A Novel Single-Gated Strained CMOS Architecture: COSMOS

AHMAD Al-AHMADI and SAVAS KAYA*,
School of EE&CS, Ohio University, Athens, OH 45701, USA,
E-mail: kaya@ohio.edu,

***Abstract*— We present a simulation study of a novel CMOS device architecture capable of building complementary logic operation using only a single gate stack. The new architecture, named complementary orthogonal stacked MOS (COSMOS), places the n and p-MOSFETs perpendicular to one another under a single gate. As a result of concurrent vertical and lateral integration, the COSMOS architecture can lead to dramatic savings in active device area of a conventional static CMOS pair, as well as significant reductions in RC device parasitics. We demonstrate how the COSMOS devices may be built, operated and optimized for symmetric operation, also verifying logic NOT operation via 3D device simulations. COSMOS architecture appears to have peculiar scaling trends such as increasing threshold at reduced gate dimensions. The increase in drive voltages lead to faster operation at the expense of higher static leakage and loss of noise margins.**

I. Introduction

The Si nanoelectronic engineering have recently reached a level of capability, which make 3D processing on silicon-on-insulator (SOI) substrates not only possible [1], [2], but also a necessity in order to surmount practical limitations of conventional planar CMOS [3]. Thus, device engineers are presented with a multitude of options in exploring new designs, as evident in the proliferation of alternative architectures, including multi-gate MOSFETs, Schottky MOSFET and Tunneling MOSFET. While these structures have unique features superior to conventional MOSFETs, nonetheless, they merely aim to replace the bulk devices in traditional CMOS circuitry. In other words, they do not offer significant paradigm shifts in design and layout, thus still retaining the redundancy inherent to CMOS operation, namely building two devices even though only one operates at a given stable output.

We have recently proposed [4] that a symmetrically operating CMOS device pair may be built under a single gate structure by a simple choice of device layout and channel engineering parameters. The new architecture, named complementary orthogonally stacked MOS (COSMOS), places the n- and p-MOSFETs perpendicular to one another under a single gate, integrating them vertically as well as horizontally (see Fig.1). Thus COSMOS can eliminate the intrinsic redundancy in CMOS, leading to dramatic savings (>50%) in device active area of conventional CMOS static-logic layout, and reduction in RC device parasitics associated with building and wiring two sets of devices for a single Boolean output function. We demonstrated, through the use of device simulations, that the COSMOS gates can operate with reasonable performance in the recent feasibility study. However, only a single COSMOS gate length was considered in this earlier work [4], with no information on the scaling capability of this exciting structure. In the present study, we evaluate the scaling potential of COSMOS architecture and develop guidelines as to how to optimize this novel architecture. We point out that the compact nature of COSMOS layout, closely coupling device geometry in vertical and horizontal directions, results in scaling rules different from those typically found in tradition CMOS. We also discuss layout considerations important for the large scale implementations of COSMOS

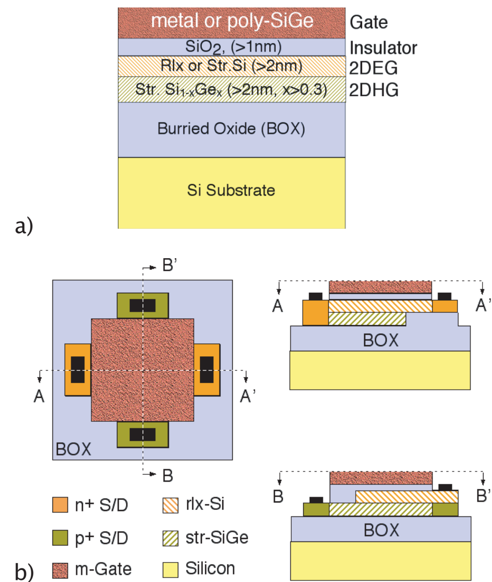

Fig. 1. (a) Generic layer structure,(b) top and cross-sectional view of the device geometry, needed for successful integration and symmetric operation of the proposed COSMOS devices.

II. COSMOS Architecture

In the proposed COSMOS architecture, the two channels are integrated under the same gate with the two active areas rotated by 90° to allow access to two independent sets of source and drain contacts. Thus there is only one active field and a single gate stack in Fig.1, thereby dramatically reducing both the total device (>50%) area and wiring parasitics. An important aspect of COSMOS is the creation and spatial isolation of two channels by the use of strained Si/SiGe heterostructure technology, which allow electrons and holes to be confined to separate layers. Both channels must be undoped and extremely

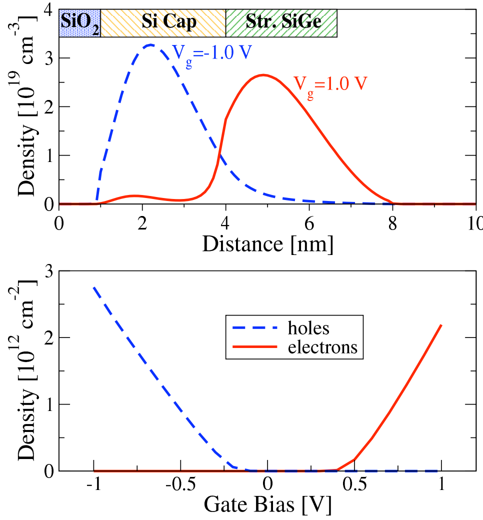

Fig. 2. Self-consistent simulation of 1D electron and hole density profiles (top) in a sample COSMOS layer design with 3nm Si cap and 4nm $Si_{0.7}Ge_{0.3}$ channel, and a mid-gap metal gate (ϕ_m=4.67eV. Carrier density against gate voltage obtained using similar simulations are shown at the bottom

thin (2 to 8nm) to facilitate threshold control and minimize parasitic conduction. Furthermore, the gate stack must be engineered to concurrently tune the two thresholds by the choice of a single barrier height and oxide thickness. This aspect is especially exciting since a move to metal-gate CMOS has been delayed mainly by the complexity in the integration of two separate gate metals in a single process.

The COSMOS devices are based on strained silicon-on-insulator (SSOI) substrate technology, which has recently become available due to advances in strain layer growth and wafer bonding techniques [5], [6]. There are already several examples of strained Si/SiGe heterostructures on SOI wafers with extremely thin dual channels having symmetric electron and hole mobilities[7], [8]. These reports, and general evolution of SOI technology, imply that COSMOS layer structure can be built using existing material and technological toolsets. The only non-standard feature is the inclusion of two etch-steps to prevent the parasitic p-i-n diode conduction by the partial removal of complementary channels near the drain end of each device (e.g. partial removal of Si cap in the pMOS device).

III. COSMOS Operation

To explain COSMOS operation, we show in Fig.2 the electron and hole profiles at two extreme cases of the gate voltage (± 1V) in an example structure with 3nm Si electron channel and 4nm strained SiGe hole channel capped with a 1nm SiO2 insulator and mid-gap metal gate. The two distributions are comparable in terms of amplitude, and are largely localized to their respective channels, eliminating parallel conduction possibility. Similar Poisson-Schrodinger solutions show (lower plot in Fig.2) that symmetric operation with a low (0-0.5V) threshold voltage is possible in COSMOS structures. The actual threshold voltage of the device requires a 3D simulation study as will be discussed in the next session.

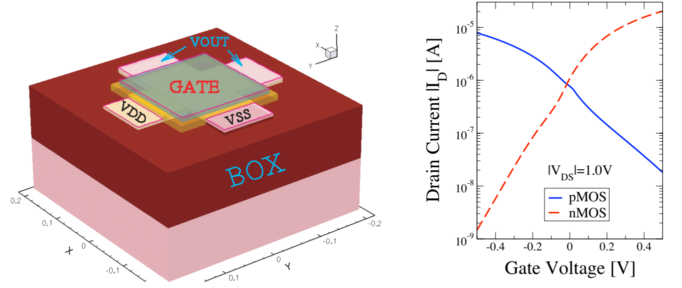

Fig. 3. (a) 3D view of a W=L=200 nm COSMOS gate and (b) 3D Simulation of (Id-Vg) characteristics of the 36 nm COSMOS devices. Note that in 3D simulations quantum mechanical corrections are omitted to save time. The non-uniform sub-threshold slope is due to p-i-n parasitic diode inherent to COSMOS operation. tSi=2nm and tSiGe=3nm,30 % Ge, mid-metal gate, and 1nm SiO_2.

To illustrate how a single gate may be used to control two MOSFETs, we show in Fig.3 the I-V characteristics of COSMOS transistors obtained from 3D drift-diffusion simulations. To save time in demanding 3D simulations, no quantum mechanical corrections were used. The IV characteristics of the two MOSFETs are almost comparable, with the small difference in ON state current resulting from the fact that the strained SiGe mobility parameters are not accurate in the present model of the TCAD simulator. The non-uniform sub-threshold slope is as a result of p-i-n leakage contribution to actual transistor current. It is clear that, by tailoring the SiGe hole mobility and choosing an appropriate vertical design, a symmetrically operation COSMOS gate is possible.

A COSMOS inverter can be easily realized by the appropriate connections to the low (V_{SS}) and high (V_{DD}) rail voltages, and by connecting the two drains with a very short metallization, as shown in Fig.3. This illustrates the immense potential of the present architecture for lowering RC parasitics, especially in the context of digital static CMOS applications. In the following transient drift-diffusion simulations, the quantum mechanical corrections are omitted to speed up the 3D solutions, the load capacitor is assumed to be 1fF and a 36nm COSMOS device is used. As evident from Fig.4, the logic inverter is fully functional at (± 0.5V) supply voltage with acceptable delay($\sim$ $100ps$). We also observe a small amount of static leakage as a result of higher sub-threshold slope of pMOSFET, augmented by leakage from p-i-n parasitic device. While the structure tested here is not fully optimized, these results are sufficient to verify the potential of COSMOS devices in logic circuits.

IV. COSMOS Scaling and Optimization

A. Vertical Scaling

Symmetrical threshold control in COSMOS can be obtained by the choice of a number of different vertical design parameters. To elucidate this point further, we plot in Fig.5, the dependence of threshold in COSMOS layers on varying levels of Ge content and thickness in the strained SiGe layer. Fig.5 indicates that accurate optimization of the density in individual channels is possible [9], providing ample room for designing COSMOS devices capable of symmetrical operation. Moreover, safety margin against parasitic hole channel formation

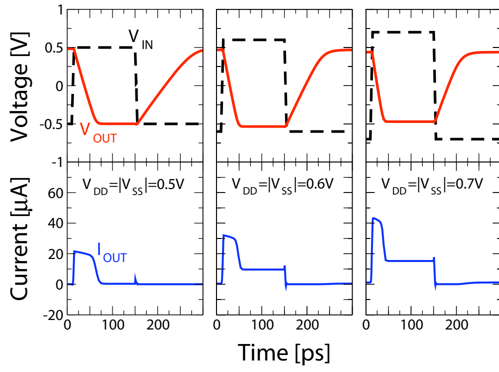

Fig. 4. The transient response of a 40 nm COSMOS inverter gate obtained from 3D drift-diffusion simulations at different drive voltage (logic) levels. The logic NOT function is succesfully implemented in this particular layer structure (tSi=2nm and tSiGe=3nm, 30% Ge, mid-metal gate, and 1nm SiO2)..

in the Si cap is sufficient for low rail-voltages found in sub-100 nm circuits. The overlap of carrier distributions in the two channels along the main device axes is not problematic, since only one channel exists at a given time. This would be a problem only for conduction via the parasitic p-i-n diode, which is inhibited by the under and over-etch regions of the channels as explained above. Since the reduction of channel thickness below 5nm significantly lowers mobility due to excessive interface scattering, vertical spread of carriers along the major transport axes is beneficial in this case for both types of devices. Besides the choice of Ge content and channel thickness, gate work-function and gate insulator thickness are two additional design parameters [4] which can be used to fine tune the threshold in COSMOS structure. Thus there are various avenues of optimization in COSMOS to adjust thresholds and mobility.

B. Lateral Scaling

To investigate COSMOS lateral scaling, we show in Fig.6a the simulated thresholds of COSMOS pairs at various gate dimensions. We find a completely reversed gate scaling trend in COSMOS devices, with smaller gates having larger thresholds and lower ON currents. This seemingly counter-intuitive scaling behavior results from the cross-shaped active area of the transistor. Cross-shape implies reciprocal behavior, i.e.$(W/L)_{nMOS}$= $(L/W)_{pMOS}$, where L and W refers to gate length and width along the major transport axis of each transistor. Hence, the reduction of gate length in one device leads to a narrower channel in the other, resulting in the opposite trend in device currents. While a square geometry is generally desirable, a small departure from this may be conveniently utilized in IC design to fine-tune small differences in the currents of the two MOSFETs. The lowering of ON current with gate scaling means that the devices with smaller gate dimensions get progressively slower as can be seen from Fig.6b. At the same time, the static leakage is reduced at lower geth lengths. Thus there is an optimum gate dimension that trades-off active device area with speed and static leakage for a given layer structure.

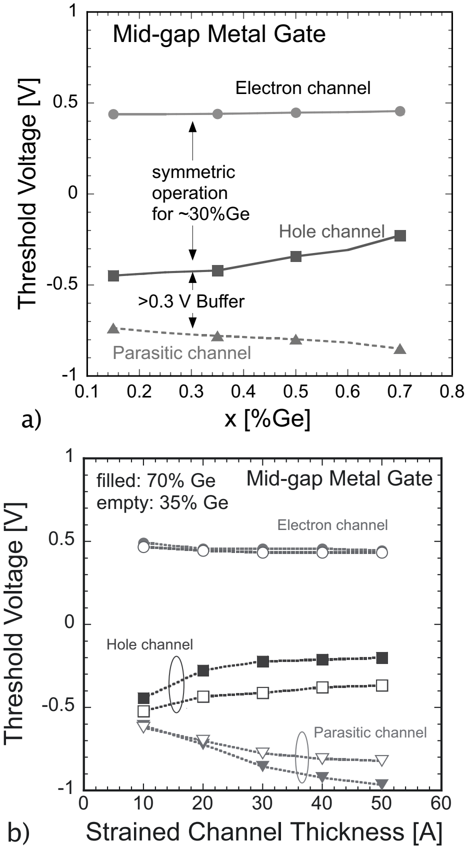

Fig. 5. Dependence of simulated threshold voltage in the COSMOS layer structure on a) Ge content and b) strained layer thickness. All parameters are same as Fig.3 unless otherwise indicated.

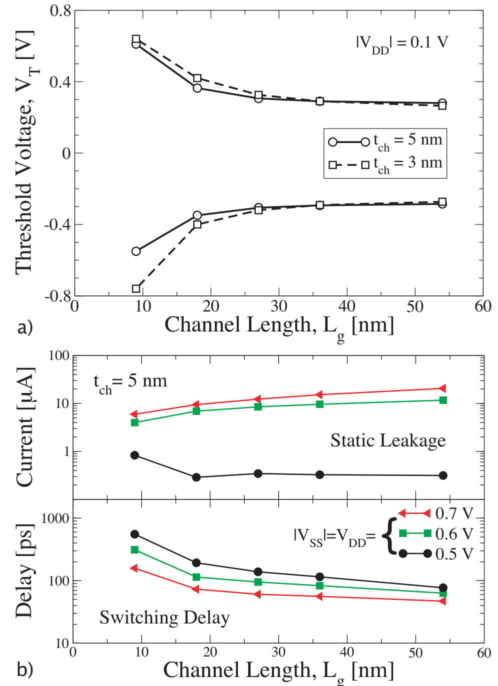

Fig. 6. Dependence of a) threshold voltage of COSMOS device pairs (total tch=5 nm) and b) their average delay and static leakage on effective channel length. Note that the scaling characteristics of the device is opposite to conventional MOSFETs due to reciprocal coupling of device dimensions, i.e. $(W/L)_{nMOS}$= $(L/W)_{pMOS}$.

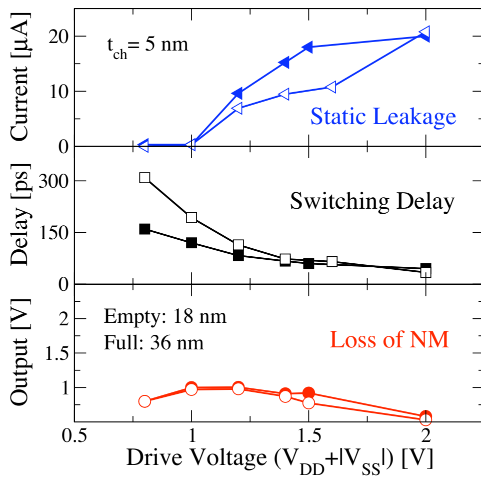

Fig. 7. Dependence of simulated (3D) performance of the COSMOS inverters on drive voltages. For rail voltage difference more than 1V, the leakage and noise margins quickly deteriorates. All device parameters are same as Fig.3.

C. Voltage Scaling

ITRS envisages rail voltages less than 1.0 V for sub-50nm era, progressively getting as low as 0.5V [9]. We have already seen in Fig.6b that heavy-bias conditions result in increase in both speed and leakage. Fig.7 shows in detail how COSMOS logic inverters respond to use of different rail voltages at two different gate lengths (18 and 36nm) and 5nm total channel (2nm Si cap) thickness. It appears that for rail voltages larger than 0.5V, the static leakage and noise margin performance quickly deteriorates, while the gate delay does not improve substantially. As expected from above discussion, the 18nm gate performs worse than longer (36nm) counterpart in terms of delay, even though excess leakage is reduced. We also see that at sufficiently high rail voltages (±1.0V) 18nm COSMOS gives way to additional leakage current associated with drain-induced barrier lowering (DIBL) effect. It is evident that typical rail voltages for a given COSMOS device geometry will have an upper limit that appears to be (±0.5V) for devices considered in this work. Therefore COSMOS is most suitable for low-power applications, which places greater importance in active area saving than speed.

V. Logic Circuits and Layout

The efficient use of Si area and layout in COSMOS circuits become more obvious as the complexity of logic circuits increases. For example, in Fig.8, we provide a suggested layout for a two-input static-CMOS NOR circuit constructed using two COSMOS gates. The serially-connected p-MOSFETs have common drain/source, which can be seperated in an ASIC implementation. In this case, the final metalization may be utilized to hard-wire dense collection of COSMOS gates to any logic function desired at the expense of higher interconnect parasitics. Moreover, the orthogonal nature of COSMOS mean that, between a NOR and NAND implementation, the same layout may be used if appropriate routing is chosen. However, in such a dense circuit layout, the routing is likely to become even a bigger problem. We are currently considering

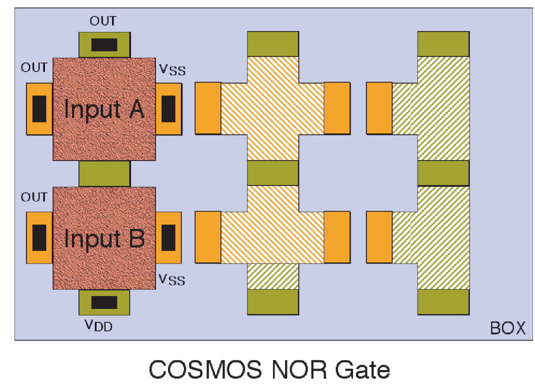

Fig. 8. An example layout for a two-input NOR geometry using only two COSMOS gates. Peel-off diagrams, in the center and on the right, allow to see the Si/SiGe layer structure under the input gates

alternative approaches in gate geometry of COSMOS that can alliviate routing concerns as well as reducing leakage. One such option is to use assymetrical cross geometry, with the center of crossed moved toward the drains. This would allow a larger area for access to gate from via cuts, and also increase the leakage path for the aforementioned parasitic p-i-n diode. We shall evaluate various layout options and their impat on device performance in a later publication.

VI. Conclusions

We presented the principles of operation of a novel CMOS architecture, termed COSMOS, which is capable of static logic operation using only a single gate and active area. In the proposed COSMOS architecture, the MOSFET pair are orthogonally integrated under a single gate by a careful engineering of dual electron-hole channels in strained Si/SiGe on insulator substrates. The design guidelines for symmetrical operation of COSMOS devices are explained using 1D and 3D simulations. The scaling of COSMOS is unique and complicated by the reciprocal dependence of gate dimensions on n and p MOSFETs, with lower ON current at smaller gate lengths. Using efficient routing techniques, very-dense logic circuits may be implemented in the proposed architecture. COSMOS appears to be especially suitable for low-power applications, as it is limited by the static leakage from parasitic p-i-n diode at high supply voltages.

References

[1] R Chau *et al.*,, *IEDM Tech Dig.*,45-48 (2000).
[2] P A Packan, *Science.*, **285**, 2079 (1999).
[3] H-S P Wong, *IBM J Res and Dev.*, **46**, 133-168 (2002).
[4] S. Kaya, *IEEE Trans, Nanoelectr submitted, also presented at Silicon Nanoelectronic Workshop.*, Honolulu, HI 7-8, 13-14(2004).
[5] T Mizuno *et al., IEEE Trans. Electr. Dev.*, 988-994(2003).
[6] K Rim *et al., IEDM Tech Dig.*, 49 (2003).
[7] M L Lee and E A Fitzgerald, *Appl. Phys. Lett.*, **83**, 4202-4204 (2003).
[8] Z Cheng *et al., Semic. Sci. and Tech.*, L48-L51 (2004).
[9] Semiconductor Industry Association (SIA), "International Technology Roadmap for Semiconductors" *http://public.itrs.net/Reports.htm.*,2003