# Using Density-Gradient Theory to Model Sb-Based p-Channel FETs

M.G. Ancona, J.B. Boos and B.R. Bennett
Electronics Science and Technology Division
Naval Research Laboratory
Washington, DC USA
ancona@estd.nrl.navy.mil

***Abstract*—Density-gradient theory is discussed as a tool for modeling Sb-based p-channel FETs. The theory's methods and approximations are reviewed with emphasis given to the phenomenological treatment of the quantum confinement. The theory is then illustrated by using it to analyze FETs having InSb, GaSb and InGaSb channels, and to project their scaling characteristics.**

***Keywords-hole transport; antimonides; device scaling; density-gradient theory.***

## I. INTRODUCTION

Device modeling has long had an important role in electronics as a means for projecting the performance of new materials, and as a tool for engineering optimization. With the aggressive scaling of devices that has occurred in recent years, the workhorse method of diffusion-drift (DD) theory has become increasingly questionable as quantum and high-field transport effects have become ever more dominant. This fact has led many researchers to emphasize microscopic methods with either a classical (*e.g.*, Boltzmann/Monte Carlo) or quantum mechanical (*e.g.*, NEGF) foundation. These methods are attractive because they provide a physically well-founded basis for treating the "new" physics of the scaled devices. However, they also suffer from a number of drawbacks with the main one being they are generally too computationally intensive for routine use in engineering. Other disadvantages are that they tend to provide an uneven treatment (*e.g.*, providing an excellent representation of the quantum mechanics but then treating scattering with a simple relaxation time), they often lapse into phenomenology with their physics content compromised by insufficient knowledge of real materials and devices (*e.g.*, about the scattering rates especially deep in the band, the degree of strain relaxation, or the precise impurity profiles), and they do not interface easily with other methods (*e.g.*, in devices in which quantum effects are important only in a small region of the device). In the face of these realities, it seems imprudent to focus solely on the microscopic methods, and premature to give up on the "traditional" device modeling approaches. This is the general philosophy of the present paper where we use methods that have the same macroscopic/continuum basis of DD theory but that extend that theory in various ways. In particular, we focus on ultra-scaled heterojunction FETs and on their description using a version of density-gradient (DG) theory that is essentially a quantum-corrected DD theory [1].

As an application area, this paper engages a topic of recent interest, namely p-channel FETs formed of compound III-V semiconductor materials. Work on such FETs has been motivated by the possibility of creating a high-performance III-V analog of Si CMOS [2]. Ongoing efforts by several groups have seen good progress in raising the III-V hole mobility by focusing on the antimonides, *i.e.*, InSb [3], GaSb [4] and their alloys [5], and by utilizing confinement and strain much as is done with p-channel Si technology [6]. The highest reported room-temperature hole mobility to date in a III-V material is about 1500cm$^2$/V-sec for InGaSb [5]. For the purposes of our macroscopic modeling work, the problem of hole transport is especially attractive for two reasons. First, efforts to model such hole transport using microscopic methods are far less well developed than for the corresponding n-channel FETs. And second, because the hole mobilities are relatively low, high-field transport phenomena will tend not to dominate except when the gate lengths are extremely short, and therefore the simplified scattering-dominated version of DG theory employed in this paper is more likely to suffice.

## II. DENSITY-GRADIENT APPROACH

### A. Governing Equations

Generically, DG theory is a classical field theory. As such, the hole gas is assumed characterized by a set of densities of which the most important are the charge density $qp$, the current density $qp\mathbf{v}_\mathrm{p}$, and the stress. These densities are constrained by the classical laws of charge conservation, momentum balance, and electrostatics, which for scattering-dominated situations may be written as [1]:

$$\text{(1a)} \qquad \frac{\partial p}{\partial t} + \nabla\cdot\left(p\mathbf{v}_\mathrm{p}\right) = 0 \qquad \frac{p\mathbf{v}_\mathrm{p}}{\mu_\mathrm{p}} + p\nabla\psi + p\nabla\phi_\mathrm{p}^{\mathrm{DG}} = 0$$

$$\text{(1b)} \qquad \nabla\cdot\left(\varepsilon_d\nabla\psi\right) = q\left(N_A - p\right)$$

where the stress has been expressed in terms of a chemical potential $\phi_\mathrm{p}^{\mathrm{DG}}$ and all other quantities take their usual meanings. That momentum conservation is reduced to the simple force balance in $(1a)_2$ is a consequence of the assumption of strong scattering and is a direct statement that ballistic phenomena are not encompassed by these equations.

The defining assertion of DG theory is that the stress in the hole gas depends not only on the gas density as in DD theory, but also on the density-gradient. This latter dependence makes the hole gas response "non-local" and can thereby provide a first-order representation of the one-particle quantum non-locality associated with the deBroglie wavelength. When formulated in terms of the chemical potential, the lowest-order relationship is

$$(2)\quad \phi_p^{DG} = \phi_p^{DD}(p) - \frac{2}{r}\nabla\cdot\left(b_p \nabla r\right) \quad \text{where} \quad b_p = \frac{\hbar^2}{12 m_p^{DG}}$$

$\phi_p^{DD}$ is the ordinary chemical potential of DD theory and $r \equiv \sqrt{p}$. The coefficient $b_p$ gauges the strength of the DG effect, and clearly when $b_p$ vanishes, the theory reduces to ordinary DD theory. Also evident from its formula, the size of $b_p$ is set by the DG effective mass $m_p^{DG}$ (or the corresponding tensor) which has been shown to equal the ordinary density-of-states effective mass $m_p$ in the high temperature limit [7].

To use the foregoing differential equations to analyze device problems, one needs to formulate appropriate boundary value problems using boundary conditions that have been discussed elsewhere [1]. For this paper only steady situations are considered so the time derivative in $(1a)_1$ always vanishes.

### B. *Representation of Quantum Confinement*

Previous work applying the DG equations to quantum confinement has found that DG theory can often be predictive, accurately capturing the behavior with the DG effective mass taking its theoretical value $m_p$ [7]. And although in numerous other circumstances DG theory loses this predictive capacity, it has been found that the theory can still function as a remarkably accurate phenomenology for fitting quantum mechanical solutions with $m_p^{DG}$ used as the fitting parameter.

The quantum confinement situations considered in this paper are generally extreme ones because, as noted in the Introduction, the imposed confinement and strain are the tools used to enhance the hole mobility and are chosen as large as possible in order to maximize the splitting between the heavy and light hole bands. Given this, our approach must necessarily be phenomenological. To determine $m_p^{DG}$ we therefore generate quantum mechanical hole density profiles for various quantum wells using the NEXTNANO program [8] as shown in Figs. 1a-c. The well/barrier materials studied correspond to the highest mobility situations in the literature for each material system; these are: InSb/$Al_{0.35}In_{0.65}Sb$ (Fig. 1a, 5nm well, 2.3% strain), GaSb/$AlAs_{0.24}Sb_{0.76}$ (Fig. 1b, 7.5nm well, 1.2% strain) and $In_{0.41}Ga_{0.59}Sb/Al_{0.75}Ga_{0.25}Sb$ (Fig. 1c, 7.5nm well, 2.1% strain). Also shown in the Figures are the best-fit DG results in which the single value of the perpendicular DG effective mass shown in Table 1 was used in each material for all densities. Although the agreements are not perfect, it seems clear that the DG approach does provide an excellent representation over the range studied. These values of the DG effective mass are used in all subsequent calculations.

TABLE I. SELECTED MATERIAL CONSTANTS

| Quantity | InSb | GaSb | InGaSb |
|---|---|---|---|
| LF mobility ($cm^2$/V-sec), well | 1230 | 1350 | 1500 |
| LF mobility, barrier | 50 | 50 | 50 |
| Saturation velocity (cm/sec) | $8x10^6$ | $8x10^6$ | $8x10^6$ |
| DG effective mass (⊥) | 0.03 | 0.045 | 0.048 |
| Schottky barrier (eV) | 0.78 | 1.03 | 1.0 |
| Band gap (eV), well | 0.17 | 0.62 | 0.45 |
| Band gap (eV), barrier | 0.78 | 1.66 | 1.43 |
| Valence band offset (eV) | 0.21 | 0.64 | 0.43 |
| Dielectric constant, well | 17.7 | 11.6 | 16.5 |
| Dielectric constant, barrier | 15.7 | 15.7 | 13 |

### C. *Representation of Quantum Tunneling*

Quantum tunneling can potentially play a role in the subject devices by providing pathways for significant leakage currents to flow through the Schottky gate and/or from source to drain.

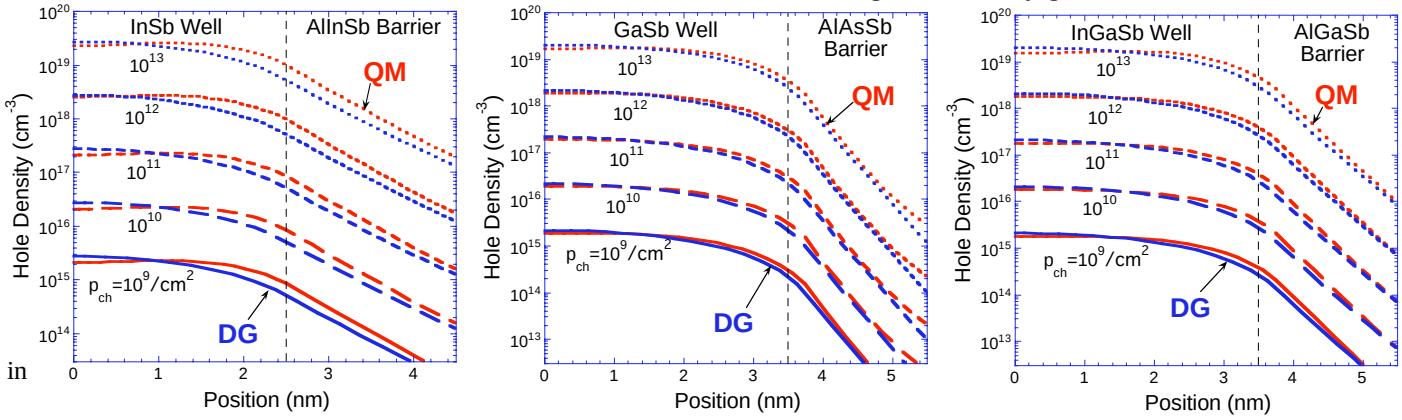


Fig. 1a-c. Density profiles across the quantum wells studied in this paper as simulated by quantum mechanics and fit by DG theory. The DG effective masses determined by these fits appear in Table 1.

While DG theory is believed to be capable of modeling such tunneling [9], the scattering-dominated theory outlined in Sect. II.A is inappropriate because the tunneling is almost always in an elastic (or quasi-elastic) regime. Nevertheless, the theory of Sect. II.A does include a scattering-dominated tunneling process that, while unphysical, may still enable a useful phenomenology as discussed in [10]. This approach is utilized in our modeling, largely for qualitative purposes, and the parallel component of the mass tensor is chosen accordingly.

### *D. Other Effects*

To provide a simple representation of the transport we assume the mobility model:

$$\frac{1}{\mu_p^{tot}} = \frac{1}{\mu_p^{LF}}\sqrt{1+\left(\mu_p^{LF}|E_{||}|/v_p^{sat}\right)^2} + \frac{1}{\mu_p^{SS}}\left|\frac{E_\perp}{E_{SS}}\right|^4 \quad (3)$$

where the first term is a commonly used expression for velocity saturation, and the second term represents surface scattering. Some of the constants appearig in this expression are given in Table I. The surface scattering model was chosen to fit data of [3] under strong negative bias (with $\mu_p^{SS}$ = 100cm$^2$/V-sec and $E_{SS}$ = 3x10$^5$V/cm); how well this term represents the other materials is unknown.

Generation-recombination processes may have to be considered for the materials of interest, and especially for InSb with its arrow gap. For this paper, however, we shall neglect these processes and this is reflected in the right-hand side of $(1a)_1$ being zero. We believe this justified even for InSb because the strong confinement acts to increase the effective gap significantly, and because of the very small generation volume.

### *E. Numerical Implementation*

Inserting $(1a)_2$ into $(1a)_1$ eliminates the velocity as a variable so that $p$ and $\psi$ are then the unknowns. For numerical purposes, we have found it convenient to convert $p$ to a "Slotboom variable" $\varphi$ defined by $p = N_A \exp(-2q\varphi/k_BT)$. In addition, a generalized chemical potential $\Phi_p^{DG} \equiv \phi_p^{DG} + \psi$ is introduced as a third variable so that each numerical variable ($\varphi$, $\psi$ and $\Phi_p^{DG}$) then satisfies a second-order PDE. These equations were solved using the Comsol Multiphysics program, and the code was found to work quite well with two standard caveats. First, the grid has to be sufficiently fine so as to adequately resolve the physics. And second, the initial guess needs to be "good enough". Our approach to obtaining the initial guess was to solve the problem first with a much smaller mass than the actual one so as to aid convergence by weakening the singular perturbation of the DG term.

## III. Simulation Examples

### *A. Device Geometry*

The device geometry studied in this work is based on an InSb device design of [3] that had a gate length of 40nm. The basic geometry is shown in Fig. 2 where the barrier thickness is 10nm and the gate lengths range from 20-300nm. The spacing of the p+-cap regions and the gate size are chosen with high frequency performance in mind (though T-gates have not been studied). The GaSb and InGaSb devices are assumed to have iden-tical designs except for the thicker channels that experimentally were important for achieving the highest mobilities [4,5].

A sample DG solution is illustrated in Fig. 3 for the InSb FET of [3] with a 40nm gate length. In the 2D plot, the simulated hole density at pinch-off is shown with the quantum confinement of the carriers in the channel being evident. From a number of such solutions at various gate biases, the drain characteristics can be assembled as shown in Fig. 4 for $V_D$ = -0.5V. In the plot we show the experimental I-V data from [3] as well as simulated curves for the comparable InGaSb and GaSb devices. For the InSb, simulation and experiment are seen to be in good agreement with the drop in calculated drain current at higher negative bias being due to the turn-on of the Schottky barrier under forward bias. And the leakage current seen under positive bias in the InSb FET is readily shown to be associated with the low valence band offset and resulting current flow in the barrier regions. Leakage currents associated with source-to-drain tunneling are not seen even when the parallel component of the DG effective mass tensor is made very small.

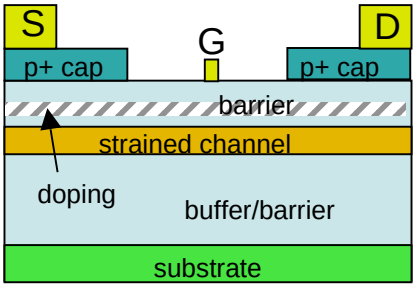


Fig. 2. Schematic of the device structure.

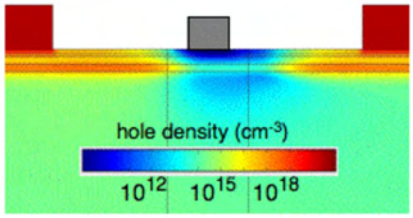


Fig. 3. Simulated hole density in a 40nm InSb FET at pinch-off.

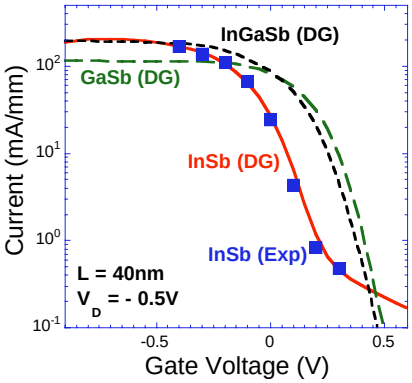


Fig. 4. Simulated drain characteristics of InSb, GaSb and InGaSb FETs plus experimental points for InSb from [3].

For a more general comparison of the three types of antimonide devices, we consider their perfomance with $V_{CC}$ = 0.5V as was assumed in [3]. To aid in the comparisons, we follow [2] in calculating certain benchmarks like the on-off ratio $(I_{ON}/I_{OFF})$, the gate delay $(C_G V_D/I_{ON})$, and the energy-delay-per-unit-width $(C_G V_D/I_{ON} \, x \, C_G V_D^2)$. One informative plot is the gate delay versus the on-off ratio in which each point corresponds to a different choice for the threshold voltage [2]. The results for 40nm FETs made of each material is shown in

Fig. 5 along with the operating point of the InSb device of [3]. All the devices have similar minimum values of gate delay, however, the simulations show that the high leakage level of the InSb device greatly limits one's ability to reach on-off ratios much above 100, whereas the other materials allow ratios of $10^4$ and even $10^5$ to be reached.

To study how the behavior of these heterostructure FETs depends on gate length one needs a scaling strategy. For this paper, the scaling strategy is based on Fig. 5 and its tradeoff between gate delay and on-off ratio. In such a plot, one can define an optimum as occurring at the "knee" of the characteristic, where further improvements in on-off ratio become quite costly in terms of gate delay (if they can be achieved at all). Using similar plots for each each material and each gate length allows us to map out a sequence of "optimum" threshold voltages in each case, and in this way to obtain a projection of the scaling behavior. The results are shown in Figs. 6 and 7 for the energy-delay product and on-off ratio, respectively. In Fig. 6, data from [4] is also included and is seen to be in reasonable agreement with the simulation. One basic point illustrated by Fig. 6 is the general advantage the antimonide devices have over silicon technology in terms of energy-delay product. In addition, the plots in Figs. 6 and 7 suggest that the InSb devices can provide the fastest switching but only with a substantial penalty in on-off ratio that likely would make them less suitable for logic applications. Alternatively, if on-off ratios of around 100 were sufficient for one's application, then the operating points of the InGaSb and GaSb devices could be adjusted so as to give them faster operation. Further such simulations that together allow for an overall evaluation of the three Sb-based material systems as p-channel FETs will appear in a future publication.

## IV. Final Remarks

In this paper we have outlined a density-gradient theory approach to modeling Sb-based p-channel FETs. The theory's basic equations have been presented and their main underlying assumptions have been noted. A key step in such modeling is the phenomenological representation of the hole gas that allows the strong quantum confinement to be described and with strain effects included. The theory was then illustrated by using it to analyze particular FET designs that had InSb, GaSb and InGaSb channels with state-of-the-art transport properties. Lastly, simulations were presented that indicate a route to evaluating the scaling properties of these FETs.


## Acknowledgment

The authors thank the Office of Naval Research for partial support of this work.


## References


[1] M.G. Ancona and H.F. Tiersten, "Macroscopic physics of the silicon inversion layer," *Phys. Rev. B* 35, 7959 (1987).

[2] R. Chau, "Opportunities and challenges of emerging nanotechnologies for future high-speed and low-power logic applications," *IEEE/CSICS Tech. Dig.* 17 (2005).

[3] M. Radasavljevic *et al.*, "High-performance 40nm gate length InSb p-channel compressively strained quantum well field effect transistors for low-power ($V_{CC}$=0.5V) logic applications," *IEDM Tech. Dig.*, 727 (2008).

[4] B.R. Bennett, M.G. Ancona, J.B. Boos, C.B. Canedy and S.A. Khan, "Strained GaSb/AlAsSb quantum wells for p-channel field effect transistors," *J. Cryst. Growth* 311, 47 (2008).

[5] B.R. Bennett, M.G. Ancona, J.B. Boos and B.V. Shanabrook, "Mobility enhancement in strained p-InGaSb quantum wells," *Appl. Phys. Lett.* 91, 042104 (2007).

[6] M.L. Lee, E.A. Fitzgerald, M. Bulsara, M.T. Currie and A. Lochtefeld, "Strained Si, SiGe and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors," *J. Appl. Phys.* 97, 011101 (2005).

[7] M.G. Ancona and G.J. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor," *Phys. Rev.* B39, 9536 (1989).

[8] S. Birner *et al.*, "Nextnano: general-purpose 3-D simulations," *IEEE Trans. Elect. Dev.* 54, 2137 (2007).

[9] M.G. Ancona and A. Svizhenko, "Density-gradient theory of tunneling: physics and verification in one-dimension," *J. Appl. Phys.* 104, 073726 (2008).

[10] J.R. Watling, A.R. Brown, A. Asenov, A. Svizhenko and M.P. Anantram, Simulation of direct source-to-drain tunneling using the density-gradient formalism: Non-equilibrium Green's function calibration," *SISPAD*, 267 (2002).


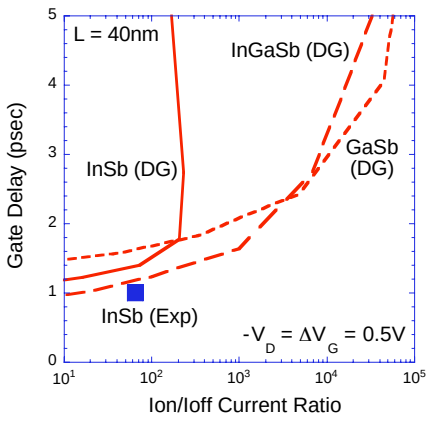


Fig. 5. Simulated gate delay versus on/off ratio for 40nm InSb, GaSb and InGaSb pFETs along with the operating point for the experimental device of [4].

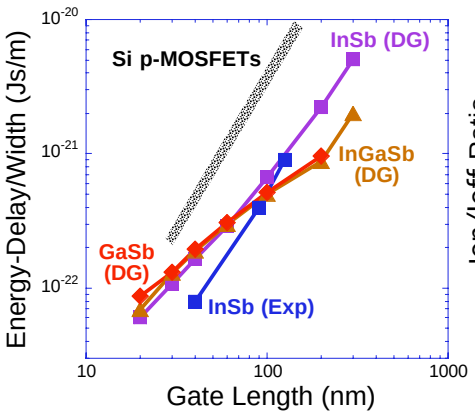


Fig. 6. Simulated energy-delay versus gate length for the InSb, InGaSb and GaSb FETs. Also shown are the data from [4].

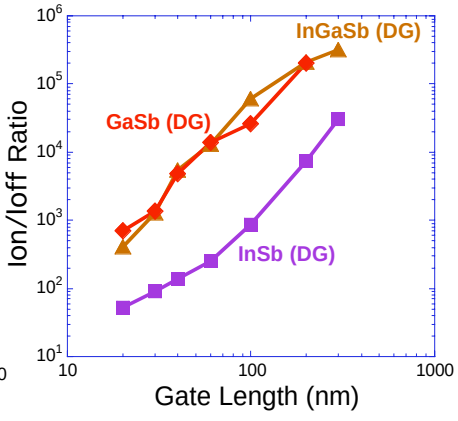


Fig. 7. Simulated on-off ratios for the three types of FETs with varying gate lengths.