Advanced NEMS-based Power Management for 3D Stacked Integrated Circuits

Marius Enachescu, George Voicu, Sorin Dan Cotofana
Delft University of Technology, The Netherlands
{M.Enachescu, G.R.Voicu, S.D.Cotofana}@tudelft.nl

***Abstract*—In this paper we introduce a novel power management architecture for 3D Through Silicon Vias based integration technology. Our approach relies on the synergy of two new technological developments as follows: (i) we utilize a Nano-ElectroMechanical (NEM) device, the Suspended Gate FET (SG-FET), as sleep transistor; and (ii) we make use of the 3D potential by placing the sleep transistor (the entire power management infrastructure) on a dedicated tier of the 3D stacked Integrated Circuit. Due to the extreme low leakage current of the SG-FET our proposal results in 2 orders of magnitude static power reduction, when compared with equivalent counterparts based on traditional CMOS devices. The SG-FET power switch requires about 4x more area when compared to bulk CMOS, however, due to the 3D integration which allows for heterogeneous dies to be stacked, the power gating devices can be placed to a low cost dedicated layer, which also results in a substantial IR-drop reduction with minimum impact on leakage.**

I. Introduction

In the nano era context power consumption has become one of the most important issues in Integrated Circuit (IC) design altogether with variability/reliability related phenomena like electro-migration, IR-drop, crosstalk, and gate delay variations [1]. Leakage power, that was irrelevant in technologies above 130nm, has emerged as an important constraint factor as technology node shrinks. To mitigate the leakage power component in circuit design, many low power techniques, e.g., power gating, body bias, multi threshold voltage devices, frequency scaling [2], have been proposed and implemented in practical designs.

A straightforward and successful way to reduce static power consumption associated with inactive hardware is to apply a power shut-off technique (Power Gating) to inactive (sleeping) Functional Units (FUs) in a design implemented in Multi-Threshold CMOS (MTCMOS) technology. To this end Switch (Sleep) Transistors (ST) are placed between the power/ground rails and the FU. When the STs are active power is supplied through them to the FU providing normal operation conditions, and, respectively, when the STs are open the power supply is cut-off to the gated functional unit, hence the leakage power of the entire FU (gated block) is reduced to the leakage power of the STs. The drawbacks of the solution are: (i) area and propagation delay penalty incurred by the addition to the gated block of STs, isolation and state retention cells; and (ii) undesired IR drop voltage on the STs. Thus, along with power gating benefits a number of design issues and power-performance-area trade-offs need to be addressed [3].

For certain specific low activity battery-operated embedded systems used in applications, e.g., environment monitoring sensors, biomedical implants, the sleep mode power, though having a low value, adds up to an energy figure that represents a significant fraction from the total energy consumed by the chip [4], [5]. For this reason new ways to further reduce the leakage power while mitigating the area and delay penalty are needed. Various enhancements for MTCMOS technologies have been proposed to suppress the sleep transistor leakage current [6]. Moreover it has been suggested that nano-electro-mechanical FETs (NEM-FETs) are viable candidates to replace the High-V_t FET based sleep transistor [7], [8]. However, although fabrication feasibility of simple hybrid CMOS-NEMS circuits has been reported in [9], further technological enhancements are needed for such an approach to become a potentially viable industrial solution.

In this paper we propose a novel 3D power management approach that attempts to alleviate some issues associated with the NEMS utilization as sleep transistor in CMOS power gated integrated circuits. Our proposal, graphically depicted in Fig. 1, relies on a 3D arrangement of the power gated functional units and power management circuits. In this way we allocate a dedicated tier in the stack for the sleep transistors and the afferent control logic. The structure in Fig. 1 can, in principle, be extended such that more levels of active logic and power management tiers are intermixed. While this structure can be also utilized in conjunction with MTCMOS technology we propose to use Suspended Gate FETs (SG-FET) as sleep transistors due to their extremely low off currents (leakage). The main implications of this approach are as follows: (i) the fabrication process complexity is simplified as we do not make use of a hybrid CMOS and SG-FET technology; (ii) the static power consumption is reduced with up to 2 orders of magnitude due to the extreme low leakage current of the SG-FETs; and (iii) the IR-drop is reduced when compared with the 2D implementation, which has a positive effect on timing.

We note inhere that CMOS-based power gating in 3D stacked circuits has been previously addressed in [10] by studying the influence of TSV placement on the power/ground network noise. Our paper however discusses the feasibility of using SG-FET based power gating in 3D stacked circuits and investigates power management architectures and techniques possible in 3D integration, along with potential design trade-offs. To the best of our knowledge, this paper is the first one to

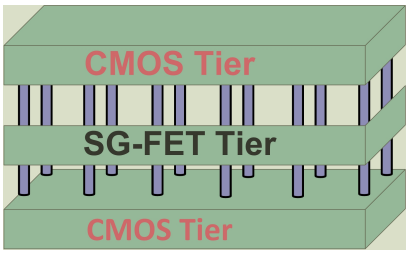


Figure 1. 3D Stacked SG-FET Based Power Management Architecture.

take advantage of the heterogeneous nature of vertical stacked ICs in order to provide a power management architecture based on SG-FET.

To evaluate the practical implications of our proposal we analyze area, IR-drop and leakage for a simplified OpenSPARC T1 processor Execution Unit implemented with power gating using classic CMOS and SG-FET based Sleep Transistors (STs) in 2D and 3D stacked approaches. Our experiments indicate that the SG-FET based implementation offers up to 2 orders of magnitude reduction of the static power consumption, compared with equivalent counterparts based on traditional CMOS devices, at about 4x more area. Additionally due to 3D integration the SG-FET based solution results in 50% IR-drop reduction with minimum impact on leakage.

The remainder of the paper is organized as follows: In Section II a brief introduction is provided on SG-FET including its basic operation. Section III describes the proposed 3D power management architecture and Section IV presents the results of an evaluation case-study and its associated methodology. Finally, concluding remarks are made in Section V.

II. Suspended Gate FET

The Suspended Gate FET (SG-FET) described in [11] is a rather complex device with a 3D geometry as presented in Fig. 2b. Essentially speaking the device behaves like an electromechanical switch which responds to gate bias changes as follows. When the gate voltage (V_G) is low the gate-oxide capacitance is in series with the air-gap capacitance (Fig. 2a) resulting in low electrostatic coupling of the gate to the channel thus in a negligible drain current (I_D). If V_G increases the situation remains unchanged until it reaches a certain "on" voltage case in which the electrostatic force cannot be compensated anymore by the mechanical restoring force and the suspended gate (beam) snaps onto the gate oxide, thus turning on the device. This is called *pull-in effect* and corresponds to a sudden I_D increase. After the pull-in, the I_D increase with V_G is comparable with the one of a standard MOSFET. On the other way around when V_G is decreased from some high value I_D starts decreasing until at a certain V_G value the system becomes unstable due to combined electro-mechanical force and the beam is pulled-out. This causes an abrupt I_D decrease due to a large decrease in capacitance and it is called *pull-out effect*.

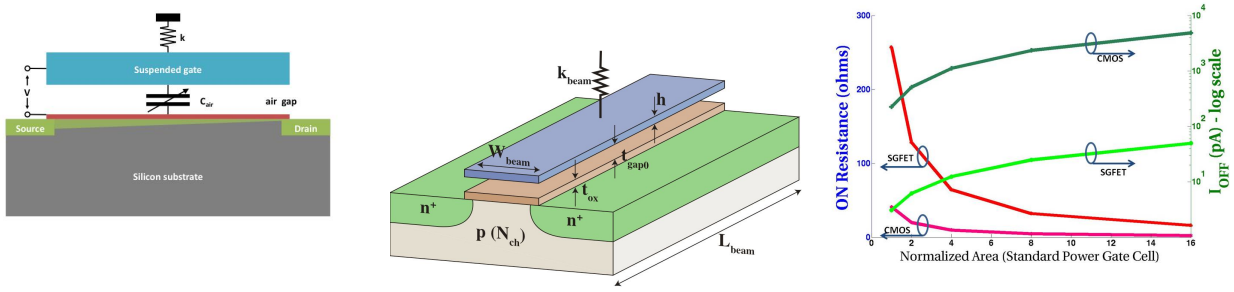
As indicated in [8] SG-FET devices can potentially replace High-V_t sleep transistors due to their ultra low leakage characteristics. Note that for power gated designs two major features are desirable for the sleep transistors: (i) low sub-threshold leakage current to minimize the static power consumption; and (ii) low "on" state resistance to minimize the voltage difference between the virtual and the real power supply nodes. A comparison between the "on" state resistance R_{ON} and "off" state leakage current I_{OFF} of SG-FET and 90 nm High-V_t CMOS based Sleep Transistors (STs) is presented in Fig. 2c for different area in terms of standard cells. The R_{ON} values are computed for an IR-drop of 10 mV over the ST, and the I_{OFF} values consider a 1.2 V power supply. One can observe that when the sleep transistor area is increasing the SG-FET ST R_{ON} tends to become equal with the CMOS ST R_{ON}, while SG-FET ST I_{OFF} (leakage) is about 2 orders of magnitude smaller than the one of the CMOS counterpart.

III. Power management in 3D Stacked ICs

One of the emergent solutions to achieve tight chip integration is to use 3D stacking technology with Through Silicon Vias (TSVs) as interconnects between the stacked dies [12]. TSVs are relative large metal vias (<10 µm diameter, <100 µm length) that are passing completely through the silicon die. The immediate advantage of vertically stacking dies is the interconnect latency reduction, which lately became the dominant latency factor [13]. Moreover, dies implemented in different technologies can be part of the same stack, which opens novel system integration avenues and tradeoffs. The side effect of this performance improvement is the increase in power density in the dies placed in the middle of the stack.

Current power gating architectures place the STs as either a ring surrounding the gated block or as columns throughout the gated block [14]. In this paper we introduce a power management architecture that capitalizes on the heterogeneity of 3D stacked systems. The proposed architecture makes use of NEMS technology dies containing SG-FET STs placed between dies containing the actual power gated circuits, as illustrated in Fig. 3 for a 2-die stack. By moving the bulky STs to a different die, precious area surrounding/inside the gated blocks previously allocated to them in the planar case can be reclaimed. Furthermore, interconnect length inside/between gated blocks is reduced, thereby potentially resulting in increased performance. Moreover this approach can make use of the extreme low SG-FET leakage without requiring an expensive and currently not available SG-FET CMOS hybrid fabrication technology.

One of the key issue in the proposed approach relates to the way the TSV electrical characteristics [15] influence the power gating efficiency and granularity. The "on" state voltage drop-out across the TSV and their behavior at power-on need to be addressed. Copper TSVs exhibit low resistivity, approximately 0.2 Ω for a 5 µm diameter and 20 µm length, hence negligible when compared with the power switch R_{ON}. The TSV pillar is isolated through an oxide layer from the silicon substrate thus the formed parasitic capacitance, typically 40 fF, has the predominant effect on the TSV propagation delay, and



(a) Schematic representation of an SGFET and the associated air-gap capacitance C_{air}. V is the potential difference between the gate and the upper surface of the gate dielectric and k is the effective spring constant of the gate.

(b) 3D schematic: t_{ox} - the thickness of the gate oxide, h - the thickness of the suspended gate, W_{beam} - the width of the beam, L_{beam} - the length of the beam, t_{gap0} - the gap between the oxide and the suspended gate, k_{beam} - the lumped linear spring constant of the beam.

(c) R_{ON} and I_{OFF} for SG-FET and 90 nm High-V_t CMOS Sleep Transistors.

Figure 2. Suspended Gate FET.

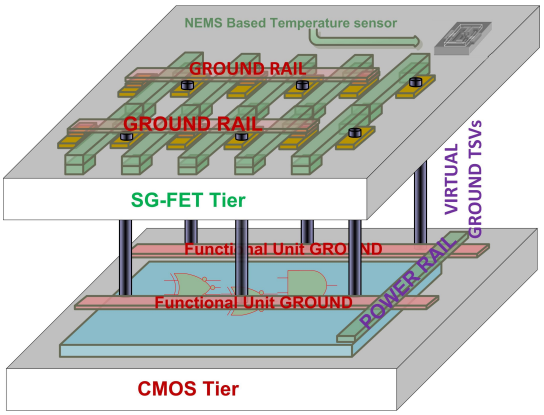


Figure 3. Detailed Representation of 3D Stacked SG-FET Based Power Management Architecture.

potentially limits the power-on time. This capacitance adds up to the gated circuit capacitance, therefore the TSV has a direct impact on the granularity at which the power gating technique can be optimally applied. This capacitance combined with the TSV minimum footprint constraint make the scaling down of the gating granularity being limited by the TSV characteristic instead of the power switch ones.

The power management controller and the always-on cells (for isolation and state-retention) can be placed either on the NEMS die (SG-FET based logic gates were successfully simulated in [16]), either on the logic die containing the gated blocks. As indicated in Table I, the power switches area is lower than the one of the associated gated logic, thus it is advantageous from the silicon area point of view to place the SG-FET die between two computing dies and/or to unify all the power management related blocks on a dedicated die. However, the latest solution infers using a hybrid CMOS SG-FET die processing technology, which can affect the yield. Alternatively, having a single process SG-FET die with a greater yield decreases the manufacturing cost. This cost can also be reduced by designing a generic power gating die consisting of a regular array of SG-FET power switches and CMOS always-on cells that can be reused for any logic designs just by connecting the TSVs to the stacked dies. The power management controller, which is usually chip-dependent, can be placed on the logic die.

To increase the reliability, we suggest to also include a thermal management mechanism by placing NEMS temperature sensors on the same die with the SG-FET sleep transistors. In this way the power management controller can determine when a processing core becomes too hot and signals the operating system to migrate the threads running on that core to other cooler cores. Once the migration is complete the power controller can turn off the core and allow it to cool down.

IV. Experimental Setup and Results

To preliminary evaluate the practical impact our proposal we performed a case study on the 2D and 3D implementations of a simplified OpenSPARC T1 processor [17] Execution Unit containing the ALU and the Shifter. We implemented the design in a commercial 90 nm Low Power MTCMOS technology using Cadence Encounter Digital Implementation 9.1 Low Power Flow. The design is successfully signed-off at 333 mHz. To evaluate the power savings we first implemented a reference design with no power gating mechanisms (the entire design is always powered-on), then we placed the Shifter in a switchable power domain, with 90 nm High-V_t CMOS power switches and SG-FET switches, respectively. Four layouts are investigated: the reference one without power gating (REF), 2D layouts with power gating with classic High-V_t CMOS (2DHVT) and SG-FET (2DSGF) and a two-tier 3D layout with one logic die and a NEMS die with SG-FET power switches (3DSGF).

To dimension the sleep transistors we first had to determine the Maximum Instantaneous Current (MIC) of the Shifter in the routed design which is 4.73 mA. Using the Finite Element Modeling characterization of SG-FET in [16], we sized the power switches in both 2D cases according to [18], in order to achieve a target IR-drop under 20 mV on them for the

TABLE I. RESULTS

Implementation type		Area (μm²)			Power switch parameters						Total power (mW)	
		Shifter	Power switches	Power switch overhead (%)	ON - average IR-drop (mV)			OFF - I_{OFF} (nA)			ON	OFF
					BC	TC	WC	BC	TC	WC		
REF		4487	N/A	N/A	N/A	N/A	N/A	1.08×10^3	2.93×10^2	2.84×10^3	6.731	N/A
2DHVT		5081	594.72	13.25	14.26	19.09	25.82	1.25	8.62	2.22×10^2	6.733	5.9
2DSGF		6626	2139	47.67	N/A	17.12	N/A	N/A	3.01×10^{-2}	N/A	6.73	5.867
3DSGF	ST die	N/A	4482	99.95, but different die	N/A	8.61	N/A	N/A	6.0×10^{-2}	N/A	6.73	5.93
	Logic die	4487	N/A									

measured MIC. The total number of required SG-FET/High-V_t CMOS standard cells are presented in Table II. Our 3D stacked floorplan takes advantage of the extra area available and increases the area of SG-FET based ST up to the Shifter unit one, in order to equally distribute the MIC on the STs and to mitigate the IR-drop effect.

TABLE II. POWER SWITCHES DIMENSIONING

Power gating standard cell	Channel width W(μm)	I_{ON} at $V_{DS} = 20\,\text{mV}$ (mA)	I_{OFF} at $V_{DS} = 1.2\,\text{V}$ (A)	Number of cells
High-V_t based	52	513	8.58×10^{-10}	10
SG-FET based	7.5	51	4.47×10^{-13}	93

Accurate dynamic power and rail analysis with the switchable power domain functioning and turned off were performed using Cadence VoltageStorm to evaluate the power gating efficiency. The switching activity was set to 0.2. Since a dynamic characterization of the SG-FET device was not available at that time, we generated a power grid library containing only the static port view model of it, based on the R_{ON}, $I_{D_{sat}}$ and I_{OFF} values from the FEM analysis of the device. The static port power-grid view of the High-V_t CMOS power switch was used for consistency. For the High-V_t CMOS power switch we ran a multi-mode multi-corner analysis in typical (V_{DD} = 1.2 V, T = 25 °C), best (V_{DD} = 1.32 V, T = 0 °C), and worst case (V_{DD} = 1.08 V, T = 125 °C) conditions; the SG-FET analysis was done only in the typical case.

Table I shows that when using 2D High-V_t CMOS based implementation the Shifter leakage current is reduced by 34x with a cost of 19 mV IR-drop and 13.25% area increase, and negligible total power penalty. By using SG-FET based ST we further reduce by 286x the Shifter leakage, with a further addition in area cost of 34.43%, for roughly the same IR-drop. Furthermore, by placing the SG-FET based STs on a separate die and increase its area according to the Shifter footprint, the IR-drop is halved with a minimum effect on leakage.

V. CONCLUSION

In this paper we introduced a novel Suspended Gate FET (SG-FET) based power management architecture for 3D Through Silicon Vias based integration technology. Our proposal demonstrates that 3D integration is an effective way to make use of the SG-FET characteristics, i.e., abrupt switching and extreme low leakage, in effective NEMS-based power/energy/thermal management.

REFERENCES

[1] "International technology roadmap for semiconductors," ITRS, Design, 2007. http://www.itrs.net/

[2] N. Weste, *CMOS VLSI design : a circuits and systems perspective*, 4th ed. Boston: Addison Wesley, 2011.

[3] H. Jiang, M. Marek-Sadowska, and S. R. Nassif, "Benefits and costs of power-gating technique," in *ICCD 2005, IEEE International Conference on Computer Design: VLSI in Computers and Processors*, 2005, pp. 559–566.

[4] H. S. Won *et al.*, "An MTCMOS design methodology and its application to mobile computing," in *Low Power Electronics and Design, 2003. ISLPED'03. Proceedings of the 2003 International Symposium on*, 2003, pp. 110–115.

[5] G. Panic, Z. Stamenkovic, and R. Kraemer, "Power gating in wireless sensor networks," in *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, 2008, pp. 499–503.

[6] *Low Power Methodology Manual*. Boston, MA: Springer US, 2007. http://www.springerlink.com/index/10.1007/978-0-387-71819-4

[7] H. F. Dadgour and K. Banerjee, "Design and analysis of hybrid NEMS-CMOS circuits for ultra low-power applications," in *Proceedings of the 44th annual Design Automation Conference*, 2007, pp. 306–311.

[8] M. Enachescu, S. Cotofana, A. Genderen, D. Tsamados, and A. Ionescu, "Can SG-FET replace FET in sleep mode circuits?" *Nano-Net*, pp. 99–104, 2009.

[9] N. Abele *et al.*, "Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor," in *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 2006, pp. 479–481.

[10] Y. Xu *et al.*, "On-line MPSoC scheduling considering power gating induced Power/Ground noise," in *2009 IEEE Computer Society Annual Symposium on VLSI*, Tampa, Florida, USA, 2009, pp. 109–114.

[11] A. M. Ionescu *et al.*, "Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture," in *Quality Electronic Design, 2002. Proceedings. International Symposium on*, 2002, pp. 496–501.

[12] W. Davis *et al.*, "Demystifying 3D ICs: the pros and cons of going vertical," *IEEE Design and Test of Computers*, vol. 22, no. 6, pp. 498–510, 2005.

[13] J. A. Davis *et al.*, "Interconnect limit on gigascale integration in the 21st century," *Proceedings of the IEEE*, vol. 89, no. 3, pp. 305–324, 2002.

[14] "TSMC reference flow 7.0," 2010. http://www.tsmc.com

[15] G. Katti *et al.*, "3D stacked ICs using cu TSVs and die to wafer hybrid collective bonding," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2010, p. 1–4.

[16] D. Tsamados *et al.*, "Finite element analysis and analytical simulations of suspended Gate-FET for ultra-low power inverters," *Solid-State Electronics*, vol. 52, no. 9, pp. 1374–1381, 2008.

[17] "OpenSPARC t1 processor," http://www.opensparc.net/opensparc-t1/index.html, 2010. http://www.opensparc.net/opensparc-t1/index.html

[18] M. Anis, S. Areibi, and M. Elmasry, "Design and optimization of multithreshold cmos (mtcmos) circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 10, pp. 1324–1342, 2003.