Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors

Y. Sun,[a] S. E. Thompson, and T. Nishida
Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116200, Gainesville, Florida 32611-6200

(Received 8 December 2006; accepted 21 March 2007; published online 18 May 2007)

A detailed theoretical picture is given for the physics of strain effects in bulk semiconductors and surface Si, Ge, and III–V channel metal-oxide-semiconductor field-effect transistors. For the technologically important in-plane biaxial and longitudinal uniaxial stress, changes in energy band splitting and warping, effective mass, and scattering are investigated by symmetry, tight-binding, and $\mathbf{k}\cdot\mathbf{p}$ methods. The results show both types of stress split the Si conduction band while only longitudinal uniaxial stress along ⟨110⟩ splits the Ge conduction band. The longitudinal uniaxial stress warps the conduction band in all semiconductors. The physics of the strain altered valence bands for Si, Ge, and III–V semiconductors are shown to be similar although the strain enhancement of hole mobility is largest for longitudinal uniaxial compression in ⟨110⟩ channel devices and channel materials with substantial differences between heavy and light hole masses such as Ge and GaAs. Furthermore, for all these materials, uniaxial is shown to offer advantages over biaxial stress: additive strain and confinement splitting, larger two dimensional in-plane density of states, smaller conductivity mass, and less band gap narrowing. © *2007 American Institute of Physics*.
[DOI: 10.1063/1.2730561]

I. INTRODUCTION

After 40 years of gate length reduction, planar metal-oxide-semiconductor field-effect transistor (MOSFET) scaling has ended at a gate length of ~20–30 nm depending on off-state leakage requirements. However, MOSFET performance improvements are continuing via strained channels which are being adopted in nearly all 90, 65, and 45 nm logic, communication, and consumer technologies.[1–12] The discovery of strained Si to enhance MOSFET performance originated in the research of growing Si layers on SiGe buffers in the early 1980s,[13,14] but the physical mechanism for the enhancement can be traced back to the critical work on deformation potentials by Bardeen and Shockley[15,16] and experimental measurements of piezoresistance by Smith.[17] Recently a detailed review of the history and current progress in the experimental research on high-mobility biaxially strained Si, SiGe, and Ge channel MOSFETs was provided by Lee *et al.*[18] Chidambaram *et al.* also reviewed the properties of strained Si with an emphasis on its manufacturability and scalability in complementary metal-oxide-semiconductor (CMOS) technology.[19] Together with the recent theoretical work by Takagi[20] and Fichetti[21] on *n*-channel MOSFETs (*n*MOSFETs) and the work by Oberhuber and Vogl[22] and Fischetti[23] on *p*-channel MOSFETs (*p*MOSFETs), a fairly complete theoretical framework on biaxial stress-enhanced mobility has emerged. However, since to date only uniaxial stress has been used in production of nanoscale MOSFETs, a comprehensive picture of the physics of both biaxial and uniaxial stressed Si and potential future channel materials is needed.

In this article, we elucidate the physics underlying the strain effects on semiconductors and MOSFETs by reviewing and examining the problems at different levels, with the focus on uniaxial[5,7,24,25] and biaxial stress.[26–30] We first discuss how strain alters the crystal symmetry and band structure. Next, the tight-binding picture and $\mathbf{k}\cdot\mathbf{p}$ method are used to explain conduction and valence band shift and strain altered band curvature (warping). Finally, we investigate the effects of strain on group IV and III–V channel MOSFETs including electric confinement and two-dimensional (2D) density of states (DOS). In this article, the channel direction of a MOSFET is taken along ⟨110⟩ (which for commercial technologies is generally but not always the case), and the wafer is (001) surface-oriented (currently the case for all commercial technologies). The term "in-plane" refers to the x-y plane and the "out-of-plane" direction is along z.

II. BRIEF HISTORY OF STRAINED SEMICONDUCTORS

Strain is a relatively old topic in semiconductor physics. In semiconductors, strain results from phonon-induced lattice vibrations, lattice mismatched film growth, and applied external stress. To account for the conduction band energy shift with strain, deformation potential theory was developed by Bardeen and Shockley[15,16] to model the coupling between acoustic waves and electrons in solids. Next, Herring and Vogt[31] in 1955 used deformation potentials to model transport in strained semiconductors. In deformation potential theory, the strain induced band edge shift is proportional to the strain tensor

[a]Electronic mail: yksun@tec.ufl.edu

 © 2007 American Institute of Physics
 TSMC EXHIBIT 1061

TABLE I. Deformation potentials in conduction band valleys. After Herring and Vogt (see Ref. 31).

	Γ	100	110	111
$\Xi_1=\partial\varepsilon/\partial u_{xx}$	Ξ_d	$\Xi_d+\Xi_u$	$\Xi_d+\Xi_u-\frac{1}{2}\Xi_p$	$\Xi_d+\frac{1}{3}\Xi_u$
$\Xi_2=\partial\varepsilon/\partial u_{yy}$	Ξ_d	Ξ_d	$\Xi_d+\Xi_u-\frac{1}{2}\Xi_p$	$\Xi_d+\frac{1}{3}\Xi_u$
$\Xi_3=\partial\varepsilon/\partial u_{zz}$	Ξ_d	Ξ_d	$\Xi_d-\Xi_u+\Xi_p$	$\Xi_d+\frac{1}{3}\Xi_u$
$\Xi_4=\partial\varepsilon/\partial u_{yz}$	0	0	0	$\frac{1}{3}\Xi_u$
$\Xi_5=\partial\varepsilon/\partial u_{zx}$	0	0	0	$\frac{1}{3}\Xi_u$
$\Xi_6=\partial\varepsilon/\partial u_{xy}$	0	0	$\frac{1}{3}\Xi_p$	$\frac{1}{3}\Xi_u$

$$\Delta E=\sum_{ij}\Xi_{ij}e_{ij}, \tag{1}$$

where Ξ are the deformation potentials. Deformation potential theory is still the primary method to model the band shift and warping via band calculations.[22,32]

The deformation potentials are constrained by symmetry at different conduction band valleys as summarized by Herring and Vogt[31] which is reproduced in Table I. Due to symmetry, cubic crystals have at most three independent deformation potentials at any **k** point for a specific energy level. The constants Ξ_d, Ξ_u, and Ξ_p characterize the conduction band valleys where Ξ_d is due to pure dilation (hydrostatic strain, i.e., change in size but not shape), Ξ_u is related to a pure shear involving a uniaxial tension along a major axis and a symmetrical compression along the other two minor axes (with no volume change), and Ξ_p is required in ⟨110⟩ valleys by symmetry for describing the shift due to a shear in the (001) plane containing the valley axis. At the Γ point, another set of symbols are commonly used: a_c, a_v, b, and d, where $a_c=\Xi_{d\Gamma}$, and a_v, b, and d are three independent valence band deformation potentials.

The first experimental data for strain-enhanced mobility in *n*- and *p*-type bulk Si and Ge was measured by Smith in 1954.[17] Using deformation potentials, Herring and Vogt[31] ascribed the electron mobility change to "electron transfer" and intervalley scattering rate change caused by valley energy shift. This is essentially the same physics that explains strain-enhanced mobility in Si *n*-channel MOSFETs at present. The exact quantification of these effects is still controversial because of a lack of agreement on the different scattering parameter sets.[20]

The picture of strain-enhanced hole mobility is more complicated. Because of strong band warping,[24] strain effects on valence bands cannot be simply explained by band shifts using deformation potentials and were studied later when more accurate band calculation approaches such as the $\mathbf{k}\cdot\mathbf{p}$ method[33,34] became available. Pikus and Bir[35] and Kleiner and Roth[36] constructed a strain Hamiltonian in terms of the angular momentum derived by symmetry considerations. Using this Hamiltonian in 1963, Hasegawa[37] and Hensel and Feher[38] systematically studied the valence band effective masses and deformation potentials in strained Si and compared them with cyclotron resonance experiments. In order to determine the valence band parameters, they calculated the strain splitting and mass change under several stress conditions. Hasegawa found that under ⟨001⟩ compressive stress, the light-hole (LH) mass becomes larger and the inverse of it varies linearly with stress due to the mixing with the split-off-hole band, and the heavy-hole (HH) mass does not change to the first order with strain. Hensel and Feher showed that under ⟨001⟩ or ⟨111⟩ compressive stress, the LH band shifts up (becomes the top valence band) and the HH band shifts down in energy. Although they only concentrated on the zone center ($k=0$) and did not discuss the strain effects on carrier transport, they revealed the key factors that affect the hole mobilities in semiconductors—band splitting and warping, mass change, and consequently the change of DOS which alters band occupation and phonon scattering.

To model strain effects in semiconductors, accurate deformation potentials are required to calculate both band shift and band warping via band calculations. Deformation potentials have been measured by many researchers: Hensel[39] using cyclotron resonance, Balslev[40] and Bhargava and Nathan[41] using photoluminescence, and Pollak and Cardona[42] using piezoelectroreflectance. In all these experiments, the authors studied the strain splitting-induced spectrum difference under various stresses. However, the hydrostatic deformation potentials for separate bands are difficult to measure and remained uncertain (e.g., $-10.7\rightarrow 1.13$ eV for Si conduction band) until recently.[32] Lim *et al.*[43] investigated the change in the gate tunneling currents of Si *n*MOSFETs induced by externally applied mechanical stress as a function of gate bias, and determined both the hydrostatic and shear deformation potentials in the Si conduction band, which were consistent with the theoretical results obtained by Fischetti[32] from pseudopotential calculations.

Following three decades of work on piezoresistance, strain altered bulk mobility, and deformation potentials, strain has emerged as a technique to enhance MOSFET performance. Wesler *et al.*[44] in 1992 demonstrated a strained Si *n*-channel MOSFET (which used biaxial stress) with electron mobility enhanced by 2.2 times. Nayak *et al.*[45] reported a biaxial stressed Si *p*-channel MOSFET in 1993 with 1.5× enhanced hole mobility. The idea of using longitudinal uniaxial stress for improving device performance in MOSFETs was activated by the investigations of Ito *et al.*[46] and Shimizu *et al.*[47] in the late 1990s who introduced high stress capping layers[48] deposited on MOSFETs to induce channel stress. Gannavaram *et al.*[49] proposed SiGe in the source and drain area for higher boron activation and reduced external resistance, which also furnished a technically convenient means to employ uniaxial channel stress. Larger mobility enhancement and smaller threshold voltage shifts were observed for both holes and electrons using uniaxial stress and the theoretical basis has recently become understood.[50] As a result, uniaxial stress has become standard in very large scale integrated device design.[9–12,51]

III. STRAIN EFFECTS ON CRYSTAL SYMMETRY

Advantageous strain reduces crystal symmetry, thus symmetry is an intuitive means to study the strain effects in semiconductors. Due to the commutation between symmetry operations and the crystal Hamiltonian, crystal symmetry determines the symmetry of the band structure. Strain which

lowers the crystal symmetry lifts symmetry-determined band degeneracies. The breaking of crystal symmetry also causes band warping from symmetry restrictions.

Starting with unstrained semiconductors, based on symmetry, all group IV and III–V semiconductors have their valence band maxima at the Γ point with HH and LH bands degenerate and the split-off band lower in energy. However, because of their different polar and covalent properties, the conduction band structures vary. Si has its conduction band minima along the Δ directions at about $0.85(2\pi/a)$, where a is the crystal lattice constant, while Ge has its conduction band minima at the L point, and direct gap III–V semiconductors such as GaAs have their conduction band minimum at the Γ point. The GaAs conduction band is singly degenerate, while the degeneracy of the Si and Ge conduction band comes from the star degeneracy which is sixfold for Si and fourfold for Ge. The different positions in the Brillouin zone have different symmetries. The group of k at the Γ point is isomorphic to the point group of the lattice so the Γ point has full crystal symmetry. The points at the Δ axis have the C_{4v} symmetry, which means that the ⟨100⟩ axis is a fourfold rotation axis which is also contained in four parallel mirror planes.

Any strain can be decomposed into a hydrostatic strain and two types of shear strain.[52] One type of shear strain is related to the change of lengths along the three axes and the other is related to the rotation of the axes of an infinitesimal cube. For cubic crystals, since the hydrostatic strain does not break crystal symmetry and, hence, only shifts energy levels without lifting band degeneracy, it is unimportant for enhanced mobility. Large hydrostatic strain is generally undesirable due to yield issues resulting from strain relaxation, band gap narrowing, and MOSFET threshold voltage shifts.[53] It is the shear strain that is important to the device engineer since it reduces crystal symmetry, leading to both degeneracy lifting and band warping.

We first consider in-plane biaxial stress, where the traceless shear strain is

$$\frac{1}{3}\begin{pmatrix} e_{xx}-e_{zz} & 0 & 0 \\ 0 & e_{xx}-e_{zz} & 0 \\ 0 & 0 & -2(e_{xx}-e_{zz}) \end{pmatrix}. \tag{2}$$

This shear strain lowers the crystal symmetry by shortening (biaxial tensile stress) or elongating (biaxial compressive stress) the z-direction lattice spacing with respect to the x or y direction, which is shown schematically in Fig. 1(a). Due to the difference of the z direction with respect to the other two directions, the conduction band minima of Si along the z axis splits away from those along the x and y direction. The sixfold degenerate valleys then splits in energy into the Δ_2 and Δ_4 valleys. The case for the Ge conduction band is different. The biaxial stress has the same effect on the four L valleys, so the L valleys do not split. For GaAs, the conduction band edge lies at the Γ point, and the effect of the biaxial stress is to shift the band edge. The effects of both the hydrostatic and shear strain on band shift and splitting for biaxial stress are schematically shown in Fig. 1(b).

In contrast, longitudinal uniaxial stress along the ⟨110⟩

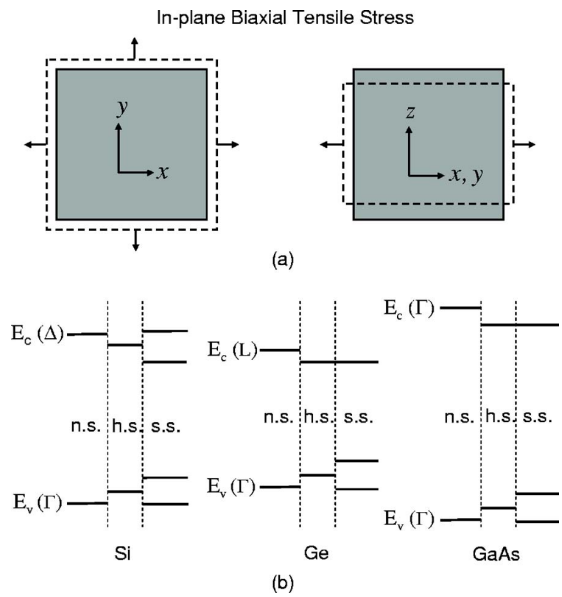

FIG. 1. (a) Cubic crystals under in-plane biaxial tensile stress. Under this type of stress, the x-y plane is still a square, but the x, y-z plane becomes a rectangle. (b) Diagrams of band splitting of Si, Ge, and GaAs under in-plane biaxial tensile stress, where "n.s." labels the unstrained band, "h.s." labels the band shifts with hydrostatic strain, and "s.s." labels the band splitting with shear strain.

direction lowers the crystal symmetry in a different manner which is shown in Fig. 2(a). In this case, the resulting strain contains both types of shear strain and can be written as

$$\frac{1}{3}\begin{pmatrix} e_{xx}-e_{zz} & 0 & 0 \\ 0 & e_{xx}-e_{zz} & 0 \\ 0 & 0 & -2(e_{xx}-e_{zz}) \end{pmatrix} + \begin{pmatrix} 0 & e_{xy} & 0 \\ e_{xy} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3}$$

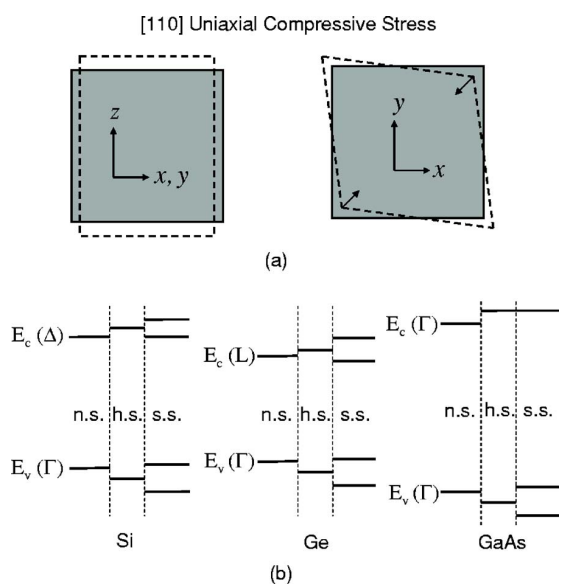

FIG. 2. (a) Cubic crystals under uniaxial [110] compressive stress. Under this type of stress, the x-y plane becomes a rhombus, and the x, y-z plane becomes a rectangle. (b) Diagrams of band splitting of Si, Ge, and GaAs under uniaxial compressive stress.

Compared to the in-plane biaxial stress case where the z direction is still a fourfold rotation axis, under uniaxial stress, the z axis only has a twofold rotation symmetry due to the second shear term. In the x-y plane, the x and y axes are no longer high symmetry directions, and the plane is mirror symmetrical with respect to the two $\langle 110 \rangle$ diagonals which are now not equivalent to each other. Similar to biaxial stress, the Si conduction band splits into Δ_2 and Δ_4 valleys due to the nonequivalency between the z and x and y axes. However, the splitting for the Ge conduction band is different. It splits into two twofold L valley groups, since the uniaxial stress distinguishes the $\langle 111 \rangle$ directions according to their projections in the x-y plane. The diagrams of the band shift and splitting of Si, Ge, and GaAs under uniaxial stress are schematically shown in Fig. 2(b).

The more important difference between uniaxial and biaxial stress is the band warping which is largely responsible for the larger electron and hole mobility enhancement observed with uniaxially stressed Si.[23,24,54] Band warping by breaking of symmetry can be understood using the principle that the band structure with respect to a band extremum in the Brillouin zone has the same symmetry of this point. When applied to the in-plane (x-y plane) band structures, it indicates that the 2D equienergy contour of the top valence band has the symmetry of a square under biaxial stress and a rhombus under uniaxial stress,[24] which can be approximated by a circle and an ellipse, respectively, for energies less than the strain splitting energy based on perturbation theory.[55] In certain conditions, the minor axis of the ellipse may be along the channel direction and thus favorable to the mobility since it results in a small conductivity mass.[24] Uniaxial stress unlike biaxial stress also warps the conduction band in diamond structure and zinc blende semiconductors. The symmetry of the z axis under uniaxial stress is no longer C_{4v} but C_{2v} with the x-y plane symmetrical only with respect to the two $\langle 110 \rangle$ diagonals. By referring to Fig. 2(a), due to the nonequivalency of the diagonals, the in-plane energy contour shall approximate an ellipse with the major and minor axes along the two diagonal directions. This has been confirmed by pseudopotential calculations.[54] Applying stress along a low symmetry axis causes more destruction of crystal symmetry and results in greater band warping than stress along a high symmetry axis. This is the reason why the valence bands of uniaxially stressed Si is more asymmetrical than that of biaxially stressed.

The valence band warping along the out-of-plane direction can be further understood by the rotation symmetry change about the z axis. In diamond structure and zinc blende semiconductors, the valence bands along $\langle 001 \rangle$ are classified by angular momentum due to the fourfold rotation symmetry. When coupled with electron spin, the eigenstates at the Γ point split into two groups classified by $J=3/2$ and $J=1/2$. The HH ($J=3/2$, $M_J=\pm 3/2$) and LH ($J=3/2$, $M_J=\pm 1/2$) bands are degenerate at the Γ point and are eigenstates of the angular momentum. For biaxial stress, the HH and LH bands have negligible warping because they do not mix since the rotation symmetry about the z axis is unchanged. However, $\langle 110 \rangle$ uniaxial stress changes the rotation symmetry about the z direction to twofold. The unperturbed HH and LH states are no longer eigenstates of the angular momentum and thus have strong mixing. This leads to the following results: (i) under biaxial stress, even though the degeneracy is lifted and the HH or LH band shifts up or down, their band curvature along $\langle 001 \rangle$ is barely changed and (ii) under uniaxial stress, in addition to the degeneracy lifting, mixing between the HH and LH bands significantly warps the bands. Symmetry breaking also causes some states to anticross in some directions. For example, in the valence bands in cubic semiconductors, if strain breaks the rotation symmetry along a direction, the HH and LH bands will mix. Under certain circumstances, when two bands whose wave functions are admixed by HH and LH states are close to each other, they will avoid crossing. This anticrossing takes place evidently along the [110] direction under [110] uniaxial stress and causes strong band warping.

Considerations from symmetry alone can give useful qualitative insights into strain effects on electronic structures. However, it does not predict the details of how the energy bands change with strain, e.g., identifying which state ascends or descends under a certain type of stress. For a simple yet more detailed qualitative description, the tight-binding theory is very useful and straightforward.

IV. STRAIN IN TIGHT-BINDING BAND PICTURE

Tetrahedral crystals including the diamond structure and zinc blende compound semiconductors have been studied using the tight-binding method extensively by many authors.[56–60] Strain changes the relative positions of atoms which compose the crystal, and thus affects the spatial interaction between atoms. The tight-binding method treats band structure by considering the microscopic, interatomic interaction, thus it provides a natural means to understand the physics of strain effects on semiconductor band structures. In band calculations using the tight-binding method, the quantum mechanical nature of bonding is retained, ensuring a simple and straightforward description of the physics of strain effects. The band shifts and splitting and effective mass changes can all be understood in this framework.

There are two atoms in a primitive cell for tetrahedral crystals. In a sp^3 tight-binding model, the conduction and valence bands are formed from the valence electron orbitals, with each atom in the primitive cell contributing an s, p_x, p_y, and p_z orbital, and the Hamiltonian is 8×8 without inclusion of the spin-orbit interaction. Using this model, Chadi and Cohen[57] and Harrison[58] studied the band structures of various diamond structure and zinc blende semiconductors. Their results show that the band structures of both diamond structure and zinc blende semiconductors are overall very similar. Along the $\langle 001 \rangle$ direction, the valence bands are formed by p bonding states with the HH band composed of the in-plane orbitals (e.g., p_x and p_y orbital) and the LH band composed of the out-of-plane orbitals (e.g., p_z orbital), as shown in Fig. 3. With inclusion of the spin-orbital interaction,[61,62] the HH wave function is circularly polarized in the perpendicular plane and the LH wave function is oriented along the quantization axis. This polarization of the wave function for the valence bands is particularly important when studying the

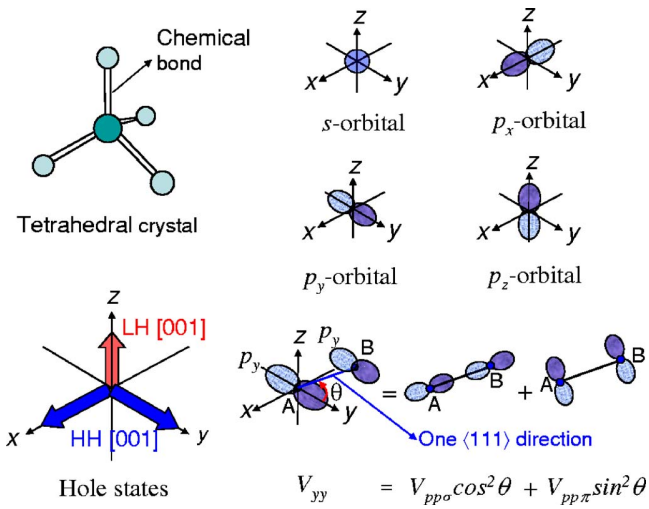

FIG. 3. The schematic diagram for the chemical bonds which connect one atom to its four nearest neighbors in a tetrahedral crystal, the s, p_x, p_y, and p_z orbital, the interaction between two p_y orbitals, and the wave function polarization for the HH and LH states.

strain effects on band structures. For direct gap zinc blende semiconductors, the conduction band is composed of the s antibonding states, while for Si, due to the p admixture from Γ to X the wave function of the conduction band edge is primarily composed of the p antibonding states. By using a small set of atomic functions as basis states, the high energy bands are described less satisfactorily because they are more delocalized and thus consist of more atomic functions. For higher precision, a sp^3s^* model[59,63,64] or more complicated $sp^3s^*d^5$ model[65,66] can be employed. Neither of these models significantly alters the composition of the wave function for the valence bands, but the conduction band for Si will change slightly due to the mixing from the excited atomic states.

A straightforward way to picture the spatial interaction between atoms is the chemical bonds. As shown in Fig. 3, in bulk tetrahedral semiconductors, every atom connects with its four nearest neighbors by four bonds, each of which points toward a ⟨111⟩ direction of the cube which circumscribes the tetrahedron. In every bond, p_x, p_y, p_z, and s each contributes 25%. Considering only the p orbitals which are directionally oriented unlike the s orbital, the projection of the bonds along one principal direction (⟨100⟩, ⟨110⟩, or ⟨111⟩) composes the LH state and the projection on the perpendicular plane composes the HH state. Strain rotates the bonds or changes the bond lengths which alters the weight of orbitals composing the bonds, causing orbital rehybridization and variation of the spatial interaction.

The effects of hydrostatic strain on the bands can be first understood by the tight-binding picture. If all the V constants[67] increase by the same ratio $(1+\delta)^{-2}$, where δ is the hydrostatic dilation,[58] the energy eigenvalues shift, but the band curvature is approximately the same as that of the initial k point at $k'=(1-\delta)k$, i.e., the energy level of the dilated crystal is scaled by $(1-\delta)$ in k. Therefore, the band warping is mainly caused by shear strain.

The shear strain for biaxial stress does not change the

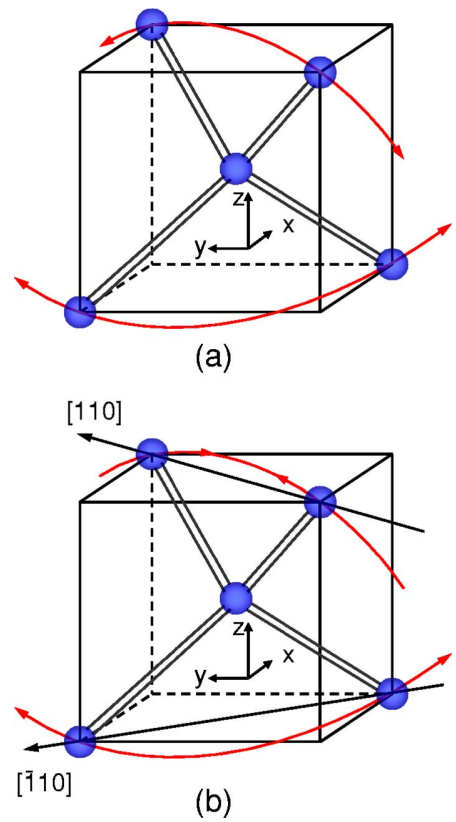

FIG. 4. (a) Under biaxial tension, four bonds rotate toward the x-y plane. (b) Under [110] uniaxial compression, two bonds are pushed together along [110] and two bonds are pulled apart along [$\bar{1}$10].

bond lengths to the first order of $\alpha=(e_{xx}-e_{zz})/3$ [see Eq. (2)], but alters the bond angles. The four bonds all rotate toward or away from the x-y plane depending on the sign of α. For biaxial tension ($\alpha>0$), the situation is illustrated in Fig. 4(a)). In this case, all the four bonds are equivalent and rotate toward the x-y plane. With such a rotation, the weight of the p_x and p_y orbitals in the bonds increases and that of the p_z orbital decreases. Along the [001] direction, this results in increased overlap integrals between in-plane orbitals and lowered HH bands and decreased overlap integrals between the p_z orbitals and a weakened LH band which ascends in energy (becomes the top valence band). This can be confirmed by checking the interaction constants V_{ii},[56–58] where $i=x$, y, z, between orbitals

$$V'_{xx} = V'_{yy} = \frac{1}{3}V_{pp\sigma}(1+2\alpha) + \frac{2}{3}V_{pp\pi}(1-\alpha) \tag{4}$$

and

$$V'_{zz} = \frac{1}{3}V_{pp\sigma}(1-4\alpha) + \frac{2}{3}V_{pp\pi}(1+2\alpha), \tag{5}$$

where $V_{pp\sigma}$ and $V_{pp\pi}$, determined by the bond length, are the interaction constants of the σ- and π-type bond coupling, respectively. Because $V_{pp\sigma}>0$ (for Si, $V_{pp\sigma}=4.47$ eV) and $V_{pp\pi}<0$ (for Si, $V_{pp\pi}=-1.12$ eV), in-plane biaxial tension enhances the interactions between p_x and p_y orbitals and weakens the overlap between p_z orbitals. Along the [100] and [010] direction, the LH band is lowered because it is com-

posed mainly of the p_x and p_y orbitials, respectively, and the topmost bands are HH bands. Therefore, the top valence band under biaxial tension is LH-like out-of-plane and HH-like in-plane. This is the reason why biaxial tensile stress is not desirable for hole mobility enhancement,[24,28] which we will elaborate in Sec. VIII.

The wave function of the Si conduction band along the Δ axis is composed of s and parallel p states, e.g., the conduction band state along the [001] direction is composed of s and p_z states. Biaxial tension lowers the two valleys along the z axis and raises the other four in the x-y plane due to both the changed overlap integrals between the p orbitals and the fact that the conduction band of Si is composed of antibonding p states.

The shear strain for [110] uniaxial stress contains two terms [see Eq. (3)]. The first shear term gives the same results as the biaxial stress and determines the band splitting and shift along the three $\langle 100 \rangle$ axes. For uniaxial compression, this term results in a top valence band with out-of-plane HH-like and in-plane LH-like characters. It does not distinguish the two in-plane $\langle 110 \rangle$ directions. However, the second shear term creates the in-plane anisotropy by changing both the bond angles and bond lengths. For the four nearest-neighbors connected by the bonds, two of them lie along the [110] direction and the other two lie along the $[\bar{1}10]$ direction. Under compressive stress, the bonds that project along the [110] direction are shortened and pushed together and the bonds that project along the $[\bar{1}10]$ direction are elongated and pulled apart, which is illustrated in Fig. 4(b). Despite the shortened bond length which increases the interaction between the origin atom with the atoms along [110], the rotation of the two bonds increases the weight of the perpendicular orbitals and reduces that of the parallel orbitals with respect to the [110] direction. Thus, lowered HH states and a raised LH state along [110] result. Along $[\bar{1}10]$, the situation is reversed because compression along [110] is equivalent to tension along $[\bar{1}10]$. The fact that the two bonds are pulled apart increases the parallel projection and reduces the perpendicular projection. The valence band splitting and shifts under [110] uniaxial compressive stress then favor enhanced hole mobility by reduced conductivity mass and still large DOS in the ground valence band.

For both biaxial and uniaxial stress, compression shifts the LH band up and the HH band down along the compressive stress axis [for biaxial tensile stress, the compressive stress axis is along z, see Fig. 1(a)]. This occurs because compression rotates the bonds toward the plane perpendicular to the stress axis and makes the interatomic interaction in the perpendicular plane stronger than that parallel to the stress axis.

The simple tight-binding picture presented in this section explains the strain effects on the band splitting and shifts in a qualitative way. Using a sophisticated and refined tight-binding method, a quantitative description of the band splitting and effective mass change due to stress can also be obtained.[65,66,68–70] For most cases, an entire band picture is not necessary because in applications to semiconductors or MOSFETs, electrons or holes are generally only located at the conduction or valence band edges. To obtain quantitative or semiquantitative results of the strain effects on band structure or charge transport, the $\mathbf{k}\cdot\mathbf{p}$ method can be employed, which usually treats the band structure around a special point in k space.

V. STRAIN IN k·p FRAMEWORK

The $\mathbf{k}\cdot\mathbf{p}$ method is based on perturbation theory and substantial symmetry considerations.[34,71–73] It also employs various empirical parameters such as the band gap. Although it cannot provide a microscopic picture, it uses a small set of basis vectors and treats strained band structures with high precision. Other band models such as the tight-binding and pseudopotential methods need to treat the strained crystal as a new system, but owing to Pikus and Bir,[35] the strain Hamiltonian has a very simple form in the $\mathbf{k}\cdot\mathbf{p}$ framework. This is important not only in three-dimensional (3D) cases, but in 2D systems as well where the $\mathbf{k}\cdot\mathbf{p}$ method is proven to be efficient[22] for solving the Schrödinger-Poisson equation to obtain the self-consistent field and subband structures. In this section, calculations under the $\mathbf{k}\cdot\mathbf{p}$ framework are used to validate the insight gained in the symmetry and tight-binding discussions.

According to Pikus and Bir,[35] the correspondence between the strain Hamiltonian and the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian is

$$k_i k_j \leftrightarrow e_{ij}, \tag{6}$$

and the total Hamiltonian is given by $H = H_{kp} + H_{\text{strain}}$. The strain Hamiltonian is formally identical to the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian due to the fact that p^2 and e_{ij} are both second rank tensors.

First consider the GaAs and Si conduction bands which are singly degenerate at the band minima. For GaAs, the total Hamiltonian for the conduction band is

$$H = \frac{\hbar^2 k^2}{2m^*} + a_c(e_{xx} + e_{yy} + e_{zz}), \tag{7}$$

where a_c is the conduction band deformation potential. For Si, we need two deformation potentials in the conduction band from symmetry considerations, and the total Hamiltonian is

$$H = \frac{\hbar^2 (k_l - k_0)^2}{2m_l^*} + \frac{\hbar^2 k_t^2}{2m_t^*} + a_l e_l + a_t e_t \tag{8}$$

or written in the conventional way

$$H = \frac{\hbar^2 (k_l - k_0)^2}{2m_l^*} + \frac{\hbar^2 k_t^2}{2m_t^*} + \Xi_d \mathrm{Tr}(e_{ij}) + \Xi_u e_l, \tag{9}$$

where Ξ_d and Ξ_u are the dilation and uniaxial deformation potentials at the Si conduction band edge, $\mathrm{Tr}(e_{ij})$ stands for the trace of the strain tensor, and e_l (e_t) is the longitudinal (transverse) strain component. Along [001],

$$e_l = e_{zz} \tag{10}$$

and

$$e_t = e_{xx} + e_{yy}. \tag{11}$$

For a singly degenerate band such as the conduction band for GaAs or Si, strain shifts the energy but does not

change the energy dispersion in the linear regime. The off-diagonal terms in the strain tensor such as e_{xy} do not contribute to the band shift or band warping in this model (see also Table I). Including more bands into the system, the warping can be obtained through the coupling from the other strain altered bands. However, for semiconductors such as Si, because the conduction and valence band edges are not at the same k point, it is difficult to treat the conduction and valence bands at the same time, although there indeed exist full zone $\mathbf{k}\cdot\mathbf{p}$ models.[71,74,75]

The valence bands for all diamond and zinc blende structure semiconductors can be calculated using the Luttinger Hamiltonian.[34,73] The Luttinger Hamiltonian is suitable for semiconductors whose band gap is much larger than the split-off energy, and thus the valence bands (including the HH, LH, and split-off band) can be treated alone. Using the basis classified by the angular momentum (e.g., $|3/2, 3/2\rangle$, $|3/2, 1/2\rangle$, $|3/2, -1/2\rangle$, $|3/2, -3/2\rangle$, $|1/2, 1/2\rangle$, and $|1/2, -1/2\rangle$, where the first quantum number labels the total angular momentum, and the second quantum number labels the angular momentum projection at one $\langle 100\rangle$ axis), the Hamiltonian is represented by

$$\begin{bmatrix} -P-Q & L & -M & 0 & \frac{1}{\sqrt{2}}L & -\sqrt{2}M \\ L^{+} & -P+Q & 0 & -M & \sqrt{2}Q & -\sqrt{\frac{3}{2}}L \\ -M^{+} & 0 & -P+Q & -L & -\sqrt{\frac{3}{2}}L^{+} & -\sqrt{2}Q \\ 0 & -M^{+} & -L^{+} & -P-Q & \sqrt{2}M^{+} & \frac{1}{\sqrt{2}}L^{+} \\ \frac{1}{\sqrt{2}}L^{+} & \sqrt{2}Q^{+} & -\sqrt{\frac{3}{2}}L & \sqrt{2}M & -P-\Delta & 0 \\ -\sqrt{2}M^{+} & -\sqrt{\frac{3}{2}}L^{+} & -\sqrt{2}Q^{+} & \frac{1}{\sqrt{2}}L & 0 & -P-\Delta \end{bmatrix}, \tag{12}$$

where

$$P = \frac{\hbar^2}{2m_0}\gamma_1(k_x^2 + k_y^2 + k_z^2), \tag{13a}$$

$$Q = \frac{\hbar^2}{2m_0}\gamma_2(k_x^2 + k_y^2 - 2k_z^2), \tag{13b}$$

$$L = \frac{\hbar^2}{m_0}\sqrt{3}\gamma_3(k_x - ik_y)k_z, \tag{13c}$$

$$M = -\frac{\hbar^2}{2m_0}\sqrt{3}[\gamma_2(k_x^2 - k_y^2) - 2i\gamma_3 k_x k_y], \tag{13d}$$

and Δ is the split-off energy, and γ_1, γ_2, and γ_3 are the Luttinger parameters. In this Hamiltonian, the only difference between the $\langle 110\rangle$ and $\langle 100\rangle$ directions is that HH couples with the LH band through the Luttinger parameter γ_2 in $\langle 100\rangle$ while they couple each other through γ_3 in $\langle 110\rangle$. Generally since $\gamma_2 < \gamma_3$ and are close to each other, the band repulsion along $\langle 110\rangle$ is slightly stronger than along the $\langle 100\rangle$ direction, so the HH mass is larger and LH mass is smaller in the $\langle 110\rangle$ direction. Due to the similarity of symmetry between Si, Ge, and zinc blende semiconductors, the band structures are similar. However, the small split-off energy of Si enhances the interaction between the top bands and the split-off band so that the HH band of Si is more warped than Ge and GaAs. This is reflected by the difference between the values of γ_2 and γ_3. If $\gamma_2 = \gamma_3$ (which is called the spherical approximation), then the valence bands are isotropic. The valence band structure of unstrained Si obtained from Hamiltonian Eq. (12) is shown in Fig. 5(a) together with the effective masses labeled where only the HH and LH bands are included. For convenience, the Luttinger parameters and deformation potentials of Si, Ge, and GaAs are listed in Table II.

When the energy range of interest is sufficiently smaller than the split-off energy, it is usual to assume the coupling from the split-off band can be ignored. In this case, the electronic structures of the HH and LH bands can approximately be described by a 4×4 Hamiltonian [the upper-left 4×4 matrix block in the Hamiltonian Eq. (12)]. By diagonalizing the 4×4 Hamiltonian, the analytic expressions for the effective masses for both the HH and LH bands along $\langle 001\rangle$ and $\langle 110\rangle$ are given as follows.[37,38]

$\langle 001\rangle$:

$$\frac{m_{\mathrm{hh}}^*}{m_0} = \frac{1}{\gamma_1 - 2\gamma_2}, \tag{14a}$$

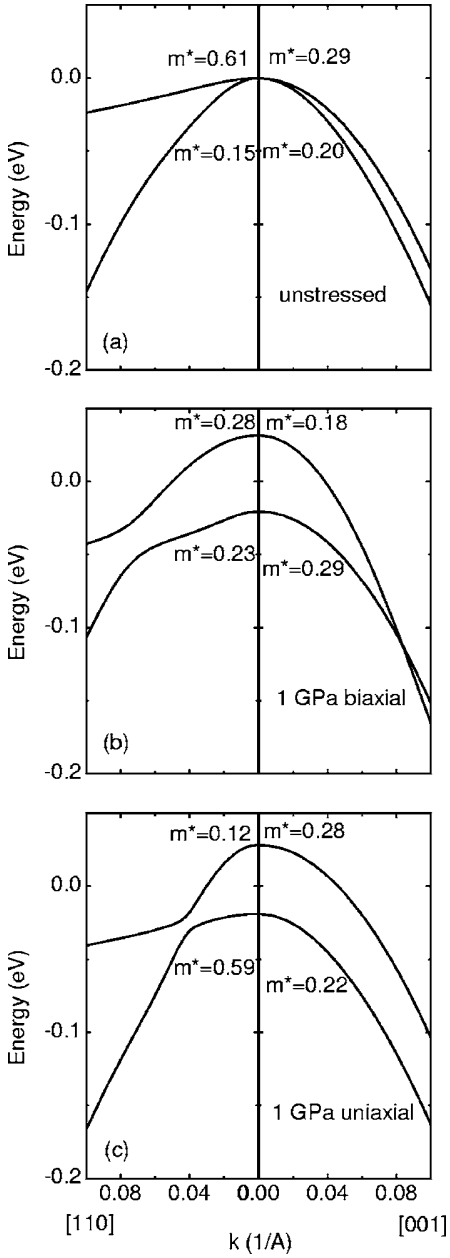

FIG. 5. Si valence band structures for (a) unstressed, (b) 1 GPa biaxial tensile stressed, and (c) 1 GPa uniaxial compressive stressed Si. Only the top two bands are shown and effective masses are labeled in units of free electron mass.

$$\frac{m_{\mathrm{lh}}^*}{m_0} = \frac{1}{\gamma_1 + 2\gamma_2}. \tag{14b}$$

$\langle 110 \rangle$:

$$\frac{m_{\mathrm{hh}}^*}{m_0} = \frac{1}{\gamma_1 - \sqrt{\gamma_2^2 + 3\gamma_3^2}}, \tag{15a}$$

$$\frac{m_{\mathrm{lh}}^*}{m_0} = \frac{1}{\gamma_1 + \sqrt{\gamma_2^2 + 3\gamma_3^2}}. \tag{15b}$$

In a general strain case, every term in the valence band Hamiltonian Eq. (12) has a strain counterpart. The corresponding strain terms are defined as the following:

TABLE II. Band parameters and deformation potentials for Si, Ge, and GaAs.

	γ_1 [a]	γ_2 [a]	γ_3 [a]	m_D^{hha}	m_D^{lha}	a_v [b] (eV)	b (eV)	d (eV)
Si	4.22	0.39	1.44	0.53	0.16	2.46	−2.1 [c]	−4.8 [c]
Ge	13.4	4.24	5.69	0.35	0.043	1.24	−2.9 [d]	−5.3 [d]
GaAs	6.98	2.06	2.93	0.62	0.074	1.16	−2.0 [e]	−4.8 [e]

[a]See Ref. 76
[b]See Ref. 77. In calculating the valence band structures shown in Fig. 5 and 6, parameter $a_v=0$ is taken for the valence bands because it shifts the valence bands uniformly.
[c]See Ref. 78.
[d]See Ref. 79.
[e]See Ref. 80.

$$P_e = -a_v(e_{xx} + e_{yy} + e_{zz}), \tag{16a}$$

$$Q_e = -\frac{b}{2}(e_{xx} + e_{yy} - 2e_{zz}), \tag{16b}$$

$$L_e = -d(e_{xz} - ie_{yz}), \tag{16c}$$

$$M_e = \frac{\sqrt{3}}{2}b(e_{xx} - e_{yy}) - ide_{xy}. \tag{16d}$$

The term P_e occurs in every diagonal term in the strain Hamiltonian so its only role is to shift the whole valence bands uniformly without warping the bands. In fact, $(e_{xx} + e_{yy} + e_{zz})$ is exactly the hydrostatic strain part. Q_e corresponds to the diagonal shear strain terms and M_e corresponds to the off-diagonal shear strain terms in Eq. (3). L_e is only nonvanishing when there is stress along [101] or [011]. Around the Γ point, if the coupling from the split-off band is neglected, through diagonalizing the 4×4 Hamiltonian, the analytical expression for the top two bands can be obtained as

$$E(k) = -P - P_e \pm \sqrt{|Q + Q_e|^2 + |L + L_e|^2 + |M + M_e|^2}. \tag{17}$$

However, strain introduces extra coupling between the HH, LH, and split-off hole bands, and the 4×4 Hamiltonian does not suffice for a realistic description even for semiconductors which have relatively large split-off energy. The strain splitting and effective mass obtained can be remarkably different from those obtained by the 6×6 Hamiltonian in highly strained semiconductors.[37,81] Thus, the analytical results obtained using the 4×4 Hamiltonian can only serve for qualitative discussions. The figures shown in this section are all plotted using the results obtained through the 6×6 Hamiltonian.

For in-plane biaxial stress, $e_{ij}=0$ for $i\neq j$, and $e_{xx}=e_{yy}\neq e_{zz}$, thus $L_e=M_e=0$. When neglecting the coupling from the split-off band, all the off-diagonal strain terms vanish, and thus the band warping along [001] is small. In fact, the HH effective mass is negligibly affected and the LH mass will change slightly due to the nonvanishing coupling from the split-off band around the Γ point. The shear strain shifts the HH and LH bands oppositely in energy. We need to note that Q_e along [100] or [010] is different from that along

[001]. Along [001], $Q_e=-b(e_{xx}-e_{zz})$, and along [100] or [010], $Q_e=-b/2(e_{zz}-e_{xx})$. Therefore, the trend of relative band shift is opposite between [001] and [100] or [010], but is the same for [100] and [010]. The HH and LH band shift and splitting of Si under in-plane biaxial tension ($e_{xx}=e_{yy}>0$ and $e_{zz}<0$) is shown in Fig. 5(b). Under biaxial tension, the top band along [001] is LH-like, and in the *x*-*y* plane, it is HH-like. The band warping and shifts are both consistent with the results obtained by Hasegawa[37] and Hensel and Feher.[38]

The band structure for [110] uniaxial compressive strained Si is shown in Fig. 5(c). For [110] uniaxial stress, $e_{xx}=e_{yy}\neq e_{zz}$, and $e_{xy}\neq 0$. In this case, $L_e=0$ and $M_e\neq 0$. The two strain terms are the diagonal term $(e_{xx}+e_{yy}-2e_{zz})$ and the off-diagonal term e_{xy}. At the Γ point, although the diagonal shear term does not couple the HH and LH bands in the [001] direction, the e_{xy} term couples them. This occurs because the uniaxial stress reduces the symmetry in the *x*-*y* plane. The band shift is determined by both Q_e and M_e. Under [110] uniaxial compression ($e_{xx}<0$ and $e_{zz}>0$), the HH-like band rises and the LH-like band lowers in energy along [001], which is mainly determined by the diagonal shear strain term Q_e. Due to the coupling between the HH and LH bands at the Γ point, the LH effective mass becomes larger and the HH effective mass becomes smaller along [001]. In the *x*-*y* plane, the top band effective mass along [110] is smaller than the unstressed value at $0.12m_0$, and the effective mass along $[\bar{1}10]$ is very large ($\sim m_0$). In Secs. VII and VIII, we will see that the effective mass along [001] determines the band shift under electric confinement and the in-plane effective masses determine the conductivity mass and 2D DOS mass for (001) surface MOSFETs. The change of these three mass values with strain ultimately determines the strain-induced mobility enhancement.

Strain has similar effects on the valence band structure for both group IV and III–V semiconductors. For comparison between different semiconductors and between different stresses, Fig. 6 shows the 3D equienergy surfaces of Si, Ge, and GaAs top valence bands with no stress, 1 GPa biaxial tensile, and 1 GPa uniaxial compressive stress computed using the full 6×6 Hamiltonian. The top valence band is especially important since with relatively large strain, holes primarily occupy the top valence band, and this determines the hole mobility. The unstressed top valence band of Si is remarkably different from Ge and GaAs because of more warping, while the stressed band characters are similar. The top bands of Si, Ge, and GaAs are all LH-like in the *z* direction and HH-like in-plane for biaxial tensile stress, and LH-like in the [110] direction and HH-like in the $[\bar{1}10]$ and *z* direction for uniaxial compressive stress.

In the 2D MOSFET channels, the conductivity mass and DOS of the hole gas in the inversion layer of the *p*-channel MOSFETs are both determined by the in-plane energy dispersion. Since electric confinement along the *z* direction does not significantly affect the energy dispersion in the *x*-*y* plane for (001) surface MOSFET devices, the in-plane energy dispersion under strain can be discussed in the bulk case. From the 4×4 Hamiltonian while neglecting the coupling from the split-off band, the energy dispersion in the small *k* limit ($\hbar^2k^2/2m_0 \ll \Delta E$, with ΔE the strain splitting) for the biaxial stress case is[38,82]

$$E(k) = -P_e \mp Q_e - \left(\frac{\hbar^2}{2m_0}\right)[(\gamma_1 \pm \gamma_2)k_t^2 + (\gamma_1 \mp 2\gamma_2)k_z^2]. \tag{18}$$

Then, for biaxial stress, the top and lower bands are both ellipsoids with the energy contours in the *x*-*y* plane being circles. The top band under biaxial tension corresponds to the lower sign, and the in-plane bands are HH-like while the out-of-plane band is LH-like. The energy dispersion with [110] uniaxial stress also depends on the compliance parameters of the specific semiconductor besides the band parameters through the relation between α and β, where we define $\alpha=e_{xx}-e_{zz}$, and $\beta=e_{xy}$. The dispersion expression for the top two bands under $[1\bar{1}0]$ uniaxial stress is

$$\begin{aligned} E(k) = &-P_e \pm \sqrt{b^2\alpha^2 + d^2\beta^2} \\ &- \frac{\hbar^2k^2}{2m_0}\gamma_1 \mp \frac{\hbar^2k_1^2}{2m_0}\frac{(\sqrt{3}\gamma_3 d\beta + \gamma_2 b\alpha)}{\sqrt{b^2\alpha^2 + d^2\beta^2}} \\ &\pm \frac{\hbar^2k_2^2}{2m_0}\frac{(\sqrt{3}\gamma_3 d\beta - \gamma_2 b\alpha)}{\sqrt{b^2\alpha^2 + d^2\beta^2}} \pm \frac{2\gamma_2 b\alpha k_z^2}{\sqrt{b^2\alpha^2 + d^2\beta^2}}, \end{aligned} \tag{19}$$

where k_1 is the *k* vector along [110] and k_2 is the *k* vector along $[\bar{1}10]$. At the small *k* limit, both the top and lower bands are ellipsoids, with the three axes along $[\bar{1}10]$, [110], and [001]. Under uniaxial compression, the energy contour for the top band in the *x*-*y* plane is an ellipse with the major axis along $[\bar{1}10]$ and the minor axis along [110]. The 2D *x*-*y* energy contours for unstressed, biaxial tensile stressed and [110] uniaxial compressive stressed Si are shown in Fig. 7, where it clearly shows that for small energy, the 2D contours for biaxial stress are close to circles and those for uniaxial stress are close to ellipses, which are consistent with the earlier discussions and also consistent with the results from the symmetry considerations. Obviously, in the large *k* limit where the strain splitting energy can be neglected, the band structure of the unstrained bulk case is recovered, which can be seen by inspecting Eq. (17) or Fig. 7.

VI. STRAIN ON BULK CARRIER TRANSPORT

With the knowledge of strain-induced band shifts, splitting, and effective mass changes under biaxial and uniaxial stress, the effects of strain on carrier mobility can be investigated qualitatively in a simple way through the Drude model $\mu=e\tau/m^*$, where μ is the carrier mobility, τ is the momentum relaxation time, and m^* is the conductivity effective mass. Strain changes the conductivity mass m^*, and also the DOS of bands, which consequently affects the momentum relaxation time. A variety of scattering mechanisms contribute to momentum relaxation among which phonon scattering dominates at room temperature for lightly doped semiconductors. In this section, we discuss the strain effects on the bulk carrier transport properties. At low gate voltages where the electric confinement is weak, the bulk case is a

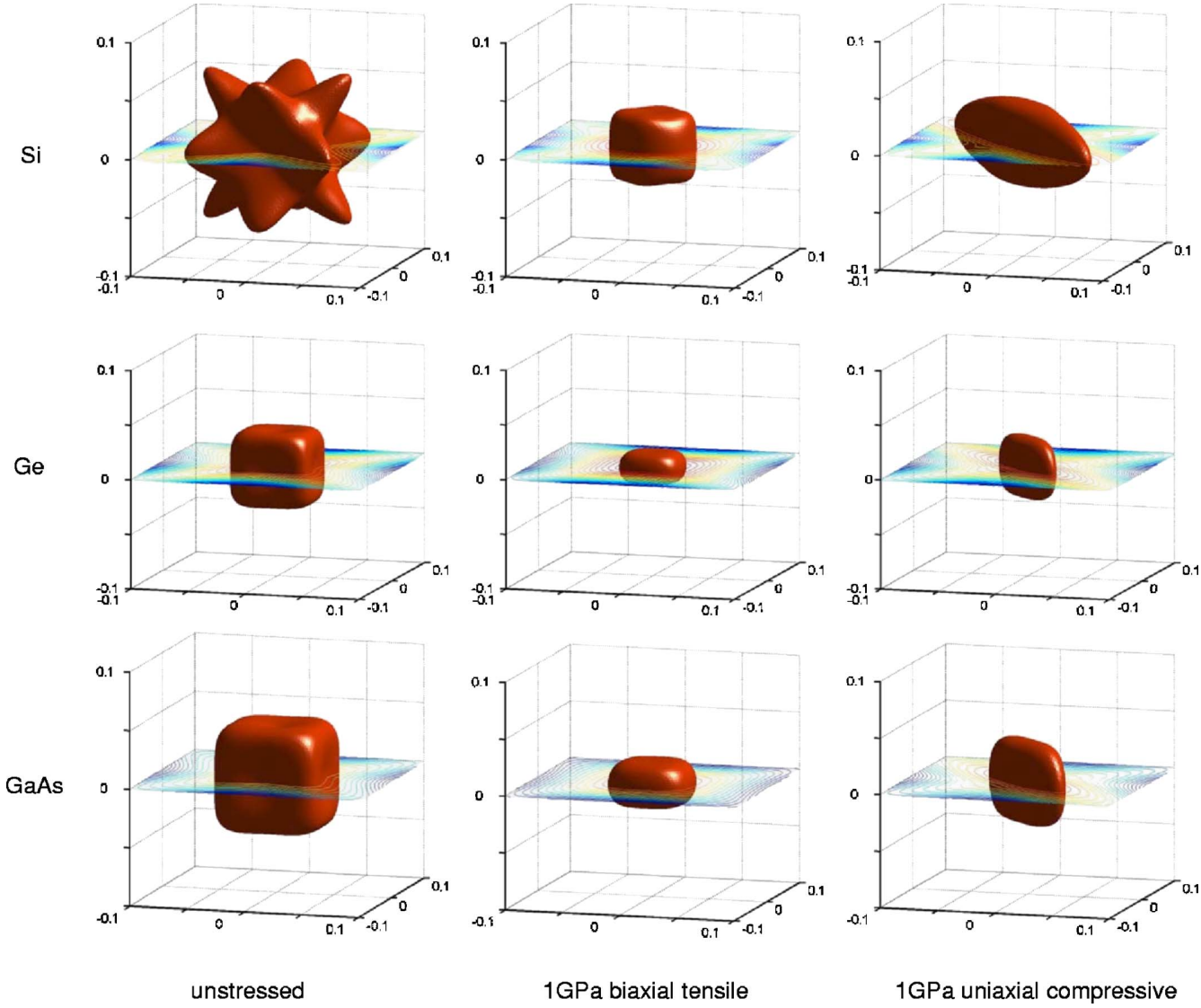

FIG. 6. The 3D equienergy surfaces for the top valence bands of Si, Ge, and GaAs with no stress, 1 GPa biaxial tensile and uniaxial compressive stress. The energy plotted is 25 meV. For biaxial tensile stress, the bands are LH-like in the z direction and HH-like in the [110] direction. For uniaxial compressive stress, the bands are HH-like in the z direction and LH-like in the [110] direction.

close approximation to the MOSFET channel. The studies of bulk transport also provide the formalism for studying 2D electron/hole gases in Sec. VIII.

Since the conduction band of III–V band materials such as GaAs is singly degenerate, their electron mobility will not be affected significantly by strain. Electron mobility enhancement in strained Si results from the lifting of degeneracy of the six conduction band valleys. The splitting of the Δ_2 valleys from the Δ_4 valleys and the preferential population of electrons in the Δ_2 valleys under in-plane biaxial tensile or [110] uniaxial tensile stress result in more electrons with lighter effective mass available for carrier transport, since the electron effective mass perpendicular to the Δ_2 axis is lighter. The maximum electron mobility enhancement due to mass change alone in strained Si is $3m_l/(2m_l+m_t)=1.36$ at the large stress limit when only the Δ_2 valleys are occupied.

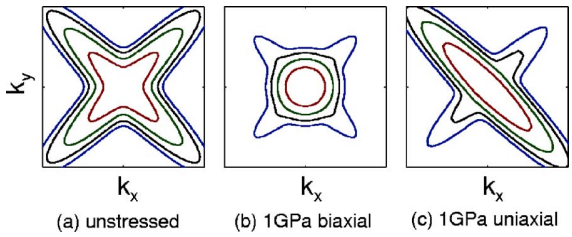

FIG. 7. Top band 2D energy contours for (a) unstressed, (b) 1 GPa biaxial tensile stressed, and (c) uniaxial compressive stressed Si. The energy interval between contours is 25 meV.

The valley degeneracy lifting in strained Si also alters phonon scattering. For moderate to large stress when the strain splitting is larger than k_BT and comparable to the intervalley phonon energy, electron repopulation becomes considerably less effective, and the strain-induced intervalley scattering reduction becomes dominant. In the Si conduction band, both intravalley (acoustic phonon scattering only, optical phonon is forbidden) and intervalley (both acoustic and optical phonon) scattering contribute. At low temperature,

acoustic phonon intravalley scattering dominates, but at room temperature, intervalley scattering is about twice as strong as intravalley scattering.[83] Intravalley scattering is enhanced due to the electron repopulation, while the intervalley scattering is reduced due to valley splitting-induced decrease of scattering joint DOS. With sufficiently large splitting, the intravalley scattering is slightly increased while the scattering from Δ_2 valleys to Δ_4 valleys is totally suppressed. The intervalley scattering rate then is reduced by about ~80% assuming the *f*- and *g*-type scattering have about the same strength. This assumption was used in the theoretical work of Takagi *et al.*[20] and yielded good agreement with experiments. However, this is an empirical attempt to explain the strain-enhanced electron mobility. Strain should not alter the *f* and *g* scattering strength, and for Si, the scattering strengths are not the same.[84–86] Nevertheless, if we stick to this assumption and consider both the conductivity mass and phonon scattering change, the bulk electron mobility enhancement is ~2.5, consistent with experimental observations.[28,29]

The physics of strain-enhanced hole mobility is complicated due to valence band complexity. For unstrained semiconductors, both heavy holes and light holes contribute to hole transport due to band degeneracy. More sophisticated models such as the Boltzmann equation should be applied to quantify the hole transport, which takes into account the energy-dependent relaxation process and considers the anisotropic valence bands. A simple estimate can be made for the hole mobility of bulk unstrained semiconductors limited by phonon scattering. The DOS of either the HH or LH band can be written as

$$N(E) = \frac{4\pi(2m_D^*)^{3/2}}{h^3}\sqrt{E}, \tag{20}$$

where m_D^* is the DOS effective mass for the HH or LH band. For a hole density of p, the density of the HH is $m_D^{\mathrm{hh}\,3/2}p/(m_D^{\mathrm{hh}\,3/2}+m_D^{\mathrm{lh}\,3/2})$ and the density of the LH is $m_D^{\mathrm{lh}\,3/2}p/(m_D^{\mathrm{hh}\,3/2}+m_D^{\mathrm{lh}\,3/2})$. For scattering in either the HH or LH band, the overlap factor is[87]

$$I^2(\mathbf{k},\mathbf{k}') = \frac{1}{4}(1+3\cos^2\theta_{\mathbf{k}}) \tag{21}$$

and for scattering between the HH and LH bands

$$I^2(\mathbf{k},\mathbf{k}') = \frac{3}{4}\sin^2\theta_{\mathbf{k}}, \tag{22}$$

where $\theta_{\mathbf{k}}$ is the angle between the initial and final state vectors.

At room temperature or above, optical phonon scattering is dominant in Si and Ge,[88,89] and polar optical phonon scattering dominates in III–V compound semiconductors. Considering only optical phonon scattering, the scattering rate is given by[90]

$$1/\tau(\mathbf{k}) \propto \int d\mathbf{k}'\frac{C_{\mathbf{k}',o}^2 I^2}{\hbar\omega_o}\left(n(\omega_o)+\frac{1}{2}\mp\frac{1}{2}\right) \times\delta(E_{\mathbf{k}'}-E_{\mathbf{k}}\mp\hbar\omega_o), \tag{23}$$

where $C_{\mathbf{k}',o}^2$ is the electron-phonon coupling term, $n(\omega_o)$ is the density of the optical phonons with energy $\hbar\omega_o$, and the upper (lower) sign corresponds to the phonon absorption (emission) process. The δ function complies with the conservation of energy and the integration of it gives the joint DOS for phonon scattering. Assuming most holes are located at the Γ point, which is approximately the case at low field and low hole density, the phonon emission process can be neglected because the initial hole energy is lower than the phonon energy $\hbar\omega_o$. The angular dependence of the scattering rate is from the overlap factor. Integration of I^2 over the range $-\pi\leq\theta\leq\pi$ gives the same value for expressions of I^2 in either Eqs. (21) or (22). Thus the heavy holes and light holes have the same scattering rate. For Si and Ge, the electron-optical phonon coupling can be assumed independent of wave vector $\mathbf{q}=\mathbf{k}'-\mathbf{k}$. For both heavy holes and light holes

$$(1/\tau)_{\mathrm{nonpolar}} \propto \frac{C^2 n(\omega_o)\pi}{\hbar^2\sqrt{\omega_o}}\left[\left(\frac{2m_D^{\mathrm{hh}}}{\hbar}\right)^{3/2}+\left(\frac{2m_D^{\mathrm{lh}}}{\hbar}\right)^{3/2}\right]. \tag{24}$$

For polar semiconductors, the electron-phonon coupling is

$$C_{q,o}^2 = A\frac{q^2}{(q^2+q_0^2)^2}, \tag{25}$$

where q_0 is the reciprocal Debye screening length, A is a constant determined by the semiconductor intrinsic properties, and the scattering rate is

$$(1/\tau)_{\mathrm{polar}} = An(\omega_o)\pi\sqrt{\hbar\omega_o}\left[\frac{(2m_D^{\mathrm{hh}}/\hbar^2)^{5/2}}{(2m_D^{\mathrm{hh}}\omega_o/\hbar+q_0^2)^2} + \frac{(2m_D^{\mathrm{lh}}/\hbar^2)^{5/2}}{(2m_D^{\mathrm{lh}}\omega_o/\hbar+q_0^2)^2}\right]. \tag{26}$$

The overall hole mobility is

$$\mu = \frac{e\tau}{p}\left(\frac{p_{\mathrm{lh}}}{m_{\mathrm{lh}}}+\frac{p_{\mathrm{hh}}}{m_{\mathrm{hh}}}\right), \tag{27}$$

where m_{hh} and m_{lh} are the conductivity masses determined by the device geometry and are different from the DOS masses. The conductivity mass in bulk semiconductors is difficult to obtain analytically because of the strong anisotropy in the valence bands and the occupation in 3D k space. In Sec. VIII, we will introduce a simple estimation method to evaluate the conductivity mass in the 2D case.

The difficulty of analyzing the strain-enhanced hole mobility lies in the band warping caused by strain. Warping mixes the HH and LH characters in the valence bands as seen in Figs. 5(b) and 5(c). Splitting also changes the hole distribution. At moderate stress (~1 GPa), the strain-induced splitting is small compared to the optical phonon energy but large enough so that the conducting holes mostly reside in the top band. Process-induced production stresses[7,51] or lab-

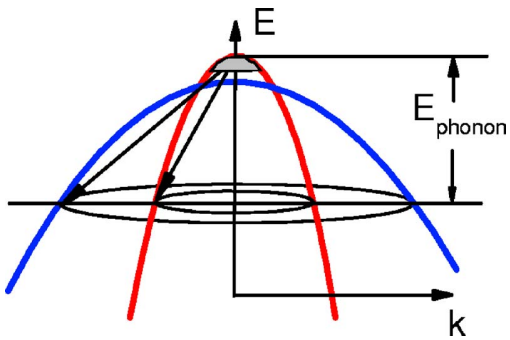

FIG. 8. Diagram for the valence bands with a small strain splitting which is much less than the phonon energy. Arrows show phonon scattering. Optical phonon scattering changes little. Mobility is enhanced because more light holes take part in conduction.

realized stresses by wafer bending method[43,50,91] are typically ~1 GPa, resulting in valence band splitting for Si of about 20–30 meV. The optical phonon energy for Si is over 60 meV. In this case, the phonon scattering rate is slightly reduced but can be considered unchanged for simplicity. Thus, the mobility enhancement can be considered to be induced mainly by hole redistribution among the split bands. This situation is schematically illustrated in Fig. 8. Neglecting changes in scattering, the enhancement factor can be expressed as the ratio of the conductivity masses

$$\frac{\mu'}{\mu} = \frac{m_c}{m_c'} = \frac{m_{\rm hh} m_D^{\rm hh3/2} + m_{\rm lh} m_D^{\rm lh3/2}}{m_{\rm top}'(m_D^{\rm hh3/2} + m_D^{\rm lh3/2})}, \tag{28}$$

where $m_{\rm top}'$, the hole conductivity mass of the topmost valence band in strained semiconductors, also constrained by the device geometry, may not be the same as the unstrained value [e.g., see Figs. 5(b) and 5(c)].

In large strain cases where the strain splitting is greater than the phonon energy, the interband phonon scattering is totally suppressed and scattering only takes place within the top band which now can be well approximated by an ellipsoid. The DOS mass for the top band is then $m_D^{\rm top} = (m_a m_b m_c)^{1/3}$ where m_a, m_b, and m_c are the three masses along the three axes of the ellipsoid. Under in-plane biaxial stress, the effective mass in the x-y plane is close to isotropic, but under uniaxial stress, the effective mass is highly anisotropic, and the three mass values are different. The scattering rate is then

$$1/\tau' \propto \frac{C^2 n(\omega_o)\pi}{\hbar^2\sqrt{\omega_o}}\left(\frac{2m_D^{\rm top}}{\hbar}\right)^{3/2} \tag{29}$$

and

$$\frac{\mu'}{\mu} = \left(\frac{m_D^{\rm hh3/2} + m_D^{\rm lh3/2}}{m_D^{\rm top3/2}}\right) \times \left[\frac{m_D^{\rm hh3/2} + m_D^{\rm lh3/2}}{(m'/m_{\rm hh})m_D^{\rm hh3/2} + (m'/m_{lh})m_D^{\rm lh3/2}}\right], \tag{30}$$

where m' is the effective mass along the channel direction under strain.

In the situation where the strain splitting is smaller than the phonon energy, the splitting energy can be taken into account in the scattering rate approximately by considering that far away from the Γ point, the band recovers its unstrained curvature. Thus at relatively large k, the top band recovers HH-like and the second band LH-like character. The DOS can still be approximated using the density-of-states effective masses, with the only difference being the two bands are split by energy ΔE. For Si and Ge, the scattering rate for holes residing at the Γ point at the top band can be estimated by

$$1/\tau' \propto \frac{C^2 n(\omega_o)\pi}{\hbar^{3/2}}\left[\frac{1}{\sqrt{\hbar\omega_o}}\left(\frac{2m_D^{\rm hh}}{\hbar}\right)^{3/2} + \frac{\sqrt{\hbar\omega_o - \Delta E}}{\hbar\omega_o}\left(\frac{2m_D^{\rm lh}}{\hbar}\right)^{3/2}\right]. \tag{31}$$

This equation only differs from the unstrained case by the scattering contribution from the LH band being reduced due to the strain splitting in the second term. We can see that when $\hbar\omega_o \gg \Delta E$, this difference is small. For Si, $\hbar\omega_o \sim 62$ meV, and assuming $\Delta E = 20$ meV, $\Delta\tau/\tau \sim 2.5\%$. This means that the mobility enhancement in Si at the small strain approximation can be considered to be solely from the charge redistribution and consequently the conductivity effective mass change. The optical phonon energy in Ge ($\hbar\omega_o = 35$ meV) is less than that in Si, but the LH mass is much smaller than the HH mass, so the mobility enhancement due to the scattering rate reduction can also be neglected at the low strain limit.

Next we consider the phonon scattering in polar semiconductors. Because of the different electron-phonon coupling of the polar semiconductors compared to Si and Ge, under certain circumstances, the splitting can make the polar optical phonon scattering more effective. With strain splitting, the scattering rate becomes

$$1/\tau' \propto \frac{2\pi A n(\omega_o)}{\hbar\omega_o}\left\{\frac{m_D^{\rm hh}}{\hbar^2}\frac{(2m_D^{\rm hh}\hbar\omega_o/\hbar^2)^{3/2}}{(2m_D^{\rm hh}\hbar\omega_o/\hbar^2 + q_0^2)^2} + \frac{m_D^{\rm lh}}{\hbar^2}\frac{[2m_D^{\rm lh}(\hbar\omega_o - \Delta E)/\hbar^2]^{3/2}}{[2m_D^{\rm lh}(\hbar\omega_o - \Delta E)/\hbar^2 + q_0^2]^2}\right\}. \tag{32}$$

If $\Delta E = 0$, then the case without strain is recovered. We can see that the first term does not change and the second term corresponds to the DOS of the LH band at energy $\hbar\omega_o - \Delta E$. If the hole density is small and the screening effect can be neglected, the second band contribution to the phonon scattering actually increases. This is in contrast with Si and Ge because of the different electron-phonon coupling for GaAs which is inversely proportional to q^2 in the low carrier density approximation. This effect is schematically illustrated in Fig. 9.

For GaAs, $m_D^{\rm hh} = 0.50 m_0$, $m_D^{\rm lh} = 0.076 m_0$, and $\hbar\omega_o = 34$ meV. For a 400 MPa stress, ΔE is ≃20 meV and $\Delta\tau/\tau$ is ~−16%. This indicates that due to strain splitting, the scattering rate actually increases by 16%, which will increase the resistivity and thus lower the mobility. The screening becomes strong as the hole density increases. In the strong screening limit $\{q_0^2 \gg [2m_D^{\rm lh}(\hbar\omega_o - \Delta E)/\hbar^2]\}$, reduction of the DOS dominates, and the scattering decreases with strain

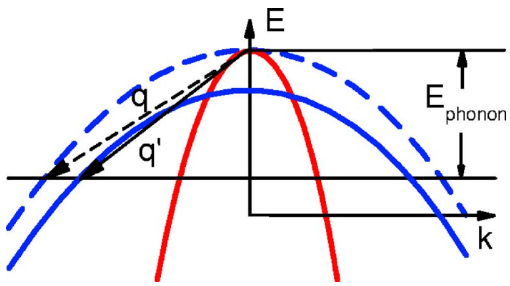

FIG. 9. Diagram of phonon scattering in GaAs valence bands. Due to the strain splitting, q decreases for scattering to the second band which enhances the electron-phonon coupling strength and enhances the phonon scattering.

splitting. In the large strain limit when the splitting is larger than the optical phonon energy, the mobility change shall follow the discussion for Si and Ge in the large strain limit, i.e., scattering reduction is mainly caused by reduced joint DOS.

From the earlier discussion, the hole mobility enhancement is predominantly determined by the conductivity mass change at the low stress limit and is determined by both the conductivity mass and phonon scattering rate reduction in large strain cases. The assumption that all holes reside in the top band only holds for low hole densities. If the hole density is high enough or the strain is small enough that the Fermi energy is much larger than the splitting energy, the redistribution actually can be neglected and consequently there is no pronounced effect on mobility.

In the normal inversion operation mode, carriers in MOSFET channels are 2D electron/hole gases. Two new features must be considered for the carrier transport: (i) electric confinement and (ii) semiconductor-oxide interface scattering. Due to the quantum confinement, the band structure then depends both on the confinement and strain splitting in strained MOSFET devices. For quantitative purposes, interface scattering must also be taken into consideration in addition to the phonon scattering. In the next section, the properties of electric confinement and interface scattering are discussed.

VII. ELECTRIC CONFINEMENT AND SURFACE ROUGHNESS SCATTERING

The motion of carriers in MOSFET channels is restricted in the gate direction (z direction) and quantized, leaving only a 2D k vector which still characterizes the Bloch wave motion in a plane (x-y plane) normal to the confining potential. An energy band which extends in the k_z direction in bulk materials is now split into a series of subbands. The properties of the subbands and transport properties of the inversion and accumulation layers at semiconductor-insulator interfaces have been reviewed in detail in Ref. 92. Quantum confinement in the z direction also reduces the crystal symmetry. Because of the different effective masses in the confinement direction, the six conduction valleys in Si split into Δ_2 and Δ_4 valleys. The degeneracy of the Ge conduction band is not altered by confinement. The valence bands in all group IV and III–V semiconductors split into two sets of subbands with the ground state of HH "on top." The subband splitting induced by quantum confinement can be additive or subtractive to the strain-induced splitting.

There are many quantitative methods to address the 2D conduction band structures such as the triangular potential well approximation,[93] Hartree self-consistent variational method[94] or Hartree–Fock approximation,[95] etc. In addition to determining the z dependent confining potential, subband structures are also required to be computed for the valence bands. The $\mathbf{k}\cdot\mathbf{p}$ method can be employed to calculate the subband structures in the triangular well approximation[96] or the confining potential and the subband structures can be obtained through a self-consistent procedure.[97] For simplicity, here we only use the triangular well approximation to qualitatively discuss the subband splitting for both the conduction and valence bands. At high substrate doping concentrations (current technological interest), the triangular potential well approximation gives satisfactory results.[23,98]

In the triangular potential well approximation, the potential is given by

$$V(z) = \begin{cases} eFz & z > 0 \\ \infty & z \leq 0 \end{cases}, \tag{33}$$

where F is the effective field along the z direction. By matching the boundary conditions, the subband energies are

$$E_n = r_n \left(\frac{\hbar^2}{2m_z}\right)^{1/3} (eF)^{2/3}, \tag{34}$$

where r_n are the roots for the equation $A(-r)=0$: $r_0=2.338$, $r_1=4.087$, $r_2=5.520$, $r_3=6.787$, $r_4=7.944\ldots$, and $A(-r)$ is the Airy function. Specifically, the ground state energy is

$$E_0 = 2.338\left(\frac{\hbar^2}{2m_z}\right)^{1/3} (eF)^{2/3}. \tag{35}$$

Under an effective field of magnitude 1 MV/cm, $(\hbar^2/2m_z)^{(1/3)}(eF)^{(2/3)}=72.5/(m_z/m_0)^{1/3}$ meV. The splitting between the ground state Δ_2 valley ($m_z=0.92$) and Δ_4 valley ($m_z=0.19$) of Si is $\sim$121 meV. For a 2D system, the DOS is reduced to a step function of energy and given by

$$D(E) = \frac{m}{\pi\hbar^2}, \quad E > E_0$$
$$= 0, \; E < E_0, \tag{36}$$

where m is the DOS mass in the $x-y$ plane. For a subband with a DOS mass m^*, $D(E)=4.19(m^*/m_0)\times 10^{14}/(\mathrm{cm}^2\,\mathrm{eV})$. For an electron density of $1\times 10^{13}/\mathrm{cm}^2$ (approximately corresponding to an effective field of 1 MV/cm in state-of-the-art devices), $\sim$80% of the electrons occupy the Δ_2 valleys which have a DOS mass of $2\times 0.19m_0$. Since confinement already favors populating Δ_2 valleys, strain-induced electron repopulation is not significant in modern devices.

The confining potential also restricts the carriers to a thin layer close to the oxide/semiconductor interface, where they experience interface scattering, which is not present in bulk materials. Interface scattering results from two sources: (i) Coulomb scattering by the fixed charges in the oxide or by trapped charges at the semiconductor-oxide interface and (ii)

interface-roughness scattering due to the deviation of the interface from an ideal plane. Interfacial charge scattering, like impurity scattering, is only dominant at low temperature. Surface roughness scattering, however, is believed to play an important role affecting carrier mobility at room temperature. Because surface roughness scattering is caused by the uneven charge potential at the interface, it is elastic and scatters an electron/hole to a state with the same energy. At present, difficulties remain in treating surface roughness scattering precisely, but according to the calculations by Fishetti[23] and Sun,[97] surface roughness scattering is about one order of magnitude weaker than phonon scattering at room temperature in relaxed *p*-Si channels.

Surface roughness scattering increases with the gate voltage due to reduced channel thickness (increased confinement) and may decrease with strain according to Fischetti.[21] At a fixed gate voltage, applying stress shifts the subbands and modifies the DOS, which changes both phonon and surface roughness scattering rate. In the next section, we first discuss the phonon-limited mobility, and then take surface roughness scattering into consideration as a correction.

VIII. STRAINED *N*- AND *P*-CHANNEL MOSFETS

Strain effects in MOSFETs are determined by the coaction of both electric confinement and strain. First, confinement and strain splitting can be additive or subtractive. Second, strain alters the in-plane 2D band structures, leading to changes of both conductivity and DOS effective masses. For studying the strain effects on MOSFET channels, we need to concentrate on three aspects of the strain altered subband structure: (i) the out-of-plane effective mass, which determines the magnitude of the energy level shift under the applied gate voltage of the MOSFET; (ii) the conductivity effective mass along the [110] channel direction; and (iii) the energy contours in the $k_x - k_y$ plane, which determines the 2D DOS for a given subband. In this section, the hole mobility is first studied, followed by a simple discussion of the electron mobility.

The valence bands under electric confinement split into two sets of subbands according to their effective hole masses along the confinement direction z following Eq. (34). For Si, the HH and LH effective masses do not differ significantly along [001] and thus the confinement splitting between the two sets of subbands of Si is much smaller than that of Ge and GaAs, which have a much larger difference between the HH and LH effective masses. For a confining effective field of 1 MV/cm, the splitting for Si is ~30 meV, compared to ~180 meV for Ge and ~145 meV for GaAs. The magnitude of confinement splitting consequently determines the importance of interband scattering. Due to the large confinement splitting and relatively small optical phonon energy, interband scattering is not important in Ge and GaAs.

For biaxial tensile stress, the strain-induced subband splitting is subtractive to the confinement-induced splitting because the top band in the [001] direction has a LH-like character while the second band has a HH-like character as shown in Fig. 5(b), and thus the two sets of subbands approach each other with stress as shown in Fig. 10 (energy zero taken at the valence band edge). For Si, due to the small difference between the HH and LH masses, and reduced net splitting between the subbands, the conductivity mass and phonon scattering both increase in the low stress range. Thus, the hole mobility for Si has an initial degradation with biaxial tensile stress.[28,99] Only after the two sets of subbands cross over each other and the phonon scattering reduction dominates due to reduced joint DOS, does the hole mobility significantly increase. This trend agrees with our theoretical results[99] and experiments by Rim *et al.*[28] Due to the large confinement splitting and the dramatic in-plane mass decrease for the ground subband because of the large HH and LH mass difference, Ge and GaAs may not have an initial mobility degradation when applying biaxial tensile stress. But the stress required to make the two sets of subbands cross over to offset the confinement splitting is large and may not be practically feasible.

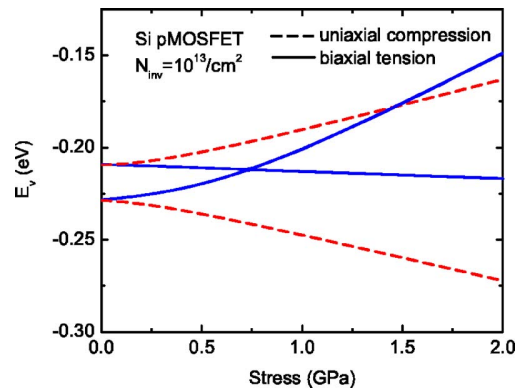

FIG. 10. Splitting between the ground states of the two sets of Si valence subbands vs stress with inversion hole density of $10^{13}/cm^2$. Energy zero is taken at the bulk valence band edge. Strain and confinement splitting are additive for uniaxial compressive stress and are subtractive for biaxial tensile stress.

The drawbacks and difficulties of using biaxial tensile stress to enhance the hole mobility can be circumvented by using [110] uniaxial compressive stress. For longitudinal compression, the top band is HH-like and second band is LH-like in the [001] direction, and thus the strain and confinement splitting are additive as shown in Fig. 10. Additional strain-induced splitting of the subbands reduces the phonon scattering. The in-plane mass change with uniaxial stress is also favorable for hole mobility. With longitudinal uniaxial compression, the hole effective mass along [110] becomes smaller than the unstressed LH mass. The hole effective mass along $[\bar{1}10]$ becomes very large and thus also ensures a large 2D DOS mass in the ground subband. This results in more light holes available for current conduction. A much higher hole mobility enhancement factor is expected for uniaxial longitudinal compression than in the biaxial tensile stress case and there is no mobility degradation as in the biaxial stress case at the low stress range for Si.

Next we give a rough estimate for the hole mobility enhancement factor in the large stress limit. For this, we first inspect the conductivity mass of the unstressed in-plane HH and LH subbands of Si. The in-plane HH band is strongly anisotropic as shown in Fig. 7. The heavy holes are located

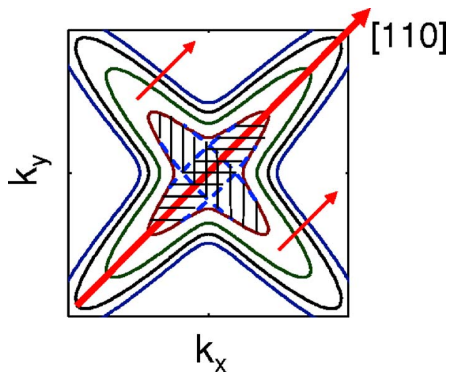

FIG. 11. The in-plane 2D energy contour for Si HH subbands. The shaded areas are the two ellipses to represent the enclosed area by energy contours.

in the four wings enclosed by the energy contours. As a simple approximation, these four wings can be represented by two partly overlapped ellipses as shown in Fig. 11. The heavy holes in the two longitudinal wings contribute with the longitudinal mass and those in the two transverse wings contribute with the transverse mass. For unstrained Si, the holes are evenly distributed in the two ellipses, the average conductivity mass is $m_c = 2m_l m_t/(m_l + m_t)$. The longitudinal mass is $0.61m_0$ as we already obtained in Fig. 5(a), and the transverse mass can be obtained by expanding Eq. (17) using the condition $k_t \ll k_l$, which gives $m_0/m_t = \gamma_1 - (7\gamma_2^2 - 3\gamma_3^2)/\sqrt{\gamma_2^2 + 3\gamma_3^2}$, so $m_t = 0.16m_0$. Therefore, the conductivity mass for heavy holes in Si is about $0.25m_0$. This value is also checked numerically by averaging the contributions to the mobility of holes inside the energy contour and found to be very accurate. This method can also be applied to Ge and GaAs, which gives HH conductivity masses of $0.15m_0$ and $0.26m_0$ for Ge and GaAs, respectively. The LH band is close to isotropic, and the conductivity mass of the light holes can be represented by the mass along the channel direction, i.e., $0.15m_0$, $0.042m_0$, and $0.081m_0$ for Si, Ge, and GaAs, respectively. The 2D DOS mass of a subband can be approximately represented by the corresponding 3D DOS mass.

For estimation convenience, we can neglect the confinement splitting between the HH and LH subbands for hole mobility of unstressed Si. This approximation is valid for low electric fields because of weak confinement, and up to an electric field of 0.6 MV/cm (which corresponds approximately to a hole density of $1 \times 10^{13}/\text{cm}^2$) because the confinement splitting is small and the optical phonon energy is large for Si. With this approximation, the mobility has contributions from both heavy holes and light holes and the scattering rate is proportional to the DOS contributed by both sets of subbands. When the stress-induced splitting is large enough so that the mobility is dominated by the ground subband, the mobility enhancement factor is

$$\frac{\mu'}{\mu} = \left(\frac{e\tau'}{m_c'}\right) \Big/ \left(\frac{e\tau}{m_c}\right) = \frac{m_c}{m_c'} \times \frac{m_D}{m_D'}, \tag{37}$$

where the primed values represent the stressed parameters. We can see that in Si the mobility enhancement due to biaxial tensile stress results from the scattering reduction ($m_D'/m_D < 1$), and is compromised by the conductivity mass increase ($m_c'/m_c > 1$). By using the effective mass values discussed above for unstressed Si and assuming in-plane parabolic bands for strained Si, the phonon limited hole mobility enhancement factor at large biaxial stress limit is estimated to be about 1.8–1.9. This enhancement is close to the calculations by Takagi *et al.*,[20] Vogelsang *et al.*,[100] and the experiment by Rim *et al.*[28] but less than the 2.5 obtained by some other authors.[27,30,101] For unstressed Ge and GaAs, the ground subband is HH due to large confinement splitting and relatively small optical phonon energy, and thus m_c and m_D in Eq. (37) are values for the HH band only. Using Eq. (18) and parameters from Table II, the stressed ground subband mass values at large stress limit can also be obtained. The phonon limited hole mobility enhancement factors at large biaxial tensile stress limit for Ge and GaAs are then about 4.2 and 2.3, respectively.

Under longitudinal uniaxial compressive stress, the top band along $[\bar{1}10]$ has a large effective mass compared to that along [110]. At the large stress limit, the contour of the top band in the energy range of interest in the *x*-*y* plane is an ellipse as discussed previously. Because of the strong band warping in the *x*-*y* plane, the stressed effective masses have a large deviation from those obtained from Eq. (19). We calculated the effective mass values at the Γ point under 1 GPa uniaxial stress using the 6×6 Hamiltonian and used them to approximate the effective masses at the large stress limit. This approximation is justifiable since stress larger than 1 GPa tends to warp the in-plane subbands at higher k and does not alter the contour a few $k_B T$ from the valence band edge. Using Eq. (37), the phonon limited hole mobility enhancement factors at the large uniaxial compressive stress limit for Si, Ge, and GaAs are found to be about 4, 16, and 14, respectively. For Si, this enhancement factor is reasonable even without considering surface roughness scattering, since the phonon limited mobility in Si is much lower than surface roughness limited mobility[23,97] and strain effects on hole mobility are dominated by mass change and phonon scattering reduction. However, since surface roughness in Ge is stronger[102] and the phonon limited mobility is much higher than in Si, the effect of surface roughness scattering cannot be neglected. Surface roughness scattering varies according to the stress-altered joint DOS. The DOS change for Ge is about four times at large uniaxial stress. The mobility enhancement factor limited by both phonon and surface roughness is expected to be significantly smaller than 16. We can expect the same situation to occur for GaAs channels, because devices based on compound semiconductors have poorer interfaces than Si devices. The hole mobility enhancement for Si, Ge, and GaAs *p*MOSFETs was calculated for large uniaxial stress by employing the six band $\mathbf{k} \cdot \mathbf{p}$ method and self-consistently solving the Schrödinger-Poisson equation to find the confining potential. The enhancement factors as a function of stress are shown in Fig. 12. Surface roughness scattering is taken into account following Ref. 23 with surface roughness parameters for Ge and GaAs taken to be the same as for Si. Experimental data from Thompson *et al.*[24] and Wang *et al.*[103] for uniaxially stressed Si are also shown in Fig. 12. Besides the higher enhancement factor for Ge and GaAs than Si, it is also important to note that the hole mo-

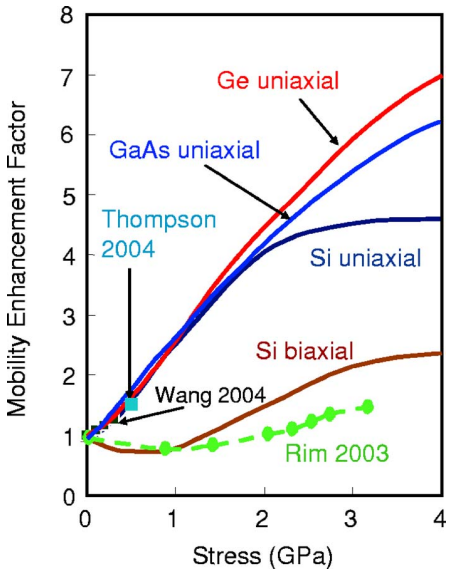

FIG. 12. Hole mobility enhancement factors of Si, Ge, and GaAs as a function of stress. The dashed line is the experimental observation of hole mobility enhancement vs biaxial stress by Rim *et al.* (see Ref. 28). Square points are experimental data by Thompson *et al.* (see Ref. 24) and Wang *et al.* (see Ref. 103).

bilities of Ge and GaAs increase steadily with stress up to 4 GPa, while the hole mobility of Si saturates at about 2 GPa. For the technologically important stresses of 1–2 GPa, Ge shows similar enhancement as Si. However, the unstressed hole mobility of Ge is ~3× higher than Si.

In summary for *p*MOSFETs, the mobility enhancement for biaxial stress is due to the reduction of scattering under high stress, while for uniaxial stress, the mobility enhancement results from both conductivity mass decrease at low and moderate stress, and scattering reduction at high stress. Hole mobility enhancement is more pronounced in semiconductors which have larger differences between the HH and LH masses. Strain enhanced mobility for state-of-the-art metal-insulator-semiconductor FET structures which use high-κ gate insulators[104,105] will follow the same physics discussed in this section.

For Si *n*MOSFETs, in-plane tensile stress (e.g., the biaxial tensile and longitudinal tensile stress) leads to additive strain and confinement splitting. The electron mobility enhancement mainly results from the interband scattering suppression due to the large confinement-induced valley splitting at normal operation fields. This leads to the enhancement for $I_{d\mathrm{sat}}$, the drive current at the saturation region, being approximately the same as that for $I_{d\mathrm{lin}}$, the drive current in the linear region for short channel Si *n*MOSFETs, while for short channel Si *p*MOSFETs, the enhancement for $I_{d\mathrm{sat}}$ is about 1/2 as that for $I_{d\mathrm{lin}}$.[106,107] For Ge, both the electric confinement along ⟨001⟩ and in-plane biaxial stress do not split the conduction valleys. In contrast, the uniaxial tensile stress along the ⟨110⟩ will remove the valley degeneracy and lowers the two *L* valleys in the (110) plane perpendicular to the stress axis, resulting in electron mobility enhancement due to electron repopulation. Stress is not expected to have significant effects on GaAs *n*-type device mobilities due to the single conduction band degeneracy and relatively large band gap. For narrow gap direct band materials such as InAs or InSb, [110] uniaxial compressive stress may lower the mass along the [110] direction due to band warping. Our calculation based on an eight-band Kane's model[108] shows that 1 GPa uniaxial compressive stress reduces the electron effective mass for InAs along the [110] direction by about 10%.

IX. III–V SEMICONDUCTORS: BENCHMARKING TO STRAINED SI

While strain introduced into Si CMOS technology is one approach for mobility enhancement, other possible means include finding new channel materials with high mobility such as compound semiconductors. III–V high electron mobility transistors[109] (HEMTs) and metal-semiconductor field effect transistors[110] (MESFETs) have been studied for decades because of their high electron mobilities compared to bulk Si and have recently gained renewed industry interest.[111] However, HEMTs or MESFETs are not compatible with the current Si technology and suffer from a low turn-on voltage and large gate leakage. To date, III–V MOSFETs for logic applications have had limited success. In this section, we investigate the properties of compound semiconductors relevant to CMOS applications as a potential option to replace strained Si channel MOSFETs.

Compared to strained Si, most III–V semiconductors with similar band gaps have electron effective masses that are several times smaller. The heavy-hole masses are typically similar, but the light-hole masses are much lighter. Thus, the hole mobilities of III–V semiconductors and Si are comparable since the heavy hole dominates *p*-type semiconductor transport, but the Si electron mobility is several to tens of times smaller. For comparison, the effective masses and mobilities for Si, Ge, and several direct gap III–V semiconductors are listed in Table III.

Then the question arises: Are III–V channel MOSFETs the next evolution for the semiconductor industry post-strained Si?[99,112,113] To answer this question, we shall first assume a surface channel structure since it seems inconceivable that a buried channel device could have adequate short channel control and drive current to compete with the 20–30 nm gate length Si devices already in production. With this assumption, we can make side-by-side comparisons between some promising III–V semiconductors with Si and strained Si.

The historical issue for III–V materials is the semiconductor/insulator interface. Unlike the Si/SiO_2 interface, which has very low defect densities, III–V semiconductors do not have a good native oxide, and for a variety of lab-grown dielectrics, the interface contains a high density of interfacial states.[114–119] Recently, Bell laboratories[120–123] and Freescale[124–126] reported a high quality Ga_2O_3 film on GaAs with unpinned Fermi level in the surface channel and successful enhancement and depletion mode operation of GaAs MOSFETs.

Assuming the III–V semiconductor/insulator interface issue can be solved, it is still uncertain if III–V channels offer advantages over strained Si. First, the small electron effective mass in III–V materials leads to a low DOS and unde-

TABLE III. Band and transport parameters for Si, Ge, and some III–V semiconductors.

	E_g (eV)	m_n	μ_n (cm^2/V s)	m_{hh}/m_{lh}	μ_h (cm^2/V s)	g_{mn} (Norm.)	$\Delta E_{\Gamma L}$ [c] (eV)	$m_l/m_t(L)$ [c]
Si	1.12	0.92/0.19	1450	0.53/0.15	500	1	⋯	⋯
Ge	0.67	1.59/0.082	3900	0.33/0.043	2270	0.92	⋯	⋯
InSb	0.17	0.014	7.7×10^4	0.45/0.016	850	1.9	0.51	1.56/0.094
InAs	0.35	0.024	$2-3.3\times10^4$	0.41/0.026	100–450	1.29	0.72	1.56/0.094
GaSb	0.73	0.041	3750	0.40/0.05	680	0.28	0.084	0.95/0.11
InP	1.34	0.08	5370	0.6/0.089	150	0.77	0.59	1.9/0.15
GaAs	1.42	0.063	9200	0.5/0.076	400	1.03	0.29	1.9/0.075
Strained Si	1.08	⋯	2900[a]	⋯	2200[b]	2	⋯	⋯

[a]See Ref. 26.
[b]See Ref. 99.
[c]These two columns list the L valley data for selected III–V semiconductors. The other data in the table are taken from Refs. 134 and 135.

sirable quantum effects.[127] Since the conduction band for direct gap compound semiconductors is approximately spherical, the DOS mass is approximately equal to the electron effective mass. The small effective mass in the z direction results in less confinement leading to the inversion charge density peaking deeper in the substrate, which reduces capacitance. Neglecting gate depletion, the gate capacitance (per unit area) is approximately given by $C_g = \varepsilon_{ox}/t_{eff}$, where ε_{ox} is the permittivity of the oxide, $t_{eff}=t_{ox}+t_{inv}$ is the effective oxide thickness, with t_{ox} being the physical oxide thickness, and t_{inv} the inversion layer thickness contribution, which accounts for the inversion electron distribution into the substrate. The inversion electron concentration versus depth is shown in Fig. 13 for InSb, InAs, GaAs, and unstrained and strained Si as computed by solving the Schrödinger equation and Poisson's equation self-consistently. To illustrate the dependence of the charge distribution on electron effective mass, only the occupation of the Γ valley is assumed for the three compound semiconductors. InSb has the lightest electron effective mass and the electron distribution in the substrate extends to several hundreds of angstroms, which adds an unacceptably large t_{inv} to t_{eff}. The electron distribution of unstrained Si has contributions from electrons in both Δ_2 and Δ_4 valleys. The out-of-plane masses of both the Δ_2 and Δ_4 valleys are much larger than that of GaAs, which has the heaviest electron effective mass among the three III-V semiconductors shown in Fig. 13, and thus the electron distribution in Si is much closer to the surface. Strain further populates the Δ_2 valley with larger out-of-plane mass, leading to greater confinement.

A low DOS and small confinement mass has significant performance issues for inversion MOSFETs. Both degrade the transconductance $g_m = W\mu C_g V_D/L \propto \mu/t_{eff}$ by requiring a larger gate voltage-swing to change the charge density,[127,128] where W, L, V_D, μ are the device width, channel length, drain voltage, and channel mobility, respectively. For example, for an inversion charge density of $10^{13}/\text{cm}^2$, the Fermi energy is over 1 eV above the conduction band edge for InSb and InAs, but only 0.26 eV for Si, as illustrated in Fig. 14. The transconductance is approximately proportional to DOS and mobility as given by

$$g_m = \frac{dI_d}{dV_g} \propto \frac{dn_e\mu}{dV_g} \propto \mu\frac{dn_e}{dV_g} \propto \mu\frac{dm_D E_F}{dV_g} \propto \mu m_D. \tag{38}$$

The normalized nMOSFET transconductances are shown in Table III. From the table, it is apparent that no III–V materials exceed the transconductance of strained Si.

The second issue for III–V semiconductors is the potential for electron population in low mobility valleys. At a high channel charge density (e.g., $>10^{13}/\text{cm}^2$) which is typical for state-of-the-art Si technology, the high Fermi energy in III–V channels resulting from the low DOS in the Γ valley and large shift of the Γ valley with confinement causes the occupation of energetically close conduction valleys which are usually characterized by lower electron mobilities. For

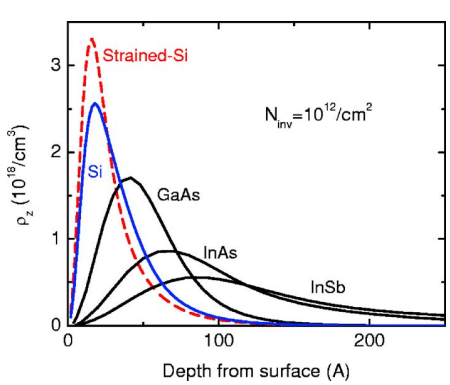

FIG. 13. Electron concentration as a function of the depth into the substrate from the channel surface for Si, GaAs, InAs, and InSb.

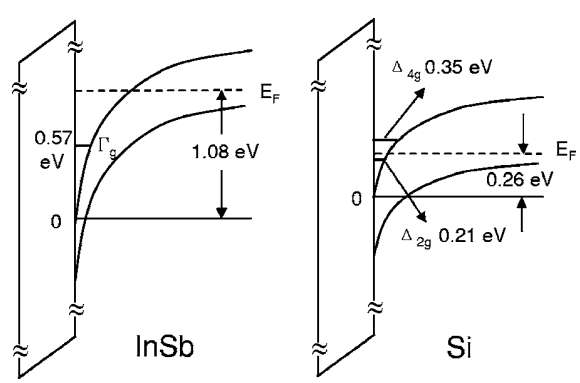

FIG. 14. Band bending for InSb and Si for the same inversion electron density of $10^{13}/\text{cm}^2$. The subscript letter "g" labels the ground subband for each conduction valley.

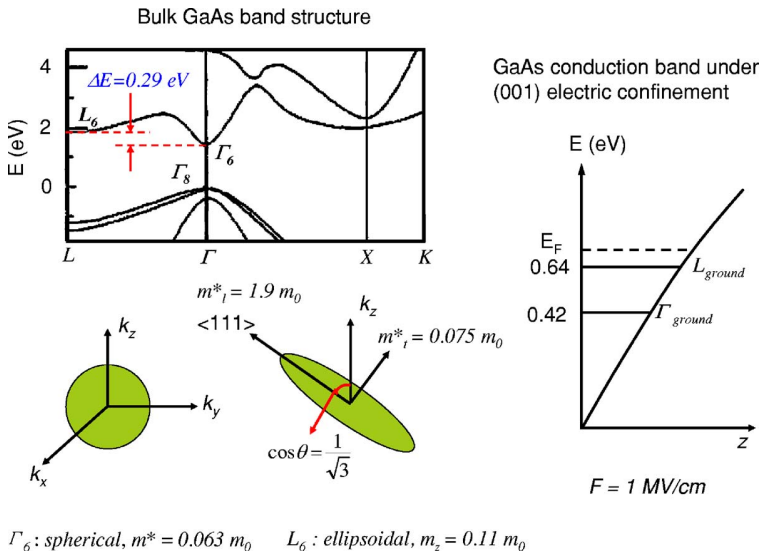

FIG. 15. GaAs band structure reproduced after Chelikowsky and Cohen (see Ref. 129) and the diagrams of the conduction band Γ valley and L valley, with the diagram showing the ground states for Γ valley and L valley under an electric field of 1 MV/cm. Occupation of electrons in the L valley is considerable.

GaAs, the electron effective mass in the Γ valley is about $0.063m_0$ and $\Delta E_{\Gamma L}$, which is the valley splitting between Γ and L, is about 0.29 eV. Under electric confinement, due to the fact that the energy level shift is inversely proportional to $m_z^{1/3}$, and m_z in the L valley is larger, the energy subbands originating from the Γ valley shift more than the subbands originating from the L valley. Thus, the two sets of subbands become closer. Due to the small DOS mass in the Γ valley, the occupation of the L valley becomes highly probable. The large mass in the L valley increases the electron effective mass and at the same time greatly enhances intravalley and intervalley phonon scattering. Therefore, the electron mobility is expected to be pronouncedly lowered at high inversion charge density. This situation can be schematically shown in Fig. 15 for GaAs where the electron ground state is calculated using the triangular well potential approximation. The Fermi energy is above the ground subband of the L valley, resulting in significant occupation of the L valley.

The third issue is the strong nonparabolicity present especially in narrow gap semiconductors such as InSb and InAs, which have extremely small electron masses and high electron mobilities. The nonparabolicity of the conduction band is due to the interaction between the conduction and valence bands caused by the energetic proximity. Away from the Γ point, the effective mass increases greatly with energy. When the device size is small and the charge density is high, the nonparabolicity-induced effective mass increase is significant. The nonparabolicity is inversely proportional to the band gap and can be described by one single parameter α, namely, $\hbar^2k^2/2m^*=E(1+\alpha E)$. The parameter α used in the literature for GaAs and Si are both about 0.5 eV^{-1},[130–132] but for InSb, α=4.1 eV^{-1}.[133–135] The nonparabolic conduction band of InSb is shown in Fig. 16 together with the parabolic approximation for comparison. The nonparabolocity effect reduces the electron mobility compared to the values listed in Table III which only considers the electron occupation at the band edge.

The electron conductivity mass versus inversion layer electron density is shown in Fig. 17 for InSb, InAs, GaAs, unstressed Si, and 1 GPa biaxial tensile stressed Si. For the three compound semiconductors, the occupation of the L valleys is also considered. For low doped Si, the variation of the conductivity mass with inversion electron density results from the two types of conduction valleys (in-plane and out-of-plane valleys). With 1 GPa biaxial tensile stress, the inversion charge mostly occupies the out-of-plane valleys which are characterized by lower, nearly constant conductivity mass in the channel direction. In contrast, InSb, InAs, and GaAs all show increases of conductivity masses with inversion electron density. This increase of conductivity mass results from the stronger nonparabolicity and the occupation of higher conduction valleys. In InSb and InAs, the nonparabolocity-induced conductivity mass increase contributes greatly, while for GaAs, the increase mainly comes from the L-valley occupation.

The fourth issue is the leakage current in narrow gap semiconductor MOSFETs where subthreshold leakage and the band-to-band tunneling (BTBT) will ultimately limit the minimum device size, which, for many applications, is the dominant metric since this directly affects cost per transistor.

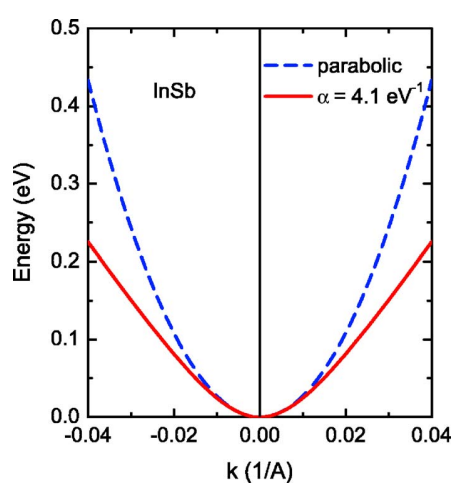

FIG. 16. Nonparabolic conduction band for InSb with parameter α =4.1 eV^{-1} and the parabolic conduction band with effective mass m =$0.014m_0$.

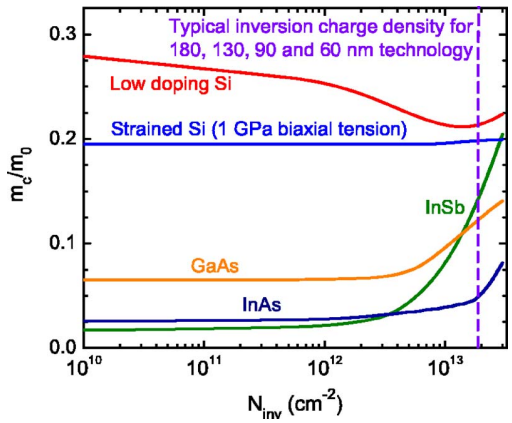

FIG. 17. Electron conductivity mass as a function of the inversion layer charge density for Si, InSb, InAs, and GaAs. Both nonparabolicity and occupation of the L valleys are considered for the three compound semiconductors.

The off-state minority carrier density is approximately given by $n_{\text{off}}=Ne^{-E_g/k_BT}$ where N is the effective DOS. For MOSFETs with large band gap channel materials, the off-state minority carrier density is sufficiently low, but MOSFETs with narrow band gap channel materials such as InSb have a high minority carrier density in the off-state, which leads to large subthreshold leakage. The low source-to-drain barrier in narrow gap channels also enhances the source to drain direct tunneling ("I" in Fig. 18), further increasing the off-state leakage current.

In addition to the direct tunneling, two types of BTBT are also important. One is from the source to the drain ("II" in Fig. 18), and the other is from drain to substrate ("III" in Fig. 18). The BTBT effect is not significant for wide gap materials due to the tunneling probability dependence on the band gap and electron effective mass[136,137] (generally the smaller the band gap, the smaller the electron effective mass), but is severe for narrow gap semiconductors. The investigation by Pethe *et al.*[138] revealed that the BTBT currents from source to drain for InSb and InAs channel nMOSFETs are 10^4-10^5 times larger than Si nMOSFETs.[138] One possible structure to reduce the BTBT leakage is depicted in Fig. 19, where a thin layer of a narrow gap material is grown on a relatively large gap material to construct a composite channel.[139] The narrow gap layer is undoped to minimize the impurity scattering. Since the inversion charge is strongly confined to the surface, with the band offset further enhancing the confinement, most charge resides in the narrow gap layer. This reduces the band bending in the narrow gap layer. Due to the reduction of the junction area of the drain with the narrow gap material, the drain to substrate BTBT tunneling is greatly reduced. This structure with strained-Ge/Si composite channel has been verified by Monte Carlo simulations to have $>3.5\times$ mobility enhancement and $10\times$ BTBT reduction.[139] However, the major drawback of this structure is the integration complexity, and defects or dislocations from the heteroepitaxy, which will likely impact yield, performance, or reliability.

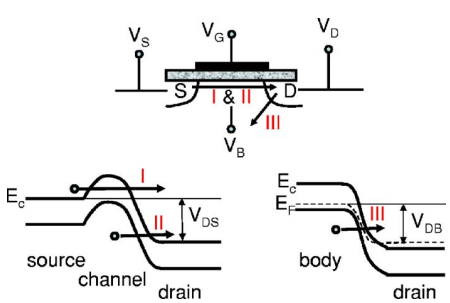

FIG. 18. Electron tunneling mechanisms important in narrow gap channels. "I" labels the source to drain direct tunneling, "II" labels the source to drain BTBT process, and "III" labels the drain to substrate BTBT process.

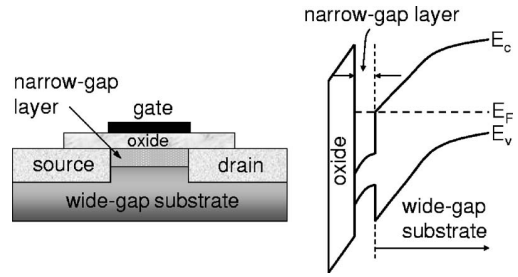

FIG. 19. A possible MOSFET structure that uses composite channel to enhance the mobility and reduce the BTBT current.

Based on the earlier issues and discussions in Secs. VI and VIII, strain is not an effective means to improve the electron mobility in direct gap III–V nMOSFETs. For III–V channel pMOSFETs, strain needs to be employed to make III–V materials competitive with strained Si. When strain is employed, it further changes the device leakage. Strain effects on leakage need to be considered due to the key importance of leakage on power dissipation.

X. STRAIN EFFECTS ON LEAKAGE CURRENT

Various leakage mechanisms contribute to the total leakage, with important components being subthreshold, gate oxide tunneling, gate induced drain, and reverse biased junction BTBT leakage. Strain alters the properties of current leakage due to the following effects: (i) strain-induced subband shift and realignment; (ii) strain-induced band warping; (iii) strain-induced band gap change; and (iv) strain effects on the oxide. Strain-induced subband shift changes the potential barrier between the oxide and semiconductor, and the subband realignment and warping changes the DOS and charge repopulation, modifying the joint DOS of the tunneling process. Band warping also varies the effective mass for tunneling, and thus alters the tunneling probability. Due to the different shift of conduction and valence bands by strain (both hydrostatic and shear), the strained semiconductor band gap is altered, which leads to a shift of the intrinsic Fermi level and carrier density, and to a threshold voltage shift. Tunneling processes dependent on the band gap such as BTBT are also altered. Except for biaxial stress which is introduced via the substrate and only strains the channel, process-induced uniaxial stress strains the channel, gate electrode, and insulator. Strain-induced changes of the gate insulator include the shift of the insulator conduction and valence bands and also the variation of the insulator layer thickness. To simplify the

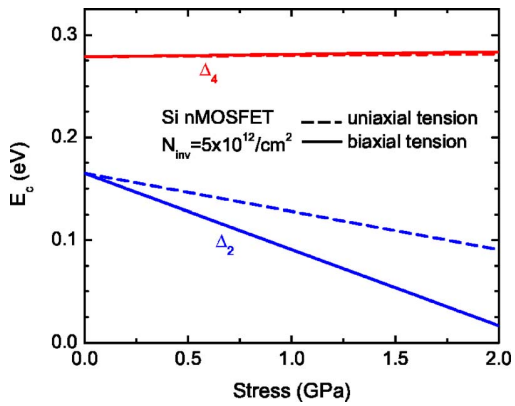

FIG. 20. Si conduction band valley splitting vs stress.

discussion, strain effects on the gate insulator are neglected, in order to concentrate on strain induced changes to semiconductors.

Because the *n*MOSFET confines the electrons and the *p*MOSFET confines holes, the conduction and valence subband shifts need to be considered separately. For Si *n*MOSFETs, under in-plane biaxial and uniaxial tensile stress, confinement splits the Δ_2 and Δ_4 valleys which shift differently under stress with the Δ_2 valleys descending and Δ_4 valleys rising which are shown in Fig. 20 (the bulk conduction band edge taken to be at zero energy). The energies of the top two hole subbands of Si under biaxial tensile and longitudinal uniaxial compressive stress are shown in Fig. 10. The relative ordering of the subbands (e.g., HH or LH character) can be distinguished by detailed investigation described in previous sections. The ground subband shifts up and the second subband shifts down under uniaxial compressive stress, while under biaxial tensile stress, the ground subband shifts down and second subband shifts up.

The qualitative band gap shift with stress can be obtained from Figs. 1(b) and 2(b) or by inspecting both the conduction and valence band shift from Figs. 10 and 20. The bulk band gap change with stress for Si, Ge, and several III–V semiconductors is shown in Fig. 21 where in-plane biaxial tensile and longitudinal uniaxial compressive stress are considered. Uniaxial compressive stress decreases the band gap slightly, while biaxial tensile stress is seen to reduce the band gap greatly for all semiconductors shown.

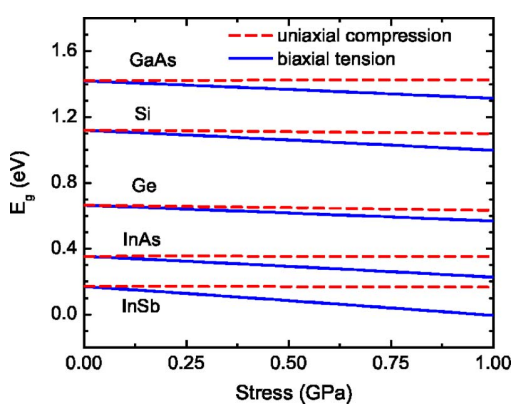

FIG. 21. Band gap change vs stress for Si, Ge, GaAs, InAs, and InSb.

With the knowledge of the subband shift and band gap change with stress, the effects of stress on current leakage can be predicted and understood. First, we inspect the gate oxide tunneling. With the increasing splitting and descending shift of the Δ_2 valleys in Si *n*MOSFETs, the barrier height increases. Furthermore, the increased population in the Δ_2 valley increases the average m_z. Since the tunneling probability decreases exponentially with the square root of barrier height[55] and also decreases with m_z, tensile stress reduces the gate oxide tunneling leakage current.[43] The tunneling for strained Si *p*MOSFETs is more complicated because the top two subbands are close to each other. However, the downward shift of the ground subband with the biaxial tensile stress will increase gate leakage. Yang *et al.*[91] studied the tunneling current dependence on stress for Si *p*MOSFETs, with full consideration of the charge occupation and carrier lifetime in each subband, and found that the gate oxide tunneling current increases for biaxial tensile stress and decreases for uniaxial compressive stress. The magnitude of tunneling current changes for both biaxial and uniaxial stress is typically small, being <5% for 300 MPa stress.

The threshold voltage is also altered by stress due to changes to the semiconductor band gap and is important to account for in performance benchmarking.[53] The band gap change causes both the extrinsic and intrinsic Fermi levels to shift. This shift is approximately proportional to the band gap change as is the threshold voltage change. The DOS change in both the conduction and valence bands also affects the Fermi energy and this also shifts the threshold voltage, but this is a smaller effect. The threshold voltage shifts for uniaxial and biaxial tensile-stressed Si *n*MOSFETs have been investigated by Lim *et al.*[53] who found that the threshold voltage shift induced by uniaxial is significantly smaller than that by biaxial stress, which is consistent with Fig. 21.

The BTBT current change by strain comes from both the band gap change and the effective mass change. From Kane's model, the leakage current density can be expressed as[140]

$$J_{\mathrm{BTBT}} = \frac{\sqrt{2m}e^3 F V_{\mathrm{app}}}{\sqrt{E_g} 4\pi^3 \hbar^2} \exp\left(-\frac{4\sqrt{2mE_g^3}}{3\hbar e F}\right), \tag{39}$$

where m is the effective mass of electron, E_g is the band gap, V_{app} is the applied reverse bias between drain and substrate, and F is the electric field at the junction. For process-induced stress, it is reasonable to assume that both sides of the drain/substrate junction are strained. If we only consider the band gap change, then for Si *p*MOSFETs under 2 GPa biaxial tensile stress, the band gap reduction is about 0.25 eV, which induces a leakage current increase of about 10^6 times, approximating the effective mass as $0.3m_0$, and assuming a doping density for both drain and substrate of $10^{18}/\mathrm{cm}^3$ and junction electric field of 1 MV/cm. This increase is very large and undesirable. However, the band gap reduction for 2 GPa uniaxial compression is about 0.05 eV and the BTBT leakage current increases by about only 20 times.

XI. CONCLUSION

In conclusion, physical insight of strain effects on MOSFET mobility may be gained from a combination of symmetry, tight-binding, and $\mathbf{k}\cdot\mathbf{p}$ considerations. In-plane tensile stresses, including both biaxial and longitudinal tensile stress, are suitable for Si *n*MOSFET mobility enhancement, while uniaxial longitudinal compressive stress is favorable to improve *p*MOSFET performance. The uniaxial longitudinal compressive stress creates both a small conductivity mass and a large 2D DOS mass which is also important to populate the carriers in the ground subband. The hole mobility enhancement by uniaxial longitudinal stress is larger in semiconductors which have a greater difference between the HH and LH masses. For Si, the estimated maximum hole mobility enhancement factor is about 4.5, but for Ge and GaAs, this number is both around 10. The leakage induced by stress limits the application of stress in small devices. Biaxial tensile stress induces much greater BTBT leakage than uniaxial compressive stress, posing a severe problem for application in scaled devices. Considering the inversion layer confinement, low DOS, and large leakage in III–V channel materials, III–V materials face obstacles to overtake strained Si in high density and high performance logic circuit applications.

ACKNOWLEDGMENTS

This work was supported by the Applied Materials Foundation, Advanced Micro Devices (AMD), Cypress Semiconductor, IBM, Intel Foundation, Semiconductor Research Corporation (SRC), Texas Instruments, Taiwan Semiconductor Manufacturing Co. (TSMC), United Microelectronics Corporation (UMC), and the National Science Foundation (NSF) under Grant No. ECS-0524316.

[1]V. Chan *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2003**, 381.
[2]T. Ghani *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2003**, 1161.
[3]F. Arnaud *et al.*, VLSI Symp. Tech. Dig. **2004**, 10.
[4]P. R. Chidambaram *et al.*, VLSI Symp. Tech. Dig. **2004**, 48.
[5]R. Khamankar *et al.*, VLSI Symp. Tech. Dig. **2004**, 162.
[6]S. Hao *et al.*, VLSI Symp. Tech. Dig. **2004**, 14.
[7]S. E. Thompson *et al.*, IEEE Electron Device Lett. **25**, 191 (2004).
[8]E. Leobandung *et al.*, VLSI Symp. Tech. Dig. **2005**, 126.
[9]N. Collaert *et al.*, VLSI Symp. Tech. Dig. **2006**, 52.
[10]H. Shang *et al.*, VLSI Symp. Tech. Dig. **2006**, 56.
[11]C. Ortolland *et al.*, VLSI Symp. Tech. Dig. **2006**, 78.
[12]X. Chen *et al.*, VLSI Symp. Tech. Dig. **2006**, 60.
[13]H. M. Manasevit, I. S. Gergis, and A. B. Jones, Appl. Phys. Lett. **41**, 464 (1982).
[14]R. People, J. C. Bean, D. V. Lang, A. M. Sergent, H. L. Stormer, K. W. Wecht, R. T. Lynch, and K. Baldwin, Appl. Phys. Lett. **45**, 1231 (1984).
[15]W. Shockley and J. Bardeen, Phys. Rev. **77**, 407 (1950).
[16]J. Bardeen and W. Shockley, Phys. Rev. **80**, 72 (1950).
[17]C. S. Smith, Phys. Rev. **94**, 42 (1954).
[18]M. L. Lee, E. A. Fitzgerald, M. T. Bulsara, M. T. Currie, and A. Lochtefeld, J. Appl. Phys. **97**, 011101 (2005).
[19]P. R. Chidambaram, C. Bowen, S. Chakravathi, C. Machala, and R. Wise, IEEE Trans. Electron Devices **53**, 944 (2006).
[20]S. Takagi, J. L. Hoyt, J. J. Welser, and J. F. Gibbons, J. Appl. Phys. **80**, 1567 (1996).
[21]M. V. Fischetti, F. Gámiz, and W. Hänsch, J. Appl. Phys. **92**, 7320 (2002).
[22]R. Oberhuber, G. Zandler, and P. Vogl, Phys. Rev. B **58**, 9941 (1998).
[23]M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, J. Appl. Phys. **94**, 1079 (2003).
[24]S. E. Thompson, G. Sun, K. Wu, J. Lim, and T. Nishida, Tech. Dig. - Int. Electron Devices Meet. **2004**, 221.
[25]S. Pidin *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2004**, 213.
[26]J. Welser, J. L. Hoyt, S. I. Takagi, and J. F. Gibbons, Tech. Dig. - Int. Electron Devices Meet. **1994**, 373.
[27]M. T. Currie, C. W. Leitz, T. A. Langdo, G. Taraschi, D. A. Antoniadis, and E. A. Fitzgerald, J. Vac. Sci. Technol. B **19**, 2268 (2001).
[28]K. Rim *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2003**, 49.
[29]J. W. Jung, M. L. Lee, S. F. Yu, E. A. Fitzgerald, and D. A. Antoniadis, IEEE Electron Device Lett. **24**, 460 (2003).
[30]C. W. Leitz, M. T. Currie, M. L. Lee, Z. Y. Cheng, D. A. Antoniadis, and E. A. Fitzgerald, J. Appl. Phys. **92**, 3745 (2002).
[31]C. Herring and E. Vogt, Phys. Rev. **101**, 944 (1956).
[32]M. V. Fischetti and S. E. Laux, J. Appl. Phys. **80**, 2234 (1996).
[33]W. Shockley, Phys. Rev. **78**, 173 (1950).
[34]J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[35]G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1974).
[36]W. H. Kleiner and L. M. Roth, Phys. Rev. Lett. **2**, 334 (1959).
[37]H. Hasegawa, Phys. Rev. **129**, 1029 (1963).
[38]J. C. Hensel and G. Feher, Phys. Rev. **129**, 1041 (1963).
[39]J. C. Hensel, Solid State Commun. **4**, 231 (1966).
[40]I. Balslev, Phys. Rev. **143**, 636 (1966).
[41]R. N. Bhargava and M. I. Nathan, Phys. Rev. **161**, 695 (1967).
[42]F. H. Pollak and M. Cardona, Phys. Rev. **172**, 816 (1968).
[43]J. Lim, X. Yang, T. Nishida, and S. E. Thompson, Appl. Phys. Lett. **89**, 073509 (2006).
[44]J. Welser, J. L. Hoyt, and J. F. Gibbons, Tech. Dig. - Int. Electron Devices Meet. **1992**, 1000.
[45]D. K. Nayak, J. C. S. Woo, J. S. Park, K. L. Wang, and K. P. MacWilliams, Appl. Phys. Lett. **62**, 2853 (1993).
[46]S. Ito *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2000**, 247.
[47]A. Shimizu, K. Hachimine, N. Ohki, H. Ohta, M. Koguchi, Y. Nonaka, H. Sato, and F. Ootsuka, Tech. Dig. - Int. Electron Devices Meet. **2001**, 433.
[48]R. Arghavani, L. Xia, H. M'Saad, M. Balseanu, G. Karunasiri, A. Mascarenhas, and S. E. Thompson, IEEE Electron Device Lett. **27**, 114 (2006).
[49]S. Gannavaram, N. Pesovic, and C. Ozturk, Tech. Dig. - Int. Electron Devices Meet. **2000**, 437.
[50]K. Uchida, R. Zednik, C. Lu, H. Jagannathan, J. McVitie, P. C. McIntyre, and Y. Nishi, Tech. Dig. - Int. Electron Devices Meet. **2004**, 229.
[51]S. E. Thompson *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2002**, 61.
[52]P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors: Physics and Materials Properties*, 3rd ed. (Springer, Berlin, 2001).
[53]J. Lim, S. E. Thompson, and J. G. Fossum, IEEE Electron Device Lett. **25**, 731 (2004).
[54]K. Uchida, T. Krishnamohan, K. C. Saraswat, and Y. Nishi, Tech. Dig. - Int. Electron Devices Meet. **2005**, 129.
[55]R. Shankar, *Principles of Quantum Mechanics*, 2nd ed. (Kluwer Academic, New York, 1994).
[56]J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
[57]D. J. Chadi and M. L. Cohen, Phys. Status Solidi B **68**, 405 (1975).
[58]W. A. Harrison, *Electronic Structure and the Properties of Solids* (Dover, New York, 1989).
[59]P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids **44**, 365 (1983).
[60]C. Priester, G. Allan, and M. Lannoo, Phys. Rev. B **37**, 8519 (1988).
[61]D. J. Chadi, Phys. Rev. B **16**, 790 (1977).
[62]T. B. Boykin, Phys. Rev. B **57**, 1620 (1998).
[63]T. B. Boykin, G. Klimeck, R. C. Bowen, and R. Lake, Phys. Rev. B **56**, 4102 (1997).
[64]G. Klimeck, R. C. Bowen, T. B. Boykin, C. S. Lazaro, T. A. Cwik, and A. Stoica, Superlattices Microstruct. **27**, 77 (2000).
[65]N. Cavassilas, F. Aniel, K. Boujdaria, and G. Fishman, Phys. Rev. B **64**, 115207 (2001).
[66]T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, Phys. Rev. B **66**, 125207 (2002).
[67]Energy integrals used in Chadi and Cohen's work (see Ref. 57). In Slater and Koster's work (see Ref. 56) and Harrison's book (see Ref. 58), symbols *E* were used instead of *V*, and they differ by a factor of 4.
[68]N. Cavassilas, F. Aniel, and G. Fishman, Nanotech 2002 Vol. 2: Technical Proceedings of the 2002 International Conference on Computational Nanoscience and Nanotechnology, 2002, pp. 411–414.
[69]T. B. Boykin, G. Klimeck, and F. Oyafuso, Phys. Rev. B **69**, 115201 (2004).

[70]J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, Phys. Rev. B **57**, 6493 (1998).
[71]M. Cardona and F. Pollak, Phys. Rev. **142**, 530 (1966).
[72]B. A. Foreman, Phys. Rev. B **48**, 4964 (1993).
[73]J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).
[74]D. Rideau, M. Feraille, L. Ciampolini, M. Minondo, C. Tavernier, H. Jaouen, and A. Ghetti, http://arxiv.org/PS_cache/cond-mat/pdf/0607/0607510.pdf.
[75]C. Mailhiot and D. L. Smith, Phys. Rev. B **33**, 8360 (1986).
[76]P. Lawaetz, Phys. Rev. B **4**, 3460 (1971).
[77]C. G. Van de Walle, Phys. Rev. B **39**, 1871 (1989).
[78]L. D. Laude, F. H. Pollak, and M. Cardona, Phys. Rev. B **3**, 2623 (1971).
[79]M. Chandrasekhar and F. H. Pollak, Phys. Rev. B **15**, 2127 (1977).
[80]I. Vurgaftman and J. R. Meyer, J. Appl. Phys. **89**, 5815 (2001).
[81]C. Y.-P. Chao and S. L. Chuang, Phys. Rev. B **46**, 4110 (1992).
[82]*Physics of Optoelectronic Devices*, edited by S. L. Chuang (Wiley, New York, 1992).
[83]D. Long, Phys. Rev. **120**, 2024 (1960).
[84]C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).
[85]D. K. Ferry, *Semiconductors* (Macmillan, New York, 1991).
[86]H. Miyata, T. Yamada, and D. K. Ferry, Appl. Phys. Lett. **62**, 2661 (1993).
[87]J. D. Wiley, Phys. Rev. B **4**, 2485 (1971).
[88]F. Szmulowicz, Appl. Phys. Lett. **43**, 485 (1983).
[89]F. Szmulowicz, Phys. Rev. B **28**, 5934 (1983).
[90]B. K. Ridley, *Quantum Processes in Semiconductors*, 4th ed. (Clarendon, Oxford, 1999).
[91]X. Yang, J. Lim, G. Sun, K. Wu, T. Nishida, and S. E. Thompson, Appl. Phys. Lett. **88**, 052108 (2006).
[92]T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[93]F. Stern, J. Appl. Phys. **39**, 1930 (1968).
[94]F. Stern, Phys. Rev. B **5**, 4891 (1972).
[95]R. B. Langhlin, J. D. Joannopoulis, and D. J. Chadi, Phys. Rev. B **21**, 5723 (1980).
[96]M. V. Fischetti and S. E. Laux, J. Appl. Phys. **80**, 2234 (1996).
[97]G. Sun and S. E. Thompson (unpublished).
[98]M. V. Fischetti and S. E. Laux, Phys. Rev. B **48**, 2244 (1993).
[99]S. E. Thompson, G. Sun, Y. S. Choi, and T. Nishida, IEEE Trans. Electron Devices **53**, 1010 (2006).
[100]T. Vogelsang and K. R. Hofmann, Appl. Phys. Lett. **63**, 186 (1993).
[101]M. L. Lee and E. A. Fitzgerald, J. Appl. Phys. **94**, 2590 (2003).
[102]T. Low *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2004**, 151.
[103]E. Wang *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2004**, 147.
[104]Meet the world's first 45 nm processor, http://www.intel.com/technology/silicon/45nm_technology.htm.
[105]IBM advancement to spawn new generation of chips, http://www-03.ibm.com/press/us/en/pressrelease/20980.wss.
[106]P. Bai *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2004**, 657.
[107]K. Mistry *et al.*, VLSI Symp. Tech. Dig. **2004**, 50.
[108]E. O. Kane, J. Phys. Chem. Solids **1**, 249 (1957).
[109]T. Mimura, S. Hiyamizu, T. Fujii, and K. Nanbu, Jpn. J. Appl. Phys. **19**, L225 (1980).
[110]R. C. Eden, GaAs IC Symposium 25th Annual Technical Digest, San Diego, CA, 2003, p. 7.
[111]SRC promotes III–V CMOS to push beyond silicon, http://www.fabtech.org/content/view/1659.
[112]S. E. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. T. Bohr, IEEE Trans. Semicond. Manuf. **18**, 26 (2005).
[113]S. E. Thompson, S. Suthram, Y. Sun, G. Sun, S. Parthasarathy, M. Chu, and T. Nishida, Tech. Dig.–Int. Electron Device Meet., 2006 (to be published).
[114]M. Schulz and E. Klansmann, J. Appl. Phys. **18**, 169 (1979).
[115]C. C. Shiue and C. T. Sah, Phys. Rev. B **19**, 2149 (1979).
[116]N. Suzuki, T. Hariu, and Y. Shibata, Appl. Phys. Lett. **33**, 761 (1978).
[117]R. P. H. Chang and J. J. Coleman, Appl. Phys. Lett. **32**, 332 (1978).
[118]T. Mimura, N. Yokoyama, and M. Fukuta, Appl. Phys. Lett. **34**, 642 (1979).
[119]J. Nishizawa and I. Shiota, Inst. Phys. Conf. Ser. **4**, 289 (1980).
[120]M. Passlack, M. Hong, J. P. Mannaerts, S. N. G. Chu, R. L. Opila, and N. Moriya, Tech. Dig. - Int. Electron Devices Meet. **2005**, 383.
[121]M. Passlack, M. Hong, J. P. Mannaerts, R. L. Opila, S. N. G. Chu, N. Moriya, F. Ren, and J. R. Kwo, IEEE Trans. Electron Devices **44**, 214 (1997).
[122]M. Hong, F. Ren, J. M. Kuo, W. S. Hobson, J. Kwo, J. P. Mannaerts, J. R. Lothian, and Y. K. Chen, J. Vac. Sci. Technol. B **16**, 1398 (1997).
[123]Y. C. Wang *et al.*, IEEE Electron Device Lett. **20**, 457 (1999).
[124]Freescale "solves" GaAs gate oxide problem, http://www.compoundsemiconductor.net/articles/news/10/1/25/1.
[125]K. Rajagopalan, R. Drooopad, J. Abrokwah, and M. Passlack, CS MANTECH Conference, Vancouver, British Columbia, Canada, 2006, p. 119.
[126]K. Rajagopalan, R. Droopad, J. Abrokwah, P. Zurcher, P. Fejes, and M. Passlack, IEEE Electron Device Lett. **28**, 100 (2007).
[127]M. V. Fischetti and S. E. Laux, IEEE Trans. Electron Devices **38**, 650 (1991).
[128]A. Rahman, G. Klimeck, and M. Lundstrom, Tech. Dig. - Int. Electron Devices Meet. **2005**, 601.
[129]J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B **14**, 556 (1976).
[130]X. M. Weng, J. Phys.: Condens. Matter **8**, 6229 (1996).
[131]D. M. Riffe, J. Opt. Soc. Am. B **19**, 1092 (2002).
[132]K. Yeom, H. Jiang, and J. Singh, J. Appl. Phys. **81**, 1807 (1997).
[133]P. Junker, U. Kops, and U. Merkt, Phys. Rev. B **49**, 4794 (1994).
[134]Semiconductor on NSM, http://www.ioffe rssi ru/SVA/NSM/Semicond/.
[135]*Data in Science and Technology: Semiconductors-Group IV Elements and III-V Compounds*, edited by O. Madelung (Springer, Berlin, 1991).
[136]Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices* (Cambridge University Press, New York, 1998).
[137]K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, Proc. IEEE **91**, 305 (2003).
[138]A. Pethe, T. Krishnamohan, D. Kim, S. Oh, H. C. P. Wong, Y. Nishi, and K. C. Saraswat, 16th Biennial University Government Industry Microelectronics Symposium, San Jose State University, San Jose, CA, 2006.
[139]T. Krishnamohan, D. Kim, Y. Nishi, and K. Saraswat, E-MRS 2006 Spring Meeting-SYMPOSIUM B: From Strained Silicon to Nanotubes: Novel channels for field effect devices, Nice, France, 2006, p. 147.
[140]E. O. Kane, J. Appl. Phys. **32**, 83 (1961).

View publication stats