Chapter 5: Resolving the True Band Gap in ZrNiSn: The Effect of the Weighted Mobility Ratio in Bipolar Semiconductors

5.1 - Introduction

Temperature-dependent electronic transport measurements serve as the primary characterization tool for most thermoelectric materials. With regards to the electronic band structure, these measurements can be used to determine things like the effective mass, deformation potential, band offset (ΔE) in the case of two-band systems with the same type of carrier (as discussed in Chapter 4 for the IV-VI materials), or band gap (E_g) for systems where the bipolar effects are large. Particularly in the case of band gap, these estimates are useful for comparing experimental and theoretical (DFT) calculations. Further, understanding the band gap is important for modelling electronic transport properties, primarily to distinguish and quantify band engineering effects. Half Heusler's (HH) have received a lot of attention lately because of their flexibility with regards to composition (XYZ where X=Ti, Zr, Hf, etc., Y=Ni, Co, etc., Z=Sn, Sb, etc.), good electronic properties, earth-abundant elements, and reasonable thermoelectric performance. In this chapter, I explore the relationship between the maximum measured temperature-dependent thermopower ($|S|$), which can be used to estimate the band gap using $E_g = 2e|S|_{max}T_{max}$ (the Goldsmid-Sharp band gap), and the electron-to-hole weighted mobility ratio, A [51]. While the Goldsmid Sharp gap estimate is useful, mainly due to its simplicity and ease of application, it gives conflicting results in ZrNiSn and other systems that have large differences in electron and hole weighted mobility (defined as $\mu_0 m^{*3/2}$). By combining optical measurements with the estimated Goldsmid-Sharp band gap in n-type and p-type samples, I am able to resolve the true band gap as well obtain an estimate for the electron-to-hole weighted mobility ratio in ZrNiSn. I also discuss the origins of the large difference in electron and hole weighted mobility in the context of existing literature results and the inherent disorder associated with HH materials.

As an extension of my findings in the ZrNiSn HH system, the second section of this chapter is dedicated to a more thorough theoretical investigation of the Goldsmid-Sharp band gap relationship in the limit of narrow gap materials, where the full Fermi statistics should be considered. I present a chart that can be used to quickly estimate deviations in the $2eS_{max}T_{max}$ relation from the true band gap as a function of the maximum Seebeck coefficient and the majority-to-minority carrier weighted mobility ratio, A. I also discuss what deviations in this relation can occur as the band gap is narrowed significantly and the Seebeck rollovers occur while the Fermi level is degenerate.

5.2 - ZrNiSn Half Heusler Thermoelectric Materials– Resolving the True Band Gap

5.2a - Introduction

Half-Heusler (HH) compounds with the general formula *X*NiSn (*X*= Ti, Zr, Hf) have generated significant interest as a promising class of materials for thermoelectric applications because of their high thermopower, reasonable mobility, earth-abundant elements, and good performance [212, 213]. The numerous possibilities to manipulate each of the three lattice sites provide an excellent opportunity to influence the electronic and thermal transport properties. In the case of XNiSn, isoelectronic alloying on the *X* site [214] or partial substitution on the Ni site [215] can lead to thermal conductivity reduction owing to lattice strain and mass fluctuations by point-defect scattering of phonons. Beyond simply alloying, a wide range of HH composites have utilized phase separation, which can result in remarkably low thermal conductivities and even enhanced electrical properties [216-219]. In fact, much of the work on the HH's has involved phase-separating alloys which benefit greatly from the increased phonon scattering from nano/microstructure features. The *X*NiSn system is the most extensively studied *n*-type HH material [23, 212, 213, 219-225] showing good *zT*s of 0.6-1.4 (usually optimizing at high temperatures, T>900 K) [223].

Much of the existing literature focuses on the thermoelectric characterization of *n*-type *X*NiSn, which show the best thermoelectric performance in comparison to their *p*-type analogs. The most promising *p*-type HHs are found in the *X*CoSb system (*X*= Ti, Zr, Hf) [213, 226-228]. However, for the construction of thermoelectric modules, it is desirable for the *n*-type and *p*-type materials to have similar chemical, thermal, and mechanical properties. In order to be most compatible with high-performance HH *n*-type materials that are currently being explored for modules [229], a suitable *p*-type analog in the *X*NiSn system is desirable. Some work has already been done to this end; for example: Xie *et al.* [230] studied the substitution of Ni by Co and demonstrated that a conversion from *n*- to *p*-type behavior for the *X*NiSn system could be achieved. Likewise, Horyn' *et al.* [231] studied the effects of substituting Sc for Ti or Zr; the resulting compounds showed promising positive thermopower at room temperature.

The electronic transport properties of semiconducting materials, such as thermopower and electrical conductivity, are reflections of the electronic energy band structure in *k*-space and the electronic states which make up the valence and conduction bands. A crucial parameter that determines the thermoelectric performance is the band gap. For narrow band-gap semiconductors, the onset of bipolar conduction (both electrons and holes contributing) occurs at lower temperatures than their higher gap analogs. The compensating Seebeck coefficient of the minority carriers leads to a peak in the thermopower, limiting the temperature-dependent zT values.

In this work, I aim to extend the existing studies of *p*-type *X*NiSn HH compounds by thoroughly investigating Sc substitutions in $Zr_{1-x}Sc_xNiSn$. Through Sc substitution, we successfully doped ZrNiSn to be *p*-type and then measured its thermoelectric properties up to 850 K. I estimated the band gap size using the commonly used Goldsmid-Sharp formula:

$$E_g = 2e|S|_{max}T_{max}$$ Equation 5-1

where T_{max} is the temperature at which the maximum of the thermopower ($|S|_{max}$) occurs [51]. E_g here is referred to as the Goldsmid-Sharp band gap and is obtained from high temperature thermopower measurements. In this work, synthesized ZrNiSn HH samples give a discrepancy between the values obtained from our *p*-type samples (~0.05 eV) and the *n*-type results from the literature (>0.23 eV), both of which are different from the *ab-initio* calculated result of 0.5 eV [232]. With the aid of the optical measurements, we used the Goldsmid-Sharp formula to resolve this apparent discrepancy by considering a large difference in weighted mobility between electrons and holes. I would like to acknowledge work from Jennifer Schmitt, who was an equally contributing coauthor on the work that this section is based upon [52].

5.2b - Results

X-ray Diffraction

Figure 5-1 shows a representative *XRD* pattern for $Zr_{0.95}Sc_{0.05}NiSn$, which conforms to the well-defined cubic structure typical for the HH phase containing negligible amounts of Sn metal. The calculation of the lattice parameter and the crystal structure refinements were done using the Rietveld method, which yielded a lattice parameter of a = 6.1148 Å for the undoped sample, which is in agreement with the values reported in the literature [233]. The inset in Figure 5-1 shows a roughly monotonic increase with increasing Sc amount, which agrees well with Vegard's law. The increase serves as evidence of the substitution of slightly larger Sc atoms (radius = 1.60 Å) for Zr (radius = 1.55 Å). Each of these observations is consistent with previous synchrotron results on $Ti_{1-x}Sc_xNiSn$ compounds synthesized in a similar way [232].

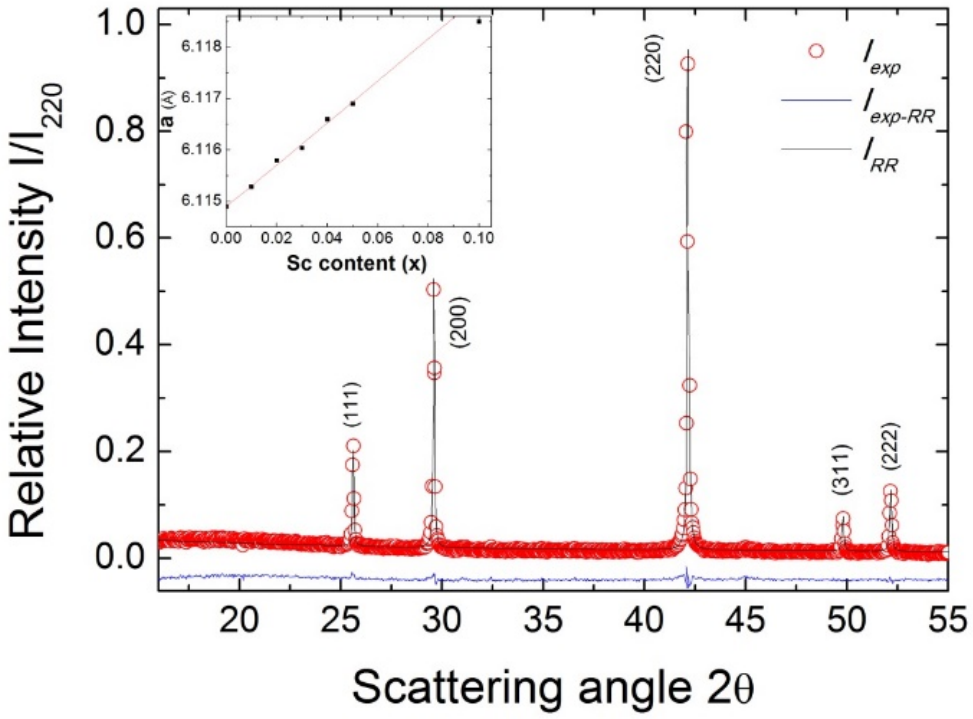

Figure 5-1 Powder XRD pattern for the $Zr_{0.95}Sc_{0.05}NiSn$ sample (I_{exp}) with the Rietveld refinement (I_{RR}) and the difference profile (I_{exp-RR}). The inset shows the dependence of the lattice parameter a (Å) for $Zr_{1-x}Sc_xNiSn$ versus Sc content (x). The straight line represents the linear fit to Vegard's law.

Electrical Transport Properties

The electrical properties of the $Zr_{1-x}Sc_xNiSn$ (x = 0 ,…, 0.10) solid solution are shown in Figure 5-2a. At room temperature, the samples exhibit high electrical resistivity, which decreases with increasing temperature for all the samples, indicating semiconducting behavior. For high substitution levels, both the room temperature value and the temperature dependence of the resistivity decreases, pointing to a shift of the Fermi level towards the valence band as a result of an increasing *p*-type carrier concentration.

It can be seen from Figure 5-2b that for increasing Sc content in the $Zr_{1-x}Sc_xNiSn$ solid solution, the Hall mobility ($\mu_H = \sigma \cdot R_H$ in units of mobility) is suppressed (and eventually becomes positive) by the addition of holes. The parent compound, *n*-type ZrNiSn, exhibits the highest mobility, with a value at room temperature of 25 $cm^2/ V \cdot s$, which is a typical value for *n*-type

*X*NiSn-based systems [234, 235]. This value is still significantly lower than 150 $cm^2/ V \cdot s$ for Bi_2Te_3 or 900-1400 $cm^2/ V\text{-}s$ for the lead chalcogenides [58, 236]. If we assume that each replacement of Zr by Sc leads to one hole, the room temperature electron concentration should be completely compensated when the Sc concentration exceeds ~5%. At low Sc content, the samples with $0 \leq x \leq 0.04$ possess a negative value for $\sigma \cdot R_H$, which is consistent with a substitutional doping explanation.

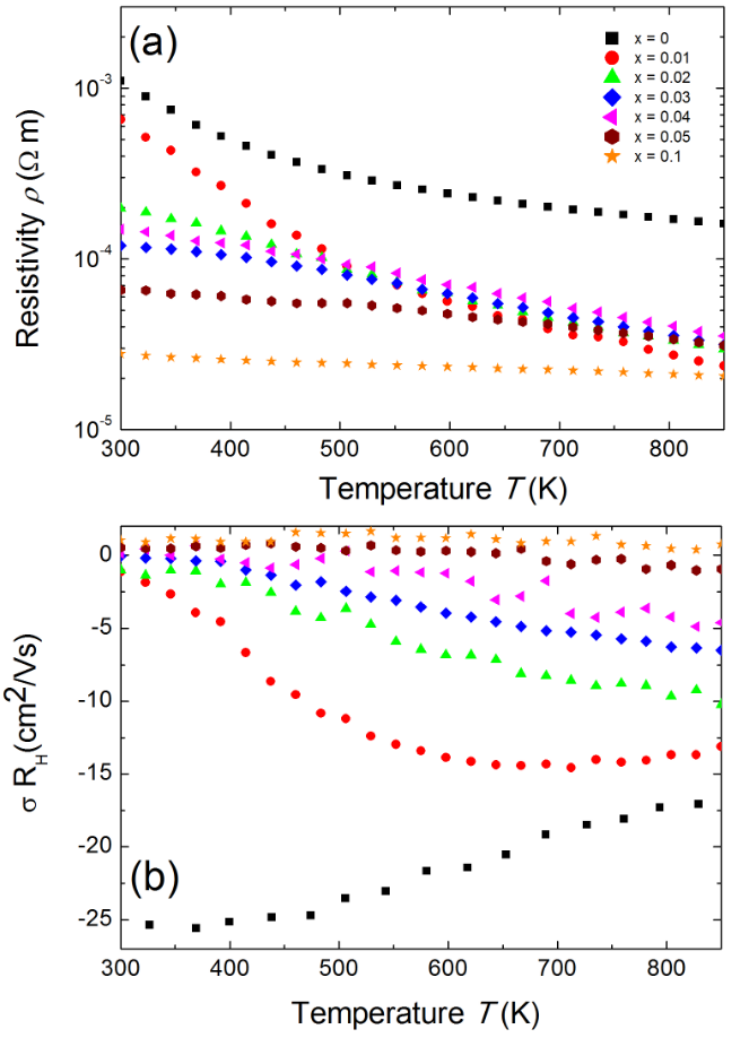

Figure 5-2: Temperature dependence of a) the electrical resistivity ρ and b) the $\sigma \cdot R_H$ product (in units of mobility) for the $Zr_{1-x}Sc_xNiSn$ solid solution. The true Hall mobility at 300 K is largest for the intrinsic, n-type ZrNiSn sample (Figure 2b). With increasing Sc content, the magnitude of $\sigma \cdot R_H$ decreases as a result of the influence of the low mobility holes.

With regards to the origin of the low mobility in the samples which show p-type character, we should consider how the Hall coefficient varies as a function of carrier mobility and concentration when both electrons and holes are present (in the bipolar region) [121]:

$$\boldsymbol{R_H = \frac{n_p\,\mu_p^2 - n_n\mu_n^2}{e(n_p\,\mu_p - n_n\mu_n)^2}} \qquad \text{Equation 5-2}$$

Here, n_n and n_p are the electron and hole concentrations, respectively, and μ_n and μ_p are the electron and hole mobilities, respectively. If $\mu_n > \mu_p$, the sign for the Hall coefficient R_H can be negative even if $n_p > n_n$. In order to obtain a positive value for the Hall coefficient, $n_p\,\mu_p^2$ needs to exceed $n_n\mu_n^2$. Because the individual carrier contributions, n_n and μ_n, are unknown and difficult to determine, $n_H = \frac{1}{R_H e}$ can be used assuming a single carrier type, but it will result in a value that is greater than either the true value for n_p or n_n.

The Seebeck coefficient (S) for the $Zr_{1-x}Sc_xNiSn$ solid solution are presented in Figure 5-3. The pure ZrNiSn compound, without any Sc doping, displays a large negative Seebeck coefficient at room temperature, indicating a significant *n*-type defect concentration in the intrinsic ZrNiSn. With increasing Sc content, the value of the thermopower decreases as compensating *p*-type defects (Sc^{3+} on Zr^{4+} sites) are added to the naturally *n*-type material. It can be seen from the change in sign of the Seebeck coefficient from negative to positive that holes become the predominant charge carriers. This sign change occurs at lower Sc contents than observed for the sign change of $\sigma \cdot R_H$ (Figure 5-2b), indicating that in this region of mixed conduction, the holes probably outnumber electrons. However, because the electrons are more mobile, the Hall coefficient does not change sign until x > 0.05 (note that the mobilities are squared, in Equation 5-2), whereas the Seebeck coefficient is weighted by mobility to the first power [121]. As observed in the Hall coefficient measurements, ambipolar conduction of both electrons and holes is most likely responsible for the decrease in the thermopower at high temperatures.

In addition to indicating the onset of bipolar conduction, the maximum of the thermopower can be used for the estimation of the Goldsmid-Sharp band gap, E_g, according to Equation 5-1 [51], where we obtained values on the order of 0.05 eV. This is much smaller than the band gap suggested by Aliev *et al.* from electrical resistivity measurements (0.18 eV) [237]. The discrepancy between the estimation from our thermopower data and the literature estimates for the size of E_g in these compounds will be discussed in detail later and is the main topic of this chapter.

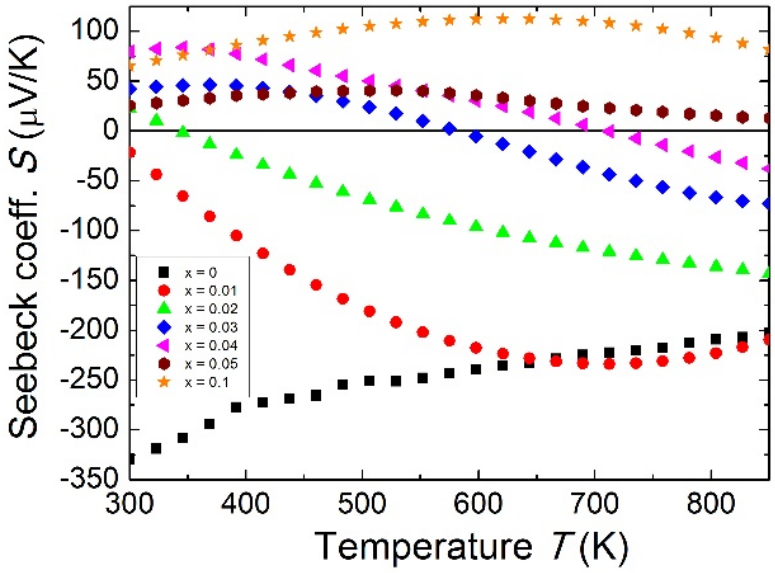

Figure 5-3: Temperature dependence of the Seebeck coefficient for the $Zr_{1-x}Sc_xNiSn$ solid solution. The Seebeck coefficient shows a rollover due to ambipolar conduction.

The temperature-dependent total thermal conductivity (κ_{tot}) is shown in Figure 5-4. The room temperature thermal conductivity is reduced by 40% for $Zr_{0.9}Sc_{0.1}NiSn$ as compared with the undoped ZrNiSn. Because this is accompanied by a decrease in resistivity (Figure 5-2a), it is clear that this must be due to scattering of phonons as a result of increased disorder in the material, which reduces the lattice thermal conductivity (which is much greater than the electronic thermal conductivity at room temperature for all Sc compositions). Above room temperature, we can see an increasing thermal conductivity for the doped samples, which is consistent with the existence

of both electrons and holes (which is also shown by the other transport properties in Figure 5-2b and Figure 5-3). This effect occurs at lower and lower temperatures as the Sc content is increased. Although the bipolar effects are less pronounced for the Seebeck coefficient measurements, thermal conductivity is quite sensitive to bipolar conduction because it is affected both by the decrease of the resistivity (increasing κ_e) and by the increasing magnitude of the bipolar term of the thermal conductivity (κ_b).

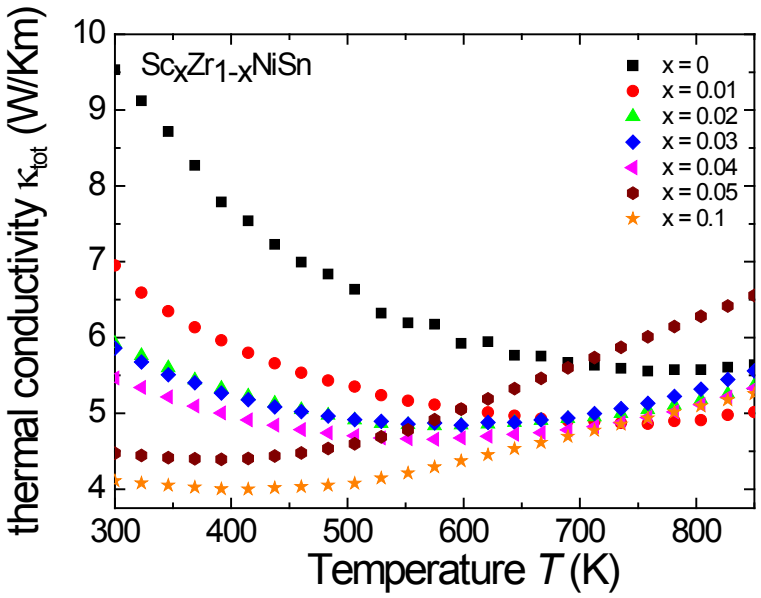

Figure 5-4 The total thermal conductivity of the $Zr_{1-x}Sc_xNiSn$ solid solution as a function of temperature and Scandium content.

Optical Absorption Edge

Optical properties were also measured to get information about the band structure, as displayed in Figure 5-5. Undoped ZrNiSn was measured in diffuse reflectance at room temperature; the indirect optical band gap was extracted using the Tauc extrapolation method (as discussed in Chapter 2) [46, 95]. The estimated value of the indirect optical band gap was 0.13 eV. Aliev et al. measured ZrNiSn samples previously using optical techniques, and they reported a minimum in the absorption coefficient of approximately 2000 cm^{-1} (~0.25 eV)—a value quite a

bit larger than the 0.13 eV measured in this work. The most likely reason for this discrepancy is due to the large difference in room temperature resistivity.

It should be understood that the properties of ZrNiSn seem to depend significantly upon the method of synthesis for the sample. I performed similar DRIFTS measurements on intrinsic ZrNiSn samples synthesized by another collaborator using levitation melting/spark plasma sintering (Xie et al. [230, 238]), which yielded a minimum in the absorption coefficient around 0.2 eV (which agrees with Aliev's results). However, I attribute the higher absorption edge to a higher carrier concentration (which increases free-carrier absorption and possibly induces a Burstein Moss shift, see Chapter 3) than those samples presented here (synthesized using arc-melting). ZrNiSn from both Xie et al. and Aliev et al. show a much lower room temperature resistivity (~$10\ m\Omega - cm$) in comparison to samples from this thesis chapter (intrinsic ZrNiSn showed resistivity near $100\ m\Omega - cm$). The lower resistivities from Xie, Aliev et al. resulted in free carrier absorption, which pushed the minimum absorption coefficient to higher energies.

Band gaps extracted from the temperature dependent resistivity both in this work and from Aliev et al. show values near 0.18 eV [237]. The large range of different values for the band gap obtained from different methods (Goldsmid-Sharp, temperature-dependent resistivity, optical, and DFT) suggest that a self-consistent model for this material should be developed. The remainder of this section will investigate the origins of the spread in values as well as the nature of the valence band. Recent literature suggests that the valence band is composed of Ni-states which arise from disorder within the material (specifically Ni disorder onto the vacancy site) [239], which will be discussed in detail later.

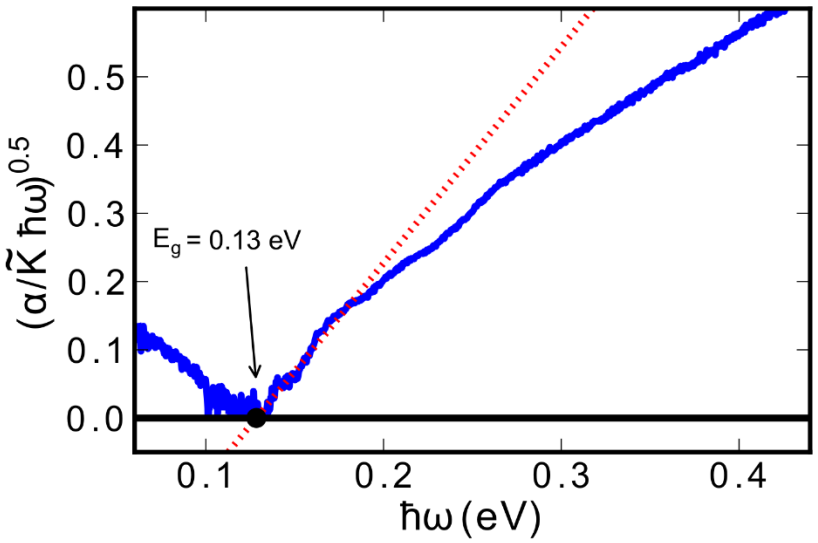

Figure 5-5: Optical diffuse reflectance data plotted as the indirect band gap transformation of the Kubelka Munk function for pure ZrNiSn. A linear fit (red dotted line) was used to estimate the band gap by extrapolating to zero absorption, indicating that the band gap is ~0.13 eV.

5.2c - Discussion

When considering the transition from heavily doped to intrinsically semiconducting behavior ($E_{g,thermal}$< ~5 $k_B T$), the Seebeck coefficient is an important indicator of the excitation of minority charge carriers across the band gap. In between these two regions, the thermopower will reach a maximum (as shown in Figure 5-3), which can be used to estimate the band-gap via the Goldsmid-Sharp formula (Equation 5-1). The thermopower band gaps for the $Zr_{1-x}Sc_xNiSn$ solid solution are compared to those of a series of *n*-type *X*NiSn samples from the literature, as shown in Figure 5-6.

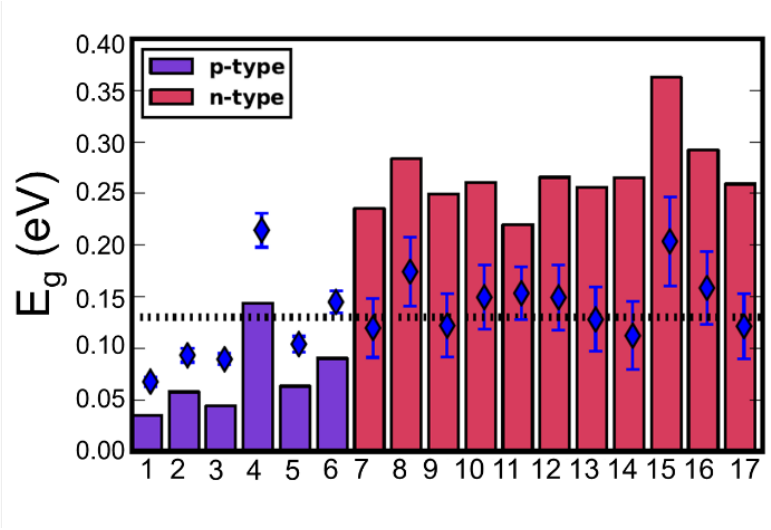

Figure 5-6: Estimation of the band gap for different n*-type (red bars) and* p*-type (purple bars) half-Heusler compounds using the Goldsmid-Sharp formula (E_g=2e$S_{max}T_{max}$) in units of eV* versus *the numerated HH compounds. The dotted line indicates the optical measured band gap of 0.13 eV for pure ZrNiSn from DRIFTS. The diamonds represent the estimated true band gap value that yields the experimental* S_{max} *and* T_{max} *when using the full relation derived by Goldsmid and Sharp for an electron-to-hole weighted mobility ratio of* A=*5 : 1)* $Sc_{0.03}Zr_{0.97}NiSn$*, 2)* $Sc_{0.04}Zr_{0.96}NiSn$*, 3)* $Sc_{0.05}Zr_{0.95}NiSn$*, 4)* $Sc_{0.1}Zr_{0.9}NiSn$*, 5)* $ZrCo_{0.08}Ni_{0.92}Sn$ *[230], 7)* $Zr_{0.75}Hf_{0.25}NiSn$ *[240], 8)* $Zr_{0.5}Hf_{0.5}NiSn$ *[240], 9)* $Zr_{0.25}Hf_{0.75}NiSn$ *[240], 10)* $HfNiSn$ *[240], 11)* $Ti_{0.95}Hf_{0.05}NiSn$ *[241], 12)* $Ti_{0.95}Hf_{0.05}NiSn_{0.995}Sb_{0.005}$ *[241], 13)* $Ti_{0.95}Hf_{0.05}NiSn_{0.99}Sb_{0.01}$ *[241], 14)* $Ti_{0.95}Hf_{0.05}NiSn_{0.98}Sb_{0.02}$ *[241] , 15)* $Hf_{0.75}Zr_{0.25}NiSn_{0.99}Sb_{0.01}$ *[242], 16)* $Hf_{0.5}Ti_{0.25}Zr_{0.25}NiSn_{0.99}Sb_{0.01}$ *[242], and 17)* $Hf_{0.25}Ti_{0.5}Zr_{0.25}NiSn_{0.99}Sb_{0.01}$ *[242].*

A large difference in the Goldsmid-Sharp thermopower band gap is apparent when comparing Sc-doped *p*-type samples from this work and *n*-type literature results; both are much smaller than those estimated by *ab initio* calculations (~0.5 eV) [232, 233]. Here, we note that even though X (in XNiSn samples) varies through Zr, Hf, and Ti for n-type samples in Figure 5-6, the band structures and band gaps are expected to be the same (for DFT calculations) [243]. Although the large difference in Goldsmid-Sharp band gap between p-type and n-type samples might lead one to the conclusion that the choice of dopant will affect the size of the gap, it is important to consider the limitations of the simple Goldsmid-Sharp band gap estimation and the parameters that might affect the results. Following the derivation from Goldsmid and Sharp, the Seebeck coefficient for a mixed semiconductor is given as:

$$\boldsymbol{S} = \frac{\mathbf{1}}{\mathbf{1}+\frac{\boldsymbol{\sigma_n}}{\boldsymbol{\sigma_p}}}\left(\boldsymbol{S_p} + \frac{\boldsymbol{\sigma_n}}{\boldsymbol{\sigma_p}}\boldsymbol{S_n}\right) \qquad \text{Equation 5-3}$$

where $S_{n,p}$ and $\sigma_{n,p}$ represent the thermopower and the electrical conductivities of the two different carrier types, respectively. Goldsmid and Sharp derived a simple model assuming classical statistics [51], which estimates the electron-to-hole conductivity ratio as

$$\frac{\boldsymbol{\sigma_n}}{\boldsymbol{\sigma_p}} = \boldsymbol{A}\, \boldsymbol{exp}(\boldsymbol{\eta_n} - \boldsymbol{\eta_p}) \qquad \text{Equation 5-4}$$

where η_n and η_p are the electron and hole dimensionless chemical potentials ($\eta = \xi/k_BT$). *A* is defined as the weighted mobility ratio:

$$\boldsymbol{A} = \frac{\boldsymbol{\mu_{0,n}} \boldsymbol{N_{v,n}}}{\boldsymbol{\mu_{0,p}} \boldsymbol{N_{v,p}}}\left(\frac{\boldsymbol{m^*_{b,n}}}{\boldsymbol{m^*_{b,p}}}\right)^{\mathbf{3/2}} \qquad \text{Equation 5-5}$$

where $\mu_{0,n,p}$ is the electron or hole mobility parameter, $N_{v,n,p}$ is the valley degeneracy of the electron or hole pocket (both are N_v=3 in this case), and $m^*_{b,np}$ is the single valley effective mass (not including the degeneracy of the valence or conduction bands). By substituting Equation 5-5 into Equation 5-4 and taking the classical limit for the thermopower and determining its maximum, Goldsmid and Sharp obtained:

$$\boldsymbol{S} = \frac{\boldsymbol{k_B}}{\boldsymbol{e}} \frac{\mathbf{2}\left(\eta_{\boldsymbol{n}} - \boldsymbol{\lambda} - \mathbf{2}\right)\left(\frac{\boldsymbol{E_g}}{\boldsymbol{k_BT}} + \mathbf{2}\boldsymbol{\lambda} + \mathbf{3}\right) - \left(-\eta_{\boldsymbol{n}} - \frac{\boldsymbol{E_g}}{\boldsymbol{k_BT}} - \boldsymbol{\lambda} - \mathbf{2}\right)}{\mathbf{1} + \mathbf{2}\left(\frac{\boldsymbol{E_g}}{\boldsymbol{k_BT}} + \mathbf{2}\boldsymbol{\lambda} + \mathbf{3}\right)} \qquad \text{Equation 5-6}$$

And

$$\boldsymbol{A}\exp\left(2\eta_n+\frac{E_g}{k_BT}\right)=\left(\frac{E_g}{k_BT}+2\lambda+3\right)\left[1\pm\sqrt{1-\frac{1}{\left(\frac{E_g}{k_BT}+2\lambda+3\right)^2}}\right] \quad \text{Equation 5-7}$$

where $\frac{E_g}{k_BT}$ is the dimensionless band gap, and λ is the scattering exponent ($\tau=\tau_0\left(\frac{E}{k_BT}\right)^{\lambda-1/2}$, where $\lambda=0$ is assumed here, which represents acoustic phonon scattering). In light of the large mobility difference between the *n*-type undoped sample (~25 cm^2/V·s) and the heavily Sc-doped samples (*p*-type mobility approximately ~1-2 cm^2/V·s), a weighted mobility ratio (*A*) that is larger than unity should be considered. In order to investigate this, we first gathered estimates of the Goldsmid-Sharp band gap for a series of *p*- and *n*-type ZrNiSn samples, as shown by the bars in Figure 5-6. A clear separation is observed between *p*-type samples (~0.05 eV) in purple and *n*-type ones in red (>~0.2 eV). For reference, the results from the optical measurements (Figure 5-5) are also plotted as a dashed line at ~ 0.13 eV in Figure 5-6. By numerically solving Equation 5-6 and Equation 5-7 with *A* = 5 (a reasonable value considering the large mobility difference between electrons and holes), I determined the required true band gap that would yield the experimentally measured S_{max} and T_{max}, the results of which are shown as blue diamonds in Figure 5-6 for each of *n*- and *p*-type ZrNiSn-based HH materials. All of these values are close to the measured optical results of 0.13 eV, indicating that the large weighted mobility ratio between electrons and holes is likely responsible for the difference in the Goldsmid-Sharp thermopower band gap when comparing *n*- and *p*-type samples.

Although Goldsmid and Sharp state that even with a large mobility ratio value (*A* = *10*), the simple band gap estimatie should hold to within ~ 20%, I note that they assume a band gap of 10 k_BT_{max} to derive this value. This is a poor assumption for most materials, as the band gap in our system (and many others) is less than 5 k_BT_{max}. In order to illustrate this point, I have plotted

$2eS_{max}T_{max}/E_g$ versus S_{max} for a mobility ratio parameter $A = 5$ using Equation 5-6 and Equation 5-7 (Figure 5-7). The red line is the result for the *n*-type doping and the blue line displays the result for *p*-type doping. Data points in their corresponding colors were taken from their Goldsmid-Sharp band gap ratio to the optical band gap measured in this work (0.13 eV). The displayed error bars assume an error of ±10% in the thermopower and the optical band gap. Figure 5-7 shows a positive deviation to $2eS_{max}T_{max}/E_g$ (relative to 1.0) for the *n*-type samples, whereas we observe a negative deviation for *p*-type ones. Qualitatively, this is because the more mobile electrons dominate the Seebeck coefficient equation because it is weighted by their conductivity; this forces the thermopower of the *p*-type samples to roll over at lower values, whereas *n*-type samples maintain higher thermopower at higher temperatures.

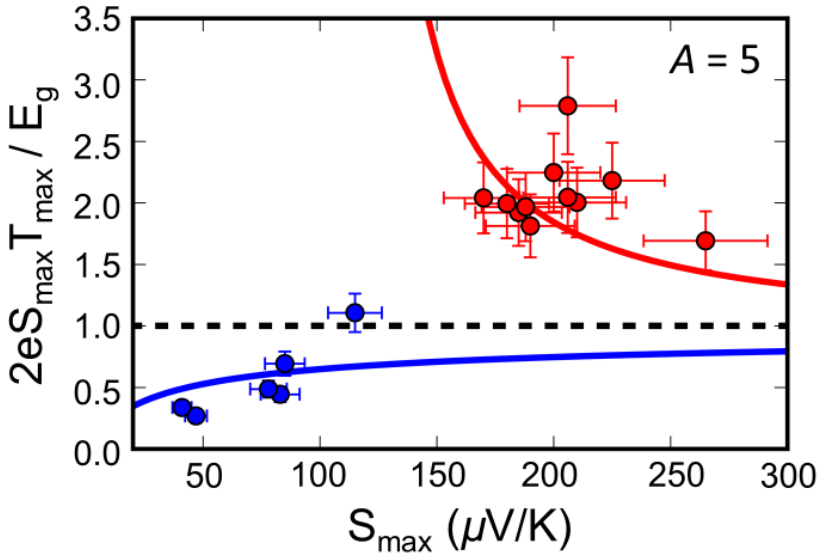

Figure 5-7: A plot of the ratio of the Goldsmid-Sharp band gap (2eSmaxTmax) to the true band gap for different p- and n-type half-Heusler compounds in red and blue, respectively, for an electron-to-hole weighted mobility ratio of A=5. The solid lines use Goldsmid and Sharp's full derived equation (Equation 5-6) to estimate this ratio. Experimental points use the observed maximum Seebeck coefficient (Smax) and temperature (Tmax) and the optical band gap (Eg,optical). An error of ±10% was assumed for the thermopower measurements and the band gap estimations.

To probe the supposition that $A \approx 5$ in ZrNiSn, an estimate for the weighted mobility ratio was obtained using a single parabolic band model as fit from the measured Seebeck coefficient and resistivity. Rather than fitting the effective mass and μ_0 (or equivalently E_{def}), in the absence

of Hall data the weighted mobility (or in our case reliable estimates of the carrier concentrations in either the valence or conduction band), σ_0 can be fit. σ_0 is the conductivity pre-factor (a constant that does not depend on Fermi level that scales the conductivity) which can be expressed as:

$$\boldsymbol{\sigma_0 = n_0 e \mu_0 = \frac{\sqrt{2}(k_B T)^{3/2} e}{\pi^2 \hbar^3} \mu_0 m_d^{*\,3/2}} \qquad \text{Equation 5-8}$$

which defines the relationship between the weighted mobility ($\mu_0 m_d^{*\,3/2}$) and σ_0. This can be fit in a similar way to the effective mass for a single parabolic band (as detailed in Chapter 2) by first using the measured Seebeck coefficient to estimate the reduced chemical potential (η), then solving for σ_0 using the measured conductivity and the relation:

$$\boldsymbol{\sigma = \sigma_0 F_0(\eta)} \qquad \text{Equation 5-9}$$

The weighted mobility ratio (*A*) can then simply be defined as the ratio of the σ_0 terms of *n*-type and *p*-type samples. I gathered data from both this thesis and XNiSn literature results (with both donor and acceptor doping) to estimate σ_0 values as a function of the measured electrical conductivityl the results are shown in Figure 5-8. In order to minimize the effect of bipolar conduction, it is desirable to estimate σ_0 for the samples with their Fermi levels as far from the opposing band as possible; this is accomplished in Figure 5-8 by taking our estimates for $\sigma_{0,CB}$ and $\sigma_{0,VB}$ for samples with the highest electrical conductivities (far left or far right of the plot respectively). This yields 1.1 and 0.15 $(m\Omega - cm)^{-1}$ for $\sigma_{0,CB}$ and $\sigma_{0,VB}$ respectively. I estimate a weighted mobility ratio given these values of σ_0 of $\frac{1.1}{0.15} \approx 7$, which agrees well with our estimate of 5 when considering the maximum Seebeck coefficients in Figure 5-6 and Figure 5-7. I should note that the exact values of σ_0 may depend on the method of synthesis (ideally we would obtain $\sigma_{0\,n,p}$ for data that all used the same synthesis procedure).

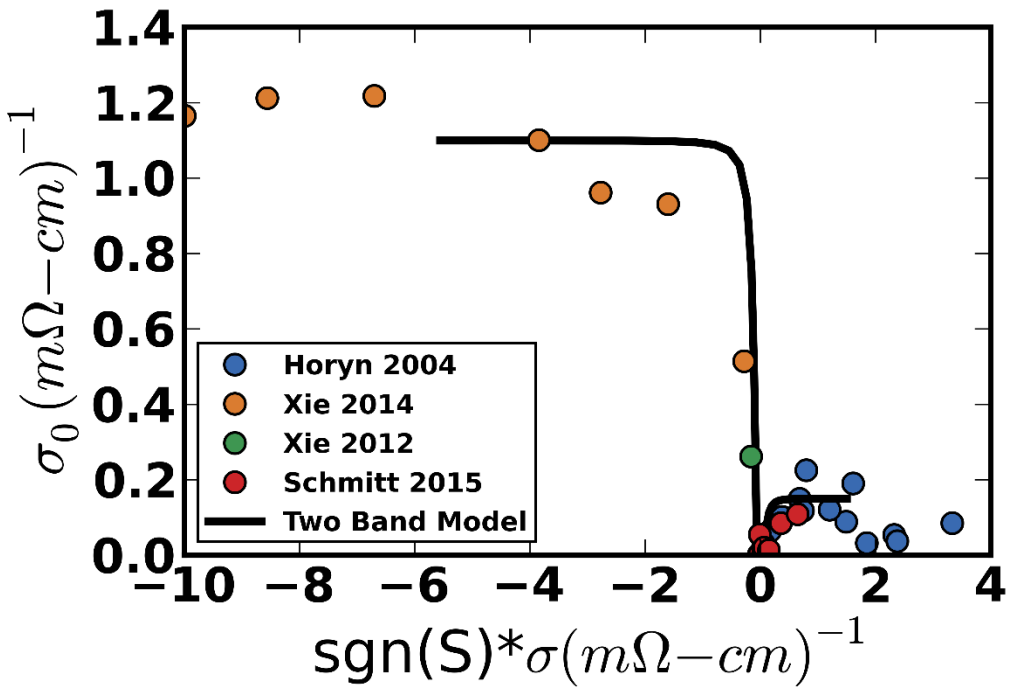

Figure 5-8: Estimated conductivity prefactor using the single parabolic band model for a series of ZrNiSn samples across the p-type to n-type transition for several literature results [52, 230, 231, 238] and results from this thesis. On the x-axis I have multiplied the measured conductivity by the sign of the measured Seebeck coefficient in order to represent the sign change. I have plotted these results along with a two-band (valence/conduction) model that uses a $\sigma_{0,CB} = 1.1$ and a $\sigma_{0,VB} = 0.15$ and a band gap of 0.128 eV (as measured optically).

5.2d - Electronic Structure Origin of the Weighted Mobility

Recent HH literature provides a self-consistent explanation as to the origin of the valence band states, which we have found to yield low weighted mobility. XNiSn materials upon addition of one more Ni atom per unit cell form the full-Heusler analog (XNi_2Sn). This crystal structure is known to be stable and produces a metallic material. Experimental evidence using synchrotron XRD and hard x-ray photoemission spectroscopy (HAXPES) confirms the existence of both in-gap states [239, 244] and up to ~10% of the Ni atoms disordering onto vacancy sites [239, 245]. Defect calculation results [246-249] confirm this theory, and the observed narrow gaps obtained from transport are sometimes attributed to these states [248, 250] in the literature.

With regards to the optical properties, the proposed picture of in-gap Ni states induced by disorder is consistent with the results obtained here. Even hydrogenic impurities, which contain a

single energy level rather than a band of states, can result in an absorption edge below the fundamental one which shows similar energy dependence to indirect transitions [45]. In the case of ZrNiSn, though, the Ni-vacancy interstitial defects result in an addition of a continuous band of states within the gap [249], which can also lead to appreciable absorption below the fundamental edge (as shown in $Zn_{1-x}Mn_xO$ [251]). The narrow gap observed both optically and electronically imply that these "in-gap" Ni-interstital states simply compose the valence band from which we observe optical transitions and bipolar effects. Further, the electronic properties and weighted mobility ratio could also be explained in this context. Specifically, the inherently disordered impurity band would likely have lower mobility than the conduction band states which are likely a result of the underlying HH framework and are probably less effected by the disorder.

While the very small mobility in *p*-type ZrNiSn may be disheartening, it beneficially results in suppressed bipolar effects in the *n*-type ZrNiSn system. Thus, the low mobility minority carrier enables the *n*-type material to maintain a high thermopower at high temperatures, despite its narrow band gap. This is contrary to the high band gaps found in other *p*-type HHs, like TiCoSb or the recently identified $FeV_{0.6}Nb_{0.4}Sb$, where the band gap estimated from the Goldsmid-Sharp formula is around 0.57 eV and 0.4 eV, respectively [22, 252].

5.2e - Conclusions

In this section, the thermoelectric transport properties of Sc-substituted ZrNiSn HH solid solutions were systematically studied. The substitution of Zr by Sc led to the successful introduction of holes into the system, resulting in a *p*-type material with a maximum thermopower of +115 μV/K at 650 K. Owing to the introduction of holes into the system, the Seebeck coefficient became positive and increased with increasing temperature, reaching a maximum as the higher mobility *n*-type carriers were thermally activated. Generally, the transport properties are dominated by the high mobility of the electrons over that of the lower mobility holes, which can be seen from the $R_H\sigma$ product (in units of mobility). Both *p*-type data from this work and *n*-type

literature data for the thermopower gap deviated significantly from the optical measurements (0.13 eV), but I have shown that this can be simply explained by a large difference in the weighted mobility between electrons and holes. A high electron-to-hole weighted mobility ratio leads to a suppression of the bipolar effect in the thermoelectric transport properties, which is essential for high zT values in *n*-type *X*NiSn (*X* = Ti, Zr, Hf) HH compounds.

5.2f - Methods

The solid solution $Zr_{1-x}Sc_xNiSn$ (x = 0, … , 0.10) was prepared by arc melting of stoichiometric amounts of Zr (99.99%), Ni (99.999%), Sn (99.999%), and Sc (99.999%) in an Ar atmosphere on a water-cooled crucible. To ensure compositional homogeneity, samples were flipped and remelted five times. The as-cast samples were annealed in evacuated quartz tubes at 1220 K for 7 d, followed by quenching in ice water to ensure the crystalline order. The crystal structure of the samples was studied by *X*-ray diffraction (*XRD*) on a Siemens D5000 diffractometer using Cu Kα radiation (wavelength of 1.5418 Å). The powder *XRD* patterns of all samples showed that they were a single phase with cubic $C1_b$ structure [233].

The total thermal conductivity was calculated from the thermal diffusivity (D) with $\kappa = C_P D d$, where C_p is assumed to be the Dulong–Petit heat capacity and d the density calculated from the molar mass and the lattice parameter for each sample obtained from XRD.

5.3 - Extending the Weighted Mobility Ratio/Goldsmid-Sharp Gap Analysis to Low Gap Materials

5.3a - Introduction

In the previous section, I have shown how to obtain an accurate estimate of the band gap using optical and electronic properties in ZrNiSn, which has a large difference in electron and hole weighted mobility. In this section, I extend the analysis beyond the non-degenerate limit to probe its applicability for narrow-gap semiconductors using full Fermi statistics. I also explore some of the assumptions made by Goldsmid and Sharp to show when the relation breaks down and what the expected error might be.

As mentioned in the previous section, band gap is an important parameter in thermoelectric materials because at high temperatures, bipolar excitation of carriers across the gap leads to a drastic drop in thermoelectric efficiency. This can be seen by plotting zT versus temperature for a good (representative) thermoelectric material (Figure 5-9 inset), the zT will rise until reaching a peak value after which it decreases. Since the peak zT values are often the metric by which materials are compared, it is worthwhile to understand the origins of the peak and what factors can influence it. Typical degenerate thermoelectric semiconductors display thermopowers which rise linearly with temperature to a maximum (Figure 5-9) followed by a decrease. Because the Seebeck coefficient is squared in the formula for zT, a maximum in the thermopower also results in a maximum in the temperature dependent zT.

It is well known that the origin of the thermopower peak is most often related to the onset of bipolar conduction which involves thermal excitation of both electrons and holes across the band gap. The contribution to the overall Seebeck coefficient by both the positive and negative charge carriers can be described by the conductivity weighted average (analogous to Equation 5-3):

$$S = \frac{\sigma_p S_p - \sigma_n |S_n|}{\sigma_p + \sigma_n} \qquad \text{Equation 5-10}$$

Because the minority carriers are (by definition) fewer in number, they will also have higher thermopower contributions (Seebeck coefficient is inversely proportional to carrier concentration). However, at low temperatures the population of minority carriers is small (low σ), meaning that they will not contribute much to the overall *S*. At higher temperatures, though, a broadening Fermi distribution leads to an exponential increase in minority carrier conductivity, resulting in a reduction (and therefore peak) in the thermopower [51].

The strength of bipolar conduction is determined by the value of the semiconductor band gap. Goldsmid and Sharp developed an analytical expression relating the band gap and the maximum thermopower, $|S|_{max}$, and the temperature at which it occurs (T_{max}) in the bipolar regime: $E_g = 2e|S|_{max}T_{max}$ (Equation 5-1) [51]. This simple method of estimating the band gap is ubiquitous in the thermoelectrics community because temperature dependent Seebeck coefficient is so commonly measured. Figure 5-9 shows a calculated temperature dependent Seebeck coefficient and corresponding *zT* (inset) for a valence (VB) and conduction band (CB) model with a band gap of 0.13 eV at various carrier concentrations. We can see that the Goldsmid-Sharp band gap formula accurately predicts the maximum thermopower over a wide range of carrier concentrations and temperatures (as demonstrated by the dashed line in Figure 5-9, which serves as a good upper bound for the thermopower at a particular temperature regardless of extrinsic doping concentration). For example, Bi_2Te_3 has a band gap of E_g~0.13 eV [253, 254] at room temperature (T_{max} = 300K), yielding a maximum Seebeck coefficient near 230 μV/K—a reasonable estimate [255-258].

While the Goldsmid-Sharp gap serves as a quick estimate of the band gap in a given material, it is important to understand where deviations might occur when using this analysis. In Goldsmid-Sharp's full equation (Equation 5-5), the weighted mobility ratio, *A*, is an important

parameter that governs how closely the Goldsmid-Sharp gap reflects the true gap. We can also consider *A* as the ratio of the quality factors, *B* (Equation 2-6), between the majority and minority bands. In the previous section, I showed that the Goldsmid-Sharp band gap can deviate by 50-100% of the optical band gap value mainly due to the large *A* values (*A*=5 for ZrNiSn) and an ϵ_g which is much less than 10 (<5 k_BT in ZrNiSn) [52]. Further, the non-degenerate limit (as assumed by Goldsmid and Sharp) is not applicable for many common thermoelectric materials: in Bi_2Te_3 [259, 260], PbSe [65, 261], and PbTe [28, 34], where the value of the band gap is actually 5-6 k_BT at T_{max}. In this section, I investigate the effects of a narrow ϵ_g by replacing Maxwell-Boltzmann statistics, applied in Goldsmid-Sharp's derivation, with Fermi-Dirac statistics which can more accurately represent semiconductor processes in narrow-gap ($\epsilon_g < 10$), doped materials. Ultimately, I present a chart that can be used to quickly estimate the relationship between the true gap and the Goldsmid-Sharp band gaps depending on *A* and S_{max}.

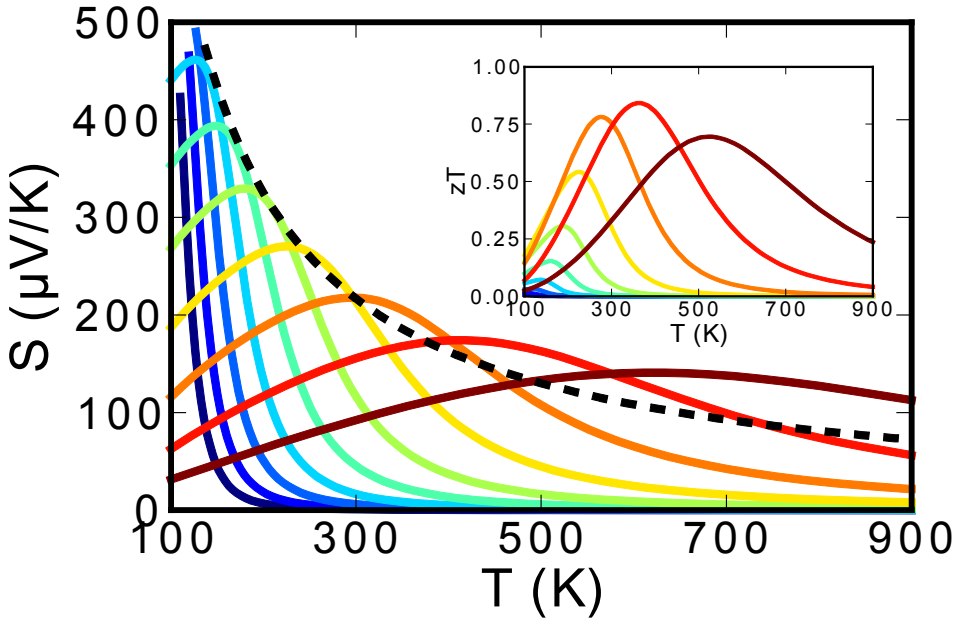

Figure 5-9: Calculated temperature dependent Seebeck coefficient and zT (inset) for various defect concentrations (N_A=p-n=1×10^{15} cm^{-3} in blue to 1×10^{20} cm^{-3} in red) for two parabolic bands with a band gap of 0.13 eV, m^*=1.0 m_0 , and $\mu_0(300\ K) = 820\ cm^2/Vs$ (VB and CB). Dashed line indicates the Goldsmid-Sharp band gap: $S = E_g/2eT$. The lattice thermal conductivity was estimated as $\kappa_L(T) = 1.7\ (300/T)$ W/m-K, and the following Umklapp scattering is used for the zT estimate.

5.3b - Theory

The Seebeck coefficient in a multi band (valence/conduction) system can be expressed by Equation 5-3 [62]. In the interest of maintaining general relationships applicable for either p or n-type materials, we chose to use the majority and minority carrier labels by changing the n- and p-type terms in Equation 5-3 into the majority (maj) and minority (min) charge carrier contributions; S_{maj}, and S_{min}, and σ_{maj} and σ_{min} are the majority and minority carrier Seebeck coefficients and conductivities, respectively. In the case of a primarily p-type material, the majority carrier will be holes. While Goldsmid and Sharp proceed assuming Maxwell-Boltzmann, non-degenerate statistics, $\frac{\sigma_{maj}}{\sigma_{min}} = A\exp\left(\eta_{maj} - \eta_{min}\right)$ (Equation 5-4), where $A = \frac{\mu_{0,maj}N_{v,maj}}{\mu 0_{,min}N_{v,min}}\left(\frac{m^*_{b,maj}}{m^*_{b,min}}\right)^{3/2}$, we will consider the Fermi integral solution to the Boltzmann transport equation (assuming scattering by acoustic phonons and parabolic bands) as presented in Chapter 2 for parabolic bands. In this context, the electrical conductivity ratio between the majority and minority carriers can be written as a function of the dimensionless chemical potential η in terms of the Fermi integral, F_j (as expressed in Chapter 2):

$$\frac{\boldsymbol{\sigma_{maj}}}{\boldsymbol{\sigma_{min}}} = \boldsymbol{A}\frac{\boldsymbol{F_0(\eta_{maj})}}{\boldsymbol{F_0(\eta_{min})}} \qquad \text{Equation 5-11}$$

Table 5-1 Description of the three different methods of estimating the maximum thermopower in this work.

Method Name	Criterion for Maximum	Statistics
Goldsmid-Sharp	$dS/d\eta = 0$	Maxwell-Boltzmann
Fermi	$dS/d\eta = 0$	Fermi
Exact	$dS/dT = 0$	Fermi

In order to find the maximum thermopower, several methods can be used as differentiated in Table 5-2. The derivation of the Goldsmid-Sharp band gap does not explicitly find the maximum

in thermopower with temperature; rather, Goldsmid-Sharp find a maxima with respect to reduced chemical potential, $dS/d\eta = 0$ which is equivalent to dS/dT=0 when $d\eta/dT$ is much larger than $\frac{d(\epsilon_g)}{dT}$ as pointed out by Goldsmid and Sharp [51]. In this work, the "Fermi" method also assumes $dS/d\eta = 0$, as in the "Goldsmid-Sharp" method, but it uses Fermi-Dirac rather than Maxwell-Boltzmann statistics. We can test the $dS/d\eta = 0$ approximation by performing a full temperature dependent calculation of the Seebeck coefficient: the "Exact" method. This is accomplished by applying a charge counting balance, $N_A - N_D = p - n$ (as described initially in Chapter 2, and in a later section of this chapter), at various temperatures, where N_A and N_D are the number of electron acceptors and donors, respectively (the difference of which was used as an input parameter), and p and n are the number of holes and electrons, respectively (for simplicity we have assumed that $A_{m^*} = \frac{m^*_{b,maj}}{m^*_{b,min}} = 1$, but we discuss the alternative later in this section; Figure 5-15). The full temperature-dependent, numerically calculated results ("Exact" method) will be presented along with the simpler $dS/d\eta$=0 solutions using both Maxwell-Boltzmann ("Goldsmid-Sharp") and Fermi-Dirac ("Fermi") statistics.

5.3c - Results

First, in order to probe the applicability Golsmid-Sharp's assumption of Maxwell-Boltzmann (non-degenerate) statistics, Figure 5-10 considers a weighted mobility ratio of A=1. Figure 5-10a shows the chemical potential dependent Seebeck coefficient (with η=0 being the valence band edge, $\eta = \epsilon_g = 5$ being the conduction band edge). As expected, the "Goldsmid-Sharp" result overlaps well with the "Fermi" result for chemical potentials in the gap ($0 < \eta < 5$), but deviations begin for chemical potentials of about 1.5 k_BT from either band edge, which become larger as the chemical potential becomes degenerate (chemical potential within the band, $\eta < 0$ or $\eta > \epsilon_g$). Upon varying the band gap, the value of chemical potential (η_{max}) that yields $dS/d\eta = 0$ can be obtained; the results are plotted in Figure 5-10b. For E_g>6 k_BT, η_{max} yield the same value

for the “Goldsmid Sharp” and “Fermi” methods, but deviations occur at smaller ϵ_g. Figure 5-10c shows the magnitude of the maximum Seebeck coefficient predicted using the three different methods (as shown in Table 5-2). The result is useful for estimating the maximum attainable thermopower at a given temperature (which would be set to T_{max}) for a material which has an electron-to-hole weighted mobility ratio (A) near 1.0 and a known band gap.

Figure 5-10d quantifies the effectiveness of the $2e|S|_{max}T_{max}$ estimate for band gap at different ϵ_g for the three cases of interest: the $dS/d\eta = 0$ models using both the “Fermi” and “Goldsmid-Sharp” methods, as well as the $dS/dT = 0$ (or “Exact”) case. For large ϵ_g, the “Fermi” and “Goldsmid-Sharp” solutions ($dS/d\eta = 0$) converge to $2e|S|_{max}T_{max}/E_g$ very near 1.0 (although the exact value is ~5% less at $\epsilon_g = 10$). However, as the band gap becomes small, (ϵ_g<~5), $2e|S|_{max}T_{max}/E_g$ increases for all three methods. The divergence for small gaps is a consequence of increasingly degenerate chemical potentials which yield the maximum thermopower. Experimentally, this would be observed for heavily doped samples that do not reach a maximum thermopower until very high temperatures (these details will be discussed thoroughly in a later section).

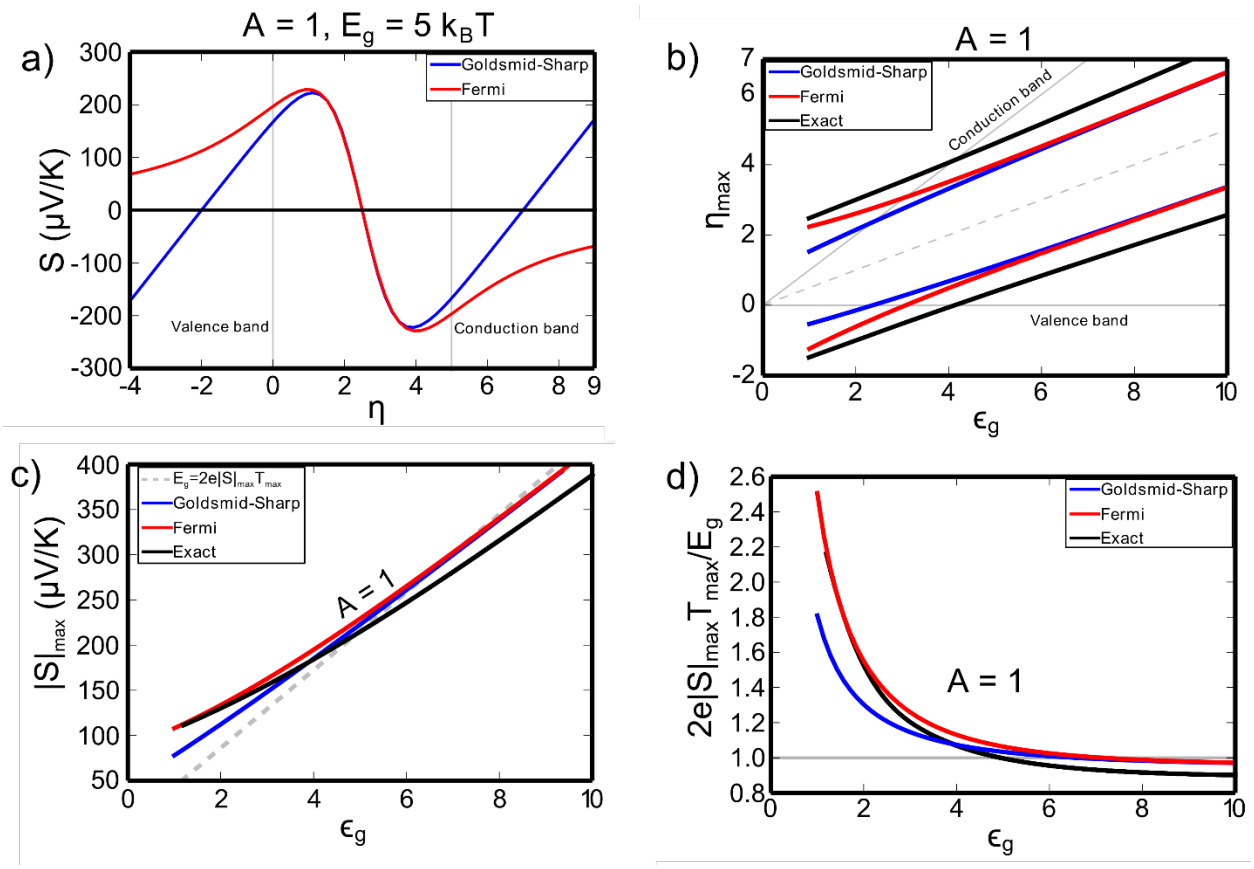

Figure 5-10: Result of the "Exact", "Fermi", and "Goldsmid-Sharp" methods calculated assuming ϵ_g= 5 and A=1 for a) Seebeck coefficient as a function of η. b) The reduced chemical potential which yields the maximum thermopower as a function of reduced band gap (at the rollover temperature) for A=1. c) Maximum attainable thermopower as a function of the reduced band gap (at A=1) for each method. d) the ratio of the $2e|S|_{max}T_{max}$ estimate to E_g as a function of the dimensionless band gap ϵ_g. The dashed line in c) represents the Goldsmid-Sharp band gap equation result. The dashed line in Figure (b) represents the halfway point between the VB and CB, and the solid grey lines represent the position of the valence and conduction bands. All calculations are done for a majority-to-minority carrier weighted mobility ratio of A=1.0.

The weighted mobility ratio (*A*) can also lead to deviations in the Goldsmid-Sharp band gap as illustrated in the previous section for ZrNiSn. While Bi_2Te_3 has similar majority and minority carrier weighted mobility [262, 263], other systems such as Si, Ge, and others [264] are believed to have values that exceed two (5 in the case of ZrNiSn). Figure 5-11 shows the *A*=5 and *A*=1/5 results for temperature dependent Seebeck coefficient ("Exact" method), analogous to Figure 5-9. It is clear that for *A*>1, the maximum Seebeck coefficient is larger than the Goldsmid-Sharp band gap result (dashed line) and that the converse is true for *A*<1.

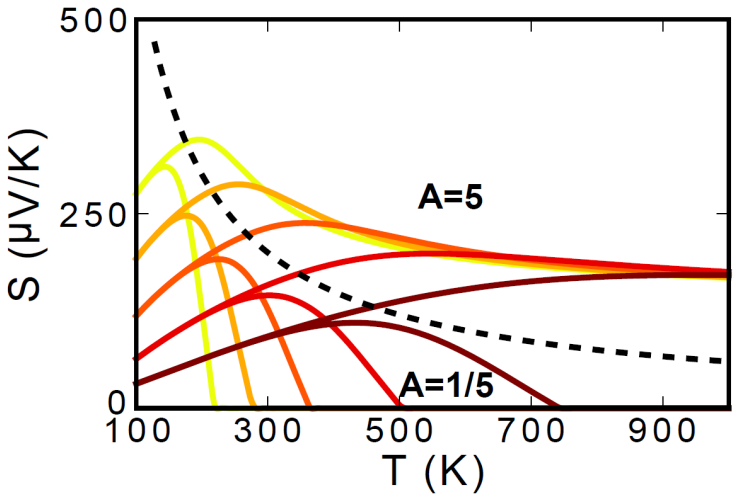

Figure 5-11: Temperature dependent Seebeck coefficient for various carrier concentrations (corresponding to Figure 1) for A=5 and A=1/5. Dashed line represents the Goldsmid-Sharp band gap: $S_{max} = E_g/2eT_{max}$ for various values of the carrier concentration; calculation details are the same as for Figure 5-9.

In order to illustrate the effect of an increasing weighted mobility ratio, the η-dependent Seebeck is plotted for ZrNiSn (ϵ_g~5 at room temperature) in Figure 5-12a. We see that the magnitude of the maximum Seebeck coefficient obtained for p-type ZrNiSn (A=1/5) is significantly lower than that for n-type ZrNiSn (A=5). The effect of having an A different from one is that the magnitude of the maximum Seebeck coefficient ($|S|_{max}$), as well as the temperature where it occurs (T_{max}), is increased for the carrier type with higher weighted mobility, while those of the lower weighted mobility carrier are decreased. So in a system like ZrNiSn, the n-type material maintains a high thermopower to much higher temperatures than might be expected from its narrow band gap (~0.13 eV) and therefore can reach an impressive zT from 0.5 – 1.0 [215, 265]. On the other hand, the p-type ZrNiSn prematurely experiences reduced thermopower due to compensating high-mobility electrons. Figure 5-12b shows the value of the maximum Seebeck coefficient for the three methods, a clear split in the A=5 and A=1/5 is observed. An alternative representation is shown in Figure 5-12c; here it is obvious that $2e|S|_{max}T_{max}/E_g$ is larger than 1.0 for all values of $\epsilon_g$when A=5, while it is less than 1.0 for all ϵ_g for A=1/5 (except at quite low ϵ_g, the reasons for which will be discussed later). Figure 5-12d shows how $2e|S|_{max}T_{max}/E_g$ increases with increasing A value; larger deviations are observed as E_g/k_BT becomes smaller. In comparison

with Goldsmid-Sharp's conclusion that only ~20% deviation is observed for *A*=10, we find that 50-100% errors in the estimated gap can be obtained for ϵ_g values reasonable for many relevant thermoelectric materials (ϵ_g~3-5).

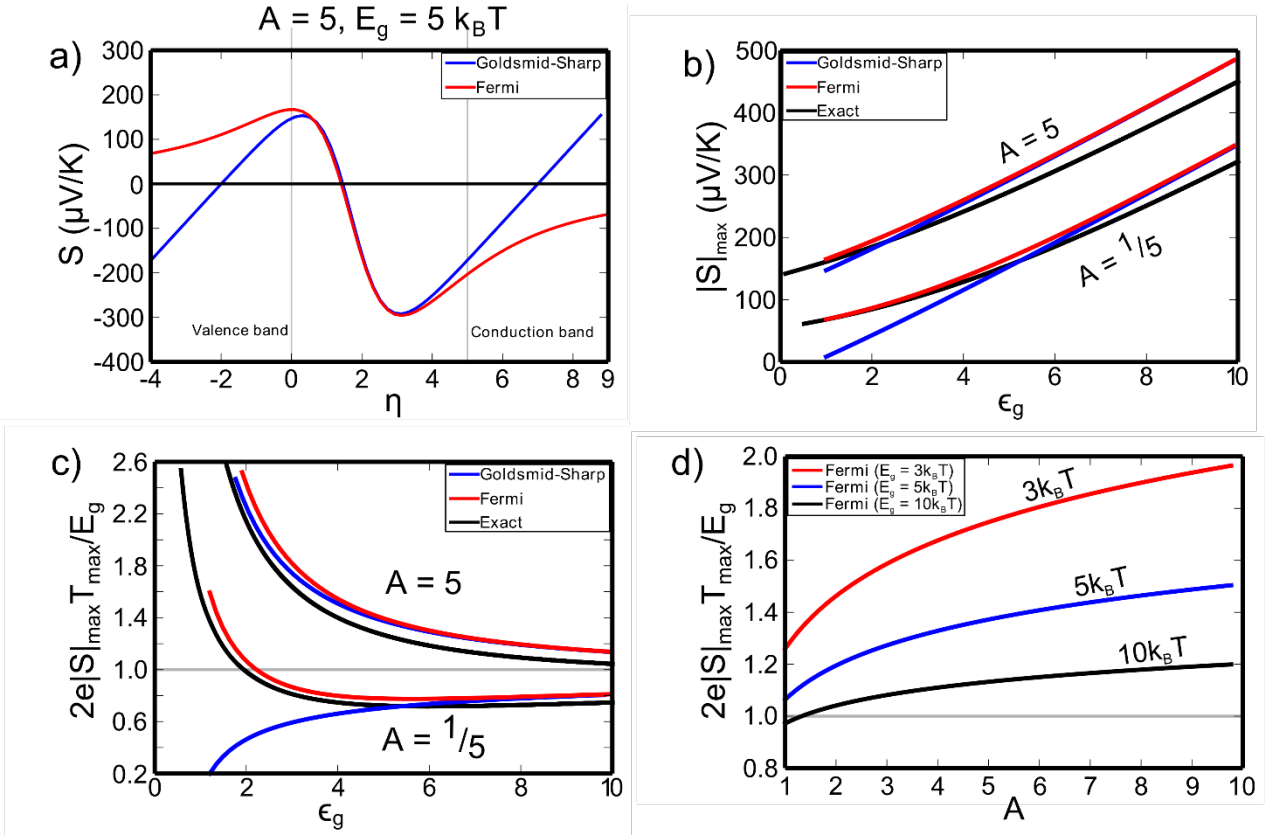

Figure 5-12: Result of the "Exact", "Fermi", and "Goldsmid-Sharp" methods assuming ϵ_g = 5 and A=5 (weighted mobility ratio) for a) Seebeck coefficient as a function of η,. b) The ratio of the 2e|S|maxTmax estimate to the actual model Eg as a function of ϵ_g, and c) the same ratio with respect to the weighted mobility ratio A at different Eg values (3, 5, and 10 kBT as indicated on the figure).

While the Goldsmid-Sharp band gap has proven to be a simple and useful estimate for the real band gap, it is not without its limitations. In this section, I have shown several cases for which this simple approximation breaks down. Figure 5-13 shows the deviation between the Goldsmid-Sharp band gap and the true band gap for a wide variety of these parameters. Ultimately, we observe that the magnitude of the deviation is largest for materials with large differences between the weighted mobility of electrons and holes ($A \neq 1$). From an experimental perspective, $A \neq 1$ will result in a larger value of $2e|S|_{max}T_{max}$ for the higher weighted mobility species, and a lower value for the one with lower weighted mobility. In the case of ZrNiSn, the

more mobile electrons (*A*=5) result in an observation of about a five-fold difference in the p-type (~0.05 eV) and the n-type (~0.25 eV) Goldsmid-Sharp band gaps as shown in the previous section[52].

Figure 5-13 can be useful in determining either an unknown *A* value for a material if the true band gap is known, or it can show the expected deviations of the Goldsmid-Sharp band gap relative to the true band gap for a given *A* value. For instance, in the case of n-type ZrNiSn with $2e|S|_{max}T_{max}/E_g$ = 2.1 (using $E_{g,optical}$=0.13 eV [52]) and observed maximum Seebeck coefficient (~200 *μ*V/K), we determine *A*~5. Alternatively, if the *A* is known, one can (based on the magnitude of the maximum Seebeck coefficient) obtain an estimate a value for $2e|S|_{max}T_{max}/E_g$ from Figure 5-13, which can be used to estimate the true band gap (as described in the numbered list below).

1. Measure temperature dependent thermopower and obtain a maximum
2. Calculate the Goldsmid-Sharp band gap: $E_g = 2e|S|_{max}T_{max}$
3. If $|S|_{max}$<150 μV/K, be aware that the true E_g may significantly differ from $2e|S|_{max}T_{max}$ (see below)
4. For $|S|_{max}$>150 μV/K, estimate the majority-to-minority carrier weighted mobility ratio, *A*.
5. Find the $2e|S|_{max}T_{max}/E_g$ ratio (*r*) from Figure 5-13 that is consistent with that *A* and S_{max} value to then calculate the corrected E_g=$2e|S|_{max}T_{max}/r$

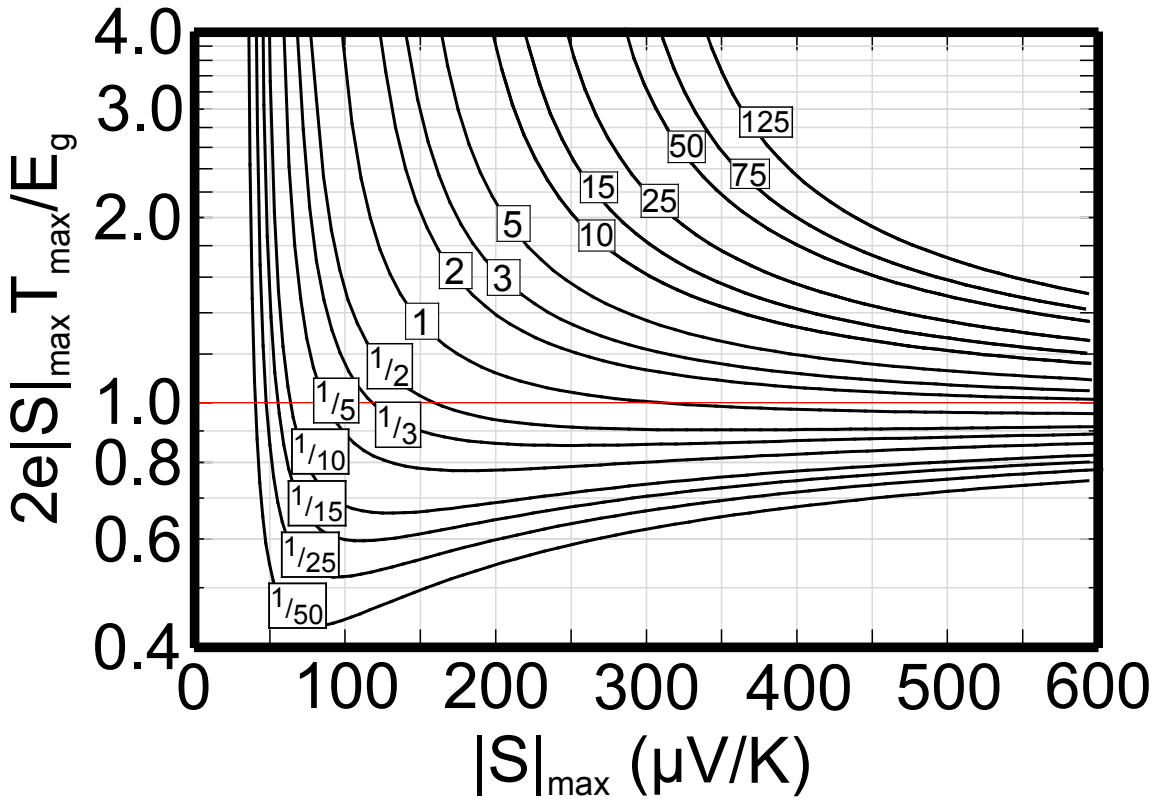

Figure 5-13: The ratio of the $2e|S|_{max}T_{max}$ estimate to the actual model E_g as a function of thermopower for a wide variety of A and S_{max}. A values are noted in a rectangular box laid on top of each black solid lines.

S_{max}<150 μV/K describes the degenerate crossover that leads to the upward trend in Figure 5-10, Figure 5-12, and Figure 5-13 mentioned previously for low values of S_{max}. For degenerate, heavily doped samples (η_{max} in the majority band) E_g/k_BT_{max} becomes a poor metric for describing the bipolar effects; rather, we believe the thermal band gap ($\frac{E_{g,thermal}}{k_BT} = \epsilon_g + \eta$) is the relevant parameter that describes the gap. This effect is even more pronounced as *A* is decreased because the lower mobility majority carrier requires a chemical potential deep within the band (large η) to mitigate the effects of a highly mobile minority carrier (see Figure 5-12b). In order to show the effects of degeneracy, I determined when the band gap that yielded a maximum thermopower and corresponded to a degenerate chemical potential (η_{max} within the band—see Figure 5-10b). I generate the analogous "Engineer's Guide" figure for determining the effectiveness of the Goldsmid-Sharp band gap (as a predictor of the thermal gap) as a function

of the maximum Seebeck coefficient, as shown in Figure 5-14. For many thermoelectrically relevant maximum Seebeck coefficients, the crossover to degenerate behavior (as indicated by a kink in $2eS_{max}T_{max}/E_{g,thermal}$) occurs. For nearly all A values, the kink (which means that degenerate η is required to reach $|S|_{max}$) occurs for Seebeck coefficients that are approximately 150 μV/K (which corresponds to a band gap of ~3 k_BT for the A=1 case as shown in Figure 5-10b). While the result seems to diverge for heavily-doped samples, nearly degenerate samples (S_{max}<150 μV/K) still give reasonable results in terms of relating the Goldsmid Sharp gap to the thermal band gap (rather than the true value) for the A=1 case.

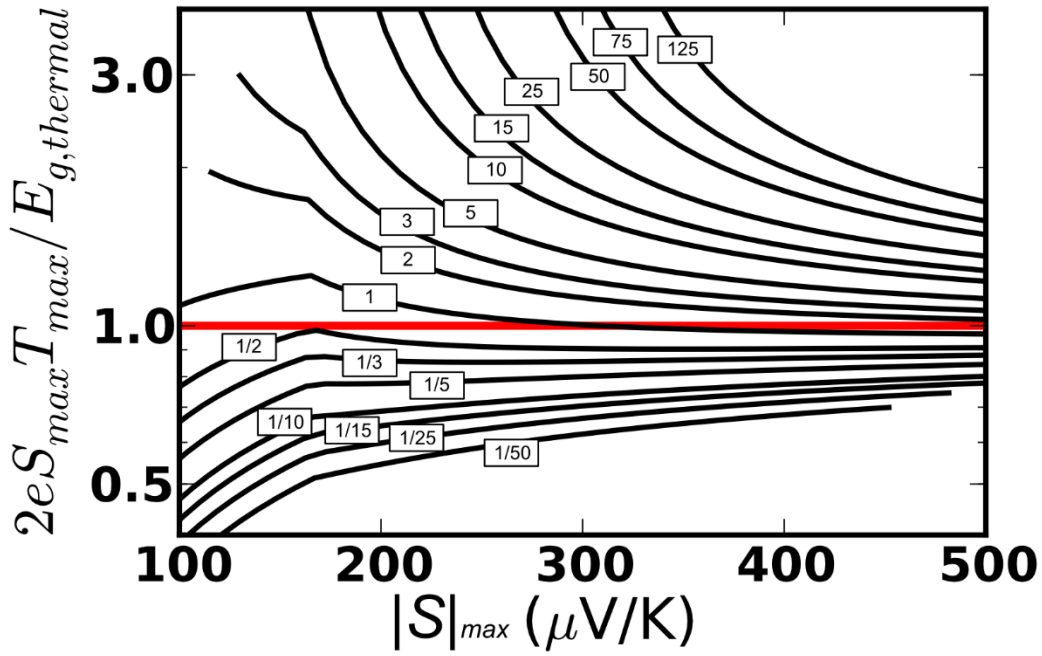

Figure 5-14: This engineer's guide is analogous to Figure 5-13; however, the predicted band gap in this case is ratioed to the thermal band gap ($\epsilon_g + \eta$, for degenerate or simply ϵ_g otherwise) instead of the true gap.

In terms of the weighted mobility ratio, up until this point we have not considered whether the difference in the weighted mobility ratio is due to the mobility (i.e., deformation potential) or effective mass: $A = A_{\mu}A_{m^*}^{3/2} = \left(\frac{\mu_{maj}}{\mu_{min}}\right)\left(\frac{m^*_{d,maj}}{m^*_{d,min}}\right)^{3/2}$. For the η-dependent Seebeck coefficient (and

therefore the Goldsmid-Sharp and Fermi methods which use $\frac{dS}{d\eta} = 0$ as a criteria for a maximum), these separate pieces are always coupled: $S = \frac{1}{1+\frac{\sigma_{maj}}{\sigma_{min}}}\left(S_{min} + \frac{\sigma_{maj}}{\sigma_{min}} S_{maj}\right) = \frac{1}{1+A_{\mu}A_{m^*}^{3/2}\frac{F_0(\eta_{maj})}{F_0(\eta_{min})}}\left(S_{min} + A_{\mu}A_{m^*}^{3/2}\frac{F_0(\eta_{maj})}{F_0(\eta_{min})} S_{maj}\right)$; but this is not the case for the charge neutrality equation. The charge neutrality equation can be expressed as $N_A - N_D = p - n$, which upon substituting the single parabolic band expression for *n* (Equation 2-8):

$$\boldsymbol{N_A - N_D = \frac{(2k_BT)^{\frac{3}{2}}}{m_{d,p}^{*\,3/2} 2\pi^2 \hbar^3}\left(F_{1/2}(\eta_p) - A_{m^*}^{3/2} F_{1/2}(\eta_n)\right)} \qquad \text{Equation 5-12}$$

While the "Fermi" and "Goldsmid-Sharp" methods do not explicitly use the charge neutrality equation, since they determine the maximum thermopower from $\frac{dS}{d\eta} = 0$, the "Exact" method uses charge neutrality to determine the maximum via $\frac{dS}{dT} = 0$. In order to probe the effects of a varying effective mass ratio (while keeping the overall *A* value equal to either 1 or 5), I have solved the *dS/dT*=0 "Exact" method for a variety of cases as shown below. First, I solved for the *A*=1 case, meaning that the majority and minority carriers have the same weighted mobility, but I vary $A_{m^*} = \frac{m_{maj}^*}{m_{min}^*}$ to smaller and larger values (while also varying A_{μ} to keep *A*=1) as shown in Figure 5-15a. On increasing the effective mass of the majority carrier band ($A_{m^*} > 1$), $2eS_{max}T_{max}/E_g$ increases slightly but is still close to the same chemical potential and S_{max} as in the $A_{m^*} = 1$ case. On the other hand, when the majority carrier band has a significantly lower effective mass ($A_{m^*} < 1$), the shift away from the $A_{m^*} = 1$ case is larger. The reduced chemical potential position (Figure 5-15b) seems to indicate that the reason for the shift in S_{max} is likely due to a shift in η_{max} towards the majority carrier band edge. This makes sense in the context of an increasing minority carrier effective mass which would push η away from the minority carrier band to maintain the same carrier concentration.

For the *A*=5 case (Figure 5-15c,d), a similar trend is observed. However, when the majority carrier has a much lower weighted mobility than the minority (*A*=1/5), the result is not affected as significantly by a varying A_{m^*}. In the case where the majority carrier has higher mobility (*A*=5), though, if the majority carrier effective mass is much smaller ($A_{m^*} < 1$) then reductions in the maximum Seebeck coefficient can occur (similar to the *A*=1 case). For these studies we have considered A_{m^*} of 2 or 5 (or the inverse), which is a feasible range of values for real semiconductors. In summary, we should expect larger, negative deviations in the Goldsmid-Sharp band gap if the majority carrier has higher weighted mobility (as in the *A*=5 case) but significantly lower effective mass ($A_{m^*} < 1$). Nearly negligible deviations occur in the other cases (either *A*<1 or $A_{m^*} > 1$).

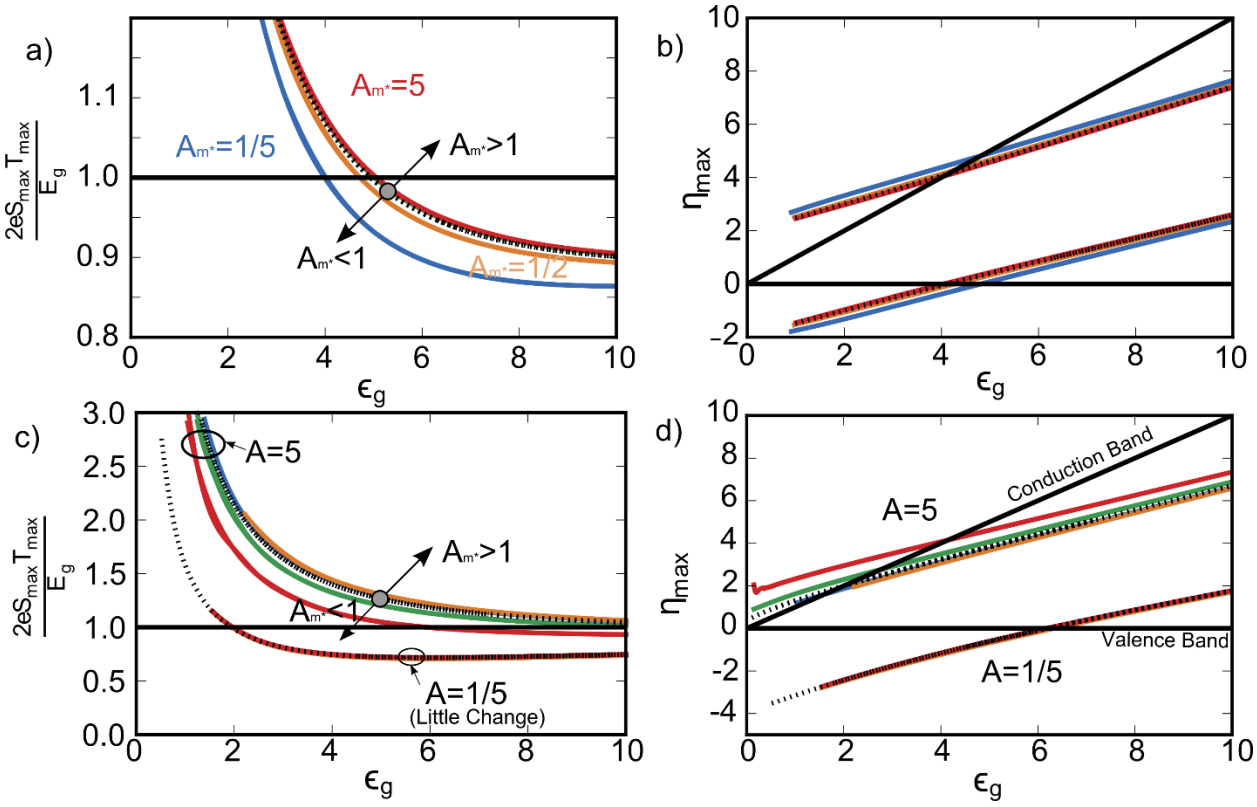

Figure 5-15: Comparison of results when the origin of A derives from a changing effective mass ratio (rather than purely mobility related). a) Goldsmid sharp effectiveness ($2eS_{max}T_{max}/E_g$) for A=1 with varying A_{m^*}. b) The varying chemical potential position for the parameters in a). c) $2eS_{max}T_{max}/E_g$ for A=5 and A=1/5 as a function of a changing contribution due to A_{m^*}. d) The resulting shift in the chemical potential that yields the maximum Seebeck coefficient as a function of reduced band gap for the A=5 and A=1/5 (colors correspond to c).

5.3d - Conclusions

In this section, I have developed the relevant theory for extending the Goldsmid-Sharp band gap to narrow gap semiconductors using the full Fermi statistics and an exact solution of the temperature-dependent transport integrals. I have investigated the effect of the weighted mobility ratio and narrow band gaps on the results, and I have shown several examples where the deviations can be quite large. These results should be generally applicable to any semiconductors (although the Engineer's guide should be rederived for alternative scattering mechanisms), but they should be particularly useful for the thermoelectrics community, which routinely measures the temperature-dependent Seebeck coefficient and utilizes the Goldsmid-Sharp band gap as a descriptor of the material's band structure. This analysis takes this one step further and allows us to more thoroughly describe discrepancies in this estimate.

5.4 - Conclusions

In summary, the Goldsmid-Sharp band gap ($E_g = 2e|S|_{max}T_{max}$) is an extremely useful tool for obtaining an estimate for a material's band gap through temperature dependent Seebeck measurements. However, in the case of ZrNiSn, these estimates yielded drastically different results when considering n-type and p-type material (a factor of 5 difference in the band gap). While most researchers understand that this is not an exact estimate, it is important to understand when and why the simple relation can break down and to what extent. In this work, we show that large deviations can occur for several reasons: a breakdown of Maxwell-Boltzmann statistics (used to derive the Goldsmid-Sharp band gap) for materials with narrow gaps, or materials with very large (or small) majority-to-minority carrier weighted mobility ratio (A). We use these analyses to explain the seemingly strange result in ZrNiSn in the context of a significantly larger weighted mobility for electrons than holes, which is likely a result of the nature of the valence band states which seem to be a result of Ni disorder. Because bipolar conduction is detrimental to thermoelectric performance, results from this work using the Goldsmid-Sharp gap could be used

to validate strategies for suppressing bipolar effects beyond altering band gap and doping [21, 110, 260, 266], but also by other methods (including nanostructures [11, 267-269]) given that both n-type and p-type samples are obtainable.

I would like to acknowledge the contributions of my collaborators, Jennifer Schmitt [52] and Hyun-Sik Kim [50], who were equally contributing coauthors on the paper's published based on work from this thesis chapter.