-time.

It becomes clear that even though partial solutions for intra-die process variability and reliability issues are being worked out, solutions that can deal with the combined impact of time-dependent variability have not gained attention yet by the research community. On the other hand, both effects manifest themselves as parametric drifts in the timing and the energy consumption of the devices. Their combined impact can also be described as time-dependent variability. For any solution to be adequate, especially for real-time embedded systems, it will have to deal with the run-time temporal shifts in the performance metrics of the devices and circuits.

6 A Paradigm Shift in System Design Solutions

One of the main reasons why the existing solutions are breaking down in the case of time-dependent variability is that they try to tackle both the functional and the parametric issues at the circuit level with clear performance constraints on meeting the target clock period. This means that all the system components are designed so as to be functional and satisfy the frequency performance constraints with minimal performance variations to achieve maximum parametric yield. This forces the designers to design for the worst-case, since all the components should meet the common clock period constraint. In reality, the performance of each system component will follow a statistical distribution if margins are not embedded in its design, see [27] for a case study on on-chip memories. Some components will be faster than the mean performance and some will be slower, due to the nature of the statistics of their performance. This variation is not exploited in state-of-the-art techniques dealing with variability issues at the system level. Instead all the components are designed to have a predictable performance, even though this incurs a significant energy overhead. Meeting the constraints of low energy, low cost and real-time performance for maximum yield will become impossible with the conventional techniques, if the magnitude of uncertainty due to time-dependent variability increases. A paradigm shift will be required both in the design of the circuits and at system level design to overcome these limitations.

Current commercially available design and modeling flows are just starting to incorporate SSTA techniques to incrementally reduce the required design margins. For the transition to technology nodes where time-dependent variability becomes prominent, these flows will have to be extended significantly. Specifically, new statistical techniques will have to be developed to cover two main holes of the existing techniques. The first hole is the lack of dynamic energy calculation in the existing SSTA techniques, currently they can only estimate timing and static energy consumption. Total energy consumption is an extremely important metric for the design of battery-powered embedded systems, even more important than timing in some cases. The second required extension is a move to a higher abstraction level [28]. SSTA today deals with combinational or sequential logic blocks. Systems, however, are heterogeneous in their composition, memories and other IP blocks take up a very significant part of the die. Statistical techniques should move one level higher and they should be able to provide complete modeling of the entire die and an estimation of the timing, dynamic and static energy consumption as well as parametric yield for the complete system. An initial attempt to cover this gap has been outlined in [29]. A Variability and Reliability Aware Modeling (VRAM) framework exhibiting all the fore-mentioned attributes is required, which can be used in parallel to the existing design flow. A potential instance of such a framework can be seen in Fig. 7. It will aid designers in characterizing the impact of random variability and degradation mechanisms on the specific design and evaluate whether the impact on the design performance and quality metrics is severe. Such a framework would enable the quantitative evaluation of the magnitude of the potential problem and supply all the relevant information for designers to decide whether the problem is significant and which solutions are appropriate.

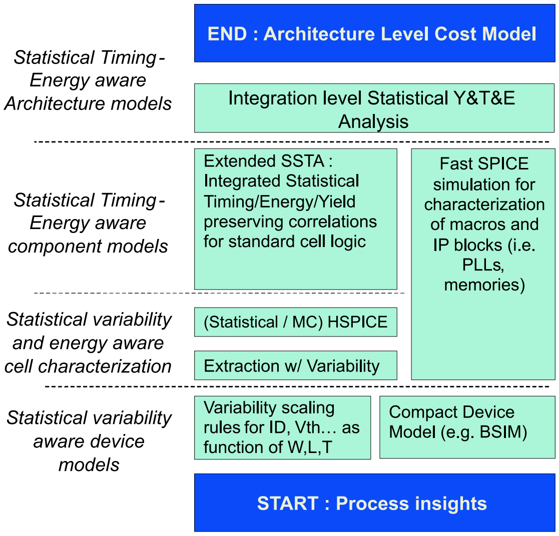
END : Architecture Level Cost Model
Statistical Timing-Energy aware Architecture models
Integration level Statistical Y&T&E Analysis
Statistical Timing-Energy aware component models
Extended SSTA : Integrated Statistical Timing/Energy/Yield preserving correlations for standard cell logic
Fast SPICE simulation for characterization of macros and IP blocks (i.e. PLLs, memories)
Statistical variability and energy aware cell characterization
(Statistical / MC) HSPICE
Extraction w/ Variability
Statistical variability aware device models
Variability scaling rules for ID, Vth… as function of W,L,T
Compact Device Model (e.g. BSIM)
START : Process insights

Fig. 7. An instance of a complete modeling flow for propagating variability and reliability information from the technology level to the complete system level.

If the problem is deemed significant enough to require a solution, one of the necessary steps is to separate the functional issues from the parametric issues, like performance and energy consumption. Circuit designers should deal with making circuits that are robust enough to remain functionally correct independently of the degree of time-dependent variability impact, because it may be impossible anyhow to fully characterize that at design time. The previous section has already outlined a number of existing methods for tackling functional issues. Solutions for functional degradations due to reliability based on redundancy and other techniques that enhance robustness are already available. Another example of a circuit level technique to design robust SRAMs cells under variability can be found in [30]. Asynchronous logic is another way of implementing functionally robust circuits against time-dependent variability. The parametric constraints can be ignored at this phase in favor of finding a functional solution for larger uncertainty ranges. This approach relieves the circuit designers of the pressure to meet performance requirements; the target is to design functional circuits under potentially extreme time-dependent variability with minimal overhead in energy consumption and delay. The only additional requirement from the circuit designers is that they should equip their designs with circuit level configuration/tuning parameters, which can trade-off performance for energy consumption at the circuit level, see [31] for an example.

Meeting the performance and energy budget constraints is the responsibility of the system itself. Only when the exact impact of time-dependent variability on the performance of the individual components and the short-term performance constraints are known, can an optimal solution be found. This implies that the actual performance of all the components will have to be measured after fabrication and at regular intervals via in-situ monitors in order to implement the required system observability. In a second step, if the actual performance of some components is lower that the required local timing constraint, the system should be able to influence it via the supplied tuning parameter. A very popular system level tuning parameter in current electronic systems is V_{dd} scaling. By lowering the supply voltage a system or component can decrease its energy dissipation while also reducing its performance and vice versa. But V_{dd} scaling is losing its efficiency due to the reduction of the voltage headroom, thus the required tuning parameters should be designed in the circuits to be more effective. An additional advantage of circuit level parameters is their local scope which is necessary in order to compensate for random variability, as opposed to parameters of global scope like V_{dd} scaling. Such tuning parameters, which we call knobs, provide the necessary controllability over the performance of the individual components and the system overall. The existence of knobs and monitors (K&M) in all, or a few critical, system components along with a simple algorithm for the knob control enables the system to find at run-time the optimal configuration setting for each of the components in order to minimize any given cost function, like timing violations or excessive energy consumption. This eliminates the need for allocating large design time margins so as to make sure that components always meet the most aggressive timing constraints, which is common practice today. Figure 8 illustrates an example architecture which utilizes configurable memories, monitors and an instance of a hardware controller for the tuning of the memories.

If this simple control algorithm does not provide enough range in the timing axis for mitigating the impact of time-dependent variability on performance, a more elaborate solution is needed which involves a more complex control algorithm. Namely the timing constraints can be moved from the level of a clock cycle to the level of application deadlines. Given that the components are designed without unnecessary design margins, their average performance will be faster than the one of components with margins but much more unpredictable. Some components will be faster than the clock frequency and some will be slower. Even though some components will violate the nominal frequency target, the average performance of all the components could still be faster than the target. Thus,

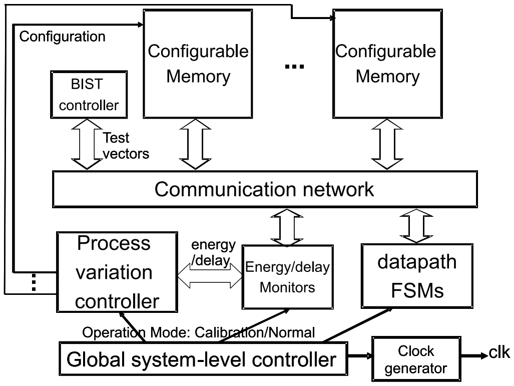

Fig. 8. Instance of a system architecture employing configurable components (memories), monitors for in-situ measurements and a controller for tuning the components.

over a number of cycles the application deadline can still be met, even though some clock level "deadlines" will be violated. This solution does not require system designers to resort to asynchronous logic. The conventional synchronization boundaries can be preserved as long as the clock frequency can be slowly adapted to the speed of the slowest component that is used at each moment in time. This can be achieved via dynamic frequency scaling or fine grain frequency islands, similar to the Globally Asynchronous Locally Synchronous principle. In combination with the use of the knobs that can fine-tune the component performance, a solution that globally meets the application deadline constraints can be achieved.

Energy consumption minimization is equally important to meeting the real-time performance constraints for embedded systems. It is mainly influenced by two factors, time-dependent variability and design margins. Variability introduces side-effects like unnecessary switching overhead and additional standby energy and its impact can only be partially mitigated at the technology level, so the system will have to live with these overhead situations. The second source of additional energy consumption is the design margins themselves. Designers control the magnitude of the margins; separating the functional from the parametric issues will allow the use of smaller margins which will result into more energy efficient system implementations.

A solution method based on the above principles has been outlined in [32] and an implementation in [33]. It is based on the assumption that the performance unpredictability is not completely tackled at the component level. When this unpredictability can be tackled at the circuit level with acceptably small overheads it makes sense to provide circuit solutions. But in many cases the resulting overheads are unacceptably high, especially in energy. In that case the system has to be exposed to the performance unpredictability to enable a reduction of the circuit energy and delay overhead due to the margins, by providing system level solutions for variability. The individual component performance and energy consumption is measured after fabrication by on-chip monitors and relevant component level configuration options are adapted by the system in order to meet the real-time performance requirements of the application with minimal energy and area overhead. In [32] the solution is only activated after processing to increase the initial processing yield. But once it is in place the same approach can be used infrequently, e.g. once every few seconds, to check whether the energy or timing is lower for the alternative path. This is still a reactive approach though and it will not solve all degradation problems in a fully optimal way. But the big advantage is that is it not that difficult to implement in existing design flows. Future work should look at more optimal global paradigm shifts. A further extension of this technique tackling the impact of time-dependent variability in the context of dynamic application has been reported in [34]. It uses the concept of application scenarios to handle the unpredictability coming from the intrinsic dynamism of the application or the user interaction.

In summary, time-dependent variability will require a paradigm shift in the design of electronic systems in order to benefit from the area scaling opportunities offered by technology scaling without excessive energy and performance overheads. A shift toward Technology-Aware Design solutions, which take into account the process imperfections early in the design cycle, will be required to design and fabricate embedded systems that will meet the constraints in all four major cost criteria: energy consumption, real-time performance, area/cost and yield/guaranteed lifetime.

7 Conclusions

Scaling to sub 45 nm technology nodes changes the nature of reliability effects from abrupt functional problems to progressive degradation of the performance characteristics of devices and system components. Process technology can no longer alleviate their impact on the performance and energy consumption at the design level. Moreover, existing design flows cannot evaluate this impact due to the lack of modeling tools, let alone provide adequate solutions. Tackling time-dependent variability will necessitate a paradigm shift for embedded system design in order to meet the power, timing and cost constraints with acceptable yield and life-time guarantees.

Acknowledgments

The authors would like to acknowledge the insightful suggestions and contributions of Guido Groeseneken, Ben Kaczer, Christophe Bruynseraede, Zsolt Tokei, Robin Degraeve, Michele Stucchi, Philippe Roussel, Satya Munaga, Pol Marchal and Bart Dierickx.

References

1. Maex, K., Stucchi, M., Bamal, M., Grossar, E., Dehaene, W., Papanikolaou, A., Miranda, M., Catthoor, F.: Technology aware design and design aware technology. Proc. of Intl. Conf. on Integrated Circuit Design and Technology, 77-81 (2005)
2. Groeseneken, G., Degraeve, R., Kaczer, B., Roussel, R.: Recent trends in reliability assessment of advanced CMOS technologies. Intl. Conf. on Microelectronic Test Structures, 81–88 (2005)
3. Reddy, V., Krishnan, A., Marshall, A., Rodriguez, J., Natarajan, S., Rost, T., Kirshnan, S.: Impact of Negative Bias Temperature Instability on Digital Circuit Reliability. Intl. Reliability Physics Symp., 248–254 (2002)
4. Bruynseraede, C., Tokei, Z., Iacopi, F., Beyer, G., Michelon, J., Maex, K.: The impact of scaling on interconnect reliability. Intl. Reliability Physics Symp., 7–17 (2005)
5. Tokei, Z., Li, Y., Beyer, G.: Reliability challenges for copper low-k dielectrics and copper diffusion barriers. J. of Microelectronics Reliability, 1436–1442 (2005)
6. Schroder, D., Babcock, J.: Negative bias temperature instability: road to cross in deep submicron silicon semiconductor manufacturing. J. Applied Physics, **94**, 1–18 (2003)
7. Stathis, J.: Physical and predictive models of ultrathin oxide reliability in CMOS devices and circuits. IEEE Trans. on Device and Materials Reliability, **1**, 43–59 (2001)
8. Kaczer, B., Degraeve, R., O'Connor, R., Roussel, P., Groeseneken, G.:Implications of progressive wear-out for lifetime extrapolation of ultra-thin SiON films. Intl. Electron Devices Meeting, 713–716 (2004)
9. International Technology Roadmap for Semiconductors, Interconnect chapter, www.itrs.net (2005)
10. Croon, J., Storms, G., Winkelmeier, S., Pollentier, I., Ercken, M., Decoutere, S., Sansen, W., Maes, H.: Line Edge Roughness: Characterisation, Modelling and Impact on Device Behaviour. Intl. Electron Device Meeting, 307-310 (2002)
11. Ramadurai, V., Rohrer, N., Gonzalez, C.: SRAM operational voltage shifts in the presence of gate oxide defects in 90 nm SOI. Intl. Reliability Physics Symp., 270–273 (2006)
12. Kaczer, B., Degraeve, R., Crupi, F., De Keersgieter, A., Groeseneken, G.:"Understanding nMOSFET characteristics after soft breakdown and their dependence on the breakdown location. European Solid-State Device Research, 139–142 (2002)
13. Kaczer, B., Degraeve, R., Rasras, M., De Keersgieter, A., Van de Mieroop, K., Groeseneken, G.: Analysis and modeling of a digital CMOS circuit operation and reliability after gate oxide breakdown: a case study. Microelectronics Reliability, **42**, 555-564 (2002)
14. Wang, H., Miranda, M., Catthoor, F., Dehaene, W.: On the combined impact of soft and medium gate oxide breakdown and process variability on the parametric figures of SRAM components. Intl. Wsh. on Memory Technology, Design and Testing, 71–76 (2006)
15. Carter, J., Ozev, S., Sorin, D.: Circuit-Level Modeling for Concurrent Testing of Operational Defects due to Gate Oxide Breakdown. Design Automation and Test in Europe, 300–305 (2005)
16. Avellan, A., Krautscneider, W.: Impact of soft and hard breakdown on analog and digital circuits. IEEE Trans. on Device and Materials Reliability, **4**, 676–680 (2004)

17. Najm, F.: On the need for statistical timing analysis. Design Automation Conf., 764–765 (2005)
18. Mani, M., Orshansky, M.: A new statistical optimization algorithm for gate sizing. Intl. Conf. on Computer Design, 272–277 (2004)
19. Tschanz, J., Kao, J., Narendra, S., Nair, R., Antoniadis, D., Chandrakasan, A., De, V.: Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage. IEEE Journal of Solid-State Circuits, **37**, 1396–1402 (2002)
20. Viswewariah, C.: Death, taxes and failing chips. Design Automation Conference, 343–347 (2003)
21. Keutzer, K., Orshansky, M.: From blind certainty to informed uncertainty. Wsh. on Timing Issues in the Specification and Synthesis of Digital Systems (TAU), 37–41 (2002)
22. Kang, K., Paul, B., Roy, K.: Statistical timing analysis using levelized covariance propagation. Design Automation and Test in Europe, 764–769 (2005)
23. Austin, T., Blaauw, D., Mudge, T., Flautner, K.: Making typical silicon matter with Razor. IEEE Computer, **37**, 57–65 (2004)
24. Sparso, J., Furber, S.: Principles of Asynchronous Circuit Design: A Systems Perspective. Kluwer Academic Publishers (2001)
25. Iyer, R., Nakka, N., Kalbarczyk, Z., Mitra, S.: Recent advances and new avenues in hardware-level reliability support. IEEE Micro, **25**, 18–29 (2005)
26. Brahme, D., Abraham, J.: Functional testing of microprocessors, IEEE Trans. on Computers, **C-33**, 475–485 (1984)
27. Wang, H., Miranda, M., Dehaene, W., Catthoor, F., Maex, K.: Impact of deep submicron (DSM) process variation effects in SRAM design. Design Automation and Test in Europe, 914–919 (2005)
28. Blaauw, D., Chopra, K.: CAD tools for variation tolerance. Design Automation Conference, 766 (2005)
29. Papanikolaou, A., Grabner, T., Miranda, M., Roussel, P., Catthoor, F.: Yield Prediction for Architecture Exploration in Nanometer Technology Nodes: A Model and Case Study for Memory Organizations. Intl. Symp. on HW/SW Co-design and System Synthesis, 253–258 (2006)
30. Grossar, E., Stucchi, M., Maex, K., Dehaene, W.: Statistically aware SRAM memory array design. Intl. Symposium on Quality Electronic Design, 25-30 (2006)
31. Wang, H., Miranda, M., Papanikolaou, A., Catthoor, F., Dehaene, W.: Variable tapered Pareto buffer design and implementation allowing run-time configuration for low power embedded SRAMs. IEEE Trans. on VLSI Systems, **13**, 1127–1135 (2005)
32. Papanikolaou, A., Lobmaier, F.., Wang, H., Miranda, M., Catthoor, F.: A system-level methodology for fully compensating process variability impact of memory organizations in periodic applications. Intl. Symp. on HW/SW Co-design and System Synthesis, 117–122 (2005)
33. Papanikolaou, A., Starzer, F., Lobmaier, F., Miranda, M., Catthoor, F., Huemer, M.: A system architecture case study for efficient calibration of memory organizations under process variability. Wsh. on Application-specific Processors, 42–49 (2005)
34. Sanz, C., Papanikolaou, A., Miranda, M., Prieto, M., Catthoor, F.: System-level process variability compensation on memory organizations of dynamic applications: a case study. Intl. Symp. On Quality Electronic Design, 376–382 (2006)