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

Mark Rodwell[1], W. Frensley[2], S. Steiger[4], E. Chagarov[3], S. Lee[4], H. Ryu[4],
Y. Tan[4], G. Hegde[4], L Wang[3], J. Law[1], T. Boykin[5], G. Klimek[4], P. Asbeck[3], A. Kummel[3], J. N. Schulman[6]
[1]ECE Department, University of California, Santa Barbara, 93106, USA, [2]University of Texas, Dallas,
University of California, San Diego, [4]Purdue University, [5]University of Alabama, Huntsville, [6]The 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 $I_d = qn_s v_{inj} W_g$ per unit gate width W_g, requiring both high sheet carrier concentrations n_s and high injection velocities v_{inj}. 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 materials[2] have very low transverse m_t and very high longitudinal mass m_l. 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:1, 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_τ. $C_{gs,f}$ 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 $n_s = (C_{g-ch} / L_g W_g)(V_{gs} - V_{th})/q$, where the gate-channel capacitance $C_{g-ch} = [1/C_{ox} + 1/C_{depth} + 1/C_{DOS}]^{-1}$ is the series combination of dielectric $C_{ox} = \varepsilon_{SiO_2} L_g W_g / T_{eq}$, wavefunction depth $C_{depth} = \varepsilon_{semi} L_g W_g / T_{inv}$ (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, $C_{dos} = q^2 g (m_\parallel m_\perp)^{1/2} L_g W_g / 2\pi\hbar^2$, 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 transport[5] and $E_f - E_{well} >> kT$, $v_{inj} = (4/3\pi)(2(E_f - E_{well})/m_\parallel)^{1/2} = (4/3\pi)(2qC_{g-ch}(V_{gs} - V_{th})/m_\parallel C_{dos})^{1/2}$. We scale by maintaining constant v_{inj} while reducing $C_{ox} / L_g W_g$, $C_{depth} / L_g W_g$, and $C_{dos} / L_g W_g$ 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 $I_d / W_g = J_0 \cdot K_1 \cdot ((V_{gs} - V_{th})/1\text{V})^{3/2}$, where $J_0 = (4/3\pi)(2q/m_0)^{1/2}(1\text{V})^{3/2}(q^2 m_0 / 2\pi\hbar^2)$ = 84 mA/μm and $K_1 = n \cdot (m_\perp / m_o)^{1/2}(1 + (C_{dos,o} / C_{equiv}) \cdot g \cdot (m_\perp^{1/2} m_\parallel^{1/2} / m_o))^{-3/2}$ is the normalized current density. $C_{equiv} = [1/C_{ox} + 1/C_{inv}]^{-1}$ is C_{depth} and C_{ox} in series, while $C_{dos,o} = q^2 m_0 L_g W_g / 2\pi\hbar^2$. The transit time $\tau_{ch} = Q_{ch} / I_d = K_2 (L_g / v_0)(1\text{V}/(V_{gs} - V_{th}))^{1/2}$, where the normalized delay is $K_2 = (m_\parallel / m_0)^{1/2}(1 + C_{dos,o} g (m_\perp m_\parallel)^{1/2} / m_0 C_{equiv})^{1/2}$ and $v_0 = (4/3\pi) \cdot (2q \cdot (1\text{ Volt})/m_0)^{1/2}$ $= 2.5 \cdot 10^7$ cm/s. Given one isotropic valley ($m_\perp = m_\parallel = m^*$, g =1) and 1 nm *total* equivalent dielectric thickness EOT (i.e. $C_{equiv} = \varepsilon_{r,SiO_2} L_g W_g / (1\text{ nm})$), 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 FETs[1], 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 $E_F - E_{well} = n_s (dn_s / dE_f)^{-1} = n_s (2\pi\hbar^2 / g m_\perp^{1/2} m_\parallel^{1/2})$ also is large, once $(E_F - E_\Gamma) \geq (E_L - E_\Gamma)$ 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 $m^{*\prime} = m^*(1 + 2\alpha(E_{well} - E_\Gamma))$ 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 AlSb 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 Al_2O_3 and 0.34 nm AlSb lie between the well and gate, giving T_{eq} =0.37 nm. Under strong inversion $C_{g-ch} / L_g W_g \cong 2.4$ μF/cm^2, far below $C_{ox} / L_g W_g$ =9 μF/cm^2, and the high-mass L-valleys fill for $n_s > 2.4 \cdot 10^{12}$ 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_t / m_o (0.062-0.1) and high m_l / 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_l. Using instead a (111)

wafer, the L[111] valley is oriented vertically, and shows low transport masses ($m_{\|} = m_{\perp} = m_t$) and high confinement mass ($m_q = m_l$). The L$[\bar{1}11]$, L$[1\bar{1}1]$, and L$[11\bar{1}]$ minima show high[6] transport mass $(m_t + 8m_l)/9$ in one in-plane direction, but low confinement mass[6] $m_q = 9m_t m_l/(m_t + 8m_l)$. 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 = $3m_t m_l/(m_t + 2m_l)$ quantization mass. In appropriate thin wells, the X and L$[\bar{1}11]$, L$[1\bar{1}1]$, and L$[11\bar{1}]$ 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$[\bar{1}11]$, L$[1\bar{1}1]$, and L$[11\bar{1}]$ 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 $V_{gs} - V_{th}$ =300 mV, and moderately higher n_s does not populate heavy valleys. In inversion, $C_{g-ch}/L_g W_g \cong 4$ μF/cm^2. The benefit over the (100) design is larger in the ballistic case.

In InGaAs, GaAs, and GaSb, the L-valley m_l is >25:1 larger than the Γ-valley mass, hence T_{well} 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 $V_{gs} - V_{th}$ =300 mV the separation is 56 meV. L$[\bar{1}11]$, L$[1\bar{1}1]$, and L$[11\bar{1}]$ 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 $7.8 \cdot 10^{12}$ cm^{-2} with $V_{gs} - V_{th}$ ~300 mV; moderately higher n_s does not populate heavy valleys. $C_{g-ch}/L_g W_g \cong 4$ μF/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 $C_{well} = \varepsilon L_g W_g / T_{pitch}$, where T_{pitch} is the well pitch. With thin wells, and low m_t, C_{dos} can be increased 1.5:1 to 2.2:1.

In these FETs, the X and L$[\bar{1}11]$, L$[1\bar{1}1]$, and L$[11\bar{1}]$ valleys play a similar role to the L and X valleys in {100} Γ-transport FETs. For high v_{inj} it is sufficient to avoid populating high-$m_{\|}$ 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 $E_n = \hbar^2 \pi^2 n^2 / 2 m_l T_{pitch}^2$ 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_{equiv} = $\varepsilon_{r,SiO_2} L_g W_g / (0.3\,\text{nm})$, we find I_d/W_g ~4 mA/μm 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 $[01\bar{1}]$ for the most rapid increase in L $[11\bar{1}]$ and $[1\bar{1}1]$ energies with lateral confinement. L$[\bar{1}11]$ has low $m_{\|}$ and is populated without degrading v_{inj}.

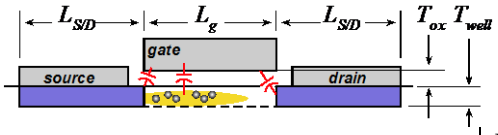
A final design uses a planar (110)-oriented channel, with the L$[\bar{1}11]$ and $[11\bar{1}]$ minima projecting into the transport plane with a low transport mass $m_{\|} = m_t$ in the $[1\bar{1}0]$ direction, a high perpendicular $m_{\perp} = (m_l + 2m_t)/3$ in the [001] direction, and moderate $m_q = 3m_l m_t/(m_l + 2m_t) \approx 3m_t$. The L$[\bar{1}11]$ and L$[1\bar{1}1]$ minima have higher $3m_l m_t/(m_l + 2m_t)$ mass in the $[1\bar{1}0]$ direction, and have low $m_q = m_t$. Quantization favors population of the desired [111], $[11\bar{1}]$ 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* basis[9] 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. .
[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
[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)

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



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

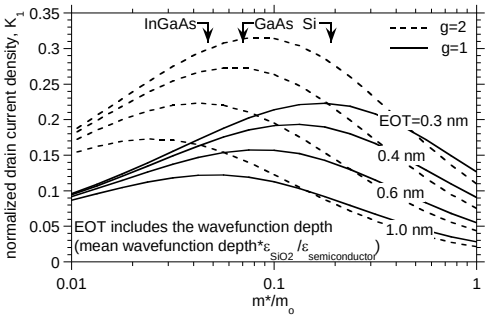


Figure 1: FET normalized drive current K_1 where $I_d/W_g = (84\,\mathrm{mA}/\mu\mathrm{m})\cdot K_1 \cdot ((V_{gs}-V_{th})/1\mathrm{V})^{3/2}$, and g is the # of valley minima.

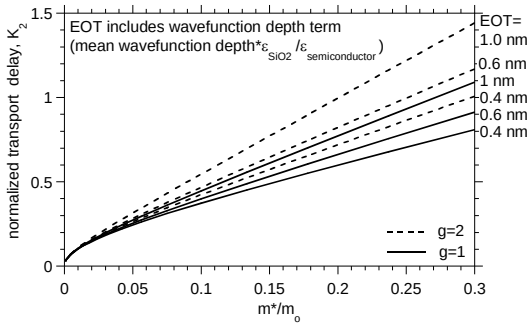


Figure 2: FET normalized transit time K_2 where $\tau_{ch} = Q_{ch}/I_D = K_2 \cdot (L_g/2.5\cdot10^{7}\,\mathrm{cm/s}) \cdot (1\mathrm{V}/(V_{gs}-V_{th}))^{1/2}$. EOT includes dielectric thickness and wavefunction depth.

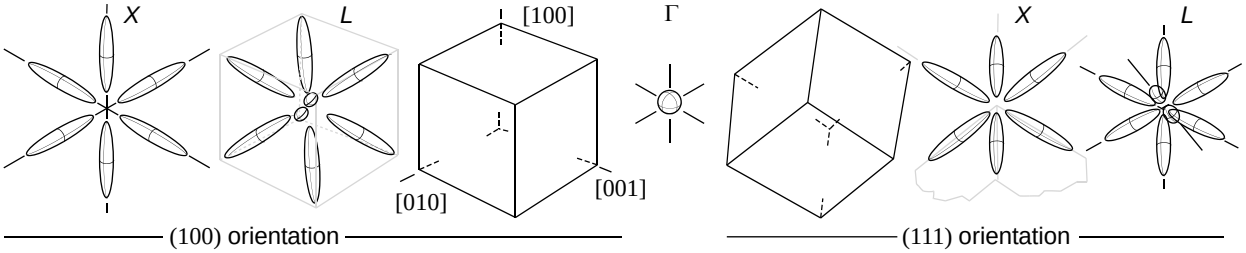


Figure 3: Γ, L, and X-valley orientations for (100)- and (111)-oriented wafers

Table 2: Parameters of Γ, L, and X-valleys for several suitable semiconductors

		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 $\Delta \sim 0.85\cdot\langle 100\rangle$		

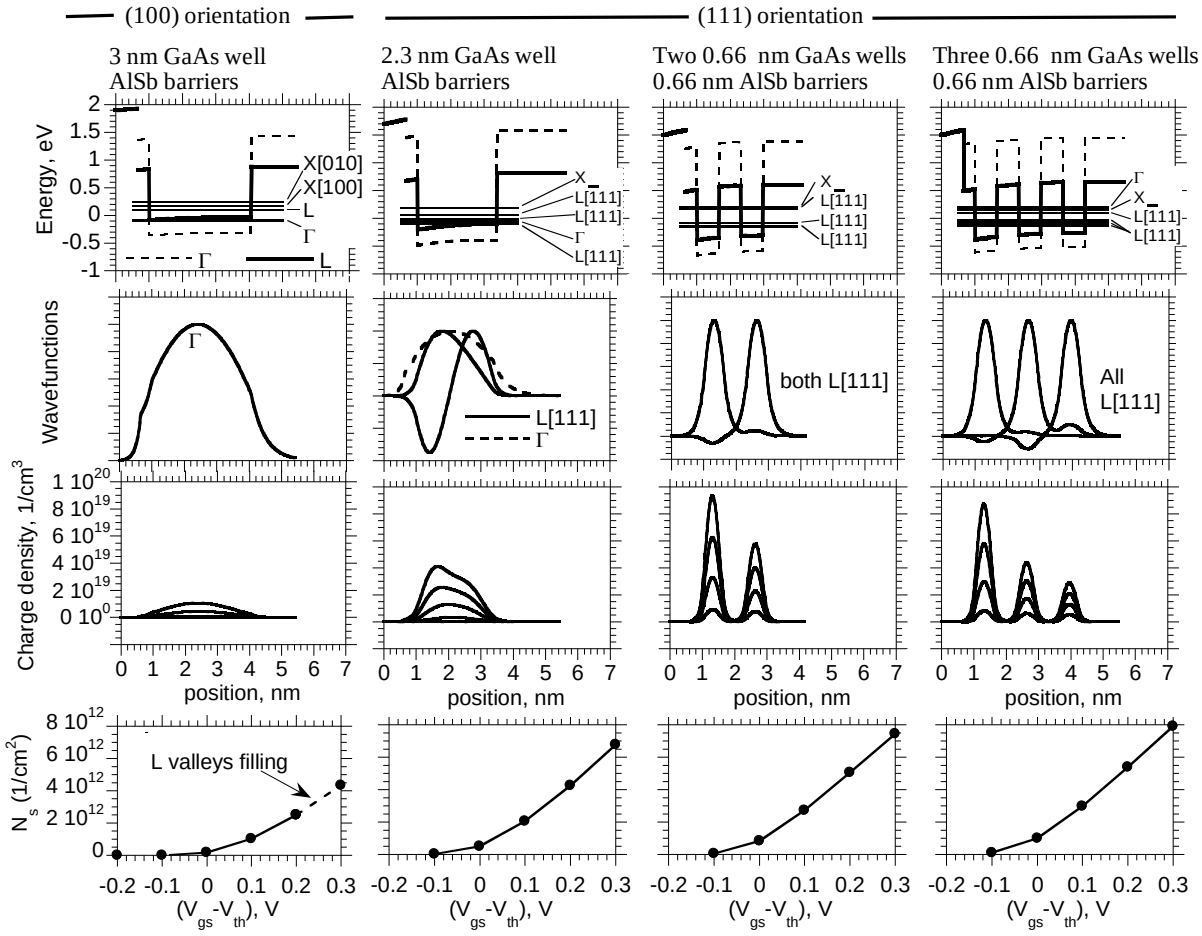


Figure 3. Simulation of Γ, Γ-L, and multiple-L valley FETs: quantized states, wavefunctions, charge density, and sheet carrier concentrations vs. bias. Well energies and charge densities calculated using the effective mass approximation and assuming parabolic bands. The gate dielectric is 0.3 nm Al_2O_3. 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.

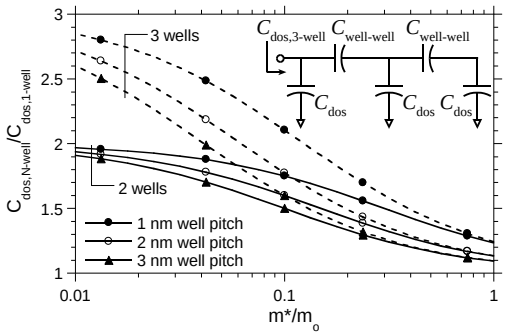


Figure 4: Increase in density of states for 2-L-well and 3-L-well FET channels as a function of carrier effective mass and well-well pitch.

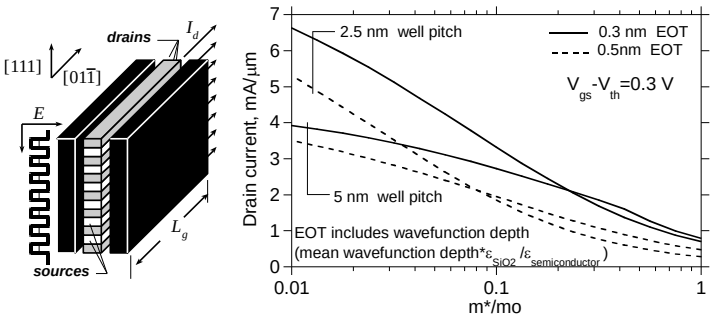


Figure 5: Computed drain current drain current density, per gate, of a FINET with (111)-oriented multiple quantum wells as a function of carrier effective mass and well-well pitch.