29.5 Digital Detection of Oxide Breakdown and Life-Time Extension in Submicron CMOS Technology

Mustafa Acar, Anne-Johan Annema, Bram Nauta

University of Twente, Enschede, Netherlands

In modern CMOS technologies reliability issues limit the maximum operating voltage of transistors. This prevents the integration of efficient power amplifiers (e.g., audio or RF) since stacked devices are needed to prevent breakdown, which reduces efficiency. Transistor reliability is strongly related to operating voltages; higher voltages result in faster degradation and hence in lower reliability and shorter life time. Degradation can be monitored by oxide degradation, threshold voltage-shifts and mobility reduction.

An approach is introduced to extend the lifetime of high-voltage analog circuits in CMOS technologies based on redundancy, like that known for DRAMS. A large power transistor is segmented into N smaller ones in parallel. If a sub-transistor is broken, it is removed automatically from the compound transistor. The principle is demonstrated in an RF CMOS Power Amplifier (PA) in standard 1.2V 90nm CMOS.

A single oxide breakdown event (OBD) shows up as a sudden increase in oxide-leakage, which can be modeled as the sudden formation of a resistance of a few $k\Omega$ from gate to drain, to source or to bulk [1]. Due to the large size of power transistors, the relative effect of a single OBD on the performance is small; consequently defining reliable operation time as "the time to the first oxide breakdown of the gate dielectric in a transistor" [1] is overly strict [2]. However, in order to have a reasonable overall lifetime, normally an extra margin against degradation is built into the design, i.e., non-degraded power transistors have to be able to deliver more power than needed in the case of zero OBD, so they still function properly after a few OBDs.

A number of OBDs is acceptable for power circuits and stochastic properties can even be used to increase lifetime. In our system each power transistor (width W) is segmented in N parallel transistors (width W/N), each with OBD monitoring circuitry and pre-driver including enable/disable functionality, as shown in Fig. 29.5.1. During operation, the number of OBDs in each power transistor (segment) is monitored by measuring the oxide-resistance; the number of OBDs is approximately proportional to the gate conductance. Upon detection of too many OBDs in a power transistor in a segment, i.e., oxide conductance that is too high, that segment is shut down.

During operation some OBDs occur, mainly in the most heavily stressed parts of a circuit, here, in the (segmented) power transistors. In case of a weak spot in a transistor, the breakdowns occur mainly in that transistor, and only that segment is shut down. If no weak spots are present, the distribution of the OBDs will be spatially uniform across all segments, while the total number of OBDs increases with operation time. Now statistics helps to increase lifetime.

If M breakdowns are spatially uniformly distributed across N segments, the distribution of OBDs in one segment is binomial (Gaussian-like): some segments have more breakdowns than others. For example, 107 uniform OBDs across 16 sections, the average OBD count in a segment is 6.7. The segment with the most breakdowns, however, has on average 11.6 OBDs. Elimination of that segment decreases the total number of OBDs in active segments much more than proportionally. For the above example, Fig. 29.5.2 shows the probability distribution both over all of the 16 segments, and in the segment that ends up with the most OBDs. Mathematically, the difference between the medians of OBDs in any segment and in the "fullest" segment is $\Delta_{median} \approx (\sqrt{2\pi} \sigma_w)^{-1} (1 - 2/N) + 0.4) \sigma_w$ where σ_w is the standard deviation of the number of OBDs averaged over N segments: more segments (N) and more variance is beneficial.

To demonstrate the effect of stochastic lifetime extension an RF PA was designed in 90nm CMOS, with thin gate oxide, $V_{dd}=1.2V$, and

each power transistor segmented in 16 sections ($N=16$). Figure 29.5.3 shows Spectre simulation results of the output power (P_{OUT}) and power added efficiency (PAE) of the Class-E PA as a function of the number of OBDs. The x-axis can be interpreted as a time axis if the OBDs are evenly distributed over time. End-of-life (EOL) is defined as either $PAE < 30\%$ or $P_{OUT} < 80mW$. The curves marked with "16 segments" correspond to always using all 16 segments, with EOL around 110 OBDs. The curves marked "8+8 segments" correspond to using half the segments and replacing all these at EOL detection with the remaining 8 segments. This redundancy type of operation (with the same total area) has EOL after about 190 OBDs: a 70% increase in lifetime. For the curves marked with "16-15-14-13 segments" the PA starts with all 16 segments and sequentially shuts down the worst segment; at EOL 3 segments are disabled. EOL is at about 400 OBDs: a 260% increase in lifetime.

Figure 29.5.4 shows the circuit schematic of one PA segment; shown in grey are the pre-driver and the cascoded output stage (M1 and M2). The four operation modes are: RF operation, disabled, and OBD-measurement for each of the transistors M1 and M2. Control signals C_n are listed in Fig. 29.5.4 for the various modes. OBD-measurement is done by determining the gate oxide conductance (about proportional to the number of OBDs). The breakdown measurements are done in three steps: gate-conductance in M1 is sensed in one step while M2 is measured in two steps.

OBDs for M1 and M2 were quasi-continuously measured in one segment (by continuously cycling through the 4 modes) to verify the monitoring function. Figure 29.5.5 shows measured sense-voltage levels in the 3 breakdown measurement modes as a function of stress time under accelerated conditions. The times at which an OBD occurs are marked with a flash on the x-axis. The schematics in Fig. 29.5.5 show the effective circuit for the 4 operation modes; the measured OBD-paths during measure-modes are indicated with a resistor+flash. For the first 6 breakdown events in the segment a slightly degraded output voltage shape resulted, while after the 7th breakdown (at 20000s in Fig. 29.5.5) the total PA fails to operate correctly. This is detected and the segment is switched off.

The RF performance of 1 out of the 16 segments is shown in Fig. 29.5.6: the measured P_{OUT} , drain efficiency DE and PAE at 900MHz are shown as a function of input power. Measurements were done on-wafer without harmonic tuning (non class-E operation), using an RF choke to bias the open drain from a 1.2V supply voltage. The measured PAE of about 30% at 900MHz is comparable to state-of-the-art implementations in CMOS PAs [1,3] and shows that RF performance is not significantly degraded by the OBD detection circuitry. The lack of harmonic tuning in our measurements prevents waveshaping characteristics of high efficiency PAs (e.g., Class-E); resulting in the difference between simulations (Class-E) and measurement results (non-Class-E). The die photo is shown in Fig. 29.5.7; including pads and control logic it occupies $3.0mm^2$.

The lifetime of power systems can be increased significantly by segmenting a power system into many lower-power segments in parallel – all with their own breakdown detection circuitry – and switching off the segments with the most oxide breakdowns. The implementation presented only has a minor effect on DE and PAE .

Acknowledgements:

The support of G. Wienk, H. de Vries, and D. Leenaerts is gratefully acknowledged.

References:

- [1] B. Kaczer, et al., "Impact of MOSFET Gate Oxide Breakdown on Digital Circuit Operation and Reliability", *IEEE Trans. Electron Devices*, pp. 500-506, May 2002.
- [2] L. Larcher, et al., "Oxide Breakdown after RF Stress: Experimental Analysis and Effects On Power Amplifier Operation", *IEEE IRPS*, pp. 283-288, Mar. 2006.
- [3] S. M. Rezaul Hasan, "A High Efficiency 3GHz 24-dBm CMOS Linear Power Amplifier for RF Application", *IWSOC'05*, pp. 503-507, July 2005.

Figure 29.5.1: Left: a conventional cascode PA; right: the segmented PA to increase lifetime.

Figure 29.5.2: Simulated probability mass function for 107 breakdown events, uniformly distributed over 16 segments: the median and mean of the fullest segment is much higher than those over all segments.

Figure 29.5.3: Simulated P_{out} and PAE as a function of the number of breakdowns (operation time) for 3 situations: continuously using 16/16 segments resulting in EOL=110 events; constant-area redundancy using 8/16 + 8/16 segments yielding EOL=190 events (+70% lifetime); switching off the worst segment using 16-15-14-13 segments for which EOL=400 events (+260% lifetime).

Figure 29.5.4: Circuit schematic of one of the 16 PA segments: a conventional RF PA with oxide breakdown monitoring circuitry; digital control logic and decoders not shown.

Figure 29.5.5: Measured breakdown-indicating sense-voltage as a function of stress time, under accelerated conditions. For each of the 4 operation modes the corresponding effective circuit configuration of the segment is shown.

Figure 29.5.6: Measured performance of 1 out of the 16 PA segments, at 900 MHz with-out harmonic tuning.

Continued on Page 633

Figure 29.5.7: Micrograph of the demonstration vehicle in 90nm: the total PA is subdivided into 16 segments with their own OBD monitor circuitry. Digital control used for addressing each segment's state; total chip area is 3.0mm².