DOI: 10.1002/201800075

Article type: Full Paper

Database-driven Materials-selection for Semiconductor Heterojunction Design

Ethan P. Shapera and *André Schleife**

E. P. Shapera

Department of Physics, University of Illinois at Urbana-Champaign,

Urbana, IL 61801, USA

Prof. A. Schleife

Department of Materials Science and Engineering,

Frederick Seitz Materials Research Laboratory,

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign,

Urbana, IL 61801, USA

e-mail: schleife@illinois.edu

Keywords: Band alignment, Heterojunction, Light-emitting diode, Solar cell, Materials database

Heterojunctions are at the heart of many modern semiconductor devices with tremendous societal impact: Light-emitting diodes shape the future of energy-efficient lighting, solar cells are promising for renewable energy, and photoelectrochemistry seeks to optimize efficiency of the water-splitting reaction. Design of heterojunctions is difficult due to the limited number

of materials for which band alignment is known and the experimental as well as computational difficulties associated with obtaining this data. We show that band alignment based on branch-point energies is a good and efficient approximation that can be obtained using data from existing electronic-structure databases. Errors associated with this approach are comparable to those of expensive first-principles computational techniques as well as experiment. Branch-point energy alignment is then incorporated into a framework capable of rapidly screening existing online databases to design semiconductor heterojunctions. The method is showcased for five different prototype cases: transport layers are successfully predicted for CdSe- and InP-based LEDs as well as for novel $CH_3NH_3PbI_3$- and nanoparticle PbS-based solar absorbers. In addition, Cu_2O as a possible hole-transport layer for solar cells is examined. The framework addresses the challenge of accomplishing fast materials selection for heterostructure design by tying together first-principles calculations and existing online materials databases.

1 Introduction

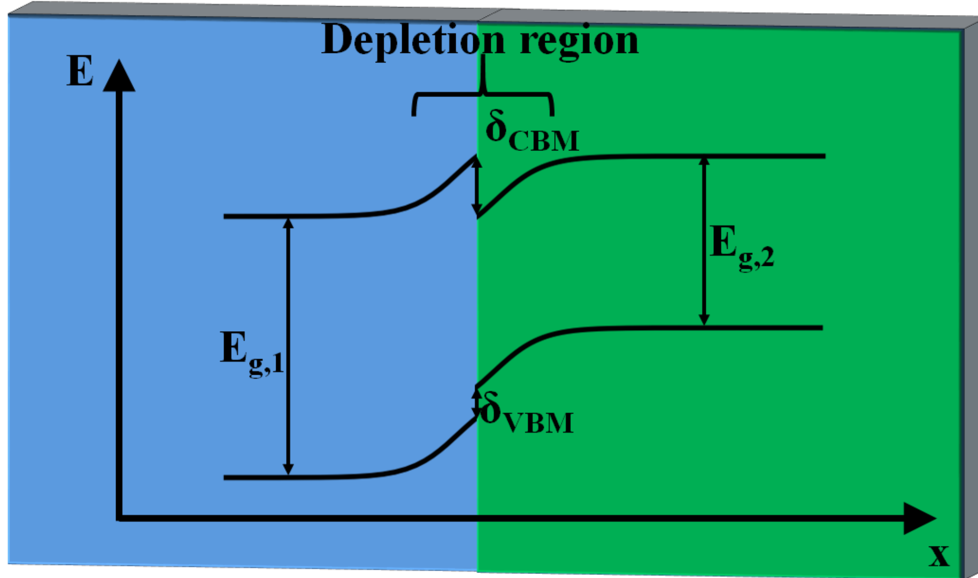

Figure 1: Schematic energy diagram of two semiconductors (shown in blue and green) with different band gaps $E_{g,1}$ and $E_{g,2}$ in contact. Far from the interface, outside the depletion region, electronic bands match those of the bulk materials. Inside the depletion region charge recombination due to different doping profiles causes band bending. At the interface, band-gap mismatch is accommodated via valence- and conduction-band offsets, δ_{CBM} and δ_{VBM}, that are determined both by the band-gap difference and energy alignment.

At the interface of two semiconductors, where bulk band structures of both materials merge into each other, an electronic transition region forms: Band bending due to different doping profiles across the depletion region extends significantly into both materials, while band-edge discontinuities are confined to not more than a few atomic layers near the interface (see e.g. Ref. 1 and the schematic in Figure1). These discontinuities, also known as valence- and conduction-band offsets δ_{VBM} and δ_{CBM}, naturally occur at the interface of materials with different band gaps. While their signs decide whether the interface acts

as a barrier or conductor for electrons or holes, their magnitudes determines how good of a barrier/conductor the interface is.

For any specific *ideal* interface, these offsets are determined by intrinsic material properties of the semiconductors in contact: The band-gap difference of the two materials, that is usually well-known for many semiconductors and insulators, fixes the *sum* of valence- and conduction-band offsets (see **Figure 1**). The relative energy positions of valence-band maxima (VBM) and conduction-band minima (CBM) at the interface are referred to as “band alignment” (or “energy alignment”) and are much harder to study both experimentally and computationally. As a result, band alignment is generally less well-known for many pairs of semiconductors. At the same time, the dependence of band alignment on intrinsic properties of the involved materials turns the design of heterojunctions with specific alignment into an interesting materials-design or materials-selection optimization problem.

Solving this ubiquitous problem of successfully designing heterojunctions with specific energy-level alignment has numerous applications in developing technologies of tremendous societal impact since it is as important as designing the active semiconductor material itself, when optimizing overall efficiency. For example, in the field of solid-state lighting, work on a (*n*-doped GaN)–(InGaN)–(*p*-doped GaN) heterostructure produced a blue light-emitting diode (LED) with greater than 50 % quantum efficiency.[2] For this, InGaN quantum wells were grown on the polar $(20\bar{2}\bar{1})$ surface of doped GaN to achieve large wave-function overlap between layers over a wide range of current densities. However, it is the band alignment between adjacent layers that determines the depth of the central quantum well: Smaller offsets facilitate flat-band conduction between layers. Another example are organic white LEDs, that use band alignment to ensure charge transfer from transport layers into the active layer while simultaneously blocking transfer of carriers in the reverse direction.[3] Finally, in the context of photoelectrochemistry, Hara *et al.* demonstrated a novel ($ZnRh_2O_4$)–(Ag)–($AgSbO_3$) heterostructure that is capable of both the splitting of water molecules and the reduction of hydrocarbons using visible-light illumination.[4] H_2 is

produced by a catalyzed reaction on the $ZnRh_2O_4$ surface and O_2 is produced on the $AgSbO_3$ surface. The central Ag layer mediates charge transfer between both separated surfaces and the relative positions of the valence and conduction bands force charge carriers to flow in the correct directions to drive the water-splitting reaction. It is, thus, not surprising that Yan *et al.* recently searched for anode materials for water splitting using band gaps and band alignment as filtering criteria.[5]

These examples impressively illustrate an imminent need for purposeful and efficient design of valence- and conduction-band offsets to match desired criteria that is even more exacerbated with the current advent of materials design: Excellent heterojunctions are particularly necessary to turn *new* active electronic materials into specific applications. Initially, optimal electron- and hole-transport layer (ETL and HTL) materials, that allow translating a new electronic material, discovered using materials design approaches, into devices such as photovoltaic absorbers or light-emitting layers, are unknown. Historically, experimentation based on trial and error has been used to solve this problem. However, the lack of a fast and reliable approach for purposeful selection of materials for semiconductor heterojunctions undeniably hampers progress towards better and more efficient semiconductor electronic devices. Being able to identify suitable band alignment to design semiconductor heterostructures around a fixed active electronic material by screening a large number of candidate materials is a desirable, but currently *unsolved*, challenge.

Screening a large number of material combinations for band offsets is very difficult in experiment, because sample-preparation conditions can sensitively affect quality and orientation of surfaces and interfaces. Carefully controlled experimental conditions similar to those in the semiconductor industry are needed, slowing down the process, rendering it expensive, and, thus, impractical to study large portions of the materials search space. In principle, computational screening of a large number of candidates is a promising alternative. Unfortunately, computational screening for candidate materials with specific band offsets is not straightforward either, since information on the band gap and on band alignment is needed, both of which are challenging, again leading to slow and expensive computational

studies: While highly accurate band gaps can be computed using techniques such as many-body perturbation theory,[6] hybrid exchange-correlation functionals,[7] time-dependent DFT,[6] or quantum Monte-Carlo calculations,[8] these schemes come at much larger computational cost than DFT and, hence, are not straightforwardly applicable to tens of thousands of materials. Quantum-mechanical first-principles calculations can also be applied to quantify band alignment at semiconductor interfaces by computing either absolute energy positions of atomic core-electron levels,[9] electronic transition levels of hydrogen impurities,[10] or the alignment of vacuum levels for the materials in contact.[11] While being successful and accurate, the disadvantage of these techniques is, again, their high computational cost.

Here we tackle the challenge of rapid screening for and identification of candidate materials for ETLs and HTLs that fulfill specific application-driven band offset criteria, by introducing a computational data-driven framework. Thanks to recent trends in materials science, data on atomic geometries and lattice parameters is readily available online, e.g. from ground-state density functional theory (DFT) calculations [12-15] and experiment.[16, 17] In this work, based on an existing, computationally inexpensive technique [18-20] we first approximate band alignment exclusively using bulk quasiparticle (QP) valence and conduction bands as critical input. This significantly reduces computational cost compared to the more accurate but more complicated first-principles techniques discussed above. We then modify this approach and interface it directly with a currently existing, digitally accessible electronic-structure database, which allows us to avoid additional band-structure calculations, again cutting computational cost significantly. Ultimately, this combination of techniques constitutes a computationally feasible, albeit approximate, route towards rapid and efficient materials selection for heterojunction design that is applicable to screen very large numbers of materials for suitable band alignment. Application of our framework enables more rapid development of widely used devices with higher efficiency, including semiconductor LEDs and novel photovoltaic absorbers for solar cells, and helps guiding the search for better photocatalytic water splitters. It can also facilitate fundamental research, e.g. by identifying charge accumulation near surfaces or yet unknown 2D electron gases.

In the following, we first describe our framework and then discuss verification and validation based on band alignment using explicit DFT calculations and comparison to the existing experimental and computational literature. We then integrate our framework into a materials-selection workflow and demonstrate that it can be successfully used for heterostructure design by applying it to LEDs and solar cells, as two important examples: ETLs and HTLs are proposed for CdSe and InP LEDs as well as solar cells using an organo-metal halide ($CH_3NH_3PbI_3$) and PbS nanoparticles as absorber or Cu_2O as HTL. All underlying input and output data and codes are made freely available online.[21,22] We envision that this addresses the long-standing problem of computing band alignment at a rapid rate and will significantly facilitate materials design of semiconductor heterojunctions. [18,20,23-26]

2 Methods

2.1 Band offsets

Valence- and conduction-band offsets (see Figure 1) are determined both by the band gaps of the semiconductors in contact, and by the relative energy alignment, which requires a common absolute energy level for both materials. In experiment, band offsets can be measured using X-ray photoemission spectroscopy to determine the position of the valence-band emission edge as a function of doping density and extrapolating that data to obtain E_{BP}.[27] First-principles techniques can be used to compute energy positions of atomic core-electron levels,[9] electronic transition levels of hydrogen impurities,[10] or vacuum levels for the materials in contact[11] and either of these can be used as absolute energy levels for band alignment. However, large simulation cells with tens or hundreds of atoms[28,29] (either including defects/impurities or material slabs and vacuum) and QP energies are needed[20] to accurately compute offsets from first principles. Such calculations are expensive even for specific *individual* interfaces, which is why this information cannot be produced in a high-throughput fashion, if at all, for large numbers of materials.

In the literature another approach exists[18-20] that uses the branch-point energy (E_{BP}), also referred to as charge neutrality level or effective mid-gap energy, as a universal energy level for band alignment, assuming negligible interface dipoles.[1] The E_{BP} can be entirely traced back to the bulk band structure of a given semiconductor.[18] Even though this is an approximation, it is advantageous in the context of this work, since it neglects complicated structural details of the interface that are oftentimes entirely unknown or hard to purposefully design in practice, or both. In order to predict band alignment, we then rely on an approximation to compute E_{BP},[18-20] that only requires the *dispersion* of Kohn-Sham eigenvalues, for which DFT results are oftentimes in good, albeit not perfect, agreement with experiment. Despite the approximate nature, it was shown in the literature and is verified here that the overall error bars of this technique are comparable to those of the first-principles vacuum-level alignment approach.[20]

Initially, E_{BP} was computed from energies at isolated **k** points[23] and, later, as a Brillouin zone (BZ) average based on the first of Baldereschi's special points.[30] Finally, a sum over a full Monkhorst-Pack[31] **k**-point grid was used to compute E_{BP} according to[19]

$$E_{BP} = \frac{1}{2N_{\boldsymbol{k}}}\left(\frac{1}{N_c}\sum_{c_i}^{N_c} \varepsilon_{c_i}^{\mathrm{QP}}(\boldsymbol{k}) + \frac{1}{N_v}\sum_{v_i}^{N_v} \varepsilon_{v_i}^{\mathrm{QP}}(\boldsymbol{k})\right) \tag{1}$$

In Equation 1 $N_{\mathbf{k}}$ is the number of points in the **k**-point mesh and the energies $\varepsilon_{v_i}^{\mathrm{QP}}(\mathbf{k})$ and $\varepsilon_{v_i}^{\mathrm{QP}}(\mathbf{k})$ describe the QP band structure of a material. N_c and N_v are the number of conduction and valence bands, respectively, included in the average of Equation 1 and these are parameters of this approach. Previously they were computed from the sum N_e of the numbers of *s* and *p* valence electrons per formula unit of the material, which led to an uncertainty on the order of 0.2 eV for E_{BP}.[19] Here we optimized N_c and N_v as free parameters by calculating the mean absolute error between the calculated and experimental E_{BP} for the 21 semiconductors listed in **Table 1**, for which high-quality experimental data is available. Using N_c=N_e/8 and N_v=N_e/4, a mean absolute difference of 0.22 eV from the experimental

values in Table 1 was found. Halving N_c and N_v produced a mean absolute difference of 0.31 eV, doubling these numbers led to a mean absolute difference of 0.73 eV, and using equal numbers for N_c and N_v also increased the mean absolute difference. Thus, in this work we use the pymatgen open-source library[32] to compute the sum N_e of *s* and *p* valence electrons and then set N_c=N_e/8 and N_v=N_e/4 as discussed above.

In this work, we carry out the summation in Equation 1 to validate against results computed using a reduced sum and data from an online database, as discussed below. To this end, we perform explicit DFT calculations for all materials listed in Table 1 using the Vienna *Ab-Initio* Simulation Package[33,34] (VASP) and the projector-augmented wave method[35] is used to describe the electron-ion interaction. The generalized-gradient approximation (GGA) of Perdew, Burke, and Ernzerhof (PBE) [36] is used to describe exchange and correlation. Kohn-Sham wave functions are expanded into a plane-wave basis with an energy cutoff set to 550 eV, which is just slightly larger than what is used by Materials Project.[12,32] Relaxed crystal structures are chosen identically to those used on Materials Project. Γ-centered **k**-point grids are selected by requiring E_{BP} to be converged within < 0.1 eV, which is achieved with $16\times16\times16$ grids for all materials listed in Table 1. Input and output files of these calculations are made accessible using the Materials Data Facility. [21,22]

2.2 Online materials databases

Computing band offsets based on E_{BP} spares us from high computational cost caused by large simulation cells, since the E_{BP} is computed entirely based on bulk band structures. One single DFT calculation per material is enough to evaluate the expression given by Equation 1. However, instead of restricting our materials search to only a small number of candidates and perform electronic-structure calculations within DFT for those, or performing large numbers of DFT calculations using e.g. high-throughput computing, we here pursue a different novel and promising strategy towards accelerating

materials design: The increasing availability of publicly accessible online databases, that have compiled experimental or computational data for numerous properties of tens to hundreds of thousands of materials, allows us to significantly speed up our search. In particular, the desirable balance of low computational cost and high accuracy of DFT[37] has led to projects that contain structural information and detailed electronic-structure data such as band structures and densities of states[12-15, 38] for materials that were synthesized before and are reported, e.g. in the Inorganic Crystal Structure Database.[39] Such DFT-based databases have previously been used to design scintillator materials,[40] to study optical effects in calcites,[41] to find trends in total energy and enthalpy of formation,[42] for Li-ion battery design,[43] to screen electrocatalytic materials,[44] and to identify novel solar-cell absorbers[45] to name a few applications most closely related to our work. However, due to the large computational cost of explicit first-principles calculations of band alignment, none of the currently existing online databases includes band offsets, nor the data needed for either of the above-mentioned first-principles approaches to compute those. This currently prevents direct search for electron- or hole-transport layer materials based on band offsets tabulated in databases.

In this work, we use the E_{BP} to compute band alignment and we render the screening of thousands of material combinations possible by directly incorporating electronic-structure data that is readily accessible in online DFT databases. Here, we specifically rely on explicit **k**-point dependent band-structure information from Materials Project,[12, 46] a currently existing and freely available online database with a Python API. Materials Project to date contains band structures for at least 66,676 materials, 45,148 of which are non-metals (according to our classification discussed below) with band structures available. We emphasize, however, that our computational framework is independent of the specific database, as long as electronic-band structure and crystal-structure data is available.

After interfacing our approach with Materials Project we are able to consider all materials as possible ETLs and HTLs for which this database provides band-structure data. In the following, we focus on the 45,148 non-metals, which also illustrates why high efficiency is particularly desirable when screening

for suitable materials for heterojunction design: The parameter-space becomes huge. Using the non-metals on Materials Project to build three-component heterostructures, where no two components are identical and the active material is in contact with exactly one electron- and one hole-transport layer, leads to a number of $45{,}148 \times {}^{45{,}147}C_2 \approx 4.6\times10^{13}$ possible permutations. Fixing the active material still leaves $\approx 10^9$ possibilities to find ETLs and HTLs. Direct computational simulation of all possible multi-component heterostructures is infeasible due to the rapidly increasing number of combinations of materials. Hence, even if our database- and E_{BP} -based framework is approximate in nature and, thus, has limited accuracy, its extreme speed makes it nevertheless promising e.g. to narrow the vast search space and to identify possible candidate materials.

2.3 Quasiparticle energies

Unfortunately, the vast majority of band structures stored in existing materials databases and, in particular, many, albeit not all, band structures obtained from Materials Project suffer from the notorious band-gap problem of DFT. As a consequence, there are currently only very few studies[47, 48] that explicitly use electronic-structure data available in databases for design of excited-state properties of functional materials. Since DFT data is not well suited to provide accurate QP energies that are directly comparable to experiment, it also does not provide $\varepsilon_{v_i}^{\mathrm{QP}}(\mathbf{k})$ and $\varepsilon_{v_i}^{\mathrm{QP}}(\mathbf{k})$ that are needed to evaluate Equation 1. This deficiency can be overcome by using more accurate computational techniques that better describe the electron-electron interaction,[6-8] however, these come at much higher computational cost.

In this work, we use a two-tiered approach to determine a scissor shift that rigidly shifts conduction bands to produce accurate band gaps: First, whenever available, the band gap is corrected to the experimental value. Values are taken from literature sources and must be provided manually by the user in a data file, such that users can readily incorporate their own band-gap data. For instance, some groups have used machine learning techniques to quickly improve estimation of the DFT band gaps compared to experimental values.[49-51] In this work, we obtain 327 experimental semiconductor band gaps from

Refs. 52-55. When no experimental data is available, we employ the linear band-gap correction described in Ref. 40. Both cases shift E_{BP} by half the band-gap correction, indicating that accurate data for band gaps is needed, since it linearly affects branch-point energy alignment. Combining our calculation of E_{BP} with this two-tiered gap correction allows for quick alignment while significantly improving accuracy over DFT band gaps.

Finally, all band structures available in Materials Project are provided only for bulk materials. In addition to the QP correction discussed above, when forming nanocrystals, the band structure is altered due to quantum confinement of electrons, since confining electrons to smaller regions increases the spacing between adjacent energy levels. In order to study band alignment with PbS nanoparticles, we approximate the change to the band structure as what would be found from a spherical potential well. To this end, we apply a scissor shift to widen the band gap using the Brus equation, which considers the effect of quantum confinement on the energy levels of a quantum dot. In this case,

$$\Delta E_{\mathrm{g}} = \frac{\hbar^2 \pi^2}{2R^2}\left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right) \tag{2}$$

where R is the radius of the nanocrystal and m_e^* and m_h^* are the electron and hole effective masses.[56] Here we focus on the case of PbS nanocrystals as active layer, and we use a radius of 5 nm as this is a typical scale discussed in Ref. 57. The electron and hole effective masses m_e^* and m_h^* are readily calculated from the Materials Project band curvatures at the CBM and VBM, respectively, using parabolic fits to the band structures. In cases where effective masses are direction dependent, values are averaged.

2.4 Ranking

No single commonly accepted computational criterion has been established for ranking the quality of multilayer heterostructures. Since in this work we focus on band alignment and the emerging transport

properties, we define a ranking criterion by considering a Drude model of semiconductors. Within this model, the resistivity of a material is

$$\rho = \frac{m^*}{ne\tau^2} \tag{3}$$

with carrier effective mass m^*, carrier density n and charge e, and carrier scattering time τ. For a material of thickness L, the resistance is $R=L\rho$. Assuming that the carrier density and scattering time are constant and that the thickness of the ETL, HTL, and active layer are equal, the conductance C of the entire heterojunction, based on Drude model conductivity, can be defined as figure of merit as

$$C = R_T^{-1} \propto \left(m^*_{e,\mathrm{ETL}} + m^*_{h,\mathrm{HTL}} + \frac{1}{2}\left(m^*_{e,\mathrm{Active}} + m^*_{h,\mathrm{Active}} \right) \right)^{-1}. \tag{4}$$

Electrons travel through the entire thickness L of the ETL and holes through the thickness L of the HTL. In the active layer, electrons and holes only travel from where the photon is absorbed (solar cell) or to where the electron and hole recombine (LED). The factor of $\frac{1}{2}$ for the active-layer contributions arises from electrons and holes that, on average, would each travel through half of the active layer.

3 Results and Discussion

3.1 Branch-point energies: High-symmetry paths

In order to compute E_{BP} for band alignment using data from Materials Project, we first investigate the error introduced by reducing the **k**-point sum in Equation 1 from one over the full Brillouin zone (FBZ) to one *only* over path segments that connect high-symmetry points (HSP). This is necessary since Materials Project currently only provides band structures along high-symmetry paths through the Brillouin zone, instead of the fully converged grid required in Equation 1. Summing only over the band-structure **k**-point path provided by Materials Project to compute E_{BP} turns Equation 1 into

$$E_{BP} = \frac{1}{2L_{\text{path}}} \sum_{j}^{N_{seg}} L_j \sum_{\mathbf{k}\in j} \frac{1}{N_{\mathbf{k},j}} \left(\frac{1}{N_c} \sum_{c_i}^{N_c} \varepsilon_{c_i}^{\text{QP}}(\boldsymbol{k}) + \frac{1}{N_v} \sum_{v_i}^{N_v} \varepsilon_{v_i}^{\text{QP}}(\boldsymbol{k}) \right) \quad (5)$$

where *Nseg* is the total number of segments connecting beginning and end of a band-structure plot (e.g. X−L−Γ−M corresponds to *Nseg*=3) and $N_{\mathbf{k},j}$ is the number of **k** points in the *jth* segment. L_{path} is the total length of the path, i.e., the sum of lengths of all segments, $L_{\text{path}} \sum_{j}^{N_j} L_j$. L_j values are calculated as the **k**-space distance between the end points of each segment.

Material	E_g	$E_{BP,MP}$	$E_{BP,FBZ}$	$E_{BP,Exp}$	$E_{BP,Theor}$[10]	$E_{BP,Theor}$[20]
Si	1.59[52]	0.49	0.39	0.36[58]	0.34	0.56
Ge	0.67[52]	-0.06	-0.04	0.09[59]	-0.25	0.01
CdO	1.91[52]	2.47	2.88	1.3±0.1[60]	–	–
KBr	7.56[52]	4.51	4.70	–	–	–
MgO	7.67[52]	5.21	5.48	–	–	–
AlAs	2.17[52]	0.90	0.96	0.92[58]	1.05	1.04
AlP	2.49[52]	1.25	1.25	1.13[58]	1.74	1.49
AlSb	1.73[52]	0.49	0.51	0.53[58]	0.08	0.19
CdSe	1.77[52]	1.56	1.71	1.83[61]	1.24	1.63
CdTe	1.56[52]	1.16	1.27	–	0.61	0.75
GaAs	1.43[52]	0.71	0.90	0.52[58]	0.46	0.45
zb-GaN	3.03	2.05	2.37	–	–	–
GaP	2.49[52]	0.77	0.81	0.83[58]	1.01	0.93
GaSb	0.76[52]	0.20	0.26	0.16[58]	-0.25	0.33
InAs	0.37[52]	0.36	0.46	0.50[58]	0.19	0.38
InP	1.38[52]	0.81	0.90	0.86[58]	0.65	0.82

InSb	0.21[52]	0.10	0.16	0.22[58]	-0.34	0.25
ZnS	3.73[52]	2.09	2.22	2.05[58]	2.04	2.23
ZnSe	2.71[52]	1.71	1.84	1.48[58]	1.45	1.75
ZnTe	2.27[52]	1.17	1.25	1.09[61]	0.50	0.71
HfO_2	5.9[53]	2.88	2.92	2.62[58]	–	–
AlN	5.33[52]	2.99	3.02	–	2.90	–
CdS	2.49[52]	1.92	1.97	–	1.70	1.97
wz-GaN	3.39[52]	2.38	2.40	2.51[58]	–	–
InN	0.70[54]	1.40	1.45	1.64±0.1[62]	1.91	
ZnO	3.38[52]	3.42	3.42	3.2[63]	3.49	3.01
Al_2O_3	7.96[52]	5.29	5.27	–	–	–
$LaAlO_3$	5.6[55]	2.47	2.48	–	–	–
PbI_2	2.26[52]	1.48	1.35	–	–	–
SiO_2	8.9[64]	5.36	5.43	4.9[58]	4.37	–

Table 1: Numerical data for E_g and E_{BP} (in eV) using the different approaches discussed in the text for 30 semiconductors (experimental results are available for 21 of them). "MP" stands for Materials Project. All values of EBP are given relative to the valence band maxima. $E_{BP,MP}$ are calculated along the high symmetry paths using Materials Project band structures. All listed materials are shown with experimental gaps except *zb*-GaN, for which we show the linearly corrected[40] DFT gap.[12, 46]

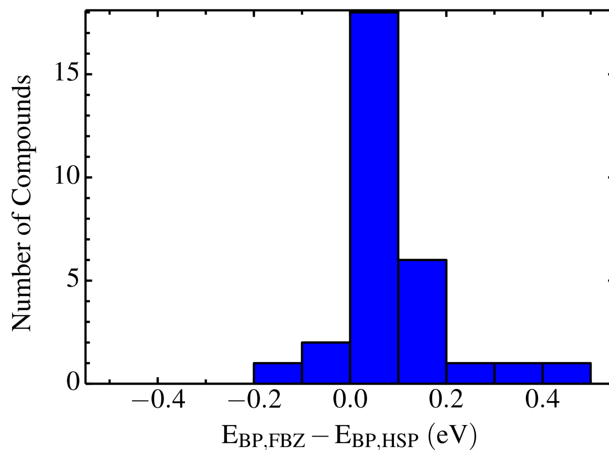

Figure 2: Difference of E_{BP} when computed from the full Brillouin zone (FBZ) and for high-symmetry **k**-point paths (HSP) only. The explicit numerical data is given in Table 1 as $E_{BP,MP}$.

Figure 2 compares E_{BP} computed using FBZ and HSP sampling from Materials Project data for the 30 common semiconductors listed in Table 1. This illustrates that summing only over high-symmetry paths produces errors in the band alignment of less than 0.4 eV for 29 of the 30 semiconductors studied here (only CdO has a larger difference of 0.41 eV), compared to FBZ sampling. In addition, the distribution is very narrow, such that the mean absolute error (MAE) is only 0.10 eV. In particular, 27 of the calculated E_{BP} are still within 0.2 eV of FBZ sampling. Such a small error introduced by our approach is tolerable for computing E_{BP} since it has the advantage that it can immediately be applied to existing, digitally available band-structure data along high-symmetry lines without any additional computational cost or performing additional DFT calculations. Thus, this technique is directly applicable to aligning valence and conduction bands of materials exclusively using the data stored in Materials Project and the extension to other databases[51, 65] is straightforward.

3.2 Branch-point energies: Validation against experiment and first-principles data

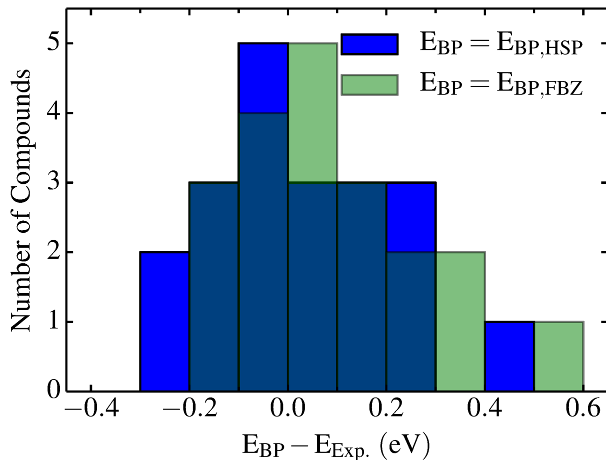

Figure 3: Difference of E_{BP} between experiment and HSP sampling using Materials Project data and, for comparison, FBZ sampling using our own VASP calculations. The explicit numerical data is given in Table 1.

Nevertheless, our approach to compute E_{BP} using band-structure data from Materials Project is approximate and, in particular, it neglects any interface-specific details of band alignment. For this reason, we validate our results against more accurate, less approximate first-principles data from Refs. 10 and 20. In addition to a dependence on the actual interface, experimental values for EBP also depend on sample preparation and quality. Thus, we expect a spread in the resulting experimental values and also validate MAEs against experimental data from Refs. 58-63.

To this end, **Figure 3** shows the difference between E_{BP} obtained using our technique and experimental data for 21 semiconductors (all rows with experimental results in Table 1). The distribution is centered around deviations of ≈0.1 eV, with the largest outlier showing 1.17 eV deviation, and a mean absolute error of 0.19 eV. 19 of the 21 materials show an absolute deviation of ≤0.3 eV compared to experiment. Two materials show particularly large differences: CdO (1.17 eV, not shown in plot) and SiO_2 (0.46

eV). For comparison, Figure 3 illustrates again that switching to FBZ sampling does not change the distribution significantly, and the MAE in this case amounts to 0.22 eV.

Vacuum-level based band alignment for 17 of the 30 materials in Table 1 was investigated from first principles[20] using QP energies from the $GW\Gamma^1$ approximation to determine the ionization potential and electron affinity, leading to a MAE of 0.28 eV for EBP relative to experimental values listed in Table 1. Using our approach based on Materials Project data, we find a MAE of 0.12 eV for these same materials[20] at tremendously reduced computational cost. Similarly, hydrogen-level alignment[10] leads to a MAE of 0.32 eV compared to experiment, while our method for the materials studied in Ref. 10 has a MAE of 0.31 eV.

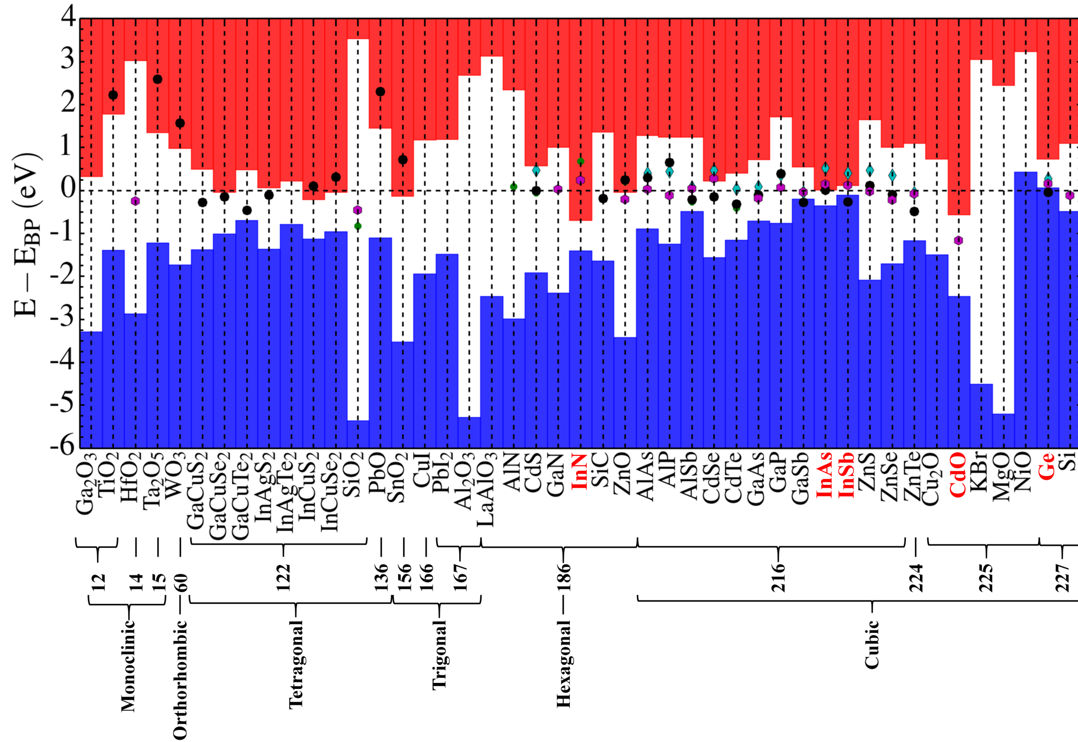

Figure 4: Band alignment for a set of 45 common semiconductors, arranged by crystal symmetry group number with crystal system shown below. E_{BP}, computed for the high-symmetry path from Materials Project using Equation 5, is used as energy zero. Blue and red bars indicate valence and conduction bands, respectively, the white space in between is the band gap. Materials labeled in red are predicted to be gapless within DFT, prior to application of the gap correction. Green dots are computational results from Ref. 10. Cyan diamonds are computational results from Ref. 20. Purple hexagons are experimental results, see Table 1 for references. Black circles are experimental results from Ref. 66. E_{BP} was calculated from the data in Refs. 10, 20, and 66 by designating Si as a reference material. E_{BP} was then set to 0.49 eV above the Si VBM, matching the E_{BP} calculated from the Materials Project band structure of Si. E_{BP} values for the remaining materials were found from the energy difference between the reference E_{BP} for Si and each material's VBM.

Next, in **Figure 4** we compare our results for E_{BP} for a larger set of 45 common semiconductors to literature data. As can be seen from the various data points that cluster around the 0 eV line, our approach

agrees reasonably well with other techniques and experiment. More specifically, the purple dataset illustrates all experimental data of Table 1 and shows a MAE of 0.20 eV with respect to our data. For the materials listed in Ref. 66, shown as black circles in Figure 4, the MAE between our computed values and experiment is 0.52 eV.

Figure 4 also shows that for two materials with heavy elements, Ta_2O_5 and PbO, the branch-point energies disagree with measured values by more than 2.5 eV. Our first-principles results show that inclusion of spin-orbit coupling in the band-structure calculations for these materials lowers the branch-point energies by 0.03 eV and 0.06 eV, respectively, and, hence, has no significant influence. In the case of five materials (CdO, PbO, SiO_2, Ta_2O_5, and TiO_2) for which computed and experimental EBP also agree poorly, we use QP energies computed using the HSE06 hybrid functional[67-69] and apply a scissor shift to correct the band gaps to the experimental values. Use of the HSE06 functional improves agreement only in the case of Ta_2O_5, by increasing E_{BP} by 0.02 eV. For all other examined materials the calculated E_{BP} is further from the experimental value and the largest shift is observed in PbO in which E_{BP} is decreased by 0.19 eV. Since in both the GGA and HSE06 calculations the band gaps are corrected to the same experimental values, the observed shifts of E_{BP} are due to changes in band dispersion. More specifically, the HSE06 functional better reproduces the energy dependence of QP corrections, which leads to larger shifts the higher or lower states are in energy, compared to states near the VBM and CBM.[24, 70] Overall this shows that both spin-orbit coupling and QP energy corrections to the band structure are ruled out as sources of disagreement between the experimental and computed values. In a future study, we plan to explore this disagreement for specific materials more systematically, by performing first-principles vacuum-level alignment calculations. However, this goes beyond the scope of the present work. Furthermore, we note that comparison to individual experimental data points is difficult, as can be seen from the scattering in Figure 4, highlighting the above-mentioned sensitivity of band alignment to specific experimental conditions. More specifically, in the case of CdO, Piper *et al.*[71]

reported charge accumulation on the surface, which is consistent with our result of a branch-point energy laying above the CBM.

In summary, this leads us to conclude that our technique is promising because of its overall satisfactory accuracy that is comparable to other standard techniques used in the literature. Given its extreme speed and applicability we can now compute E_{BP} and band alignment exclusively using electronic-structure data available in online databases, which enables a fast initial screening for very large numbers of materials. However, the agreement is not perfect and the error bars discussed above need to be kept in mind for applications, e.g. by refinement using more accurate but more expensive follow-up first-principles calculations for the most promising candidate materials. We note that we also explored a suggestion by Moench (Ref. 1), which uses the electronegativity difference of the two materials forming the interface to compute an interface-dipole correction. Unfortunately, we found that this approach cannot systematically improve the results and MAEs for the compounds shown in Fig. 4 indicated either no improvement, or even increased upon inclusion of this dipole correction.

3.3 Materials selection: Heterojunction design framework

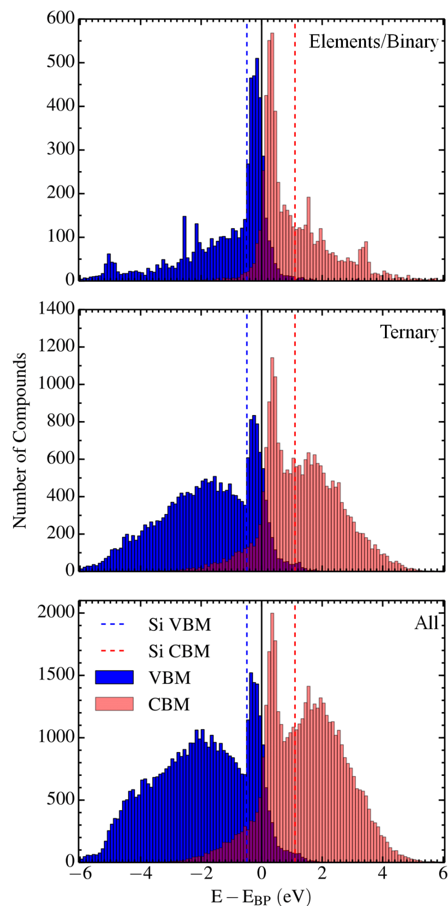

Figure 5: Distributions of VBM and CBM using branch-point energy alignment and band-gap correction for 5,824 elemental or binary semiconductors (top), 21,202 three-element semiconductors (middle), and all 45,148 semiconductors (bottom) in Materials Project. Only filtering to remove metals has been performed. The VBM (blue dashed line) and CBM (red dashed line) of silicon are included for reference.

Using branch-point energies to align VBM and CBM of all 45,148 non-metals on Materials Project, leads to the distribution depicted in **Figure 5**. This figure shows that there are at least 500, and in many cases more, materials available for any branch-point energy between 0 and 3.5 eV below the CBM as well as 0 and 4.3 eV above the VBM. This illustrates that many possible candidate materials are identified by our approach for a given active material in a heterojunction and desired band alignment.

At the same time, Figure 5 suggests a possible rough band-alignment design criterion: Among the elemental and binary materials, the distributions of the VBMs and CBMs peak at −0.2 eV and 0.3 eV, respectively. For the ternary and all materials, the distributions become bimodal with secondary peaks at ≈−2.2 eV and ≈ 1.5 eV. Hence, for applications requiring the VBM or CBM to be near EBP, elemental or binary materials should be considered first, while many more materials with three or more elements are available for structures requiring the VBM or CBM to be further from EBP. This guidance is beneficial when several criteria need to be fulfilled simultaneously, for instance when lattice matching is needed for heterostructure design.

Furthermore, Figure 5 illustrates that only a small fraction of materials has a branch-point energy that is located outside the band gap. As discussed above for CdO, it can be interpreted as an indicator for surface-charge accumulation if the branch-point energy occurs either in the valence or in the conduction band. Examples for materials where this occurs, as predicted by our approach, include $ZnGeO_3$, SnO_2, $AlCuO_3$, and In_2O_3. Indeed, charge accumulation has been measured experimentally in In_2O_3 by King *et al.*[72] Another material where charge accumulation was reported previously is InAs[73, 74] and our approach predicts the branch-point energy 0.01 eV underneath the CBM. Similarly, the branch-point energy lying below the VBM may indicate hole accumulation and predicted example materials include Sr_2VO_4, Ti_2O_3, $TiCl_3$, and GeAs. Extracting this information from our results provides a feasible route to further verify branch-point energy alignment experimentally and also to identify materials that show a tendency for either intrinsic *n*- or *p*-type behavior.

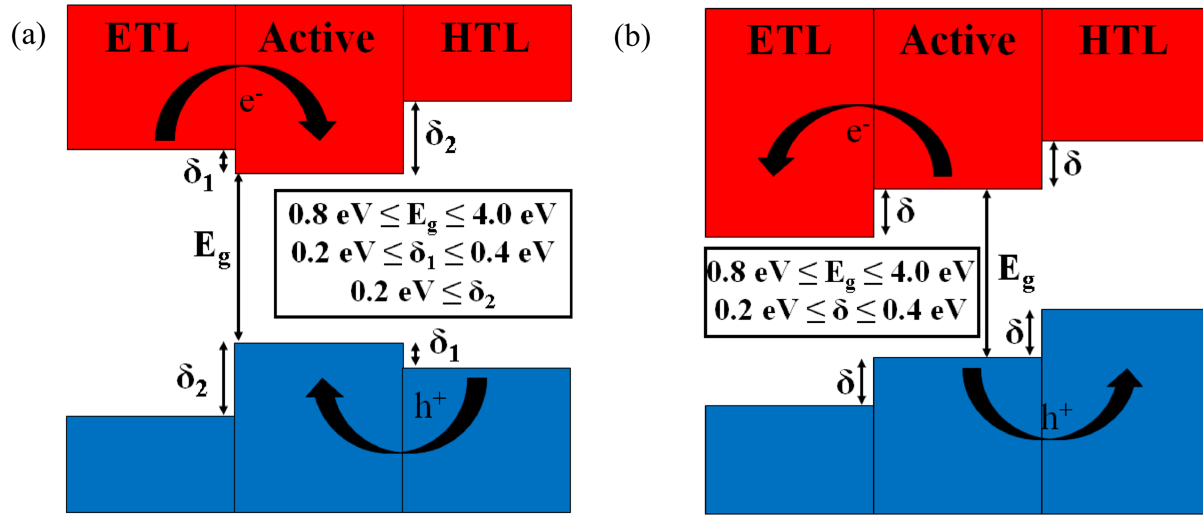

Figure 6: Band structure configurations utilized in this work to screen for (a) LEDs (CdSe, InP) and (b) solar cells ($CH_3NH_3PbI_3$, PbS nanocrystal, Cu_2O). Black arrows indicate direction of intended carrier flow in devices for each structure. Blue and red bars indicate valence and conduction bands, respectively, the white space in between is the band gap. The offsets needed to block flow are obtained by considering charge carriers to follow a Boltzmann distribution. At an operating temperature of 60 °C, an offset of 0.2 eV is sufficient to block 99.9% of thermal carriers. Limiting the size of offsets for bands relevant to conduction prevents the accumulation of charge at the interfaces.[75]

In the following, however, we apply our technique for computing band alignment to study LEDs and photovoltaic cells as proof of concept applications for heterojunction design. The design in both is similar,[76] in that an active component, the emitter in an LED or the absorber in a solar cell, is selected first, based on application- or fabrication-driven criteria. The shape and material of the active component must be chosen to maximize photon absorption/emission, charge separation, and charge transport. This layer is then sandwiched by ETL and HTL materials (schematically shown in **Figure 6**) that usually can be picked more freely, but must transport charge carriers without recombination.[77, 78] Most importantly,

ETL and HTL must demonstrate efficient electron/hole transport between the photolayer and other connected circuitry and act as barriers in the opposite direction. To study LEDs in this work, we pick CdSe and InP as emitter materials, since both are commonly used in practice which makes them ideal test cases. For solar cells we focus on a new and promising hybrid perovskite $CH_3NH_3PbI_3$ absorber and also study PbS nanocrystals. Finally, we look at a variation of the approach by studying solar cells but fix the HTL material as Cu_2O.

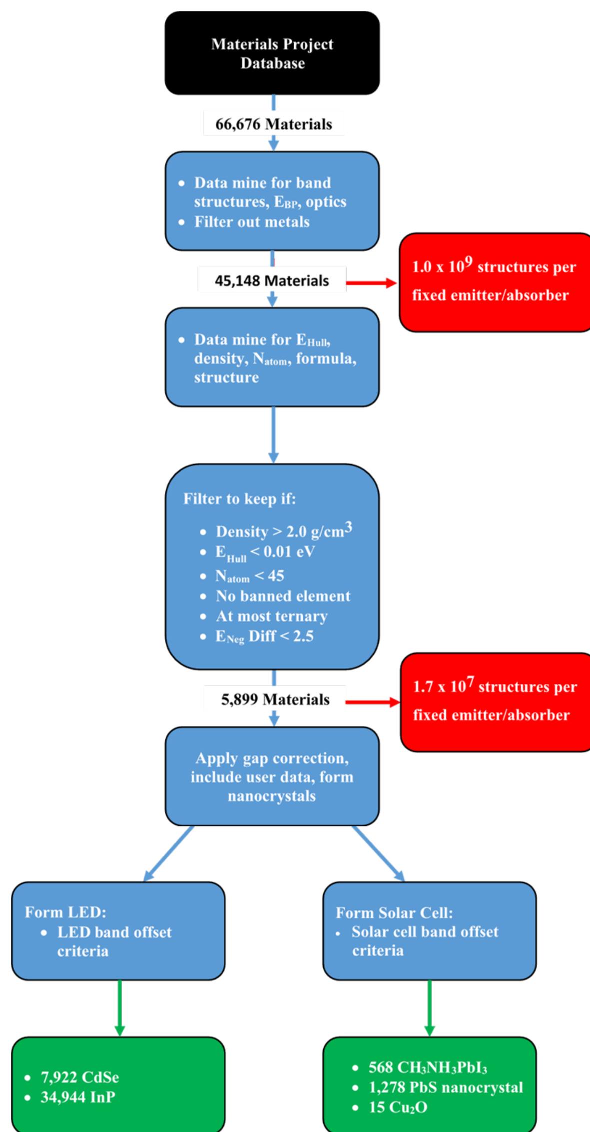

Figure 7: Flow chart used to filter for ETL and HTL materials for LEDs and solar cells. Listed in the bottom two boxes are the numbers of predicted heterostructures for each composition.

In order to design semiconductor heterojunctions, we introduce a scheme that uses "filters" to identify materials that are relevant out of all non-metallic materials available in Materials Project. DFT electronic band structures are queried sequentially based on their "Materials Project ID" using the pymatgen library. The band structures are then used to compute E_{BP}, E_{CBM}, E_{VBM}, m_e^*, and m_h^*. The overall approach is visualized diagrammatically in **Figure 7** and explained in the following.

Since we focus on search for ETL and HTL materials, the first step (see Figure 7) is to filter out metals. To this end, the pymatgen library averages the energies of each band and selects the band with the highest average energy below the Fermi energy as the highest valence band and lowest average energy above the Fermi energy as the lowest conduction band. The library then provides the maximum of the highest-energy valence band as E_{VBM} and the minimum of the lowest energy conduction band as E_{CBM}. In general, metals are characterized by $E_{CBM} - E_{VBM} \leq 0$ eV and semiconductors are gapped, i.e. $E_{CBM} - E_{VBM} > 0$ eV eV. However, since DFT underestimates band gaps, small-gap semiconductors can appear gapless in the Materials Project data or negative values of $E_{CBM} - E_{VBM}$ can occur, e.g. for indirect-gap semiconductors. For this reason we apply the gap correction scheme first and then retain materials with $E_{CBM} - E_{VBM} > 0$ eV as semiconductors. We use the condition $E_{CBM} - E_{VBM} \leq 0$ eV (after gap correction) to designate metals and to exclude them from further consideration in this work.

We note that, in general, this approach has difficulty distinguishing between small-gap semiconductors which appear gapless due to the DFT band-gap error and metals with flatter band dispersion near the Fermi level. Using the criterion described above, we attempt to not accidentally exclude semiconductors, however, ultimately, this needs to be fixed manually: Spuriously included metals must be removed by the user. In the particular heterostructure alignments studied in this work, the issue of distinguishing between metals and small-gap semiconductors is avoided since the offset criteria (see Figure 6) effectively constrain the band gaps of the transport layers, that are adjacent to a fixed active layer, to within a certain range. As an example, consider a CdSe emitter LED with band alignments shown

in Figure 6(a). CdSe has an experimental band gap of 1.77 eV. The band alignment criteria then require both the ETL and HTL to have band gaps of at least 1.87 eV. For a materials to satisfy $Eg \geq 1.87$ eV, it would either need to have an experimental band gap of at least 1.87 eV or if an experimental gap is not available, a DFT calculated band gap of at least 0.71 eV prior to linear correction.

The next stage (see Figure 7) applies filters to the remaining non-metals. Materials that are not solid under ambient conditions (e.g. Materials Project ID mp-20066 corresponds to CO_2) were filtered out using the density stored in Materials Project, only keeping those with density >2 g/cm^3. Unstable materials are removed by retaining only those with energies above Hull, E_{Hull}, that are smaller than 0.01 eV. While stability requires E_{Hull}=0 eV, we allow for materials with small nonzero E_{Hull}, due to possible errors of the total energy deriving from the approximate treatment of exchange and correlation within DFT. The specific value is readily adjustable by the user. Viability of fabrication is taken into account by imposing that (i) compounds contain at most three different atomic species, and (ii) unit cells contain at most 45 atoms. Pauling electronegativity differences were used to exclude water soluble compounds. Our filter allows a maximum difference in electronegativity of 2.5 between the most and least electronegative element in a material. Finally, compounds which contain radioactive elements (Ac, Th, U, Np, Pu, Tc, Pm) or noble gases are filtered out as well. After applying all filters, band gaps are corrected using the two-tiered approach discussed above. In the final stage before constructing heterostructures users may specify compositions and radii of nanocrystals and in this case, the band gap and E_{BP} is further corrected using Equation 2.

This entire process, schematically shown in Figure 7, is implemented as jupyter notebook,[79] which is compatible with the pymatgen library, and allows users to easily alter, add, or remove any of the filters. It also makes selecting the number of materials and the desired target band offsets (discussed below) between materials very easy. At the same time, using this code to down-select the maximum possible number of 10^9 heterostructures discussed earlier did not take longer than 10 minutes on a single CPU.

This notebook is also available in the Materials Data Facility.[21, 22] Some of the inaccuracies of the input to our framework, such as the linear band-gap correction and the neglect of interface induced perturbations of the atomic and electronic structure, can be addressed using the following multi-step procedure: First, we apply branch-point alignment to reduce the number of candidate materials from many thousands to a manageable count. For this step, band-alignment criteria should be adjusted to be not too restrictive, to not exclude too many promising candidates. Once the list has been reduced to tens to hundreds of materials, more accurate but more expensive first-principles calculations can then be applied to refine the band-gap information as well as to study atomic rearrangements and perturbations of the bulk electronic structure near the interface.

3.4 Materials selection: Heterojunction design results

The remaining candidate list comprises of 1,731 binary compounds or 5,899 materials if ternaries are included. As discussed above, E_{CBM}, E_{VBM}, E_{BP}, E_{BP}, m_e^*, and m_h^* for each compound were determined. The energies of the VBM and CBM relative to E_{BP} are used for determining offsets between materials. The effective masses determine the conductive figure of merit in Equation 4. We then use this information to design possible three-component heterostructures as showcase applications and to demonstrate our framework. Two heterojunction LEDs, two photovoltaic absorber materials, and one photovoltaic HTL material are used as examples in this work.

Literature data is not consistent on the issue of exact absolute positions of bands that are ideal for ETLs and HTLs in these devices and we use the band-alignment criteria illustrated in **Figure 6**. They require the CBM of the ETL to be between 0.2 eV and 0.4 eV above the CBM of the active layer. These offset values are taken from Ref. 80 which measured an offset of 0.2 eV would produce a near ideal $I{\sim}V^2$ at low applied bias with increasing offset requiring larger biases to function efficiently. In

experiment cases were reported where CBM of the ETL is below[81] or above[82] the CBM of the active material. This can be readily adjusted in the code. In the example of a Cu_2O HTL solar cell, no explicit condition is placed on the band gap of the active layer. Instead, these follow from the band gap of Cu_2O and the chosen alignment criteria: For Cu_2O, with a band gap of 2.23 eV,[52] the offset criteria shown in Figure 6 imply that the active layer must have a band gap between 2.03 eV and 2.43 eV. Band gaps in this range would allow for absorption of part of the visible spectrum and, even if not optimal, make this a reasonable test case for photovoltaic applications in this work.

Example	Binary	Ternary
LED (CdSe)	264	7,922
LED (InP)	1,764	34,944
Solar Cell ($CH_3NH_3PbI_3$)	144	568
Solar Cell (PbS nanoparticle R = 5 nm)	212	1,278
Solar Cell (Cu_2O HTL)	4	15

Table 2: Number of possible three-component heterostructures that satisfy all criteria applied for our showcase applications. The ternary column includes all binaries.

For all five examples, the number of heterostructures which pass all filters and satisfy the band-alignment criteria in Figure 6 are given in **Table 2**. **Table 3** and **Table 4** show several promising examples, either based on their figure of merit or since they have been successfully used in the experimental literature before. The entire list of results is available on the Materials Data Facility.[21, 22] As would be expected, significantly more heterostructures are predicted if ternary compounds are allowed in addition to pure elements and binaries. The number predicted further depends on how many materials in Materials Project have the correct band alignment and the range of band offsets allowed. The band alignment conditions in this work are stricter for the solar cells than LEDs. As a result, there are fewer predicted solar cells than LEDs.

Example	ETL	Active	HTL
LED	Ca_3N_2, Mg_3N_2, $ZnSeO_4$	CdSe	CdS,[83] WO_3,[84] MoO_3,[85] SiC[86]
LED	ZnSe,[87] CuI, $NaBiS_2$[88]	InP	GaSe, V_2O_5,[89] $SrCuO_2$[90]
Solar Cell	WO_3, In_2S_3[91]	$CH_3NH_3PbI_3$	Mn_3O_4, Cr_2O_3[92]
Solar Cell	PdS, Fe_3Si[93]	PbS, R = 5 nm	MnP, Cr_3S_4
Solar Cell	$Ca_2Cu_2O_5$	CdS[94]	Cu_2O

Table 3: List of promising examples picked from the final results. Materials listed here were chosen due to either high figures of merit, Equation 4, or because they appear in the experimental literature (indicated by the corresponding reference) performing the listed function in heterostructures.

Example	ETL	Active	HTL
LED	CdS[95]	CdSe	LaB_6
LED	$Ba_3(SnP_2)_2$	InP	TePb
Solar Cell	CdSe[96]	$CH_3NH_3PbI_3$	$KAuI_3$
Solar Cell	Cs_3Bi	PbS R = 5 nm	Sb_3Pd_8
Solar Cell	CuBr	PdI2	Cu_2O

Table 4: Top-ranked heterostructures using the figure of merit in Equation 4. Materials followed by a reference are found in the experimental literature performing the listed function in heterostructures.

In the following, we discuss the examples listed in Tables 3 and 4 as potentially viable heterojunctions. Starting with ETLs, we find ZnSe to be suitable for the InP LED and it has indeed been used in experiment for electron conduction.[97] $NaBiS_2$ is the third-highest ranked ETL for InP LEDs and it appears as a possible battery anode material in the literature.[88] We predict CdSe to be the highest performing ETL for $CH_3NH_3PbI_3$ and it has also been used in experiment for a perovskite solar cell.[96] Another example is In_2S_3, which was identified by our approach and, using our figure of merit, is the fourth highest ranked ETL for $CH_3NH_3PbI_3$ solar cells. It has been reported as a viable ETL for $CH_3NH_3PbI_3$.[91] In addition, we predict CdS to be an ETL for CdSe and it has appeared in the literature before in this context.[95]

Finally, for PbS nanocrystals the fourth ranked ETL is Fe_3Si and it is identified as a transport layer in magnetic tunnel junctions owing to its large spin polarization of electrons near the Fermi energy.[93] These examples clearly demonstrate that our framework can successfully identify important ETLs that were implemented in practice.

Among the HTLs listed in Table 3, amorphous SiC has been used for hole conduction before.[86] For CdSe LEDs, the commonly used HTL materials WO_3[84] and MoO_3[85] are both predicted to have suitable band alignment to act as HTLs. For the LED test cases studied in this work, most of the top-ranked (based on the figure of merit) ETLs or HTLs, such as $Ba_3(SnP_2)_2$, LaB_6, and TePb, are not reported in the experimental literature. While this could be due to the simplicity of our figure of merit used for ranking, or the approximations used in our approach, it may also mean that they are actually better ETLs or HTLs but have yet to be tested. It would, therefore, be most interesting to investigate their viability in experiment to test our predictions. Finally, for the Cu_2O based solar cell, our algorithm predicts PdI_2 as an active layer (see Table 4), but it does not appear in the literature as a solar cell absorber. It has an indirect band gap of 2.17 eV and a direct band gap of 2.18 eV, using the linear band-gap correction. DFT results on Materials Project predict flat bands near the VBM and CBM with high electronic densities of states. In addition, PdI_2 would not be viable for large scale application due to the high cost of Pd. Contrary, the predicted second highest figure of merit structure for Cu_2O solar cells contains CdS as an absorber. CdS has been reported as a solar cell absorber in the literature with a Cu-containing HTL.[98]

Conversely, some experimentally known ETLs or HTLs are not present in our list. The reasons for this can be, generally, grouped into three different categories, that will be discussed next:

(i) Band gaps: As discussed above, accurate band gaps are needed, because they directly enter the branch-point energy and, thus, the alignment and selection criteria (see Figure 6). While this is a problem in general, it is not a problem intrinsic to our approach, since we explicitly allow the user to provide

accurate band-gap data and only resort to a simple correction scheme if such data is absent. Incorporating more reliable band-gap data into our scheme is straightforward.

(ii) Branch-point alignment: For TiO_2, which is known as an ETL,[99] the band gap is corrected to the experimental value of 3.2 eV, however it is still not listed as viable here for CdSe- or InP-based LEDs. The reason is that our branch-point energy alignment predicts TiO_2 to have a higher VBM than CdSe. This contradicts Ref. 66 which measures the VBM of TiO_2 to be 2.21 eV below the VBM of CdSe and the CBM of TiO_2 in the gap of CdSe, making TiO_2 a viable ETL for a CdSe LED. The band alignment predicted by our approach incorrectly implies that holes could escape the active layer through the TiO_2 layer, lowering device efficiency. We also do not identify TiO_2 as a suitable ETL for InP because the CBM of TiO_2 is 1.2 eV above the CBM of InP, based on branch-point energy alignment. In Ref. 66, the CBM of TiO_2 is found to be 1.03 eV below the CBM of InP and the VBM of TiO_2 2.82 eV below the VBM of InP. We plan to investigate the origin of disagreement between branch-point energy alignment and experiment, e.g. for TiO_2, in more detail. One possible reason is charge transfer at interfaces of semiconductors[1] and in future work we will explore this in more detail. Independent of this, the user can easily incorporate their own, potentially more accurate, band-alignment data into our framework and still use it for fast identification of candidate materials.

(iii) Alignment criteria: While MoO_3 and WO_3 are predicted to be HTLs for CdSe LEDs, neither is predicted suitable for use with InP. Both are excluded because their VBM offsets are more than 0.4 eV below the VBM of InP: The VBMs of WO_3 and MoO_3 occur 0.92 eV and 1.03 eV below the VBM of InP, respectively. Similarly, for another experimentally known ETL, ZrO_2,[100] the branch-point energy alignment computed in this work predicts a higher energy CBM and lower energy VBM than CdSe and InP, which would make it possible to act as an ETL. However, the CBM of ZrO_2 is 2.59 eV above the

CBM in CdSe and 2.37 eV above the CBM in InP and, thus, outside of our allowed ranges (see Figure 6). CuO, as another viable HTL,[89] is omitted even if the gap is adjusted to the experimental value of 1.35 eV.[101] For a material to be a transport layer according to our criteria (see Figure 6), its gap must be at least 0.4 eV larger than the active material, but the experimental gaps are 1.77 eV in CdSe and 1.38 eV in InP.[52] NiO[102] is excluded due to its VBM being 1.99 eV above the VBM of CdSe and 1.24 eV above the VBM of InP. Again, this is easy to address and while in this work we follow the recommendation of Ref. 80 for band-alignment criteria, these ranges can be readily adjusted to the user's discretion.

A combination of these reasons explains why the (ZnO)–(PbS nanocrystal)–(NiO) solar-cell heterostructures reported in Ref. 57 are not identified in our work: Using band alignments calculated here, the CBM and VBM of ZnO are 0.05 eV and 3.01 eV below the PbS CBM and VBM, respectively. The PbS CBM and VBM are 3.14 eV and 0.84 eV below the CBM and VBM of NiO, respectively. Our prediction of the PbS nanocrystal CBM lying above the ZnO CBM is consistent with the alignment found in Ref. 57. However, alignments between PbS and NiO differ significantly with Ref. 57 calculating the PbS VBM to be within the possible ranges of NiO VBM energies, while our alignment predicts the NiO VBM to be 0.84 eV above the PbS VBM. In addition, there is a further issue that the VBM of NiO is 0.34 eV above the CBM of the PbS nanocrystal. With such a configuration, electrons would be able to flow from the NiO valence states into the PbS conduction states. Such a configuration would not be suitable for a solar cell, but could be applied to constructing a semiconductor laser. In all these cases, merely adjusting the alignment of the bands of the three materials would not be sufficient to meet the band offset conditions described in Figure 6 due to the implicit requirements for the material band gaps discussed above: According to Ref. 57, the PbS nanocrystals have a band gap of 0.50 eV. In order to satisfy the corresponding alignment criteria in Figure 6, the ETL and HTL band gaps must be between 0.30 eV and 0.70 eV, but ZnO and NiO have band gaps of 3.38 eV and 2.81 eV, respectively.[52] Hence, the experimentally studied heterostructure is also excluded due to our choice of alignment criteria.

Overall, the different examples discussed above clearly show that several of the materials identified by our framework have successfully been used before in practice, which illustrates that it is successful for heterojunction design. Several, so far untested, possible candidates are predicted and require experimental testing to judge whether they would work in practice and, potentially, even improve over state-of-the-art materials. Three sources of uncertainty are identified here as reasons for why several materials that are used in practice do not occur in our results and we discuss steps to fix those.

4 Conclusions and Outlook

We described, validated, and demonstrated a computational framework for fast screening of large numbers of materials in order to achieve efficient computational design of semiconductor heterojunctions. Its computational efficiency derives from (i) approximating band alignment using branch-point energies and (ii) directly incorporating band-structure data from the Materials Project online database. We confirm for common semiconductors that branch-point energy alignment leads to results with mean absolute errors relative to experiment that are comparable to those of more accurate but more expensive first-principles techniques. We further show that this approach can be directly applied to existing, freely available band-structure data without a significant sacrifice in accuracy. This eliminates the need to perform additional electronic-structure calculations for materials that are already part of Materials Project and the extension to other databases is straightforward.

We then use this approach in a materials selection framework to predict three-component heterostructures as showcase applications: Heterojunctions are designed for CdSe and InP based LEDs, for hybrid perovskite $CH_3NH_3PbI_3$ and PbS nanocrystal photovoltaic absorbers, and Cu_2O as hole-transport material for photovoltaics. Each of these takes only minutes on a single CPU and we rank results using the Drude model conductivity as figure of merit. Comparing our predictions to literature data shows that we are able to successfully identify electron- and hole-transport layers that are used in practice. Cases that are not successfully predicted are traced back to one of three different causes and strategies are

outlined to address these shortcomings. Several candidates are predicted that are currently not used in heterojunctions and we encourage experimental work to measure performance of those materials in the context of semiconductor heterostructures.

The code developed in this work is implemented as freely available jupyter notebook, which makes it readily applicable for other users. It can be used to directly design desired semiconductor heterostructures, or, at least, to reduce the vast candidate search space. It is applicable to existing data in materials databases and, in addition, it is immediately compatible with new high-throughput density-functional theory work. Since our framework efficiently ties together first-principles calculations and existing databases for an important application, we are optimistic that it can tremendously accelerate the search for novel high-efficiency LEDs, solar cell devices, photocatalytic water splitters, or 2D electron gases confined to semiconductor interfaces. Our framework contributes to solving the grand challenge of efficient heterojunction design and, as such, we envision that it will be used in research with broad societal impact. In addition, we see great potential for addressing fundamental scientific problems, including the search for materials with surface-charge accumulation or intrinsic *n*- or *p*-type behavior.

Acknowledgments

The authors are grateful to M. Shim, A. L. Ferguson, and P. Gorai for helpful discussions, to S. Biesboer, B. Cabinian, A. Erwin, and H. Wu for the first version of the Python code that this study builds on, and J. Leveillee for providing band structure data of $CH_3NH_3PbI_3$. This material is based upon work supported by the National Science Foundation (awards: CBET-1437230 and DMR-1555153). Part of this work was supported by the Materials and Manufacturing Graduate Student Fellowship of the National Center for Supercomputing Applications. This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

Received:
Revised:
Published online:

[1] W. Mönch, *Semiconductor Surfaces and Interfaces* Springer **2001**.

[2] D. F. Feezell, J. S. Speck, S. P. DenBaars, and S. Nakamura, *J. Disp. Technol*. **2013**, 9, 190.

[3] X. Gong, S. Wang, D. Moses, G. C. Bazan, and A. J. Heeger, *Adv. Mater.* **2005**, *17*, 2053.

[4] Y. Hara, T. Takashima, R. Kobayashi, S. Abeyrathna, B. Ohtani, and H. Irie, Appl Catal. *B-Environ.* **2017**, *209*, 663.

[5] Q. Yan, J. Yu, S. K. Suram, L. Zhou, A. Shinde, P. F. Newhouse, W. Chen, G. Li, K. A. Persson, J. M. Gregoire, and J. B. Neaton, *Proc. Natl. Acad. Sci*. **2017,** *114*, 3040.

[6] G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys*. **2002**, *74*, 601.

[7] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, *Phys. Rev. B* **1996**, *53*, 3764.

[8] L. K. Wagner and D. M. Ceperley, *Rep. Prog. Phys.* **2016,** *79*, 094501.

[9] S.-H. Wei and A. Zunger, *Appl. Phys. Lett*. **1998**, *72*, 2011.

[10]C. G. Van de Walle and J. Neugebauer, *Nature* **2003,** *423*, 626.

[11] R. Anderson, *Solid State Electron*. **1962**, *5*, 341.

[12] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, *APL Materials* **2013**, *1*, 011002.

[13] R. H. Taylor, F. Rose, C. Toher, O. Levy, K. Yang, M. B. Nardelli, and S. Curtarolo, *Comp. Mater. Sci.* **2014**, *93*, 178.

[14] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, *JOM* **2013**, *65*, 1501.

[15] C. Ortiz, O. Eriksson, and M. Klintenberg, *Comput. Mater. Sci.* **2009**, *44*, 1042.

[16] F. Allen, G. Bergerhoff, and R. Sievers, Crystallographic Databases, *International Union of Crystallography*, Chester, UK **1987**.

[17] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, *ACTA Crystall. B-Stru.* **2002**, *58*, 364.

[18] J. Tersoff, *Phys. Rev. B* **1984,** *30*, 4874.

[19] A. Schleife, F. Fuchs, C. Rödl, J. Furthmüller, and F. Bechstedt, *Appl. Phys. Lett.* **2009**, *94*, 012104.

[20] Y. Hinuma, A. Grüneis, G. Kresse, and F. Oba, *Phys. Rev. B* **2014**, *90*, 155405.

[21] B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke, and I. Foster, *JOM* **2016**, *68*, 2045.

[22] E. Shapera and A. Schleife, **2018**, 10.18126/M28K8M.

[23] F. Flores and C. Tejedor, *J. Phys. C Solid State* **1979**, *12*, 731.

[24] A. Schleife, C. Rödl, F. Fuchs, J. Furthmüller, and F. Bechstedt, *Phys. Rev. B* **2009**, *80*, 035112.

[25] M. Stengel, P. Aguado-Puente, N. A. Spaldin, and J. Junquera, *Phys. Rev. B* **2011,** *83*, 235112.

[26] C. G. Van de Walle and R. M. Martin, *Phys. Rev. B* **1986**, *34*, 5621.

[27] P. D. C. King, T. D. Veal, P. H. Jefferson, S. A. Hatfield, L. F. J. Piper, C. F. McConville, F. Fuchs, J. Furthmüller, F. Bechstedt, H. Lu, and W. J. Schaff, *Phys. Rev. B* **2008**, *77*, 045316.

[28] B. Höffling, A. Schleife, F. Fuchs, C. Rödl, and F. Bechstedt, *Appl. Phys. Lett.* **2010**, *97*, 032116.

[29] B. Höffling, A. Schleife, C. Rödl, and F. Bechstedt, *Phys*. *Rev*. *B* **2012**, *85*, 035305.

[30] M. Cardona and N. E. Christensen, *Phys. Rev. B* **1987**, *35*, 6182.

[31] H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* **1976**, *13*, 5188.

[32] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, *Comput. Mater. Sci*. **2013**, *68*, 314.

[33] G. Kresse and J. Furthmüller, *Comp. Mater. Sci.* **1996**, *6*, 15.

[34] G. Kresse and D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758.

[35] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.

[36] J. P. Perdew and Y. Wang, *Phys. Rev. B* **1992**, *45*, 13244.

[37] W. Kohn and L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

[38] M. Scheffler and C. Draxl, "Computer center of the max-planck society, garching, the nomad repository," **2014**.

[39] G. Bergerhoff, I. D. Brown, and F. H. Allen, *International Union of Crystallography* **1987**, *360*, 77.

[40] W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson, and S. Curtarolo, *ACS Comb. Sci.* **2011**, *13*, 382.

[41] K. M. Poduska, L. Regev, E. Boaretto, L. Addadi, S. Weiner, L. Kronik, and S. Curtarolo, *Adv. Mater.* **2011**, *23*, 550.

[42] A. M. Deml, R. O'Hayre, C. Wolverton, and V. Stevanović, *Phys. Rev. B* **2016**, *93*, 085142.

[43] M. N. Obrovac and V. L. Chevrier, *Chem. Rev.* **2014**, *114*, 11444.

[44] J. Greeley, T. Jaramillo, J. Bonde, I. Chorkendorff, and J. Nørskov, *Nat. Mater.* **2006**, *5*, 909.

[45] L. Yu and A. Zunger, *Phys. Rev. Lett.* **2012**, *108*, 068701.

[46] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, *Comput. Mater. Sci.* **2011**, *50*, 2295.

[47] I. E. Castelli, F. Hüser, M. Pandey, H. Li, K. S. Thygesen, B. Seger, A. Jain, K. A. Persson, G. Ceder, and K. W. Jacobsen, *Adv. Ener. Mater.* **2015**, *5*, 1400915.

[48] S. Korbel, M. A. L. Marques, and S. Botti, *J. Mater. Chem. C* **2016**, DOI:10.1039/C5TC04172D.

[49] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, and A. Tropsha, *Nat. Commun.* **2017**, *8*, 15679.

[50] J. Lee, A. Seko, K. Shitara, K. Nakayama, and I. Tanaka, *Phys. Rev. B* **2016**, *93*, 115104.

[51] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, *Comp. Mater. Sci.* **2012**, *58*, 227.

[52] W. Strehlow and E. Cook, *J. Phys. Chem. Ref. DATA* **1973**, *2*, 163.

[53] A. Alkauskas, P. Broqvist, and A. Pasquarello, *Phys. Rev. Lett.* **2008**, *101*, 046405.

[54] T. Matsuoka, H. Okamoto, M. Nakao, H. Harima, and E. Kurimoto, *Appl. Phys. Lett.* **2002**, *81*, 1246.

[55] P. Peacock and J. Robertson, *J. Appl. Phys.* **2002**, *92*, 4712.

[56] L. Brus, *J. Phys. Chem.* **1986**, *90*, 2555.

[57] B.-R. Hyun, J. J. Choi, K. L. Seyler, T. Hanrath, and F. W. Wise, *ACS Nano* **2013**, *7*, 10938.

[58] W. Moench, Electronic properties of semiconductor interfaces, *Springer*, Berlin, **2004**.

[59] A. Dimoulas, P. Tsipas, A. Sotiropoulos, and E. K. Evangelou, *Appl. Phys. Lett.* **2006**, *89*, 252110.

[60] L. Piper, P. Jefferson, T. Veal, C. McConville, J. Zúñiga-Pérez, and V. Muñoz-Sanjosé, *Superlattices Microst.* **2007**, *42*, 197.

[61] D. Mourad, J.-P. Richters, L. Gérard, R. André, J. Bleuse, and H. Mariette, *Phys. Rev. B* **2012**, *86*, 195308.

[62] Mahboob, I., Veal, T. Á., McConville, C. Á., Lu, H., & Schaff, W. Á, *Phys. Rev. Lett.* **2004**, *92*, 036804.

[63] P. D. C. King, T. D. Veal, P. H. Jefferson, J. Zúñiga-Pérez, V. Muñoz-Sanjosé, and C. F. McConville, *Phys. Rev. B* **2009**, *79*, 035203.

[64] Z. Weinberg, G. Rubloff, and E. Bassous, *Phys. Rev. B* **1979**, *19*, 3107.

[65] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, *JOM* **2013,** *65*, 1501.

[66] M. Shim, private communication, **2017**.

[67] J. Heyd, G. E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **2003**, *118*, 8207.

[68] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, *J. Chem. Phys.* **2006**, *125*, 224106.

[69] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán, *J. Chem. Phys.* **2006**, *124*, 154709.

[70] A. Schleife, M. D. Neumann, N. Esser, Z. Galazka, A. Gottwald, J. Nixdorf, R. Goldhahn, and M. Feneberg, *New J. Phys.* **2018**, *20*, 053016.

[71] L. F. J. Piper, L. Colakerol, P. D. C. King, A. Schleife, J. Zúñiga-Pérez, P.-A. Glans, T. Lear-

altmonth, A. Federov, T. D. Veal, F. Fuchs, V. Muñoz-Sanjosé, F. Bechstedt, C. F. McConville, and K. E. Smith, *Phys. Rev. B* **2008**, *78*, 165127.

[72] P. King, T. Veal, D. Payne, A. Bourlange, R. Egdell, and C. McConville, *Phys. Rev. Lett.* **2008**, *101*, 116808.

[73] D. C. Tsui, *Phys. Rev. Lett.* **1970**, *24*, 303.

[74] M. Noguchi, K. Hirakawa, and T. Ikoma, *Phys. Rev. Lett.* **1991,** *66*, 2243.

[75] M. Akazawa, T. Matsuyama, T. Hashizume, M. Hiroki, S. Yamahata, and N. Shigekawa, *Appl. Phys. Lett.* **2010**, *96*, 132104.

[76] T. Yao and S.-K. Hong, eds., "The pin diode circuit designers' handbook," Microsemi-Watertown, **1998**.

[77] D. Chen, A. Nakahara, D. Wei, D. Nordlund, and T. P. Russell, *Nano Lett.* **2011**, *11*, 561.

[78] Y. Wang, T. Sun, T. Paudel, Y. Zhang, Z. Ren, and K. Kempa, *Nano. Lett.* **2012**, *12*, 440.

[79] F. Pérez and B. E. Granger, *Comput. Sci. Eng.* **2007**, *9*, 21.

[80] B. Ruhstaller, S. Carter, S. Barth, H. Riel, W. Riess, and J. Scott, *J. Appl. Phys.* **2001**, *89*, 4575.

[81] A. K. Chandiran, M. Abdi-Jalebi, M. K. Nazeeruddin, and M. Grätzel, *ACS Nano* **2014**, *8*, 2261.

[82] J. M. Caruge, J. E. Halpert, V. Wood, V. Bulovic, and M. G. Bawendi, *Nat. Photon*. **2008**, *2*, 247.

[83] A. W. Schill, C. S. Gaddis, W. Qian, M. A. El-Sayed, Y. Cai, V. T. Milam, and K. Sandhage, *Nano Lett.* **2006**, *6*, 1940.

[84] R. Lampande, G. W. Kim, J. Boizot, Y. J. Kim, R. Pode, and J. H. Kwon, *J. Mat. Chem. A* **2013**, *1*, 6895.

[85] P. R. Brown, R. R. Lunt, N. Zhao, T. P. Osedach, D. D. Wanger, L.-Y. Chang, M. G. Bawendi, and V. Bulovic, *Nano Lett*. **2011**, *11*, 2955.

[86] A. Nazarov, Y. N. Vovk, V. Lysenko, V. Turchanikov, V. Scryshevskii, and S. Ashok, *J. Appl. Phys.* **2001**, *89*, 4422.

[87] Y. Guo, H. Wang, B.-L. Gu, and Y. Kawazoe, *J. Appl. Phys.* **2000**, *88*, 6614.

[88] H. Fei, Z. Feng, and X. Liu, *Ionics* **2015**, *21*, 1967.

[89] Z. B. Wang, M. G. Helander, M. T. Greiner, J. Qiu, and Z. H. Lu, *Phys. Rev. B* **2009**, *80*, 235325.

[90] C. Ahn, J.-M. Triscone, N. Archibald, M. Decroux, R. Hammond, T. Geballe, O. Fischer, and M. Beasley, *Science* **1995**, *269*, 373.

[91] Y. Hou, X. Chen, S. Yang, Y. L. Zhong, C. Li, H. Zhao, and H. G. Yang, *Nano Energy* **2017**, *36*, 102.

[92] S.-W. Choi, A. Katoch, J.-H. Kim, and S. S. Kim, *ACS Appl. Mater. Inter.* **2014**, *6*, 17723.

[93] T. Harianto, K. Sadakuni, H. Akinaga, and T. Suemasu, *Jpn. J. Appl. Phys*. **2008**, *47*, 6310.

[94] R. Kannan, C. M. A., P. C. L., A. Sally, H. F. S., K. James, Y. David, R. Manuel, M. Wyatt, N. Rommel, W. James, and D. Anna, *Prog. Photovoltaics* **2003**, *11*, 225.

[95] Z. Song, G. Tong, H. Li, G. Li, S. Ma, S. Yu, Q. Liu, and Y. Jiang, *Nanotechnology* **2017**, *29*, 025401.

[96] X. Zeng, T. Zhou, C. Leng, Z. Zang, M. Wang, W. Hu, X. Tang, S. Lu, L. Fang, and M. Zhou, *J. Mater. Chem. A* **2017**, *5*, 17499.

[97] J. C. Egues, *Phys. Rev. Lett.* **1998**, *80*, 4578.

[98]I. Repins, M. A. Contreras, B. Egaas, C. DeHart, J. Scharf, C. L. Perkins, B. To, and R. Noufi, *Prog. Photovoltaics* **2008**, *16*, 235.

[99] J. v. d. L. N.-G. Park and A. J. Frank, *J. Phys. Chem. B* **2000**, *104*, 8989.

[100] N. Tokmoldin, N. Griffiths, D. D. C. Bradley, and S. A. Haque, *Adv. Mater.* **2009**, *21*, 3475.

[101] F. P. Koffyberg and F. A. Benko, *J. Appl. Phys.* **1982**, *53*, 1173.

[102] J.-M. Caruge, J. E. Halpert, V. Bulović, and M. G. Bawendi, *Nano Lett*. **2006**, *6*, 2991.