Encountering Gate Oxide Breakdown with Shadow Transistors to Increase Reliability

Claas Cornelius[1], Frank Sill[2], Hagen Sämrow[1], Jakob Salzmann[1], Dirk Timmermann[1], Diógenes da Silva[2]

[1] Institute of Applied Microelectronics and Computer Engineering, University of Rostock
Richard-Wagner-Str. 31, 18119 Rostock, Germany

[2] Department of Electrical Engineering, Federal University of Minas Gerais
Av. Antônio Carlos 6627, CEP: 31270-010, Belo Horizonte (MG), Brazil

Email: claas.cornelius@uni-rostock.de or franksill@ufmg.br

ABSTRACT

Device scaling has enabled continuous performance increase of integrated circuits. However, severe reliability and yield concerns are arising against the background of nanotechnology. Traditionally, most causes and countermeasures were solely considered manufacturing issues, but lately, we have seen a shift towards operational reliability issues. Though, besides intense research on soft-errors and system-level approaches very little effort is put into low-level design solutions in order to enhance lifetime reliability. Hence, we demonstrate that redundant transistor insertion does improve system reliability significantly as regards Time-Dependent Dielectric Breakdown (TDDB). Furthermore, we introduce an algorithm which identifies the transistors being most vulnerable to TDDB. Subsequently, redundant transistors (called shadow transistors) are inserted at the previously identified instances. Lastly, we argue for applying high threshold voltage devices for the redundant transistors. Finally, we present results for a set of benchmark circuits and prove the combined approach successful. The enhanced designs were on average 41.8 % more reliable compared to the initial designs in respect of TDDB at the price of moderately increased power consumption and delay.

Categories and Subject Descriptors

B.8.1 [**Performance and Reliability**]: Reliability, Testing, and Fault-Tolerance; B.8.2 [**Performance and Reliability**]: Performance Analysis and Design Aids; C.4 [**Performance of Systems**]: – Design studies, Fault tolerance, Performance attributes, Reliability, availability, and serviceability

General Terms

Algorithms, Design, Performance, Reliability.

Keywords

Logic design, Gate oxide breakdown, Modeling, Redundancy, Nanotechnology, Transistor, Organic computing.

SBCCI'08, September 1–4, 2008, Gramado, Brazil.

1. INTRODUCTION

Performance of integrated circuits benefits from aggressive scaling of device sizes. By contrast, manufacturing yield and lifetime reliability decrease at alarming pace while the physical limits of nanotechnology are being approached. That is, reduced design and process error margins due to material defects and imperfections have to be dealt with. Additionally, miniaturization results in increased susceptibility to different kinds of failures during operation. Of major concern for reliability, performance and power consumption is the gate oxide which is the dielectric isolating the transistor input and the conducting channel. The thickness of gate oxide is less than 20 Å in current technologies and due to the small number of atomic layers in the oxide, non-ideal scaling of supply voltage and the increase of the electric field, gate oxide has become highly vulnerable. In combination with the rising transistor count per die, the probability for a system failure due to gate oxide breakdown grows with every new technology generation and is a crucial problem for design, manufacturing and operation of integrated systems.

Gate oxide breakdown is defined as the point in time when a conducting path is created between the gate and the substrate of a transistor [20]. This conducting path originates from two different mechanisms in respect of time. Firstly, abrupt damage because of extreme overvoltage ─ e.g. due to Electro-Static Discharge (ESD). Secondly, a rather slow destruction over time called Time-Dependent Dielectric Breakdown (TDDB). The physical root causes for TDDB are still subject of discussion in the research community whereas the thermochemical model and anode hole injection seem to be the most relevant effects [23]. However, during operation charge traps start to form in the gate oxide which eventually overlap to form a conducting path between gate and substrate. Once formed, a chain of events occurs: Conduction leads to increased current flow, current leads to heat, heat leads to thermal damage and thermal damage leads to even more charge traps which increase the conduction again. This positive feedback loop accelerates the gate oxide breakdown and ends with a failing transistor [4][23].

According to ITRS, tool assisted insertion of reliability into the design will become a key priority because systems are becoming too complex to be fully functionally tested [15]. Hence, we present a new approach for the automatic insertion of redundant transistors to encounter TDDB and to increase lifetime reliability. Thereto, we analyze various netlists and identify the most vulnerable transistors. At those instances, redundant transistors are added in parallel ─ called **shadow transistors** hereafter. This does not solely in-

crease lifetime reliability but also raises manufacturing yield because certain production failures can be compensated (e.g. stuck-open transistors).

2. RELATED WORK

Traditionally, reliability and yield engineering has been considered only as an issue of manufacturing. However, especially against the background of nanotechnology, reliability engineering has gained an increased attention by circuit designers as well. Established low-level approaches for defect yield enhancement are static reconfiguration of redundant modules and layout modification. In [1] both approaches are implemented to increase the yield of an adder structure. That is, static reconfiguration is achieved by the use of laser fuses for disconnecting defective bit slices and connecting spare ones. This is also common practice in nowadays memory manufacturing. Furthermore, layout modifications rearrange parts of the layout in order to optimize the design, e.g. duplication of vias, widening and length-minimization of wires, area compaction or layer reassignment [2].

A further issue is parametric yield loss where in contrast to defect yield all devices might work functionally correct but not within the given design requirements. The root cause is increased parameter variations in nanotechnology which means that desired feature dimensions cannot be reproduced repeatedly. Thus, even small variations of critical parameters like dopant concentration, channel length or gate oxide thickness have huge impact on performance, power consumption and reliability. Formerly, chips out of specification were rejected during manufacturing test. However, methods for sensing process perturbations have been proposed, e.g. phase detection, Modified Vernier Delay Line [5][22]. With such sensing in hand, adaptive design techniques can be applied to shift leakage currents and frequency into the required design constraints. Body biasing [22] is the most widespread technique, but other approaches to compensate for leakage and charge sharing have been proposed as well [5]. In recent years, research focus has shifted towards causes of soft errors, and applicable countermeasures have been investigated elaborately. Soft-errors are caused by radiation events (like neutrons, α-particles, electromagnetic interference) that result in charge variations of internal nodes and, thus, transient failures. Miscellaneous techniques have been published to avoid or recover from soft-errors. For instance, extra RC-filters and additional capacitances at vulnerable internal nodes have been proposed [10]. Another approach was published in [9] where debug resources of the scan path (redundant latches) are used to detect soft-errors resulting in very little design overhead.

Concluding, albeit the proposed techniques improve yield and soft-error resiliency, they do not tackle lifetime reliability which is already a major concern in integrated system design. Intrinsic reliability issues due to TDDB, electromigration and thermal cycling are additionally intensified by non-ideal scaling, increased transistor count, power densities and adaptive processing [19]. Therefore, dynamic reliability management is investigated which aims at postponing device failures by means of microarchitectural awareness of lifetime reliability [18]. A very different assumption is made in [17], assuming that device failure cannot be prevented and has to be resolved. Thereto, redundant transistors are added in parallel to vulnerable transistors while keeping the performance constant. The analysis is limited to yield enhancements under constant performance and considers only stuck-open faults. In this work we expand the idea of redundant transistors which we call **shadow transistors** hereafter. Our contribution will show that shadow transistors can deal with yield defects as well as lifetime reliability due to gate oxide breakdown. Additionally, we will introduce an algorithm that achieves enhanced reliability figures at the cost of minor degradation of delay and power. Furthermore, we show that applying high threshold voltage devices as shadow transistors increases reliability significantly at marginal extra costs.

3. APPROACH

The basic concepts and underlying assumptions used in this work are described in the following sections.

3.1 Fundamentals of Reliability

The term Reliability *R(t)* is to be understood as the probability of a system to perform as desired until time instance *t*. For instance, $R(t_x) = 0.8$ states that there is an 80 % chance that the system is still running at time t_x. Furthermore, the failure rate λ expresses the probability that a system fails in a given time interval. Assume that 10 out of 100 systems fail in a given year, hence, an individual system will fail with a probability of 10 % in that same year. For most cases, a constant failure rate λ is assumed over the useful system lifetime [7] so that reliability *R(t)* can be expressed by an exponential function (see equation 1). Closely related to the rather probabilistic expressions of reliability and failure rate is the Mean Time To Failure (MTTF), which is the average time that a system runs until it fails. Hence, it is equal to the expected lifetime and is expressed as the inverse of the constant failure rate λ (see equation 2).

$$R(t) = e^{-\lambda t} \tag{1}$$

$$MTTF = \int_0^\infty R(t)dt = \frac{1}{\lambda} \tag{2}$$

Even though, these equations are used in most calculations, it needs to be noted that the assumption of a constant failure rate λ is only valid for the regular lifetime. Thus, infant mortality as well as wear-out mechanisms are excluded and are for the most part described by Weibull distributions [7]. As we want to identify single transistors in a netlist that are susceptible to TDDB, we need to apply a quantitative measure for each transistor. Therefore, we calculate the $MTTF_{TDDB}$ in a first step which is the average time until a transistor fails due to TDDB. The applied equation is based on experimental work performed at IBM [24]:

$$MTTF_{TDDB} \propto \left(\frac{1}{V_{DD}}\right)^{a-bT} e^{\frac{X+\frac{Y}{T}+ZT}{kT}} \tag{3}$$

where V_{DD} denotes the supply voltage, *T* is the absolute temperature in Kelvin, *k* is Boltzmann's constant and a, b, X, Y as well as Z are fitting parameters. However, to be able to compare different designs and implementations, we also require a single quantitative measure for the reliability of a system which has to include all transistors and all failure mechanisms under investigation. Therefore, we adopt the model of summation [21] which requires making two assumptions:

1) All failure mechanisms as well as all transistors are independent and the whole system fails if any transistor fails.
2) Each of the failure mechanisms has a known lifetime model.

Then, the MTTF of the system can be formulated by means of the sum of the individual MTTFs:

$$MTTF_{sum} = \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{MTTF_{ij}}} \tag{4}$$

Here, $MTTF_{ij}$ denotes the MTTF for the i[th] failure mechanism (in this work only TDDB is considered) of the j[th] transistor. For example, $MTTF_{14}$ is the MTTF of the 4[th] transistor in respect to TDDB (with TDDB being the first failure mechanism).

3.2 Origination of the approach

The presented work originated from a discussion about a research project on organic computing (SPP-1183). In the following, Organic Computing (OC) is described very briefly and interesting similarities to reliable chip design are discussed. Organic computing describes the idea to analyze techniques found in nature and to emulate them within the field of computing [13]. Thereby, graceful degradation and the so called self-X behaviors are the most cited characteristics. The self-X characteristics include amongst others self-organization, self-adaptation and self-protection which are concepts likewise found in the design and operation of integrated circuits [8]. Systems with such approaches implemented, promise to be very robust provided that a certain level of redundancy is available.

The similarities for reliable systems shall be clarified by means of a wireless sensor network which generally includes numerous OC-characteristics. Such a network consists of plenty of sensor nodes that compute a joint task (e.g. measuring some physical parameter above a geographic stretch). Within the network topology, there is a certain fraction of sensor nodes that act as so called cluster heads. That is, these nodes administrate local clusters of sensors, aggregate data from the neighbors and transmit the data across other cluster heads towards the data sink [11]. The resulting topology can be seen as a directed acyclic graph similar to a netlist (see figure 1). It is obvious that the cluster heads are of increased importance for system reliability. There, the critical cause for failure is the limited power resource/battery (due to data aggregation and transmission) like TDDB is the critical cause in our work. Two solutions can be applied:

1) Adding redundancy: Wireless sensor networks include implicit redundancy as data can be transmitted along different routes and every sensor node can become a cluster head at any time. By comparison, only static redundancy can be added into a netlist when considering circuit design (such results are presented in section 4.2). This does not hold true for networks-on-chip where redundant computation and communication is at hand [3].
2) Reducing susceptibility: In the case of sensor networks the easiest solution is to increase battery capacity which delays the failure of the cluster heads. Likewise, circuit failures can be delayed by applying transistors with thicker gate oxide. The resulting implications for reliability, power and performance will be discussed in section 4.3.

Finally, failure of a subsystem needs to be detected in order to self-adapt to the situation. Such adaptation as regards sensor networks can be the rerouting of data or the new assignment of cluster heads. In terms of circuit design, such approaches would be Dynamic Voltage/Frequency Scaling (DV/FS) to adapt the system to declining performance or the shut down of defect redundant computation units. Such mechanisms will finally allow graceful system degradation rather then early and abrupt system failure.

3.3 Setup

For this work we performed reliability investigations on two different abstraction levels. Firstly, we simulated basic logic gates on transistor level with HSpice to demonstrate that shadow transistors can compensate for impaired transistors due to TDDB (presented in section 4.1). Thereto, we used modified predictive transistor models from a 65 nm technology ─ called *L-V*$_t$*/To* [16]. Furthermore, basic logic gates were characterized in terms of power consumption and delay for the various cases that one or several shadow transistors can be inserted. Secondly, we simulated various benchmark circuits based on the previously characterized logic gates and investigated the impact of shadow transistors on the critical design parameters. For the simulations of the different mechanisms of gate oxide breakdown, we chose equivalent circuits that can be used in standard CAD environments [14]. Figure 2 a) depicts for instance the electrical defect model of gate-channel breakdown. Here, the transistor under examination is split into two transistors in series whereas $w = w_1 + w_2$ holds for the gate widths. By varying the ratio of w_1 and w_2, the location of breakdown can be modeled. Moreover, the extent of breakdown can be represented by different values for the resistance R_{GC}. Hence, R_{GC} close to zero is referred to as a hard breakdown whereas $R_{GC} >> 0$ is called a soft breakdown.

Similarly, oxide breakdown between the gate overlap regions and the source/drain extensions are modeled by the equivalent circuit depicted in figure 2 b) ─ called gate-drain and gate-source breakdown, respectively. Unlike the gate-channel model, the location of impact can not be reproduced but is of no concern for this breakdown mechanism. In the literature, several more elaborated models exist but the chosen ones are fully sufficient for our purpose [12]. For our approach we had to extend equation 3 because the equation holds only true across all transistors if equal input signal probabilities are assumed (i.e. the input is a logical 1 for 50 % of the time). However, this is not realistic and it should be considered that TDDB depends on the absolute voltage level at the gate [20]. For instance, if no voltage is supplied to the gate of an NMOS (gate voltage $V_g = 0$), there is no electric field and TDDB does not occur. On the other hand, if $V_g = V_{DD}$ an electric field and TDDB are present and the $MTTF_{TDDB}$ is smaller compared to the general case with equal input probability. The same applies for PMOS, but with inverse voltage levels ─ the critical case is $V_g = 0$. Considering this additional aspect, we obtain the MTTF of the system as equation 5:

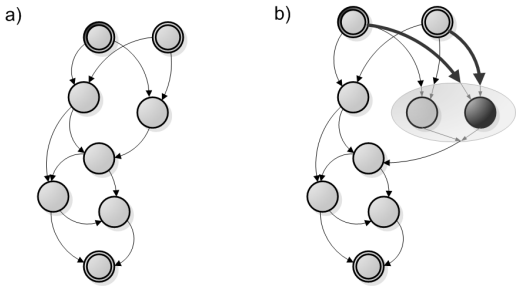

Figure 1. A netlist drawn as a directed acyclic graph: a) original and b) with added redundancy.

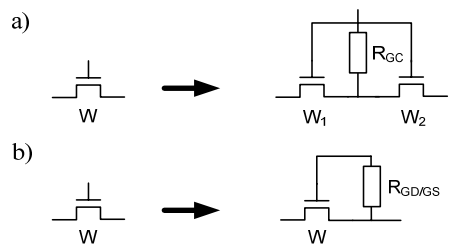

Figure 2. Underlying models for dielectric breakdown: a) gate-channel and b) gate-drain (gate-source).

$$MTTF_{system} = \frac{1}{\sum_{i=1}^{\#PMOS} \frac{1}{\frac{0.5}{1-P_{1,i}} MTTF_{L-Vt/To,i}} + \sum_{i=1}^{\#NMOS} \frac{1}{\frac{0.5}{P_{1,i}} MTTF_{L-Vt/To,i}}}$$

where *#PMOS* and *#NMOS* are the numbers of PMOS and NMOS transistors, respectively and $P_{1,i}$ is the probability of a logic 1 at pin *i*. $MTTF_{L\text{-}Vt/To}$ is the reference value for the chosen technology with equal input probability (calculated based on equation 3). Now, a quantitative measure can be calculated to express the reliability of the system. However, the equation represents a series failure system which is not the case for shadow transistors which are in parallel to their counterparts. Lastly, equation 6 considers the serial and parallel components separately:

$$MTTF_{system,ST} = \frac{1}{\sum_{i=1}^{\#PMOS,ST} \frac{1}{\frac{0.5}{1-P_{1,i}} MTTF_{ST}} + \sum_{i=1}^{\#NMOS,ST} \frac{1}{\frac{0.5}{P_{1,i}} MTTF_{ST}} + \frac{1}{MTTF_{system,w/o,ST}}}$$

with *#PMOS,ST* and *#NMOS,ST* being the number of shadow transistors (including their counterparts), $MTTF_{ST}$ accounting for the parallel structure of the shadow transistors and $MTTF_{system,w/o,ST}$ representing the MTTF for the rest of the circuit. For the sake of simplicity, we define the temperature T, supply voltage V_{DD} and the gate oxide areas to be constant across the benchmark circuits. For this simplification, the probability of gate oxide breakdown depends only on the signal probability and the transistor type.

3.4 Algorithm

The primary idea of our work is to identify the transistors that are most susceptible to TDDB and to add shadow transistors in parallel at those instances. This does not reduce the failure rate of the initial transistor but adds an additional path into the netlist (see figure 3 b). Thus, the probability that both transistors fail is significantly smaller and, consequently, the probability for system failure is reduced as well. The different steps for the insertion of shadow transistors are described in the following:

1) Appropriate input signal probabilities are applied to the design that is to be enhanced. Whereas signal probability (SP) does not stand for the probability of the signal to change but for the probability of the signal being a logical 1. For instance, 0.8 denotes that the signal is a logical 1 for 80 % of the time.
2) With the input probabilities as stimulus, the netlist is simulated and a signal probability is assigned to each pin of the netlist.
3) Based on the signal probabilities and the technology values for $MTTF_{TDDB}$, those transistors most vulnerable to TDDB can be identified.
4) After that, shadow transistors are added to the netlist wherever the signal probabilities exceed a limit defined by the critical threshold C_{crit}. As the damaging voltages are different for the gate inputs of NMOS and PMOS (see section 3.3), the following two conditions are applied:

```
FOR ALL i DO
  IF (SPi < Ccrit) THEN ADD_SHADOW_PMOS
  IF (SPi > (1-Ccrit)) THEN ADD_SHADOW_NMOS
END FOR
```

5) Lastly, the netlist is analyzed anew to gather design parameters like delay, power consumption and reliability. These new and the initial parameters give valuable hints for adapting the critical threshold C_{crit} to the desired needs and given constraints.

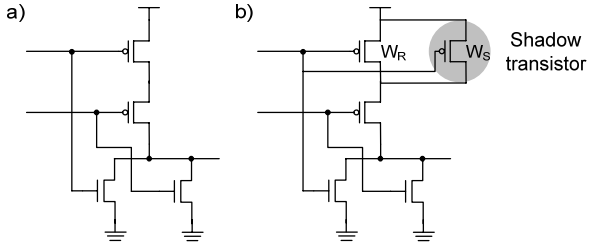

Figure 3. Example of a) a standard NOR2 and b) a NOR2 with a shadow transistor for the upper PMOS.

4. RESULTS AND DISCUSSION

This chapter presents the attained findings and discusses its implications.

4.1 Gate functionality and TDDB

As the positions of oxide breakdown are randomly distributed over the dielectric area, we considered a uniform distribution of the breakdown location along the channel length of our model [6]. Furthermore, we altered the resistance of the TDDB models in a reasonable range from 0 to 5.2 kΩ. Accordingly, we simulated the miscellaneous scenarios for inserting shadow transistors across the different logic gates and parameter sets. An example of a NOR2 is depicted in figure 3 b) with a redundant shadow transistor in parallel to the upper PMOS. The gate lengths of all shadow transistors were set to be identical to their according counterparts ($W_R = W_S$). It is worth to note that even though we concentrate on TDDB, the results are also valid for abrupt oxide breakdown (e.g. due to ESD).

The results for a centered gate-channel breakdown are plotted in figure 4 for the NOR2 gates shown in figure 3 a) and b). It can be observed that the gates with shadow transistor are rarely being affected except for the delay increase towards $R_{GC} = 0$. By contrast, the unprotected NOR2 does not even reach the correct logic levels for all cases. Furthermore, the delay increases exponentially towards the y-axis and exhibits already for soft breakdowns (R_{GC} rather high) an increased delay that can hardly be compensated in subsequent gates ─ thus, the unprotected NOR2 has to be considered defect. Unfortunately, the same conclusions do not apply to gate-drain/source breakdown where some scenarios also reveal a strong delay increase or do not reach the correct state at all. Let us consider an inverter with a complete gate-drain breakdown at the PMOS. Hence, the input signal is directly mapped to the output. In the case of an incoming 0 and a shadow transistor in parallel to the PMOS, the driving gate and the shadow transistor compete for the output whereas the result depends on the driving strength of the participating transistors.

Two more things need to be considered for logic gates with any kind of gate oxide breakdown. Firstly, such gates do not reach the full voltage swing which can mostly be recovered due to the level

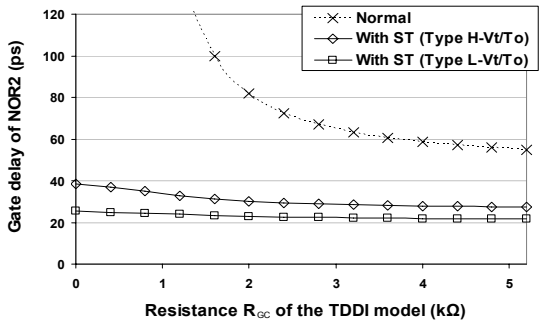

Figure 4. Results of different NOR2 versions in the presence of oxide breakdown above the center of the channel.

restoring characteristic of CMOS. Secondly, such gates increase in delay which can result in timing errors rather than logical errors. However, if adaptive techniques are applied (e.g. DFS), the defect module could work further on at reduced performance. The results of the miscellaneous simulations exhibited quantitatively identical behavior so that it can be concluded that logic gates with shadow transistors are fully functional in the presence of gate-channel breakdown. Though, with respect to gate-drain/-source breakdown shadow transistors increase reliability but can not compensate all scenarios. Thus, unreliable transistors do not necessarily lead to failing system behavior but system reliability depends on the degree of unreliable transistors, their position in the logic composition and on the degree of system-level adaptation.

To be able to evaluate different benchmark circuits in the next step, the logic gates with shadow transistors were also characterized without the failure models. In table 1, a couple of results extracted from these simulations are given exemplarily. The delays refer to a testbench with a fan-out of 4 (FO4) and also consider the slow-down of the fan-in gates due to the increased input capacitance. Thus, the impact on performance in a chain of gates is relatively small because the larger input capacitance is mostly compensated by the improved driving strength. Alongside the values for an inverter and a NOR2, the mean values of delay increase are given for the subset of two cases where all NMOS, respective PMOS, transistors of a gate receive a shadow transistor.

Table 1. Delay penalties for a few exemplary cases with shadow transistors (ST)

	INV	NOR2	Exemplary Mean Delays	
Normal	20.53 ps	28.77 ps	$t_{d,Ø}$	
With ST	21.52 ps	28.90 ps	$t_{d,Ø,NMOS}$ * 1.15	$t_{d,Ø,PMOS}$ * 1.3
			All NMOS resp. PMOS redundant	

4.2 Design enhancements

After it was shown that shadow transistors do compensate for a wide range of gate oxide breakdowns and after the logic gates were characterized with and without shadow transistors, we were able to implement several benchmark circuits. Figure 5 depicts the improvement of MTTF for different designs in relation to an unprotected design. The critical threshold C_{crit} was set to 0.3 for these figures which resulted on average in 22.8 % additional shadow transistors. Furthermore, the results for random insertion of STs are also given in the figure for comparison. Thereby, the same amount of shadow transistors was added to each design as with our algorithm. The figures in the diagram prove both the approach of inserting shadow transistors and the chosen algorithm successful. Independent from the algorithm, inserting shadow transistors raises the expected lifetime of an integrated circuit in respect of oxide breakdown. Moreover, our selective algorithm improves the MTTF for the given critical threshold C_{crit} on average by 13.9 %. Thus, the random insertion with 8.8 % improvement on average is clearly outperformed at no further costs.

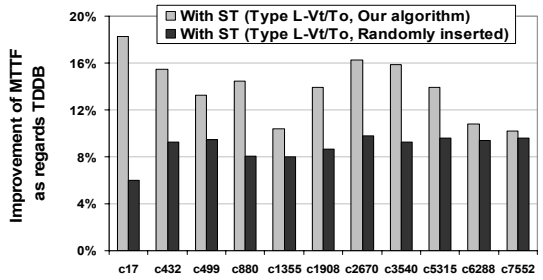

Figure 5. Improvement of MTTF for designs with shadow transistors inserted randomly and with our algorithm.

The improvement of the MTTF can be adjusted by adapting the critical threshold C_{crit}. As the distribution of signal probabilities across the transistors is not uniform, there is no linear correlation between C_{crit} and the MTTF. However, it holds true that larger C_{crit} adds more shadow transistors and thus raises reliability. The tendency of an increase is also valid for the power consumption and the system delay. Finally, the modified designs also inherently increase manufacturing yield as errors such as stuck-open transistors are also compensated.

4.3 Applying Dual-V_t

It is a known correlation that issues like gate leakage and gate oxide breakdown worsen with thinner gate oxides. For our goal of improved reliability, the correlation holds also true vice versa. That is, if we apply devices with thicker gate oxide, the expected lifetime increases as well. Fortunately, the chosen transistor models also include *H-V_t/To* transistors for this case [16]. Such transistors have thicker gate oxide and higher threshold voltage, and thus, they are slower and exhibit less leakage currents compared to the standard *L-V_t/To* types. Because the root causes of oxide breakdown are still under discussion, there exists no defined relation between oxide thickness and the MTTF. However, in [18] and [20] simple rules of thumb are deduced which state that MTTF increases by an order of magnitude when the gate oxide is made 0.22 nm thicker [18] (see equation 7), respectively 0.1 nm [20]. We refer to the more conservative value and calculate the difference in MTTF for the two transistor types as follows:

$$10^{\frac{\Delta t_{ox}}{0.22}} \quad \rightarrow \quad \frac{MTTF_{H-Vt/To}}{MTTF_{L-Vt/To}} = 10^{\frac{0.15}{0.22}} = 4.81 \qquad (7)$$

where *Δtox* is the difference in gate oxide of the two transistor types (in nm) and $MTTF_{H\text{-}Vt/To}$ as well as $MTTF_{L\text{-}Vt/To}$ are the corresponding MTTFs. Furthermore, the constant 0.15 is the difference of the gate oxides in our technology (1.85 nm ─ 1.7 nm). To evaluate the differences to the previous benchmark circuits for standard transistors (figure 5), we chose the same premises. This means that the same unprotected designs were chosen as starting point (including solely standard transistors). And the same algorithm was chosen for inserting the shadow transistors of type *H-V_t/To*, i.e. the same amount of redundancy was added for each design. However, in contrast to the previous case, the gate widths of the shadow transistors were not equally sized to their counterparts (W_R != W_S, see figure 3 b). They were rather sized to have the same input capacitance which promises to lead to similar figures for delay and power consumption. This could be proven by transistor-level simulations where only a small impact on delay could be observed. Though, it

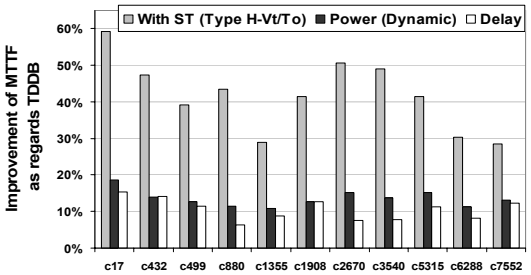

Figure 6. Enhanced results due to employment of H-Vt/To transistors as ST and the impact on power and delay.

needs to be considered that in the case of oxide breakdown in the original transistor, the setup with the high threshold shadow transistor performs slower due to smaller driving strength. This can also be seen in figure 4 where the two curves behave quantitatively similar but with a certain offset in delay. Lastly, the MTTF for the redundant pair has to be recalculated as the simple case of equal transistor types does not hold true here:

$$MTTF_{ST} = \int_0^\infty 1 - \left(1 - e^{\frac{t}{MTTF_{L-Vt/To}}}\right) \cdot \left(1 - e^{\frac{t}{MTTF_{H-Vt/To}}}\right) dt$$
$$= MTTF_{L-Vt/To} + MTTF_{H-Vt/To} - \frac{1}{\frac{1}{MTTF_{L-Vt/To}} + \frac{1}{MTTF_{H-Vt/To}}} \quad (8)$$

Figure 6 presents the results for the combined approach of shadow transistors with high threshold voltage transistors and the developed algorithm. Firstly, an impressive additional improvement in terms of MTTF can be observed with up to 59 % in the best case and 42 % on average (in contrast to 18.3 %, respective 13.9 % for standard shadow transistors). Though, dynamic power consumption increased by 13.5 % and the delay by 10.6 % on average. Nonetheless, it should be mentioned that the design parameters can be adapted in a wide range by changing the critical threshold C_{crit} accordingly. For instance, for a larger critical threshold of 0.5, the following mean values were achieved: MTTF 91.6 %, power 30.7 % and delay 20.0 %.

5. CONCLUSIONS

This contribution identified the need for improvements of lifetime reliability and the potential for approaches on transistor and gate-level design. One of the earlier findings stated that redundant transistors enhance manufacturing yield in the presence of stuck-open faults. We delved into this approach and demonstrated ─ based on comprehensive transistor level simulations ─ that redundant transistors can additionally compensate for a wide range of failures due to gate oxide breakdown. Motivated by this finding, we characterized numerous logic gates with redundant transistors (called shadow transistors) and implemented an algorithm that identifies those transistors of a netlist that are most vulnerable to Time-Dependent Dielectric Breakdown (TDDB). Furthermore, we argued for applying high threshold devices in place of the standard shadow transistors to further improve reliability. Finally, we implemented several benchmark circuits and proved our combined approach to be successful. For roughly 20 % of additional transistors, reliability with respect to TDDB was improved up to 59 % and 42 % on average compared to the unprotected netlist. However, delay and power consumption also exhibited moderate increase of 10.6 % and 13.5 %, respectively. Future efforts aim at integrating further mechanisms of lifetime failure into our models of reliability. Moreover, it is planned to include variable constraints for delay and power consumption to trade off design parameters more elaborately.

6. REFERENCES

[1] Chen, Z. and Koren, I., "Techniques for Yield Enhancement of VLSI Adders", In Proc. of ASAP, 1995.

[2] Chiluvuri, V. and Koren, I. "Layout-Synthesis Techniques for Yield Enhancement", In Trans. of Semiconductor Manufacturing, vol. 8, no. 2, 1995.

[3] Cornelius, C., Sill, F. and Timmermann, D., "Modeling the Power-Reliability Tradeoff in On-Chip Networks", In Proc. of Symposium Maritime Elektrotechnik, 2007.

[4] Crook, D., "Method of Determining Reliability Screens for Time Dependent Reliability Breakdown", In Proc. of Intern. Reliability Physics Symposium, 1979.

[5] Datta, R. et al., "Adaptive Design for Performance-Optimized Robustness", In Proc. of DFT, 2006.

[6] Degraeve, R., Kaczer, B., De Keersgieter, A. and Groeseneken, G., "Relation between Breakdown Mode and Breakdown Location in Short Channel nMOSFETs and its Impact on Reliability Specifications", In Proc. of IRPS,2001.

[7] Koren, I. and Krisha, C., "Fault-tolerant Systems", Morgan Kaufmann, 2007.

[8] Kephart, J. and Chess, D., "The Vision of Autonomic Computing", In Computer, vol. 36, no.1, 2003.

[9] Mitra, S. et al., "Robust System Design with Built-In Soft-Error Resilience", In Computer, vol. 38, 2005.

[10] Omana, M., Rossi, D. and Metra, C., "Latch Susceptibility to Transient Faults and New Hardening Approach", In Trans. on Computers, vol. 56, no. 9, 2007.

[11] Reichenbach, F., Bobek, A., Hagen, P., Timmermann, D., "Increasing Lifetime of Wireless Sensor Networks with Energy-Aware Role-Changing", In Proc. of SelfMan, 2006.

[12] Renovell, M., Gallière, J., Azaïs, F. and Bertrand, Y., "Modeling the Random Parameters Effects in a Non-Split Model of Gate Oxide Short", In Journal Electronic Testing, vol. 19, no. 4, 2003.

[13] Schmeck, H., "Organic Computing – A New Vision for Distributed Embedded Systems", In Proc. of ISORC, 2005.

[14] Segura, J., Benito, C., Rubio, A. and Hawkins, C., "A Detailed Analysis of GOS Defects in MOS Transistors: Testing Implications at Circuit Level", In Proc. of Test Conference on Driving Down the Cost of Test, 1995.

[15] Semiconductor Industry Association (SIA), "International Technology Roadmap for Semiconductors", Release 2007, Published on-line: http://www.itrs.net/.

[16] Sill, F., You, J. and Timmermann, D., "Design of Mixed Gates for Leakage Reduction", In Proc. of GLSVLSI, 2007.

[17] Sirisantana, M., Paul, B. and Roy, K., "Enhancing Yield at the End of the Technology Roadmap", In Trans. of Design&Test of Computers, vol. 21, no. 6, 2004.

[18] Srinivasan, J. et al., "RAMP: A Model for Reliability Aware Microprocessor Design", In IBM Research Report, RC23048, 2003.

[19] Srinivasan, J., Adve, S., Bose, P. and Rivers, J., "The Impact of Technology Scaling on Lifetime Reliability", In Proc. of DSN, 2004.

[20] Stathis, J., "Reliability Limits for the Gate Insulator in CMOS Technology", In IBM Journal of Research and Development, 2002.

[21] Trivedi, K., "Probability and Statistics with Reliability, Queuing, and Computer Science Applications", Prentice Hall, 1982.

[22] Tschanz, J. et al., "Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage", In JSSC, vol. 37, 2002.

[23] Vogel, E. et al., "Reliability of Ultra-Thin Silicon Dioxide Under Combined Substrate Hot Electron and Constant Voltage Tunneling Stress", In Trans. of Electron Devices, vol. 47, no. 6, 2000.

[24] Wu, E. and Sune, J., "Power-Law Voltage Acceleration: A Key Element for Ultra-thin Gate Oxide Reliability", In Microelectronics Reliability, 2005.