Advances in half-Heusler alloys for thermoelectric power generation

Robert J. Quinn and Jan-Willem G. Bos

Half-Heusler alloys are leading candidate materials for large scale thermoelectric power generation. In the past decade, substantial improvements have been made to the established n-types based on $X_4\text{NiSn}$, and exciting new p-types based on $X_5\text{FeSb}$ and $\text{ZrCoBi}$ have been discovered. This has been achieved through careful experimental work and by increased fundamental insight. In particular, advances in materials synthesis and compositional optimisation guided by empirical modelling have been key to extracting more performance from half-Heuslers alloys. This review summarises recent progress for the main classes of half-Heusler alloys, including their fundamental electronic and thermal properties, processing, and thermoelectric performance. It also incorporates some of the expanding body of work focused on translating materials performance ($zT > 1$ at 773 K is now routinely possible) into generator devices.

1. Introduction

The world is currently undergoing a transition from an energy economy based on non-renewable fossil fuels to a system making use of a range of renewable generators including solar panels, wind turbines and hydroelectric plants, supplemented by nuclear energy. While major changes are taking place in terms of primary energy generation, work on improving the overall efficiency of the world’s energy infrastructure is important, which can be achieved by improving electricity transmission, improving device efficiency, and recovering waste energy lost as heat. Thermoelectric generators (TEG’s) turn a temperature gradient into an electrical current, allowing waste heat to converted back into electricity. The widespread application of TEG’s would lead to a reduction in the energy production required by primary generators. TEG’s can also be used as primary generators, in scenarios where maintenance is difficult, or a long lifetime is required.

TEG’s have temperature ranges in which they operate most effectively, due to the underlying material properties. The only commercial TEGs based on $\text{Bi}_2\text{Te}_3$ for example operate well from 300–500 K, but their effectiveness begins to decrease at higher temperatures. For application involving higher temperatures a different material is required, such as PbTe and GeTe used in the Mars Perseverance rover. However, for large scale terrestrial applications, other requirements such as elemental abundance...
and cost are important.9,10 Other factors are scalability of materials processing, mechanical properties, temperature stability, matching n- and p-type compositions and ease of electrical and thermal contacting within the TEG devices.4

Of all investigated bulk thermoelectric materials, the half-Heusler (HH) alloys hit most of these target criteria.11–16 In terms of outright performance, they are characterised by good electrical properties, making them suitable for power generation application, and are only heldback by relatively high lattice thermal conductivities, which limit the overall heat to electricity conversion efficiency. Due to their good compromise between engineering properties and TE performance, HH alloys have received significant attention as potential thermoelectric materials. Initial investigations into these materials started in the late 1990s and identified n-type and p-type phases based on XnNiSn and XnCoSb (Xn = Ti, Zr and Hf) as the most promising compositions.17–29 However, the difficulty in getting reproducible results and pushing zT above one meant that other materials systems, in particular skutterudites and tellurides gained more prominence.6,7

For designing and quantifying the potential of a thermoelectric material, the figure of merit (zT) is used:3

\[
zT = \frac{S^2 \sigma}{k} \quad (1)
\]

where \(S\) is the Seebeck coefficient, \(\sigma\) is the electrical conductivity and \(k\) is the total thermal conductivity. \(zT\) is maximised when electrical transport is high (large power factor \(S^2\sigma\)) and when thermal transport is low (small \(k\)). \(k\) is made up of a lattice component \(k_l\) due to vibrations, an electronic component \(k_{el}\) due to heat carried by the charge carriers and in some cases a bipolar term \(k_{el,0}\) for semiconducting systems. Within a given material, optimisation of \(zT\) involves adjusting the carrier concentration to find the optimal balance between a maximal \(S^2\sigma\) and minimal \(k_{el}\), whilst atomic substitutions (alloying) or nanostructuring can be used to reduce \(k_l\).6

The performance of a material is fundamentally determined by its electronic and phonon bandstructures. Another metric which describes this balance is the quality factor (\(B\)), originally established in works by Goldsmid, Chasmar and Stratton:30–32

\[
B \propto \frac{\mu_w}{k_l} \quad (2)
\]

where \(\mu_w\) is the weighted carrier mobility. In its simplest case, this is given by \(\mu_w = \mu_0 n^{3/2}\), where \(\mu_0\) is the carrier mobility in the limit of zero carrier concentration and \(n^{3/2}\) is the density of states effective mass. \(B\) gives a measure of the quality of a thermoelectric material independent of the requirement to optimise the carrier concentration \(n\).30

Having a large \(zT\) is not the only metric that needs to be considered when selecting a good thermoelectric material, and it has been noted that a reduction in \(zT\) to improve the power output can be beneficial.33–35 A singular focus on \(zT\) ignores the importance of \(S^2\sigma\) for power generation, and on this metric the HH alloys are leading. As discussed below the HH alloys have intrinsic fairly large \(k_l\) values, and in terms of the quality factor, \(B\), their good performance derives from the large \(\mu_w\). This contrasts with many chalcogenides where high \(B\) values are driven by glass-like \(k_l\) values.6

Up to 2014, the main research focus was on developing n-type materials based on XnNiSn and p-type materials based on XnCoSb. This involved developing strategies for alloying and carrier doping and resulted in \(zT\) \(\sim 1\) and \(zT\) \(\sim 0.7\) at 773 K for the n-types and p-types respectively.14 Grain size reduction through high energy ball milling was also established as a route to lower \(k_l\) but was found to require small average grain sizes 0.2–0.3 µm to be effective.14,36,37 The other route for optimisation that gained interest in the early 2010s was the use of phase segregation to embed favourable micro- and nanostructures for reduction of \(k_l\). This focused in particular on segregation of X-site elements (Ti, Zr and Hf)38–40 and excess metals, leading to incorporation of full-Heusler domains.41–44

Recent advances include the rise of p-type compositions XnFeSb (Xn = V, Nb and Ta) and ZrCoBi, each with \(zT > 1\) at 773 K.46–48 It also includes the widespread use of single parabolic band (SPB) and Callaway modelling,15 coupled to first principles calculations, which have placed understanding of the HH alloys on a much firmer footing. This manuscript will review these main recent developments and will conclude with a brief overview of the integration of HH materials into working TEGs, which is attracting more and more interest.

This review is structured into 3 main sections. Section 2 will discuss general structural and processing aspects, and fundamental electronic and thermal properties of HH alloys. Section 3 gives an overview of key recent materials developments for the main classes of HH alloys. Section 4 gives a brief overview of incorporation into TEG devices.

2. General trends in half-Heusler materials

This section will describe some of the general trends in HH materials. This is focused on crystal structure and compositions, an overview of observed \(zT\) and temperature-averaged \(zT_{av}\) values and general trends in electronic and thermal properties of these materials.

2.1. Crystal structure

HH alloys have a typical composition XYZ, where X is an early transition metal (Sc, Ti, V groups), Y is a late transition metal (Fe, Co, Ni groups) and Z is a heavy main group metal (typically Sn or Sb, occasionally Ge, Pb or Bi).39 The structure can be described as a combination of zinc blende (YZ) and rock-salt (XZ) structures, with a \(F3\overline{1}m\) (216) cubic space group, as shown in Fig. 1.19,50 HH materials with a valence electron count of 18 are semiconducting, as each element has either a filled or empty valence shell. For example in TiNiSn, the formal oxidation states are Ti\(^{+4}\) (3d\(^{0}\)), Ni\(^{+3}\) (3d\(^{6}\)) and Sn\(^{+4}\) (5s\(^{2}5p\(^{6}\)) and in NbFeSb, they are Nb\(^{+3}\) (4d\(^{9}\)), Fe\(^{+2}\) (3d\(^{9}\)) and Sb\(^{+3}\) (5s\(^{2}5p\(^{5}\)). Hence, each element has an empty/full shell and the oxidation states are balanced, signalling a stable composition.49–51 Similarly, one can
consider the electronic structure to be related to III–V semiconductors, with an 18 electron \([\text{YZ}]^{2-}\) zinc blende sublattice charge balanced by the electropositive metal \(X^{+}\). This electron counting does not fully guarantee a semiconducting state, with HH phases and with higher average atomic number sometimes found to be semimetallic (e.g. HfCoSb), due to the closing of the band gap. It should also be noted that d-orbitals from the X and Y metals form the dominant contributions to the valence and conduction band, so the simple Zintl-model, with separate ionic \(X^n+\) and covalent \([\text{YZ}]^{2-}\) substructures, does not fully capture the chemical bonding. First principles calculations confirm that the crystal structure is stabilised by strong covalent bonding in the YZ sublattice but also by stabilisation linked to the “ionic” XZ rock-salt sublattice, where (intuitively) site inversion would carry a large electrostatic penalty. Defect energy calculations confirm this picture with XZ and Y/Z site inversion being costly, and the lowest energy defect generally being occupancy of the interstitial tetrahedral site. Full occupancy of the interstitial site leads to the \(\text{XY}_2\text{Z}\) full-Heusler structure as shown in Fig. 1. As indicated by defect energy calculations, inversion of X and Y is uncommon (e.g. <1% in \(\text{V}_{0.8}\text{Co}_0\text{Sb}\)) but there is evidence that this occurs for \(\text{V}_{0.8}\text{Co}_0\text{Sb}\) and \(\text{VFeSb}\), where V/Fe and V/Co are (almost) adjacent in the periodic table and hence likely to have similar site preferences.

2.2. Synthesis

HH materials are intermetallic compounds and are usually prepared from elemental precursors, which span a range of melting points, e.g. 1855 °C for Zr and 232 °C for Sn. Given the presence of refractory elements, melting has historically been the favoured route for initial reaction. Most frequently this is done by arc-melting of elemental pieces in an argon atmosphere, although levitation melting has become more common in recent years. The main advantage of levitation melting is that the sample does not touch the sample environment (usually a Cu plate in arc melting) and the presence of eddy currents that improve the homogeneity of the melt. Arc-melting has been noted to contain different microstructures in different sections of the final ingot. A linked practical advantage of melt-based routes is that larger elemental pieces can be used avoiding issues with surface oxidation that can be significant in powder metallurgy. Nevertheless, powder-based routes can give good performing samples, but care has to be taken with precursors, and the lower temperatures also place limitations on grain growth. Other means of achieving reaction include microwave synthesis and self-propagating combustion synthesis.

After initial reaction it is common to anneal the sample, typically between 900–1000 °C, to equilibrate the sample and increase homogeneity. In the past couple of years however, what may be called ‘out-of-equilibrium’ processing has gained much prominence. Here, melting is followed by ball milling and rapid consolidation at high temperature and under applied pressure, often using spark plasma sintering (SPS), but no long duration heating is used. This can be a favourable route that maintains embedded nanostructures that would be annealed out at high temperatures. In some instances, most notably for ZrCoBi, the melting step is removed altogether, leading to an efficient synthetic protocol. Compared to the initial literature of the early 2000s, ball milling after HH phase formation has gained prominence and is now nearly always used. Reasons include promoting sample homogeneity and achieving a good particle size distribution that facilitates consolidation into dense ingots, whilst very small grain sizes are also beneficial for thermal conductivity. Using direct mechanical alloying of elemental precursors has been found to narrow the grain size distribution in a range of HH materials, with HH and FF alloys forming during the milling stage. Ball milling needs to be carried out under an inert atmosphere, including the handling of the resulting fine powders, to protect the high surface area samples from oxidation.

Whether or not there is an annealing stage, all samples need to be in a highly dense form before property measurement. While samples can be cut from a melted piece with high density, it is more common for the powdered sample to be consolidated using either hot-pressing or SPS, with much of the improvement in thermoelectric performance in recent years arising due to more careful consolidation of samples. In particular, SPS has become more prominent and enables very rapid densification. In conventional hot pressing, uniaxial pressure is applied to the sample in a graphite die which is
2.3. Electronic transport properties

HH thermoelectric materials are small band gap semiconductors, with calculated band gaps typically ranging between 0.5–1 eV. The conduction band minimum typically lies at the X-point, while the valence band will either exist at the Γ-point or the L-point, and in some cases, both can be effectively degenerate, as illustrated for the ZrCoBi system (Fig. 2). In addition, the W-point can sometimes be close to the other valence band maxima. From this perspective, p-type materials have a larger potential for band-engineering to improve the thermoelectric properties, as degenerate band extrema allow for an increase in the density of states effective mass ($m^*_{\text{DOS}}$):

$$m^*_{\text{DOS}} = (N_v)^2 m^*_b \left( \frac{N_v}{\sum_i N_i N'_i} \right)$$

(3)

where $N_v$ is the total carrier pocket degeneracy (where $N_0$ is the number of bands contributing and $N_v$ is the number of symmetry related pockets in the first Brillouin zone) in the electronic bandstructure and $m^*_b$ is the band effective mass. Paradoxically, for good electronic performance a large $m^*_{\text{DOS}}$ is desirable for large $S$, but a low $m^*_b$ is desirable for high $\sigma$. On balance a smaller $m^*_b$ is better overall, yielding large $\mu^*_{\text{w}}$.

As shown in Table 1, $m^*_{\text{DOS}}$ for HH materials can be quite large, in some cases exceeding 10 $m^*_b$, with $m^*_b$ typically between 1.5–3.0 $m^*_e$, which requires large amounts of doping to reach optimal $n_H$ up to orders of magnitude of $10^{20}$–$10^{21}$ cm$^{-3}$. HH materials typically have $\mu_H$ on the order of 10 cm$^2$ V$^{-1}$ s$^{-1}$, with examples of single crystal ZrNiSn as large as 70 cm$^2$ V$^{-1}$ s$^{-1}$ and Ta$_{0.81}$CoSb with $\mu_b$ as low as 1.4 cm$^2$ V$^{-1}$ s$^{-1}$.86,87 Here, $n_H$ and $\mu_H$ are the carrier concentration and mobility from Hall measurements.

As collection of Hall data has become more frequent, so have detailed investigations into the dominant carrier scattering mechanisms. For semiconductors at low doping concentrations ($n_H < 10^{17}$ cm$^{-3}$) $\mu_H$ is expected to be limited by ionised impurity (II) scattering, which has a characteristic $\mu_H \sim T^{-1.5}$ behaviour.88 At high doping concentration ($n_H > 10^{18}$ cm$^{-3}$), II scattering is minimal due to screening from the high number of charge carriers, and it is common to find scattering by phonons (acoustic phonon scattering; APS) to be dominant. This has a characteristic $\mu_H \sim T^{-1.5}$ temperature dependence for semiconductors and $\mu_H \sim T^{-1}$ in the metallic limit.88 A final intrinsic effect is alloy scattering (AS) due to size difference between alloying elements, with similar $n_H$ dependence to APS but a different characteristic $\mu_H \sim T^{-0.5}$ dependence.89,90 AS has been identified as the dominant carrier scattering mechanism in a number of HH phases,91,92 with recent work indicating that this can be minimised by using alloying elements with similar sizes (minimising lattice strain).93 In other systems, the data suggests a dominant APS dependence, with the large $S^2\sigma$ in HH materials attributed to a fundamentally weak coupling of electrons with acoustic phonons, leading to high $\mu$ in spite of large $m^*_{\text{DOS}}$.94

Electronic transport can also be affected by “extrinsic” scattering contributions arising from the materials microstructure.95

In particular, the impact of grain boundary (GB) resistances has attracted much recent interest.96–98 The most successful electronic transport models treat these materials as two-phase composites with bulk and GB phase contributions. The GB phase is usually modelled as a tunnel barrier (due to charge depletion in the GB region), accurately reproducing $\sigma(T)$
in a range of thermoelectrics, including Mg$_2$Sb$_2$ and SrTiO$_3$.96,97

GB scattering is particularly relevant for p-type NbFeSb with a very sensitive processing dependence of $S^2\sigma$, caused by changes in $\sigma(T)^{98}$ (also Section 3.1). It has also been observed in p-type ZrCoSb, where GB scattering is found to be substantial for low $n_H$ and small grain sizes (Section 3.4).100

The recent availability of ZrNiSn single crystals has enabled in-depth investigation of the electronic properties. This includes collection of ARPES data on ZrNiSn that revealed a large intrinsic band gap ~0.66 eV, comparable to DFT values (~0.5 eV) and unlike the ~0.15 eV values usually found from transport data due to in-gap states resulting from interstitial Ni.101 A comparison of single crystal and polycrystalline ZrNiSn$_{1-x}$Sb$_x$ samples reveals a complex carrier scattering phase diagram. With increasing $n_H$, a crossover from II to GB scattering to APS and AS occurs near 300 K, whilst at high temperature the transition is from polar optical (PO) phonon scattering to APS and AS. These transitions occur near $n_H = 3-5 \times 10^{20}$ cm$^{-3}$ and are a consequence of the effective screening of II, GB and PO scattering at higher carrier concentrations.87

2.4. Thermal transport properties

Compared to other state-of-the-art thermoelectric materials, which benefit from ultralow (~< 1 W m$^{-1}$ K$^{-1}$) $\kappa$, the HH family of materials typically have large $S^2\sigma$ and relatively high $\kappa$ values. This is partly due to large $\kappa_1$ as outlined above, but also due to large electronic ($\kappa_{el}$) values, consistent with the large reported $S^2\sigma$ values of these materials. Within the HH thermoelectric literature, experimental $\kappa(T)$ data is typically treated using the Debye–Callaway model to describe the temperature dependence, with $\kappa_1(T)$ defined as:91,104,105

$$\kappa_1 = \frac{k_B}{2\pi v_s} \left( \frac{k_B T}{\hbar} \right)^3 \int_{0}^{\infty} \frac{x^4 e^x}{\tau_{ph} x (e^{x} - 1)} \, dx \quad (x = \frac{\hbar \omega}{k_B T}) \tag{4}$$

where $v_s$ is the velocity of sound, $\theta_B$ is the Debye temperature and $\tau_{ph}^{-1}$ is the total phonon scattering rate. Some publications also make use of the simplified Slack model.104,106

Table 2: Overview of 300 K velocity of sound data for well-studied HH compositions. Here, $a_{HH}$ is the lattice parameter, $v_i$ and $v_{any}$ are the longitudinal, transverse, and average velocities of sound, $\Theta_B$ is the Debye temperature and $\gamma$ is the Grüneisen parameter. The latter two were obtained from $v_i$ and $v_{any}$ data as outlined in Appendix A. $\kappa_{mklapp}^{bulk}$ is the 300 K lattice thermal conductivity for defect-free compositions, calculated using the Debye–Callaway model and assuming only Umklapp scattering. In many cases, experimental acoustic velocity of sound data is available (AC), other sources include calculations (DFT, materials project, MP) and for HfNiSn from mechanical measurements.

<table>
<thead>
<tr>
<th>Composition</th>
<th>Source</th>
<th>$a_{HH}$ (Å)</th>
<th>$v_i$ (m s$^{-1}$)</th>
<th>$v_{any}$ (m s$^{-1}$)</th>
<th>$v_s$ (m s$^{-1}$)</th>
<th>$\Theta_B$ (K)</th>
<th>$\gamma$</th>
<th>$\kappa_{mklapp}^{bulk}$ (W m$^{-1}$ K$^{-1}$)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFeSb</td>
<td>DFT</td>
<td>5.769</td>
<td>5773</td>
<td>2974</td>
<td>3330</td>
<td>393</td>
<td>1.90</td>
<td>19.7</td>
<td>236</td>
</tr>
<tr>
<td>NbFeSb</td>
<td>AC</td>
<td>5.950</td>
<td>5597</td>
<td>3099</td>
<td>3452</td>
<td>396</td>
<td>1.65</td>
<td>32.9</td>
<td>123</td>
</tr>
<tr>
<td>TaFeSb</td>
<td>AC</td>
<td>5.938</td>
<td>—</td>
<td>—</td>
<td>2907</td>
<td>334</td>
<td>—</td>
<td>—</td>
<td>47</td>
</tr>
<tr>
<td>TaCoBi</td>
<td>DFT</td>
<td>6.033</td>
<td>4775</td>
<td>2736</td>
<td>3039</td>
<td>343</td>
<td>1.53</td>
<td>29.9</td>
<td>113</td>
</tr>
<tr>
<td>ZrCoBi</td>
<td>AC</td>
<td>6.188</td>
<td>4613</td>
<td>2558</td>
<td>2849</td>
<td>314</td>
<td>1.64</td>
<td>33.4</td>
<td>48</td>
</tr>
<tr>
<td>HfCoBi</td>
<td>DFT</td>
<td>6.194</td>
<td>4155</td>
<td>2335</td>
<td>2617</td>
<td>288</td>
<td>1.57</td>
<td>24.6</td>
<td>125</td>
</tr>
<tr>
<td>TiCoBi</td>
<td>AC</td>
<td>5.882</td>
<td>5691</td>
<td>3230</td>
<td>3591</td>
<td>416</td>
<td>1.56</td>
<td>35.6</td>
<td>141</td>
</tr>
<tr>
<td>ZrCoSb</td>
<td>AC</td>
<td>6.070</td>
<td>4588</td>
<td>3144</td>
<td>3493</td>
<td>392</td>
<td>1.53</td>
<td>37.9</td>
<td>141</td>
</tr>
<tr>
<td>HfCoSb</td>
<td>AC</td>
<td>6.039</td>
<td>4703</td>
<td>2709</td>
<td>3088</td>
<td>340</td>
<td>1.51</td>
<td>34.1</td>
<td>141</td>
</tr>
<tr>
<td>TiNiSn</td>
<td>AC</td>
<td>5.930</td>
<td>5426</td>
<td>3062</td>
<td>3406</td>
<td>392</td>
<td>1.58</td>
<td>29.0</td>
<td>237</td>
</tr>
<tr>
<td>ZrNiSn</td>
<td>AC</td>
<td>6.110</td>
<td>5498</td>
<td>3149</td>
<td>3498</td>
<td>390</td>
<td>1.53</td>
<td>37.4</td>
<td>237</td>
</tr>
<tr>
<td>HfNiSn</td>
<td>Mech.</td>
<td>6.080</td>
<td>4191</td>
<td>2503</td>
<td>2770</td>
<td>311</td>
<td>1.38</td>
<td>31.4</td>
<td>238</td>
</tr>
<tr>
<td>V$_{0.9+}$CoSb</td>
<td>AC</td>
<td>5.795</td>
<td>5437</td>
<td>2902</td>
<td>3242</td>
<td>381</td>
<td>1.78</td>
<td>21.0</td>
<td>191</td>
</tr>
<tr>
<td>Nb$_{0.8}$CoSb</td>
<td>AC</td>
<td>5.902</td>
<td>5581</td>
<td>2987</td>
<td>3336</td>
<td>385</td>
<td>1.77</td>
<td>26.2</td>
<td>183</td>
</tr>
<tr>
<td>Ta$_{0.1}$CoSb</td>
<td>MP-31459</td>
<td>5.960</td>
<td>4650</td>
<td>2323</td>
<td>2607</td>
<td>297</td>
<td>1.00</td>
<td>12.8</td>
<td>239 and 240</td>
</tr>
<tr>
<td>NbCoSn</td>
<td>DFT</td>
<td>5.946</td>
<td>5625</td>
<td>3016</td>
<td>3368</td>
<td>386.09</td>
<td>1.762</td>
<td>26.7</td>
<td>196</td>
</tr>
<tr>
<td>TaCoSn</td>
<td>DFT</td>
<td>5.948</td>
<td>4955</td>
<td>2761</td>
<td>3077</td>
<td>352.30</td>
<td>1.6268</td>
<td>31.6</td>
<td>196</td>
</tr>
</tbody>
</table>
on calculating $k_t$ from the phonon bandstructure exists,\textsuperscript{107-113} it is more typical to use approximations based on calculated elastic properties or experimental acoustic properties to estimate $\nu_s$, $\theta_l$, and $\tau_{ph}$. The most fundamental description of $\tau_{ph}$ considers Umklapp scattering, which can be written as:\textsuperscript{105}

$$\frac{1}{\tau_U} = 2\gamma k_B T \frac{G\sigma_D}{\nu_0}$$

where $\gamma$ is the Grüneisen anharmonicity parameter, $G$ is the Shear modulus and $\nu_0$ is the volume per atom. Assuming only Umklapp scattering and grain boundary scattering equivalent to grain size of 100 $\mu$m, $k_t$ (300 K) for HH materials can be calculated to be around 20–40 W m$^{-1}$ K$^{-1}$ as shown in Table 2 and Fig. 3. Experimental velocity of sound values obtained directly from acoustic measurements have been used where possible and used to calculate $\gamma$ and $G$ (see Appendix). The volume per atom has been calculated using the reported lattice parameter ($a_{lat}$).

As the scattering rate, $\tau_{ph}^{-1}$ increases, $k_t$ is expected to decrease, which is illustrated well in Fig. 3a, with the outliers being ZrCoBi and HfCoSb due to their low $\theta_l$. TaFeSb also has low $\theta_l$ and is another outlier, but is not included in Fig. 3, since only the average velocity of sound was reported.\textsuperscript{47} In the plot of $k_t$ vs. $\gamma$ in Fig. 3b, there is a larger spread due to lower $\nu_s$ for TiNiSn and $\nu_{Bi2-1}CoSb$ (yielding a larger $\gamma$) versus their second row analogues. This is ultimately corrected for by a larger $G$, giving a relatively smooth dependence in Fig. 3a. Naturally these values only represent an upper limit for $k_t$ and the measured experimental values don’t get close to 40 W m$^{-1}$ K$^{-1}$. Within the Debye–Callaway model, phonon scattering can be expected to arise due to grain boundaries, the existence of point defects and electron–phonon interactions. Even small effects can have a large effect on $k_t$. In ZrNiSb for example, introducing a point defect scattering term $\Gamma = 0.02$ is enough to reduce $k_t$ from 37 W m$^{-1}$ K$^{-1}$ to 17 W m$^{-1}$ K$^{-1}$, which is the value expected from a 2% Hf impurity (which is typical in commercial elemental Zr). Small amounts of interstitials are also sufficient to embed this amount point defect disorder and have been linked to strongly suppressed $k_t$ values in $\nu_0$NiSn and $\nu_0$CoSb systems.\textsuperscript{63,114}

In addition to $k_t$, HH thermoelectrics usually have significant $\kappa_{el}$ due to the large carrier concentrations and bipolar transport at high temperatures. The total $k_{el}$ is given by:\textsuperscript{115}

$$k_{el} = (L_n\sigma_n + L_p\sigma_p) T + \frac{\sigma_n\sigma_p}{\sigma_n + \sigma_p} \left( S_p - S_n \right)^2 T$$

Here, the first term is the Wiedemann–Franz ($\kappa_{WF}$) electronic thermal conduction due to the flow of electron ($\sigma_n$) and hole ($\sigma_p$) currents, which simplifies to $\kappa_{WF} = LnT$ if the Lorenz numbers ($Ln$) for electrons and holes are identical.\textsuperscript{115} The second term is the bipolar thermal conductivity ($\kappa_{bi}$), which is linked to carrier recombination due to the presence of opposing electron and hole currents.\textsuperscript{115} Here $S_{n/p}$ are the partial Seebeck coefficients of the n- and p-type carriers. The measured $S$ is given by:\textsuperscript{115}

$$S = \frac{S_n\sigma_n + S_p\sigma_p}{\sigma_n + \sigma_p}$$

$\kappa_{bi}$ can be significant at high temperatures for materials with small band gaps, where a large number of minority carriers can be easily excited. From eqn (7), this also leads to a simultaneous degradation of $S$, leading to a rapid drop in the thermoelectric figure of merit, $zT$. In the context of the HH thermoelectrics, this situation applies to the n-type $\nu_0$NiSn materials, which have small band gaps, $E_g \sim 0.5$ eV from DFT calculations and $\sim 0.15$ eV from transport and optical measurements.\textsuperscript{116} Most p-types have much larger band gaps and are not strongly affected.

### 2.5. Overview of materials efficiency: $zT$ and $zT_{avg}$

The most commonly used metric to quantify the thermoelectric performance of a single material is the materials figure of merit ($zT$). A larger $zT$ is expected to translate into higher efficiency in

![Fig. 3](https://example.com/fig3.png)

**Fig. 3** Plots of calculated lattice thermal conductivity ($k_t$) at 300 K (see Section 2.4), assuming the presence of only Umklapp scattering (i.e. for the defect free parent materials), versus (a) the variables determining the magnitude of Umklapp scattering and (b) the acoustic Grüneisen parameter. HfCoSb and ZrCoBi have lower $k_t$ due to low $\nu_s$ (see Table 2).
a thermoelectric device. However, a more appropriate metric of performance is the temperature average of $zT$ over the gradient the TEG operates under. For this review we have used the following equation:\textsuperscript{117,118}

$$zT_{\text{avg}} = \frac{\int_{T_C}^{T_H} zTdT}{T_H - T_C}$$  \hspace{1cm} (8)

Fig. 4 shows $zT$ and $zT_{\text{avg}}$ for some of the most promising HH materials found in the literature. The best n-type materials based on $X_{\mu}\text{NiSn}$ achieve large $zT$ at 600–900 K (Fig. 4a), reaching a peak due to a gradual decline in $S(T)$ and increase in $\kappa(T)$ at high temperatures. The n-type $X_{\mu}\text{CoSb}$ and $X_{\mu,\alpha}\text{CoSb}$ examples shown have a lower $zT$, but they maintain a growth in $zT$ up to 1100 K, potentially making them more suitable at higher temperatures than the $X_{\mu}\text{NiSn}$ systems.

The best p-type samples all show a gradual increase in $zT$ at least up to 1200 K, at their best achieving $zT \sim 1.5$. The overall lower $zT$ for the p-types means that where $zT_{\text{avg}}$ would approach 1 at 800 K for the best n-type materials, it takes until 1200 K to reach similar $zT_{\text{avg}}$ for p-type HH phases. From Fig. 4c and d, it is clear that the p-type $X_{\mu}\text{CoSb}$ materials have been overtaken rapidly by the new wave of $\text{ZrCoBi}$ and $X_{\mu}\text{FeSb}$ systems.

3. Recent developments in half-Heusler materials

In the following an overview of the main recent materials development will be given, starting with the new p-type materials, followed by good n-types and then other compositions of interest.

3.1. $X_{\mu}\text{FeSb}$ ($X_{\mu} = V, Nb$ and $Ta$) (p-type)

In 2014, promising p-type thermoelectric properties were reported for ($V_{0.6}Nb_{0.4})_{1-x}Ti_{x}\text{FeSb}$, with moderately large $S^2$ around 3 mW m$^{-1}$ K$^{-2}$ and quite small $\kappa_1$ below 3 W m$^{-1}$ K$^{-1}$, yielding peak $zT_{1200K} = 0.8$ for $x = 0.2$.\textsuperscript{119} The good performance in this series was attributed to large $m_{CH} = 10m_e$ and consequently large $S(T)$, combined with significant X-site alloy point defect scattering suppressing $\kappa_1$. This ‘heavy hole’ behaviour is the main characteristic of the $X_{\mu}\text{FeSb}$ family (and for the other p-type HH materials) and arises due to a doubly degenerate $L$-point valence band maxima with $N_v = 8$ (Table 1).\textsuperscript{119,120} Consequently, high doping levels between $10^{21}$–$10^{22}$ cm$^{-3}$ are required to optimise the thermoelectric performance at high temperatures, which was not achieved in this initial report. It only took one year for the same group to report $zT_{1200K} = 1.5$ for the composition $XNiSn$ systems have large low temperature $zT$, but their increase slows above 800 K due to onset of bipolar transport linked to the relatively small band gap.

Fig. 4 Literature $zT$ values of the best n-type and p-type half-Heusler thermoelectric materials. Panels (a) and (b) show $zT$ and the average $zT_{\text{avg}}$ of n-type samples as a function of temperature.\textsuperscript{142,148,172,174,183,191,242} Panels (c) and (d) show $zT$ and the average $zT_{\text{avg}}$ of p-type samples.\textsuperscript{47,48,75,120,121}
This material removes all V from the system, resulting in a decrease in \( m_{\text{DoS}} \) to 6.9 \( m_e \) reducing the doping required to reach the optimal \( S^2 \sigma \), with the increased \( \mu_4 \) resulting in a near doubling of \( S^2 \sigma \) to between 5–6 mW m\(^{-1}\) K\(^{-2}\). In addition to p-type doping the material, Hf substitution introduces X-site alloy point defect scattering, maintaining low \( \kappa_l \). Further success was found in TaFeSb, for which the composition \( \text{Ta}_{0.74} \text{V}_{0.1} \text{Ti}_{0.16} \text{FeSb} \) uses the same approach of combined carrier doping and phonon scattering using Ti, with a small fraction of V used to further suppress \( \kappa_l \) and push \( zT_{973} \) from 1.39 to 1.52.\(^{17}\) The largest peak \( zT \) reported for \( \text{X}_n \text{FeSb} \) materials comes from the \( \text{Nb}_{1-x} \text{Ta}_{0.4} \text{Ti}_{0.6} \text{FeSb} \) series, with both \( x = 0.36 \) and 0.4 reaching \( zT_{1200K} = 1.6 \),\(^{48}\) As can be seen in Fig. 4c however, the magnitude of \( zT \) is lower for the \( \text{Nb}_{1-x} \text{Ta}_{0.4} \text{Ti}_{0.6} \text{FeSb} \) series closer to room temperature (leading to a lower \( zT_{300K} \) in Fig. 4d), perhaps due to the weaker mass disorder provided by Nb.

Perhaps the most impressive feature of this family is the ability to achieve extremely large \( S^2 \sigma \) between 5–6 mW m\(^{-1}\) K\(^{-2}\) over a wide temperature range.\(^{36–48,121–123}\) As first reported in 2016, careful processing allows for \( S^2 \sigma_{300K} > 10 \) mW m\(^{-1}\) K\(^{-2}\) to be achieved in \( \text{Nb}_{0.93} \text{Ta}_{0.08} \text{FeSb} \), using a hot pressing temperature of 1100 °C to eliminate the parasitic electrical resistance from grain boundary scattering observed with lower pressing temperatures, as illustrated in Fig. 5.\(^{29}\) This was later also demonstrated for the analogous systems with Zr and Hf dopants.\(^{123}\) In addition to the very large \( S^2 \sigma \), the \( \kappa_l(T) \) achieved in \( \text{Ta}_{0.74} \text{V}_{0.1} \text{Ti}_{0.16} \text{FeSb} \) is extremely low for a HH material, starting at \( \kappa_l_{300K} = 2.3 \) W m\(^{-1}\) K\(^{-1}\) and decreasing to \( \kappa_l_{1200K} = 1.6 \) W m\(^{-1}\) K\(^{-1}\), with mass disorder from Ti/V substitution proving to be effective at suppressing \( \kappa_l \).\(^{47}\)

A range of synthetic routes have been used to prepare \( \text{X}_n \text{FeSb} \) HH materials. The initial reports on \( \text{V}_{0.6} \text{Nb}_{0.4} \text{FeSb} \) and \( \text{Nb}_{1-x} \text{Hf}_{x} \text{FeSb} \) used levitation melting, mechanical milling, SPS and a short annealing step at a low temperature of 750 °C.\(^{119,121}\) The TaFeSb based samples were prepared directly from the elements using ball milling (in a shaker mill), followed by direct current hot-pressing (similar to SPS) for only two minutes at 850 °C, with no subsequent annealing step.\(^{47}\) This report does not establish whether the HH phase forms during the milling step, or whether reaction takes place during hot pressing with the assistance of finely dispersed powder precursors. In spite of the well documented benefit on \( \sigma_{\text{RT}} \) of a 1100 °C densification temperature for \( \text{Nb}_{1-x} \text{Ta}_{x} \text{FeSb} \) (Fig. 5), highly doped \( \text{Ta}_{1-x} \text{Ti}_{x} \text{FeSb} \) samples seem to suffer no detriment from the relatively low 850 °C pressing temperature (although \( x < 0.06 \) samples seem to suffer from grain boundary effects).\(^{37}\) This suggests that these systems have a different grain boundary chemistry. Note that due to the higher melting point of Ta compared to Nb, \textit{a priori} smaller grain sizes and a more pronounced impact on \( \sigma_{\text{RT}} \) might have been expected. However, this does not appear to be the case.

To conclude, the \( \text{X}_n \text{FeSb} \) materials are blessed with naturally occurring electronic properties and thermal transport, which can be easily suppressed using well established mass disorder techniques.

3.2. ZrCoBi (p-type)

In 2018, \( \text{ZrCoBi}_{0.65} \text{Sn}_{0.35} \) was reported to have the then largest \( zT_{973K} = 1.42 \) for any p-type HH phase.\(^{48}\) This composition benefits from a large \( N_e = 10 \), which arises due to degenerate \( \Gamma \) and \( L \) (Table 1) valence band maxima, which is an inherent property of the stoichiometric ZrCoBi phase (Fig. 2). This convergence translates to large \( m_{\text{DoS}} \), with the reported Hall data suggesting that \( m_{\text{DoS}} \) increases from around 9 \( m_e \) to 13 \( m_e \) with Sn substitution up to \( x = 0.2 \) in the ZrCoBi\(_{1-x}\)Sn\(_x\) solid solution (Table 1).\(^{48,86}\) A combination of Sn and Sb alloying is successful in suppressing \( \kappa_l \) and optimising the carrier concentration, with the parent composition already having a markedly lower \( \kappa_l \) than many HH phases due to the high mass of Bi and a low velocity of sound, \( v_s \) (Table 2). As was the case for TaFeSb,\(^{47}\) the ZrCoBi samples were prepared \textit{via} ball milling the constituent elements and then direct current hot pressing for 5 minutes at 900 °C.\(^{48}\) In this report the HH phase was found to form directly after the milling step. No further experimental work has emerged on p-type ZrCoBi materials since, although two separate computational investigations of ZrCoBi materials also find near degenerate \( \Gamma \) and \( L \) point valence band maxima in TiCoBi
and HfCoBi.\textsuperscript{124,125} It was discovered, however, that n-type ZrCoBi is also a promising thermoelectric material, reaching $zT_{973K} = 1$ for ZrCo$_{0.6}$Ni$_{0.4}$Bi$_{0.85}$Sb$_{0.15}$ again utilising a ball-milling to pressing synthetic protocol.\textsuperscript{86} As before Bi/Sb alloying is used to suppress $\kappa_l$, but interestingly Co/Ni alloying also reduces $\kappa_l$ significantly, despite minimal expected mass and strain differences. The reported $m^*_\text{DoS} = 7 m_e$ is quite large considering the low band degeneracy of the X-point, suggesting a large $m_b$, consistent with the reported low mobility. In the previous year n-type ZrCo$_{1.25}$Pd$_{0.75}$Bi was reported with substantially lower $zT$, due to a combination of significantly larger $\kappa_l$ and a much lower $m^*_\text{DoS} = 4.4 m_e$.\textsuperscript{126} The samples in this report were prepared using arc-melting, ball-milling and rapid uniaxial hot-pressing.

3.3. X$_V$CoSb (X$_V$ = Ti, Zr and Hf) (p-type)

Before the rapid emergence of the high $zT$ p-type X$_V$FeSb and ZrCoBi materials, X$_V$CoSb materials were by far the most promising candidate p-type HH phases. $zT_{1073K} = 1.0$ had been achieved for Ti$_{0.2}$Hf$_{0.8}$CoSb$_{0.8}$Sn$_{0.2}$ in 2012, with the same achieved in a 2013 follow up report on Ti$_{0.1}$Zr$_{0.9}$Hf$_{0.1}$CoSb$_{0.8}$Sn$_{0.2}$, with both utilising X-site alloying and milling to suppress $\kappa_l$.\textsuperscript{127,128} Once again, these p-type compositions benefit from high $m^*_\text{DoS}$ due to the doubly degenerate L-point valence band maximum (Table 1), although $S'\sigma$ did not exceed 3.0 mW m$^{-1}$ K$^{-2}$ in these materials, with a fundamentally low $\mu_H$ in this system. The modest $S'\sigma$ is a serious drawback for these compositions, as there are few clear-cut ways to improve the electronic properties and $zT$, whereas a range of strategies are available to suppress $\kappa_l$. In 2015 $zT$ was pushed up to 1.15 at 973 K for the composition Ti$_{0.2}$Hf$_{0.8}$CoSb$_{0.8}$Sn$_{0.2}$, by optimising $\mu_H$ and X-site alloying, with phase segregation into Ti-rich and Hf-rich regions credited with further suppression of $\kappa_l$.\textsuperscript{129} Another important detail in this paper is while $m^*_\text{DoS} \sim 10.7 m_e$ remains largely constant in the Ti$_{1-x}$Hf$_x$CoSb solid solution, $\mu_H$ is around 3 times lower in TiCoSb than HfCoSb, which makes the expensive element Hf key to high performance.\textsuperscript{75} Studies into the stability of the phase segregation revealed both $zT$ and microstructure to remain unchanged upon repeated temperature cycling.\textsuperscript{129,130} Another drawback of TiCoSb is that it appears to be more sensitive to oxidation than other HHs.\textsuperscript{131-137}

Since 2015, there hasn’t been any increase in $zT$ in the p-type X$_V$CoSb materials, although work on these materials continues.\textsuperscript{138,139} An extensive investigation into the Zr$_{1-x}$Hf$_x$CoSb$_{1-y}$Sn$_y$ system determined that the temperature dependence of hole mobility is consistent with alloy scattering ($\mu_H \sim T^{-0.5}$), in stark contrast to the X$_V$FeSb materials, which have temperature dependences consistent with acoustic phonon scattering.\textsuperscript{92}

The above reports generally make use of arc-melting to react elemental pieces, although a 2021 report eliminated the melting step, simply ball-milling elemental precursors and hot-pressing at 1000 °C.\textsuperscript{140} This study equated the record with $zT_{973K} = 1.1$ for this family with the composition Ti$_{0.6}$Hf$_{0.4}$CoSb$_{0.83}$Sn$_{0.17}$, with suppressed $\kappa_l$ and $S'\sigma$.\textsuperscript{140} While this is a positive result, it also illustrates that $\kappa_l$ can’t be suppressed further using Ti without also suppressing the electronic performance.

3.4. X$_V$CoSb (X$_V$ = Ti, Zr and Hf) (n-type)

The X$_V$CoSb materials are unusual in that good $zT$ values are possible in both n- and p-type doped compositions. The n-type materials have the X-point conduction band minimum shared amongst all 18 electron HH phases (Fig. 2). Despite the fact that X$_V$CoSb materials have long been known to be intrinsically n-type,\textsuperscript{141} reports on competitive $zT$ have only emerged in the last 4–5 years.\textsuperscript{93,142} Initially, Ni substitution was used to dope Zr$_{0.5}$Hf$_{0.5}$CoSb, with $zT_{973K} = 1$ achieved in Zr$_{0.5}$Hf$_{0.5}$Co$_{0.8}$Ni$_{0.2}$Sb, with moderate $S'\sigma$ values exceeding 3.25 mW m$^{-1}$ K$^{-2}$ and once again suppressed $\kappa_l$ due to X-site alloying.\textsuperscript{142} A $m^*_\text{DoS} \sim 6 m_e$ was confirmed in subsequent reports on other n-type compositions, with (Zr$_{1-x}$Hf)$_{1-y}$Nb$_x$CoSb having $m^*_\text{DoS} \sim 6.5 m_e$.\textsuperscript{93,100} Although the Nb doped materials don’t exceed $zT = 1$, the initial report used Callaway analysis to show that the X-site alloying in these materials suppresses $\kappa_l$ without causing any reduction in $\mu_H$. This is consistent with the earlier observation that minimising the size mismatch between dopant/allloying elements helps maintain good $\mu_H$ values,\textsuperscript{93} whilst mass disorder effectively reduces $\kappa_l$. These studies have also shown that n-type TiCoSb is substantially more promising than its p-type counterpart,\textsuperscript{139} with the composition Ti$_{0.6}$Hf$_{0.4}$Ti$_{0.1}$CoSb reaching $zT_{973K} = 0.9$.\textsuperscript{143}

A subsequent study on (Hf$_{0.3}$Zr$_{0.7}$)$_{0.94}$Nb$_{0.06}$CoSb established that $\sigma_{\text{room}}$ can be heavily suppressed by grain boundary scattering, introduced by intense ball-milling used to reduce grain size.\textsuperscript{100} This leads to semiconducting behaviour near 300 K, while at high temperatures degenerate semiconducting behaviour is observed. Hence, as was seen for NbFeSb, care must be taken when processing HH phases to prevent grain boundary scattering of charge carriers. At present all n-type X$_V$CoSb materials reported use a melting, milling and pressing procedure, but this may simply reflect the novel nature of these compositions and a lack of detailed processing studies.

3.5. X$_V$NiSn (X$_V$ = Ti, Zr and Hf) (n-type)

The X$_V$NiSn compositions are very much the original HH thermoelectric,\textsuperscript{17,18,20,21,24-27,144-147} with the 2005 report of $zT_{\text{room}} = 1.3$ driving a significant amount of interest in HH thermoelectrics.\textsuperscript{148} As shown in Fig. 4, $zT$ values from this report have never been fully matched, contrary to the expected evolution of material optimisation, which sees a gradual increase in $zT$ over the years.

The X$_V$NiSn phases have some unique properties relative to the other HH alloys. The X$_V$NiSn materials have the lowest reported $m^*_\text{DoS} \sim 3 m_e$ amongst the high $zT$ materials, which arises due to a combination of low $N_s = 3$ and low $m^*_b \sim 1.7 m_e$ (Table 1). This means that lower doping levels are required to optimise $S'\sigma$, and in several instances the best performance comes for ~1–2% atomic substitution,\textsuperscript{148-150} contrasting with 15–20% required for the p-types.

An additional factor to consider is the interstitial site and the full-Heusler XNi$_2$Sn phase.\textsuperscript{41-44,54,57,71,72,151-166} The interstitial site is the vacant 8-coordinate (tetrahedral coordination to X and Z) site in the material (sometimes referred to as the Wyckoff 4d or tetrahedral site), which is fully occupied in the
FH structure (Fig. 1). Even stoichiometric \(X_{1/2}NiSn\) materials commonly form with a few percent occupation of the interstitial site, with an \(X_{0.5}Sn\) binary phase to compensate. Preparing compositions with intentional metal excess, \(X_{0.5}Ni_{1-x}Sn\) is possible, although there is a limit to \(x\) past which a FH phase begins to form. \(X_{0.5}Ni_{1-x}Sn\) investigations in the early 2010s. The presence of interstitial Ni is also cited as reducing the observed electronic band gap.116

**X-Site phase segregation.** Initial success in reproducing the large reported \(zT\) values came from work focusing on X-site phase segregation. This is based on the limited mutual solidity of Ti and Zr/Hf, whereas Zr and Hf with a more similar size are fully miscible. Reports by Schwall and Balke, and Populoh et al. yielded \(zT_{\text{30K}} = 1.2\) with the low \(k_1\) underpinning the good performance linked to phase segregation. Phase segregation due to limited solubility between Ti, Zr and Hf is expected from first-principles, with only the solid solution between ZrNiSn and HfNiSn expected to be stable at 300 K, although full Ti/Zr(Hf) solubility is predicted to be possible above 600 K. This phase segregation behaviour has been observed routinely in these materials through powder diffraction and SEM analysis. Phase segregation is not simply a case of the formation of two distinct phases, with Ti alloyed compositions containing several HH phases discernible by diffraction. This can be ascribed in large part to low expected diffusion rates of the \(X_{0.5}\) site atoms, leading to regions with kinetically trapped compositions resulting from the melting stage. Interestingly, SEM and SXRD data for \(Ti_{0.5}Zr_{0.5}Hf_{0.5}\) NiSn prepared by a powder route (never exceeding 850 °C) shows an increasing homogenisation on the \(X_{0.5}\) site with increasing Cu content, which was attributed to the presence of a molten Cu-Sn phase that improves the reaction kinetics during synthesis.

The approach which was able to reduce \(k_1\) the most used a convoluted synthetic protocol consisting of arc-melting, induction-melting, annealing, an undisclosed “densification aid”, ball-milling and finally hot-pressing. This work also aimed to exploit X-site phase segregation by initial heating above the spinodal line and subsequent heat treatment just below to force nanoscale segregation. However, the paper itself does not provide direct microscopy confirmation. Nevertheless, the composition \(Ti_{0.5}Zr_{0.25}Hf_{0.25}\) NiSn (+DA) was reported to have a very low \(k_1\) of 1.8 W m\(^{-1}\) K\(^{-1}\) at 300 K, while maintaining a large \(S^2\sigma\), enabling \(zT_{2.2K} = 1.5\). Outside of this report, the group is able to consistently achieve \(zT_{\text{30K}} > 1\) with the assistance of a “densification aid” that is considered to improve electrical contacting of the grains. In the composition \(Ti_{0.5}Zr_{0.1}Ni_{0.92}Sn\) for example, the addition of the “densification aid” increases \(\sigma\) at 700 K from \(2.5\) W m\(^{-1}\) K\(^{-2}\) to \(4\) W m\(^{-1}\) K\(^{-2}\), and coupled to a low \(k_1\) this enables \(zT > 1\).

**Interstitial metals and FH inclusions.** Investigations into the nature of excess Ni in these materials has progressed since our 2014 review. Calculations investigating the \(X_{0.5}NiSn/X_{0.5}Ni_{1-x}Sn\) solid solution have explored the stability of interstitial Ni and experimental work has been performed to find the limits of interstitial occupancy. In well-annealed samples \((T > 800 °C)\), the solubility of interstitials is highest in TiNiSn \((x_{\text{max}} = 0.06–0.08 \text{ at } 900 °C)\), whilst \(x_{\text{max}} \leq 0.03\) for ZrNiSn and HfNiSn. The solubility of Ni interstitials increases with temperature, consistent with the entropy stabilised nature of this defect. Full-Heusler (nano)inclusions can be stabilised by (rapid) cooling from the melt, but exposure to \(T > 800 °C\) leads to the gradual dissolution of FH inclusion, and conversion to interstitials and distinct (macroscopically segregated) FH phases that are discernible in diffraction. Due to the link to processing \((i.e.,\) cooling from the melt), the experimental \(4d\) site occupancies can far exceed equilibrium values, with electronic properties recorded for ZrNi\(_{1-x}\)Sn with \(x > 0.1\). Unpublished neutron powder diffraction data from our group confirms \(x \leq 0.18\) in arc-melted ZrNi\(_{1-x}\)Sn ingots, although this decreases to \(x = 0.03\) when annealed at 900 °C (in agreement with the phase diagram), with the excess Ni forming FH regions visible by diffraction.

**Property measurements.** Property measurements have confirmed the significance of interstitials on the thermoelectric properties, with excess Ni suppressing \(k_1\) by up to 70% in the TiNi\(_{1-x}\)Sn and ZrNi\(_{1-x}\)Sn system. Callaway analysis demonstrates that Ni/Cu interstitials have a similar impact to vacancies with additional disorder contribution due to bond formation. For the Ti\(_{1-x}\)Ta\(_{x}\)Ni\(_{0.92}\)Sn system, the intentional Ni deficiency (with the aim of obtaining stoichiometric TiNiSn) leads to an increase in both \(k_1\) and \(\mu_{\text{H}}\) with the improved electronic properties leading to a net increase in \(zT\) (Fig. 6). At low temperatures, the Ni deficiency leads to a substantial increase in \(\sigma\), while at higher temperatures \(S\) is enhanced, leading to a minimum of 35% increase in \(S^2\sigma(T)\) across the entire temperature range. Using appropriate processing, excess Ni can also be stabilised as FH (nano)inclusions in the HH matrix, acting an extrinsic source of phonon and electron scattering. Chai et al. reported success in preparing composite ZrNi\(_{1-x}\)Sn with significantly improved thermoelectric properties relative to the stoichiometric composition, attributed largely to a dispersion of nanoscale FH inclusions. DTA analysis suggested the FH phases ‘dissolve’ into the HH matrix above 800 °C, and reform with a smaller domain size when cooled. These nanoscale FH inclusions suppress \(k_1\) by up to 40%, leading to a ~40% increase in \(zT_{\text{30K}}\) from 0.54 to 0.75, compared to the first heating cycle. In addition to these melt-based routes, the Poudeu group has reported extensively on noncomposite HH samples prepared using a high-temperature solid-state route. Carrier filtering was reported to improve \(S^2\sigma\) and reductions in \(k_1\) were observed, with a highest \(zT = 0.8\) at 800 K.

Recently it was found that W nano inclusions are highly effective in reducing \(k_1\) by 20% while simultaneously enhancing \(S^2\sigma\) by up to 33%. Hf\(_{0.6}Zr_{0.4}NiSn_{0.99}\)B\(_{0.01}\) + 5 wt% W has peak \(zT_{77K} = 1.4\). This would seem to be a successful approach. \(zT\) and \(\sigma\) at 800 K.
the Ti$_{1-x}$Ta$_x$Ni$_{0.92}$Sn system was not addressed, first principles calculations on the interstitial Cu-doped materials suggest the appearance of a second conduction band minimum in the electronic bandstructure, supporting the idea that the increase in $m^*_{\text{DoS}}$ is caused by to an increase in $N_e$. 

Given the high volume of publications on the $X_V$NiSn system, a number of protocols outside of melting, ball-milling and hot-pressing/SPS exist, with TiNiSn being the target of most experimentation. Microwave synthesis has reported for TiNiSn, achieving a peak $zT_{740K} \approx 0.45$. In 2019 Zilber et al. successfully grew TiNiSn single crystals from a Sn flux, with lower $n_H$ and larger $S$ than typically reported in polycrystalline samples. Similarly ZrNiSn$_{1_x}$Si$_y$ crystals can be grown and extracted by centrifuging, with improved $\mu_H$ compared to high quality polycrystalline samples.

Self-propagating synthesis for exothermic processes allows for the heat produced from initial reaction to drive the reaction to completion, as has been demonstrated for ZrNiSn. In a custom pressing setup, the heat released from the reaction (initiated by Tungsten electrode) was enough to densify the pellet, with a $T_{\text{max}} = 1668 \text{ K}$ achieved by the self-propagating reaction.

The large literature pool also allows for comparisons to be made in how the HH phases form, which is important when planning microstructure and defects. ZrNiSn and HfNiSn are known form congruently from the molten state, while TiNiSn prefers to form the TiNi$_2$Sn FH phase. Similarly TiNiSn decomposes when heated above 1450 K. The rapid solidification achieved by melt-spinning is claimed to be able to suppress formation of TiNi$_2$Sn from the molten state. It remains the case that melting, milling and hot-pressing/SPS give the best thermoelectric performance.

Overall, for the $X_V$NiSn compositions, it remains challenging to optimise the thermoelectric properties. This is related to the tendency for significant deviations from nominal stoichiometry, phase segregation and the small electronic bandgap, which leads to significant bipolar transport at high temperatures. Elimination of interstitial Ni is desirable as this would increase the electronic bandgap, hence reducing the minority carrier conduction and improving the high temperature $zT$.

### 3.6. $X_{V\alpha}$CoSb ($X_V = V$, Nb and Ta) (n-type)

Perhaps the most significant recent discovery for the n-type HH materials is promising thermoelectric performance in the nominally 19 electron $X_V$CoSb compositions. As a stoichiometric phase, these materials are expected to be metallic with the possibility of ferromagnetism. In 2016, the material VCoSb was synthesised with $S(T)$ significantly larger than what is typical of a fully metallic sample, although $n_H$ was quite large at $\approx 1.6 \times 10^{22} \text{ cm}^{-3}$. The properties of this phase and the subsequently investigated NbCoSb and TaCoSb phases were discovered to be due to X-site vacancies, which reduce the overall electron count without depleting electronic states in the valence band, which is easily understood through the valence electron counting $X^{10}[YZ]^-$ description introduced earlier.
To reach a theoretically semiconducting state one would need to reach the composition $X_{0.8}V_{0.2}CoSb$ (Fig. 7). Each of the $X_rCoSb$ phases forms with a different vacancy fraction, with the $V$ phase forming preferentially as $V_{0.87}CoSb$, the Nb phase as $Nb_{0.85}CoSb$, while the Ta phase forms close to being ‘fully’ vacant at highly degenerate while $Ta_{0.81}CoSb$ is close to being a semiconductor.\(^{59}\) There is some level of control that can be achieved, as demonstrated in $Nb_{0.8}CoSb$, which can be doped between $0.5-6.5 \times 10^{21}$ cm$^{-3}$ by varying the nominal vacancy fraction.\(^{183}\) One complication to doping control is the dependence of $\delta$ on the annealing temperature.\(^{190}\)

Relative to the $X_rNiSn$ materials, $m^*_D$ for these materials is quite large, ranging from $10 \ m_e$ for $V_{0.87}CoSb$ to $5.9 \ m_e$ for $Ta_{0.81}CoSb$ (Table 1).\(^{59,183,191}\) The lower magnitude of $\mu_H$ ($\sim 4$ cm$^2$ V$^{-1}$ s$^{-1}$ for $Nb_{0.8}CoSb$) may point to an increase in $m^*_e$, but calculated bandstructures for both $NbCoSb$ and $Nb_{0.8}CoSb$ suggests more bands can contribute than the singly degenerate $X$-point minimum that is commonly observed (Fig. 7).\(^{188}\) Assuming a doubly degenerate conduction band as shown in Table 1, $m^*_e$ for $Nb_{0.8}CoSb$ is calculated to be $\sim 2.33 \ m_e$, which is a large value compared to n-type $X_rNiSn$.

One unusual behaviour in this family is the very low $k_1$ in $V_{0.87}CoSb$ below $2 \ W \ m^{-1} \ K^{-1}$, which would expected to be larger than the $Nb$ and $Ta$ phases due its lower formula mass.\(^{59,191}\) The relative acoustic behaviour of the $V_{0.87}CoSb$ and $Nb_{0.8}CoSb$ materials have been confirmed to be similar, with no significant effect caused by vacancies observed through experiment, which would point to a reduced $k_1$, with measured $\nu_s$ in the $Nb_{0.8}CoSb$ varying by no more than $\pm 2\%$ between $0 < \delta < 0.2$.\(^{183,191}\) As of yet there is no clear explanation for this extremely low value. One can note that the experimental acoustic properties point to a large $\gamma$ for these materials (Table 2), which correlates with strong Umklapp scattering and low $k_1$ as shown in Fig. 3. One structural feature of interest in these phases is the potential ordering of the $X_r$ site vacancies.\(^{192-195}\) Over large distances the vacancies are disorder, but at the nanoscale the vacancies are short-range ordered (SRO), offering a route to reduce $k_1$.\(^{192}\) Furthermore, Monte Carlo simulated vacancy ordering coupled to DFT supercell calculations points to widening of the electronic CBM (increase in $m^*_e$) relative to the ordered $Nb_{0.8}CoSb$ and the presence of Co-defect states in the band gap, with electron diffraction simulations confirming existence of SRO in real samples.\(^{193}\) A more detailed discussion on the vacancy ordering in these materials can be found in the 2021 review by Xia et al.\(^{12}\)

The majority of reports on the $X_{0.8}CoSb$ materials have utilised a similar synthetic protocol, with melted elemental pieces being optionally ball-milled and then consolidated by hot-pressing/ SPS.\(^{183,184}\) However, a powder route consisting of solid-state reaction and hot pressing achieves similar performance.\(^{59}\) Finally, a recent report on $V_{0.85}Ti_{0.1}CoSb$ used a direct ball milling and consolidation approach and achieved $zT$ values larger than other reported $V_{0.8}CoSb$ samples.\(^{191}\) This suggests that for the $V$ based materials, these relatively straightforward powder based routes may be more feasible than for $X_rNiSn$, where the best performing samples are invariably prepared using melting routes.

### 3.7. Other HH phases

The best performing new and existing materials have been described above, but exploration of new phases has turned up some interesting results. One recent HH family which has emerged is the $X_rCoSn$ ($X_r = Nb$ and $Ta$) phases, which are intrinsically n-type materials.\(^{187,196-198}\) The bandstructure of $NbCoSn$ appears contain a doubly degenerate $X$-point conduction band minimum, giving $N_v = 6$, which appears to be partially responsible for $m^*_D = 6.1 \ m_e$ in the $NbCoSn_{1-x}Sb_x$ series (Table 1).\(^{197,198}\) At present the largest $zT$ is found for X-site alloying between $Nb$ and $Ta$, yielding $zT_{1000K} = 0.75$ for $Nb_{0.4}Ta_{0.6}CoSn_{0.94}Sb_{0.06}$.\(^{196}\) Better low temperature performance has been observed when using $Pt$ as both n-type dopant and phonon scatterer in $NbCoSn_{0.5}Pt_{0.5}Sn$, due to the large mass difference between $Co$ and $Pt$.\(^{197}\) While it would be useful to use $V$ for X-site alloying, the $VCoSn$ phase does not appear to readily form.\(^{131}\) While there have been reports of the phase being observed by powder diffraction, the evidence presented is less than convincing.\(^{199,200}\) It has been suggested that this
composition is difficult to reach through normal processing routes, but might be made through non-equilibrium processes.\textsuperscript{201} Finally, p-type doping has also been demonstrated using scandium doping in Nb\textsubscript{1-x}Sc\textsubscript{x}CoSn, however these samples have relatively poor performance, with the largest $zT_{773K} = 0.13$ reported for Nb\textsubscript{0.95}Sc\textsubscript{0.05}CoSn.\textsuperscript{202}

An established family of HH alloys that has recently gained renewed interest are the RNiSb phase (R = lanthanide, Sc).\textsuperscript{203–206} As synthesised, these materials are p-type and are characterised by spontaneous Ni deficiency (as much as 15\%).\textsuperscript{207} They also have small band gaps ($E_g \sim 0.3$ eV), leading to substantial bipolar transport at low temperatures, ultimately limiting the maximum achievable $zT$ values. A large study comparing $R =$ Sc, Dy, Ho, Er, Tm and Lu, confirmed the intrinsic p-type behaviour with semiconductor like $S(T)$ and $\rho(T)$.\textsuperscript{205} Typical peak $S^2/\rho = 1–1.5$ mW m\textsuperscript{-1} K\textsuperscript{-2} for the RNiSb materials with highest reported $zT \sim 0.3$ at 650 K.\textsuperscript{205,206,209} Careful analysis of the bipolar transport, suggested much higher electron than hole mobilities.\textsuperscript{204} Subsequent attempts at n-type doping exploiting Te in ScNiSb\textsubscript{1-x}Te\textsubscript{x} yielded large $S^2/\rho = 4$ mW m\textsuperscript{-1} K\textsuperscript{-2}, demonstrating the potential of n-type RNiSb systems. Substitution of Te is also correlated to an increasing $m^*_D$, providing another example of the significant impact of a doping element on the underlying bandstructure.\textsuperscript{206} No $zT$ values were reported, but the chemical flexibility for substitution on the Z site and potential to manipulate the Ni deficiency (concentration and potentially SRO through heat treatment), suggest that these materials are promising for further exploration.

Another interesting development is the exploration of “double HH” phases, which combine an equal mixture of 17 electron and 19 electron HH phases.\textsuperscript{112} For example the model material proposed in 2019 was [TiFeSb]\textsubscript{0.5}[TiNiSb]\textsubscript{0.5}, each of which forms a HH-like phase independently. Experimental realisation of [TiFeSb]\textsubscript{1-x}[TiNiSb]\textsubscript{x} was successful, showing that VEC < 18 were p-type and VEC > 18 were n-type as would be expected, and that isoelectronic alloying using Hf could be used to reduce $\kappa$.\textsuperscript{210} Some unusual points emerged, such as the large reduction in $\kappa$ just above and below $x = 0.5$ and massive apparent $m^*_D$ values up to 23.8 $m_e$, which was later confirmed by a separate group.\textsuperscript{211} The emergence of defective HH phases and double-HH phases has provided an extended range of potential compositions to explore, which has allowed for the application of computational screening, beyond the common style of reporting the first principles properties of stoichiometric XYZ HHs.\textsuperscript{112,212} The emergence of these phases coincide with the development of effective band structure techniques, that allow supercell bandstructures to be compared to stoichiometric parent phases.\textsuperscript{213,214}

### 4. Recent developments in half-Heusler thermoelectric generators

There has been an increased research effort in recent years aimed at implementing HH materials into TEG modules and systems.\textsuperscript{215–219} It is now more common for “materials” studies to include work on prototype devices,\textsuperscript{121,122,179} while an increasing number of manuscripts are dedicated to electrical and thermal contacting of HH materials.\textsuperscript{220–224} Most of this published work is from research institutions but there has been industrial work producing HH materials and generators, most notably from GMZ energy in the USA, and the Isabellen Hutte Heusler company and Vacuum Schmelzen in Germany. A vital aspect in translating materials to TEG performance is the electrical and thermal contacting. In particular, the requirements for electrical contacting are strict in order to maintain as much of the intrinsic performance of the HH materials. The effective $zT$ for a “contacted” material is given by:\textsuperscript{4}

$$zT_{\text{avg}} = \left[ \frac{L}{L + 2\rho_e \sigma} \right] zT_{\text{avg}}$$

This leads to strict criteria on permissible contact resistances, e.g. for a $L = 2$ mm height leg, $\rho_e < 5 \, \mu \Omega \, \text{cm}^2$ to keep losses <5–10\%. In addition to a low contact resistance, the contacts also need to be stable under operating conditions, i.e. no chemical reaction with the TE materials, often necessitating the use of barrier layers.\textsuperscript{8} In terms of reviewing the literature, comparison of published results on TEG modules is difficult as they depend on dimensions of the legs, measurement conditions, heat flow through the legs and minimising heat losses. An overview of reported power densities and efficiencies for a range of HH based TEG modules is given in Table 3.

<table>
<thead>
<tr>
<th>n/p-type materials</th>
<th># couples (dimensions mm\textsuperscript{3})</th>
<th>$zT$ (n/p-type@T/K)</th>
<th>Power density (W cm\textsuperscript{-2})/(AT/K)</th>
<th>Efficiency (%)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-scale melt X\textsubscript{0.9}NiSn/X\textsubscript{0.9}CoSb</td>
<td>7 (2.5 \times 2.5 \times 3)</td>
<td>0.8/0.5@773 K</td>
<td>3.2 (525 K)</td>
<td>5</td>
<td>215</td>
</tr>
<tr>
<td>Milled nano-X\textsubscript{0.9}NiSn/X\textsubscript{0.9}CoSb</td>
<td>1 (1.8 \times 1.8 \times 2)</td>
<td>1.0/0.9@773 K/973 K</td>
<td>5.3 (500 K)</td>
<td>7</td>
<td>216 and 241</td>
</tr>
<tr>
<td>Self-propagating X\textsubscript{0.9}NiSn/X\textsubscript{0.9}CoSb</td>
<td>28 (1 module)</td>
<td>1.0/0.9@773 K</td>
<td>4.0 (500 K)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>400 modules</td>
<td></td>
<td>2.8 (500 K)</td>
<td>2.1</td>
<td>217</td>
</tr>
<tr>
<td>X\textsubscript{0.9}NiSn/X\textsubscript{0.9}FeSb</td>
<td>8</td>
<td>1.0/0.9@773 K</td>
<td>2.4 (680 K)\textsuperscript{ab}</td>
<td>8</td>
<td>217</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.7 (680 K)</td>
<td>9.5</td>
<td>—</td>
</tr>
<tr>
<td>X\textsubscript{0.9}NiSn/X\textsubscript{0.9}FeSb</td>
<td>8</td>
<td>1.0/1.0@973 K/1200 K</td>
<td>2.1 (655 K)\textsuperscript{a}</td>
<td>8.3</td>
<td>218</td>
</tr>
<tr>
<td>X\textsubscript{0.9}NiSn/X\textsubscript{0.9}FeSb</td>
<td>1</td>
<td>1.0/1.0@773 K/973 K</td>
<td>6.2 (500 K)</td>
<td>—</td>
<td>122</td>
</tr>
<tr>
<td>X\textsubscript{0.9}NiSn/X\textsubscript{0.9}FeSb</td>
<td>1 (1.5 \times 1.5 \times 2.4)</td>
<td>1.0/1.0@773 K/973 K</td>
<td>8.6 (570 K)</td>
<td>6.2</td>
<td>224</td>
</tr>
<tr>
<td>TiSn\textsubscript{0.9}NiSn/FeSb</td>
<td>18 (1.75 \times 1.75 \times 4.0)</td>
<td>0.4/0.4@873 K</td>
<td>3.1 (575 K)</td>
<td>—</td>
<td>219</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Surface area of module, not TE material. \textsuperscript{b} This study covers a range of leg lengths and surface areas, showing the inverse correlation between efficiency and power density.
One of the first papers to address upscaling and device manufacture originated from the Fraunhofer Institute in 2014.\textsuperscript{215} This work used vacuum melting and casting to prepare kilogram scale ingots of optimal n-type Zr\textsubscript{0.4}Hf\textsubscript{0.6}NiSn\textsubscript{0.98}Sb\textsubscript{0.02} and p-type Zr\textsubscript{0.5}Hf\textsubscript{0.5}NiSb\textsubscript{0.8}Sn\textsubscript{0.2}. The HH materials had $zT$ $\sim$ 0.8 (n-type) and $\sim$ 0.5 (p-type) at 773 K, respectively. TEG modules containing 7 n/p-couples were fabricated using Cu contacts and a brazing process. These were exposed to repeated thermal cycling up to 873 K and thermal shock without degradation of performance. The power density was 3.2 W cm\textsuperscript{-2} (area of thermoelectric) from a 527 K gradient, with the cold side near room temperature. The efficiency was estimated to be $\sim$ 5\% with an average TEG module $zT$ $\sim$ 0.4.

Zhang et al. demonstrated a 1 kW TEG for use in automotive applications.\textsuperscript{216} The HH materials used were based on milled nanostructured X\textsubscript{IV}NiSn and X\textsubscript{IV}CoSb with $zT = 1$ (773 K) and $zT = 0.9$ (973 K), respectively. The production process and different stages in the TEG module development are illustrated in Fig. 8. A Ag–Cu brazing alloy was used to bond the legs directly to the Cu leads through heating in vacuum at 825 °C. Unicouple generators produced 5.3 W cm\textsuperscript{-2} (500 K gradient), while a module consisting of 28 couples generated 4 W cm\textsuperscript{-2} (500 K gradient). The reduced power density of the module is caused by parasitic losses, attributed to thermal losses between the inert atmosphere steel encapsulation and TEG device (Fig. 8).

The final TEG that was embedded in the exhaust stream of a heavy-duty vehicle used an impressive number of 400 TEG unicouple pairs. This study covered a range of leg geometries, where the ratio of leg height over area ($L/A$) can be used to favour either efficiency (increase $L/A$) or power density (decrease $L/A$). Limiting measured values in a 680 K gradient are $\sim$ 2.4 W cm\textsuperscript{-2} ($\sim$ 8\% efficiency) and $\sim$ 1.7 W cm\textsuperscript{-2} (9.5\% efficiency). Simulations suggest an upper limit of $\sim$ 3.1 W cm\textsuperscript{-2} and 10.5\% for the power density and efficiency, respectively.\textsuperscript{217} However, it is worth noting that the performance of these modules is limited by relatively large contact resistances, $\sim$ 30 $\mu$2 cm\textsuperscript{-2} for both n- and p-type legs. A final noteworthy study focused on TiNiSn (hence avoiding the use of Hf) as the n-type material and NbFeSb as the p-type. Despite a low $zT_{\text{373K}}$ $\sim$ 0.4 for both n- and p-types, a fabricated TEG with 16 pairs showed a respectable power density of $\sim$ 3.1 W cm\textsuperscript{-2} from a 575 K gradient, reflecting the large $S^2\sigma$ of the HHs.\textsuperscript{219}

5. Conclusions

Significant progress has been made in the development of HH thermoelectrics with the emergence of new high-performing p-type materials based on X\textsubscript{IV}FeSb and ZrCoBi, and further development of existing n-types based on X\textsubscript{IV}NiSn. Application of these materials in TEG modules has led to the demonstration of promising power outputs and efficiencies. Looking ahead there are a number of areas that deserve attention. Further work on translating materials into TEGs is needed, including underpinning work on electrical and thermal contacting and stability studies. A related area is the development of scalable synthetic routes to mass produce HH alloys with good performance. These more applied activities should be underpinned by further materials development work as there continues to be significant potential for further improvements. One area of interest is the elimination of parasitic effects, for example grain
boundary resistances, enabling the intrinsic materials performance to be extracted. Another area is new design concepts, in particular building on the much-improved understanding of the electronic structure,225–227 and the emergence of new classes of more complex HH alloy compositions.112 The complex HH alloys behave opposite to the conventional XYZ compositions: they have low thermal conductivities but also relatively poor power factors. Uniting large power factors and low thermal conductivities remains the key challenge in HH materials development. The largely harmonic bonding limits the strength of phonon–phonon scattering, leaving structural disorder as the most plausible route to achieve low $\kappa_l$. A possible interesting route is to focus on grain boundary engineered composite samples, where the boundary region is exploited to suppress phonon transmission and reduce $\kappa_l$. This has been demonstrated for graphene decorated skutterudites and titanium oxides,228,229 whilst dislocation networks have been explored in Bi$_2$Te$_3$.230 In principle, these approaches can be applied to HH alloys, and the key challenge is to find appropriate additives and processing routes to manipulate the microstructure.

In terms of homogenous materials, advances might come from the design of materials with favourable bandstructures, enabling large power factors at lower doping levels, hence limiting the electronic thermal conductivity.227 One possible route to enhance $S$ without affecting $\sigma$ is the introduction of resonant states, leading to a “spike” in the electronic density of states. This concept has been explored for X$_9$NiSn with some evidence of a favourable impact on the power factor and thermoelectric performance.231,232 As already remarked, the similar conduction band structures for XYZ HH compositions limits opportunities for band engineering in the $n$-types. In the $p$-types, there is more diversity and hence more opportunity for bandstructure engineering. ZrCoBi and NbFeSb are both systems that naturally benefit from large valley and orbital degeneracies. Alloying has been widely explored to reduce the lattice thermal conductivity but can also lead to substantially different bandstructures. Recent work using unfolded supercell calculations has shown that new bands can emerge in substituted HH compositions (e.g. in vacancy systems,170 for interstitials58 and complex compositions233) and this is another possible route for performance improvements. This is an area where theory could lead experiments by identifying new compositions with promising electronic bandstructures.

To conclude, great progress has been made over the past decade, driven by improvements in materials preparation and through better understanding from first principles and semi-empirical modelling. Many avenues of research remain open for exploration in the future and some of these will undoubtedly lead to better materials performance. The current generation of materials is already good enough for niche applications in thermoelectric waste heat recovery. Improved performance, or increased penalties for carbon emissions, may lead to increased industrial interest for large scale application in thermoelectric power generation.

Author contributions

RJQ undertook the data analysis based on the literature and drafted figures and tables. JWGB supervised the work and acquired funding. Both authors contributed to writing the article.

Conflicts of interest

There are no conflicts of interest to declare.

Appendix A – supplemental equations

DFT computational papers will typically report the bulk ($B$) and shear ($G$) moduli, from elastic tensors ($C_{ij}$). The crystallographic density ($\rho$) is simply obtained following a sufficiently converged geometry optimisation of the unit cell. The longitudinal ($v_l$) and transverse ($v_t$) can be expressed as:$^{243,244}$

$$v_l = \left(\frac{3B + 4G}{3\rho}\right)^{\frac{1}{2}}$$  \hspace{1cm} (S1)

$$v_t = \left(\frac{G}{\rho}\right)^{\frac{1}{2}}$$  \hspace{1cm} (S2)

Experimentally, most reports on the acoustic properties by thermoelectric groups report both $v_l$ and $v_t$ alongside the average velocity of sound ($v_s$). The subscripts used to denote the sound velocities can vary between reports (the most common issue being transverse versus shear velocity of sound). Calculation of $v_s$ is typically achieved using the following expression:$^{244}$

$$v_s = \left[\frac{1}{3} \left(\frac{1}{v_l} + \frac{2}{v_t}\right)\right]^{\frac{1}{2}}$$  \hspace{1cm} (S3)

Although alternative averaging equations are sometimes used,$^{245}$ this equation is biased heavily towards the longitudinal contribution. The Debye temperature ($\theta_D$) can be calculated using the above values:$^{246}$

$$\theta_D = \frac{h}{k_B} \left(\frac{6\pi^2}{V_0}\right) v_s$$  \hspace{1cm} (S4)

where $V_0$ is the volume per atom. This equation can be found in a number of different forms, expressing $V_0$ in terms of the number of atoms, using Planck's constant and moving around the $2\pi$ factor, or even moving $v_s$ inside of the brackets. The final elastic property required to estimate Umklapp scattering is the Gruneisen anharmonicity parameter ($\gamma$), which quantifies bond anharmonicity. Different $\gamma$ can be found in the literature, but given the information we have available the acoustic $\gamma$ is used, calculated using the equivalent expressions:$^{113,245}$

$$\gamma = \frac{3}{2} \left(1 + \nu\right)$$  \hspace{1cm} (S5)

where $\nu = \frac{1 - 2\left(\frac{v_t}{v_l}\right)^2}{2 - 2\left(\frac{v_t}{v_l}\right)^2}$
where $\nu$ is the Poisson ratio. A more involved procedure to determine $\gamma$ uses the phonon bandstructure to calculate $\gamma(q)$, which can be done through DFT.\cite{247}

Within the SPB model, HH materials are typically treated with acoustic phonon scattering as the dominant carrier scattering mechanism. For clarity, the SPB model uses the following equations with only acoustic phonon scattering:\cite{91,248}

$$F_i(\eta) = \int_0^\infty \frac{e^z}{1 + \exp[z - \eta]} \, dz$$ \hspace{1cm} (S6)

$$S = \frac{k_B}{e} \left( \frac{2F_i}{F_0} - \eta \right)$$ \hspace{1cm} (S7)

$$n_H = \frac{8\pi}{3} \left( \frac{2m_{\text{DoS}}k_B T}{h^2} \right) \frac{3}{2} F_0^2$$ \hspace{1cm} (S8)

$$\mu^H = \frac{\pi e h^4}{\sqrt{2}(k_B T)^3} \frac{N_0}{E_{\text{ddT}} m^3/2 m^2_{\text{DoS}}} \frac{1}{F_0}$$ \hspace{1cm} (S9)

$$\mu^H_\text{al} = \frac{8e h^4}{3\sqrt{2}(k_B T)^3} \frac{N_0}{E_{\text{ddT}} x(1 - x)} \frac{1}{m^3/2 m^2_{\text{DoS}}} \frac{1}{F_0}$$ \hspace{1cm} (S10)

where $\eta$ is the reduced Fermi energy, $n_H$ and $\mu_H$ are the Hall concentration and Hall mobility determined directly from Hall measurements. $\eta$ is typically solved numerically from $S$. $m_{\text{DoS}}$ is typically determined from a series of samples at fixed temperature by plotting $S$ vs. $n_H$ in a Pisarenko plot and fitting for a fixed $m_{\text{DoS}}$, although a single datapoint can be used. The weighted mobility ($\mu_w$) is a measure of the fundamental mobility of a material independent of $n_H$ and can be defined as:\cite{248,249}

$$\mu_w = \mu_0 \left( \frac{m_{\text{DoS}}}{m_e} \right)^{3/2} \frac{N_0}{m^2_{\text{DoS}}}$$ \hspace{1cm} (S11)

where $\mu_0$ is the $n_H = 0$ limit within the SPB model. From this relationship, multiple highly dispersive bands are preferable to a single low dispersion band with the same $m^2_{\text{DoS}}$. A relationship compatible with $S(T)$ and $\sigma(T)$ is expressed as:\cite{249}

$$\mu_w = \frac{3h^3}{8ne(2m_e k_B T)^2} \left[ e \left( \frac{S_k^{\text{DoS}}}{S_k^{\text{DoS}} + 1} \right) \frac{3 |S|}{\pi^2 k_B/e} + \frac{1}{1 + e^{-S_k^{\text{DoS}}}} \left( \frac{S_E}{S_k^{\text{DoS}} + 1} \right) \right]$$ \hspace{1cm} (S12)

### Acknowledgements

The EPSRC (EP/N01717X/1) and Leverhulme Trust (RPG-2020-177) for funding and Heriot-Watt University is for a James Watt Scholarship for RJQ.

### References
