Conference paper

ESD protection circuit design for ultrasensitive IO applications in advanced sub-90nm CMOS technologies

International Symposium on Circuits and Systems 2005

This paper presents a protection strategy for ultra-sensitive IO's containing thin gate oxides, while combining two complementary ESD design approaches: 1. Low-voltage diode-chain triggered SCR clamps that allow for efficient voltage clamping. 2. Active-Source-Pump circuits applied for effective expansion of narrow ESD design windows for ultra-thin GOX protection. The focus of the paper is on the ASP schemes while some RF aspects will be covered as well.

ESD protection circuit design for ultra-sensitive IO applications in advanced sub-90nm CMOS technologies

Markus Mergens[3], Geert Wybo[1], Benjamin Van Camp[1], Bart Keppens[1], Frederic De Ranter[1], Koen Verhaege[1]
[1)] Sarnoff Europe, Aalter, Belgium
mmergens@sarnoffeurope.com

Phil Jozwiak[2], John Armer[2], Christian Russ[3]
[2)] Sarnoff Corporation, Princeton, NJ, USA
[3)] Formerly with Sarnoff now with Infineon Technologies, Munich, Germany

Abstract - This paper presents a protection strategy for ultra-sensitive IO's containing thin gate oxides, while combining two complementary ESD design approaches: 1. Low-voltage diode-chain triggered SCR clamps that allow for efficient voltage clamping. 2. Active-Source-Pump circuits applied for effective expansion of narrow ESD design windows for ultra-thin GOX protection. The focus of the paper is on the ASP schemes while some RF aspects will be covered as well.

I. INTRODUCTION

With further downscaling of feature size in advanced CMOS technologies, protection of thin and ultra-thin gate oxides (GOX, tox < 2nm) becomes increasingly challenging [1]-[4]. This critical trend is corroborated in Figure 1 showing Transmission Line Pulse (TLP: 100ns square pulse) results for CMOS technologies down to 65nm [3].

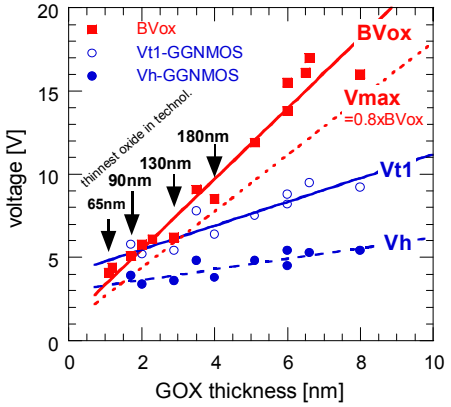

Figure 1 Transient breakdown of NMOS GOX (area <40u^2) and corresponding GGNMOS trigger voltage Vt1 and holding voltage Vh as a function of physical GOX thickness in various CMOS technologies (0.18u - 0.065u). Technology advancement results in a continuous narrowing of the ESD design window for input protection.

The plot shows a rapid decay of the transient GOX breakdown voltage BVox as a function of the physical GOX thickness. As a result of the strong BVox decrease, the ESD design window for input GOX protection, cf. Figure 2, is dramatically narrowing down in advanced technologies. The upper limit V_{max} (considered for ESD design on IC level) is derived from BVox by subtracting a safety margin of typically about 20%, to account for process fluctuations. The reliability of the V_{max} level is experimentally verified by TLP endurance tests ruling out potential MOS characteristic degradation. Typical design window spans V_{max}-VDD range between 2.5V to 3.5V for ultra-thin gate inputs in sub-90nm CMOS technologies.

Grounded-Gate NMOS (GGNMOS) transistors are most widely applied for ESD protection design. Figure 1 also illustrates the evolution of the critical GGNMOS snapback parameters (i.e. parasitic NPN trigger voltage Vt1 and holding voltage Vh indicated in Figure 2) with technologies advancement. Apparently, both these values come close to or even exceed BVox and V_{max}. As a consequence of the insufficient GGNMOS clamping capability, pure (parasitic) bipolar protection of ultra-thin GOXs is not feasible anymore in these very advanced technologies.

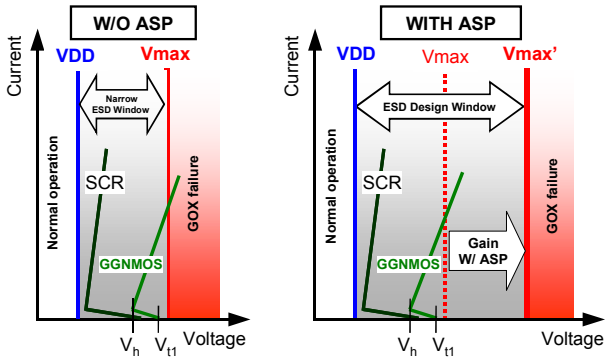

Figure 2 Schematic ESD Design Window including typical SCR and GGNMOS TLP-IV curve. Design window expansion by ASP is indicated.

The following complementary protection concepts to solve the narrow ESD window issue will be discussed in the paper:

1. A diode chain triggered SCR (DTSCR) [2]. These elements are capable to protect ultra-thin gate oxides due to a sufficient transient voltage clamping.

2. Active Source Pump (ASP) schemes for effective ESD design window expansion (cf. Figure 2: V_{max}'>V_{max}) will be introduced [3]. These schemes are in particular useful for ultra-thin GOX protection and RF pins, where local clamps would introduce an extremely high parasitic capacitance

compromising high-speed performance. Moreover, as will be demonstrated experimentally, the ASP scheme adds significant design flexibility on IC-level for narrow ESD design windows.

II. DIODE-TRIGGERED SCR (DTSCR)

The DTSCR uses a diode chain trigger to latch the SCR during ESD stress conditions at sufficient trigger current injection into the gates G1 or G2 of the SCR.

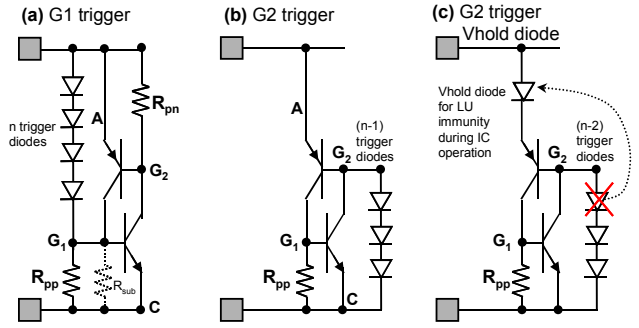

Figure 3 G1-triggered DTSCR (a): forward bias of SCR G1-Cathode junction. In CMOS-SCR an intrinsic connection to the substrate (Rsub) is present. G2-triggered DTSCR (b/c): forward bias of G2-Anode junction. Vh diode for LU immunity replaces trigger diode (c).

Triggering can either be accomplished by forward biasing the inherent SCR G1 - Cathode junction, cf. Figure 3 (a), or alternatively the G2 - Anode diode, cf. Figure 3 (b), or both simultaneously. The number of trigger diodes (n) must be chosen sufficiently high such that the chain does neither leak nor trigger the SCR during normal circuit operation. Conversely, the SCR ESD trigger voltage is increasing with n. A reasonable design trade-off between low leakage during normal operation ("maximize n") and low ESD trigger voltage ("minimize n") can be achieved for low-voltage applications with supply or signal voltage smaller than 1.8V. Here, the trigger voltage can be tuned sufficiently below the transient BVox and V_{max}: $Vt1 \sim (n+1)\ 0.8V < V_{max}$ [2]. With n=2 trigger diodes e.g. in a G2 scheme we would obtain $Vt1 \sim 2.4V$. Applying the DTSCR as a power-clamp, the SCR holding voltage must be increased above VDD for latch-up immunity. A trigger diode can be moved from the trigger chain to the SCR anode, cf. Figure 3 (c), adjusting also its width. This adds approximately 1V to the total SCR holding voltage.

III. ACTIVE-SOURCE-PUMP ASP SCHEMES

Figure 4 (left) illustrates the general ASP concept here for a simple input inverter including generic placeholders for primary IO ESD protection (diodes or local clamps). The ASP scheme consists of source-pump devices (i.e. relatively small protection elements) inserted between IO line and internal MOS source node in series with a source impedance Z_S. During ESD stress conditions, one ASP circuit is activated at a safe voltage level below V_{max} for the respective critical stress case (NMOS: positive ESD to IO vs. VSS; PMOS: positive ESD on VDD vs. IO). As a result, the ASP circuit starts to inject minor amounts of ESD current into the source impedance Z_S as indicated for IO vs. VSS stress in Figure 4. The main ESD current is conducted thru the primary ESD protection. Due to the resulting voltage drop across Z_S, the NMOS source potential V_S rises and reduces the total ESD voltage across the endangered gate-source oxide $V_{GS}=V_{IN}-V_S$, which often represents the most sensitive IC part. Thus, a higher gate ESD voltage $V_G=V_{max}$' can be tolerated before the maximum limit V_{max} directly across the GOX is reached. In other words, the ASP element and the source impedance Z_S act as a transient voltage divider similar to secondary clamp schemes.

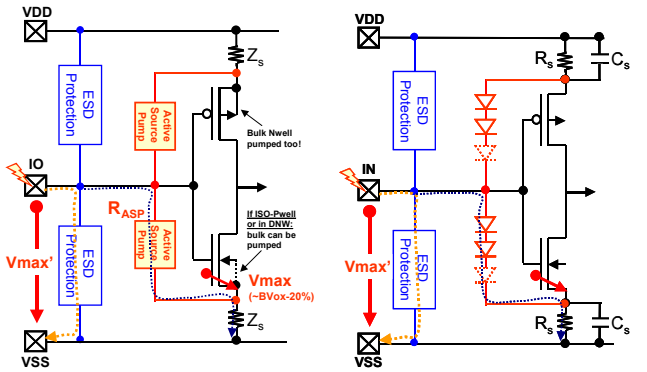

Figure 4 Generic scheme for Active-Source-Pump technique with ASP circuit and source impedance for input inverter stage (left). Specific ASP implementation with diode-chain ASP and source resistor (right). The parallel capacitance C_S is optional for RF noise suppression. The diode orientation is defined to protect against the worst-case stress polarities.

A specific ASP design example is presented in Figure 4 (right) employing a diode chain ASP circuit in conjunction with a poly source resistor $R_S \sim 10\Omega$. An optional parallel source capacitance C_S can be used for resistor noise suppression for instance in a RF-LNA. ESD stress conditions define a maximum C_S to avoid a too low overall source impedance. For C_S=2.5pF we can estimate an order of magnitude for the ESD impedance $Zs=1/\omega C_S$: $Z_S=8K\Omega$ for HBM (duration T=100ns) and $Z_S=80\Omega$ for CDM (T=1ns). Both values would be sufficiently high for source pumping. In some RF applications, a source matching inductor $Z_S = L_S$ can be present. The inductance $Z_S=\omega L_S$ as well as the parasitic inductor resistance R_L can be exploited for ASP integration. Calculating a rough order of magnitude for Z_S with typically L=5nH results in $Z_S=0.30\Omega$ for HBM and $Z_S=30\Omega$ for CDM. The inductance would become effective for fast CDM only whereas for HBM stress one would need to potentially exploit the parasitic resistance R_L.

The specific ASP diode design is applicable in low-voltage domains (e.g. Vin=VDD≤1.2V) where 3 diodes can sufficiently limit the normal operation leakage current. Moreover, for minor ESD currents, the ASP diodes can be relatively small in size, thus introducing a limited parasitic capacitance (~10fF) only to RF pins. On the other hand, the ASP resistance must be sufficiently small to allow for an efficient voltage divider and thus a design window increase. The maximum permissible Z_S would depend on the specific IC application as will be discussed below.

As shown in Figure 5, alternatively to the Z_S impedance in the conventional ASP scheme, an active element as for instance a second MOS transistor can be used. These cascodes are frequently present in LVDS inputs.

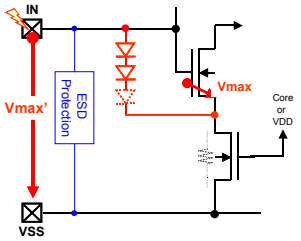

Figure 5 Typical cascoded NMOS configuration including ASP diode chain for source pumping of the jeopardized upper NMOS gate.

Input MOS bulk connection

If possible it is advisable that the ASP scheme is also applied to simultaneously pump the MOS bulk and thus increase the gate-bulk breakdown accordingly. In the PMOS this can easily be achieved by connecting the "isolated" Nwell to Z_S instead of directly to VDD. For the NMOS transistor, the Pwell would need to be isolated (as in SOI technologies) by deep-Nwell application if available without any additional costs. *But is it really mandatory to also pump the bulk for GOX protection?*

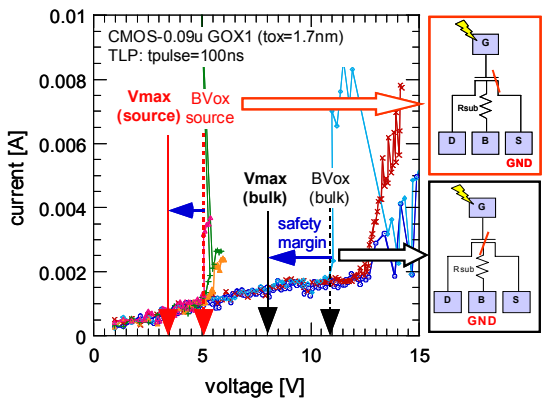

Figure 6 Transient NMOS GOX (90nm-CMOS: tox=1.7nm, W/L=20u/0.5u) breakdown for positive gate-source TLP stress and gate-bulk TLP stress. BVox to bulk is significantly higher than to the source.

The following phenomenon shown in Figure 6 for a plain dual-well CMOS technology demonstrates why bulk pumping can be abandoned. The transient gate-to-bulk oxide breakdown occurs at drastically higher voltage levels as compared to the gate-to-source/drain breakdown. As illustrated by the TLP data in Figure 6, the source-side stress reveals the characteristically low failure level of approximately BVox(G-S)~5V (V_{max}~4V), whereas the gate-bulk breakdown appears at more than a factor of two higher values BVox(G-B)>12V for the lowest breakdown of multiple GOX samples. The leakage current evolution was also monitored after each stress pulse but is not shown in the plot. The leakage remained constant up to the voltage level, where BVox indicates a considerable stress current increase thru the ruptured GOX.

The physical mechanism responsible for this breakdown phenomenon is the thin bulk depletion layer below the gate occurring for inversion-type of stress i.e. a positive pulse to the NMOS gate. The depleted bulk region considerably increases the effective bulk resistance. The resulting voltage drop across the bulk resistance during ESD boosts the gate-bulk breakdown significantly. Note that the same effect does not occur for negative NMOS gate stress due to the missing depletion region under accumulation stress. As a consequence, for negative gate stress bulk and source breakdown will show similarly low BVox as experimentally verified. But generally this ESD stress case is not critical.

To harness the discussed breakdown phenomenon for ESD protection design it is highly recommended a) to thoroughly analyze the behavior as shown in Figure 6, b) to introduce a sufficiently large safety margin for V_{max}(GB) regarding BVox(GB) to account for potential BVox fluctuation, and (c) to verify V_{max} experimentally as discussed above.

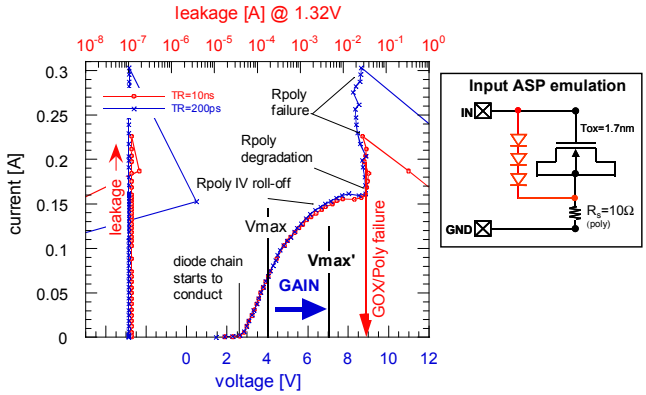

Figure 7 TLP data of the input emulation structure in 90nm-CMOS (see inset) applying a diode chain ASP and poly source resistor of R_S=10Ω. Vmax expansion: >75% including conservative safety margin.

IV. Experimental Analysis of ASP Input

The right inset in Figure 7 represents a dual-pad structure for ultra-thin GOX NMOS input emulation including the corresponding ASP circuit in Figure 4 (right). Here stacked diodes are inserted between IN pad and the shorted drain/source/bulk which in turn are connected by a poly resistor of R_S=10Ω to a ground pad. The TLP analysis results for positive pulse to IN vs. GND indicate that at approximately 2.7V the diode chain in series with R_S starts to conduct current resulting in a pumping of the source/drain node. At elevated currents, the linear IV curve regime bends into the typical IV roll-off of a poly resistor under high current conditions. This thermal resistance increase of R_S generally does not degrade the structure but leads to a more efficient voltage divider allowing for a larger V_{max}'. As indicated by the constant leakage current evolution up to almost 9V TLP, the upper ESD design margin limit can be increased significantly. Above this point, the GOX may be damaged (slight leakage increase). Subsequently, the poly resistor is blown open as evidenced by a rapid leakage drop and TLP voltage increase. Defining a V_{max}'=7V with

sufficient safety margin regarding the actual failure point, corresponds to a significant 75% increase of the GOX V_{max}=4V, cf. Figure 6. The measurement results were verified for slow (TR=10ns) and fast TLP rise times (TR=200ps), respectively, hence ensuring that ASP also reacts for fast CDM transients.

V. ASP IC IO Application Examples

Protecting ultra-thin GOXs within RF inputs is challenging since low capacitance diode concepts, e.g. dual-diode in Figure 8, must replace high capacitive local clamps. Moreover, pi-type protection schemes require input gate resistors to separate secondary from primary protection elements. This however significantly compromises the RF performance due to detrimental gate delay and added noise.

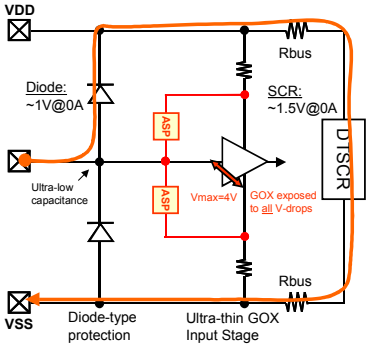

Figure 8 RF input applying diode-type protection in conjunction with ASP scheme and DTSCR power clamp

For worst-case ESD stress conditions RF-IN vs. VSS indicated in Figure 8, the VDD diode, the DTSCR power clamp, and the bus resistance Rbus are involved in the ESD current path. The ultra-sensitive GOX of the NMOS is exposed to the resulting voltage drops. Considering the diode cut-in voltage ~1V and the SCR holding voltage ~1.5V, these elements start to conduct considerable ESD currents at approximately ~2.5V. As a result, there is only (4V-2.5V) = 1.5V margin left to shunt for instance 1.5A (~2kV-HBM). This would mean that less than 1Ω total resistance R_{SCR}+R_{diode}+R_{BUS} was permissible to successfully protect the input. The required increase of protection elements would lead to excessive area consumption as well as to a higher diode capacitance at the RF pin. On IC-level a higher power clamp repetition rate would be necessary to limit the bus resistance Rbus.

Figure 9 shows the TLP data for the input configuration and stress case visualized in Figure 8 utilizing the diode chain ASP scheme in Figure 4 (right) and a DTSCR power clamp. The triggering and holding voltage correspond to the sum of DTSCR and VDD diode voltage drop. Evidently, the ASP scheme allows for a significantly higher voltage across the NMOS gate oxide due to V_{max}'=7V. Therefore, the effective ESD performance is improved from approximately It2=1.5A (intercept of IV with V_{max} w/o ASP) to It2=2.5A corresponding to the failure level of the DTSCR itself.

Normal operation interference of the ASP circuit within the input stage (neglecting minor ASP capacitance) is caused by source de-biasing due to the $I_{DS}R_S$ voltage drop. This results roughly in a 10% I_{DS} reduction [3] and in a slight shift of the operation point. The lowered drive performance and shift in DC switch point can be compensated by a slight increase of the MOS widths. The power gain loss caused by R_S must be traded off with ESD for the specific application.

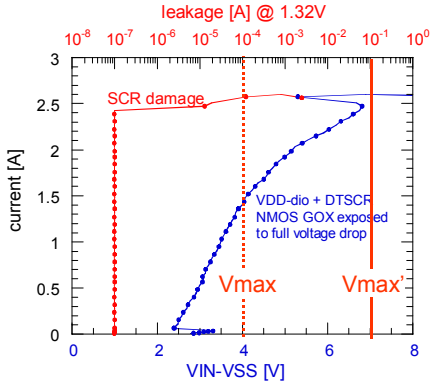

Figure 9 TLP data of dual-diode protection scheme and DTSCR power clamp with diode ASP scheme protecting thin GOX.

In general, the normal operation interference of the ASP circuit should be considered during the actual IO design phase. Embedding the ASP scheme into a comprehensive "RF-ESD co-design" concept allows simultaneously trading-off and optimizing ESD and RF performance. Moreover, the additional introduction or exploitation of circuit elements (capacitors, inductors, cascaded NMOS etc) for ASP can be explored more effectively.

VI. Conclusions

This paper presents a successful protection strategy of highly sensitive IO's containing ultra-thin gate oxides. The approach combines low-voltage triggered DTSCR clamps with an active-source-pump (ASP) circuit technique for effective ESD design window expansion. The significant advancement is achieved by minor ESD current injection into a source impedance effectively reducing the most critical gate-source ESD voltage. The excellent SCR clamping capabilities in conjunction with the margin boost (here approximately 75%) allow for largely enhanced ESD performance and design flexibility. This also supports low-capacitance RF-ESD input solutions. Normal operation RF interference must be traded off with ESD performance.

References

[1] A. Salman et al. *ESD-Induced Oxide Breakdown on Self-Protecting GG-nMOSFET in 0.1-um CMOS Technology.* TDMR, Sept. 2003, pp. 79.

[2] M. Mergens et al. *Diode-Triggered SCR (DTSCR) for RF-ESD Protection of BiCMOS SiGe HBTs and CMOS Ultra-Thin Gate Oxides.* IEDM 2003, pp. 515.

[3] M. Mergens et al. *Active-Source-Pump (ASP) Technique for ESD Design Window Expansion and Ultra-Thin Gate Oxide Protection in Sub-90nm Technologies*. CICC 2004, pp. 251.

[4] J. Wu et al. *Breakdown and Latent Damage of Ultra-Thin Gate Oxides under ESD Stress Conditions*, EOS/ESD 2000, pp. 287.

About Sofics

Sofics (www.sofics.com) is the world leader in on-chip ESD protection. Its patented technology is proven in more than a thousand IC designs across all major foundries and process nodes. IC companies of all sizes rely on Sofics for off-the-shelf or custom-crafted solutions to protect overvoltage I/Os, other non-standard I/Os, and high-voltage ICs, including those that require system-level protection on the chip. Sofics technology produces smaller I/Os than any generic ESD configuration. It also permits twice the IC performance in high-frequency and high-speed applications. Sofics ESD solutions and service begin where the foundry design manual ends.

Our service and support

Our business models include

- Single-use, multi-use or royalty bearing license for ESD clamps
- Services to customize ESD protection
  - Enable unique requirements for Latch-up, ESD, EOS
  - Layout, metallization and aspect ratio customization
  - Area, capacitance, leakage optimization
- Transfer of individual clamps to another target technology
- Develop custom ESD clamps for foundry or proprietary process
- Debugging and correcting an existing IC or IO
- ESD testing and analysis

Notes

Version

May 2011

Sofics BVBA
Groendreef 31
B-9880 Aalter, Belgium
(tel) +32-9-21-68-333
(fax) +32-9-37-46-846
bd@sofics.com
RPR 0472.687.037