Design Considerations and Quantum Confinement effect in Monolithic ε-Ge/$In_xGa_{1-x}As$ Nanoscale FinFETs Down to N5 Node

Rutwik Joshi, Sengunthar Karthikeyan and Mantu K. Hudait, *Senior Member, IEEE*

Abstract— In this work, we have studied the effect of material parameters (indium (In) composition and doping), geometrical parameters (channel length L, fin width W, aspect ratio AR), and quantum confinement (QC) on the performance and operability of a ε-Ge/$In_xGa_{1-x}As$ hybrid CMOS system. In this system, the In compositional $In_xGa_{1-x}As$ and tensile strained Ge (ε-Ge) grown on the $In_xGa_{1-x}As$ layer, were used as n- and p-channel FinFETs, respectively. The In composition in $In_xGa_{1-x}As$ layer (lattice matched with graded $In_xAl_{1-x}As$ buffer) determines the amount of tensile strain in Ge. This hybrid system utilizes the benefits of metamorphic ($In_xGa_{1-x}As$/$In_xAl_{1-x}As$) as well as pseudomorphic (ε-Ge/$In_xGa_{1-x}As$) heteroepitaxy to create high performance tunable complementary devices, suitable for 0.5 V CMOS operation. The device metrics such as, threshold voltage, on-current (I_{ON}), off-current (I_{OFF}), subthreshold-swing (SS), and drain induced barrier lowering (DIBL), and their dependence on material and geometrical parameters were evaluated using self-consistent analytical solvers scaled down to the N5 node. At these scaled dimensions, this hybrid system demonstrated ultra-low leakage current and SS for the n-FinFET and p-FinFET of 10 pA/µm, 27 nA/µm, 85 mV/dec and 95 mV/dec, respectively. With the effect of QC, we identify a transition fin width (W_T) associated with scaling of alternate channel FinFETs, at which the performance is optimum and below W_T, the benefits of scaling are diminished. Moreover, this hybrid system has a potential to find applications in optoelectronic and RF systems as well as high-performance computing.

***Index Terms*— Tensile strained Germanium, InGaAs, FinFET, Quantum Confinement, low power, CMOS, 5 nm.**

I. Introduction

THE progression of Moore's law over the past decades has been possible through iterative optimization of Si MOSFETs through transistor scaling and implementation of new device architectures. Benefits of aggressive scaling are outweighed by the issues such as, leakage associated with quantum-mechanical tunneling through the gate oxide, and short channel effects (SCE) such as degradation of drain induced barrier lowering (DIBL) and subthreshold swing (SS). Beyond this, further scaling has been possible using the 3D FinFET architecture. Along with the reduction of device footprint, this architecture provides an excellent electrostatic control over the fully-depleted channel. As further scaling approaches atomic dimensions, one must carefully look for high-mobility, low-bandgap material-based solutions along

Manuscript received June 15, 2022. The review of this paper was arranged by Editor xxxxx.
Rutwik Joshi, Sengunthar Karthikeyan and Mantu K. Hudait are with the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA (mantu@vt.edu).
Digital Object Identifier.

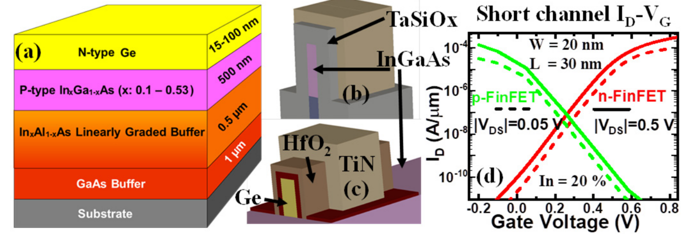


Fig. 1. (a) The co-integrated stack and, schematic representation of proposed hybrid CMOS using monolithically co-integrated (b) $In_xGa_{1-x}As$ n-FinFET and (c) ε-Ge p-FinFET. (d) I_D-V_G characteristics demonstrating matched response.

with high-κ gate dielectrics [1]-[3] to provide low-power and high-speed operation. If feasible, one should avoid the layer transfer and bonding for cost-performance benefits. The advantages of utilizing a hybrid system for the creation of high-performance complementary devices, which can be experimentally realizable with minimal alteration of device design, should be explored. One such hybrid system is the metamorphic $In_xGa_{1-x}As$/$In_xAl_{1-x}As$ for n-channel FinFET and pseudomorphic tensile strained ε-Ge/$In_xGa_{1-x}As$ for p-channel FinFET, suitable for low-power operation. In addition, this hybrid system has potential applications in CMOS circuits [4-6], optoelectronics [7], and RF systems [8]. Fig. 1(a) shows the starting epitaxial stack, which uses $In_xAl_{1-x}As$/GaAs metamorphic buffer followed by Ge/$In_xGa_{1-x}As$ pseudomorphic growth. The ε-Ge p-FinFET and the $In_xGa_{1-x}As$ n-FinFET can be fabricated side-by-side using this epitaxial stack to form high performance CMOS circuits. Recently, we have evaluated the merits of this ε-Ge/$In_xGa_{1-x}As$ CMOS architecture, utilizing ε-Ge for p-FinFET and $In_xGa_{1-x}As$ for n-FinFET, in the vicinity of experimentally proven feature sizes. This system offered an excellent device and circuit performance as well as application specific tunability [9]. A rigorous study of device metrics such as, V_{TH}, I_{ON}, I_{OFF}, SS, and DIBL and their dependence on material parameters (*i.e.*, indium composition, doping) and geometrical parameters (*i.e.*, channel length L, fin width W, and fin aspect ratio AR) using self-consistent analytical solvers, which capture realistic physics of nanoscale devices, is required to characterize architectures scaled down to the N5 node [10].

To ascertain the performance benefits of this hybrid CMOS system for low-power applications at ultra-scaled feature sizes, here, we investigate the CMOS operability constraints and merits of monolithically co-integrated $In_xGa_{1-x}As$ as n-FinFET (Fig 1b) and ε-Ge as p-FinFET (Fig 1c), which can be fabricated using the heteroepitaxial stack shown in Fig. 1(a). The cumulative effects of In composition and doping (N_{CH}) as well as geometrical parameters on the device performance were evaluated at forward looking feature sizes down to the N5 node. The results were displayed through contours as a means to effectively visualize the relative trends of the metrics for a large

number of parameter combinations, while also providing an estimate of the variability due to changes in process, material and design. The short channel I_D-V_G characteristics with channel length of 30 nm and an In composition of 20 % is shown in Fig. 1(d), demonstrating the performance of the ε-Ge as p-FinFET (*green*) and $In_xGa_{1-x}As$ as n-FinFET (*red*) at V_{DS} = 0.5 V (*solid lines*) and V_{DS} = 0.05V (*dashed lines*). Furthermore, in nanoscale multi-gate devices, the quantum confinement (QC) of inversion charge results in an additional capacitance in series with the oxide capacitance due to the charge centroid. This effect manifests in FETs as an increase in V_{TH} creating a challenge for device operability and scaling. The increase in V_{TH} results in a drop in I_{OFF} and a corresponding increase in I_{ON}/I_{OFF} ratio, while reducing SS and DIBL due to improved electrostatic control owing to QC. This performance consideration is valid only down to a specific width, beyond which the device transitions back to charge dominated regime. Here, we quantify this transition width (W_T) in the context of CMOS operability to be < 6 nm for $In_xGa_{1-x}As$ n-FinFET and ~ 20 nm for ε-Ge p-FinFET and illustrate the corresponding transition in design considerations for ultra-scaled CMOS devices.

II. Growth, Simulation Methodology and Device Parameters

The device simulations were performed using Synopsys Sentaurus TCAD [11] considering a 3D numerical solver while accounting for the classical transport and QC effects within the fin structure. The ε-Ge p-FinFET model is calibrated with the experimental results [12], and the n-FinFET is calibrated using a lattice matched $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As$ heterostructure, where $TaSiO_x$ is used as a high-κ dielectric [13]. Further details of the model calibration with experimental data for the $In_xGa_{1-x}As$ n-FinFET and the ε-Ge p-FinFET, as well as the strain calculation in the ε-Ge fin are discussed elsewhere [9].

A transistor is said to be operable for CMOS applications when the charges in the channel can be modulated considerably within the operating voltage range (here, $V_{GS} = V_{DS} = 0.5$ V). This implies that the transistor is normally-off and the V_{TH} lies within the supply voltage range, and the I_{ON}/I_{OFF} ratio is > 1000 [14]. The device performance can be characterized by SS and DIBL, which indicate the leakage and the V_{TH} change due to the effect of drain bias, respectively. The SS and DIBL should be as small as possible, where the SS can be as low as ~ 60 mV/dec at 300 K and DIBL can approach ~ 0 mV/V. This hybrid material system can be grown on Si or GaAs substrate

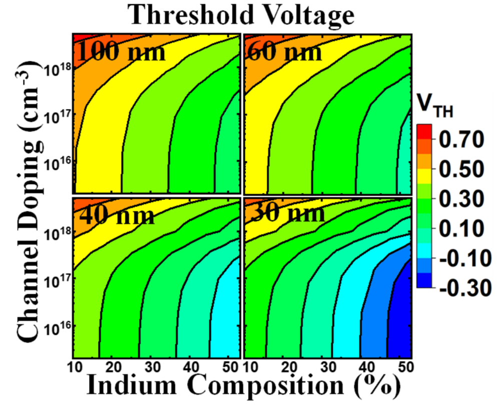


Fig. 2. Effect of N_{CH} and In composition on V_{TH} for $In_xGa_{1-x}As$ n-FinFET at channel lengths of 100 nm, 60 nm, 40 nm, and 30 nm.

using a graded III-V buffer. The heteroepitaxial growth of ε-Ge on $In_xGa_{1-x}As/In_xAl_{1-x}As$ using a solid source molecular beam epitaxy can mitigate the lattice mismatch induced defects and dislocations, thus providing a virtually defect-free active layer [6]. The top-most layer of the $In_xGa_{1-x}As$ on graded $In_xAl_{1-x}As$ buffer acts as the channel for the n-FinFET and has no thickness constraints, thus providing a large flexibility in the fin height H and AR for the n-FinFET. The epitaxial Ge layer grown on $In_xGa_{1-x}As$ active layer will experience a tensile strain corresponding to In composition. This ε-Ge layer can be grown defect-free only up to the critical thickness, limiting the fin height and aspect ratio of the ε-Ge p-FinFET, and in general the critical thickness reduces with increase in tensile strain [15].

III. Material Properties: Alloying & Doping

The choice of In composition and N_{CH} directly affects the material properties, and thus application specific selection of these parameters can lead to a significant impact on the designed CMOS circuit/system. In addition, channel length scaling introduces SCE (*i.e.*, V_{TH} roll-off, increase in SS and DIBL), which further affect the optimum region of operation in terms of In composition and N_{CH}. The cumulative effect of N_{CH} and In composition on the device performance and operability is magnified at shorter channels and is shown in Fig. 2-3 ($In_xGa_{1-x}As$ n-FinFET) and Fig. 4-5 (ε-Ge p-FinFET) for a fixed AR (H/W) = 1. The N_{CH} in the range of 1×10^{15} cm^{-3} to 1×10^{19} cm^{-3} for the $In_xGa_{1-x}As$ n-FinFET is selected considering two important restrictions on the doping in epitaxially grown $In_xGa_{1-x}As$: (i) higher N_{CH} (> 1×10^{17} cm^{-3}) is experimentally easier to incorporate [16] and provides increased immunity to SCE and an improved I_{ON}/I_{OFF} ratio, but will also result in

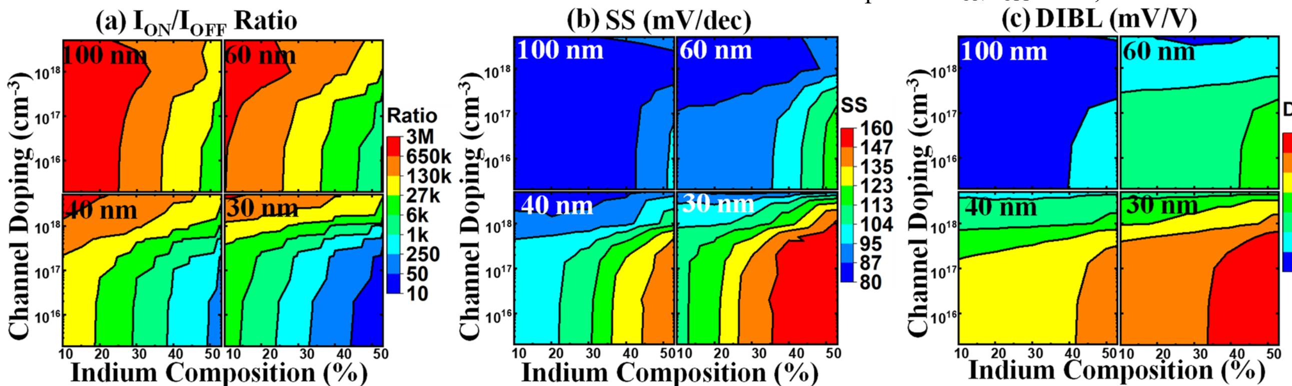


Fig. 3. Effect of N_{CH} and In composition on the (a) I_{ON}/I_{OFF} ratio, (b) SS and (c) DIBL for $In_xGa_{1-x}As$ n-FinFET at L = 100 nm, 60 nm, 40 nm, 30 nm.

increased impurity scattering induced mobility degradation along with dopant variability effects; (ii) lower N_{CH} (~ $1x10^{15}$ cm^{-3}) results in an operation tending to normally-on behavior as seen through the negative V_{TH} in Fig. 2.

A wide range of In compositions have been reported in $In_xGa_{1-x}As$ layers grown heteroepitaxially on III-V substrates [6][17]. Epitaxial growth of ε-Ge on $In_xGa_{1-x}As$ having In composition greater than 53 % will result in > 3.3 % tensile strain in Ge, severely degrading the critical thickness of ε-Ge thus limiting the fin height of the p-FinFET [15]. At such high tensile strain, the Ge bandgap reduces to 0.1 eV, making the ε-Ge layer unsuitable for use as a p-channel FET. Furthermore, ε-Ge/$In_xGa_{1-x}As$ band-alignment transitions from type-I to type-II or III for higher In compositions, resulting in a broken gap alignment, leading to leakage [6]. Tensile strain enhances the Ge p-FinFET performance in the form of bandgap lowering (and consequent reduction in V_{TH} of a p-channel transistor) and increasing hole mobility. The In composition in the $In_xGa_{1-x}As$ layer underlying the Ge layer (*see* Fig. 1a) determines the amount of tensile strain in Ge, and a lower In composition results in a lower lattice mismatch between Ge and $In_xGa_{1-x}As$, resulting in a lower tensile strain. At In compositions below 10 % (the tensile strain is ~ 0.3 %), the enhancement of the material properties of Ge is insignificant. At such low tensile strain amount, no appreciable reduction in V_{TH} can be observed due to negligible change in the bandgap. Tensile strain also improves hole mobility in Ge due to the upward movement of the light-hole (LH) band. As the In composition increases, the LH band moves upward and starts to dominate over the heavy-hole (HH) band. Due to the higher density of states (DOS) of the HH band, to see a significant improvement in net hole mobility gain, the LH band has to move up considerably which is possible at higher In compositions. In addition, the choice of In composition directly affects the performance of the $In_xGa_{1-x}As$ n-FinFET and at very low In compositions, the bandgap of $In_xGa_{1-x}As$ is very large (~ 1.4 eV at x ~ 1%), which results in a high V_{TH} for the n-FinFET, resulting in inoperability at 0.5 V CMOS. Furthermore, the electron mobility of $In_xGa_{1-x}As$ increases with In compositions, and an appreciable increase can be observed in the In composition range of $0.10 < x < 0.55$. Considering these factors, $0.10 \leq x \leq 0.53$ is chosen as the optimum range for this study to ensure CMOS operability at 0.5 V supply as well as experimental realizability of such hybrid system. To ensure operability (normally-off and I_{ON}/I_{OFF} ratio > 1000), the low bandgap ε-Ge p-FinFET is more sensitive to N_{CH} variation and requires a higher doping, and thus the range $2x10^{18}$ cm^{-3} to $1x10^{19}$ cm^{-3} [17] was selected for this study, as discussed below.

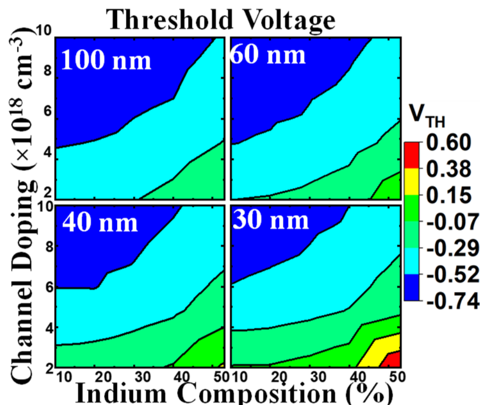


Fig. 4. Effect of In composition and N_{CH} on the V_{TH} of ε-Ge p-FinFET at channel lengths of 100 nm, 60 nm, 40 nm, and 30 nm.

A. Effect of alloying and doping on $In_xGa_{1-x}As$ n-FinFET

The In composition affects the $In_xGa_{1-x}As$ channel resulting in spectrum of tunable material properties to utilize for the n-FinFET. A higher In composition results in a lower bandgap in $In_xGa_{1-x}As$ as well as higher electron mobility. The lower bandgap results in a reduced V_{TH} while the higher electron mobility enhances the I_{ON}; whereas, a higher N_{CH} increases the V_{TH}. Thus, the In composition and N_{CH} act as competing mechanisms affecting the device performance in opposite ways. Fig. 2 shows the variation of V_{TH} with In composition and N_{CH} for L = 100 nm, 60 nm, 40 nm, and 30 nm. For operation of the n-FinFET at the desired supply voltage of 0.5 V, it is essential that the V_{TH} be well within this range, thus the region shown in *red* (low In, *high* N_{CH}) and in *blue* (*high* In, *low* N_{CH}) render the device inoperable due to high V_{TH} and low V_{TH}, respectively. The $In_xGa_{1-x}As$ n-FinFET shows an excellent I_{ON}/I_{OFF} ratio for a wide range of In compositions and N_{CH}, as shown in Fig. 3(a), indicating suitability for ultra-low power applications. At high In of 53 %, the bandgap of $In_xGa_{1-x}As$ decreases significantly, resulting in much higher I_{OFF}. Due to the increase in electron mobility with increasing In composition in $In_xGa_{1-x}As$, the I_{ON} increases, but the increase in I_{OFF} is much more severe, resulting in an overall degradation in I_{ON}/I_{OFF} ratio. There exists a trade-off between the leakage and the drive strength for the $In_xGa_{1-x}As$ n-FinFET. For In = 10 % with L = 40 nm, I_{OFF} is ~ 8 pA/µm and becomes ~ 20 nA/µm at In = 40 %, indicating a 2500× increase in I_{OFF}. On the other hand, the I_{ON} is ~2 µA/µm at In = 10 % and becomes 135 µA/µm at In =40 %, indicating a 67.5× increase in I_{ON}. Although the device can now drive larger loads faster due to high I_{ON}, the I_{OFF} has increased drastically resulting in a decrease in I_{ON}/I_{OFF} ratio as seen in Fig. 3(a).

Approaching shorter gate lengths is expected to exacerbate SCE as the L is varied from 100 nm to 30 nm. Fig. 3(b) and (c) show the SS and DIBL as a function of In and N_{CH}, respectively. The long channel devices remain immune to SCE and this can be seen by the *blue-contours* in Fig. 3(b) and (c) indicating low values of SS and DIBL. Furthermore, it is seen that the performance of short channel devices is more sensitive to variations in In composition and doping compared to long channel devices. With increasing In composition, the I_{OFF} increases significantly, resulting in higher leakage and a hence a degraded SS as seen in Fig. 3(b). The shift in the contours toward the *top-left corner* in Fig. 3, indicates that a higher N_{CH} is required to mitigate the SCE, thus improving DIBL and I_{ON}/I_{OFF} ratio at a channel length of 30 nm. Additionally, in device configurations having high In and low N_{CH} (see, *bottom-right corners* in Fig. 3), the SCE dominate the device performance, as seen through the degradation of SS and DIBL. Further length scaling is possible through reduction of W and corresponding improvement in electrostatic control. Based on the results shown in Fig. 2 and 3, the optimum N_{CH} of $5x10^{17}$ cm^{-3} is used for further analysis. At this N_{CH}, the $In_xGa_{1-x}As$ n-FinFET shows excellent CMOS operation and performance at 0.5 V.

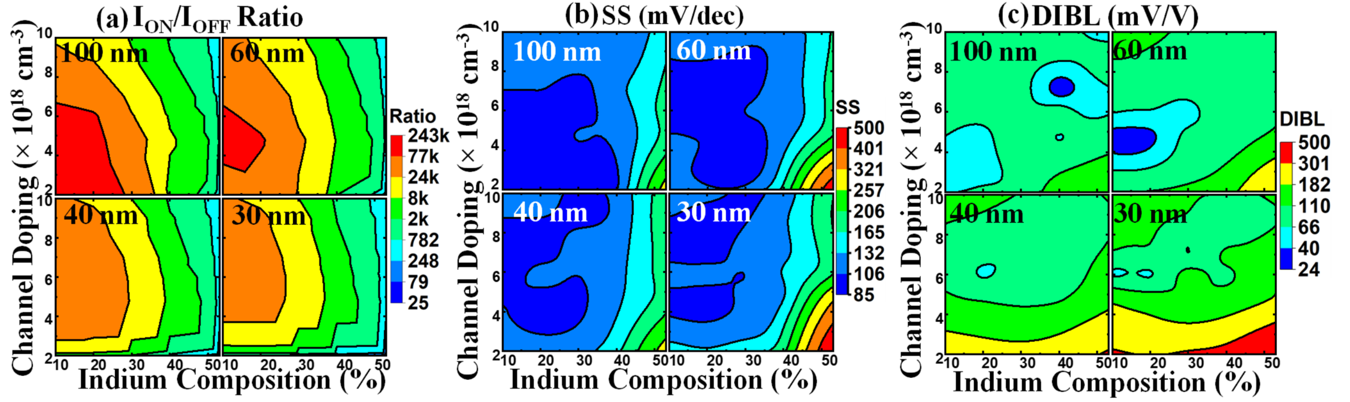


Fig. 5. Effect of In composition and N_{CH} on the (a) I_{ON}/I_{OFF} ratio, (b) SS and (c) DIBL of ε-Ge p-FinFET at L = 100 nm, 60 nm, 40 nm and 30 nm.

B. *Effect of alloying and doping on ε-Ge p-FinFET*

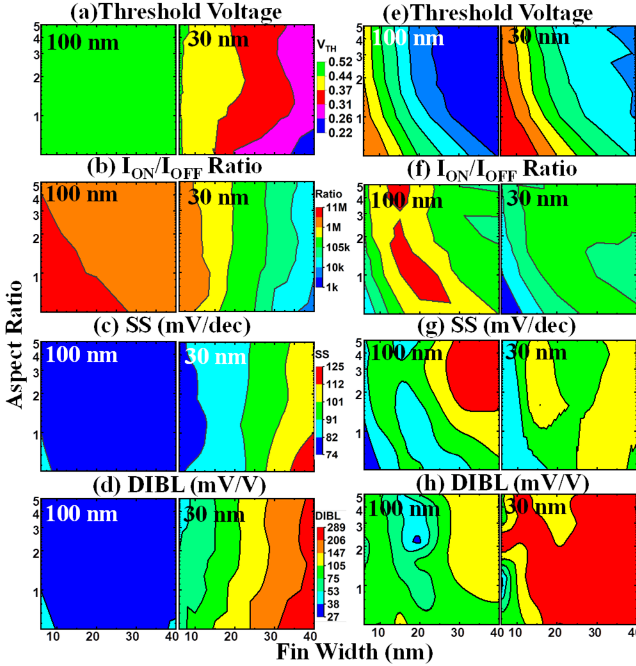
The In composition in $In_xGa_{1-x}As$ indirectly creates a spectrum of properties through the tensile strain in Ge (bandgap and hole mobility). The hole mobility increases and the bandgap decreases with increase in tensile strain in Ge [9], resulting in an increase in I_{ON} and I_{OFF}. Also, the absolute leakage in ε-Ge is expected to be higher than $In_xGa_{1-x}As$ as the bandgap is lower in Ge, resulting in a lower I_{ON}/I_{OFF} ratio. Fig. 4 and 5 show the interplay between the competing variables, In composition and N_{CH}, and their effect on device performance. A relatively low N_{CH} has the advantage of reduced impurity scattering but results in low V_{TH}, as shown in Fig. 4 and reduced I_{ON}/I_{OFF} ratio, as shown in Fig. 5(a). With increasing In composition, the bandgap of Ge reduces, resulting in lower I_{ON}/I_{OFF} ratio. Thus, warranting a careful optimization of the Ge p-FinFET channel doping and In composition in order to obtain the desired operability and performance. Devices with channel length of 30 nm and 40 nm require a higher N_{CH} to mitigate the degradation of DIBL, SS, and I_{ON}/I_{OFF} ratio due to length scaling. The SS, DIBL, and I_{ON}/I_{OFF} ratio at lower N_{CH} degrade more rapidly with increasing In composition due to increased SCE, as seen through the SS and DIBL variation in Fig. 5(c) and (d), respectively. The SS indicates the switching Due to the low bandgap nature of Ge, along with a realistic interface trap density, to maintain normally-off operation with a good I_{ON}/I_{OFF} ratio for p-FinFET, high N_{CH} as well as a large work function metal are needed. The TiN gate metal work-function of 4.5 eV was selected here to provide a higher I_{ON}/I_{OFF} ratio and V_{TH} and reduce the requirement of high N_{CH} (exotic gate metals can relax this dependency further). For optimum p-FinFET design, the device performance *i.e.*, $V_{TH} < 0$, I_{ON}/I_{OFF} ratio > 1000 as well as low SS and DIBL should be achieved while keeping N_{CH} as low as possible (*here*, N_{CH} of $6x10^{18}$ cm^{-3} is chosen). Moreover, a high N_{CH} can mask the benefits obtained due to QC such as V_{TH} and I_{ON}/I_{OFF} ratio improvement as discussed below.

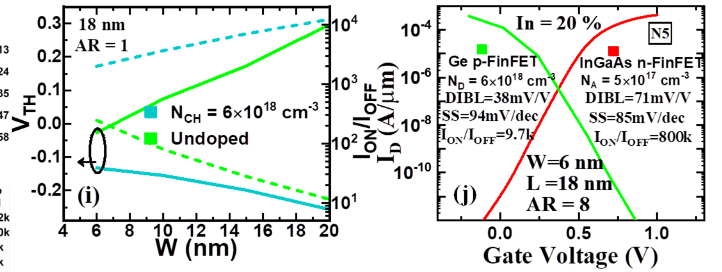
IV. Impact of Fin Scaling: Quantum Confinement

Device geometrical parameters of the n- and p- FinFETs have a considerable effect on their operability as CMOS devices. Here, we have analyzed the effects of channel geometrical parameters for experimentally validated choices of fin length and width [12]-[14],[17],[18] and extended the predictions to scaled dimensions up to 6 nm, which corresponds to the N5 node [19]. From Fig. 2-5 a significant shift in contours can be observed in I_{ON}/I_{OFF} ratio, SS and DIBL while transitioning from L = 60 nm to 40 nm, indicating a transition point (~ 50 nm) to short channel regime. This transition point is in agreement with experimentally demonstrated $In_xGa_{1-x}As$ n-FinFETs with L = 50 nm [18]. The effect of fin width scaling and AR on the performance of long channel (100 nm) and short channel (30 nm) devices is shown in Fig. 6(a-d) for the $In_xGa_{1-x}As$ n-FinFET and Fig. 6(e-h) for the ε-Ge p-FinFET, while keeping the In composition and tensile strain in Ge fixed at 20 % and 1.3 %, respectively. The effect of QC on the performance is clearly different between the two devices, and its dependence on N_{CH} is seen in Fig. 6(i) and (j) for ε-Ge p-FinFET.

A. *Effect of fin width and AR on the $In_xGa_{1-x}As$ n-FinFET*

The V_{TH} of long channel (100 nm) $In_xGa_{1-x}As$ n-FinFET is high, and nearly invariant to fin width and aspect ratio as seen in Fig. 6(a). Note that there exists a small increase in V_{TH} (< 10 mV over 40 nm of width scaling) due to QC of inversion charges, but is not visible in the contour. Fig. 6(b) shows that the *lower-left* corner which corresponds to a narrow fin provides a much higher I_{ON}/I_{OFF} ratio than wider fins in the *upper-right* corner. This ~ 10× improvement in I_{ON}/I_{OFF} ratio is attributed to the small increase in V_{TH}. Consequently, the 100 nm device shows little variation of SS (~ 80 mV/dec) and DIBL (~ 30 mV/V) with fin width and aspect ratio, as shown in Fig. 6(c) and (d), respectively, as these metrics have approached their lowest possible values in this simulation framework. For short channel device (30 nm), the V_{TH} increase is evident for narrow W owing to the QC effect [20], which results in gradual increase in I_{ON}/I_{OFF} ratio (Fig. 6(b)) as well as gradual decrease in SS and DIBL (Fig. (c) and (d)). The benefit of QC can be seen through the gradual improvement of the $In_xGa_{1-x}As$ n-FinFET performance with W scaling. The relative impact of QC on the device performance is amplified in the 30 nm device compared to the 100 nm device owing to increased SCE. Thus, the $In_xGa_{1-x}As$ n-FinFET provides optimum low-power CMOS performance for the 100 nm and 30 nm channels for fin width down to 6 nm, indicating suitability for the ultra-scaled N5 node. Higher aspect ratio (i.e., taller fin) is seen to provide a boost in I_{ON} thus providing a better drive strength, but the increase in I_{OFF} (due to a small decrease in V_{TH}) results in a nearly constant I_{ON}/I_{OFF} ratio. The electrostatic control over the fin shaped channel is determined by the ability of the two vertical (each side of the fin) and one horizontal (top of the fin) gates. A good electrostatic control implies that these gates are





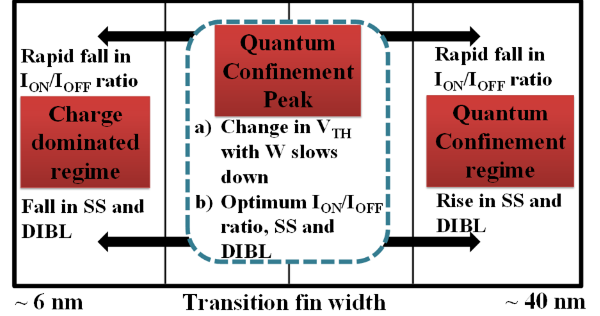


Fig. 6. Effect of W (6 nm to 40 nm) and AR (0.5 to 5) on the operability and performance of (a-d) $In_xGa_{1-x}As$ n-FinFET and (e-h) ε-Ge p-FinFET for long and short channel devices. (i) The transition from QC to charge dominated regime is seen through the variation of V_{TH} (solid) and I_{ON}/I_{OFF} ratio (dash-dash) for ε-Ge p-FinFET with W for undoped and doped channels for ultra-scaled channel length. (j) The predicted performance at the N5 node for In composition of 20 % and corresponding tensile strain of 1.3 % in Ge. (k) Transition between QC and charge dominated regimes for scaled FinFETs having doped fins.

able to invert the channel with a minimal applied gate bias. Each gate is able to deplete and consequently invert a portion of the channel. A large number of charges in the fin are depleted provided the fin is narrow, and making the fin taller (*i.e.*, increasing AR) is not expected to improve the electrostatic control in the form of SS, DIBL and I_{ON}/I_{OFF} ratio as the fraction of depleted charges remains the same (~ 1). This trend is clearly visible in Fig. 6(b), (c) and (d), implying that the aspect ratio does not affect the SS, DIBL and I_{ON}/I_{OFF} ratio in a significant way for the $In_xGa_{1-x}As$ n-FinFET.

B. *Effect of fin width and AR on ε-Ge p-FinFET*

The ε-Ge p-FinFET experiences multiple physical mechanisms such as (i) strain induced bandgap lowering, (ii) strain induced hole mobility enhancement, (iii) QC due to narrow width, (iv) high charge density, and (v) dopant induced bandgap narrowing. These affect the performance significantly resulting in complex contours seen in Fig. 6(e-h) compared to $In_xGa_{1-x}As$ n-FinFET. An interesting transition in the I_{ON}/I_{OFF} ratio, SS and DIBL occurs in the ε-Ge p-FinFET with reduction in fin width due to two competing phenomena: QC and charge reduction, as discussed below (Fig. 6(e-h)).

C. *Quantum confinement in narrow fins and performance at the N5 node*

In FinFETs, there are two fin width dependent phenomena, as indicated above, which compete to modulate the channel charges. QC of inversion charges in the fin causes a shift in the inversion charge peak away from the oxide-channel interface resulting in reduced interface scattering for real devices. This results in an additional capacitance in series with the oxide capacitance, resulting in a higher V_{TH} as well as an improvement in SS, DIBL and I_{ON}/I_{OFF} ratio [21][22]. The shift in V_{TH} due to QC, increases with decrease in effective mass and channel thickness [21][23]. Thus, the effect of QC is more dominant in $In_xGa_{1-x}As$ than in Ge, due to much lower effective mass [24]. These trends can be clearly seen in Fig. 6(a-d) for the $In_xGa_{1-x}As$ n-FinFET as its performance is solely determined by the dominant QC effect. The other mechanism affected by fin width scaling is the reduction in volume charges. Due to this mechanism, it is expected that the V_{TH} reduces in both the devices (n- and p-channel FinFET), indicating a better electrostatic control (improved SS and DIBL). Corresponding to the V_{TH} reduction, the I_{OFF} increases, thus reducing the I_{ON}/I_{OFF} ratio. The relative impact of this mechanism is governed by the charge density which is determined by N_{CH}. Thus, this mechanism will have an effect on the electrostatic control and leakage of the FinFET for moderate-to-high N_{CH}. The dependence of V_{TH} on W can be visualized using an approximate relation given by,

$$V_{TH} = V_{TH_i} + \frac{qN_{CH}\,w_d}{C_{ox}} + \Delta\varphi_{QM}(W) \qquad (1)$$

where, $\Delta\varphi_{QM}(W)$ is derived using Schrödinger-Poisson equations [23] assuming an undoped channel, and w_d is the depletion width (for scaled fins $w_d \cong W$). For highly doped channels, the second term (charge effect) dominates resulting in a steady fall in V_{TH} with decreasing fin width for a fixed aspect ratio. The third term (QC effect) is independent of doping and has an exponential rise with decreasing W [23]. Thus, for intrinsic/undoped/low-doped fins (*here*, $In_xGa_{1-x}As$ n-FinFET), the third term corresponding to the QC effect dominates resulting in gradual performance improvement governed by QC (Fig. 6(a-d)), down to W = 6 nm where the V_{TH} becomes higher than 0.5 V. For devices which utilize materials having low bandgap (*here*, ε-Ge p-FinFET), a significant N_{CH} or an exotic gate metal work function are two ways to ensure a normally-off

operation with a decent I_{ON}/I_{OFF} ratio, each of them having their own trade-offs. Due to the high N_{CH}, neither of the two terms from Eq. 1 can be ignored. As these two terms compete, V_{TH} will have different gradients depending on which term dominates at the particular value of W. For the value of N_{CH} = 6x10^{18} cm^{-3} selected here for the ε-Ge p-FinFET, the QC term is able to slow down the decrease of V_{TH} with W due to charge reduction, but not completely offset it. Thus, a nearly monotonous decrease in V_{TH} is observed in Fig. 6(e). On further inspection, $\left|\frac{d(V_{TH})}{dW}\right|$ which indicates the rate of V_{TH} decrease, slows down owing to the QC term. Furthermore, QC results in better electrostatic control resulting in reduced I_{OFF} [22]. This results in a gradual rise in I_{ON}/I_{OFF} ratio down to a specific width as seen in Fig. 6(b).

We predict that there exists a specific fin width associated with the transition between the charge dominated device operation and the QC dominated regime. Here, this transition width (W_T) is defined as the maximum fin width at which all metrics (I_{ON}/I_{OFF} ratio, SS and DIBL) approach a peak. Below W_T, the device is governed by the charge term (rapid decrease in V_{TH}) as well as punch through associated with reduced separation between the depletion regions under each gate, thus the I_{ON}/I_{OFF} ratio drops with further scaling. The SS and DIBL also reduce gradually due to the QC effect [22] down to the W_T. Below W_T, the effect of QC is masked by the large charge density. Initially, the punch through degrades the SS and DIBL [22], upon further scaling the charge volume reduction dominates, improving SS and DIBL (Fig. 6(e) and (f)). This indicates that for low bandgap channel based FinFETs utilizing high N_{CH}, there exists a small window of scaled fin widths (~ 10 nm), where the SS and DIBL improve in-spite of punch through effect, and the I_{ON}/I_{OFF} ratio continues to degrade. Furthermore, due to the significant charge density as well as QC resulting in shifts in relative strengths of the two effects, there exists a strong dependency on the aspect ratio compared to the $In_xGa_{1-x}As$ n-FinFET. This can be seen as peaks in I_{ON}/I_{OFF} ratio, SS and DIBL for the ε-Ge p-FinFET (Fig. 6 (f-h)). Thus, a proper choice of aspect ratio as well as W is crucial for CMOS operability and performance of the ε-Ge p-FinFET. These peaks occur due to the two competing mechanisms discussed above. Ignoring the peaks (no effect of QC), the SS and DIBL can be improved up to their physical limits provided the fin width is scaled well below ~ 6 nm, which adds more process complexity. In doing so, serious degradation in V_{TH} and I_{ON}/I_{OFF} ratio renders the device inoperable. Hence, with width scaling in a charge dominated device, there exists a tradeoff between device operability (V_{TH} and I_{ON}/I_{OFF} ratio) and device performance (SS and DIBL). For an AR=1, the material parameters (tensile strain and N_{CH}) of the ε-Ge p-FinFET demonstrated here are such that W_T occurs in the vicinity of ~ 20 nm (Fig. 6(f-h)), whereas the $In_xGa_{1-x}As$ n-FinFET has W_T well below 6 nm (Fig. 6(b-d)). This implies that the $In_xGa_{1-x}As$ n-FinFET can benefit from scaling down to ~ 6 nm, up to the fin width where V_{TH} goes beyond 0.5 V. For devices which can perform well in terms of all the metrics with intrinsic/undoped channels (*here*, $In_xGa_{1-x}As$ n-FinFET), the W_T can approach to sub-nm dimensions. Such an architecture can benefit from scaling for almost all realistic feature sizes (down to W_T) as the QC is able to offset the charge reduction effect.

The QC effect is offset by the significant charge density, resulting in a reversal of the trend between V_{TH}, I_{ON}/I_{OFF} ratio, SS and DIBL. This inverted trend between the electrostatic control and the I_{ON}/I_{OFF} ratio over a short window is opposite to that of QC effect. The trend is further evident as shown Fig. 6 (i), where the effect of channel doping and QC is visible. The device with high doping (6x10^{18} cm^{-3}) operates in the charge dominated regime (*solid blue*) and the QC is not able to offset the V_{TH} reduction with W (*see* Eq. 1). The device with undoped channel (*solid green*) shows a steady rise in V_{TH} with decreasing fin width owing to QC and the W_T approaches sub-nm range. In spite of the disadvantages, in this scenario, a high doping may be selected to get an I_{ON}/I_{OFF} ratio > 1000 and normally-off operation. It is desirable to have W_T equal or less than the minimum feature size realizable during photolithography, in order to completely utilize the benefits of scaling. This transition width depends on the volume of charges to be inverted within the channel as well as the QC gate capacitance associated with each gate, and the relative strengths of these two phenomena. Only a representative understanding of their contributions can be judged from Eq. 1. A detailed numerical analysis of the 3D geometry (*i.e.*, AR, W) based on quantum-corrected and experimentally calibrated models as shown in Fig. 6 estimates this intricate dependency. Similar to the $In_xGa_{1-x}As$ n-FinFET, the V_{TH} of Ge p-FinFET (as shown in Fig. 6e) is nearly independent of the AR but the electrostatic control (SS, DIBL and I_{ON}/I_{OFF} ratio) depends on the relative magnitudes of the two competing phenomena (*i.e.*, QC and charge reduction). At the transition point ~ 20 nm in Fig. 6(f-h), the peak values of SS, DIBL and I_{ON}/I_{OFF} ratio will occur, but the absolute value of the peak will be AR dependent as the two competing mechanisms are strong functions of device geometry. Thus, W_T can prove to be a critical parameter for future investigations in novel high mobility and low effective mass materials as channel for multi-gate devices. Furthermore, this system shows excellent CMOS operability and performance at the scaled N5 node, as shown in Fig. 6 (j). In addition, Fig. 6 (k) shows the transition in design consideration between QC and charge dominated regimes for scaled ε-Ge/$In_xGa_{1-x}As$ FinFETs while highlighting the performance tradeoffs.

V. Conclusion

The performance trends of a monolithically co-integrated hybrid CMOS system, where ε-Ge for p-FinFET and underlying $In_xGa_{1-x}As$ layer for n-FinFET, were investigated in the context of dimension scaling, material parameters and the effect of QC. This hybrid system utilizes the benefits of experimentally realistic metamorphic ($In_xGa_{1-x}As/In_xAl_{1-x}As$) and pseudomorphic (ε-Ge/$In_xGa_{1-x}As$) heteroepitaxy to create high performance complementary devices, suitable for 0.5 V CMOS operation. The In composition in $In_xGa_{1-x}As$ layer determines the tensile strain and device performance in Ge p-FinFET. The device metrics (V_{TH}, I_{ON}, I_{OFF}, SS, and DIBL) have a strong interdependent relationship with material and geometrical parameters. Based on these findings, we define a unique transition fin width (W_T) associated with QC, at which the performance is optimum and below W_T, the advantages of fin scaling are diminished. The ε-Ge p-FinFET and $In_xGa_{1-x}As$ n-FinFET exhibited this W_T in the vicinity of 20 nm and well

below 6 nm, respectively. The design considerations for devices operating on either side of W_T vary drastically. This transition width can enable better assessment of scaling performance of future material and device solutions for feature sizes approaching sub-nanometer range. Furthermore, this system showed excellent performance at the N5 node indicating suitability for applications such as low-power logic, RF and optoelectronics while being experimentally feasible.

VI. Acknowledgement

The authors thank John K. Ghra (Systems Administrator for ECE, Virginia Tech) for assistance in computational services.

VII. References

[1] T. Irisawa *et al*., "3D integration of high mobility InGaAs nFETs and Ge pFETs for ultra low power and high performance CMOS," *IEEE SOI-3D-Subthreshold Microelectronics Tech. Unified Conf. (S3S)*, Oct. 2013, pp. 1-2, doi: 10.1109/S3S.2013.6716513.

[2] W. Rachmady *et al*., "300mm Heterogeneous 3D Integration of Record Performance Layer Transfer Germanium PMOS with Silicon NMOS for Low Power High Performance Logic Applications," *IEDM Tech. Dig.*, Dec. 2019, pp. 29.7.1-29.7.4, doi: 10.1109/IEDM19573.2019.8993626.

[3] P. Goley and M. K. Hudait, "Germanium Based Field-Effect Transistors: Challenges and Opportunities," *Materials*, vol. 7, no. 3, pp. 2301-2339, Mar. 2014, doi: 10.3390/ma703230.

[4] S. Takagi *et al*., "III-V/Ge CMOS Device Technologies for High Performance Logic Applications," *ECS Trans.*, vol. 53, no.3, pp. 85-96, Nov. 2013, doi: 10.1149/05303.0085ecst.

[5] M. Heyns *et al.*, "Advancing CMOS beyond the Si roadmap with Ge and III/V devices," *IEDM Tech. Dig.,* Dec. 2011, pp. 13.1.1-13.1.4, doi: 10.1109/IEDM.2011.6131543.

[6] M. Clavel, P. Goley, N. Jain, Y. Zhu and M. K. Hudait, "Strain-Engineered Biaxial Tensile Epitaxial Germanium for High-Performance Ge/InGaAs Tunnel Field-Effect Transistors," *IEEE J. Electron Dev. Soc.*, vol. 3, no. 3, pp. 184-193, May 2015, doi: 10.1109/JEDS.2015.2394743.

[7] M. K. Hudait *et al.*, "Design, Theoretical, and Experimental Investigation of Tensile-Strained Germanium Quantum-Well Laser Structure," *ACS Appl. Electron. Mater*., vol. 3, no. 10, pp. 4535–4547, Oct. 2021, doi: 10.1021/acsaelm.1c00660.

[8] Nadine Collaert, Heterogeneous III-V/CMOS Technologies for Beyond 5G RF Front-end Modules, Microwave Journal January 2020. https://www.microwavejournal.com/articles/33299-heterogeneous-iii-vcmos-technologies-for-beyond-5g-rf-front-end-modules

[9] R. Joshi, S. Karthikeyan and M. K. Hudait, "Monolithically Co-integrated Tensile Strained Germanium and InxGa1-xAs FinFETs for Tunable CMOS Logic," *IEEE Trans. on Electron Dev.*, 2022, doi: 10.1109/TED.2022.3181112.

[10] K. J. Kuhn, "Considerations for Ultimate CMOS Scaling," *IEEE Trans. on Electron Dev*, vol. 59, no. 7, pp. 1813-1828, July 2012, doi: 10.1109/TED.2012.2193129.

[11] *TCAD Sentaurus Device Manual Release: H-2010.03,* Synopsys Inc., Mountain View, CA, USA, 2013.

[12] M. J. H. van Dal *et al.*, "Germanium p-Channel FinFET Fabricated by Aspect Ratio Trapping," *IEEE Trans. Electron Dev*., vol. 61, no. 2, pp. 430-436, Feb. 2014, doi: 10.1109/TED.2013.2295883.

[13] M. Radosavljevic *et al.*, "Electrostatics improvement in 3-D tri-gate over ultra-thin body planar InGaAs quantum well field effect transistors with high-K gate dielectric and scaled gate-to-drain/gate-to-source separation," *IEDM Tech. Dig.*, Dec. 2011, pp. 33.1.1-33.1.4, doi: 10.1109/IEDM.2011.6131661.

[14] A. V. Thathachary *et al*., "Impact of Sidewall Passivation and Channel Composition on InxGa1-xAs FinFET Performance," *IEEE Electron Device Letters*, vol. 36, no. 2, pp. 117-119, Feb. 2015, doi: 10.1109/LED.2014.2384280.

[15] R. People and J. C. Bean, "Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x} /Si strained-layer heterostructures," *Appl. Phys. Lett.*, vol. 47, no. 3, pp. 322–324, Aug. 1985. doi: 10.1063/1.96206.

[16] Schubert, E. (1993). Doping in III-V Semiconductors (Cambridge Studies in Semiconductor Physics and Microelectronic Engineering). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511599828

[17] A. Agrawal *et al.*, "Enhancement mode strained (1.3%) germanium quantum well FinFET (WFin=20nm) with high mobility (μ Hole=700 cm^2/Vs), low EOT (~0.7nm) on bulk silicon substrate," *IEDM Tech. Dig.,* Dec. 2014, pp. 16.4.1-16.4.4, doi: 10.1109/IEDM.2014.7047064.

[18] C. Convertino, C. B. Zota, D. Caimi, M. Sousa and L. Czornomaz, "InGaAs FinFETs 3D Sequentially Integrated on FDSOI Si CMOS with Record Performance," *2018 48th European Solid-State Device Research Conference (ESSDERC)*, pp. 162-165, doi: 10.1109/ESSDERC.2018.8486862.

[19] IEEE International Roadmap For Devices and Systems (IRDS) (2021). More Moore [Online]. Available: https://irds.ieee.org/

[20] Y. Choi, D. Ha, T. King and C. Hu, "Nanoscale ultrathin body PMOSFETs with raised selective germanium source/drain," *IEEE Electron Dev. Lett.*, vol. 22, no. 9, pp. 447-448, Sept. 2001, doi: 10.1109/55.944335.

[21] J. B. Chang et al., "Scaling of SOI FinFETs down to fin width of 4 nm for the 10 nm technology node," Symposium on VLSI Technology, 2011, pp. 12-13.

[22] X. He et al., "Impact of aggressive fin width scaling on FinFET device characteristics," IEDM Tech. Dig., pp. 20.2.1-20.2.4, doi: 10.1109/IEDM.2017.8268427.

[23] V. P. Trivedi and J. G. Fossum, "Quantum-mechanical effects on the threshold voltage of undoped double-gate MOSFETs," *IEEE Electron Dev. Lett.*, vol. 26, no. 8, pp. 579-582, Aug. 2005, doi: 10.1109/LED.2005.852741.

[24] S. -E. Huang, C. -L. Yu and P. Su, "Investigation of Fin-Width Sensitivity of Threshold Voltage for InGaAs and Si Negative-Capacitance FinFETs Considering Quantum-Confinement Effect," *IEEE Trans. on Electron Dev.*, vol. 66, no. 6, pp. 2538-2543, June 2019, doi: 10.1109/TED.2019.2907994.