Solid state dielectric screening versus band gap trends and implications

Ram Ravichandran, Alan X. Wang*, John F. Wager

School of EECS, Oregon State University, Corvallis, OR 97331-5501, USA

ABSTRACT

High-frequency (optical) and low-frequency (static) dielectric constant versus band gap trends, as well as index of refraction versus band gap trends are plotted for 107 inorganic semiconductors and insulators. These plots are describable via power-law fitting. Dielectric screening trends that emerge from this analysis have important optical and electronic implications. For example, barrier lowering during Schottky emission, phonon-assisted or Fowler-Nordheim tunneling, or Frenkel-Poole emission from a trap is found to be significantly more pronounced with increasing band gap due to a reduction in the optical dielectric constant with increasing band gap. The decrease in the interface state density with increasing band gap is another optical dielectric constant trend. The tendency for a material with a wider band gap to be more difficult to dope is attributed to an increase in the ionization energy of the donor or acceptor dopant, which in turn, depends on the optical dielectric constant and the effective mass. Since the effective mass for holes is almost always larger than that for electrons, p-type doping is more challenging than n-type doping in a wide band gap material. Finally, the polar optical phonon-limited mobility depends critically upon the reciprocal difference of the optical and the static dielectric constant. Consequently, electron and hole mobility tend to decrease with increasing band gap in a polar material.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Dielectric constant is a fundamental property of a solid-state material. It accounts for a reduction (screening) of the electric field within a material due to an internal rearrangement of charge (polarization) upon application of an external electric field. Screening is often classified as either static or dynamic, as characterized by a low-frequency (static) or a high-frequency (optical) relative dielectric constant denoted herein as \( \varepsilon_{SR} \) or \( \varepsilon_{\infty R} \) respectively. Many optical and electronic phenomena intimately depend upon dielectric screening.

In the development of our atomic solid-state energy (SSE) scale [1–3], we have discovered several interesting dielectric constant versus band gap trends exhibited by a large number of semiconductors and insulators in the SSE data base [2]. The goal of the work reported herein is to (i) discuss these trends by data mining the SSE data base, and (ii) explore how these trends can be gainfully employed in optical and electronic assessment of solid state semiconductor and insulator materials.

2. Index of refraction trends

Formulation of refractive index versus band gap (n-E\(_G\)) relationships has a long and rich history [4–12], as has been periodically reviewed [13–15]. Table 1 highlights four proposed two-parameter n-E\(_G\) relations. These relations are used to fit 107 n-E\(_G\) pairs obtained from the SSE data base (see Supporting Information), as shown in Fig. 1. Using the SSE dielectric constant data base, we approximate the index of refraction using \( n = \left[ \varepsilon_{\infty R} \right]^{1/2} \) instead of employing the more precise expression \( n = \left[ \varepsilon_{\infty R} + \kappa^2 \right]^{1/2} \), where \( \kappa \) is the extinction coefficient. This approximation is of questionable viability only for small band gap semiconductors. However, we find that the n-E\(_G\) regression fitting parameters are almost identical regardless of whether or not the extinction coefficient correction is employed.

Returning to Fig. 1, it is clear that the power-law, hyperbolic, and logarithmic models are all capable of reasonably describing the n-E\(_G\) trend, while the linear model is a poor descriptor of the data. From a coefficient of determination (R\(^2\)) perspective, the power-law
model ($R^2 = 0.90$) is slightly better than the hyperbolic model ($R^2 = 0.89$) and is somewhat better than the logarithmic model ($R^2 = 0.82$), at least for the data set used to construct Fig. 1.

In selecting between use of the power-law or the hyperbolic $n$-$E_G$ model, a compelling attribute of the hyperbolic model is its physical basis since it derives from the Penn model [8,9,16]. Penn asserts that the dielectric properties of a semiconductor or insulator can be envisaged as arising from two isotropic bands (valence and conduction) separated in energy by an average band gap, $E_{\text{Penn}}$, such that

$$n = \sqrt{\varepsilon_{\infty R}} = \sqrt{1 + \left(\frac{\hbar \omega_p}{E_{\text{Penn}}}\right)^2} \quad (1)$$

where $\hbar \omega_p$ is the valence electron plasma energy. Comparing Eq. (1) to the hyperbolic model included in Table 1, $a = \hbar \omega_p = 15.2 \text{ eV}$, and $b = E_{\text{Penn}} - E_G = 3.9 \text{ eV}$. Moreover,

$$\omega_p = \sqrt{\frac{q^2 N_V}{m_0 \varepsilon_0}} \quad (2)$$

where $q$ is electronic charge, $N_V$ is the valence electron density, $m_0$ is the electron rest mass, and $\varepsilon_0$ is the free space dielectric constant. Using Eq. (2) in conjunction with $\hbar \omega_p = 15.2 \text{ eV}$ yields $N_V = 1.7 \times 10^{23} \text{ cm}^{-3}$.

Thus, the hyperbolic fit to the $n$-$E_G$ data shown in Fig. 1 and summarized in Table 1 indicates that the $n$-$E_G$ trend is consistent with a simple two-band Penn model in which $N_V = 1.7 \times 10^{23} \text{ cm}^{-3}$ and $E_{\text{Penn}} - E_G = 3.9 \text{ eV}$. If $E_{\text{Penn}} - E_G$ is equally partitioned between the valence and conduction bands, then the Penn gap is positioned ~1.95 eV below or above the valence band maximum or conduction band minimum, respectively. This constant positioning trend suggests that the Penn gap is established by energies sufficiently removed from the band edges that an appreciable density of states is reached so that an adequately large joint density of states is achieved for strong optical absorption. This interpretation differs somewhat from the perspective of Penn and others [17] who consider the Penn gap to be an average band gap essentially establishing the centroid of the valence and conduction band joint density of states.

Although the physical foundation upon which the hyperbolic model rests makes it an attractive $n$-$E_G$ data fitting model option, we typically employ the power-law model to describe $n$-$E_G$ and other trends in the remainder of this contribution. We do this since (i) the power-law model provides an accurate description of $n$-$E_G$ and other trends considered herein, and (ii) power-law fitting can be accomplished directly without having to precondition the data for linear regression (e.g., linear regression fitting of the index of refraction to the Penn model requires plotting $[n^2 - 1]^{0.5}$ versus $E_C$).

### 3. Dielectric constant trends and implications

A power-law fit to a plot of the high-frequency (optical) relative dielectric constant as a function of band gap ($\varepsilon_{\infty R}$-$E_G$) is given in Fig. 2. Since the index of refraction, $n$, is taken to be equal to the square root of $\varepsilon_{\infty R}$, the power-law fit included in Fig. 1 is simply a square-root version of Fig. 2. Thus, $R^2 = 0.90$ for both of the power-law fits shown in Figs. 1 and 2. The insert included in Fig. 2 displays a log-log plot of the data. A straight-line fit to a log-log plot is supporting evidence for the viability of the power-law. Additionally, a log-log plot is useful for visualizing the variability of the data.

A power-law fit to a plot of the low-frequency (static) relative dielectric constant as a function of band gap ($\varepsilon_{SR}$-$E_G$) is indicated in Fig. 3. The much smaller $R^2 = 0.48$ in Fig. 3 compared to $R^2 = 0.90$ in Fig. 2 reveals that $\varepsilon_{SR}$ exhibits much more variability with respect to $E_G$ than does $\varepsilon_{\infty R}$. Nevertheless, Fig. 3 demonstrates a clear tendency for $\varepsilon_{SR}$ to decrease with increasing $E_G$. Comparing the log-log plot insert of Fig. 3 to the insert of Fig. 2 confirms that the variability of $\varepsilon_{SR}$ is indeed much greater than the variability of $\varepsilon_{\infty R}$.

Next we discuss implications of these $\varepsilon_{SR}$-$E_G$ and $\varepsilon_{\infty R}$-$E_G$ trends. We begin with barrier lowering. A variety of electronic phenomena involve the reduction of an energetic barrier upon application of an electric field (see insert of Fig. 4 for an example involving electron injection from a metal into the conduction band of an insulator).
The extent of barrier lowering is quantitatively given by

$$\Delta E_{\text{BARRIER}} = \sqrt{\frac{q^2}{\pi \varepsilon_\infty}}$$

where $\xi$ is the electric field and $\varepsilon_\infty$ is the high-frequency (optical) dielectric constant. A plot of barrier lowering energy versus band gap ($\Delta E_{\text{BARRIER}}$-$E_g$), as obtained from Eq. (3), is shown in Fig. 4 for an applied electric field of 1 MV/cm. Several things are worth noting with respect to Fig. 4. First, barrier lowering is more pronounced at large band gaps because dynamic screening (i.e., $\varepsilon_\infty R$) decreases with increasing band gap. Second, the high-frequency (optical) dielectric constant should be used to model screening associated with barrier lowering since the time associated with an electron transiting through the barrier is very short [18]; this is often not recognized in the literature. Third, the 1 MV/cm electric field assumed in Fig. 4 is an excessively large field for narrow band gap materials, and is thus more pertinent to the case of wider band gap insulators. Fourth, barrier-lowering considerations may be applicable to Schottky emission over a barrier [18,19], phonon-assisted tunneling or Fowler-Nordheim tunneling through a barrier [20], or Frenkel-Poole emission from a trap [18,19].

The interface parameter, $S$, is a unitless figure-of-merit bounded by zero and one, i.e., $0 \leq S \leq 1$, that is often used in the assessment of the electronic properties of a semiconductor or an insulator surface or interface [21,22]. The limit $S = 1$ corresponds to an ideal case in which the surface state density is zero, whereas $S = 0$ implies that the interface state density is infinitely large. $S$ is often estimated using [21].

$$S = \frac{1}{1 + 0.1(\varepsilon_\infty R - 1)^2}$$

where $\varepsilon_\infty R$ is the high-frequency (optical) relative dielectric constant. A plot of interface parameter versus band gap ($S-E_g$), as obtained using Eq. (4), is shown in Fig. 5. The trend is quite clear; a narrow band gap semiconductor tends to be strongly non-ideal, possessing a very small interface parameter. In contrast, a very wide band gap insulator is nearly ideal, with $S$ approaching 1.

Induced gap state theory asserts that the minimum surface state density for a semiconductor or insulator is [22].

$$D_{\text{SSmin}} = \frac{C_i}{q} \left( \frac{1}{S} - 1 \right)$$

where $C_i$ is the dipole capacitance density, given by

$$C_i = \frac{\varepsilon_\infty d}{\delta}$$

where $\varepsilon_\infty$ is the high-frequency (optical) dielectric constant and $\delta$ is a dipole thickness, typically assumed to be $\delta = 0.4$ nm. A plot of minimum surface state density versus band gap ($D_{\text{SSmin}}$-$E_g$), as obtained using Eqs. (5) and (6), is shown in Fig. 6. There is a clear tendency for the surface state density to decrease with increasing band gap. This trend is consistent with the interface parameter trend witnessed in Fig. 5.

Extrinsic doping refers to the introduction of substitutional impurities in order to modulate the carrier concentration of a semiconductor. Donor impurities provide n-type doping, increasing the concentration of electrons in the conduction band. Acceptor impurities result in p-type doping, increasing the concentration of holes in the valence band. The ionization energy of a dopant, $E_i$, is the energy separating the dopant in its ground state from its relevant band edge, i.e., the conduction band minimum or the valence band maximum for a donor or acceptor, respectively. Typically the ionization energy of a dopant is estimated using the hydrogen-atom model [19].

$$E_i = \frac{m^*}{(\varepsilon_\infty R)^2} 13.6 \text{ eV}$$

where $m^*$ is the conductivity relative effective mass and $\varepsilon_\infty R$ is the high-frequency (optical) relative dielectric constant (both are
Effective mass versus band gap (m*-E_g) trends for selected semiconductors and insulators are shown in Fig. 7. Plots of dopant ionization energy versus band gap (E_I-E_G) trends, obtained using Eq. (7) and the effective mass fits given in Fig. 7, are shown in Fig. 8 for electron and hole doping. The hydrogen-atom model predicts that the ionization energy increases with increasing band gap for both n- and p-type doping. However, since hole effective masses tend to be larger than electron effective masses, E_I increases more abruptly with increasing E_G for hole (acceptor) doping than for electron (donor) doping. Thus, it is typically more difficult to accomplish p-type doping than n-type doping in a wide band gap material. Doping is challenging due to carrier freeze-out in which the ionization energy is so large that most dopants remain un-ionized at room temperature. The red shading in Fig. 8 identifies the approximate range of band gap above which room temperature carrier freeze-out is likely to become problematic. These hydrogen-atom-model-based E_I-E_G trends suggest that donor and acceptor doping will be challenging when E_C > ~3 eV and E_C > ~2 eV, respectively. It is impossible to establish a more precise band gap doping limit due to variability in m* and ε∞R. Note that this large ionization energy, room temperature carrier freeze-out explanation for why doping in a wide band gap material is difficult differs from the explanation normally put forth involving self-compensation [23–27]. Basically, the self-compensation mechanism asserts that for a wide band gap material it is often energetically more favorable to create a self-compensating defect than to modulate the Fermi level closer to the relevant band edge.

All of the dielectric constant implication examples considered up to now have involved dynamic screening since the high-frequency (optical) dielectric constant, ε∞, is employed. The low-frequency (static) dielectric constant, ε_S, is relevant to this next example. Debye length is a measure of the physical extent of dielectric screening [19]. The extrinsic Debye length is given by

\[
L_{De} = \sqrt{\frac{\varepsilon_S k_B T}{q^2 N}}
\]

where N is the carrier concentration induced by extrinsic doping. Fig. 9 shows that L_{De} decreases with increasing E_C. This trend is a simple consequence of decreasing static dielectric screening with increasing E_C, as evident from Fig. 3. However, a bit of reflection reveals that much of the trend shown in Fig. 9 is incorrect, or at least very misleading. Recall from our discussion of doping ionization energy trends that E_I becomes so large at higher E_C’s that it is impossible to dope a wide band gap material. Thus, it is improper to invoke an extrinsic Debye length for wide band gap materials that are undopable.

An alternative approach is to use the intrinsic Debye length,
Intrinsic carrier concentration versus band gap ($n_i - E_g$) plot for narrow band gap semiconductors. This type of free carrier screening is likely to dominate for semiconductors with band gaps between $\sim 0.5$ eV to $\sim 3$ eV, as shown in Fig. 10b. The Debye length pertains to free carrier screening. Intrinsic free carrier screening is the same type of screening that occurs in a metal. Perfect dielectric screening occurs in a metal since the rearrangement of free carriers allows the electric field to be excluded from the interior of the metal. As the band gap increases from zero, the intrinsic carrier concentration decreases exponentially with band gap. Thus, dielectric screening is no longer perfect in a narrow band gap semiconductor. When the semiconductor band gap exceeds $\sim 0.5$ eV, the intrinsic carrier concentration is so small that free carrier screening is no longer relevant. However, free carrier screening continues to be important in larger band gap materials as long as carriers (electrons or holes) are available via doping. This type of free carrier screening is likely to dominate for semiconductors with band gaps between $\sim 0.5$ eV to $\sim 3$ eV, beyond which a material is difficult to dope. However, this range of band gap is somewhat anomalous since the electric field is not constant across extrinsic semiconductors due to the possibility of accumulation, depletion, and/or inversion. Finally, for materials with band gaps greater than $\sim 3$ eV, dielectric screening is no longer due to free carrier screening. Rather, wide band gap dielectric screening is exclusively due to polarization of the lattice.

Figure 9. Extrinsic Debye length versus band gap ($L_{De} - E_g$) plot at room temperature and for a carrier concentration of $10^{16}$ cm$^{-3}$ for 107 inorganic semiconductors and insulators. Insert is a log-log plot.

\begin{equation}
L_{Di} = \sqrt{\frac{\epsilon_0 k_B T}{q^2 n_i}}
\end{equation}

where $n_i$ is the intrinsic carrier concentration, given by Ref. [19].

\begin{equation}
n_i = \sqrt{N_C N_V} e^{-\frac{E_g}{k_B T}}
\end{equation}

where $N_C$ and $N_V$ are the conduction and valence band effective density of states, respectively. If the conduction and valence band density of states effective masses are both assumed to be equal to 0.5, then $N_C = N_V = 8.9 \times 10^{18}$ cm$^{-3}$ at room temperature. Making this assumption allows us to calculate $L_{Di}$ as a function of $E_g$, as indicated in Fig. 10a. Only small band gap semiconductors are considered in Fig 10a since $L_{Di}$ is extraordinarily large for larger band gaps. This means that intrinsic carrier screening is not relevant for materials with band gaps greater than $\sim 0.5$ eV. The reason for this is that the intrinsic carrier concentration becomes negligible when $E_g > -0.5$ eV, as shown in Fig. 10b.

The final dielectric constant implication example involves both $\epsilon_\infty$ and $\epsilon_S$. Non-elemental semiconductors and insulators are polar materials. Electron or hole mobility in a polar solid tends to be dominated by polar optical phonon (POP) scattering [28,29]. The POP mobility is determined from the momentum relaxation time [28].

\begin{equation}
\tau_m(E) = \left\{ \frac{q^2 \omega_0}{4 \pi \varepsilon_0 h \sqrt{2/E/m^*}} \right\} \left[ N_0 \sqrt{1 + \frac{h\omega_0}{E}} + (N_0 + 1) \sqrt{1 - \frac{h\omega_0}{E}} \right]^{-1} \left( \frac{E}{h\omega_0} \right) \frac{h\omega_0 N_0}{E} \sinh^{-1} \left( \frac{E}{h\omega_0} \right)^{1/2} + h\omega_0 (N_0 + 1) \sinh^{-1} \left( \frac{E}{h\omega_0} - 1 \right)^{1/2} \right\}^{-1}
\end{equation}

where $q$ is electronic charge, $\omega_0$ is optical phonon angular frequency, $\epsilon_\infty$ is high-frequency (optical) dielectric constant, $\epsilon_S$ is low-frequency (static) dielectric constant, $m^*$ is conductivity effective mass, and $N_0$ is the phonon density as calculated using Bose-Einstein statistics. For a non-degenerate semiconductor or insulator with spherical constant energy surfaces, the average momentum relaxation time is given by Ref. [30].
\[
\tau_m = \frac{\int_0^\infty \tau_m E^{3/2} \exp[-E/k_B T] dE}{\int_0^\infty E^{3/2} \exp[-E/k_B T] dE}
\]

(12)

where \(k_B\) is Boltzmann’s constant. Finally, the POP mobility is obtained via

\[
\mu_{\text{POP}} = \frac{q \tau_m}{m^*}
\]

(13)

Note that four physical properties must be specified in order to calculate \(\mu_{\text{POP}}\), i.e., \(\varepsilon_\infty, \varepsilon_S, u_0,\) and \(m^*\). Using Eqs. (11)–(13), the POP mobility is calculated as a function of band gap for 35 n-type and 30 p-type semiconductors and insulators (see Supplementary Information), as shown in Fig. 11. The POP mobility clearly decreases with increasing band gap, for both electrons and holes. Much of this mobility degradation with respect to increasing band gap appears to arise from the term in Eq. (11) involving the reciprocal difference of the optical and static dielectric constants. In an attempt to quantify this tendency, a polar optical phonon (POP) coupling parameter can be defined [29].

4. Conclusions

High-frequency (optical) and low-frequency (static) dielectric constants and the index of refraction all tend to decrease in approximately a power-law fashion with increasing band gap. These dielectric screening trends have important implications with respect to field-induced barrier lowering, interface state density, impurity doping, Debye length, and polar optical phonon scattering-limited mobility.

Fig. 11. Polar optical phonon mobility versus band gap (\(\mu_{\text{POP}}-E_G\)) plots for (a) electrons and (b) holes.

Fig. 12. Polar optical phonon mobility versus polar optical phonon coupling parameter (\(\mu_{\text{POP}}-\mu_{\text{POP}}\)) plots for (a) electrons and (b) holes.

Acknowledgments

The work of R.R. and J.F.W. was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract no. DE-AC36-08GO28308 and the Center for Inverse Design, an Energy Frontier Research Center.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.optmat.2016.07.027.

References