Solid-State Electronics 54 (2010) 1349–1358

Contents lists available at ScienceDirect

Solid-State Electronics

journal homepage: www.elsevier.com/locate/sse

Scaling projections for Sb-based p-channel FETs

M.G. Ancona *, B.R. Bennett, J.B. Boos

Electronics Science and Technology Division, Naval Research Laboratory, Washington, DC, USA

ARTICLE INFO

Article history:
Received 20 January 2010
Received in revised form 19 June 2010
Accepted 24 June 2010
Available online 23 July 2010
The review of this paper was arranged by Prof. S. Cristoloveanu

Keywords:
Hole transport
Antimonides
p-Channel
Device scaling
Density-gradient theory
Complementary circuits

ABSTRACT

Numerical device modeling is used to study p-channel FETs with InSb, GaSb and InGaSb channels. To be as realistic as possible, the basic parameters are chosen to be those measured experimentally in state-of-the-art high-mobility materials, and where possible, predictions are compared against published data. Confinement effects are captured in the simulations using the density-gradient description of quantum transport. The emphasis is on projecting scaling properties and ultimate performance, with key issues being short-channel effects, the importance of source-drain leakage current, power considerations and p^+-cap design. Although important, issues related to gate leakage current and gate stack design are not well addressed by modeling, and so are not considered in detail. With III–V complementary circuits and high-speed, low-power applications in mind, the general conclusion is that among the antimonide-based pFETs, InGaSb devices provide the best balance of speed and power dissipation.

Published by Elsevier Ltd.

1. Introduction

As a method for minimizing standby power, the CMOS approach of mainstream digital silicon electronics that employs matched n-channel and p-channel transistors provides a near-ideal solution. Of course one of the prime advantages of silicon is the facility with which it can implement complementary circuits, readily allowing the requisite enhancement-mode n- and p-channel transistors to be integrated in the same channel material, and with an easily grown, high-quality insulator in SiO_2. And this has been a major reason why silicon has dominated semiconductor electronics for many years. Recently, however, the quest for ever-better performance as envisioned by the ITRS roadmap has had researchers looking into the possibilities of CMOS circuits that utilize III–V materials, either alone or in hybrid form with germanium [1]. The main advantage of these alternative materials is their potential for higher speed at low voltage. Of course the III–Vs lack silicon's "natural fit" with CMOS – some of the main problems are the complications of heterogeneous materials integration, the lack of a good gate insulator, the generally lower material quality, the difficulties of achieving enhancement-mode, and the characteristically low values of the hole mobility in the III–Vs compared with their electron mobilities (see Table 1) – but recently significant advances have been made and the notion of III–V complementary circuitry is becoming more plausible [1]. A primary area of progress has been in III–V hole mobility, and the purpose of the present paper is to explore the scaling potential of these new pFETs, and to compare the relative performance of several candidate III–V materials.

Regularly achievable electron mobilities in the channels of III–V HEMTs at room temperature are much higher than in Si or Ge, e.g., in InAs they are in the range of 20–30,000 cm^2/V sec [2]. Such numbers have made the III–V semiconductors very attractive for fabricating high-speed n-channel transistors for use in analog circuits. In addition, the trend toward higher mobility with smaller bandgap (see Table 1) has made the III–V materials a route to achieving higher speed at lower voltage and on-state power for a given geometry as illustrated in Fig. 1 [3]. By contrast, the relatively poor hole transport properties and material quality of the III–V semiconductors has in the past yielded experimental p-channel mobilities that were limited to the range 100–300 cm^2/V sec [4]. The last few years have, however, brought encouraging progress. The starting point for these advances has been a focus on the antimonides that along with Ge have the most intrinsic promise as seen from Table 1. Pushing the hole mobilities even higher has then come via three strategies: (i) improving material quality, (ii) imposing strong confinement, and (iii) employing high levels of strain. The better quality materials are largely the products of long-standing programs on antimonide growth by molecular beam epitaxy at the QinetiQ Corp. (for InSb) and at the Naval Research Laboratory (for GaSb and InGaSb). When confinement and strain act on such materials, the hole mobilities are enhanced because, as in Si and SiGe [5], the degeneracy of the light- and heavy-hole bands is split, thereby raising the proportion of carriers in the

* Corresponding author.
E-mail address: ancona@estd.nrl.navy.mil (M.G. Ancona).

0038-1101/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.sse.2010.06.022

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Scaling projections for Sb-based p-channel FETs	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Electronics Science and Technology Division,4555 Overlook Avenue SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Numerical device modeling is used to study p-channel FETs with InSb, GaSb and InGaSb channels. To be as realistic as possible, the basic parameters are chosen to be those measured experimentally in state-ofthe-art high-mobility materials, and where possible, predictions are compared against published data. Confinement effects are captured in the simulations using the density-gradient description of quantum transport. The emphasis is on projecting scaling properties and ultimate performance, with key issues being short-channel effects, the importance of source-drain leakage current, power considerations and p+-cap design. Although important, issues related to gate leakage current and gate stack design are not well addressed by modeling, and so are not considered in detail. With III?V complementary circuits and high-speed, low-power applications in mind, the general conclusion is that among the antimonide- based pFETs, InGaSb devices provide the best balance of speed and power dissipation.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**10**	

Table 1
Bulk coefficients of various semiconductors at room temperature. Source: *Handbook Series on Semiconductor Parameters*, vol. 1, eds. M. Levinshtein, S. Rumyantsev, M. Shur, (World Sci., Singapore, 1996).

Semiconductor	Bulk electron mobility (cm^2/V sec)	Bulk hole mobility (cm^2/V sec)	Band gap (eV)
Si	1400	450	1.12
Ge	3900	1900	0.66
GaP	250	150	2.24
GaAs	8500	400	1.43
GaSb	3000	1000	0.67
InP	5400	200	1.29
InAs	40,000	500	0.33
InSb	77,000	850	0.16

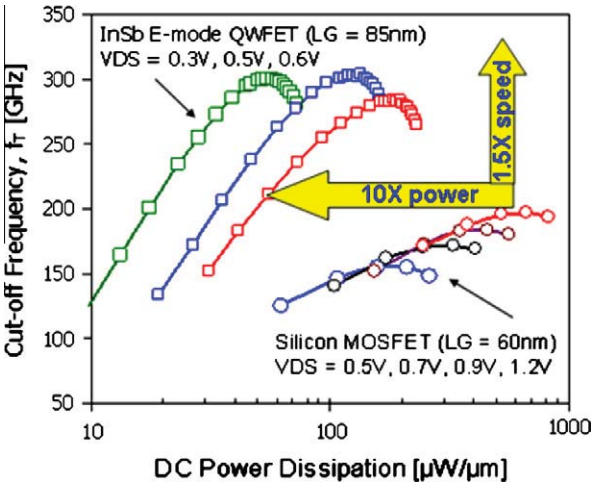

Fig. 1. Plot of cut-off frequency versus dc on-state power from [3] that compares the rf performance of scaled n-channel InSb and Si devices. The former are seen to be better than the latter in terms of both speed and power.

higher mobility heavy-hole band and lowering the density of final states for scattering. Using these strategies with quantum well thicknesses in the range of 5–10 nm and biaxial compressive strains of 1–2%, p-channel mobilities have been raised into the range of 1200–1500 cm^2/V sec as summarized in Table 2 and more thoroughly in [6–9]. Moreover, there is potential for further increases, particularly if one could reach higher strains (especially in GaSb where the best results to date have had strains of only 0.8–1.2% [8]), or could exert *uniaxial* strains as has been done to advantage in SiGe [10] and explored theoretically for the antimonides in [11]. And even the present mobility levels, which are better than the best results in Si (though not Ge), could be sufficient for a future III–V CMOS technology. From the perspective of matching the n- and p-channel devices, the balance is especially favorable if one looks at important high-field and on-state measures like the saturated drift velocity and the density-of-states. With respect to the velocity, for example, the best unity-gain cut-off frequency f_T for a 30 nm n-channel InAs devices is 628 GHz [12], while for a 40 nm p-channel InSb devices it is only a factor of 4 smaller at 140 GHz [6].

Table 2
Material Properties for Three Different Channels.

Quantity	InSb/ (Al,In)Sb	GaSb/ Al(As,Sb)	(In,Ga)Sb/ (Al,Ga)Sb
Mole fraction	Al = 0.35	As = 0.24	In = 0.41, Al = 0.75
Biaxial strain in well (%)	1.9	1.2	2.1
LF hole mob. (cm^2/Vs), well	1230	1350	1500
LF hole mobility, barrier	50	50	50
Saturation velocity (cm/s)	8×10^6	8×10^6	8×10^6
HH/LH effective masses (⊥)	0.263/0.015	0.25/0.044	0.26/0.032
DG effective mass (⊥)	0.04	0.06	0.057
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

A second advantage of the antimonides is that, as in Si technology, the same material can serve as both an n-channel and a p-channel. To see this we plot in Fig. 2 the cross-channel band diagrams for the InSb, GaSb and InGaSb channels that have achieved the highest mobilities and whose parameters are summarized in Table 2. The calculations were performed using the NEXTNANO program [13] and the particular well/barrier combinations are a 5 nm InSb quantum well with $Al_{0.35}In_{0.65}Sb$ barriers [6], a 7.5 nm GaSb quantum well with $AlAs_{0.24}Sb_{0.76}$ barriers [7], and a 7.5 nm $In_{0.41}Ga_{0.59}Sb$ quantum well with $Al_{0.75}Ga_{0.25}Sb$ barriers [8] where the well thicknesses have been chosen to be consistent with experimental results. The favorable band alignments are evident with the possibility of good confinement of both electrons and holes seen in all three cases, and with the smallest band offsets in InSb. We note that good quality n- and p-channels in the same material layer have already been demonstrated experimentally for InGaSb [14].

The main purpose of this paper is to develop an understanding of the comparative performance and scaling properties of pFETs made with InSb, GaSb or InGaSb as the channel materials. In doing so it should be noted that we do not consider in detail issues related to the gate stack (e.g., MIS designs, gate leakage currents, self-aligned source/drains, etc.), primarily because the relevant non-idealities are less amenable to modeling. Our approach is largely computational, and as reviewed briefly in the next section, the numerical device modeling is done in the density-gradient approximation. The basis for our technology comparisons is then discussed in Section 3. The main results and interpretations appear in Section 4, and the paper closes with some final comments in Section 5.

2. Modeling approach

In modeling the Sb-based pFETs it is important to include the effect of the strong quantum confinement. For this purpose we use the density-gradient (DG) description [15], an approach to quantum transport theory that sacrifices accuracy in the quantum mechanical representation for gains in computational efficiency and facility in modeling other real-world device complications such as complex geometry, self-consistent electrostatics, sophisticated mobility models, generation-recombination phenomena, etc. Because the equations of DG theory have been discussed both in general, and as applied specifically to the problem of hole transport in antimonide pFETs [16], our coverage here can be brief.

For steady-state situations, the relevant differential equations of DG theory can be written as [15]:

$$\nabla \cdot (p\mathbf{v}_p) = 0 \quad p\mathbf{v}_p/\mu_p + p\nabla\psi + p\nabla\varphi_p^{DG} = 0 \tag{1a}$$

$$\nabla \cdot (\varepsilon_d \nabla\psi) = q(N_A - p) \tag{1b}$$

where the generalized chemical potential is given by

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

φ_p^{DD} is the ordinary chemical potential of DD theory, $r \equiv \sqrt{p}$, and the coefficient b_p gauges the strength of the DG effect [15]. When the

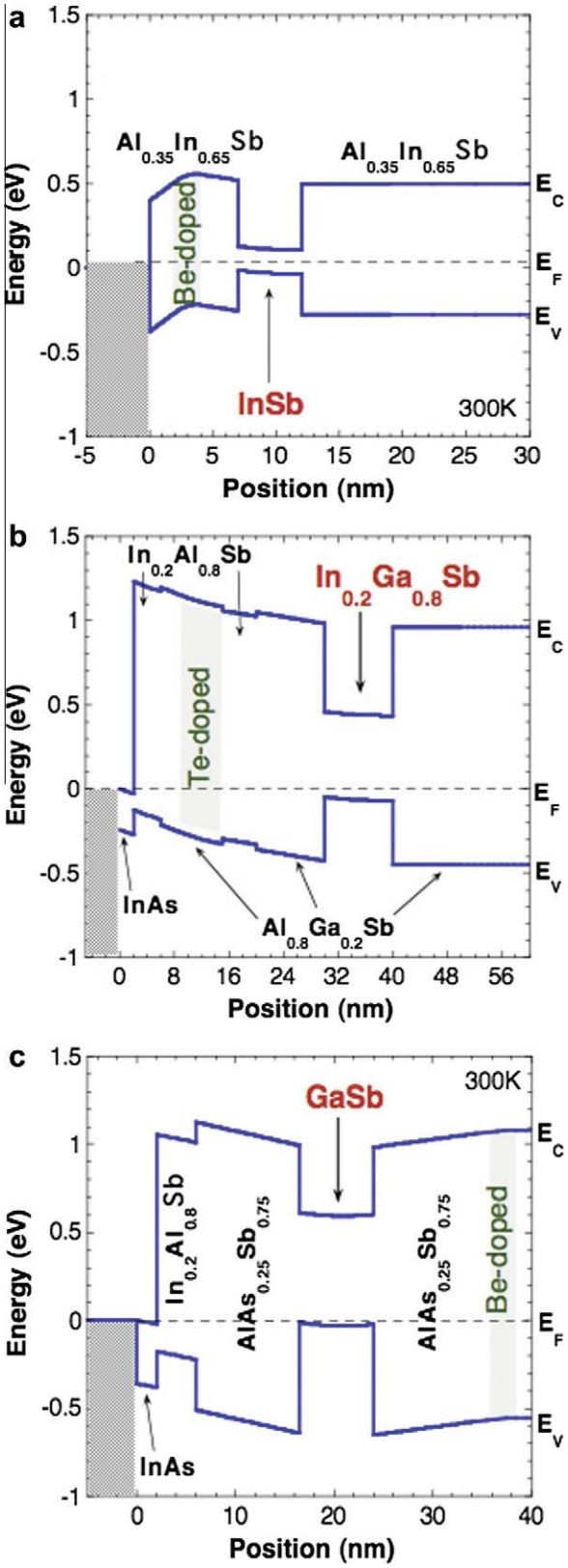

Fig. 2. (a–c) Band diagrams for the structures of Table 2 as simulated with the 8×8 k. p method using the NEXTNANO program [12]. The possibility of channel confinement for both electrons and holes in these structures is evident.

DG effective mass m_p^{DG} is taken to be the ordinary density-of-states effective mass, the foregoing theory is often quite accurate in describing quantum confinement [17]. However, as we shall see, this is not true for the strongly confined problems of interest in this paper. In such circumstances, a number of authors have shown that the theory can still function as a remarkably accurate phenomenology (including for multi-dimensional problems [18]) simply by using m_p^{DG} as a fitting parameter [17]. This approach is used in the present work, and as done in [18] we take m_p^{DG} to be a tensorial quantity composed of perpendicular and parallel components, $m_{p\perp}^{DG}$ and $m_{p||}^{DG}$, respectively.

The perpendicular mass is the crucial one for us as it quantifies the response to the confinement. To determine "best" values for this mass we curve-fit the hole density profiles predicted by quantum mechanics. The InSb, GaSb and InGaSb channels summarized in Table 2 and depicted in Fig. 2a–c are modeled, and the density profiles as computed by NEXTNANO [13] using the k·p method are shown (solid curves) in Fig. 3a–c for a variety of densities. Also shown in these figures (dashed curves) are the best-fit DG results in which a *single* value of the perpendicular DG effective mass for each material has been used for all densities. While not perfect, these fits are excellent and show that the DG approach is providing an accurate representation. The mass values used in each case, and then throughout this paper, are listed in Table 2 along with the light-hole (LH) and heavy-hole (HH) density-of-states masses of the quantum well materials. That the DG mass is depressed well below the dominant HH mass is due both to the contribution of the LH mass and to the strong confinement as discussed in [17]. Note also that in Fig. 3a–c the largest discrepancies are observed for InSb, which is not surprising given the narrowness of the well and the small valence band offset.

Regarding the parallel mass, as has been explored especially by the modeling group in Glasgow [18], it is possible to use the parallel DG effective mass $m_{p||}^{DG}$ in DG theory to represent both lateral confinement effects and source-drain tunneling. The latter application is intriguing, but the incorrect assumption implicit in $(1a)_2$ of scattering-dominated tunneling [19] makes the approach entirely phenomenological, and accurate only when calibrated by quantum mechanics [18]. We do not do this here because source-drain tunneling does not appear to make a significant contribution even for the shortest gate lengths considered (20 nm). Evidence for this claim is given in Fig. 4 where we plot the simulated I–V characteristics for a 20 nm gate InGaSb FET with $m_{p||}^{DG}$ varied as a parameter and find that only for effective masses with unrealistically low values (below about 0.01) does appreciable tunneling leakage current flow. This plot also shows that the lateral confinement has very little influence on the predicted I–V. We therefore feel safe in simply assuming a fixed value for $m_{p||}^{DG}$ $(0.053m_e)$ for all problems considered in this paper.

Other essential parameters for our simulations relate to the mobility models for the holes in the channel and barrier materials considered. With respect to the channels, as described in [16] we include terms in the mobility models representing velocity saturation and surface scattering:

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

Estimates for the low-field hole mobilities for each channel material come from Hall measurements and are as listed in Table 2, and the saturated drift velocity in all the materials is assumed to take the estimated bulk value of 8×10^6 cm/s. The parameter values in the surface scattering term were chosen to fit data of [6] for InSb pFETs under strong negative bias (with $\mu_p^{SS} = 100\ \mathrm{cm^2/V\ sec}$ and $E_{SS} = 3 \times 10^5$ V/cm); in the absence of better information, these values are also used for the other channel materials. It should also be noted that in using (3) we are asserting that effects such as ballistic transport and injection efficiency, possibly important for extremely

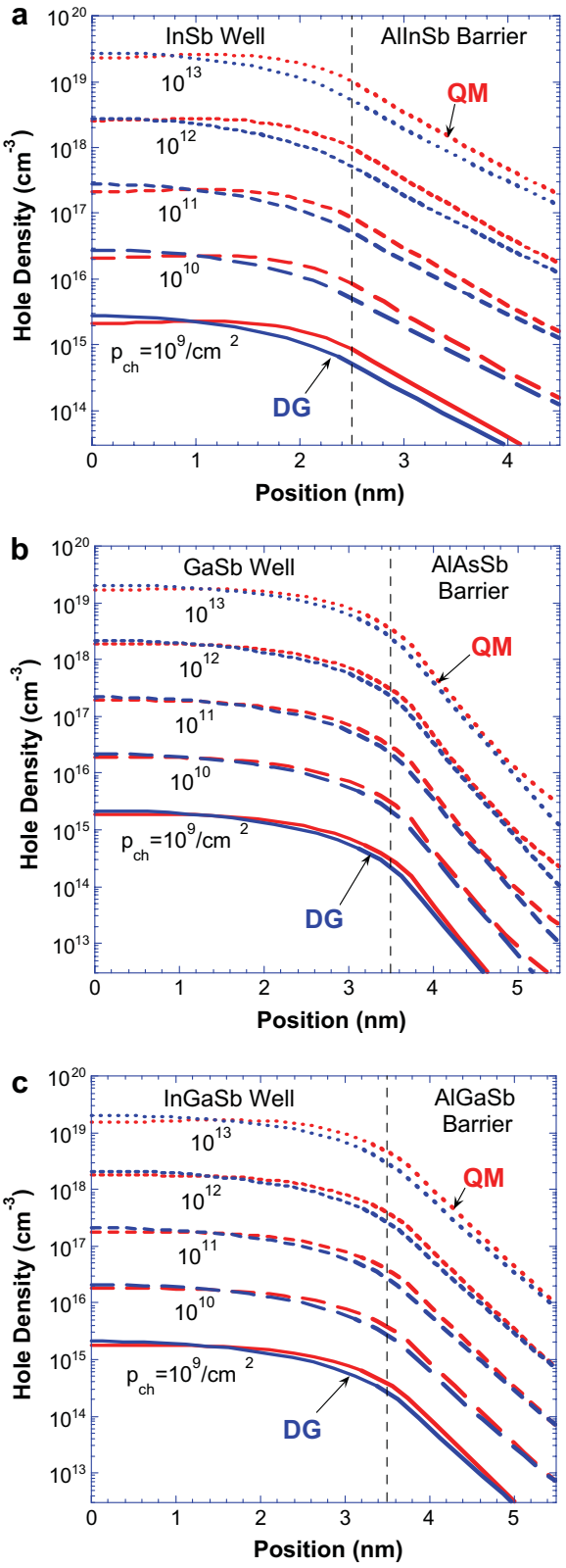

Fig. 3. (a–c) Density profiles across the quantum wells studied in this paper as simulated by quantum mechanics and fit by DG theory. The perpendicular components of the DG effective masses as determined by these fits are given in Table 2.

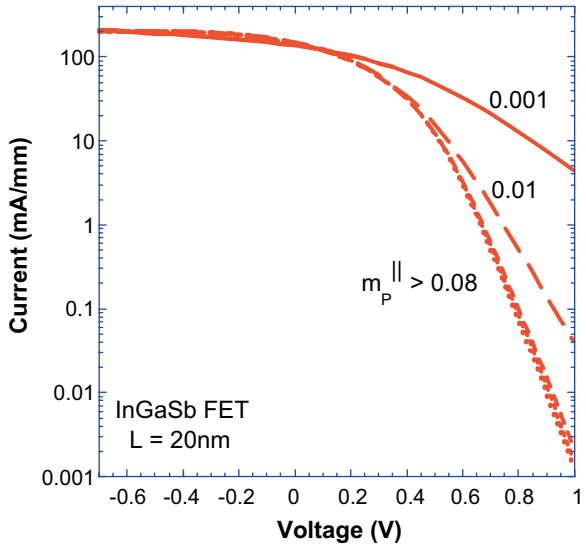

Fig. 4. Simulated I–V curves for a 20 nm InGaSb FET with varying parallel DG effective masses and showing that the parallel mass is likely not too important for this paper.

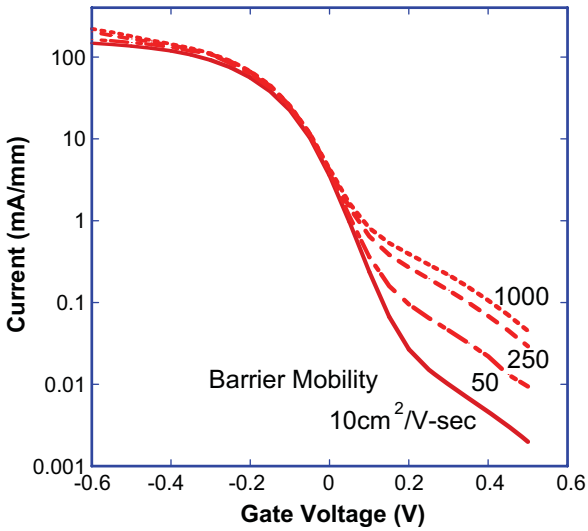

Fig. 5. Simulated I–V curves for a 40 nm InSb FET with varying hole barrier mobility and showing its importance for modeling gate leakage in the InSb devices.

short channels, may be neglected for the gate lengths of interest here. Barrier transport is most germane to understanding source-drain leakage, an issue that is especially important for low valence-band-barrier InSb pFETs (see Fig. 2a) where it dominates gate current at short gate lengths [6]. This leakage current is illustrated in the simulated I–V curves in Fig. 5 where the barrier mobility is treated as a parameter. The particular value of the barrier mobility as given in Table 2 (50 cm^2/V sec) has been chosen so as to fit the source-drain leakage current observed in [6]; this same value is arbitrarily used for the other barrier materials as well.

Lastly, generation–recombination processes may have to be considered for the materials of interest, and especially for InSb with its narrow gap [20]. For this paper, however, we neglect these processes as 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 [21], and because of the very small generation volume.

The final element of the description is a set of consistent boundary conditions, all of which take well-known forms. It is important to note that, among the choices made, we assume *ideal* contacts

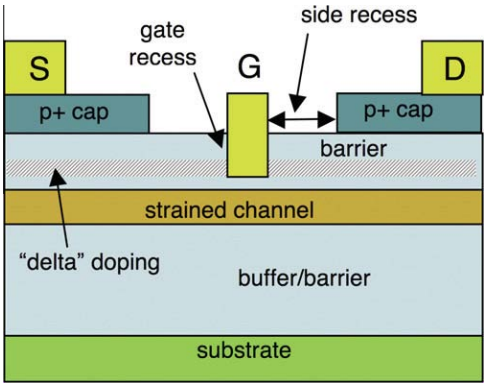

Fig. 6. Schematic of the generic device structure simulated in this paper.

with zero contact resistance; that this is often inappropriate for III–V materials means that our results should be regarded as "best-case". With all the boundary conditions prescribed, boundary value problems relevant to the device situations of interest can be formulated and solved. The generic device structure studied in this work is the field effect transistor depicted in Fig. 6. The channels and barrier materials are the three discussed earlier and specified in Table 2. A "delta" doping is often included in the structures in a normal frontside configuration as shown in Fig. 6. In addition, doped p^+ caps are incorporated with a variable side-recess spacing separating them from the gate. Lastly, the gate is sometimes vertically recessed as shown in Fig. 6 in order to enhance the charge control. The particular gate-to-channel spacings considered were 10 nm (with no gate recess) and 4 nm (with a gate recess). The gate lengths ranged from 0.5 μm down to 20 nm.

For this paper, the boundary value problems are solved numerically using the Comsol Multiphysics finite element package [22]. A sample DG solution is shown in Fig. 7 for a non-recessed InSb FET of [5] with a 40 nm 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. 8 for V_D=−0.5 V. In the plot we show the experimental I–V data from [5] as well as simulated curves for the comparable InGaSb and GaSb devices. For the InSb device, simulation and experiment are seen to be in good agreement, though of course to a large extent this is a consequence of curve-fitting, e.g., the choices for the parameters in (3).

3. Basis of technology comparisons

As noted in the Introduction, the III–V pFET technology discussed herein is envisioned as supplying components for complementary circuits. The likely application space is low-power/high-speed electronics, e.g., for wireless applications, and could be implemented either as pure III–V at relatively low levels of integration (say 10^3–10^4 transistors), or as hybrids with much denser but slower silicon devices. The primary advantage offered by the antimonides is the ability to sustain high-speed at low operating voltages and for this paper we therefore fix V_{DD} at −0.5 V. This of course strongly impacts both the static and dynamic power dissipation. It also imposes strong requirements (which we do not consider in detail) on the control of the threshold voltages and a variety of other quantities.

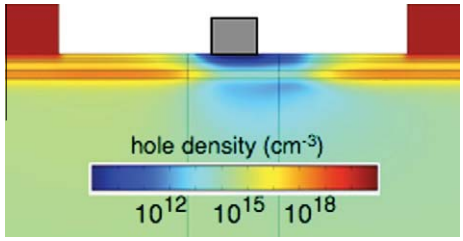

Fig. 7. Simulated hole density in a 40 nm InSb FET at pinch-off.

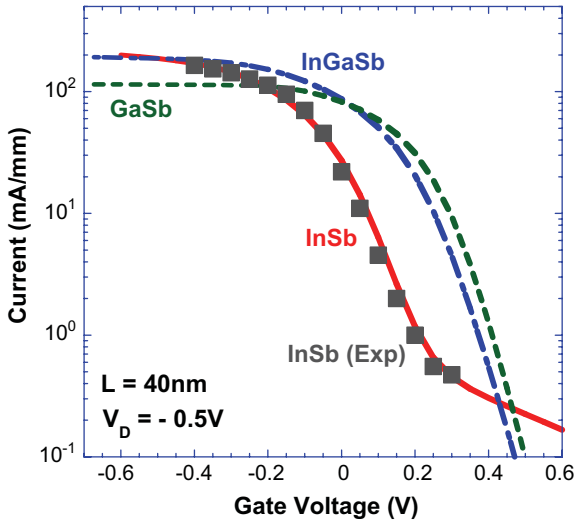

Fig. 8. Simulated drain characteristics of InSb, GaSb and InGaSb pFETs plus experimental data for InSb from [6].

For a general comparison of the three types of antimonide pFETs under consideration, we follow [23] in utilizing certain benchmarks relevant to digital circuits such as the on–off ratio ($I_{rat} \equiv I_{ON}/I_{OFF}$), the gate delay ($\tau_D \equiv C_G V_{DD}/I_{ON}$), the dynamic-energy-delay-product ($E_D\tau_D \equiv C_G V_{DD}/I_{ON} \times C_G V_{DD}^2$), the subthreshold slope ($S \equiv \min\{dV_G/d\log(I_D)\}$), and the drain-induced barrier lowering ($\mathrm{DIBL} \equiv dV_T/dV_D$). To gauge the devices from the perspective of standby power, we also consider the static power dissipation ($P_S \equiv I_{OFF}V_{DD}$), and a new benchmark defined by $f_D \equiv \log(I_{rat})/(2\pi\tau_D)$. The latter quantity is a characteristic frequency that emphasizes the trade-off between a short gate delay and a high on–off ratio, and is thus a sort of (inverse) static energy-delay product.

A crucial choice for our technology comparisons is the "proper" selection of the threshold voltage. Once known, we shall employ a "2/3–1/3" rule [23] as depicted in Fig. 9 to define I_{ON} and I_{OFF}, and from thence (with C_G estimated by simulation) determine the various other benchmarks. From Fig. 9 it should be clear that, as V_T is varied, there is a general trade-off between speed and power, e.g., a more positive V_T will cause I_{rat} to improve (increase) as I_{OFF} gets exponentially smaller, while at the same time τ_D will degrade (increase) as I_{ON} drops. A convenient way of optimizing this trade-off

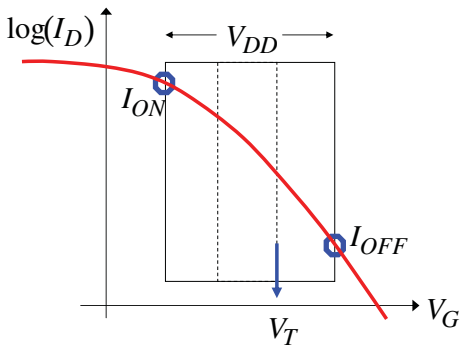

Fig. 9. A depiction of the use of the "2/3–1/3" rule to relate V_{DD}, V_T, I_{ON} and I_{OFF}.

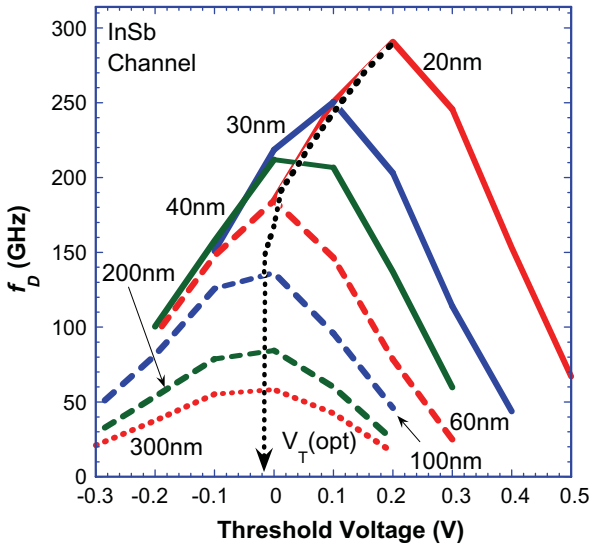

Fig. 10. The simulated figure-of-merit f_D versus V_T for non-recessed InSb pFETs with various gate lengths. The dotted curve highlights the "optimum" V_T's at each gate length as determined by this plot.

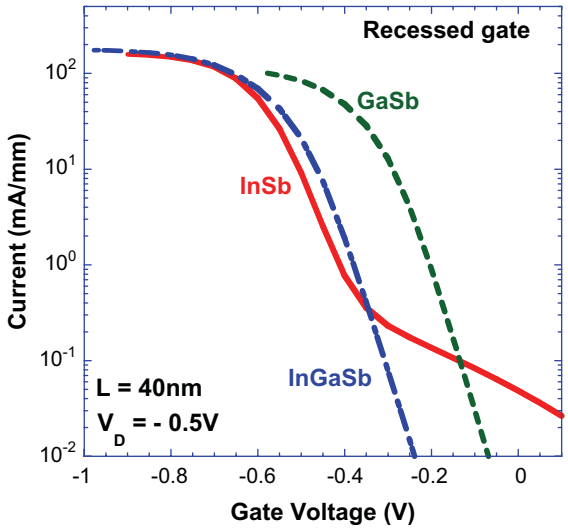

Fig. 11. Simulated transfer curves for 40 nm InSb, GaSb and InGaSb pFETs with recessed gates and showing enhancement-mode operation.

and thereby finding the "best" V_T is to maximize f_D.[1] A sample plot of computed f_D's as a function of the choice of V_T for non-recessed InSb pFETs with various gate lengths is shown in Fig. 10. Fixing V_T by the maxima of these curves, we observe that there is a "long-channel" regime in which the optimum V_T is essentially constant, and then deviations from that value occur as the channels get shorter (see further discussion in Section 4). Again, once the V_T values have been selected, then all of the benchmarks may be computed and the technology comparisons made.

As is well known, unless the semiconductor materials of interest have been extensively characterized, numerical device modeling is at best of semi-quantitative value, and is probably of most use for parametric studies. For this reason, experimental checks are of great importance. To date the only experimental scaling study of p-channel Sb devices in the literature is that on non-recessed InSb pFETs in [6]. A first basic agreement between modeling and their experiments was the transfer curve in Fig. 7 for a 40 nm device. Where possible, we include comparisons with the data from [6] elsewhere in this paper, and in general find that these tend to support the validity of our simulations. Of course, as noted earlier, our simulations involve a significant amount of curve-fitting, and so our modeling methods, parameters and results should be viewed more as a means to greater understanding than as definitive models and predictions.

4. Simulation results and discussion

In the scaling of heterostructure transistors, a crucial parameter is the aspect ratio formed of the gate length divided by the gate-to-channel spacing. To achieve good scaling the usual rule-of-the-thumb is that one should maintain an aspect ratio in the range 3–6 [24,25]. Hence, for a 20 nm device one would need a gate-to-channel spacing of roughly 4–6 nm.[2] In typical designs this implies the use of an etched gate recess. As is well known, the gate recess also makes possible enhancement-mode devices both by reducing the distance over which the voltage drops, and by eliminating any front-side modulation doping in the gate region (see Fig. 6). In Fig. 11 we show simulated I–V curves for recessed pFET designs analogous to the non-recessed devices of Fig. 8, and with a gate-to-channel spacing of 4 nm. The characteristics look quite similar to the earlier ones (because of the 40 nm gate length) except that the devices are now clearly enhancement-mode. In the remainder of this paper we compare various simulated characteristics and benchmarks for antimonide-based pFETs using both the non-recessed and recessed device designs.

As studied by Kim and co-authors [26], another important parameter in the design of heterostructure transistors is the side-recess spacing between the gate and the doped cap layers (see Fig. 6). On the one hand, as shown in Fig. 12 for a 40 nm recessed-gate InGaSb pFET, as the side-recess distance becomes large, the increase in access resistance degrades the transconduc-

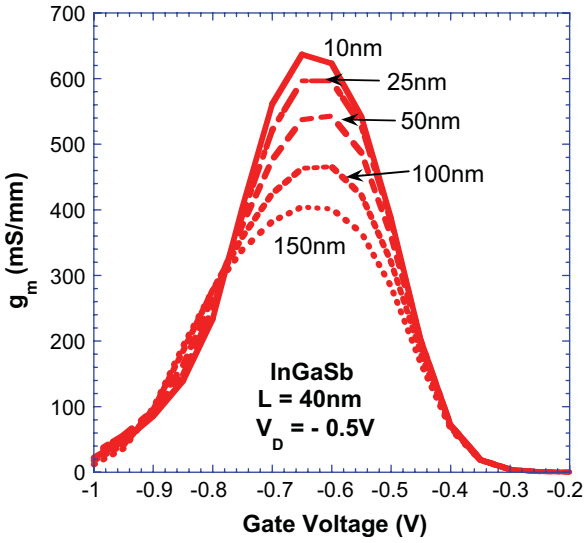

Fig. 12. Simulated effect of the side-recess width on the transconductance of 40 nm InGaSb pFETs with recessed gates. The degradation seen is due to increased access resistance.

[1] This procedure is of course just one convenient way of optimizing the devices. What is actually "optimum" depends on the application, and other choices that favor a different balance between speed and power could readily be preferred.

[2] For convenience, the gate-to-channel spacings mentioned in this paper refer to the distance between the gate and the *top* of the channel. However, from the perspective of electrostatic control, a better measure is the distance to the *center* of the channel.

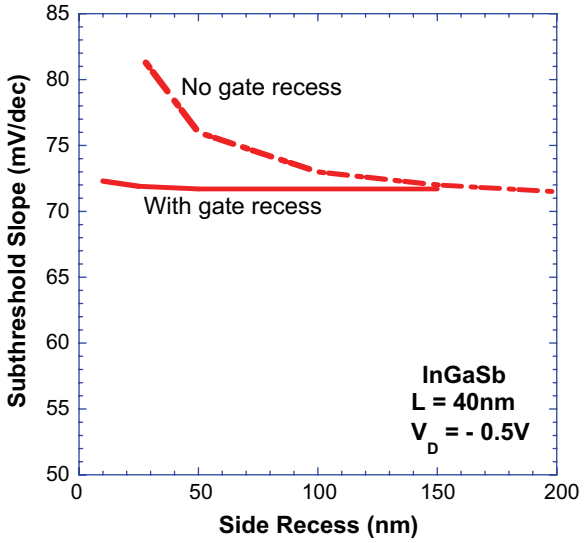

Fig. 13. Simulated degradation of subthreshold slope with decreasing side-recess width of 40 nm InGaSb pFETs. The dependence on aspect ratio shows this to be a short-channel effect.

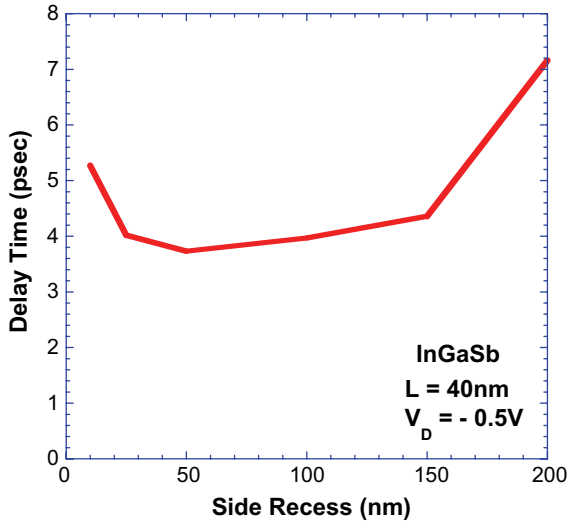

Fig. 14. Simulated effect of side-recess width on delay time for 40 nm InGaSb pFETs with recessed gates. The rise in delay time for small side-recess widths is due to increased gate capacitance.

tance. On the other hand, as the side-recess distance becomes small two deleterious effects enter. First, the nearby p^+-cap can influence the electrostatic barrier shape and thereby degrade the subthreshold slope. This is seen in Fig. 13 where simulated results for 40 nm InGaSb pFETs both with and without recessed gates are plotted; that the impact of the p^+-cap is much greater without a gate recess shows this phenomenon to be a short-channel effect. The second consequence of having a small side-recess distance is an increased gate capacitance to the proximal p^+-cap, which in turn degrades the delay time as shown in Fig. 14 for 40 nm InGaSb pFETs. In this figure the delay times are elevated because the full capacitance for a particular (and overly large) gate has been used in the calculation in order to include fully the p^+-cap effect. Based on this plot, for the remainder of this paper we fix the side-recess spacing at 50 nm.

Fig. 15a is a summary plot of the threshold voltages as a function of gate length as determined by the method described in Section 3 for the three antimonide technologies both with and without the gate recess. A companion plot in Fig. 15b gives the corresponding values for the figure-of-merit f_D in each case. As was noted earlier, Fig. 15a shows the non-recessed designs to be depletion-mode whereas the recessed pFETs are enhancement-mode. Also, much as in Fig. 10, in each case there exists a clear long-channel regime in which the threshold voltage is essentially constant, and then deviations from this value occur when the channel length becomes short enough. As expected, inclusion of recessed gates greatly improves the short channel performance by virtue of the higher aspect ratio and better electrostatic control. With non-recessed designs the long-channel regime ends at around 60–80 nm, whereas for recess designs it extends down to 30–40 nm. Of all the devices, those with InSb channels show the least short-channel effects, presumably because of their narrower (5 nm) channel widths and thus larger aspect ratios. With regard to the figure-of-merit f_D plotted in Fig. 15b, without a gate recess all the devices provide roughly the same level of performance, but when the gate recess is included there is a clear ordering with InGaSb being best and GaSb being worst. The reasons for this will be discussed further below.

A scaling plot of the delay times τ_D (obtained assuming the V_T's of Fig. 15a) for the various antimonide pFETs is shown in Fig. 16. For non-recessed devices the delay times for all the antimonide de-

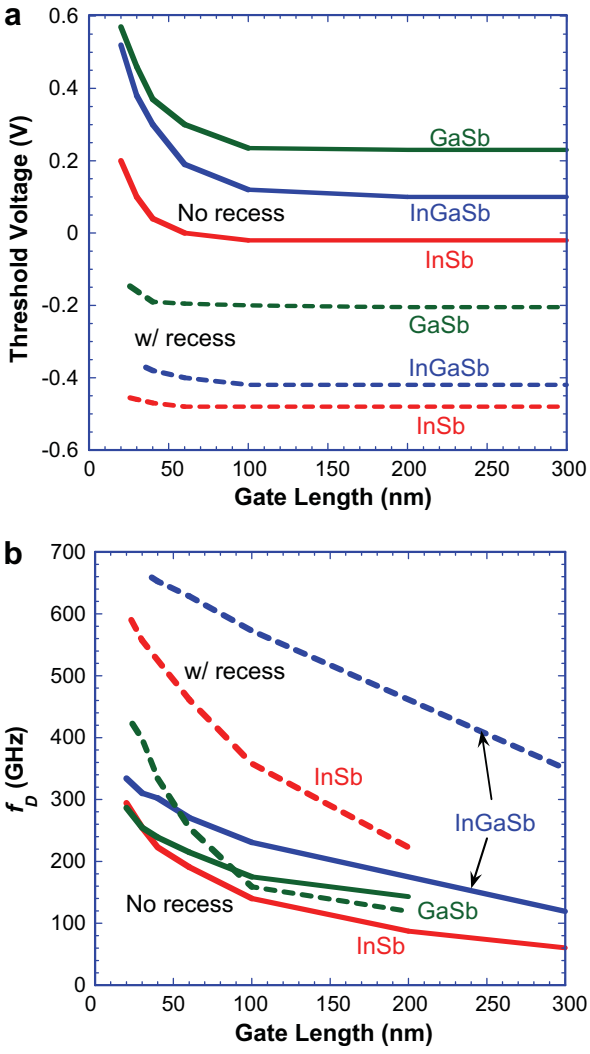

Fig. 15. DG simulated (a) V_T's and (b) f_D's versus gate length as determined by the method depicted in Fig. 10 for InSb, GaSb and InGaSb pFETs both with and without recessed gates.

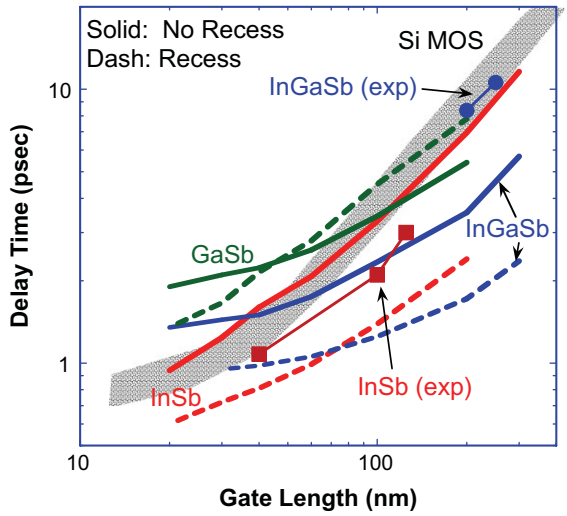

Fig. 16. Scaling plot of the simulated delay times for InSb, GaSb and InGaSb pFETs both with and without recessed gates. Shown for comparison are experimental values for InSb [6], InGaSb [6], and Si PMOS with the latter operated at higher V_{DD}.

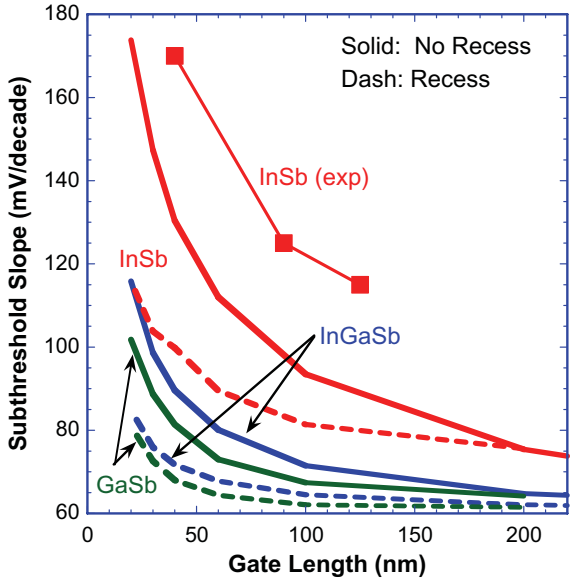

Fig. 17. Scaling plot of the simulated subthreshold slopes for InSb, GaSb and InGaSb pFETs both with and without recessed gates. Shown for comparison are experimental values for InSb [5].

vices are seen to be comparable. However, just as in Fig. 15b, in the case of recessed devices the indium-containing pFETs are better with the InGaSb being somewhat better for longer channels and the InSb devices, with their better electrostatic control, performing best for short channels. Included also in Fig. 16 are some experimental data for non-recessed InSb devices [6] and InGaSb devices [7]. The agreement with experiment is reasonably good for the InSb devices, but less so for the InGaSb devices which had a relatively high contact resistance. Finally, although less comparable because V_{DD} is generally significantly higher (0.8 V or higher), a broad line appears in Fig. 16 to indicate the rough level of performance obtained by Si PMOS devices.

We next look closer at short-channel effects as reflected in the subthresold slopes and the DIBL for the various antimonide devices. Fig. 17 plots the calculated subthreshold slopes versus gate length and, generally speaking, follows the anticipated trends. However, in contrast to Fig. 15a, here the excessive source-drain leakage in the InSb device makes its deviations from ideality greatest. The measured subthreshold slopes from [6] for InSb devices are also shown in Fig. 17, and are seen to be in reasonable agreement with simulation.[3] The corresponding scaling plot for the DIBL appears in Fig. 18, and the comparisons with the modeling are much the same.

For the last technology comparisons we evaluate the various antimonide devices with respect to power/energy consumption. A scaling plot of the dynamic energy-delay product is given in Fig. 19. The results are analogous to those of Fig. 16 with the most important point being the basic one that, from a dynamic energy perspective, the recessed antimonide devices are about a factor of 10 better than silicon technology because of the former's possibility of higher speed at lower voltages.

With respect to static power, as emphasized in the Introduction, the basic motivation for our investigating III–V pFETs is as a pathway to III–V CMOS and to minimizing standby power. For this paper where we ignore gate leakage (or equivalently, assume the ideal case of zero gate leakage) the standby power is set entirely by source-drain leakage. The associated power per unit width and its scaling characteristics are plotted in Fig. 20 for the various antimonide devices. Generally speaking, because of their relatively

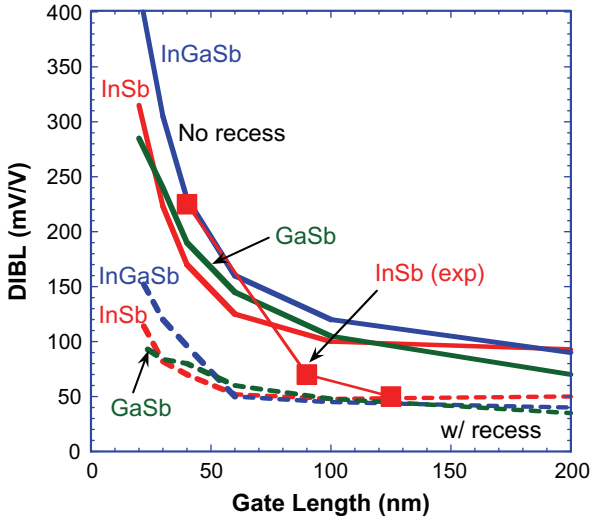

Fig. 18. Scaling plot of the simulated DIBL for InSb, GaSb and InGaSb pFETs both with and without recessed gates. Shown for comparison are experimental values for InSb FETs with unrecessed gates [6].

high leakage currents, the InSb devices are one to two orders of magnitude worse than the InGaSb and GaSb devices. To give context to these values, in Fig. 20 we also show ITRS roadmap values [27] out to the 22 nm node for the source-drain off-state power (again ignoring gate leakage) for three different technology scenarios all operating at around V_{DD} = 1 V: High-Performance Logic, Low Operating Power, and Low Standby Power. Clearly, the III–V devices have potential for the coming years, though unless effective insulator [28] and contact [29] technologies are available, this potential will not be realized.

5. Final remarks

As the "endgame" of the electronics revolution as envisioned by the ITRS roadmap [27] approaches, the semiconductor industry is

[3] In silicon devices interface traps are known contributors to the subthreshold slope, but given the high quality of the III–V heterointerfaces this effect is not likely of importance here.

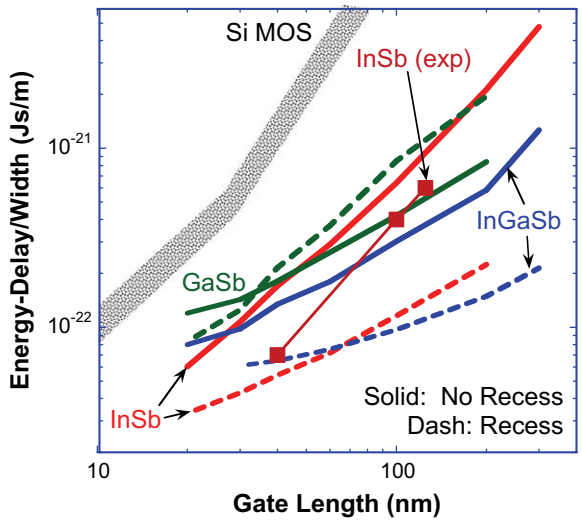

Fig. 19. Scaling plot of the simulated energy-delay for InSb, GaSb and InGaSb pFETs both with and without recessed gates. Shown for comparison are experimental values for InSb [6] and Si PMOS with the latter operated at higher V_{DD}.

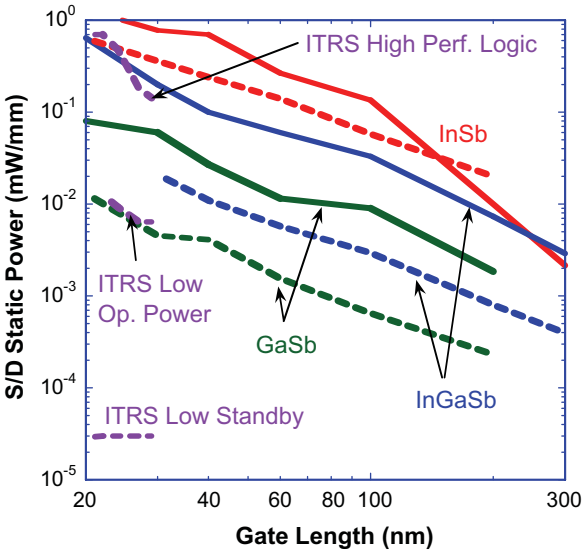

Fig. 20. Scaling plot of the static power dissipation for InSb, GaSb and InGaSb pFETs both with and without recessed gates. Shown for comparison are the ITRS projected requirements for three different technology scenarios out to the 22 nm node.

increasingly faced with the limitations of Si/SiO_2 devices and with pressure to consider alternatives. High-performance III–V materials have long been a possibility, but a variety of factors have always limited their use to niche applications. In the area of power dissipation, despite the III–V advantage in speed at low voltage, their unsuitability for complementary circuits has made high standby power a critical limitation. III–V CMOS has been difficult to achieve for a variety of reasons, with a main one being the lack of a well-matched p-channel device. But it may be that this situation is changing given recent progress in the fabrication of high-mobility III–V FETs based on antimonide-containing semiconductors [1]. To explore the performance possibilities of such transistors, in this paper we have carried out numerical simulations of pFETs fabricated with the new, high-quality InSb, GaSb and InGaSb transistors.

The numerical calculations were performed using the density-gradient approximation to quantum transport theory. Particular FET designs were studied both with and without recessed gates, and with InSb, GaSb and InGaSb channels that had state-of-the-art transport properties. Emphasis was placed on the scaling behavior and on ultimate performance limits. In general, our simulations showed the pFETs to be quite competitive with silicon. Of the three antimonide technologies studied, the InGaSb-channel devices seem to provide the best compromise between speed and power. From a speed perspective, InSb had the edge when the FETs were aggressively scaled, however, in our simulations this was mainly due to their better aspect ratio.

Our work points up several aspects of the antimonide pFETs that need further development/improvement. First, higher mobility material in thinner channels is advantageous. This might be achieved by further improving the quality of the materials and interfaces. In addition, one might work to boost the biaxial strain levels (especially in GaSb) and/or to exploit uniaxial strain [9,10]. Second, a robust approach to making very thin insulated gates with self-aligned source/drains and good threshold control is essential [28]. Third, an approach to low-resistance contacts, possibly using regrowth techniques [29], is needed. And finally, while the band diagrams in Fig. 2a–c and the work of [12] demonstrate that the antimonides share the attribute of allowing both n-channel and p-channel transistors to be fabricated in the same channel material, a well-integrated III–V CMOS must be demonstrated, characterized and optimized.

Acknowledgment

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

References

[1] See, e.g., Heyns M, Tsai W, editors. MRS Bulletin entitled: Ultimate scaling of CMOS logic devices with Ge and III–V materials; July 2009.
[2] Bennett BR, Magno R, Boos JB, Kruppa W, Ancona MG. Solid-State Electron 2005;49:1875.
[3] Datta S, et al. 85nm gate length enhancement and depletion mode InSb quantum well transistors for ultra high speed and very low power digital logic applications, IEDM Tech Dig; 2005 [Fig. 14].
[4] Chan YJ, Pavlidis D. Single and dual p-doped channel $In_{0.52}Al_{0.48}/In_xGa_{1-x}As$ ($x = 0.53,0.65$) FETs and the role of doping. IEEE Trans Electron Dev 1992;39:466.
[5] Lee ML, Fitzgerald EA, Bulsara M, Currie MT, Lochtefeld A. Strained Si, SiGe and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors. J Appl Phys 2005;97:011101.
[6] Radasavljevic M et al. High-performance 40 nm gate length InSb p-channel compressively strained quantum well field effect transistors for low-power ($V_{CC} = 0.5V$) logic applications. IEDM Tech Dig 2008;727.
[7] Bennett BR, Ancona MG, Boos JB, Canedy CB, Khan SA. Strained GaSb/AlAsSb quantum wells for p-channel field effect transistors. J Cryst Growth 2008;311:47.
[8] Bennett BR, Ancona MG, Boos JB, Shanabrook BV. Mobility enhancement in strained p-InGaSb quantum wells. Appl Phys Lett 2007;91:042104.
[9] Liao C, Cheng KY. Hole mobility in pseudomorhpic InGaSb quantum well modulation doped with carbon. J Vac Sci Technol 2010;B28:C3C29.
[10] Thompson SE, Sun G, Choi YS, Nishida T. Uniaxial-process-induced strained-Si: extending the CMOS roadmap. IEEE Trans Electron Dev 2006;53:1010.
[11] Nainani A, Raghunathan S, Witte D, Kobayashi M, Irisawa T, Krishnamohan T, et al. Engineering of strained III–V heterostructures for high hole mobility. IEDM Tech Dig 2009;857.
[12] Kim D-H, del Alamo JA. 30nm InAs Pseudomorphic HENTs on an InP substrate with a current-gain cutoff frequency of 628GHz. IEEE Electron Dev Lett 2008;29:830.
[13] Birner S et al. Nextnano: general-purpose 3-D simulations. IEEE Trans Electron Dev 2007;54:2137.
[14] Bennett BR, Ancona MG, Champlain JG, Papanicolaou NA, Boos JB. Demonstration of high-mobility electron and hole transport in a single InGaSb well for complementary circuits. J Cryst Growth 2009;312:37.
[15] (a) Ancona MG, Tiersten HF. Macroscopic physics of the silicon inversion layer. Phys Rev B 1987;35:7959;
(b) Ancona MG, Iafrate GJ. Quantum correction to the equation of state of an electron gas in a semiconductor. Phys Rev 1989;B39:9536.
[16] Ancona MG, Boos JB, Bennett BR. Using density-gradient theory to model Sb-based p-channel FETs. In: Proc SISPAD; 2009.
[17] Ancona MG. Density-gradient theory analysis of electron distributions in heterostructures. Superlatt Microstr 1990;7:119.

[18] Brown AR, Martinez A, Seoane N, Asenov A. Comparison of density gradient and NEGF for 3D simulation of a nanowire MOSFET. In: Proc 2009 Spanish conf elect dev; 2009. p. 140.
[19] Ancona MG. Macroscopic description of quantum-mechanical tunneling. Phys Rev 1990;B42:1222.
[20] Passlack M. Off-state current limits of narrow bandgap MOSFETs. IEEE Trans Electron Dev 2006;53:2773.
[21] Bolognesi CR, Dvorak MW, Chow DH. Impact ionization suppression by quantum confinement: effects on the dc and microwave performance of narrow-gap channel InAs/AlSb HFETs. IEEE Trans Electron Dev 1999;46:826.
[22] See http://www.comsol.com.
[23] Chau R, Datta S, Doczy M, Doyle B, Jin B, Kavalieros J, et al. Benchmarking nanotechnology for high-performance and low-power logic transistor applications. IEEE Trans Nanotechnol 2005;4:153.
[24] Das MB. A high aspect ratio design approach to millimeter-wave HEMT structures. IEEE Trans Electron Dev 1985;32:11.
[25] Awano Y, Kosugi M, Kosemura K, Mimura T, Abe M. Short-channel effects in subquarter-micrometer-gate HEMTs: simulation and experiment. IEEE Trans Electron Dev 1989;36:2260.
[26] Kim D-H, del Alamo JA, Lee J-H, Seo K-S. The impact of side-recess spacing on the logic performance of 50 nm $In_{0.7}Ga_{0.3}As$ HEMTs. In: Proc 18th IEEE IPRM conf; May 2006. p. 177–80.
[27] 2008 ITRS Roadmap. <http://www.itrs.org>.
[28] Radosavljevic M et al. Advanced high-K gate dielectric for high-performance short-channel $In_{0.7}Ga_{0.3}As$ quantum well field effect transistors on silicon substrate for low power logic applications. IEDM Tech Dig 2009;319.
[29] Singisetti U et al. $In_{0.53}Ga_{0.47}As$ channel MOSFETS with self-aligned InAs source/drain formed by MEE regrowth. IEEE Electron Dev Lett 2009;30:1128.