Topical Review

Thermoelectricity in correlated narrow-gap semiconductors

Jan M Tomczak

Institute of Solid State Physics, TU Wien, A-1040 Vienna, Austria
E-mail: jan.tomczak@tuwien.ac.at

Received 20 November 2017, revised 20 February 2018
Accepted for publication 27 February 2018
Published 10 April 2018

Abstract

We review many-body effects, their microscopic origin, as well as their impact on thermoelectricity in correlated narrow-gap semiconductors. Members of this class—such as FeSi and FeSb₂—display an unusual temperature dependence in various observables: insulating with large thermopowers at low temperatures, they turn bad metals at temperatures much smaller than the size of their gaps. This insulator-to-metal crossover is accompanied by spectral weight-transfers over large energies in the optical conductivity and by a gradual transition from activated to Curie–Weiss-like behaviour in the magnetic susceptibility. We show a retrospective of the understanding of these phenomena, discuss the relation to heavy-fermion Kondo insulators—such as Ce₃Bi₄Pt₃ for which we present new results—and propose a general classification of paramagnetic insulators. From the latter, FeSi emerges as an orbital-selective Kondo insulator. Focussing on intermetallics such as silicides, antimonides, skutterudites, and Heusler compounds we showcase successes and challenges for the realistic simulation of transport properties in the presence of electronic correlations. Further, we explore new avenues in which electronic correlations may contribute to the improvement of thermoelectric performance.

Keywords: thermoelectricity, strongly correlated electrons, skutterudites, Heusler compounds, Kondo insulator, intermetallics, FeSi, FeSb₂

(Some figures may appear in colour only in the online journal)
2.2.1. FeSi, other B20 compounds, and a comparison to oxides
2.2.2. The marcasites: FeSb₂, FeAs₃ & Co
2.2.3. The Kondo insulator Ce₃Bi₄Pt₃

2.3. Many-body models
2.3.1. Periodic Anderson model (PAM)
2.3.2. Hubbard models

2.4. Many-body electronic-structure theory
2.4.1. Theoretical description of experimental observables of FeSi
2.4.2. A microscopic understanding: spin fluctuations and Hund’s physics
2.4.3. Realistic many-body calculations for Ce₃Bi₄Pt₃

3.1. General considerations
3.1.1. Optimizing thermoelectric performance
3.1.2. Electronic correlations and thermoelectricity: friends or foes?
3.1.3. Limits of a purely electronic picture of thermoelectricity
3.1.4. Electronic correlations: a challenge for ab initio transport methodologies

3.2. Silicides: FeSi, RuSi and their alloys
3.2.1. Simulations of the thermopower
3.2.2. Influence of correlation-induced incoherence
3.2.3. Doped FeSi
3.2.4. A comparison of FeSi with RuSi

3.3. Marcasite di-antimonides: FeSb₂, CrSb₂ & Co
3.3.1. Polycrystalline samples—physics akin to FeSi?
3.3.2. Single crystals—do correlations cause the large thermopower?
3.3.3. Empirical evidence for a phonon-drag thermopower
3.3.4. Comparison to FeAs₂

3.3.5. Constraints to a theory of FeSb₂: correlation of features in transport observables
3.3.6. Modelling the phonon-drag in FeSb₂: evidence for important role of in-gap states
3.3.7. Existence and possible origin of in-gap states
3.3.8. Outlook

3.4. Skutterudite antimonides: a focus on CoSb₃
3.4.1. Experimental and theoretical knowledge
3.4.2. Doped CoSb₃

3.5. Heusler compounds: a focus on Fe₂VA₁
3.5.1. Experimental knowledge
3.5.2. Theoretical results

4. Conclusions and outlook

Acknowledgments
References

Avant-propos

Thermoelectricity is the striking ability of some materials to generate an electrical voltage when subjected to a temperature gradient (and vice versa). Prospective applications are multifarious, ranging from radiation and temperature sensors, vacuum gauges, to Peltier coolers and power generators. However, the current efficiency of thermoelectric conversion is insufficient for commercially viable coolers and generators on a larger scale. Indeed, for these applications, thermoelectric devices are so far limited to niche products for which flexibility in design, reliability, and maintenance-free operation trump aspects of cost. Examples are low-performance fridges, waste-heat recovery modules and power supplies for space probes.

Nowadays, most commercially used thermoelectrics are doped intermetallic narrow-gap semiconductors, built from post-transition metals, pnictogens, and chalcogens, e.g. Pb(Se,Te) and Bi₁₋ₓGeₓ, while at high temperatures group-IV element alloys, Si₁₋ₓGeₓ, are deployed. These materials combine good powerfactors S²/ρ (determined by the thermopower S and the resistivity ρ) with reasonably low thermal conductivities κ to reach a favourable thermoelectric conversion efficiency (measured by the figure of merit ZT = S²T/(κρ)). Indeed doped semiconductors present the optimal compromise between good conduction (ρ small for high carrier density) and a high thermopower (S large for low carrier densities), see figure 33. Much recent thermoelectric research has been devoted to reducing contributions from phonons to the thermal conduction via micro-/nano-/super-structuring or by finding crystal structures with intrinsically low κ. Here, we will instead focus on the powerfactor S²/ρ, which is controlled by the electronic structure of a material (see however section 3.3.6).

1 The prime example are the radiisotope thermoelectric generators aboard Voyager 1 that have been operating since 5 September 1977 and will continue to power the spacecraft—now well into interstellar space—until at least 2025.
One class of materials whose electronic structures give rise to a particularly extensive—and so far largely untapped—panoply of functionalities are correlated materials. In these compounds electrons interact strongly with each other, causing a high sensitivity of even fundamental properties to perturbations, such as changes in temperature, pressure, doping, and applied external fields. These stimuli can induce unusual behaviours, e.g. metal–insulator transitions, various types of long-range order, and, in particular, large response functions. These phenomena are, both, a great challenge for fundamental research and a harbinger for technical developments in sensors, switches, transistors and, indeed, thermoelectric devices.

Electronic correlation effects occur in systems in which orbitals of small radial extent—thus particularly those with maximal angular quantum number: $3d$, $4f$—are partially filled and thus strongly affect low-energy properties. In consequence, correlations effects are ubiquitous among many families of compounds, such as transition metals, their oxides, silicides, pnictides, chalcogenides, as well as lanthanide and actinide based materials. By exploiting various aspects of correlation effects numerous new functionalities were realized or have been proposed. Among them are new transparent conductors [1], oxide electronics [2], such as the Mott transistor [3, 4], electrode materials [5], eco-friendly pigments [6], or intelligent window coatings [7, 8].

Naturally the question arises as to whether correlation effects could also be used to overcome limitations of present thermoelectrics, all of which (are above) uncorrelated narrow-gap semiconductors. This question constitutes one of the interests in correlated narrow-gap semiconductors—the subject of this article. The review is organized as follows:

1. **Introduction**

In section 1 we give an overview of characteristic signatures of electronic correlation effects in narrow-gap semiconductors and detail in which material classes they can be found. Examples include transition-metal-based intermetallics, skutterudites, and Heusler compounds. Since basic observables in some of the identified compounds bear a cunning empirical resemblance to that of heavy-fermion Kondo insulators, we give in section 1.1 a quite extensive comparison of both types of materials, focusing on a contrasting juxtaposition of the prototype compounds FeSi and Ce$_3$Bi$_4$Pt$_3$. 

2. **Theories of correlated narrow-gap semiconductors**

A key part of this review is dedicated to the description and microscopic understanding of the electronic structure of correlated narrow-gap semiconductors. We begin section 2 with a review of early phenomenological proposals that advocate the importance of electronic correlation effects. Again, we focus the discussion on protagonist materials, such as FeSi, FeSb$_2$, and Ce$_3$Bi$_4$Pt$_3$. Section 2.2 presents the band-structure point of view for these materials. While highlighting the deficiencies of effective one-particle theories in describing the physics of our materials of interest, these first principle techniques provide key insights into the chemistry of bonding, and complexities related to crystal fields and other multi-orbital effects. Section 2.2.3 includes new material on the Kondo insulator Ce$_3$Bi$_4$Pt$_3$. Guided, or at least inspired, by realistic band-structure calculations, the physics of electronic correlations was investigated within many-body models, such as that due to Hubbard, Kanamori and Gutzwiller, as well as the periodic Anderson model. Works that are directly linked to our compounds of interest are reviewed in section 2.3. Next, we discuss in section 2.4 the current understanding of the physics of correlated narrow-gap semiconductors from the perspective or realistic many-body simulations. In section 2.4.1, we first showcase results from state-of-the-art dynamical mean-field methodologies for the example of FeSi. Indeed, a vast array of observables can be simulated in quantitative congruence with experiment: photoemission spectra, optical conductivities, neutron spectra, and—as discussed in section 3.2—transport properties, including the thermopower. After such validation of the theoretical framework, we discuss in section 2.4.2 the microscopic insights that have been extracted from the numerical calculations. In section 2.4.3 we further include new realistic many-body simulations for the Kondo insulator Ce$_3$Bi$_4$Pt$_3$. For paramagnetic insulators with partially filled $d$ or $f$-shells we then (i) introduce a microscopic classification from the dynamical mean-field perspective (section 2.5.1), and (ii) propose a simple measure for the occurrence of Kondo-characteristics in said systems (section 2.5.3). Equipped with these tool, we then compare the ab initio results for FeSi and Ce$_3$Bi$_4$Pt$_3$ (section 2.5.4) and propose an orbital-selective Kondo-scenario for FeSi. Finally, section 2.7 discusses the potential influence of the lattice degrees of freedom onto the electronic structure.

3. **Thermoelectricity**

After a brief introduction to thermoelectricity in section 3.1, we review in section 3.1.2 the potential influence of electronic correlations onto thermoelectric performance. In section 3.1.3 we investigate the impact of basic many-body renormalizations in semiconductors and motivate that the thermopower has an upper bound if excitations are coherent. Effects of incoherence (finite lifetimes) are studied in section 3.1.4 and demonstrated to be beyond semi-classical Boltzmann approaches. Then, we review experimental findings and theoretical simulations for transport and thermoelectric properties for representative materials among silicides (section 3.2), antimonides (section 3.3), skutterudites (section 3.4), and Heusler compounds (section 3.5). In particular, we discuss the impact of finite lifetimes and effective mass renormalizations for FeSi, the proposal that the colossal thermopower of FeSb$_2$ is driven by the so-called phonon-drag effect, as well as the pitfalls of electronic structure theory and the influence of defects and disorder in CoSb$_3$ and Fe$_2$VAl.

4. **Conclusions and outlook**

Finally, we summarize the covered material, point out open challenges, and identify possible directions of future research.
Table 1. Collection of experimental data of prototypical narrow-gap semiconductors. $\Delta_{\text{dir}}$, $\Delta_{\text{indir}}$, $\Delta_S$: direct (optical), indirect (transport), and spin gap; $S_{\text{max}}$, $(S^2/\rho)_{\text{max}}$: peak values of the thermopower and the powerfactor at indicated temperatures. ‘not act.’ = no activated behaviour up to largest $T$ measured.

| Structure/material | $\Delta_{\text{dir}}$ (meV) | $\Delta_{\text{indir}}$ (meV) | $\Delta_S$ (meV) | $S_{\text{max}}$ (mV K$^{-1}$) | $(S^2/\rho)_{\text{max}}$ (µW K$^{-2}$ cm$^{-1}$) | See section |
|---------------------|-----------------------------|-------------------------------|-----------------|-----------------------------|-----------------------------------------------|------------|
| B20 structure       |                             |                               |                 |                             |                                               |            |
| FeSi                | 60–73 [9]                   | 50–70 [10, 11]                | 50 [12, 13]     | 0.5@50 K [14]              | 42.1@62 K [14]                                | 1.1, 2.2, 1, 1.2 |
|                     | [15, 16]                    |                               |                 | 1.2@32 K [17]              |                                               | 2.4, 3.2   |
| RuSi                | 200–400                     | 310 [18]                      | not act. [19]   |                             |                                               | 2.2.1      |
|                     | [18, 20]                    | 260 [21]                      |                 | 0.27@286 K [21]            | 8.7@347 K [21]                                | 1.1, 3.2   |
| Marcasites          |                             |                               |                 |                             |                                               |            |
| FeSb$_2$            | 130 [22]                    | 30 [23]                       | 32 [12]         | −45 [23]                   | 2300@12 K [23, 24]                           | 1.1, 2.2.2  |
|                     | [37–48 [25]                 |                               |                 | −18 [26] −25 [27]          | 8000@28 K [28]                               | 3.3        |
| RuSb$_2$            | 790 [22]                    | 290 [24] 330 [29]            | not act. [24]   | −3.1@15 K [24]             | 40@32 K [24]                                 | 1.1, 2.2.2  |
| FeAs$_2$            | 200–220                     | 200 [24]                      |                 | −7@12 K [24]               | 153@30 K [24]                                | 1.1, 2.2.2  |
| FeP$_2$             | 370–400                     | not act. [32]                 |                 |                             |                                               |            |
| CrSb$_2^a$          | 50 [33], 70 [29]            | 50 [33]                       | −4.5@18 K [33]   | 46@24 K [33]               |                                               | 1.1, 3.3   |
| CoGa$_3$-type structure |                         |                               |                 |                             |                                               |            |
| FeGa$_3$            | 600 [34]                    | 260 [35]                      | 300–400 [36]    | −0.35@350 K [37]           | 0.003@100 K [37]                             | 1.1, 3.3   |
|                     | 400–500 [37, 38]            | 290–450 [39]                  |                 | −0.45@300 K [40]           | 0.2@400 K [40]                               |            |
|                     | [36, 39, 41]                |                               |                 | −15@12 K [41]              |                                               |            |
| RuGa$_3$            | 650 [34]                    | 320 [35]                      | 0.3@566 K [35]  | 3.8@920 K [35]             |                                               | 1.1        |
| Skutterudites       |                             |                               |                 |                             |                                               |            |
| CoSb$_3$            | 230 [42]                    | 230–700 [42, 43]              | not act. [44]    | 0.2@300 K [45]             | 9.5@625 K [46]                               | 1.1, 3.4   |
|                     | [45, 47]                    |                               |                 | −0.4@450 K [48]            |                                               |            |
| Hesseur compounds   |                             |                               |                 |                             |                                               |            |
| Fe$_2$VAl           | 100–200                     | 90–130 [49, 50]               | 270 [51]        | −0.14@200 K [50]           | 3.3@300 K [52]                               | 1.1, 3.5   |
|                     | [53, 54]                    |                               |                 | doped:68@300 K [55]        |                                               |            |
| Oxides              |                             |                               |                 |                             |                                               |            |
| LaCoO$_3$           | 100 [56, 57]                | 100 [56, 57]                  | 30 [56]         | 1.2 [58] 33@80 K [59]      | 0.136@80 K [59]                              | 2.2.1, 2.4.2, 2.5 |
| Kondo insulators    |                             |                               |                 |                             |                                               |            |
| Ce$_3$Bi$_4$Pt$_3$  | 37 [60]                     | 8–12 [61, 62]                 | 12 [63]         | 0.055@23 K [64]            | 1.67@30 K [64]                               | 1.2, 2.2.3  |

$^a$ CrSb$_2$ is antiferromagnetic below $T_N = 273$ K [65, 66].

1. Introduction

Correlated narrow-gap semiconductors are systems with small band-gaps in which many-body effects play an important role. In the following section 1.1 we will give an overview over what typical signatures of such correlation effects are and in which kind of materials one can expect to find them. In section 1.2 we will give a more detailed account of the intriguing physical observations, but focus the discussion on a one-to-one comparison of a prototypical correlated narrow-gap semiconductor, FeSi, and the archetypical Kondo insulator Ce$_3$Bi$_4$Pt$_3$, with which it shares many empirical similarities.

1.1. Correlated narrow-gap semiconductors: a brief overview

A paradigmatic class of materials with strong electronic correlations are 3$d$ transition metal oxides [67]. They boast extremely rich phase diagrams, indicative of competing energy scales. Indeed, the electronic structure of these oxides is dominated by the interplay of a relatively strong ionic character of bonding (causing large crystal-field splittings), and strong electron-electron interactions (Mott and Hund physics). As a result—while they can be induced by small changes in external conditions—insulating phases in transition metal oxides have comparatively large gaps ($\Delta \gtrsim 0.5$ eV). Therefore, this review does not mention oxides, with the notable exception of LaCoO$_3$. In transition metal intermetallics, on the other hand, the bonding has a much stronger covalent character (see sections 2.2.1 and 2.5 for a discussion of ionic versus covalent compounds). Also, in most insulating intermetallics the gap is determined by bonding and/or anti-bonding states formed by the d-orbitals of a transition metal and the p-orbitals of a group 13–15 element. Such hybridization gaps in covalent insulators are typically rather narrow ($\Delta \lesssim 0.5$ eV, see table 1 for examples). Insulating intermetallic compounds that contain transition metal atoms with partially filled d-shells constitute the pool of materials that we focus on in this review. We will group materials into classes according to their crystal-structure and discuss the electronic structure and thermoelectric effects of representative members in sections 2 and 3, respectively. To be specific, the chosen prototypical classes and representative compounds are:

- cubic B20-structure: FeSi, RuSi (see sections 1.2, 2.2.1 and 3.2)

[Table 1 and sections of text about correlated narrow-gap semiconductors and other classes of materials are included here.]
The following gives a brief overview of basic experimental findings for the above classes of correlated narrow-gap semiconductors. Besides, for the above materials, we collect in table 1 a list of experimental data, such as the direct (optical), indirect, and spin gaps, as well as peak amplitudes of thermopowers and powerfactors. In that list, we have further included a single oxide: LaCoO3. While exhibiting similarities in some physical observables (see figures 1 and 2), the microscopic physics underlying them are pronouncedly different than in all other compounds listed here. These differences will be discussed in sections 2.2.1 and 2.4.2 as well as section 2.5.

1.1. Signatures of correlation effects. Correlated narrow-gap semiconductors display a number of intriguing properties. In particular the temperature-dependence of transport, spectroscopic and magnetic quantities is highly unusual.

**Charge degrees of freedom.** Basically all materials considered here exhibit one or more temperature regimes in which the resistivity follows activated behaviour, as expected for conventional semiconductors, see figure 1. In systems with comparatively large gaps ($\Delta \gtrsim 200$ meV), e.g. FeAs2, RuSb2, or FeGa3, there can be several regions that are amendable to an activation-law fit with gaps of different sizes. The largest of the latter will give a description of the activation regime that is highest in temperature and is identified as the fundamental gap, oftentimes in concurrence with the direct gap extracted from optical experiments. The smaller gaps are then interpreted as arising from impurity or defect states inside the intrinsic gap. The different activated regimes may be connected by regions in which the slope of the resistivity is positive, albeit in absence of any metallic characteristics in optical or photoemission spectra.

The situation is different for the materials on our list that have the narrowest gaps ($\Delta \lesssim 100$ meV), e.g. FeSi or FeSb2. These systems behave as renormalized but coherent semiconductors at low temperatures. However, activated behaviour persists only up to a temperature $T^{*}$ that is still significantly smaller than the respective gap $\Delta$: $k_b T^{*} \ll \Delta$. Here, we arbitrarily define $T^{*}$ as the temperature where the relative deviation from activated behaviour, $(\rho - \rho_{act})/\rho_{act}$, exceeds 3%, see figure 4(a) for the case of FeSi. There, the resistivity clearly surpasses the exponential decay above $\sim T^{*} = 160$ K and even develops a positive slope, while $k_b T^{*} = 14$ meV $\ll \Delta \approx 50$ meV. Concomitantly, optical [9, 16, 22, 69–71], Raman [72], photoemission [73–77] and tunnelling [78] spectroscopy witness in FeSi and FeSb2 a clear crossover to a metallic phase, see figures 5(a) and (b). These results are in clear defiance of a mere thermal activation across a fundamental gap. Moreover, with changing temperature, transfers of spectral-weight in the optical conductivity occur over energy-scales much larger than the size of the gap [9, 16, 22, 71], suggestive of many-body effects [79] that are controlled by large interactions $U$ rather than the size of the gap, $U \gg \Delta$. This will be discussed in more detail in section 1.2.

Besides the intriguing temperature dependence, also the energies of one-particle excitations, such as the gap, are beyond expectations from band-theory: In the case of FeSi, the low temperature gap as well as band-dispersions are largely overestimated by density functional theory (DFT [80, 81]) (see section 2.2), indicative of notably enhanced masses $m' / m_{\text{band}}$ caused by dynamical correlations.

The mentioned effects are most pronounced for the materials that have the smallest gaps. Within a given class, these are the systems built from a 3$d$ transition metal and a group 13–15 element with large atomic radius. Indeed, correlation effects, and thus band-width and gap renormalizations, are expected to be less pronounced for transition metals with more extended 4$d$-orbitals. For comparisons of 3$d$-transition metal systems to their 4$d$ homologues, see sections 2.2.1 and 2.2.2 for the cases of silicides and antimonides, respectively.

Also at very low temperature deviations to activated behaviour occur in the resistivities, see figure 1. Indeed, they show pronounced tendencies towards saturation for basically all materials considered here. Commonly, this behaviour is ascribed to the presence of impurity or defect states inside the gap that pin the
chemical potential and supply residual conduction. This scenario is supported by the presence of different activation regimes in some compounds (see above), as well as a strong sample dependence of low-temperature transport observables. However, finite lifetimes of intrinsic valence and conduction states—ubiquitous in correlated materials—provide an alternative mechanism for resistivity saturation, as will be discussed in section 3.1.4.

The importance of electronic correlation effects in narrow-gap intermetallic semiconductors with open 3d-shells is further strengthened by the following: doping the discussed stoichiometric compounds yields metallic phases in which mobile charge carriers exhibit largely enhanced effective masses. Some examples are:

- FeSi1-xAlx [82–85] yields masses 14 times larger than the free electron value, as extracted from, both, the specific heat and the susceptibility for x = 1%–5% [83].
- Fe1-xCoSi has masses up to m*/me = 30 [70], as extracted from the specific heat.
- FeSb2-xTe yields m*/me = 10–30 as extracted from the thermopower [86].
- FeSb2-xSn, exhibits for x > 0 masses m*/me = 10–15 as extracted from the specific heat [87].
- CoSb3’s Hall coefficient [88], as well as the thermopowers of Co1-xNi3Sb3 [89, 90] and CoSb3 with Yb-fillings [42] display enhanced masses: m* ≈ 2–5me.
- Fe2VAI: the residual in-gap density in typical (possibly off-stoichiometric) samples causes a metallic specific heat from which effective masses enhanced by a factor of five have been extracted [91].

**Magnetic properties.** The intriguing insulator-to-metal crossover in the small-gap materials is accompanied by large changes in the magnetic response, see figure 2. Starting from low temperatures, the uniform spin susceptibility of some systems rises to quite large absolute values. Following activated behaviour, the increase in the response is largest for the materials with the smallest fundamental gaps. Beyond that, one can surmise that electronic correlations—greater in 3d than 4d transition metal-based compounds—play a role in the prefactor of the exponential rise: indeed while FeAs2 and RuSi, as well as FeGa3 and RuGa3, have pairwise comparable gaps (see table 1), only the 3d compounds show an activated susceptibility (within the available experimental temperature window). In the theory (section 2, see in particular section 2.6), we suggest that this distinction is mainly driven by the difference in the Hund’s rule coupling.

Still more intriguing is that the magnetic susceptibilities of FeSi, FeSb2, and CrSb2 reach a maximum—at a temperature that we shall call T*max—beyond which they realize a Curie–Weiss-like decay. Together with the insulator-to-metal crossover, this signature in the spin-response is the most salient and unexpected property of correlated semiconductors.

While all materials considered here are paramagnetic, many of them are situated in direct proximity to a spin-ordered phase. Magnetism can be induced by isoelectronic substitutions that expand the lattice, off-stoichiometry, as well as doping. Examples include:

- FeSi1-xGe x: ferromagnetic metal above x = 0.25 [96, 97].
- Fe1-xRuSi: ferromagnetic metal for x = 0.1 below 15 K [98].
- Fe1-xCoSi: ferromagnetic metal for 0.05 ≤ x ≤ 0.8 [98–102].
- Fe1-xCoSb2: weak metallic ferromagnetism for 0.2 ≤ x ≤ 0.4 [25, 103].
- Fe1-xCrSb2: antiferromagnetic insulator for 0.2 ≤ x ≤ 1 [66].
- Fe2VAI orders ferromagnetically via a multitude of small changes with respect to stoichometry: Fe2+V1-xAl [49, 104–106], Fe2+V1-xCr1-Al [107], Fe2+VAI1-x [108], Fe2+V1+3Al [109].

Figure 2. Overview: magnetic susceptibilities. Shown is a selection of susceptibilities for some materials from table 1. FeSi: Jaccarino et al [12], Takagi et al [13], RuSi: Buschinger et al [19] and Hohl et al [21] stated the intrinsic χ to be basically temperature-independent up to 320 K, FeSb2: Koyama et al [92], RuSb2, FeAs2: Sun et al [24], CrSb2: Sales et al [33], FeGa3, RuGa3: Gamza et al [36] (field H || ab), CoSb3: Morelli et al [93], LaCoO3: Yamaguchi et al [56]. For Fe2VAI the 27Al NMR relaxation rate of Lue and Ross [51] is shown. Dotted lines for FeAs2 and FeGa3 are activation-law fit extrapolations. CrSb2 is antiferromagnetic below 273 K; a complex magnetic order has also been suggested for FeGa3 [36, 94].

3 The onset of a peculiar antiferromagnetic order in CrSb2 below 273 K [65, 95] has no visible signature in the susceptibility.
4 With the exception of CrSb2 below its Néel temperature TN = 273 K, and possibly FeGa1 [36, 94].

---

2See section 3.3.6 for a modelling of such impurity contributions to the conductivity in FeSb2.
Details can be found in section 1.2 (FeSi), section 2.2.2 (FeSb2), and section 3.5 (Fe2VAl).

### 1.1.2. Thermoelectricity

The observation that warrants the title of this review is that correlated narrow-gap semiconductors exhibit large thermopowers. For stoichiometric compositions, the largest response is usually achieved at rather low temperatures when these systems are in their insulating regime, see figure 3. Indeed, the typical temperature dependence of the thermopower is a dome-like curve. The thermopower of a generic, coherent insulator behaves roughly as ∝ 1/T (see section 3.1.3). In our systems of interest, this decay with rising temperature is accelerated by the insulator-to-metal crossover (see figure 38 in section 3.2 for the case of FeSi). On the low-temperature side, the thermopowers vanish towards absolute zero, as expected from the third law of thermodynamics (see figure 35 in section 3.1.4 for the influence of finite lifetimes on this decay).

Some of the considered materials have competitive powerfactors $S^2\sigma$ in a temperature-range that is useful for Peltier-refrigeration, or even waste-heat recovery (see sections 3.4 and 3.5 for skutterudite and Heusler compounds, respectively). However, most of the correlated narrow-gap semiconductors discussed here (notably those which exhibit the most pronounced correlation effects), have large powerfactors only at very low temperatures (see section 3.2 for FeSi and section 3.3 for FeSb2). An application of these stoichiometric compounds in thermoelectric devices is thus only conceivable for sensor or cooling technologies at cryogenic temperatures [111]. However, owing to large lattice thermal conductivities, no viable alternative for liquid-He-cooling is on the horizon. The impact of correlation effects onto thermoelectric properties, as well as individual materials will be discussed in section 3.

### 1.1.3. Other applications

Besides thermoelectricity, there are a number of proposals for the use and relevance of the unconventional physics of correlated narrow-gap semiconductors, both, in fundamental science as well as for technological applications. Examples include:

- **Spin transport electronics:** the very large anomalous Hall effect in Fe1−xCoSi [112] is a harbinger for spintronic applications.
- **Astrophysics:** from infrared absorption measurements, FeSi has been hinted to occur in circumstellar dust shells [113], as previously suggested by calculations [114]. Ferrarotti et al [113] proposed to exploit the strongly temperature-dependent properties of FeSi as a ‘thermometer’ to probe environmental temperatures near distant stars.
- **Earth and planetary science:** FeSi (B20 or B2 structure) is possibly a relevant composition in the core of the Earth and other terrestrial (rocky) planets [115–117]. Its presence near the core–mantle boundary might explain the anomalously high electrical conductivity of this region [118].
- **CrSb2 and FeSb2 have been considered as high-capacity anode materials in lithium-based batteries [119, 120].**
- **Heusler materials have applications in, e.g. spintronics, solar cells, or magneto-calorics [121, 122].**

---

**Figure 3.** Overview: thermopowers. Shown is a selection of thermopowers for materials from table 1. FeSi: Sales et al [110], RuSi: Hohlf et al [21], FeSb2, FeSb, FeAs: Sun et al [24], CrSb2: Sales et al [33], FeGa3, RuGa3: Wagner-Reetz et al [41], CoSb3: Dyck et al [90], Fe2VAl: Knapp et al [52], LaCoO3: Seneris-Rodriguez and Goodenough [58]. For more experimental data see figures 4(c) and 36 for FeSi, figure 42 for FeSb2, figure 46 for CoSb3, figure 47 for Fe2VAl.

---

**Signatures of correlation effects in narrow-gap semiconductors**

- small upper bound for activation laws in the resistivity $T_p^* \ll \Delta/k_B$, and the magnetic susceptibility $T_x^{\text{max}} \ll \Delta/k_B$.
- insulator-to-metal crossovers for $k_T T \ll \Delta$.
- enhanced paramagnetism at low $T$, Curie–Weiss-like behaviour for $T > T_x^{\text{max}}$ (if experimentally accessible).
- transfer of optical spectral weight over energies much larger than the gap $\Delta$.
- proximity to (spin) ordered phases.
- large effective masses under doping.
- large thermopowers at low temperatures.
1.2. Correlated narrow-gap semiconductors versus Kondo insulators

Early on it has been noted [84, 125–134] that the anomalies observed in the most prominent of d-electron narrow-gap semiconductors are strikingly reminiscent of the physics of heavy-fermion semiconductors. The latter, also known as Kondo insulators [125], are systems with partially filled f-shells that show insulating behaviour at low temperatures in the absence of any magnetic long-range order. In the standard (Doniach) picture [135], the local moment of the f-electrons fluctuates freely at high-temperatures, giving rise to a Curie–Weiss susceptibility. Concomitantly, light (s, p) conduction states are decoupled from the localized (f) states and their density is finite at the Fermi level, causing metallic behaviour. Upon lowering the temperature the local moments get screened by the conduction electrons via the Kondo effect, and the resulting hybridization between f- and conduction states leads to the liberation of mobile charges with large effective masses. In Kondo insulators, however, these heavy charges lead to a completely filled band, i.e. the chemical potential happens to fall into the hybridization gap, quenching spin and charge excitations at low temperatures.

In this section, we will detail empirical similarities and differences between such f-electron heavy-fermion Kondo insulators and the d-electron based correlated narrow-gap insulators that are the main subject of this review. We will focus the juxtaposition mainly on a direct comparison of two insulators that are the main subject of this review. We will present a survey of results for the minimal many-body setups that contain the salient features of correlated semiconductors and Kondo insulators—the two-covalent-band Hubbard model and the periodic Anderson model, respectively. This comparison in the context of reductionist models is extended in section 2.6. Finally, we turn in section 2.4 to realistic many-body electronic structure calculations: results for FeSi are reviewed in sections 2.4.1 and 2.4.2, while section 2.4.3 presents new findings for Ce3Bi4Pt3. The comparison on these two systems culminates in section 2.5.

1.2.1. Charge degrees of freedom. We begin our one-to-one comparison of FeSi and Ce3Bi4Pt3 with properties related to the charge degrees of freedom: the resistivity, the optical conductivity and photoemission spectra.

Insulator-to-metal crossover. As shown in figure 4(a) the resistivity of FeSi and Ce3Bi4Pt3 both follow activation laws at intermediate temperatures—60–120 K for FeSi and 30–150 K for Ce3Bi4Pt3—with charge gaps of about 59 meV and 8 meV, respectively. At lower temperatures, the resistivity of both compounds increases more slowly than expected from thermal activation. This tendency towards resistivity saturation is commonly interpreted as caused by impurities or other defects, but it can also be a sign for the presence of other residual bulk scattering mechanisms (see the discussion in section 3.1.4). More important in our context are the deviations at high temperatures, where the resistivity surpasses the activation law. In the case of FeSi, the slope in temperature, δρ/δT, even becomes positive, clearly indicative of metallic behaviour. This has led some authors to propose that the value of the charge gap in these systems is temperature-dependent. Indeed, it is common practice to collect deviations from activated behaviour into a non-constant Δ(T). Doing this, e.g. Hundley et al extracted for Ce3Bi4Pt3 a gap from the resistivity and the thermopower that halves in size when going from 50 K up to 200 K [140], in congruence with slave-boson calculations for the periodic Anderson model [141]. Without complementary observables, it is, however, difficult to ascertain whether this scenario is realized in Ce3Bi4Pt3 or FeSi.

The metallization process is more pronounced in spectral and optical properties: figure 5(a) displays the change in the FeSi photoemission intensity of Klein et al [75] relative to a high-temperature reference spectrum (see the figure caption for details, and [73, 74, 76, 77] for other photoemission works). Figure 5(b) shows similarly processed data for Ce3Bi4Pt3 from Takeda et al [139]. At low temperature, both materials exhibit a pronounced depletion at the Fermi level. In the case of FeSi (left), this gap is flanked by notably sharp features in the valence band. These are particularly evident in angle-resolved spectra [73, 77]. For Ce3Bi4Pt3, changes are much less pronounced. Indeed, as seen in figure 24, also the valence part of the theoretical many-body spectrum of this f-compound is rather broad at all temperatures.

The charge gap at low temperatures is equally visible in the optical conductivity of FeSi, shown in figure 5(b), left. In the case of FeSi, the direct (optical) gap probed by dipolar transitions is comparable to the indirect gap extracted from the resistivity. In Ce3Bi4Pt3, figure 5(b), right, the optical gap is notably bigger (Δind ≈ 37 meV = 429 K ×kB [60]) than the indirect gap (Δdir ≈ 7.6 meV), as expected in a hybridization gap scenario in the spirit of the periodic Anderson model (see section 2.6).

With growing temperature the gap in the photoemission and the optical spectrum gets increasingly filled. Indeed, while the gap edges soften in FeSi, there is no discernible trend in the photoemission spectrum towards a gap closure by virtue of the valence peak moving towards the Fermi edge. In the optical conductivity, the absorption feature around 100 meV does slightly move towards lower energies [16, 69], but it is clearly not responsible for the rising conductivity in the dc limit. In the case of Ce3Bi4Pt3, the first shoulder in the optical spectrum even moves to slightly higher energies upon increasing temperature. In all, this suggests that in both materials,

---

6 Spectra obtained more recently for heavy-fermion insulators with more f-electrons feature much narrower valence-state peaks, see, e.g. the results for SmB6 [142–146].

7 See also figure 2.5 in [147], and results for electronic Raman scattering in [72].
the gap does not close by the displacement of quasi-particle states, but instead it is filled by incoherent spectral weight as temperature rises. This picture is further supported by differential-conductance measurements from point-contact spectroscopy for Ce₃Bi₄Pt₃ [148] and tunnelling spectroscopy for FeSi [78], that also see practically temperature-independent peak positions.

At \( T^\ast = 160 \text{ K} \) FeSi reaches \( \sigma(\omega = 0) \approx 2 \cdot 10^3 / (\Omega \text{cm}) \), the conductivity of a bad metal: FeSi metallizes at a temperature that is smaller than the low-temperature gap \( \Delta/k_B = 685 \text{ K} \) by a factor of more than four. Ce₃Bi₄Pt₃ surpasses the same value of the conductivity already around 60 K, which is slightly smaller than the coherence temperature \( T^\ast_{\rho} = 100 \text{ K} \) that we defined above. These two temperatures are comparable to the indirect gap \( \Delta_{\text{ind}}/k_B \approx 88 \text{ K} \), but they are still much smaller than the direct gap \( \Delta_{\text{dir}}/k_B \approx 429 \text{ K} \). Consequently, the crossover to a metallic state is, in both systems, qualitatively far beyond thermal excitations across a fundamental gap.

Figure 4. Transport and magnetic properties: the correlated narrow-gap semiconductor FeSi (left) versus the Kondo insulator Ce₃Bi₄Pt₃ (right). (a) Resistivity. FeSi (left): Bocelli et al [10]; Ce₃Bi₄Pt₃ (right): Katoh et al [64]. Dashed (blue) lines are activation law fits with gap \( \Delta ; T^\text{max}_\rho \) indicates the peak temperature of the susceptibility (see figure 4(b)); at \( T^\ast_{\rho} = 160 \text{ K}/120 \text{ K} \) (FeSi/Ce₃Bi₄Pt₃) the relative deviation of the resistivity to the shown activation law exceeds 3%. (b) Magnetic susceptibility. FeSi (left): data combined from Jaccarino et al [12] and Takagi et al [13]; Ce₃Bi₄Pt₃ (right): Hundley et al [61]; dashed (blue) lines are activation law fits \(~\exp(-\Delta_s/k_BT)\)/\(T\) in the local moment picture; dashed (red) lines indicate Curie–Weiss-like behaviour, \( \chi = N\alpha\mu_B^2 / (3k_B(T - \theta)) \). In between, \( \chi \) peaks at \( T^\text{max}_\chi = 500 \text{ K} \) (75 K) for FeSi (Ce₃Bi₄Pt₃). (c) Thermopower. FeSi (left): Wolfe et al [14], Sun et al [123], Sales et al [17, 110]; Ce₃Bi₄Pt₃ (right): Hundley et al [124] and Katoh et al [64].
Topical Review

Spectral weight transfers. One of the powers of optical spectroscopy is the existence of sum-rules that allow for a quantitative analysis of spectral weight transfers. Indeed, the integral over the optical conductivity,

$$F(\Omega, T) = \int_{\Omega_0}^{\Omega} d\omega \Re \sigma(\omega, T)$$  \hspace{1cm} (1)

yields, for $\Omega \to \infty$, a value, $F(\infty, T) = \frac{\hbar^2 n e^2}{2m}$, that is solely determined by the total density $n$ of carriers participating in optical absorption and the bare electron mass $m$ [149]. Therewith, the integral of spectral weight is independent of external parameters, such as temperature, or pressure. In other words, the spectral weight that is lost in the low-energy optical conductivity, as FeSi and Ce$_3$Bi$_4$Pt$_3$ become insulating upon cooling, must be shifted to higher energies. The frequency $\Omega = \Omega_c$ above which the integral $F(\Omega_c, T)$ becomes independent of the external parameter sets the energy scale over which the reshuffling of spectral weight occurs. This scale can give clues about the mechanisms that underlie the evolution of the system for changing conditions: In a conventional semiconductor, thermal activation leads to a finite electron (hole) population in the conduction (valence) band, that is however restricted to the direct vicinity of the band edges. Changes in inter-band transitions in the optical conductivity are thus limited to a few $k_B T$ above the charge gap $\Delta$, implying $\Omega_c \geq \Delta$. The same applies, e.g. for gap-opening phase transitions owing to density-waves (examples are Cr [150] or NdNiO$_3$ [151]).

In materials with strong electronic correlations, the energy scale $\Omega_c$ required to reach a T-independent $F(\Omega_c, T)$ can be much larger. In Mott–Hubbard systems, for example, spectral weight transfers occur over a frequency range set by the Hubbard $U$ interaction [79, 152]. The latter can be of the order of, or larger than, the bandwidth. In case it triggers a metal-insulator transition in a multi-orbital system, the fundamental gap $\Delta$ can be much smaller than $U$. Examples with $\Omega_c \gg \Delta$ are, e.g. the transition metal oxides VO$_2$ and V$_2$O$_3$ [153–155].

In the case of FeSi, the energy needed to recover the transferred spectral weight with changing temperature was controversial [9, 69, 70, 156], until accurate ellipsometry measurements [16] ruled in favour of a $\Omega_c$ much larger than the charge gap. The situation is similar for the Kondo insulator Ce$_3$Bi$_4$Pt$_3$; there, the reshuffling of spectral weight occurs over an energy range of $\sim$250 meV [126] corresponding to more than six times the size of the optical gap [60].

**Figure 5.** Spectroscopic observables: the correlated narrow-gap semiconductor FeSi versus the Kondo insulator Ce$_3$Bi$_4$Pt$_3$. (a) Photoemission spectra. FeSi (left): Klein et al [75]; Ce$_3$Bi$_4$Pt$_3$ (right): Takeda et al [139]. In both cases the photoemission intensity was (i) divided by the Fermi function, and, (ii) a fit of the resulting spectra at the highest temperature (120 K for FeSi, 300 K for Ce$_3$Bi$_4$Pt$_3$) was subtracted to emphasize relative changes. (b) Optical conductivity. FeSi (left): ellipsometry data from Menzel et al [16] and normal-incidence infrared spectroscopy from Damascelli et al [69]; Ce$_3$Bi$_4$Pt$_3$ (right): data from Bucher et al [60]. FeSi (Ce$_3$Bi$_4$Pt$_3$) surpasses $2 \cdot 10^3/(\Omega \text{cm})$ at around 160 K $\approx T^*_\rho (60 K \lesssim T^*_\rho = 120 K)$.

---

**Spectral weight transfers.** One of the powers of optical spectroscopy is the existence of sum-rules that allow for a quantitative analysis of spectral weight transfers. Indeed, the integral over the optical conductivity,

$$F(\Omega, T) = \int_{\Omega_0}^{\Omega} d\omega \Re \sigma(\omega, T)$$  \hspace{1cm} (1)

yields, for $\Omega \to \infty$, a value, $F(\infty, T) = \frac{\hbar^2 n e^2}{2m}$, that is solely determined by the total density $n$ of carriers participating in optical absorption and the bare electron mass $m$ [149].

Therewith, the integral of spectral weight is independent of external parameters, such as temperature, or pressure. In other words, the spectral weight that is lost in the low-energy optical conductivity, as FeSi and Ce$_3$Bi$_4$Pt$_3$ become insulating upon cooling, must be shifted to higher energies. The frequency $\Omega = \Omega_c$ above which the integral $F(\Omega_c, T)$ becomes independent of the external parameter sets the energy scale over which the reshuffling of spectral weight occurs. This scale can give clues about the mechanisms that underlie the evolution of the system for changing conditions: In a conventional semiconductor, thermal activation leads to a finite electron (hole) population in the conduction (valence) band, that is however restricted to the direct vicinity of the band edges. Changes in inter-band transitions in the optical conductivity are thus limited to a few $k_B T$ above the charge gap $\Delta$, implying $\Omega_c \geq \Delta$. The same applies, e.g. for gap-opening phase transitions owing to density-waves (examples are Cr [150] or NdNiO$_3$ [151]).

In materials with strong electronic correlations, the energy scale $\Omega_c$ required to reach a T-independent $F(\Omega_c, T)$ can be much larger. In Mott–Hubbard systems, for example, spectral weight transfers occur over a frequency range set by the Hubbard $U$ interaction [79, 152]. The latter can be of the order of, or larger than, the bandwidth. In case it triggers a metal-insulator transition in a multi-orbital system, the fundamental gap $\Delta$ can be much smaller than $U$. Examples with $\Omega_c \gg \Delta$ are, e.g. the transition metal oxides VO$_2$ and V$_2$O$_3$ [153–155].

In the case of FeSi, the energy needed to recover the transferred spectral weight with changing temperature was controversial [9, 69, 70, 156], until accurate ellipsometry measurements [16] ruled in favour of a $\Omega_c$ much larger than the charge gap. The situation is similar for the Kondo insulator Ce$_3$Bi$_4$Pt$_3$; there, the reshuffling of spectral weight occurs over an energy range of $\sim$250 meV [126] corresponding to more than six times the size of the optical gap [60].

---

That do not require Kramers–Kronig transforms to be applied to reflectivity data obtained for a restricted frequency range.
1.2.2. Spin degrees of freedom.

Magnetic susceptibility and magnetic neutron scattering. Figure 4(b) displays the uniform magnetic susceptibilities for FeSi [12, 13] (left) and Ce$_3$Bi$_4$Pt$_3$ [61] (right). At low temperatures spin excitations in both compounds are gapped and the responses have an activation-type behaviour\(^9\). Without any further microscopic insight, it is however not obvious to deduce the origin of enhanced or temperature-induced magnetism from the experimental data. In figure 4(b) we have deliberately fitted both low temperature susceptibilities with the same expression, \(\chi(T) \propto \exp(-\Delta_s/2k_BT)/T\), which corresponds to the behaviour of local moments. The applicability of this form and further details will be discussed in section 2.1. Here, we only note that in this picture the extracted spin gap—\(\Delta_s = 50\) meV for FeSi and \(\Delta_s = 7\) meV for Ce$_3$Bi$_4$Pt$_3$—is of similar magnitude than the indirect charge gap \(\Delta_{indir}\) obtained from fitting the resistivity, yet much smaller than the direct, optical gap \(\Delta_{dir}\)\(^10\). Using instead a simplified band picture to extract the spin-gap from the susceptibility yields, via \(\chi(T) \propto \exp(-\Delta_s/(2k_BT))\) a larger value \(\Delta_s = 96\) meV for FeSi, advocating, in all, the hierarchy \(\Delta_{indir} < \Delta_s < \Delta_{dir}\). This finding suggests that a non-trivial electronic structure is realized in these systems. Indeed in a band insulator, activation of spin and charge excitation are governed by the same energy scale \(\Delta_s\). In a Mott insulator, on the other hand, a finite charge gap is accompanied by gap-less spin fluctuations (\(\Delta_s = 0\)). More insight will be discussed in the context of many-body models, see section 2.3.

At larger temperatures, after going through a maximum at \(T^\text{max}\), see figure 4(b), the activated behaviour in \(\chi(T)\) gives way to a crossover to Curie–Weiss-like decay

\[
\chi(T) = \frac{N_A\mu_{\text{eff}}^2}{3k_B(T-\theta)}.
\]

This crossover in the magnetic response is seen for Ce$_3$Bi$_4$Pt$_3$ also in the Knight shift\(^15\). For FeSi we find \(\mu_{\text{eff}} = 2.6\ \mu_B\), which is close to the behaviour expected for local moments with \(S = 1\): \(\mu_{\text{eff}}/\mu_B = g_S \sqrt{S(S+1)} = 2\sqrt{2} \approx 2.8\) (\(g_S = 2\)). In the case of Ce$_3$Bi$_4$Pt$_3$, we extract \(\mu_{\text{eff}} = 2.53\ \mu_B\). This value is very close to the fluctuating moment of isolated Ce$^{3+}$ ions: \(\mu_{\text{eff}}/\mu_B = g_J \sqrt{J(J+1)} = 2.54\) (\(J = 5/2\), \(g_J = 0.857\)). The Curie–Weiss temperatures are negative in both cases, \(\theta = -289\ K\) (\(\theta = -103\ K\)) for FeSi (Ce$_3$Bi$_4$Pt$_3$), suggestive of antiferromagnetic interactions. In the Kondo lattice picture, this is in line with the Ruderman–Kittel–Kasuya–Yosida (RKKY) coupling that competes with the Kondo screening [135]. Ce$_3$Bi$_4$Pt$_3$ (and other Kondo insulators) are believed to be in the strong J-coupling regime, where the RKKY coupling fails to impose long range magnetic order because of a dominating Kondo effect\(^11\). In the case of FeSi, the negative Curie–Weiss temperature is contrasted with results from magnetic neutron spectroscopy results of Shirane et al [160] and Tajima et al [161] (see figure 20 and the discussion in section 2.4.1). These measurements revealed a strongly peaked magnetic cross section for \(q = 0\) at \(T^\text{max} = 500\ K\) [160] and above [161], indicative of ferromagnetic fluctuations\(^12\).

Magnetoresistance and high-field experiments. In a Kondo insulator, the formation of Kondo singlets below the coherence temperature leads to the gapping of spin- and charge excitations. These singlets, formed by a strongly renormalized hybridization between conduction electrons and f-levels, can be broken up by a large enough magnetic field \(H\). To do so, the Zeeman energy needs to be comparable to the spin gap, i.e. \(g_f\mu_B H \approx \Delta_s\), where \(g_f\) is the gyromagnetic ratio, and \(J\) the electron’s total angular momentum. In the simplest picture, such a field \(H\) will induce a phase transition from the Kondo phase to a light (conduction) metal with fully polarized f-electrons\(^13\).

In Ce$_3$Bi$_4$Pt$_3$ a partial polarization of the f-states is signalled by a large negative magnetoresistance (MR), \((\rho(H) - \rho(0))/\rho(0)\), measured below 50 K in fields up to 10 T [166]. With a spin gap of about 12 meV [63], the critical field \(H_c\) needed to metallize Ce$_3$Bi$_4$Pt$_3$ (Ce$^{3+}$: \(J = 5/2\)) is however much larger. The estimate from the above formula is 40 T. Reaching such fields, the insulator-to-metal transition has indeed been observed in measurements of the resistivity [167] (\(H_c \sim 50\ T\)), as well as the specific heat [168] (\(H_c \sim 10\ T\)). These findings support the Kondo picture for Ce$_3$Bi$_4$Pt$_3$.

In electron derived narrow-gap semiconductors, the much larger size of the spin gap prohibits a clear picture in magnetic fields of conventional size. In fields up to 9T experiments find a large positive MR for FeSb$_2$ [23, 26, 128]. For FeSi, even the sign of the MR strongly depends on the experiment. The samples of highest quality —according to the residual-resistance ratio criterion—are those of Paschen et al [11] that yield a negative magnetoresistance below 50 K in a field of 7 T\(^14\). The notable sample dependence, as well as a positive contribution to the magnetoresistance could be linked to the presence of impurities. Alternatively, a positive MR could be due to the proximity to a ferromagnetic instability\(^15\).

\(^{11}\) Indeed, a ferromagnetic state can also be reached by doping, Fe$_{1-x}$Co$_x$Si [99–102], expansion of the lattice by chemical pressure, FeSi$_{1-x}$Ge$_x$ [96, 97], as well as ultra-high magnetic fields [162, 163] (see the discussion below).

\(^{12}\) Note however, that numerical calculations for the Kondo lattice model [164] and the periodic Anderson model [165] suggest that with an applied magnetic field the spin gap collapses and a transverse AF order develops, while the charge gap remains finite up to higher fields.

\(^{13}\) The MR reported by Ohta et al [160] for FeSi is positive up to 16 T for all temperatures, but samples with lower RRR show a less positive MR; Lisunov et al [176] found a positive signal up to 35 T; more recently, Sun et al [123] reported for 8 T a sign change from MR > 0 at low temperatures to MR < 0 above 70 K that interestingly correlates with a maximum (minimum) in the Nernst (Hall) coefficient, while the absolute value of the magnetoresistance is much smaller than in older experiments.

\(^{14}\) While the size of the charge gap is insensitive to, e.g. dopings with up to 3% cobalt [70], dopings (\(T = \text{Co, Rh}\)) and also isoelectronic substitutions (\(T = \text{Ru}\)) of 10%, Fe$_{x} \text{Fe}_0.1 \text{Si}$, induce itinerant ferromagnetism [98], concomitant with a large positive MR [98, 101, 171].
Using ultra-high magnetic fields of up to spectacular 450T as produced by an explosively pumped flux compression generator, Kudasov et al. [162, 163] found the resistivity of FeSi to continuously decrease by several orders of magnitude. While at \( T = 77 \, K \) the semi-conductor to metal crossover was continuous also in the magnetic signal, a sample cooled to \( T = 4.2 \, K \) displayed a jump to a magnetic moment of 0.95 \( \mu_B \) at a field of 355 T [163], suggestive of a ferromagnetic metallic state. Moreover, the field-dependent resistivity was found to be strongly non-exponential, indicative of a Zeeman splitting beyond the rigid-band picture. Indeed, a conventional collapse of the experimental spin-gap of \( \Delta_s \approx 50 \, \text{meV} \) (see table 1), is naively expected at 435 T (using \( S = 1 \) and \( g = 2 \)), i.e. for fields larger than found in experiment.

1.2.3. Interplay of charge and spin degrees of freedom.

Correlation in the temperature dependence of charge and spin observables. Despite different values for the charge and spin gap, electrical and magnetic properties are intimately linked in the Kondo picture: when Kondo coherence is destroyed as temperature rises, the \( f \)-charge decouples from the conduction electrons and its moment begins to fluctuate freely. Comiconitantly, conduction electron density transfers to lower energies and fills the hybridization gap. That both effects occur on the same temperature scale in Ce\(_3\)Bi\(_4\)Pt\(_3\) has been beautifully illustrated by Bucher et al. [60]. In figure 6 we reproduce their analysis and extend it to FeSi.

Fitting the local moment expression, \( \chi(T) = \frac{N}{4m} \mu_{\text{eff}}(T) \exp(-\frac{\Delta_{\text{s}}}{k_B T}) \), to the experimental susceptibility, a temperature-dependent effective moment \( \mu_{\text{eff}}(T) \) is extracted. At high temperature we recover, as mentioned before, for both compounds, values that are close to the respective single ion limit (see section 1.2.2). For Ce\(_3\)Bi\(_4\)Pt\(_3\):

\[
\mu_{\text{eff}}/\mu_B = \sqrt{J(J+1)}g_J = 5.92 \quad \text{as expected for Ce}^{3+} (J = 5/2, g_J = 0.857) ; \text{for FeSi } \mu_{\text{eff}}/\mu_B = \sqrt{S(S+1)}g_s = 2.8 \text{ supposing } S = 1 \text{ (} g_s = 2 \text{). Upon lowering temperature, however, the magnitude of these effective moments decreases. In Ce}_3\text{Bi}_4\text{Pt}_3 this observation is associated with the Kondo effect: conduction electrons begin to screen the } f \text{-moment. The concomitant gapping of the density of states is heralded by the loss of optical weight at low energy. The amount of charge participating in this transfer can be obtained from the spectral weight integral of (1). By tracking the integral for a cut-off frequency } \Omega \text{ that roughly delimits the regime in which weight is lost, with respect to a base temperature } T_0 \text{ we can compute the amount of the charge transferred away from the Fermi level to energies above } \Omega :}
\]

\[
\Delta n(T) = \Delta n(\Omega, T, T_0) = \frac{2m}{h\pi e} \left[ F(\Omega, T) - F(\Omega, T_0) \right].
\]  

Using \( T_0 = 300 \, K \) and \( \Omega = 37 \, \text{meV} \) for Ce\(_3\)Bi\(_4\)Pt\(_3\) [60], and \( \Omega = 100 \, \text{meV} \) for FeSi (see figure 5), we see that the amount of charge \( \Delta n(T) \) that participates in conduction upon heating indeed correlates with the emergence of the effective local moment \( \mu_{\text{eff}}(T) \). Unfortunately, there are no results for the optical conductivity at high enough temperatures to reach the Curie–Weiss-like regime for FeSi. It would be interesting to see whether the ferromagnetic-like fluctuations evidenced in magnetic neutron scattering above 300K [160, 161] induce, for FeSi, a deviation from the current analysis which is based on the local moment picture.

In fact, from our (admittedly somewhat arbitrary) definition of characteristic temperatures for the charge and spin response, \( T_{\mu}^* \) and \( T_{\chi}^{\text{max}} \), we note a quantitative difference between FeSi and Ce\(_3\)Bi\(_4\)Pt\(_3\). In the latter, the temperature below which Kondo screening sets in, \( T_{\mu}^{\text{trans}} \), is slightly smaller than the crossover temperature \( T_{\mu}^{\text{max}} \) of the conductivity (see...
is the value of the gap the same in both moves, currently unavailable.

...controlled Mott metal–insulator transition in V2O3 [67, 180].

If, however, the band-gap originates from hybridization, i.e. at least one side of the gap consists of bonding/anti-bonding states, the opposite tendency can be found. In fact, since also the transfer integrals at the origin of the bonding/anti-bonding splitting are enhanced by pressure, the gap typically increases. This behaviour is indeed found for, e.g. FeSi [96, 173, 175, 181, 182], FeSb2 [176, 177] and Ce3Bi4Pt3 [178, 179] as extracted from the activated behaviour of the resistivity and the magnetic susceptibility. As depicted in figure 7, the increase of the gap scales roughly linearly with the applied pressure and is notably larger (on a relative scale) for the Kondo insulator Ce3Bi4Pt3[16,17].

It would be interesting to extend the above analysis of charge and spin gap—$\Delta_{\text{dirac}}, \Delta_{c}$—to the coherence scale $T_{\chi}^{max}$, as the dependence of $T_{\chi}^{max}$ on the value of the gap the same in both compounds? Or is the crossover to Curie–Weiss-like decay in the $d$-electron compounds different, as it is proposed to be driven by Hund’s rather than Hubbard physics (see section 2.6)? Answering these questions requires, however, measurements of FeSi under pressure up to temperatures above $T_{\chi}^{max}$, currently unavailable.

Isoelectronic substitutions: Another perturbation of the gap is achieved by isoelectronic substitutions. One route is the isoelectronic replacement of ligands[18]. Here, the most prominent example is the alloy FeSi1−xGe. While the end-member compounds of this series are isoelectronic as well as isostructural, their electronic and magnetic properties are very different. Contrary to FeSi, which is insulating at low temperature and does not show any signs of magnetic order, FeGe is metallic and shows spin helical long-range order [188, 189].

Isotropic substitutions: Another route of isoelectronic substitution is the replacement of the $3d$-transition metal(lanthanoid) in FeSi(Ce3Bi4Pt3) by their homologues one row down in the periodic table. The interesting case of Fe1−xRuSi will be discussed in section 3.2.4. Interchanging 4f orbitals with 5f ones is unfortunately not possible in Ce3Bi4Pt3, as the actinoid thorium prefers the $5f^66d^2?5s^2$ rather than the $5f^66d^2?5s^2$ configuration.

---

16 In view of other Kondo insulating systems, a word of caution is in order. While for Ce-based compounds, the above rationale holds, the gap in hole-like Kondo insulators is actually suppressed under pressure, as shown for SmB6 [183] or YbB2 [184]. Interestingly, a critical pressure of 6 GPa suppresses the gap in SmB6 and magnetic long-range order with a saturated moment of 0.8μB develops below 12K [185]. It was suggested that the opposing trends i.e. Ce3Bi4Pt3 and SmB6 are rooted in the different signs for the change in unit-cell volume upon adding/removing an electron to the system (ion-size difference of the $f^2, f^{n+1}$ configurations) [186].

17 For a study of doped FeSi under pressure, see [102, 187].
Doping: Finally, charge and spin excitations can be altered by doping. We first discuss non-isoelectronic substitutions in FeSi. Substituting up to 3% of Fe with Co causes only a minute reduction of, both, the direct and indirect charge gap [70]. Moreover no anomalies in thermodynamic and transport quantities have been observed, pointing towards the persistence of paramagnetism. The specific heat, however, was found to increase roughly linearly with doping, reaching 7.6 mJ mol$^{-1}$ K$^{-2}$ at $x = 3\%$. Assuming as origin an increased density of states at the Fermi level, Chernikov et al. [70] extracted a surprisingly large effective mass ratio $m^*/m_e \approx 30$.

Co-concentrations beyond 5\% lead to a transition to a ferromagnetic metallic state [99–102]. This is quite notable, since the end compound, CoSi, is a diamagnetic semimetal. Indeed for $x > 0.8$, Fe$_{1-x}$Co$_x$Si becomes again paramagnetic. Lacerda et al. [130] showed that the magnetic susceptibility of electron doped Fe$_{1-x}$Co$_x$Si and FeSi$_{1-y}$P$_y$ are qualitatively akin, prompting the authors to conclude that magnetism in the Co-substituted samples is not due to the moment of Co. By comparing Co-doping with Fe$_{1-x}$Rh$_x$Si and Fe$_{1-x}$Ru$_x$Si—which become ferromagnetic above a critical doping—Paschen et al. [98] substantiated that magnetism in Fe$_{1-x}$Co$_x$Si does not arise from local moments on Co. Indeed the 4$d$-orbitals orbitals of Rh have too large a radial extent to allow for local magnetism. That also isoelectronically substituted Fe$_{1-x}$Ru$_x$Si orders ferromagnetically points to an influence of the unit-cell volume onto magnetism (as in FeSi$_{1-y}$Ge$_y$, see above); that the ordered moment for insulating Ru-substituted samples is the smallest among the considered alloys indicates that (electron doping-induced) metallicity boosts magnetism in FeSi [98], further strengthening the itinerant picture. The effect of the unit-cell volume onto magnetism is corroborated by pressure experiments that see the disappearance of long-range order in doped samples above a critical pressure [187]; also, epitaxially strained samples have a largely different critical doping level than the doped bulk [218].

It is crucial to note the influence of cobalt-doping onto the spin-gap $\Delta_S$ and the crossover temperature $T_{C\rightarrow M}^{\text{max}}$ at which the magnetic susceptibility peaks before assuming a Curie–Weiss–like decay: the introduction of Co may affect the impurity-derived low-temperature upturn in $\chi$. Otherwise, the magnetic susceptibility of Fe$_{1-x}$Co$_x$Si for concentrations up to 10\% looks (above $T_C$) remarkably similar to that of FeSi, albeit with a constant upwards shift [130]. This means that both $\Delta_S$ (as extracted from an activation-law fit) and, apparently, also $T_{C\rightarrow M}^{\text{max}}$ are basically unaltered by doping.

The added carriers only introduce a $T$-independent Pauli-like contribution, congruent with the above findings from specific heat measurements. As will be discussed below, this is a marked difference to the behaviour of doped Ce$_3$Bi$_2$P$_2$T$_3$.

Alternatively also the ligand can be substituted. Hole doping can be achieved, e.g. in FeSi$_{1-x}$Al$_x$ [82–85]. DiTusa et al. [20] Other transition metal substitutions that support the B20 crystal-structure (but will not be discussed here) include electron doping with Ni [213–215], and hole doping by replacing Fe with Mo [216], or Cr [213, 217].

21 Additional electrons also induce metallic ferromagnetism in Fe$_3$VAl (see figure 48) and in FeSb$_2$ [25, 103].

22 Note, however, a further complication: these thin films exhibit a rhombohedral distortion.

---

19 Yet, CeNiSn is often referred to as a ‘failed Kondo insulator’. In fact, the system is thought to be rather a Kondo semimetal [209, 210] owing to an anisotropic hybridization amplitude [211, 212].
interpreted their results such that FeSi is a renormalized realization of silicon, in the same sense as proposed for Kondo insulators: Doping the system leads to a (possibly disordered) Fermi liquid ground state, albeit with largely enhanced effective masses. Indeed, Al-doping causes the charge gap— as extracted from fitting the resistivity to an activation law above 100 K—to rapidly decrease, while a metallic slope appears at lower temperatures already above about \( x = 1\% \) [82, 84]. From the specific heat coefficient, a considerable effective mass \( m^* / m = 14 \) was extracted [83]. Concomitantly, as in the case of Co-doping, a Pauli-like contribution appears in the magnetic susceptibility. The electrons introduced by Al affect neither \( T^{\text{max}}_\chi \) nor the (impurity-derived) Curie–Weiss upturn at very low temperatures. As a consequence, contrary to the charge gap, the spin gap remains roughly constant up to 10\% Al [84], indicating a non-trivial interplay of charge and spin degrees of freedom.

Hole-doping the prototypical Kondo insulator Ce\(_2\)Bi\(_2\)Pt\(_3\) by substituting La for Ce notably suppresses the electrical resistivity [61] and leads to a finite specific heat at low-temperature [61, 219]. From the latter, much enhanced effective masses of up to \( m^* / m \approx 100 \) at \( x = 30\% \) have been extracted [220]. However, doping also largely modifies the spin degrees of freedom: in fact the peak temperature, \( T^{\text{max}}_\chi \), consistently moves to lower temperatures with increasing doping [61, 220]. Indeed, as expected from, e.g. the Coqblin–Schrieffer model [221], the linear coefficient of the specific heat \( \gamma \) scales with \( 1/T^{\text{max}}_\chi \). This dependency is realized in (Ce\(_{1-x}\),La\(_x\))\(_2\)Bi\(_2\)Pt\(_3\) above 25\% La [219]: Ce ions then behave as impurities in a metallic host. In polycrystalline samples, additionally, the overall magnitude of \( \gamma \) decreases [220], contrary to the Pauli-like increase seen in doped FeSi (see above). Moreover, the spin-gap decreases notably with \( x \) [222]. As discussed by Kwei et al [223], there is a direct correlation between the magnetic response in the local moment regime and the change in lattice constant along the (Ce\(_{1-x}\),La\(_x\))\(_2\)Bi\(_2\)Pt\(_3\) series: indeed the square of the effective moment \( \mu^2 \propto T \cdot \gamma \) increases linearly with the shrinking lattice constant towards the Ce \( (x = 1) \) end-member. This suggests that the change in lattice constant tracks the number of \( f \)-electrons. Interestingly, also the metallic \( f^2 \)-compound Pt\(_2\)Bi\(_2\)Pt\(_3\) has a larger volume than its Ce cousin [126]. Also in other heavy fermion compounds, e.g. rare-earth chalcogenides, (mixed) valency and lattice constants are found to correlate [224]. In Ce\(_2\)Bi\(_2\)Pt\(_3\) the valency that realizes Kondo insulating volume coincides with the lowest equilibrium volume.[23]

1.2.5. Thermoelectricity. Finally, FeSi and Ce\(_2\)Bi\(_2\)Pt\(_3\) also exhibit empirical similarities in their thermoelectric response. As evident in figure 4(c) both compounds develop large thermopowers at low temperatures. Combined with the resistivity data, figure 4(a), these yield quite notable powerfactors, see table 1. This is expected for insulating systems which possess a significant particle-hole asymmetry. In an asymmetric, coherent semiconductor the thermopower roughly behaves as \( 1/T \); see section 3.1.3. Effects of incoherence or other mechanisms towards metallization will quench the thermopower. In the photoemission spectra of FeSi discussed above (see figure 5(a)) notable spectral weight appears at the Fermi level starting at 100 K. In this range, the thermopower is already below 100 \( \mu \text{V/K} \). At \( T^* \) (see section 1.2 and, there, figure 5(b)), the thermopower has already become negligible in FeSi. Using the same measure, the decay in Ce\(_2\)Bi\(_2\)Pt\(_3\) is

23 Also in some \( d \)-electron narrow-gap semiconductors, the smallest lattice constant is found for insulating phases, see, e.g. figure 48 (top): the lattice constant of Fe\(_2\)VAL is minimal for perfect stoichiometry. Deviations cause the volume to expand, while the system metatizes (see the specific heat in panel (c)) and develops long-range magnetic order (see panel (b)). In FeSi, off-stoichiometry Fe\(_{1-x}\)Si\(_x\) does not show any significant change in the lattice parameter for small \( x \) [225], while electron doping via Fe\(_{1-x}\)Co\(_x\)Si increases the lattice [112]. However, in the latter potential effects of the valency onto the unit-cell volume are masked by the contraction induced by the smaller atomic radius of Co.

24 While spin–orbit coupling effects are certainly smaller in \( 3d \) transition-metal-based narrow-gap semiconductors, let us mention that also FeSi, RuSi, FeGe, and other compounds crystallizing in the non-symmorphic B20 structure (see the discussion of their band-structures in section 2.2.1) allow for non-trivial topological effects, such as non-vanishing electronic Berry phases [229]. Also, see the review by Martins et al [230] for the interplay of correlation effects and the spin–orbit coupling in \( 4d \) and \( 5d \) systems.
slower: at its \( T'_m \), the thermopower is still about a third of its maximum value.

While the insulator-to-metal crossover with rising temperature is antagonistic to a large thermoelectric response, it is difficult to assess, on the basis of the experimental data alone, whether the metallization causes the quenching of the thermopower. As will be discussed in section 3.2 (see, in particular, figure 38), theoretical calculations strongly suggest an affirmative answer.

Response of doped samples. Besides the difference between a metallic and an insulating thermoelectric response, the thermopower—being a measure for particle-hole asymmetry—is extremely sensitive to doping. In the case of FeSi, Ce\(_3\)Bi\(_4\)Pt\(_3\), and other narrow-gap semiconductors both effects conspire: Introduced electrons or holes may not only change the character of preponderant charge carriers but they also suppress the insulating state. Indeed, the thermopower of the Kondo insulator Ce\(_3\)Sb\(_2\)Pt\(_3\) is extremely sensitive to doping both via ligand alloying (Ce\(_3\)Cu\(_{1-x}\)Pt\(_3-x\)-Sb\(_4\) [231]), as well as lanthanoid substitution (Nd\(_{x}\)Ce\(_{3-x}\)-Pt\(_3\) [232]). Both kinds of doping quickly suppress the thermopower. The same applies to FeSi: as seen in figure 39 stoichiometric FeSi sticks out to FeSi: as seen in figure 39 stoichiometric FeSi sticks out among transition metal mono-silicides TMSi (TM = Cr, Mn, Fe, Co) and their alloys. Also doping FeSi on the ligand site, e.g. hole-doping via FeSi\(_{1-x}\)-Al\(_x\) [84], suppresses the thermopower. As will be explained in section 3.1.3 electron-doping FeSi is expected to drive the thermopower \( \delta \) through a sign-change and realize a negative response of a magnitude comparable to that of stoichiometric FeSi. Indeed, this is seen when substituting some Co for Fe, figure 39. Realizing electron-doping on the ligand site might have the advantage of less perturbing the low-energy electronic structure in terms of disorder. Since FeSi\(_{1-x}\)-P\(_x\) can be synthesized in the B20 structure [130]\(^{25}\), it would be quite interesting to perform thermoelectric measurements on it.

1.2.6. Summary. We end this section with a list of empirical similarities and differences of FeSi and Ce\(_3\)Bi\(_4\)Pt\(_3\), as well as some proposals for future experiments that could extend the presented comparison.

Empirical similarities of FeSi and Ce\(_3\)Bi\(_4\)Pt\(_3\).

- Absence of long-range order at all \( T \) in stoichiometric samples.
- Charge and spin excitations are gapped at low-\( T \) with activated behaviour in resistivity and magnetic susceptibility.
- Spin gap tends to be smaller than the direct charge gap: \( \Delta_s < \Delta_{\text{dir}} \).
- Crossover to a metallic state for \( T \ll \Delta_{\text{dir}}/k_B \); towards the metallization, the gap does not shrink in size, but it gets filled by incoherent weight.
- Spectral weight transfers in the optical conductivity over scales much larger than the direct charge gap \( \Delta_{\text{dir}} \).

\(^{25}\) At ambient conditions FeP has the regular MnP-type structure (space group Pnma) [235].

Empirical differences between FeSi and Ce\(_3\)Bi\(_4\)Pt\(_3\).

- Proximity to different spin-ordered phases: non-magnetic \( d \)-electron narrow-gap semiconductors have a propensity for nearby ferromagnetic instabilities\(^{26}\). Long range order is likely of itinerant origin, with enhancements from the Hund’s-rule coupling (see [236–239], and the discussion in section 2.6), and can be reached under volume expansion or doping. Some Kondo insulators can instead be driven to an antiferromagnetic (RKKY-mediated) instability by (inverse) physical or chemical pressure, changes in the Kondo coupling by isoelectronic substitutions with different radial quantum number, or doping. Moment-carrying charges tend to be localized.
- Interplay of charge and spin degrees of freedom under doping: while doping largely changes the activated behaviour in the conductivity and induces a conducting state at low-temperatures in both FeSi and Ce\(_3\)Bi\(_4\)Pt\(_3\), the effect on the magnetic response is markedly different. Indeed doping only mildly affects the spin-gap \( \Delta_s \) and crossover temperature \( T^* \) in FeSi, while in Ce\(_3\)Bi\(_4\)Pt\(_3\) both are quite sensitive to the introduction of holes.

Experiments to further the current understanding.

- High temperature optical conductivity measurements of FeSi to access spectral weight transfers; does the potentially different microscopic origin of Curie–Weiss-like decay in the susceptibility (local versus itinerant fluctuations) result in a disparate interplay of charge and spin degrees of freedom above \( T^* \)?
- Dependence of \( T^* \) in FeSi and Ce\(_3\)Bi\(_4\)Pt\(_3\) under pressure (or FeSi\(_{1-x}\)-Ge\(_x\)): does quasi-particle coherence follow the same scaling with \( \Delta \) in both materials?
- Measurements of the magnetic susceptibility of Fe\(_{1-x}\)-CoSi and FeSi\(_{1-x}\)-Al\(_x\) up to \( T^* \) and beyond: we stated that \( T^* \) is seemingly unaffected by doping. However, this was a statement made by an educated extrapolation. Actual measurements of \( \chi \) only extend to 400 K, which is still slightly below \( T^* \).

\(^{26}\) One exception is Fe\(_{1-x}\)-CrSb which is an antiferromagnetic insulator for 0.2 \( \leq x \leq 1 \) [66].
2. Theories of correlated narrow-gap semiconductors

This section deals with theoretical efforts to understand the peculiar behaviour of correlated narrow-gap semiconductors. First (section 2.1) we will discuss early phenomenological scenarios that invoke electronic correlation effects of some sort to explain the experimental observations in the prototypical compounds FeSi and FeSb$_2$. Next (section 2.2) we review the merits and failures of band-theory applied to several classes of relevant materials, and include new calculations for the Kondo insulator Ce$_3$Bi$_4$Pt$_3$. Based on the insights of band-theory, solutions of many-body models have shed light onto the qualitative impact of correlation effects in these systems. These are recapitulated in section 2.3. Building on this knowledge, realistic many-body calculations (section 2.4) were able to quantitatively reproduce, for the prototypical FeSi, the signatures of correlation effects in diverse experimental observables. Importantly, the congruence with experiments allowed for the deducing of microscopic insights (section 2.4.2) beyond what was included in the reductionist models studied before. Indeed, the emerging physical picture refutes some earlier scenarios, while reconciling others. Crucially, the Hund’s rule coupling was identified as a new key ingredient to the physics observed in FeSi. Further, new realistic many-body calculations for Ce$_3$Bi$_4$Pt$_3$ (section 2.4.3) are used to discuss potential microscopic differences between $d$-based narrow-gap semiconductors and Kondo insulators. With this insight, we performed new model calculations (section 2.6) that will put the essence of correlated narrow-gap semiconductors into a wider context. Finally, section 2.7, is devoted to the role played by lattice degrees of freedom and effects of the electron–phonon coupling in systems such as FeSi.

2.1. Early phenomenological scenarios

The endeavour to deduce a phenomenological scenario from experimental findings for FeSi was pioneered by Jaccarino et al in their seminal work [12]. There, they modelled in particular

Electron-doped FeSi$_{1-x}$P$_x$: it would be interesting to extend the work of Lacerda et al [130] to magnetoresistance and thermopower measurements. The former could substantiate the claim of an itinerant nature of magnetism in Fe$_{1-x}$Co$_x$Si. The latter is motivated by the finding that Co-doping yields a large $n$-type thermopower [233] in Fe$_{1-x}$Co$_x$Si. Does FeSi$_{1-x}$P$_x$ perform better (because the transition-metal site is less perturbed)?

Dependence of $T^\text{max}_\chi$ in Fe$_{1-x}$Ru$_x$Si: is coherence controlled by the fundamental gap or the Hund’s rule coupling? (See the discussion of the resistivity of the substitution series in section 3.2, and the model section 2.6 on Hund’s physics.)

2.1.1. Density-of-states models. A regular semiconductor has dispersive valence and conduction bands, separated by a gap $\Delta$ (top); narrow-band DOS ($W < \Delta$) as proposed by Jaccarino et al [12] (middle); Gaussian DOS model as proposed by Mani [240] (bottom).
structure. Spin excitations are thus governed by the same energy scale, \( \Delta \), as charge excitations, stating that in band-insulators the charge and the spin gap are equal. Using a constant density of states for the low-temperature susceptibility of FeSi, we find \( \Delta = 96 \text{ meV} \). However, the fit is of worse quality than the one shown in figure 4(b, left) that uses a spin model (see paragraph below). In order to account for a maximum in \( \chi(T) \) at \( T^{\text{max}} \) within the one-particle framework, Jaccarino et al. [12] and later Mandrus et al. [127] proposed a strong renormalization of the bandwidth, as schematically shown in figure 9 (middle). Indeed the finite bandwidth \( W \) can lead to anomalies in thermodynamic and transport properties. Via (4) a good fit of the experimental susceptibility is achieved for \( \Delta \approx 80 \text{ meV} \) and \( W = 65 \text{ meV} \) [127]. Also the electronic contributions to the specific heat and the thermal expansion coefficient were faithfully modelled with parameters of similar magnitude [127]. In this picture, the DOS has to exhibit features that are narrower than the gap itself, \( W < \Delta \), as realized, e.g. for the \( f \)-electron density in Kondo insulators. It was however concluded that the renormalization effects necessary to reach such an electronic structure for FeSi are largely beyond conventional band-theory [127] and deemed rather unrealistic [12]. Finally, Mani [240] proposed a DOS-model with a Gaussian density, see figure 9 (bottom). The Gaussian tails of the excitations are supposed to describe a finite in-gap density arising from disorder-derived localized states, but could also, qualitatively, account for an incoherence-induced (yet temperature independent) broadening. This model was shown to yield a reliable fit to the conductivity under pressure for pristine FeSi, as well as for the FeSi\(_{1-x}\)Ge\(_x\) data of [175], foreshadowing the influence of quasi-particle incoherence seen in more sophisticated approaches discussed later in this section.

### 2.12. Spin models

A scenario opposite to the extended Bloch band picture is that of localized moments. For an excited state of spin \( S \) separated from the ground state by a spin gap \( \Delta_s \), one expects in the limit \( k_B T \ll \Delta \), a magnetic susceptibility

\[
\chi = N_\text{Av} \mu_B^2 \frac{S(S+1)}{2} \exp\left(\frac{-\Delta_s}{k_B T}\right),
\]

where \( N_\text{Av} \) is Avogadro’s constant and \( g_s \) the gyromagnetic ratio. Using the standard value \( g_s = 2 \), Jaccarino et al. [12] found for FeSi a spin gap \( \Delta_s = 69 \text{ meV} = 795 \text{ K} \times k_B \) for a spin state \( S = 1 \), while the best fit was achieved for \( S = 1/2 \), \( \Delta = 65 \text{ meV} = 750 \text{ K} \times k_B \), albeit with an enhanced \( g_s \approx 4 \). From the data in figure 4(b) we find \( \Delta_s = 50 \text{ meV} \) and \( S(S+1)/2 + g_s^2 = 14.6 \), which yields for \( g_s = 2 \) a value \( S = 0.8 \), to give a satisfactory description. Above 700 K, as discussed in section 1.2.2 (see again figure 4(b, left)), a Curie–Weiss law, \( \chi = N_\text{Av} \mu_B^2 g_s^2 / (3k_B(T - \theta)) \), gives a faithful representation of the experimental data, with a fluctuating moment \( \mu_s \) that is actually close to the single-ion limit of iron for \( S = 1 \), where \( \mu_s / \mu_B = g_s \sqrt{S(S+1)/2} = 2.6 \mu_B \) for \( g_s = 2 \). This good phenomenological description of the different temperature regimes in the susceptibility notwithstanding, further insights are needed to establish a conclusive microscopic picture. In fact the presented local-moment scenario is at odds with magnetism under doping or lattice expansion, established to be of itinerant origin (see the discussion in section 1.2.2). Also in the case of FeSb\(_2\), inelastic neutron spectra point towards an itinerant origin of magnetic fluctuations [241].

#### 2.1.3. Other scenarios

Varma [242] stressed the importance of going beyond a description in terms of the Kondo-lattice model, to include effects of itinerancy and charge fluctuations for both FeSi and heavy-fermion materials. He proposed that a temperature-induced mixed valence could be at the origin of the anomalous behaviour of FeSi.

Misawa and Tate [243] reproduced the experimental susceptibility of FeSi in an itinerant model when going beyond the Pauli expression, (4), by including a scattering amplitude in the Fermi-liquid picture. The inclusion of scattering effects allowed to employ the model density-of-states of [12, 127] and shown in figure 9 (middle), albeit in a more realistic parameter regime, namely \( W > \Delta \).

Finally, different variants of spin-fluctuation theory were invoked to explain magnetic and thermal properties of FeSi: Takahashi and Moriya [244] discussed FeSi as a nearly itinerant ferromagnetic semiconductor with a temperature-induced local moment; Evangelou and Edwards [245] moreover included important effects of a temperature-dependent density-of-states self-consistently and found qualitative agreement with experiment.

### 2.2. Band-structures and many-body perturbation theory

In this section, we will review applications of \textit{ab initio} electronic structure methods to correlated narrow-gap semiconductors, as well as to the prototypical Kondo insulator Ce\(_3\)Bi\(_4\)Pt\(_3\). We will discuss results from effective one-particle theories, such as DFT [80, 81], hybrid functionals, and static mean-field DFT \(+ U\) [246]. These methods provide very valuable insights into the complexities and chemistry of real materials, such as the nature of bonding (ionic versus covalent), crystal-field splittings, multi-orbital effects, as well as qualitative information on instabilities towards (itinerant) magnetism. Such results are instrumental for constructing effective low-energy many-body models, that will be discussed in section 2.3. Going beyond density functional based methods, we also include here some results from Hedin’s GW method [247–249], a many-body perturbation theory, as well as constrained random phase approximation (cRPA) [250] estimates of local (Hubbard \( U \), Hund’s \( J \)) interactions for use in low-energy many-body models.

#### 2.2.1. FeSi, other B20 compounds, and oxides

FeSi and FeGe as well as their \( 4d \) homologues RuSi and RuGe crystallize in the so-called B20 structure. Despite being cubic (space group P2\(_1\)3) the compounds’ crystal-structure is quite complex: there are four formula units per unit-cell, each Ru (Fe) has seven Si (Ge) neighbours, and the absence of inversion symmetry causes a Dzyaloshinskii–Moriya-derived spin spiral structure in the magnetically ordered phase of

\[\text{FeSi} \quad (4\text{P}) \quad \text{for a constant density of states} \]

\[\rho(\epsilon) = \rho_0 \left[ \theta(\epsilon - \Delta/2) + \theta(-\epsilon - \Delta/2) \right] \]

and \( \Delta(\epsilon) = \epsilon^2 \left[ \theta(-\epsilon - \Delta/2) \sqrt{-\epsilon - \Delta/2} + \theta(\epsilon - \Delta/2) \sqrt{-\epsilon + \Delta/2} \right] \].

\[\text{For example, } \rho(\epsilon) = \rho_0 \left[ \theta(\epsilon - \Delta/2) + \theta(-\epsilon - \Delta/2) \right], \quad \text{and } \Delta(\epsilon) = \epsilon^2 \left[ \theta(-\epsilon - \Delta/2) \sqrt{-\epsilon - \Delta/2} + \theta(\epsilon - \Delta/2) \sqrt{-\epsilon + \Delta/2} \right] \]
FeSi and FeGe: insights from band theory. The described distortion is crucial for the insulating ground-state in the silicides: indeed, as shown in figure 10 (left), rock-salt FeSi is metallic within DFT. Nonetheless the DOS at low energies is quite small, as the bands that cross the Fermi level are quite dispersive. With increasing distortion the Fermi level crossings are avoided and an indirect gap of 100–200 meV is opened [251, 253, 254, 258]. The atomic displacements of the distortion conserve the cubic Bravais lattice but still reduce the point group to tetrahedral. As a consequence the transition metal 3d-orbitals split into a low-lying \( z^2 \) and two doublets \( x^2-y^2 \) and \( xy \) and \( xz, yz \). The latter two are sufficiently separated in energy so that—for a 3d\(^{5}\) configuration for Fe/Ru—the three lowest orbitals are completely filled and the paramagnetic DFT solutions become insulating.

For the silicides, DFT correctly finds this paramagnetic insulator to be the ground-state. As discussed by Mazurenko et al [254] for FeSi, the band-gap does not primarily originate from crystal-fields, but from distortion-induced changes in the hybridization between the iron atoms and their environment. To illustrate this, we have constructed maximally localized Wannier functions for a subspace consisting of the Fe-3d and Si-3s and 3p orbitals of a generalized gradient approximation (GGA) band-structure calculation. The use of this atom-centred local basis allows to empirically study the influence of inter-atomic hybridizations onto the electronic structure. Indeed, we can scale all hybridizations between each iron atom and all other atoms in the unit-cell with a factor \( \alpha \leq 1 \). Figure 11(a) displays the results: akin to the (direct gap of the) periodic Anderson model, the gap in FeSi changes linearly with the strength \( \alpha \) of the hybridization. In fact, below \( \alpha \lesssim 0.75 \) the system becomes metallic. Hence FeSi can be categorized as a covariant hybridization-gap semiconductor [251, 254, 259–261]. However, there are qualitative differences to the case of canonical Kondo insulators. An individual scaling of inter-atomic hybridizations in FeSi (results not shown) divulgates that the gap is mainly driven by inter-iron hybridization— as opposed to hybridizations between localized orbitals and conduction-electron bands as in, e.g. Ce\(_3\)Bi\(_4\)Pt\(_3\) (see section 2.2.3). In fact, Mazurenko et al [254] investigated the spread of their Wannier functions and found that the ones centred on the iron atoms have substantial weight at neighbouring iron sites—in stark contrast to localized f-orbitals.
in Kondo insulators. As a consequence the highest valence and lowest conduction bands have different orbital characters. Instead, we will see in section 2.2.3 that in DFT calculations for Ce3Bi4Pt3, the dominant character is the same on both sides of the gap. This distinction between these two systems will be further discussed in section 2.6 for prototypical many-body models, and in section 2.5.4 in the context of realistic many-body calculations.

At this point, we find it instructive to compare the covalently bonded hybridization-gap insulator FeSi to the oxide LaCoO3. Both compounds share, at first glance, several empirical similarities, such as the insulator-to-metal transitions with rising temperature and a strongly non-monotonous magnetic susceptibility (see figure 2). Here, we limit the comparison to the ground-state as described within band-structure methods. To obtain an insulating Kohn–Sham spectrum for LaCoO3, the inclusion of exchange effects beyond the local-density approximation (LDA) or GGA functionals is required. Figure 11(b) displays the band-structure obtained using the mBJ functional [262]. There, the gap, \( \Delta_{\text{mBJ}} \approx 1.9 \text{ eV} \), vasty overestimates the experimental finding of 100 meV [263] (see table 1). In the shown DFT solution, the nominal six \( d \)-electrons of cobalt fully populate the \( t_{2g} \) orbitals that are well separated from the thus empty \( e_g \) states, resulting in a low-spin configuration. In order to elucidate the dominant origin of the \( t_{2g} - e_g \) splitting, we have performed the same analysis as for FeSi, i.e. we have constructed a local Wannier basis and scaled down the hybridizations of the cobalt atoms with their environment. Contrary to FeSi, this procedure causes the gap to in LaCoO3 to increase. Indeed the visible shrinking of the \( e_g \)-dispersion trumps the slight downshift of its centre-of-mass. Therefore, the gap-formation in this setup is dominated by local crystal-fields, as previously observed by Křápek et al [263]. For a discussion of how the covalent versus ionic gap-formation impacts finite-temperature properties, including electronic correlation effect, see section 2.4.2 for a sequel of the FeSi-LaCoO3 comparison and section 2.5 for a discussion of the bigger picture.

We now return to band-theory results for FeSi: incorporating Hubbard-\( U \)-like interactions via static mean-field LDA + \( U \) calculations, Anisimov et al [265] proposed that FeSi could be driven through a first-order transition to a ferromagnetic metal by applying a magnetic field of \( B_c \approx 170 \text{T} \). The field would induce a low-spin (paramagnetic insulator) to high-spin (ferromagnetic metal with a moment of \( 1\mu_\text{B/Fe} \)) transition. They further argued that the anomalous behaviour of FeSi in the absence of any field could be propelled by the proximity (in total energy) to the critical end point, \( T_c \approx 280 \text{ K} \), of that transition. This interpretation gives \textit{ab initio} support to the early theory that FeSi is a nearly ferromagnetic system (see [244, 245] and section 2.1 above). Subsequently, measurements in ultra-high fields (see the discussion in section 1.2.2) confirmed the existence of a discontinuous metamagnetic semiconductor-to-metal transition with \( B_c \approx 355 \text{T} \) and \( T_c < 77 \text{ K} \) [162, 163].

Allowing for spin-polarization, iron germanide, FeGe, is predicted to be a ferromagnetic metal with an ordered moment of \( \sim 1 \mu_\text{B} \) within DFT [198, 266–268], DFT + \( U \) [269], or hybrid functional methods [267], in congruence with experiment [188, 189, 270]. Also some of the observations of the substitutational series FeSi\(_{1-x}\)Ge\(_x\) (see section 1.2.3 for details of experimental findings) are qualitatively captured within band-theory: Yamada et al [266] found a transition from a non-magnetic-insulator to ferromagnetic-metal for \( x \approx 0.5 \) in DFT. Using DFT supercell calculations, Jarlborg [198] found ferromagnetism above \( x = 0.3 \), suggested to arise from a combination of an increased volume and substitutional disorder. Within LDA + \( U \), Anisimov et al [269] found a first order transition to a ferromagnetic metal at \( x = 0.4 \) for a reasonable value of \( U \). These findings are in rough agreement with the experimental critical substitution \( x \approx 0.25 \) [97].

If the dominant control parameter in the FeSi\(_{1-x}\)Ge\(_x\) series was the unit-cell volume, applying pressure to FeGe should suppress magnetism and drive the system insulating.
Indeed, FeSi has a 12.5% smaller unit-cell than FeGe, which can roughly be reached at 25 GPa [197]. In experiment [198] the Curie temperature is indeed suppressed at a critical pressure of $P_c \approx 19$ GPa. However, even above this pressure a residual conductivity persists, so that it was suggested to arise from zero-point motion [198]. Neglecting the latter effect, band theory finds a pressure-driven phase transition from a ferromagnetic-metal to a non-magnetic-insulator. The critical pressure varies however by a factor of at least four, depending on the functional used in DFT (LDA: $P_c \leq 10$ GPa, PW91: $P_c = 40$ GPa) [267]. The best agreement with experiment is found when optimising the crystal structure with GGA and use LDA for the electronic structure ($P_c = 18$ GPa) [268].

Also some aspects of the non-isoelectronic substitution series Fe$_{1-x}$Co$_x$Si [99, 100] are qualitatively captured within band-theory. Using the virtual crystal approximation in DFT, Morozumi and Yamada [271] found a transition to a half-metallic ferromagnet for finite dopings up to 50%. These results are consistent with the experimentally evidenced linear increase of the magnetic moment with $x$ up to 30%. Indeed, as was proven experimentally by Paschen et al [98], the spontaneous moment in Fe$_{1-x}$Co$_x$Si does not originate from moments localized on the Co-dopants, but is of itinerant origin, and thus, in principle, amenable to band-theory via Stoner physics.

At higher dopings, the theoretical moment [271] in Fe$_{1-x}$Co$_x$Si decreases less quickly than the experimental one [100], probably for lack of treating the randomness of the alloy properly. Indeed, the theoretical magnetic structure is very sensitive to disorder [272], before the diamagnetic metallic state of the end member CoSi is reached.

Finally, other band-structure intricacies of B20-compounds include the recent findings of Kühler et al [229] of a non-vanishing electronic Berry phase in the insulating silicides FeSi, RuSi, and OsSi, corresponding to a crystalline chirality that heralds the possibility of realizing a macroscopic electric polarization in these systems.

In all, for the discussed properties, band-structure methods give very valuable information, as well as an often semi-quantitative description. We now turn to physical properties where this is not the case.

**FeSi: Failures of band-theory.** Standard band-structure methods work at zero temperature. Therefore, they cannot access the unusual spectral and magnetic behaviour of FeSi at finite temperatures. Yet, we can compare band-theory results to the experimental properties at low-temperatures.

Spectral properties: While DFT-related methods in principle only produce an auxiliary Kohn–Sham spectrum, it has become common practice to compare the latter to the excitation spectrum of the solid. In the case of FeSi, the Kohn–Sham band-gap is notably overestimated, which is in stark contrast to the usual underestimation of band-gaps in conventional semiconductors. Indeed while activation law fits e.g. to the resistivity extract $\Delta_{\text{indir}} = 59$ meV (see figure 4(a), left),

\[
\Delta_{\text{indir}} = 59 \text{ meV}
\]

Figure 12. Angle-resolved photoemission spectrum (ARPES) of FeSi. (a) left: ARPES intensity obtained from a He-I source (21.23 eV) from the (100) surface at $T = 10$ K; right: DFT bands (white dashed lines) superimposed on an intensity plot obtained from fitting a phenomenological self-energy correction on top of DFT to the experimental data: $\Sigma(\omega) = g_0 \frac{\omega + i \gamma_l}{\omega + i \gamma_l + \frac{\mu}{\omega + i \gamma_l} - l \delta}$ with $g_0 = 10.5$ eV$^2$, $\gamma_l = 6.5$ eV, $g_l = 0.0093$ eV$^2$, $\gamma_l = 0.1103$ eV. (b) energy-distribution curve at $k_{\parallel} = 0$ A$^{-1}$. Reprinted figure with permission from [264]. Copyright 2008 by the American Physical Society.

DFT finds a value about twice as large: $\Delta_{\text{indir}}^{\text{GGA}} = 130$ meV [251, 253, 254, 258, 259].

As will be discussed in the theory (section 2.4.1), the discrepancy between band-theory and experiment is rooted in dynamical electronic correlation effects that renormalize the excitation spectrum. On an empirical level, this physics can be particularly well illustrated on the basis of photoemission [74] and angle-resolved photoemission spectroscopy (ARPES) [73, 75, 77, 264] experiments. Comparing the photoemission intensity with band-structure results, see figure 12, one witnesses (i) a notable shrinking of the charge gap, and (ii) a narrowing of the bands near the Fermi level. This renormalization of bands has been analysed in terms of phenomenological self-energies $\Sigma(\omega)$ [73–75, 77, 264] that allow for a Brinkman-Rice-like bandwidth-narrowing but neglect orbital- and momentum dependencies. Performing a low-energy expansion, $\Sigma(\omega) = (1 - 1/Z)\omega - i\gamma\omega^2 + \mathcal{O}(\omega^3)$, of the self-energy form used by Klein et al [264] yields at $T = 10$ K a substantially reduced quasi-particle weight $Z = 0.5$ and a scattering coefficient $\gamma = 7/eV$. Despite the simplicity of the Ansatz, the resulting effective mass $m^* / m_{\text{DFT}} = 1/Z = 2$ is quite close to the value found in realistic many-body approaches (discussed in section 2.4.1).

 Nonetheless, DFT calculations give good estimates for the change of the gap with the unit-cell volume, $d \ln \Delta / d \ln V \approx -6$, see [181]. Also note that there is a non-negligible dependence of the charge gap on the exchange-correlation potential used, e.g. $\Delta_{\text{DFT}} = 85$ meV, congruent with the gap originating from hybridization effects, see below.
Ground-state properties: Also in the true realm of DFT, ground-state properties, there are findings that are at odds with experiment, or are at least quite unusual: the equilibrium unit-cell volume of FeSi (experimentally: $V_0 \approx 89.6$ Å$^3$ [273, 274]) comes out slightly too small in DFT. While overbinding is expected for the LDA functional ($V_0 = 84$ Å$^3$ [275]), surprisingly also the typically underbinding GGA underestimates the volume ($V_0 = 88.96$ Å$^3$ [275], $V_0 = 88.1$ Å$^3$ [276]). More dramatic are the discrepancies in the bulk modulus: while DFT yields values above 200 GPa (LDA: $B_0 = 260$ GPa, GGA: $B_0 = 220$ GPa [275]), the experimental values scatter in the range $B_0 = 115$–185 GPa [117, 273, 277–279] with, however, a single outlier of 209 GPa [115]. These findings may be interpreted as arising from electronic correlation effects that decrease participation of charges in the bonding. More electron-lattice properties will be discussed in section 2.7.

The 4$d$ homologue: RuSi. The 4$d$-analogue of FeSi, RuSi, is a semiconductor with a medium-sized gap of 260–310 meV as inferred from resistivity measurements [18, 21] and ~400 meV from optical spectroscopy [18, 20]. As mentioned in the introduction, there are no anomalies in the temperature dependence of spectroscopic or magnetic observables of RuSi.

It is plausible that the extended 4$d$-orbitals of Ru make for a less correlated electronic structure that is, then, more amenable to band-theory. In fact, DFT yields a band-structure similar in shape to FeSi, see figure 13 for a comparison within GGA. Yet the dispersions are enhanced and a larger indirect gap is present: 230 meV [253, 280–282] without, 170 meV with the inclusion of spin–orbit coupling. Thus, contrary to FeSi, the gap is smaller within band theory than in experiment, as is usually the case for uncorrelated semiconductors. The origin of the gap-underestimation is typically attributed to an insufficient treatment of the exchange part [283] of the DFT exchange-correlation functional. An *ab initio* many-body approach that excels, among others, at describing band-gaps is Hedin’s GW method [247–249]. Here, we employ the so-called quasi-particle self-consistent variant QSGW [284] to RuSi [30]. QSGW was shown to yield improved band-gaps [286], particularly for sp-semiconductors, while it overestimates gaps in some d-electron systems [286] (see also the discussion of FeSb$_2$ and related materials below). As can be seen in figure 13, the screened exchange self-energy from the GW indeed widens the band-gap; we find a value of about 370 meV, slightly above experimental values. We will discuss the thermopower of RuSi and the substitution series Fe$_1$–Ru$_3$ in section 3.2. Model calculations in section 2.6 will be used to rationalize the different behaviours of FeSi and RuSi.

2.2.2. The marcasites: FeSb$_2$, FeAs$_2$ & Co. The family of FeSb$_2$ [30, 65, 287], FeAs$_2$ [30], FeP$_2$ [31, 32], CrSb$_2$ [65], their 4$d$-analogues RuSb$_2$ [288], RuAs$_2$ [289], as well as their 5$d$ osmium-based members [290] crystallize in the regular FeS$_2$ marcasite structure (orthorhombic space group Pnnm) with two formula units per unit cell.

In this structure, the transition metal ions are surrounded by distorted pnictogen octahedra, that share corners along the c-axis. According to ligand field theory, the transition metal 3$d$-orbitals split into the lower $t_2g$ and higher-lying $e_g$ orbitals. The existence of two different transition metal–pnictogen distances causes the $t_2g$ to further split into a lower doublet $\Delta$ and a higher single orbital $\Xi$. In the ionic picture, the transition metal would be in a 3$d^4$ low-spin configuration, with an empty $\Xi$ orbital [129, 291]. From this point of view, a (thermal) population of the $\Xi$ orbital may cause the metallization [291], as well as—via a spin-state transition [128, 129]—the enhanced paramagnetism in FeSb$_2$. As will be briefly discussed in section 2.4.2, such a scenario was shown to be realized in LaCo$_3$.

DFT. Subsequent band-structure calculations by Madsen *et al* [292], however, suggested a more covalent rather than ionic picture of bonding in the marcasites. Indeed a stabilization of those 3$d$-orbitals pointing towards the ligands was evidenced, which in particular causes a lowering of the $e_g$ bands. As a consequence the transition metal configuration of all iron-, ruthenium-, and osmium-based marcasites from above is close to $d^9$ rather than $d^8$ (see also the discussion surrounding figure 25).

In FeSb$_2$ and RuSb$_2$ the stabilization of the $e_g$ orbitals with respect to ligand field theory occurs to the extent that density functional theory in fact yields a metallic ground-state [87, 292–294], with small electron pockets halfway between the $\Gamma$ and $Z$ high symmetry points, and corresponding hole pockets at all corners, $R$, of the orthorhombic Brillouin zone, see figure 14(a) for FeSb$_2$.

It is important to notice that these pockets are of different orbital characters [292, 293] (see also the discussion for 30 In QSGW a hermitianized self-energy is fed back into the band-structure code so as to eliminate any dependence on the DFT starting point. Here we use a full-potential linearized muffin-tin orbital (FPLMTO) method [285].

![Figure 13. Comparison of FeSi and RuSi. Shown is the band-structure of the isostructural and isoelectronic compounds FeSi (red, dashed) and RuSi (dark blue) in the GGA. The indirect charge gaps are $\Delta_{\text{indir}} = 127$ meV and $\Delta_{\text{indir}} = 172$ meV for FeSi and RuSi, respectively. In RuSi the inclusion of the spin–orbit coupling leads to a notable splitting of the bands. Many-body corrections within QSGW (light blue) boost the gap to $\Delta_{\text{indir}} = 370$ meV, slightly above the experimental $\Delta_{\text{indir}} = 260$–310 meV.](image-url)
FeSb\(_2\): in the global coordinate system, the electron pocket is dominantly of \(z^2\) and the hole of \(x^2 - y^2\) character \([293]\)\(^{31}\).

With FeSb\(_2\) and RuSb\(_2\) being insulators at low temperatures, band-structure methods fail qualitatively to give the correct ground-state spectrum (contrary to the case of FeSi). This shortcoming notwithstanding, conventional band-theory does reproduce some experimental findings for iron antimonide even on a quantitative level: the unit-cell volume and bulk-modulus \([129]\) are indeed well captured \([295]\)\(^{32}\).

It is instructive to further compare FeSb\(_2\) to its isostructural, isoelectronic relatives FeAs\(_2\) and FeP\(_2\). The LDA band-structure of the former is shown in figure 14(b), for the latter compound, see \([295, 296]\). With respect to FeSb\(_2\), the chemical pressure of the larger As atoms is almost isotropic, and the \(cr/a\) ratio remains virtually constant\(^{32}\). Consequently, the bands of FeAs\(_2\) are similar to those of FeSb\(_2\), albeit a finite gap opens at the Fermi level. We find \(\Delta_{\text{adir}} = 225\) meV within LDA and \(\Delta_{\text{indir}} = 275\) meV in GGA \([294]\), only slightly larger than the experimental 200–220\(\) meV \([24, 30]\). Below, we will discuss why—in moderately correlated materials—such good agreement for band-gaps is often fortuitous. First, however, we will review works in which methodologies beyond DFT were employed to yield an insulating ground-state for FeSb\(_2\).

\(^{31}\)In a local coordinate system in which the local projection of the \(d\)-block of the Hamiltonian is as diagonal as possible (see \([294]\) for details), the electron pocket is mainly of \(d_{xy}\) character, and the hole pocket is formed by the now degenerate \(d_{xy}\) and \(d_{xy}\) orbitals.

\(^{32}\)We use \(a = 5.3\) \(\AA\), \(b = 5.98\) \(\AA\), \(c = 2.88\) \(\AA\) \([30]\); as a function of external pressure, the \(cr/a\) ratio slightly decreases \([129]\).

**Figure 14.** Density functional and many-body perturbation theory. Band-structures of FeSb\(_2\) and FeAs\(_2\) as obtain from LDA (full lines) and Hedin’s \(GW\) approximation (dashed lines). Note that, contrary to LDA, \(GW\) correctly finds FeSb\(_2\) to be an insulator. (a) FeSb\(_2\). (b) FeAs\(_2\). Reprinted figure with permission from \([294]\), Copyright 2010 by the American Physical Society.

**Table 2.** \(GW\) scattering rates. Particle-hole-asymmetry of the scattering amplitude \(\Gamma\) within the \(GW\) approximation, as extracted by fitting the average \(d\)-orbital self-energy (in the Kohn–Sham basis) to \(\Delta\Sigma(|\omega| < 5\) eV\) = \(\Omega\omega^2\). Reprinted table with permission from \([294]\), Copyright 2010 by the American Physical Society.

| \(\Gamma\) (eV\(^{-1}\)) | FeSb\(_2\) | FeAs\(_2\) |
|------------------------|------------|------------|
| \(\omega < 0\)         | 0.15       | 0.08       |
| \(\omega > 0\)         | 0.02–0.05  | 0.02–0.03  |

**Hybrid functional methods.** That band-gaps are underestimated in DFT methods is owing to an insufficient treatment of exchange effects. One way to improve upon this, is to include fractions of exact exchange via so-called hybrid functionals \([297]\). Using the B3PW91 functional \([283]\) for the \(d\)-orbitals of the iron atoms, a paramagnetic solution with a much too large gap of 600 meV (900 meV) was found for FeSb\(_2\) (FeAs\(_2\)) \([294]\). See also figure 45 that shows the electronic structure of FeSb\(_2\) in the presence of an antisite defect, using the modified Becke–Johnson exchange potential \([262]\).

**GW approximation.** To put the latter findings as well as dynamical effects of electronic correlations into context, Hedin’s (non-self-consistent) \(GW\) approximation \([247]\) was applied to FeSb\(_2\) and FeAs\(_2\) \([294]\). The resulting one-particle dispersions—computed from the \(GW\) self-energy \(\Sigma\) via \(\epsilon_{GW} \approx Z^{-1} \int \rho_{\text{LDA}}(\epsilon) + \Re \Sigma_{\epsilon} \rho_{\text{LDA}}(\epsilon) + \Re \Sigma_{\epsilon} \rho_{\text{LDA}}(\epsilon + \mu)\)

—are shown in figure 14 in comparison to LDA results: in FeSb\(_2\) a small gap opens, in agreement with experiment, while \(GW\) mildly increases the gap in FeAs\(_2\), slightly deteriorating the congruence with experiment.

To analyse the \(GW\) band-structures, we note that, quite generally, there are two competing effects \([298]\) in electronic structures beyond DFT: (i) the inclusion of exchange...
Table 3. Gaps and interactions in the marcasites. Tabulated are the indirect charge gaps from experiment (see table 1 for references), LDA and QSGW. Moreover the Hubbard U and Hund’s J for the transition-metal d orbitals are given. These have been obtained by applying cRPA on top of the QSGW electronic structure in a maximally-localized Wannier setup [300] that includes transition metal d and pnictogen p-orbitals. Screening has been eliminated in a window [−3, +3]eV around the Fermi level. Shown are d-orbital averages.

|        | FeSb$_2$ | FeAs$_2$ | RuSb$_2$ |
|--------|----------|----------|----------|
| $\Delta_{\text{indir}}^{\text{exp}}$ (meV) | 30       | 200–220  | 290–330  |
| $\Delta_{\text{indir}}^{\text{LDA}}$ (meV) | —        | 220      | —        |
| $\Delta_{\text{indir}}^{\text{QSGW}}$ (meV) | 150      | 430      | 430      |
| $U_{\text{RPA}}^{\text{cRPA}}$ (eV)     | 3.9      | 3.4      | 2.6      |
| $J_{\text{RPA}}^{\text{cRPA}}$ (eV)     | 0.63     | 0.63     | 0.44     |

Figure 15. Comparison of FeSb$_2$, FeAs$_2$, and RuSb$_2$ within QSGW. Panel (a) shows the QSGW DOS for FeSb$_2$ (red), FeAs$_2$ (green), and RuSb$_2$ (blue). Panel (b) displays the corresponding band-structures. (a) Density of states. (b) Band-structure.

QSGW and cRPA. For a more in-depth comparison of FeSb$_2$, FeAs$_2$, and RuSb$_2$ we performed new QSGW [284] and cRPA [250] calculations. The QSGW DOS and band-structures are shown in figures 15(a) and (b), respectively, while table 3 summarizes key results. There, we also include estimates for the values of the Hubbard U and Hund’s J interactions (see caption for computational details). In all marcasites studied here, the self-consistency increases the gap with respect to the above one-shot GW calculations. In FeSb$_2$ $\Delta_{\text{indir}}$ reaches 150 meV—a value five-times as large as in experiment. Also in the case of FeAs$_2$, the gap is significantly overestimated, while QSGW yields a more reasonably sized gap for RuSb$_2$.

The tendency to overestimate band-gaps (see also the case of RuSi above) is rooted in two deficiencies: (i) insufficient screening of the Coulomb interaction within RPA [286, 301] (e.g. neglecting particle-hole correlations), (ii) an underestimation of dynamical self-energy effects in the perturbative (first order) GW approximation. Both problems can be addressed by combining GW with dynamical-mean-field theory (DMFT) [302, 303] in so-called GW+DMFT [304] methods (see [305–307] for recent reviews, and [308, 309] for the QSGW+DMFT variant).

These sources of error notwithstanding, we can still analyze the trends in our results and speculate about the reasons why these three materials behave very differently in experiment (see figures 1 and 2 for the resistivities and magnetic susceptibilities, respectively). To do so, we further compute interaction matrix elements within cRPA, see table 3. Our estimates for the Hubbard U and Hund’s J are quite similar for FeSb$_2$ and FeAs$_2$, while significantly smaller for RuSb$_2$. From the point of view of the charge gap, however, FeAs$_2$ and RuSb$_2$ contribute to the self-energy widens gaps and bandwidths [247, 283], and (ii) dynamical renormalizations, that (to linear order) give rise to the quasi-particle weight $Z$, reduce gaps and band-widths [260, 299].

When applying GW to FeSb$_2$ the competition of both effects is noticeable: while occupied bands are visibly narrowed and renormalized towards the Fermi level, unoccupied states are moved up in energy, resulting in the opening of a gap. Turning off dynamical renormalizations, i.e. setting the quasi-particle weight to unity $Z = 1$, yields a larger gap of about 200 meV. From this observation, one can learn two things: (i) the ab initio screened exchange (SEX) included in the GW approach yields a smaller gap-enhancement than the ad hoc admixtures of bare exchange in hybrid functional approaches; (ii) effects of dynamical renormalizations substantially shrink the band-gap [260, 299] with respect to a (screened) Hartree–Fock-like reference. Within GW, values of $Z \approx 0.5$ ($Z \approx 0.6$) are found for FeSb$_2$ (FeAs$_2$) [294]. For the case of FeAs$_2$ one could thus claim that, LDA finds a good value for the indirect gap merely because of an error cancellation, namely the joint neglecting of both exchange and dynamical correlations.

Concomitant with the linear slope of the real-part giving rise to Z, also the imaginary part of the self-energy $\Sigma$ is Fermi-liquid-like, i.e. quadratic in frequency. Interestingly, however, the scattering rate is notably asymmetric with respect to the Fermi level, see table 2.

33 The bare (i.e. unscreened) interaction is slightly larger in the arsenide than the antimonide [294], in line with trends of the Coulomb interaction under external pressure [310, 311]. Indeed the ionic radius of As is larger than Sb and causes chemical pressure. The slightly smaller value of U for FeAs$_2$, on the other hand, is mostly owing to using the same energy window for which screening is eliminated. Since the band-width of FeAs$_2$ is larger than that of FeSb$_2$ there are more excitations outside that window and contribute to the screening. In RuSb$_2$ the dominant change with respect to the iron compounds is the larger extend of the 4d-orbitals, resulting in smaller interaction matrix elements.
In (b) the occupied 4f orbitals are removed from the valence band, and transferred into the core, i.e. it is treated as purely atomic and no longer participate in the bonding. The unhybridized f-level appears as a delta peak at the Fermi level: the resulting density of states (DOS) is metallic.

RuSb$_2$ can be loosely grouped together, while FeSb$_2$ is the odd one out.

Therefore it can be speculated that the difference in physics between FeSb$_2$ and FeAs$_2$ is mainly controlled by the different sizes of the gaps. This is congruent with the observation that observables, such as the resistivity and the magnetic susceptibility, are akin in shape, albeit their characteristic temperature scales differ by the ratio of the respective gaps, $\Delta(\text{FeAs}_2)/\Delta(\text{FeSb}_2)$. In this scenario one would expect that the activated behaviour of the magnetic susceptibility of FeAs$_2$ [24] crosses over to a Curie–Weiss-like decay at $T^*_\text{max} \approx 2200$ K—which is, however above the melting point of the material.

The susceptibility of RuSb$_2$, on the other hand, does not exhibit any signs of activated behaviour [24] (see figure 2), despite having a gap comparable in size to that of FeAs$_2$. Here the difference lies in the vastly disparate interactions. Given the propensity for ferromagnetic order in these systems, one can surmise that the dominant control parameter of spin-fluctuations is the Hund’s rule coupling $J$. Since the latter is smaller by one third in ruthenium antimonide, the prefactor of the activated behaviour is expected to be significantly smaller. Model calculations presented in section 2.6 substantiate these claims (see also section 2.4.2 for the influence of the $J$ onto the mass enhancement of FeSi).

2.2.3. The Kondo insulator Ce$_3$Bi$_4$Pt$_3$. The treatment of 4f-states in effective one-particle theories such as density functional theory is very poor: strongly localized states contribute too much to the bonding as their hybridization with valence and conduction electrons is largely overestimated. Moreover, interaction-driven multiplet effects are completely neglected. Nonetheless, even for Ce$_3$Bi$_4$Pt$_3$, we can gain some interesting qualitative insights from band theory.

Ce$_3$Bi$_4$Pt$_3$ is a cubic intermetallic that crystallizes in the body-centred Y$_3$Sb$_4$Au$_3$ structure [313] with non-symmorphic space-group I-43d, two formula units per unit-cell of lattice constant $a = 10.051$ Å [61, 223]. Ce atoms sit in the Wyckoff position 12a (3/8, 0, 1/4), Pt occupies site 12b (7/8, 0, 1/4), and Bi is located at $(u,u,u)$. Since no value for $u$ has been reported in the literature, we follow [312] and use the ideal position, $u = 1/12$.

Density functional calculations for this compound were pioneered by Takegahara et al [312] within the local density approximation. This work established that Ce$_3$Bi$_4$Pt$_3$ has an insulating ground-state within band theory. In the data shown here, we employed the generalized gradient approximation (GGA) with the Perdew–Burke–Ernzerhof (PBE) functional, and include (unless otherwise specified) spin–orbit coupling for all atoms within the wien2k [314] implementation. As seen from the DOS in figure 16 and the band-structure in figure 17, Ce$_3$Bi$_4$Pt$_3$ is insulating within DFT, with a weakly indirect gap: we find $\Delta_\text{dir} = 141$ meV, $\Delta_\text{indir} = 131$ meV$^{34}$.

On the conduction state side the gap is delimited by a single, rather dispersive band that dips down through $\Delta = 140$ meV. In (b) the occupied 4f electron was removed from the valence and transferred into the core, i.e. it is treated as purely atomic and does no longer participate in the bonding. The unhybridized f-level appears as a delta peak at the Fermi level: the resulting DOS is metallic. Since no refined atomic positions have been reported, we follow [312] and use for Bi the ideal position $\frac{u}{12}$. Orbital characters have been obtained from an angular decomposition within atomic spheres of radius 2.5 Å. Other weight is mainly accounted for by interstitial regions.

(a) Ce$_3$Bi$_4$Pt$_3$: an insulator. (b) 4f$^1$ removed from valence: a metal.

---

$^{34}$The gap within LDA is slightly smaller, we find $\Delta_\text{dir} = 110$ meV, while legacy calculations [312] gave only 30 meV.
the bulk of unoccupied $f$-states around the $\Gamma$-point (see also figure 18(b)). Among the valence states, there is roughly one charge carrier of 4$f$-character per Ce atom. As will be illustrated below, in a purely atomic scenario the 4$f$ configuration would lead to a half-filled level, and thus a metallic band-structure. Hybridizations with other orbitals—most notably the Pt-5$d$ and Bi-6$p$ (see figure 16(a))—lead to the formation of dispersive bonding/anti-bonding states, separated by a hybridization gap. The total number of valence electrons in Ce$_3$Bi$_4$Pt$_3$ is even and thus allows for an insulating solution. Indeed, the valence $f$-electrons of Ce then reside in the bonding bands filling them completely, causing the chemical potential to fall inside the gap. As an illustration, we performed a calculation in which the 4$f$ configuration of each Ce atom was moved from the valence into the core sector, i.e. it was treated as completely localized. The situation is thus qualitatively akin to the generic case of the periodic Anderson model (PAM, see section 2.6) with the interaction turned off for Pt (green), and (3) Ce$_3$Bi$_4$Pd$_3$ (blue). These calculations illustrate that—a change in the spin–orbit coupling along the substitution series mostly affects the valence states, while important modifications on the conduction side of the hybridization gap results from the change in the radial character (Pt-5$d$ to Pb-4$d$). We use the same lattice constant for Ce$_3$Bi$_4$Pt$_3$ and Ce$_3$Bi$_4$Pd$_3$ as motivated in [62].

![Figure 18. Comparison of Ce$_3$Bi$_4$Pt$_3$ and Ce$_3$Bi$_4$Pd$_3$: importance of radial character and spin–orbit coupling. Panel (a) shows the DOS for (1) Ce$_3$Bi$_4$Pt$_3$ (red), (2) fictitious Ce$_3$Bi$_4$Pt$_3$ with the spin–orbit coupling turned off for Pt (green), and (3) Ce$_3$Bi$_4$Pd$_3$ (blue). Panel (b) displays the band-structure of Ce$_3$Bi$_4$Pt$_3$ (red) and Ce$_3$Bi$_4$Pd$_3$ (blue). These calculations illustrate that—the change in the spin–orbit coupling along the substitution series mostly affects the valence states, while important modifications on the conduction side of the hybridization gap results from the change in the radial character (Pt-5$d$ to Pb-4$d$). We use the same lattice constant for Ce$_3$Bi$_4$Pt$_3$ and Ce$_3$Bi$_4$Pd$_3$ as motivated in [62]. (a) DOS. (b) Band-structure.](https://example.com/figure18.png)
displaced towards the Fermi level; the gap as apparent in the DOS shrinks, while the actual gap value increases. The relative stabilization of the noble-metal $d$ states is mirrored also in their filling: within atomic spheres of the same radius (2.5 Å), Pt in Ce$_2$Bi$_4$Pt$_3$ harbours 7.6 $d$-electrons, while in Ce$_2$Bi$_4$Pd$_3$ there are in fact 8.2. Thus, for conduction states the relevant change along the substitution series is the radial character of the noble metal orbitals.

In all, these band-structure considerations suggest that, the bare (i.e. unrenormalized) electronic structure in the substitution series Ce$_2$Bi$_4$(Pd$_{1-x}$Pt$_x$)$_3$ is not only controlled by the spin–orbit coupling but also by the radial character of ligand orbitals. As far as the $d$-$f$ hybridization is concerned, the effect of the latter is in fact dominating and, as a consequence, the Kondo coupling is not constant along the series.

2.3. Many-body models

The above band-structures account for the structural and multi-orbital complexity of the materials under consideration. From them, one can construct [250, 316, 317]—or at least take inspiration for—reductionist many-body models that are to be solved with more sophisticated techniques, allowing for a more proper treatment of electronic correlation effects and finite temperatures.

The models that have hitherto been used to describe correlation effects in FeSi can be grouped into two classes: those based on the PAM, and those related to Hubbard models [35]. We certainly do not aim for completeness in this section and strictly focus on studies and parameter regimes that are directly linked to either FeSi or Ce$_2$Bi$_4$Pt$_3$. Beyond the literature reviewed here, a contrasting juxtaposition of the two quintessential models will be presented in section 2.6.

2.3.1. Periodic Anderson model (PAM). The PAM is believed to contain the essence of the physics relevant in heavy-fermion materials. It describes a periodic array of atomic levels ($f$-states) subjected to a local Coulomb interaction $U$ and a hybridization to a dispersive band of non-interacting ($c$) electrons:

$$H = \sum_{k\sigma} \epsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \epsilon_f \sum_{i,\sigma} n_{i\sigma} + \sum_{k\sigma} V_k (c_{k\sigma}^\dagger f_{k\sigma} + f_{k\sigma}^\dagger c_{k\sigma}) + U \sum_{i} n_{i\uparrow} n_{i\downarrow}. \tag{5}$$

Here $c$, $f$ ($c^\dagger$, $f^\dagger$) are the creation and annihilation operators of the conduction (atomic) states, and we have used a mixed representation where $k$ labels the momentum in the Brillouin zone, while $i$ indexes a lattice site; the spin is denoted by $\sigma$; $\epsilon_k$ is the dispersion of the decoupled $c$-electrons, $\epsilon_f$ the location of the $f$-level, $V_k$ the hybridization between the two electron species. Finally, $n_{i\sigma} = f_{i\sigma}^\dagger f_{i\sigma}$ is the local number operator of the $f$-electrons, and $U$ their on-site Coulomb repulsion. In the non-interacting limit, $U = 0$, the symmetric PAM ($\epsilon_f = 0$, $\epsilon_k$ particle-hole symmetric, e.g. hyper-cubic lattice with only nearest neighbour hopping, $\mu = U/2$) with local hybridization $V_k = V$, is gapped for finite $V$. The direct gap is proportional to the hybridization, $\Delta_{\text{dir}} \propto 2V$, while the indirect gap can be much smaller, one finds $\Delta_{\text{indir}} = \sqrt{4\epsilon_f^2 + 4V^2 - 2V^2}$, where $D$ is half the bandwidth of the (unhybridized) $c$-electron (see figure 27 for a specific case).

In the context of Ce$_2$Bi$_4$Pt$_3$, Riseborough [318] studied the PAM within mean-field theory and found a renormalization of the hybridization $V$, that leads—in the symmetric, half-filled case—to a shrinking of the low-temperature gap. Further he obtained spin susceptibilities in good qualitative agreement with experimental findings for Ce$_2$Bi$_4$Pt$_3$. Sanchez-Castro et al. [141] used a slave boson technique in the strong coupling limit, and applied it to the PAM with a parabolic conduction band. Owing to the particle-hole asymmetry of their model, the authors could in particular study the temperature dependence of the thermopower and the Hall coefficient. It was found that these quantities had opposite signs at low temperatures, congruent with experimental findings for Ce$_2$Bi$_4$Pt$_3$ (thermopower hole-like [124], Hall coefficient electron-like [126] as commonly found also in other heavy-fermion materials, e.g. CeCu$_2$Si$_2$ [319]).

DMFT [302, 303] was first applied to the symmetric PAM by Jarrell et al. [320, 321]. They evidenced that spectra, susceptibilities and thermodynamic quantities behave like heavy-fermion metals at high temperatures (large masses, Curie–Weiss law), and like insulators at low temperatures (activated behaviours). The DMFT approach was in particular shown to capture the competition between the Kondo effect and the magnetic RKKY interaction [321]. Indeed, the DMFT phase-diagram of the half-filled symmetric PAM on the infinite Bethe lattice (see Sun et al. [322] and Rozenberg [323]) exhibits antiferromagnetic long-range order above a second order critical line given by $U_c \propto V^2$ for large $U$. These findings extend Doniach’s ideas [135] for the Kondo lattice to the PAM.

Subsequently, also Rozenberg et al. [79] followed the Kondo insulator route and interpreted experimental findings for FeSi and Ce$_2$Bi$_4$Pt$_3$ in terms of DMFT calculations for the PAM. They showed that the PAM yields, for $U > 0$, the same hierarchy of energy scales, $\Delta_{\text{indir}} < \Delta_{\text{dir}} < \Delta_{\text{eff}}$, that was evidenced for FeSi and Ce$_2$Bi$_4$Pt$_3$ experimentally (see section 1.2). Rozenberg et al. in particular modelled the optical conductivity: congruent with experiments for both compounds (see section 5(b)), the direct gap $\Delta_{\text{dir}}$ is virtually independent of temperature, yet, with rising temperature, it is filled with incoherent spectral weight as the local moment of the $f$-states begin to fluctuate freely [37]. A similar observation was made by

---

35 We shall not discuss the Kondo lattice model that is often invoked for heavy-fermion intermetallics, since effects of itinerancy are certainly of importance in FeSi.

36 See however [64] where a positive Hall signal was found.

37 See also [324] for optical conductivities, and [325] for susceptibilities in the strong coupling limit.
Mutou and Hirashima [326] for the (f-contribution to the) spin-excitation spectrum.

Having in mind FeSi, Figueira et al [327] studied the PAM with a slave-boson approach and computed transport observables. The authors found good qualitative agreement with experimental measurements at low temperatures, when using an ad hoc scattering rate. This finding advocates the presence of sizable many-body renormalizations for the dispersion near the Fermi level.

That DMFT provides an accurate picture of the symmetric PAM was motivated by Tanasković et al [328]. Employing cellular DMFT, they demonstrated the absence of non-local (yet short-range) correlations in regimes sufficiently above the model’s Neél temperature. This dominance of local correlation physics also serves to motivate the accurateness of, e.g. our realistic calculations for the Kondo insulator Ce$_3$Bi$_4$Pt$_3$ in section 2.4.3.

Having in mind anisotropic Kondo insulators, Yamada and Ono [329] studied the PAM with non-local hybridizations. They evidenced, among others, that the uniform and the local magnetic susceptibility virtually coincide down to $T_{\text{max}}$. Further, they studied how interaction effects suppress fluctuations of charge, while enhancing spin fluctuations with respect to the spin and charge susceptibilities that coincide in the non-interacting ($U = 0$) limit.

A model that goes beyond the PAM in that it allows for the correlated band to have finite dispersions was introduced by Continentino, Japiassu and Troper [330]. Indeed the model was designed to encompass both, systems like Ce$_3$Bi$_4$Pt$_3$ (in which the correlated states are close to atomic-like) and compounds such as FeSi (in which d-orbital-derived bands have a finite dispersion), through a scalable hopping amplitude. The model was solved using an equation-of-motion technique in the strong coupling limit. The authors found that, contrary to the symmetric PAM, a critical hybridization was needed to yield an insulating spectrum. It was speculated that this is the reason why covalent insulators are not that common among d-electron systems.

While the PAM exhibits notable similarities to the one-band Hubbard model [331, 332], the band-structure results from the preceding section motivate that narrow-gap semiconductors, such as FeSi and FeSb$_2$, require a description in terms of a multi-band Hubbard model. These are the subject of the next section.

2.3.2. Hubbard models. The first multi-band Hubbard model

$$H = - \sum_{kL\sigma} \epsilon_k c_{kL\sigma}^\dagger c_{kL\sigma} + \sum_{kL \neq L' \sigma} V_{kL\sigma}^{\daggerL'} (c_{kL\sigma}^\dagger c_{L'\sigma} + c_{kL\sigma}^\dagger c_{L'\sigma}^\dagger)$$

$$+ \sum_{dL \neq \sigma'} U_{dL\sigma}^\dagger n_{dL\sigma} n_{L\sigma'}$$

(6)

devised for studying FeSi is owed to Fu and Doniach [333]. They considered two bands ($\epsilon_k = -\epsilon_k^2$) on a hyper-cubic lattice that hybridize locally ($V_k = V$). This setup yields an insulating density of states in the non-interacting limit with a gap that is direct, $\Delta_{\text{dir}} = 2V$. Turning on the Hubbard $U$ interaction (i) renormalizes downwards the bare gap and (ii) introduces a scattering rate. The latter rises with temperature and causes one-particle excitations to broaden, to the extent that incoherent spectral weight appears at the Fermi level. Using an approximate expression for the magnetic susceptibility, good agreement was found with experiment. The model of Fu and Doniach [333] contains the minimal ingredients to describe the essence of correlated narrow-gap semiconductors.

Using a DOS inspired from ab initio calculations, Kuneš and Anisimov [260] solved a one-band Hubbard model with DMFT and stressed the importance of the covalent nature of the hybridization gap in systems like FeSi and FeSb$_2$. Further, they computed optical conductivities and obtained a resistivity that accounts for the non-monotonous temperature dependence shown in figure 4(a), left. Finally, the authors also computed within DMFT the local and uniform magnetic susceptibilities of their model with and without doping. It was found that susceptibilities were much enhanced with respect to the non-interacting system, with a Curie–Weiss-like tail emerging at high temperatures, or, in the doped case, throughout the considered temperature range. Interestingly, the effective fluctuating moment was shown to be notably momentum-dependent—indicating that a picture of spatially localized moments (which is at the origin of Curie–Weiss behaviour) is not appropriate in this setting. Notably, the local susceptibility was found to be much larger than the uniform one (we will come back to this point in section 2.4.1).

Sentef et al [299] studied within DMFT a half-filled two-band model similar to that of Doniach and Fu [333] and investigated the interaction-temperature phase diagram as well as more formal aspects of correlation effects in covalent insulators. The Hubbard $U$ was found to drive the system through a first order transition from a correlated covalent insulator to a Mott insulator. The coexistence region terminates in a critical end point at finite temperatures, above which the hybridization insulator and the Mott insulator are continuously connected via a crossover through a bad-metal phase. While in the non-interacting limit, spin and charge gap coincide, their degeneracy was found to be lifted at finite $U$, with the spin gap decreasing faster than the charge gap upon approaching the insulator-to-insulator transition, in congruence with the experimental gaps of FeSi and Ce$_3$Bi$_4$Pt$_3$ (see section 1.2.

The effect of correlation effects on the optical conductivity of FeSi was also studied by Urasaki and Saso [334] using perturbation theory for a two-band Hubbard model, finding qualitative agreement with the experimental temperature-dependence. They also found remarkable agreement for the thermopower when using a realistic density-of-states [335].

Inspired from their band-structure results (see section 2.2.), Mazurenko et al [254] constructed an effective three-band model to mimic FeSi, and performed one-band DMFT calculations for its density-of-states. In agreement with previous works and experimental observations, they found an effective mass enhancement of about two. In particular, however, Mazurenko et al studied the doped system and rationalized the appearance of ferromagnetism in Fe$_{1-x}$Co$_x$Si in terms of the Stoner criterion, arriving at the conclusion that magnetism
is of itinerant origin. Later, Yang et al. [336] solved the full three-band model of Mazurenko et al., and made the important observation that the Hund’s rule coupling $J$ plays a significant role in stabilizing ferromagnetism.38

Thus, as for experimental observables, also the minimal models for FeSi (two-band Hubbard model) and Ce$_3$Bi$_4$Pt$_3$ (PAM) are exhibiting a quite similar behaviour. This begs the question whether the physics underlying both the models and the actual compounds is in fact akin, or whether the microscopic origin can be distinguished. We will try to answer this question in the following sections using available and new realistic many-body calculations.

### 2.4. Many-body electronic-structure theory

While many-body models have enlightened our understanding about many physical phenomena in correlated systems, they cannot account for all the complexity encountered in real materials. This concerns in particular multi-band effects, such as crystal-fields, multiplet structures, ligand hybridizations, spin–orbit coupling, etc. Given the enormous sensitivity of even fundamental properties of correlated materials, such details do matter. Hence, at times, it is advantageous to turn around the usual modus operandi in which reductionist models are solved before using state-of-the-art methods for realistic many-body calculations. Indeed, most model setups for FeSi and others have used rather unspecific dispersions, at times even particle-hole symmetric ones (that result in a vanishing thermopower). Also for Kondo insulators, the importance of realistic electronic structures was noted already early on [337]. Still, the influence of the Hund’s rule coupling—a genuine multi-orbital effect—was hitherto mentioned in only a very few works.

In this section we will discuss recent ab initio DFT+DMFT calculations for FeSi that build on realistic band-structures and Coulomb interactions. We will first show (section 2.4.1) that for a considerable panoply of physical observables which probe different (yet interlinked) degrees of freedom, excellent agreement is found with respect to experiment. With this validation of the theoretical setup, the microscopic ingredients of the theory can be analysed (section 2.4.2), allowing for a better fundamental understanding of the physics relevant to correlated narrow-gap semiconductors. In section 2.4.3, we will complement these results for FeSi with new calculations for Ce$_3$Bi$_4$Pt$_3$ to highlight microscopic similarities and differences. In section 2.6 the thus gained insight will be used to go back to model-setups and scan relevant parameter regimes so as to put FeSi and Ce$_3$Bi$_4$Pt$_3$ into the wider context of related compounds.

#### 2.4.1. Theoretical description of experimental observables of FeSi

In this section, we will review results from realistic many-body (DFT+DMFT) calculations for FeSi for spectral and optical observables, as well as magnetic susceptibilities [261]. These theoretical findings compare favourably with experimental results from photoemission and optical

---

38For the importance of the Hund’s rule coupling for metallic ferromagnetism, see, e.g. [236–239].
spectroscopies, as well as neutron spectroscopy and other magnetic probes. A discussion of transport observables (resistivity, thermopower, etc) can be found in the thermoelectricity section (section 3.2). The methodology employed for the following results is the realistic extension of DMFT \cite{[302, 303]}, dubbed DFT + DMFT \cite{[315]}, as implemented by Haule et al \cite{[338]}. We refer to the original works \cite{[253, 261]} for technical details.

**Spectral properties.** Figure 19 (top left) displays the DFT + DMFT spectral function for various temperatures. At low temperatures (blue) FeSi is gapped, with a charge gap of $\sim 50$ meV that is narrowed by about a factor of two as compared to band-theory (see figure 13), in congruence with photoemission spectroscopy (see figures 12 and 5(a)). As temperature rises, peaks in the spectral function broaden and incoherent spectral weight spills into the gap and the system increasingly metallizes. Similar spectra have subsequently been obtained by Yanagi and Ueda \cite{[339]} using a perturbative approach.

**Optical properties.** The crossover to a (badly) metallic state is also quantitatively reproduced in the optical conductivity, as shown in figure 19 (right). As noted already in section 1.2, the metallic response is not caused by the gap edges shifting to the Fermi level. Indeed, the first maximum in the absorption does not move substantially with temperature\footnote{We note, however, that the trend in experiment and theory is opposite: the peak moves up with $T$ in theory, while it moves down in experiment. We attribute this to the expansion of the lattice with increasing temperature, which tends to reduce the hybridization gap. The calculations were done at constant volume.}.

The transfer of spectral weight is analysed in more detail in figure 19 (bottom). According to (3) $\Delta n(\omega, T, T_0)$ describes the redistribution of spectral weight up to the energy $\omega$ with respect to a base temperature $T_0 < T$. A positive value of $\Delta n(\omega, T, T_0)$ thus indicates that, for $T > T_0$, there is a net transfer of spectral weight to energies lower than $\omega$. An intersection of $\Delta n$ with the $x$-axis would correspond to a full recovery of spectral weight as imposed by

$$\frac{\partial}{\partial T} \Delta n(\omega, T, T_0) = -\omega$$

\footnote{Note that the latter has been multiplied by a factor of 30.}

**Magnetic properties.** The inset of figure 19 (top left) shows a comparison of the local spin-susceptibility as obtained from DFT + DMFT and the Pauli susceptibility according to (4). The local susceptibility has the expected characteristic temperature dependence as seen in experiment (growing with $T$ at low-$T$, Curie–Weiss-like decay above 500 K, see figure 4(b)). The Pauli expression, as computed from the DFT + DMFT spectral function is very low in magnitude\footnote{We note, however, that the trend in experiment and theory is opposite: the peak moves up with $T$ in theory, while it moves down in experiment. We attribute this to the expansion of the lattice with increasing temperature, which tends to reduce the hybridization gap. The calculations were done at constant volume.} and is, moreover, a monotonously increasing function of temperature, and thus does not account for the observed physics. We further note

---

**Figure 20.** FeSi: neutron spectroscopy. Magnetic (quasi) elastic structure factor displaying scattering near the $q$-points $(1 + \xi, 1 + \xi, 0)$ and $(1 + \xi, 1 + \xi, 1 + \xi)$. Experimental data from Shirane et al \cite{[160]} and Tajima et al \cite{[161]} obtained at 300 K with 41 meV neutrons at $\Delta \omega = 0$. Theory: LDA + DMFT $\Delta \omega = 2$ meV at 290K. Note that the intensities have been arbitrarily scaled.

**Figure 21.** FeSi: coherence of excitations. Shown is the imaginary part of the DFT + DMFT self-energy at zero frequency, $\Im \Sigma_L(\omega = 0)$, as a function of temperature. For the orbital components $L = x^2-y^2, xy, xz, yz$, the inverse lifetime scales with $T^2$. Reproduced with permission from \cite{[261]}.
that the local susceptibility has a magnitude smaller by a factor of ~10 as compared to the measured uniform magnetic susceptibility (see figure 4(b)). To make things worse, previous (one-band) model calculations [260] found that the local susceptibility is larger than the uniform susceptibility, potentially aggravating congruency with experiment. In the periodic Anderson model, the local and uniform susceptibility coincide above \( T_{\text{max}} \) [329]. This impasse is solved as follows. (i) Experimentally: as mentioned before, magnetic neutron spectroscopy experiments [160, 161] evidenced a ferromagnetic susceptibility, i.e. a peak at \( q = 0 \) and equivalent points [41]. This finding strongly suggests that the uniform susceptibility should instead be much larger than the \( q \)-averaged (=local) susceptibility. (ii) Theoretically: using the formalism of Park et al [342], we compute the structure factor \( S(q, \omega \rightarrow 0) \) from our DFT+DMFT data [343]. The agreement with the experimental neutron results as shown in figure 20 is excellent and clearly evidences an enhancement of the susceptibility at ferromagnetic wavevectors. Kranuch et al [344] analysed the shown elastic structure factor of Tajima et al [161] with a phenomenological paramagnetic scattering function, \( S(q, \omega = 0) = \frac{k_B T}{\hbar} \frac{\gamma_0}{1 + (\omega / \chi_0)^2} \), with the static susceptibility \( \chi_0 \), a damping factor \( \gamma \), and a correlation length \( \xi \). Our theoretical simulation reproduces their finding of \( \xi \approx 3.6 \) Å. This is a quite remarkable result, since the correlation length is therewith larger than the distance 2.75 Å between neighbouring iron atoms. This ferromagnetic enhancement in particular implies a uniform susceptibility that is significantly larger than the local one. This qualitative change as compared to previous single-band calculations (see section 2.3) advocates the importance of genuine multi-band effects in FeSi. It is in particular the influence of the Hund’s rule coupling that favours tendencies towards ferromagnetism.

Indeed, using realistic many-body setups solved with a perturbative technique, Yanagi and Ueda [339] discussed the different magnetic ground-states of the isostructural and isoelectronic FeSi and FeGe (see section 2.2.1). In particular they evidenced a strong dependence of the magnetic ordering temperature and moment in FeGe on the strength of the Hund’s rule coupling \( J \). The important influence of \( J \) for spectral properties is discussed in the next paragraph.

2.4.2. A microscopic understanding: spin fluctuations and Hund’s physics. Having validated the theoretical setup through quantitative congruence with experiments, one can take a step back, and analyse the microscopic origin of the evidenced behaviour.

Figure 22. FeSi: influence of the Hund’s rule coupling \( J \). (a) Scattering rate \( \Im \Sigma(\omega = 0) \), (b) effective mass \( m^*/m \), (c) spectral function for different Hund’s \( J \) at constant Hubbard \( U = 5.0 \) eV at \( T = 116 \) K, using only density–density interactions. Reproduced with permission from from [261].

\[ \frac{\Delta}{\Delta_T} \sim -\frac{1}{2} \chi_0 \sim k_B T. \]

What controls the strength of quasi-particle incoherence in FeSi? In the many-body model of Fu and Doniach [333] and others, see section 2.3, the local Coulomb interaction—the Hubbard \( U \)—was identified as the driving force behind the metallization crossover. In the realistic calculations (that use \( U = 5 \) eV and \( J = 0.7 \) eV [43]), however, changes in \( U \) do not play a role.

\[ \Delta \sim k_B T. \]

41 Interestingly, extracted fluctuating moments and linewidths of FeSi are reminiscent of results for the metallic ferromagnetic Hsuier compound \( \text{Pd}_2\text{MnSn} \) well above its Curie temperature [341].

42 Details will be published elsewhere. Note that the same formalism has recently also been successfully applied to another intermediate valence compound, namely CePd$_3$ [343].

43 In rough accordance with constrained DFT results [347].
not yield substantial modifications of the spectral function. Indeed, the large bandwidth \((\gtrsim 10 \text{ eV})\) of the Fe-3d and Si-3s,3p conglomerate dominates the Hubbard \(U\) for reasonable values of the latter. However, a significant dependence of many-body-renormalizations on the Hund’s coupling \(J\) was evidenced \([261]\): As seen in figure 22(a) the scattering rate at \(T = 116 \text{ K}\) dramatically changes with \(J\). In fact, when going from the realistic value \(J = 0.7 \text{ eV}\) of iron compounds \([348]\) up to only \(J = 0.8 \text{ eV}\), \(\Im \Sigma(\omega = 0)\) more than doubles. Decreasing \(J\) to 0.6 eV, on the other hand, causes \(\Im \Sigma(\omega = 0)\) to almost vanish.

Concomitantly, also renormalizations encoded in the real-parts of the self-energy strongly depend on \(J\): the mass enhancement—in DMFT: \(m^*/m = 1/\Im\Sigma\)—increases notably with \(J\), see figure 22(b), while remaining at a moderate magnitude \((< 2)\)\(^{44}\). Combined, the effects of \(J\) in the self-energy cause largely different spectra: as seen in panel (c) of figure 22 both the width of the pseudogap at \(T = 116 \text{ K}\) as well as the amount of spectral weight at the Fermi level is drastically dependent on the Hund’s rule coupling.

What is the origin of this sensitivity? First, in FeSi there is no direct competition between the Hund’s coupling \(J \sim 0.7 \text{ eV}\) and crystal-field splittings \(\delta = O(0.3 \text{ eV})\). That \(J > \delta\) indicates that a high-spin state is always favoured. This clear hierarchy also rules out that a spin-state transition is at the origin of FeSi’s anomalous behaviour (see also below). Second, such a strong impact of \(J\) was previously found also for other iron-based materials \([350–352]\), adatoms \([353]\), as well as for some 4d-systems \([354]\) (see also the recent \([355, 356]\)). The general mechanism is investigated in \([351, 354, 354, 357–361]\): a large Hund’s \(J\) constrains electrons in different orbitals towards having the same spin, therewith ‘orbitally blocking’ \([351]\) the Kondo interaction. As a consequence the coherence temperature is shown to be suppressed exponentially \([357]\), resulting in a larger \(\Im \Sigma\) at a given temperature. For these systems the name Hund’s metal has been coined \([351]\). Besides the effective mass enhancement, also spin- and charge fluctuations in these Hund’s systems have genuine characteristics, as discussed for FeSi in the next section.

\(^{44}\)The mass enhancement is albeit larger than in other Fe–Si binary compounds, e.g. metallic \(\alpha\)-FeSi\(_2\) \([349]\).

**Figure 23.** FeSi: DMFT histograms. The bar-charts display the probability distribution of the system being in an eigenstate of the effective iron atom, resolved into states of occupation \(N\) and spin-state \(S\) for (a) \(T = 116 \text{ K}\) and (b) \(T = 1161 \text{ K}\). (a) \(T = 116 \text{ K}\); (b) \(T = 1161 \text{ K}\). \([253]\) 2013 © Springer Science+Business Media Dordrecht 2013. Adapted by permission from Springer.

**The spin and charge state.** Above, we have discussed the momentum-structure of the magnetic susceptibility, and explained the reasons why the local component is smaller than the measured uniform susceptibility. Besides the differentiation in momentum-space, the spin response also has an interesting structure in the time domain: the static local spin susceptibility, \(\chi_{\omega\omega}(\omega = 0) \sim \frac{1}{2} \int \text{d}r \langle S_z(\tau)S_z(0) \rangle\), displays, as discussed above, a strong and non-monotonous temperature dependence (see inset of figure 19 top left). \(\chi_{\omega\omega}(\omega = 0)\) describes the time-averaged response. Complementarily, we can compute the instantaneous response \(\lim_{\tau \rightarrow 0} \langle S_z(\tau)S_z(0) \rangle\) for \(\mu_{\text{eff}} = 2.6 \mu_B\) which is close to the single-ion limit in case of a \(S = 1\) configuration (see discussion in section 1.2).

We can analyse this in more detail: figure 23 shows the quantum Monte Carlo histogram \([362, 363]\) for the DFT+DMFT calculation at low (left) and high (right) temperature. The bars indicate the probability distribution of the many-body wave-function for being in an eigenstate of the effective iron atom, decomposed into the number of particles \(N\) and the spin state \(S\). We observe that:

(i) The decomposition onto \(N\) and \(S\) is almost independent of temperature. The scenario of a temperature-induced moment \([160]\) is hence not realized in our theory. Also, as anticipated above, a spin-state-transition does not occur. This situation has to be contrasted with systems in which the crystal-field splitting and the Hund’s rule coupling are of comparable size \([363]\). For example, spin-state crossovers have been identified to underlie the pressure-induced (low-spin)/insulator-to-(high-spin)/metal transition in the B8 phase of) MnO \([364]\), or the anomalous behaviour in the magnetic susceptibility of LaCoO\(_3\) \([263, 365–368]\) (see figure 2 or \([56, 58]\)). Indeed, in LaCoO\(_3\) x-ray absorption spectroscopy estimates a crystal-field splitting between (at low-\(T\) fully occupied) \(t_{2g}\) and (empty) \(e_g\) orbitals of \(\delta \approx 0.7 \text{ eV}\). This value corresponds to
typical estimates of the Hund’s $J$ in transition-metal based compounds. In LaCoO$_3$, the competition between $\delta$ and $J$ is tipped by a thermal activation of the high-spin state (in cooperation with the expansion of the lattice). As shown by Krápek et al [263] this leads to a pronounced temperature-dependence of the DMFT histogram for both charge- and spin-resolved probabilities—in qualitative difference to the case of FeSi. An important factor that distinguishes FeSi from LaCoO$_3$—the degree of ionic versus covalent bonding—will be discussed in a broader context in section 2.5.

(ii) The iron valence of $\langle N \rangle \approx 6.2$ is consistent with the covalent band-picture of section 2.2.1 (see also figure 25).

(iii) The spin state $S = 1$ has the overall largest probability at all temperatures, consistent with the above discussed effective moment.

(iv) The probability distribution is not dominated by only few configurations of $N$ and $S$. This indicates strong fluctuations at short time-scales. In fact, the system has a strongly intermediate valence, as the charge variance is of order $1 : \delta N = (\langle N \rangle - \langle N \rangle)^2 \approx 0.93$. However, as the variance is insensitive to temperature, athermally induced mixed valence [242] can be excluded. Also spin fluctuations at short time-scales, $\delta S = (\langle S \rangle - \langle S \rangle)^2 \approx 0.33$, are large.

The above analysis shows that the physics of correlations in FeSi is very different from the Mott–Hubbard physics in, say, transition metal oxides [67]. There, the Hubbard $U$ quenches charge fluctuations and reduces the time-scale associated with them, while the time averaged local spin susceptibility is strongly enhanced. In fact, in a Mott insulator, charge fluctuations are basically absent, while the local moment of charge carriers fluctuates freely, giving rise to a Curie–Weiss susceptibility. In FeSi, on the contrary, spin fluctuations are short-lived: The (time-averaged) local spin susceptibility is only moderately enhanced, while the short-time (energy-averaged) fluctuating magnetic moment is strongly enhanced. Contrary to Mott insulators, fluctuating magnetic moments in systems like FeSi do not require the quenching of charge fluctuations. This is indeed a hallmark of Hund’s metals (see the preceding section). Still, FeSi is an example where the Hund’s metal concept exists for an only poorly metallic state, in which the hybridization of the iron atoms with their surrounding is weak at low energies (see also the discussion of the DMFT hybridization function in section 2.4.3).

The physical picture. With the above details of the realistic many-body simulations, one can address the fundamental picture of FeSi: (a) what is the microscopic origin of the crossover in the magnetic susceptibility? (b) How is the latter linked to the emergence of incoherent spectral weight. In other words, what is the relation between the crossovers for the spin and charge degrees of freedom? (c) What is the difference, if any, to the physics of Kondo insulators?

At low temperatures, FeSi is a band-like semiconductor: The band gap in the one-particle spectrum inhibits, both, spin and charge fluctuations at finite time-scales. This results in (i) a small, activated (uniform) spin susceptibility, and (ii) a lack of phase-space available for electronic scattering, resulting in only a small imaginary part of the self-energy. As such, the system can be approximately described by an effective one-particle Hamiltonian that is diagonalizable in momentum space, stating that $k$ is a good quantum number.

With increasing temperature (beyond $T \approx T_{\text{max}}^k$) a fluctuating moment develops at the iron sites, which is a manifestation of the single-ion physics that dominates at low temperatures when spatial correlations become short-ranged [369]. In this limit, the index conjugate to the momentum, the lattice site $i$, becomes a good quantum number. The fluctuating moment thus provides a link to the real-space lattice, and the description in terms of Bloch-bands breaks down as $k$ is no longer a good quantum number. As a consequence, the one-particle spectrum decomposes over many momenta, or, equivalently acquires a finite lifetime for a given momentum. This lifetime is encoded in the imaginary parts of the self-energy and leads to the appearance of incoherent spectral weight that eventually fills the gap. We stress that despite the notion of locality of the moments on the iron sites, magnetic fluctuations are ferromagnetic (see figure 20) due to the strong on-site Hund’s rule coupling.

This scenario qualitatively reconciles the behaviour seen for spectral properties in the pioneering model of Fu and Doniach [333] with the seminal spin-fluctuation theory of Takahashi and Moriya [244]: in the DFT+DMFT picture, the crossovers in the spin and charge response in FeSi are intimately linked to each other. In the microscopic details, however, the realistic many-body calculations [253, 261] revealed notable differences to previous model calculations [254, 260, 333]. It is likely that the scenario for FeSi reviewed here is also relevant for other narrow-gap intermetallics. For example, it might be expected that—once issues related to exchange contributions to the self-energy (see section 2.2.2)—are overcome—the microscopic behaviour of FeSb$_2$ is qualitatively akin to that of FeSi.

The above discussions answers the first two questions, (a) and (b), raised at the beginning of this section. In order to answer—in section 2.5—the third question, (c), we performed realistic many-body calculations for Ce$_3$Bi$_4$Pt$_3$ that are discussed below.

2.4.3. Realistic many-body calculations for Ce$_3$Bi$_4$Pt$_3$. For a direct comparison of FeSi to heavy-fermion-based Kondo insulators, we performed DFT+DMFT calculations for the prototypical compound Ce$_3$Bi$_4$Pt$_3$. Figure 24 displays the spectral function obtained for different temperatures. As was the case in FeSi, lowering temperature results in a depletion of spectral weight at the Fermi level. The width of the emerging gap, $\Delta \lesssim 10$ meV, is consistent with the experimental indirect charge gap of 8–12 meV (see figures 4(a) 5(a) and table 1). With respect to band-structure results (see section 2.2.3) the renormalization of the gap corresponds to an effective mass-enhancement of about a factor of 10. Also the temperature scale, $\lesssim 100$ K, around which the metallization occurs in the spectral function is congruent with both photoemission [139] and optical spectra [60] (see figure 5).
coherence–incoherence crossover is encoded in the imaginary parts of the self-energy (not shown). As was the case for FeSi (see figure 21), $\Sigma(\omega = 0) \propto T^2$ for orbitals relevant for spectral weight near the Fermi level. These changes in spectral properties are accompanied by a rearrangement of the spin degrees of freedom. Indeed, an analysis of the DMFT magnetic susceptibility (not shown) is compatible with the standard picture of the unlocking of local moments above a coherence temperature $T^\text{max}_\chi$. The next section will put these results into a broader context by discussing differences between FeSi, Ce$_2$Bi$_4$Pt$_3$, and LaCoO$_3$.

2.5. Kondo insulators, transition-metal intermetallics and oxides: a categorization

2.5.1. The DMFT perspective on insulators. In order to illustrate qualitative differences in narrow-gap semiconductors with partially-filled correlated orbitals, it is convenient to take a look under the hood of the DMFT method [302, 303]. There, the periodic solid is mapped onto an effective (e.g. Brinkmann-Rice [373]) scenario of a Mott insulator: spectral weight is destroyed (Z → 0) by diverging effective masses, as realized e.g. in YTiO$_3$ [374], V$_2$O$_3$ [345, 375], or thin films of SrVO$_3$ [4].

This is the case of Ce$_3$Bi$_4$Pt$_3$, see below. By taking a look under the hood of the DMFT method [302, 303], we can distinguish46:

| Type / property | $H(k)$ | $-\Delta \Sigma$ | $\Delta H^\text{loc}$ | $-\Delta \Delta$ | Examples |
|-----------------|--------|-----------------|--------------------------|---------------|-----------|
| Band insulator  | Gapped | 0               | Large                    | 0             | LaCoO$_3$, FeSi |
| Kondo insulator | Gapped | 0               | Small $\propto$           | $\Sigma_{\Delta}$ | Ce$_2$Bi$_4$Pt$_3$, FeSi |
| Mott insulator  | Metallic | $\propto$          | Small 0                  |                | YTiO$_3$, thin films |
|                |        |                 |                          |               | SrVO$_3$ |

Here, $\mu$ is the chemical potential, $H(k)$ the one-particle Hamiltonian, and $H^\text{dd}_{\Delta}$ its local ($k$-summed) component in the correlated $d$-subspace that encodes local crystal-fields. $\Sigma$ is the lattice self-energy which is approximated as the impurity self-energy $\Sigma_{\Delta}$ in the $d$-subspace, $\Sigma_{dd} = \Sigma_{\Delta}$ and is zero otherwise. Finally, $\Delta(\omega)$ is the so-called hybridization function that describes the coupling of the effective correlated atom to the bath.

Using (7), we can characterize the origin of an insulating state, i.e. a vanishing of $\Sigma G^\text{loc}(\omega = 0)$, through its manifestations in $H(k)$, $\Sigma(\omega)$, $H^\text{loc}_{\Delta}$, and $\Delta(\omega)$. In this DMFT perspective, we can distinguish46:

Three prototypes of paramagnetic insulators (see also table 4).

1) **Mott insulators**: $\Sigma$ diverges at low energies (at least for some orbitals), while $H(k)$ is metallic. This is the Brinkmann-Rice [373] scenario of a Mott insulator: spectral weight is destroyed ($Z \rightarrow 0$) by diverging effective masses, as realized e.g. in YTiO$_3$ [374], V$_2$O$_3$ [345, 375], or thin films of SrVO$_3$ [4].

2) **Ionic band insulators**: here, both $\Sigma_{\Delta}$ and $\Delta_\Delta$ vanish at the Fermi level at low $T$, while $H^\text{loc}_{\Delta} + \Delta_\Sigma(0) + \Delta_\Sigma(0) - \mu$ is finite for all orbitals. In other words, it is local crystal-fields encoded in $H^\text{loc}_{\Delta}$ (possibly enhanced by static renormalizations $\Delta_\Sigma(0)$) that are responsible for the gapping. This is the case of LaCoO$_3$ where the chemical potential falls between the (at low temperatures) filled $t_{2g}$ and the empty $e_g$ orbitals (see section 2.2.1 and below). Also the dimerized M1 phase of VO$_2$, a correlation-enhanced Peierls insulator [376], will fall into this category.

3) **Kondo insulators (covalent band insulators)**: in these systems $\Delta$ diverges at low energies to suppress spectral weight, while $\Sigma$ is benign and crystal-fields in $H^\text{loc}_{\Delta}$ are mostly irrelevant for realizing the insulator. This is the case of Ce$_2$Bi$_4$Pt$_3$, see below.

We employed DFT+DMFT with a CTQMC solver as implemented in [338] with rotationally invariant interactions parametrized by $U = 5.5$ eV and $J = 0.68$ eV (similar to what was used previously for elemental Ce [338, 370–372] and other Ce-based compounds [6, 338, 343]), spin-orbit coupling, charge self-consistency, and nominal double-counting. Details will be published elsewhere.

Note that, here, we limit the discussion to paramagnetic insulators without spin, charge, or structural orders.
Figure 25. Ionic versus covalent bonding. Shown is a collection of materials distinguished by the degree of covalent bonding for their partially-filled correlated orbitals (3/4d or 4f). As a rough measure, we use the relative deviation of the DFT-occupation of these orbitals, $n_{d}^{\text{DFT}}$, from the expectation in the purely ionic picture, $n_{d}^{\text{ionic}}$. For some materials different ionic states are indicated. The gapping of correlated electrons in materials on the right (left) are dominated by hybridizations $\Delta(\omega)$ (local crystal-fields $H_{\text{loc}}^{\text{cr}}$), respectively. Using this proxy, Kondo-insulating characteristics prevail in compounds on the right and become less pronounced towards the left. DFT occupations were determined from a character projection inside the muffin-tin spheres (of typical radii) using wien2k. The dependence on the exchange-correlation potential (LDA, PBE, mBJ) is negligible for all practical purposes. The y-axis shows the experimental (indirect) charge gap (see table 1).

Since at least the last two scenarios do not exclude each other, it may be difficult to provide a clean-cut categorization for a given material. Before looking at ab initio results for FeSi and Ce$_3$Bi$_4$Pt$_3$, let us thus illustrate some characteristics of Kondo insulators, as well as the grey area between purely covalent and ionic insulators first for a toy model, then from the DFT perspective.

2.5.2. A toy-model. Let us first focus on the left side of (7): From the lattice point of view the difference between ionic and covalent orbitals is hidden in $H(k)$. This is seen in a simple toy model in the quasi-particle picture. Consider:

$$H(k) = \begin{bmatrix} \epsilon_{k} & V \\ V^{*} & \epsilon_{k}^{p} \end{bmatrix}, \quad \Sigma(\omega) = \begin{bmatrix} \Sigma(\omega) & 0 \\ 0 & 0 \end{bmatrix}$$

i.e. one correlated band $\epsilon_{k}$ coupled to a ligand state $\epsilon_{k}^{p}$ via a hybridization $V$, that for simplicity we have assumed to be local. Many-body renormalizations are described by a self-energy, for which we take a Fermi-liquid-like form $\Sigma(\omega) = (1 - 1/Z)\omega - i0^{+}$ that neglects lifetime effects.

The first goal is to illustrate generic behaviours of the hybridization function for Kondo insulators. To this end, we consider a setup in the spirit of the PAM to mimic Ce$_3$Bi$_4$Pt$_3$: $\epsilon_{k} = 0$ is an atomic-like f-level, renormalized by Z, and $\epsilon_{k}^{p} = \epsilon_{k}$ are the non-interacting conduction electrons of atoms other than Ce. Assuming particle-hole symmetry yields $\mu = 0$ and $\sum_{k} \epsilon_{k} = 0$, implying the absence of any crystal-fields. The bare one-particle spectrum is then gapped, with the direct gap determined by $2V$. The interacting dispersions are given by $\omega_{k} = \epsilon_{k}/2 \pm \sqrt{\epsilon_{k}^{2} / 4 + ZV^{2}}$, stating that only the inter-atomic hybridization (hence also the direct gap) is renormalized by electron-electron interactions.

Equivalently, this can be seen in the hybridization function $\Delta(\omega) = \omega - (G_{\text{loc,dd}}(\omega))^{-1} \Sigma(\omega)$. For the left side of (7) we have $G_{\text{loc,dd}} = \sum_{k} Z(\omega - Z\epsilon_{k} + i0^{+})$ with a momentum-dependent quantity $\Delta_{k}(\omega) = \frac{|V|^{2}}{\omega - \epsilon_{k} + i0^{+}}$. $\Delta_{k}(\omega)$ in particular diverges at the conduction-electron positions. In order to match the right-hand side of (7), $G_{\text{DFT}}^{\text{loc}}(\omega) = Z(\omega - Z\Delta(\omega) + i0^{+})$, the DMFT hybridization function $\Delta(\omega)$ has to also diverge, here—because of particle-hole symmetry—at the Fermi level. Consequently, the divergence of $\Delta$ is a pure one-particle effect. Instead of being a signature of the Kondo effect, $\Delta$ here merely encodes inter-atomic hybridizations that couple the effective correlated site to the DMFT bath. From this band-structure perspective, the influence of many-body effects is to renormalize down the strength of hybridization $\Delta \rightarrow Z\Delta$. Further, effects of incoherence (neglected here) will, when iterating to self-consistency, mitigate the divergence above a coherence temperature (see the ab initio results in section 2.5.4).

The model can also shed some insights onto the differences of FeSi and LaCoO$_3$. We model both compounds by two decoupled copies of the above model: the first copy, $\Delta_{\text{loc}}$, and the second copy, $\Delta_{\text{loc}}^{p}$. In the first copy, we then obtain for the left side of (7) $G_{\text{loc,dd}}^{\text{loc}}(\omega) = \sum_{k} Z(\omega - Z\epsilon_{k} + i0^{+}) - Z\Delta_{k}(\omega) + i0^{+}$ with $\Delta_{k}(\omega) = \frac{|V|^{2}}{\omega - \epsilon_{k} + i0^{+}}$. Now, there are two degrees of freedom that can cause insulating behaviour: For small crystal-fields, $\delta$, insulating behaviour requires the same mechanism as in the above model for Ce$_3$Bi$_4$Pt$_3$: a pole-like structure in $\Delta_{k}$ (and, correspondingly, in the DMFT hybridization function $\Delta$). Indeed, for $\delta = 0$ the band-gap is proportional to $V$ (at least for $Z = 1$). Empirically this situation corresponds to FeSi within DFT, as illustrated by figure 11(a). In the opposite limit, i.e. a vanishing hybridization $V = 0$, insulating behaviour requires that the crystal-field $\delta/2$ exceeds the half-bandwidth of $\epsilon_{k}$. This is the prevailing mechanism in LaCoO$_3$, as shown within DFT in figure 11(b). From this point of view, one can call LaCoO$_3$ a mainly ionic insulator, and FeSi a dominantly covalent, i.e. Kondo-like insulator. Naturally, there exists a continuous spectrum between the pure prototypes of Kondo-like and (ionic) band insulators. Also, the hybridization function of an ionic system can still be strongly peaked outside the charge-gap, encoding charge-fluctuations at higher energies.

2.5.3. A DFT-based proxy to distinguish ionic from covalent insulators. As motivated by the preceding model as well as section 2.2, interesting insights can be gained from one-particle band-structures even in the case of Kondo systems. Indeed, DFT-based hybridization functions $\Delta(\omega)$ can divulge trends in gap-sizes or Kondo temperatures among families of compounds. For a recent discussion of band-structure-derived hybridization functions for Ce-compounds, see [377].
Here, we are interested in determining—without performing costly many-body calculations—whether an insulating material is rather a Kondo or a (ionic) band insulator. For paramagnetic insulators with partially filled $d$/$f$-shells we propose—as a crude indicator for Kondo insulating behaviour—the relative deviation of DFT-occupancies with respect to nominal expectations in a purely ionic picture: $\left( n_{\text{DFT}}^d - n_{\text{ionic}}^d \right) / n_{\text{DFT}}^d$. This proxy can hence be seen as a measure for charges of correlated orbitals to participate in covalent bonding and could, as a matter of fact, serve as a descriptor in high-throughput studies. In figure 25 we have collected a number of relevant compounds. The Kondo insulator Ce$_3$Bi$_4$Pt$_3$ appears on the right-most side, as it does not donate all of its Ce-4$f^1$ electron to more electronegative atoms. The oxide LaCoO$_3$, on the other hand, is found much towards the left. Indeed the Co-valence within DFT is rather close to that of a Co$^{3+}$ ion. The intermetallic semiconductors FeSi and FeSb$_2$ rank in the middle of the diagram, pointing to the influence of, both, covalent and ionic characteristics.

2.5.4. Comparison of the ab initio results of FeSi and Ce$_3$Bi$_4$Pt$_3$. We are now in the position to analyse and compare the realistic many-body simulations for FeSi and Ce$_3$Bi$_4$Pt$_3$: figures 26(a) and (b) display the respective DFT+DMFT spectral functions, $A(\omega)$, at low temperatures, resolved into relevant orbital contributions. Panels (c) and (d) show how the imaginary parts of the corresponding hybridization functions, $\Delta(\omega)$, evolve with temperature.

In the local spectral function, we see a pronounced difference between Ce$_3$Bi$_4$Pt$_3$ (left) and FeSi (right): in the heavy-fermion material, there is no orbital polarization. Indeed, both, the valence and the conduction side of the low-temperature gap are dominated by the correlated 4$f$-orbitals. Furthermore, the $J = 5/2$ multiplets are basically degenerate in what concerns the gap-formation. In particular, all of them have their centre of mass above the Fermi level. The gapping thus occurs similarly for all $J = 5/2$ multiplets.

From the DMFT perspective, this situation is described by the development of a strong peak in the hybridization function for the $J = 5/2$ components. In the many-body language, this is the building up of the Kondo effect. It restores band-like coherence below a characteristic temperature through the formation of singlet states between 4$f$-electrons of the effective Ce-atom and conduction electrons from its surrounding. Seen from the low temperature (band-structure) perspective, the large $\Delta(\omega)$ is merely a way to encode—within the single-site impurity construction—the effect of inter-atomic (i.e. out-of-impurity) hybridizations, that, with growing temperature, are rendered ineffective through the emergence of local spin-fluctuations. Note that, unlike for the simple model in section 2.5.2, the enhancement of $\Delta(\omega)$ occurs at the lower edge of the charge gap since Ce$_3$Bi$_4$Pt$_3$ is strongly particle-hole asymmetric.

Contrary to Ce$_3$Bi$_4$Pt$_3$, the orbital character of the valence and conduction states are not the same in FeSi. Indeed, the
occupied spectrum is dominated by the Fe-3$d_{x^2-y^2}$ + 3$d_{xy}$ orbitals, while unoccupied states derive mainly from Fe-3$d_{xz}$ + 3$d_{yz}$ weight. Still, all orbital components are individually gapped. We can analyse the origin of for the different orbitals from the DMFT perspective: We see that insulating behaviour of the 3$d_{xz}$ + 3$d_{yz}$ components is realized by a hybridization function that increasingly vanishes inside the gap: there is (averaged over time) no spontaneous exchange of particles of these orbitals with the surrounding of the effective iron atom. Feature in $\Delta(\omega)$ that mark charge fluctuations only appear for energies well above gap. In the categorization of section 2.5.1, these orbitals are hence trivially band-insulating.

The situation for the 3$d_{x^2-y^2}$, 3$d_{xy}$ components, however, bear all the marks of a Kondo-insulator, as seen in Ce$_3$Bi$_4$Pt$_3$: These orbitals’ spectral weight accumulates to the most part of one side of the Fermi-level, while on the other side a sharp low-energy feature develops in $\Delta(\omega)$ that suppresses states and opens a gap. In conclusion, we can call FeSi an orbital-selective Kondo insulator.

In the light of this discussion, there are three main differences between FeSi and Ce$_3$Bi$_4$Pt$_3$: (i) the low-temperature gap in Ce$_3$Bi$_4$Pt$_3$ is purely of Kondo-insulating nature, whereas in FeSi the origin of the gapping depends on the orbital character. (ii) For orbitals in FeSi that suppress quantum fluctuations through the Kondo effect, the relevant non-local hybridization dominantly links these orbitals to other iron atoms. In Ce$_3$Bi$_4$Pt$_3$, the gap-inducing hybridization connects the ‘impurity’ Ce-4$f$-5/2 orbitals to uncorrelated, dispersive states of Bi and Pt. (iii) In the case of FeSi, the origin of the destruction of asymptotic freedom with rising temperature has multi-orbital (Hund’s rule) characteristics.

2.6. Covalent insulators versus Kondo insulators: Hubbard versus Anderson

The above discussed ab initio many-body results for prototype materials of correlated narrow-gap semiconductors and Kondo insulators allowed for a quantitative description of

Figure 27. Non-interacting electronic structure of the models (9) and (10). Shown are (a) the dispersions and (b) the DOS of the two-band Hubbard model (right), (9), and the PAM (left), (10), in the non-interacting limit ($U = 0$). Hubbard model: $t = V = 0.25$ eV ($\eta = 1$), $\Delta_{0\text{dir}} = 50$ meV. PAM: $t = 0.25$ eV, $V = 0.3$ eV. The Fermi level corresponds to zero energy. (a) Non-interacting band-structure. Dispersion along the diagonal ($k, k, k$)-direction. Insets show a zoom to a low-energy window $[-150 : 150]$ meV. (b) DOS.
We want to answer the question why, \( f - U = V \eta_2 = \cos(\gamma) = \cos(\beta) \) on the Hubbard \( \epsilon \langle \rangle \) related compounds such as FeAs \( 2 \) or RuSb \( 2 \) and RuSi display temperatures. In section 2.2 we already surmised why closely related compounds such as FeAs \( 2 \) or RuSb \( 2 \) and RuSi display less or no signatures of correlation effects. Here, we will put this analysis on a firmer footing, by studying the tendencies in both the crossover temperature in the susceptibility \( T_\chi^{\text{max}} \). We want to answer the question why, among correlated intermetallic semiconductors only FeSi, FeSb \( 2 \), and CrSb \( 2 \) display a crossover from activated behaviour to a Curie–Weiss-like decay at experimentally accessible temperatures. In section 2.2 we already surmised why closely related compounds such as FeAs \( 2 \) or RuSb \( 2 \) and RuSi display less or no signatures of correlation effects. Here, we will put this analysis on a firmer footing, by studying the tendencies in both the crossover temperature \( T_\chi^{\text{max}} \) and the peak value \( \chi(T_\chi^{\text{max}}) \) of the susceptibility for varying parameters. The models we consider are (i) a variant of a two-band Hubbard model, and (ii) the PAM.

2.6.1. The model Hamiltonians.

(i) The two-band Hubbard model. We employ the Hamiltonian

\[
H = \sum_{k, \sigma} (\epsilon_k + \delta/2) (c_{k,1}^{\dagger} c_{k,1,\sigma} - c_{k,2}^{\dagger} c_{k,2,\sigma}) \\
+ \sum_{k, \sigma} (V_{k} c_{k,2,\sigma}^{\dagger} c_{k,1,\sigma} + V_{k} c_{k,1,\sigma}^{\dagger} c_{k,2,\sigma}) \\
+ \sum_{iL \sigma} U_{iL, \sigma} n_{iL, \sigma} n_{iL', -\sigma} + \sum_{iL \sigma} (U_{iL, \sigma} - J) n_{iL, \sigma} n_{iL, \sigma} 
\]  

at half-filling, using a cubic lattice with nearest neighbour hopping, \( \epsilon_k = -2t \sum_{i=1}^{3} \cos(k_i a) \), a local crystal-field \( \delta \), and a purely non-local inter-band hybridization \( V_k = \eta_0 \epsilon_k \). The interaction is limited to density-density terms, and we use the cubic parametrization \( U_{LL'} = U - 2J(1 - \delta_{LL'}) \) with the Hubbard \( U \) and the Hund’s rule coupling \( J \). In the non-interacting limit \( (U = 0) \) the system has a direct charge gap \( \Delta^0 = \delta \frac{\eta}{\sqrt{1 + \eta^2}} \); see figures 27(a) and (b), right) for the dispersion and the density of states, respectively. We note that the orbital characters of the valence and conduction states at a given momentum are different, as was found for FeSi and FeSb \( 2 \) within band-theory, see section 2.2.47.

(ii) The periodic Anderson model. We study the PAM, (5), in its particle-hole symmetric \( (\epsilon_f = 0) \) variant with local hybridization, \( V_k = V \):

\[
H = \sum_{k, \sigma} \epsilon_k c_{k, \sigma}^{\dagger} c_{k, \sigma} + V \sum_{k, \sigma} (c_{k, \sigma}^{\dagger} f_{k, \sigma} + f_{k, \sigma}^{\dagger} c_{k, \sigma}) + U \sum_{i} n_{i, \sigma}^{f} n_{i, \bar{\sigma}}^{f} 
\]  

(10)
on the 3d cubic lattice, \( \epsilon_k = -2t \sum_{i=1}^{3} \cos(k_i a) \), at half-filling. The non-interacting density of states states are displayed in figures 27(a) and (b), left), for a choice of parameters as indicated in the caption. While the bare direct gap is given by \( \Delta^0_{\text{dir}} = 2V \); the bare indirect gap is much smaller, \( \Delta^0_{\text{indir}} = \sqrt{D^2 - 4V^2 - D} \), with \( D = 6t \) the half-bandwidth of the (unhybridized) c-electrons.

Note that the orbitally resolved DOS of the above defined Hubbard (periodic Anderson) model neatly captures the

47 See, e.g. [378, 379] for related two-band Hubbard models that explore topological effects via the inclusion of effective spin–orbit couplings.

48 For the interesting physics of incoherence in the asymmetric PAM see e.g. [380].
2.7. Electron-lattice effects

The realistic many-body calculations reviewed in the preceding sections were based on the Born–Oppenheimer approximation, i.e. there was no interplay between electronic and phononic degrees of freedom. Computations were even done for a fixed, temperature-independent lattice constant. Given the sensitivity of correlated materials to minute changes in conditions, accounting for the volume expansion in FeSi [274] might have noticeable results. In fact, its omission was blamed for slight discrepancies between the experimental and theoretical optical conductivity (see section 2.4.1). This section is devoted to such electron-lattice effects: we will review experimental evidence for large electron–phonon couplings in FeSi and FeSb2, discuss proposals for the consequences thereof for the electronic and phononic subsystems, as well as parse related ab initio calculations. The influence of phonons onto the thermopower via the phonon-drag effect will be discussed

 qualitative aspects of the gap in FeSi (Ce3Bi4Pt3). As in Ce3Bi4Pt3, both, valence and conduction states in the PAM are dominated by the f-contribution. In the two-band Hubbard model, valence and conduction bands are instead of different orbital character, akin to FeSi—compare figures 27(a) to figures 26(a) and (b).

2.6.2. Dynamical mean-field theory results. We solve both models within DMFT, using the *w*2dynamics package [381]. Here, we do not allow for any long-range order, and leave the study of ordered phases for future work. As a proxy for the experimental crossover in the spin response, we consider the maximum, $T_{\chi}^{\text{max}}$, in the local (DMFT) susceptibility $\chi^\text{loc}(0) = \frac{g^2}{\beta} \int_0^\beta \frac{d \tau}{\tau} \Sigma_L(\tau) \Sigma_L(0)$ [49]. We limit the sum over indices $L, L'$ to correlated orbitals and use $g = 2$.

Figure 28 displays (a) $T_{\chi}^{\text{max}}$ and (b) $\chi(T_{\chi}^{\text{max}})$ as a function of the Hubbard $U$ interaction for the PAM (left) and the Hubbard model (right), for different values of the bare gap $\Delta^0$. For both models, $T_{\chi}^{\text{max}}$ quickly becomes inaccessible high to experiments as the gap size grows. This rationalizes, why enhanced susceptibilities cannot be seen in experiments for, e.g. FeAs2 ($\Delta_{\text{FeAs}} \sim 200 \text{ meV} \gg \Delta_{\text{FeSb2}} \sim 30 \text{ meV}$).

As to the influence of an increased Hubbard $U$ onto the crossover temperature, it grows slightly for the Hubbard model, while it decreases strongly for the PAM. In the Hubbard model context—the dependence of the magnetic response on $J$. As displayed in figure 29, both the crossover temperature $T_{\chi}^{\text{max}}$ and the peak value of the (local) susceptibility strongly depend on the Hund’s coupling. Indeed the sensitivity on $J$ is notably more pronounced than the dependence on $U$ (see above). The role of the Hund’s coupling is to strongly enhance the susceptibility and to push the Curie–Weiss–like decay to much lower temperatures. This tendency explains that while strong signatures of correlation effects are visible in the 3d-compound FeSi, the 4d-compounds RuSi and RuSb2 with their much smaller Hund’s rule $J$ (see table 3), exhibit neither enhanced paramagnetism, nor Curie–Weiss–like behaviour.

Interestingly, in the Hubbard model, the magnitude (but not the crossover temperature) depends strongly on the strength of the hybridization: for a constant gap $\Delta^0 = \delta \cdot \eta / \sqrt{1 + \eta^2}$, $\chi(T_{\chi}^{\text{max}})$ is larger for smaller $\eta$ (and thus larger $\delta$), i.e. for weaker (and less dispersive) hybridizations. This is congruent with the observation that $\chi^\text{loc}$ is much larger in the case of the PAM than for the two-band Hubbard model.

In all, strong signatures of correlation effects in the magnetic response of narrow-gap semiconductors is favoured by a small gap, weak hybridizations, and a large Hund’s rule coupling. These prerequisites may explain why the discussed anomalous paramagnetic response is relatively rare. Indeed, contrary to, e.g. ubiquitous mass enhancements, the occurrence of Curie–Weiss-like susceptibilities in 3d-electron based semiconductors has so far only been established for FeSi, FeSb2, and CrSb2.
2.7.1. Thermal conductivity. Let us briefly discuss the thermal conductivity $\kappa$ of FeSi [17, 132, 384, 385] and FeSb$_2$ [23, 24, 386, 387], both of which—according to a Wiedemann–Franz analysis—are dominated by the lattice contribution for all measured temperatures ($T < 300$ K). Interestingly, *ab initio* lattice dynamics calculations of the stoichiometric materials largely overestimate the thermal conductivity for, both, FeSb$_2$ [388] and FeSi, while simulations for semimetallic CoSi [389] compare well to experiments [389, 390]. Even more striking is the observation that in FeSi the thermal conductivity is larger in polycrystalline samples than for single crystals by a factor of two below 100 K. This is rather counter-intuitive, considering that boundary scattering is expected to suppress phonon-driven heat-conduction. Finding smaller carrier densities in polycrystalline samples, Sales et al [17] suggested that strong electron–phonon scattering is at the origin of the suppression of thermal conduction in single crystals. In the case of FeSb$_2$, the behaviour is more conventional with thermal conduction significantly increasing for larger granularities [391, 392] or sample sizes [27], see also figure 42.

2.7.2. Phonon structure and electron–phonon coupling. Some lattice properties such as the bulk modulus have already been discussed in conjunction with failures of band-theory for FeSi in section 2.2.1. In that compound, unusual behaviour is also seen in other static lattice properties, such as elastic constants that exhibit an anomalous softening above $\sim 100$ K [394].

We now turn to dynamical lattice properties. Individual phonon modes can be studied by Raman and infrared spectroscopy. Results for FeSi [16, 69, 382, 395–397] and FeSb$_2$ [382] indicate a notable softening of phonon modes and a strongly increasing phonon-width (shown in figure 30) upon heating above $\sim 100$ K. These effects were interpreted as arising—via the electron–phonon coupling—from the metalization of FeSi above $\sim 100$ K. This crossover provides for electron–phonon scattering events—absent in an insulator—and reduce the lifetime of phonons. An analysis of the phonon lineshapes in Raman spectra by Racu et al [396] indeed pointed electron–phonon and not phonon–phonon interactions to be at the origin of the temperature-induced broadening.

The phonon density of states can be accessed from inelastic neutron scattering (INS) and nuclear resonant inelastic scattering (NRIS) experiments. In a seminal work [393], Delaire et al presented a comprehensive INS study for FeSi in comparison to CoSi. They found (see figure 31) that the phonon softening in the metallic CoSi, which is of moderate 4% (in the range of 10–750 K), can be fully accounted for by the temperature-induced volume expansion. Indeed, using experimental thermal expansion data, the experimental softening is congruent with computations in the quasi-harmonic approximation, and more sophisticated *ab initio* molecular dynamics simulations give similar results. In the case of FeSi, an inspection of the thermal expansion would suggest a phonon softening of a magnitude, 5%, comparable to CoSi. Instead, Delaire et al [393] found in FeSi a softening three times as large: 14%. At low temperatures, the phonon dispersions of FeSi are in reasonable agreement with DFT calculations [17, 393]. Above $\sim 100$ K, molecular dynamics simulations still give a good description, while the quasi-harmonic approximation now fails to account for the observed softening of phononic dispersions, strongly suggesting a mechanism linked to the electronic degrees of freedom.

Parshin et al [398] used the NRIS technique and resolved the fine structure of the partial phonon DOS associated with the vibration of iron atoms. Moreover, they were able to extract phonon linewidths: congruent with the Raman and optical spectroscopy, these increase notably above 150 K. The empirical connection between phonon energies and temperature-induced changes in the electronic dispersion has been illustrated by Delaire et al, see figure 32:
If there is no notable rearrangement of the electronic structure, the thermal activation of phonons provides scattering. Indeed, via the electron–phonon coupling, lattice vibrations may conversely modify these scenarios suggesting a large influence of the electronic structure. Thus, modifications of modes is governed by thermal expansion of the lattice and is well described by the quasi-harmonic (QH) approximation. The increased coupling between electrons and phonons induces an anomalous softening, conversely, if the electronic density at the Fermi level decreases with temperature, e.g. as a consequence of a stiffening of the electronic structure are shown: (A) no notable changes between low and high temperatures, (B) crossover from a narrow-gap semiconductor at low temperature to a metallic state at high temperature, (C) a crossover from a coherent metal to an incoherent metal with rising temperature. On the right are shown the expected implications for the phonon dispersions in the presence of electron–phonon coupling: (A) the softening of modes is governed by thermal expansion of the lattice and is well described by the quasi-harmonic (QH) approximation. (B) the decreased coupling between electrons and phonons leads to a hardening of modes. Reproduced with permission from [393].

(1) If there is no notable rearrangement of the electronic density of states at the Fermi level, phonon energies are modified only through the thermal volume expansion, and the quasi-harmonic (QH) approximation captures the associated mode softening.

(2) If the system evolves from a gapped electronic spectrum at low temperature to a state with finite weight at the Fermi level at high T, an anomalous softening beyond the QH approximation occurs, in conjunction with a broadening of phonon linewidths. The former is caused by electronic screening of inter-atomic potentials, while the latter originates from electron–phonon scattering. This behaviour is seen in FeSi (see above) and FeSb2 [399], as well as Nb3Sn where the low-temperature gap is owing to superconductivity [400]. Similar effects are also seen when metallizing FeSi through doping [401].

(3) Conversely, if the electronic density at the Fermi level decreases with temperature, e.g. as a consequence of a (heavy) Fermi liquid losing coherence, a stiffening of phonons is expected as force constants get unscreened. This phenomenon was indeed observed in V-based A15 compounds [402], as well as bcc V-based alloys [403].

2.7. Inhibition of vibrations onto the electronic structure. While these scenarios suggest a large influence of the electronic structure onto the phonons, lattice vibrations may conversely modify the electronic structure. Indeed, via the electron–phonon coupling, the thermal activation of phonons provides scattering channels that may decrease the lifetime of electronic excitations. Given that FeSi metallizes through the occurrence of incoherent spectral weight at the Fermi level, lattice vibrations can at least accelerate the crossover to the metallic state.

Looking at the literature of FeSi, the interplay of electrons and phonons could be interpreted as a chicken-and-egg question: does the phonon system merely react to the (correlation-induced) evolution of the electronic and magnetic degrees of freedom, or are lattice vibrations themselves a major driver of the metallization? Both standpoints have been forwarded on the basis of experimental and theoretical studies:

Thermal disorder. Jarlborg [259, 404–406] pioneered the view that lattice properties are predominantly responsible for the insulator-to-metal crossover in FeSi. Using a static supercell approach with a Gaussian distribution of atomic displacements, he evidenced a gap closure at around 300 K, found an increasing magnetic susceptibility of reasonable magnitude (at least when including exchange enhancement factors of 3.5–6), a sensible specific heat, as well as spectral weight transfers in the optical conductivity over a range of more than 3.5 eV. This thermal disorder picture was later supported by the molecular dynamics (MD) simulations of Delaire et al. [393]. Moreover, investigating the effect of thermal disorder on inter-atomic force constants, Stern and Madsen [276] concluded that thermal disorder in conjunction with the thermal expansion can explain most of the anomalous softening of acoustic phonon branches.

A few comments are in order: (i) in the MD calculations [393] FeSi metallizes only for temperatures reaching the order of $k_B T \approx \Delta_{\text{DFT}}/2$, congruent with the mean-square average of atomic displacements being controlled by $k_B T$. Experimentally, FeSi already metallizes for temperatures corresponding to a much smaller fraction of the gap. (ii) The thermal disorder induces changes in the electronic structure on all energy scales, explaining why, in this scenario, optical weight is redistributed over large energy scales. However, the optical spectral weight redistribution $\Delta n$ (see section 2.4.1) is, in the thermal disorder picture, a monotonously increasing function of energy up to 3.5 eV [405], contrary to experimental findings (see figure 19)51. (iii) The thermal disorder picture is unlikely to explain the occurrence of the Curie–Weiss-like decay in the magnetic susceptibility at high temperatures.

Magneto-elastic coupling. An interesting proposal comes from Kranich et al. [344]: using inelastic neutron and neutron resonant spin-echo spectroscopy, the authors evidence that in FeSi the relative change with temperature of both the phonon linewidth and the phonon intensity neatly tracks the relative evolution of $3k_B T \cdot \chi(T)$, with $\chi(T)$ being the uniform magnetic susceptibility. The latter can be interpreted—in a local moment picture—as the square of an effective fluctuating magnetic moment $\mu_{\text{eff}}(T)$ (see the discussion in sections 1.2.2 and 50).

50 The metallization of a Kondo insulator via the electron-phonon coupling was studied theoretically e.g. in the Holstein–Kondo lattice model by Nourafkan and Nafari [407].

51 See also [155] for a system with large electron–phonon coupling, where optical spectral weight transfers have been quantitatively described without accounting for thermal disorder.
and 2.1.2). This empirical observation strongly suggests a direct link between the emergence of (ferro)magnetic fluctuations and the phonon self-energy. Using density functional perturbation theory in conjunction with a phenomenological broadening of the electronic DOS to mimic an incoherence-driven metallization the authors computed phonon linewidths. Finding a significantly too small scattering rate (1 meV versus the experimental 3–5 meV), they concluded that a metallization in the absence of magnetic fluctuations cannot account for the observed phonon line-broadening. As far as the phonon energies are concerned, Krannich et al [344] found that the softening (of phonons at the R point in the phonon Brillouin zone) can be grouped into three temperature regimes: (i) at low temperatures the quasi-harmonic approximation applies, (ii) above 100 K a regime of substantially increased softening occurs, (iii) above 300 K the softening resumes a rate similar to regime (i). This behaviour is shown schematically in figure 32 (B, right) and suggestively coincides with the insulating, crossover, and metallic regime witnessed in spectroscopies probing charge degrees of freedom as well as transport properties.

The phonon spectator picture. Parshin et al [410, 411] performed nuclear inelastic scattering experiments of FeSi as a function of temperature and under pressure. The rationale behind this experiment is to disentangle the effects of (i) an increasing amplitude of atomic vibrations with raising temperature from (ii) the thermally induced volume expansion. Loosely speaking, pressure is used to reduce the volume to a value that is otherwise realized at lower temperatures.

As previously evidenced [411] the lowest peak in the phonon DOS has a sudden increase when heating above \( T_c \sim 180 \) K. Parshin et al interpret \( T_c \) as the critical temperature of the electronic subsystem, i.e. the characteristic scale for which the metallization occurs. Applying pressure to FeSi at room temperature, a phonon DOS comparable to that at \( T_c \sim 180 \) K (and ambient pressure) is obtained at \( P \sim 2 \) GPa [410]. Comparing the two cases—(i) \( T = 180 \) K, \( P \sim 0 \) GPa and (ii) \( T = 300 \) K, \( P \sim 2 \) GPa—the crucial observation is that what is similar at both conditions is the unit-cell volume, while the mean-square amplitude of atomic vibrations, \( \langle u \rangle \), is largely different. This finding has led Parshin et al to dismiss the thermal disorder picture (which is driven by \( \langle u \rangle \)) and to assign the change in unit-cell volume (instead of the phonons) as the dominant lattice contribution to the insulator-to-metal crossover in FeSi. In that sense, the influence of the lattice is a static one, while phonons are mere spectators, not actors, to the metallization crossover. It should be noted, however, that the electronic degrees of freedom in the cases (i) and (ii) are not quite identical: While pressure reduces the magnetic susceptibility, it is unlikely that the value of \( \gamma \) at \( T = 180 \) K and ambient pressure is reached by applying 2 GPa at \( T = 300 \) K [173]. Moreover, while pressure increases the gap in activation-law fits [172, 175], pressure does not undo the effects of increasing temperature from 180 to 300 K in the resistivity [175]. Also, it should be noted that from the electronic point of view (see the optical conductivity in figure 5(b)), as well as the Raman linewidth (see figure 30), the reference temperature \( T_C = 180 \) K is already far above the metallization threshold.

Synthesis. In most of this section we have described theories that focus on the aspect of electronic correlation effects to describe the anomalous electronic, optical and magnetic properties of certain narrow-gap semiconductors. In this section we have instead reviewed experimental evidence, as well as some theoretical investigations that point to strong signatures of the electronic crossover in the phonon subsystem. To which degree the changes in the lattice degrees of freedom are a consequence of the electronic crossover, or vice versa, is an unresolved issue. It is the author’s opinion that effects of electronic correlations and the electron–phonon coupling, to an extent, conspire in causing the insulator-to-metal crossover. Yet, the anomalous magnetic behaviour in FeSi and related systems may signal that the prevailing role is played by electronic correlation effects. A full description of the coupled electron–phonon system is, however, still lacking. Indeed the interplay of electronic correlations and lattice degrees of freedom beyond the Born–Oppenheimer approximation is a challenging avenue for future research.

Theories of correlated narrow-gap semiconductors: key points

- Signatures of correlation effects in diverse observables can be quantitatively captured with realistic many-body simulations.
- Insulator-to-metal crossovers are linked to incoherence induced by (ferromagnetic) spin fluctuations.
- Lifetimes and effective masses in d-electron-based insulators are propelled by Hund’s physics.
- Kondo-insulating (covalent) behaviour can coexist with the suppression of quantum fluctuations by local crystal-field (ionic) physics. Indeed, we propose an orbital-selective Kondo insulator scenario for FeSi.
- Important interplay of electronic and lattice degrees of freedom.

3. Thermoelectricity

The topic of this section is thermoelectricity in correlated narrow-gap semiconductors. We begin section 3.1 with a brief general introduction into the field of thermoelectrics: we introduce relevant quantities, and review in section 3.1.1 current efforts to improve thermoelectric performance. More details can be found in excellent books, e.g. [412, 413], or review articles [414–420]. Section 3.1.2 then gives an overview over potential impacts of electronic correlations onto thermoelectricity. To explore the effects of simple many-body renormalizations onto the thermopower of a semiconductors, section 3.1.3 discusses an instructive analytical model. As a by-product of the latter, it is recapitulated that the thermopower owing to electronic

Note that also in the Kondo insulator YbB\(_2\) a coupling between magnetic and phononic properties was suggested [408, 409]: while no perceptible phonon softening occurs above the system’s coherence temperature \( T^* \approx 50 \) K, the phonon intensity of modes (whose symmetry is thought to be commensurate with that of relevant magnetic excitations [409]) significantly drops around \( T^* \).
diffusion has an upper bound in a coherent semiconductor. Including effects of incoherence—ubiquitous in correlated materials—is shown to dramatically change transport properties, to the effect that semi-classical Boltzmann approaches fail to describe them, as discussed in section 3.1.4.

This prelude sets the stage for the discussion of thermoelectric effects in the material classes that are the subject of this review. As in the previous sections, we shall pick representative materials for each class and review experimental findings and theoretical efforts to describe them. Each subsection will discuss and highlight different aspects of electronic correlations and their impact on thermoelectricity:

- Silicides, section 3.2: effects of incoherence and effective masses onto the thermopower.
- Marcasites, section 3.3: limits of a purely electronic picture of thermoelectricity, the phonon-drag effect, couplings of phonons to charge carriers with large effective masses.
- Skutterudites, section 3.4: large consequences of unreliable electronic structure calculations.
- Heuslers, section 3.5: materials on the verge of thermoelectric particle-hole symmetry, impact of defects and disorder.

Besides the materials discussed here, there are numerous other classes of compounds that are of interest from the thermoelectric point of view, while at the same time probably hosting at least moderate correlation effects. Among these materials are half-Heuslers [421], other 4d transition-metal intermetallics, pnictides, and chalcogenides [422] not covered here, complex oxides, sulfides, and selenides [423].

### 3.1. General considerations

Thermoelectricity is the effect of a voltage drop $\Delta V$ occurring across a sample that is subjected to a temperature gradient $\Delta T$ or vice versa. The most common mechanism of thermoelectricity is the net diffusion of electrons (or holes) from the hot toward the cold regions of the sample. The displaced charges build up an electrical field that counteracts the diffusion. The trace goes over orbital indices $\alpha$ and $\beta$.

$S = \Delta V / \Delta T$, also known as the thermopower. The materials’ properties that govern the efficiency of a thermoelectric conversion process—leading to electricity generation, or refrigeration—can be gathered into the so-called dimensionless figure of merit $ZT$.

$ZT = \frac{S^2 \sigma T}{\kappa}$

Besides the materials discussed here, there are numerous other classes of compounds that are of interest from the thermoelectric point of view, while at the same time probably hosting at least moderate correlation effects. Among these materials are half-Heuslers [421], other 4d transition-metal intermetallics, pnictides, and chalcogenides [422] not covered here, complex oxides, sulfides, and selenides [423].

$ZT$ is the thermoelectric figure of merit. It is apparent that an optimization of the latter requires satisfying conflicting demands. This dilemma is illustrated in figure 33 where we display the typical dependency of relevant quantities on the carrier concentration. The conductivity is typically large in insulators ($S \propto \Delta T / \Delta T$) and small in metals (where, at low $T$, $S \propto T$). The conductivity, on the other hand, is large in metals and exponentially suppressed in insulators. Hence, the power factor $S^2 \sigma$ is largest in between, at densities corresponding to degenerate semiconductors. The thermal conductivity, $\kappa = \kappa_e + \kappa_h$, consists of an electronic and a phononic part. While in metals the electronic contribution is preponderant, in the doped semiconductor regime the lattice contribution by far dominates the detrimental thermal conduction and severely limits $ZT$.

The conductivity and the thermopower can be expressed, e.g. in the Kubo linear response formalism. One finds [294, 425–427]

$S_{\alpha \beta} = -\frac{k_B}{e|A_0|} \rho_{\alpha \beta}$

$\sigma_{\alpha \beta} = \frac{2\pi e^2}{h} |A_0|$ (11)

where the response functions are

$A_0 = \int d\omega \beta\rho(\omega - \mu) f(\omega) \left( \frac{\partial f}{\partial\omega} \right) \sum_k Tr \left[ \upsilon_k^\alpha A_0(\omega) \upsilon_k^\beta A_0(\omega) \right]$. (12)

Here, $k_B$ is the Boltzmann constant, $\beta = 1/(k_B T)$ the inverse temperature, $f(\omega)$ the Fermi function, and $\mu$ the chemical potential. The trace goes over orbital indices $L$ and $L'$ and the Fermi velocities, $\upsilon_k^L$, and the spectral functions, $A_0(\omega)$, are matrices in orbital space. Further, $\upsilon_k^L$ has an orientational $\alpha$-dependence, which turns both $\sigma$ and $S$ into tensors, whose elements depend on the Cartesian directions $\alpha$, $\beta$ of the temperature gradient and the electrical field, respectively.

### 3.1.1. Optimizing thermoelectric performance.

To increase thermoelectric efficiencies two main avenues are currently pursued [428–431] 

53 The minimal realization of an actual thermoelectric device consists of a thermocouple made of two materials. Then the overall figure of merit becomes $ZT = \frac{k_B}{e|A_0|} \rho_{\alpha \beta}$ for the average temperature $T$ [424]. Nonetheless, an individual material’s $ZT$ still gives a good indication as to its potential thermoelectric performance.

### Figure 33.

Optimizing thermoelectric performance. Shown is the typical carrier-density dependence of quantities relevant for thermoelectric devices at a given temperature: the conductivity $\sigma$, the thermopower $S$, the powerfactor $S^2 \sigma$, the thermal conductivity $\kappa$, and the thermoelectric figure of merit $ZT$. 

54 If neglecting electron-hole interactions, so-called vertex-corrections. 

55 For interesting proposals that are beyond this classification see, e.g. the recent articles [432, 433].
Phonon engineering. Following the ‘phonon-glass, electron crystal’ paradigm [434], the dominant lattice thermal conduction is sought to be limited by disordered unit-cells [429], structuring samples on the nano [414, 428] and the mesoscale [435], introducing geometric constraints via superlattices [436–439] or nanowires [440, 441], or including loosely bound ‘rattling’ atoms/modes [442–446]. The global rationale is to introduce phonon scattering channels on length scales that are long compared to the electronic mean-free path. Therewith, while reducing detrimental thermal conduction, the electronic powerfactor ideally is unhampered. This strategy of phonon engineering is bounded by the physics of the amorphous limit, yet it has led to significant progress, yielding $\kappa T$ values in excess of 2. A further route to reduce $\kappa$ are enhanced phonon-phonon scatterings owing to anharmonicities [447, 448] linked to, e.g. ferroelectric lattice instabilities [449–452].

Band engineering. The second major strategy is the enhancement of the power factor, i.e. an optimization of the electronic degrees of freedom\(^56\). Besides reaching an optimal compromise between the thermopower and the conductivity (see figure 33), the crucial observation is that the thermopower (somewhat akin to the Hall coefficient) is sensitive to the type of charge carriers. Indeed $S$ vanishes for particle-hole symmetric systems and increases the more asymmetric the electronic spectrum is. A large particle-hole asymmetry can be achieved by a variety of mechanisms, e.g.

- impurity levels that introduce sharp (=high-mass) density-of-states peaks on one side of the Fermi level [455, 456],
- band convergence, i.e. a high degeneracy of selective excitations [42, 457–459],
- low-dimensional bands in three-dimensional crystals [460, 461] that lead to large effective masses. This paradigm achieves the ideas of Hicks and Dresselhaus [462] without the need for electronic confinement via quantum-well structures,
- ‘pudding mold’ band-structures as proposed by Kuroki and Arita [463] (see also [464, 465]). Here, the thermopower is enhanced by a large particle-hole asymmetry in the Fermi velocities.

The first three points advocate a large asymmetry of the excitation spectrum $A_{\kappa}(\omega)$, while the fourth emphasizes the role of transition matrix elements $\delta_{\kappa}$. In the simplest picture there is, however, a compensation between the spectral and the velocity route towards large $S^2 \sigma$. Commonly the Fermi velocity is approximated by the group velocity, $v(\mathbf{k}) = 1/\hbar \delta_{\mathbf{k}}(\mathbf{k})$\(^57\). Then it is apparent that flat bands, that create sharp spectral features, have a small velocity (at least in some direction), which in turn hampers the response functions (in that direction)\(^58\).

\(^{56}\) See, however, section 3.3 and e.g. [453, 454] for a phonon-mediated contribution to the thermopower.

\(^{57}\) This is known as the Feierls approximation, see e.g. [155, 466].

\(^{58}\) As a nice illustration let us mention the red pigments Pb(Se,Te), Sn(Se,Te), and Bi$_2$Te$_3$. These materials combine good powerfactors $S^2 \sigma$ with reasonably low thermal conductivities $\kappa$ to reach a favourable thermoelectric conversion efficiency. Moreover, the intrinsically low thermal conduction in, e.g. lead telluride [449] alloys [467] could be combined with the above mentioned band-convergence [457], mesoscale phonon engineering [435], as well as with impurity states [455].

3.12. Electronic correlations and thermoelectricity: friends or foes? The propensity of correlated materials to exhibit large response functions begs the question whether electronic correlation effects could provide means to improve upon the conventional thermoelectric materials just mentioned—all of which are uncorrelated narrow-gap semiconductors. Several scenarios suggest an affirmative answer [412, 423]:

On a model level [425, 427, 468–471], as well as for realistic calculations of metallic compounds [472–475], correlation effects were shown to enhance the thermopower. In fact, the low-temperature thermopower in correlated metals is boosted by $1/\kappa T$, with the quasi-particle weight $\kappa < 1$ [476]. However, since the conductivity scales with $\kappa^2$ [294], the powerfactor is actually independent of $\kappa^2$\(^59\).

Noting that the thermopower is a measure for the entropy per carrier [424, 425, 470, 475, 477, 478], there are further sources of a large thermoelectric response: indeed, in the limit of large temperatures, the thermopower is entirely determined by the number of possible microstates (degeneracy) $g$,

$$
S \xrightarrow{T \to \infty} \frac{1}{e} \left( \frac{\partial \Omega}{\partial N} \right)_{E,V} = -\frac{k_b}{|e|} \frac{\partial \log g}{\partial N}
$$

where $s$ is the entropy of the system, $N$ the number of particles, and $E$ and $V$ indicate that the derivative is taken at constant energy and volume. Equation (13) is called Heikes formula [479]. In correlated materials, the degeneracy is enhanced by the typically large number of spin- and orbital degrees of freedom. In fact, the large thermopower in cobaltates has been proposed to originate from such entropic contributions [480, 481]. Heikes-type formulae were demonstrated to also hold for other oxides [482].

Large effects of entropy are also expected near phase transitions. Indeed, large thermopower enhancements have been evidenced near structural instabilities. The prime example is copper selenide, Cu$_2$Se, a compound whose transition from a low-temperature semi-conducting $\alpha$-phase to a superionic $\beta$-phase is accompanied by a lambda-shaped thermopower anomaly [483, 484]. Similar effects have been evidenced also in other semiconductors with mixed ionic and electronic conduction, e.g. Ag$_{10}$Te$_2$Br$_3$ [485] or AgCuS [486], where structural phase transitions induce sudden changes in the ionic mobility. In this context let us highlight the recent work of Sun et al [432], who evidenced that a large temperature-dependence in electronic mobilities induces pronounced features in the thermopower as well as in the Nernst coefficient. A strong

\(^{59}\) In fact, this is different from the semi-classical picture, where $\sigma \propto m^{* \to 1}$, i.e. $\sigma \propto Z$ for a momentum-independent self-energy.
variation of the electronic mobility could be engineered by fabricating a junction between materials of vastly different mobilities. Large mobility gradients could, however, also be realized in individual correlated materials in the vicinity of ordering instabilities, metal-insulator transitions or other phenomena that cause abrupt changes in electronic relaxation rates. This mechanism of mobility gradients was suggested to be active for the Ni-doped skutterudite CoSb₃ [432] (see section 3.4), as well as heavy-fermion compounds such as CeRu₂Al₁₀ [432] and CeCu₂Si₂ [319]. Also the recent observation of colossal thermopower anomalies in doped double perovskites [487] could be interpreted in this vein. Whether critical phenomena in the multifarious structural, magnetic, and electronic phase transitions in correlated materials may lead more generally to useful behaviours in the thermopower is yet an open question.

Further, large thermopowers occur when the excitation spectrum is strongly peaked. Indeed as suggested by Mahan and Sofo [488], ‘the best thermoelectric’ has a very narrow peak within a couple of k_B T around the Fermi level. Besides the paths of band-structure engineering on the one-particle level mentioned above, weakly dispersive, narrow features can be generated by many-body enhancements of effective masses. Particle-hole selective renormalizations can moreover result in an increased asymmetry, benefiting thermoelectricity, as suggested by Haule and Kotliar [473]. See e.g. table 15 for the notable particle-hole asymmetric scattering rate in FeSb₂ and FeAs₂.

Many correlated narrow-gap semiconductors do indeed exhibit large thermoelectric effects: FeSi, FeSb₂, CrSb₂, and Fe₃VAl all have intrinsically large thermopowers and power-factors (see table 1). Yet, being insulators, this large response occurs at rather low temperatures (see the next section). The same applies to Kondo insulators (see section 1.2). Although the undoped materials are hence unsuited for standard thermoelectric applications, such as waste-heat recovery at elevated temperatures or Peltier refrigeration down from ambient conditions, they do have potential for cooling applications or sensors at cryogenic temperatures [111], and could significantly reduce the need for liquid helium. The figure of merit of correlated narrow-gap semiconductors and Kondo insulators, are however miniscule, because of their strong lattice thermal conductivity. To address this issue, it has been suggested [489] to combine the aspect of correlation effects with that of geometric constraints mentioned earlier.

Conventional thermoelectrics are degenerate narrow-gap semiconductors. For correlated narrow-gap semiconductors, however, doping typically does not help achieving high power-factors at larger, more generally useful temperatures (see however sections 3.4 and 3.5). A likely reason for this drawback are the reduced electronic lifetimes that accompany the beneficial mass enhancements. In fact, the scattering rate \( \Gamma = -3\Omega(\omega = 0) \) is often not small compared to the gap. As discussed in section 2.4.1 for FeSi, this leads to an incoherence-induced metalization at intermediate temperatures, and doping yields a full metal rather than a degenerate semiconductor.

3.1.3. Limits of a purely electronic picture of thermoelectricity. We find it instructive to discuss here a simple analytical model [294] of a correlated but coherent narrow-gap semiconductor. By this we mean a system in which effective masses are renormalized and quasi-particle weights reduced, while excitations remain coherent (scattering rate \( \Gamma \ll 1 \)). The model serves several purposes: (i) it provides a minimal understanding of the coherent low-temperature regimes of systems such as FeSi, (ii) it shows the limits of applicability of band-like transport calculations, and (iii) it establishes an upper bound for the purely electronic thermopower in insulators.

Our starting point is (11) and (12). We consider a half-filled system of two bands, whose non-interacting dispersions are electron gas-like (quadratic in three dimensions), \( \epsilon_k^\pm = \pm \Delta_0/2 \pm \frac{\hbar^2 k^2}{2m^*_c}, \) for both (v)alence and (c)onduction states, and separated by a crystal-field gap \( \Delta_0 \). Using the Peierls approximation [155, 466] yields Fermi velocities \( v_F = \frac{\hbar}{\mathcal{Z} k_B T}, \) Correlation effects shall be encoded in a momentum-independent self-energy of the form \( \Sigma_{\nu}^c(\omega) = \hbar \Sigma^c_{\nu}(0) + (1 - 1/\mathcal{Z}^c)\omega - i\nu \mathcal{Z}. \) Then the interacting dispersions will be given by

\[
\xi_{\nu,p} = \pm \Delta_0/2 \pm \frac{\hbar^2 k^2}{2m^*_c},
\]

where \( \Delta = Z\Delta_0 \) is the renormalized charge gap, \( m^*/m_0 = 1/Z \) the effective mass of the carriers, and the origin of the chemical potential is chosen as the mid-gap point. Each excitation carries the reduced spectral weight \( Z^{c,\nu} \), and has an inverse lifetime/scattering rate \( \nu^c,\nu \).

In the limit of a coherent system (\( \Gamma \ll 1 \)) and a large gap (\( \beta |\Delta/2 - \mu| > 1 \)), (12) can be solved analytically to give

\[
A_n^{c,\nu} = (1)^\nu \frac{3\lambda}{\sqrt{2\pi/\beta}} e^{-\beta \Delta/2} \left\{ |\beta(\pm \mu - \Delta/2)|^n - \frac{5}{2} n \right\}
\]

where all carrier specific parameters have been collected in
\[ \lambda_{c,v} = \frac{Z^2 m^*_v}{m_0} \frac{5/2}{\Gamma} e^{\pm \beta \mu}. \]  

(16)

Then, the thermopower acquires the form

\[ S = \frac{1}{|e|T} \left( \mu - \frac{\Delta}{2} \delta \lambda \right) - \frac{5}{2} \frac{k_B}{|e|} \delta \lambda \]  

(17)

where we have defined the asymmetry parameter \( \delta \lambda \) by

\[ \delta \lambda = \frac{\lambda^c - \lambda^0}{\lambda^c + \lambda^0} \]  

(18)

which itself depends on \( T \) and \( \mu \). Hence, a large thermopower is achieved by an interplay of the gap, \( \Delta \), the chemical potential, \( \mu \), and the particle-hole asymmetry, \( \delta \lambda \). The latter is itself influenced by the bare masses \((m^0_v, m^0_e)\), the chemical potential \((\mu)\), and by the many-body effects of encoded in mass enhancements \((m^*_v, m^*_e)\), scattering amplitudes \((\Gamma_{v,c})\), and quasi-particle weights \((Z_{c,v})\).

**Low-temperature behaviour.** Let us have a look at the typical behaviour at low temperatures, where in systems such as FeSi or FeSb2, coherent transport is expected. In figure 34 we display the coefficient of the \(1/(|e|T)\) term, namely \( \mu - \Delta \delta \lambda/2 \), as a function of the chemical potential inside the gap, \(-\Delta/2 \leq \mu \leq \Delta/2\), for different temperatures \(T\).

In the particle-hole symmetric case, parameters for the valence and the conduction band are equal, i.e. \(m_v^* = m_e^*\), and the asymmetry parameter becomes \( \delta \lambda = \tanh(\beta \mu) \). Then, the thermopower vanishes for all temperatures, since \( \mu = 0 \). However, small deviations from particle-hole symmetry, that could, e.g. be caused by impurity states pinning the chemical potential, can result in large thermopowers. In fact, the thermopower in narrow-gap semiconductors is quite sensitive to defects or off-stoichiometries, as also indicated by often notable sample dependencies (see, e.g. sections 3.4 and 3.5).

Larger thermopowers are possible in case the electronic structure is strongly particle-hole asymmetric. Figure 34 displays a case, where the effective mass of the conduction electrons is much higher than that of holes in the valence band. Then, a large positive thermopower is attainable. This situation is for example realized in FeSi. Doping the system with electrons—i.e. moving \( \mu \) to the right in figure 34—will drive the system through a sign change in \( S \), to realize a negative thermopower albeit of smaller absolute magnitude. Experimentally this is seen, e.g. for Co-doped FeSi, see figure 39 (right). There, undoped \((x = 0, \text{red})\) FeSi has a large positive thermopower at low temperatures. Adding electrons \((x = 0.05, \text{blue})\) switches the sign and yields an overall smaller \( S \).

As seen in figures 36 and 39 (right), the thermopower falls off quicker than \(1/7\) above its peak value. This is potentially indicative of a breakdown of the simplifications made in the current, simple model. Most notably, assuming the scattering rate to be vanishingly small might no longer be a viable approximation. This issue will be discussed more in the next section 3.1.4, as well as for FeSi in section 3.2.

**An upper bound for the diffusive thermopower in semiconductors.** Within the framework considered here, the thermopower cannot be arbitrarily large. Indeed, the asymmetry parameter in (18) is bounded: \(|\delta \lambda| \leq 1\). With the further constraint of the chemical potential remaining inside the gap, \(|\mu| < \Delta/2\), the thermopower of the coherent semiconductor thus obeys

\[ |S(T)| \leq \Delta/2 + 5/2k_B. \]  

(19)

In the coherent regime, correlation effects can hence enhance the thermopower by selective renormalizations (of, e.g. the effective mass). The fundamental limit (19), which simply corresponds to fully eliminating either the electron or hole contribution to thermoelectric transport, can, however, not be surmounted. While derived for a two-band model, also effects of degeneracies cannot change this fact. Through similar considerations, Goldschmidt and Sharp [490] actually suggested that the charge gap of an insulating system could be read off from the peak value \( S_{\text{max}} = S(T_{\text{max}}) \) of the thermopower for which they assessed: \( S_{\text{max}} \approx \Delta/(2k_T) \).

In section 3.3 (see in particular figure 41) we will encounter systems that violate the upper bound (19). As discussed there, it is believed that in these materials a major part of the thermopower comes from the so-called phonon-drag effect that allows to circumvent the bound of the purely electronic thermopower considered here.

### 3.1.4. Electronic correlations: a challenge for ab initio transport methodologies

In currently used thermoelectric materials electronic correlations do not play a significant role\(^{60}\). This allowed the successful use of band-structure methods to describe their electronic structure, and Boltzmann-derived methodologies [491] to compute their transport properties. Applications of Boltzmann codes—such as those of [492, 493]—are diverse and include theoretical studies of individual materials [494–496], theoretical support for experimental studies [497, 498], and even high-throughput studies for thermoelectrics [458, 499–504]. These simulations indeed reach predictive character: it was, e.g. suggested that PbSe could be a low-cost alternative to the commonly used, but expensive PbTe [505]. This proposal was supported by subsequent experiments [506].

In materials with sizable electronic correlation effects, however, both band-theory and the standard Boltzmann approach become inaccurate. While we discussed the implications of many-body effects onto the electronic structure in section 2, we focus here on the challenges that correlation effects pose for transport methodologies. For this, we compare the Kubo approach (see (11)) with the commonly used semi-classical Boltzmann approach in the relaxation time approximation. The latter can be obtained from the Kubo formalism in the limit of a vanishing scattering rate, \( \Gamma \rightarrow 0 \). Then, e.g. the conductivity and the thermopower of (11) simplify to

\[ \sigma_{\alpha\beta} = \frac{e^2}{\Gamma} \sum_{kl} \left( \frac{\partial f}{\partial \omega} \right)_{\omega = \epsilon_{kl}} v_{\alpha k}^0 v_{\beta l}^0 \]  

(20)

and

---

\(^{60}\) This does not mean that all these materials are well-described by DFT. Indeed band-gaps in the undoped parent compounds are consistently underestimated.
and have omitted interband contributions. Then, the Boltzmann resistivity becomes independent of the scattering rate $\Gamma$.

As an illustration of the importance of finite lifetime effects, we show in figure 35 a comparison for a generic case of a renormalized narrow-gap semiconductor supplemented with a temperature-independent scattering rate. At large temperature the Kubo and the Boltzmann approach yield similar resistivities, that follow activation laws, characteristic of coherent insulators. However, already below 200–400 K (for the given parameters), strong deviations between the two approaches occur: while the Boltzmann resistivities continue to grow exponentially, the rise in the Kubo result slows down and $\rho$ tends to level off. Trends toward resistivity saturation are commonly witnessed experimentally in correlated narrow-gap semiconductors (see figure 1 for some examples). Usually this situation is interpreted as indicative of impurity or defect derived in-gap states. Here, we propose that a finite (residual) scattering rate for intrinsic charge carriers provides an alternative scenario for resistivity saturation in the Kubo result, on the other hand, largely depends on the scattering rate. Towards absolute zero, $S$ vanishes as required by the third law of thermodynamics. At intermediate temperatures, the thermopower is non-monotonous in temperature.

Turning to the thermopowers of the model, figure 35 (right), we see large differences between the two approaches: the Boltzmann result is independent of the scattering rate, see (21), and diverges inverse proportional to $T$, as expected for a fully coherent semiconductor (see also section 3.1.3). The Kubo result, on the other hand, largely depends on the scattering rate. Towards absolute zero, $S$ vanishes as required by the third law of thermodynamics. At intermediate temperatures, the thermopower is non-monotonous in temperature and assumes a shape common to that seen in experiments, see figure 3. Interestingly, for some temperatures, the Kubo thermopower can be larger than the Boltzmann result. Peak values of the thermopower, however, increase with growing coherence (smaller scattering rate $\Gamma$).

The resulting powerfactors are shown in figure 35 (right). While $S^2\sigma$ is larger for small scattering rates, it is interesting to note that properly accounting for effects of finite lifetimes via the Kubo formalism yields larger values than the semiclassical Boltzmann approach.

\[ S_{\alpha\beta} = \frac{e}{T} \sum_{KL} (\epsilon_{KL} - \mu) \left( \frac{-\partial f}{\partial \epsilon} \right)_{\omega=e_{KL}} \frac{\epsilon_{KL} v_{KL}^\beta v_{L\beta}^\alpha}{\sum_{KL} \left( \frac{-\partial f}{\partial \epsilon} \right)_{\omega=e_{KL}} \epsilon_{KL} v_{KL}^\alpha v_{L\alpha}^\beta} \]  

(21)

where, for simplicity, we assumed a momentum- and band-independent scattering rate $\Gamma_{KL} = \Gamma$ and have omitted interband contributions. This expression is a mere guide to the eye and is independent of the scattering rate $\Gamma$. The resulting powerfactors are shown in figure 35 (right).
These model calculations thus show: (i) a quantitative description of transport properties in regimes of non-negligible lifetime effect requires going beyond the Boltzmann approach (for metallic systems in which accounting for correlation effects proved crucial, see e.g. [464, 474, 475, 511]), (ii) finite lifetime effects can actually yield a better thermoelectric performance than would be anticipated from semi-classical approaches. Consequently, high-throughput screenings [499–503] based on DFT and Boltzmann approaches might overlook promising thermoelectric materials.

3.2. Silicides: FeSi, RuSi and their alloys

The electronic structure of the prototypical correlated narrow-gap semiconductor FeSi was discussed in section 2, and its properties compared to Kondo insulators in section 1.2. Here, we focus on the thermopower of the material, that was already shown in figures 3 and 4(c). In figure 36 we again display the experimental results, now in comparison to theoretical simulations.

3.2.1. Simulations of the thermopower. At temperatures below 100 K, i.e. when FeSi is in its coherent regime (see section 2.4.1), its thermopower can be reproduced by a slightly hole-doped and renormalized band-structure. This is illustrated in figure 36, where we plot a thermopower obtained from band-theory doped with 0.1% holes per iron (see also the work of Jarlborg [259]) and supplementing the dispersions with an effective mass of two so as to narrow the band-gap to the experimental value (see section 2.4.1). The agreement of this band-theory-derived thermopower with experiment, however, worsens significantly when temperature rises. Indeed, above 100 K the experimental thermopower decreases quickly in absolute magnitude and changes sign twice. Good agreement with experimental results is found, when computing the thermopower of FeSi within Kubo’s linear response formalism (see section 3.1.4) from the realistic many-body electronic structure discussed in section 2.4.1. In particular, the sign changes as a function of temperature, which indicate the transition between hole ($S > 0$) and electron ($S < 0$) dominated transport, are captured. This non-monotonousness in the thermopower was already heralded by the moving of the chemical potential, as seen in the right inset of figure 19 (top left): starting from low temperatures, the chemical potential moves down, therewith reducing the hole contributions to the thermopower, before passing, at around 120 K, the point of thermoelectric particle-hole symmetry, below which the thermopower becomes negative. At higher temperatures yet, the trend reverses and the thermopower changes sign again. Besides the thermopower, also for the resistivity, and the powerfactor derived from both of them, there is good agreement between experiment and the many-body simulations, see figure 37. Whether the many-body picture for FeSi [261] also accounts for the compound’s Hall [26, 84, 512] and its intriguing Nernst coefficient [26], is an open question.

63 That tiny amounts of holes improves congruence might not be a question of introducing charge, but simply of changing the particle-hole asymmetry.
3.2.2. Influence of correlation-induced incoherence. Having validated the accuracy of transport simulation in the many-body framework, we can elucidate the origin of the quenching of the thermopower above 100 K. Is it a consequence of the chemical potential moving, or does the onset of incoherent conduction destroy the voltage drop across the sample? In figure 38 we therefore compare the original DFT thermopower (from figure 36) with a calculation in which lifetime effects, i.e. $\Im \Sigma(\omega)$, are frozen to their values at $T = 77 K$. We see that the resulting thermopower still switches sign, but remains much larger on an absolute scale. Hence, the moving of the chemical potential has a large influence on what is the dominant type of carriers (electrons or holes), but it is the onset of incoherent spectral weight at the Fermi level above 100 K (see the spectral function in figure 5(a)) that is the main culprit for the substantial decrease in the thermoelectric signal. In that sense, short lifetimes, i.e. a substantial $\Im \Sigma(\omega)$, are antagonistic to thermoelectric performance in semiconductors.

3.2.3. Doped FeSi. If incoherent spectral weight at the Fermi level kills the thermopower, than so does doping. Indeed, while the thermopower of FeSi is robust with respect to different samples and preparation techniques, any disturbance away from stoichiometry vastly decreases the response. The uniqueness of pure FeSi was impressively illustrated by Sakai et al [233] who studied the thermopower of monosilicides and their alloys: in the continuous phase-diagram of temperature versus composition, see figure 39, FeSi sticks out as an island of largely enhanced thermoelectricity. The above discussed effects of incoherence explain why the narrow (red) strip of large and positive thermopower is restricted to low temperatures. Its confinement with respect to doping (the chemical potential) can be understood from the simple model, presented in section 3.1.3. Indeed, in an insulator with conduction electrons heavier than valence holes, large positive thermopowers are possible, as indicated in figure 34 that shows the coefficient of the 1/T behaviour. Moving the chemical potential to the left (hole doping) continuously decreases the thermopower which, however, remains positive. A trend similar to that shown here for Fe$_{1-x}$Mn$_x$Si, is realized in ligand-substituted FeSi$_{1-x}$Al$_x$ [84, 513].

3.2.4. A comparison of FeSi with RuSi. We find it instructive to compare FeSi to its isostructural and iso electronic 4$d$ homologue RuSi, and discuss the series Fe$_{1-x}$Ru$_x$Si in view of its thermopower [64]. As discussed in section 2.2.1 in conjunction with band-structure results, ruthenium silicide, RuSi, is a semiconductor with an indirect gap of 260–310 meV (obtained from resistivity measurements [18, 21]) and an optical gap of $\sim$400 meV [18, 20]. Interestingly, the trend in the size of the gap in the alloy series Fe$_{1-x}$Ru$_x$Si is not monotonous with $x$ [515], see figure 40: in fact up to a ruthenium concentration of 6% the charge gap is found to decrease with respect to FeSi. At higher concentrations the gap then starts to grow, passes the initial value of FeSi at around 15% ruthenium, and augments further up to the value of stoichiometric RuSi.

With the insight into FeSi (section 2.4.1, figure 38, and section 2.6), this behaviour was understood as follows [253]. In fact, there are, in Fe$_{1-x}$Ru$_x$Si, two opposing tendencies: (i) ruthenium has a larger atomic radius than iron. Thus, with increasing Ru content, the hybridization gap will shrink as the lattice expands. This effect is immediate [515], and wins for low ruthenium concentrations. (ii) With $x$ there is a crossover in the dominant orbital character of low-energy excitations, namely from a 3$d$ to a 4$d$ radial distribution (see also the discussion in Ce$_3$Bi$_3$Pt$_3$, Pd, in section 2.2.3). As a consequence of the greater extension of the latter, interaction matrix elements, such as the Hubbard $U$ and the Hund’s $J$, will have smaller values (see [310, 311], as well as the comparison of $U$ and $J$ of FeSb$_2$ and RuSb$_2$ in table 3 in section 2.2). As discussed in section 2.4.1, and elaborated on in the model calculations of section 2.6, the Hund’s rule coupling drives the strong correlation physics in FeSi. This is the reason why

Increasing $\mu$ (electron doping) beyond the maximum in figure 34 causes the thermopower to decrease abruptly, change sign, and realize negative values of smaller magnitude. This is indeed, what is found experimentally, as can be inferred from figure 39 and [385] for Co-doping, and is seen in [110] also for Ir-doping.

![Figure 39](image.png)

**Figure 39.** Thermopower of monosilicides. The thermopower of the 3$d$ monosilicides and their alloys is shown as colour code in the phase-diagram spanned by composition and temperature (left). For two vertical cuts, corresponding to stoichiometric FeSi and Fe$_{0.95}$Co$_{0.05}$Si, the thermopower is shown as a function of temperature (right). The behaviour at low-temperatures is congruent with the simple model shown in figure 34. Reproduced with permission from [233]. © 2007 The Physical Society of Japan.
correlation effects are near absent in RuSi, in congruence with the fact that band-structure theory underestimates its gap—instead of the large overestimation for FeSi, see section 2.2.1. The effect of diluting the Hund’s rule coupling in Fe$_{1-x}$Ru$_x$Si with growing $x$ becomes preponderant for larger ruthenium concentrations, diminishes effects of band-narrowings, and, thus, causes the gap to increase. The thermopower of RuSi, see figure 36, while not achieving the very large values of FeSi at low temperatures, is noteworthy in size over an extended temperature regime, with $S \sim 250 \mu$V K$^{-1}$ from 100 to 500 K. As discussed in section 3.1.3, the thermopower of a coherent insulator is controlled by the size of the gap $\Delta$ and the particle-hole asymmetry, while it cannot exceed $\Delta/T$. RuSi, having a larger gap than FeSi, exhibits a thermopower that indeed surpasses the envelope function, $\Delta_{\text{FeSi}}/T$, of FeSi, for 200 K $\leq T \leq 500$ K, as indicated in figure 36. However, RuSi obeys its own boundary, $\Delta_{\text{RuSi}}/T$, and the distance between the largest possible and the actual thermopower is in fact much larger than is the case for FeSi. Hence, the particle-hole asymmetry (called $\delta \lambda$ in section 3.1.3) is smaller in the 4d compound.

The dilution of the Hund’s rule coupling upon replacing Fe with Ru will, besides the decrease in effective masses, also diminish effects of incoherence. This should push the quenching of the thermopower (see figure 38) to higher temperatures. Therefore, isovalent, yet non-isoazimuthal and non-isovolume substitutions in correlated narrow-gap semiconductors might produce larger thermopowers in the temperature regime of 100–300 K. Besides the greater coherence, the loss of thermoelectric particle-hole asymmetry ($\delta \lambda$) when departing from pure FeSi, is partly compensated by the enlarged gap $\Delta$ (for $x > 0.15$). While the power factor $S^2 \sigma$ of stoichiometric FeSi reaches favourable $40 \mu$W/(K$^2$ cm) at around 60 K, see figure 37, a notable improvement beyond 100 K can be expected in Fe$_{1-x}$Ru$_x$Si for $x > 0$. Finally, the substitution will also reduce the thermal lattice conductivity $\kappa$, yielding a better figure of merit. Nonetheless, it is unlikely that Ru-substitution alone will result in a useful $ZT$, let alone a good price-performance ratio.

### 3.3. Marcasite di-antimonides: FeSb$_2$, CrSb$_2$ & Co

The material that has particularly fuelled the interest in thermoelectric effects in correlated narrow-gap semiconductors is FeSb$_2$. Indeed, its thermopower has rightly been called ‘colossal’ by Bentien et al. [23], who found values up to $-45$ mV K$^{-1}$ at around 10 K, and the powerfactor is the largest ever measured$^{65}$ with up to 8 mW K$^{-2}$ cm$^{-1}$ at 28 K [28]. While basic transport, optical and magnetic properties of FeSb$_2$ have been discussed in section 1.1 and its electronic structure in section 2.2.2, we here make the connection to its thermoelectric properties, and review works that try to explain the unusual behaviour. Here, one has to make a clear distinction between polycrystalline samples and single crystals. Indeed, while the temperature dependence of the thermopower is qualitatively similar for both cases, the colossal magnitude of $S$ occurs only in single crystals. The thermopower measured for polycrystalline samples [86, 87, 517, 518] instead never exceeds 1 mV K$^{-1}$, see, e.g. figure 42 (left).

#### 3.3.1. Polycrystalline samples—physics akin to FeSi$^2$

Where it peaks, the thermopower of FeSb$_2$ is negative, indicative of preponderant contributions from electrons. This is opposite to the hole-like, positive thermopower of FeSi, discussed above. This finding is in line with the particle-hole asymmetry evidenced in electronic structure calculations (see, e.g. the DOS in figure 15). Indeed spectral features are much sharper and of higher density for states below the Fermi level than above it. When described in terms of an effective model in the spirit of section 3.1.3, this asymmetry translates into effective masses that are larger for valence than for conduction bands. Assumming long electronic lifetimes at low temperatures, the qualitative behaviour of the thermopower of FeSb$_2$ can be discussed in terms of figure 34. There, to account for the hierarchy of masses, $m_v^* > m_c^*$ instead of the shown $m_v^* < m_c^*$, the figure needs to be mirrored with respect to both, the x- and the y-axis. Then it is expected that electron-doping leads to a gradual decrease of the thermopower that however remains negative. This is indeed what is found experimentally, see.

$^{65}$ With the possible exception of elemental bismuth, where, when combining results from different experiments (see data compiled in [516]), $S^2 \sigma$ can reach symbolic 81 mW K$^{-2}$ cm$^{-1}$ at 3 K.
e.g. the case of FeSb$_2$-$\text{Te}_x$ in [86, 387]. Hole-doping, on the other hand, quenches the thermopower more rapidly and leads to a sign change already at very low doping, see, e.g. [87] for the case of FeSb$_2$-$\text{Sn}_x$. A quantitative simulation of thermoelectric effects in polycrystalline samples of FeSb$_2$ is still outstanding. However, it can be expected that once the obstacles to describe the electronic structure of the compound (discussed in section 2.2.2) are overcome, a computation of (intrinsic) transport properties will come into reach.

3.3.2. Single crystals—do correlations cause the large thermopower? It is in single crystals [23, 27, 28, 386] that the thermopower of FeSb$_2$ was found to be unusually large, indeed ‘colossal’, with, however, notable variations on the quality [157] and size [27, 392] of samples (see also below). Given the pronounced effects of electronic correlation effects in this system (discussed in section 1.1), it was proposed that also the thermoelectric response owes to many-body effects.

Indeed, extending the comparison of FeSb$_2$ with its isostructural homologues FeAs$_2$ and RuSb$_2$ (see the discussion in section 2.2.2 and also section 2.6) to thermoelectric properties, one notes that only the strongly correlated FeSb$_2$ boosts an unusually large thermopower [24, 386, 519]. Also, the thermoelectric properties of doped FeSb$_2$, e.g. the mentioned FeSb$_{2-x}$Te$_x$ are indeed that of a correlated metal, i.e. the low-temperature thermopower is linear in $T$, with an enhancement factor, via the effective mass, of the order of 15 [386]. Moreover, the shape of the thermopower of FeSb$_2$ can actually be modelled [28, 519] with phenomenological expressions that include many-body renormalizations [294]. However, the modelling failed to account for the colossal magnitude of the thermopower. Here, it is instructive to examine the experimental results in terms of the model that was discussed in section 3.1.3. Since, at low temperatures, FeSb$_2$ is a coherent semiconductor (see section 1.1), the conditions to apply the band-like model are met. In this coherent, purely electronic picture—which, on top, neglects so-called vertex corrections—the thermopower cannot exceed $\Delta / T$, where $\Delta$ is the charge gap. In figure 41 we plot the experimental thermopowers of several narrow-gap semiconductors, scaled with $T / \Delta$. Any value of $|S| / T / \Delta$ in excess of unity indicates a violation of the theoretical upper bound of the thermopower. The antimonides FeSb$_2$ and CrSb$_2$, subject of this section, indeed defy the constraint. In the case of FeSb$_2$ the violation amounts to almost one order of magnitude [66]. Hence, to explain the large thermopower in the antimonides, either the omitted vertex corrections (describing, e.g. excitonic effects) somehow strongly boost the response, or, alternatively, an additional non-electron-diffusive mechanism is at work in these materials.

3.3.3. Empirical evidence for a phonon-drag thermopower. As a potential candidate the so-called phonon-drag effect [520, 521] has been suggested for FeSb$_2$ [294]. This electron–phonon scattering mediated enhanced electron-drift has been evidenced in pure, conventional semiconductors, such as p-type germanium [522], or indium antimonide [523], but also in pure metals and dilute alloys [524]. In the following we will summarize empirical evidence that motivates the presence of the phonon-drag in FeSb$_2$ and hints at why it is negligible in FeAs$_2$ which obeys the $\Delta / T$-rule.

When cooling FeSb$_2$ below its insulator-to-metal crossover, optical spectroscopy witnesses a large change in phonon lifetimes, suggesting an important electron-lattice coupling [22, 71] (see the discussion for FeSi in section 2.7). Also polarized Raman scattering experiments indicate notable electron–phonon couplings, that are strongly temperature dependent below 40K [525] (see figure 30). Further, we note that, as can be expected for a non-electronic mechanism of the thermopower, the magneto-thermopower of FeSb$_2$, $MT = [S(B) - S(0)] / S(0)$ is very low for samples with large $S(B = 0)$ [23]. Interestingly, the large low-temperature thermopower in FeSb$_2$ is also accompanied by a notable increase in the nuclear spin-lattice relaxation rate [526].

Also, the mentioned dependence of the thermopower onto the type of samples can be consistently interpreted in favour of the phonon-drag scenario. Indeed, a decrease in the phonon-meanfree path by scatterings of non-electronic origin is expected to lower phonon-drag contributions to the thermopower, as well as lattice contributions to the thermal conductivity. This naturally explains why the low-temperature thermopower found in polycrystalline samples [87], polycrystalline thin films [517] and nano-particular samples [527] is significantly lower than in single crystals. That all types of samples yield a comparable response at higher temperatures [526] suggests that the phonon-drag contribution to the thermopower decreases with the emergence of phonon-phonon or umklapp scattering, as does the lattice thermal conductivity (see figure 42).

The trend of the thermopower with the phonon-mean-free path has been put onto a much firmer footing by studies in which the granularity or size of samples was varied in a controlled fashion: Pokharel et al [392] found that the thermopower of nano-composites roughly grows logarithmically with the average grain size. Takahashi et al [27] studied rod-shaped single crystals of

---

Note that in figure 41 we have used the thermopower from [26], where the peak-value is $-18 \text{ meV K}^{-1}$, which is less than half of the record $-45 \text{ meV K}^{-1}$ found in [23].
for \( \text{FeAs}_2 \) \[24\], the specific heat of \( \text{FeSb}_2 \) will be larger than Debye temperatures of 348 K for \( \text{FeSb}_2 \) \[24, 87\], and 510 K for \( \text{FeAs}_2 \). Indeed, the thermal conductivity is found to be even larger in the arsenide. The charge carrier concentration at temperatures beyond this regime at around 150 K, where the entropy reaches \( R \log 2 \), owing to a second and larger hump in the specific heat, and in congruity with the susceptibility \[92, 128\]. Indeed \( \text{FeSb}_2 \) becomes paramagnetic above 100 K \[24, 87\], and a Curie-like downturn appears at temperatures above 350 K \[92\], whereas the susceptibility of \( \text{FeAs}_2 \) is flat up to 350 K \[24\].

The charge degrees of freedom (excitonic or charge ordering features) or, indeed, an electron–phonon effect. No such feature appears in the specific heat of \( \text{FeAs}_2 \) \[528\] for a computational study) is likely to be larger in the antimonide: the electronic contribution to the specific heat in the antimonide: the electronic contribution to the specific heat \( C_s \) is a measure for the entropy per carrier. Hence the phonon contribution \( S_{\text{ph}} \) to the thermopower will roughly scale with \( \frac{C_s}{n} \). With Debye temperatures of 348 K for \( \text{FeSb}_2 \) \[24, 87\], and 510 K for \( \text{FeAs}_2 \) \[24\], the specific heat of \( \text{FeSb}_2 \) will be larger than that of \( \text{FeAs}_2 \). The charge carrier concentration at temperatures where the thermopower is maximal, on the other hand, is larger for \( \text{FeSb}_2 \), \( n \sim 8 \times 10^{14} \text{ cm}^{-3} \) for the sample with largest \( S \) in \[386\], whereas \( n \sim 5 \times 10^{14} \text{ cm}^{-3} \) for \( \text{FeAs}_2 \). Also the electron–phonon coupling—the least accessible quantity (see Diakhate \[528\] for a computational study) is likely to be larger in the antimonide: the electronic contribution to the specific heat of \( \text{FeSb}_2 \) exhibits a notable feature below 50 K \[24\]. That there is no analogue in the spin response, may indicate that the features’s contribution to the entropy is associated with either the charge degrees of freedom (excitonic or charge ordering effects) or is, indeed, an electron–phonon effect. No such feature appears in the specific heat of \( \text{FeAs}_2 \) \[528\]. Together, this suggests a much smaller phonon-drag effect in \( \text{FeSb}_2 \).

Comparing the thermopower of the antimonide and the arsenide (see figure 1 in \[24\]) one notes, however, that the Seebeck coefficient of \( \text{FeAs}_2 \) becomes larger than that of \( \text{FeSb}_2 \) above 35 K. This might indicate—if the phonon-drag picture holds—that the effective electron–phonon coupling in \( \text{FeSb}_2 \) has sufficiently decreased (by Umklapp and phonon–phonon scattering) so that the thermopower is now dominated by the electronic degrees of freedom, i.e. the larger gap in \( \text{FeAs}_2 \) causes a larger response. Indeed, an approximate simulation \[294\] of purely electron-diffusive contributions to the thermopower yielded good results for \( \text{FeAs}_2 \) above 15 K, and, for \( \text{FeSb}_2 \), in the interval 35–50 K which is high enough for phonon-contributions to have decayed and low enough to not cross the insulator-to-metal transition of the compound.

### 3.3.4. Comparison to \( \text{FeAs}_2 \)

At first glance, phonon related properties, including the phonon-drag effect, might be expected to be important also for the isoelectronic and isostructural \( \text{FeAs}_2 \). Indeed, the thermal conductivity is found to be even larger in the arsenide \[24\]. Yet, the maximal thermopower of \( \text{FeAs}_2 \) is much smaller than that of \( \text{FeSb}_2 \), and, in particular, compatible with a purely electronic picture (see figure 41). The following might explain this disparity: the thermopower is a measure for the entropy per carrier. Hence the phonon contribution \( S_{\text{ph}} \) to the thermopower will roughly scale with the lattice specific heat \( C_l \) times the electron–phonon coupling constant \( g \), divided by the electron density: \( S_{\text{ph}} \propto g C_l / n \). With Debye temperatures of 348 K for \( \text{FeSb}_2 \) \[24, 87\], and 510 K for \( \text{FeAs}_2 \) \[24\], the specific heat of \( \text{FeSb}_2 \) will be larger than that of \( \text{FeAs}_2 \). The charge carrier concentration at temperatures where the thermopower is maximal, on the other hand, is larger for \( \text{FeSb}_2 \), \( n \sim 8 \times 10^{14} \text{ cm}^{-3} \) for the sample with largest \( S \) in \[386\], whereas \( n \sim 5 \times 10^{14} \text{ cm}^{-3} \) for \( \text{FeAs}_2 \). Also the electron–phonon coupling—the least accessible quantity (see Diakhate \[528\] for a computational study) is likely to be larger in the antimonide: the electronic contribution to the specific heat of \( \text{FeSb}_2 \) exhibits a notable feature below 50 K \[24\]. That there is no analogue in the spin response, may indicate that the features’s contribution to the entropy is associated with either the charge degrees of freedom (excitonic or charge ordering effects) or is, indeed, an electron–phonon effect. No such feature appears in the specific heat of \( \text{FeAs}_2 \) \[528\]. Together, this suggests a much smaller phonon-drag effect in \( \text{FeSb}_2 \).

### 3.3.5. Constraints to a theory of \( \text{FeSb}_2 \): correlation of features in transport observables

For the development of a microscopic understanding, it is paramount to establish as many empirical constraints as possible, against which a candidate theory can be tested. In the case of \( \text{FeSb}_2 \) the comprehensive work of Sun \[26, 519\] demonstrated that there is a striking inter-dependence of diverse transport observables. As depicted in figure 43, there are (at least) three temperatures at which the resistivity, the thermopower, and the Hall and Nernst coefficient display distinct features: Inflection points in the thermopower coincide with extrema in the Nernst coefficient, while the thermopower is maximal for inflection points in both the resistivity and the Nernst coefficient. Likewise, at these characteristic temperatures features appear in the Hall coefficient, as well as the magnetoresistance (see figure 44(d)). This intimate linkage between electrical, magnetic, and thermolectric quantities is a harbinger for a common microscopic origin, posing severe constraints on possible explanations of the colossal thermopower.

### 3.3.6. Modelling the phonon-drag in \( \text{FeSb}_2 \): evidence for important role of in-gap states

Realizing that the characteristic temperatures just mentioned all lie far below the insulator to metal crossover in \( \text{FeSb}_2 \), it is reasonable to assume electronic lifetimes to be long. In that case, one
The congruence of the temperature profile can be explained by the fact that in this scenario, the thermopower and the Nernst coefficient have the opposite symmetry under a particle-hole transformation when compared to their usual purely electronic counterparts: in the phonon-drag scenario the Seebeck coefficient is even, while the Nernst coefficient is odd, see table S9.

In the Nernst coefficient, the proposed phonon-coupling to the density of in-gap states circumvents Sondheimer cancellation [529], without requiring multi-band effects. Previously, it had been proposed that a strongly energy-dependent scattering rate $\Gamma(\epsilon)$ causes the colossal Nernst signal [26]. For a scattering rate of the form $\Gamma(\omega) \propto \omega^s$, the exponent can be determined by $r = -e/k_B \times \nu/(R_H \sigma)$. The phonon-drag formalism formally yields a large exponent $r = -292.9$, in reasonable agreement with [26], albeit in the absence of strongly dynamical electronic correlation effects.

### 3.3.7 Existence and possible origin of in-gap states.

To strengthen the unconventional phonon-drag picture, we here review experimental findings regarding in-gap states and their nature in FeSb$_2$, as well as mention attempts to pinpoint their origin using realistic electronic structure calculations.

In in-gap states in FeSb$_2$ have been held responsible for activation-law regimes with distinct energies in the material’s resistivity [23, 30], as well as the low-temperature upturn in the magnetic susceptibility [92]. Using pressure-dependent magneto-transport measurements, Takahashi et al [177], disentangled the band-gap energies and found strong evidence for the smallest gap to originate from an impurity band.

Contrary to FeSi [508], also low-temperature nuclear magnetic resonance measurements of both FeSb$_2$ and RuSb$_2$ (as well as FeGa$_3$ [530]) by Gippius et al [526, 531] show clear indications of localized in-gap states, suggested to...
demonstrated that even the smallest iron concentrations in that may suggest Fe-richness. Moreover, Fuccillo et al. further supported by the binary Fe accessed temperatures. The proposed origin of in-gap states, which shows a low-temperature upturn of less than that the expected magnetic signature of the required s law, it was shown the phonon-drag amplitude. Using Curie’s law, it was shown that the expected magnetic signature of the required excess iron is consistent with the experimental magnetic susceptibility, which shows a low-temperature upturn of less than $5 \cdot 10^{-5}$ emu mol$^{-1}$ for all accessed temperatures. The proposed origin of in-gap states is further supported by the binary Fe–Sb phase diagram that may suggest Fe-richness. Moreover, Fuccillo et al. demonstrated that even the smallest iron concentrations in Ru$_{1-x}$Fe$_x$Sb$_2$ causes the appearance of the low temperature characteristics of FeSb$_2$, suggesting a possible link to the transition-metal stoichiometry. Finally, recent transport measurements of Sanchala et al. showed that excess iron (or, equivalently, Sb-deficiency) has a large impact onto the peak value of the thermopower.

Figure 44. Boltzmann modelling of transport functions of FeSb$_2$. (a) Electronic structure consisting of a valence band (VB), a conduction band (CB), and three in-gap states (B.1–B.3) at energies $E_i = 11.4, 19.8, 24.3$ meV. The black line indicates the temperature dependence of the chemical potential. (b) Conductivity $\sigma$ including contributions from all states. (c) Phonon-drag thermopower owing to an injection of momentum into the in-gap states. (d) Conductivity contributions of the in-gap states in comparison to the magneto resistance MR. (e) The Hall coefficient, (f) the phonon-drag Nernst coefficient. The grey vertical lines are guides to the eye indicating maxima in the MR. Reprinted figure with permission from [453]. Copyright 2015 by the American Physical Society, with experimental data from Sun et al. [26].

be located slightly below the conduction band and to be of $S = 1/2$ character. This finding is consistent with the strong reaction of the in-gap states to magnetic fields (see the magnetoresistance in figure 44(d)). Also, Storchak et al. interpreted their muon spin rotation spectra as showing the presence of in-gap states, proposed to be of spin-polaron nature.

Modelling several defect and impurity scenarios from first principles, Battiato et al. found the off-stoichiometry Fe$_{1+x}$Sb$_{2-x}$ to be a likely candidate for the emergence of in-gap states. Only for this substitution, electronic states were found to emerge inside the fundamental gap. Figure 45 shows a simulation of a single anti-site defect: Fe$_{25}$Sb$_{47}$. Indeed two impurity bands appear close to the conduction states. The size of the supercell corresponds to $x \approx 0.04$ in Fe$_{1+x}$Sb$_{2-x}$. This concentration is, however, much larger than the estimated number of defect, $x \approx 4.5 \cdot 10^{-5}$, that is needed to account for the phonon-drag amplitude. Using Curie’s law, it was shown [453] that the expected magnetic signature of the required excess iron is consistent with the experimental magnetic susceptibility, which shows a low-temperature upturn of less than $5 \cdot 10^{-5}$ emu mol$^{-1}$ for all accessed temperatures. The proposed origin of in-gap states is further supported by the binary Fe–Sb phase diagram that may suggest Fe-richness. Moreover, Fuccillo et al. demonstrated that even the smallest iron concentrations in Ru$_{1-x}$Fe$_x$Sb$_2$ causes the appearance of the low temperature characteristics of FeSb$_2$, suggesting a possible link to the transition-metal stoichiometry. Finally, recent transport measurements of Sanchala et al. showed that excess iron (or, equivalently, Sb-deficiency) has a large impact onto the peak value of the thermopower.

3.3.8. Outlook. At first glance, the phonon-drag scenario for FeSb$_2$ appears a disappointment on several levels. However, a closer inspection actually reveals new paths for future research.

(i) Phonon engineering. The phonon-drag contribution to the thermopower $S_{\text{ph}}$ and the thermal lattice conductivity $\kappa$ are inter-linked, see e.g. figure 42. Maximizing the first while minimizing the latter hence means satisfying antagonistic effects. The usual electron-crystal–phonon-glass paradigm is therefore likely to be unsuitable: an undifferentiated decrease in the phonon mean-free path to lower $\kappa$ will equally diminish the phonon-drag effect. Any attempt at phonon engineering has therefore to be selective. It is a priori uncertain whether a separation between the phonon modes that dominate the phonon-drag and those propelling thermal conduction is possible, let alone whether they can be independently manipulated. In a pioneering work, Zhou et al. demonstrated the feasibility of a phonon filter that achieves just that. The authors showed that in the case of silicon the phonon-drag induced figure of merit, $ZT$, could be boosted by more than one order of magnitude. The key observation is that phonon modes contributing to the phonon-drag have typically a larger mean-free path, as well as notably lower energies than modes dominating thermal conduction. It is then proposed that nano-clusters can be used as impurities to preferentially scatter high-energy phonons, thus acting as a low-pass filter. These findings provide a new path for optimizing the phonon-drag mechanism in thermoelectric materials for applications at temperatures up to some fraction of their Debye temperatures.

(ii) Occurrence in other materials. The phonon-drag scenario proposed for FeSb$_2$ relies on an unconventional mechanism, in which momentum is transferred to in-gap states. However, defect states are notoriously difficult to control (see the large sample dependence of $S$). Therefore it could be expected that technical applications are cumbersome to realize for this system and that the microscopic mechanism might be non-universal, i.e. not found in other materials. Yet, despite the large variation in the thermopower, the signatures of a narrow band inside the fundamental gap is consistently found around the same energy, not just in other transport and spectroscopic probes, but also in the electronic specific heat. This may indicate that the impurity band is an intrinsic feature of the FeSb$_2$ compound or crystal structure. The magnitude
of the phonon-drag thermopower will, however, depend on minute details, and the interplay of position of in-gap states, the fundamental gap, the electron–phonon coupling, as well as phonon energies. It seems plausible that the inter-period substitutions Ru for Fe, or As for Sb, thus sufficiently perturb the phonon system (in addition to changes in the electronic structure, see section 2.2.2), so as to destroy the conditions for a large phonon drag in FeAs2 and RuSb2. A substitution that changes the mass-balances much less is replacing Fe with Cr. Interestingly, the isostructural CrSb2—despite its quite different electronic structure (d4 configuration in antiferromagnetic CrSb2 versus d6 in paramagnetic FeSb2)—also violates the upper bound of electronic diffusive thermoelectricity [33], see figure 41. Therefore it is tempting to propose that the material realizes the same phonon-drag mechanism as in FeSb2. Indeed, also in CrSb2 the existence of a donor impurity band below the conduction states has been suggested [33]. A further system could be FeGa3 [41]. As mentioned before, nuclear magnetic resonance measurements evidenced in-gap states there as well [530].

Also, more recent thermopower data [41], while not violating Δ/T, do exhibit a characteristically sharp peak of up to −15 meV K−1 at 12 K, as well as a pronounced sample, and interestingly, polarization-dependence in the thermoelectric response. In conclusion, the phonon-drag effect in intermetallic narrow-gap semiconductors with open d-electron shells might be more prevalent than initially thought. Further experimental investigations, in particular magnetoresponse and Nernst effect measurements for CrSb2 and FeGa3 are called for.

(iii) Correlation effects. In the scenario for FeSb2 electronic correlation effects do not seem to play any role. Indeed, the phonon-drag mechanism does not inject momentum into the correlation-effects hosting valence and conduction carriers. Instead phonons couple to in-gap states. As remarked by Takahashi et al [27], this, in fact, makes the phonon-drag effect particularly large: Indeed, according to the phenomenological treatment due to Herring [521], the phonon-drag contribution to the thermopower reads

\[ S_{\text{ph}} = \pm \alpha \frac{\tau_{\text{ph}} v}{m} \]

Here, \( \tau_{\text{ph}} \) and \( v_{\text{ph}} \) are the phonon lifetime and velocity, respectively, and \( \alpha \) a parameter related to the electron–phonon coupling strength that determines the fraction of the phonon momentum-loss/transfer. Crucially, the \( S_{\text{ph}} \) is inversely proportional to the mobility of the electrons (−) or holes (+) that dominate the regular electron diffusive thermopower. In the semi-classical picture, \( \mu = e\tau/m^* \), where \( \tau \) is the electronic lifetime and \( m^* \) the charge carriers’ effective mass. Hence, large effective masses enhance the phonon-drag. From cyclotron resonance spectra, a mass \( m^* = 5.4 m_e \) was indeed extracted for the impurity states in FeSb2 [27]. Masses of this order of magnitude are, however, not at all uncommon for intrinsic carriers in correlated materials. This suggests that strongly correlated electrons may actually lead to enhanced thermopowers and powerfactors also in semiconductors—with a little help of phonons.

Hence, instead of wrapping up and shelving the FeSb2 puzzle, the reviewed recent theoretical and experimental insights actually open up new directions in the field of thermoelectrics. While most likely not relevant to thermoelectricity in FeSb2, studying its colossal thermopower may have led to the unravelling of a new paradigm: the combination of the time-honoured phonon-drag effect and the physics of strongly correlated electrons.

Lessons from FeSb2: colossal thermopower caused by unconventional phonon-drag

- FeSb2 and CrSb2 violate upper bound (19) of purely diffusive thermopower.
- Correlation of features (extrema, inflection points) in diverse transport observables constrain microscopic modelling.
- Phonon-drag coupling to extrinsic in-gap states gives a consistent description of all transport quantities.
- Potential path for large thermopowers: coupling phonons to correlated electrons with large effective masses: enhanced phonon-drag thermopowers; advanced phonon-engineering to reduce detrimental thermal conduction while leaving beneficial phonon-drag intact.

3.4. Skutterudite antimonides: a focus on CoSb3

Skutterudites are the prototypical compounds in which the phonon-glass, electron-crystal paradigm [434] can be realized [444]. These materials crystallize in the CoAs3-structure (cubic space-group \( \text{Im}3 \)) harbouring four formula units. Crucially, the structure has voids at the 2a Wyckoff positions, that can be filled with rare earth, alkali metal, alkaline earth, or other elements, that are only weakly bound to the skutterudite host. As a consequence, the localized fillers exhibit large anharmonic vibrations that diminish the
propagation of phonons, and thus heat. Besides the possibility to reach low thermal conductivities, skutterudite systems exhibit notably large powerfactors. For example, $S^2\sigma$ of CoSb$_3$ reaches $9.5 \mu W/(K^2cm)$ at 625 K [46] (see table 1). Over the years, this led to a continuous optimization of the thermoelectric figure of merit. $ZT \sim 1$ at 800 K was achieved in the year 2000 [536], multiple fillings boosted the value to $\sim 1.6$ in 2011 [537], and $\sim 1.9$ was reached in 2016 [538].

3.4.1 Experimental and theoretical knowledge. To start, experimental results for this compound scatter notably [43]. For example, $S^2\sigma$ of CoSb$_3$ reaches 9.5 $\mu W/(K^2cm)$ at 625 K [46] (see table 1). Over the years, this led to a continuous optimization of the thermoelectric figure of merit. $ZT \sim 1$ at 800 K was achieved in the year 2000 [536], multiple fillings boosted the value to $\sim 1.6$ in 2011 [537], and $\sim 1.9$ was reached in 2016 [538].

Providing an all-encompassing overview of the research in skutterudites is far beyond the scope of this work. Instead we will focus on a single (unfilled) compound, CoSb$_3$, to highlight challenges in the ab initio modelling of its thermoelectric properties.

Figure 46. Thermopower and band-structure of CoSb$_3$. (a) Thermopower of CoSb$_3$ from Mandrus et al [45], Nolas et al [445], Leszczynski et al [543], Kawaharada et al [47], Kuznetsov et al [48], and of a lightly electron-doped (0.08% Te) sample from Caillat et al [88]. A comparison is shown to Boltzmann transport calculations of Wee et al [544] based on an LDA band-structure with the carrier density adjusted to the value extracted from the Hall coefficient of the n-type sample of Caillat et al [88]. (b) Thermopower of lightly Ni-doped CoSb$_3$, Experimental data from Sun et al [432]. The simulated thermopowers use the Kubo code from [261] with a constant scattering rate $\Gamma = 8$ meV and foot on the LDA and mBJ band-structures shown in (d). There the doping is simulated in the virtual crystal approximation (VCA). (c) LDA and QS$_\text{GW}$ band-structure of stoichiometric CoSb$_3$, (b) and (d) Adapted from [432]. CC BY 4.0. (a) CoSb$_3$: thermopower. (b) Co$_{0.999}$Ni$_{0.001}$Sb$_3$: thermopower. (c) CoSb$_3$: band-structure. (d) Co$_{0.999}$Ni$_{0.001}$Sb$_3$: band-structure.

320 meV, while Shubnikov–de Haas measurements [539] and photoemission experiments [540] suggest gap values as low as 30–40 meV. Also the thermopower is highly dependent on the sample preparation [541], to the extent that both $p$ and $n$-type samples have been measured [43, 45, 47, 48, 88, 445, 542, 543], see figure 46(a). From the Hall coefficient ($R_H > 0$ [90, 542]) as well as electronic structure calculations (DOS larger for conduction states, see figure 46(c)) one would naively expect a positive thermopower for the stoichiometric compound. One might surmise that the origin of these variations lies in the loosely packed crystal structure which makes skutterudite compounds prone to diverse types of defects. Therefore, also the very small gap values extracted from Shubnikov–de Haas and photoemission experiments might be linked to impurity bands.

These experimental uncertainties are met with substantial variations in theoretical results for the perfectly stoichiometric compound. Using experimental atomic positions yields a direct gap in the Kohn–Sham spectrum of CoSb$_3$ that varies from 40–60 meV (LDA) [545, 546], over 170 meV (ACBN0) [547] to 310 meV (mBJ) [432]. For the band-structure of

70 However, it has been noted that even some of the unfilled skutterudites have intrinsically low thermal conductivities, see, e.g. [535] for a discussion of FeSb$_3$.

71 The valence count, however, varies little, there are $\sim 7.6 d$-electrons on each Co-site (wien2k with muffin-tin radius of 2.5$a_0$).
slightly electron-doped CoSb₃ within LDA and mBJ, see figure 46(d) and its discussion below. On top of this, also the equilibrium volume largely depends on the functional used in the relaxation: one finds a difference of about 7% when relaxing the structure within LDA or GGA [546]. Using these theoretical structures, large variations in the gap size are found for a given functional [544]. The strong link between electronic and structural properties is also congruent with quantum molecular dynamics simulations [548]. The latter evidence a substantial temperature-dependence in the electronic structure in CoSb₃, and might explain the temperature-induced band-convergence seen in experiments [42].

Within DFT it is the treatment of exchange contributions that largely affects the size of the gap. Using quasi-particle self-consistent QSGW, we find an even larger gap of 730 meV, see figure 46(e). This increase in the size of the gap with respect to DFT-LDA occurs despite the presence of significant band-narrowings. Indeed, within QSGW the valence states are renormalized towards the Fermi with a factor of $m_e^*/m_{e\text{LDA}} \sim 1.8$. Due to the perturbative nature of the GW approach, it tends to underestimate effective masses in compounds with open 3d-shells (see also the discussion on di-antimonides in section 2.2.2). This finding thus indirectly advocates the presence of substantially enhanced effective masses in CoSb₃. Experimentally, masses of 2–5$m_e$ were indeed extracted from the Hall coefficient [88], or from the thermopower when doping with Ni [89], and 1–3$m_e$ when adding Yb fillers [42]. Given the competition [298, 549] of exchange self-energies (widen gaps) and mass enhancements (shrink gaps), as well as moderately large values of effective masses, an accurate modelling of the electronic structure of skutterudites requires a methodology that treats exchange and correlations beyond DFT, and correlations beyond the perturbative GW. Viable approaches will be mentioned in section 4.

Despite issues related to defect-derived in-gap states and the putative presence of electronic correlation effects, several transport observables have been successfully modelled from first principles: Guo et al [550] computed the thermal lattice conductivity of unfilled and filled CoSb₃ and found good agreement with experiments. Also thermoelectric simulations have been performed with some success: using the Boltzmann response formalism on top of band-structure calculations, several groups [544, 546, 551–553] reproduced experimental thermopowers when adjusting the carrier density to values extracted from the experimental Hall coefficients, see figure 46(a) for an example.

3.4.2. Doped CoSb₃. While for thermoelectric applications, doping CoSb₃ by introducing filler atoms is the way to go, let us here nonetheless highlight the intriguing case of lightly Ni-doped CoSb₃ [89, 90, 432]. Indeed introducing 3d-transition-metal-derived carriers to the skutterudite has pronounced effects on transport and magnetic properties even for very low concentrations [90, 432]: up to 1% Ni, the resistivity remains insulating, while the magnetic susceptibility develops a Curie-tail at low temperatures. The Hall coefficient becomes negative already for the tiniest amounts of Ni. Concomitantly, as seen in figure 46(c) for 0.1% Ni, also the thermopower is negative, with the exception of a pronounced positive peak below 25 K, that is accompanied by a large peak in the Nernst coefficient [432].

To first approximation, one can try to model the electronic structure of Co₁₋ₓNiₓSb₃ within the virtual crystal approximation (VCA). The resulting LDA and mBJ band-structures are shown in figure 46(c) for $x = 0.001$. These bands are very close to those without doping, albeit shifted rigidly downwards. Crucially the mBJ electronic structure is notably more particle-hole asymmetric than the LDA one, owing especially to a rearrangement of conduction bands, as well as the increased gap below the conduction states. The corresponding thermopower has been computed within the Kubo formalism, using a constant scattering rate and fixing the temperature dependence of the chemical potential by the requirement of charge neutrality (instead of adjusting the carrier density to experimental values [544, 546, 552]). With the chemical potential residing inside the conduction bands, the LDA and mBJ-based thermopowers shown in figure 46(c) are metallic. We note that—as expected—the visibly more particle-hole-asymmetric mBJ-band-structure translates into a larger thermopower. Vis-à-vis experiment, the mBJ-based thermopower approaches the experimental result above 100 K. In fact—if the electronic structure is reliable and lifetime effects are minor (see section 3.1.4)—congruence between simulation and experiment is expected at high enough temperatures when putative donor-band contributions can be neglected. This quantitative congruence suggests that the mBJ approach gives a qualitatively correct description of the intrinsic electronic structure of (lightly doped) CoSb₃. The qualitative discrepancy below 100 K, however, point to sizable contributions to the thermopower from effects not contained in the (intrinsical) band-structure. Indeed, using a phenomenological band-structure that includes a spin–orbit splitting (see also [551]) and several impurity bands, Kajikawa [554] successfully fitted the transport observables of Sun et al [432]. In the grander scheme of things, the onset of the extrinsic low-temperature region can be viewed as a transition into a regime of vastly different electronic mobility [432]. As discussed in section 3.1.2, such mobility gradients can be a source of enhanced thermoelectric effects. Indeed, the large Nernst coefficient in Co₀.₉₉₉Ni₀.₀₀₁Sb₃ can be quantitatively explained in this scenario [432].

Lessons for the skutterudite CoSb₃:

- Unreliable electronic structures: large dependence on DFT functional.
- Indications for competition of exchange and correlation effects beyond DFT.
- Need for electronic structure calculations to include defects.
- Potential path for large thermoelectric effects: engineering of mobility gradients.

---

72 At larger dopings the system becomes metallic with mobile charge carriers having effective masses enhanced by a factor of two to five [89, 90], similar to CoSb₃ with Yb-fillings [42].
3.5. Heusler compounds: a focus on Fe$_2$VAI

The Heusler class of compounds displays an enormous richness in physical phenomena, with applications in spintronics, solar cells, magneto-calorics, and thermoelectrics [121, 122]. Here, we will limit the discussion to the so-called full-Heusler structure X$_2$YZ which crystallizes in the cubic L2$_1$ structure (space group Fm3m). We again pick a representative of this subclass, that is both of fundamental and of potential technological interest: Fe$_2$VAI.

3.5.1. Experimental knowledge. The material is a semiconductor with a narrow gap of 90–130 meV [49, 50], 100–200 meV [53, 54], or 210–280 meV [51] as extracted from transport, optical, and nuclear magnetic resonance (NMR) measurements, respectively. Indeed, the resistivity of Fe$_2$VAI, figure 47(a) has a shape that seems typical for the materials investigated in this review. A closer inspection, see figure 1, reveals, however, that the resistivity increase upon lowering temperature is orders of magnitude smaller than in other materials with similar gap sizes. Indeed, despite insulating characteristics in the resistivity, photoemission spectroscopy [55, 68, 558] sees a metallic edge at the Fermi level, and — the activated behaviour of the spin-lattice relaxation rate notwithstanding—also NMR [51] suggests the presence of a residual density of states. This disparity between transport (and thermodynamic) quantities on the one hand, and spectroscopic probes on the other has been suggested to derive from off-stoichiometry [559, 560], anti-site disorder [561] or weak localization [558].

In the spin sector, Fe$_2$VAI has been characterized as paramagnetic down to lowest temperatures. Early specific heat measurements suggested metallic contributions from carriers with effective electronic masses of the order of 20–70 times the band mass [562–564]. Together this prompted the suggestion of possible 3d-orbital-based heavy fermion behaviour [562]. Subsequently, however, Lue et al [91] established that these finding were owing to the presence of extrinsic magnetic clusters [565]. Extracting the ‘intrinsic’ contribution to the electronic specific heat from measurements in a varying magnetic field, however, still yielded masses enhanced by a factor of five [91]. The intrinsic magnetic susceptibility is much flatter with no signs of towards activated behaviour up to 300 K [560, 566, 567]. NMR relaxation rates, however, do grow exponentially above 300 K [51], yet without any signs towards a crossover to Curie–Weiss-like behaviour is seen up to 600 K, see figure 2.

In the chemical phase-space—given, e.g. by the pseudo-binary alloy series (Fe$_{1-x}$V)$_x$Al or the off-stoichiometry Fe$_2$V$_{1-x}$Al$_{1-x}$—the full-Heusler composition Fe$_2$VAI is in many respects distinguished. This is illustrated in figure 48: as a function of the number of valence electrons, stoichiometric Fe$_2$VAI realizes (a) the minimal equilibrium volume, (b) the minimal specific heat, is (c) in direct proximity to ferromagnetic long range order, and (d) is placed close to thermoelectric particle-hole symmetry.

The property that both electron and hole doping increases the volume of Fe$_2$VAI is reminiscent of the Kondo insulator Ce$_3$Bi$_4$Pt$_3$ (see the discussion in section 1.2.3). However, the physical origin behind this phenomenon is very different: in Ce$_3$Bi$_4$Pt$_3$ the volume is indeed linked to the valence count [126]. In the case of Fe$_2$VAI the counter-intuitive volume decrease upon substituting iron with larger vanadium atoms has been suggested to originate from a gain in cohesive energy. Indeed, Kato et al [568] have shown that long range structural order in (Fe$_{1-x}$V)$_x$Al consistently increases when going from DO$_2$-Fe$_2$Al to the L$_2$$_1$-Fe$_2$VAI ternary Heusler composition. This finding roots in the specifics of the Heusler crystal-structure: there are two inequivalent iron sites in Fe$_2$Al that, among others, have different nearest neighbour distances. Vanadium, being larger than iron, preferentially substitutes the larger site, which gets completely filled when reaching the L$_2$$_1$ composition [569]. For excess vanadium, the smaller iron site gets populated and the lattice necessarily expands.73

That also the specific heat is minimal for Fe$_2$VAI, indicates that the ternary Heusler compound is an insulating island immersed in a metallic sea of adjacent alloy compositions. This finding is corroborated by high-temperature extrapolations of the magnetic susceptibility that find Pauli contributions to be minimal for stoichiometric Fe$_2$VAI [560]. On top of that, Fe$_2$VAI is in proximity to

---

73 In this respect the substitution with vanadium is unique. Indeed, substituting with the larger Ti the volume increases continuously, while for Cr—which has a radius only slightly larger than V—realizes a minimal volume at x ≈ 0.2 in (Fe$_{1-x}$Cr)$_x$Al [569].
ferromagnetic long-range order that is reached by, both, excess iron (e.g. for $x \geq 0.05$ in Fe$_2V_1-x$Al [104]), and excess vanadium [109]. Indeed, stoichiometric Fe$_2$VAl represents an exact zero of the Slater–Pauling function, $M = (N - 24)\mu_B$, for the spontaneous ferromagnetic moment [121, 122]. Here, $N$ is the total number of valence electrons per formula unit. Since here $N = 24$, it follows that Fe$_2$VAl is non-magnetic.

Also thermopower measurements place Fe$_2$VAl into a tight spot. Indeed, there is a tremendous sample dependence in $S$, as shown in figures 47(b) and 48. While the thermopower most of the time reaches large values, the sign of the response varies from experiment to experiment. Note that the thermopower is much more sample dependent than the resistivity shown in figure 47(a), as it is very sensitive to the balance of electron and hole contributions. As discussed in section 3.1.3 in the model context, this sensitivity suggest that intrinsic Fe$_2$VAl is placed close to thermoelectric particle-hole symmetry, for which $S$ exactly cancels but acquires very large values in the immediate vicinity. Moreover, the trend in the off-stoichiometry of the composition series Fe$_2V_1-x$Al is unconventional, see figure 48: adding in this way electrons (holes) to the system causes a more hole-like (electron-like) response, in even qualitative defiance of the rigid band picture of carrier doping. Similarly in the Hall coefficient, $R_H > 0$ for Fe$_2$VAl$_{0.05}$, while $R_H < 0$ for Fe$_2$VAl$_{0.05}$ [571]. These counter-intuitive findings may be linked again to the unconventional effects of off-stoichiometry onto the crystal structure, see figure 48(a). On the other hand, doping the Heusler compound by substitutions with a fourth element, yields (at small enough doping) the standard form of (17), as indicated by the model curve in figure 49.

The large thermopower in conjunction with the low resistivity values in Fe$_2$VAl-based systems are indicative of large thermoelectric powerfactors. Indeed, Fe$_2$VAl$_{1-x}$Si, with $x = 0.10$ shows a large (n-type) power factor of $S^2/\rho = 5.4$ mW K$^{-2}$ m$^{-1}$ in the range of $T = 300-400$ K [572]. Also Fe$_2V_1-x$Al$_{1-x}$ reaches a very favourable (n-type) powerfactor of 6.8 mW K$^{-2}$ m$^{-1}$ for $x = 0.05$ and p-type 4.2 mW K$^{-2}$ m$^{-1}$ for $x = -0.03$ [576]. These value fall into the same performance class as standard Bi$_2$Te$_3$-based thermoelectrics for which $S^2/\rho \sim 4-5$ mW K$^{-2}$ m$^{-1}$ in the same temperature range. However, the corresponding figures of merit are small: at 300 K, E.g. $ZT \approx 0.13$ for Fe$_2$VAl$_{0.5}$Si$_{0.01}$ [573], or $ZT \approx 0.14$ for Fe$_2$VAl$_{0.02}$Si$_{0.03}$ [556]. Indeed the thermal conductivity of stoichiometric Fe$_2$VAl reaches 28 W K$^{-1}$ m$^{-1}$ at 300K [573], which is an order of magnitude worse than in Bi$_2$Te$_3$.

3.5.2. Theoretical results.

Electronic structure. Standard band-structure methods characterize Fe$_2$VAl as a non-magnetic semi-metal, with a
correlation-induced changes near the Fermi level. In particular, the spectral weight at the Fermi level was shown to be strongly dependent on temperature, causing a metallization in DFT-LDA (light blue) with QSGW (green) that yields a gap of 210 meV. Left and right show different energy windows.

Figure 50. Band-structure of Fe₂VAl. Compared are metallic DFT-LDA (light blue) with QSGW (green) that yields a gap of 210 meV. Left and right show different energy windows.

hole-pocket at Γ and an electron-pocket at X and a low-density pseudogap of 100–200 meV in the DOS [562–564, 577], see the LDA results in figure 50. Improving on the exchange part in DFT, Fe₂VAl becomes a semiconductor with an indirect gap of 1.4 for vanadium states that were identified to dominate 150 meV at low temperatures and a mass-enhancement factor see figure 51, a strongly renormalized pseudogap of about 210 meV. Contrary to exchange-improved DFT (see, e.g. [578, 580], or 340–620 meV (B1-WC) [578, 579]. Therefore, we compare in figure 50 results of LDA (metallic) to that of the many-body perturbation theory QSGW75, which yields a gap of ~210 meV. Contrary to exchange-improved DFT (see, e.g. [578, 580]), the conduction states are not merely shifted upwards rigidly in QSGW, but experience a narrowing of their bandwidth, see figure 50(b). Interestingly, the effect is largest for the V-t₂g-derived band that disperses from below 2eV at the L or X-point to around 6eV at Γ.

Such band-width narrowings are a consequence of dynamical correlation effects. Treating the latter within DFT + dynamical mean-field theory, Kristanovski et al [581] recently investigated the end member compounds (x = 0, x = 1) of the Feₓ₂₋ₓVₓ₋ₓAl series76. For Fe₂VAl they found, see figure 51, a strongly renormalized pseudogap of about 150 meV at low temperatures and a mass-enhancement factor of 1.4 for vanadium states that were identified to dominate correlation-induced changes near the Fermi level. In particular, the spectral weight at the Fermi level was shown to be strongly dependent on temperature, causing a metallization with increasing temperature, reminiscent of what happens in FeSi (see section 2.4.1)77.

Disorder and magnetism. An alternative suggestion to explain that insulator-like transport properties coinciding with metallic photoemission spectrum resorts to issues of defects and disorder. Indeed, the discussed DMFT calculations build on a metallic GGA band-structure and incorporate correlation effects originating from local interactions. Longer range interaction lead, however, to significant exchange contributions in the self-energy that open a charge gap as discussed above for hybrid functionals and GW (see also the cases of FeSb₂ in section 2.2.2 and CoSb₃ in section 3.4). Starting from an insulating host, diverse types of defects and disorder introduce states inside the gap [578].

Besides changes in the density-of-states at low-energies, deviations from perfect Fe₂VAl also strongly affect magnetic properties. As mentioned above, stoichiometric Fe₂VAl verifies the Slater–Pauling rule for the magnetic moment of full-Heusler compounds [121, 122], namely M = (N – 24)μB = 0, with N = 24 the total number of valence electrons per formula unit. Deviations from the stoichiometric valence count induces a spontaneous moment in experiments (see above). This is reproduced by band-theories—based on supercell setups or the coherent potential approximation—that see ferromagnetic order induced by excess vanadium [585, 586], excess iron [586], as well as anti-site defects (V for Fe, or Fe for Al) [578]. Among point defects, anti-site substitutions were shown to have the lowest formation enthalpies [587]. In this picture, Fe₂VAl is an intrinsic semiconductor whose puzzling transport and spectral properties are dominated by extrinsic effects.

In all, the electronic-structure of Fe₂VAl is not yet properly understood. Indeed, as was the case for skutterudites (see section 3.4) the current Heusler compound presents a very challenging case, in which correlation and exchange effects beyond DFT-LDA need to be considered. Viable approaches will be mentioned in section 4. On top of this, the interplay of correlation effects and defects need to be considered, as pioneered in different contexts, e.g. by Lechermann et al [588, 589].

Thermoelectricity. From the theory side, the large powerfactors in Fe₂YZ Heusler compounds has been rationalized to originate from an effective low-dimensionality in the electronic structure [461]. By this mechanism, it was shown that

75 Which in particular does not depend on empirical parameters such as the fraction of Hartree–Fock admixtures.

76 For DMFT calculations for another compounds in the binary Fe–Al system, see Gall et al [582]; for magnetic properties of half-metallic Heusler compounds within DMFT, see, e.g. [583, 584] and references therein.

77 Kristanovski et al [581] use U = 3.36 eV and J = 0.71 eV for both Fe and V atoms, while constrained DFT yields Uₓ₋ₓ = 4 eV, Uₓ = 1.5 eV [579], and our constrained RPA for a maximally localized Wannier setup consisting of Fe-3d and V-3d orbitals gives Uₓ₋ₓ = 2.5–3.3 eV, Jₓ₋ₓ = 0.4–0.6 eV, Uₓ = 1.3–1.8 eV, Jₓ = 0.15–0.3 eV, where the ranges owe to an intra-shell orbital dependence.
bulk semiconductors can realize highly directional conduction otherwise only found in actual low-dimensional systems.

Further, computing transport properties of Fe$_2$VAl from first principles, using the semi-classical Boltzmann approach in the relaxation time approximation helped shedding some light onto the controversy of the compounds’ electronic structure. Do et al [579] computed transport properties on top of their DFT (metallic with pseudogap) and DFT + U (insulating) electronic structure. A comparison with experimental results might elucidate whether from a thermoelectric point of view the pseudogap or the fully gapped scenario is more adequate. Using the pseudogap electronic structure, no thermopower values above ±50 μV K$^{-1}$ can be realized at 300K, even when considering (rigid band) doping [579, 590, 591]. This finding is at odds with experiments for ‘stoichiometric’ Fe$_2$VAl, see figure 47(b). The gapped electronic structure can qualitatively reproduce both the magnitude and the shape of the experimental thermopower [578, 579], however only for densities that seem to be at odds with those extracted from the experimental Hall coefficient. Do et al [579] suggested that the impact of defects needs to be included for a consistent description of Fe$_2$VAl. Indeed, it was shown by Blic and Ghosez [578] and subsequently by Bandaru et al [592] that antisite defects largely impact the thermopower, and can easily flip its sign. While the powerfactor in the presence of such defects slightly decreases [578], ZT gets significantly enhanced in the simulations [592] due to a notable reduction of thermal conduction.

Lessons for the Heusler compound Fe$_2$VAl:

- Large, qualitative uncertainties in electronic structure.
- Need for treatment of both exchange and correlations beyond DFT.
- Interplay of intrinsic and extrinsic (defects, disorder) states crucial for thermoelectric properties.

4. Conclusions and outlook

In this article we have reviewed signatures of electronic correlation effects, their microscopic origin, and their impact on thermoelectricity in correlated narrow-gap semiconductors.

The salient characteristic of these systems is the large and unconventional temperature dependence in physical observables. In defiance of mere thermal activation across a fundamental gap, the most prominent members of these materials display an insulator-to-metal crossover at temperatures corresponding to energies of only a fraction of the gap size. Concomitantly, the magnetic susceptibility first grows to large values, before fluctuations disemboque into a Curie–Weiss-like decay.

In the microscopic picture emerging from realistic many-body calculations, the evolution of charge and spin degrees of freedom are intimately linked: the unlocking of spin fluctuations on transition-metal atoms cause the Brillouin zone momentum to seize being a good quantum number. The widening decomposition onto momentum-states corresponds to a diminishing lifetime of electronic excitations at a given momentum, causing incoherent spectral weight to spill into the gap, eventually filling it to realize a bad metal phase. Contrary to Mott insulators, the magnetic susceptibility is not dominated by local spin fluctuation. In fact, the main driving force of many-body effects in the considered systems was identified to be the Hund’s rule coupling $^{78}$ entailing fluctuations to have ferromagnetic characteristics. The interplay of the Hund’s rule coupling and the hybridization gap was shown to explain the large variation of correlation signatures witnessed among various materials.

New realistic many-body simulations for Ce$_3$Bi$_2$Pt$_3$ and the study of the Kondo-insulating nature of its $J = 5/2$ states allowed for an in-depth comparison of heavy-fermion and d-electron-based semiconductors. Based on a quite general classification of paramagnetic insulators from the dynamical mean-field perspective, we suggest an orbital-selective scenario for FeSi: while some iron-3d orbitals are Kondo-insulating, in others quantum fluctuations are suppressed by a more conventional gapping of non-local hybridizations.

As to the influence of correlation effects onto thermoelectricity, details matter, but simplistically one could summarize that (selective) mass enhancements (via the real-parts of the self-energy) are beneficial for a large thermopower, while (global) incoherence (described by the imaginary parts of the self-energy) is detrimental. We pointed hence out that many-body renormalizations that increase the particle-hole fluctuations on transition-metal atoms cause the Brillouin zone corresponding to energies of only a fraction of the gap size. Weiss–Concomitantly, the magnetic susceptibility first grows to large values, before fluctuations disemboque into a Curie–Weiss-like decay.

In the microscopic picture emerging from realistic many-body calculations, the evolution of charge and spin degrees of freedom are intimately linked: the unlocking of spin fluctuations on transition-metal atoms cause the Brillouin zone momentum to seize being a good quantum number. The widening decomposition onto momentum-states corresponds to a diminishing lifetime of electronic excitations at a given momentum, causing incoherent spectral weight to spill into the gap, eventually filling it to realize a bad metal phase. Contrary to Mott insulators, the magnetic susceptibility is not dominated by local spin fluctuation. In fact, the main driving force of many-body effects in the considered systems was identified to be the Hund’s rule coupling $^{78}$ entailing fluctuations to have ferromagnetic characteristics. The interplay of the Hund’s rule coupling and the hybridization gap was shown to explain the large variation of correlation signatures witnessed among various materials.

As to the influence of correlation effects onto thermoelectricity, details matter, but simplistically one could summarize that (selective) mass enhancements (via the real-parts of the self-energy) are beneficial for a large thermopower, while (global) incoherence (described by the imaginary parts of the self-energy) is detrimental. We pointed hence out that many-body renormalizations that increase the particle-hole asymmetry are to be looked for, while coherence (excitation lifetimes) could be increased by diluting the atoms that host the dominant Hund’s rule coupling.

We further elaborated on new paths to exploit electronic correlations in thermoelectrics: in one proposal many-body effects have been suggested for the design of mobility gradi-
dominating unwanted thermal conduction, heralding a new
direction in the phonon-engineering route towards high-per-
performance thermoelectrics.

The faithful simulation of thermoelectric properties of the
silicides, antimonides, skutterudites, and Heusler compounds
discussed here, exposed several challenges for the future.

(i) Electronic structure: in all materials discussed here
effects of defects and disorder are a particular
concern, especially when doping the stoichiometric
parent compounds. Accounting for these effects in the
presence of electronic correlation effects is a challenging,
yet necessary endeavour for the future.

(ii) Defects and disorder: as in more conventional semicon-
ductors, effects of defects and disorder are a particular
concern, especially when doping the stoichiometric
parent compounds. Accounting for these effects in the
presence of electronic correlation effects is a challenging,
yet necessary endeavour for the future.

(iii) Electron–phonon effects: future studies must address
the impact of electronic correlations onto the phonon-
structure, the interplay of electronic incoherence and
thermal-disorder, as well as the phonon-drag effect.

Acknowledgments

The author gratefully acknowledges discussions on topics
related to this review with H Aoki, M Aronson, M Battiato, S
Bühler-Paschen, O Delaire, A Galler, M Gamza, A Georges,
K Haule, K Held, G Kotliar, J Kuněš, F Lechermann, G K H
Madsen, E Maggio, J Mravlje, Hyowon Park, C Petrovic, J
Pflaum, A Prokofiev, G Sangiovanni, F Steglich, Peijie Sun, P
Tomes, A Toschi, Wenhu Xu, I Zaliznyak, V Zlatić, Z Zhong,
as well as M Battiato, S Bühler-Paschen, K Held, J Kuněš for
comments on parts of the manuscript. We thank Dr A Hausoel
for providing pre-release functionalities and support of the
w2dynamics package. This work has been supported by the
Austrian Science Fund (FWF) through project ‘LinReTraCe’ P
30213-N36. Some calculations were performed on the Vienna
Scientific Cluster (VSC).

ORCID iDs

Jan M Tomczak  https://orcid.org/0000-0003-1581-8799

References

[1] Zhang L et al 2015 Nat. Mater. 15 204
[2] Ngai J, Walker F and Ahn C 2014 Ann. Rev. Mater. Res. 44 1–17
[3] Schweritz R, Zabko P, Lezama I G, Ono S, Morpurgo A F, 
Catalan G and Triscone J M 2010 Adv. Mater. 22 5517–20
[4] Zhong Z, Wallerberger M, Tomczak J M, Taranto C, 
Parragh N, Toschi A, Sangiovanni G and Held K 2015 
Phys. Rev. Lett. 114 246401
[5] Moyer J A, Eaton C and Engel-Herbert R 2013 Adv. Mater. 
25 3578–82
[6] Tomczak J M, Prouvost L V, Vaugier L, Georges A and 
Biermann S 2013 Proc. Natl Acad. Sci. USA 110 904–7
[7] Li S Y, Niklasson G and Granqvist C 2012 Thin Solid Films 
520 3823–8
[8] Tomczak J M and Biermann S 2009 Phys. Status Solidi B 
246 1996 (Scientific Highlight of the Month of the Ψt 
Network, no. 88, August 2008)
[9] Schlesinger Z, Fisk Z, Zhang H T, Maple M B, DiTusa J and
Aeppli G 1993 Phys. Rev. Lett. 71 1748–51
[10] Bocelli S, Marabelli F, Spolenak R and Bauer E 1996 Mat. 
Res. Soc. Symp. Proc. 402 361
[11] Paschen S, Felder E, Chernikov M A, Degiorgi L, Schwer H, 
Ott H R, Young D P, Sarrao J L and Fisk Z 1997 Phys. Rev. 
B 56 12916–30
[12] Jaccarino V, Wertheim G K, Wernick J H, Walker L R and 
Arajs S 1967 Phys. Rev. 160 476–82
[13] Takagi S, Yasuoka H, Ogawa S and Wernick J H 1981 
Phys. Soc. Japan 50 2539–46
[14] Wolfe R, Wernick J H and Haszko S E 1965 Phys. Lett. 
19 449–50
[15] Damascelli A, Schulte K, van der Marel D, Fisk Z, Zhang H T, 
Maple M B, DiTusa J and Aeppli G 1993 Phys. Rev. 
B 56 12916–30
[16] Schlesinger Z, Fisk Z, Zhang H T, Maple M B, DiTusa J and 
Aeppli G 1993 Phys. Rev. Lett. 71 1748–51
