CMOS beyond Si: Nanometer-Scale III-V MOSFETs

Jesús A. del Alamo, Xiaowei Cai, Jianqiang Lin, Wenjie Lu, Alon Vardi, Xin Zhao

Microsystems Technology Laboratories, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA; alamo@mit.edu

***Abstract*—CMOS electronics constitutes the backbone of our information society. The extraordinary success of CMOS is based on the unique scaling properties of the Si MOSFET. Over the last decade, increasing concerns have been expressed about the ability of the Si MOSFET to continue its historical scaling path. This threatens the future of Moore's Law and its trifecta of benefits: cost, power and performance. For about a decade now, III-V compound semiconductors have been under intensive research in an effort to push Moore's Law beyond the point that Si can reach. How far have we come in this quest? This paper summarizes the state of the art of III-V compound semiconductor MOSFETs for future CMOS.**

Keywords—InGaAs, CMOS, Moore's Law, MOSFET, III-V, compound semiconductors

I. INTRODUCTION

For over 50 years, Moore's Law has been the guiding principle that has propelled CMOS logic technology to become the backbone of our information society. Moore's Law, as traditionally stated, refers to the doubling of transistor density in microchips every two years [1]. An amazing aspect of MOSFET physics is that *"smaller is better."* Footprint scaling in Si MOSFETs has brought along commensurate improvements in current drive, capacitance and switching delay [2]. This has yielded triple dividends in terms of cost, power and performance that have fueled the industry for over 50 years.

All this seems threatened now. In recent times, continued MOSFET scaling has required the introduction of an increasing diversity of new materials, new processes and new device structures. As a consequence, the cost per unit area of semiconductor has accelerated in the last few years [3]. Faster transistor density scaling has managed to somehow mitigate this problem. At the same time, supply voltage scaling, needed in order to manage power dissipation, has put a dent in performance improvements (performance essentially refers to current density in the ON state) if not brought it to a complete halt. So, in recent times, instead of triple benefits, it has become common to refer to "cost, power, performance… choose two!"

It is in this context that III-V compound semiconductors entered the scene in the mid 2000's [4,5]. While III-V High Electron Mobility Transistors (HEMT) have long ruled many high-frequency applications [6], III-V MOSFETs were largely irrelevant. This was due to Fermi level pinning at the oxide-semiconductor interface that prevented the modulation of the surface potential in the channel by the gate. All this changed with the finding that Atomic Layer Deposition of a high-permittivity dielectric on top of III-V semiconductors yields a high-quality interface with an unpinned Fermi level [7]. From this moment, progress in III-V MOSFET performance made dramatic strides [8,9]. Today, III-V planar MOSFETs, FinFETs and Nanowire FETs have been demonstrated. How far III-V MOSFETs have come and what are their prospects and challenges for future logic are the topics that this paper reviews.

II. THE CASE FOR III-V MOSFETS FOR CMOS

The potential of certain III-V compound semiconductors for high speed and high frequency applications has been known for some time. III-V High-Electron Mobility Transistors (HEMT) and Heterojunction Bipolar Transistors (HBT) have long held the record in figures of merit that are relevant for high frequency applications [6,10]. As a result, these devices are widely used in the front-end communication subsystems of many consumer, infrastructure and defense applications.

Among common III-Vs, InGaAs has emerged in the last few years as a promising choice for future n-channel MOSFETs. The mobility of electrons in InGaAs can easily exceed that of Si by a factor of 5 to 10, depending on composition. The electron velocity at the virtual source, a more relevant parameter for logic MOSFETs, is a factor of 2 to 3 higher than Si [11,12]. For p-channel devices, InGaSb under compressive strain exhibits a hole mobility comparable to Ge, both being the highest among common semiconductors [13].

Theoretically, the superior electron transport properties of InGaAs, as compared to strained Si, should translate into a significant advantage in logic performance metrics. Recent projections for deeply scaled nanowire MOSFETs give InGaAs an advantage of 37% in terms of ON current, a factor of 2.3 in switching time and a 40% lower switching energy as compared with similar Si devices operating at V_{DD}=0.5 V [14]. These are projections of intrinsic device performance, i.e., without considering contact resistance and interconnect load. They are obtained from a Non-Equilibrium Green Functions quantum transport tool that includes realistic band structures. These projections assume ballistic transport for InGaAs but 60% ballisticity for strained Si. Separate multi sub-band Monte Carlo simulations indicate that electron-phonon and alloy scattering reduce the ON current in InGaAs by less than 10%,

justifying the full ballisticity assumption [14]. Ballisticity of short-channel Si MOSFETs has been experimentally verified to be around 60% [15]. This difference in ballisticity is crucial as the assumption of fully ballistic transport for both channel materials largely eliminates the advantage of InGaAs due to its lower quantum capacitance [16,17].

Superior carrier transport properties, when compared with Si, constitute a necessary but not sufficient requirement in a MOSFET technology that aims to surpass Si. An excellent oxide-semiconductor interface is also critical. For over 25 years, this was the stumbling block of III-Vs. All oxide deposition techniques on III-Vs resulted in Fermi level pinning at the oxide/semiconductor interface that prevented the effective modulation of the surface potential. This was the result of unavoidable native oxides. Atomic Layer Deposition (ALD) changed this. In the early 2000's, it was found that in the first stages of ALD of a dielectric on a III-V surface, the native oxides were scavenged before a fresh oxide was deposited. This yielded a very low interface density with a Fermi level that was able to move across the entire bandgap. For InGaAs, interface state densities in the 10^{11} $cm^{-2}.eV^{-1}$ range have been reported [18]. Similarly, in InGaSb, a lowest interface state density of $3x10^{11}$ $cm^{-2}.eV^{-1}$ has also been realized [19].

Low contact resistivity ohmic contacts is another essential element of a high-performance MOSFET technology. 0.1 $\Omega.\mu m^2$ is often given as a maximum tolerable value for scaled logic CMOS. Great strides have also taken place in recent times in this area. In InGaAs, refractory ohmic contacts with a resistivity of a few tenths of $\Omega.\mu m^2$ have been reported [20]. In the InGaSb system, ohmic contacts are often made to a surface cap layer of p^+-InAs. This is to prevent exposure of the highly reactive InGaSb surface. Ni-based contacts to InAs have been demonstrated with contact resistivities in the low $\Omega.\mu m^2$ range [21].

This progress in important aspects of device technology bode well for our ability to realize the intrinsic potential of III-V compound semiconductors for future logic applications. Recent device demonstrations are encouraging but more work is needed. The next four sections provide an overview of the state of the art among III-V planar MOSFETs, FinFETs, and Vertical Nanowire MOSFETs based on InGaAs and InGaSb, with emphasis on the research of the authors. The key issue of Si integration is not discussed in this paper.

III. InGaAs Planar MOSFETs

Planar MOSFETs, per se, are of limited interest because they have poor scaling potential. However, the planar device architecture constitutes a superb platform to develop core fabrication processes and to identify and build models for new and relevant device physics.

The last few years have seen great progress in planar InGaAs MOSFETs. Fig. 1 sketches a planar self-aligned InGaAs Quantum-Well MOSFET structure investigated at MIT [22]. This is prototypical of similar efforts elsewhere [23-26]. In the last year, this device structure has yielded the highest transconductance of any III-V FET, including HEMTs, a value of 3.45 mS/μm at V_{DS}=0.5 V [22]. A key objective of our effort was to develop a Si compatible fabrication process in the front end. This is accomplished through the use of refractory ohmic contacts and gate metal, extensive use of RIE and avoidance of lift-off.

This device structure has yielded not only excellent performance but also insight into important transport and reliability physics. An important contribution was the identification of band-to-band tunneling and its multiplication by the current gain of a parasitic floating-base lateral bipolar transistor that, similar to floating-body Silicon-on-Insulator (SOI), lurks inside the MOSFET [27]. This is responsible for the excessive drain current in the OFF-state in tightly designed devices.

Separately, we have shown the enormous importance of ballistic transport in scaled InGaAs MOSFETs. In particular, we have estimated the mean-free path of our transistors at around 80-160 nm (depending on carrier degeneracy) at room temperature [28]. This means that in the sub-50 nm regime, ballistic effects are paramount. One implication of this is that conventional assumptions about transistor modeling fail. An example is the standard approach to extract the sum of the source and drain resistances in MOSFETs, R_{sd}, which is based on measurements of the ON resistance at strong positive gate bias in devices of different gate lengths and extrapolation to zero gate length, that is $R_{sd}=R_{on}(L_g=0)$ [28]. In a transistor family with prominent ballistic effects, $R_{on}(L_g=0)$ includes the ballistic resistance associated with ballistic transport in the channel and as a result, overestimates R_{sd} [28]. Proper accounting of ballistic resistance has allowed us to extract an excellent value of R_{sd}=74 ohm.μm in our self-aligned InGaAs MOSFETs [28].

Self-aligned planar InGaAs MOSFETs have also presented us with opportunities to study important reliability issues of relevance in any device structure. An unexpected one emerged from the use of F species in RIE. This is device instability associated with the reversible F passivation of Si dopants in InAlAs. F^- ions are known to passivate the Si donors in n-type InAlAs resulting in a loss of carrier concentration [29]. At the same time, F^- moves in the presence of electric fields. No permanent damage results but the device characteristics are rendered unstable [30]. The solution to this problem consists of eliminating n-InAlAs from the device structure. An n-type doped InAlAs ledge is often introduced to facilitate the linkage between the extrinsic source and drain and the intrinsic device. Substituting n-InAlAs by n-InP, not only eliminated the F-induced instability but also increased the electron concentration in the linkage region resulting in a much reduced access resistance and a record transconductance [31].

To assess progress, a benchmark of InGaAs and Si planar MOSFET transconductance is shown in Fig. 2. The Si data is from Intel's technology. These types of comparisons are never entirely fair but the staggering recent progress of InGaAs MOSFETs is evident.

IV. InGaAs FinFETs

The FinFET is the current state of the art Si logic MOSFET [32]. In InGaAs, numerous demonstrations of FinFETs have been reported [33-38]. In nearly every case, the fins are

prepared by RIE. Future scalable FinFETs must have fin widths in the 5-7 nm range. Yes, until recently, there had not been InGaAs FinFET demonstrations with fin widths below 15 nm.

At MIT, we have developed etching technology for III-Vs that comfortably reach sub-10 nm wide fins with smooth and highly vertical sidewalls [39-41]. We accomplish this through a combination of RIE and digital etch [42]. Using this technology, we have demonstrated InGaAs FinFETs with fin widths down to 7 nm [40,41]. Device sketches along the fin direction and across the intrinsic portion of the device are shown in Fig. 3. A feature unique to this process is that the HSQ mask that is used to define the fins in the RIE process is left in place. This makes this device a double-gate MOSFET as opposed to a tri-gate MOSFET. While theoretically not optimum, in practice this leads to a robust fabrication process with large process tolerance and yielding scaled devices. Our most aggressively scaled FinFET features a fin width of 7 nm and a gate length of 30 nm and exhibits a g_m of 900 μS/μm (normalized by conducting gate periphery) [41].

While well-behaved scaling has been demonstrated for devices with fin widths in the 20-22 nm range, width scaling below ~ 20 nm is far from ideal [41]. We observe a rapid degradation of transconductance and ON resistance as the fin width scales down. Fin-width scaling also does not result in significant improvement in the subthreshold characteristics, unlike what is to be expected. The origin of this anomalous scaling behavior is at the present not understood. Several sources associated with the etched sidewalls are of concern: excessive interface state density, surface roughness, loss of semiconductor stoichiometry and line-edge roughness.

Nevertheless, the results obtained by our team represent a significant advancement of the state of the art. Fig. 4 benchmarks the peak transconductance obtained in InGaAs FinFETs against that of Si FinFETs. In this figure, the transconductance is normalized by the fin footprint. This is a figure of merit that balances the twin goals of transistor density and performance. Our devices (red stars) are the first demonstrations of InGaAs FinFETs with sub-15 nm fin widths and channel aspect ratios greater than 2. We have doubled the transconductance of earlier InGaAs FinFETs but it still lags that of Si devices. The difference in voltage (0.5 V for InGaAs, 0.8 V for Si) accounts for some of that gap, but a gap nevertheless remains. This graph makes it clear that there is substantial work ahead before the intrinsic potential of InGaAs is realized in the FinFET architecture.

V. InGaAs Vertical Nanowire FETs

The transistor architecture with the ultimate scaling potential is the Vertical Nanowire (VNW) MOSFET. The gate-all-around geometry provides the most effective channel charge control [43]. In addition, the VNW MOSFET is a very compelling architecture because, with carrier flow in the perpendicular direction to the wafer surface, footprint scaling is decoupled from gate, contact, and spacer length scaling. This affords device design freedom that is not available in any other MOSFET geometry.

At MIT we are exploring InGaAs VNW MOSFETs etched by a combination of RIE and digital etch. Recent improvements in our etching techniques have yielded InGaAs nanowires down to 5 nm in diameter with vertical walls and an aspect ratio in excess of 40, as shown in Fig. 5 [44]. We build VNW MOSFETs on these structures using a process that largely follows earlier demonstrations [45,46,47]. This process yields devices with diameters between 30 and 50 nm. In our latest results, we have demonstrated 40 nm diameter single-nanowire transistors with a channel length of 80 nm and a peak g_m of 720 μS/μm at 0.5 V [48]. In combination with a saturated subthreshold swing of 80 mV/dec, this is one of the best balanced VNW-MOSFET ever demonstrated.

Fig. 6 benchmarks InGaAs and Si VNW MOSFETs as a function of nanowire diameter. g_m is normalized by nanowire periphery. Most results are at 0.5 V. InGaAs VNW MOSFETs compare well with their Si counterparts. Scaling to sub-10 nm diameters in both types of devices still needs to be demonstrated.

VI. InGaSb FinFETs

InGaSb is the most attractive III-V compound semiconductor for p-channel transistors. The hole mobility in InGaSb is relatively high and, as in Si and Ge, it increases quickly with the application of compressive stress [13]. There have been a number of demonstrations of p-type InGaSb MOSFETs that have established the viability of this device concept [49,50].

In our group at MIT, we have been working for several years towards demonstrating InGaSb FinFETs. This required the development of a reactive ion etching technology for thin fins with high aspect ratio [21]. We have also developed an ohmic contact technology for p+-InAs (the preferred cap for InGaSb MOSFETs) that yields contact resistivities in 10^{-8} ohm.cm^2 range [51]. Recent p-type FinFET results are quite promising, although properly turning off the devices remains a significant challenge [52]. Underlying this problem is the high reactivity of antimonides which oxidize easily. This makes it difficult effective smoothing and passivating of the etched sidewalls using chemical treatments. In a significant recent development, we have recently demonstrated a benign digital etch scheme for InGaSb [44]. This should be instrumental in achieving InGaSb FinFETs with improved subthreshold characteristics.

VII. Conclusions

To summarize, there has been great progress in the last few years in planar, fin and nanowire InGaAs n-channel MOSFETs. Planar devices now exhibit extraordinary performance. FinFETs with fin widths in the target range have been demonstrated but their performance is still disappointing. Vertical nanowire FETs with diameters in the desired range have yet to be realized. InGaSb p-type MOSFETs and FinFETs, while promising, are still at their infancy. In all 3-dimensional III-V device architectures, many extrinsic issues remain to be understood. This makes it difficult at this time to project the performance of future scalable devices.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation (E3S STC Award 0939514), Defense Threat Reduction Agency (HDTRA1-14-1-0057), Semiconductor Research Corporation (2016-LM-2655), Samsung Electronics, Lam Research and Korea Institute of Science and Technology. III-V transistor and test structure fabrication at MIT is carried out in the fab facilities of the Microsystems Technology Laboratories and the Electron Beam Lithography Facility.

REFERENCES

[1] G. E. Moore, Proc. 1965 IEEE Int. Electron Dev. Meeting, p. 11.

[2] J. A. del Alamo, "Integrated Microelectronic Devices," Pearson, 2017.

[3] W. M. Holt, Proc. 2016 ISSCC, p. 8.

[4] S. Datta et al., Proc. 2015 IEEE Int. Electron Dev. Meet., p. 763.

[5] D. H. Kim and J. A. del Alamo, Proc. 2006 IEEE Int. Electron Dev. Meeting, p. 837.

[6] J. A. del Alamo, Proc. 2011 Comp. Semic. Manuf. Techn. Conf. p. 17.

[7] P. D. Ye et al., IEEE Electron Device Lett. 24, p. 209 (2003).

[8] H. Riel, L.-E. Wernersson, M. Hong, and J. A. del Alamo, MRS Bull. 39, p. 668 (2014).

[9] J. A. del Alamo, D. A. Antoniadis, J. Lin, W. Lu, A. Vardi, and X. Zhao, J. Electron Dev. Soc. 5, p. 205 (2016).

[10] [10] W. Snodgrass, W. Hafez, N. Harff, and M. Feng, Proc. 2006 IEEE Int. Electron Dev. Meet., p. 595

[11] D. H. Kim, J. A. del Alamo, D. A. Antoniadis, and B. Brar, Proc. 2009 IEEE Int. Electron Dev. Meeting, p. 861.

[12] J. A. del Alamo, Nature 479, p. 317 (2011).

[13] B. R. Bennett, M. G. Ancona, and J. B. Boos, MRS Bull. 34, p. 530 (2009).

[14] M. Rau, E. Caruso, D. Lizzit, D. Esseni, A. Schenk, L. Selmi and M. Luisier, Proc. 2016 IEEE Electron Devices Meeting, p. 758 (the switching energy given in the original manuscript was in error, un updated value received from the author is given in the present manuscript).

[15] A. Majumdar and D. A. Antoniadis, "Analysis of Carrier Transport in Short-Channel MOSFETs." IEEE Trans. Electron Dev. 61, p. 351 (2014).

[16] S. H. Park, Y. Liu, N. Kharche, M. S. Jelodar, G. Klimeck, M. S. Lundstrom and M. Luisier, IEEE Trans. Electron Dev. 59, p. 2107 (2012).

[17] R. Kim, U. E. Avci, and I. A. Young, Proc. 2015 IEEE Int. Electron Dev. Meet., p. 875.

[18] N. Waldron et al., Proc. 2015 IEEE Int. Electron Dev. Meeting, p. 799.

[19] A. Nainani et al., J. Appl. Phys. 109, p. 114908 (2011).

[20] J. Lin, D. A. Antoniadis, and J. A. del Alamo, IEEE Trans. Electron Dev., 63, p. 1020 (2016).

[21] W. Lu, J. K. Kim, J. F. Klem, S. D. Hawkins and J. A. del Alamo, Proc. 2015 IEEE Int. Electron Devices Meeting, p. 819.

[22] J. Lin, X. Cai, Y. Wu, D. A. Antoniadis, and J. A. del Alamo, IEEE Electron Dev. Lett. 37, p. 381 (2016).

[23] Y. Sun et al., Proc. 2014 IEEE Int. Electron Dev. Meet., p. 582.

[24] S. Chang et al., Proc. 2013 IEEE Int. Electron Dev. Meet, p. 417.

[25] S. Lee et al., IEEE Electron Dev. Lett. 35, p. 621 (2014).

[26] M. Egard et al., Proc. 2011 IEEE Int. Electron Dev. Meet., p. 303.

[27] J. Lin, J., D. A. Antoniadis, and J. A. del Alamo, IEEE Trans. Electron Dev., 62, p. 1448 (2015).

[28] J. Lin, Y. Wu, J. A. del Alamo and D. A. Antoniadis, IEEE Trans. Electron Dev., 63, p. 1464 (2016).

[29] N. Hayafuji, Y. Yamamoto, N. Yoshida, T. Sonoda, W. Takamiya, and S. Mitsui, Appl. Phys. Lett., 66, p. 863 (1995).

[30] N. Hayafuji, Y. Yamamoto, T. Ishida, and K. Sato, Appl. Phys. Lett., 69, p. 4075 (1996).

[31] X. Cai, J. Lin, D. A. Antoniadis, and J. A. del Alamo, Proc. 2016 IEEE Int. Electron Dev. Meet., p. 63.

[32] S. Natarajan et al., Proc. 2014 IEEE Int. Electron Dev. Meet., p. 71.

[33] Y.Q. Wu, R. Wang, T. Shen, J.J. Gu, and P.D. Ye, Proc. 2009 IEEE Int. Electron Dev. Meet., p. 331.

[34] M. Radosavljevic et al., Proc. 2011 IEEE Int. Electron Dev. Meet., p. 765.

[35] T.-W. Kim et al., Proc. 2013 IEEE Int. Electron Dev. Meeting, p. 425.

[36] V. Djara et al., Proc. 2015 Symp. VLSI Tech., p. 176.

[37] R. Oxland et al., IEEE Electron Dev. Lett., 37, p. 261 (2016).

[38] C. B. Zota, F. Lindelow, L.-E. Wernersson, and E. Lind, Proc. 2016 Int. Electron Dev. Meet., p. 55.

[39] X. Zhao and J. A. del Alamo, IEEE Electron Dev. Lett., 35, p. 521 (2014).

[40] A. Vardi, J. Lin, W. Lu, X. Zhao and J. A. del Alamo, Proc. 2016 Symp. VLSI Tech., p. 136.

[41] A. Vardi and J. A. del Alamo, IEEE Electron Dev. Lett., 37, p. 1104 (2016).

[42] A. Vardi, J. Lin, W. Lu, X. Zhao, and J. A. del Alamo, Proc. 2017 Comp. Semic. Manuf. Tech. Conf. (to be published).

[43] C. Zhang and X. Li, IEEE Trans. Electron Dev., 63, 223 (2016).

[44] W. Lu, X. Zhao, D. Choi, S. El Kazzi, and J. A. del Alamo, IEEE Electron Dev. Meet., 38, p. 548 (2017).

[45] K. Tomioka, M. Yoshimura, and T. Fukui, Nature, 488, p. 189 (2012).

[46] K. -M. Persson et al., Proc. 2012 IEEE Dev. Res. Conf., p. 195.

[47] X. Zhao, J. Lin, C. Heidelberger, E. A. Fitzgerald, and J. A. del Alamo, Proc. 2013 IEEE Int. Electron Dev. Meet., p. 695.

[48] X. Zhao, C. Heidelberger, E. A. Fitzgerald, and J. A. del Alamo, Proc. 2017 Comp. Semic. Week (to be presented).

[49] A. Nainani et al., Proc. 2010 IEEE Int. Electron Dev. Meet., p. 138.

[50] K. Takei et al., Nano Lett. 12, 2060 (2012).

[51] L. W. Guo, W. Lu, B. R. Bennett, J. B. Boos, and J. A. del Alamo, IEEE Electron Dev. Meet., 36, p. 546 (2015).

[52] W. Lu, S. Mack, B. R. Bennett, and J. A. del Alamo, Proc. 2017 Comp. Semic. Week (to be presented).

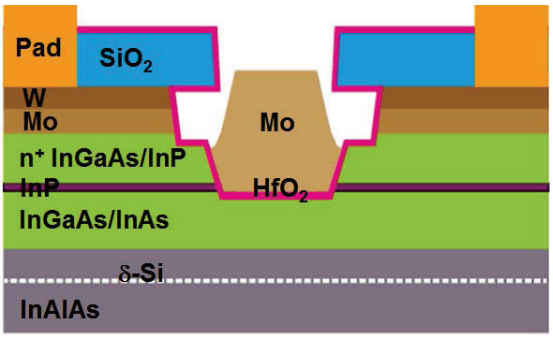

Figure 1: Cross-sectional schematic of a self-aligned InGaAs Quantum-Well MOSFET fabricated by a contact-first, gate-last process [22].

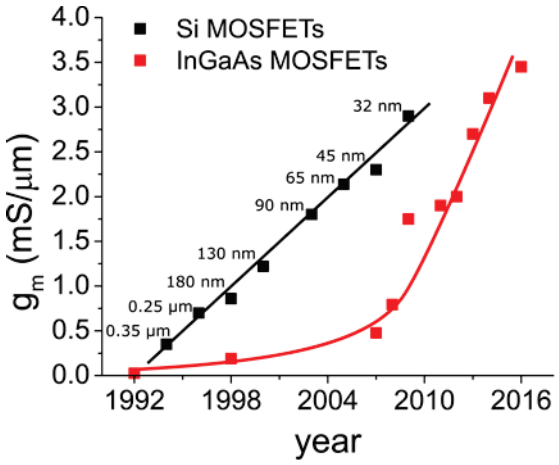

Figure 2: InGaAs planar MOSFET transconductance as a function of year of demonstration. For reference, g_m of Intel's Si planar n-MOSFETs are included.

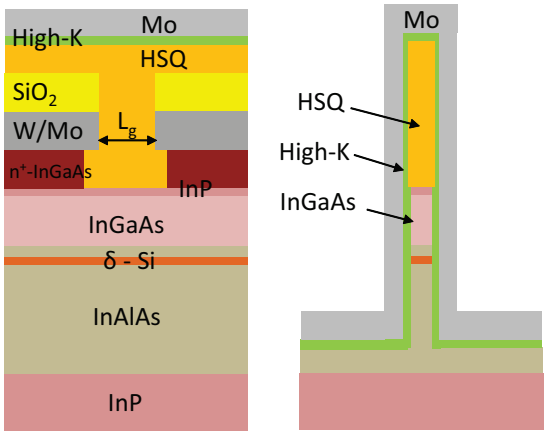

Figure 3: Cross-section schematic of self-aligned InGaAs FinFET fabricated at MIT. Left: along fin; right: across fin [40,41].

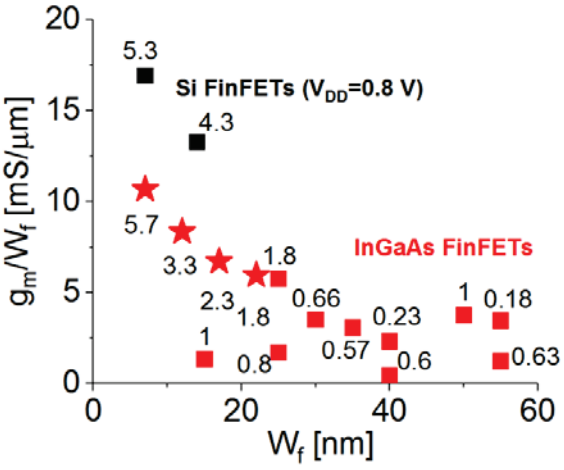

Figure 4: Transconductance per fin footprint as a function of fin width for InGaAs FinFETs and state-of-art Si FinFETs. The number next to each data point represents the channel aspect ratio. The red stars are devices recently fabricated at MIT [41].

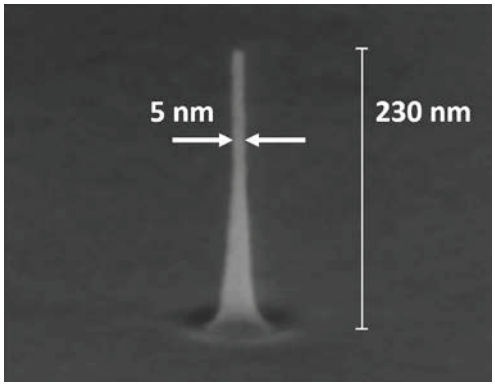

Figure 5: 5 nm diameter InGaAs nanowire fabricated by reactive ion etching and alcohol-based digital etch [44].

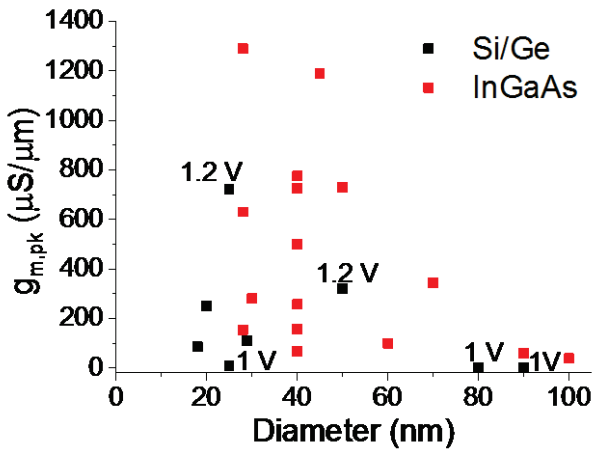

Figure 6: Transconductance of InGaAs and Si/SiGe vertical nanowire n-type MOSFETs as a function of nanowire diameter. Unless indicated, V_{DD}=0.5 V.