Energy and momentum conservation during energetic-carrier generation and recombination in silicon

Yi Lu and Chih-Tang Sah
Florida Solid-State Electronics Laboratory, University of Florida, Gainesville, Florida 32611
(Received 27 January 1995)

The threshold energies of electrons and holes to generate electron-hole pairs, and energetic holes and electrons, by the interband impact and Auger mechanisms are computed analytically from energy and momentum conservation and found to be nearly equal to the minimum values based on energy conservation alone as was recognized by Kane as due to the intervalley impact generation pathway of electron-hole pairs. Experimental thresholds and energy dependency of the carrier generation efficiencies appear to support the analytic theory. The experimental data are obtained using the nearly monoenergetic primary electrons from Fowler-Nordheim tunneling through a thin gate oxide to generate secondary electron-hole pairs in a crystalline silicon substrate, or secondary hot holes in the polysilicon gate or thin crystalline silicon surface layer of p- and n-type channel silicon-gate/silicon-oxide/silicon field-effect transistors.

I. INTRODUCTION

Energetic electrons and holes in semiconductors can create energetic secondary electrons and holes via impact and Auger collisions. These carrier generation processes have been investigated for several decades due to their intriguing fundamental physics and controlling influence in applications: they limit the performance of semiconductor photon counters and voltage reference diodes and they reduce the reliability or operating life of present and future generations of submicrometer transistors and integrated circuits operated at low voltages (<3.3 V).

The two solid arrows in Fig. 1 (labeled by the four participating particles 1, 2, 3, 4) depict the two electron transitions during the *impact* generation of a secondary electron-hole pair (labeled 3 and 4) by a primary energetic electron (labeled 1). The two broken arrows are the inverse transitions during the *Auger* generation of an energetic hole (4') by interband Auger recombination of an electron (1') with a hole (2') whose recombination energy excites a hole (3') to a higher kinetic energy (4'). A simple model to analyze the interband impact generation rate of electron-hole pairs in Si was proposed by Wolff,[1] who did not consider the effect of the energy band structure. Later, Shockley[2] and Baraff[3] gave improved analyses by including an energy-independent ionization mean free path. Kane[4,5] then undertook a Monte Carlo calculation using the real energy band structure of silicon, and arrived at many results including a simple picture of intervalley impact transition to account for the near-energy-gap (near-E_G) value of the threshold electron energy. Recently, the pair generation rate by impact was further analyzed by Drummond and Moll[6] using Kane's "scattering rate" theory. The inverse problem, the Auger recombination, has been analyzed numerically by Laks, Neumark, and Pantelides,[7] who computed the total Auger recombination rate over the entire bandwidth by numerically evaluating the nine-dimensional transition-probability integral.

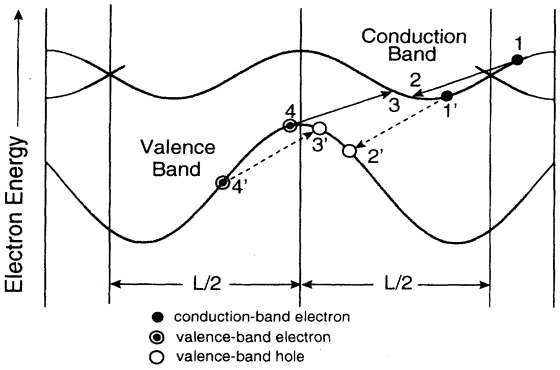

FIG. 1. Transition energy band diagram for impact (solid arrows) and Auger (dashed arrows) generation of secondary electrons and holes by an energetic primary electron, 1 and 1', in silicon.

There appear to be no experimental data to support Kane's near-E_G impact threshold energy. There are no theoretical and experimental data on the impact and Auger rates and thresholds of generating an energetic hole or electron at a specified final energy. The modifications of the threshold energies by the complex energy band structure of a real semiconductor (see Fig. 1 for Si's multivalley conduction bands) and by Bragg reflections or umklapp processes have not been taken into account in previous analytical theories. This paper will attempt to demonstrate that the complex-band and umklapp effects are important and may be estimated from an analytical solution which can be derived in closed form using a slight simplification of the band structure and Kane's random-k approximation, i.e., parabolic $E-k$ for the multiple valleys of the conduction band and spherical $E-k$ for the two warped upper valence bands, using ex-

0163-1829/95/52(8)/5657(8)/$06.00

perimental effective masses near the band edges and the free-electron mass at high energies. The theoretical results are then compared with experiments, showing rather good agreement and hence may serve as a starting point for a more complete analytical transition-rate theory.

II. THEORETICAL THRESHOLD ENERGIES

Using Fig. 1 the energy and momentum conservation relationships are given by

$$\phi_E = E_1 - E_2 - E_3 - E_4 - E_G = 0 , \tag{1}$$

$$\boldsymbol{\phi}_p = \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{p}_3 + \mathbf{p}_4 + \mathbf{p}_h = \mathbf{0} . \tag{2}$$

The ellipsoidal energy band assumption gives

$$E_i = \sum_\alpha \mathbf{p}_i^2/2m_{i\alpha} = \sum_\alpha \hbar^2 \mathbf{k}_{i\alpha}^2/2m_{i\alpha} \qquad (i=1,2,3,4,\ \alpha=x,y,z) , \tag{3}$$

where $m_{i\alpha}=m_1=0.916m$ or $m_{i\alpha}=m_t=0.191m$ near the conduction-band edge and $m_{i\alpha}=m$ at high energies, i.e., $E_i \geq E_{\rm th} > E_G$. The umklapp process introduces the reciprocal lattice vector in (2), $\mathbf{p}_h=\hbar\mathbf{K}_h$, where $\mathbf{K}_h$ is the translation vector of the reciprocal lattice. The component of the electron momentum will be denoted by $p_{i\alpha}=p_i\cos\theta_{i\alpha}$ where $\alpha=x,y,z$. p_i is the magnitude and $\cos\theta_i$ is its directional cosine. Then the momentum conservation law (2) of impact generation is given by

$$\phi_{p\alpha} = p_{1\alpha} - p_{2\alpha} - p_{3\alpha} + p_{4\alpha} + (p_{h\alpha} + p_{1\alpha}^m - p_{2\alpha}^m - p_{3\alpha}^m + p_{4\alpha}^m) = 0 , \tag{4}$$

where $p_{h\alpha}^m=\hbar k_{h\alpha}^m$, and $k_{h\alpha}^m$ are the conduction-band valley locations $(2\pi/a)\langle 0.81,0,0\rangle$. The $\mathbf{k}_1$ angle dependence of the primary particle's threshold energy E_1 to impact generate an electron-hole pair (3 and 4 in Fig. 1) can be obtained by variation calculus using $\partial E/\partial p_i=0$ where $E=E_1+\sum(\alpha)\lambda_{p\alpha}\phi_{p\alpha}+\lambda_E\phi_E$. This variation gives a quadratic equation in p_1,

$$ap_1^2 + bp_1 + c = 0 , \tag{5}$$

where

$$a = \sum_\alpha \tfrac{1}{2}\cos^2\theta_{1\alpha}[m_{1\alpha}^{-1} - (m_{2\alpha}+m_{3\alpha}+m_{4\alpha})^{-1}] , \tag{6}$$

$$b = \sum_\alpha (p_{n\alpha} + p_{1\alpha}^m - p_{2\alpha}^m - p_{3\alpha}^m + p_{4\alpha}^m) \times \cos\theta_{1\alpha}/(m_{2\alpha}+m_{3\alpha}+m_{4\alpha}) , \tag{7}$$

$$c = -E_G - \frac{(p_{n\alpha}+p_{1\alpha}-p_{2\alpha}-p_{3\alpha}+p_{4\alpha})^2}{2(m_{2\alpha}+m_{3\alpha}+m_{4\alpha})} . \tag{8}$$

The $\mathbf{k}_1$ angle dependences of the threshold energy E_1 are shown in polar coordinates in Fig. 2 for eight of the many calculated intravalley and intervalley pathways with or without umklapp. The numerical values of the parameters and the assumptions are the following: silicon lattice constant $a=5.43095$ Å, conduction-band valleys at the six locations $(2\pi/a)\langle 0.81,0,0\rangle$, ellipsoidal electron

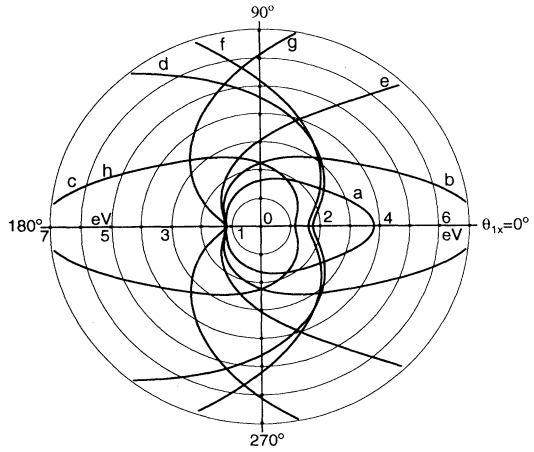

FIG. 2. Angular dependence of threshold energy of total electron-hole pair generation rate by electron impact in silicon. The initial and final valleys of the impacting (1 and 2) and impacted (3 and 4) electron and the umklapp vectors of the eight curves $a-h$ are given in the text.

effective masses of $0.916m$, $0.191m$, and $0.191m$ at and near the band edge, free-electron mass $1.0m$ when $E_i \geq E_{\rm th} > E_G$, spherical hole effective mass of $0.599m$, and indirect energy gap of $E_G=1.122$ eV. The reciprocal lattice vector of the diamond or silicon lattice is

$$\mathbf{K}_h = (2\pi/a)[h_x\mathbf{i}_x + h_y\mathbf{i}_y + h_z\mathbf{i}_z] \equiv (2\pi/a)[(n_1+n_2+n_3)\mathbf{i}_x + (n_1-n_2+n_3)\mathbf{i}_y + (n_1+n_2+n_3)\mathbf{i}_z]$$

where h_x, h_y, h_z, n_1, n_2, and n_3 are integers and zero and only certain $\mathbf{K}_h$'s have a nonvanishing Fourier component of the periodic potential due to the destructive interference from the two atoms in the diamond or silicon primitive unit cell. The characteristics of the eight threshold energy curves E_1 vs θ_{kl} have the following valleys for electrons 1,2,3, umklapp vector (h_x,h_y,h_z), and minimum threshold energy and directions. They are written as (1,2,3; h_x,h_y,h_z; E_1, angles).

$$\begin{aligned}
&a:\ (+x,-x,-x;\ -2,0,0;\ 1.1725\ \text{eV},\ 180°) ,\\
&b:\ (+x,-x,+x:\ 0,0,0;\ 1.1512\ \text{eV},\ 180°) ,\\
&c:\ (-x,+x,-x;\ 0,0,0;\ 1.1512\ \text{eV},\ 0) ,\\
&d:\ (+x,-x,-x;\ -4,0,0;\ 1.7532\ \text{eV},\ 0) ,\\
&e:\ (+x,+y,-y;\ 0,0,0;\ 1.1860\ \text{eV},\ 180°) ,\\
&f:\ (+x,+y,-y;\ -2,0,0;\ 1.6037\ \text{eV},\ 0) ,\\
&g:\ (+x,+z,-z;\ 0,0,0;\ 1.1860\ \text{eV},\ 180°) ,\\
&h:\ (+x,+x,+x;\ 0,0,0;\ 1.1815\ \text{eV},\ 0) .
\end{aligned} \tag{9}$$

These results show that the lowest threshold energy, $E_{1\,\rm min}$ is 1.1512 eV from the two identical inter-opposite-valley nonumklapp backscattering pathways given by

curves *b* and *c* in Fig. 2 whose one-dimensional (1D) view is depicted in Fig. 1. There are also other low-threshold nonumklapp backscattering pathways which are of the interdiagonal-valley type such as the two identical pathways whose cross-sectional curves in two planes are curves *e* and *g*. Their threshold minimum is 1.1860 eV. These intervalley impact generation thresholds are nearly equal to the energy gap, 1.122 eV, instead of the $3E_G/2$ used by most previous investigators. The intervalley transition is the cause as was recognized by Kane in his Monte Carlo calculations.[5]

In addition to the uncertainty in the interband impact threshold energy for total electron-hole pair generation just analyzed, there has been recent debates on the existence of a threshold voltage or threshold electron (or hole) kinetic energy to generate or build up positive (or negative) charges in thin SiO_2 films. Aside from the fundamental interest concerning the quantum-mechanical pathways that could be responsible for the creation of *positive* charges by *electrons,* the generated positive oxide charges have been shown to limit the operational lifetime of metal-oxide-silicon transistors and integrated circuits. (See Ref. 8 and references cited.) This lifetime limitation, from positive-charge buildup during electron transport, becomes increasingly critical as the transistor dimension decreases to less than 0.25 μm and its gate oxide film becomes thinner than ~ 6 nm. Our recent experimental results indicated the existence of electron kinetic-energy thresholds to build up positive oxide charge during Fowler-Nordheim electron tunneling through the thin oxide.[8] The threshold voltage data measured by us at different bias-voltage polarities on several different transistor structures agreed remarkably well with the analytical prediction based on energy conservation alone.[8] Thus it is desirable to provide a deeper understanding of the underlying reason using the energy-momentum conservation analysis just described. This is presented as follows for two pathways, the impact and Auger.

Consider first the interband impact mechanism that occurs in the heavily phosphorus-donor-doped (denoted by n^+) polycrystalline silicon gate (*n*-type gate) of an *n*-type channel metal-oxide-semiconductor (MOS) structure built on a *p*-type silicon substrate. Its energy band diagram and the dominant transition pathways (indicated by arrows) are shown in Fig. 3(a) when a positive voltage is applied to the $n^{\pm}$-type Si gate, labeled *G*, relative to the substrate silicon, labeled *X,S,D*. The current through the thin SiO_2 layer, J_G, is dominated by Fowler-Nordheim electron tunneling, from the *n*-type inversion layer on the *p*-type Si substrate to the SiO_2 conduction band, through the thin triangular SiO_2 potential barrier of the n^+-type gate/SiO_2/*n*-type channel structure. While passing through the SiO_2 film, these tunneled electrons are accelerated by the oxide electric field and gain kinetic energy, labeled $E_{k(\text{th-impact})}$. Phonon scattering energy loss in oxide films thinner than the mean free path is negligible. As the electrons enter the n^+-type gate, they fall off the n^+-type gate/SiO_2 potential cliff and gain an additional kinetic energy equal to the electron potential barrier

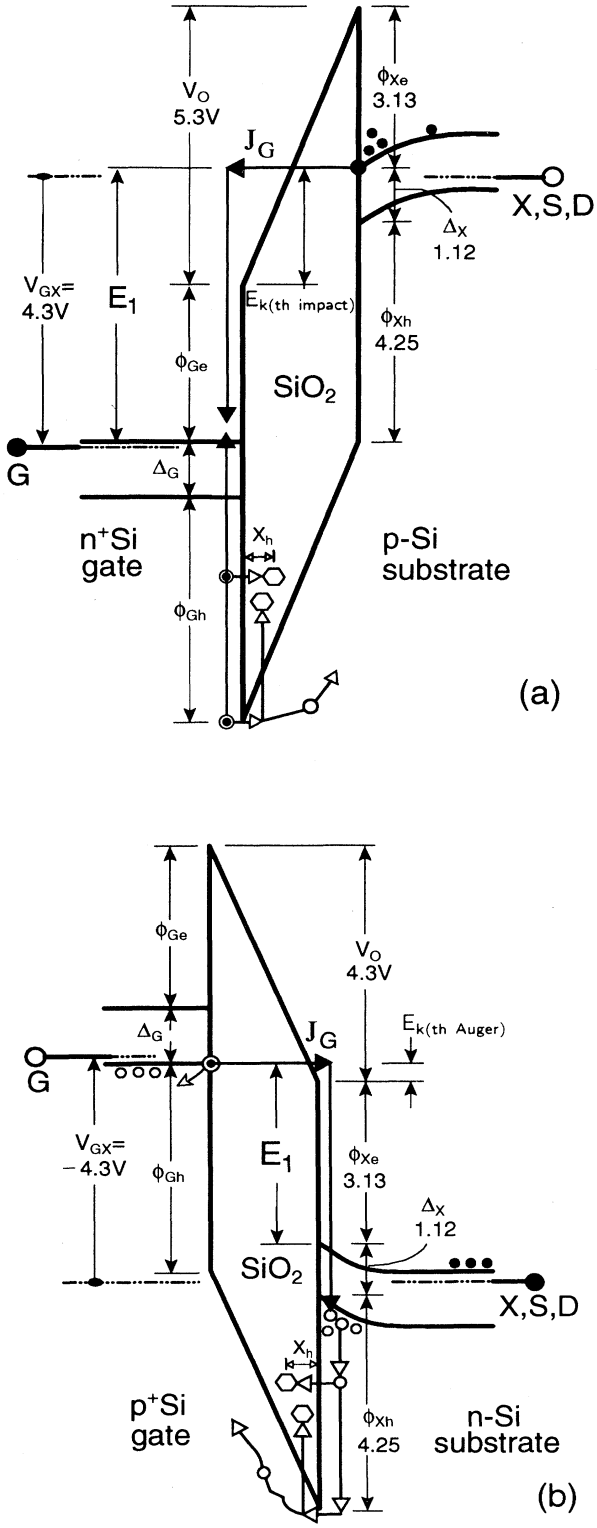

FIG. 3. Energy band diagrams to compute the threshold energy and voltage to build up positive oxide charge in (a) n^+-type gate/*n*-type MOS transistor during $+V_G$ stress ($V_{O\,th}=5.37$ V), and (b) p^+-type gate/*p*-type MOS transistor $-V_G$ stress ($V_{O\,th}=4.25$ V).

height at the n^+-type gate/SiO_2 interface, ϕ_{Ge}. These energetic electrons can now generate electron-hole pairs in the n^+-type gate via the interband impact mechanism. The electron-impact-generated hole with kinetic energy greater than the n^+-type gate/SiO_2 hole barrier height, $\phi_{Gh}=4.25$ eV, can then be back injected into the oxide valence band from the n^+-type gate by surmounting the 4.25 eV SiO_2/Si hole barrier. These injected holes can then be captured by oxide hole traps [hexagons in Fig. 3(a)] to increase the positive oxide charges.

Consider next the Auger pathways to generate hot holes, which is the inverse of the impact pathway just described. This is shown in Fig. 3(b) for a p^+-type gate/SiO_2/n-type Si MOS structure during negative gate voltage bias. The p^+-type gate is a polycrystalline silicon layer doped with a high concentration of boron acceptor impurity. In this pathway, valence-band electrons in the p^+-type gate tunnel through the SiO_2 barrier into the inverted p-type channel of the n-type Si substrate. The tunneled electrons then recombine with the inversion-layer holes via the Auger mechanism which generates energetic holes to carry off the recombination energy. The generated holes with kinetic energy greater than the SiO_2/Si hole barrier height ($\phi_{Xh}\simeq 4.25$ eV) can then be back injected into the oxide valence band and captured by oxide hole traps to increase positive oxide charges.

Based on energy conservation alone, Fig. 3(a) (intentionally drawn to be near the minimum threshold bias condition) shows that the minimum oxide voltage drop and the minimum electron kinetic energy required to create a 4.25 eV hot hole by the interband impact mechanism are 5.37 V and $5.37-3.13=2.14$ eV, respectively.[8] Similarly, energy conservation applied to Fig. 3(b) (also drawn near the minimum threshold, i.e., $E_{K(\text{th-Auger})}\simeq 0$) gives 4.25 V minimum oxide voltage drop and 0 eV minimum kinetic energy to create a 4.25 eV hot hole by the interband Auger mechanism. These theoretical energy-conservation-alone thresholds are in excellent agreement with experiments.[8] The agreements suggest two possibilities: (1) momentum conservation is unimportant (following Kane's assumption of random k via bulk phonon scattering in a large volume,[4,5] or supposing that the rough SiO_2/Si interface and the polycrystallinity of the silicon gate tend to randomize the momentum) or (2) as suggested by Fig. 2, some energy-momentum conservation intervalley backscattering pathways are responsible in these very thin layers whose thicknesses are less than the scattering mean free path and the size of the structural imperfections. An analysis of (2) is now given using the variational procedure. The momentum and energy conservation equations are listed as follows.

Both impact and Auger generation of a hot hole at $E_4=E_T$.

$$\phi_{p\alpha}=p_{1\alpha}-p_{2\alpha}-p_{3\alpha}+p_{4\alpha}$$
$$+(p_{h\alpha}+p^m_{1\alpha}-p^m_{2\alpha}-p^m_{3\alpha}+p^m_{4\alpha})=0\ , \tag{10}$$

$$\phi_T=E_4-E_T=\sum(\alpha)p^2_{4\alpha}/2m_{4\alpha}-E_T\ . \tag{11}$$

Impact generation.

$$\phi_E=E_1-E_2-E_3-E_4-E_G=0\ . \tag{12}$$

Auger generation.

$$\phi_E=E_1+E_2+E_3-E_4+E_G=0\ . \tag{13}$$

The momentum of the electron is $p_{i\alpha}=p_i\cos\theta_i$ ($\alpha=x,y,z$) where p_i is the magnitude of the momentum, $\cos\theta_i$ is its directional cosine, $p_{h\alpha}$ is the reciprocal lattice vector in the appropriate direction, and $p^m_{i\alpha}$ is the location of the valley minimum. The final hot-hole energy E_4 is predetermined as E_T which in this case is the SiO_2/Si hole barrier height, $E_4=E_T=\phi_{Xh}$ (or ϕ_{Gh})$=4.25$ eV. Taking the variation $\partial E/\partial p_{i\alpha}=0$, where

$$E=E_1+\lambda_E\phi_E+\sum(\alpha)\lambda_{p\alpha}\phi_{p\alpha}+\lambda_T\phi_T\ ,$$

We have the following.

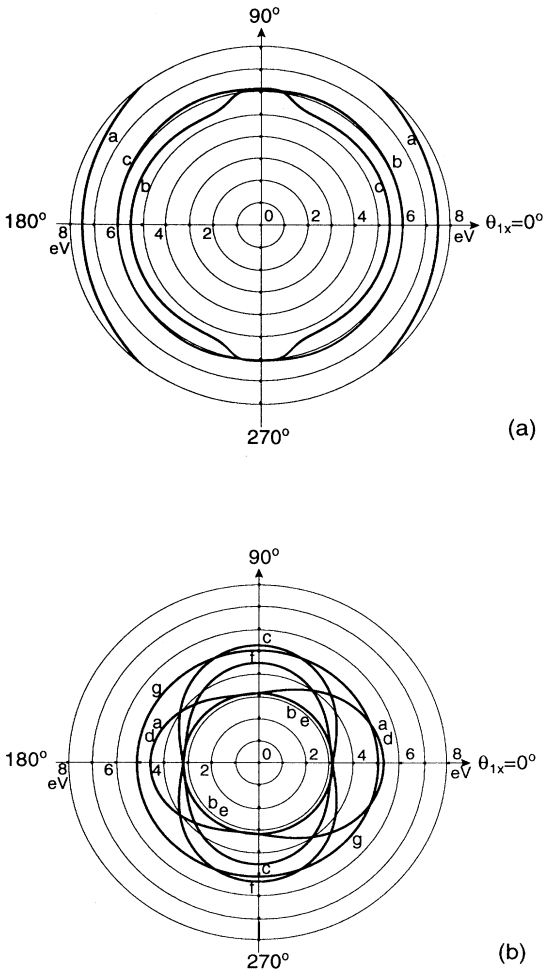

FIG. 4. Angular dependence of the primary electron threshold energy for generating 4.25 eV hot holes (a) by electron impact on a valence electron and (b) by Auger recombination with two holes.

Impact generation.

$$\sum_{\alpha}\frac{1}{2m_{4\alpha}}\left[\frac{(m_{2\alpha}+m_{3\alpha})X-m_{1\alpha}}{m_{1\alpha}}p_1\cos\theta_{1\alpha}+\hbar Q_\alpha\right]^2 -4.250=0 . \quad (14)$$

$$\sum_{\alpha}\frac{m_{1\alpha}+(m_{2\alpha}+m_{3\alpha})X^2}{2m_{1\alpha}^2}p_1^2\cos\theta_{1\alpha}^2-5.372=0 . \quad (15)$$

Auger generation.

$$\sum_{\alpha}\frac{1}{2m_{4\alpha}}\left[\frac{(m_{2\alpha}+m_{3\alpha})X-m_{1\alpha}}{m_{1\alpha}}p_1\cos\theta_{1\alpha}+\hbar Q_\alpha\right]^2 -4.250=0 , \quad (16)$$

$$\sum_{\alpha}\frac{m_{1\alpha}+(m_{2\alpha}+m_{3\alpha})X^2}{2m_{1\alpha}^2}p_1^2\cos\theta_{1\alpha}^2-3.128=0 , \quad (17)$$

where $Q_\alpha=p_{h\alpha}+p^m_{1\alpha}-p^m_{2\alpha}-p^m_{3\alpha}+p^m_{4\alpha}$ and $X=\lambda_E/(\lambda_E-1)$. The four significant figures reflect the consistency with the input, $E_G=1.122$ eV.

The numerical solutions give the threshold energy E_1 required to generate 4.25 eV hot holes as a function of the direction of $\mathbf{k}_1$. Many pathways are computed and a few lowest-energy ones are displayed in Fig. 4(a) for impact generation and in Fig. 4(b) for Auger generation. The same parameter values as those used for computing the pair generation threshold in Fig. 2 are used, except the effective masses. For the impact pathway, they are $m_{1x}=m_{1y}=m_{1z}=1.0m$ (high energy), $m_{2x}=m_{3x}=0.916m$, $m_{2y}=m_{2z}=m_{3y}=m_{3z}=0.191m$, and $m_{4x}=m_{4y}=m_{4z}=m_{\text{hole spherical}}=1.0m$ (high energy). For the Auger pathway, the masses are $m_{1x}=m_{1y}=m_{1z}=1.0m, m_{2h}=m_{3h}=0.54m$, and $m_{4h}=1.0m$. The electron-electron-electron valleys (impact) or electron-hole-hole (Auger) electron valley and hole band, umklapp vector, $E_{1\,\min}$, and directions are listed as follows. They are represented by (1,2,3; h_x,h_y,h_z; $E_{1\,\min}$, k_1 angles).

Impact generation pathways of 4.25 eV hot holes [Fig. 4(a)].

a: $(+x,-x,+x;\ 0,0,0;\ 7.484\,15\text{ eV},\ 0,180°)$,

b: $(+x,-x,-x;\ -2,0,0;\ 5.372\,00\text{ eV},\ 120°,240°)$, (18)

c: $(-x,+x,+x);\ 2,0,0;\ 5.372\,00\text{ eV},\ 60°,300°)$.

Auger generation pathways of 4.25 eV hot holes [Fig. 4(b)].

a: $(+x,0,0;\ 0,0,0;\ 3.128\,01\text{ eV},\ 97°,263°)$,

b: $(+y,0,0;\ 0,0,0;\ 3.154\,91\text{ eV},\ \text{isotropic})$,

c: $(+z,0,0;\ 0,0,0;\ 3.128\,01\text{ eV},\ 187°,353°)$,

d: $(-x,0,0;\ 0,0,0;\ 3.128\,01\text{ eV},\ 97°,263°)\equiv a$, (19)

e: $(-y,0,0;\ 0,0,0;\ 3.154\,91\text{ eV},\ \text{isotropic})\equiv b$,

f: $(-z,0,0;\ 0,0,0;\ 3.128\,01\text{ eV},\ 7°,173°)$,

g: $(+y,0,0;\ 0,\underline{2},0;\ 5.084\,70\text{ eV},\ \text{isotropic})$.

The impact 4.25 eV hot-hole generation results in (18) show that the minimum primary electron kinetic energy in the very thin oxide $E_{k\,(\text{th-impact-min})}$ and the minimum oxide voltage drop $V_{\text{O-th-min}}$ are $E_{k\,(\text{th-impact-min})}=2.242$ eV and $V_{\text{O-th-min}}=5.372$ V, respectively, involving inter-diagonal-valley umklapp 60° backscattering. These are equal to the energy-conservation-only value, $\phi_{Gh}+\Delta_G=4.25+1.122=5.372$ V.[8]

The Auger 4.25 eV hot-hole generation results in (19) shows several anisotropic lowest-energy pathways all with a minimum oxide voltage drop of 3.128 V or zero electron kinetic energy (3.128−3.13=0.002 eV=0 due to truncation error) which is identical to that predicted by energy conservation.[8] The backscatter angles at the minimum energies are 7° and 97°. There are also two identical isotropic pathways, b and e, whose minimum threshold energy is only slightly higher, 3.154 91 eV.

III. THEORETICAL GENERATION EFFICIENCIES

Threshold voltages or energies of these transitions and current pathways are generally difficult to measure due to noise and the very low and vanishing currents near the threshold. Extrapolation to the threshold from currents measured at higher voltages or energies is usually made. A knowledge of the theoretical kinetic-energy dependence or applied-voltage dependence of the electron-hole pair and hot-hole generation rates would be most helpful to guide the extrapolation of the experimental data. A simple analytical solution for each pathway just described can be derived if it is assumed that the energy dependence is dominated by the density of states and that the scattering matrix elements are nearly constant in the threshold energy range (Kane's random-k and scattering-rate models[5]). These assumptions can be tested by comparing the analytical results against the experimental energy or voltage dependences. For this purpose, analytical transition-probability rates are derived using the Fermi golden rule and spherical effective masses for all states. This can be improved by using the threshold model just analyzed: multivalley ellipsoidal effective masses for electrons near the conduction-band edge and spherical effective mass for holes near the valence-band edge. The results are listed as follows with the constant coefficient defined by

$$A=n_1^*[(2m^*)^{9/2}/(8\pi^6\hbar^9)](2\pi M^2/\hbar)(\pi/8) , \quad (20)$$

where n_1^* is the tunneled electron density given by the measured oxide gate current I_G, M is the matrix element assumed to be independent of energy ($M=M_I$ for impact and $M=M_A$ for Auger with $M_A\gg M_I$ probably), and m^* is the effective mass which is an appropriate combination of the electron and hole effective masses.

Total electron-hole pair impact generation rate by energetic electron.

$$\partial p/\partial t=A\tfrac{16}{105}(E_1-E_{\text{th-pair}})^{7/2} , \quad (21)$$

where $E_{\text{th-pair}} \simeq \Delta_X \equiv E_{\text{Si-gap}}$ when extrapolated from experimental data.

Hot-hole impact generation rate by energetic electron.

$$\partial p^*/\partial t = A\tfrac{16}{105}(E_1 - \Delta_G - \phi_{Gh})^{7/2} . \tag{22}$$

Hot-hole Auger generation rate by energetic electron.

$$\partial p^*/\partial t = A(P_2^2/N_V)(kT)^{7/2}\int_0^\infty (y_0+y)^{1/2}y^2\exp(-y)dy \tag{23}$$

$$\simeq A(P_2^2/N_V)2(kT)^3[E_1+E_G-\phi_h+(\tfrac{15}{16})^2\pi kT]^{1/2} \quad \text{for } y_0>0 \tag{23a}$$

$$= A(P_2^2/N_V)(kT)^{7/2}\exp(y_0)(\sqrt{\pi}/2)[\tfrac{15}{4}-3y_0+y_0^2] \quad \text{for } y_0<0 , \tag{23b}$$

where the normalized energy is $y_0 \equiv (E_1+\Delta_X-\phi_{Xh})/kT$ and $y_0<0$ is the subthreshold thermal tail region. The energy conservation threshold is

$$E_{\text{th 4.25 Auger}} \simeq \phi_{Xh} - \Delta_X = 4.35 - 1.1122 = 3.24 \text{ eV} .$$

P_S is the hole surface concentration in the p-type channel on the n-type substrate and $P_S \sim (V_{\text{oxide}})^{1+\delta}$.

The positive-oxide-charge buildup rate, $\partial q_{OT}/\partial t$, is then the product of (22) or (23), the fraction of hot holes backscattered into the oxide valence band, $T_{4'4}$ (~ 0.5), and the fraction of the oxide valence-band holes captured by the oxide hole traps while drifting through the oxide in the high oxide electric field ($\sim 20\%$). With these estimates, then $\partial q_{OT}/\partial t \simeq 0.1\times(\partial p^*/\partial t)$.

The $\frac{7}{2}$-power energy dependence of the total electron-hole pair generation rate (21) was evident even in the Monte Carlo calculation carried out by Kane,[5] which can also be extrapolated to give a threshold nearly equal to the Si energy gap 1.2 eV, rather than 1.8 eV (about $3E_G/2$). The analytical solution (21) was also given by Alig, Bloom, and Struck[9] in a slightly different form; they attributed the derivation to R. Klopfenstein. To our knowledge, the hot-hole generation rates (22) and (23) by energetic electron collision via impact and Auger processes have not appeared in the literature previously.

IV. EXPERIMENTAL RESULTS

In experiments, the generation efficiency is readily measured. The total pair generation efficiency by a monoenergetic electron in an n^+-type gate G/p-type channel transistor is given by

$$\eta_{\text{pair}} = \text{Eq. (21)}/J_G = (\partial p/\partial t)/J_G = (I_D+I_S)/I_G \propto (E_1-E_G)^{7/2} ,$$

The energy-conservation threshold

$$E_{\text{th 4.25 impact}} = \Delta_G + \phi_{Gh} = 1.122 + 4.35 = 5.472 \text{ eV}$$

is used in (22).

where I_D+I_S is the total hole current generated by the tunneled electrons. Similarly, the positive-oxide-charge buildup efficiency in either a MOS capacitor or transistor structure is given by

$$\eta_O^+ \equiv (dq_{OT}/dt)_0/J_G \propto (\partial p^*/\partial t)/J_G \propto (E_1-E_G-\phi_h)^{7/2}$$

for the impact mechanism and

$$P_S^2(E_1+E_G-\phi_h)^{1/2} \simeq V_O^2(E_1+E_G-\phi_h)^{1/2}$$

for the Auger mechanism.

The experimental generation efficiencies and threshold voltages were measured for both the MOS capacitor and transistor structures previously described.[8] The total pair generation efficiency by electron impact and its threshold, described by Fig. 2 and Eqs. (1)–(9), was measured using the well-known current separation property of a three-terminal transistor structure: the emitter, base, and collector currents in a bipolar junction transistor;[10] and the gate, source-drain, and substrate currents in a MOS field-effect transistor.[11] The n^+-type gate/p-type channel MOS transistor structure was used to exploit two unique properties: (i) the gate current is dominated by Fowler-Nordheim electron tunneling through the thin gate oxide, from the n^+-type gate to the p-type channel when the applied gate voltage is negative [8]; and (ii) the energy distribution of the energetic primary electrons that generate the secondary electron-hole pairs in the p-type channel is nearly monoenergetic because the combined energy dependence of the Fowler-Nordheim rate and the band-edge density of states results in a very sharp energy filter,[12] and because collision broadening during post-tunneling acceleration in the oxide is negligible due to the very thin SiO_2 compared with the electron mean free path in the oxide which maintains nearly ballistic transport.[13] Thus the hole current flowing through the p-type channel and measured by the sum of the drain and source

terminal currents is the gate current reduced by the pair generation efficiency, $I_D+I_S=\eta I_G$, and the substrate current is $I_X=(1+\eta)I_G$. The experimental efficiency η is plotted in Fig. 5(a) as a function of the electron energy E_1. It is replotted as $\eta^{2/7}$ in Fig. 5(b) as suggested by the analytical theory (21), following the classic eyeball (straight-line) threshold analysis method. Indeed, a straight line is observed with an intercept near the energy gap, 1.2 eV, as predicted by the analytical solutions given in (9), rather than the extrapolation made by previous investigators of ~1.8 eV ($3E_G/2$) which would be deduced from the η-E_1 plot in Fig. 5(a). This discovery is actually also contained in the decade-old data of Ref. 14 where η was plotted vs E_1 [triangles in Fig. 5(a)] and a threshold of 1.8 eV was suggested. These data are replotted as $\eta^{2/7}$ vs E_1 in Fig. 5(b) which gives slope and threshold nearly identical to our results.

Experiments were also carried out to measure the positive-oxide-charge generation efficiency and threshold by tunnel-electron-generated hot holes. Both MOS capacitor and transistor structures with n^+-type gate or p^+-type gate and n-type substrate or p-type substrate were employed. The experimental data of the impact and Auger hot-hole generation pathways[8] are plotted in Fig. 6(a) as η vs E_1. They are then replotted in Fig. 6(b), in which an appropriate theory-based function of η is plotted vs E_1 in order to give a straight line. For impact generation of the 4.25 eV hot holes, the function is $\eta^{2/7}$. For Auger generation, it is $(\eta N_V^2/P_S^2)^2$ where N_V is the effective density of states of the silicon valence-band edge and P_S, or P_2 in (23), is the volume density of holes at the SiO_2/Si interface of the p-type surface channel. P_S is computed from standard MOS electrostatic theory and may be approximated by V_{oxide} as an initial estimate. It is evident that the results of the data plotted in Fig. 6(b) are consistent with the simple analytical theories (22) and (23). The extrapolated thresholds of $E_{1\text{-min}}$, 3.1 eV for Auger and ~5 eV for impact, are consistent, respectively, with the theoretical values of 3.128 eV given by (19) and 5.372 eV given by (18).

It is noted from Fig. 6(a) that in the lower-energy range the impact generation efficiency of 4.25 eV hot holes is

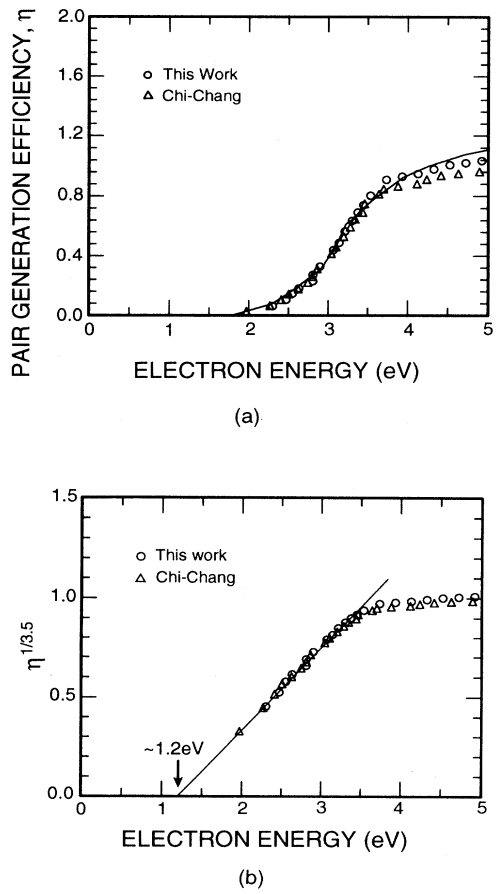

FIG. 5. Experimental total electron-hole pair generation efficiency by an energetic electron in an n^+-type gate/p-type MOS transistor. (a) η and (b) $\eta^{2/7}$. The oxide thicknesses are 3.5 and 5.5 nm for data below and above 3.5 V, respectively.

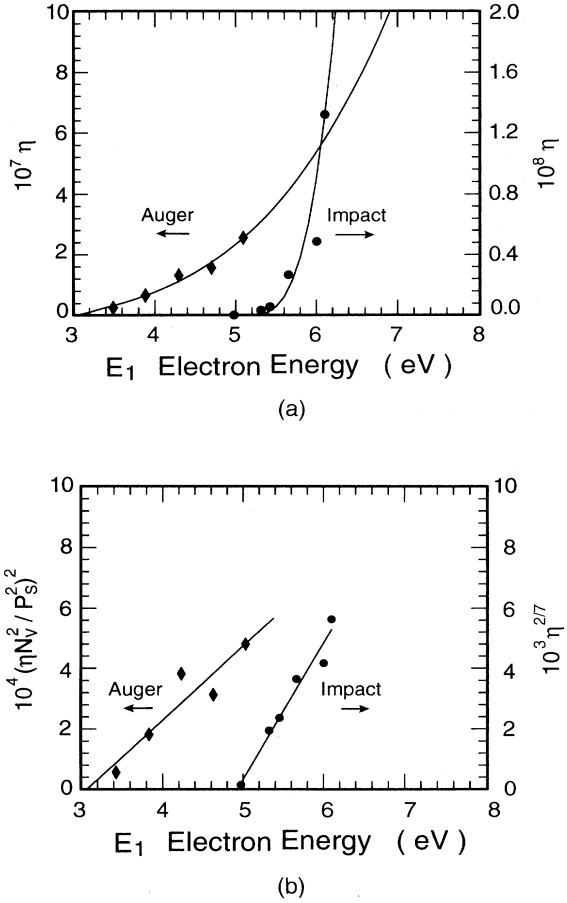

FIG. 6. Experimental positive-oxide-charge buildup efficiency during Fowler-Nordheim electron tunneling through the oxide via impact and Auger generation of hot holes. (a) η and (b) reduced η, $(\eta N_V^2/P_S^2)^2$ for Auger and $\eta^{2/7}$ for impact.

substantially smaller than the Auger generation efficiency. This is due to the smaller number of particles participating in the impact transition than in the Auger transition. This is depicted by the factor $\frac{16}{105}$ in (22) for the impact hot-hole generation rate. The impact rate increases sharply and dominates at higher energies due to the rapidly rising density-of-states factor from the multiple particles during impact, $(E_1-E_G-\phi_h)^{7/2}$ in (22). The larger Auger rate than impact rate for hot-hole generation is consistent with two other experimental observations: (i) a very small negative pair generation efficiency ($\eta<0$), indicating a hole current flowing into the drain-and-source terminals or the p-type channel, occasionally observed to rise above large background noise near the threshold E_G of Fig. 5, which is consistent with Auger recombination of the tunnel-injected electron with a hole, reducing the hole concentration and causing a reversal of hole current flow, and (ii) the leveling off of η above 3.5 eV in Fig. 5, suggesting a steady-state balance between Auger recombination and impact generation of the holes that are measured by the hole current flowing in the p-type channel.

V. SUMMARY

In summary, we have shown that intervalley electron transition in a silicon semiconductor is the fundamental reason that makes threshold energies and energy dependences of the generation rate of secondary electron-hole pairs and hot holes by nearly monoenergetic primary electrons, computed from energy-momentum conservation, nearly equal to the values computed from energy conservation alone. The experiment-theory agreements of the threshold energy and efficiency versus energy suggest that the random imperfections and phonon scattering are not significant to cause momentum randomization in a thin silicon dioxide film and interfacial layers with a silicon substrate and a covering polycrystalline silicon film.

ACKNOWLEDGMENTS

This work is partially supported by the Semiconductor Research Corporation and by Intel Foundation (Yi Lu).

[1]P. A. Wolff, Phys. Rev. **95**, 1415 (1954).
[2]W. Shockley, Solid-State Electron. **2**, 35 (1961).
[3]G. A. Baraff, Phys. Rev. **128**, 2507 (1962).
[4]E. O. Kane, J. Phys. Soc. Jpn. Suppl. **21**, 37 (1966).
[5]E. O. Kane, Phys. Rev. **159**, 624 (1967).
[6]W. E. Drummond and J. L. Moll, J. Appl. Phys. **42**, 5556 (1971).
[7]D. B. Laks, G. F. Neumark, and S. T. Pantelides, Phys. Rev. B **42**, 5176 (1990).
[8]Y. Lu and Chih-Tang Sah, J. Appl. Phys. **76**, 4724 (1994).
[9]R. C. Alig, S. Bloom, and C. W. Struck, Phys. Rev. B **42**, 565 (1980).
[10]W. Shockley, Bell Syst. Tech. J. **28**, 435 (1949).
[11]W. Z. Weinberg, W. C. Johnson, and M. A. Lampert, Appl. Phys. Lett. **25**, 42 (1974).
[12]Russell D. Young, Phys. Rev. **113**, 110 (1959).
[13]M. V. Fischetti, D. J. DiMaria, L. Dori, J. Batey, E. Tierney, and J. Stasiak, Phys. Rev. B **35**, 4404 (1987).
[14]Chi Chang, C. Hu, and R. Brodersen, J. Appl. Phys. **57**, 302 (1985).