III-V/Ge Channel Engineering for Future CMOS

M. A. Wistey[a,b], U. Singisetti[a], G. J. Burek[a], E. Kim[c], B. J. Thibeault[a], A. Nelson[a], J. Cagnon[b], Y.-J. Lee[d], S. R. Bank[e], S. Stemmer[b], P. C. McIntyre[c], A. C. Gossard[b], M. J. W. Rodwell[a]

[a] Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106-9560, USA
[b] Materials Department, University of California, Santa Barbara, California 93106-5050, USA
[c] Materials Department, Stanford University, Stanford, California 94305, USA
[d] Intel Corporation, Technology Manufacturing Group, Santa Clara, CA 95054, USA
[e] University of Texas at Austin, Austin, Texas 78712, USA

As silicon CMOS reaches its scaling limits, alternative materials become more attractive. Dielectric thickness and parasitic resistance and capacitance do not scale well, so "more than Moore" scaling is required even to keep up with Moore's Law. Replacing Si MOSFET channels on a short time scale (3-6 years) raises significant challenges for any proposed material or device structure. New materials must be compatible with Si CMOS fabrication. $In_{1-x}Ga_xAs$ based MOSFETs offer higher carrier velocities than Si, plus contact resistivities below 10^{-8} Ω-cm^2, mature processing, and straightforward heterostructure confinement for vertical scaling, and additional degrees of freedom in composition and heterostructure for future scaling. Self-aligned source-drain regrowth places contact metal within 30nm of the channel, reducing access resistance. Here we demonstrate InGaAs channels with self-aligned regrowth of source/drain contacts. This work led to depletion mode InGaAs MOSFETs with peak transconductance of 0.24 mS/μm.

1. Introduction

Silicon-based electronics is one of the most successful technologies in history. More than 10^{19} transistors are made each year, at costs of a nanodollar per transistor. These gains have been possible because the physical size of transistors has been scaling to progressively smaller dimensions, conforming to Moore's Law. Until recently, operating frequency also scaled with the inverse of gate length because electrons could traverse shorter distances in less time at the same velocity. Other semiconductors such as III-V's have higher electron velocities, but the native oxide on silicon, SiO_2, is both passivating and insulating. These properties are particularly important for MOS field effect transistors (MOSFETs) because current flows along the surface of the semiconductor. However, if the SiO_2 or similar oxynitride (SiO_xN_y) is thinner than a few nanometers, electrons can tunnel through it, causing leakage currents. Starting with the 45 nm technology node, SiO_2 gate dielectrics are being replaced by high-k gate dielectrics such as HfO_2 in CMOS field effect transistors. The higher dielectric constant produces the same charge in the channel but allows a physically thicker dielectric, reducing tunneling. Because new dielectrics already need to be employed on silicon, the simplicity of SiO_2 is

no longer an advantage, and this is an appropriate time to consider new materials that could replace Si channels (1,2,3).

Proposed new channels must meet a number of conditions. First, the new channel must offer a substantial performance improvement over silicon. Second, new channels must be growable on Si and compatible with CMOS processing since Si remains the substrate of choice for most electronics. Third, regardless of material, the new structure must reduce parasitic resistance, capacitance, and leakage currents compared to comparable Si CMOS and be scalable to future generations.

This paper examines the requirements and prospects for future FET channels that could be implemented beyond the 32nm scaling generation and be useful for the next two decades. Section 2 examines the scaling laws that motivate the introduction of new channel materials, including parasitics and fundamental limits that diminish the benefits of scaling in MOSFETs and high electron mobility transistors (HEMTs). Section 3 discusses difficulties in scaling. Section 4 presents one particular channel material (InGaAs) that meets these requirements. Finally, Section 5 shows the technical developments that have resulted in the first scalable, InGaAs MOSFETs.

2. Basic FET Scaling Laws

Field effect transistors include MOSFETs, HEMTs, MESFETs, FinFETs, and more. Regardless of details, all FETs obey a simple principle: applying an electric field across an insulating barrier either enhances or depletes a given conducting channel, thus turning current on or off. Thus most FETs follow certain basic scaling laws. Many are outlined in Dennard and elsewhere (4,5), but we repeat essentials here for emphasis. We assume highly scaled FETs by confining our analysis to injection of carriers at the source assuming high fields and constant velocity *v,* whether from thermal injection or saturated velocity. A simplified FET longitudinal cross section is shown in Figure 1.

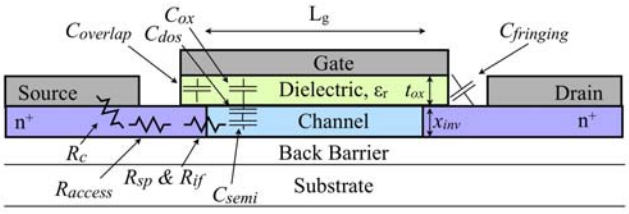

Figure 1. Simplified FET.

FET scaling laws under the assumptions of constant voltage and constant velocity are shown in Table 1. Constant-voltage scaling attempts to increase the transistor bandwidth by a factor of a without changing any external currents, voltages, or resistances. Because electrons travel with finite velocity, reducing the transit time from source to drain by $1/a$ requires that the gate length L_g be reduced by $1/a$. Similarly, all RC delays must be reduced by $1/a$. We consider the implications for several of these parameters below.

Effective oxide thickness: $EOT \sim 1/a$. As L_g decreases, the drain exerts more control over the channel. To prevent short-channel effects such as drain induced barrier lowering (DIBL), subthreshold current, and punchthrough, the channel must be brought under better control of the gate. Horizontal scaling thus requires vertical scaling. The effective dielectric thickness (EOT) must be scaled either by reducing its physical thickness t_{ox} or by increasing its dielectric constant ε_r (6,7).

TABLE I. Elementary FET scaling laws in the constant voltage, constant velocity limit.

Parameter	Symbol	Units	Law	Comments
Gate length	L_g	μm	$1/a$	--
Gate width	W	μm	$1/a$	--
Effective oxide thickness $= t_{ox}\varepsilon_{SiO2}/\varepsilon_r$	EOT	nm	$1/a$	Tunneling limits scaling
Channel or inversion thickness	t_{inv}	nm	$1/a$	
Channel lateral electric field	E	V/m	a	
Substrate doping	N_a , N_d	$1/cm^3$	a^2	Back barrier relaxes this
Gate dielectric capacitance $\sim \varepsilon_{SiO2}L_gW/EOT$	C_{ox}	fF	$1/a$	EOT scaling constraints → scales faster than $1/a$
Semiconductor capacitance	C_{semi}	fF	$1/a$	Scales faster than $1/a$
DOS capacitance[A]	C_{DOS}	fF	$1/a$	With given material, falls faster than $1/a$
Total gate-channel capacitance $= [1/C_{ox}+1/C_{semi}+1/C_{DOS}]^{-1}$	$C_{g\text{-}ch}$	fF	$1/a$	Given C_{semi}, C_{DOS} limits, falls faster than $1/a$
Carrier velocity in channel	v	cm/s	1	Increases if degenerate carrier population.
Transconductance $\partial I_{ds}/\partial V_g \sim C_{g\text{-}ch}v/L_g$	g_m	mS	1	Falls because $C_{g\text{-}ch}$ scales faster than $1/a$
Fringing capacitance	$C_{fringing}$	fF	$1/a$	$1/a$
Overlap capacitance	$C_{overlap}$	fF	$1/a$	EOT scaling constraints → scales slower than $1/a$
Gate delay $= C_{total}V_{dd}/I_d$	τ_g	ps	$1/a$	Fails to scale as $1/a$ given g_m scaling failure
On-state resistivity	R'_{TOT}	Ω-cm	$1/a$	$R_{TOT}=R'_{TOT}/W$=constant
Quantized conductance resistivity	R'_{QW}	Ω-cm	$C_{ox}^{-1/2}$	Scales slowly at best
Contact resistivity	ρ_c	$\Omega\text{-cm}^2$	$1/a^2$	Must drop faster
Extrinsic access resistivity	R'_{access}	Ω-cm	$1/a$	Must drop faster to compensate R'_{QW}
Carrier transit time	τ_r	s	$1/a$	
Transit frequency	f_t	GHz	a	
Voltage	V	V	1	1
Current density	I_d/W	mA/μm	a	Limited g_m →less than a

Notes: (A) Assumes parabolic bands and lowest valley.
(B) Depends on barrier heights and m*.

Channel thickness: $t_{inv} \sim 1/a$. Similarly, the channel thickness must be scaled as well, increasing confinement to control short channel effects (8). For inversion-type MOSFETs, this requires increasing the substrate doping N_a by a factor of a^2, a technological challenge. Furthermore, high p-type doping in the body adjacent to n^{++} doped source-drain regions may lead to increased source/body capacitance and leakage by tunneling currents to the body. Heterostructures, FinFETs, or back barriers can reduce these effects and provide strong confinement without such high body/channel doping.

Source/drain doping: $N_{s,d} \sim 1/a^2$, shallower doping profile. Current conservation requires that the electron density at the source side of the channel equal or exceed the electron density just inside the channel, to prevent source exhaustion. Although the total resistance of the source must remain constant, it must provide a times more electrons in a layer $1/a$ as thick. Thus the source/drain doping (cm^{-3}) must also increase by a^2 and do so in a very shallow layer to avoid spreading resistance. This poses difficulties for traditional ion implantation, which distributes dopants broadly over some depth.

Capacitances: Overlap capacitance $C_{overlap} \sim x_j W/EOT$ scales properly as $1/a$ if all dimensions and EOT scale.

Contact resistance: $R_c \sim 1/a^2$. Source/drain contact areas shrink by $1/a^2$, so to maintain constant resistance, the contact resistivity ρ_c (per unit area) must scale as $1/a^2$ as well. This presents a demanding challenge. The problem has been partly mitigated in silicon CMOS by self-aligned silicides (salicides, 9), but an equivalent process has yet to be

discovered in compound semiconductors. In either case, future scaling requires elimination of surface Schottky barriers in a reliable and repeatable process flow.

We have seen that ideal scaling laws reveal significant challenges, particularly for dopant concentration, doping layer thickness, contact resistance, access resistance, and effective oxide thickness. Unfortunately, even these projections are too optimistic. A number of parasitics fail to scale well, if at all.

3. Non-scaling Parasitics

Non-channel Resistances. The total, external on-state resistance is

$$R_{\mathrm{TOT}}(\Omega) = \left(2\rho_{\mathrm{c}}/L_{pad} + 2R_{\mathrm{access}} + 2R_{\mathrm{sp}} + 2R_{\mathrm{if}} + R_{\mathrm{ch}}\right)/W \qquad [1]$$

R_{TOT} must remain constant while W decreases by $1/a$, and contact and access resistivities (per cross section area) must scale by $1/a^2$. The source extension and the channel may contribute to interface resistivity R_{if}, spreading resistivity R_{sp}, or both, depending on geometry. These presumably do not scale, though they might be reduced by improvements in doping geometry or growth technique. Access resistance may be decreased by making thicker doped regions in the source and drain, but this increases the relative contributions of spreading resistance and source starvation. We now examine the others in turn.

Source doping density must scale as a^2 to prevent source exhaustion. This increase in doping would seem to satisfy the bulk access resistivity scaling requirement since resistivity scales inversely with doping: $\rho=1/qn\mu$. However, doping also reduces mobility (μ) due to both carrier-carrier scattering and ionized impurity scattering. Thus the bulk resistivity in the source, ρ_{access}, scales more slowly than doping density $1/N_d$. Indeed, series parasitics already contribute nearly half the on-state resistance in 45 nm MOSFETs (10); R_{access} does not scale well.

Quantized conductance presents an additional, fundamental resistance. The contact resistance between an infinite contact and a thin channel is 12.9 kΩ per conducting mode in the channel, accounting for spin degeneracy. The number of conducting modes equals the number of half-wavelengths that can fit into the width. Assuming parabolic bands and degenerate carrier statistics, this leads to a fundamental contact resistivity of (11):

$$R_q W = \frac{\pi\hbar}{q^2}\frac{\lambda_F}{2} = \frac{\pi\hbar}{q^2}\sqrt{\frac{2}{n_s}} = (52\ \Omega - \mu\mathrm{m})\sqrt{\frac{10^{13}/\mathrm{cm}^2}{n_s}} \qquad [2]$$

Note that this resistivity scales only slowly with electron density ($n_s^{-1/2}$). Even if n_s scales as a^2 as above, this still leads to only a linear scaling of R_{sp} with a. It was shown above that resistivities must scale as $1/a^2$, so extrinsic contact and access resistivities must be reduced much faster than the simple scaling laws in order to keep the total resistance constant.

Channel Resistance R_{ch}. In very thin channels, interface roughness scattering becomes important. In the limit of infinite barrier quantum wells, roughness scattering causes mobility to drop as the sixth power of well width, $\mu \propto t_{qw}^6$ (12,13). This inhibits ballistic transport in scaled FETs. Finite barriers decrease the dependence somewhat (14), and high carrier densities also partially screen the carriers from changes in the local potential, improving mobility (15). Interface roughness can also scatter and concentrate LO phonons, reducing electron saturated drift velocity (16). As we shall see in Section 4, preliminary experiments with InGaAs quantum wells show approximately a second-order

dependence $\mu \propto t_{qw}^2$ at carrier densities near $1x10^{13}$ cm^{-3}. Thus even with screening, channel resistance is expected to rise sharply for channels below about 5 nm.

In addition to scattering along the length of the channel, backscattering at the start of the channel contributes to channel resistance. The average velocity in the channel at the source is reduced from the thermal velocity as follows (17):

$$\langle v(0) \rangle \approx \frac{1-r}{1+r} v_{th}, \text{ with backscattering coefficient } r \approx \frac{l}{l+\lambda} \qquad [3]$$

and $v(0)$ is the velocity at the start of the channel, v_{th} is the thermal velocity, λ is the mean free path between backscattering events, and $l \approx kT/qE$ is the distance over which the channel potential falls by kT/q, approximate in degenerate conditions. In the ballistic limit, r=0 and equation [3] reduces to $v(0)=v_{th}$. As mentioned above, roughness scattering rises sharply for thin channels, so λ decreases, increasing r and reducing $v(0)$. Thus source/channel backscattering does not scale.

Capacitance and Transconductance. Transconductance follows $g_m=v \cdot C_{g-ch}/L_g$, and L_g scales as $1/a$. So to maintain constant g_m, the gate-channel capacitance

$$\frac{1}{C_{g-ch}} = \frac{1}{C_{ox}} + \frac{1}{C_{semi}} + \frac{1}{C_{dos}} \qquad [4]$$

should decrease no faster than $1/a$. The oxide capacitance $C_{ox} \sim \varepsilon_{SiO2} \cdot L_g W/\text{EOT}$ imposes a scaling limit unless EOT can be scaled, otherwise C_{ox} falls faster than $1/a$. Similarly, the finite density of states defines an effective capacitance of

$$C_{DOS} = \frac{2\pi\hbar^2}{q^2 m^*} L_g W \qquad [5]$$

which imposes a scaling limit unless effective mass m^* can be scaled. Finally, because of quantum confinement, the lowest bound state energy level rises as $E_1 \approx \hbar^2\pi^2/2m^* t_{inv}^2$ in deep wells. This may increase E_1 to the point where carriers are poorly confined, especially if we wish to use valleys with small effective mass m^*. This is shown conceptually in Figure 2. Because the centroid of the wavefunction remains farther from the gate, $C_{semi} \sim \varepsilon_{chan} \cdot L_g W/c$ decreases faster than $1/a$. All three components of C_{g-ch} thus scale faster than $1/a$, so the transconductance g_m decreases.

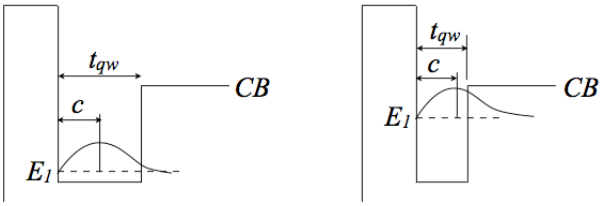

Figure 2. Conduction band diagram and wavefunction of lowest confined state. In narrow channels, the wavefunction penetrates the lower energy back barrier, so the distance from the oxide interface to the wavefunction centroid does not scale proportionally with quantum well thickness t_{qw}. Square wells are shown, but triangular wells show the same qualitative effect.

Delay. As mentioned in the introduction, gate oxides have lagged several generations of scaling. Even with high-k dielectrics, the effective oxide thickness (EOT=$t_{ox}\varepsilon_{SiO2}/\varepsilon_r$) cannot be scaled indefinitely. To preserve mobility and prevent interdiffusion, an interface layer appears to be necessary (18,19,20). Yet a single layer of these interface oxides contributes roughly 0.5 nm to EOT. If EOT cannot be reduced in successive scaling generations, then parasitic capacitances dominate. For example, the fringing capacitance $C_{fringing}/W \sim \varepsilon_r \varepsilon_0$ is nearly constant with scaling, so to reduce the unloaded gate delay

$$\tau_g \sim C_{tot}V_{dd}/I_D \sim (C_{fringing} + C_{overlap} + C_{g-ch})/g_m$$
$$= C_{fringing}/g_m + C_{overlap}/g_m + L_g/v \qquad [6]$$

the transconductance per unit gate width g_m/W must be progressively increased. Given upper limits on either C_{dos} or C_{ox}, g_m/W cannot be increased. Similarly, short channel effects and lower limits on C_{ox} mean that L_g cannot be decreased. Therefore scaling τ_g depends on increasing the carrier velocity v.

Carrier density. HEMTs present additional difficulties for scaling because the higher current density requires higher n_s. The higher n_s populates higher energy states in the well, which increases leakage through the top heterojunction barrier. Thus additional gate dielectrics may be needed above the top barrier. (See Ref. 23 and references therein for examples.)

To summarize the difficulties with scaling, increasing FET bandwidth requires more than simple dimensional shrinkage. Contact and access resistivities (per unit area) and doping must scale faster than $1/a^2$, and in ever-shallower junction depths. Quantized conductance and source-channel backscattering scale poorly at best, and interface scattering gets worse with scaling. Effective oxide thickness is unlikely to be scaled much below 1 nm EOT. Because of all these difficulties, dramatic reductions in resistance and improved g_m must come soon from new materials and device geometries.

4. Growth and Fabrication Challenges

Self-aligned InGaAs MOSFETs are an excellent candidate to meet the above challenges. InGaAs has a higher electron velocity and mobility and smaller electron effective mass than Si, which increases the current density for the same geometry and voltage. The main goals for this work were to verify and maintain high mobility channels, reduce contact resistance, and create a self-aligned geometry for minimum access resistance. We now examine each of these in turn.

To verify that InGaAs mobility would persist in a narrow quantum well, the structure shown in Figure 3(a) was grown. Resulting Hall mobilities approached 1000 $cm^2V^{-1}s^{-1}$ at electron concentrations of $1x10^{13}$ cm^{-2} in 5 nm thick channels. However, there was some question whether these mobilities represented actual electrons in the channel, or whether electrons were traveling outside the channel. Simulations using nextnano3 software (21) assuming parabolic bands showed that the large majority of electrons were in the channel, as shown in the figure. Nonparabolic corrections would only increase this concentration. Furthermore, for conductors in parallel, the less resistive path dominates, suggesting that these high mobilities represented the channel. For further experimental verification, a bulk Hall sample consisting of 100nm InAlAs of the same bulk doping level with an InGaAs cap was grown. This sample showed a mobility of just 346 $cm^2V^{-1}s^{-1}$. Since channel mobilities are substantially higher than this value, as shown in Figure 3(b), we conclude that most of the electrons are not traveling in the InAlAs layer, but are confined to the channel.

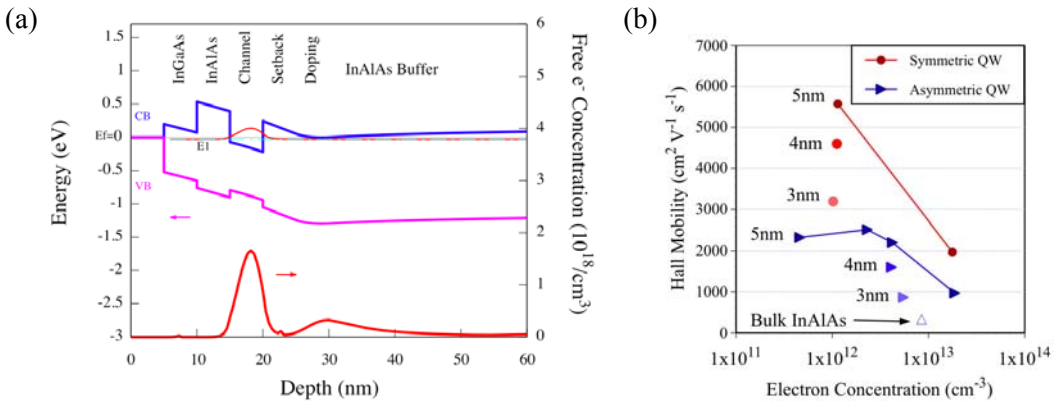

Figure 3. (a) Modeled band structure and electron concentration for 5 nm InGaAs channel with 5 nm doped InAlAs layer. (b) Hall mobilities at 300K for quantum well channels of various thickness, doped asymmetrically below (closed triangles, as in (a)) or symmetrically above and below (circles) the channel. Also shown is bulk InAlAs of equivalent doping (open triangle).

Like other III-V materials, InGaAs has been widely used in HEMTs, with mature processing techniques and high performance. Kim reported HEMTs with higher drive currents than 65 nm CMOS (with SiO_2) for the same gate leakage (22). But replacing CMOS with HEMTs will require attention to the difficulties above. In particular, the channel must be brought even closer to the gate to preserve g_m, from short channel effects, but the InAlAs barrier is already as leaky as SiO_2 at 0.5V drive voltages (22). Thus an additional, higher barrier would help further scaling, i.e. a MOSFET (23). In addition, injection through the wide bandgap top barrier adds 50 ohm-micron of access resistance, plus a total of 110 Ω-micron lateral resistance. These may be reduced by bringing the source/drain contacts closer to the gate and eliminating the barrier above the source/drain.

Contact resistance can be reduced several ways, particularly by increased doping (24). High doping also reduces source exhaustion and spreading resistance at the source extension. Molecular beam epitaxy (MBE) provides active doping levels of $n=5x10^{19}$ cm^{-3} (Si) and $p=1x10^{20}$ cm^{-3} (C) with no annealing required. These densities are higher than available by ion implantation and avoid the damage and crystal disordering from implantation as well. In addition, by evaporating Mo onto n^{++} doped InGaAs under UHV without breaking vacuum, non-annealed contacts with resistivities below $5x10^{-9}$ Ω-cm^2 have been demonstrated (25).

There are no known equivalents to salicides for III-V materials, but the same advantages can be achieved by a self-aligned regrowth and metal deposition. Figure 4(a) compares a typical geometry on a III-V FET with a self-aligned regrowth FET. While classical FETs rely on diffusion or implantation of dopants for contacts and/or lithographically defined contacts some distance from the gate, the regrown structure permits high concentrations of dopants in a shallow layer immediately touching the channel. Unlike conventional III-V metal contacts using liftoff processes, we used blanket deposition followed by a height-selective etch technique to define self-aligned source/drain metal (26). This further decreased the total access resistance since metal lies within nanometers of the channel, limited only by the thickness of the gate sidewalls.

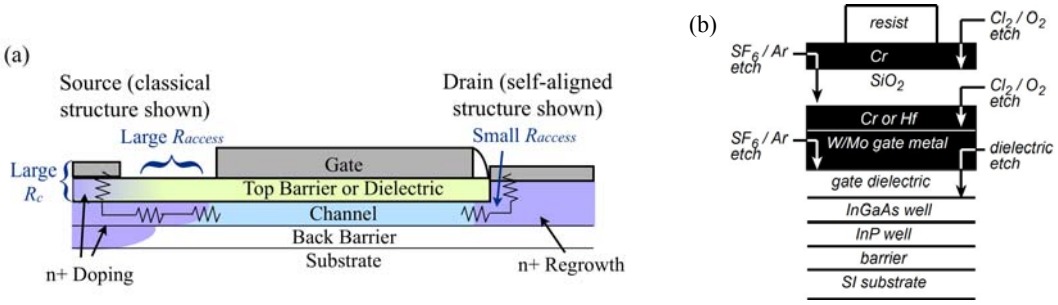

Figure 4. (a) Comparison between source/drain geometry on a classical III-V FET (details may vary) and a self-aligned regrowth FET. Not to scale. (b) Alternating selective etches are used to define gate stack without damaging channel.

High quality regrowth by MBE requires a damage-free stating surface, adequate surface cleaning procedures, and proper growth techniques. The channel was protected from oxidation by a 130 nm arsenic cap and shipped under vacuum for high-k deposition. A gate-first process that applied the ALD Al2O3 dielectric before any processing prevented damage during gate fabrication. Progressively gentler, alternating selective etches were used to approach the surface without damage, as shown in Figure 4(b). After the gate metal was defined, the gate was encapsulated in SiO_2 and SiN_x sidewalls to protect the MBE chamber from contamination and promote selective-area growth only in the source-drain regions. The high-k dielectric was then etched to expose the source and drain, and the channel was wet etched to undercut the sidewalls for regrowth. To clean the exposed semiconductor for regrowth, the wafer was exposed to UV ozone for 30 minutes followed by a 60s 1:10 HCl:H_2O oxide removal dip and 60s DI water rinse, then immediately loaded into UHV. After baking at 200 °C overnight, the wafer was cleaned using thermally cracked hydrogen at $1x10^{-6}$ Torr for 30 minutes at 380-420 °C to remove remaining surface oxides.

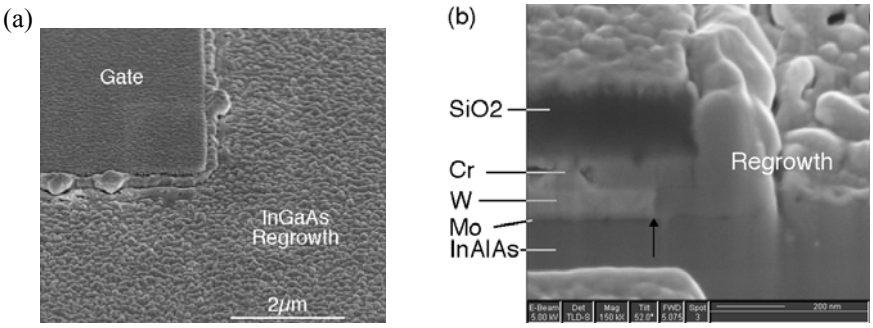

Figure 5(a) Scanning electron micrograph (SEM) of high temperature regrowth, showing preferential growth along base of gate but polycrystalline growth elsewhere. (b) Focused ion beam cross section through similar regrowth gate on deeply undercut gate with exposed metal. Note extent of backfilling by regrowth under the SiO_2 gate overhang (arrow).

MBE regrowth can exhibit moats or gaps near surface features, or else preferential growth along gates at high temperatures or if gate metal is exposed, as shown in Figure 5. On the other hand, a variation of migration enhanced epitaxy (MEE) produced smooth films with nearly perfect interfaces, with no apparent gaps next to the gate, as shown in Figure 6. For these growths, a small, continuous As_2 flux was used to prevent InGaAs decomposition at the high growth temperatures used, 540-560 °C. The In, Ga, and Si shutters were simultaneously opened for 1 monolayer deposition, followed by a 10-15 second soak under As_2 flux. The As flux was low enough to permit a Group III rich 4x2

reconstruction to appear in reflection high energy electron diffraction (RHEED) during each InGaAs pulse, but high enough that the reconstruction switched to an As-rich 3x1 or 2x4 pattern during the arsenic soak. Because In desorbs at these temperatures, 10-34% excess In was deposited with each cycle, calibrated by x-ray diffraction of similar samples grown at the same temperature. The high temperature and low arsenic flux permitted a very high Group III surface diffusion length. Typical As_2 fluxes ranged from $8x10^{-7}$ Torr for growth at 490°C to $2x10^{-6}$ Torr for growth at 540°C.

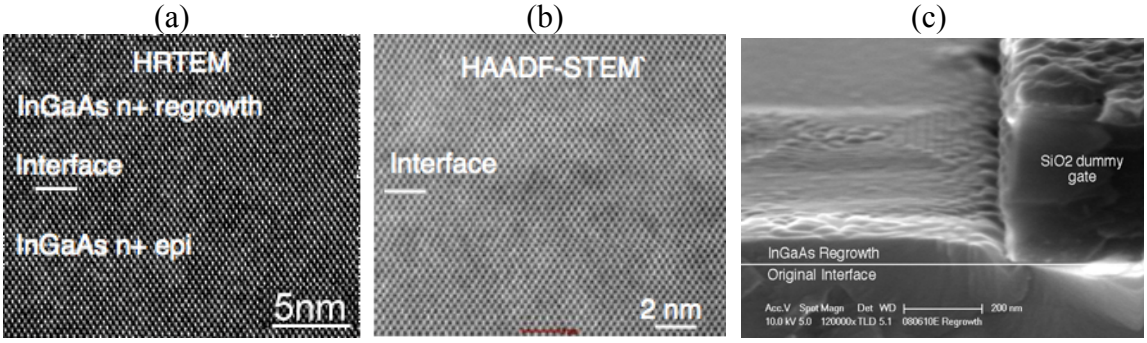

Figure 6 Regrowth. (a) High resolution transmission electron microscopy (HRTEM) of regrowth interface. (b) Chemically sensitive high angle annular dark field scanning TEM. (c) SEM of cleaved face of an SiO_2 dummy gate, showing smooth, continuous regrowth up to the gate.

At 0.7V gate overdrive, we can expect up to $n=1x10^{13}$ cm^{-2} in the absence of traps. As mentioned in Figure 3(b), Hall samples show μ~1000 $cm^2V^{-1}s^{-1}$, ~80x lower than bulk InGaAs, but this represents only a 10% penalty in I_d (27). Carrier velocity is more relevant than mobility in channels, and the small m^* produces higher injection velocities, whether thermal or saturated drift velocity. On the other hand, high mobility is relevant and advantageous in the source/drain access regions.

A major advantage of InGaAs is its small m_e^*, which increases the injection velocity of electrons. The small m^* also means a smaller density of states, but this is somewhat offset by nonparabolic bands. Taken together, $m^*/m_0=0.1$ turns out to be an optimal effective mass for EOT=0.4-1.0 nm (27).

Finally, InGaAs growth is compatible with Si. High quality III-V's are already being grown on Si and Ge ([28],[29],32). Ge buffer layers from methylated precursor gases eliminate strain from lattice mismatch at room temperature (30,31). The buffer layers need only be tens of nm thick to reach $<10^{-6}$ cm^{-2} dislocation densities. Pre-annealing the Si or Ge surface before III-V growth to form double-height atomic steps eliminates antiphase domains, so vicinal substrates are not required (32). Also, the Ge buffer provides a high mobility channel for PMOS devices on the same wafer (30,33). InGaAs FETs maintain the same planar structure as conventional CMOS, greatly reducing capital investment and risk, while providing even greater benefits in new structures such as FinFETs.

5. InGaAs Depletion-Mode MOSFETs

Prototype InGaAs MOSFETs were fabricated using the above techniques. The epitaxial structure up through the InGaAs channel shown in Figure 7(a) was grown by Intelligent Epitaxy Technology and capped with arsenic. The wafer was transported to an ALD chamber, where the arsenic cap was desorbed and 5nm of Al_2O_3 was deposited. The wafer was unloaded, and the gate stack and SiN_x sidewalls were patterned. The exposed

Al_2O_3 over the source and drain was etched in dilute KOH developer, then the InGaAs channel was etched to provide a recess under the sidewalls. Then the wafer was HCl cleaned and loaded into an MBE chamber for regrowth. After hydrogen cleaning, 50 nm of n++ InGaAs were grown by MEE, followed by 20 nm Mo.

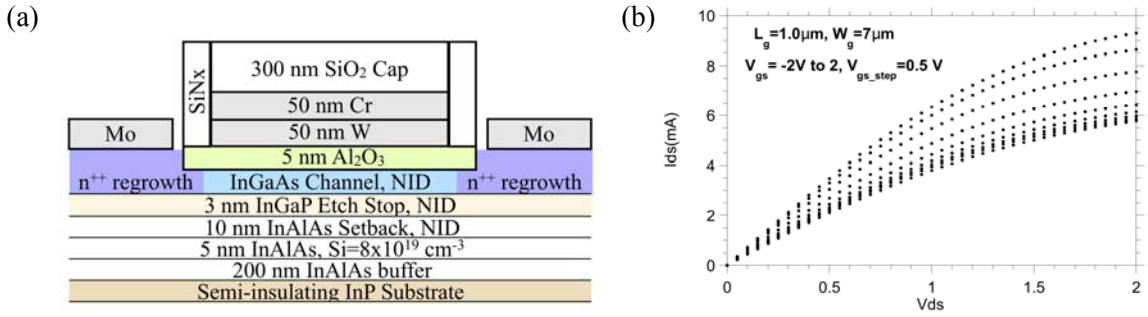

Figure 7. (a) Layer structure of depletion mode, self-aligned Al_2O_3/InGaAs MOSFET. Not to scale. (b) Drain current vs. V_{ds} for -2<Vgs<2.

Figure 8 shows a typical MOSFET after regrowth and Mo deposition. Roughness in the far field was from Mo grains; no large-scale roughening from strain relaxation was apparent. Although the oblique view SEM appears to show a gap, this was merely the shadow of the edges of the gate in the top surface Mo. Cleaved views (not shown) do not show gaps in the underlying InGaAs. In fact, the appearance of a gap indicates that the SiNx sidewalls are intact, covering the gate metal in the region below the 300-400 nm SiO_2 cap. If the sidewalls were porous or missing, then amorphous growth would have nucleated along the sides of the gate as in Figure 5.

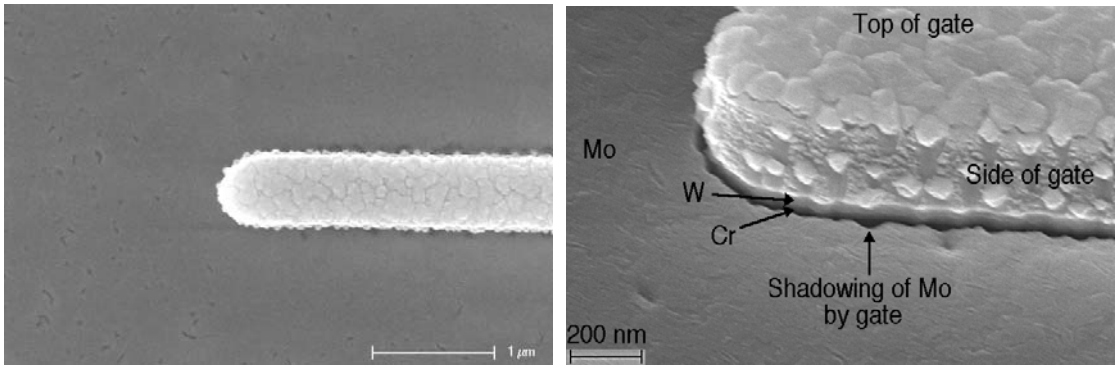

Figure 8. SEM of MOSFET after regrowth and Mo source/drain contact metal but before isolation etch and contact pads. (left) Top view. (right) Oblique view.

One challenge with most III-V devices is that surface and interface states tend to pin the Fermi level midgap, depleting carriers. Even a small number of traps under the high-k would deplete the source under the sidewalls if the sidewalls were too thick. To assure that the MOSFETs could be turned on regardless of surface depletion, and to prevent ionized impurity scattering, high doping was used with a setback layer significantly larger than necessary. As a result, these MOSFETs could not be turned off, as shown in Figure 7(b). Equivalent Hall samples of the same doping, [Si]=$8x10^{19}$ cm^{-3}, produced n=$3.6x10^{19}$ cm^{-3} in bulk and n_s=$1x10^{13}$ cm^{-2} sheet carrier concentrations in quantum wells. Enhancement-mode devices will be reported elsewhere (34).

The total parasitic resistance can be estimated by studying devices at high gate voltages. As figure 9(a) shows, channel resistance ceases to be a function of Lg below

1µm, indicating a large series resistance. The resistance at V_{ds}=0.4V vs. L_g is plotted in Figure 9(b), indicating that at L_g=0, a large series resistivity of 0.71 kΩ-µm remains. Possible causes include poor vertical regrowth interface, surface depletion in the recess under the lip of the gate, polycrystalline nucleation at the base of the gate, incomplete underfill (gaps) under the gate, lack of dopant atoms near the gate, or Fermi level pinning due to local strain-induced dislocations near the gate, caused by In-rich conditions. Experiments to determine the cause of this high resistance are currently underway.

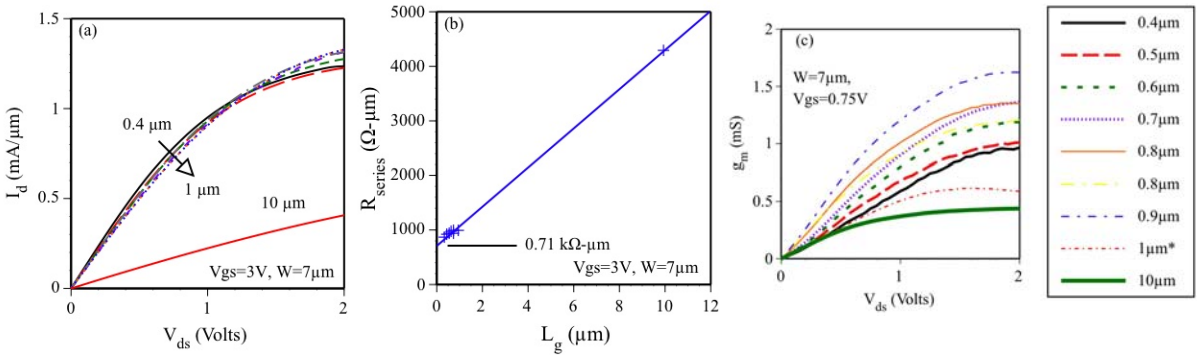

Figure 9. (a) Id vs. Vds at Vgs=3V, with Lg=0.4, 0.5, 0.6, 0.7, 0.8, 1, and 10 µm. (b) Series resistance vs. gate length. (c) Transconductance vs. V_{ds} at V_{gs}=0.75V.

Despite the high series resistance, the transconductance of these transistors was as high as 1.7 mS (or 0.24 mS/µm) for 0.9 µm long devices at Vds=2V and Vgs=1V. Many devices show comparable gm, as shown in Figure 9(c). These results are promising for future MOSFETs.

6. Conclusion

Self-aligned InGaAs MOSFETs are a promising replacement for Si n-channels as CMOS reaches fundamental scaling limits. If high-k gate dielectrics can be incorporated, then surface-channel InGaAs MOSFETs should have advantages in current density and low parasitics. Basic scaling laws show that contact and access resistances must be reduced as the square of scaling factor. Several parasitics do not scale, requiring an even greater improvement in resistance and transconductance to continue scaling. This requires not only new, high electron velocity materials such as InGaAs but also self-aligned regrowth geometries that have never been used in III-V FETs to date. These techniques have led to the first scalable designs for III-V MOSFETs. Depletion-mode InGaAs MOSFETs demonstrated current densities above 1mA/µm and peak transconductance of 0.24 mS/µm.

Acknowledgments

A portion of this work was performed in the UCSB nanofabrication facility, part of the NSF funded NNIN network. This work also made use of MRL Central Facilities supported by the MRSEC Program of the National Science Foundation under award No. DMR05-20415.

References

1. A. Delabie, D. P. Brunco, T. Conard, P. Favia, H. Bender, A. Franquet, S. Sioncke, W. Vandervorst, et al. *J. Electrochem. Soc.* **155**, H937-H944 (2008).

2. C.-W. Cheng and E. A. Fitzgerald, *Appl. Phys. Lett.*, **93**, 031902 (2008).
3. R. J. W. Hill, R. Droopad, D. A. J. Moran, X. Li, H. Zhou, D. Macintyre, S. Thoms, O. Ignatova, A. Asenov, K. Rajagopalan, P. Fejes, et al., *Electron. Lett.*, **44**, (2008).
4. R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, *IEEE J. Solid State Circuits*, **SC-9**, 256 (1974).
5. D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, *Proc. IEEE*, **89**, 259 (2001).
6. Y. Awano, M. Kosugi, K. Kosemura, T. Mimura, and M. Abe, *IEEE Trans. Electron Devices*, **36**, 2260, (1989).
7. G. H. Jessen, R. C. Fitch, J. K. Gillespie, G. Via, A. Crespo, D. Langley, D. J. Denninghoff, M. Trejo, and E. R. Heller, *IEEE Trans. Electron Devices*, **54**, 2589, (2007).
8. K. Uchida et al., *IEDM Technical Digest*, p. 47-50 (2002)
9. J. P. Gambino, E. G. Colgan, *Mater. Chem. and Phys.*, **52**, 99 (1998)
10. Arghavani, A.M. Noori, A. Gelatos, A. Dkandelwal, S. Gandikota, S. Felch, and S. E. Thompson, in *Semiconductor Fabtech*, 35th Edition (2007).
11. S. Datta, F. Assad, M. S. Lundstrom, *Superlattices and Microstructures*, **23**, 771 (1998).
12. A. Gold., *Solid State Commun.,* **60**, 531 (1986).
13. A. Gold, *Phys. Rev. B,* **35**, 723 (1987).
14. U Penner *et al* 1998 *Semicond. Sci. Technol.* **13** 709-713 (1998).
15. R. Gottinger, A. Gold, G. Abstreiter, G. Weimann, and W. Schlapp, *Europhys. Lett.,* **6,** 183 (1988).
16. R. Gupta, N. Balkan, and B. K. Ridley, *Phys. Rev. B* **46**, 7745 (1992).
17. M. Lundstrom, IEEE Trans. Electron. Devices, **49**, 133 (2002).
18. T. D. Lin, H. C. Chiu, P. Chang, L. T. Tung, C. P. Chen, M. Hong, J. Kwo, W. Tsai, Y. C. Wang, *Appl. Phys. Lett.*, **93**, 033516 (2008).
19. D. A. Buchanan, *IBM J. Res. and Dev.*, **43**, 245 (1999).
20. Gutowski M, Jaffe JE, Liu CL, et al., *Appl. Phys. Lett.,* **80**, 1897 (2002)
21. [Online] http://www.nextnano.de/nextnano3
22. D.-H. Kim and J. A. del Alamo, *IEEE Trans. Electron* Devices, **55**, 2546 (2008).
23. P. J. Hansen, L. Shen, Y. Wu, A. Stonas, Y. Terao, S. Heikman, D. Buttari, T. R. Taylor, S. P. DenBaars,, U. K. Mishra, R. A. York, and J. S. Speck, *J. Vac. Sci. Technol. B*, **22**, 2479 (2004).
24. A. K. Baraskar, M. A. Wistey, V. Jain, U. Singisetti, G. Burek, B. Thibeault, A. Gossard, and M. Rodwell, 36th Conference on the Physics and Chemistry of Surfaces and Interfaces, Santa Barbara, California, USA (2009).
25. A. K. Baraskar, M. A. Wistey,, V. Jain, U. Singisetti, G. Burek, B. J. Thibeault, Y.-J. Lee, A. C. Gossard, and M. J. W. Rodwell, submitted to 2009 Electronic Materials Conference.
26. G. J. Burek, M. A. Wistey, U. Singisetti, A. Nelson, B. J. Thibeault, S. R. Bank, M. J. W. Rodwell, and A. C. Gossard, *J. Cryst. Growth* (in press).
27. M. Rodwell, M. Wistey, U. Singisetti, G. Burek, A. Gossard, S. Stemmer, R. Engel-Herbert, Y. Hwang, Y. Zheng, C. Van de Walle, et al., *IEEE 2008 International Conf. Indium Phosphide and Related Materials*, (2008).
28. L. A. Eyres, P. J. Tourreau, T. J. Pinguet, C. B> Ebert, J. S. Harris, M. M. Fejer, L. Becouarn, B. Gerard, E. Lallier, *Appl. Phys. Lett.*, **79**, 904 (2001).
29. K. Chilukuri, M. J. Mori, C. L. Dohrman, E. A. Fitzgerald, *Semicond. Sci. Technol.*, **22**, 29 (2007).
30. D. Q. Kelly, I. Wiedmann, J. P. Donnelly, S. V. Joshi, S. Dey, and S. K. Banerjee, D. I. Garcia-Gutierrez and M. José-Yacamán, *Appl. Phys. Lett.*, **88**, 152101 (2006).
31. M. A. Wistey, Y-Y. Fang, J. Tolle, A. V. G. Chizmeshya, and J. Kouvetakis, *Appl. Phys. Lett.*, **90**, 082108 (2007).
32. I. Németh, B. Kunert, W. Stolz, K. Volz, *J. Cryst. Growth*, **310**, 1595 (2008).
33. H. Shang, H. Okorn-Schmidt, K. K. Chan, M. Copel, J. A. Ott, P. M. Kozlowski, S. E. Steen, S. A. Cordes, H.-S. P. Wong, E. C. Jones, W. E. Haensch, *IEEE International Electron Devices Meeting 2002*, 441 (2002).
34. U. Singisetti, et al., submitted to 2009 Electronic Materials Conference