SOFICS
SOLUTIONS FOR ICS

Conference paper

Novel Design of Driver and ESD Transistors with Significantly Reduced Silicon Area

EOS/ESD symposium 2001

This paper presents three novel design techniques, which combined fulfill all major requirements posed on large driver and Electro Static Discharge (ESD) protection transistors: minimum area consumption, good ESD robustness and optimized normal operation. Transistors protecting 5V/um2 Human Body Model (HBM) were demonstrated. Significant silicon area reduction was demonstrated in deep-sub micron CMOS, ranging from 0.35um down to 0.13um CMOS. This novel design solution follows standard design flows and does not require any process modifications.

Novel Design of Driver and ESD Transistors with Significantly Reduced Silicon Area

Koen G. Verhaege*, Markus Mergens, Christian Russ, John Armer, Phillip Jozwiak

* Sarnoff Europe, Brugse Baan 188A, B-8470 Gistel, Belgium, kverhaege@sarnoffeurope.com

Sarnoff Corporation, 201 Washington Road, Princeton NJ-08543, USA, kverhaege@sarnoff.com

Abstract – This paper presents three novel design techniques, which combined fulfill all major requirements posed on large driver and Electro Static Discharge (ESD) protection transistors: minimum area consumption, good ESD robustness and optimized normal operation. Transistors protecting 5V/um^2 Human Body Model (HBM) were demonstrated. Significant silicon area reduction was demonstrated in deep-sub micron CMOS, ranging from 0.35um down to 0.13um CMOS. This novel design solution follows standard design flows and does not require any process modifications.

1. Introduction

It is well established that on-chip protection is necessary in order to prevent significant production loss. As such, the industry has developed methods and techniques to provide ICs with the required ESD robustness. The foremost important measure is the introduction of ballast resistance in those devices that will absorb the multi-Ampere ESD currents. Ballast resistance avoids current crowding and thus distributes the stress current more uniformly in the semiconductor devices. [1-11] An example of an NMOS transistor with insufficient ballasting leading to early ESD failure is shown in Figure 1.

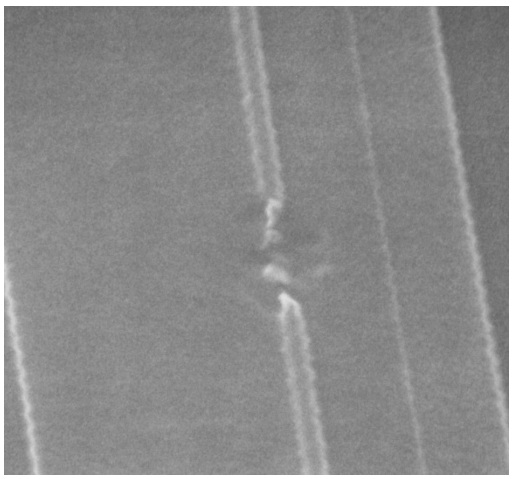

Figure 1 – *Fully silicided NMOS transistor, with (too) low active ballasting resistance showing a non-uniform failure (center of picture) due to early current collapsing under ESD stress conditions.*

For decades, the proper ballast resistance was formed in the active silicon [1-11]. It was inserted as 'drain-contact-to-gate-spacing' (DCGS – see Figure 2) the NMOS transistors, which are intrinsically the most sensitive ESD devices. Typical DCGS values are 4-5um in 0.5um CMOS, 2-3um in 0.25um CMOS.

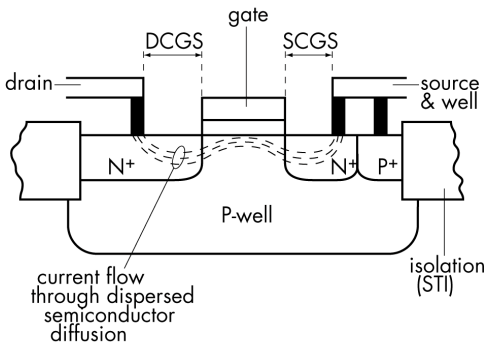

Figure 2 *– Active ballasting in a NMOS transistor introduced by Drain-Contact-to-Gate-Spacing (DCGS), to enhance uniform current flow under high current (several Amperes) ESD conditions*

In advanced technologies, the silicidation of the semiconductor diffusions reduces the ballast resistance provided by DCGS with at least a factor of 10, causing major setback in IC ESD protection performance [2,8,10-16]. Therefore, the active ballasting was re-introduced by the addition of a "silicide block" process step. This is accomplished by patterning the wafer with a blocking layer (typically nitride) to selectively cover areas where no silicidation is wanted. The most important drawback of this solution is the need for additional

processing steps and the associated costs: about $20 or more per each 8-inch wafer.

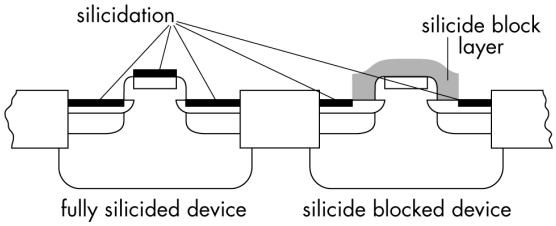

Figure 3 *– Fully silicided device (left) with shorted active area ballast and Silicide blocked device (right) with re-introduced active ballast, but at the expense of additional processing and masks.*

The fundamental issue with ballast resistance, as required in IO drivers and ESD transistors, is that it consumes significant silicon real estate. As IC cores shrink with the ever-shrinking technology geometries, the IO transistor areas stayed at the 0.5um CMOS node. Every dimension scales except the large drive and/or protection transistors. Since silicon real estate and thus die and IO size is the single most cost sensitive element in IC production, there is a constant need in the industry to reduce it [10, 17-19].

2. Novel Device Design and Layout Techniques

This paper presents a universal, uncomplicated design technique that delivers superior ESD performance while reducing the area significantly. It is compatible with the base-line logic process.

2.1. Back-End Ballast and Segmentation

Advanced deep sub-micron technologies have contact, via and interconnect resistances in the order of several Ohms per instance (e.g., a single contact). This allows building ballast resistance networks with these basic process elements. This technique is called Back-End Ballasting as it uses the back-end process to create the ballast resistance [17]. Of particular interest is the silicided poly material that is sufficiently resistant and ESD robust (Figure 4). Generally, silicided poly is at least two orders of magnitude more resistive than the typical metal layers, and it can absorb fairly high levels of ESD current.

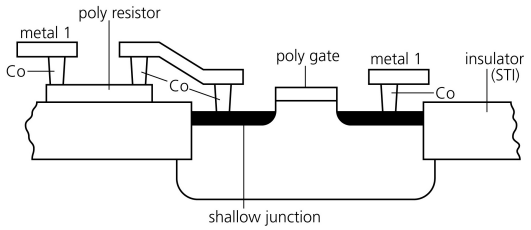

Figure 4 *– Poly Back-End-Ballast replacing active area ballasting to avoid current crowding under ESD stress conditions.*

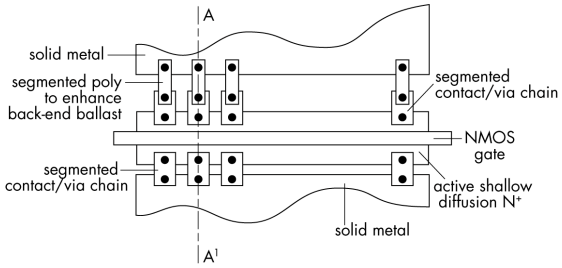

Figure 5 *– Segmentation of the (back-end) ballast resistive network allows for additional micro-ballasting.*

Segmenting the current path between e.g. the pad metal and the active silicided device will further enhance the ballast mechanism (Figure 5). The segmentation can be maintained through several layers of metals, vias and a single contact to the active device. Using a silicided poly layer will create sufficient and effective ballast. Constructing vertical meander metal-interconnect chains is an additional option.

Every segment can easily provide ballast up to several tens of Ohms, while the combination of all parallel ballast segments can still provide a very low total series resistance in the device and insure a high ESD current capability. Also, segmentation and back-end ballasting do not introduce mechanical stress beyond the normal process windows.

The proposed design technique includes a very effective negative feedback mechanism that will de-focus any local increase in current. When at any point in the active silicided device there is an onset of current focus and current crowding, the voltage across the resistive “feeding” segment must increase above the voltage across the neighboring segments. This is due to the local increase in current. Since all segments are short circuited by metal at one end and resistively connected at the silicided device end, the voltage must increase

across the other segments and/or the current must decrease in the "feeding" segment. Therefore, crowding current is de-focused instantly using the universal principle of Ohms Law.

The performance of Back-End-Ballast with Segmentation (BEB) is silicon proven as illustrated in Figure 6, which shows a TLP [24] I-V curve for a 0.25um CMOS NMOS transistor including leakage current evolution. While a fully silicided NMOS device without BEB exhibits instant failure, the BEB device demonstrates a robustness of 10mA/um. Such ESD performance in fully silicided devices had not been demonstrated before in the industry.

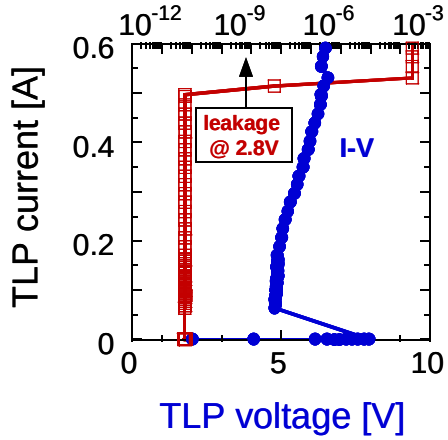

Figure 6 *– TLP data for a BEB fully silicided 0.25um gate length, 50um wide grounded gate NMOS. The leakage evolution indicates a 500mA ESD performance.*

To ensure that the obtained results were reliable, several tens of devices were tested across multiple 0.25um, 0.18um and 0.13um CMOS technologies. Also, endurance tests were set-up. No failure or any leakage shift was found during endurance testing with up to 1000 TLP pulses at It2–10%. When after the endurance test the stress level was increased, the devices failed at the assumed It2 (failure) level.

The keys to BEB are its use of back-end resistive elements and its segmentation of the ballast resistance in parallel segments. This replaces the current industry standard design approach for ESD robustness, which was the use of active area spacing between the drain contact and the gate. The BEB technique is straight-forward to implement.

- BEB uses standard, readily available process elements
- BEB does not require process modification
- BEB does not require additional or different masks

The chief benefits of BEB are two: it allows for the elimination of a process step (silicide blocking); and it allows significant area savings through another innovation: Merged Ballast Circuit layout (MBC), described below.

2.2. Merged Ballast Circuit layout

The Merged Ballast Circuit layout is a major departure from the typical circuit layout approach. With MBC the areas occupied by the ballast resistance are merged into shared areas. This is impossible with active silicon ballasting.

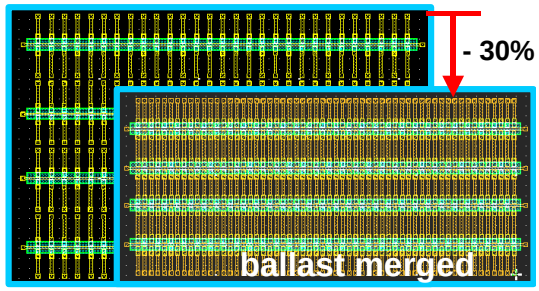

Figure 7 *– Comparison of 4-finger NMOS with conventional poly with BEB design to NMOS with merged ballast technique of adjacent fingers. Significantly compact layout is achieved: ballast poly stripes of adjacent drain and source fingers, respectively, are combined within the same area. Area savings of more than 30% are feasible.*

However, since BEB design builds the ballast from isolated segmented back-end ballast segments, drain and source ballast area can be shared with the ballast resistors of the neighboring fingers, as illustrated in Figure 7. In order to apply this approach, the contact pitch is increased to two times design rule minimum pitch. It has been confirmed in silicon that the double pitch does not deteriorate the ESD performance.

3. Novel Circuit Design Techniques

So far, this paper has focused on micro-ballast resistance to ensure uniform current distribution in a single transistor. However, macro-ballast resistance is also needed to ensure multi parallel transistor performance under ESD.

Due to the (bipolar) snapback behavior, which is statistically distributed, mostly only one arbitrary finger is triggered when ESD stress is applied to a multi-finger transistor. The risk occurs that one single triggered finger may fail before enough voltage is built up again to trigger a neighboring parallel finger.

To reach a uniform current conduction state in the snapback mode, the so-called 'uniformity condition' must be fulfilled (see Figure 8). This means that the single finger failure voltage (Vt2) must be higher then the parasitic bipolar trigger voltage (Vt1) at the onset of snapback, or also: Vt2 > Vt1.

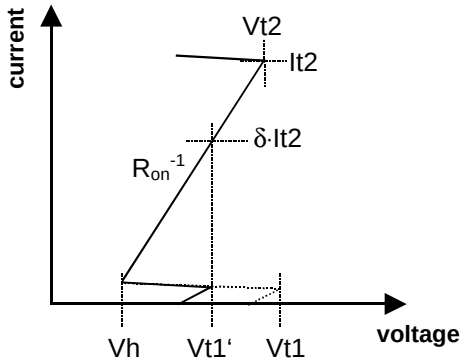

Figure 8 *– Sketch of a generic snapback IV curve indicating characteristic single-finger parameters: triggering voltage Vt1, MFT voltage Vt1', bipolar snapback holding voltage Vh, failure voltage Vt2, failure current It2, MFT current δ·It2 with $0 \le \delta \le 1$. The dynamic on-resistance Ron is inversely proportional to the IV slope in the linear high current region.*

Figure 9 shows the snapback TLP-IV curves of regular fully-silicided multi-finger ggNMOS transistors (0.18um-CMOS technology) where the gate width was varied from W = 1x to 12x50um. For these structures, the uniformity condition is not satisfied. Therefore, all structures reveal the same failure level, indicating that only one finger conducts the TLP current until the single segment fails.

To meet the uniformity condition and thus to improve the ESD robustness of these multi-finger devices, one has two options:
- increase Vt2 and/or,
- reduce Vt1

Vt2 can be increased by adding ballast resistance to the structure, e.g. by introducing silicide blocking, back-end ballasting, or N-well resistor extensions into NMOS-type protection structures. A major drawback of the macro-ballasting approach (or 'Ron-engineered MFT', cf. below) is the increase of the clamping voltage caused by the higher dynamic on-resistance as well as higher power dissipation for the same ESD current. Moreover, a significantly larger silicon real-estate is required to provide the ballast resistance.

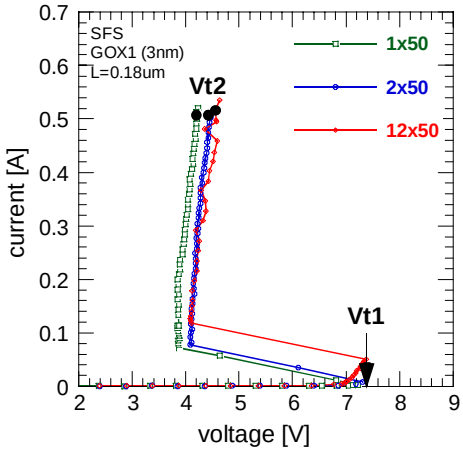

Figure 9 *– Snapback IV-curves of regular fully-silicided ggNMOS devices (W=1, 2, 12x50um) indicating no multi-finger width-scaling behavior of the ESD performance.*

Vt1 can be decreased by both 'static' and 'dynamic' design techniques. For example, using a Zener diode breakdown to pull up a gate (or base) of a NMOS (parasitic npn) device is a static design technique, which is largely independent of other conditions than static voltage bias [20-22]. In general, when a certain I-V bias is reached, the Vt1 reduction goes into effect quasi independent of any dynamic parameter.

A dynamic approach is typically based on a RC circuit, which will provide temporary bias to a gate (or base). Well-known designs are the gate-coupled NMOS base coupling (or substrate pumping), and dynamic triggering [23]. For these approaches dynamic boundary conditions must be fulfilled, which leads to several challenges and issues for successful dynamic trigger implementations. The concerns are:
- Rise-time dependent triggering, while a real life ESD event can have a widely varying rise time

- Providing proper R [kOhm] and C [pF] values, within the constraints of the process technology
- Correct turn-on and turn-off to avoid hot carrier exposure during (dynamic) gate bias, but at the same time ensure sufficient bias to trigger the entire multi-finger transistor

A dynamic design should not be too narrowly defined to work for just one small rise time window and can lead to tester-to-tester correlation issues. Furthermore, due to the significant capacitance added to the protection structure, these multi-finger triggering schemes disqualify for RF I/O applications.

3.1. Novel Multi-Finger Turn-on circuits

A novel approach was developed to fulfill the uniformity condition Vt2 > Vt1: Multi-Finger Turn-on (MFT).

The generic MFT method is to derive a multi-finger turn-on bias signal from a non-uniform conduction situation. This bias signal is transferred to inactive device segments by transfer circuits to activate gate and/or bulk bias turn-on schemes. This will ensure that the uniformity condition is satisfied after the first triggering event at regular Vt1. A modified *MFT uniformity condition* could be expressed as *Vt1' < Vt2* where Vt1' corresponds to the trigger voltage of those fingers that did not trigger initially at Vt1 – the MFT voltage (see also Figure 9). Therefore, we also refer to the novel device type as '*Vt1'-engineered MFT*' as opposed to the '*Ron-engineered MFT*' where multi-finger triggering is reached by pure ballast resistance implementation only.

One distinguishes different turn-on mechanisms:
- *Subsequent turn-on:* bias signals are propagated from finger to finger. Fingers turn on subsequently.
- *Simultaneous turn-on:* the bias signal is fed to all fingers at the same time. Fingers can turn-on simultaneously (if snapback is 'eliminated').

These types differ mainly in their transfer circuits to provide the turn-on bias and in the applicability within self-protecting output drivers.

An important feature, as discussed in the next sections, is that MFT designs are static and some feature an 'auto-timed' current balancing mechanism:
- *Static auto-on:* the MFT effect is derived from a static bias, which is largely independent of other conditions. When a certain ESD current level is reached in the active device part, the MFT goes in effect. No dynamic boundary conditions exist.
- *Static auto-off:* in some implementations the MFT bias is only present as long as there is non-uniform conduction state. When the device uniformly carries the ESD current, the bias vanishes. This is particularly advantageous in view of hot carrier exposure when a gate biasing technique is used.

3.2. Soft-grounded-gate NMOS MFT

The equivalent circuit of a gate-bulk coupled or soft-grounded-gate NMOS MFT (s-ggNMOS) device is depicted in Figure 10. This particular gate-biasing technique is based on a substrate potential pick-up, which ties the gates to the local substrate (bulk) rather than wiring it hard to ground by metal as in a ggNMOS device. During normal operation, the bulk is safely grounded, hence disabling NMOS action and avoiding interference with the normal IC functionality.

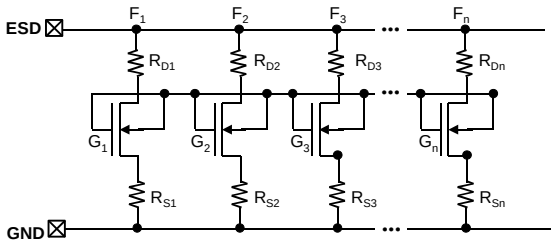

Figure 10 *– Equivalent circuit of a soft-grounded-gate NMOS MFT for simultaneous gate-biasing and multi-finger turn-on.*

In case of an ESD event, however, the drain-bulk junctions of all NMOS transistor fingers are driven into avalanche breakdown and impact generated holes are locally injected into the substrate thus lifting its potential. Eventually, at a self-bias of approximately 0.8V, the avalanche process turns on the parasitic BJT within one (or more) arbitrary finger(s). During subsequent snapback of the activated device segment, the substrate is further pumped with holes. The resulting positive potential due to avalanche breakdown and snapback operation is applied to the gate and indirectly also to the bulk ties of other fingers.

Due to gate-coupling and bulk-coupling effects, the MFT voltage of inactive device parts is reduced (Vt1' < Vt1) enabling multi-finger turn-on of the entire device. Note, the resistors R_{D1} and R_{S1} might be required for micro-ballasting of single finger as well as macro-ballasting to increase Vt2 and to further support multi-finger triggering by fulfilling Vt1' < Vt2. An example is given in Figure 11.

The triggering voltage Vt1 reflects the snapback of an initially triggered finger without trigger voltage reduction. By driving an arbitrary finger into snapback, the turn-on bias for another inactive segment is generated. The significantly reduced MFT voltage Vt1' is slightly higher than the holding voltage. A second finger can turn on when a sufficient but small voltage drop across the active device segment is provided. By the same effect all fingers are activated successively (see inset Figure 11) with increasing current. The resulting uniform conduction state and ideal width scaling are reflected in the failure current of It2 = 2A, which corresponds to a normalized ESD performance of It2 = 10mA/um.

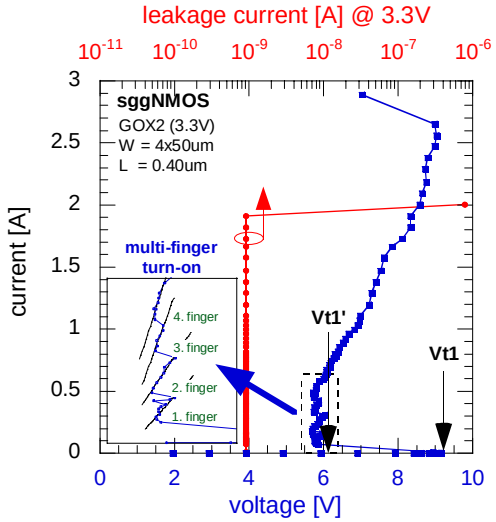

Figure 11 *– TLP snapback IV curve of a soft-grounded-gate NMOS MFT (7nm GOX2, W = 4x50um, L = 0.40um) including leakage current evolution. Inset: close-up of triggering regime indicating successive finger turn-on with increasing current at a reduced MFT voltage Vt1'.*

3.3. Domino-type MFT

Figure 12 depicts the domino-type MFT for subsequent finger triggering. This configuration consists of n NMOS transistor fingers in parallel who will turn-on one after the other in analogy to falling domino blocks.

The drain and source resistors R_{Di} and R_{Si} are implemented for single-finger micro-ballasting and will also support multi-finger triggering by macro-ballasting. The resistors $R_{Si,MFT}$ in each source finger are used to generate a potential for activation of the MFT voltage reduction mechanism (Vt1' < Vt1). If single finger ESD performance is uncritical, only the source resistors $R_{Si,MFT}$ are required to achieve homogeneous current flow in MFTs. The MFT mechanism for the domino-type device is explained in detail in the following.

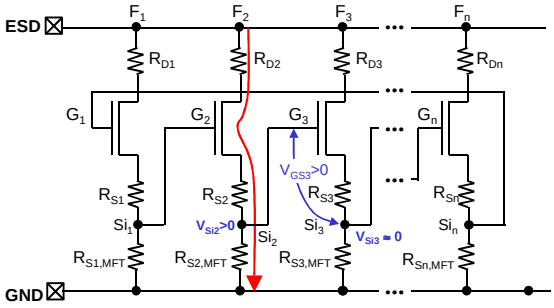

Figure 12 *– Equivalent circuit of domino-type NMOS MFT (source-gate-coupled) for subsequent finger triggering indicating the function of the device.*

After one arbitrary finger triggers as indicated by the arrow at finger F_2, the initial stress current is solely conducted by this device segment. As a result, a source potential V_{Si2} builds up at the internal source node Si2 due to the voltage drop across $R_{S2,MFT}$. These source resistor elements can be regarded as ESD current sensors, which are implemented into each finger, firstly, to sense ESD events, and, secondly, to provide the signal to turn on inactive fingers - in the case of the domino-type MFT the adjacent finger. Moreover, these resistors act as regular macro-ballasting elements by increasing the dynamic on-resistance in the single elements. This combined purpose allows extremely area efficient MFT layouts. The potential V_{Si2} is applied to the NMOS gate of finger F_3 by connecting the gate to the internal source node Si_2. As long as F3 is inactive and no current flows through this finger, V_{Si3} will be zero and thus a positive gate-source bias V_{GS3} will exist. This will eventually drive the finger into NMOS operation, which results in a reduction of the

parasitic NPN triggering voltage due to the well-known gate-coupling effect. Note, even for sub-threshold gate bias, generally a significant Vt1 reduction occurs as will be demonstrated below.

By sufficiently decreasing Vt1 towards the holding voltage, eventually the inactive finger F3 turns on. The same mechanism transfers the internal source signal at Si_3 to the gate of F4, thus triggering this finger. By creating this domino-effect of subsequently triggered fingers, eventually the entire structure is forced into a homogeneous conduction state.

The domino MFT design technique has two major advantages:

- The MFT mechanism is *'static'* and driven instantly by current, not time. Thus, no timing issues as for instance for dynamic gate-coupling schemes can occur as described above.
- The MFT mechanism is *'auto-timed'* – it only acts during a limited (i.e. short) time interval, until uniform current conduction is reached. Evidently, in this state, the current is equally distributed to all fingers such that all internal source nodes are on the same potential. As a result, no gate-source bias exists, which prevents the structure from excessive hot-carrier exposure.

Therefore, the MFT technique does not only enhance multi-finger triggering, it also provides a dynamic balancing bias between fingers that would conduct different amounts of current.

Figure 13 shows the TLP measurement results for a domino-type MFT. After initial device triggering at Vt1 ≈ 7.5V, the MFT mechanism reduces the turn-on voltage of subsequently activated fingers to roughly Vt1' ≈ 5.5V. By this, eventually all 16 fingers are forced into uniform current conduction, which results in an optimum failure current of It2 = 16 x 50um x 10mA/um = 8A.

This excellent result indicates again a perfect linear width-scaling behavior. Moreover, HBM stress tests up to the pre-charge voltage tester limit of 8kV could not damage this device.

Both the MFT implementations as shown before can only be employed as pure ESD protection clamps because of the gate connection to ground. Though not described here, domino-type MFT designs can be adapted to make self-protecting NMOS drivers. In the next two sections, other MFT schemes will be presented, which can be applied as self-protecting NMOS drivers as well.

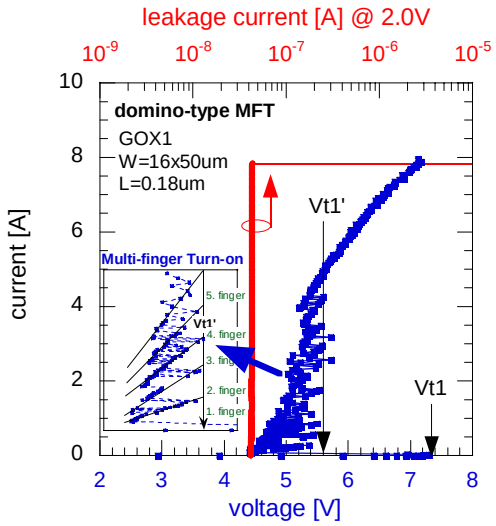

Figure 13 *– TLP snapback IV curve of a domino-type NMOS MFT (3.5nm GOX1, W = 16x50um, L = 0.18um) including leakage current evolution. Inset: close-up of triggering regime indicating successive finger turn-on with increasing current at a reduced MFT voltage Vt1'.*

3.4. NMOS-type and diode-type MFT

The schematic in Figure 14 shows a different type of gate driven triggering and current balancing MFT. The active approach employs small NMOS transistors N_{gi}. These elements control simultaneously the bias of the common gate line connected to all gates G_i of the ESD-current conducting NMOS fingers N_i in parallel.

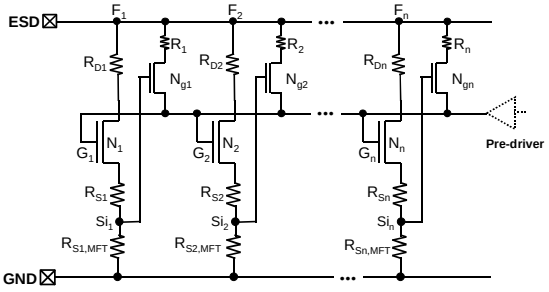

Figure 14 *– Equivalent circuit of NMOS-type MFT for simultaneous gate-biasing and multi-finger turn-on.*

The advantage of the presented circuitry is that it can be applied in NMOS output drivers since gate

bias from a pre-driver stage can be supplied, whereas the ESD triggering elements $R_{Si,MFT}$ and N_{gi} do not compromise regular circuit operation. The principle approach of sensing an ESD conduction state of an arbitrarily triggered finger and of generating a signal for the turn-on of inactive segments is equivalent to the method presented above for the domino-type MFTs.

Again, MFT source resistors $R_{Si,MFT}$ are inserted in each source finger to derive a bias signal during ESD current conduction. The potential at the internal source node Si of an initially triggered finger is used to positively bias the gate of the small NMOS transistor N_{gi}. Note, these NMOS devices act as voltage followers, i.e. they transfer the source potential to the common MFT gate as soon as V_{Si} exceeds the threshold voltage V_{th}. Thus, the common MFT gate is simultaneously biased to a gate voltage $V_{G,MFT} = V_{Si} - V_{th}$. This reduces the MFT voltage Vt1' of all parasitic BJTs simultaneously as opposed to the subsequent approach of the domino-type MFT. If the snapback effect is removed by gate-coupling, the technique forces all fingers simultaneously into the BJT on-state.

To achieve a sufficient gate bias, $R_{Si,MFT}$ should be designed to provide more than $2 \cdot V_{th}$ internal source voltage for an ESD current well below the failure-level It2 of the single finger. This results in an efficient MFT gate-bias above V_{th}, which ensures a strong gate-coupling effect, i.e. MFT voltage Vt1' reduction.

During normal operation, regular MOS currents through the NMOS device do not interfere with circuit behavior since only negligible voltages occur at the internal source node.

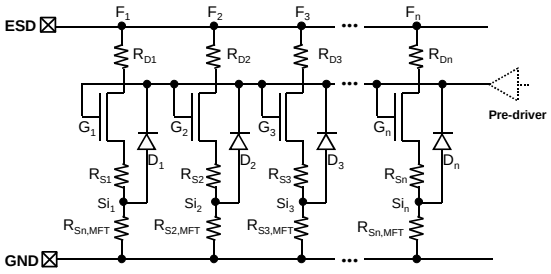

Figure 15 *– Equivalent circuit of diode-type MFT for simultaneous multi-finger turn-on.*

One more MFT driver compatible approach is depicted in Figure 15. The principle function of simultaneous gate-biasing is identical to the NMOS-type MFT. In this configuration, however, minimum sized diodes $D_1 \ldots D_n$ are employed as transfer circuits between the internal source nodes and the common MFT gate line.

After finger F_i conducts ESD stress current, the corresponding diode D_i becomes forward biased. The other diodes will be reverse biased, thus preventing charge loss from the gate line to ground. In accordance to the NMOS-type MFT, a small diode current can charge all MFT gates simultaneously. In case of efficient gate-bias and smooth transition to the BJT on-state (no snapback), eventually all inactive fingers will be turned on simultaneously.

The diode-type implementation is also compatible with normal MOS operation and thus can be exploited as self-protecting drivers.

To achieve an efficient gate-coupling effect, again an MFT gate-bias higher than the NMOS threshold voltage V_{th} is targeted. Therefore, the MFT resistor $R_{Si,MFT}$ must provide the diode built-in voltage plus MOS threshold V_{th}, i.e. $V_{Si} = V_{diode} + V_{th}$. Shottky diodes could be used to reduce V_{diode} to ca. 0.3V.

The diode-type MFT maintains only a small gate-source bias in the homogeneous conduction state. This balancing concept was referred to as auto-on/off.

In Figure 16, TLP analysis data of a driver compatible (NMOS-type) MFT (0.18um-CMOS) is presented.

Evidently, the trigger spikes of subsequently turned on fingers observed for the previous MFT devices are not visible for this configuration. This fact indicates a very efficient MFT gate biasing and thus gate-coupling effect, which reduces the MFT voltage to the ultimate minimum of roughly the holding voltage, i.e. $Vt1' \approx Vh$. Consequently, all fingers go simultaneously and smoothly into the BJT on-state without snapback.

In accordance to the domino-type MFT, again an excellent failure current of It2 = 8A (10mA/um) is reached. This verifies again an ideal performance scaling with device width. HBM tests of the NMOS-type MFT could not damage the device for pre-charge voltages of 8kV – the tester limit.

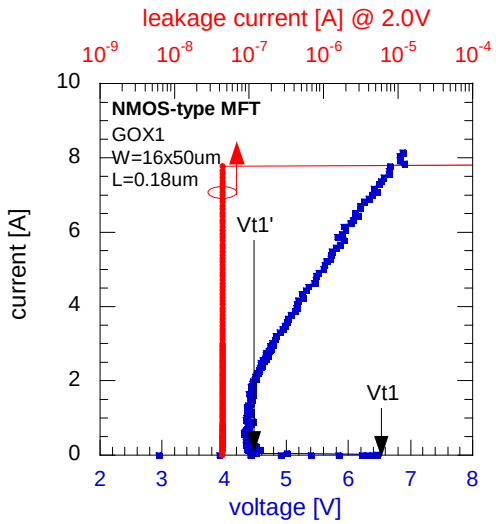

Figure 16 *– TLP snapback IV curve of a NMOS-type NMOS MFT (3.5nm GOX1, W = 16x50um, L = 0.18um) including leakage current evolution. No triggering spikes of subsequently triggered fingers are visible due to a minimum MFT voltage Vt1'≈ Vh.*

4. MFT design

4.1. MFT circuit design

According to Figure 8, multi-finger turn-on is reached in MFT devices when the reduced single-finger MFT voltage Vt1' is lower than the voltage at second breakdown Vt2, i.e. the modified uniformity condition Vt1' < Vt2 must be fulfilled. To accomplish first-silicon success for MFT design under area optimized conditions, it is beneficial to measure the gate-coupling effect, i.e. the MFT voltage Vt1' as a function of the gate-source bias. The corresponding data as extracted from static snapback IV curves of an NMOS transistor (7nm gate, 0.18um-CMOS) is presented in Figure 17. Obviously, a significant reduction of Vt1 occurs already at sub-threshold gate-bias.

In order to describe this behavior analytically, a linear fit is applied to the curve in the relevant gate-source voltage regime. The linear approximation for Vt1' = f(V_{GS})

$$Vt1' = \gamma \cdot Vt1 + \alpha \cdot \left\{ \frac{Vt1 - Vh}{Vth} \right\} \cdot V_{GS}$$

contains two parameters, to be experimentally determined from the slope (α) and the y-axis intersection (γ). From the linear fit in Figure 17, the values γ = 1.1 and α = 0.83 are extracted.

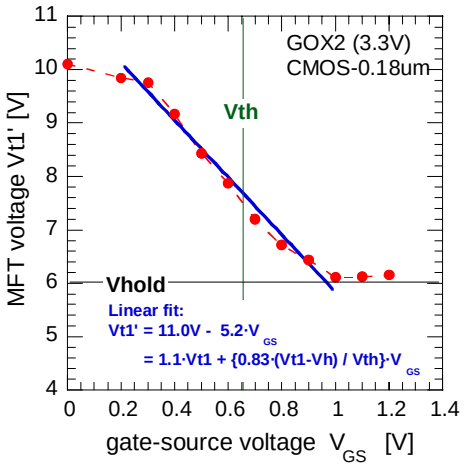

Figure 17 *– MFT voltage Vt1' as a function of the gate-source bias in a 3.3V-NMOS device (0.18um-CMOS) as extracted from static snapback IV curves; a significant Vt1 reduction is already reached for sub-threshold gate bias V_{GS} < Vth*

The gate-source voltage V_{GS} in a domino-type MFT (Figure 11) is generated by the voltage drop across $R_{S,MFT}$ at a current-level of δ·It2 required for turn-on of the next finger, i.e.

$$V_{GS} = R_{S,MFT} \cdot I_{t2} \cdot \delta$$

The failure voltage Vt2 at a failure current It2 can be described as

$$Vt2 = V_h + R_{on} \cdot I_{t2} = V_h + \left(R_{ballast} + R_{S,MFT} \right) \cdot I_{t2}$$

Here, $R_{ballast} = R_D + R_S$ corresponds to the sum of ballast resistance at the drain and source side of the single NMOS fingers. Combining the last three equations inserted into the modified MFT uniformity condition Vt1' < Vt2 results in

$$R_{ballast} + R_{S,MFT} \cdot \left\{ 1 + \alpha \cdot \delta \cdot \frac{V_{t1} - V_h}{V_{th}} \right\} > \frac{1}{I_{t2}} \left(\gamma \cdot V_{t1} - V_h \right)$$

For the values of the regular fully-silicided ggNMOS structures in Figure 9 (Vt1 = 7.5V, Vh=4V, Vth = 0.52V, target value It2 = 0.5A for a 50um-wide device, and δ = 0.75 for subsequent

finger triggering at $\delta \cdot It2 = 0.38A$) the previous equation simplifies to

$$R_{ballast} + \mathbf{4.2} \cdot R_{S,MFT} > 8.5\Omega \quad \textit{for a single finger}$$

Three crucial facts can immediately be derived from this relation:

- The relation directly demonstrates and quantifies one major advantage of MFT implementation compared to regular ballasting approaches: the MFT resistor $R_{S,MFT}$ is approximately a factor of 4 more efficient for multi-finger triggering than plain ballast resistance $R_{ballast}$ implementation. This provides tremendous direct area savings since less ballast area is required for scalable multi-finger devices. Indirect area savings can be significant due to superior voltage clamping capabilities accomplished by a lower dynamic on-resistance Ron of the protection device.
- The formula provides straightforward guidelines for MFT resistance design.
- The equation underlines that a safe MFT design with regard to process fluctuations (e.g. BEB resistance values) can be easily realized by a well-defined $R_{S,MFT}$ without increasing the total on-resistance significantly.

4.2. MFT merged ballast circuit layout

Figure 18 illustrates a concept of a fully functional, metal 1 - routed ESD protection design. Basically, poly resistor arrays are employed to form the segmented drain and source back-end-ballast (BEB). In addition, on the source-side an MFT source-potential pick-up contact is inserted. This splits the resistor into a segmented ballast resistor part R_{Sn} and the MFT source resistor $R_{Sn,MFT}$ in accordance to the MFT schematics presented before. Even with only poly exploited for BEB, area efficient ESD protection elements can be introduced due to the highly efficient $R_{Sn,MFT}$ resistor.

The Merged Ballast Circuit (MBC) layout, as introduced in prior sections, is illustrated in Figure 19.

Combining the advantage of efficient MFT ballasting resistance implementation with the merged ballast circuit technique in a 0.18um-CMOS technology, an ESD area performance of up to $5V_{HBM}/um^2$ could be realized. To our knowledge, this value represents a new milestone for NMOS-type ESD protection. Current industry solutions achieve approximately $1\text{-}1.5V_{HBM}/um^2$.

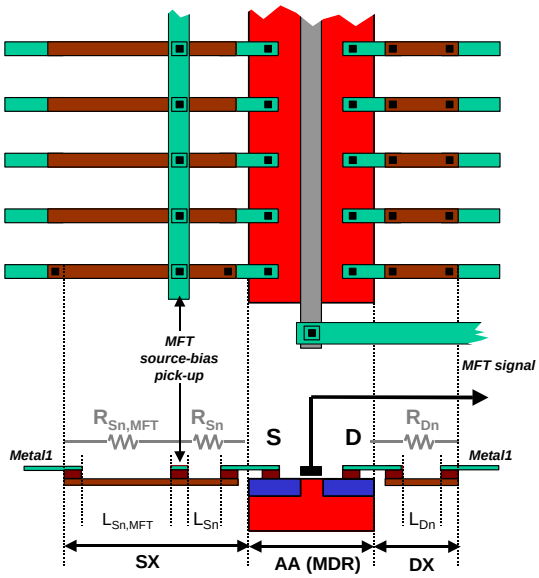

Figure 18 *– Layout and cross-section sketch of optimized MFT design using poly resistor arrays for segmented back-end-ballast R_{Dn}, R_{Sn} and MFT source resistance $R_{Sn,MFT}$ implementation.*

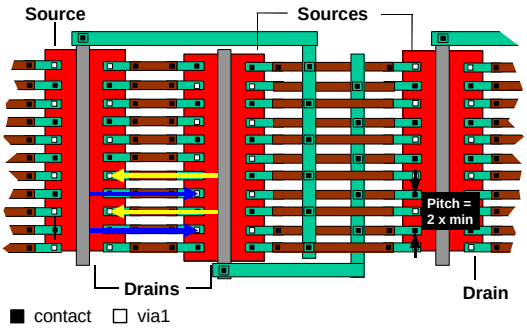

Figure 19 *– Layout sketch of domino-type MFT demonstrating the merged ballast circuit (MBC) technique used to significantly compact layout: ballast poly stripes of adjacent drain and source fingers, respectively, are combined within the same area.*

5. Full I/O Implementation

Figure 20 compares an original silicide-blocked (bi-directional, over-voltage tolerant) I/O cell design (left) to the corresponding layout applying fully-silicided MBC MFT design (middle) as self-protected cascoded NMOS devices and MBC PMOS drivers. The right hand side shows the most aggressive area reduction implementation.

Due to the highly area efficient MFT/MBC implementations, total area savings of 30% and more are demonstrated within this I/O cell.

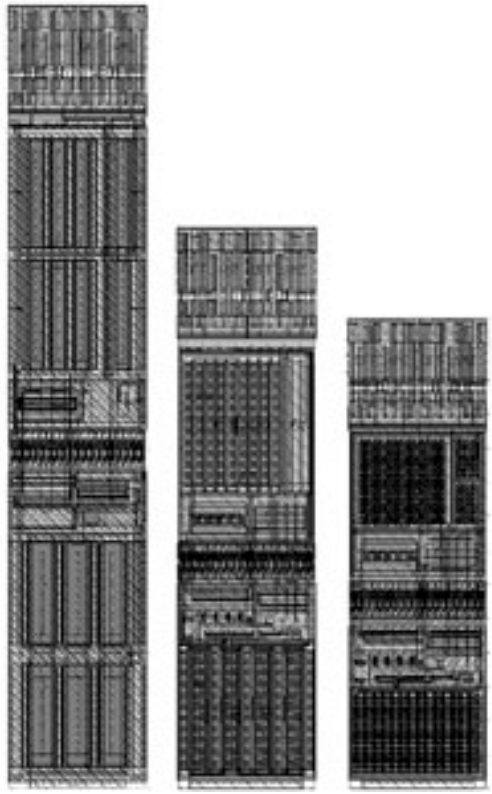

Figure 20 *– I/O cell of original silicide-block design (left), fully-silicided MFT design applying MBC (middle) and most aggressive BEB-MFT-MBC design (right). The direct comparison indicates dramatic area savings of 30% to 40% possible with the new techniques.*

6. Conclusions

Area Reduction and Performance Implications

The chief advantage of the presented innovations in IO drive and ESD design – BEB, MFT and MBC – is their ability to yield significant area reductions. They achieve up to 5V per um^2 HBM, which is 3 to 5 X the industry average.

BEB and MFT techniques in conjunction with a highly area efficient merged ballast circuit (MBC) produce a number of significant advantages in addition to area reduction.

- Linearly scalable multi-finger ESD performance
- Minimum ballasting resistance on I/O drivers and circuits, thus
  - area consumption associated with ballast resistance implementation is greatly reduced
  - faster I/O's can be designed
- MFT resistance RS,MFT is approximately 4 times more efficient than regular ballast resistance implementation, helping to achieve area reduction
- MBC further reduces area consumption while maintaining normal and ESD performance
- No drawbacks in timing and transient interference under normal and ESD operation
- Straightforward analytical design, easily incorporated into any CMOS multi-finger protection structure
- Parasitic capacitance is significantly reduced compared to silicide-blocked NMOS, due to minimum active area (drain-bulk junction); works with minimized dynamic on-resistance to produce a drastic speed increase in self-protecting output-driver MFT devices
- Bulk or combined bulk/gate-coupled MFT scheme can be applied to technologies where gate-coupling is not efficient to reduce Vt1 towards a low MFT Vt1'

This combination of economic and technical benefits promises to make the presented technology the new standard for IC design.

Acknowledgements

The authors gratefully acknowledge their Sarnoff co-workers: Wayne Fisher for FA work and SEM pictures; Girija Kolluri, Laura Housel and Russel Moen for device layout and design support; Les Avery for fruitful discussions; Lou Ann Wingerter for the preparation of some real art-work; Dana Brock for administrative support; Bill Mayweather and Kathy Schad for management support. The authors also acknowledge fruitful discussions and valuable suggestions by Dr. Warren Anderson, Compaq, Dr. Jeremy C. Smith and Dr. Charvaka Duvvury from Texas Instruments and Dr. Tim Maloney from Intel.

References

[1] A. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits*, Wiley and Sons, 1995, ISBN 0 471 95481 0.

[2] A. Amerasekera, et al., *The Impact of Technology Scaling on ESD Robustness and Protection Circuit Design*, EOS/ESD 1994, pp. 237-245.

[3] A. Palella, H. Domingos, *A Design Methodology for ESD Protection Networks*, EOS/ESD 1985, pp. 24-40.

[4] C. Diaz, et al., *Source Contact Placement for Efficient ESD/EOS Protection in Grounded-Substrate MOS IC's*, US Patent filed, 1993.

[5] C. Diaz, et al., *Studies of EOS Susceptibility in 0.6um nMOS ESD I/O Protection Structures*, EOS/ESD 1993, pp. 83-91.

[6] C. Duvvury, et al. *Achieving Uniform nMOS Device Power distribution for Sub-Micron ESD Reliability*, IEDM 1992, pp. 131-134.

[7] C. Duvvury, et al., *ESD Design Considerations for ULSI*, EOS/ESD 1985, pp. 45-48.

[8] D. Krakauer, et al. *ESD Protection in a 3.3V Sub-Micron Silicided CMOS Technology*. EOS/ESD Symposium 1992, pp. 250-257.

[9] G. Krieger, *Non-Uniform ESD Current Distribution due to Improper Metal Routing*, EOS/ESD 1991, pp. 104-109.

[10] HB. Park et al., *A Novel NMOS Transistor for High Performance ESD Protection Devices in a 0.18um CMOS Technology, Utilizing Salicide Process*, EOS/ESD 2000, pp.407-412.

[11] T. Polgreen, et al. *Improving the ESD Failure Threshold of Silicided NMOS Output Transistors by Ensuring Uniform Current Flow*, EOS/ESD Symposium 1989, pp. 167-174.

[12] A. Amerasekera, et al., *Correlating Drain Junction Scaling, Salicided Thickness, And Lateral NPN Behavior With The ESD/EOS Performance Of A 0.25um CMOS Process*, IEDM 1996, pp. 893-896.

[13] C. Duvvury, et al. *Substrate Pump NMOS for ESD Protection Application*. EOS/ESD Symposium 2000, pp. 18-28.

[14] G. Notermans, et al. *The Effect of Silicide on ESD Performance*. IRPS 1999, pp. 154-158.

[15] J.C. Smith, *An Anti-Snapback Circuit Technique for Inhibiting Parasitic Bipolar Conduction*, EOS/ESD 1999, pp. 62-69

[16] K. Bock, et al., *Influence Of Well Profile And Gate Length On ESD Performance Of A Fully Silicided 0.25um CMOS Technology*, EOS/ESD 1997, pp. 308-315

[17] K. Verhaege and C. Russ. *Wafer Cost Reduction through Design of High Performance Fully Silicided ESD Devices*. EOS/ESD Symposium 2000, pp. 18-28.

[18] C. Russ, M. Mergens, K. Verhaege, J. Armer, P. Jozwiak, G. Kolluri, L. Avery. *GGSCRs: GGNMOS Triggered Silicon Controlled Rectifiers for ESD Protection in Deep Sub-Micron CMOS Technologies*. Proceedings of EOS/ESD Symposium 2001, Portland, USA.

[19] M. Mergens, K. Verhaege, C. Russ, J. Armer, P. Jozwiak, G. Kolluri, L. Avery. *Multi-Finger Turn-on Circuits and Design Techniques for Enhanced ESD Performance and Width-Scaling* Proceedings of EOS/ESD Symposium 2001, Portland, USA.

[20] L. Avery, *Low trigger voltage SCR protection device and structure*, US Patent 5,072,273

[21] L. Avery, *SCR Electrostatic Discharge protection for integrated circuits*, US Patent 5,343,053

[22] L. Avery, *Electrostatic Discharge Protection Circuit for a NMOS or Lateral NPN transistor*, US Patent 5,519,242

[23] C. Duvvury, C. Diaz, *Dynamic Gate Coupling of NMOS for Efficient Output ESD Protection*, IRPS 1992, pp141-150.

[24] T. Maloney et al., *Transmission Line Pulsing Techniques for Circuit Modeling of ESD Phenomena*, EOS/ESD 1985, pp. 49-54

About Sofics

ESD SOLUTIONS AT YOUR FINGERTIPS

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