Relativistic quasiparticle self-consistent electronic structure of hybrid halide perovskite photovoltaic absorbers

Federico Brivio, Keith T. Butler, Aron Walsh, and Mark van Schilfgaarde

1) Centre for Sustainable Chemical Technologies and Department of Chemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK
2) Department of Physics, Kings College London, London WC2R 2LS, UK

Abstract: Solar cells based on a light absorbing layer of the organometal halide perovskite CH$_3$NH$_3$PbI$_3$ have recently surpassed 15% conversion efficiency, though how these materials work remains largely unknown. We analyse the electronic structure and optical properties within the quasiparticle self-consistent $GW$ approximation. While this compound bears some similarity to conventional $sp$ semiconductors, it also displays unique features. Quasiparticle self-consistency is essential for an accurate description of the band structure: band gaps are much larger than what is predicted by the local density approximation (LDA) or $GW$ based on the LDA. Valence band dispersions are modified in a very unusual manner. In addition, spin-orbit coupling strongly modifies the band structure and gives rise to unconventional dispersion relations and a Dresselhaus splitting at the band edges. The average hole mass is small, which partially accounts for the long diffusion lengths observed. The surface ionisation potential (workfunction) is calculated to be 5.7 eV with respect to the vacuum level, explaining efficient carrier transfer to TiO$_2$ and Au electrical contacts.

PACS numbers: 88.40.-j, 71.20.Nr, 72.40.+w, 61.66.Fn

I. INTRODUCTION

One of the most promising third-generation photovoltaic technologies is based on metal-organic halide perovskites. The materials physics of inorganic (ABX$_3$) perovskites is well developed; however, the replacement of the inorganic cation by an iso-electronic organic moiety provides an opportunity for tuning the chemical bonding and optical response. We apply a range of electronic structure techniques to calculate and predict how the rich and unusual physics of these materials accounts for their widely reported success as absorber layers in solar cells.

It has been established that similar to traditional dielectric perovskites, these hybrid analogues have a range of accessible polymorphs with variations in the tilting and rotation of the BX$_6$ polyhedra in the lattice. A large family of hybrid perovskites have been reported with inorganic networks ranging from 1–3 dimensions. However, the methylammonium (MA) cation (i.e. CH$_3$NH$_3^+$) has been widely applied, resulting in the highest-performance perovskite-structured solar absorbers. The polar MA cation can also be replaced by ammonium (NH$_4^+$) as a smaller non-polar analogue.

A large number of density functional theory (DFT) studies have been reported that examine the electronic properties of hybrid perovskites, while a relativistic treatment based on local or semi-local exchange-correlation functionals results in severe band gaps underestimations. Both approaches are insufficient to describe the complexity of the electronic structure of these hybrid semiconductors, with large errors expected in predicted properties such as carrier effective mass and dielectric function.

An alternative approach is the $GW$ formalism, which can be used to correct errors in the one-electron Kohn-Sham eigenvalues within a many-body quasiparticle framework. Here we employ quasiparticle self-consistent $GW$ theory (QS$GW$) to study the electronic structure of CH$_3$NH$_3$PbI$_3$ and NH$_4$PbI$_3$, including the effect of spin-orbit coupling (SOC) on both the kinetic energy and electron self-energy $\Sigma$ (see the Appendix).

As Pb and I are heavy elements, SOC is large and has a major effect on spectral properties. SOC predominantly modifies the kinetic energy; however, in this case relativistic effects are large enough to induce a modest reduction in $\Sigma$ as well, in contrast to the vast majority of semiconductors, e.g. elemental Sn. As a consequence of large relativistic effects, the conduction and valence bands near the band extrema deviate strongly from parabolic behavior. Effective masses are no longer constant, but depend on doping, temperature, and the property being measured. Average effective masses are nevertheless light, and the dielectric constants large, accounting for the long diffusion lengths that have been recently reported.

In many respects these perovskites are similar to conventional $sp$ semiconductors: conduction and valence bands near the Fermi level have $sp$ character, and local- and semi-local- density approximation (LDA) to DFT systematically underestimate the band gap $E_G$ because they do not include spatial non-locality in the exchange-correlation potential. There are other significant points...
of departure: in sharp contrast to tetrahedral semiconductors, DFT also poorly describes valence band dispersions. This surprising result, which we discuss further below, indicates that the usual explanations invoked to account for deficiencies in DFT’s description of semiconductors are not sufficient here.

We show that there is a strong feedback between dielectric response and quasiparticle levels, as occurs for CuInSe$_2$ [24]. Thus self-consistency in GW is essential: $E_G$ calculated from $G^{LDA}W^{LDA}$, i.e. LDA as the starting Hamiltonian, picks up only a little better than half the gap correction to the LDA. Moreover, the QSGW and LDA valence bands, which the LDA describes reasonably well in tetrahedral semiconductors, are significantly different. These differences underscore the limitations of density-functional based approaches (LDA, hybrid functionals, or $G^{LDA}W^{LDA}$) in describing the properties of these materials. QSGW does not depend on the LDA: self-consistency renders it more reliable and universally applicable than other forms of GW, which will be important for in silico design of hybrid systems. Moreover, QSGW can determine some ground-state properties, e.g. the charge density and electric field gradient. Errors in QSGW tend to be small and highly systematic; most notably there is a tendency to slightly overestimate semiconductor band gaps. Limited data is available for organic-inorganic halide perovskites, but at least for CH$_3$NH$_3$PbI$_3$ the universal tendency found in other materials is consistent with recent measurements.

Finally, based on the workfunction calculated for CH$_3$NH$_3$PbI$_3$ within DFT (including an estimate for quasiparticle corrections) we show that band alignments are consistent with efficient electron transfer to TiO$_2$ and Au electrical contacts.

II. RESULTS

Optimisation of the crystal structures of NH$_4$PbI$_3$ and CH$_3$NH$_3$PbI$_3$ have recently been reported in DFT using the PBEsol exchange-correlation functional. Atomic forces were converged to within 5 meV/Å, and the bond lengths are in good agreement with experiment. The representative (100) configuration of MA is considered here. Lattice vectors of these perovskites are approximately cubic (a = 6.29 Å and 6.21 Å for the MA and NH$_4$ perovskites, respectively), with small distortions of the simple cubic ones. The valence band maximum and conduction band minimum falls close to a zone boundary point, the analogue of the R point ($\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$) in cubic symmetry. We denote this point as R in the remainder of the paper.

A. Band structure

The QP band structures for CH$_3$NH$_3$PbI$_3$ and NH$_4$PbI$_3$, with colours denoting the orbital character of the states, are shown in Fig. 1. The ions within the inorganic (PbI$_3$)$^-$ cage have formal electronic configurations of Pb: 5d$^{10}$6s$^2$6p$^6$ and I: 5p$^5$. As can be seen from the colour coding, the valence band maximum consists of approximately 70% I 5p and 25% Pb 6s (the Pb 6s forms a band centred around $-8$ eV), while the conduction band consists a mixture of Pb 6p and other orbitals. The molecular units CH$_3$NH$_3$ and NH$_4$ form $\sigma$ bonds deep in the valence band. They are essentially dispersionless: they do not hybridise with the cage until energies exceed $E_F+5$ eV. Thus their interaction with the host is largely electrostatic and structural; they provide charge compensation to the PbI$_3^-$ cage.

The results presented in Table I demonstrate the various contributions to the band energies around the fundamental gap. The contribution from SOC ($\sim 1$ eV), is extraordinarily large, of the order of the gap itself; so large that screening is enhanced. As a result there is a smaller, but nonetheless non-negligible contribution of SOC to the electron self-energy ($\Sigma=iGW$), apparent from the difference between ‘SO($T$)’ and ‘SO($\Sigma$)’. Furthermore, Table I emphasises the importance of the feedback between W and QP when calculating the band structure of these materials is consistent with recent measurements. Light dashed gray lines show corresponding bands in the LDA. The dispersionless state near $-5$ eV corresponds to a molecular level of methylammonium. In QSGW this state is pushed down to $-7.7$ eV. The dispersion of the highest valence bands is very poorly described by the LDA, as described in the text.
systems. The ‘GW’ gap is based on a perturbation of the LDA gap; and it is significantly smaller. Because the LDA gap is too small, \( W \) is overscreened, and GW underestimated. The role of feedback is important in other semiconductors: it is particularly strong in InN and Cu(In,Ga)Se\(_2\). In the latter case the interplay between \( W \) and \( E_G \) was shown explicitly by comparing functionals that did or did not include the dependence of \( W \) on band structure. \W and the gap correction is not a function of the fundamental gap alone: all the bands (including valence band dispersions) shift in a nontrivial manner. To reliably determine the electronic structure including the fundamental gap, self-consistency is essential.

Recent measurements place the room temperature band gap of NH\(_3\)CH\(_3\)PbI\(_3\) at 1.61 eV,\(^{27}\) which falls slightly below the QSGW result. Some tendency for QSGW to overestimate gaps is expected. In any case theory and experiment cannot be compared to better than 0.1 eV resolution for several reasons. There are small issues with \( k \)-point convergence,\(^{28}\) and with the shape of local wave functions determined by solving a scalar relativistic equation rather than the Dirac equation (see the Appendix). On the experimental side there may be some temperature dependence of the gap given the structural flexibility of the material;\(^{29}\) this has yet to be explored.

<table>
<thead>
<tr>
<th></th>
<th>DFT</th>
<th>PBEsol</th>
<th>LDA</th>
<th>Expt(^{27})</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH(_3)CH(_3)PbI(_3)</td>
<td>1.38</td>
<td>1.46</td>
<td>0.53</td>
<td>1.61 (RT)</td>
</tr>
<tr>
<td>NH(_3)PbI(_3)</td>
<td>1.20</td>
<td>1.13</td>
<td>0.09</td>
<td>-</td>
</tr>
<tr>
<td>QSGW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO=0</td>
<td>SO(T)</td>
<td>SO((\Sigma))</td>
<td>GW(^{0.5})</td>
<td>Expt</td>
</tr>
<tr>
<td>CH(_3)NH(_3)PbI(_3)</td>
<td>2.73</td>
<td>1.78</td>
<td>1.67</td>
<td>1.27</td>
</tr>
<tr>
<td>NH(_3)PbI(_3)</td>
<td>2.30</td>
<td>1.36</td>
<td>1.38</td>
<td>0.76</td>
</tr>
</tbody>
</table>

**TABLE I:** Fundamental band gaps (in eV) of CH\(_3\)NH\(_3\)PbI\(_3\) and NH\(_3\)PbI\(_3\) calculated at varying levels of approximation. Top rows show DFT results using PBEsol (reported in Ref. 18) and the Barth-Hedin LDA functional. First columns show that semilocal and local functionals generate similar gaps. ALS (column ‘+SO’) strongly reduces the gap. GW gaps are shown without spin-orbit coupling (SO=0), with ALS added to a fixed potential, modifying the kinetic energy only (SO(T)), and included in the QSGW self-consistency cycle (SO(\(\Sigma\))). Column ‘GW\(^{0.5}\)’ is similar to SO(\(\Sigma\)) but \( G \) and \( W \) are generated from the LDA. (In this calculation the full \( \Sigma \) matrix was generated, not just the diagonal part as is customary. A Z factor of 1 was used to take partial account of self-consistency, which brings the gap in better agreement with the QSGW result; see Appendix A in Ref. 30). An error of order 0.1 eV might be associated with the treatment of SOC; see the Appendix. Room temperature (RT) band gap data is only available for NH\(_3\)CH\(_3\)PbI\(_3\).\(^{27}\)

Fig. 1 also shows the LDA energy bands. Remarkably, the LDA badly underestimates not only the gap but poorly describes the dispersion in the valence bands. The tendency for LDA to underestimate band gaps is traditionally associated with the energy cost for an excited electron-hole pair. The exchange-correlation potential should distinguish between a neutral excitation (e.g. a hole shifting from one \( k \) point in the valence band to another) and one where charge is separated (excitation of an electron-hole pair). Such a distinction is problematic for a local potential, which is by necessity the same for all electrons. Such an error is seen in the present case, as the LDA gap is too small. As Fig. 1 clearly shows, LDA and QSGW valence bands also deviate strongly from one another. Note in particular the states at \( R \) between 0 and \(-2\) eV. This behaviour allows us to deduce that the hopping matrix elements between I \( 5p \) (and to some extent Pb 6s) states are poorly described by the LDA.

**B. Carrier effective mass**

Typically light hole masses are too small in the LDA because, according to \( k \cdot p \) theory, \( m^*_h \propto E_G/V^2 \); \( V \) is a matrix element of the gradient operator between the conduction and valence bands. \( m^*_h \) is expected to be too small because \( E_G \) is underestimated; indeed for traditional narrow gap tetrahedral semiconductors, the proportionality between \( m^*_h \) and \( E_G \) is reasonably well obeyed. The LDA predicts the light hole mass to be too small while other masses (which do not couple to the nearest conduction band) are reasonably described. For example, GaAs has a gap similar to CH\(_3\)NH\(_3\)PbI\(_3\), and LDA underestimates it by a comparable amount (~1 eV). Following expectations, the LDA underestimates the light hole mass in GaAs by a factor of ~3. But for CH\(_3\)NH\(_3\)PbI\(_3\) the situation is reversed: the LDA overestimates \( m^*_h \) even while it severely underestimates \( E_G \). For NH\(_3\)PbI\(_3\) the LDA and QSGW masses are comparable, but only because the LDA gap is very small.

Spin-orbit coupling greatly complicates both valence and conduction bands within \( k_B T \) of the band edges. We focus on the two valence bands of NH\(_3\)CH\(_3\)PbI\(_3\), as these are the ones that govern transport in hole-based devices. As these two bands approach the \( R \) point, they must merge to the same value by symmetry. However, they approach the \( R \) point with a linear dispersion in some directions; as a consequence the dispersions in the upper and lower bands are non-analytic. The upper band is maximum in some directions but increases with a linear slope in the \( \pm [112] \) direction. Two maxima form near \( R \pm 0.005 \times [112] \), Fig. 2(a), arising from Dresselhaus spin-orbit coupling, which is even more pronounced in the lower conduction band. The lower band has a single maximum at \( R \), and its constant energy surfaces deviate only modestly from spheres for \( k \) near \( R \) (Fig. 2(b)). Yet, the right panels of Fig. 2 show \( E_v(k)=E(R)−E(k) \) deviates markedly from a parabolic dispersion. This has important consequences for the device behavior of this material. Fig. 2(c) shows that, provided \( E_v >10 \) meV, the band dispersion can be expressed approximately as a \( k \) dependent mass:

\[
\frac{\hbar^2 k^2}{2m} = m^*(k) E_v(k), \quad \frac{m^*(k)}{m} = m_0[1 + \alpha E_v(k)] \quad (1)
\]

with \( m_0^* \sim 0.12 \) and \( \alpha \) independent of \( |k| \) but dependent
on orientation. For $E_v$, large enough, the upper valence band can also be described by an effective mass of roughly the same size. The lower conduction band exhibits a similar behaviour, with $m^*_v \sim 0.15$. The small masses explain how these materials can exhibit high mobility and long diffusion lengths. Bands of NH$_4$PbI$_3$ differ in important details from NH$_3$CH$_3$PbI$_3$ (the influence of SOC is less pronounced), but the basic structure is similar. Both the band gap and effective masses are reduced relative to NH$_3$CH$_3$PbI$_3$, as can be seen directly by inspecting Fig. 1.

C. Optical and dielectric response

It is known that these compounds strongly absorb visible light. We confirm this through a random phase approximation (RPA) calculation of $\alpha(\omega)$ from the imaginary part of the macroscopic dielectric function $\epsilon_M(\omega) = \epsilon_\infty + (\epsilon_G(\omega) + \epsilon_\infty)(\omega \rightarrow 0, \omega)$, and $\alpha$ is somewhat smaller than – but comparable to – that of GaAs. This explains why very thin layers of the hybrid perovskites have been found to give high photovoltaic efficiencies. Indeed combined with the low carrier effective masses, the resulting electron-hole diffusion lengths exceed the typical film thickness.

Some static ($\epsilon_0$) and high-frequency ($\epsilon_\infty$) dielectric constants are shown in Table II. These values were unusually sensitive to the $k$-point sampling density and require dense meshes for convergence. Those calculated by density-functional perturbation theory (e.g. PBEsol) leave out SOC and thus get fortuitously good band gaps. As a result $\epsilon_\infty$ is not so different from the QS$GW$ case (which includes SOC). Contributions to $\epsilon_0$ from lattice polarization are significantly larger than seen in typical tetrahedral semiconductors (compare $\epsilon_0$ to $\epsilon_\infty$). For NH$_3$CH$_3$PbI$_3$, the average value of the static dielectric constant, including quasi-particle (QP) corrections from Table II, of 24.1 is in very good agreement with permittivity measurements of 23.3. These values exclude contributions from orientational disorder of the methylammonium ions, which is the subject of further study.

D. Surface ionisation potential

In order to place the electronic bands on an absolute energy scale, we have aligned the quasiparticle energies with respect to the vacuum level of a non-polar (110) termination of the perovskite, generated using METADISE$^{34}$. We take the Pb 1s core level as an energy reference and use a planar average of the electrostatic potential following the standard procedure.$^{35,36}$ The slab model consisted of four perovskite layers with the dipole of the methyl ammonium cations aligned parallel to the surface termination, which ensure no macroscopic electric field. The resulting ionisation potential is 5.7 eV (5.9 eV from LDA, which is corrected by the $GW^0$ self energy), with a corresponding electron affinity of 4.0 eV. These values are in good agreement with initial photoe-
III. CONCLUSION

We have explored the electronic structure of two key hybrid halide perovskites. Relativistic and many-body corrections are shown to be essential for a quantitative description of the bulk properties important for photovoltaics: band gap, band dispersion, effective mass and dielectric response. These organic-inorganic materials display quantum mechanical behaviour atypical of traditional semiconductors, which begins to explain their remarkable performance in mesoporous and thin-film solar cells.

ACKNOWLEDGMENTS

We thank L.M. Peter and H. J Snaith for useful discussions, and acknowledge membership of the UK’s HPC Materials Chemistry Consortium, which is funded by EPSRC grant EP/F006746. F.B., K.T.B and A.W. are funded by the DESTINY ITN (316494), EPSRC (EP/J017361/1) and the ERC (277757), respectively.

APPENDIX: SIMPLIFIED TREATMENT OF SPIN-ORBIT COUPLING IN QSGW

Aryasetiwan and Biermann\textsuperscript{38} developed a formalism for $GW$ with spin-dependent interactions. Rather than proceed with a completely non-collinear treatment, we take advantage of the fact that $\lambda \mathbf{L} \cdot \mathbf{S}$ is relatively small, and moreover that the non-collinear part of the eigenfunctions is unimportant for these semiconductors. We present a simplified treatment that generates results essentially as good as adding $\lambda \mathbf{L} \cdot \mathbf{S}$ non-perturbatively to the scalar Dirac Hamiltonian for M-PbI$_3$ compounds.

Partitioning $\mathbf{L} \cdot \mathbf{S}$ into components, the non-interacting QSGW Hamiltonian reads

$$H_0 = H_0(\lambda=0) + \lambda \mathbf{L}^2 \mathbf{S}^2 + \lambda (\mathbf{L}^+ \cdot \mathbf{S}^- + \mathbf{L}^- \cdot \mathbf{S}^+)$$

The first two terms are spin-diagonal and can be diagonalized non-perturbatively in the same manner as $H_0(\lambda=0)$. The eigenvalues $\epsilon_i$ and eigenfunctions $\psi_i$ contain the $\mathbf{L}^2 \mathbf{S}^2$ portion of $\mathbf{L} \cdot \mathbf{S}$, keeping $\psi_i$ spin diagonal. The latter two terms, when treated exactly, further shift the $\epsilon_i$ and also introduce spin off-diagonal parts to the $\psi_i$. We allow the former but omit the latter.

The lowest order of correction to the eigenvalues is second order and we follow the spirit of second-order perturbation theory. Let $\delta \epsilon_{ij}$ be the initial splitting in $\epsilon_i$ and $\epsilon_j$, $\Delta_{ij} = |\epsilon_i - \epsilon_j|/2$. If $H_{ij}^{+\pm}$ couples $i$ and $j$, $\Delta_{ij}$ increases by $\delta \Delta_{ij} = |H_{ij}^{+\pm}|^2/|\epsilon_i - \epsilon_j|$, in lowest order.

Second-order perturbation theory can be problematic when $\epsilon_i \rightarrow \epsilon_j$. We instead obtain $\delta \Delta_{ij}$ from

$$\delta \Delta_{ij} = \sqrt{\Delta_{ij}^2 + |H_{ij}^{+\pm}|^2} - |\Delta_{ij}|$$

This expression is exact if $i$ and $j$ are isolated from the rest of the system. The final expression (the net shift for each $\epsilon_i$ is obtained by summing over each $ij$ pair) is nevertheless correct only to second order because terms involving three or more states are not included.

We carefully tested our quasi-perturbative approach in the LDA or LDA+$U$ context for a wide range of materials, e.g. Fe, Sn, Au, GdN, Pu, and the perovskites addressed in this manuscript. In all cases except Pu ($Z=94$) the difference between the perturbation expression resulted in $\epsilon_i$ very close to $\mathbf{L} \cdot \mathbf{S}$ treated non-perturbatively. For CH$_3$NH$_3$PbI$_3$, for example, $E_G$ changed by less than 0.01 eV. Self-consistency carried through with both approaches generate a slight difference in density, but no significant difference in the $\epsilon_i$.

Tests of the adequacy of the quasi-perturbative $\mathbf{L} \cdot \mathbf{S}$ in the QSGW were performed as follows: self-consistency was reached with $\mathbf{L} \cdot \mathbf{S}$ included quasi-perturbatively, and for a given $\Sigma$, the quasiparticle levels with $\mathbf{L} \cdot \mathbf{S}$ calculated non-perturbatively were compared to the perturbative treatment. As in the LDA case, negligible differences were found for all compounds studied except for Pu, where modest differences were found. As in the LDA case, the nonperturbative treatment generated a slight change in density. Since the $\epsilon_i$ are reliably determined, it is unlikely that a better treatment of $\mathbf{L} \cdot \mathbf{S}$ (non-collinear eigenfunctions) will further affect $\Sigma$ appreciably in these compounds. On the other hand, fully relativistic treatment might affect $H_0$ a little, since the relativistic radial functions vary as $r^\gamma$ for small $r$, where $\gamma^2 = \kappa^2 - (2Z/c)^2$, $\kappa$ playing the role of the $l$ quantum number. $\gamma$ reduces the scalar relativistic case only when $c \rightarrow \infty$. A better treatment of the small-$r$ behavior of the partial waves modifies spin orbit splitting of the $p$ levels for Pb by about 10%, which is not included here.

An appreciable effect of $\mathbf{L} \cdot \mathbf{S}$ on $\Sigma$ is observed only for compounds with large-$Z$ constituents. For semiconductors as heavy as Sn ($Z=50$), and for metals as heavy as Au ($Z=79$), the effect of $\mathbf{L} \cdot \mathbf{S}$ on $\Sigma$ appears to be very small. But for the iodide perovskites studied here, $\mathbf{L} \cdot \mathbf{S}$ has a noticeable effect on $\Sigma$ (Table1) because of the interplay between $E_G$ and $\epsilon$ present in semiconductors but not in metals.

Careful convergence checks were made of the eigenfunction and product basis. A $4 \times 4 \times 4$ $k$-mesh was used for the M-PbI$_3$ compounds. A 1-shot calculation with a $6 \times 6 \times 6$ mesh as perturbation to the QS$_{\text{GW}}$ result showed that the $k$-converged gap is about 0.1 eV larger than what is reported in Table I.