Chapter 2 Impact Ionization and Hot-Electron Injection in MOSFETs Derived Consistently from Boltzman Transport

Floating-gate MOS devices [48] rely on hot-electron injection and tunneling phenomena to control the amount of charge stored on the floating gate. In this traditional mode of operation, employed in EEPROM and Flash ROM memories [49, 50], the device is used simply as a nonvolatile, charge-storage node. Another application of this device is in analog silicon synapses, where the device stores analog nonvolatile charge while performing a single-quadrant multiplication on the applied voltages and adapting the floating-gate charge as a function of the applied voltages [35, 36]. Optimization of silicon synapses for particular circuit and system applications, and of floating gate structures in general, necessitates a thorough understanding of the physical mechanisms involved in device operation. To explain hot-electron injection in SiO_2, one must understand how electrons obtain sufficiently high energies that they overcome the silicon–silicon-dioxide barrier.

Shockley introduced the concept of a *lucky-electron* in modeling the impact-ionization process in silicon and germanium [51]. According to the lucky-electron model, the probability (P_{nc}) that an electron does not experience collisions, and thus does not lose energy, is proportional to an exponential function of the ratio of distance that the electron travels (x) and its mean free path (λ):

$$P_{nc} \propto e^{-x/\lambda}. \tag{2.1}$$

Scientists employed this lucky-electron model to explain hot-electron injection in MOS capacitors [52]: Shockley's simple theory was in good agreement with the data over a wide range of dopings and temperatures. The same model was also employed to explain hot-electron injection and impact ionization currents in MOS transistors biased above threshold [53, 54]. Tam and colleagues observed a correlation between the hot-electron injection currents and the impact-ionization currents in MOS transistors [53], but provided no physical connection between these two phenomena. Although the lucky-electron theory provides simple explanations, it does not give insights on the actual physical processes, and it cannot be related to an underlying hot-electron distribution function.

A second modeling approach assumes an a priori hot-electron distribution function [52, 54]. Usually, the distribution function (f) as a function of energy (E) and an effective temperature (T_c) is

$$f(E) = e^{-E/kT_c}. \tag{2.2}$$

However, the results of Monte Carlo simulations [55, 56] suggest that, at the energies at which impact ionization and hot-electron injection occur, a constant electron temperature is not a good approximation. Baraff [57, 58] proposed an analytic model for impact ionization deriving from a spatially-uniform, Boltzmann transport equation. His model fits impact-ionization data from p-n junctions, and in some regimes, agrees with the lucky-electron theory.

Unfortunately, we need a model derived from a spatially-varying Boltzmann transport equation, because the distribution function in the drain-to-channel depletion region in a MOS transistor is far from spatially uniform for significant impact-ionization and hot-electron-injection currents. When the impact-ionization and hot-electron-injection currents in a MOS transistor are significant, the distribution function in the drain-to-channel depletion region is far from spatially uniform; therefore necessitating a model derived from a spatially-varying Boltzmann transport equation. As a result of the added complexity in the Boltzmann transport equation little progress has been made towards analytically relating the description of the electron distribution functions to the measured data of hot-electron-injection currents and impact-ionization

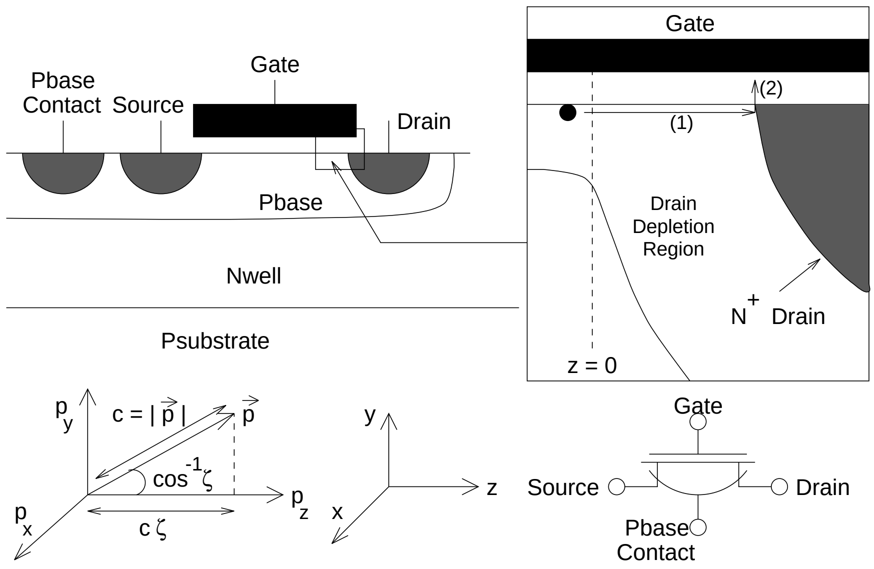

Figure 2.1: Cross section of the MOSFET device we used to measure the hot-electron effects. It uses a moderately doped ($1\text{x}10^{17}\text{cm}^{-3}$) substrate to achieve a high threshold voltage which allows hot-electron injection for bias current levels in subthreshold. The higher doping is consistent with a 0.2μm channel length CMOS process; therefore the effects are directly applicable to modern processes although the device was fabricated in a 2μm process. The $n$well isolates the moderately doped substrate region from the surrounding substrate, and allows measurement of substrate current. Holes resulting from impact ionization are measured at the $p$base contact. The hot-electron injection process is identical for the FET with or without the isolating $n$well. Inset: the electron is accelerated through the drain depletion region (1), and when it gains energy greater than the Si–SiO_2 barrier, the electron is injected over the Si–SiO_2 barrier to the floating-gate (2). Lower Left: Circuit symbol for the nFET with the $p$base implant.

currents. For a level of description beginning with electron distribution functions, the only models that agree with experimental data of both hot-electron injection and electron tunneling are numerical Monte Carlo models [56].

In this chapter, I develop a quantitative model of the impact ionization and hot-electron processes, that is derived consistently from a single spatially varying hot-electron distribution function. The hot-electron transport in the drain-to-channel depletion region is modeled using the spatially varying Boltzmann transport equation. I also show an analytical relationship between the impact ionization in the drain-to-channel depletion region of an MOS transistor and hot-electron injection on the gate.

I verified the model experimentally using the MOS transistor structure with a moderately high substrate doping (1×10^{17} cm^{-3}) shown in Fig. 2.1. I simultaneously measured substrate and gate currents from the device biased in subthreshold. Under subthreshold biasing conditions, the surface potential is fixed; therefore, the high field effects are confined to the drain-to-channel depletion region where carrier concentrations are at low concentrations. I shall show analytically that, at the energies where impact ionization and hot-electron injection occur, a constant electron temperature is not a valid assumption. The analytic model of impact-ionization and hot-electron injection currents agrees well with experimental data.

This chapter is structured as follows. Section 2.1 formulates a Boltzmann Transport Equation and makes reasonable device approximations to model the transport in the drain-to-channel depletion region. Section 2.2 solves analytically for the distribution function in a self-consistent manner. I solve for the average trajectory taken by a hot electron as a function of position through the depletion region. This determines the average electron energy and direction as a function of position. In this coordinate system, phonon collisions act to diffuse the distribution function as a function of position, and impact ionization collisions remove electrons from participating in hot-electron injection. A new collision operator for impact ionization is proposed, which is necessary to simultaneously fit the hot-electron injection and impact ionization data. Section 2.3 uses this distribution function to calculate the probabilities of impact ionization and hot-electron injection as a function of channel current, drain voltage, and floating gate voltage. I compare my analytical model to measurements of hot-electron injection and electron impact ionization; the model simultaneously fits both sets of experimental data. A discussion follows in Section 2.4.

2.1 Electron Transport in the Drain to Channel Depletion Region

The Boltzmann Transport Equation determines the spati-temporal evolution of an electron distribution function under low to moderately high fields. Even though it can be argued that its use in this work is questionable because of the high electric fields involved, I believe that it is a good starting point. Monte Carlo simulations of impact ionization in silicon [56] indicate that the collision times in the high energy regime of concern here, are at most a factor of two larger than 10fs. In other words, the time between collisions is on the same order of magnitude as the time of a collision. The Boltzmann transport equation is still approximately valid when the average time between collisions is on the order of 10fs; therefore, its use is justifiable. The alternate and more rigorous approach is to model this problem using quantum transport techniques [59]; this will be the natural followup to this work. In addition, Section 2.4 will discuss the effects of zeroth-order quantum effects, like collision broadening.

2.1.1 Boltzmann Transport Under an Applied Electric Field

The density of an electron gas in 6-dimensional space is defined by the distribution function, $f(x, y, z, p_x, p_y, p_z)$, where p_x, p_y, and p_z are the components of momentum in x, y, and z coordinates. Integrating $f()$ over all six coordinates equals a constant, which without loss of generality is is set equal to one. The evolution of this distribution function under the influence of applied fields is described by the Boltzmann transport equation [60]

$$\frac{\partial f}{\partial t} + \vec{v} \cdot \nabla f + q\vec{\mathcal{E}} \cdot \nabla_{\vec{p}} f = S(f), \tag{2.3}$$

where $\vec{v}$ is the average velocity vector, $\vec{\mathcal{E}}$ is the applied electric field, and the function $S(f)$ is defined as the difference of the rates of electrons being scattered into and out of this six dimensional space. The collision operator, $S(f)$, if formulated as [60]

$$S(f) = \left(\frac{\partial f}{\partial t}\right)_{scatterin} - \left(\frac{\partial f}{\partial t}\right)_{scatterout}. \tag{2.4}$$

2.1.2 Simplifications for Hot-Electron Transport in the MOSFET Drain-to-Channel Region

We can simplify the general Boltzmann transport formulation by making the following simplified approximations to the carrier transport in a subthreshold MOSFET. This work is concerned with hot electron injection in applications where the time scales involved are in the μs range or slower. Since scattering times are on the order of ps and fs, we can safely assume that the electron distribution function has sufficient time to reach its steady state. The electron gas is due to several electrons in time, although at any given point in time there will be very few electrons in the drain-to-channel depletion region.

I will assume that the 2D confinement in the MOS inversion layer only slightly perturbs the density of states, and to a first order does not affect our solutions. The electric field from the charge on the gate confines the electrons to a region near the substrate surface. A triangle barrier of height 2kT and a width of 10nm is just on the edge of the regime where the continuum approximation for the energy levels is still valid. Monte Carlo calculations in [61] indicate this may not always be the case.

Under normal operating conditions there is no electric field in the x or width dimension of the transistor, and therefore the electric field vector has a component only in the z direction, which I write as

$$\vec{\mathcal{E}} = (\mathcal{E}_x, \mathcal{E}_y, \mathcal{E}_z) = (0, 0, \mathcal{E}) \tag{2.5}$$

The electrostatic potentials in the y-direction result in small electron oscillations near the Si-SiO_2 interface; therefore, the electron's y-velocity distribution is uniform over a given region in x and z, and the electron's maximum y-displacement is negligible. Since collisions restore electron distributions to equilibrium, the electron distribution in the y direction will remain uniform. I will further discuss the electron confinement and the implications of the y-direction electron distribution in Section 2.4. Under the above simplifying assumptions, (2.3) simplifies to

$$\left(v_z \frac{\partial}{\partial z} + q\mathcal{E} \frac{\partial}{\partial p_z} \right) f = S(f) \tag{2.6}$$

where ∂p_y and ∂y terms dropped out due to the confinement of the electrons in the y direction, and ∂x term dropped out due to the uniform distribution function along the width of the device.

Following Baraff's approach [58], I choose a coordinate system that orthogonalizes the total momentum and the average direction of that momentum. By transforming the momentum variables into polar coordinates (see Fig. 2.1), we get

$$\zeta \frac{c}{m^*(c)} \frac{\partial f}{\partial z} + \zeta q\mathcal{E} \frac{\partial f}{\partial c} + q\mathcal{E} \frac{1-\zeta^2}{c} \frac{\partial f}{\partial \zeta} = S(f) \tag{2.7}$$

where c is the magnitude of the average momentum vector, and ζ is the cosine of the angle of $\vec{p}$ and the z axis. ζ describes the degree of anisotropy of the distribution function. I define E, the electron energy, as

$$E(c) = \frac{c^2}{2m^*(c)} \tag{2.8}$$

where $m^*(c)$ is the effective mass of the electron. Substituting (2.8) into (2.7) we can get the following expression for the simplified BTE in this energy coordinate system as

$$\frac{\partial f}{\partial z} + q\mathcal{E} \frac{\partial f}{\partial E} + q\mathcal{E} \frac{1-\zeta^2}{\zeta E} \frac{\partial f}{\partial \zeta} = \frac{m^*(c)}{\zeta c} S(f). \tag{2.9}$$

The collision operators discussed in the next section are given in terms of $\frac{m^*(c)}{c} S(f)$, which effectively allows the formulation of a model that does not address explicitly the details of band structure. In the two subsections, we will define the collision operators in terms of $\frac{m^*(c)}{c} S(f)$, which will allow us to solve for the distribution function without knowing the detailed silicon bandstructure.[1]

[1] Canceling out the effects of the bandstructure may limit the predictive power of this model. This insight by Karl Hess is appreciated.

2.1.3 Collision Operators

The collision operator, (2.4), was introduced to describe electron scattering. The main scattering mechanisms in silicon are:

- **Phonons** are quanta of lattice vibrations [60], and are divided in two groups. *Acoustical phonons* are low energy phonons whose energy is proportional to the magnitude of their momentum vectors. *Optical phonons* have energies much larger than acoustical phonons, and these phonons have nearly constant energy as a function of its momentum. This energy is designated by E_R.

- **Ion-impurity** scattering is an elastic scattering mechanism, where a collision does not result in a change in energy of the electron, but only a change in momentum direction. At high energies, the deflections are either small in angle or lead to impact ionization, therefore we will neglect this mechanism except to explicitly model the impact ionization collisions.

- **Impact ionization** collisions occur when a high-energy carrier (electron or hole) can give up sufficient energy to liberate another electron into the conduction band, which also creates a hole in the valence band. This mechanism has been considered in several references [51, 56, 57, 62, 63].

- **Electron–electron** scattering occurs only in devices with a high density of electrons; therefore, this mechanism is negligible in the depletion regions and for subthreshold current levels because of the low density of electrons [64] .

Since only optical phonon collisions and impact ionization processes will sufficient affect the electron's energy, my model of the collision process will include these two mechanisms.

We shall proceed by first deriving a model for optical phonon collisions. Figures 2.2 and 2.3 graphically shows the effect of a phonon collision. I model optical phonon

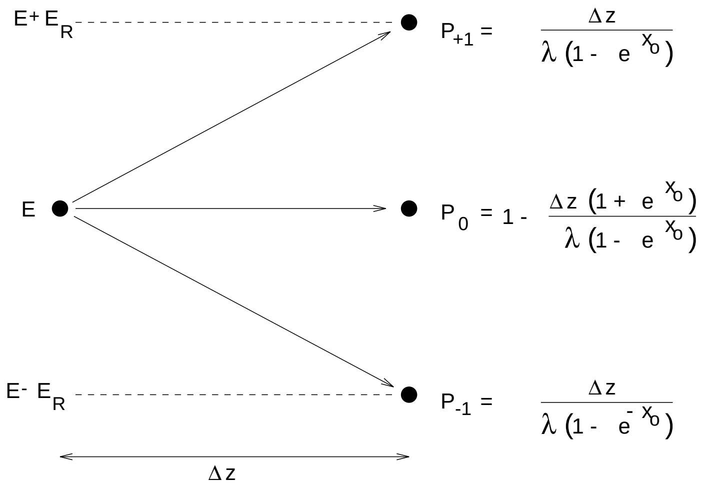

Figure 2.2: Graphical representations of the phonon collision operator. An electron can either gain or lose energy due to a phonon. The transition probabilities of gaining or losing energy due to a phonon are different. The operator is based on a mean free path of a collision; therefore the longer an electron goes in a region, the more likely it will gain or lose energy due to a phonon collision. Gaining or losing a phonon will alter the electron's momentum as well as its energy.

scattering as

$$S(f)_{op}\left[\lambda(e^{\frac{E_R}{kT}}-1)\frac{m^*(E)}{c(E)}\right] = \begin{array}{l}\left(1+\frac{E_R}{E}\right)^{1/2}\left(e^{\frac{E_R}{kT}}f(E+E_R)-f(E)\right)+ \\ \left(1-\frac{E_R}{E}\right)^{1/2}\left(f(E-E_R)-e^{\frac{E_R}{kT}}f(E)\right),\end{array} \tag{2.10}$$

which was derived elsewhere [60]. Assuming no other forces on the electron, the phonons change the distribution function after the electron travels a small distance (Δz) away. In Fig. 2.2, I graphically show the equation:

$$f(z,E) = f(z+\Delta z,E)\left(1-\frac{\Delta z}{\lambda}\frac{1+e^{x_o}}{1-e^{x_o}}\right)+\frac{\Delta z}{\lambda}\frac{e^{x_o}}{1-e^{x_o}}f(z+\Delta z,E-E_R)$$
$$+\frac{\Delta z}{\lambda}\frac{1}{1-e^{x_o}}f(z+\Delta z,E+E_R), \tag{2.11}$$

where $x_0 = \frac{E_R}{kT}$; the different terms for absorbing and emitting a phonon arise from the phonon density of states. The interaction with a phonon increases or decreases the electron energy by E_R, which I set to the known silicon value of 63meV [51]. On

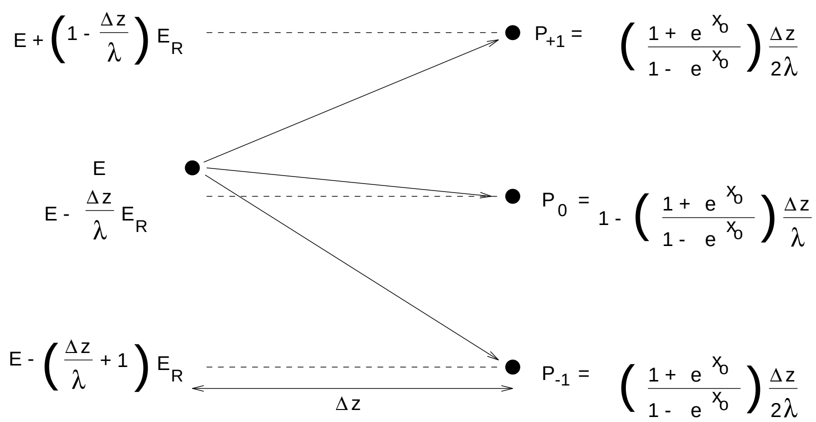

Figure 2.3: A second way to graphically represent the phonon collision operator is considering an electron either gaining or losing energy due to phonons with equal strength along a particular coordinate system.

the average, an electron makes a collision with a phonon in a distance λ, or the mean free path of the electron; I use $\lambda = 6.5$nm because it agrees with my experimental data and with previous results [51, 57].

To gain more insight on the physical mechanisms involved, I simplify (2.10) by performing a Taylor series expansion for $E >> E_R$

$$S_{op}(f)\left[\frac{m^*(E)}{c(E)}\right] \approx F(T)\frac{E_R^2}{2\lambda}\frac{\partial^2 f}{\partial E^2} + \frac{E_R}{\lambda}\frac{\partial f}{\partial E}, \tag{2.12}$$

where $F(T)$ is

$$F(T) = \frac{e^{\frac{E_R}{kT}} + 1}{e^{\frac{E_R}{kT}} - 1}, \tag{2.13}$$

which is nearly equal to one at room temperature ($T = 300K$). This expansion is valid to fourth order in derivatives in energy, and is very nearly equal to the original expression for energies greater than a few E_R. In Fig. 2.3, I show this transformation graphically, by modeling the asymmetric phonon scattering event as an average energy loss (corresponding to $\frac{\partial f}{\partial E}$) and a symmetric scattering event around this resulting energy (corresponding to $\frac{\partial^2 f}{\partial E^2}$). A similar expansion and simplification has been done for polar optical phonons [65].

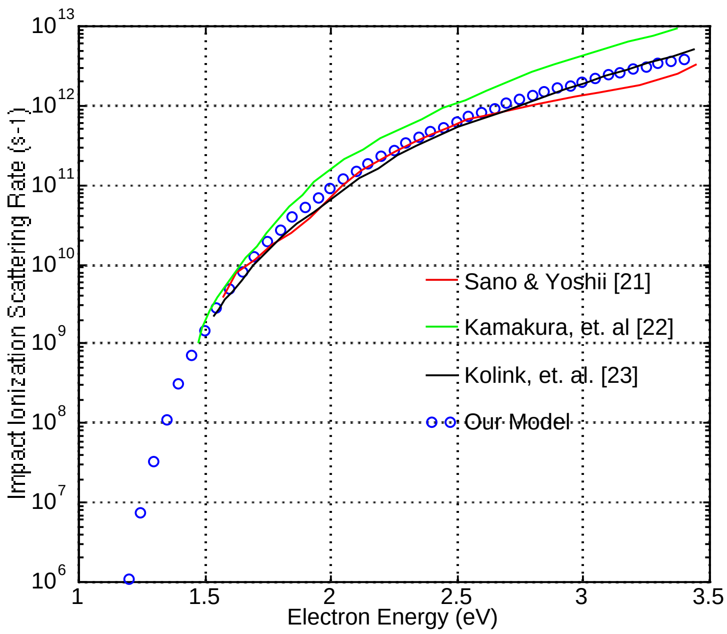

Figure 2.4: Plot of previous calculations of impact-ionization rate versus electron energy in silicon, and our derived impact-ionization rate from our measured impact-ionization and hot-electron injection data. I have assumed a constant velocity, since our model measures the impact-ionization mean-free length. Our measured data is directly related to $L(E)$ and not τ_{ion}.

Next, I will derive the collision operator for impact ionization. The collision operator for impact ionization has the general form of

$$S(f)_{ion} = -\frac{f}{\tau_{ion}(E)} = -\frac{c(E)}{m^*(E)}\frac{1}{L(E)}f, \tag{2.14}$$

where τ_{ion} is the mean free time for an impact ionization collision, and L(E) is the mean free path, which is a function of the electron energy. The first impact-ionization models were of the form [51, 57, 62]

$$\frac{1}{\tau_{ion}} \propto H(E - E_{th}), \tag{2.15}$$

where $H()$ is the unit step function or the "Heavyside Operator", and E_{th} is the threshold energy for impact ionization. Keldish [63] proposed the impact-ionization model:

$$\frac{1}{\tau_{ion}} \propto (E - E_{th})^2 H(E - E_{th}). \tag{2.16}$$

Wolfe [62] used E_{th} equal to 2.3eV; other researchers [51, 57, 58] used E_{th} equal to the bandgap of silicon ($E_{th} = 1.1\text{eV}$) since the bandgap is the minimum amount of energy required for an impact-ionization event. Monte Carlo simulations [56] of impact ionization and hot-electron injection using (2.16) showed that a small ionization rate and $E_{th} = 1.1\text{eV}$ best fit the data; the resulting ionization rate was nearly constant only for energies greater than 2.3eV.

Recently, several researchers put considerable effort in calculating the impact-ionization collision times directly from band-structure calculations, and then using these results with some success in Monte Carlo simulations of hot-electron phenomena [66, 67, 68]. Figure 2.4 shows several numerically computed impact-ionization rates. Based on these calculations and our experimental measurements of impact ionization and hot-electron injection, we propose the following energy dependence for the impact-ionization mean free length:

$$L(E) = l_{ion} \exp\left(\sqrt{\frac{119rmeV}{E - .95\text{eV}}}\right). \tag{2.17}$$

Figure 2.4 shows our functional form with these three numerically calculated models, where l_{ion} is 1.81×10^{-2} nm. I have assumed a constant velocity of 8.1×10^6 cm/s in converting from $L(E)$ to τ_{ion}, since our measured data is directly related to $L(E)$ and not τ_{ion}. This functional form is a curve fit to experimental data of $L(E)$ derived from my experimental measurements of hot-electron-injection and impact-ionization currents in Section 2.3.

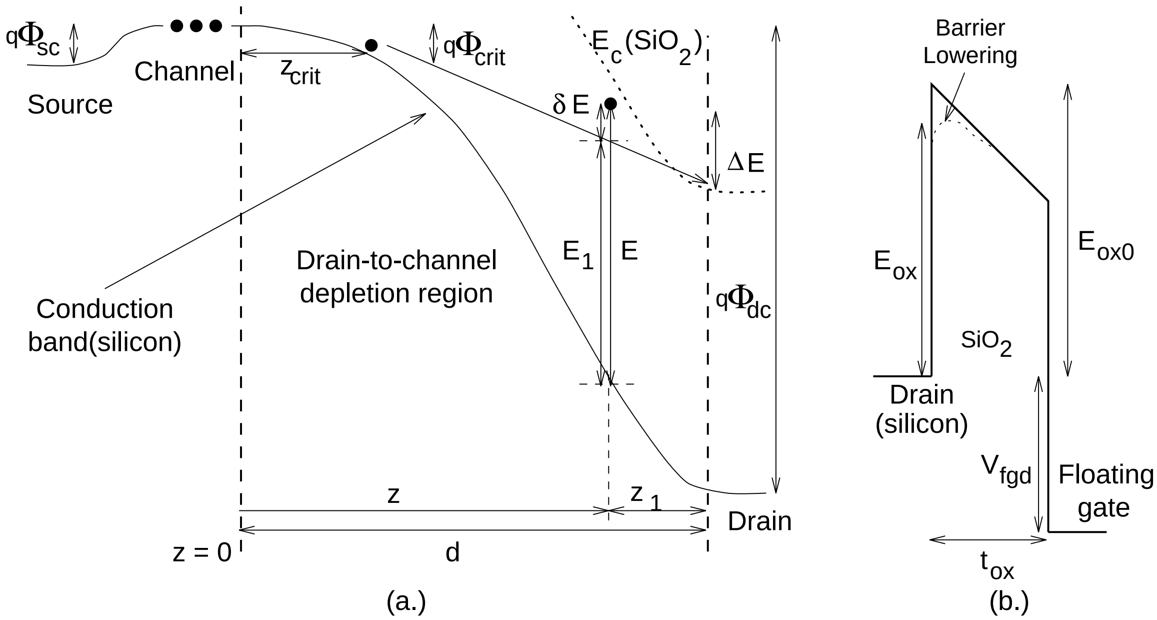

Figure 2.5: Band diagram illustrating hot-electron injection in a MOSFET biased in subthreshold. The appropriate variables in the Boltzmann transport equation and its variable transformations are shown on the graphs. (a) Band diagram along the surface of the Si-SiO_2 barrier. This region is the lowest local potential in either material; therefore the electrons are most likely to travel along this path. This region corresponds to path (1) in the inset in Fig. 2.1. (b) Band diagram at the drain edge from the substrate to the gate. This region corresponds to path (2) in the inset in Fig. 2.1.

2.2 Solving the Electron Distribution Function

This section solves for $f(z, E, \zeta)$ from the Boltzmann transport equation. Under the simplifying assumptions of the previous section, the resulting Boltzmann transport equation for $f(z, E, \zeta)$ can be written as

$$\frac{\partial f}{\partial z} + \left(q\mathcal{E} - \frac{E_R}{\lambda\zeta} \right) \frac{\partial f}{\partial E} + q\mathcal{E}\frac{1-\zeta^2}{\zeta E}\frac{\partial f}{\partial \zeta} = F(T)\frac{E_R^2}{2\lambda\zeta}\frac{\partial^2 f}{\partial E^2} - \frac{f}{L(E)} \tag{2.18}$$

with the parameters defined in the previous section. I begin by seeking a solution for the energy and momentum angle of an average electron as a function of position. I define the energy of an average electron as $E = E_1(z)$ and we define the average momentum angle of the average electron as $\zeta = \zeta_1(z)$. I then transform (2.18) around this average electron energy path. When impact ionization does not occur, we will see that the distribution function satisfies a diffusion equation along this path. When impact ionization occurs, we will see that the solution is a product of the solution to a diffusion equation and a loss term due to impact ionization.

Figure 2.5 shows the band diagram of a subthreshold MOSFET, with definitions for the variables used in this section. I define the drain-to-channel potential, Φ_{dc}, as $V_d - \Psi$, where Ψ is the channel potential and V_d is the voltage at the drain. I define d as the width of the drain-to-channel depletion region. Let δE be the difference of the electron energy from $E_1(z)$:

$$\delta E = E(z) - E_1(z). \tag{2.19}$$

Let $\delta\zeta$ be the difference in the electron momentum angle from $\zeta_1(z)$:

$$\delta\zeta = \zeta - \zeta_1(z). \tag{2.20}$$

2.2.1 Average Electron Energy and Momentum Direction

This subsection determines the average electron energy and the average direction of the momentum as a function of z. To do so, we must solve (2.18) in the new energy and momentum frame of reference of $(z, \delta E, \delta\zeta)$ as defined by (2.19) and (2.20). We can perform this transformation through a Jacobian matrix transformation of (2.18) into this new coordinate system. The transformation is

$$\begin{aligned}
\begin{bmatrix} \frac{\partial f}{\partial E} \\ \frac{\partial f}{\partial \zeta} \\ \frac{\partial f}{\partial z} \end{bmatrix} &= \begin{bmatrix} \frac{\partial \delta E}{\partial E} & \frac{\partial \delta\zeta}{\partial E} & \frac{\partial z}{\partial E} \\ \frac{\partial \delta E}{\partial \zeta} & \frac{\partial \delta\zeta}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \\ \frac{\partial \delta E}{\partial z} & \frac{\partial \delta\zeta}{\partial z} & \frac{\partial z}{\partial z} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \delta E} \\ \frac{\partial f}{\partial \delta\zeta} \\ \frac{\partial f}{\partial z} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{dE_1(z)}{dz} & -\frac{d\zeta_1(z)}{dz} & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \delta E} \\ \frac{\partial f}{\partial \delta\zeta} \\ \frac{\partial f}{\partial z} \end{bmatrix}.
\end{aligned} \tag{2.21}$$

After the coordinate transformation, the operators on the right-hand side of (2.18) become:

$$\left(q\mathcal{E}(z) - \frac{E_R}{\zeta\lambda}\right)\frac{\partial f}{\partial E} \rightarrow \left(q\mathcal{E}(z) - \frac{E_R}{\lambda} - \frac{dE_1(z)}{dz}\right)\frac{\partial f}{\partial \delta E}, \tag{2.22}$$

$$q\mathcal{E}\frac{1-\zeta^2}{\zeta E}\frac{\partial f}{\partial \zeta} \rightarrow \left(q\mathcal{E}\frac{1-\zeta^2}{\zeta E} - \frac{d\zeta_1(z)}{dz}\right)\frac{\partial f}{\partial \delta\zeta}. \tag{2.23}$$

I will choose $E_1(z)$, $\zeta_1(z)$ such that the right-hand side of these two terms equal zero. With the appropriate initial conditions, choosing the functions $E_1(z)$, $\zeta_1(z)$ in this way yields reasonable approximations of the average electron energy rigorously defined as the integral of the distribution function over all energies and momentum angles. We can eliminate the dependence on the δE partial derivative in (2.22) by setting $E_1(z)$ as

$$\frac{dE_1(z)}{dz} = q\mathcal{E}(z) - \frac{E_R}{\lambda\zeta}, \tag{2.24}$$

and we can eliminate dependence on the $\delta\zeta$ partial derivative in (2.23) by setting $\zeta_1(z)$ as

$$\frac{d\zeta_1(z)}{dz} = q\mathcal{E}(z)\frac{1-\zeta^2}{\zeta E}. \tag{2.25}$$

The approach followed so far in solving the transport equation is known as "the method of characteristics" and is described in [69, 70]; it is also similar to the numerical method presented in [71]. The functions $E_1(z)$, $\zeta_1(z)$ are called the *characteristics* since they are the flow lines for the hyperbolic P.D.E. operator on the left-hand side.

This model assumes a boundary condition that the electron energy is above the conduction band, that is that the electron has positive kinetic energy. The characteristic functions are only valid in the region inside the boundary conditions. If we assume that the electron has no initial energy, then (2.24) is not valid until its right-hand side is greater than zero since the electron cannot have negative energy in the conduction band. Therefore, I consider this analysis valid when the electrons are accelerated by a sufficient electric field to be 'freed' from the low energies around the conduction band, as seen in Fig. 2.6. I define the breakaway field, $\mathcal{E}_{crit}$ as the minimum electric field at which the electron gains energy at the same rate as it loses

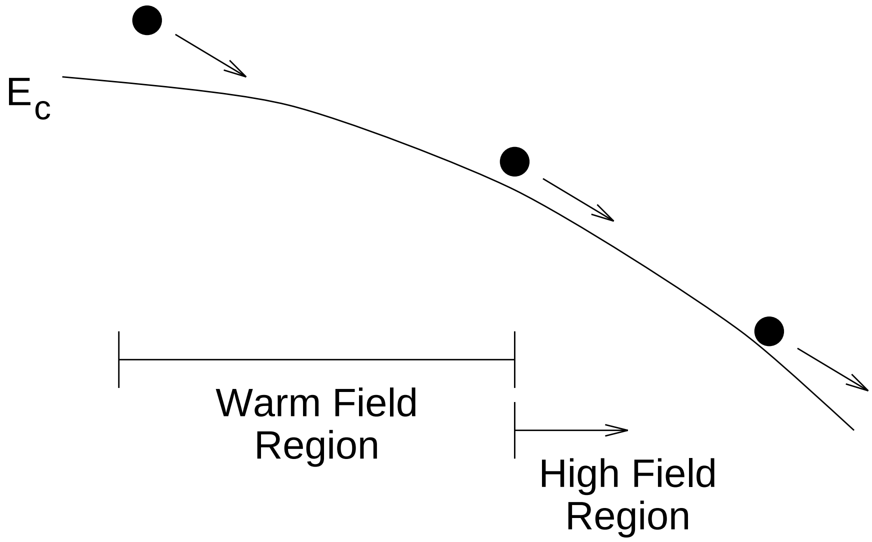

Figure 2.6: Illustration of the electron characteristic paths at different electric fields. In the warm field region, the restoring force due to collisions is larger than the force due to the electric field. In the high field region, the force due to the electric field is larger than the restoring force due to collisions. When an electron reaches an electric field larger than the restoring force due to collisions it can gain significant energies above the conduction band.

energy to phonon collisions. The breakaway field is expressed as $\frac{E_R}{q\lambda}$, which for our parameters is 9.7V/μm. This result is consistent with *streaming* behavior obtained for spatially-uniform electric fields of 10V/μm from Monte Carlo studies [59]; the resulting steady-state average electron energy was 250meV, which is consistent with my approximation. I define the distance from the channel end of the depletion region to this breakaway field as z_{crit}, as illustrated in Fig. 2.5. The average electron analysis will be applied to the region beyond z_{crit}. The physics of the electrons leaving the conduction band is analogous to ballistic transport in the sense that the electrons do not make enough phonon collisions to restore the electrons back to the conduction band [72, 73]. Typically ballistic transport is considered in semiconductor devices where device dimensions are on the order of the mean free path length [72, 73].

For this discussion, I approximate a constant number of electrons per unit length leaving the conduction band throughout the depletion region after z_{crit}. In a more accurate model, the number of electrons per unit length will increase initially due to the larger window of ζ that the electrons can escape the conduction band, and decrease eventually because at some point a majority of electrons will have left the conduction band. Also, before the electrons reach the breakaway field, the electron distribution function will be a function of the transport in the 'warm' field region. The effects of these phenomena will be the subject of another study.

We now return to the solution for the characteristic equations (2.24) and (2.25). The solution for $\zeta_1(z)$ from (2.25) is (derived in Appendix A.1)

$$\left(1-(\zeta_1+\delta\zeta)^2\right)\left(\frac{E_1(z)+\delta E}{E_a}\right)\left(\frac{1-(\zeta_1+\delta\zeta)}{1+(\zeta_1+\delta\zeta)}\right)^{\frac{E_R}{q\mathcal{E}\lambda}}=1, \tag{2.26}$$

where E_a is a parameter dependent upon the initial conditions. When $E_1(z)+\delta E$ becomes large, $\zeta_1(z)+\Delta\zeta$ goes to one; therefore once the electrons reach higher energies they are more likely to be pointed along the field direction. Since the field direction has the smallest required electric field to free an electron from the conduction band, the electrons also start directed along the field direction. Because of these two facts, I can safely approximate ζ as near one for energies above the conduction band. In Section 2.4, I will consider the distribution function away from $\zeta = 1$.

Assuming that $\zeta \approx 1$, $E_1(z)$ can be obtained by integrating (2.24) in the region from z_{crit} to z. The solution for $E_1(z)$ is

$$E_1(z) = qV(z) - qV(z_{crit}) - E_R\frac{z - z_{crit}}{\lambda}, \tag{2.27}$$

as performed in Appendix A. The energy the electron gains after reaching z_{crit} is the difference of the potential from z_{crit} to drain and the number of phonon collisions made in the distance $z - z_{crit}$.

2.2.2 Solution of the Distribution function

This section obtains solutions for the distribution function, first for the simpler case of low energies, and then for all energies. These solutions are obtained by solving the differential equations along the average electron trajectory. I begin with the

Boltzmann Transport Equation for $\zeta \approx 1$ given in terms of E:

$$\frac{\partial f}{\partial z} + \left(q\mathcal{E} - \frac{E_R}{\lambda}\right)\frac{\partial f}{\partial E} = F(T)\frac{E_R^2}{2\lambda}\frac{\partial^2 f}{\partial E^2} - \frac{f}{L(E)}; \tag{2.28}$$

I will make the transformation to the ΔE coordinate system later in this section. I will assume that $1/L(E)$ has a negligible effect on the solution for sufficiently low energies, because $L(E)$ changes rapidly for energies between 1eV and 2eV. For electron energies where $1/L(E)$ is negligible, (2.28) reduces to the following lossless partial differential equation:

$$\frac{\partial f}{\partial z} + \left(q\mathcal{E} - \frac{E_R}{\lambda}\right)\frac{\partial f}{\partial E} = F(T)\frac{E_R^2}{2\lambda}\frac{\partial^2 f}{\partial E^2}. \tag{2.29}$$

For energies where $1/L(E)$ is not negligible, we will factor out the solution of (2.29) from (2.28). I accomplish the factorization by defining $f(z, E)$ in terms of two functions

$$f(z, E) = g(z, E)a(z, E). \tag{2.30}$$

I define $g(z, E)$ as the solution to (2.29); therefore the solutions of the lossless region continue to apply into this region, but are simply attenuated by the function $a(z, E)$ due to impact ionization. Since this procedure is valid for all values of $1/L(E)$, I require that $a(z, E) = 1$ everywhere for the lossless region. The attenuation function $a(z, E)$ can be substituted in (2.28), which results in the following partial differential equation for $a(z, E)$ in terms of z and E (please see detailed derivation in Appendix A.2.1):

$$\frac{\partial a}{\partial z} + \left(q\mathcal{E} - \frac{E_R}{\lambda} - F(T)\frac{E_R^2}{\lambda}\frac{\frac{\partial g}{\partial E}}{g}\right)\frac{\partial a}{\partial E} = F(T)\frac{E_R^2}{2\lambda}\frac{\partial^2 a}{\partial E^2} - L(E)a. \tag{2.31}$$

If electrons are uniformly generated in z, as in the classical MOS capacitor experiments [52], then the Boltzmann transport equation which models this phenomena is (2.31). This measured data followed a lucky-electron model; I also observe a lucky-electron model response from our device in this configuration.

Let us consider the solution of (2.29) by first changing from E to δE coordinates

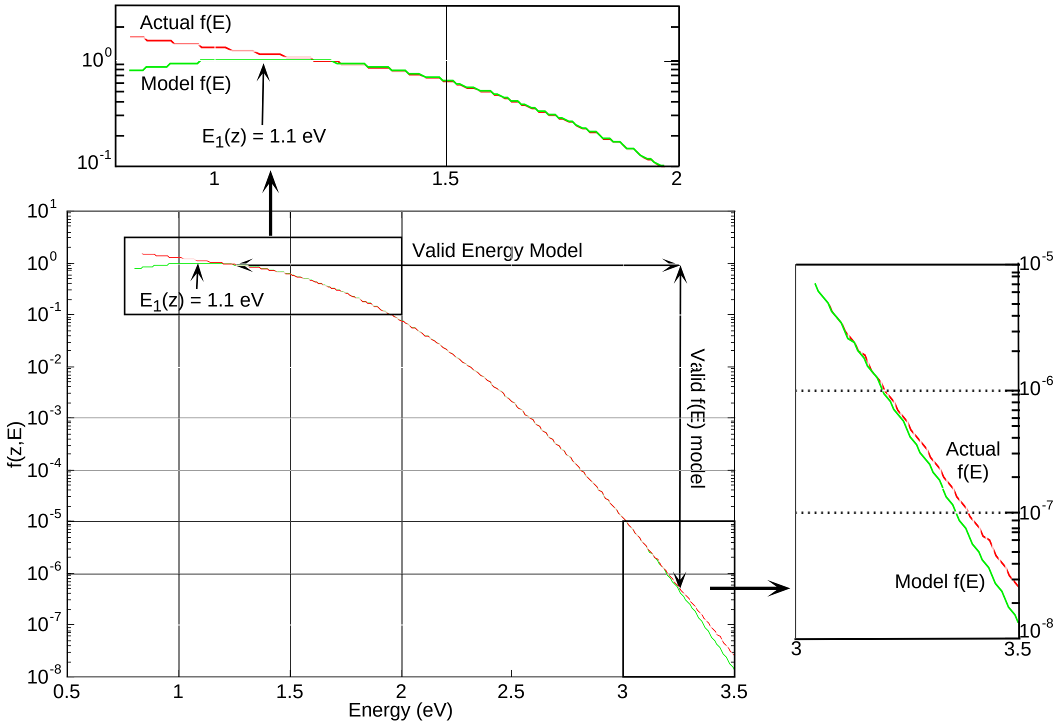

Figure 2.7: Picture of the distribution function for an electron in the drain-to-channel beyond $z = z_{crit}$. This figure compares my Gaussian approximation model versus the solution using the exact boundary condition. For large positive energies, the distribution function does not change as fast as the Gaussian, but rather at a slope around kT_c.

to get

$$\frac{\partial f}{\partial z} = F(T)\frac{E_R^2}{2\lambda}\frac{\partial^2 f}{\partial \delta E^2}, \tag{2.32}$$

which is equivalent to the well-known diffusion equation in space and energy coordinates. To solve (2.32), we need to determine the boundary and initial conditions. The first boundary condition is determined by noting that the solution of (2.32) is bounded in the energy and space coordinates. To obtain additional boundary/initial conditions we assume that the electron leaving at z_{crit} dominates the behavior of hot-electron injection and impact-ionization for a wide range of drain voltages. We can write the solution to (2.32) with this approximation as

$$f(z, \delta E) = \exp\left(-\frac{\lambda}{z - z_{crit}}\left(\frac{\delta E}{2E_R}\right)^2\right), \tag{2.33}$$

which when converted back to E coordinates yields

$$f(z,E) = \exp\left(-\frac{\lambda}{z - z_{crit}}\left(\frac{E - E_1(z)}{2E_R}\right)^2\right). \tag{2.34}$$

The distribution function, f, diffuses from its initial state in energy δE as it evolves in z. Note that the width of the diffusion solution increases as the square root of the position, and therefore the number of phonon collisions.

Figure 2.7 shows the resulting distribution function for the complete model with all three boundary conditions and for the approximate Gaussian model of an impulse at $z = z_{crit}$ for $E = 0$. The Gaussian approximation has two limitations:

1. If the electron distribution function at z_{crit} is approximated by a Boltzmann distribution with an 'effective' temperature, T_c, due to the warm-field transport, then the following equation gives an additional initial condition

$$f(z = z_{crit}, E = \delta E) = e^{-\frac{E}{kT_c}}. \tag{2.35}$$

2. If I model a constant number of electrons per unit length leaving the conduction band, the resulting second boundary condition is

$$f(z, E = 0) = H(z - z_{crit}). \tag{2.36}$$

Including these two effects increases greatly the complexity of an analytic solution.

I now proceed to simplify (2.31) and solve for the attenuation function $a(z,E)$. Since L(E) is extremely small until the electron energy is larger than the bandgap, the energy boundary condition is nearly at the silicon bandgap. One might expect that the second-order derivative in E is small since a constant solution is as diffuse as possible. I show in Appendix A.2.2, that the $F(T)\frac{E_R^2}{\lambda}\frac{\partial^2 g}{\partial E^2}$ term is a second-order perturbation in $\frac{E_R}{q\mathcal{E}(d)l_{ion}}$, and the $\frac{\partial g}{\partial E}/g$ term is a small quantity in the ranges of interest. For this device, the region of significant impact ionization collisions corresponds to the region of the largest electric fields; therefore $q\mathcal{E}\lambda$ will be much larger than E_R.

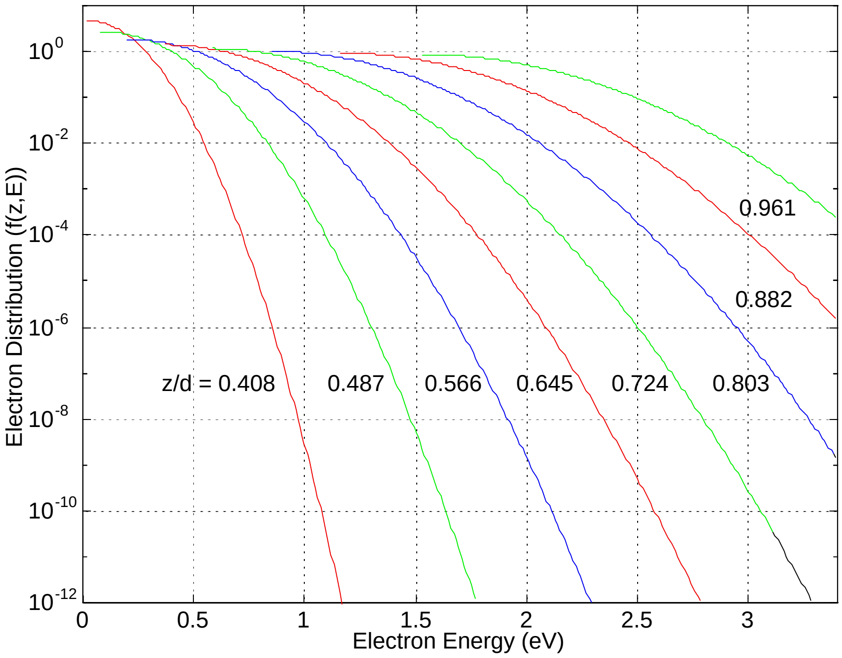

Figure 2.8: Plot of $f(z,E)$ as a function of energy (E) for several locations in the drain-to-channel depletion region for Φ_{dc} = 4.0V, λ =6.5nm, and $N_a = 10^{17}\mathrm{cm}^{-3}$. For these parameters, $z_{crit} = 0.36d$. I normalized $f(z,E)$ over energy; normalizing over position would simply change the global scale factor. The width of the drain-to-channel depletion region is d.

By ignoring these terms, I reduce the problem to the form

$$\frac{\partial a}{\partial z} + \left(q\mathcal{E} - \frac{E_R}{\lambda}\right)\frac{\partial a}{\partial E} = -L(E)a. \tag{2.37}$$

with a constant boundary condition at energies near the bandgap.

To obtain a simple form solution for (2.37), I make the following two approximations. First, if L(E) has no spatial or electric field dependence, then with $a(z, E = 0) = 1$, there is no gradient in z. Second, if $q\mathcal{E}(d) - \frac{E_R}{\lambda}$ is nearly constant at the drain edge, then the distribution function is nearly uniform as a function of position. This approximation is good since most of the impact-ionization occurs near the drain edge, and our measurements are sensitive to the effects at the drain edge. Therefore, I assume that the partial derivative of a with respect to z is zero. With these approximations, we can solve (2.31) by inspection:

$$a(z,E) = \exp\left(-\frac{1}{(q\mathcal{E}(z)\lambda - E_R)}\int_{E=0}^{E}\frac{\lambda}{L(E)}dE\right). \tag{2.38}$$

The solution to $f(z,E)$ can be written as

$$f(z,E) = \exp\left(-\frac{\lambda}{z - z_{crit}}\left(\frac{E - E_1(z)}{2E_R}\right)^2\right) a(z,E). \tag{2.39}$$

Figure 2.8 shows a plot of the above distribution function versus electron energy at several positions in the drain-to-channel depletion.

2.3 Comparing Theory with Experiment

This section compares experimental data for impact ionization and hot electron injection measured in MOSFETs to model calculations that employ the derived distribution functions of the previous sections. More specifically I relate the theoretical distribution functions with the measurable quantities of substrate current and gate currents at the device terminals.

In the next five subsections, I will show measured data of hot-electron injection and impact ionization, and derive device models to relate our data to our model of the electron distribution function. Section 2.3.1 describes the structure and operation of the MOSFET used to measure hot-electron-injection and impact-ionization currents. The depletion region is characterized by its Φ_{dc} and its doping profile; therefore, Section 2.3.2 presents experimental data on the Early voltage to show that the channel doping profile can be approximated by a step junction. Section 2.3.3 derives the distribution function at the drain edge of the depletion region, because the hot-electron injection and impact ionization currents are primarily functions of the distribution at the drain edge. Section 2.3.4 presents the device model of hot-electron-injection current and show the supporting measured data. Section 2.3.5 presents the device model of impact-ionization current and show the supporting measured data.

2.3.1 Device Structure and Bias Condition

Figure 2.1 shows the cross-section of the n-type MOSFET that I use to measure hot-electron injection and impact ionization. I place the MOSFETs in a moderately-doped ($1 \times 10^{17}\mathrm{cm}^{-3}$) substrate to achieve a high threshold voltage. These devices are fabricated through the MOSIS foundry; the moderately doped substrate is formed by the $p$base implant layer, normally employed to form the base of an npn bipolar transistor. The $n$well is used to isolate the $p$base layer from the substrate, to facilitate the measurement of the substrate current due to the electron multiplication. The hot-electron-injection properties of the devices are identical with or without the isolating $n$well.

Since I operate MOSFET transistors in subthreshold, the transport in the drain-to-channel depletion layer is the high-field region of interest. In the subthreshold operating region of a MOSFET, the surface potential in the channel is fixed [74], and thus the high-field phenomena does not depend on the channel current. The electrons in the drain-to-channel region are accelerated to high energies by the high electric fields in the drain-to-channel depletion region. Subthreshold current levels are sufficiently small as to not affect the surrounding electrostatics and the electron density in the channel is low which allows us to ignore the electron-electron scattering effects.

The moderately doped substrate helps the measurements in two ways. First, the high threshold voltage allows hot-electron injection in subthreshold by guaranteeing that the floating-gate to drain voltage always stays positive, and thus electrons reaching the silicon–silicon-dioxide barrier are swept to the floating-gate. Second, the higher substrate doping yields a higher hot-electron-injection and impact-ionization efficiency for a given drain-to-source voltage. The high electric fields responsible for the hot-electron injection and impact ionization in this subthreshold MOSFET structure are consistent with measurements in deep submicron FETs.

I deduce the impact-ionization current by measuring the substrate current, and the hot-electron-injection current from measurements of the gate current. I compute

Φ_{dc} from the sum of the drain-to-source voltage and the source-to-channel potential (Φ_{sc}). In a subthreshold MOSFET, source-to-channel potential is a direct function of the source current (I_s):

$$\Phi_{sc} = kT \ln \left(\frac{I_s}{I_{th}} \right) + 150\text{mV}, \tag{2.40}$$

where I_{th} is the current at threshold, and $\Phi_{sc} = 150$mV at threshold. The measured substrate and gate currents are proportional to the transistor's channel current for identical drain-to-channel and floating-gate-to-drain voltages, as confirmed by the experimental measurements in this section, and in [35].

I define impact-ionization and hot-electron-injection current efficiency as the probability that an electron starting in the channel will be involved either in an impact ionization event or will be injected in the gate, respectively. My model presently assumes that the differences in the initial and final density of states result in small perturbations to the overall solution for this calculation.

2.3.2 Drain-to-Channel Depletion Width Dependence on Φ_{dc}

The electrostatics of the drain-to-channel depletion region near the silicon–silicon-dioxide interface are probed through the channel length modulation phenomenon, which is known as the Early effect. The Early effect manifests itself as a weak dependence of the channel current on drain to source voltage and is a result of the drain-to-channel depletion region width changes with applied Φ_{dc}. I analytically model this behavior as

$$I_{source} = \frac{I_{sat}}{1 - \frac{d(\Phi_{dc})}{L}}. \tag{2.41}$$

Assuming the step-junction model in (2.48), we can model the source current as

$$I_{source} = \frac{I_{sat}}{1 - \frac{\sqrt{\frac{2\epsilon_{si}}{qN_A}}}{L}\sqrt{\Phi_{dc}}}, \tag{2.42}$$

where I_{sat} is the saturation current of the transistors at small drain-channel voltages, ϵ_{si} is the permittivity of silicon, q is the charge of an electron, and N_A is the substrate doping. We can determine the dependence of Φ_{dc} on d by measuring the FET channel current for different drain-to-source voltages. I generalize the original Early-voltage (V_o) definition as

$$V_o = I_{ds} / \frac{\partial I_{ds}}{\partial V_{ds}}, \tag{2.43}$$

where V_{ds} is the drain voltage. For a given saturation current, Φ_{sc} is nearly constant; therefore, I approximate $d\Phi_{dc} = dV_{ds}$ and I define

$$V_o(\Phi_{dc}) = I_{ds} / \frac{\partial I_{ds}}{\partial \Phi_{dc}}. \tag{2.44}$$

From our definition of V_o and (2.42), the Early voltage for a step junction is

$$V_o(\Phi_{dc}) = \sqrt{\frac{2qN_aL^2}{\epsilon}}\sqrt{\Phi_{dc}} - 2\Phi_{dc}. \tag{2.45}$$

If the drain diffusion is not a step junction, but follows the relation $d(\Phi_{dc}) = AL\Phi_{dc}^x$, then the source current is related as

$$I_{source} = \frac{I_{sat}}{1 - A\Phi_{dc}^x}, \tag{2.46}$$

and the resulting V_o is

$$V_o(\Phi_{dc}) = \frac{1}{xA}\Phi_{dc}^{(1-x)} - \frac{1}{x}\Phi_{dc}. \tag{2.47}$$

Figure 2.9 shows the measured early voltage versus Φ_{dc} for two values of gate voltage. The slope of both curves plotted in Fig. 2.9 are in close agreement with step junction behavior. However, the drain-to-channel electric field, and therefore the effective doping, changes significantly with gate voltage. For gate voltages below 7V, the junction behavior is a weak function of gate voltage and roughly corresponds to a substrate doping of 9.53 $\times 10^{16}$cm^{-3}. The effective junction doping steadily increases for larger gate voltages; from Fig. 2.9, a gate voltage of 9.72V corresponds to an effective substrate doping of 2 $\times 10^{18}$cm^{-3}. The dependence of the channel length

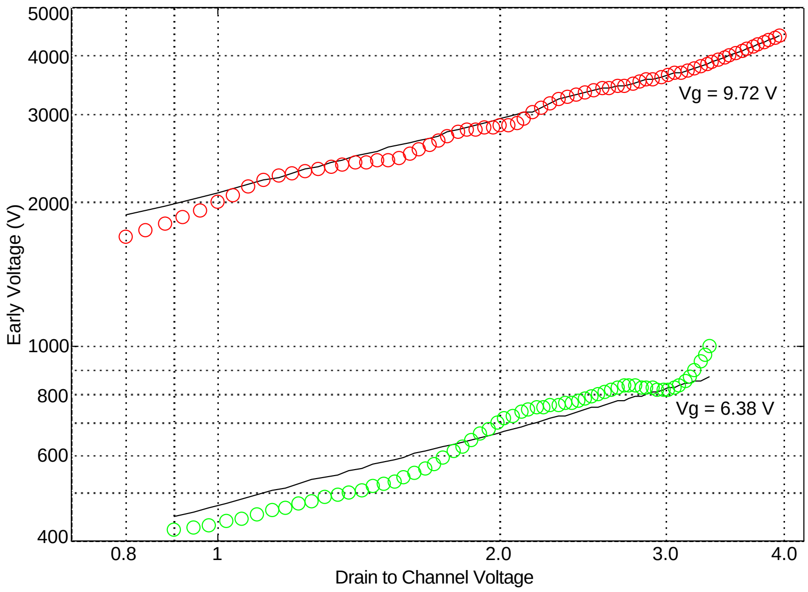

Figure 2.9: Measured Early voltage versus Φ_{dc} for two values of gate voltage. This log-log plot brings out the power law of the junction, and therefore the effective doping profile. For a particular gate voltage, the Early voltage curve is consistent with results from a step junction with $L = 16\mu$m and $N_a = 9.53 \times 10^{16}$cm^{-3}. The curves are similar to the lower V_g case up to gate voltages of 7V, and then gradually increases until it reaches the higher V_g curve. For a gate voltage of 9.72eV, the Early voltage curve is consistent with results from a step junction with $N_a = 2 \times 10^{18}$cm^{-3}.

modulation on the gate voltage does not affect the development of impact-ionization or hot-electron-injection models presented in this chapter.

2.3.3 Electron Distribution Function at the Drain Edge

The gate and substrate currents depend on the distribution function at the drain edge of the drain-to-channel depletion region. To calculate the distribution function for a given drain-to-channel potential (Φ_{dc}), I need to compute the average electron energy at the drain edge (E_{dc}). Here I compute the width of the drain-to-channel depletion region (d), the electric field at the drain edge $\mathcal{E}(z = d)$, and z_{crit} as defined in Section 2.2.

The width of the drain-to-channel depletion region is related to the applied drain

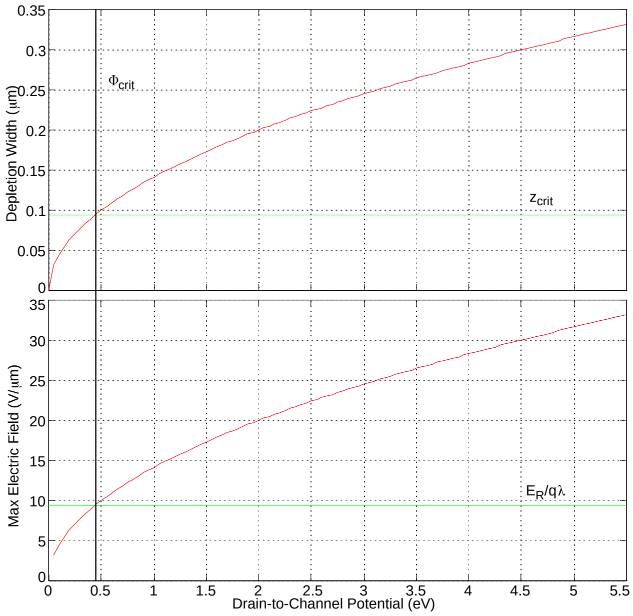

Figure 2.10: Computed parameters for the drain-to-channel depletion region. The plots show the width of the drain-to-channel depletion width, and the maximum electric field in the depletion region (at the drain edge). The substrate doping was $1 \times 10^{17}\text{cm}^{-3}$, and $\lambda = 6.5\text{nm}$. For this device, z_{crit} is approximately one third of the distance into the depletion region for $\Phi_{dc} = 3.6\text{V}$.

to channel potential (Φ_{dc}), and is given by

$$d = \sqrt{\frac{2\epsilon_{si}\Phi_{dc}}{qN_A}}, \tag{2.48}$$

where ϵ_{si} is the permittivity for silicon, q is the charge of an electron, and N_A is the substrate doping level. This relation assumes a step junction doping profile that was justified in the previous section. The electric field, $\mathcal{E}(z)$, is given by

$$\mathcal{E}(z) = \frac{qN_A}{\epsilon}z = \Phi_{dc}\frac{2z}{d^2}, \tag{2.49}$$

and the maximum electric field in the drain-to-channel depletion region, which is the electric field at d, is given by

$$\mathcal{E}(z = d) = \sqrt{\frac{2qN_a\Phi_{dc}}{\epsilon_{si}}}. \tag{2.50}$$

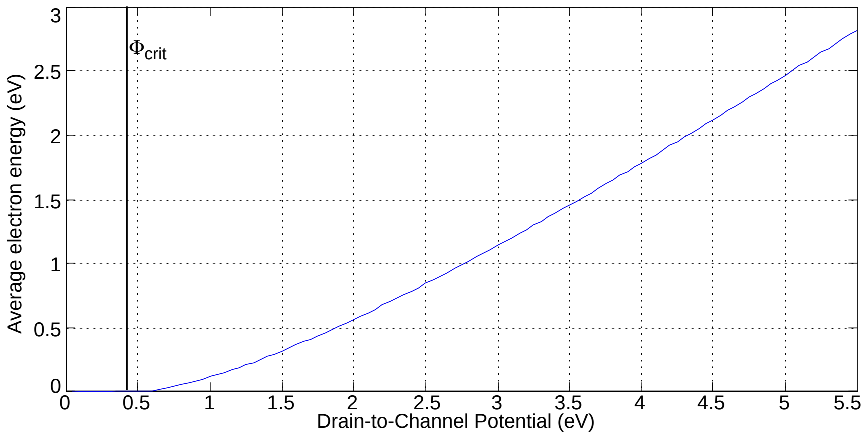

Figure 2.11: Plot of the average electron energy as a function for the drain-to-channel depletion potential. The substrate doping was $1 \times 10^{17}\mathrm{cm}^{-3}$, and $\lambda = 6.5$nm.

Recall that z_{crit} is the position where the applied electric field is equal to $\frac{E_R}{q\lambda}$; I solve for z_{crit} as

$$z_{crit} = \left(\frac{E_R \frac{d}{2\lambda}}{q\Phi_{dc}} \right) d = \frac{E_R}{\lambda} \frac{\epsilon_{si}}{qN_A}. \tag{2.51}$$

I define Φ_{crit} as the potential at z_{crit} by

$$\Phi_{crit} = \left(\frac{E_R d}{2\lambda} \right)^2 \frac{1}{\Phi_{dc}} = \left(\frac{E_R}{\lambda} \right)^2 \frac{\epsilon_{si}}{2qN_A}. \tag{2.52}$$

Figures 2.10 and 2.11 illustrates these parameters calculated as a function of Φ_{dc} for a substrate doping of $N_A = 1 \times 10^{17} cm^{-3}$. Figure 2.12 shows z_{crit} versus substrate doping.

The average electron energy, E_{dc}, is computed by integrating $\mathcal{E}(z)$ from z_{crit} to the drain edge, d, as in (2.27):

$$E_{dc} = E_1(z = d) = q\Phi_{dc} - E_R \frac{d}{\lambda} + q\Phi_{crit}. \tag{2.53}$$

By substituting d from (2.48), the average electron energy is explicitly a function of Φ_{dc}:

$$\sqrt{E_{dc}} = \sqrt{q\Phi_{dc}} - \sqrt{q\Phi_{crit}}. \tag{2.54}$$

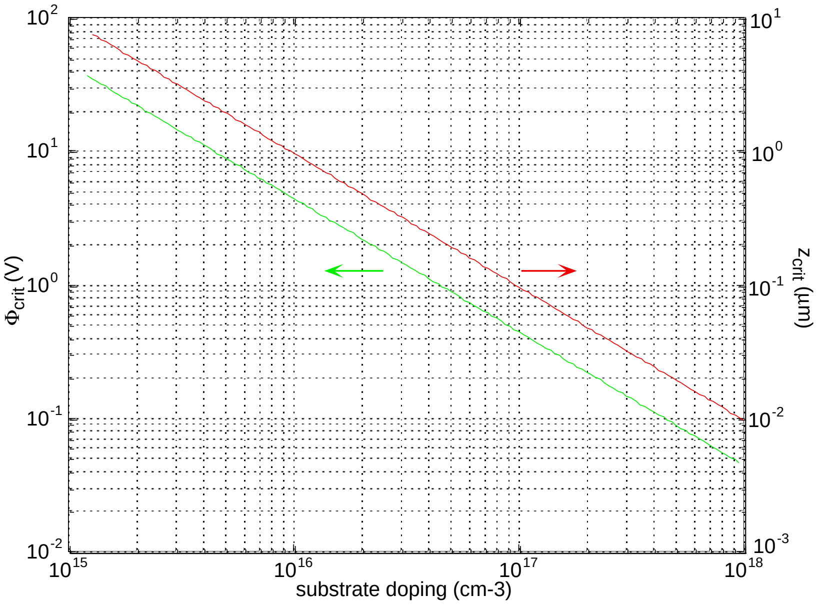

Figure 2.12: Plot of z_{crit} versus MOSFET substrate doping for $\lambda = 6.5$nm.

Figure 2.13 shows the average electron energy versus the drain-to-channel potential for several substrate dopings. The distribution function at the drain edge can be derived from (2.39) as

$$f(d,E) = \exp\left(-\frac{\lambda}{d - z_{crit}}\left(\frac{E - E_{dc}}{2E_R}\right)^2\right) a(d,E),) \tag{2.55}$$

where $a(d,E)$ is

$$a(d,E) = \exp\left(-\frac{1}{(q\mathcal{E}(d)\lambda - E_R)}\int_{E=0}^{E}\frac{\lambda}{L(E)}dE\right). \tag{2.56}$$

Functional form of $L(E)$ in (2.17) results in two distinct $a(d,E)$ regions from the definition in (2.56). For small energies, $|\log(a(d,E))|$ is much less than 1; therefore I approximate $a(d,E)$ by expanding the first exponential in (2.56) as

$$1 - a(d,E) = \frac{\int_0^E \exp\left(\frac{\lambda}{L(E)}dE\right)}{q\mathcal{E}(d)\lambda - E_R}. \tag{2.57}$$

I will use (2.57) to model low electron impact ionization efficiencies for low electron

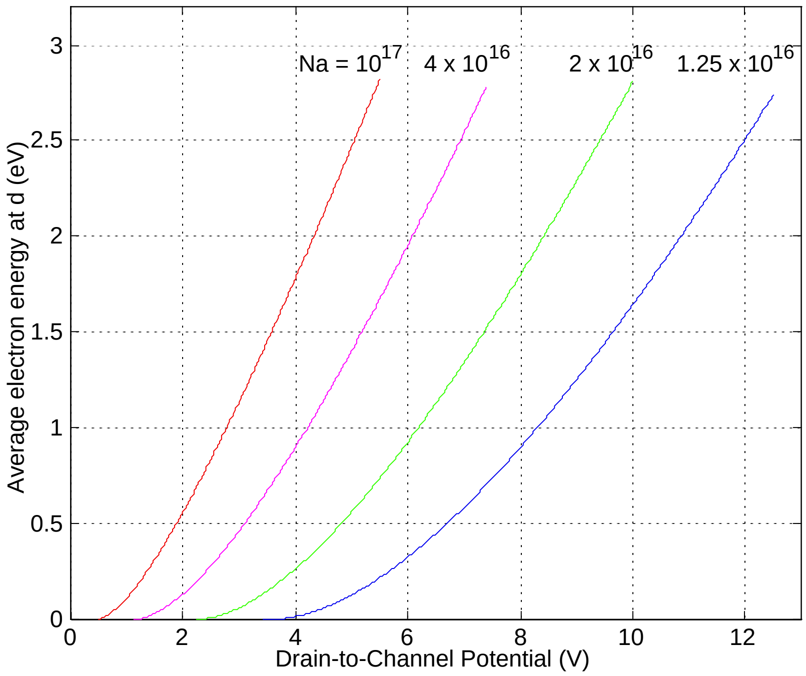

Figure 2.13: Plot of the average electron energy at the drain edge versus drain-to-channel potential for four different MOSFET substrate dopings.

energies, corresponding to I_{ion}/I_s less than 0.2. For larger energies (>2.4 eV), L(E) is a slowly varying function, and can be approximated by expanding the exponential; by expanding around 2.8eV, we can approximate $a(d, E)$ in this region as

$$a(d, E) = \exp\left(-\frac{3.29 \times 10^{-4}\frac{\lambda}{l_{ion}}(E + 3.7eV)}{q\mathcal{E}(d)\lambda - E_R}\right). \tag{2.58}$$

Later in this section, I use (2.58) to model the hot-electron injection current. Figures 2.14 and 2.15 show numerically computed plots for $a(d, E)$ and $1-a(d, E)$ as a function of energy for different values of Φ_{dc}. Figure 2.16 shows a numerically computed plot of the distribution function at d for three values of Φ_{dc}.

2.3.4 Modeling of Hot-Electron Injection in MOS Transistor

This subsection solves for the hot-electron injection (gate current) efficiency. If

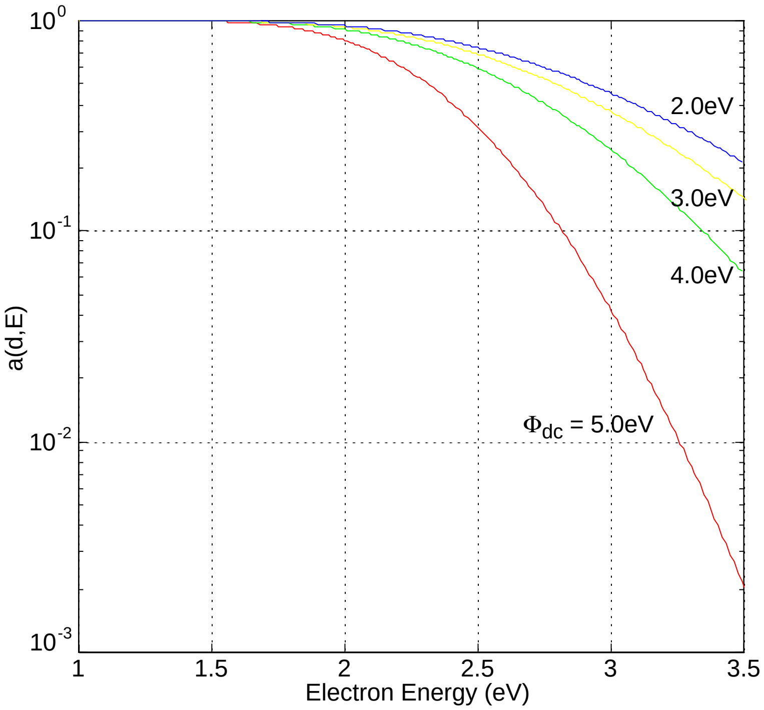

Figure 2.14: Attenuation function at the drain edge due to impact ionization collisions as a function of electron energy for five different Φ_{dc}. I calculated these curves by evaluating (2.56) by numerical integration.

Φ_{dc} and V_{fgd} are fixed, then the injection current should be proportional to the source current, because each electron makes an equal contribution to the injection current. Figure 2.17 shows the measured probability of hot-electron injection versus source current. This data was taken by sweeping the source voltage, and measuring the resulting source current and injection current for a constant floating-gate voltage and drain voltage. The floating-gate–to–drain voltage was explicitly held fixed. For subthreshold biases, the drain-to-channel voltage is fixed throughout the sweep since only the source voltage is adjusted to change the source current. For each sweep in Fig. 2.17, the injection efficiency is very nearly constant for subthreshold current levels. I obtained different drain-to-channel potentials by applying different drain voltages. For above threshold biases, Φ_{dc} is a function of both the source and drain.

Figure 2.18 shows measured data of hot-electron injection efficiencies as a function of drain-to-channel voltage for three channel currents. The experimental evidence

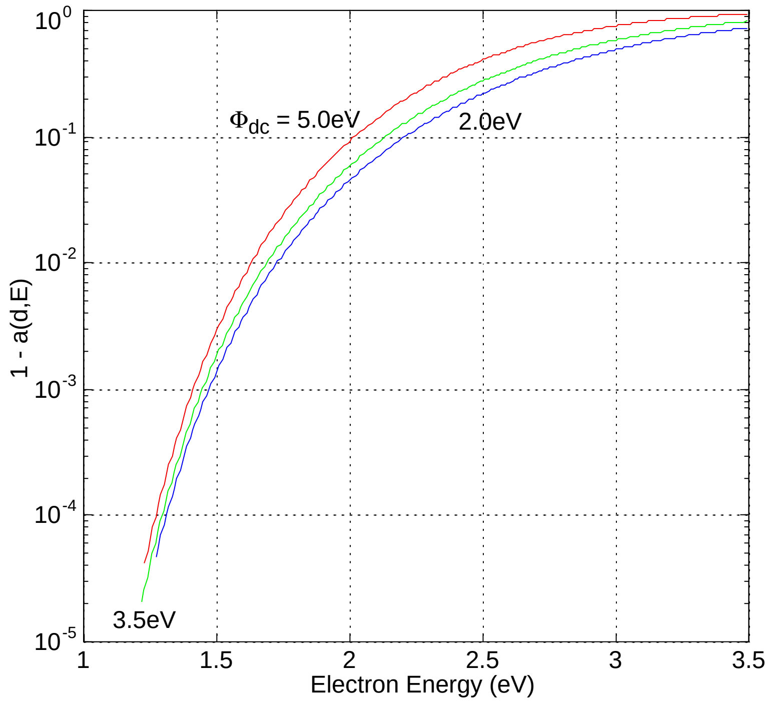

Figure 2.15: 1 - $a(z, E)$ at the drain edge due to impact ionization collisions as a function of electron energy for three different Φ_{dc}. I calculated these curves by evaluating (2.56) by numerical integration.

in Fig. 2.18 shows that the hot-electron-injection efficiency is independent of source current. The electrons go over the barrier by a similar process to thermionic emission, except that the electrons are not described by an equilibrium distribution function. The electrons which have energies higher than the SiO_2 barrier level can participate in hot-electron injection. To investigate hot-electron injection, we need to consider the electron distribution function around the SiO_2 barrier near the drain edge.

I am only considering the case where the floating gate potential is much greater than the drain potential. Two effects are important in this operating region. First, any electrons that can enter the SiO_2 region will be swept towards the floating gate by the resulting electric field. Second, most of the electrons with sufficient energy to surmount the SiO_2 barrier will be at the drain edge. The electron distribution at the SiO_2 barrier energy will be exponentially decreasing as one moves from the drain edge; therefore the hot-electron-injection probability is entirely dependent upon the

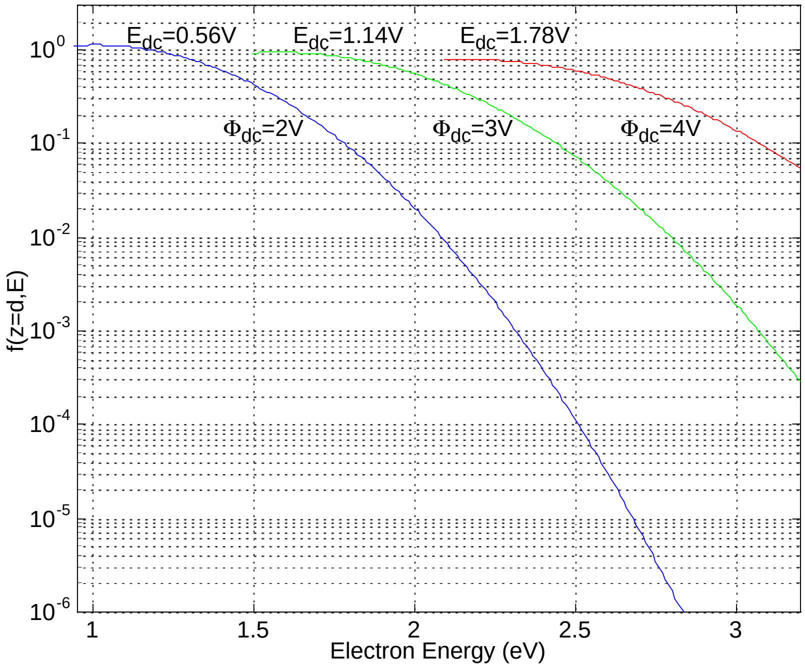

Figure 2.16: Plot of the distribution function at the drain edge as a function of energy for three drain-to-channel potentials. I calculated these curves by evaluating (2.56) by numerical integration.

distribution function at $z = d$. In my experimental setup, the floating gate-to-drain voltage (V_{fgd}) is always positive; therefore, I am certain that most of the electrons are surmounting the barrier at the drain edge. It is well-known that the actual height of the oxide barrier is a function of V_{fgd} due to the barrier lowering effects [90]. I estimate from my data, that the effective barrier height is 2.8eV, and this value is used in my models.

We now solve for the hot-electron injection efficiency. From (2.58), and substituting the values for l_{ion}, λ, and $\mathcal{E}(d)$, the resulting distribution function is

$$f(d,E) \propto \mathrm{e}^{-\frac{\lambda}{d-z_{crit}}\left(\frac{E-q\Phi_{dc}+E_R\frac{d}{\lambda}-q\Phi_{crit}}{2E_R}\right)^2} e^{-\frac{7.102}{\sqrt{E_{dc}}}}, \tag{2.59}$$

where I obtain $a(d,E)$ by expanding around an energy of 2.8eV, as described earlier in this section. Swings in the barrier heights of 100-200meV around 2.8eV add only

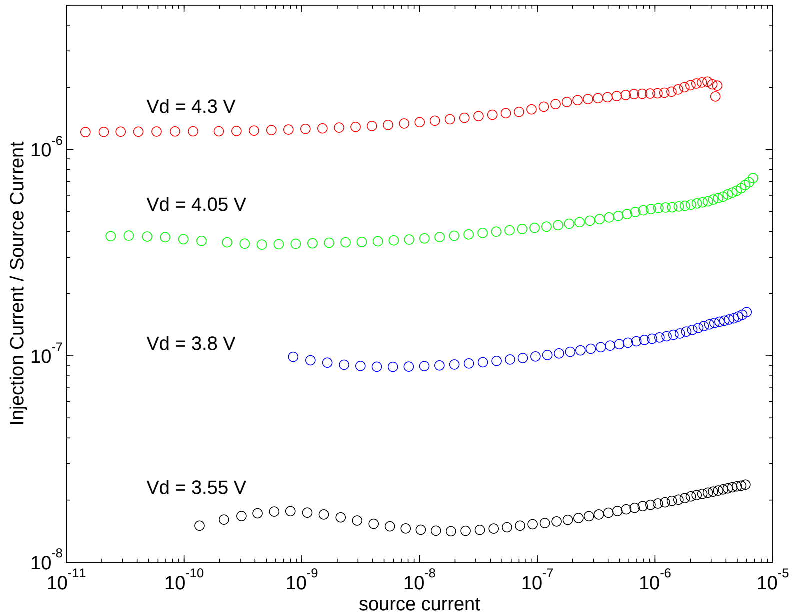

Figure 2.17: Electron injection efficiency versus source current for three values of drain voltage. The curves were obtained by holding the drain and floating-gate fixed while sweeping the source voltage. Electron injection efficiency is defined as the ratio of the injection current and the source current. For subthreshold currents, this ratio is constant for a given drain voltage since the drain-to-channel voltage is constant in this regime.

a small correction term. For electrons to inject into the oxide, they need to have a direction pointing towards the SiO_2; in the discussion, I will show that the angle distribution only weakly affects the hot-electron–injection efficiency.

The hot-electron-injection efficiency (δ) is approximated as proportional to the integral of the distribution function at the drain edge from the oxide barrier energy to the vacuum level, which I write as

$$\delta = \frac{I_{inj}}{I_s} \propto \int_{E=E_{ox}(V_{fgd})}^{E=\infty} f(d, E) dE. \tag{2.60}$$

I approximate this integral to obtain δ as

$$= B_2 e^{-\frac{\lambda}{d - z_{crit}}\left(\frac{E_{ox}(V_{fgd}) - \Phi_{dc} + E_R \frac{d}{\lambda} - \Phi_{crit}}{2E_R}\right)^2} e^{-\frac{7.102}{\sqrt{E_{dc}}}}, \tag{2.61}$$

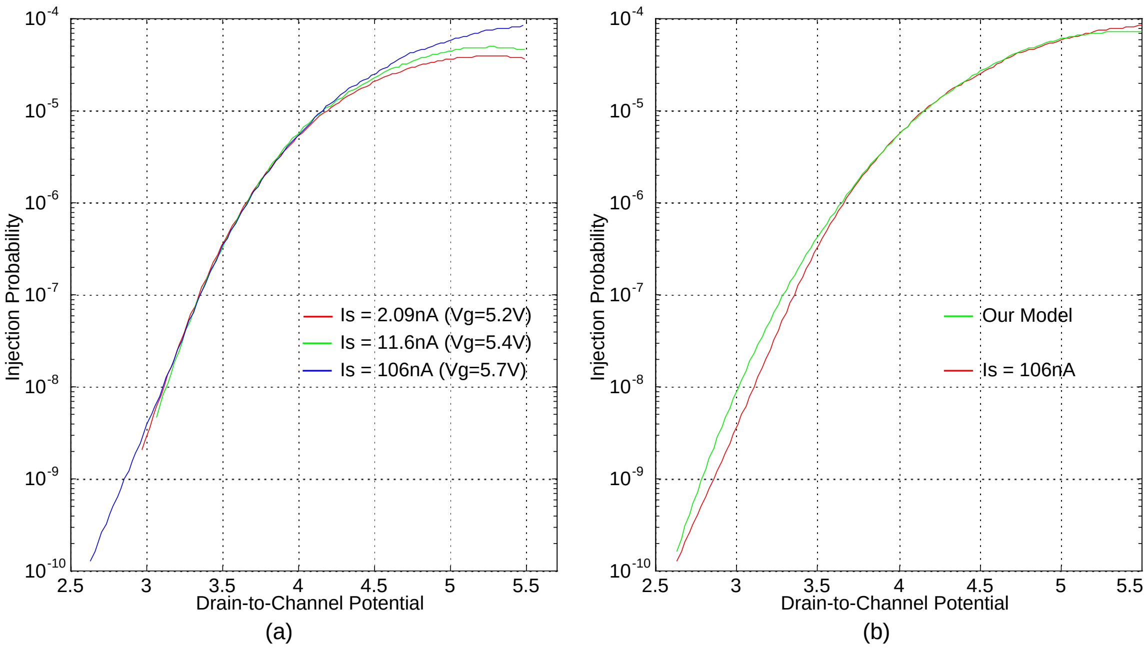

Figure 2.18: (a) Measurements of hot-electron-injection efficiency versus drain-to-channel voltage for several values of source current. I simultaneously measured the substrate and gate currents for different drain-to-channel voltages which gives us the impact-ionization and hot-electron-injection efficiency. The drain-to-channel voltage is computed from the source current and the drain-to-source voltage. For each sweep, I used a constant gate voltage to choose a particular channel current; the actual oxide barrier height changes slightly due to image force lowering, because the floating-gate–to–drain voltage is not constant. E_{ox} will roughly change by 100meV as predicted by image force lowering. (b) Measurement of hot-electron-injection efficiency versus Φ_{dc} compared with a curve fit to the analytic model in (2.61).

where $B_2 = 4.55 \text{ x } 10^{-3}$ is a measured constant of proportionality which includes the effects of the electron angles. In this expression, I explicitly show that oxide barrier energy at the drain edge, $E_{ox}(V_{fgd})$ is a function of V_{fgd}. Figure 2.18 shows (2.61) fitted to the injection efficiency data. The curve fit shows close agreement to (2.61) except at Φ_{dc} greater than 5.3V, since average-electron energy is near the energy of the silicon–silicon-dioxide barrier. The significant deviation at low Φ_{dc} is probably due to ignoring the band-structure effects.

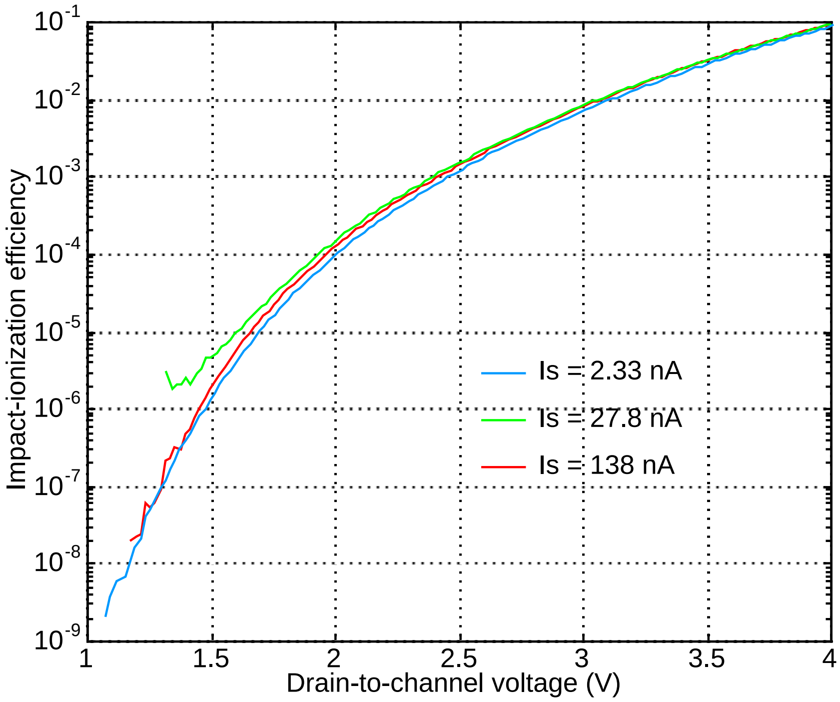

Figure 2.19: Measurements of impact-ionization efficiency vs. drain to channel voltage for three source currents (gate voltages).

2.3.5 Electron Impact Ionization

This subsection solves for the impact ionization (substrate current) efficiency (α), which is the ratio of the substrate current (I_{ion}) to the current at the source (I_s) of the transistor. Figure 2.19 shows measured data of impact-ionization efficiencies as a function of drain-to-channel voltage for several currents. The impact ionization efficiency $\alpha(E_{dc})$ can be thought of as the probability of an electron starting in the channel undergoing an impact ionization in the drain-to channel depletion region.

I calculate $\alpha(E_{dc})$ by computing the ratio of the integral of the distribution function of the electrons undergoing an impact-ionization collision to the integral of the starting distribution function:

$$\alpha(E_{dc}) = \frac{I_{ion}}{I_s} = \frac{\int_{-\infty}^{\infty} (f(0,E) - f(d,E))\, dE}{\int_0^{\infty} f(0,E) dE}. \tag{2.62}$$

To evaluate $f(0,E)$, I substitute $\int_0^\infty f(d,E)dE$ for $\int_0^\infty f(0,E)dE$ for the case of no impact ionization ($a(d,E)=1$ for all energies); this mathematical operation explicitly shows the influence of $1-a(d,E)$ on $\alpha(E_{dc})$. In the region where the electron distribution function can be approximated by a Gaussian (see (2.33)), the equation for $\alpha(E_{dc})$ becomes

$$\alpha(E_{dc}) = \frac{\int_0^\infty (1-a(d,E))\, e^{-\frac{\lambda}{d-z_{crit}}\left(\frac{E-E_{dc}}{2E_R}\right)^2} dE}{\int_0^\infty e^{-\frac{\lambda}{d-z_{crit}}\left(\frac{E-E_{dc}}{2E_R}\right)^2} dE}. \tag{2.63}$$

From (2.63) and the distribution function given by (2.55), we can solve for $\alpha(\Phi_{dc})$. I approximate this solution by expanding the function in the exponent around the function's maximum value in E; the critical energy, E_c, that maximizes this function is nearly equal to the linear function

$$E_c = E_{c1} - \frac{3}{5}\left(1 - \frac{E_{c1}}{2.85\text{eV}}\right)(E_{dc} - 0.95\text{eV}), \tag{2.64}$$

where E_{c1} is

$$E_{c1} = \left(E_R^2 \sqrt{\frac{2\epsilon_{si}(119\text{eV})(0.95\text{eV})}{q^2 N_a \lambda^2}}\right)^{2/5}. \tag{2.65}$$

By expanding the terms in the exponent of (2.63) around E_c, using the relation in (2.54), I get the approximate solution of $\alpha(\Phi_{dc})$ as

$$\alpha(\Phi_{dc}) = \exp\left(-\sqrt{\frac{119\text{eV}}{\left(1-\frac{E_{c1}}{2.85\text{eV}}\right)\Phi_{dc}}}\right)$$

$$\exp\left(\frac{\sqrt{\frac{qN_a\lambda^2}{2\epsilon_{si}}}}{\sqrt{\Phi_{dc}} - \sqrt{\Phi_{crit}}}\left(\frac{E_{c1} - \frac{3}{5}\left(1-\frac{E_{c1}}{2.85\text{eV}}\right)(E_{dc}-0.95\text{eV})}{2E_R}\right)^2\right) \tag{2.66}$$

This integration yields a smoothed version of $L(E)$. The factor in the second exponential decreases rapidly for Φ_{dc} greater than 3.5V. Note that we can use (2.63) to compute $a(d,E)$, and therefore can compute $L(E)$ from measurements of α as a function of Φ_{dc}. We calculate the $L(E)$ function plotted in Fig. 2.4 using the measurements of α as a function of Φ_{dc}.

So far, this model only models the impact-ionization current in the drain-to-channel depletion region and not the impact-ionizations in the drain regions due to the high-energy electrons entering this region. In this region, the electron transport changes in two ways. First, the electric field in the drain region is negligible. Second, the electrons start at high energies. The transport in this region is similar to other experimental setups used to measure quantum-yield by other researchers [75], and therefore, provides yet another check of my modeling from additional experimental results.

We will follow the distribution of a single electron on its (x, y, z) path through the drain. I represent the position along this path by s; this approach allows us to concentrate on the energy distribution and neglect the momentum scattering. The following Boltzmann transport equation models the electron distribution in (s, E)

$$\lambda \frac{\partial f}{\partial s} = E_R \frac{\partial f}{\partial E} + F(T) \frac{E_R^2}{2} \frac{\partial^2 f}{\partial E^2} - \frac{\lambda}{L(E)} f. \tag{2.67}$$

Our goal is to solve for the number of electrons that are removed from the distribution over all position; therefore I integrate over all s

$$f(s, E)|_{s=0}^{\infty} = E_R \frac{\partial \bar{f}}{\partial E} + F(T) \frac{E_R^2}{2} \frac{\partial^2 \bar{f}}{\partial E^2} - \frac{\lambda}{L(E)} \bar{f}, \tag{2.68}$$

where I define

$$\bar{f} = \frac{1}{\lambda} \int_0^{\infty} f(s, E) ds. \tag{2.69}$$

Since all the electrons will return to equilibrium energies at the conduction band, $f(\infty, E) = 0$ for all energies, except for $E = 0$. Also, since I am considering a single electron starting at an initial energy, E_i, $f(0, E) = \delta(E - E_i)$, where $\delta(\cdot)$ is the delta function. I simplify (2.68) by modeling the distribution function for low impact-ionization efficiencies, and by assuming a negligible second derivative in energy; both simplifications are consistent with the model level in the drain-to-channel region. The

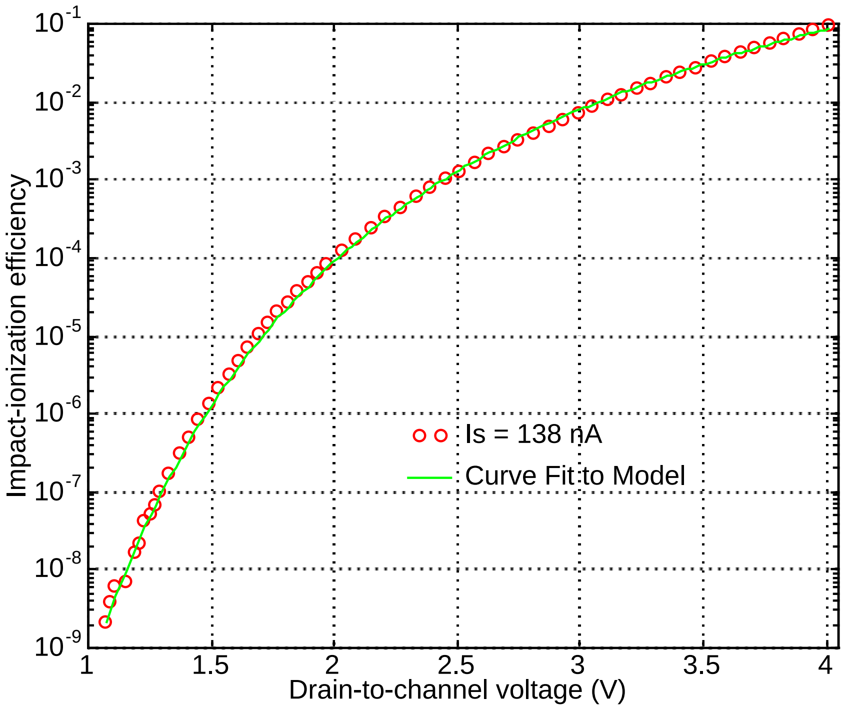

Figure 2.20: Measurement of impact-ionization efficiency versus Φ_{dc} compared with a curve fit to the analytic model in (2.71). The source voltage was held at zero and the gate voltage was held at 5.8V.

solution to this simplified form of (2.68) is

$$1 - \bar{f} = \frac{1}{E_R} \int_0^\infty \frac{\lambda}{L(E)} dE. \tag{2.70}$$

Combining the impact-ionization models in the drain and drain-to-channel regions, I model E_{dc} as

$$\alpha(E_{dc}) = \frac{\left(\frac{1}{q\mathcal{E}(d)\lambda - E_R} + \frac{1}{E_R}\right) \int_0^\infty \frac{\lambda}{L(E)} e^{-\frac{\lambda}{d - z_{crit}}\left(\frac{E - E_{dc}}{2E_R}\right)^2} dE}{\int_0^\infty e^{-\frac{\lambda}{d - z_{crit}}\left(\frac{E - E_{dc}}{2E_R}\right)^2} dE}. \tag{2.71}$$

This modification increases $\alpha(E_{dc})$, and therefore increases the effect of $L(E)$. Figure 2.19 shows experimental measurements of α versus drain-to-channel potential for three different channel currents. Figure 2.20 shows a curve fit (solid line) of (2.71) and (2.66) to the experimental data with the parameters defined in Section 2.1; the fit closely agrees with the measured data. I measured the impact ionization efficiency by assuming at most only a single impact ionization event will occur for each starting electron, which is reasonable for the low impact-ionization efficiencies of the measurements shown in Fig. 2.20.

2.4 Discussion

In the previous section, I skipped several subtle discussion points for clarity of the presentation. I will address three points in this discussion section. Section 2.4.1 discusses the effect of collision broadening on my derived phonon collision operator. Section 2.4.2 considers the effect of momentum-angle scattering due to optical-phonon collisions. Section 2.4.3 discusses how the electron obtains the necessary angle as well as energy for hot-electron injection.

2.4.1 Effect of Collision Broadening on the Phonon Collision Operator

My model of optical phonon collisions has several interesting extensions. One interesting extension is that the starting point and derivation for acoustical phonons of a particular energy or band to band scattering is similar to my results for optical phonons [60]; therefore with a different set of parameters, (2.12) can model the effects of these three collision processes. It can also be shown when all three processes are taken into account, that the largest contribution comes from the optical phonon collisions.

Second, my model of the optical-phonon collision operator changes only slightly when I model the effect of collision broadening. This result further justifies my use of semi-classical theory, since the τ for optical-phonon collisions for hot-electron injection and impact ionization are at the boundary where Boltzmann transport may not be valid. Collision broadening is due to the limited certainty about a particular energy between collisions because of the uncertainty principle, $\Delta E\tau = \hbar$, where τ is the average time between collisions. For τ of 10fs, which is the order of the smallest τ, the energy uncertainty is roughly 60meV.

I will now show the effect of collision broadening on my optical-phonon collision operator. I define the broadening function, $B(E)$, as a symmetric function around $E = 0$ that is characterized by a variance of E_d. I formulate the new optical-phonon collision operator as

$$S_{op}(f)\left[\frac{m^*(E)}{c(E)}\right] \approx -\left(e^{\frac{E_R}{kT}}+1\right) f(E)$$

$$\int_{\infty}^{-\infty}\left[B(\Delta E + E_R) + e^{\frac{E_R}{kT}} B(\Delta E - E_R)\right] f(E-\Delta E) d\Delta E. \tag{2.72}$$

This function is a convolution; by taking the Laplace transform, we transform the convolution to a product:

$$\mathcal{L}\left\{S_{op}(f)\left[\frac{m^*(E)}{c(E)}\right]\right\} \approx -\left(e^{\frac{E_R}{kT}}+1\right) f(s_E) + \left[e^{-s_E E_R} + e^{s_E E_R + \frac{E_R}{kT}}\right] B(s_E) f(s_E). \tag{2.73}$$

Since $B(E)$ is real, symmetric function, then $B(s_E)$ will also be a real, symmetric function, and therefore I can expand $B(s_E) = 1 - s_E^2 E_d^2 + O(s_E^4)$; E_d is a measure of the width of the energy uncertainty. If we Taylor expand the s_E terms to second order, we get

$$\mathcal{L}\left\{S_{op}(f)\left[\frac{m^*(E)}{c(E)}\right]\right\} \approx s_E E_R \left(e^{\frac{E_R}{kT}}-1\right) f(s_E)$$

$$+s_E^2 \frac{E_R^2}{2}\left(e^{\frac{E_R}{kT}}+1\right) f(s_E) - s_E^2 \frac{E_d^2}{2}\left(e^{\frac{E_R}{kT}}+1\right) f(s_E). \tag{2.74}$$

By taking the inverse Laplace transform of this approximate equation, we get the new collision operator as

$$S_{op}(f)\left[\frac{m^*(E)}{c(E)}\right] \approx F(T)\frac{E_R^2 - E_d^2}{2\lambda}\frac{\partial^2 f}{\partial E^2} + \frac{E_R}{2\lambda\zeta}\frac{\partial f}{\partial E}. \tag{2.75}$$

Collision broadening reduces the second derivative in energy term in in the collision operator expression; Monte Carlo simulations have seen that collision broadening increases the number of high-energy electrons [59] by decreasing the restoring force of the optical phonons. The effect of collision broadening will decrease the higher order derivative terms; the higher order derivatives are more significant at low energies and fields due to the sharper distribution functions. The effect of collision broadening will also be larger at lower energies since $\left(1+\frac{E_R}{E}\right)^{1/2}$ will deviate from 1.

2.4.2 Effect of Momentum Scattering Optical Phonons

In Section 2.1, I assumed that the phonon collisions did not affect the overall electron momentum distribution function. In my previous analysis, I have assumed that the phonon distribution is at its equilibrium level; a careful modeling must also take these effects into account. Phonons have momentum, and the total momentum involved for a phonon absorption or emission must be conserved. My previous phonon-collision operator defined in (2.10) left the distribution function unaffected in the ζ direction; this assumption is one extreme of a continuum of plausible situations. The other extreme assumes that phonons have uniform momentum angle distributions, and therefore a phonon collision will uniformly scatter the electron distribution through all angles. To precisely model this effect, one would need to know the distribution function of momentum for the phonons in the drain–to–channel depletion region; this subsection investigates the assumption that a phonon collision uniformly randomizes the momentum distribution. I model the new phonon collision operator by modifying the $f(E+E_R,\zeta)$ and $f(E+E_R,\zeta)$ terms in (2.10) as

$$f(E,\zeta) \rightarrow \frac{1}{2}\int_0^1 \left(f(E,\zeta_1)+f(z,E,-\zeta_1)\right)d\zeta_1. \tag{2.76}$$

Again, I will assume that $E \gg E_R$, and therefore $\left(1+\frac{E_R}{E}\right)^{1/2} \approx 1$. The resulting new phonon operator makes an analytical solution of the transport equations extremely difficult. This subsection will consider the resulting behavior when phonon collisions uniformly randomize the momentum distribution function.

One might wonder if this collision operator would result in a nearly isotropic distribution function. The largest electron energy is approximately equal to my previously derived average electron energy for the same z_{crit}. Hot electron injection will occur at low Φ_{dc} only if λ is substantially longer than I defined in Section 2.1; a longer λ requires fewer electron collisions in the drain–to–channel depletion region. The average electron energy for this isotropic model is nearly equal to the energy gain for the $\zeta = 0$ electron. The average electron energy of an electron starting at z and arriving at $z + \lambda$ is approximately described by

$$E_1(z + \lambda) = \sqrt{E_1(z)^2 + (q\mathcal{E}(z)\lambda)^2} - E_R. \tag{2.77}$$

This approximation overestimates the average-electron energy because electrons reaching $z+\lambda$ from z may take several paths that require more phonon collisions, resulting in fewer electrons at higher energies. As in Section 2.2, an electron must gain more energy than is lost by the phonons to leave the conduction band. We define this threshold when $E_1(z_{crit} + \lambda) = E_R$ for $E_1(z_{crit}) = E_R$; we can solve for z_{crit} as

$$q\mathcal{E}(z_{crit}) = \sqrt{3}E_R. \tag{2.78}$$

This z_{crit} is larger than I derived in Section 2.2. The average electron energy will reach a steady state when $E_1(z + \lambda) = E_1(z)$; the steady-state value of $E_1(z)$ is

$$E_1(z) = \frac{(q\mathcal{E}(z)\lambda)^2 - E_R^2}{E_R}, \tag{2.79}$$

and at $z = d$, the steady-state value is

$$E_1(z) = E_R\left(\frac{\Phi_{dc}}{\Phi_{crit}} - 1\right). \tag{2.80}$$

For the parameters given in Section 2.3, $E_1(z) \approx 600$meV for Φ_{dc} =5.5eV; λ would need to increase by more than a factor of two to match the value in Section 2.3. Also, the numerical calculations of impact ionization and hot-electron injection [76] do not give this type of distribution. Therefore it seems highly improbable that the electron distribution function remains nearly isotropic. We could argue that the relevant electrons have no collisions, as in the physical description of the lucky electron model; this hypothesis could explain the longer λ used in lucky-electron formulations. Since this model does not simultaneously fit both hot-electron-injection and impact-ioniziation currents, this hypothesis also does not satisfy the data; this hypothesis would also require the unlikely possibility that λ is a strong function of Φ_{dc}.

The above arguments show that an isotropic solution is inconsistent with the experimental data; therefore the distribution function in ζ must have bias along the field direction. Consider the case where the distribution function is near to my previous solution assuming $\zeta = 1$. If we are interested in the solution in a neighborhood near $\zeta = 1$, we can Taylor expand the distribution function in the integrals around $\zeta = 1$:

$$S_{op}(f)\left[\frac{m^*(E)}{c(E)}\right] \approx F(T)\frac{E_R^2}{2\lambda}\frac{\partial^2 f}{\partial E^2} + \frac{E_R}{\lambda}\frac{\partial f}{\partial E} - F(T)\frac{\zeta}{\lambda}\frac{\partial f}{\partial \zeta} + \frac{F(T)}{\lambda}\left(\frac{1}{6} + \frac{\zeta}{2} + \zeta^2\right)\frac{\partial^2 f}{\partial \zeta^2} \tag{2.81}$$

where $F(T)$ was defined in (2.13), which is nearly equal to one at room temperature ($T = 300K$). This expansion is valid to third order in partial derivatives, and is very nearly equal to the original expression for energies greater than a few E_R. The first order term moves electrons further towards $\zeta = 1$; this new phonon collision operator modifies the characteristic equation for $\zeta_1(z)$ in (2.25) as

$$\frac{d\zeta_1(z)}{dz} = q\mathcal{E}(z)\frac{1-\zeta^2}{\zeta E} + \frac{1}{\lambda}, \tag{2.82}$$

resulting in $\zeta_1(z)$ moving faster towards 1 than before and being nearly independent of electron energy or electric field. The resulting partial differential equation is

$$\frac{\partial f}{\partial z} + \frac{1}{\lambda}\frac{\partial f}{\partial \zeta} = F(T)\frac{E_R^2}{2\lambda\zeta}\frac{\partial^2 f}{\partial \delta E^2} + \frac{F(T)}{\lambda}\left(\frac{1}{6} + \frac{\zeta}{2} + \zeta^2\right)\frac{\partial^2 f}{\partial \zeta^2}. \tag{2.83}$$

This characteristic equation further illustrates that electrons near $\zeta = 1$ will not quickly move away from $\zeta = 1$. The diffusive second-order term is consistent with a phonon collision uniformly randomizing the electron's momentum angle.

We will now solve the distribution function around $\zeta = 1$. By transforming (2.29) with the new collision operator in (2.81) to δE coordinates, I approximately get

$$\frac{\partial f}{\partial z}+\frac{1}{\lambda}\frac{\partial f}{\partial \zeta}=F(T)\frac{E_R^2}{2\lambda\zeta}\frac{\partial^2 f}{\partial \delta E^2}+\frac{F(T)}{\lambda}\frac{\frac{1}{6}+\frac{\zeta}{2}+\zeta^2}{\zeta}\frac{\partial^2 f}{\partial \zeta^2}. \tag{2.84}$$

For purposes of analysis, I assume that the solution of (2.84) near $\zeta=1$ is close to the solution of (2.84) when evaluating the coefficients at $\zeta=1$:

$$\lambda\frac{\partial f}{\partial z}+\frac{\partial f}{\partial \zeta}=F(T)\frac{E_R^2}{2}\frac{\partial f}{\partial \delta E}+\frac{5}{3}F(T)\frac{\partial^2 f}{\partial \zeta^2}. \tag{2.85}$$

The solution in $(z,\delta E,\zeta)$ to this simplified equation for an impulse at z_{crit} is

$$f=\frac{(\zeta-1)\sqrt{4E_R^2\frac{z-z_{crit}}{\lambda}}}{12\lambda\sqrt{\pi z^3}}\exp\left(-(\frac{9A^2}{16}+\frac{3}{2})\frac{z-z_{crit}}{\lambda}\right)$$

$$\exp\left(\frac{3A}{4}(\zeta-1)-\frac{\lambda}{z-z_{crit}}\left(\frac{3}{8}(\zeta-1)^2+\frac{\delta E^2}{2E_R^2}\right)\right). \tag{2.86}$$

By taking the integral of this solution over all ζ, the solution for $\int_{\zeta=-1}^{1} f(z,E,\zeta)$ is identical to my previous solution for f assuming that $\zeta=1$.

I will make some qualitative points about the distribution function by extrapolating beyond this perturbative analysis, that is when ζ is significantly smaller than 1. First, the distribution function is diffused more in δE. Second, the average electron energy will be lower along a given $\zeta_1(z)$ path than for the $\zeta=1$ case we previously analyzed. Third, a lower ζ will result in a slower diffusion of the distribution function away from $\zeta=1$; the diffusion virtually stops near $\zeta=0$. Therefore, it is reasonable to believe that the $\zeta=1$ behavior will dominate the phenomena for most of my range of measurements, but behavior for ζ much smaller than 1 will be significant near and below $E_1(z)$, and this boundary coincides where the Gaussian model breaks down as we previously illustrated in Fig. 2.7. The best approach to justify this model will be with comparisons to full bandstructure Monte Carlo calculations similar to [76], for this device.

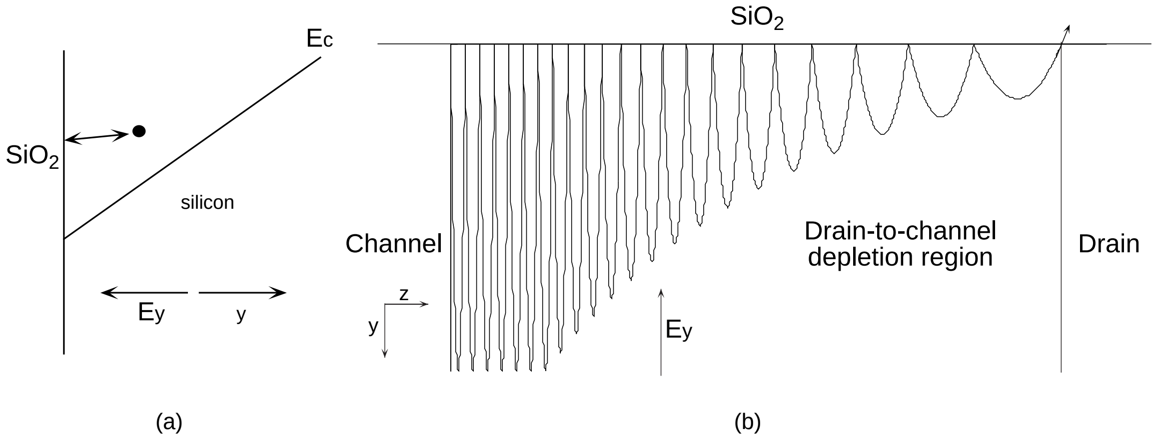

Figure 2.21: Illustration of the 2D nature of the electron transport in the drain-to-channel region. The electron elastically bumps along the barrier at the SiO_2 interface and then gets redirected back towards the interface due to the electric field. In this way, some of the electrons have the proper direction to enter the SiO_2 and eventually reach the floating gate.

2.4.3 How the Electron Gets the Necessary Angle for Impact Ionization

For an electron to reach the floating gate, it must have energy greater than the oxide barrier height and must be directed towards the SiO_2 when the electron reaches that energy. In the previous sections, I only considered how the electrons reach the energy to surmount the SiO_2 barrier; this section addresses how the electrons enter the SiO_2. An electron gains energy due to the electric field in the z direction; electrons are confined by the electric field in the y direction. In the y direction, the electron is being accelerated by the electric field towards the silicon–silicon-dioxide interface, only to be elastically reflected off of this interface back into the silicon. Therefore, the electron's position oscillates in y, and that results in a roughly uniform distribution function in y that is independent from the distribution function in z. Figure 2.21 illustrates the 2D path that the electron takes in the drain-to-channel depletion region. Previous theories by [52, 53] propose that the electrons become directed into the SiO_2 by an elastic ion collision. According to my model, the elastic-ion collisions have little effect in the y direction because randomizing the electron momentum cannot further diffuse an already uniform distribution function.