Conference paper

Active-Area-Segmentation (AAS) Technique for Compact, ESD Robust, Fully Silicided NMOS Design

EOS/ESD symposium 2003

This paper describes a layout technique to optimize the ESD performance per area for fully silicided NMOS devices by segmenting the active area of drain and source regions. Efficient multi-finger triggering is achieved by intrinsic inter-finger-coupling through the bulk enabled by compact finger design. The technique is successfully applied in a 0.13um and a 0.18um CMOS technology obtaining HBM ESD capability of up to 8.6V/um2

Active-Area-Segmentation (AAS) Technique for Compact, ESD Robust, Fully Silicided NMOS Design

Bart Keppens, Markus P.J. Mergens, John Armer (1),
Phillip C. Jozwiak (1), Gordon Taylor (1), Russ Mohn (1),
Cong Son Trinh (1), Christian C. Russ (1), Koen G. Verhaege, Frederic De Ranter

Sarnoff Europe, Brugse Baan 188A, B-8470 Gistel, Belgium,
phone: 32-59-433-360; fax: 32-59-275-916; e-mail: bkeppens@sarnoffeurope.com

(1) Sarnoff Corporation, 201 Washington Road, Princeton, NJ-08543, USA

This paper is co-copyrighted by Sarnoff Corporation and the ESD Association.

Abstract - This paper describes a layout technique to optimize the ESD performance per area for fully silicided NMOS devices by segmenting the active area of drain and source regions. Efficient multi finger triggering is achieved by intrinsic inter-finger-coupling through the bulk enabled by compact finger design. The technique is successfully applied in a 0.13um and a 0.18um CMOS technology obtaining HBM ESD capability of up to 8.6V/um^2.

I. Introduction

While IC technologies scale further down, the density of the core blocks is increased, resulting in a core shrink for a constant functionality. This same downscaling is not possible for the ESD protection devices that are needed at the IC's interface (IO's) and supplies. The ESD blocks tend even to increase in size due to higher ESD demands. Moreover, in advanced technologies the junctions get shallower hereby effectively limiting the available volume for heat dissipation, thus claiming more silicon real estate to meet the ESD requirements.

For a number of product applications, such as display drivers, high pin count FPGA's, etc., larger drivers and ESD protection elements result in a trend towards pad-limited design. For these type of IC's there is a strong need to scale down the area used by the peripheral elements while maintaining the same ESD performance. Such a shrink of the IO cell significantly reduces the total chip area, generating a high economical benefit, as evident from Table 1.

Of course, an important technical prerequisite to achieve this goal is a higher ESD performance per area in drivers and ESD protection. Several process and design options have been explored in the past to gain a higher ESD area efficiency. These include process changes, introducing additional process steps such as ESD implants in addition to silicide block [1-7], etc. as well as layout-only techniques like integrated substrate pumps, well ballast extensions, back-end ballasting, etc. [8-11].

In this paper we focus on a layout approach to maximize the performance per area of fully silicided NMOS devices namely Active Area Segmentation (AAS).

First, the paper reviews different nMOS based ESD design techniques including corresponding analysis data and the AAS layout technique is introduced. Further on, the paper presents experimental ESD results (TLP/HBM) for AAS NMOS transistors in two process technologies. The last section briefly discusses a test correlation issue between HBM and TLP results.

Table 1: Cost reduction per wafer due to smaller dies as a percentage of full wafer cost for different die sizes and IO cell height reductions.

		Die size [mm^2]		
		4x4	**8x8**	**12x12**
Dies per 8" wafer		1788	432	180
		Cost reduction per wafer		
IO cell reduction	**50 um**	4.9 %	2.5 %	1.7 %
	100 um	9.8 %	4.9 %	3.3 %
	200 um	19 %	9.7 %	6.3 %

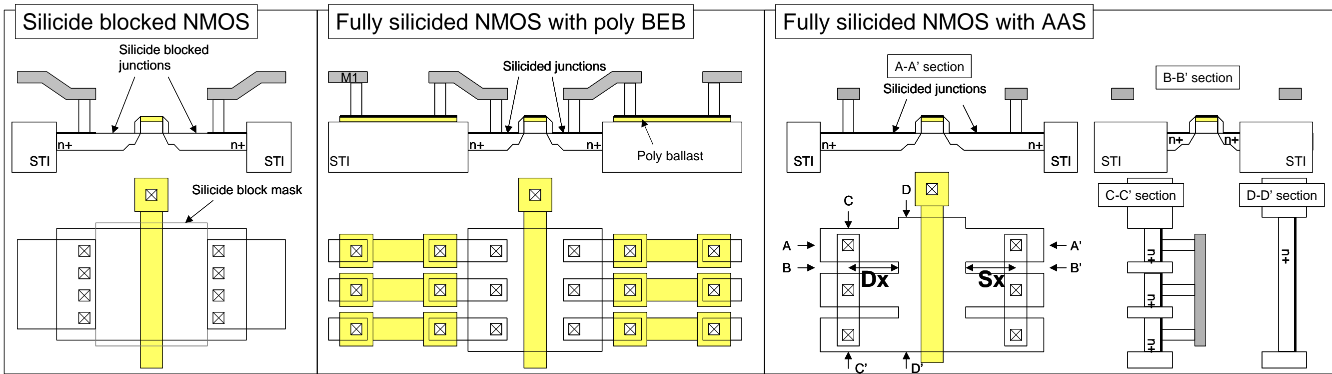

Figure 1: Cross section and layout view of the different devices: A silicide blocked device (left), a fully silicided NMOS with poly ballast (BEB, middle) and a fully silicided NMOS with Active Area Segmentation (AAS, right side).

II. Devices

The Active Area Segmentation devices are compared to silicide blocked and back-end-ballasted ggNMOS devices. An overview of all the device cross-sections and layout views is depicted on Figure 1. These device types were processed in two different CMOS technologies: TSMC-0.13um low resistive substrate bulk process and a proprietary CMOS-0.18um process using a 5.0 µm epi layer with resistivity of 10 Ohm-cm on a Si substrate of 0.015 Ohm-cm. Unless otherwise mentioned, the devices fabricated in the CMOS-0.13um technology contain 20 fingers of 20um width each (i.e. default width: W=20x20u=400u) while the devices in the 0.18um process all contain 18 fingers of 20um (W=18x20u=360u). The devices in 0.13um are all processed as thick gate oxide 'GOX2' structures with an electrical equivalent thickness of 6.6nm. In the 0.18um process thin 'GOX1' (4nm) as well as thick oxide 'GOX2' (8nm) samples were available. In both processes the devices were created in a regular as well as isolated pwell. The latter is created using a so-called Deep nwell, a layer commonly used for RF IC's to obtain better transistor isolation from a noisy substrate (See Figure 2).

III. Analysis results

TLP characteristics of reference devices, BEB and AAS elements were obtained with a Barth Electronics 4002 Transmission Line Pulse (TLP) tester with a pulse duration of 100ns and default pulse rise time of 10ns. Fast rise time measurements (200ps) were performed to study the influence of slew rate. For the TLP analysis, a first significant leakage increase is used as the failure criterion [12].

HBM analysis was performed using two different testers: in-house HBM system (a Keytek-Verifier) and a tester from an independent test house (a Keytek MK2). Results from the independent test house were consistently higher, however, showed the same trends. Only the results from the in-house tester are retained in this paper.

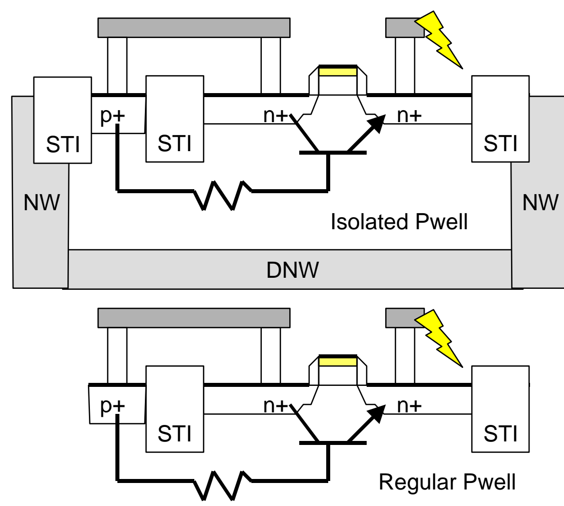

Figure 2: Cross section of NMOS devices in isolated Pwell using a deep N-well (DNW, above) and in regular Pwell (below). The DNW process option is frequently used in RF IC's to isolate nMOS transistors from a noisy substrate.

Sections A to B contain the TLP/HBM results for the 0.13um and 0.18um devices, followed by a discussion in section C. Finally, a theory to explain an observed HBM/TLP correlation issue will be presented.

A. Results on TSMC 0.13um

Figure 3 shows TLP measurement results on TSMC 0.13um, thick oxide (GOX2) silicide blocked ggNMOS devices. The devices show good TLP performance with and without ESD implant.

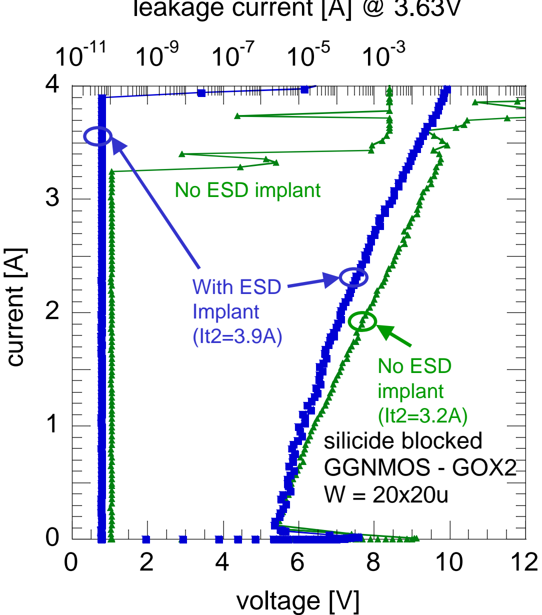

Figure 3: TLP-IV measurements (using bottom X-axis and Y axis) and corresponding leakage evolution (on top X-axis vs. Y-axis) on silicide blocked ggNMOS devices (0.13um TSMC). Both devices have a total gate width of 400um. Adding an ESD implant increases the ESD performance.

Figure 5 depicts measurements on 400um (20x20um) BEB ggNMOS devices and AAS devices with an It2 value of 2A for both.

Although these BEB and AAS It2 values are lower than the silicide blocked performance for the same gate width of W=400um (It2 Silicide_Blocked=3.2A), the resulting ESD performance per area is higher for the BEB and significantly higher for the AAS solution (Table 2 and Figure 4). Additionally the clamping behavior is much better for the BEB and AAS devices since they exhibit a much steeper curve after triggering which leaves more margin in the ESD design window (lower voltage drop in protection device for the same ESD current stress level).

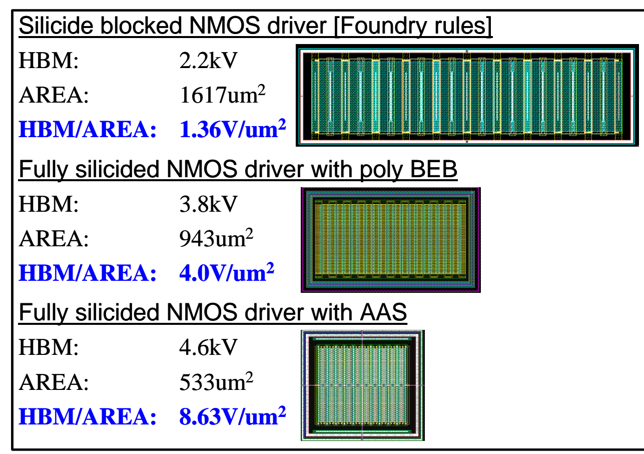

Figure 4: Area comparison between silicide blocked, BEB and AAS devices for a 400um wide NMOS output driver in TSMC 0.13um. Using the AAS layout technique a factor 6 improvement of ESD performance per area can be achieved with respect to the silicide blocked device.

Table 2: Summary of TLP and HBM results for TSMC 0.13um GOX2 ggNMOS devices with 20 fingers of 20um. ESD implant can improve the performance of silicide blocked devices visible in both TLP and HBM results. **Isolated Pwell** (values on the second line in bold) can improve the HBM performance drastically. [All values are the minimum obtained for a number of samples]

0.13um GOX2 Devices	**Area [um^2]**	**It2 10ns [A]**	**It2 200ps [A]**	**It2/area [mA/um^2]**	**HBM [kV]**	**HBM/area [V/um^2]**	**HBM/It2 factor [kV/A]**
Silicide blocked	1617	3.2		2	2.2	1.36	0.7
Silicide blocked with ESD implant	1617	3.9		2.4	6.6	4.1	1.7
BEB	943	2.0		2.1	1.2 **3.8**	1.3 **4.0**	0.6 **1.9**
AAS 1 DX=MDR SX=0	533	2.0 **2.2**	2.0	3.7 **4.1**	2.4 **4.6**	4.5 **8.6**	1.2 **2.1**
AAS 2 DX=MDR SX=MDR	720	0.7 **2.2**	2.0	1 **3.0**	2.0 **3.8**	2.8 **5.3**	2.9 **1.7**
AAS 3 DX=MDR+0.5 SX=MDR	924	0.7 **2.4**	2.2	0.75 **2.6**	0.6 **4.0**	0.6 **4.3**	0.9 **1.6**
AAS4 DX=MDR+1.0 SX=MDR	1128	0.3 **2.25**	2.2	0.27 **2.0**	0.6 **2.6**	0.5 **2.3**	2 **1.2**
AAS5 DX=MDR+2.0 SX=MDR	1536	1.4 **0.35**	2.4	0.9 **0.23**	0 **0.6**	0 **0.4**	0 **1.7**

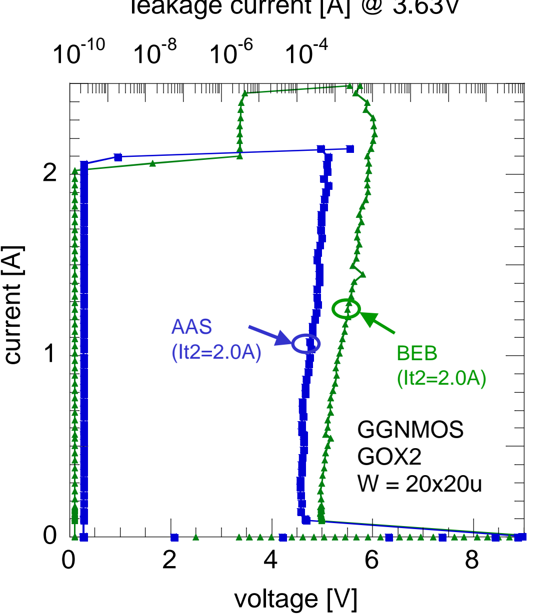

Figure 5: TLP-IV curves and corresponding leakage evolution of a ggNMOS device (0.13um TSMC) with BEB (minimum poly on source and drain side, respectively) and a device with AAS.

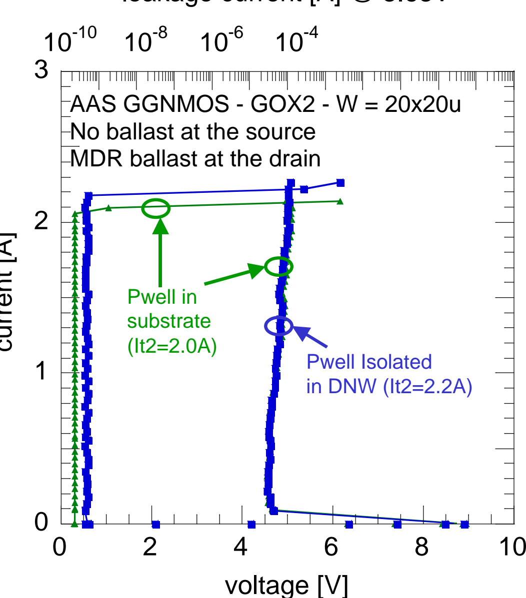

Figure 6: TLP-IV measurements on AAS ggNMOS devices (0.13um TSMC) in a regular p-substrate and in an isolated Pwell. The TLP-IV curves are identical.

Figure 6 depicts TLP results on identical ggNMOS devices using the AAS layout with different bulk material: regular pwell in p-substrate versus isolated pwell using a deep Nwell. The TLP-IV curves are identical but the device in Isolated Pwell has a slightly higher It2 value.

On Figure 7, TLP curves are shown from AAS devices with different lengths for the N+ segmentation region at the drain, all processed in isolated pwells. The device with no ballast at the source has the lowest holding voltage. The IV curves for the remaining devices are very similar and do not show a clear increase in Ron because the contact (remains unchanged) from metal1 to the N+ junction is the main contributor for the total resistance.

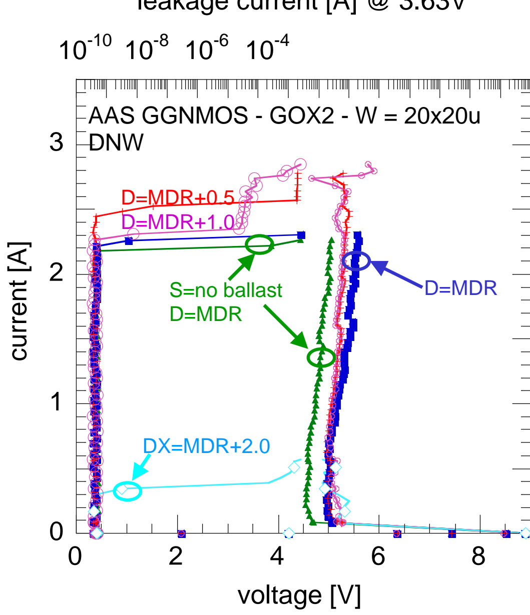

Figure 7: TLP-IV curves of AAS devices with different N+ AAS lengths (0.13um TSMC). All devices have a total gate width of 400um. No visible change in Ron slope for increased N+ ballast.

B. 0.18u CMOS

1. GOX1 - 4nm

Multi finger devices with 18 fingers of 20um were processed in a proprietary 0.18um CMOS process using GOX1 layers as well as GOX2 layers. Figure 8 shows TLP measurements on the thin gate oxide devices for different kinds of ballasting. The ESD performance per area is improved for the BEB and AAS devices (Table 3).

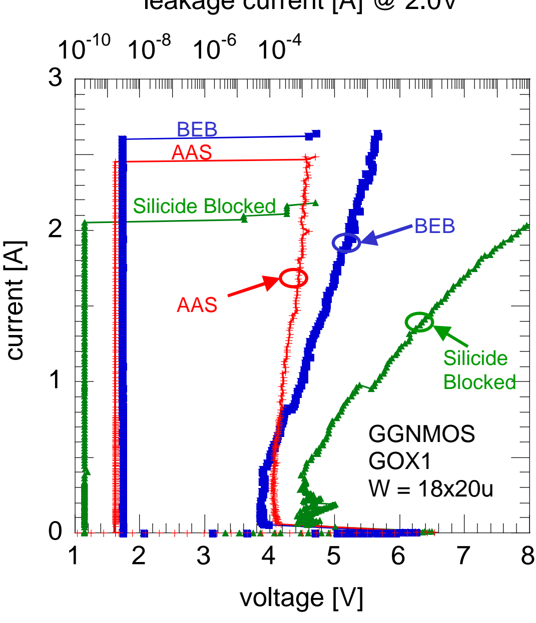

Figure 8: Comparison TLP-IV and leakage evolution plots for 0.18 GOX1 devices (W=18 x 20 um) using silicide blocked, Back-end-ballast and Active Area Segmentation NMOS devices. The ggNMOS device with AAS layout has the lowest Ron.

On Figure 9 identical AAS NMOS devices using different pwell configurations are shown. Like in the case of the 0.13um devices, there is no difference in behavior visible in the TLP results.

2. GOX2 - 8nm

NMOS devices processed in the same 0.18um process using GOX2 were studied. Figure 10 shows TLP results for silicide blocked, BEB and AAS devices.

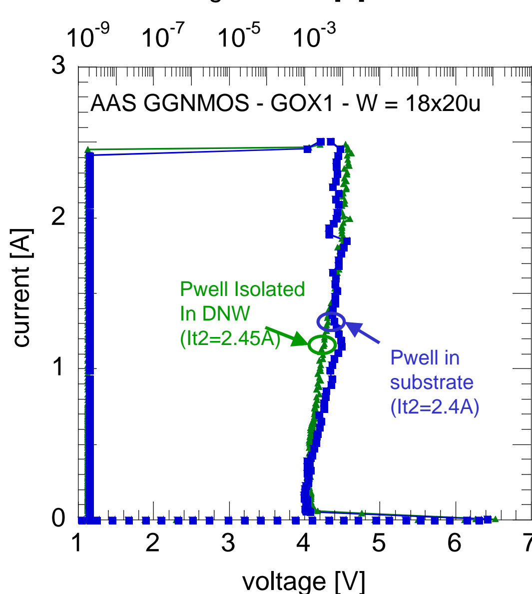

Figure 9: TLP-IV measurements on AAS ggNMOS devices in a regular p-substrate and the same device in an isolated Pwell.

Table 3: Summary of TLP results for 0.18um ggNMOS devices with 18 fingers of 20um. **Isolated Pwell** (values on the second line in bold) do not further improve the TLP performance.

0.18um GOX1 Devices	**Area [um^2]**	**It2 10ns [A]**	**It2/area [mA/um^2]**
Silicide blocked	1431	2.1	1.5
BEB	1009	2.55	2.5
AAS DX=MDR SX=MDR	740	2.4 **2.45**	3.2
0.18um GOX2 Devices	**Area [um2]**	**It2 10ns [A]**	**It2/area [mA/um2]**
Silicide blocked	1322	1.75	1.3
BEB	1066	1.5	1.4
AAS DX=MDR SX=0	584	1.62 **1.64**	2.7 **2.8**
AAS DX=MDR SX=MDR	798	1.75 **1.68**	2.2 **2.1**
AAS DX=MDR + 0.17um SX=MDR	1030	1.8	1.75
AAS DX=MDR + 0.35um SX=MDR	1160	1.95	1.68
AAS DX=MDR + 0.67um SX=MDR	1391	1.92	1.38

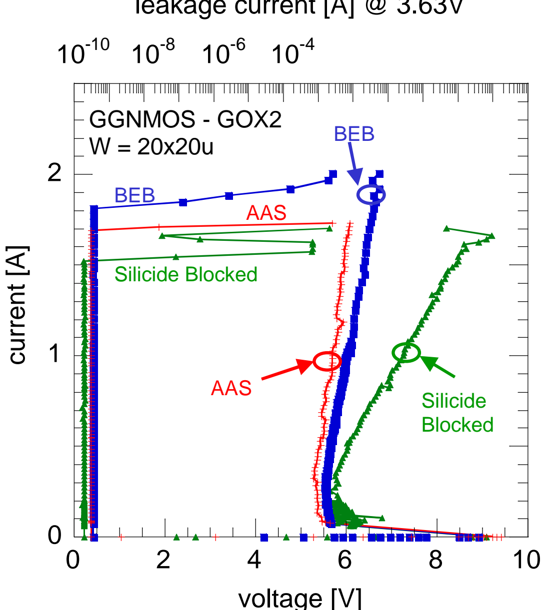

Figure 10: Comparison TLP-IV and leakage evolution plots for 0.18 GOX2 using silicide blocked, Back-end-ballast and Active Area Segmentation NMOS devices.

C. Discussion

In the case of silicide-block devices uniform, multi-finger triggering is achieved by blocking the silicide on the junctions close to the gate in every finger and thus introducing ballast resistance. Typical for silicide-blocked devices is the high Ron.

More area efficient alternative MOS design approaches are based on ballast resistance segmentation, which can be implemented in two ways:

The Back-End-Ballast (BEB) segmentation concept, reported before [8], employs "back-end" poly resistor arrays connected to drain and source, respectively. By adding series resistance in every segment the current uniformity within the finger is improved (micro-ballasting) and at the same time the dynamic on-resistance is increased (macro-ballasting) facilitating the trigger of multiple fingers through Vt2 increase above Vt1. Crucial is that this segmentation results in a current homogenizing mechanism defocusing the current at the onset of current crowding and resulting in a stable single finger ESD performance (Figure 11).

Additionally Multi-Finger-Trigger (MFT) schemes [9] can be applied to further improve the turn-on of all fingers through reduction of Vt1 for the remaining fingers using bulk and/or gate coupling techniques.

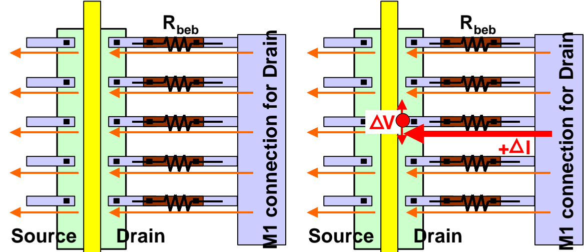

Figure 11: Current defocusing due to segmentation: a local current increase I+dI at onset of current crowding through one BEB element generates a local voltage decrease dV = R_{beb} x dI, which redistributes the current along the gate width.

Active-Area-Segmentation (AAS) represents a similar design technique as BEB. AAS applies small silicided N+ active area for ballast segmentation instead of silicided poly. The AAS is constructed by segmenting the drain and source regions. The same micro ballasting and de-focusing mechanisms works as for the BEB approach, however, for AAS mostly the contact (~10-15 Ohm) contributes to the ballasting resistance.

There is one further very substantial difference to the BEB poly method: AAS can be implemented in an extremely compact way close to minimum feature size still maintaining functional fingers. Obviously, this enhances area efficiency dramatically. Moreover, due to the small device dimensions, the base potential generated by avalanche breakdown during ESD can propagate easily from finger to finger once one "strong" device segment in a multi-finger array is triggered. Consequently, adjacent fingers can easily couple to each other through the substrate. In other words, AAS creates an intrinsic bulk-coupled MFT device.

To confirm this bulk coupling effect, variations in the length of the silicided N+ segmentation at the drain (DX-length) and source (SX length) have been studied.

Adding additional ballast at the source and the drain using active area (AAS) does not change the Ron slope drastically (Figure 7). From the TLP and HBM results (Table 2) it is clearly visible that this intrinsic bulk-coupling effect is only effective for very compact devices since a clear It2 (TLP-rise time: 10ns) and HBM value decrease are noted for less compact devices. Here, bulk coupling is not strong enough anymore due to the increased dimensions (starting from device AAS2 in Table 2). Because the minimum added ballast in the AAS devices is not sufficient to fulfill the Vt2>Vt1 requirement, the multi-finger triggering cannot be realized anymore once the efficient bulk coupling is eliminated.

By using the smallest ballast possible on the drain side (using the minimum STI island allowed) and no ballast resistance on the source side, two objectives are achieved:

First, the single fingers seem to be sufficiently strong due to segmentation to trigger into snapback and efficiently pump the substrate. Second, the resulting AAS device is compact enough such that the adjacent fingers can trigger due to a locally increased bulk-potential generated by the already triggered finger(s).

As shown in Table 2, devices in an isolated Pwell reveal a higher TLP-It2 as well as HBM thresholds. Adding a DNW under the Pwell isolates the NMOS from the substrate (Figure 2) and prevents the avalanche current to be drained away into the substrate and this facilitates direct bulk coupling between the fingers. Additionally, the trigger current It1 is reduced, as depicted in Figure 12.

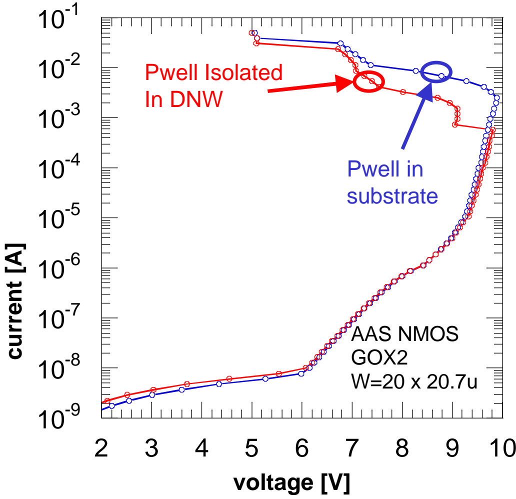

Figure 12: DC current forcing measurements on 2 identical AAS devices using different bulk material. The trigger current is drastically reduced if the NMOS is isolated from the substrate through the use of a DNW.

Here, DC measurements on AAS devices are compared for isolated and regular Pwell. For the isolated pwell device, the trigger current It1 is reduced due to a higher effective bulk resistance. The lower It1 facilitates the triggering mechanism, since less base current is required to turn on the parasitic NPN. Moreover, dV/dt trigger sensitivity is largely enhanced, i.e. the displacement current through junction capacitance can provide significant additional base drive.

When fast ESD transients are applied to the device, the performance decrease for less compact AAS transistors is not visible anymore (Table 2, TLP-10ns versus TLP-200ps). The very fast rise time reduces the Vt1 trigger point as can be seen on Figure 13, where the Vt1 decreases by 3 volt for a 200ps rise time due to the dV/dt effect for an NMOS in an isolated Pwell. For an NMOS in regular substrate the reduction is less pronounced (Figure 15: 1V).

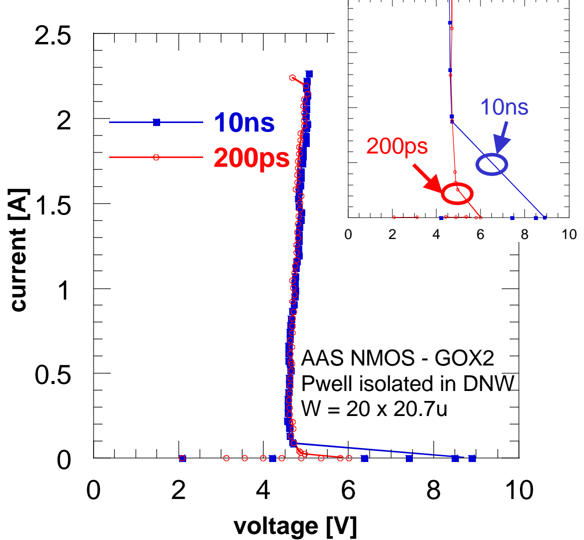

Figure 13: Influence of the pulse rise time on a 0.13um GOX2 NMOS with AAS layout in an isolated Pwell. A clear Vt1 reduction is visible, while the remaining part of the IV curve remains the same.

A fast ESD transient (large applied dI/dt) will generate a high dV/dt over the drain bulk junction capacitance (Figure 14), increasing the bulk potential by displacement current. The displacement current is flowing through the drain-bulk junction for a short period of time. Thus, less avalanche current is needed to provide base bias to trigger the NPN. The effective Vt1 can be reduced significantly below the junction breakdown.

In the event that snap-back devices rely (only) on this dV/dt effect to trigger multiple fingers there is a danger for correlation issues between different test systems (both HBM to TLP and HBM tester to tester correlation). For the AAS devices that have larger AA segments the intrinsic bulk coupling is not efficient anymore as shown by the lower performance values for both TLP-10ns and HBM. These devices show good performance for 200ps TLP, indicating that they rely only on the dV/dt effect to trigger all the fingers.

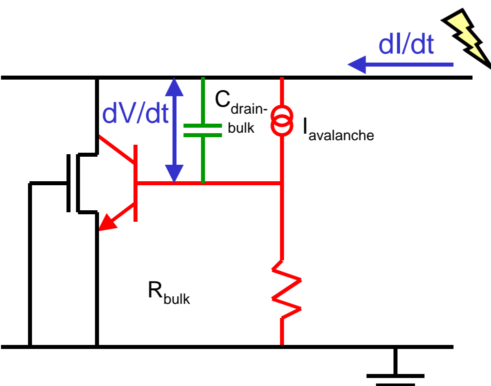

Figure 14: A fast ESD pulse will generate a high dV/dt at the device, which will couple the bulk of the NMOS to a higher potential for a short period. This displacement current allows a faster NPN turn-on because less avalanche current is needed (Vt1 reduction)

Furthermore, the dV/dt effect even increases for increasing (TLP) amplitudes when using pulses with a fixed rise time. As shown on Figure 15, a single finger ggNMOS triggers at an It1 trigger current of 1mA and a Vt1 of 9V using a rise time of 10ns. A 200ps rise time pulse will reduce the Vt1 to about 8V due to a higher slew rate. However, at 50mA TLP amplitude, the 10ns rise time pulse has the same slew rate as a 200ps rise time pulse at the 1mA current level. This effect (slew rate increase at higher amplitudes) will reduce the Vt1 trigger voltage of the device for higher TLP stress amplitudes.

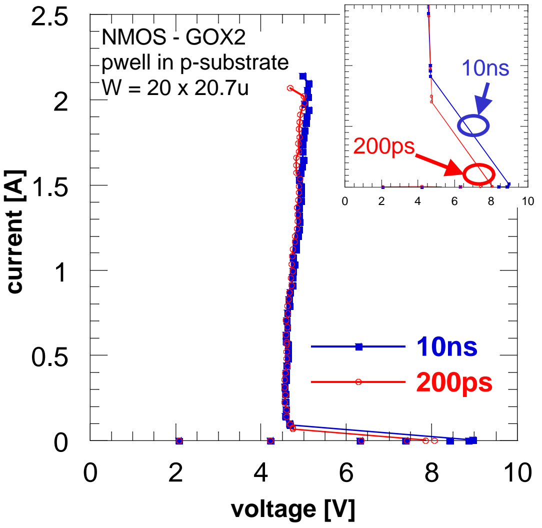

Figure 15: Influence of the pulse rise time on a 0.13um GOX2 NMOS with AAS layout in regular pwell. A clear, but smaller Vt1 reduction is visible, while the remaining part of the IV curve remains the same.

An HBM tester with a relatively small test board capacitance would show consistently higher performance levels than the tester with high $C_{testboard}$. The reason is that effectively the voltage pulse can be faster due to smaller R_{HBM} x C_{TB}.

The HBM tester results show a clear difference between devices in isolated (with DNW) and regular pwell, which points out that for devices in regular pwell not all fingers are triggered. A higher test board capacitance will slow down the applied ESD pulse thus reducing the beneficial dV/dt effect. For devices in an isolated pwell this reduction is less a problem because of the higher sensitivity to dV/dt (Figure 13 vs. Figure 15) due to a lower It1.

IV. Conclusions

In this paper a layout technique that is based on the segmentation of the active area of drain and source junctions is presented. By dividing the active area into smaller segments micro ballasting, crucial to obtain uniform current flow within single fingers and thus robust device segments, is achieved. The extremely compact design enabled by AAS, allows an efficient intrinsic bulk coupling between neighbored fingers (intrinsic bulk-MFT mechanism) required for uniform multi-finger triggering. A DNW layer to create isolated NMOS devices can enhance the direct bulk coupling mechanism significantly. For recent processes with very low ohmic and/or Epi substrates, the HBM performance and correlation between HBM and TLP results can be improved significantly using the DNW layer. In particular, for RF products the DNW is frequently an available design option.

In a number of pad-limited applications, the additional cost imposed by introduction of DNW can easily be compensated and overwritten by the IC area shrink enabled by extremely compact AAB driver and protection design.

The AAS devices have an ideal clamping behavior (low Ron) due to the low ballast resistance needed for segmentation. Moreover, due to the very compact nature of AAS devices, the drain junction capacitance can be minimized, which is beneficial for RF output drivers. The fully silicided design further supports the need for ever-faster drivers and increased drive current.

AAS multi-finger devices were successfully implemented and demonstrated in both 0.13um and 0.18um CMOS technologies. The ESD performance per area of NMOS devices could be significantly improved by a factor 6 over the values achieved with benchmark designs following the foundry rules.

Using this layout technique the area needed to satisfy the ESD performance can be considerably reduced.

The same layout technique was also proven to be successful for increasing the ESD performance for PMOS and cascoded devices in the same technologies.

V. Acknowledgements

The authors would like to thank Natarajan M Iyer, IMEC, for his valuable technical support and many suggestions to improve the structure of the paper. The authors acknowledge valuable input from Xaver Guggenmos, Tim Maloney, Charvaka Duvvury, Steve Halperin, Ajith Amerasekera, Craig Salling and Fred Kuper.

References

[1] D. Krakauer, et al., “ESD protection in a 3.3V sub-Micron silicided CMOS Technology”, EOS/ESD 1992, pp. 250-257

[2] A. Amerasekera et al., “ESD Failure Modes: Characteristics, Mechanisms and Process Influences”, IEEE TED-39, p.2, 1992

[3] A. Amerasekera et al., “The impact of Technology scaling on ESD robustness and Protection Circuit Design”, EOS/ESD 1994, pp.237-245

[4] 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

[5] 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

[6] G. Notermans et al., “The effect of Silicide on the ESD performance”, IRPS 1999, pp.154-158

[7] 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

[8] K. Verhaege et al., “Wafer cost Reduction Through Design of High Performance Fully Silicided ESD Devices”, EOS/ESD 2000, pp.18-28

[9] M. Mergens et al., “Multi-Finger Turn-on Circuits and Design Techniques for enhanced ESD performance and width-scaling”, EOS/ESD 2001, pp 1-11

[10] C. Duvvury et al., “Substrate pump NMOS for ESD protection applications”, EOS/ESD 2000, pp 7-17

[11] V. A. Vashchenko et al.,”Increasing the ESD protection capability of over-voltage NMOS structures by comb-ballasting region design”, proc. IRPS 2003, pp 261-268

[12] B. Keppens et al., “Contributions to standardization of transmission line pulse testing methodology”, proc. EOS/ESD 2001, pp. 461-467

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

As is the case with many published ESD design solutions, the techniques and protection solutions described in this data sheet are protected by patents and patents pending and cannot be copied freely. PowerQubic, TakeCharge, and Sofics are trademarks of Sofics BVBA.

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