

# Band Structure Interpolation using Optimized Local Orbitals from Linear-Scaling Density-Functional Theory

L. E. Ratcliff,<sup>1</sup> G. J. Conduit,<sup>2</sup> N. D. M. Hine,<sup>1,3,4</sup> and P. D. Haynes<sup>1,3,2,\*</sup>

<sup>1</sup>*Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, UK*

<sup>2</sup>*Theory of Condensed Matter, Cavendish Laboratory,*

*University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, UK*

<sup>3</sup>*Department of Physics, Imperial College London, Exhibition Road, London SW7 2AZ, UK*

<sup>4</sup>*Department of Physics, University of Warwick, Coventry CV4 7AL, UK*

(Dated: August 21, 2018)

Several approaches to linear-scaling Density Functional Theory (LS-DFT) which seek to achieve accuracy equivalent to plane-wave methods do so by optimizing *in situ* a set of local orbitals in terms of which the density matrix can be accurately expressed. These local orbitals, which can also accurately represent the canonical Kohn-Sham orbitals, qualitatively resemble the maximally-localized Wannier functions employed in band structure interpolation. As LS-DFT methods are increasingly being used in real-world applications demanding accurate band structures, it is natural to question the extent to which these optimized local orbitals can provide sufficient accuracy. In this article, we present and compare, in principle and in practice, two methods for obtaining band structures. We apply these to a (10, 0) carbon nanotube as an example. By comparing with the results from a traditional plane-wave pseudopotential calculation, the optimized local orbitals are found to provide an excellent description of the occupied bands and some low-lying unoccupied bands, with consistent agreement across the Brillouin zone. However free-electron-like states derived from weakly bound states independent of the  $\sigma$  and  $\pi$  orbitals can only be found if additional local orbitals are included.

PACS numbers: 71.15.Ap

## I. INTRODUCTION

Methods for inferring detailed information about the electronic structure throughout the first Brillouin zone (BZ) from first-principles calculations on a discrete mesh of Bloch wavevectors<sup>1-4</sup> have received significant attention over the last 15 years. The most popular approach to date has been the interpolative scheme based upon maximally-localized Wannier functions (MLWFs)<sup>2,5,6</sup> obtained by a post-processing of the results from standard codes<sup>7,8</sup>. These methods have been used for a wide range of applications, not just within density-functional theory (DFT) but also many-body perturbation theory within both the *GW* approximation<sup>9</sup> and Bethe-Salpeter equation<sup>10</sup> due to their importance for theoretical spectroscopy. Other notable applications include electric polarization<sup>11-13</sup>, orbital magnetization<sup>14-16</sup> and topological insulators<sup>17-19</sup>. These and others have been described in a recent review alongside the theory of Wannier functions<sup>6</sup>.

Linear-scaling or  $O(N)$  methods for first-principles electronic structure calculations<sup>20-22</sup> based on DFT now provide routine access to the total energy and atomic forces<sup>23</sup> of systems consisting of many thousands of atoms<sup>24,25</sup>, giving rise to new opportunities for applications in different fields, as discussed for example in recent reviews on the subject<sup>26,27</sup>. Methods that optimize a set of Wannier-like localized orbitals<sup>28-32</sup> are also capable of achieving accuracy equivalent to plane-waves (PWs) or other systematic basis sets<sup>31,33</sup>. This article explores the use and limitations of *in situ* optimized orbitals, often

referred to as support functions, for band structure interpolation.

Due to the large simulation cells appropriate for linear-scaling methods, the correspondingly small BZs are sampled using the  $\Gamma$ -point only, with the added benefit that the local orbitals may be chosen to be real rather than complex. However, in the contemporary DFT approach, the total energies and atomic forces computed represent averages of the valence electronic structure across the BZ and return relatively little information about the system in view of the effort invested in their calculation. Moreover, the support functions are formally related<sup>34</sup> to the canonical Kohn-Sham (KS) orbitals and therefore, in principle at least, contain far more information about the electronic structure.

It is straightforward to obtain the KS energies and orbitals from the support functions by a single diagonalization of the ground-state self-consistent KS Hamiltonian matrix. Although such a post-processing step does not scale linearly, the minimal size of the local orbital set means that its contribution to the overall computational cost remains small for systems containing up to several thousand atoms. A method for performing band structure interpolation across the BZ based upon the local orbitals would therefore provide much richer information about the electronic structure, allow a detailed assessment of the accuracy of this description and provide a basis for further calculations including (local) densities of states.

For materials which require a large number of atoms for a proper treatment, such as complex nanostructures

or defective systems requiring large supercells, the high computational cost prohibits the standard approach of extracting MLWFs from cubic-scaling DFT. The use of a support function based approach would therefore be invaluable for applications such as generating band structures of materials containing low defect concentrations and calculating quantities such as effective masses. Where applicable, supercell band structures can then be compared to those obtained from primitive cell calculations, using unfolding or projection techniques. As an example, a spectral function projection method which uses such localized orbitals to obtain bandstructures for the primitive cell<sup>35</sup>, has been used to obtain insights into the behaviour of various complex heterostructures compared with the equivalent monolayers<sup>36,37</sup>.

Despite the formal connection between support functions and MLWFs, the extent to which they form an accurate basis for band structure interpolation in practice has not previously been explored. In other words, do the approximations made in a typical calculation, such as the imposition of strict localization radii or imperfect convergence, prevent accurate band structure interpolation? In this work we answer this question, while also considering what is the best method for obtaining band structures from a support function basis.

Methods for calculating band structures fall into two main classes, each relating to one of the two equivalent statements of Bloch's theorem. Codes based upon atomic-type basis sets adopt the interpolative tight-binding approach that is also employed for Wannier interpolation<sup>2,5</sup>, whereas within PW codes, an extrapolative approach<sup>3</sup> based on  $\mathbf{k} \cdot \mathbf{p}$  perturbation theory<sup>38</sup> has been successfully implemented<sup>39</sup>. In order to compare the two methods we choose to use the basis of localized orbitals generated by the linear-scaling method implemented in the ONETEP code<sup>34,40</sup>, however similar results would be expected for related methods. In principle, the localized orbitals of ONETEP, referred to as non-orthogonal generalized Wannier functions (NGWFs)<sup>32</sup>, straddle both approaches: they form a minimal basis of local orbitals that have been optimized in terms of an underlying basis equivalent to a set of PWs. When implemented within the ONETEP code, the two approaches give similar but different results that are interpreted here in terms of a simple one-dimensional Kronig Penney model.

The paper is organized as follows: in Sec. II the relevant theory underlying the ONETEP method is outlined, and two approaches are described by which the KS energies and orbitals may be obtained at a general point in the BZ. In Sec. III these two methods are compared for a one-dimensional toy model that illustrates the advantages and disadvantages of each. Section IV discusses the results obtained from both approaches when implemented in the ONETEP code and applied to a carbon nanotube. Conclusions are drawn in Sec. V.

## II. METHODOLOGY

### A. Linear-Scaling Methods

The single-particle density-matrix (DM)  $\rho(\mathbf{r}, \mathbf{r}')$  provides a complete description of the fictitious KS system in DFT. The DM may be described in the diagonal representation provided by the canonical KS orbitals  $\{\psi_n(\mathbf{r})\}$  and associated occupancies  $\{f_n\}$  (where  $f_n = 1$  for occupied states lying below the chemical potential and  $f_n = 0$  otherwise) or more generally in terms of a set of non-orthogonal orbitals  $\{\phi_\alpha(\mathbf{r})\}$  and a so-called density kernel  $K^{\alpha\beta}$ :

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_n \psi_n(\mathbf{r}) f_n \psi_n(\mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta(\mathbf{r}'). \quad (1)$$

Note that spin and  $\mathbf{k}$ -point labels have been dropped, the former for simplicity and the latter due to the assumption of  $\Gamma$ -point sampling justified above. Eq. 1 takes the form of a similarity transformation where there is a linear transformation between the canonical and non-orthogonal orbitals. For a linear-scaling method it is necessary to choose a representation in which the non-orthogonal orbitals are also localized in space and to exploit the nearsightedness of the DM<sup>41</sup> by enforcing sparsity on the density kernel. The ground state of a system may then be found by minimizing the total energy with respect to the DM subject to the constraints of normalization and idempotency<sup>42-45</sup>.

In ONETEP, the minimization is performed both with respect to the density kernel<sup>46</sup> and the NGWFs<sup>32</sup>. Various methods exist to calculate the density kernel at linear scaling cost; in ONETEP a combination of purification<sup>42</sup> and penalty functional<sup>47</sup> approaches are used, as described in detail elsewhere<sup>46</sup>. The initial guess for the NGWFs consists of a set of fireballs (FBs)<sup>48</sup>, which are generated by self-consistently solving the KS equations for an isolated atom using the same pseudopotential and exchange correlation functional as for the full system. The NGWF optimization is then carried out by expanding them in terms of a basis set of primitive functions that have been variously known as periodic cardinal sine (psinc)<sup>49</sup>, Dirichlet<sup>50</sup> or Fourier Lagrange<sup>51</sup> functions. These functions are the discrete equivalent of the Dirac delta function, which are arranged on the grid points of a regular mesh commensurate with the simulation cell. Each NGWF is associated with an atom and is expanded in a restricted set of psinc functions whose grid points lie within spheres of a given radius centred on that atom. Fast Fourier transforms (FFTs) may then be employed as in PW codes to apply the Hamiltonian, in particular the kinetic energy<sup>52</sup>, with a reduced FFT box<sup>53</sup> used consistently to achieve linear scaling.

ONETEP has been shown to achieve linear scaling for the entire calculation with controlled accuracy comparable to that of PW pseudopotential codes<sup>54</sup>. The result is a set of NGWFs, each of which has been optimized

*in situ* according to its individual chemical environment in terms of a basis equivalent to a set of PWs. A minimal set of NGWFs may then be used successfully whilst avoiding some of the pitfalls of local orbitals such as basis set superposition error<sup>55</sup>.

Using the approach described above, the KS eigenvalues are not explicitly referenced. However, they may be recovered by a one-off diagonalization in the optimized NGWF basis at the end of a calculation. While the NGWFs are designed to accurately represent the valence states, there is no guarantee that they are capable of also representing the KS conduction states. In some cases the first few unoccupied states might still be relatively well represented by the NGWF basis, however in general the conduction state energies are significantly overestimated, with some states not captured at all<sup>54,56</sup>. In order to overcome this problem, a method has been devised wherein a second set of NGWFs are optimized to explicitly represent a select number of (bound) conduction states. The density operator,  $\hat{\rho}$ , is used to project out the valence states, so that when a sufficiently large energy shift  $\sigma$  is applied, the conduction states of interest become lower in energy than the valence states. The Hamiltonian operator therefore becomes:

$$\hat{H} \rightarrow \hat{H} - \hat{\rho} (\hat{H} - \sigma) \hat{\rho}. \quad (2)$$

Using this Hamiltonian, a second non-self-consistent calculation can be performed following a ground state calculation to obtain a set of conduction NGWFs and associated density kernel, for which the total occupancy corresponds to the requested number of conduction states. In order to reach a comparable level of accuracy, the conduction NGWF radii are typically larger than those required for valence NGWFs. Before diagonalizing the unprojected Hamiltonian, the conduction NGWFs are combined with the valence NGWFs to form a joint basis. Provided sufficiently large localization radii are used, the joint basis is capable of representing both the occupied and unoccupied KS states to a similar high level of accuracy<sup>56,57</sup>.

## B. Band Structure Calculation

For any periodic system that is invariant under translation by a lattice vector  $\mathbf{R}$ , Bloch's theorem states that any eigenstate of the Hamiltonian  $\psi(\mathbf{r})$  must satisfy

$$\psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \psi(\mathbf{r}), \quad (3)$$

where  $\mathbf{k}$  is the Bloch wave vector or crystal momentum. An equivalent statement is

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u(\mathbf{r}), \quad (4)$$

where  $u(\mathbf{r}) = u(\mathbf{r} + \mathbf{R})$  is a cell-periodic function. If  $\mathbf{G}$  is a reciprocal lattice vector such that  $e^{i\mathbf{G}\cdot\mathbf{R}} = 1$  then from the first formulation (3) it is clear that the statements

of Bloch's theorem for wave vectors  $\mathbf{k}$  and  $\mathbf{k} + \mathbf{G}$  are identical and these may be considered equivalent. Hence  $\mathbf{k}$  may be chosen from the first BZ and the band structure may be represented as periodic in reciprocal space.

The KS Hamiltonian takes the following general form (in Hartree atomic units  $\hbar = m_e = 1$ ):

$$\hat{H} = -\frac{1}{2}\nabla^2 + V(\mathbf{r}) + \sum_I |p_I\rangle E_I \langle p_I|, \quad (5)$$

where the local potential  $V(\mathbf{r})$  is cell-periodic and the third term represents the action of a non-local norm-conserving pseudopotential in separable form<sup>58</sup> in which  $I$  is a composite index running over ions and angular momentum channels,  $\{|p_I\rangle\}$  are the projectors and  $\{E_I\}$  the associated energies.

To calculate the band structure from a linear-scaling calculation, it is necessary to obtain the KS energies and orbitals at an arbitrary  $\mathbf{k}$ -point in terms of a set of real local orbitals, e.g., NGWFs optimized in a self-consistent ground-state total energy calculation using the  $\Gamma$ -point only.

### 1. NGWF Approach

We now present two approaches to calculating the band structure, one of which is inspired by Eq. 3 and related to tight-binding (TB) and Wannier interpolation, in which the wave function at a general  $\mathbf{k}$ -point with band index  $n$  is written

$$\psi_{n\mathbf{k}}^{\text{TB}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \sum_{\alpha} c_{n\mathbf{k}\alpha}^{\text{TB}} \phi_{\alpha}(\mathbf{r} - \mathbf{R}), \quad (6)$$

involving a sum over all lattice vectors  $\mathbf{R}$ . In this approach identical copies of each NGWF for each cell are made from the home cell ( $\mathbf{R} = 0$ ). The Hamiltonian for a general  $\mathbf{k}$ -point is then constructed from Eq. 5 with additional  $\mathbf{k}$ -dependent phase factors arising when a matrix element corresponds to NGWFs from neighbouring cells. This is straightforward when the NGWFs are localized within a single cell – the  $k$ -dependent Hamiltonian matrix elements are written as

$$H_{\alpha\beta}^{\text{TB}}(\mathbf{k}) = \langle \phi_{\alpha} | \hat{H} | \phi_{\beta} \rangle \prod_i \theta(k_i, r_{\alpha i} - r_{\beta i}, R_i), \quad (7)$$

where  $\mathbf{r}_{\alpha(\beta)}$  is the centre of  $\phi_{\alpha(\beta)}$  expressed in fractional coordinates, and the one-dimensional phase factors  $\theta$  take the form

$$\theta(k, r, R) = \begin{cases} 1 & |r| \leq \frac{1}{2} \\ e^{ikR} & r > \frac{1}{2} \\ e^{-ikR} & r < -\frac{1}{2} \end{cases}. \quad (8)$$

Diagonalization of this Hamiltonian yields the KS energies and orbitals, i.e., the set of expansion coefficients  $\{c_{n\mathbf{k}\alpha}^{\text{TB}}\}$ , for the wave vector  $\mathbf{k}$ .

An alternative approach relates to the second formulation of Bloch's theorem (4) which expresses the wave function as

$$\psi_{\mathbf{n}\mathbf{k}}^{\text{KP}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{n}\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\alpha} c_{\mathbf{n}\mathbf{k}\alpha}^{\text{KP}} \sum_{\mathbf{R}} \phi_{\alpha}(\mathbf{r} - \mathbf{R}), \quad (9)$$

where the expansion of  $u_{\mathbf{n}\mathbf{k}}(\mathbf{r})$  in terms of NGWFs is explicitly cell-periodic, a construction that fits naturally with a PW or psinc basis set. Indeed, this approach is successfully employed in PW calculations. As suggested by  $\mathbf{k} \cdot \mathbf{p}$  perturbation theory, the Bloch phase factor may be treated analytically to derive a  $\mathbf{k}$ -dependent Hamiltonian that acts on the cell-periodic part of the wave function

$$\hat{H}(\mathbf{k}) = -\frac{1}{2}\nabla^2 - i\mathbf{k}\cdot\nabla + \frac{k^2}{2} + V(\mathbf{r}) + \sum_I |p_I(\mathbf{k})\rangle E_I \langle p_I(\mathbf{k})|, \quad (10)$$

where the phase factor has been incorporated into the projectors, which have thus become  $\mathbf{k}$ -dependent:  $\langle \mathbf{r} | p_I(\mathbf{k}) \rangle = e^{i\mathbf{k}\cdot\mathbf{r}} p_I(\mathbf{r})$ . In contrast to Eq. 7, the Hamiltonian matrix elements do not contain any additional phase factors and are simply calculated as

$$H_{\alpha\beta}^{\text{KP}}(\mathbf{k}) = \langle \phi_{\alpha} | \hat{H}(\mathbf{k}) | \phi_{\beta} \rangle. \quad (11)$$

In short, for the TB method  $\mathbf{k}$  enters into the Hamiltonian matrix elements via phase factors associated with the NGWFs, while for the  $\mathbf{k} \cdot \mathbf{p}$  method  $\mathbf{k}$  enters into the Hamiltonian matrix elements through a  $\mathbf{k}$ -dependent Hamiltonian operator. As for the TB method, the  $\mathbf{k} \cdot \mathbf{p}$  band structure is again obtained by diagonalization.

### III. TOY MODEL

Before comparing the two approaches to calculating band structures using ONETEP, we first make use of a simple one-dimensional model. In the following we describe the setup of this toy model, i.e. the choice of potential and definition of the localized basis sets which are designed to mimic some of the features of ONETEP NGWFs. Using this model, we identify characteristic features of the two methods for cases where the basis set is incomplete using both analytic and numerical analyses.

#### A. Description of the Model

We choose to use the one-dimensional Kronig-Penney model, where the Hamiltonian is defined as

$$\hat{H} = -\frac{1}{2} \frac{d^2}{dx^2} + V(x), \quad (12)$$

with a periodic potential  $V(x) = V(x + L)$ , given by

$$V(x) = \begin{cases} V_0 & 0 \leq x < b \\ 0 & -w \leq x < 0 \end{cases} \quad (13)$$

and the lattice parameter  $L = b + w$ . The potential and lowest energy state are illustrated in Fig. 1.

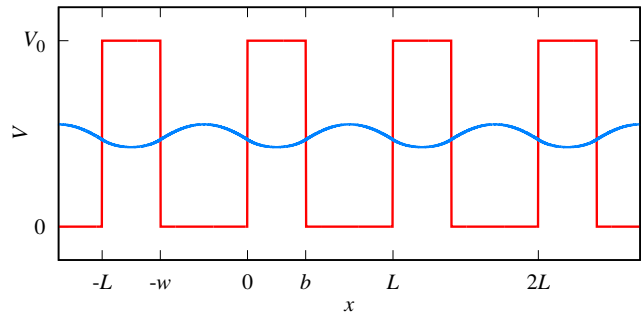


FIG. 1. (Color online) Potential (red) for the Kronig Penney model with the lowest energy state (blue) (at  $\mathbf{k} = 0$ ) superimposed.

The band structure for this model may be found by solving a transcendental equation numerically. This is used as a reference for results obtained from a fully numerical approach employing piecewise polynomial non-orthogonal basis functions (to mimic the role of NGWFs of different quality) for which Hamiltonian matrix elements may be found analytically. Basis functions are centred on the nodes of a regular grid with spacing  $a$  such that  $L = Ma$ .

The first basis is piecewise quadratic and continuous up to and including the first derivative. Basis functions only overlap their nearest neighbours. The function centred at  $x = 0$  is denoted  $N_0(x)$  and then  $N_m(x) \equiv N_0(x - ma)$  is centred at  $x = ma$ .

$$N_0(x) = \sqrt{\frac{30}{23a}} \times \begin{cases} 1 - 2(x/a)^2 & 0 \leq |x| < a/2 \\ 2(1 - |x|/a)^2 & a/2 \leq |x| < a \\ 0 & |x| \geq a \end{cases} \quad (14)$$

The second basis consists of cubic B-splines<sup>28</sup> continuous up to the second derivative but twice as wide as the first set and hence overlapping up to third nearest neighbours.

$$B_0(x) = \sqrt{\frac{140}{151a}} \times \begin{cases} 1 - (3/2)(x/a)^2 & 0 \leq |x| < a \\ + (3/4)(|x|/a)^3 & \\ (1/4)(2 - |x|/a)^3 & a \leq |x| < 2a \\ 0 & |x| \geq 2a \end{cases} \quad (15)$$

These two functions have been chosen to examine the effect of basis set quality. In particular, the optimization process for NGWFs imposes no explicit constraints on their continuity, either at the boundary or inside the localization region. Instead, continuity is desirable because it minimizes the kinetic energy, since the Fourier transform  $\tilde{f}(q)$  of a function  $f(x)$  decays as  $q^{-(n+1)}$  where  $n$  is the order of lowest discontinuous derivative. Hence the

TABLE I. Overlap, momentum and kinetic energy matrix elements for the two basis sets employed in Sec. III normalized such that  $s_0 = 1$ . By symmetry  $p_0 = 0$  and  $o_{-n} = o_n^*$  where  $o$  may be replaced by  $s$ ,  $p$  or  $t$ . Results for each basis should all be divided by the number shown in the second ( $\div$ ) column.

	$\div$	Overlap			Momentum			Kinetic			
		$s_1$	$s_2$	$s_3$	$ip_{1a}$	$ip_{2a}$	$ip_{3a}$	$t_0a^2$	$t_1a^2$	$t_2a^2$	$t_3a^2$
$N$	46	7	0	0	30	0	0	80	-40	0	0
$B$	2416	1191	120	1	1715	392	7	1680	-315	-504	-21

behaviours of the  $N$  and  $B$  bases are expected to differ qualitatively.

### B. Free-Electron Limit

First, the free-electron  $V_0 = 0$  limit is examined. In this case the exact forms of the wave functions are known:  $\psi_{nk}(x) \sim e^{i(k+nG)x}$  where  $G = 2\pi/L$ . The discrete translational symmetry (invariant under translation by any multiple of  $a$ ) of the Hamiltonian constructed using the  $N$  and  $B$  bases can similarly be exploited to write down the eigenvectors  $c_{nkm}^{\text{TB}} = e^{2\pi i(k/G+n)m/M}/\sqrt{M}$  and  $c_{nkm}^{\text{KP}} = e^{2\pi inm/M}/\sqrt{M}$  where the index  $m$  labelling the centre of the basis function is analogous to the NGWF index  $\alpha$  in Eqs. 6 and 9.

In the free-electron limit the translational invariance also means that matrix elements depend only on the separation of the basis centres, i.e., for some operator  $\hat{O}$ , the matrix elements for the  $B$  basis between functions centred at  $x = la$  and  $x = ma$ ,  $O_{lm} = o_{m-l}$  where

$$o_n = \int B_0(x) \hat{O} B_n(x) dx. \quad (16)$$

Matrix elements for the overlap  $\hat{O} \equiv 1$ , momentum  $\hat{O} \equiv -id/dx$  and kinetic  $\hat{O} \equiv -(1/2)d^2/dx^2$  operators are summarized in Table I. In this special case band structures may be calculated analytically from the expectation values of the TB and  $\mathbf{k} \cdot \mathbf{p}$  Hamiltonians in the corresponding eigenstates.

Defining  $\theta_{nk} = 2\pi(k/G+n)/M$ , the band structure for the TB approach (including overlaps up to third nearest neighbour) is given by

$$\varepsilon_n^{\text{TB}}(k) = \frac{t_0 + 2t_1 \cos \theta_{nk} + 2t_2 \cos 2\theta_{nk} + 2t_3 \cos 3\theta_{nk}}{s_0 + 2s_1 \cos \theta_{nk} + 2s_2 \cos 2\theta_{nk} + 2s_3 \cos 3\theta_{nk}} \quad (17)$$

which is a quotient of Fourier series as expected from tight-binding. This is manifestly periodic in reciprocal space, i.e.,  $\varepsilon_n^{\text{TB}}(k) = \varepsilon_n^{\text{TB}}(k + G)$ . Similarly,

$$\varepsilon_n^{\text{KP}}(k) = \frac{1}{2}k^2 + \frac{T_n + P_n|k|}{S_n} \quad (18)$$

where for basis functions that overlap up to third nearest neighbours

$$S_n = s_0 + 2s_1 \cos \theta_{n0} + 2s_2 \cos 2\theta_{n0} + 2s_3 \cos 3\theta_{n0} \quad (19)$$

and similarly for  $T_n$  in terms of  $\{t_m\}$ . For  $P_n$ :

$$P_n = 2(-1)^n i [p_1 \sin \theta_{n0} + p_2 \sin 2\theta_{n0} + p_3 \sin 3\theta_{n0}] \quad (20)$$

This takes the form of an expansion about  $k = 0$  and is not periodic in  $k$ , so that degeneracy of bands at the BZ boundary is not guaranteed. However this method does return the correct effective mass for all bands  $n$ :  $m_n^{\text{KP}}(k) = (d^2\varepsilon_n/dk^2)^{-1} = 1$ . By contrast, expanding the TB result for the lowest band ( $n = 0$ ) about  $k = 0$ ,  $\varepsilon_0^{\text{TB}}(k) \approx T_0/S_0 + k^2/[2m_n^{\text{TB}}(0)]$  where

$$m_n^{\text{TB}}(0) = \frac{S_0^2/2}{(s_1 + 4s_2 + 9s_3)T_0 - (t_1 + 4t_2 + 9t_3)S_0} \quad (21)$$

For the  $N$  basis,  $m_n^{\text{TB}}(k = 0) = 3/4$  whereas for the  $B$  basis the correct value is obtained. Hence the accuracy of the band curvature depends upon the quality of the basis used.

Both methods predict the same value for the bottom of the lowest band  $\varepsilon_0(0) = T_0/S_0$  which should vanish. This is the case for both bases used here, but is not a universal feature and depends upon the ability of a basis set to describe exactly the lowest state  $\psi_{00}(x) = 1/\sqrt{L}$ , i.e. a constant.

The qualitative difference between the two basis sets is clearly demonstrated by examining  $\varepsilon_1(0)$ . For both methods, the  $B$  basis is correct to leading order with an error of order  $M^{-6}$  whereas the  $N$  basis is wrong by a factor of  $4/3$  consistent with the error in the TB effective mass.

The methods are also distinguished by examining the energies of the lowest two bands at the BZ boundary, which should degenerate. The TB method guarantees this by construction, but for the  $\mathbf{k} \cdot \mathbf{p}$  method this is not the case, so that for a poor quality basis gaps appear at the BZ boundary. Indeed, for the  $\mathbf{k} \cdot \mathbf{p}$  method, for  $k \rightarrow k + G$ ,  $u_{k+G}(x) = u_k(x)e^{iGx}$  and thus the basis used to describe  $u_k(x)$  must also be able to describe its product with  $e^{iGx}$ . For all but a few high energy PWs at the edge of the cutoff sphere, this product is perfectly represented in a PW basis and so no such gaps arise. However for the toy model, and as will be seen in the following also for the NGWF basis, this is not expected to be the case.

To illustrate this point we have considered two types of poor quality basis set: an intrinsically lower quality basis which is nonetheless well converged in terms of the number of basis functions, i.e. basis  $N$  with  $M = 200$ , and a basis set which is of high quality but with an insufficient number of basis functions, i.e. basis  $B$  with  $M = 8$ . In terms of ONETEP, these two setups might loosely be compared to a set of NGWFs with large enough localization radii but which have been poorly optimized (e.g. loose convergence thresholds) or not optimized at all, and to NGWFs which have been well optimized but which have insufficient degrees of freedom, e.g. due to localization radii which are too small.

TABLE II. Low energy eigenvalues for the free-electron case at the  $\Gamma$ -point and BZ boundary for both basis sets compared to the analytic values, calculated with the given number of basis functions,  $M$ . Where applicable, both TB and  $\mathbf{k} \cdot \mathbf{p}$  values are included. Energies are given in Ha.

	Analytic	$N(M=200)$		$B(M=8)$	
		TB	$\mathbf{k} \cdot \mathbf{p}$	TB	$\mathbf{k} \cdot \mathbf{p}$
$\varepsilon_0(G/2)$	4.9348	6.5798	4.9348	4.9348	4.9348
$\varepsilon_1(0)$	19.7392	26.3198		19.7394	
$\varepsilon_1(G/2)$	4.9348	6.5798	11.5164	4.9348	4.9350

Using these two setups we have calculated the free-electron band structure numerically using both the  $\mathbf{k} \cdot \mathbf{p}$  and TB methods; the results are plotted in Fig. 2, with selected eigenvalues also given in Table II. The anticipated differences between the two methods can be clearly seen for the  $N$  basis, where the poor quality of the basis set results in calculated eigenvalues with energies which are increasingly too high for increasing band index. In such a case, the  $\mathbf{k} \cdot \mathbf{p}$  method preserves the correct band curvature as expected, but at the expense of unphysical gaps opening at the BZ boundary. The TB method, on the other hand, preserves the correct BZ degeneracies, but results in the incorrect band curvature. The selected number of basis functions for  $N$  is sufficiently high that the eigenvalues have converged on the analytic values, e.g. the factor of 4/3 error in  $\varepsilon_1(0)$  can be seen in Table II. The same qualitative differences between the two methods for calculating band structures are also present for the  $B$  basis, however since the basis is of much higher quality even for such a small number of basis functions, the errors in the eigenvalues are very small and thus the two methods give very similar results.

### C. Kronig-Penney Model

We now introduce a localized potential in the form of the Kronig-Penney potential with the values  $V_0 = 100$ ,  $b = 0.25$  and  $L = 1.0$ , with the aim of exploring weakly-hybridised states. We keep the same basis setup as before. It is interesting at this point to also consider the implications of supercell calculations, i.e. how the two methods for band structure calculation compare when band folding and unfolding occurs, since such effects frequently come into play for the typical system sizes studied with ONETEP. In order to investigate this point, we calculated band structures for a primitive cell of length  $L$ , as well as a supercell of length  $2L$ , containing two repeats of the primitive cell.

The resulting band structures are plotted in Fig. 3. For the primitive cell calculation, the differences between the two methods are less noticeable than for the free-electron band structure and are only really distinguishable close to the BZ boundary, where the methods give differing band curvatures, with the TB results slightly closer to the

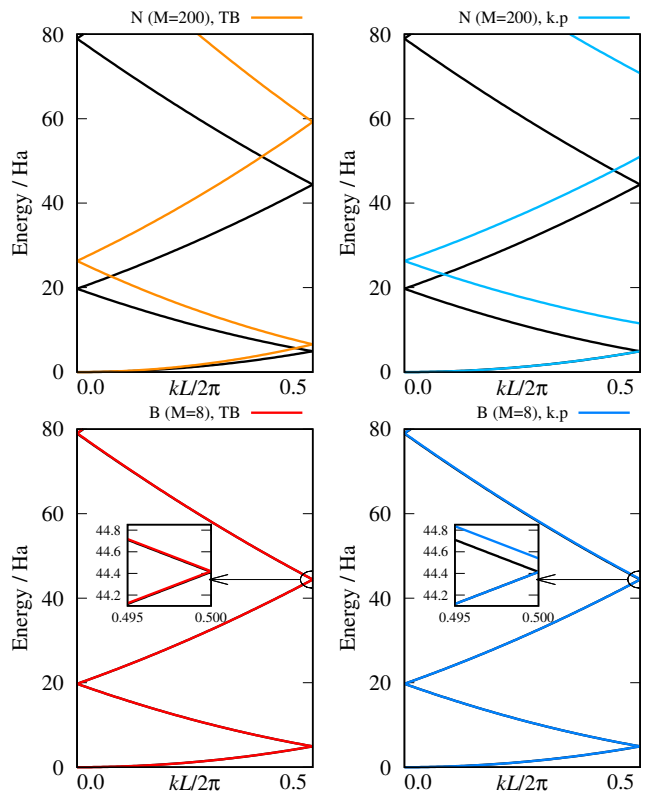


FIG. 2. (Color online) Low energy free-electron band structures for the  $N$  [top] and  $B$  [bottom] basis sets using both the TB [left] and  $\mathbf{k} \cdot \mathbf{p}$  [right] methods. The analytical solution is shown for reference (black lines).

reference. For the supercell results however, the discrepancy between the two methods is more significant, with gaps again appearing at the BZ boundary for the  $\mathbf{k} \cdot \mathbf{p}$  method. In this case we have direct access to the band structure of the primitive cell, however for a ONETEP calculation where it can be the case that only supercell calculations are accessible, an unfolding procedure would be necessary. In this case, such an unfolding would result in discontinuities in the bands when using the  $\mathbf{k} \cdot \mathbf{p}$  method.

In summary, we have compared the two methods for calculating band structures through the use of a one-dimensional model employing localized basis sets of different quality. When the basis set is of good quality, the two methods give comparable results, however for an incomplete basis set there are notable differences. Considering first the free electron case, we have demonstrated that the TB style approach imposes periodicity in reciprocal space, guaranteeing the correct band degeneracies at the BZ boundary. For the  $\mathbf{k} \cdot \mathbf{p}$  method, however, this is not the case, so that for a poor basis set, unphysical gaps appear at the BZ boundary. On the other hand, the  $\mathbf{k} \cdot \mathbf{p}$  approach gives the correct effective mass, while for the TB method the use of a low quality basis set results

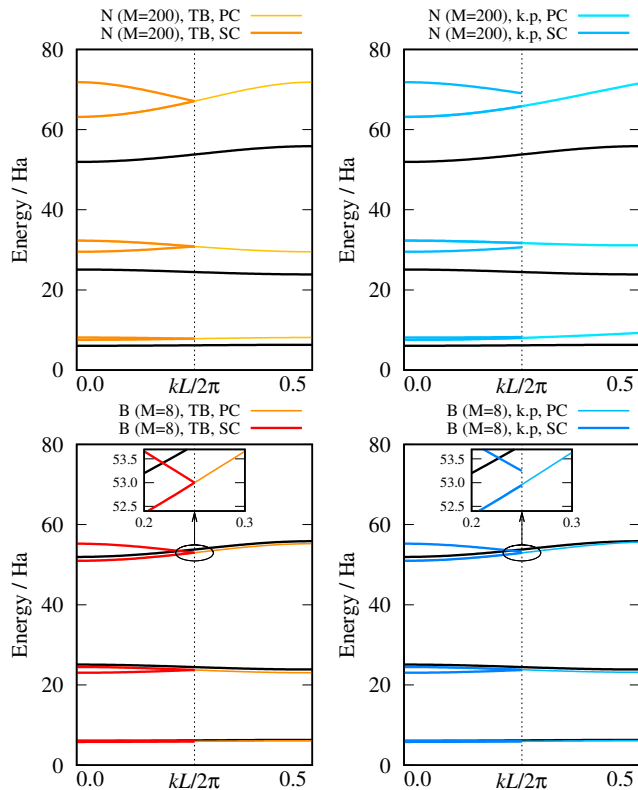


FIG. 3. (Color online) The four lowest energy bands for a Kronig Penney potential with  $V_0 = 100$ ,  $b = 0.25$  and  $L = 1.0$  for the  $N$  [top] and  $B$  [bottom] basis sets using both the TB [left] and  $\mathbf{k} \cdot \mathbf{p}$  [right] methods. The band structures have been calculated for a primitive cell (PC), as well as for a supercell (SC) comprising two primitive cells. The analytic solution is also plotted for comparison (black lines).

in incorrect band curvatures. When a Kronig-Penney potential is introduced to the model, the band structure generated with the TB method is qualitatively better, with unphysical gaps again appearing for the  $\mathbf{k} \cdot \mathbf{p}$  approach for supercell calculations. Given the above, the TB style method is the clear method of choice.

#### IV. (10,0) CARBON NANOTUBE

Having explored the differences between the two methods for band structure calculation in the context of a toy model, we now use the case study of a (10,0) carbon nanotube (CNT), which is depicted in Fig. 4, to see how the above conclusions carry over to a real system. In order to have a reference against which to compare the results of our ONETEP calculations, we have also calculated the band structure using the PW based CASTEP<sup>59</sup> DFT code. The same pseudopotential was used for both ONETEP and CASTEP, while the ONETEP psinc grid spacing was also set to be equivalent to the PW cut-off

energy of 916 eV which was used in CASTEP. The grid spacing (and thus cut-off energy) was selected to ensure the number of psinc grid points was divisible by the number of CNT repeat units, to ensure translational symmetry. The nanotube was aligned along the  $z$ -axis, and the unit cell was padded by more than 25 Å in the  $x$  and  $y$  directions to reduce interactions between periodic images. We used the PBE exchange-correlation functional<sup>60</sup>. All ONETEP calculations were at the  $\Gamma$ -point only.

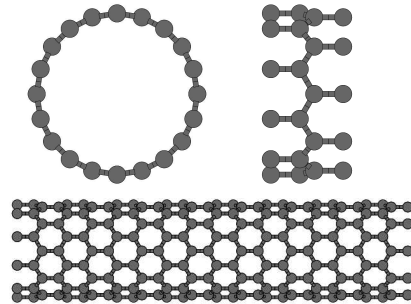


FIG. 4. Atomic structure of the (10,0) CNT. The end view [top left], primitive cell [top right] and an eight repeat unit supercell [bottom] are depicted.

In the first instance, we performed ONETEP calculations for a supercell containing eight repeat units of the CNT and thus 320 atoms. We used four NGWFs per carbon atom. The CASTEP calculation was for the primitive cell, with a Monkhorst-Pack  $\mathbf{k}$ -point mesh of  $1 \times 1 \times 8$ . The ONETEP supercell band structures were unfolded for comparison with CASTEP. As with the toy model, we wished to consider cases where the basis is of relatively low quality as well as a good calculation setup, and thus considered two scenarios where the basis would be considered poor, namely an NGWF basis which consists of unoptimized fireballs but with reasonable localization radii of 7  $a_0$ , and a well optimized NGWF basis but with small radii of 5  $a_0$ , which are respectively denoted by FB(7) and NGWF(5). This was compared with a good calculation setup with an optimized NGWF basis with radii of 7  $a_0$ , denoted by NGWF(7).

The band structures calculated using both methods and associated densities of states (DOS) are plotted in Fig. 5. Considering first only the valence states, it is clear that the unoptimized basis is poor, with the band structure showing significant differences with the PW reference, both at the  $\Gamma$ -point and across the BZ. The situation is improved upon optimizing the NGWFs, even for small radii, for which there are only quite small deviations from the PW band structure, while for the calculation with larger NGWF radii there is excellent agreement between the ONETEP and PW band structure. Importantly, the NGWF basis is capable of correctly reproducing the band structure across the whole BZ, despite having only been explicitly optimized at the  $\Gamma$ -point.

For the conduction states, although the final calcula-

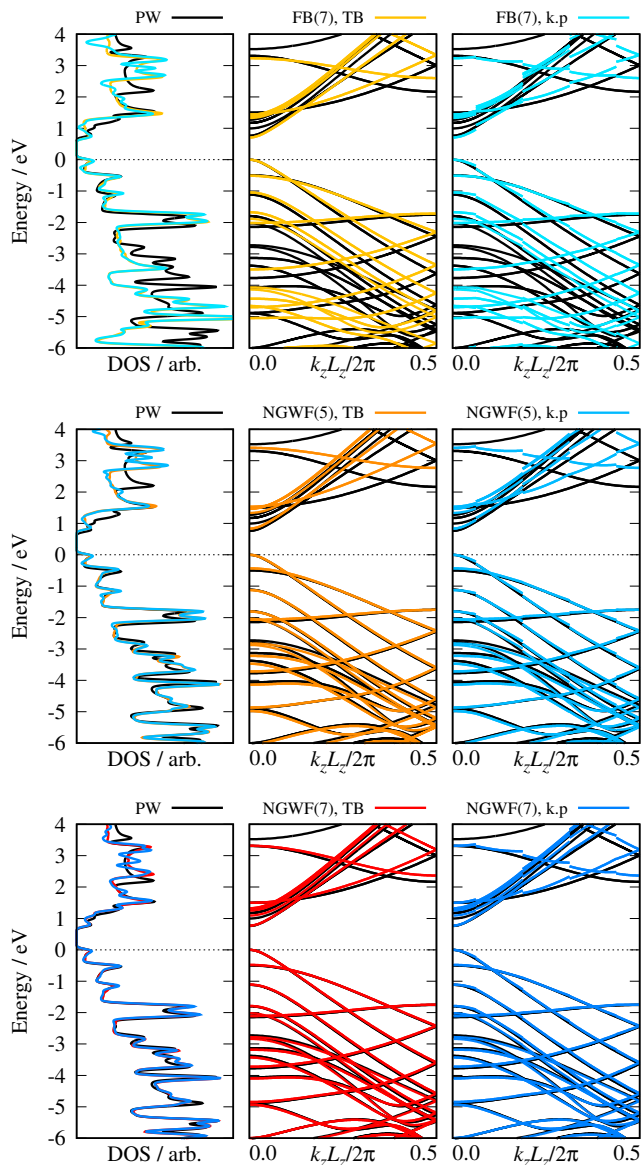


FIG. 5. (Color online) Density of states [left] and band structures calculated using the TB [middle] and  $\mathbf{k} \cdot \mathbf{p}$  [right] methods for different setups, compared to the PW reference. The ONETEP setups consist of unoptimized fireballs (FB) with NGWF radii of  $7 a_0$  [top], and optimized NGWFs with radii of  $5 a_0$  [middle] and  $7 a_0$  [bottom]. In each case there were four NGWF functions per carbon atom. For the DOS a Gaussian smearing of  $0.05$  eV was applied. Each plot has been shifted so that the highest occupied molecular orbital (HOMO) is at zero.

tion setup shows the smallest errors, in each case there is nonetheless a visible error compared to the PW result. This is unsurprising given that, as discussed above, the NGWF optimization procedure is designed to construct a basis which can accurately represent the valence states. Therefore, even though the error is reduced by increasing

the NGWF radii, this is not in itself sufficient to guarantee a good basis representation for the empty states. In the following we therefore also perform band structure calculations using a second set of conduction NGWFs.

As predicted, discontinuities are present in the band structure when using the  $\mathbf{k} \cdot \mathbf{p}$  method, although these are smallest for the good quality basis and larger for all basis sets for the conduction states. The origins of this are more clearly demonstrated in Fig. 6, where the calculation with the small NGWF radii was repeated for a smaller supercell of four units. The folded and unfolded band structures are compared, so that the correspondence between unphysical gaps at the BZ boundary and discontinuities at the corresponding  $\mathbf{k}$ -points in the unfolded band structure is clearly visible. As expected, no such gaps are present for the TB method and thus the unfolded bands remain smooth and continuous. Furthermore, the shape of the conduction bands is also better calculated with the TB method, for example for the band which is just above  $3$  eV at the  $\Gamma$  point, the local shape of the bands (i.e. between discontinuities) deviates noticeably from the PW band. This is consistent with the observations from the Kronig-Penney model, where the TB approach resulted in bands with a shape closer to the reference.

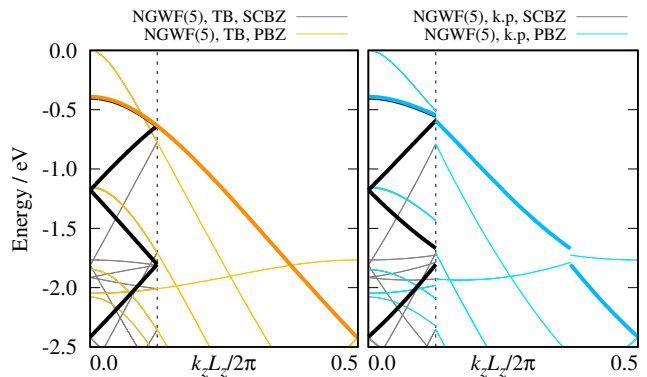


FIG. 6. (Color online) Demonstration of the origin of band discontinuities when using the  $\mathbf{k} \cdot \mathbf{p}$  method [right], compared with the continuous bands resulting from the TB method [left]. The folded supercell (SC) band structure for a supercell of four repeat units is contrasted with the unfolded band structure in the primitive Brillouin zone (PBZ). A single band has been highlighted in black (darker blue) to emphasize the presence of gaps (discontinuities) in the folded (unfolded) band. The plot has been shifted so that the HOMO is at zero.

It is also interesting to compare the ONETEP band structures with the PW reference in a more quantitative manner. To do so, we calculate the band structures over a dense sampling of  $\mathbf{k}$ -points and determine the average absolute difference between the PW and ONETEP energies, after first shifting the energy values to account for the difference in HOMO values. This average error is calculated separately for the valence states and for a



fixed number of conduction states; the values are given in Table III. Similar values were obtained when the energies were compared for only those  $\mathbf{k}$ -points that were included in the Monkhorst-Pack mesh for the CASTEP calculation. As expected, the valence errors for the unoptimized basis are significantly higher, with the error decreasing for increasing NGWF radii once the basis is optimized. In all three cases the conduction errors are significant. Interestingly, the average errors are very similar for the two methods, with  $\mathbf{k} \cdot \mathbf{p}$  worse for the conduction states but otherwise virtually identical for the valence states. Nonetheless, the TB approach is clearly the method of choice, in agreement with the conclusions drawn from the toy model calculations.

### A. Conduction States

We now investigate in more detail the conduction state band structure. In order to have the possibility of employing large NGWF radii, we increase the size of the supercell to contain 11 repeat units, otherwise the simulation setup remains the same. In the following we use only the TB method, since it has proven to be the better choice.

We compare four different calculation setups: two with only valence NGWFs and two with both valence and conduction NGWFs. These combinations were chosen to allow us to compare the varying impacts of increasing the number of valence NGWFs without performing a conduction calculation, adding conduction NGWFs, and increasing the conduction NGWF radii. For shorthand, we use the notation  $N_v \text{val}(r_v) + N_c \text{cond}(r_c)$ , where  $N_v$  ( $N_c$ ) is the number of valence (conduction) NGWFs and  $r_v$  ( $r_c$ ) denotes the valence (conduction) NGWF radii. The four setups are as follows:

1. 4 valence NGWFs with radii of  $7 a_0$ , no conduction NGWFs (4val(7))
2. 9 valence NGWFs with radii of  $7 a_0$ , no conduction NGWFs (9val(7))
3. 4 valence NGWFs with radii of  $7 a_0$  and 5 conduction NGWFs with radii of  $7 a_0$  (4val(7)+5cond(7))
4. 4 valence NGWFs with radii of  $7 a_0$  and 5 conduction NGWFs with radii of  $11 a_0$  (4val(7)+5cond(11))

The number of conduction states to optimize was selected to include all states which are less than 5 eV above the HOMO at the  $\Gamma$ -point, as calculated with CASTEP. A significant number of additional states (eight times the number actually required) were included in the conduction density kernel in an initial pre-optimization process<sup>56</sup>.

The results are plotted in Fig. 7 and the corresponding quantitative errors given in Table III. For both the

ONETEP and PW band structures and DOS, we only depict explicitly optimized conduction states. For 4val(7), the lower energy conduction bands are in relatively good agreement with the PW results, with the largest differences appearing at the BZ boundary. However the higher energy nearly free-electron states, which are equivalent to the weakly bound delocalized states that sit above and below a graphene sheet<sup>61</sup>, are completely absent. When the number of NGWFs per atom is increased to nine (9val(7)), the agreement for the low energy states is excellent and there are some free-electron like states present, however their energies are nonetheless significantly over-estimated.

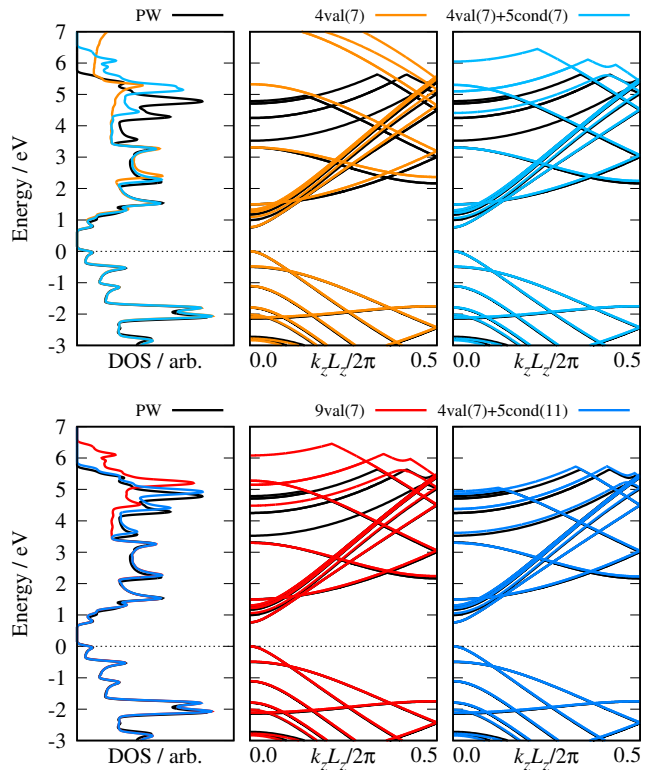


FIG. 7. (Color online) Density of states [left] and band structures calculated using the TB method for different setups, compared to the PW reference. The NGWF setups are given in the figure, where  $N_v \text{val}(r_v)$  denotes  $N_v$  valence NGWFs with radius  $r_v$  and  $N_c \text{cond}(r_c)$  denotes  $N_c$  conduction NGWFs with radius  $r_c$ . A Gaussian smearing of 0.05 eV was applied for the DOS. Each plot has been shifted so that the highest occupied molecular orbital (HOMO) is at zero.

The 9val(7) and 4val(7)+5cond(7) setups have the same number of basis degrees of freedom (same total number of NGWFs and radii) and so can be directly compared. Indeed the band structures are of similar quality, with 4val(7)+5cond(7) showing marginally better agreement on average. However when the conduction NGWF radii are increased to  $11 a_0$  (4val(7)+5cond(11)), the agreement for all states is markedly improved, with

TABLE III. Average and maximum errors across bands and  $\mathbf{k}$ -points with respect to the PW band structure, in eV. Calculation setups are the same as those used for Fig. 5 and 7. The number of CNT units, whether or not the NGWFs have been optimized, the number of valence and conduction NGWFs per atom and their respective localization radii are indicated. The valence error is that averaged across all occupied states, while the conduction error has been calculated for the number of unoccupied states for which the calculated PW energy is less than 5.0 eV above the HOMO at the  $\Gamma$  point.

Number of CNT units	Optimized NGWFs	Valence NGWFs		Conduction NGWFs		TB Error (Av., Max)		$\mathbf{k} \cdot \mathbf{p}$ Error (Av., Max)	
		Number	Radii	Number	Radii	Valence	Conduction	Valence	Conduction
8	✗	4	7.0	-	-	0.49, 1.21	0.57, 3.16	0.49, 1.22	0.59, 3.16
8	✓	4	5.0	-	-	0.04, 0.10	0.72, 3.25	0.04, 0.10	0.73, 3.26
8	✓	4	7.0	-	-	0.03, 0.08	0.52, 3.19	0.03, 0.09	0.54, 3.20
11	✓	4	7.0	-	-	0.03, 0.08	0.52, 3.19	-	-
11	✓	9	7.0	-	-	0.03, 0.08	0.21, 1.37	-	-
11	✓	4	7.0	5	7.0	0.03, 0.08	0.20, 1.34	-	-
11	✓	4	7.0	5	11.0	0.03, 0.08	0.04, 0.16	-	-

the quantitative difference with the PW band structure at a similar level as for the valence states. As with the valence band structure, it is important to note that a consistent level of accuracy is maintained throughout the BZ.

Given that the accuracy of the 9val(7) and 4val(7)+5cond(7) band structures are of similar quality, one might wonder if a conduction calculation is indeed necessary, or if it would also be possible to achieve high quality results by continuing to increase the valence NGWF radii. However, although increasing the radii might indeed improve the quality, systematic improvement is not guaranteed, particularly for the higher energy conduction states. The results obtained using the conduction approach on the other hand, are expected to be at least as good as a valence only calculation with the same number of degrees of freedom, and in the majority of cases should be significantly better.

## V. CONCLUSIONS

In summary, we have presented two methods for calculating band structures using a local orbital basis set, which are derived from  $\mathbf{k} \cdot \mathbf{p}$  perturbation theory and tight-binding respectively. The two methods were initially compared in the context of a one-dimensional model, using a Kronig-Penney potential and two different localized basis sets: a piecewise quadratic nearest neighbour basis and cubic B-splines. Subsequently, the two approaches were compared for a set of optimized local orbitals, referred to as NGWFs, obtained from the ONETEP linear-scaling DFT code for the example of a (10, 0) carbon nanotube.

For high quality basis sets, i.e. B-splines in the case of the toy model and an optimized set of NGWFs with sufficient degrees of freedom for the ONETEP calculations, band structures generated with the two methods are similar. However, for a lower quality basis, such as the near-

est neighbour basis set or unoptimized local orbitals, the  $\mathbf{k} \cdot \mathbf{p}$  style method was found to produce unphysical gaps at the BZ boundary. When unfolding band structures obtained from supercell calculations, these gaps translate to discontinuities in the band structure. This problem can arise even for a moderate quality basis set, although the discontinuities are much smaller. The tight-binding style method guarantees the correct BZ periodicity so that this problem does not occur. Furthermore when the error with respect to the plane-wave reference is assessed quantitatively, the errors are smaller than or equal to those for the  $\mathbf{k} \cdot \mathbf{p}$  style method. Therefore the tight-binding style approach is the recommended method for calculating band structures in ONETEP or similar linear scaling approaches.

The CNT band structures generated from ONETEP were compared to results from a traditional plane-wave pseudopotential calculation, both qualitatively and quantitatively. When a reasonable localized radius was used for the NGWF basis, excellent agreement was observed with the PW reference across the entire BZ for the occupied and low energy unoccupied KS states. However the higher energy free-electron-like bands were missing from the generated band structure. When a second set of NGWFs optimized explicitly for the conduction states is also included, excellent agreement is obtained for both occupied and unoccupied states, including the free-electron-like states. In short, given sufficient degrees of freedom, including the use of conduction NGWFs where necessary, the NGWF basis generated by ONETEP forms an excellent basis for band structure interpolation using a tight-binding style approach.

In the future it would be interesting to also compare the above results with ONETEP calculations where the imposition of the NGWF localization is relaxed in one or more directions, i.e. for partially localized or hybrid Wannier functions. This approach has recently been implemented in ONETEP<sup>62</sup> and is particularly applicable to 2D-periodic systems such as surfaces or interfaces. In

particular, such an approach would be expected to result in better convergence for the higher energy conduction states.

Finally, beyond the calculation of band structures, MLWFs are also widely employed as a basis for model Hamiltonian approaches, including tight-binding. Given the above conclusions, the support functions generated from linear-scaling DFT should also be highly suitable for this purpose. In the context of the material considered in this work, an example might be the derivation of tight-binding parameters for defective nanotubes. Given the large system sizes which are accessible to linear-scaling DFT, this approach would also provide an opportunity for validating effective models. For example one could directly compare large scale DFT calculations with smaller scale model Hamiltonian calculations in order to determine whether sufficient degrees of freedom had been included in the model. A similar process can be applied to semiconductors at low doping levels. For example, effective Hamiltonians derived from ONETEP calculations have been used to study sulphur-doped silicon at various

(low) defect concentration levels<sup>63</sup>.

## VI. ACKNOWLEDGMENTS

We acknowledge financial support from the UK Engineering and Physical Sciences Research Council under the following grants: L.E.R. under grant EP/P504694/1, N.D.M.H. under grant EP/G05567X/1, and L.E.R and P.D.H. under grant EP/J015059/1. The work of G.J.C. was supported by a Nuffield Undergraduate Research Bursary URB/01923/G. P.D.H. acknowledges the support of a Royal Society University Research Fellowship. We are grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1). Calculations were also performed on the Imperial College High Performance Computing Service. Underlying data for this article are available at <https://doi.org/10.5281/zenodo.1206196>.

- 
- \* p.haynes@imperial.ac.uk
- <sup>1</sup> E. L. Shirley, Phys. Rev. B **54**, 16464 (1996)
  - <sup>2</sup> N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997)
  - <sup>3</sup> C. J. Pickard and M. C. Payne, Phys. Rev. B **59**, 4685 (1999)
  - <sup>4</sup> D. Prendergast and S. G. Louie, Phys. Rev. B **80**, 235126 (2009)
  - <sup>5</sup> I. Souza, N. Marzari and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001)
  - <sup>6</sup> N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012)
  - <sup>7</sup> A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008)
  - <sup>8</sup> A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt and N. Marzari, Comput. Phys. Commun. **185**, 2309 (2014)
  - <sup>9</sup> D. R. Hamann and D. Vanderbilt, Phys. Rev. B **79**, 045109 (2009)
  - <sup>10</sup> L. X. Benedict, E. L. Shirley and R. B. Bohn, Phys. Rev. Lett. **80**, 4514 (1998)
  - <sup>11</sup> R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993)
  - <sup>12</sup> D. Vanderbilt and R. D. King-Smith, Phys. Rev. B **48**, 4442 (1993)
  - <sup>13</sup> R. Resta, Rev. Mod. Phys. **66**, 889 (1994)
  - <sup>14</sup> T. Thonhauser, D. Ceresoli, D. Vanderbilt and R. Resta, Phys. Rev. Lett. **95**, 137205 (2005)
  - <sup>15</sup> D. Ceresoli, T. Thonhauser, D. Vanderbilt and R. Resta, Phys. Rev. B **74**, 024408 (2006)
  - <sup>16</sup> I. Souza and D. Vanderbilt, Phys. Rev. B **77**, 054438 (2008)
  - <sup>17</sup> W. Zhang, R. Yu, H.-J. Zhang, X. Dai, and Z. Fang, New J. Phys. **12**, 065013 (2010)
  - <sup>18</sup> A. A. Soluyanov and D. Vanderbilt, Phys. Rev. B **83**, 235401 (2011)
  - <sup>19</sup> S. Coh, D. Vanderbilt, A. Malashevich and I. Souza, Phys. Rev. B **83**, 085108 (2011)
  - <sup>20</sup> G. Galli, Curr. Opin. Solid St. M. **1**, 864 (1996)
  - <sup>21</sup> S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999)
  - <sup>22</sup> D. R. Bowler and T. Miyazaki, Rep. Prog. Phys. **75**, 036503 (2012)
  - <sup>23</sup> N. D. M. Hine, M. Robinson, P. D. Haynes, C.-K. Skylaris, M. C. Payne and A. A. Mostofi, Phys. Rev. B **83**, 195102 (2011)
  - <sup>24</sup> N. D. M. Hine, P. D. Haynes, A. A. Mostofi, C.-K. Skylaris and M. C. Payne, Comput. Phys. Commun. **180**, 1041 (2009)
  - <sup>25</sup> D. R. Bowler and T. Miyazaki, J. Phys.: Condens. Matter **22**, 074207 (2010)
  - <sup>26</sup> L. E. Ratcliff, S. Mohr, G. Huhs, T. Deutsch, M. Masella and L. Genovese, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **7**, e1290 (2017)
  - <sup>27</sup> D. J. Cole and N. D. M. Hine, J. Phys.: Condens. Matter **28**, 393001 (2016)
  - <sup>28</sup> E. Hernández, M. J. Gillan and C. M. Goringe, Phys. Rev. B **55**, 13485 (1997)
  - <sup>29</sup> J.-L. Fattebert and J. Bernholc, Phys. Rev. B **62**, 1713 (2000)
  - <sup>30</sup> S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, S. Goedecker, J. Chem. Phys. **140**, 204110 (2014)
  - <sup>31</sup> S. Mohr, L. E. Ratcliff, L. Genovese, D. Caliste, P. Boulanger, S. Goedecker, T. Deutsch, Phys. Chem. Chem. Phys. **17**, 31360 (2015)
  - <sup>32</sup> C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez and M. C. Payne, Phys. Rev. B **66**, 035119 (2002)
  - <sup>33</sup> C.-K. Skylaris and P. D. Haynes, J. Chem. Phys. **127**, 164712 (2007)
  - <sup>34</sup> C.-K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005)
  - <sup>35</sup> G. C. Constantinescu and N. D. M. Hine, Phys. Rev. B **91**, 195416 (2015)

- <sup>36</sup> N. R. Wilson, P. Nguyen, K. Seyler, P. Rivera, A. J. Marsden, Z. P. L. Laker, G. C. Constantinescu, V. Kandyba, A. Barinov, N. D. M. Hine, X. Xu and D. H. Cobden, *Sci. Adv.* **3**, e1601832 (2017)
- <sup>37</sup> G. C. Constantinescu and N. D. M. Hine, *Nano Lett.* **16**, 2586 (2016)
- <sup>38</sup> E. Kane, *Semicond. Semimet.* **1**, 75 (1963)
- <sup>39</sup> C. J. Pickard and M. C. Payne, *Phys. Rev. B* **62**, 4383 (2000)
- <sup>40</sup> P. D. Haynes, C.-K. Skylaris, A. A. Mostofi and M. C. Payne, *Phys. Stat. Sol. (b)* **243**, 2489 (2006)
- <sup>41</sup> W. Kohn, *Phys. Rev. Lett.* **76**, 3168 (1996)
- <sup>42</sup> R. McWeeny, *Rev. Mod. Phys.* **32**, 335 (1960)
- <sup>43</sup> X.-P. Li, R. W. Nunes and D. Vanderbilt, *Phys. Rev. B* **47**, 10891 (1993)
- <sup>44</sup> M. S. Daw, *Phys. Rev. B* **47**, 10895 (1993)
- <sup>45</sup> R. W. Nunes and D. Vanderbilt, *Phys. Rev. B* **50**, 17611 (1994)
- <sup>46</sup> P. D. Haynes, C.-K. Skylaris, A. A. Mostofi and M. C. Payne, *J. Phys.: Condens. Matter* **20**, 294207 (2008)
- <sup>47</sup> P. D. Haynes and M. C. Payne, *Phys. Rev. B* **59**, 12173 (1999)
- <sup>48</sup> O. F. Sankey and D. J. Niklewski, *Phys. Rev. B*, **40**, 3979 (1989)
- <sup>49</sup> A. A. Mostofi, P. D. Haynes, C.-K. Skylaris and M. C. Payne, *J. Chem. Phys.* **119**, 8842 (2003)
- <sup>50</sup> D. Baye and P.-H. Heenen, *J. Phys. A: Math. Gen.* **19**, 2041 (1986)
- <sup>51</sup> K. Varga, Z. Zhang and S. T. Pantelides, *Phys. Rev. Lett.* **93**, 176403 (2004)
- <sup>52</sup> C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, C. J. Pickard and M. C. Payne, *Comput. Phys. Commun.* **140**, 315 (2001)
- <sup>53</sup> A. A. Mostofi, C.-K. Skylaris, P. D. Haynes and M. C. Payne, *Comput. Phys. Commun.* **147**, 788 (2002)
- <sup>54</sup> C.-K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *J. Phys.: Condens. Matter* **17**, 5757 (2005)
- <sup>55</sup> P. D. Haynes, C.-K. Skylaris, A. A. Mostofi and M. C. Payne, *Chem. Phys. Lett.* **422**, 345 (2006)
- <sup>56</sup> L. E. Ratcliff, N. D. M. Hine and P. D. Haynes, *Phys. Rev. B* **84**, 165131 (2011)
- <sup>57</sup> L. E. Ratcliff and P. D. Haynes, *Phys. Chem. Chem. Phys.* **15**, 13024 (2013)
- <sup>58</sup> L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982)
- <sup>59</sup> S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M. C. Payne, *Z. Kristall.* **220**, 567 (2005)
- <sup>60</sup> J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, **77**, 3865 (1996)
- <sup>61</sup> M. Houssa, A. Dimoulas and A. Molle (Eds.) *2D Materials for Nanoelectronics* CRC Press (Taylor & Francis Group), Boca Raton, FL (2016)
- <sup>62</sup> A. Greco, *Development and Application of First-Principles Methods for Complex Oxide Surfaces and Interfaces*, PhD Thesis, Imperial College London (2017)
- <sup>63</sup> E. G. Carnio, N. D. M. Hine and R. A. Römer, arXiv:1710.01742v2 (2018)