Understanding low-temperature bulk transport in samarium hexaboride without relying on in-gap bulk states

A. Rakoski, Y. S. Eo, K. Sun, and Ç. Kurdak

Phys. Rev. B 95, 195133 — Published 17 May 2017

DOI: 10.1103/PhysRevB.95.195133
Understanding low-temperature bulk transport in samarium hexaboride without relying on in-gap bulk states

A. Rakoski,* Y.S. Eo, K. Sun, and Ç. Kurdak
University of Michigan

We present a new model to explain the difference between the transport and spectroscopy gaps in samarium hexaboride (SmB\textsubscript{6}), which has been a mystery for some time. We propose that SmB\textsubscript{6} can be modeled as an intrinsic semiconductor with an accumulation length that diverges at cryogenic temperatures. In this model, we find a self-consistent solution to Poisson’s equation in the bulk, with boundary conditions based on Fermi energy pinning due to surface charges. The solution yields band bending in the bulk; this explains the difference between the two gaps because spectroscopic methods measure the gap near the surface, while transport measures the average over the bulk. We also connect the model to transport parameters, including the Hall coefficient and thermopower, using semiclassical transport theory. The divergence of the accumulation length additionally explains the 10-12 K feature in data for these parameters, demonstrating a crossover from bulk dominated transport above this temperature to surface-dominated transport below this temperature. We find good agreement between our model and a collection of transport data from 4-40 K. This model can also be generalized to materials with similar band structure.

I. INTRODUCTION

Samarium hexaboride (SmB\textsubscript{6}) has long eluded classification due to its unique properties, beginning with its unusual temperature-resistivity curve.\textsuperscript{1} Past research has led to the classification of SmB\textsubscript{6} as a rare-earth mixed valence compound; although it is homogeneous, the Sm ions have mixed valence of 2+ and 3+.\textsuperscript{1–3} SmB\textsubscript{6} has a simple cubic structure, and it is also free of both magnetic and structural phase transitions,\textsuperscript{4} which makes it a good candidate for studying homogenous mixed valence. One of the most significant developments was the discovery that SmB\textsubscript{6} is a Kondo insulator.\textsuperscript{4} Kondo insulators are characterized by the opening of a small gap at the Fermi energy due to hybridization between f-electrons and conduction electrons.\textsuperscript{5,6} In SmB\textsubscript{6} specifically, there are three 4f bands, one of which hybridizes with the 5d conduction electrons, opening a gap between the two hybrid bands.\textsuperscript{7}

Although the presence of the hybridization gap suggests that SmB\textsubscript{6} should exhibit insulating behavior, detailed transport results over a 40-year period suggest a much more complicated picture.\textsuperscript{1,8–12} High-quality SmB\textsubscript{6} has consistently demonstrated activated behavior from 4-40 K. Around 4 K, the activated behavior is always terminated by an unknown conduction mechanism, leading to a plateau in the temperature-resistivity curve. Many attempts have been made to explain this crossover to conductive behavior in terms of bulk effects, with a commonly accepted picture based on impurities in the material.\textsuperscript{1} A breakthrough came from the prediction that Kondo insulators can additionally be topological insulators,\textsuperscript{13–15} a class of materials that undergo a crossover at low temperature from a conventional state to a bulk insulating state with topologically protected metallic surface states.\textsuperscript{16} Indeed, recent transport\textsuperscript{12,17} and angle-resolved photoemission spectroscopy (ARPES) results\textsuperscript{18–20} have demonstrated evidence for these topologically protected metallic surface states in SmB\textsubscript{6} with a crossover temperature of approximately 4 K. In this paper, we will concentrate on temperatures from 4-40 K, where bulk transport is dominant. Although activated behavior is consistently observed in this temperature range, transport measurements including resistivity and Hall coefficient also demonstrate a feature at about 10 K.

Before discussing some of the unusual properties of SmB\textsubscript{6} that arise from its small gap and hybridized bands, it is instructive to discuss bulk transport in standard TIs. Such “standard” TIs (for example, Bi\textsubscript{2}Se\textsubscript{3} or Bi\textsubscript{2}Te\textsubscript{3}) are characterized by a bulk band gap as in conventional semiconductors, and after undergoing the crossover to the topological phase, there are an odd number of surface states located in the gap.\textsuperscript{21,22} If the topological surface states are not considered, all standard TIs can be treated as semiconductors. In this picture, charge neutrality must be enforced, so the Fermi energy (E\textsubscript{F}) is initially expected to be exactly halfway between the top of the valence band and the bottom of the conduction band. As in semiconductors, impurity states may also be present in the gap. For an n-type material, donor states would be in the gap near the conduction band, and for a p-type material, acceptor states would be in the gap near the valence band. Because charge neutrality must also be enforced, the presence of these extra states shifts E\textsubscript{F} towards the conduction band for donor states and towards the valence band for acceptor states.\textsuperscript{23}

Impurity states in semiconductors and standard TIs can be treated quantitatively using the effective mass approximation.\textsuperscript{24} In this picture, impurities are assumed to be hydrogenic, but with the substitution of effective mass for electron mass (m \rightarrow m*) and dielectric constant for vacuum permittivity (\varepsilon_0 \rightarrow \kappa\varepsilon_0). The results are an effective Bohr radius

\[ a_B^* = \frac{4\pi\varepsilon_0\hbar^2}{m*e^2} = \frac{\kappa}{m*/m} \text{(0.53 Å)} \]  \hspace{1cm} (1)
and an effective ground-state energy

\[ E^* = -\frac{m^* e^4}{2(4\pi \kappa \varepsilon_0)^2 \hbar^2} = -\frac{m^*}{\kappa^2} (13.6 \text{ eV}). \tag{2} \]

Once these parameters have been calculated, the initial assumption of hydrogenic impurities can be verified. In a donor state, for example, the extra electron must have an extent much larger than one lattice constant for the donor to be hydrogenic. If this were not the case, contributions from the donor itself would also have to be considered, and the hydrogen model could not be used. Therefore, when comparing the effective Bohr radius to the lattice constant, we must satisfy the condition \( a_B^* \gg a \) for the approximation to be valid.

The effective mass approximation has been used successfully in all standard TIs. Standard TIs exhibit residual bulk conduction after undergoing the crossover to the topological state, and this is well-understood as arising from impurity states that can be treated using the effective mass approximation. Many researchers have applied the idea of impurity states in the gap to SmB\(_6\) to try to understand the plateau at 4 K before it was thought to be a TI,\(^1\) and later to explain experimental discrepancies in the size of the Kondo hybridization gap. While the presence of the gap is well-known, transport and spectroscopic methods disagree on the size. Transport measurements, which probe the energy difference between \( E_F \) and the conduction band (the activation energy), report 3-4 meV.\(^9\)\(^12\)\(^17\) In analogy with semiconductors and standard TIs, it is expected that \( E_F \) is exactly halfway between the valence and conduction bands, suggesting that the total transport gap is 6-8 meV. On the other hand, spectroscopy and tunneling experiments measure the full gap near the surface, and they report 16-20 meV.\(^25\)\(^31\)

This discrepancy has been interpreted as arising from the presence of in-gap bulk impurity states.\(^7\)\(^25\)\(^26\)\(^28\)\(^30\)\(^32\)\(^34\) In this scenario, transport would measure the difference between the impurity state and the conduction band, yielding an incorrect result for the total gap. Since the effective mass approximation is usually so successful, it has been applied to SmB\(_6\) to understand this proposed in-gap impurity state. However, reported values\(^26\)\(^30\) of the dielectric constant \( \kappa \) range from 600-1500, and as we have seen, activation energy (which can be used to obtain effective mass) ranges from 3-10 meV. Using these values with Eqs. 1 and 2, we obtain a minimum \( a_B^* \) of 0.5 Å and a maximum \( a_B^* \) of 4 Å. Reports of the effective Bohr radius are usually in this range; for example, Sluchanko et al. report 3 Å.\(^32\) Additionally, most reports agree that the lattice constant \( a \) of SmB\(_6\) is about 4.13 Å, so we find that the condition \( a_B^* \gg a \) required to verify the effective mass approximation is not satisfied anywhere in the range of Bohr radii that can be calculated. Because the effective mass approximation fails, in-gap hydrogenic impurity states in SmB\(_6\) are not justified.

Since semiconductor theory and the effective mass approximation are successful in most cases, this result is startling. However, upon closer examination, we find that it is perhaps not completely unexpected. Because the gap in SmB\(_6\) arises due to hybridization, its band structure is very different from that of a conventional semiconductor. Unlike a semiconductor, SmB\(_6\) has a non-quadratic and asymmetric dispersion, because its band structure arises due to Kondo hybridization. Both the valence and conduction bands have nearly flat regions characterized by the localized \( f \) states as well as low-mass regions characterized by the \( d \) states. Because of this unusual composition, while the gap and band structure effects in SmB\(_6\) arise based on contributions from all the carriers, transport is dominated only by the low-mass carriers. Additionally, both bands have positive curvature, unlike in a semiconductor, where only the conduction band has positive curvature. As we will see, this has a significant effect on how we understand transport. The gap is also much smaller than that of standard semiconductors or TIs. Because of these differences in the band structure, we will see that SmB\(_6\) must be treated much more carefully than standard gapped materials. (In contrast, other hexaboride materials can be treated as standard gapped materials, and in these cases, impurity states are present within the bulk gap.\(^35\))

In the context of in-gap bulk states, we can gain insight into SmB\(_6\) by analogy with superconductors. When a material undergoes a transition to a superconducting state, some of the electrons near the Fermi energy condense into Cooper pairs. Formation of the condensate opens up a gap at the Fermi energy.\(^36\) Even though this gap is so small, tunneling measurements\(^37\) have shown that the addition of impurities to a superconductor does not destroy superconductivity (until the impurity concentration becomes sufficiently high). This suggests that the impurity states are not in the gap, or that superconductors exhibit a small and clean gap and are not vulnerable to impurity conduction. Although the mechanism for gap formation (the Kondo effect) is very different in SmB\(_6\), the gap is also much smaller than a semiconductor bulk gap. This, combined with the failure of the effective mass approximation, suggests that SmB\(_6\) should have a clean gap that is not vulnerable to impurity conduction at low temperatures. In fact, SmB\(_6\) does not exhibit residual bulk conduction experimentally, and this can be taken as evidence for a clean bulk gap.\(^12\)

Another interpretation of the gap discrepancy is that spectroscopy measures the direct gap while transport measures an indirect gap that forms during hybridization. Theoretical treatments of Kondo hybridization predict the presence of both a direct and an indirect gap,\(^6\)\(^38\)\(^39\) and researchers have also used this idea to explain the gap discrepancy.\(^11\)\(^31\) This interpretation explains the gap discrepancy while avoiding the problem of the in-gap impurity states, but we will propose an alternative explanation that is consistent with features observed in transport. As we have seen, in-gap impurity states in SmB\(_6\) are not justified by the effective mass approximation, nor are they consistent with the observation...
of no residual bulk conduction after the surface states become relevant. To achieve consistency, we propose another explanation of the gap discrepancy that does not rely on in-gap states and instead allows for a clean direct gap.

II. SIMPLIFIED DENSITY OF STATES AND BAND BENDING CALCULATIONS

A. Density of states

The dispersion in SmB$_6$ is well-known based on the success of recent high-resolution ARPES measurements.\textsuperscript{18–20} The gap forms at low temperature, when the conduction band (5$d$) hybridizes with localized states (4$f$). In SmB$_6$, there are three 4$f$ bands, and ARPES indicates that only one band participates in hybridization. A sketch of the band structure of SmB$_6$ before hybridization is shown in Fig. 1(a), and the hybridized band structure is shown in Fig. 1(b), both along the $\Gamma$ – $X$ – $\Gamma$ direction. SmB$_6$ is an insulator, so the Fermi energy is located in the gap in the hybridized band structure. As can be seen in the figure, the two unhybridized 4$f$ bands are still present.

In addition to this basic structure, the hybridized dispersion has some subtle features, which can also be observed by ARPES. For example, ARPES demonstrates an indirect gap with a valence peak about 15 meV below the Fermi energy at the $H$ point.\textsuperscript{18} This feature can also be observed at nearby energies due to intrinsic and thermal broadening. Compared to the main features of the dispersion, however, this feature is small and close to the valence band. Although ARPES cannot probe far into the conduction band at the temperatures we are considering, there are likely some similarly small features present in the conduction band. We refer to the regions in which such small features exist as the “region of non-parabolicity.”

In our model, we will use a dispersion that is simplified considerably from the actual dispersion. We neglect the small features in the region of non-parabolicity, such as the feature at the $H$-point and any similar features in the conduction band. To do this, we approximate the band structure using a piecewise function, as shown in Fig. 1(c). Here, the flat regions approximate the pieces of the hybridized dispersion that primarily come from the 4$f$ band, which we refer to as “$f$-like” states. The linear regions approximate the pieces of the hybridized dispersion that primarily come from the 5$d$ band, which we call “$d$-like” states. Additionally, the two unhybridized 4$f$ bands cannot be resolved separately from the valence band by ARPES, so we approximate them to be at the top of the valence band. Making these approximations introduces some error into the model, but the features in the region of non-parabolicity are small, so the error is not more than a few meV.

From this dispersion, we can calculate a simplified density of states (DOS). The DOS corresponding to our simplified dispersion (Fig. 1(c)) is shown in the lower inset of Fig. 2. In this figure, the peaks in the DOS correspond to the $f$-like regions of the dispersion, and these can be estimated from the size of the pockets in the SmB$_6$ Brillouin zone (BZ), shown in the upper inset of Fig. 2. The flat parts of the DOS correspond to the $d$-like regions of the dispersion. In the range $E_{\text{gap}}$, the DOS is zero, and the gap changes with temperature. Data for the gap as a function of temperature\textsuperscript{31} is shown in Fig. 2, as is the fit to this data that was used in the calculation (dashed line).

We can represent the simplified DOS using delta functions ($\delta$) for the $f$-like states and step functions ($\theta$) for
the $d$-like states:
\[
g(\epsilon) = N_{cf} \delta(\epsilon - E_C) + g_{cd} \theta(\epsilon - E_C) \\
+ N_{sf} \delta(\epsilon - E_V) - g_{sd} \theta(\epsilon - E_V) + 1
\]  
(3)

where $E_C$ is the edge of the conduction band, $E_V$ is the edge of the valence band, $N$ is a density in cm$^{-3}$, $g$ is a DOS in cm$^{-3}$ eV$^{-1}$, and the Fermi energy $E_F$ has been set to 0. The subscripts on the four factors refer to electron type and band; e.g. $N_{cf}$ is the density of states for $f$-like electrons in the conduction band ($c$). The Fermi energy, which is in the gap, is defined to be zero.

The sizes of these four terms can be estimated using ARPES data.\(^{40}\) According to this measurement, the hybridized $f$ band BZ has six half-ellipsoid Fermi pockets (upper inset of Fig. 2), and the total volume of these yields the number of filled states in the conduction band, which we will denote $n_{\text{cf}}$. This implies that
\[
N_{cf} = 3n_{\text{cf}} = 9.0 \times 10^{21} \text{ cm}^{-3}.
\]  
(4)

The remaining volume in that BZ, plus the total volume of the BZs for the two unhybridized $f$ bands, yields the valence band contribution. We will denote the volume of the BZ as $n_{\text{BZ}}$, so this implies that
\[
N_{sf} = 3n_{\text{BZ}} - n_{\text{cf}} = 7.6 \times 10^{22} \text{ cm}^{-3}.
\]  
(5)

The $d$-like states can be calculated from data above the hybridization temperature.\(^{40}\) We approximate the dispersion to be quadratic and use the usual result for a 3D quadratic DOS,
\[
g_{3D}(\epsilon) = \frac{m^*}{2\pi^2\hbar^2} \sqrt{2m^*} \epsilon
\]  
(6)

In our simplified DOS we can approximate the $d$ bands on both sides of the gap as constant. Specifically, we approximate this constant to be $g_{\text{BZ}}(E_F)$, because $E_F$ is in the gap and the gap is small. This also means that this value in both bands is about the same constant,
\[
g_{cd} \approx g_{sd} \approx g_{3D}(E_F) = 1.8 \times 10^{19} \text{ cm}^{-3} \text{ eV}^{-1}.
\]  
(7)

We note that this term is 2-3 orders of magnitude smaller than the terms in Eqs. 4 and 5. We will define this value as $g_0$ for brevity.

**B. Band structure calculation**

As we have seen, the actual and simplified dispersions as well as the DOS of SmB$_6$ can be characterized by two types of carriers. The flat regions are dominated by $f$-like carriers, and the remainder is dominated by $d$-like carriers. In this section, we will outline a self-consistent calculation used to obtain the band structure. For such a calculation, we must take all the charges into account. However, the $f$-like terms (Eqs. 4 and 5) are 2-3 orders of magnitude greater than the $d$-like coefficients (Eq. 7).

To get the total charge density, the $d$-like states require a factor of $k_B T$, so they become even smaller; because of this we neglect the $d$-like carriers for the band structure calculation. However, we will later see that transport is governed by the low-mass, $d$-like carriers.

The charge density can be calculated using usual methods for semiconductors. In semiconductors, the conduction band is nearly empty, so the Fermi-Dirac distribution $f^0(\epsilon)$ can be approximated by the Boltzmann distribution. The electron density is
\[
n = \int_{E_C}^{\infty} f^0(\epsilon) g(\epsilon) d\epsilon = n_0 e^{-(E_C - E_F)/k_B T}
\]  
(8)

where $n_0$ is the average DOS. In SmB$_6$ we use Eq. 3 for the DOS, keeping only the delta function terms. This yields electrons in the conduction band with approximate density
\[
n \approx N_{cf} e^{-(E_C - E_F)/k_B T}.
\]  
(9)

We can similarly calculate the approximate density of holes in the valence band to be
\[
p \approx N_{cf} e^{-(E_F - E_V)/k_B T}
\]  
(10)

where $E_V$ is the valence band edge. This result resembles the carrier density of a conventional semiconductor. For such semiconductors, charge neutrality, $n = p$, yields the intrinsic carrier density
\[
n = p = n_i \approx \sqrt{N_{cf} N_{cf} e^{-(E_{gap} - 2k_B T)}}.
\]  
(11)

The intrinsic picture works well for SmB$_6$ at high temperatures. However, intrinsic materials are sensitive to surface effects, and at low temperature, these become relevant. All surface charges, such as TI states, if they are present, and localized surface charges associated with oxidation on the surface, contribute. Requiring charge neutrality with the addition of the surface charges forces the Fermi energy to be pinned in place, leading to band bending in the valence and conduction bands. This possibility has been suggested by recent experimental results,\(^{41}\) but was not previously explored in depth.

To understand the effects of band bending, we perform a self-consistent calculation to obtain the band structure. In this calculation, we model the effects of band bending using a potential $\phi(z)$ of the form
\[
e\phi(z) = E_C(z) - E_{gap}/2
\]  
(12)

where the conduction band is now dependent on location $z$ in the bulk, and $E_{gap} = E_C - E_V$. We can rewrite the carrier densities in terms of this potential to obtain
\[
n(z) = N_{cf} \exp\left[ - \frac{e\phi(z) + E_{gap}/2 - E_F}{k_B T} \right]
\]  
(13)

and a similar expression for $p(z)$. Using charge neutrality again, we obtain
\[
\rho(z) = -en(z) + ep(z) = 2n_i e \sinh \left[ \frac{e\phi(z)}{k_B T} \right]
\]  
(14)
for the total charge density. We then solve for the potential across the bulk using Poisson’s equation in one dimension,

$$\frac{d^2 \phi}{dz^2} = \frac{\rho(z)}{\varepsilon} = \frac{2n_i e}{\varepsilon} \sinh \left( \frac{e\phi(z)}{k_BT} \right).$$ (15)

To solve this equation, we choose a “test sample” of thickness 200 \( \mu \)m, which is typical to a real SmB_6 sample. We define \( z = 0 \) as the center of the sample (so that \( z = \pm 100 \mu \)m are the edges). In addition, we use \( \kappa = 600 \) as the dielectric constant.\(^{20}\) To proceed with the solution, we now require boundary conditions. For the first boundary condition, we simulate the band bending effects on the surface by introducing a pinning energy \( E_{\text{pin}} \), which is defined with respect to the midpoint of the gap by \( E_{\text{pin}} = (E_C + E_V)/2 - E_F \). This describes the energy difference between the band at zero potential and the minimum of the bent band. Therefore, the boundary condition can be expressed as \( e\phi(z = 100 \mu m) = -E_{\text{pin}} \).

We can also define this pinning relative to the gap as \( E_{\text{pin}} = E_{\text{gap}}/2 - E_a \). As we will see later, when temperature is sufficiently low, \( E_a \) corresponds to the activation gap measured by transport. The pinning energy must also be the same at both edges of the sample, and to enforce this, the second boundary condition is that \( d\phi/dz = 0 \) at \( z = 0 \), the center of the sample.

Additionally, we can define a “built-in potential,” \( V_{bi} \), as is commonly done for band-bending calculations in semiconductors.\(^{21}\) \( V_{bi} \) describes the magnitude of the bending in terms of the difference between the maximum and minimum points on the conduction or valence band. Using this built-in potential, we can determine an associated length scale (the accumulation length) of Eq. 15, given by

$$l = \sqrt{\frac{2eV_{bi}}{en_i}} = \sqrt{\frac{2eV_{bi}}{e(N_{cf}N_{cf})^{1/2}} e^{E_{\text{gap}}/4k_BT}}.$$ (16)

At low temperature, when \( E_{\text{gap}} \) is large, the accumulation length is large, and at high temperature, when \( E_{\text{gap}} \) is small, the accumulation length is small.

With these parameters, solutions to Eq. 15 were found for temperatures of 4-40 K and values of \( E_{\text{pin}} \) between 4 and 7 meV. From a solution \( \phi(z) \), the conduction and valence bands can be obtained from Eq. 12, and the charge density can be obtained from Eq. 14. An example of these are shown for 8 K and \( E_{\text{pin}} = 5.5 \) meV in Fig 3. Fig. 3(a) shows the calculated conduction and valence bands, as well as the relationships among the band structure and the parameters \( E_{\text{pin}}, E_{\text{gap}}, E_a, \) and \( eV_{bi} \). We note that the valence band is always parallel to the Fermi energy than the valence band is, and this result is very different from what is observed in a standard gapped material.

This is expected, because excess charge at the surfaces leads to band bending.

Fig. 4 shows how the band structure varies with temperature, again using \( E_{\text{pin}} = 5.5 \) meV; the valence band is omitted. At 12 K, the highest temperature shown, the conduction band for the majority of the bulk is \( E_{\text{gap}}/2 \) above the Fermi energy. There is a small amount of band bending at the edges, but it does not extend very far into the bulk, as expected from Eq. 16. This means that the band structure is similar to that of a standard gapped material, except near the surface. As the temperature is lowered, however, the band bending effects begin to extend farther into the bulk. At 2 K, the lowest temperature shown, these effects completely dominate the band structure. Here, the conduction band is much closer to the Fermi energy than the valence band is, and this result demonstrates a crossover between bulk
conduction dominated by the usual bulk effects (at high temperatures) and bulk conduction dominated by surface effects (at low temperature). We can understand where the crossover occurs by comparing the charge densities and relevant length scales. At high temperatures, the bulk dominates bulk transport, and this can be characterized by the size of the sample ($t$) and the intrinsic carrier density ($n_i$). At low temperatures, the surface dominates bulk transport, and this can be characterized by the accumulation length ($l$) and the carrier density on the surface ($n_s$). From this, we can estimate the crossover as occurring when $tn_i \approx ln_s$. In our calculation, we estimate that the crossover occurs at about 10-12 K. Since the crossover depends on the relationship between the accumulation length and the sample, the crossover temperature is dependent on the thickness of the sample. For example, in a much thinner sample such as a thin film, the crossover would occur at a higher temperature than we have estimated in this calculation.

These band bending results and the crossover are able to explain the gap discrepancy between transport and spectroscopy. In our picture, spectroscopy still measures the full gap, which does not change based on where the measurement occurs in the sample. Transport, on the other hand, measures a different activation gap depending on temperature. Below the crossover, transport measures the activation gap on the surface, as shown in Fig. 3(a), but above the crossover, transport measures the average gap across the entire bulk. As a rough estimate using our model’s parameters, at 4 K the total gap is about 19 meV, and with a pinning of 5.5 meV the model yields $E_a = 4$ meV, in agreement with experiment. In this way, the gap discrepancy can be understood without using in-gap bulk impurity states.

The crossover also has interesting implications for the gap. As we have noted, the accumulation length extends through the bulk at low temperatures. As the temperature is lowered, the accumulation length diverges, and this can be thought of in analogy with semiconductors. In semiconductors, the accumulation or depletion length increases with purity, and a completely pure material would have an infinite accumulation or depletion length. Our calculation shows that the accumulation length becomes large at cryogenic temperatures in SmB$_6$, consistent with the hypothesis that the gap is clean and the bulk is truly insulating.

We note that the effects of the surface, whether topological, trivial, or both, were all included in the parameter $E_{pin}$. In terms of our simulation, various surface effects including the crystal plane, polarity, disorder, TI states, or other predicted surface effects would simply change the value of $E_{pin}$. We do not predict which effects are strongest, but changes due to such effects would change the strength of the band bending for the surface considered.

We further note that the Fermi energy pinning is not rigid, which means that fixing the conduction band permanently at $E_{pin}$ is not exact for all temperatures. Near the crossover, the Fermi energy is allowed to shift slightly. This means that competing with the surface-to-bulk crossover is a slightly shifting pinning. (If the surface effects are topological, we predict that the pinning would be more rigid than if the effects were trivial.) This type of shift does not affect the fully surface or fully bulk regions seen in the model, but if its effects were included, it would slightly change the temperature at which the crossover occurs. We have chosen not to include these effects, because the model is quite robust against changes. The band bending result is always present, and even a large variation in the parameters only slightly shifts the magnitude of the calculated effect. For example, reports of dielectric constant vary from 600 to 1500. When comparing these extremes in the calculation, the results above the crossover are exactly the same, and the results below the crossover differ only slightly.

III. CONNECTION TO EXPERIMENT

To assess the validity of our model, we must connect the results of the self-consistent calculation to measurable parameters. Specifically, we examine Hall coefficient, resistivity, and thermopower, comparing the simulation results for each to data. Although we have used a semiconductor picture, SmB$_6$ is very different from a standard semiconductor due to its non-parabolic dispersion arising from the hybridized band structure. Because of this unusual band structure, its transport properties are unique and must be considered in detail. We must consider which carriers, d-like or f-like, contribute to transport phenomena, as well as the sign of these carriers.

In the discussion of the simplified DOS above, we saw that the f-like electrons dominate when calculating the carrier density, because their contribution to the DOS is much greater than the contribution of the d-like electrons. However, the f-like electrons have a flat dispersion, which must yield zero mobility. This means they cannot con-
tribute to transport, so observed transport phenomena must be due to the $d$-like electrons. Looking back at the features of the actual band structure, we note that the $f$-like electrons do not have an exactly flat dispersion, and there is some curvature connecting the $f$- and $d$-like states. However, based on ARPES data, these features are small compared to the size of the gap. Thus the curvature of these features is small compared to the simplified band structure, so we can again neglect the effects.

We can understand this more quantitatively using semiclassical (Boltzmann) transport. First, we will consider the case of an intrinsic semiconductor in the conduction band to demonstrate the calculation, but the valence band result can be found similarly. We will then discuss modifications for the SmB$_6$ case. Using the elementary solution to the Boltzmann equation, the current due to an electric field applied in the $z$-direction is

$$j_z = - \int_{E_C}^{\infty} v_z^2 e^2 \tau E_z \left( \frac{\partial f^0(\epsilon)}{\partial \epsilon} g_C(\epsilon) \right) d\epsilon$$  \hspace{1cm} (17)

where $v_z$ is the particle velocity, $e$ is the electronic charge, $\tau$ is the scattering time, $E_z$ is the applied electric field, $g_C(\epsilon)$ denotes the DOS for the conduction band, and $\epsilon$ is the energy. We use the relaxation time approximation, where $\tau$ is independent of energy, and the equipartition theorem, $v_z^2 = v^2/3$, to rewrite Eq. 17. Then, using $j_z = \sigma E_z$, we find the conductivity,

$$\sigma \approx -\frac{e^2 \tau}{3} \int_{E_C}^{\infty} v^2 \frac{\partial f^0(\epsilon)}{\partial \epsilon} g_C(\epsilon) d\epsilon.$$  \hspace{1cm} (18)

The derivative of the Fermi-Dirac distribution is

$$\frac{\partial f^0(\epsilon)}{\partial \epsilon} = -\frac{1}{k_B T} f^0(\epsilon)(1 - f^0(\epsilon))$$  \hspace{1cm} (19)

and for a general intrinsic semiconductor, the conduction band is almost empty, so the term in parentheses can be approximated as 1. Then the conductivity becomes

$$\sigma \approx \frac{e^2 \tau}{3k_B T} \int_{E_C}^{\infty} v^2 f^0(\epsilon) g_C(\epsilon) d\epsilon.$$  \hspace{1cm} (20)

Since only electrons near the Fermi energy are mobile, we can approximate their velocity as the Fermi velocity $v_F$, and this is a constant. The remaining integral is just the usual method for calculating carrier density, so we find

$$\sigma \approx -\frac{e^2 \tau}{3k_B T} v_F^2 n$$  \hspace{1cm} (21)

where $n$ is given by Eq. 8. To further simplify, we can use the Einstein relation for semiconductors, which relates the diffusion constant, $D = \frac{v^2 \tau}{3}$, to the mobility by

$$\mu k_B T = \frac{v^2 \tau}{3}$$  \hspace{1cm} (22)

where $\mu$ is the mobility, $v$ is the average velocity, and 3 represents the number of dimensions (the right-hand side of this equation can be derived using the equipartition theorem). So we find, for average velocity $v_F$, the familiar result, written for an intrinsic semiconductor,

$$\sigma = n e \mu.$$  \hspace{1cm} (23)

In SmB$_6$, the picture is a little more involved. We will now re-derive the general result of Eq. 23 with modifications for SmB$_6$. First, we consider the carriers. Since there are two types of carriers ($f$-like and $d$-like), a small displacement of the Fermi surface due to an applied electric field is not uniform. For our simplified dispersion (Fig. 1c), say the field is being applied from right to left (so that electrons move from left to right). Then the electrons in $d$-like states on the right are mobile as they would be in a conventional semiconductor. However, the $f$-like states are filled and have zero mobility. This means that only half of the carriers in the band can move when a current is present. Therefore we must include a factor of 1/2 relative to the usual result (Eq. 19).

Additionally, since only the $d$-like carriers contribute to transport, we should only consider the carrier density $n$ coming from these. Using Eq. 8 and Eq. 7, we find that

$$n = g_0 k_B T e^{-(E_C - E_F)/k_B T}.$$  \hspace{1cm} (24)

Now we can rewrite $g_0$ using Eq. 6. Also, the carriers should still move with an average speed of $v_F$. We obtain

$$\sigma_{\text{SmB}_6} \approx \frac{e^2 \tau v_F^2}{6} \frac{m^*}{\pi^2 h^2} \sqrt{2m^* E_F} e^{-(E_C - E_F)/k_B T}$$  \hspace{1cm} (25)

and this can be further simplified using $\sqrt{2m^* E_F} = \hbar k_F$ and $v_F = \hbar k_F/m^*$ to obtain

$$\sigma_{\text{SmB}_6} \approx \frac{e^2 \tau k_F^3}{6m^* \pi^2} e^{-(E_C - E_F)/k_B T}.$$  \hspace{1cm} (26)

Next, we apply $k_F = (3\pi^2 n)^{1/3}$, which can be calculated by integrating to find the carrier density at zero temperature and rearranging. This means that $n$ in this expression is the density of filled states up to $E_F$, and according to Eq. 4, this is just $n_{\text{cell}}$. We also use $\mu = eT/m^*$ to find

$$\sigma_{\text{SmB}_6} \approx \frac{1}{2} n_{\text{cell}} \mu_d e^{-E_C/E_F} e^{-(E_C - E_F)/k_B T}$$  \hspace{1cm} (27)

where the subscript on mobility denotes that only $d$-like electrons are mobile.

This calculation can be repeated for the valence band, and the result is similar, except that the exponential is replaced by $\exp[-(E_F - E_V)/k_B T]$. To understand how the conduction and valence band contributions are related physically, we must consider the signs of the carriers in both bands. For a conventional semiconductor, the conduction band contributes electrons with positive effective mass and the valence band contributes holes with negative effective mass. These have opposite contributions to transport. In SmB$_6$, we still have electrons.
in the conduction band and holes in the valence band, but both bands have positive curvature in the $d$-like electrons. Since we are just considering the $d$-like carriers, the conduction band case is the same as that of a conventional semiconductor, electrons with positive effective mass. However, the valence band has positive curvature rather than negative as it would for a conventional semiconductor. This means that although there are holes in the valence band, they have a positive effective mass as well, so they contribute with the same sign as the electrons in the conduction band.

Although this result was found by considering the simplified band structure, again we can neglect the details of the bands. For example, the $H$-point feature in the valence band observed by ARPES\textsuperscript{18} shows up as a small bump with negative curvature in the $f$-like part of the dispersion. As discussed previously, this feature is very close to the valence band, so although it creates some curvature in the valence band, the effect is small. This means that there are holes with negative effective mass at these points in the valence band, but we assume that the curvature is large, so that these carriers have a much smaller mobility than the $d$-like carriers.

Because the dominant ($d$-like) carriers in the valence and conduction band contribute to transport with the same sign, we can return to Eq. 27 and conclude that the total conductivity for all carriers in both bands is the usual result as in Eq. 23, with $\mu = \mu_d$, and $n$ defined as

$$n = \frac{1}{2} n_{ell} e^{-(E_C - E_F)/k_B T} + \frac{1}{2} n_{ell} e^{-(E_F - E_V)/k_B T}$$

(28)

This means that we can use the usual transport relations to connect our model to the experimental results, provided that this expression is used to calculate the carrier density.

IV. TRANSPORT IN THE MODEL

A. Hall coefficient and resistivity

We can now combine the results $E_C(z)$ and $E_V(z)$ of the self-consistent calculation with Eq. 28 to define an effective carrier density

$$n_{eff} = \frac{1}{t} \int_{-t/2}^{t/2} \frac{1}{2} n_{ell} \left( e^{-|E_C(z) - E_F|/k_B T} + e^{-|E_F - E_V(z)|/k_B T} \right) dz$$

(29)

where $t$ is the thickness of the sample ($t = 200 \mu m$ for our test sample). We can then use this carrier density to compare the model to transport data. We first concentrate on the Hall coefficient ($R_H = 1/ne$) because it does not require any further parameters to be included; however, if we assume that mobility is constant, the resistivity follows the same trend. This is not a good assumption, as mobility is often temperature dependent, but the same feature around 10 K is seen in data for both Hall coefficient and resistivity.

Fig. 5 shows a plot of calculated Hall coefficient as a function of temperature for various values of $E_{pin}$. As in the band structure result of Fig. 4, we observe a crossover around 10 K, which is also seen in data for both Hall coefficient and resistivity. Below the crossover, however, there is some variation. In this region, the amount of bending influences the accumulation length, so the magnitude of the Hall coefficient changes with $E_{pin}$. As mentioned previously, in this region, the activation energy can also be determined by $E_a = E_{gap}/2 - E_{pin}$.

The calculated Hall coefficient for $E_{pin} = 5.5$ meV is plotted along a collection of data in Fig. 6. The data agrees very well above about 10-12 K, and this agree-
ment suggests that the bulk carrier density in SmB$_6$ is fairly consistent across samples. Below about 10-12 K, the data exhibits a variation of about an order of magnitude. We do not suggest a mechanism for this variation, but we do note that in this temperature range, our model suggests that the bulk conduction is dominated by the surface. Around 10-12 K, a feature is observed by all researchers, and this corresponds to the crossover discussed previously, between bulk conduction dominated by bulk effects above the crossover and bulk conduction dominated by surface effects below the crossover.

The calculated Hall coefficient from our model also demonstrates a crossover at about 10-12 K and correctly describes the shape of the data, although the magnitude of our result does not agree with the data. This disagreement was expected based on the simplifications made to the dispersion. Previously, we estimated that these simplifications yielded errors of not more than a few meV in the dispersion, but this becomes very important for the Hall coefficient calculation. We can estimate the size of the expected discrepancy in the calculated Hall coefficient by using a Boltzmann factor, $e^{\Gamma/k_BT}$, where $\Gamma$ is the approximate width of the features that were neglected. For $\Gamma$ in the range of 1 to 3 meV in a temperature range of 10 to 20 K, the model is expected to be off at least by a factor of 2 and at most by a factor of 32. Our result was consistently a factor of about 5-6 greater than the data, and this is within the expected range of the discrepancy. Based on this estimate, agreement would likely be improved by including more features of the dispersion and DOS. However, this type of refinement would require many more parameters to be introduced.

**B. Thermopower**

We also compare the model to thermopower data, because like the Hall coefficient, it does not require any further parameters. In the relaxation time approximation, thermopower for electrons in a semiconductor is given by

$$S_C = -\frac{k_B}{e} \left[ \alpha + \frac{5}{2} - \frac{E_C - E_F}{k_BT} \right]$$

(30)

where the subscript denotes the conduction band, and $\alpha$ is a constant between 0 and 2 that describes how energy is related to scattering time ($\tau \propto E^\alpha$). A similar expression can be obtained for holes in the valence band; it is important to note that the sign is positive for holes in a standard semiconductor. For a material containing both electrons and holes, these can be combined according to

$$S_{tot} = \frac{S_C \sigma_C + S_V \sigma_V}{\sigma_C + \sigma_V}.$$  

(31)

In the limit of an intrinsic semiconductor, where $n = p$, the intrinsic carrier density in Eq. 11 and the usual conductivity in Eq. 23 can be used to simplify this. Assuming a quadratic dispersion for both the valence and conduction bands, we obtain

$$S = \frac{k_B}{e} \left[ b - 1 \frac{E_{gap}}{b + 1} \frac{3}{4} \ln \frac{m_n}{m_p} \right]$$

(32)

where $b = \mu_n/\mu_p$, and the subscripts refer to electrons ($n$) and holes ($p$).\textsuperscript{32,44}

Again, the picture is slightly different in SmB$_6$. The factor of 1/2 must be accounted for once again, but based on the form of Eq. 31, it is clear that this factor cancels out for thermopower. Also, the curvature of the valence band means that the holes contribute with the same sign as the electrons. Using the conductivity derived in Eq. 27 and the analogous result for the valence band in Eq. 31, we can find the total thermopower in the model. Defining $\epsilon_v(z) = E_v(z) - E_F$ and $\epsilon_e(z) = E_F - E_v(z)$, we obtain

$$S(z) = \frac{k_B}{e} \frac{1}{k_BT} \int \frac{\epsilon_v(z) e^{-\epsilon_v(z)/k_BT} + \epsilon_e(z) e^{-\epsilon_e(z)/k_BT}}{e^{-\epsilon_v(z)/k_BT} + e^{-\epsilon_e(z)/k_BT}} dz$$

(33)

where $z$ again refers to the location in the bulk in the model. To get the total thermopower, we must integrate this expression, but since thermopower is dependent on conductivity and conductivity is dependent on $z$, it must be integrated using a form similar to that of Eq. 31. Thus, the effective thermopower across the bulk is

$$S_{eff} = \int_{-t/2}^{t/2} S(z) \sigma(z) dz$$

(34)

where $t$ is the thickness of the sample (here $t = 200 \mu m$) and $\sigma(z) = n(z) e \mu$, where $n(z)$ is given by the integrand of Eq. 29.

$S_{eff}$ was calculated as a function of temperature for various values of the pinning, shown in Fig. 7. Again, a feature around about 10 K is evident, although it is broader than the feature seen for the Hall effect. The bulk behavior in Fig. 7 is the same for all values of $E_{pin}$.
but the surface behavior and prominence of the feature varies.

Fig. 8 shows a collection of thermopower data with a fit from the model. The data are consistent at high temperatures, and in this regime, there is also excellent agreement between the data and the model. As with the Hall coefficient, this agreement is expected as bulk effects dominate in this regime. At low temperatures, the data are more diverse. We note that the crossover does not occur at the same temperature in each data set shown, but we do not propose a mechanism for this.

At high temperature, in the bulk-dominated regime, there is good agreement between the data and the model. A feature is also present in both the data and the model, although it does not occur at exactly the same temperature for each set of data. At low temperature, a deviation can be observed in the data, although it does not occur at the same point in each data set. We attribute this deviation to the manifestation of the TI surface state near 4 K, which can be observed in the data, although it does not occur at exactly the same temperature. We attribute this deviation to the surface state near 4 K, which can be added to the calculation. Near 4 K, with $E_a = 3.47$ meV, $^{12}$ is approximately

$$S_{tot} \approx \frac{eE_a/k_BT + eE_a/k_BT_c}{S_b}. \quad (36)$$

This expression is shown in Fig. 8 with a dashed line, and the result provides a better estimate of the data near 4 K than the fit from the calculation.

Our model, with the addition of the estimation of surface effects, captures the low-temperature features of the thermopower well. We note that there is not much thermopower data available in the literature, which makes it difficult to understand trends in the data as done for the Hall coefficient. Our model agrees quite well with the data from Sluchanko et al. $^{32}$, but not as well with the other data. Again, a discrepancy between the data and the model is present, and again we attribute this to neglecting the small features of the dispersion. Improved agreement could likely be attained by adding more details of the dispersion to the model.

V. CONCLUSION

We have presented a new model to understand the difference between the spectroscopy and transport gaps in SmB$_6$ without relying on in-gap bulk states. Transport measures an activation energy of 3-4 meV, or a gap of 6-8 meV, while spectroscopic methods measure a gap of 16-20 meV. This discrepancy between the two results has often been explained by introducing a localized bulk state in the gap, perhaps due to impurities, although other explanations such as the presence of an indirect gap have also been used. The effective mass approximation has been used to understand such an in-gap impurity state, because this method has been successful at describing impurity states in many gapped materials. However, we showed that the effective mass approximation fails when it is applied to SmB$_6$, suggesting that the in-gap impurity state picture is not justified.

Instead, we suggested a new way of understanding the SmB$_6$ gap using self-consistent band-bending calculations. We simplified the well-known dispersion and corresponding density of states to capture the main characteristics of SmB$_6$ and modeled SmB$_6$ as an intrinsic semiconductor. We considered the possibility of band bending, which is expected to arise from the presence of excess surface charges. A self-consistent solution for the potential was found by numerically solving Poisson’s equation with this charge density across the bulk of a test sample, with the boundary conditions simulating the strength of the bending effects at the surface and enforcing symmetry across the bulk.

The self-consistent solution was found for temperatures from 4-40 K, and from this result the band structure was calculated. In addition, the band structure result was connected to measurable transport parameters using
semiclassical transport theory modified for SmB$_6$. Specifically, we concentrated on Hall coefficient, resistivity, and thermopower, as these do not require the addition of adjustable parameters to our model. The results of these calculations demonstrated a crossover to bulk transport dominated by bulk effects (in analogy with a standard gapped material) at high temperatures, to bulk transport dominated by surface effects at low temperature. The calculated crossover temperature was 10-12 K, which accounts for a feature that has been observed in transport data near this temperature. The transport parameters were compared to data, and although there was qualitative agreement between the data and the calculated curves, their magnitudes did not agree. We attributed the disagreement to neglecting some of the small features of the actual dispersion in the model, and found that a few meV of error in the dispersion could explain the disagreement.

We also related the feature seen in transport data to the accumulation length for a semiconductor. At high temperatures, the accumulation length was much smaller than the sample size, and at low temperatures, the accumulation length was much larger than the sample size. We estimated that the crossover would occur when bulk effects in terms of sample thickness and intrinsic carrier density become comparable to surface effects in terms of accumulation length and surface charges. Because of this relationship between accumulation length and sample size, our model suggests that the crossover would occur at different temperatures for different sample thicknesses.

This suggests one straightforward way of testing our model: measuring the location of the transport feature around 10-12 K for samples of different thicknesses. For a sample thinner than 200 µm, the crossover temperature would be higher than the 10-12 K temperature calculated in our simulation, and for a thicker sample, the crossover temperature would be lower than 10-12 K. This dependence on thickness is shown in Fig. 9 for samples from 20-2000 µm in thickness, and the change in the crossover temperature is large enough to be experimentally measurable. On the other hand, testing the model using a method such as ARPES would be difficult. Since ARPES probes at and near the surface, it might be expected that changes in accumulation length could be detected. However, the ARPES penetration depth is tiny compared to the accumulation length, even when high energy photons that can penetrate farther into the bulk are used. Because of this, it would be difficult to directly image the band bending effects using ARPES.

The divergence of the accumulation length in SmB$_6$ at low temperatures suggests that the gap is clean, similar to the gap in a superconductor. This, combined with the success of our model at describing a variety of data without introducing bulk states in the gap, agrees well with the observation that there is no residual bulk conduction in SmB$_6$ below the TI crossover temperature of about 4 K. Together, these observations imply that SmB$_6$ is a true TI; it does not exhibit bulk conduction below the TI crossover temperature as all other known TIs do. This would be exciting for research in technological applications that require a clean gap and no bulk conduction. We also predict that our model could be extended to other materials that have a dispersion similar to that of SmB$_6$, including alloys of SmB$_6$.

ACKNOWLEDGMENTS

The authors thank Lu Li and Steven Wolgast for helpful discussions, as well as J.W. Allen and Jonathan Denglinger for helpful discussions and advice on improving the manuscript. Funding for this work was provided by NSF grants #DMR-1441965 and #DMR-1643145.

---

FIG. 9. Predicted thickness dependence of crossover for $E_{\text{pin}} = 5.5$ meV.

---

1. A. Menth, E. Buehler, and T. H. Geballe, Physical Review Letters 22, 295 (1969).
2. R. L. Cohen, M. Eibschtz, and K. W. West, Physical Review Letters 24, 383 (1970).
3. M. B. Maple and D. Wohlleben, Physical Review Letters 27, 511 (1971).
4. G. Aeppli and Z. Fisk, Comments Cond. Mat. Phys. 16, 155 (1992).
5. J. Kondo, Progress of Theoretical Physics 32, 37 (1964).
6. P. Coleman, Introduction to Many-Body Physics (Cambridge University Press, 2015).
7. N. F. Mott, Philosophical Magazine 30, 403 (1974).
8. J. C. Nickerson, R. M. White, K. N. Lee, R. Bachmann,
