III-V FET Channel Designs for High Current Densities and Thin Inversion Layers

Mark Rodwell¹, W. Frensley², S. Steiger⁴, E. Chagarov³, S. Lee⁴, H. Ryu⁴,

Y. Tan⁴, G. Hegde⁴, L Wang³, J. Law¹, T. Boykin⁵, G. Klimek⁴, P. Asbeck³, A. Kummel³, J. N. Schulman⁶

¹ECE Department, University of California, Santa Barbara, 93106, USA, ²University of Texas, Dallas,

University of California, San Diego, ⁴Purdue University, ⁵University of Alabama, Huntsville, ⁶The Aerospace Corporation, El Segundo

phone: 805-893-3244, fax 805-893-3262, rodwell@ece.ucsb.edu

III-V FETs are being developed for potential application in 0.3-3 THz systems and VLSI. To increase bandwidth, we must increase the drive current $I_d = qn_s v_{snj} W_g$ per unit gate width W_g , requiring both high sheet carrier concentrations n_s and high injection velocities v_{mj} . Present III-V NFETs restrict control region transport to the single isotropic Γ band minimum. As the gate dielectric is thinned, I_d becomes limited by the effective mass m^* , and is only increased by using materials with increased m^* and hence increased transit times.¹ The deep wavefunction also makes Γ -valley transport in low- m^* materials unsuitable for < 22-nm gate length (L_g) FETs. Yet, the L-valleys in many III-V materials² have very low transverse m_i and very high longitudinal mass m_i . L-valley bound state energies depend upon orientation, and the directions of confinement, growth, and transport can be chosen to selectively populate valleys having low mass in the transport direction^{3,4}. The high perpendicular mass permits placement of multiple quantum wells spaced by a few nm, or population of multiple states of a thicker well spaced by ~10-100 meV. Using combinations of Γ and L valleys, n_s can be increased, m^* kept low, and vertical confinement improved, key requirements for <20-nm L_g III-V FETs.

Consider FET scaling. To increase bandwidth γ : I, capacitances and transit delays must be reduced γ : 1 while maintaining constant voltages, currents, and resistances. In short- L_g FETs, the gate-source $C_{gs,f} \propto \varepsilon W_g$ and gate-drain $C_{gd} \propto \varepsilon W_g$ fringing capacitances are a significant fraction of the total capacitance, and limit $f_\tau \cdot C_{gs,f}$ and C_{gd} are only weakly dependent on lateral geometry, hence the $(C_{gs,f} + C_{gd})\Delta V/I_d$ delay is reduced γ : 1 only if I_d/W_g is increased γ : 1. I_d is determined by the sheet carrier concentration $n_s = (C_{g-ch}/L_gW_g)(V_{gs} - V_h)/q$, where the gate-channel capacitance $C_{g-ch} = [1/C_{ox} + 1/C_{deph} + 1/C_{DOS}]^{-1}$ is the series combination of dielectric $C_{ax} = \varepsilon_{SO_2}L_gW_g/T_{eq}$, wavefunction depth $C_{deph} = \varepsilon_{sont}L_gW_g/T_{mv}$ (T_{inv} is the wavefunction mean depth) and density of states $C_{dos} = q^2 \cdot dn_s/dE_f$ capacitances. In the ballistic case, $C_{dos} = q^2g(m_{\parallel}m_{\perp})^{1/2}L_gW_g/2\pi\hbar^2$, where g is the # of populated valleys, and m_{\parallel} and m_{\perp} the effective masses parallel and perpendicular to transport; near equilibrium, C_{dos} is 2:1 larger. Given ballistic transport⁵ and $E_f - E_{well} \gg kT$, $v_{inj} = (4/3\pi)(2(E_f - E_{well})/m_{\parallel})^{1/2}$. We scale by maintaining constant v_{inj} while reducing C_{ox}/L_gW_g , C_{deph}/L_gW_g , and C_{dos}/L_gW_g by γ :1 This requires fixed transport mass m_{\parallel} , T_{eq} and T_{inv} reduced γ :1, and C_{dos} increased γ :1 by increasing the # of valleys or the perpendicular mass.

At a given dielectric thickness T_{equiv} , there is an optimum m^* maximizing I_d . We find $I_d/W_g = J_0 \cdot K_1 \cdot ((V_{gs} - V_h)/1V)^{3/2}$, where $J_0 = (4/3\pi)(2q/m_0)^{1/2}(1V)^{3/2}(q^2m_0/2\pi\hbar^2) = 84$ mA/ μ m and $K_1 = n \cdot (m_{\perp}/m_0)^{1/2}(1 + (C_{doso}/C_{equiv}) \cdot g \cdot (m_{\perp}^{1/2}m_0)^{1/2}/m_0))^{-3/2}$ is the normalized current density. $C_{equiv} = [1/C_{ox} + 1/C_{inv}]^{-1}$ is C_{deph} and C_{ox} in series, while $C_{doso} = q^2m_0L_gW_g/2\pi\hbar^2$. The transit time $\tau_{ch} = Q_{ch}/I_d = K_2(L_g/v_0)(1V/(V_{gs} - V_h))^{1/2}$, where the normalized delay is $K_2 = (m_{\parallel}/m_0)^{1/2}(1 + C_{doso}g(m_{\perp}m_{\parallel})^{1/2}/m_0C_{equiv})^{1/2}$ and $v_0 = (4/3\pi) \cdot (2q \cdot (1 \text{ Volt})/m_0)^{1/2}$ = 2.5 · 10⁷ cm/s. Given one isotropic valley ($m_{\perp} = m_{\parallel} = m^*$, g = 1) and 1 nm total equivalent dielectric thickness EOT (i.e. $C_{equiv} = \varepsilon_{r,SiO_2}L_gW_g/(1 \text{ nm})$), highest current is obtained for $m^*/m_0 = 0.05$, while for 0.3 nm EOT, peak I_d is obtained at $m^*/m_0 = 0.2$; given one isotropic valley, low m^* gives low I_d in nm FETs¹, though low m^* reduces τ_{ch} for any EOT. Note that for Si {100} FETs^6, $m^*/m_0 = 0.19$ and g = 2.

Given a {100}-oriented FET and Γ -valley transport, the vertical confinement mass m_q and transport mass m_{\parallel} are equal. For good FET electrostatics, T_{inv} must be less than $\sim L_g/10$, and $E_{well} - E_{\Gamma} \approx \pi^2 \hbar^2 / 2m_q T_{well}^2 \approx \pi^2 \hbar^2 / 8m_q T_{inv}^2$, and is large for small L_g . Noting that $E_F - E_{well} = n_s (dn_s / dE_f)^{-1} = n_s (2\pi \hbar^2 / gm_{\perp}^{1/2} m_{\parallel}^{1/2})$ also is large, once $(E_F - E_{\Gamma}) \ge (E_L - E_{\Gamma})$ the L valleys populate and v_{inj} decreases. High n_s cannot be obtained in the Γ valleys in low- m^* materials, nor is $L_g < 20$ nm feasible; such FETs can no longer scale.

Thin wells with high carrier concentrations show increased transport mass from nonparabolic bands. Assuming $E(1 + \alpha E) = \hbar^2 k^2 / 2m^*$, the effective mass $m^* = m^* (1 + 2\alpha (E_{well} - E_{\Gamma}))$ of the bound state increases from confinement, and then further increases from the channel charge density. Further, with thin wells, and low m_q , the wavefunction extends into the barriers, further increasing the in-plane masses m_{\perp} and m_{\parallel} .

Consider a 3 nm (100) GaAs well with strained AISb barriers. The L bound states lie 177 meV above that of Γ . Equilibrium (not ballistic transport) analysis uses Schrödinger-Poisson, the effective mass approximation, and parabolic bands. 0.66 nm Al₂O₃ and 0.34 nm AISb lie between the well and gate, giving $T_{eq} = 0.37$ nm. Under strong inversion $C_{g-ch}/L_gW_g \cong 2.4 \ \mu F/cm^2$, far below $C_{ox}/L_gW_g = 9 \ \mu F/cm^2$, and the high-mass L-valleys fill for $n_s > 2.4 \cdot 10^{12} \text{ cm}^{-2}$. Under ballistic transport, C_{dos} and the maximum n_s would both decrease 2:1.

Increased C_{dos} and low m_{\parallel} can be obtained by using L valley minima alone or combined with the Γ valley. The InGaAs, GaAs, and GaSb L-valleys have low m_{i}/m_{o} (0.062-0.1) and high m_{i}/m_{o} (1.23-1.9). The L-valleys have $\langle 111 \rangle$ orientations, and transport in a (100) channel includes contributions from the high m_{i} . Using instead a (111)

wafer, the L[111] valley is oriented vertically, and shows low transport masses $(m_{\parallel} = m_{\perp} = m_i)$ and high confinement mass $(m_q = m_i)$. The L[111], [111], and [111] minima show high⁶ transport mass $(m_i + 8m_i)/9$ in one in-plane direction, but low confinement mass⁶ $m_q = 9m_i m_i/(m_i + 8m_i)$. The X valleys have $\langle 100 \rangle$ orientations, in bulk InGaAs, GaAs, and GaSb have minima well above Γ and L, and in a (111) well have low $m_q = 3m_i m_i/(m_i + 2m_i)$ quantization mass. In appropriate thin wells, the X and L[111], [111], and [111] quantized states are driven to high energies and depopulated. T_{well} can be selected to place Γ and L[111] at similar energies, doubling C_{dos} , or Γ driven in energy above L [111], and transport provided in multiple L [111] valleys.

Consider a 2.3 nm (111) GaAs well with strained AlSb barriers. m_q is large, thus the first two L [111] states are separated by only 84 meV. The Γ state lies 41 meV above the lower L [111] state; 3 valleys are populated over a 300mV range of V_{gs} . L [111], [111], [111], and [111] and X lie 175 and 288 meV above the lower L [111] state. In equilibrium simulation $n_s = 7 \cdot 10^{12}$ cm⁻² with $V_{gs} - V_{fh} = 300$ mV, and moderately higher n_s does not populate heavy valleys. In inversion, $C_{g-ch} / L_g W_g \cong 4 \ \mu F/cm^2$. The benefit over the (100) design is larger in the ballistic case.

In InGaAs, GaAs, and GaSb, the L-valley m_i is >25:1 larger than the Γ -valley mass, hence T_{wed} can be made 5:1 smaller for a given quantization energy. m_q is high in the barriers, hence multiple wells can be placed between ~ 1 nm barriers without significant well coupling hence energy redistribution. Multiple L[111] quantum wells can stacked to increase g hence C_{dos} . Consider a FET with two 0.66 nm (2 ML) (100) GaAs wells separated by strained 0.66 nm AlSb barriers. Given zero field, the two L[111] states split in energy by < 40 meV; for $V_{gs} - V_h = 300 \text{ mV}$ the separation is 56 meV. L[111], [11], and [111] and X lie 322 and 346 meV above the lower L[111] state. The Γ state is driven to high energy. In equilibrium, n_s is driven to 7.8 $\cdot 10^{12} \text{ cm}^{-2}$ with $V_{gs} - V_h \sim 300$ mV; moderately higher n_s does not populate heavy valleys. $C_{g-ch} / L_g W_g \cong 4 \ \mu \text{F/cm}^2$. The advantage over Γ {100} is greater for ballistic transport. A triple- well L[111] design gives similar results. In these FETs, the upper wells charge most strongly because of charge division between the wells' C_{dos} and the well-well capacitance $C_{well} = \epsilon L_g W_g / T_{pikh}$, where T_{pikh} is the well pitch. With thin wells, and low m_i , C_{dos} can be increased 1.5:1 to 2.2:1. In these FETs, the X and L[111], [11], and [111] valleys play a similar role to the L and X valleys in {100}

In these FE1s, the X and L[111], [111], and [111] valleys play a similar role to the L and X valleys in {100} Γ -transport FETs. For high v_{m_j} it is sufficient to avoid populating high- m_{\parallel} valleys in the low-field control region; scattering into these valleys in the drain high-field region increases the drain delay but does not reduce I_d .

High I_d/W_g can be obtained from a FINFET with multiple (111) wells forming an array with many conduction channels per unit width. Because the L[111] vertical mass m_q is large, the well pitch can be made small (2-5 nm) while maintaining small well coupling hence small vertical eigenstate energy splitting. Transport is thus 1-D with each well having L[111] valley minima energies $E_n = \hbar^2 \pi^2 n^2 / 2m_l T_{pitch}^2$ and per-state current $I_n = gq(E_f - E_n)/\pi\hbar$ and charge $Q_n = qg(2m^*(E_f - E_n))^{1/2}L_g/\pi\hbar$. From Q_n and the specified EOT, the dielectric voltage drop is found and $I_d(V_{gs})$ computed. At $V_{gs} - V_{th} = 0.3$ V, $T_{pitch} = 2.5$ nm, $m_l/m_0 = 0.06$, and $C_{equile} = \varepsilon_{r,SiO_2}L_gW_g/(0.3 \text{ nm})$, we find $I_d/W_g \sim 4$ mA/ μ am per gate. We assume a P-doped fin with separate inversion layers for each gate. Inversion layer confinement raises the L[111] states' energies relative to other minima, hence the inversion layer should be thicker than T_{pitch} and the fins oriented as [011] for the most rapid increase in L [111] and [111] energies with lateral confinement. L[111] has low m_{\parallel} and is populated without degrading v_{m_l} .

A final design uses a planar (110)-oriented channel, with the L[11] and [111] minima projecting into the transport plane with a low transport mass $m_{\parallel} = m_i$ in the [110] direction, a high perpendicular $m_{\perp} = (m_i + 2m_i)/3$ in the [001] direction, and moderate $m_q = 3m_i m_i/(m_i + 2m_i) \approx 3m_i$. The L[111] and [111] minima have higher $3m_i m_i/(m_i + 2m_i)$ mass in the [110] direction, and have low $m_q = m_i$. Quantization favors population of the desired [111], [111] minima, but the separation in energy of the undesired valleys is only moderate.

Several designs above use very thin wells and barriers. We must determine whether such layers can be grown and whether mobility is acceptable. We must refine the energy calculations. 2-4 ML GaSb and InAs wells^{7,8} have been grown. Preliminary tightbinding calculations using an sp3d5s* basis⁹ conducted for triple 1.1nm GaSb wells with 1.1nm AlSb barriers confirm the symmetry of the lowest state manifold and its expected transverse dispersion. Excited states are slightly lower than predicted by effective mass, but the design still appears viable.

We acknowledge extensive discussions with H. Kroemer, A. C. Gossard, B. Brar, and J. Albrecht. J.N.S. was supported by the Internal Research & Development program of The Aerospace Corporation.

¹ P. M. Solomon, S. E. Laux, 2001 IEEE IEDM, 2-5 Dec., Technical Digest, pp. 5.1.1 - 5.1.4

² I. Vurgaftman, J. R. Meyer, L. R. Ram-Mohan, J. Appl. Phys, Vol. 89, 11, pp 5815-5875.

³ S.E. Laux, 2004 IEDM, Technical Digest, December, pp.135 - 138

⁴ A. Rahman, G. Klimeck, T. B. Boykin, M. Lundstrom, 2004 IEDM, Dec. 13-15, Tech. digest pp. 139 - 142

⁵ K. Natori, J. Appl. Phys., vol. 76, no. 8, pp. 4879–4890, 1994.

- ⁸ T. B. Boykin, Appl. Phys. Lett. 64 (12), 21 March 1994, pp. 1529-1531
- ⁹ G Klimeck, *et al*, IEEE Transactions on Electron Devices, Vol. 54, 2079 (2007)

⁶ F. Stern, W. E. Howard, Physical Review, Vol. 163, No. 3, 15 Nov 1967, pp. 816-835

⁷ B. Brar, et al, Appl. Phys. Lett. 62 (25), pp 3303-3305, also: J. of Crystal Growth 127 (1993) 752-754

Table 1: Constant-voltage / constant-velocity FET scaling laws: changes required for γ : 1 increased bandwidth in an arbitrary circuit



parameter	law	parameter	law
gate length L_s , source-drain contact lengths $L_{S/D}$ (nm)		electron density	γ^1
		$n_{s} = (C_{g-ch} / L_{g}W_{g})(V_{gs} - V_{h}) / q (\mathrm{cm}^{-2})$	
gate width W_g (nm)	γ ⁻¹	injection velocity (m/s)	γ°
equivalent oxide thickness		$v_{injection} = (4/3\pi)(2qC_{g-ch}(V_{gs} - V_{in})/m_{\parallel}C_{dos})^{1/2}$	
$T_{_{eq}} = T_{_{ox}} \varepsilon_{_{SiO_2}} / \varepsilon_{_{oxide}} (\text{nm})$		drain current $I_d = qn_s v_{injection}$ (mA)	γ^{0}
dielectric capacitance	γ^{-1}	drain current density I_d / W_g (mA/ μ m)	γ^1
$C_{ox} = \varepsilon_{SiO_2} L_g W_g / T_{eq} (\mathbf{fF})$		transconductance $\partial I_d / \partial V_{gs}$ (mS)	γ°
wavefunction mean depth T_{inv} (nm)	γ^{-1}	gate-source, gate-drain fringing capacitances	γ^{-1}
wavefunction depth capacitance	γ^{-1}	$C_{gs,f} \propto \mathcal{E}W_g$, $C_{gd} \propto \mathcal{E}W_g$ (fF)	
$C_{depth} = \varepsilon_{semi} L_g W_g / T_{inv} (\mathbf{fF})$		S/D access resistances R_s , $R_d(\Omega)$	γ^{0}
DOS capacitance (ballistic case)	γ^{-1}	S/D access resistivities $R_s W_g$, $R_d W_g$ ($\Omega - \mu m$)	γ^{-1}
$C_{dos} = q^2 g(m_{\parallel}^* m_{\perp}^*)^{1/2} L_g W_g / 2\pi \hbar^2 \text{ (fF)}$		S/D contact resistivities $\rho_c (\Omega - \mu m^2)$	γ-2
gate-channel capacitance	γ^{-1}	temperature rise (one device, K)	$\sim W_{o}^{-1}$
$C = [1/C + 1/C + 1/C]^{-1}(\mathbf{fF})$			<u> </u>





Figure 1: FET normalized drive current K_1 where $I_d/W_g = (84 \text{ mA}/\mu\text{m}) \cdot K_1 \cdot ((V_{gs} - V_{ft})/1\text{V})^{32}$, and g is the # of valley minima.





(100) orientation

 $\tau_{di} = Q_{di} / I_D = K_2 \cdot (L_g / 2.5 \cdot 10^7 \text{ cm/s}) \cdot (1 \text{V} / (V_{gs} - V_{di}))^{1/2}.$ EOT includes dielectric thickness and wavefunction depth.



		Γ valley	X valleys*			L valleys		
material	substrate	$m*/m_0$	m_l / m_0	m_t / m_0	$E_{X} - E_{\Gamma}$	m_l / m_0	m_t / m_0	$E_L - E_{\Gamma}$
In _{0.5} Ga _{0.5} As	InP	0.045	1.29	0.19	0.83 eV	1.23	0.062	0.47 eV
InAs	InP	0.026	1.13	0.16	0.87 eV	0.65	0.050	0.57 eV
GaAs	GaAs	0.067	1.3	0.22	0.47 eV	1.9	0.075	0.28 eV
GaSb	GaSb	0.039	1.51	0.22	0.30 eV	1.3	0.10	0.07 eV
Si	Si		0.92	0.19	(negative)	*Si minima at $\Delta \sim 0.85 \cdot \langle 100 \rangle$		



Figure 3. Simulation of Γ , Γ -L, and multiple-L valley FETs: quantized states, wavefunctions, charge density, and sheet carrier concentrations vs. bias. Well energies and charge densities calculated using the effective mass approximation and assuming parabolic bands. The gate dielectric is 0.3 nm Al_2O_3 . Well charge densities are computed assuming thermal equilibrium; in the ballistic limit, C_{does} is 2:1 smaller than in equilibrium, and multiple-valley FET channels provide a proportionally larger improvement in N_s . 0.66 nm is 2 monolayers.



Figure 4: Increase in density of states for 2-Lwell and 3-L-well FET channels as a function of carrier effective mass and well-well pitch.

Figure 5: Computed drain current drain current density, per gate, of a FINET with (111)-oriented multiple quantum wells as a function of carrier effective mass and well-well pitch.