February 2013
Semiconductor Briefs

EM|QM
S·C·I·E·N·C·E

Carrier Screening

A brief literature overview of Coulomb-potential screening by free charge carriers.

By Jeffrey R. Clark, PhD

Screening of Coulomb potentials by free carriers or bound charges is a ubiquitous phenomenon that occurs in all manner of materials ranging from insulators to metals. Of particular interest is the screening of potentials by mobile carriers in semiconductors and the results that this action has on the behavior of semiconducting materials.

In insulators (dielectric materials), the screening of a potential involves polarization of the surrounding material. This polarization is, in effect, a displacement of the centroids of positive and negative charge in the atoms or molecules, resulting in the creation of numerous nearly infinitesimal dipoles [1]. Although the functional dependence of the potential is unchanged, the dielectric constant increases, reducing the potential φ.

$$\phi(r) = \frac{q}{4\pi\varepsilon r} \tag{1}$$

As (1) shows, when the dielectric constant ε is greater than ε_0, the potential decreases with respect to the free-space potential.

A much more dramatic and relevant form of screening occurs in the presence of mobile carriers, which are abundant in conductors (metals). Free (or nearly free) carriers in the metal will ideally arrange themselves to maintain a constant potential throughout. Any variation in the potential will cause the carriers to reorient. This phenomenon leads to effective screening of any Coulomb potential in the metal to within approximately the interatomic distance. Beyond simply changing a multiplicative factor as in the case of dielectrics, the functional dependence of a screened potential in a metal is drastically altered. (It can be modeled to some degree of accuracy as an exponential, as this discussion will show later.)

Between these two extremes of conductors and dielectrics are semiconducting materials. Semiconductors typically have some concentration of mobile charge carriers available for screening, and they can be expected to behave similarly to metals, albeit to a limited extent. Although screening by mobile carriers is the dominant factor in this problem, dielectric screening does also take place in semiconductors and cannot be ignored.

Qualitative Description of Screening

In a perfect crystal, electrons in a particular state indexed by the wave vector **k** are not scattered by the atoms (potentials in the lattice) [2]. Effectively, then, the ideal crystal acts as a uniform (or zero) potential with regard to electron behavior. On the other hand, impurities or defects produce potential fluctuations that can scatter electrons. Since these defects alter the behavior of the semiconductor, including its optical and electrical properties, determining the nature of the potential and, specifically in this case, how the mobile carriers are involved is important.

Qualitatively, a potential caused by a defect can be viewed as a perturbation of the energy-band diagram with respect to a spatial coordinate, as Figure 1 shows. The dip in the energy band depicted in this figure creates new lower energy states near the impurity. As a result, when free electrons are abundant, they will move to achieve a more preferred lower-energy state by "filling" this dip in the conduction band. Figure 2 shows the filling effect. The rearrangement of carriers in this scenario leads to the presence of an excess density above the uniform concentration, balancing the potential to some extent. When few mobile carriers are present, the screening is less pronounced [3].

This description applies to a positively charged impurity, where electrons screen the potential; the same action can occur around a negatively charged impurity, however. Holes will fill the "bump" in the band, screening the negative potential in the same way that electrons screen the positive potential. In a doped semiconductor, then, the band structure should not be a series of smooth curves, but will instead comprise moderately rough curves, indicating the presence of (screened) potentials owing to defects in the lattice. Figure 3 illustrates this situation approximately.

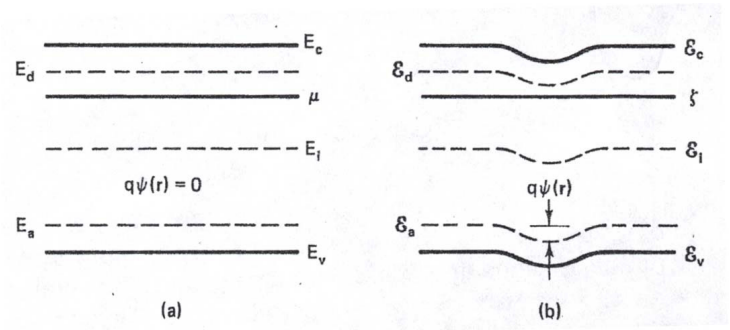

Figure 1. A perturbation of the energy band in direct space owing to the presence of an ionized impurity with electrical potential $\psi(r)$. (Image from [2].)

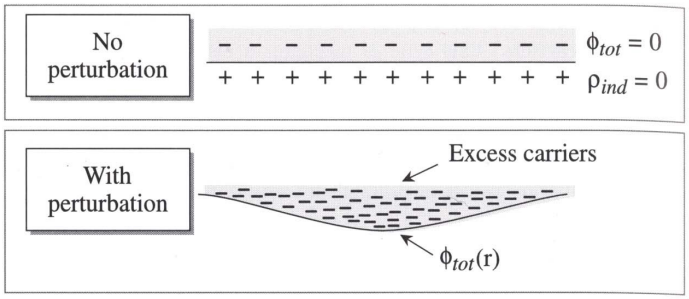

Figure 2. Movement of mobile carriers to fill the "dip" and screen the impurity potential. (Image from [3].)

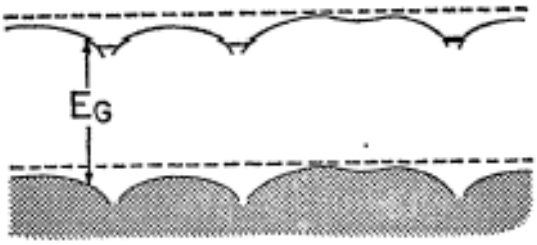

Figure 3. Variations in the energy band due to the presence of impurities or defects. (Image from [4].)

Thomas-Fermi Model

Quantitatively, one simple approach to screening uses the Thomas-Fermi model of an electron "gas." This semiclassical approach solves both the Schrödinger and Poisson equations. An original application of this model was metals [5], but it also applies roughly to semiconductors. The model requires a significant number of carriers in the conduction band (since the screening electrons are considered "free") and therefore a Fermi energy higher than the bottom of the conduction band, as Figure 4 depicts. This situation implies that the semiconductor is degenerate, and the results must be considered in this light.

The approach first requires derivation of the Fermi energy of the free-electron gas in terms of useful or known parameters [7]; this derivation uses the solution to the Schrödinger equation for a single free electron in three dimensions.

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{r}) \qquad (2)$$

The solution to this Schrödinger equation for the case of confinement to a cube of edge length L is the typical solution for an infinite potential well in three dimensions. If a periodic boundary condition is placed on the solution such that $\psi_{\mathbf{k}}(\mathbf{r})$ has a period L in all three dimensions, then the solution becomes the following:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sqrt{\frac{1}{V}}\,e^{i\mathbf{k}\cdot\mathbf{r}} \qquad (3)$$

where V is the volume of the cube (L^3). This plane-wave solution restricts the allowed components of the wave vector **k,** being k_x, k_y and k_z, to

$$k_i = \pm\frac{2n\pi}{L}, \qquad n = 0,1,2,3,... \qquad (4)$$

The use of the quantum-mechanical momentum operator allows expression of the Fermi energy E_F in terms of a Fermi wave vector $\mathbf{k}_F$.

$$E_F = \frac{\hbar^2}{2m}k_F^2 \qquad (5)$$

If a certain sphere in **k**-space contains all the ground-state free electrons, then $\mathbf{k}_F$ can be calculated by noting that the total number of free electrons, N, is the volume of the sphere multiplied by the number of states per volume element. Equation (4) requires that each volume element $(2\pi/L)^3$ in **k**-space contain only one state. Taking into account spin degeneracy, which doubles the number of states, the following equation applies:

$$2\left(\frac{4}{3}\pi k_F^3\right)\left(\frac{1}{(2\pi/L)^3}\right) = N = \frac{V}{3\pi^2}k_F^3 \qquad (6)$$

Here, V is simply L^3. Equation (6) can also be expressed in terms of the electron density, n, by dividing both sides by V, where n is N/V. Combining (5) and (6) yields the Fermi energy as a function of the electron density n.

$$E_F = \frac{\hbar^2}{2m}\left(3\pi^2 n\right)^{2/3} \qquad (7)$$

If the Coulomb potential φ from an impurity in a semiconductor is a perturbation of the Fermi energy in (7), then the direct-space expression is

$$E_F - e\phi(\mathbf{r}) = \frac{\hbar^2}{2m}\left[3\pi^2 n(\mathbf{r})\right]^{2/3} \qquad (8)$$

In this case, $e\varphi(\mathbf{r})$ is the potential energy of the impurity and $n(\mathbf{r})$ the total carrier concentration. Replacing the E_F term in

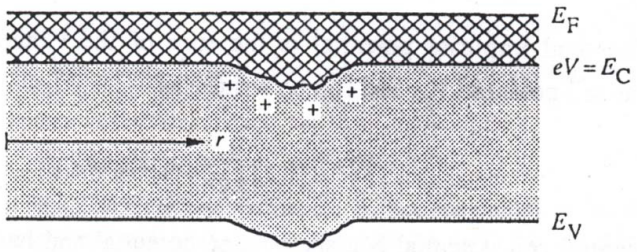

Figure 4. Perturbation of the energy band by impurities which are significantly screened by the presence of mobile carriers in a degenerate semiconductor. (Image from [6].)

(8) with the unperturbed carrier density n_0, the Poisson equation for the total potential at **r** is the following:

$$\nabla^2\phi(\mathbf{r}) = \frac{e}{\varepsilon}\left[n(\mathbf{r}) - n_0\right] \tag{9}$$

In (9), $n(\mathbf{r})$ takes account of the positive charge perturbation. Solving (8) for $n(\mathbf{r})$ and substituting the result into (9) reveals that the differential equation is nonlinear. Nonlinear-screening theory is important, but a general solution to the equation does not exist [6]. Acquiring a simplified result necessitates linearization of $n(\mathbf{r})$ using the binomial expansion and taking only the first two terms.

$$n(\mathbf{r}) = \frac{1}{3\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\left[E_F + e\phi(\mathbf{r})\right]^{3/2} \approx n_0 + n_0\frac{3e\phi(\mathbf{r})}{2E_F} \tag{10}$$

The linearization in (10) enables a solution to (9) under the assumption that the potential $\varphi(\mathbf{r})$ is spherically symmetric.

$$\nabla^2\phi(\mathbf{r}) = \frac{3e^2n_0}{2\varepsilon E_F}\phi(\mathbf{r}) = \frac{e^2\, m(3n_0)^{1/3}}{\varepsilon\pi^{4/3}\hbar^2}\phi(\mathbf{r}) \tag{11}$$

The solution to this Poisson equation is the screened potential, $\varphi(\mathbf{r})$.

$$\phi(\mathbf{r}) = \frac{Ze\,e^{-r/r_s}}{4\pi\varepsilon r} \tag{12}$$

Here, Ze is the charge of the impurity (as an integer multiple Z of the electronic charge e), and r_s is the Thomas-Fermi screening radius:

$$r_{sTF} = \pi^{2/3}\left[\frac{\varepsilon\hbar^2}{e^2 m(3n_0)^{1/3}}\right]^{1/2} \tag{13}$$

In reality, the electron mass m must be the effective mass m^* [6]. Although this expression is the result for degenerate semiconductors, nondegenerate semiconductors warrant a similar analysis. In this case, the screened potential in (12) is still be the result, but the screening length is slightly different.

$$r_{sD} = \left(\frac{\varepsilon k_B T}{e^2 n_0}\right)^{1/2} \tag{14}$$

This is the Debye screening radius.

In the degenerate case, the screening radius has no temperature dependence, but it does depend very weakly on the carrier concentration. In the nondegenerate case, the Debye screening radius does in fact depend on temperature, while also showing a much stronger dependence on carrier concentration compared with the Thomas-Fermi radius.

Figure 5 shows a general comparison of a bare and screened potential. This figure illustrates that for small values of r (less than the screening length), the two potentials are very nearly the same. Beyond the screening length, however, the screened potential deviates from the bare potential in a pronounced manner. This phenomenon occurs because as the radial distance becomes significantly less than the screening length, the exponential becomes nearly unity.

These results show a quantitative, first-order approximation of the characteristics of a screened potential in a semiconductor. Given that these results come from a linearized version of the problem, it is assumed that the potential is slowly varying. For large variations, such as in the case of a Schottky barrier, non-linear analysis of the screening must be used [6], since (10) is no longer a valid approximation.

Impurity Ionization and Critical Concentration

An impurity, with its corresponding potential, can form a hydrogen-like structure when it interacts with a mobile

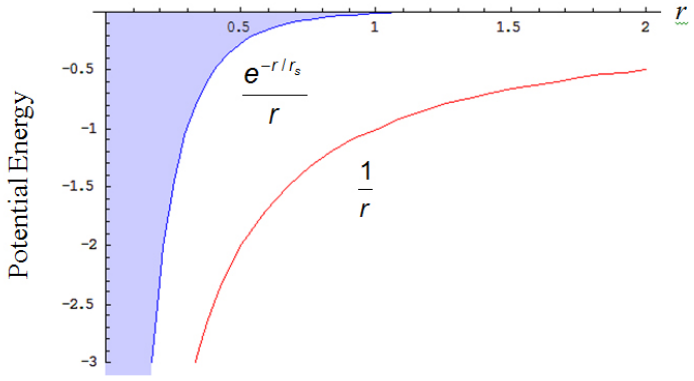

Figure 5. A comparison of the screened (blue) and unscreened (red) Coulomb potentials for a screening length of 0.25.

carrier. In this "hydrogenic" model, the ionization energy associated with, for example, an electron bound to a positively charged donor atom becomes important. Given the fact that the screened potential discussed above is much weaker than the bare potential, the ionization energy of the impurity should be lower. Thus, a donor level is closer to the conduction band, and an acceptor level is closer to the valence band, as Figure 6 illustrates. (Although most of this discussion cites the case of a positive charge screened by free electrons, a negative charge, such as an acceptor atom, screened by free holes is a perfectly legitimate scenario as well. Figure 6 also depicts the decrease in ionization energy for this case)

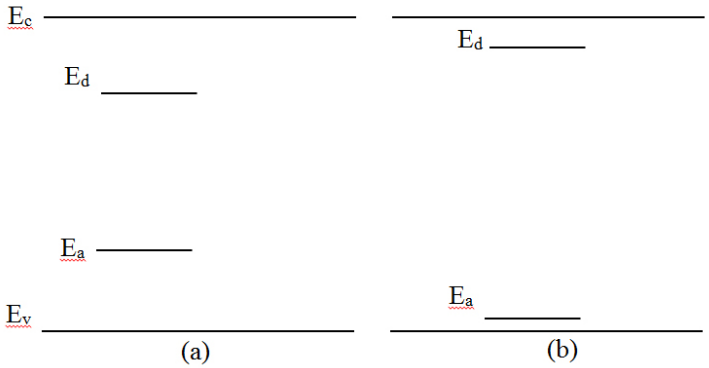

Figure 6. Qualitative donor and acceptor energy levels for bare potentials (a) and screened potentials (b).

As the figure suggests, a smaller ionization energy will tend to result in a higher mobile-carrier density, which in turn will make these carriers available to further screen the potential. This situation leads to an even lower ionization energy, suggesting that at some point the donor levels actually reach the conduction band. This is indeed the case, as the following empirical relation shows [3]:

$$E_d = E_{d0}\left[1-\left(\frac{N_d}{N_{\text{crit}}}\right)^{1/3}\right] \qquad (15)$$

Here, E_{d0} is the donor energy level for low doping, and N_{crit} is the critical doping concentration for the ionization energy to reach zero. For arsenic-doped germanium, N_{crit} is on the order of 10^{17} cm^{-3} [3]. This formulation is essentially a description of the Mott, or insulator-to-metal, transition [6]. As the doping concentration N_d increases and the ionization energy of the donors approaches zero, the semiconductor begins to take on the characteristics of a metal.

Two-Dimensional, Quasi-Two-Dimensional and Nonlinear Screening

The preceding discussion has focused primarily on impurities or defects located well within the semiconductor and has lent itself to a three-dimensional analysis. Also of interest are cases with a two-dimensional geometry. Two-dimensional analysis yields results for screened potentials significantly different from those of three-dimensional analysis. In the limit of large distances from the impurity (much greater than the screening length) in the two-dimensional case, the potential falls off as the inverse of the distance cubed [8]. This decay is much weaker than in the three-dimensional case, where the fall-off is exponential.

Typically, the ideal two-dimensional analysis is inappropriate for phenomena that occur in a three-dimensional world, so quasi-two-dimensional analysis can be more useful. It is a more complicated approach that considers a layer with finite thickness instead of an infinitesimally thin plane [8].

Quasi-two-dimensional analysis becomes useful in the case of inversion layers in semiconductor devices. Inversion layers are nearly two-dimensional sheets of mobile carriers, and potentials located therein must be treated differently than potentials in an equilibrium semiconductor. Furthermore, research has found that the linearization approach of the Thomas-Fermi method discussed above fails to accurately predict the screening of potentials in an inversion layer [9]. As a result, the approximation in (10) must be invalid, just as in the briefly mentioned case of a Schottky barrier, since the electron density varies too quickly with the potential (and vice versa).

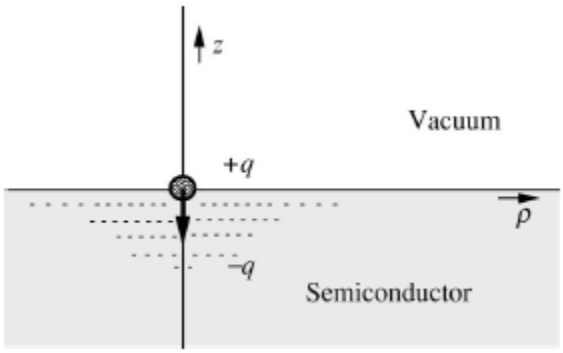

Figure 7. A potential perturbation of charge q at the surface of a semiconductor. (Image from [10].)

Impurities on the surface of a semiconductor cannot be accurately modeled using the assumption of spherical symmetry for obvious reasons. Owing to the geometry of the problem, as Figure 7 shows, the analysis has a degree of two-dimensionality due to the surface, but it cannot be wholly understood except in three dimensions. Research has shown that the radial dependence of a screened potential on the surface, when linear-screening theory is invoked, is far more complex than the simpler case of a po-

tential immersed in the semiconductor [10]. This research determined that the screened potential behaves similarly to (12) within several multiples of the screening distance. Far beyond this, however, it has $1/\rho^3$ dependence.

Also of interest are Coulomb potentials in quantum wells; Figure 8 shows several examples. Semiconductor quantum wells can be modeled as quasi-two-dimensional structures when the width of the well is small. In such cases, electrons with energies lower than the potential barriers on either side are confined to the well and are available to screen the Coulomb potentials that may be present because of an impurity in the well, for example.

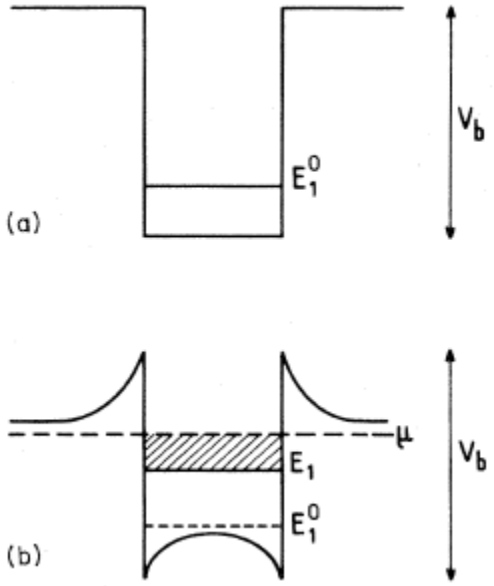

Figure 8. Quantum wells for the case of undoped semiconductors (a) and so-called modulation-doped semiconductors (b), where V_b is the barrier height. (Image from [11].)

Conclusion

This discussion has reviewed some of the basic concepts of Coulomb-potential screening by free carriers. The Thomas-Fermi model is an acceptable basis for a simple semiclassical approach to this problem. To obtain a solution without resorting to extremely complicated analysis, a linearization of the electron concentration with respect to potential energy is helpful. The derivation showed that the screened potential, on these assumptions, behaves exponentially beyond the screening radius, a length that varies in form depending on whether the semiconductor is degenerate or nondegenerate.

As an effect of screening, the ionization energy of a hydrogenic impurity is less than that of the corresponding unscreened impurity. This result led to a process whereby the free-carrier density increased as the ionization energy decreased (and doping increased), which in turn caused greater screening and an even lower ionization energy. This cycle reaches a critical point at which the impurity bands merge with the conduction band: the so-called Mott transition.

The cases of (quasi-)two-dimensional and nonlinear screening applied to quantum wells, inversion layers and Schottky barriers are more complicated. The mathematics in these cases becomes significantly more obtuse than the simpler approaches discussed above, but it is noteworthy nevertheless, as many real problems require a more rigorous approach.

Although all the specific mechanisms are beyond the scope of this discussion, the fact that impurities and other defects, especially charged ones, serve as scattering centers for electrons and holes in the semiconductor implies that screening is phenomenon critical to understanding the electrical and optical properties of the material. A rigorous mathematical approach to screening is highly complicated, but a basic understanding of the concepts necessary to this analysis, as well as some of the peripheral issues, is generally more accessible.

References

[1] C.A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons, Inc., USA, 1989.

[2] C.M. Wolfe, N. Holonyak, Jr. and G.E. Stillman, *Physical Properties of Semiconductors*, Prentice-Hall, Inc., New Jersey, 1989.

[3] J. Singh, *Electronic and Optoelectronic Properties of Semiconductor Structures*, Cambridge University Press, New York, 2003.

[4] D. Redfield, "Electric fields of defects in solids," *Physical Review, Vol. 130*, May 1, 1963.

[5] J.C. Slater and H.M. Krutter, "The Thomas-Fermi method for metals," *Physical Review, Vol. 47*, April 1, 1935.

[6] E.F. Schubert, *Doping in III-V Semiconductors*, Cambridge University Press, Cambridge, U.K., 1993.

[7] C. Kittel, *Introduction to Solid State Physics*, John Wiley & Sons, Inc., New York, 1968.

[8] T. Ando, A.B. Fowler and F. Stern, "Electronic properties of two-dimensional systems," *Reviews of Modern Physics, Vol. 54, No.2*, April, 1982.

[9] O. Hipólito and V.B. Campos, "Screening effects in n-type Si inversion layers," *Physical Review B, Vol. 19, No. 6*, March 15, 1979.

[10] M. Krčmar, W.M. Saslow and M.B. Weimer, "Electrostatic screening near semiconductor surfaces," *Physical Review B, Vol. 61, No. 20*, May 15, 2000.

[11] J.A. Brum, G. Bastard and C. Guillemot, "Screened Coulombic impurity bound states in semiconductor quantum wells," *Physical Review B, Vol. 30, No. 2*, July 15, 1984.