Arch. Rational Mech. Anal. 246 (2022) 1–60
Digital Object Identifier (DOI) https://doi.org/10.1007/s00205-022-01809-w

Body-Ordered Approximations of Atomic Properties

Jack Thomas, Huajie Chen & Christoph Ortner

Communicated by C. Le Bris

Abstract

We show that the local density of states (LDOS) of a wide class of tight-binding models has a weak body-order expansion. Specifically, we prove that the resulting body-order expansion for analytic observables such as the electron density or the energy has an exponential rate of convergence both at finite Fermi-temperature as well as for insulators at zero Fermi-temperature. We discuss potential consequences of this observation for modelling the potential energy landscape, as well as for solving the electronic structure problem.

1. Introduction

An atomistic *potential energy landscape* (PEL) is a mapping assigning energies $E(\boldsymbol{r})$, or local energy contributions, to atomic structures $\boldsymbol{r} = \{\boldsymbol{r}_\ell\}_{\ell\in\Lambda} \in (\mathbb{R}^d)^\Lambda$, where Λ is a general (possibly infinite) index set. High-fidelity models are provided by the Born–Oppenheimer PEL associated with *ab initio* electronic structure models such as tight-binding, Kohn–Sham density functional theory (DFT), Hartree–Fock, or even lower level quantum chemistry models [38,48,54,58,73,94]. Even now, however, the high computational cost of electronic structure models severely limits their applicability in material modelling to thousands of atoms for static and hundreds of atoms for long-time dynamic simulations.

There is a long and successful history of using surrogate models for the simulation of materials, devised to remain computationally tractable but capture as much detail of the reference *ab initio* PEL as possible. Empirical interatomic potentials are purely phenomenological and are able to capture a minimal subset of desired properties of the PEL, severely limiting their transferability [23,86]. The rapid growth in computational resources, increased both the desire and the possibility to match as much of an *ab inito* PEL as possible. A continuous increase in the complexity of parameterisations since the 1990s [6,7,36] has over time naturally led to

a new generation of "machine-learned interatomic potentials" employing universal approximators instead of empirical mechanistic models. Early examples include symmetric polynomials [11,80], artificial neural networks [8] and kernel methods [5]. A striking case is the Gaussian approximation potential for Silicon [4], capturing the vast majority of the PEL of Silicon of interest for material applications.

The purpose of the present work is, first, to rigorously evaluate some of the implicit or explicit assumptions underlying this latest class of interatomic potential models, as well as more general models for atomic properties. Specifically, we will identify natural modelling parameters as *approximation parameters* and rigorously establish convergence. Secondly, our results indicate that nonlinearities are an important feature, highlighting some superior theoretical properties. Finally, unlike existing nonlinear models, we will identify explicit low-dimensional nonlinear parameterisations yet prove that they are systematic. In addition to justifying and supporting the development of new models for general atomic properties, our results establish generic properties of *ab initio* models that have broader consequences, e.g. for the study of the mechanical properties of atomistic materials [15,17,32,93]. The application of our results to the construction and analysis of practical parameterisations (approximation schemes) that exploit our results will be pursued elsewhere.

Our overarching principle is to search for representations of properties of *ab initio* models in terms of *simple components*, where "simple" is of course highly context-specific. To illustrate this point, let us focus on modelling the potential energy landscape (PEL), which motivated this work in the first place. Pragmatically, we require that these *simple components* are easier to analyse and manipulate analytically or to fit than the PEL. For many materials (at least as long as Coulomb interaction does not play a role), the first step is to decompose the PEL into site energy contributions,

$$E(\boldsymbol{r}) = \sum_{\ell \in \Lambda} E_\ell(\boldsymbol{r}), \tag{1.1}$$

where one assumes that each E_ℓ is *local*, i.e., it depends only weakly on atoms far away. In previous works we have made this rigorous for the case of tight-binding models of varying complexity [14,16,17,93]. In practise, one may therefore truncate the interaction by admitting only those atoms $\boldsymbol{r}_k$ with $r_{\ell k} := |\boldsymbol{r}_k - \boldsymbol{r}_\ell| < r_{\rm cut}$ as arguments. Typical cutoff radii range from 5Å to 8Å, which means that on the order 30 to 100 atoms still make important contributions. Thus the site energy E_ℓ is still an extremely high-dimensional object and short of identifying low-dimensional features it would be practically impossible to numerically approximate it, due to the curse of dimensionality.

A classical example that illustrates our search for such low-dimensional features is the embedded atom model (EAM) [23], which assigns to each atom $\ell \in \Lambda$ a site energy

$$E_\ell^{\rm eam}(\boldsymbol{r}) = \sum_{k \neq \ell} \phi(r_{\ell k}) + F\big(\textstyle\sum_{k \neq \ell} \rho(r_{\ell k})\big).$$

While the site energy $E_\ell^{\rm eam}$ remains *high-dimensional*, the representation is in terms of three *one-dimensional* functions ϕ, ρ, F which are easily represented for example in terms of splines with relatively few parameters. Such a low-dimensional representation significantly simplifies parameter estimation, and vastly improves generalisation of the model outside a training set. Unfortunately, the EAM model and its immediately generalisations [6] have limited ability to capture a complex *ab initio* PEL. Still, this example inspires our search for representations of the PEL involving parameters that are

- low-dimensional,
- short-ranged.

Following our work on locality of interaction [14,16,17,93] we will focus on a class of tight-binding models as the *ab initio* reference model. These can be seen either as discrete approximations to density functional theory [38] or alternatively as electronic structure toy models sharing many similarities with the more complex Kohn–Sham DFT and Hartree–Fock models.

To control the dimensionality of representations, a natural idea is to to consider a body-order expansion,

$$E_\ell(\boldsymbol{r}) \approx V_0 + \sum_{k\neq\ell} V_1(\boldsymbol{r}_{\ell k}) + \sum_{\substack{k_1,k_2\neq\ell \\ k_1<k_2}} V_2(\boldsymbol{r}_{\ell k_1}, \boldsymbol{r}_{\ell k_2}) + \cdots + \sum_{\substack{k_1,\ldots,k_N\neq\ell \\ k_1<\cdots<k_N}} V_N\big(\boldsymbol{r}_{\ell k_1}, \ldots, \boldsymbol{r}_{\ell k_N}\big), \tag{1.2}$$

where $\boldsymbol{r}_{\ell k} := \boldsymbol{r}_k - \boldsymbol{r}_\ell$ and we say that $V_n(\boldsymbol{r}_{\ell k_1}, \ldots, \boldsymbol{r}_{\ell k_n})$ is an $(n+1)$-*body potential* modelling the interaction of a centre atom ℓ and n neighbouring atoms $\{k_1, \ldots, k_n\}$. This expansion was traditionally truncated at body-order *three* ($N = 2$) due to the exponential increase in computational cost with N. However, it was recently demonstrated by Shapeev's moment tensor potentials (MTPs) [80] and Drautz' atomic cluster expansion (ACE) [25] that a careful reformulation leads to models with at most linear N-dependence. Indeed, algorithms proposed in [2,80] suggest that the computational cost may even be N-independent, but this has not been proven. Even more striking is the fact that the MTP and ACE models which are both *linear models* based on a body-ordered approximation, currently appear to outperform the most advanced nonlinear models in regression and generalisation tests [66,106].

These recent successes are in stark contrast with the "folklore" that body-order expansions generally converge slowly, if at all [10,25,27,46,86]. The fallacy in those observations is typically that they implicitly assume a vacuum cluster expansion (cf. § 2.2). Indeed, our first set of main results in § 2.4 will be to demonstrate that a rapidly convergent body-order *approximation* can be constructed if one accounts for the chemical environment of the material. We will precisely characterise the convergence of such an approximation as $N \to \infty$, in terms of the Fermi-temperature and the band-gap of the material.

In the simplest scheme we consider, we achieve this by considering atomic properties $[O(\mathcal{H})]_{\ell\ell}$, where $\mathcal{H}$ is a tight-binding Hamiltonian and O an analytic

function. Approximating O by a polynomial on the spectrum $\sigma(\mathcal{H})$ results in an approximation of the atomic property $[p(\mathcal{H})]_{\ell\ell}$, which is naturally "body-ordered". To obtain quasi-optimal approximation results, naive polynomial approximation schemes (e.g. Chebyshev) are suitable only in the simplest scenarios. For the insulating case we leverage potential theory techniques which in particular yield quasi-optimal approximation rates on unions of disconnected domains. Our main results are obtained by converting these into approximation results on atomic properties, analysing their qualitative features, and taking care to obtain sharp estimates in the zero-Fermi-temperature limit.

These initial results provide strong evidence for the accuracy of a linear body-order approximation in relatively simple scenarios, and would for example be useful in a study of the mechanical response of single crystals with a limited selection of possible defects. However, they come with limitations that we discuss in the main text. In response, we consider a much more general framework, generalizing the theory of bond order potentials [55], that incorporates our linear body-ordered model as well as a range of nonlinear models. We will highlight a specific nonlinear construction with significantly improved theoretical properties over the linear scheme.

For both the linear and nonlinear body-ordered approximation schemes we prove that they inherit regularity, symmetries and locality of the original quantity of interest.

Finally, we consider the case of self-consistent tight-binding models such as DFTB [33,59,78]. In this case the highly nonlinear charge-equilibration leads *in principle* to arbitrarily complex intermixing of the nuclei information, and thus arbitrarily high body-order. However, our results on the body-ordered approximations for linear tight-binding models mean that each iteration of the self-consistent field (SCF) iteration can be expressed in terms of a low body-ordered and local interaction scheme. This leads us to propose a self-similar compositional representation of atomic properties that is highly reminiscent of recurrant neural network architectures. Each "layer" of this representation remains "simple" in the sense that we specified above.

2. Results

2.1. Preliminaries

2.1.1. Tight binding model We suppose Λ is a finite or countable index set. For $\ell \in \Lambda$, we denote the *state* of atom ℓ by $\boldsymbol{u}_\ell = (\boldsymbol{r}_\ell, v_\ell, Z_\ell)$ where $\boldsymbol{r}_\ell \in \mathbb{R}^d$ denotes the position, v_ℓ the effective potential, and Z_ℓ the atomic species of ℓ. Moreover, we define $\boldsymbol{r}_{\ell k} := \boldsymbol{r}_k - \boldsymbol{r}_\ell$, $r_{\ell k} := |\boldsymbol{r}_{\ell k}|$, and $\boldsymbol{u}_{\ell k} := (\boldsymbol{r}_{\ell k}, v_\ell, v_k, Z_\ell, Z_k)$. For functions f of the relative atomic positions $\boldsymbol{u}_{\ell k}$, the gradient denotes the gradient with respect to the spatial variable: $\nabla f(\boldsymbol{u}_{\ell k}) := \nabla\big(\xi \mapsto f((\xi, v_\ell, v_k, Z_\ell, Z_k))\big)\big|_{\xi=\boldsymbol{r}_{\ell k}}$. The whole configuration is denoted by $\boldsymbol{u} = (\boldsymbol{r}, v, Z) = (\{\boldsymbol{r}_\ell\}_{\ell\in\Lambda}, \{v_\ell\}_{\ell\in\Lambda}, \{Z_\ell\}_{\ell\in\Lambda})$.

For a given configuration $\boldsymbol{u}$, the tight binding Hamiltonian takes the following form:

(TB) *For $\ell, k \in \Lambda$ and N_b atomic orbitals per atom, we suppose that*

$$\mathcal{H}(\boldsymbol{u})_{\ell k} = h(\boldsymbol{u}_{\ell k}) + \sum_{m \notin \{\ell, k\}} t(\boldsymbol{u}_{\ell m}, \boldsymbol{u}_{k m}) + \delta_{\ell k} v_\ell \mathrm{Id}_{N_\mathrm{b}}, \tag{2.1}$$

where h and t have values in $\mathbb{R}^{N_\mathrm{b} \times N_\mathrm{b}}$, are independent of the effective potential v, and are continuously differentiable with

$$|h(\boldsymbol{u}_{\ell k})| + |\nabla h(\boldsymbol{u}_{\ell k})| \leqslant h_0 e^{-\gamma_0 r_{\ell k}}, \qquad \text{and} \tag{2.2}$$

$$|t(\boldsymbol{u}_{\ell m}, \boldsymbol{u}_{k m})| + |\nabla t(\boldsymbol{u}_{\ell m}, \boldsymbol{u}_{k m})| \leqslant h_0 e^{-\gamma_0 (r_{\ell m} + r_{k m})}, \tag{2.3}$$

for some $h_0, \gamma_0 > 0$.

Moreover, we suppose the Hamiltonian satisfies the following symmetries:

- *$h(\boldsymbol{u}_{\ell k}) = h(\boldsymbol{u}_{k \ell})^\mathrm{T}$ and $t(\boldsymbol{u}_{\ell m}, \boldsymbol{u}_{k m}) = t(\boldsymbol{u}_{k m}, \boldsymbol{u}_{\ell m})^\mathrm{T}$ for all $\ell, k, m \in \Lambda$,*
- *For orthogonal transformations $Q \in \mathbb{R}^{d \times d}$, there exist orthogonal $D^\ell(Q) \in \mathbb{R}^{N_\mathrm{b} \times N_\mathrm{b}}$ such that $\mathcal{H}(Q\boldsymbol{u}) = D(Q)\mathcal{H}(\boldsymbol{u})D(Q)^\mathrm{T}$ where $D(Q) = \mathrm{diag}(\{D^\ell(Q)\}_{\ell \in \Lambda})$ and $Q\boldsymbol{u} := (\{Q\boldsymbol{r}_\ell\}_{\ell \in \Lambda}, v, Z)$.*

Remark 1. (i) The constants in (2.2)-(2.3) are independent of the atomic sites $\ell, k, m \in \Lambda$.

(ii) Pointwise bounds on $|h(\boldsymbol{u}_{\ell k})|$ and $|t(\boldsymbol{u}_{\ell m}, \boldsymbol{u}_{k m})|$ are normally automatically satisfied since most linear tight binding models impose finite cut-off radii. Moreover, the assumption on the derivatives $|\nabla h(\boldsymbol{u}_{\ell k})|$ and $|\nabla t(\boldsymbol{u}_{\ell m}, \boldsymbol{u}_{k m})|$ states that there are no long range interactions in the model. In particular, we are assuming that Coulomb interactions have been screened, a typical assumption in many practical tight binding codes [20,68,71].

(iii) The Hamiltonian is symmetric and thus the spectrum is real.

(iv) The operators $\mathcal{H}(\boldsymbol{u})$ and $\mathcal{H}(Q\boldsymbol{u})$ are similar, and thus have the same spectra.

(v) The symmetry assumptions [84] of **(TB)** are justified in [16, Appendix A].

(vi) The entries of $\mathcal{H}(\boldsymbol{u})_{\ell k} \in \mathbb{R}^{N_\mathrm{b} \times N_\mathrm{b}}$ will be denoted $\mathcal{H}(\boldsymbol{u})^{ab}_{\ell k}$ for $1 \leqslant a, b \leqslant N_\mathrm{b}$. When clear from the context, we drop the argument $(\boldsymbol{u})$ in the notation.

The assumptions **(TB)** define a general *three-centre* tight binding model, whereas, if $t \equiv 0$, a simplification made in the majority of tight binding codes, we say **(TB)** is a *two-centre* model [38].

The choice of potential in **(TB)** defines a hierarchy of tight binding models. If $v = \mathrm{const}$, **(TB)** defines a linear tight binding model, a simple yet common model [14,16,17,70]. In this case, we implicitly assume that the Coulomb interactions have been screened, a typical assumption made in practice for a wide variety of materials [20,68,71,72]. Supposing that v is a function of a self-consistent electronic density, we arrive at a non-linear model such as DFTB [33,59,78]. Abstract variants of these nonlinear models have been analysed, for example, in [93,99]. Through much of this article we will treat $\boldsymbol{r}$, v as independent inputs into the Hamiltonian, but will discuss their connection and self-consistency in § 2.7.

For a finite system $\boldsymbol{u}$ (that is, with Λ a finite set), we consider *analytic observables* of the density of states [14,93]: for functions $O\colon \mathbb{R} \to \mathbb{R}$ that can be analytically continued into an open neighbourhood of $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$, we consider that

$$\operatorname{Tr} O\big(\mathcal{H}(\boldsymbol{u})\big) = \sum_s O(\lambda_s),$$

where (λ_s, ψ_s) are normalised eigenpairs of $\mathcal{H}(\boldsymbol{u})$. Many properties of the system, including the particle number functional and Helmholtz free energy, may be written in this form [14,16,70,93]. By distributing these quantities amongst atomic positions, we obtain a well-known spatial decomposition [14,16,35,38],

$$\operatorname{Tr} O\big(\mathcal{H}(\boldsymbol{u})\big) = \sum_{\ell\in\Lambda} O_\ell(\boldsymbol{u}) \quad \text{where} \quad O_\ell(\boldsymbol{u}) := \operatorname{tr}\big[O\big(\mathcal{H}(\boldsymbol{u})\big)_{\ell\ell}\big] = \sum_s O(\lambda_s)\big|[\psi_s]_\ell\big|^2. \tag{2.4}$$

For infinite systems, we may define $O_\ell(\boldsymbol{u})$ through the thermodynamic limit [14,16] or via the holomorphic functional calculus; see § 4.1.2 for further details.

When discussing derivatives of the local observables, we will simplify notation and write

$$\frac{\partial O_\ell(\boldsymbol{u})}{\partial \boldsymbol{u}_m} := \left(\frac{\partial O_\ell(\boldsymbol{u})}{\partial \boldsymbol{r}_m}, \frac{\partial O_\ell(\boldsymbol{u})}{\partial v_m}\right). \tag{2.5}$$

2.1.2. Local observables Although the results in this paper apply to general analytic observables, our primary interest is in applying them to two special cases. A local observable of particular importance is the electron density; for inverse Fermi-temperature $\beta \in (0, +\infty]$ and fixed chemical potential μ, we use the notation of (2.4) to define

$$\rho_\ell = F^\beta_\ell(\boldsymbol{u}) \quad \text{where} \quad F^\beta(z) := \begin{cases} \big(1 + e^{\beta(z-\mu)}\big)^{-1} & \text{if } \beta < \infty \\ \chi_{(-\infty,\mu)}(z) + \frac{1}{2}\chi_{\{\mu\}}(z) & \text{if } \beta = \infty. \end{cases} \tag{2.6}$$

Throughout this paper $F^\beta(\boldsymbol{u}) := \big(F^\beta_\ell(\boldsymbol{u})\big)_{\ell\in\Lambda}$ will denote a vector and so (2.6) reads $\rho = F^\beta(\boldsymbol{u})$.

In § 2.7, we consider the case where the effective potential is a function of the electron density (2.6) (that is, $v = w(\rho)$ for some $w\colon \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda$) which leads to the self-consistent local observables

$$\begin{cases} O^{\mathrm{sc}}_\ell(\boldsymbol{u}) := O_\ell\big(\boldsymbol{u}(\rho^\star)\big) \\ \rho^\star = F^\beta\big(\boldsymbol{u}(\rho^\star)\big) \end{cases}, \tag{2.7}$$

where $\boldsymbol{u}(\rho) := \big(\boldsymbol{r}, w(\rho), Z\big)$.

Remark 2. All the results of this paper also hold for the off-diagonal entries of the density matrix ($\rho_{\ell k} := \operatorname{tr} F^\beta\big(\mathcal{H}(\boldsymbol{u})\big)_{\ell k}$) without any additional work. This fact will be clear from the proofs. It is likely though that additional properties related to the off-diagonal decay (near-sightedness) and spatial regularity further improve the "sparsity" of the density matrix. A complete analysis would go beyond the scope of this work.

The second observable we are particularly interested in is the site energy, which allows us to decompose the total potential energy landscape into localised "atomic" contributions. In the grand potential model for the electrons, which is appropriate for large or infinite condensed phase systems [14], it is defined as

$$G^\beta_\ell(\boldsymbol{u}) := \mathrm{tr}\big[G^\beta\big(\mathcal{H}(\boldsymbol{u})\big)_{\ell\ell}\big] \quad \text{where} \quad G^\beta(z) := \begin{cases} \frac{2}{\beta}\log\big(1 - F^\beta(z)\big) & \text{if } \beta < \infty \\ 2(z-\mu)\chi_{(-\infty,\mu)}(z) & \text{if } \beta = \infty. \end{cases} \tag{2.8}$$

The total grand potential is defined as $\sum_\ell G^\beta_\ell(\boldsymbol{u})$ [14,70].

For $\beta < \infty$, the functions $F^\beta(\,\cdot\,)$ and $G^\beta(\,\cdot\,)$ are analytic in a strip of width $\pi\beta^{-1}$ about the real axis [17, Lemma 5.1]. To define the zero Fermi-temperature observables, we assume that μ lies in a spectral gap ($\mu \notin \sigma\big(\mathcal{H}(\boldsymbol{u})\big)$; see § 2.1.3). In this case, $F^\beta(\,\cdot\,)$ and $G^\beta(\,\cdot\,)$ extend to analytic functions in a neighbourhood of $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ for all $\beta \in (0, \infty]$.

In order to describe the relationship between the various constants in our estimates and the inverse Fermi-temperature or spectral gap (in the case of insulators), we will state all of our results for $O^\beta = F^\beta$ or G^β. Other analytic quantities of interest can be treated similarly with constants depending, e.g., on the region of analyticity of the corresponding function $z \mapsto O(z)$.

2.1.3. Metals, insulators, and defects As we can see from (2.4), the structure of the spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ will have a key role in the analysis. Firstly, by (**TB**), $\mathcal{H}(\boldsymbol{u})$ is a bounded self-adjoint operator on $\ell^2(\Lambda \times \{1, \ldots, N_\mathrm{b}\})$ and thus the spectrum is real and contained in some bounded interval. In order to keep the mathematical results general, we will not impose any further restrictions on the spectrum. However, to illustrate the main ideas, we briefly describe typical spectra seen in metals and insulating systems.

In the case where $\boldsymbol{u}$ describes a multi-lattice in $\mathbb{R}^d$ formed by taking the union of finitely many shifted Bravais lattices, the spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ is the union of finitely many continuous energy bands [57]. That is, there exist continuous functions, $\varepsilon^\alpha : \overline{\mathrm{BZ}} \to \mathbb{R}$, on the *Brillouin zone* BZ, a compact connected subset of $\mathbb{R}^d$, such that

$$\sigma\big(\mathcal{H}(\boldsymbol{u})\big) = \bigcup_\alpha \varepsilon^\alpha(\mathrm{BZ}).$$

In particular, in this case, $\sigma\big(\mathcal{H}(\boldsymbol{u})\big) = \sigma_{\mathrm{ess}}\big(\mathcal{H}(\boldsymbol{u})\big)$ is the union of finitely many intervals on the real line. The *band structure* $\{\varepsilon^\alpha\}$ relative to the position of the chemical potential, μ, determines the electronic properties of the system [89]. In metals μ lies within a band, whereas for insulators, μ lies between two bands in a *spectral gap*. Schematic plots of these two situations are given in Figure 1.

We now consider perturbations of a reference configuration $\boldsymbol{u}^{\mathrm{ref}} = (\boldsymbol{r}^{\mathrm{ref}}, v^{\mathrm{ref}}, Z^{\mathrm{ref}})$ defined on an index set Λ^{ref}.

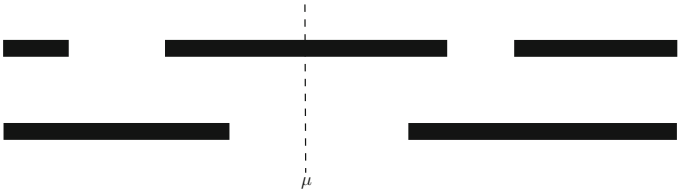


Fig. 1. Schematic plots of the spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ of a metal (top) and insulator (bottom)

Proposition 2.1. (Perturbation of the Spectrum) *For δ, $R_{\rm def} > 0$, there exists $\delta_0 > 0$ such that if $\boldsymbol{u} = (\boldsymbol{r}, v, Z)$ is a configuration defined on some index set Λ satisfying $\Lambda \setminus B_{R_{\rm def}} = \Lambda^{\rm ref} \setminus B_{R_{\rm def}}$, $\Lambda \cap B_{R_{\rm def}}$ is finite, $Z_k = Z_k^{\rm ref}$ for all $k \in \Lambda \setminus B_{R_{\rm ref}}$, and $\sup_{k\in\Lambda\setminus B_{R_{\rm def}}} \big[|\boldsymbol{r}_k - \boldsymbol{r}_k^{\rm ref}| + |v_k - v_k^{\rm ref}|\big] \leqslant \delta_0$, then*

$$\Big|\sigma\big(\mathcal{H}(\boldsymbol{u})\big) \setminus B_\delta\big(\sigma\big(\mathcal{H}(\boldsymbol{u}^{\rm ref})\big)\big)\Big| < \infty.$$

In particular, if $\boldsymbol{u}^{\rm ref}$ describes a multilattice, then, since local perturbations in the *defect core* are of finite rank, the essential spectrum is unchanged and we obtain finitely many eigenvalues bounded away from the spectral bands. Moreover, a small global perturbation can only result in a small change in the spectrum. Again, a schematic plot of this situation is given in Figure 2.

For the remainder of this paper, we consider the following notation:

Definition 1. Suppose that $\boldsymbol{u}^{\rm ref}$ is a general reference configuration defined on $\Lambda^{\rm ref}$ and $\boldsymbol{u}$ is a configuration arising due to Proposition 2.1. Then, we define I_- and I_+ to be compact intervals and $\{\lambda_j\}$ to be a finite set such that

$$\sigma\big(\mathcal{H}(\boldsymbol{u}^{\rm ref})\big) \subset I_- \cup I_+, \qquad \sigma\big(\mathcal{H}(\boldsymbol{u})\big) \subset I_- \cup \{\lambda_j\} \cup I_+ \tag{2.9}$$

and $\max I_- \leqslant \mu \leqslant \min I_+$. Moreover, we define

$$\mathsf{g} := \min I_+ - \max I_- \geqslant 0, \qquad \text{and} \tag{2.10}$$

$$\mathsf{g}^{\rm def} := \min I_+ \cup \{\lambda_j : \lambda_j \geqslant \mu\} - \max I_- \cup \{\lambda_j : \lambda_j \leqslant \mu\}. \tag{2.11}$$

The constants in Definition 1 are also displayed in Figure 2. The constant g in Definition 1 is slightly arbitrary in the sense that as long as $B_\delta\big(\sigma\big(\mathcal{H}(\boldsymbol{u}^{\rm ref})\big)\big) \subset I_- \cup I_+$ (where δ is the constant from Proposition 2.1), then there exists a finite set $\{\lambda_j\}$ as in (2.9). Choosing smaller g reduces the size of the set $\{\lambda_j\}$.

2.2. Vacuum cluster expansion

For a system of M identical particles $X_1, \dots, X_M$, a maximal body-order N, and a permutation invariant energy $E = E(\{X_1, \dots, X_M\})$, we may consider the vacuum cluster expansion,

$$E(\{X_1, \dots, X_M\}) \approx \sum_{n=0}^{N} \sum_{1\leqslant m_1<\cdots<m_n\leqslant M} V^{(n)}(X_{m_1}, \dots, X_{m_n}), \tag{2.12}$$

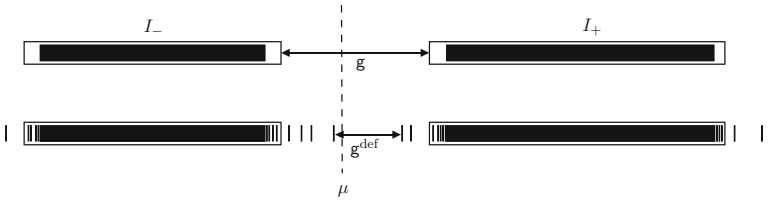


Fig. 2. Top: Schematic plot of the spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u}^{\text{ref}})\big)$ for an insulating system, together with two compact intervals I_- and I_+ as in (2.9) and the constant g from (2.10). Bottom: The spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ after considering perturbations satisfying Proposition 2.1. While the edges of the spectrum may be accumulation points for a sequence of eigenvalues within the band gap, the number of such eigenvalues bounded away from the edges is finite

where the n-body interaction potentials $V^{(n)}$ are defined by considering all isolated clusters of $j \leqslant n$ atoms:

$$V^{(n)}(X_1, \ldots, X_n) = \sum_{j=0}^{n} (-1)^{n-j} \sum_{1 \leqslant m_1 < \cdots < m_j \leqslant n} E(\{X_{m_1}, \ldots, X_{m_j}\}).$$

The expansion (2.12) is exact for $N = M$. The vacuum cluster expansion is the traditional and, arguably, the most natural many-body expansion of a potential energy landscape. However, in many systems, it converges extremely slowly with respect to the body-order N and is thus computationally impractical. An intuitive explanation for this slow convergence is that, when defining the body-order expansion in this way, we are building an interaction law for a condensed or possibly even crystalline phase material from clusters in vacuum where the bonding chemistry is significantly different. Although this observation appears to be "common knowledge" we were unable to find references that provide clear evidence for it. However, some limited discussions and further references can be found in [10,25,27,46,86].

Our own approach employs an entirely different mechanism, which in particular incorporates environment information and leads to an exponential convergence of an N-body approximation. Technically, our approximation is not an expansion, that is, the n-body terms $V^{(n)}$ of the classical cluster expansion are replaced by terms that depend also on the highest body-order N. We will provide a more technical discussion contrasting our results with the vacuum cluster expansion in § 2.6.

2.3. A general framework

Before we consider two specific body-ordered approximations, we present a general framework which both incorporates many (linear-scaling) electronic structure methods from the literature (e.g. the kernel polynomial method (KPM) [82], bond-order potentials (BOP) [26,39,55,74], and quadrature-based methods [69,87,88]), and illustrates the key features needed for a convergent scheme: To that end, we introduce the *local density of states* (LDOS) [38] which is the (positive) measure D_ℓ supported on $\sigma(\mathcal{H})$ such that

$$\int x^n \mathrm{d}D_\ell(x) = \mathrm{tr}[\mathcal{H}^n]_{\ell\ell}, \qquad \text{for } n \in \mathbb{N}_0. \tag{2.13}$$

Existence and uniqueness follows from the spectral theorem for normal operators (e.g. see [1, Theorem 6.3.3] or [92]). In particular, (2.4) may be written as the integral $O_\ell(\boldsymbol{u}) = \int O \,\mathrm{d}D_\ell$.

Then, on constructing a (possibly signed) unit measure D_ℓ^N with exact first N moments (that is, $\int x^n \mathrm{d}D_\ell^N(x) = \mathrm{tr}[\mathcal{H}^n]_{\ell\ell}$ for $n = 1, \dots, N$), we may define the approximate local observable $O_\ell^N(\boldsymbol{u}) := \int O \,\mathrm{d}D_\ell^N$, and obtain the general error estimates

$$\begin{aligned} \left|O_\ell(\boldsymbol{u}) - O_\ell^N(\boldsymbol{u})\right| &= \inf_{P_N \in \mathcal{P}_N} \left| \int \big(O - P_N\big) \mathrm{d}\big(D_\ell - D_\ell^N\big) \right| \\ &\leqslant \left\| D_\ell - D_\ell^N \right\|_{\mathrm{op}} \inf_{P_N \in \mathcal{P}_N} \left\| O - P_N \right\|_\infty, \end{aligned} \tag{2.14}$$

where $\mathcal{P}_N$ denotes the set of polynomials of degree at most N, and $\| \cdot \|_{\mathrm{op}}$ is the operator norm on a function space $(\mathcal{S}, \| \cdot \|_\infty)$. For example, we may take $\mathcal{S}$ to be the set of functions analytic on an open set containing $\mathscr{C}$, a contour encircling $\mathrm{supp}\big(D_\ell - D_\ell^N\big)$, and consider

$$\|O\|_\infty := \frac{\mathrm{len}(\mathscr{C})}{2\pi} \|O\|_{L^\infty(\mathscr{C})}.$$

Alternatively, we may consider $\mathcal{S} = L^\infty\big(\mathrm{supp}(D_\ell - D_\ell^N)\big)$ leading to the total variation operator norm.

Equation (2.14) highlights the key generic features that are crucial ingredients in obtaining convergence results:

- *Analyticity.* The potential theory results of § 4.1.5 connect the asymptotic convergence rates for polynomial approximation to the size and shape of the region of analyticity of O.
- *Spectral Pollution.* While $\mathrm{supp} D_\ell \subset \sigma(\mathcal{H})$, this need not be true for D_ℓ^N. Indeed, if $\mathrm{supp} D_\ell^N$ introduces additional points within the band gap, this may significantly slow the convergence of the polynomial approximation; cf. § 2.6.
- *Regularity of* D_ℓ^N. Roughly speaking, the first term of (2.14) measures how "well-behaved" D_ℓ^N is. In particular, if D_ℓ^N is positive, then this term is bounded independently of N, whereas, if D_ℓ^N is a general signed measure, then this factor contributes to the asymptotic convergence behaviour.

In the sections to follow, we introduce linear (§ 2.4) and nonlinear (§ 2.5) approximation schemes that fit into this general framework. Moreover, in § 2.6, we also write the vacuum cluster expansion as an integral against an approximate LDOS. In order to complement the intuitive explanation for the slow convergence of the vacuum cluster expansion, we investigate which of the requirements listed above fail.

In the appendices, we review other approximation schemes that fit into this general framework such as the quadrature method (Appendix D), numerical bond order potentials (Appendix E), and the kernel polynomial method (Appendix F).

2.4. Linear body-ordered approximation

We will construct two distinct but related many-body approximation models. To construct our first model we exploit the observation that polynomial approximations of an analytic function correspond to body-order expansions of an observable.

An intuitive approach is to write the local observable in terms of its Chebyshev expansion and truncate to some maximal polynomial degree. The corresponding projection operator is a simple example of the kernel polynomial method (KPM) [82] and the basis for analytic bond order potentials (BOP) [74]. We discuss in Appendix F that these schemes put more emphasis on the approximation of the local density of states (LDOS) and, in particular, exploit particular features of the Chebyshev polynomials to obtain a positive approximate LDOS. Since our focus is instead on the approximation of observables, we employ a different approach that is tailored to specific properties of the band structure and leads to superior convergence rates for these quantities.

For a set of $N+1$ interpolation points $X_N = \{x_j\}_{j=0}^N$, and a complex-valued function O defined on X_N, we denote by $I_{X_N} O$ the degree N polynomial interpolant of $x \mapsto O(x)$ on X_N. This gives rise to the body-ordered approximation

$$I_{X_N} O_\ell(\boldsymbol{u}) := \mathrm{tr}\big[I_{X_N} O\big(\mathcal{H}(\boldsymbol{u})\big)_{\ell\ell}\big]. \tag{2.15}$$

We may connect (2.15) to the general framework in § 2.3 by defining

$$I_{X_N} O_\ell(\boldsymbol{u}) = \int O \,\mathrm{d}D_\ell^{N,\mathrm{lin}} \qquad \text{where} \qquad D_\ell^{N,\mathrm{lin}} := \mathrm{tr} \sum_j \ell_j(\mathcal{H})_{\ell\ell}\, \delta(\cdot - x_j), \tag{2.16}$$

and ℓ_j are the node polynomials corresponding to $X_N = \{x_j\}_{j=0}^N$ (that is, ℓ_j are the polynomials of degree N with $\ell_j(x_i) = \delta_{ij}$).

Proposition 2.2. *$I_{X_N} O_\ell(\boldsymbol{u})$ has body-order at most $2N$. More specifically, there exists $(n+1)$-body potentials V_{nN} for $n = 0, \ldots, 2N-1$ such that*

$$I_{X_N} O_\ell(\boldsymbol{u}) = \sum_{n=0}^{2N-1} \sum_{\substack{k_1,\ldots,k_n \neq \ell \\ k_1 < \cdots < k_n}} V_{nN}(\boldsymbol{u}_\ell; \boldsymbol{u}_{\ell k_1}, \ldots, \boldsymbol{u}_{\ell k_n}). \tag{2.17}$$

Proof. (Sketch of the Proof.) Since (2.15) is a linear combination of the monomials $[\mathcal{H}^n]_{\ell\ell}$, it is enough to show that, for each $n \in \mathbb{N}$,

$$[\mathcal{H}^n]_{\ell\ell} = \sum_{\ell_1,\ldots,\ell_{n-1}} \mathcal{H}_{\ell\ell_1} \mathcal{H}_{\ell_1\ell_2} \cdots \mathcal{H}_{\ell_{n-1}\ell} \tag{2.18}$$

has finite body order.

Each term in (2.18) depends on the central atom ℓ, the $n-1$ neighbouring sites $\ell_1, \ldots, \ell_{n-1}$, and the at most n additional sites arising from the three-centre summation in the tight binding Hamiltonian (**TB**). In particular, (2.15) has body order at most $2N$. See § 4.2 for a complete proof including an explicit definition of the V_{nN}. □

If one uses Chebyshev points as the basis for the body-ordered approximation (2.15), the rates of convergence depend on the size of the largest *Bernstein ellipse* (that is, ellipses with foci points ± 1) contained in the region of analyticity of $z \mapsto O(z)$ [95]. This leads to a exponentially convergent body-order expansion in the metallic finite-temperature case (see § 4.1.4 for the details).

However, the resulting estimates deteriorate in the zero-temperature limit. Instead, we apply results of potential theory to construct interpolation sets X_N that are adapted to the spectral properties of the system (see § 4.1.5 for examples) and *(i)* do not suffer from spectral pollution, and *(ii)* (asymptotically) minimise the total variation of $D_\ell^{N,\mathrm{lin}}$ which, in this context, is the Lebesgue constant [95] for the interpolation operator I_{X_N}. This leads to rapid convergence of the body-order approximation based on (2.15). The interpolation sets X_N depend only on the intervals I_-, I_+ from Definition 1 (see also Figure 2) and can be chosen independently of $\boldsymbol{u}^{\mathrm{ref}}$ as long as $B_\delta\big(\sigma\big(\mathcal{H}(\boldsymbol{u}^{\mathrm{ref}})\big)\big) \subset I_- \cup I_+$.

Theorem 2.3. *Suppose $\boldsymbol{u}^{\mathrm{ref}}$ is given by Definition 1. Fix $0 < \beta \leqslant \infty$ and suppose that, either $\beta < \infty$ or $\mathsf{g} > 0$. Then, for all $N \in \mathbb{N}$, there exist constants $\gamma_N > 0$ and interpolation sets $X_N = \{x_j\}_{j=0}^N \subset I_- \cup I_+$ satisfying (2.17) such that*

$$\begin{aligned} \big|O_\ell^\beta(\boldsymbol{u}^{\mathrm{ref}}) - I_{X_N} O_\ell^\beta(\boldsymbol{u}^{\mathrm{ref}})\big| &\leqslant C_1 e^{-\gamma_N N}, \qquad \text{and} \\ \left|\frac{\partial O_\ell^\beta}{\partial \boldsymbol{u}_m}(\boldsymbol{u}^{\mathrm{ref}}) - \frac{\partial I_{X_N} O_\ell^\beta}{\partial \boldsymbol{u}_m}(\boldsymbol{u}^{\mathrm{ref}})\right| &\leqslant C_2 e^{-\frac{1}{2}\gamma_N N} e^{-\eta r_{\ell m}}, \end{aligned}$$

where $O^\beta = F^\beta$ or G^β and $C_1, C_2, \eta > 0$ are independent of N. The asymptotic convergence rate $\gamma := \lim_{N\to\infty} \gamma_N$ is positive and exhibits the asymptotic behaviour

$$\begin{aligned} &C_1 \sim (\mathsf{g} + \beta^{-1})^{-1}, \quad C_2 \sim (\mathsf{g} + \beta^{-1})^{-3}, \\ &\text{and} \quad \gamma, \eta \sim \mathsf{g} + \beta^{-1} \quad \text{as } \mathsf{g} + \beta^{-1} \to 0. \end{aligned} \tag{2.19}$$

In this asymptotic relation, we assume that the limit $\mathsf{g} \to 0$ is approached symmetrically about the chemical potential μ.

Remark 3. Higher derivatives may be treated similarly under the assumption that higher derivatives of the tight binding Hamiltonian (**TB**) exist and are short ranged.

2.4.1. The role of the point spectrum We now turn towards the important scenario when a localised defect is embedded within a homogeneous crystalline solid. Recall from § 2.1.3 (see in particular Fig. 2) that this gives rise to a discrete spectrum, which "pollutes" the band gap [70]. Thus, the spectral gap is reduced and a naive application of Theorem 2.3 leads to a reduction in the convergence rate of the body-ordered approximation. We now improve these estimates by showing that, away from the defect, we obtain improved pre-asymptotics, reminiscent of similar results for locality of interaction [17].

In that follow, we fix $\boldsymbol{u}$ satisfying Definition 1. While improved estimates may be obtained by choosing $\{\lambda_j\}$ as interpolation points, leading to asymptotic exponents

that are independent of the defect, in practice, this requires full knowledge of the point spectrum. Since the point spectrum within the spectral gap depends on the whole atomic configuration, the approximate quantities of interest corresponding to these interpolation operators would no longer satisfy Proposition 2.2.

Remark 4. This phenomenon has been observed in the context of Krylov subspace methods for solving linear equations $Ax = b$ where outlying eigenvalues delay the convergence by $O(1)$ steps without affecting the asymptotic rate [30]. Indeed, since the residual after n steps may be written as $r_n = p_n(A)r_0$ where p_n is a polynomial of degree n, there is a close link between polynomial approximation and convergence of Krylov methods.

On the other hand, we may use the exponential localisation of the eigenvectors corresponding to isolated eigenvalues to obtain pre-factors that decay exponentially as $|\boldsymbol{r}_\ell| \to \infty$.

Theorem 2.4. *Suppose* $\boldsymbol{u}$ *satisfies Definition* 1 *with* $\mathsf{g} > 0$. *Fix* $0 < \beta \leqslant \infty$ *and suppose that, if* $\beta = \infty$, *then* $\mathsf{g}^{\rm def} > 0$, *and let* $C_1, C_2, \gamma_N, \gamma, \eta$, *and* $X_N = \{x_j\}_{j=0}^N \subset I_- \cup I_+$ *be given by Theorem* 2.3. *Then, there exist* $\gamma_{\rm CT}, \gamma_N^{\rm def} > 0$ *such that*

$$\left| O_\ell^\beta(\boldsymbol{u}) - I_{X_N} O_\ell^\beta(\boldsymbol{u}) \right| \leqslant C_1 e^{-\gamma_N N} + C_3 e^{-\gamma_{\rm CT}|\boldsymbol{r}_\ell|} e^{-\frac{1}{2}\gamma_N^{\rm def} N} \tag{2.20}$$

$$\left| \frac{\partial O_\ell^\beta}{\partial \boldsymbol{u}_m}(\boldsymbol{u}) - \frac{\partial I_{X_N} O_\ell^\beta}{\partial \boldsymbol{u}_m}(\boldsymbol{u}) \right| \leqslant \left(C_2 e^{-\frac{1}{2}\gamma_N N} + C_4 e^{-\gamma_{\rm CT}|\boldsymbol{r}_\ell|} e^{-\frac{1}{2}\gamma_N^{\rm def} N} \right) e^{-\eta\, r_{\ell m}} \tag{2.21}$$

where $O^\beta = F^\beta$ *or* G^β *and* $C_3, C_4 > 0$ *are independent of* N. *The asymptotic convergence rate* $\gamma^{\rm def} := \lim_{N\to\infty} \gamma_N^{\rm def}$ *is positive and we have*

$$\begin{aligned} &\gamma^{\rm def} \sim \mathsf{g}^{\rm def} + \beta^{-1} \quad \text{as } \mathsf{g}^{\rm def} + \beta^{-1} \to 0, \\ &\text{and} \quad \gamma_{\rm CT}, \eta \sim \mathsf{g} + \beta^{-1} \quad \text{as } \mathsf{g} + \beta^{-1} \to 0. \end{aligned} \tag{2.22}$$

In these asymptotic relations, we assume that the limits $\mathsf{g}^{\rm def}, \mathsf{g} \to 0$ *are approached symmetrically about the chemical potential* μ.

In practice, Theorem 2.4 means that, for atomic sites ℓ away from the defect-core, the observed pre-asymptotic error estimates may be significantly better than the asymptotic convergence rates obtained in Theorem 2.3.

Remark 5. (Locality) *(i)* By Theorem 2.4, and the locality estimates for the exact observables O_ℓ^β [17], we immediately obtain corresponding locality estimates for the approximate quantities:

$$\left| \frac{\partial I_{X_N} O_\ell^\beta(\boldsymbol{u})}{\partial \boldsymbol{u}_m} \right| \lesssim e^{-\eta\, r_{\ell m}}. \tag{2.23}$$

(ii) We investigate another type of locality in Appendix B where we show that various truncation operators result in approximation schemes that only depend on a small atomic neighbourhood of the central site. An exponential rate of convergence as the truncation radius tends to infinity is obtained.

Remark 6. (Connection to the general framework) The fact that the exponents in Theorem 2.4 depend on the discrete eigenvalues of $\mathcal{H}(\boldsymbol{u})$ can be seen from the general estimate (2.14) applied to the approximate LDOS $D_\ell^{N,\mathrm{lin}}$ from (2.16):

- *Spectral Pollution.* We choose the interpolation points so that the support of $D_\ell^{N,\mathrm{lin}}$ lies within $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ and so spectral pollution does not play a role,
- *Regularity of* $D_\ell^{N,\mathrm{lin}}$. The total variation of $D_\ell^{N,\mathrm{lin}}$ can be estimated by the Lebesgue constant [95] for the interpolation operator I_{X_N}:

$$\begin{aligned}\|D_\ell^{N,\mathrm{lin}}\|_{\mathrm{TV}} &:= \sup_{\|f\|_{L^\infty(\sigma(\mathcal{H}))}=1} |I_{X_N} f(\mathcal{H})_{\ell\ell}| \leqslant \sup_{\|f\|_{L^\infty(\sigma(\mathcal{H}))}=1} \sup_{x\in\sigma(\mathcal{H})} |I_{X_N} f(x)| \\ &= \sup_{x\in\sigma(\mathcal{H})} \sum_j |\ell_j(x)|.\end{aligned}$$

  This quantity depends on the discrete eigenvalues within the band gap.

2.5. A non-linear representation

The method presented in § 2.4 approximates local quantities of interest by approximating the integrand $O\colon \mathbb{C}\to\mathbb{C}$ with polynomials. As we have seen, this leads to approximation schemes that are linear functions of the spatial correlations $\{[\mathcal{H}^n]_{\ell\ell}\}_{n\in\mathbb{N}}$. In this section, we construct a non-linear approximation related to bond-order potentials (BOP) [26,39,55] and show that the added non-linearity leads to improved asymptotic error estimates that are independent of the discrete spectra lying within the band gap. In this way, the nonlinearity captures "spectral information" from $\mathcal{H}$ rather than only approximating $O\colon \mathbb{C}\to\mathbb{C}$ without reference to the Hamiltonian.

Applying the recursion method [49,50], a reformulation of the Lanczos process [61], we obtain a tri-diagonal (Jacobi) operator T on $\ell^2(\mathbb{N}_0)$ whose spectral measure is the LDOS D_ℓ [91] (see § 4.3.1 for the details). We then truncate T by taking the principal $\frac{1}{2}(N+1)\times\frac{1}{2}(N+1)$ submatrix $T_{\frac{1}{2}(N-1)}$ and define

$$\Theta_N\big(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \ldots, [\mathcal{H}^N]_{\ell\ell}\big) := O^\beta(T_{\frac{1}{2}(N-1)})_{00} = \int O^\beta \mathrm{d}D_\ell^{N,\mathrm{nonlin}}, \quad (2.24)$$

where $D_\ell^{N,\mathrm{nonlin}} = \sum_s [\psi_s]_0^2 \delta(\cdot-\lambda_s)$ is a spectral measure for $T_{\frac{1}{2}(N-1)}$ (that is, (λ_s,ψ_s) are normalised eigenpairs of $T_{\frac{1}{2}(N-1)}$). By showing that the first N moments of $D_\ell^{N,\mathrm{nonlin}}$ are exact, we are able to apply (2.14) to obtain the following error estimates. The asymptotic behaviour of the exponent in these estimates follows by proving that the spectral pollution of $D_\ell^{N,\mathrm{nonlin}}$ in the band gap is sufficiently mild.

Theorem 2.5. *Suppose* $\boldsymbol{u}$ *satisfies Definition 1. Fix* $0<\beta\leqslant\infty$ *and suppose that, if* $\beta=\infty$, *then* $\mathsf{g}>0$. *Then, for N odd, there exists an open set* $U\subset\mathbb{C}^N$ *such that* (2.24) *extends to an analytic function* $\Theta_N\colon U\to\mathbb{C}$, *such that*

$$\Big|O_\ell^\beta(\boldsymbol{u}) - \Theta_N\big(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \ldots, [\mathcal{H}^N]_{\ell\ell}\big)\Big| \lesssim e^{-\gamma_N N} \quad (2.25)$$

where $O^\beta = F^\beta$ *or* G^β. *The asymptotic convergence rate* $\gamma := \lim_{N\to\infty} \gamma_N$ *is positive and* $\gamma \sim \mathsf{g} + \beta^{-1}$ *as* $\mathsf{g} + \beta^{-1} \to 0$.

Remark 7. It is important to note that $\Theta_N : U \to \mathbb{C}$ can be constructed without knowledge of $\mathcal{H}$ because, as we have seen, if the discrete eigenvalues are known *a priori*, then Theorem 2.5 is immediate from Theorem 2.4 by adding finitely many additional interpolation points on the discrete spectrum.

In particular, the fact that Θ_N is a *material-agnostic* nonlinearity has potentially far-reaching consequences for material modelling.

Remark 8. (Connection to the general framework) The fact that the exponents in Theorem 2.5 are independent of the discrete eigenvalues of $\mathcal{H}(\boldsymbol{u})$ can be seen from the general estimate (2.14) applied to the approximate LDOS $D_\ell^{N,\mathrm{nonlin}}$ from (2.24):

- *Spectral Pollution.* We show that $\left|\operatorname{supp} D_\ell^{N,\mathrm{nonlin}} \setminus \operatorname{supp} D_\ell\right|$ remains bounded independently of N and so spectral pollution only slows the convergence by at most $O(1)$ steps,
- *Regularity of* $D_\ell^{N,\mathrm{nonlin}}$. Since $D_\ell^{N,\mathrm{nonlin}}$ is a positive unit measure, we have the bound $\|D_\ell - D_\ell^{N,\mathrm{nonlin}}\|_{\mathrm{TV}} \leqslant 2$.

Remark 9. (Quadrature Method) Alternatively, we may use the sequence of orthogonal polynomials [40] corresponding to D_ℓ as the basis for a Gauss quadrature rule to evaluate local observables. This procedure, called the *Quadrature Method* [51,69], is a precursor of the bond order potentials. Outlined in Appendix D, we show that it produces an alternative scheme also satisfying Theorem 2.5.

The linear-scaling spectral Gauss quadrature (LSSGQ) method [87] is based upon this idea, albeit in the context of finite difference approximations to the DFT Hamiltonian. However, since the resulting discrete Hamiltonian in [87] is banded, the analysis of the present work may be readily applied. Therefore, Theorem 2.5 provides rigorous justification for the exponential rate of convergence for increasing body-order (number of quadrature points), complementing the intuitive explanations and numerical experiments of [87].

Since the convergence results are independent of system size, we obtain a linear-scaling method, a result that complements the intuitive explanation [87, (56)], and numerical evidence [87, Fig. 5].

Remark 10. (Convergence of Derivatives) In this more complicated nonlinear setting, obtaining results such as (2.21) is more subtle. We require an additional assumption on D_ℓ, which we believe maybe be typically satisfied, but we currently cannot justify it and have therefore postponed this discussion to Appendix C. We briefly mention, however, that if D_ℓ is absolutely continuous (e.g., in periodic systems), we obtain

$$\left|\frac{\partial}{\partial \boldsymbol{u}_m}\Big(O_\ell^\beta(\boldsymbol{u}) - \Theta_N\big(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \ldots, [\mathcal{H}^N]_{\ell\ell}\big)\Big)\right| \lesssim e^{-\frac{1}{2}\gamma_N N} e^{-\eta\, r_{\ell m}}.$$

2.6. The vacuum cluster expansion revisited

For $\ell \in \Lambda$, we denote by $\mathcal{H}\big|_{\ell;K}$ the Hamiltonian matrix corresponding to the finite subsystem $\{\ell\} \cup K \subset \Lambda$: for $k_1, k_2 \in \{\ell\} \cup K$,

$$\big[\mathcal{H}\big|_{\ell;K}\big]_{k_1k_2} := h(\boldsymbol{u}_{k_1k_2}) + \sum_{m \in \{\ell\}\cup K} t(\boldsymbol{u}_{k_1m}, \boldsymbol{u}_{k_2m}) + \delta_{k_1k_2} v_{k_1} \mathrm{Id}_{N_\mathrm{b}}. \tag{2.26}$$

For an observable O, the vacuum cluster expansion as detailed in § 2.2 is constructed as follows:

$$O_\ell^{N,\mathrm{vac}}(\boldsymbol{u}) := \sum_{n=0}^{2N-1} \sum_{\substack{k_1,\dots,k_n \neq \ell \\ k_1<\dots<k_n}} V^{(n)}(\boldsymbol{u}_\ell; \boldsymbol{u}_{\ell k_1}, \dots, \boldsymbol{u}_{\ell k_n}) \qquad \text{where} \tag{2.27}$$

$$V^{(n)}(\boldsymbol{u}_\ell; \boldsymbol{u}_{\ell k_1}, \dots, \boldsymbol{u}_{\ell k_n}) = \sum_{K \subseteq \{k_1,\dots,k_n\}} (-1)^{n-|K|} O\big(\mathcal{H}\big|_{\ell;K}\big)_{\ell\ell}. \tag{2.28}$$

Therefore, on defining the *spectral measure* $D_{\ell;K} := \sum_s \delta\big(\,\cdot - \lambda_s(K)\big)|[\psi_s(K)]_\ell|^2$ where $\big(\lambda_s(K), \psi_s(K)\big)$ the are normalised eigenpairs of $\mathcal{H}\big|_{\ell;K}$, we may write the vacuum cluster expansion as in § 2.3:

$$O_\ell^{N,\mathrm{vac}}(\boldsymbol{u}) = \int O \,\mathrm{d}D_\ell^{N,\mathrm{vac}} \qquad \text{where}$$

$$D_\ell^{N,\mathrm{vac}} := \sum_{n=0}^{2N-1} \sum_{\substack{k_1,\dots,k_n \neq \ell \\ k_1<\dots<k_n}} \sum_{K \subseteq \{k_1,\dots,k_n\}} (-1)^{n-|K|} D_{\ell;K}. \tag{2.29}$$

While $D_\ell^{N,\mathrm{vac}}$ is a generalised signed measure (with values in $\mathbb{R} \cup \{\pm\infty\}$), all moments are finite. More specifically, if we absorb the effective potential and two centre terms into the three centre summation by writing $\mathcal{H}_{k_1k_2} = \sum_m \mathcal{H}_{k_1k_2m}$, see (4.16), we have

$$\int x^j \,\mathrm{d}D_\ell^{N,\mathrm{vac}}(x) = \sum_{\substack{\ell_1,\dots,\ell_{j-1},m_1,\dots,m_j \\ |\{\ell,\ell_1,\dots,\ell_{j-1},m_1,\dots,m_j\}|\leq 2N}} \mathcal{H}_{\ell\ell_1m_1}\mathcal{H}_{\ell_1\ell_2m_2}\dots\mathcal{H}_{\ell_{j-1}\ell m_j}. \tag{2.30}$$

Equation (2.30) follows from the proof of Proposition 2.2, see (4.19). In particular, the first N moments of $D_\ell^{N,\mathrm{vac}}$ are exact. Therefore, we may apply the general error estimate (2.14) and describe the various features of $D_\ell^{N,\mathrm{vac}}$ which provide mathematical intuition for the slow convergence of the vacuum cluster expansion:

- *Spectral Pollution.* When splitting the system up into arbitrary subsystems as is the case in the vacuum cluster expansion, one expects significant spectral pollution in the band gaps, leading to a reduction in the convergence rate,
- *Regularity of* $D_\ell^{N,\mathrm{vac}}$. The approximate LDOS is a linear combination of countably many Dirac deltas and does not have bounded variation. Moreover, $D_\ell^{N,\mathrm{vac}}$ has values in $\mathbb{R} \cup \{\pm\infty\}$.

2.7. Self-consistency

Throughout this section, we suppose that the effective potential is a function of a self-consistent electron density: that is, (2.6) becomes the following nonlinear equation:

$$\rho^\star = F^\beta\big(\boldsymbol{u}(\rho^\star)\big) \tag{2.31}$$

where $\boldsymbol{u}(\rho) := \big(\boldsymbol{r}, w(\rho), Z\big)$. We shall assume that the effective potential satisfies the following:

(EP) *We suppose that* $w\colon \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda$ *is twice continuously differentiable with*

$$\big|\nabla w(\rho)_{\ell k}\big| \leqslant C e^{-\gamma_v r_{\ell k}}$$

for some $\gamma_v > 0$.

Remark 11. (i) For a smooth function $\widetilde{w}\colon \mathbb{R} \to \mathbb{R}$, the effective potential $w(\rho)_\ell := \widetilde{w}(\rho_\ell)$ satisfies (**EP**). This leads to the simplest abstract nonlinear tight binding models discussed in [93,99].

(ii) The (short-ranged) Yukawa potential defined by $w(\rho)_\ell := \sum_{m\neq\ell} \frac{\rho_m - Z_m}{r_{\ell m}} e^{-\tau r_{\ell m}}$ (for some $\tau > 0$) also fits into this general framework. This setting already covers many important modelling scenarios and also serves as a crucial stepping stone towards charge equilibration under full Coulomb interaction, which goes beyond the scope of the present work.

The main result of this section is the following: if there exists a self-consistent solution $\rho^\star$ to (2.31), then we can approximate $\rho^\star$ with self-consistent solutions to the following approximate self-consistency equation:

$$\rho_N = I_{X_N} F^\beta\big(\boldsymbol{u}(\rho_N)\big), \tag{2.32}$$

for sufficiently large N. The operator $I_{X_N} F^\beta$ is a linear body-ordered approximation of the form we analyzed in detail in § 2.4.

To do this, we require a natural stability assumption on the electronic structure problem, which was employed for example in [93,99,100]:

(STAB) *The stability operator* $\mathscr{L}(\rho)$ *is the Jacobian of* $\rho \mapsto F^\beta\big(\boldsymbol{u}(\rho)\big)$. *We say electron densities* $\rho^\star$ *solving* (2.31) *are stable if* $I - \mathscr{L}(\rho^\star)$ *is invertible as a bounded linear operator* $\ell^2 \to \ell^2$.

Remark 12. (Stability) *(i)* The stability condition of Theorem 2.6 is a minimal starting assumption that naturally arises from the analysis [93,99,100].

For example, if ρ is a stable self-consistent electron density, then there exists $\phi^{(m)} \in \ell^2(\Lambda)$ such that [93]:

$$\frac{\partial \rho_\ell}{\partial \boldsymbol{u}_m} = \Big[\big(I - \mathscr{L}(\rho)\big)^{-1} \phi^{(m)}\Big]_\ell.$$

(ii) As noted in [99] (in a slightly simpler setting), the stability condition of Theorem 2.6 is automatically satisfied for multi-lattices with ∇w positive semi-definite. In fact, in this case the stability operator is negative semi-definite.

Theorem 2.6. *For $\boldsymbol{u}$ satisfying Definition 1, suppose that $\rho^\star$ is a corresponding stable self-consistent electron density.*

Then, for N sufficiently large, there exist self-consistent solutions ρ_N of (2.32) such that

$$\left\|\rho_N - \rho^\star\right\|_{\ell^\infty} \leqslant Ce^{-\gamma_N N}, \tag{2.33}$$

where γ_N are the constants from Theorem 2.3 applied to $\boldsymbol{u}(\rho^\star)$.

Corollary 2.7. *Suppose that $\rho^\star$ and ρ_N are as in Theorem 2.6 and denote by $O_\ell^{\mathrm{sc}}(\boldsymbol{u}) := O_\ell\big(\boldsymbol{u}(\rho^\star)\big)$ a self-consistent local observable as in (2.7). Then,*

$$\left|O_\ell^{\mathrm{sc}}(\boldsymbol{u}) - I_{X_N} O_\ell\big(\boldsymbol{u}(\rho_N)\big)\right| \leqslant Ce^{-\gamma_N N},$$

where γ_N are the constants from Theorem 2.3 applied to $\boldsymbol{u}(\rho^\star)$.

In order for this result to be of any practical use, we need to solve the non-linear equation (2.32) for the electron density via a self-consistent field (SCF) procedure. Supposing we have the electron density ρ^i and corresponding state $\boldsymbol{u}^i := \boldsymbol{u}(\rho^i)$ after i iterations, we diagonalise the Hamiltonian $\mathcal{H}(\boldsymbol{u}^i)$ and hence evaluate the output density $\rho^{\mathrm{out}} = I_{X_N} F^\beta(\boldsymbol{u}^i)$. At this point, since the simple iteration $\rho^{i+1} = \rho^{\mathrm{out}}$ does not converge in general, a mixing strategy, possibly combined with Anderson acceleration [19], is used in order to compute the next iterate. The analysis of such mixing schemes is a major topic in electronic structure and numerical analysis in general and so we only present a small step in this direction.

Proposition 2.8. (Stability) *The approximate electron densities ρ_N arising from Theorem 2.6 are* stable *in the following sense: $I - \mathscr{L}_N(\rho_N)\colon \ell^2 \to \ell^2$ is an invertible bounded linear operator where $\mathscr{L}_N$ is the Jacobian of $\rho \mapsto I_{X_N} F^\beta\big(\boldsymbol{u}(\rho)\big)$. Moreover, $\big(I - \mathscr{L}_N(\rho_N)\big)^{-1}$ is uniformly bounded in N in operator norm.*

Theorem 2.9. *For $\boldsymbol{u}$ satisfying Definition 1, suppose that ρ_N is a corresponding approximate self-consistent electron density stable in the sense of Proposition 2.8. For fixed ρ^0, we define $\{\rho^i\}_{i=0}^\infty$ via the Newton iteration*

$$\rho^{i+1} = \rho^i - \big(I - \mathscr{L}_N(\rho^i)\big)^{-1}\Big(\rho^i - I_{X_N} F^\beta\big(\boldsymbol{u}(\rho^i)\big)\Big).$$

Then, for $\|\rho^0 - \rho_N\|_{\ell^\infty}$ sufficiently small, the Newton iteration converges quadratically to ρ_N.

A more thorough treatment of these SCF results is beyond the scope of this work. See [12,53,63] for recent results in the context of Hartree-Fock and Kohn-Sham density functional theory. For a recent review of SCF in the context density functional theory, see [101].

Remark 13. It is clear from the proofs of Theorems 2.6 and 2.9 that as long as the approximate scheme $F^{\beta,N}$ satisfies

$$\left|F_\ell^\beta(\boldsymbol{u}) - F_\ell^{\beta,N}(\boldsymbol{u})\right| \lesssim e^{-\gamma_N N} \qquad \text{and} \qquad \left|\frac{\partial F_\ell^\beta(\boldsymbol{u})}{\partial v_m} - \frac{\partial F_\ell^{\beta,N}(\boldsymbol{u})}{\partial v_m}\right| \lesssim e^{-\frac{1}{2}\gamma_N N} e^{-\eta r_{\ell m}},$$

then we may approximate (2.31) with approximate self-consistent solutions $\rho_N = F^{\beta,N}\big(\boldsymbol{u}(\rho_N)\big)$. In particular, as long as we have the estimate from Remark 10 (see Appendix C for the technical details), then we may use the nonlinear approximation scheme Θ_N from Theorem 2.5 in Theorems 2.6 and 2.9 . In this case, we obtain error estimates that are (asymptotically) independent of the discrete spectrum.

Remark 14. In the linear-scaling spectral Gauss quadrature (LSSGQ) method [87], a self-consistent field iteration analogous to (2.32) is proposed. In particular, with the caveats outlined in Remark 13 taken into consideration, Theorem 2.6 goes some way to rigorously justify the exponential rate of convergence observed numerically in [87, Fig. 4].

3. Conclusions and Discussion

The main result of this work is a sequence of rigorous results about body-ordered approximations of a wide class of properties extracted from tight-binding models for condensed phase systems, the primary example being the potential energy landscape. Our results demonstrate that exponentially fast convergence can be obtained, provided that the chemical environment is taken into account. In the spirit of our previous results on the locality of interaction [16,17,93], these provide further theoretical justification—albeit qualitative—for widely assumed properties of atomic interactions. More broadly, our analysis illustrates how to construct general low-dimensional but systematic representations of high-dimensional complex properties of atomistic systems. Our results, as well as potential generalisations, serve as a starting point towards a rigorous end-to-end theory of multi-scale and coarse-grained models, including but not limited to machine-learned potential energy landscapes.

In the following paragraphs we will make further remarks on the potential applications of our results, and on some apparent limitations of our analysis.

3.1. Representation of atomic properties

Our initial motivation for studying the body-order expansion was to explain the (unreasonable?) success of machine-learned interatomic potentials [5,8,80], and our remarks will focus on this topic, however in principle they apply more generally.

Briefly, given an *ab initio* potential energy landscape (PEL) E^{QM} for some material one formulates a parameterised interatomic potential

$$E(\{\boldsymbol{r}_\ell\}_\ell) = \sum_\ell \varepsilon(\boldsymbol{\theta}, \{\boldsymbol{r}_{\ell k}\}_{k\neq\ell})$$

and then "learn" the parameters $\boldsymbol{\theta}$ by fitting them to observations of the reference PEL E^{QM}. A great variety of such parameterisations exist, including but not limited to neural networks [8], kernel methods [5] and symmetric polynomials [2,25,80]. Symmetric polynomials are linear regression schemes where each basis function

has a natural body-order attached to it. It is particularly striking that for very low body-orders of four to six these schemes are able to match and often outperform the more complex nonlinear regression schemes [66,80,106]. Our analysis in the previous sections provides a partial explanation for these results, by justifying why one may expect that a reference *ab initio* PEL intrinsically has a low body-order. Moreover, classical approximation theory can now be applied to the body-ordered components as they are finite-dimensional to obtain new approximation results where the curse of dimensionality is alleviated.

Our results on nonlinear representations are less directly applicable to existing MLIPs, but rather suggest new directions to explore. Still, some connections can be made. The BOP-type construction of § 2.5,

$$\Theta_N\big(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \dots, [\mathcal{H}^N]_{\ell\ell}\big) \tag{3.1}$$

points towards a blending of machine-learning and BOP techniques that have not been explored to the best of our knowledge. A second interesting connection is to the overlap-matrix based fingerprint descriptors (OMFPs) introduced in [105] where a global spectrum for a small subcluster is used as a descriptor, while (3.1) can be understood as taking the projected spectrum as the descriptor. Thus, Theorem 2.5 suggests (1) an interesting modification of OMFPs which comes with guaranteed completeness to describe atomic properties; and (2) a possible pathway towards proving completeness of the original OMFPs.

Finally, our self-consistent representation of § 2.7 motivates how to construct compositional models, reminiscent of artificial neural networks, but with minimal nonlinearity that is moreover physically interpretable. Although we did not pursue it in the present work, this is a particularly promising starting point to incorporate meaningful electrostatic interaction into the MLIPs framework.

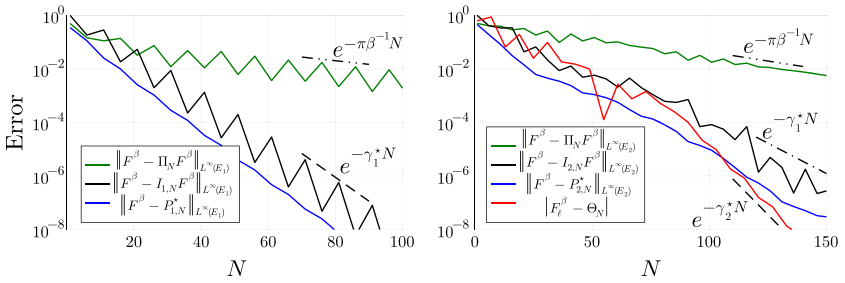
3.2. Linear body-ordered approximation: the preasymptotic regime

Possibly the most significiant limitation of our analysis of the linear body-ordered approximation scheme is that the estimates deteriorate when defects cause a pollution of the point spectrum. Here, we briefly demonstrate that this appears to be an asymptotic effect, while in the pre-asymptotic regime this deterioration is not noticable.

To explore this we choose a union of intervals $E \supseteq \sigma(\mathcal{H})$ and a polynomial P_N of degree N and note

$$\begin{aligned}\Big|\big[O(\mathcal{H}) - P_N(\mathcal{H})\big]_{\ell\ell}\Big| &\leqslant \big\|O(\mathcal{H}) - P_N(\mathcal{H})\big\|_{\ell^2\to\ell^2} = \big\|O - P_N\big\|_{L^\infty(\sigma(\mathcal{H}))} \\ &\leqslant \big\|O - P_N\big\|_{L^\infty(E)}. \end{aligned} \tag{3.2}$$

We then construct interpolation sets (Fejér sets) such that the corresponding polynomial interpolant gives the optimal asymptotic approximation rates (for details of this construction, see §4.1.5-§4.1.8). We then contrast this with a best $L^\infty(E)$-approximation, and with the nonlinear approximation scheme from Theorem 2.5. We will observe that the non-linearity leads to improved asymptotic but comparable pre-asymptotic approximation errors.



(A) $E_1 = [-1, a] \cup [b, 1]$. **(B)** $E_2 = [-1, a] \cup [c, d] \cup [b, 1]$.

Fig. 3. Approximation errors for Chebyshev projection (green), polynomial interpolation in Fejér sets on E_j (black), best $L^\infty(E_j)$ polynomial approximation (blue), and, for $j = 2$, errors in the nonlinear approximation scheme (red). We also plot the corresponding predicted asymptotic rates (from (4.5), (4.15), and Theorem 2.3). Here, we only plot data points for $N \in \{1, 6, 11, 16, \dots\}$ in the linear schemes (which captures the oscillatory behaviour), and $N \in \{1, 7, 13, 19, \dots\}$ for the nonlinear scheme (since N must be odd)

As a representative scenario we consider the Fermi-Dirac distribution $F^\beta(z) = (1+e^{\beta z})^{-1}$ with $\beta = 100$ and both the "defect-free" case $E_1 := [-1, a]\cup[b, 1]$ and $E_2 := [-1, a]\cup[c, d]\cup[b, 1]$ with the parameters $a = -0.2$, $b = 0.2$, $c = -0.06$, and $d = -0.03$. Then, for fixed polynomial degree N and $j \in \{1, 2\}$, we construct the $(N + 1)$-point Fejér set for E_j and the corresponding polynomial interpolant $I_{j,N}F^\beta$. Moreover, we consider a polynomial $P^\star_{j,N}$ of degree N minimising the right hand side of (3.2) for $E = E_j$. Then, in Figure 3, we plot the errors $\|F^\beta - I_{j,N}F^\beta\|_{L^\infty(E_j)}$ and $\|F^\beta - P^\star_{j,N}\|_{L^\infty(E_j)}$ for both $j = 1$ (Fig. 3a) and $j = 2$ (Fig. 3b) against the polynomial degree N together with the theoretical asymptotic convergence rates for best $L^\infty(E_j)$ polynomial approximation (4.15).

What we observe is that, as expected, introducing the interval $[c, d]$ into the approximation domain drastically affects the asymptotic convergence rate and the errors in the approximation based on interpolation. While the best approximation errors follow the asymptotic rate for larger polynomial degree, it appears that, pre-asymptotically, the errors are significantly reduced. We also see that the approximation errors are significantly better than the general error estimate $\|F^\beta - \Pi_N F^\beta\|_{L^\infty} \lesssim e^{-\pi\beta^{-1}N}$ where Π_N is the Chebyshev projection operator (see § 4.1.4).

Moreover, in Figure 3b, we plot the errors when using a nonlinear approximation scheme satisfying Theorem 2.5. In this simple experiment, we consider the Gauss quadrature rule $\Theta_N := \int I_{X_{\frac{1}{2}(N-1)}} F^\beta \mathrm{d}D_\ell$ where $X_{\frac{1}{2}(N-1)}$ is the set of zeros of the degree $\frac{1}{2}(N + 1)$ orthogonal polynomial (see Appendix D) with respect to $\mathrm{d}D_\ell(x) := \big(\chi_{E_1}(x) + \sum_j \delta(x - \lambda_j)\big)\mathrm{d}x$ where $\{\lambda_j\} = \{c, \frac{1}{2}(c + d), d\} \subset [c, d]$. While D_ℓ does not correspond to a physically relevant Hamiltonian, the same procedure may be carried out for any measure supported on E_1 with $\mathrm{supp}\, D_\ell \cap [c, d]$ finite. Then plotting the errors $|F_\ell^\beta - \Theta_N|$, we observe improved asymptotic convergence rates that agree with that of the "defect-free" case from Figure 3a. However,

the improvement is only observed in the asymptotic regime which corresponds to body-orders never reached in practice.

4. Proofs

4.1. Preliminaries

Here, we introduce the concepts needed in the proofs of the main results.

4.1.1. Hermite integral formula For a finite interpolation set $X \subset \mathbb{C}$, we let $\ell_X(z) := \prod_{x\in X}(z-x)$ be the correpsonding *node polynomial*.

For fixed $z \in \mathbb{C} \setminus X$, we suppose that O is analytic on an open neighbourhood of $X \cup \{z\}$. Then, for a simple closed positively oriented contour (or system of contours) $\mathscr{C}$ contained in the region of analyticity of O, encircling X, and avoiding $\{z\}$, we have

$$I_X O(z) = \frac{1}{2\pi i} \oint_{\mathscr{C}} \frac{\ell_X(\xi) - \ell_X(z)}{\ell_X(\xi)} \frac{O(\xi)}{\xi - z} \mathrm{d}\xi. \tag{4.1}$$

If, in addition, $\mathscr{C}$ encircles $\{z\}$, then

$$O(z) - I_X O(z) = \frac{1}{2\pi i} \oint_{\mathscr{C}} \frac{\ell_X(z)}{\ell_X(\xi)} \frac{O(\xi)}{\xi - z} \mathrm{d}\xi. \tag{4.2}$$

The proof of these facts is a simple application of Cauchy's integral formula, [3,95].

4.1.2. Resolvent calculus Given a configuration $\boldsymbol{u}$, we consider the Hamiltonian $\mathcal{H} = \mathcal{H}(\boldsymbol{u})$ and functions O analytic in some neighbourhood of the spectrum $\sigma(\mathcal{H})$. We define $O(\mathcal{H})$ via the holomorphic functional calculus [1]:

$$O(\mathcal{H}) := -\frac{1}{2\pi i} \oint_{\mathscr{C}} O(z)(\mathcal{H} - z)^{-1} \mathrm{d}z \tag{4.3}$$

where $\mathscr{C}$ is a simple closed positively oriented contour (or system of contours) contained in the region of analyticity of O and encircling the spectrum $\sigma(\mathcal{H})$.

The following Combes–Thomas resolvent estimate [21] will play a key role in the analysis:

Lemma 1. (Combes-Thomas) *Suppose that* $\boldsymbol{u}$ *satisfies Definition 1 and* $z \in \mathbb{C}$ *is contained in a bounded set with* $\operatorname{dist}\left(z, \sigma\left(\mathcal{H}(\boldsymbol{u})\right)\right) > 0$ *and* $\mathfrak{d} := \operatorname{dist}\left(z, \sigma\left(\mathcal{H}(\boldsymbol{u}^{\mathrm{ref}})\right)\right) > \delta$.

Then, there exists a constant $C > 0$ *such that*

$$\left|\left[(\mathcal{H}(\boldsymbol{u}) - z)^{-1}\right]_{\ell k}\right| \leqslant C_{\ell k} e^{-\gamma_{\mathrm{CT}} r_{\ell k}}, \qquad \text{where}$$
$$C_{\ell k} := 2\mathfrak{d}^{-1} + C e^{-\gamma_{\mathrm{CT}}(|\boldsymbol{r}_\ell| + |\boldsymbol{r}_k| - |\boldsymbol{r}_{\ell k}|)}$$

and $\gamma_{\mathrm{CT}} := c \min\{1, \mathfrak{d}\}$ *and* $c > 0$ *depends on* h_0, γ_0, d *and* $\min_{\ell \neq k} r_{\ell k}$.

Proof. A proof with γ_{CT} depending instead on $\operatorname{dist}\left(z, \sigma\left(\mathcal{H}(\boldsymbol{u})\right)\right)$ can be found in [16]. A low-rank update formula leads to the improved "defect-independent" result [17] where the exponent only depends on the distance between z and the reference spectrum. See [93] for an explicit description of γ_{CT} in terms of the constants γ_0, d and the non-interpenetration constant $\min_{\ell \neq k} r_{\ell k}$. □

A key observation for arguments involving forces (or more generally, derivatives of the analytic quantities of interest) is that the Combes-Thomas estimate allows us to bound derivatives of the resolvent operator:

Lemma 2. *Suppose that* $z \in \mathbb{C}$ *with* $\mathfrak{d} := \operatorname{dist}\left(z, \sigma\left(\mathcal{H}(\boldsymbol{u})\right)\right) > 0$. *Then,*

$$\left| \frac{\partial \left[(\mathcal{H}(\boldsymbol{u}) - z)^{-1} \right]_{\ell k}}{\partial \boldsymbol{u}_m} \right| \leqslant 4 h_0 \mathfrak{d}^{-2} e^{-\frac{1}{2} \min\{\gamma_0, \gamma_{\mathrm{CT}}\}(r_{\ell m} + r_{mk})}$$

where γ_{CT} *is the Combes-Thomas constant from Lemma 1 and* γ_0 *is the constant from (**TB**).*

Proof. This result can be found in the previous works [14,16,17], but we give a brief sketch for completeness.

Derivatives of the resolvent have the following form:

$$\frac{\partial (\mathcal{H}(\boldsymbol{u}) - z)^{-1}}{\partial \boldsymbol{u}_m} = -(\mathcal{H}(\boldsymbol{u}) - z)^{-1} \frac{\partial \mathcal{H}(\boldsymbol{u})}{\partial \boldsymbol{u}_m} (\mathcal{H}(\boldsymbol{r}, v) - z)^{-1}. \tag{4.4}$$

The result follows by applying the Combes-Thomas resolvent estimates together with the fact that the Hamiltonian is short-ranged (**TB**).

Assuming that the Hamiltonian has higher derivatives that are also short-ranged, higher order derivatives of the resolvent can be treated similarly [16]. □

4.1.3. Local observables Firstly, we note that $F^{\beta}(\cdot)$ is analytic away from the simple poles at $\pi\beta^{-1}(2\mathbb{Z}+1)$. Moreover, $G^{\beta}(\cdot)$ can be analytically continued onto the open set $\mathbb{C} \setminus \left\{\mu + ir : r \in \mathbb{R}, |r| \geqslant \pi\beta^{-1}\right\}$ [17]. Therefore, we may consider (4.3) with $O = F^{\beta}$ or G^{β} and a contour $\mathscr{C}_{\beta}$ encircling $\sigma(\mathcal{H})$ and avoiding $\mathbb{C} \setminus \left\{\mu + ir : r \in \mathbb{R}, |r| \geqslant \pi\beta^{-1}\right\}$. Therefore, we may choose $\mathscr{C}_{\beta}$ so that the constant $\mathfrak{d}$, from Lemma 1, is proportional to β^{-1}. Moreover, if there is a spectral gap, the constant $\mathfrak{d}$ is uniformly bounded below by a positive constant multiple of g as $\beta \to \infty$.

In the case of insulators at zero Fermi-temperature, we take $\mathscr{C}_{\infty}$ encircling $\sigma(\mathcal{H}(\boldsymbol{u})) \cap (-\infty, \mu)$ and avoiding the rest of the spectrum. Therefore, we may choose $\mathscr{C}_{\infty}$ so that the constant $\mathfrak{d}$, from Lemma 1, is proportional to g.

Following [16, Lemma 4], we can conclude that $\sigma(\mathcal{H}) \subset [\underline{\sigma}, \overline{\sigma}]$ for some $\underline{\sigma}, \overline{\sigma}$ depending on h_0, γ_0, v, d and $\min_{\ell \neq k} r_{\ell k}$. This means that, the contours $\mathscr{C}_{\beta}$ can be chosen to have finite length and, when applying Lemma 1, we have $\gamma_{\mathrm{CT}} = c \min\{1, \max\{\beta^{-1}, \mathsf{g}\}\}$.

Moreover, for all $0 < \mathsf{b} < \pi$ and bounded sets $A_{\beta} \subset A \subset \mathbb{C}$ such that

$$\operatorname{dist}(z, \{\mu + ir : r \in \mathbb{R}, |r| \geqslant \pi\beta^{-1}\}) \geqslant \mathsf{b}\beta^{-1} \qquad \text{for all } z \in A_{\beta},$$

both $F^{\beta}(\cdot)$ and $G^{\beta}(\cdot)$ are uniformly bounded on A_{β} independently of β [17, Lemma 5.2].

4.1.4. Chebyshev Projection and Interpolation in Chebyshev Points We denote by $\{T_n\}$ the *Chebyshev polynomials* (of the first kind) satisfying $T_n(\cos\theta) = \cos n\theta$ on $[-1, 1]$ and, equivalently, the recurrence $T_0 = 1$, $T_1 = x$, and $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$.

For O Lipshitz continuous on $[-1, 1]$, there exists an absolutely convergent Chebyshev series expansion: there exists c_n such that $O(z) = \sum_{n=0}^{\infty} c_n T_n(z)$. For maximal polynomial degree N, the corresponding projection operator is denoted $\Pi_N O(z) := \sum_{n=0}^{N} c_n T_n(z)$. This approach is a special case of the Kernel Polynomial Method (KPM) which we briefly review in Appendix F.

On the other hand, supposing that the interpolation set is given by the *Chebyshev points* $X = \{\cos \frac{j\pi}{N}\}_{0\leqslant j\leqslant N}$, we may expand the polynomial interpolant $I_N O := I_X O$ in terms of the Chebyshev polynomials: there exists c'_n such that $I_N O(z) = \sum_{n=0}^{N} c'_n T_n(z)$.

For functions O that can be analytically continued the *Bernstein ellipse* $E_\rho := \{\frac{1}{2}(z + z^{-1}) : |z| = \rho\}$ for $\rho > 1$, the corresponding coefficients $\{c_n\}$, $\{c'_n\}$ decay exponentially with rate ρ. This leads to the following error estimates

$$\|O - \Pi_N O\|_{L^\infty([-1,1])} + \|O - I_N O\|_{L^\infty([-1,1])} \leqslant 6\|O\|_{L^\infty(E_\rho)} \frac{\rho^{-N}}{\rho - 1}. \quad (4.5)$$

For $O^\beta = F^\beta$ or G^β, these estimates give an exponential rate of convergence with exponent depending on $\sim \beta^{-1}$. Indeed, after scaling $\mathcal{H}$ so that the spectrum is contained in $[-1, 1]$, we obtain

$$\begin{aligned} \left| O^\beta_\ell(\boldsymbol{u}) - \Pi_N O^\beta_\ell(\boldsymbol{u}) \right| &\leqslant \left\| O^\beta(\mathcal{H}) - \Pi_N O^\beta(\mathcal{H}) \right\|_{\ell^2 \to \ell^2} \\ &\leqslant \|O^\beta - \Pi_N O^\beta\|_{L^\infty([-1,1])}, \end{aligned} \quad (4.6)$$

and we conclude by directly applying (4.5). The same estimate also holds for I_N (or any polynomial).

For full details of all the statements made in this subsection, see [95].

4.1.5. Classical logarithmic potential theory In this section, we give a very brief introduction to classical potential theory in order to lay out the key notation. For a more thorough treatment, see [75] or [37,62,76,95].

It can be seen from the Hermite integral formula (4.2) that the approximation error for polynomial interpolation may be determined by taking the ratio of the size of the node polynomial ℓ_X at the approximation points to the size of ℓ_X along an appropriately chosen contour. Logarithmic potential theory provides an elegant mechanism for choosing the interpolation points so that the asymptotic behaviour of ℓ_X can be described.

We suppose that $E \subset \mathbb{C}$ is a compact set. We will see that choosing the interpolation nodes as to maximise the geometric mean of pairwise distances provides a particularly good approximation scheme:

$$\delta_n(E) := \max_{z_1,\dots,z_n \in E} \Big(\prod_{1\leqslant i<j\leqslant n} |z_i - z_j| \Big)^{\frac{2}{n(n-1)}}. \quad (4.7)$$

Any set $\mathcal{F}_n \subset E$ attaining this maximum is known as a *Fekete set*. It can be shown that the quantities $\delta_n(E)$ form a decreasing sequence and thus converges to what is known as the *transfinite diameter*: $\tau(E) := \lim_{n\to\infty} \delta_n(E)$.

We let $\ell_n(z)$ denote the node polynomial corresponding to a Fekete set and note that

$$|\ell_n(z)|\delta_n(E)^{\frac{n(n-1)}{2}} = \max_{z_0,\dots,z_n\in E:\ z_0=z} \prod_{0\leqslant i<j\leqslant n} |z_i - z_j| \leqslant \delta_{n+1}(E)^{\frac{n(n+1)}{2}}. \quad (4.8)$$

Therefore, rearranging (4.8), we obtain $\lim_{n\to\infty} \|\ell_n\|_{L^\infty(E)}^{1/n} \leqslant \tau(E)$. In fact, this inequality can be replaced with equality, showing that Fekete sets allow us to describe the asymptotic behaviour of the node polynomials on the domain of approximation.

To extend these results, it is useful to recast the maximisation problem (4.7) into the following minimisation problem, describing the minimal logarithmic energy attained by n particles lying in E with the repelling force $1/|z_i - z_j|$ between particles i and j lying at positions z_i and z_j, respectively:

$$\mathcal{E}_n(E) := \min_{z_1,\dots,z_n\in E} \sum_{1\leqslant i<j\leqslant n} \log \frac{1}{|z_i - z_j|} = \frac{n(n-1)}{2} \log \frac{1}{\delta_n(E)}. \quad (4.9)$$

Fekete sets can therefore be seen as minimal energy configurations and described by the normalised counting measure $\nu_n := \frac{1}{n}\sum_{j=1}^n \delta_{z_j}$ where $\mathcal{F}_n = \{z_j\}_{j=1}^n$.

The minimisation problem (4.9) may be extended for general unit Borel measures μ supported on E by defining the logarithmic potential and corresponding total energy by

$$U^\mu(z) := \int \log \frac{1}{|z-\xi|} \mathrm{d}\mu(\xi) \quad \text{and} \quad I(\mu) := \iint \log \frac{1}{|z-\xi|} \mathrm{d}\mu(\xi)\mathrm{d}\mu(z).$$

The infimum of the energy over the space of unit Borel measures supported on E, known as the *Robin constant* for E, will be denoted $-\infty < V_E \leqslant +\infty$. The *capacity* of E is defined as $\mathrm{cap}(E) := e^{-V_E}$ and is equal to the transfinite diameter [34]. Using a compactness argument, it can be shown that there exists an *equilibrium measure* ω_E with $I(\omega_E) = V_E$ and, in the case $V_E < \infty$, by the strict convexity of the integral, ω_E is unique [77]. Moreover, if $V_E < \infty$ (equivalently, if $\mathrm{cap}(E) > 0$), then $U^{\omega_E}(z) \leqslant V_E$ for all $z \in \mathbb{C}$, with equality holding on E except on a set of capacity zero (we say this property holds *quasi-everywhere*).

Moreover, if $\mathrm{cap}\, E > 0$, then it can be shown that the normalised counting measures, ν_n, corresponding to a sequence of Fekete sets weak-$\star$ converges to ω_E. Since $U^{\nu_n}(z) = \frac{1}{n}\log\frac{1}{|\ell_n(z)|}$, the weak-$\star$ convergence allows one to conclude that

$$\lim_{n\to\infty} \|\ell_n\|_{L^\infty(E)}^{1/n} = \mathrm{cap}(E), \quad \text{and}$$
$$\lim_{n\to\infty} |\ell_n(z)|^{1/n} = e^{-U^{\omega_E}(z)} =: \mathrm{cap}(E)e^{g_E(z)} \quad (4.10)$$

uniformly on compact subsets of $\mathbb{C}\setminus E$. Here, we have defined the *Green's function* $g_E(z) := V_E - U^{\omega_E}(z)$, which describes the asymptotic behaviour of the node polynomials corresponding to Fekete sets. We therefore wish to understand the Green's function g_E.

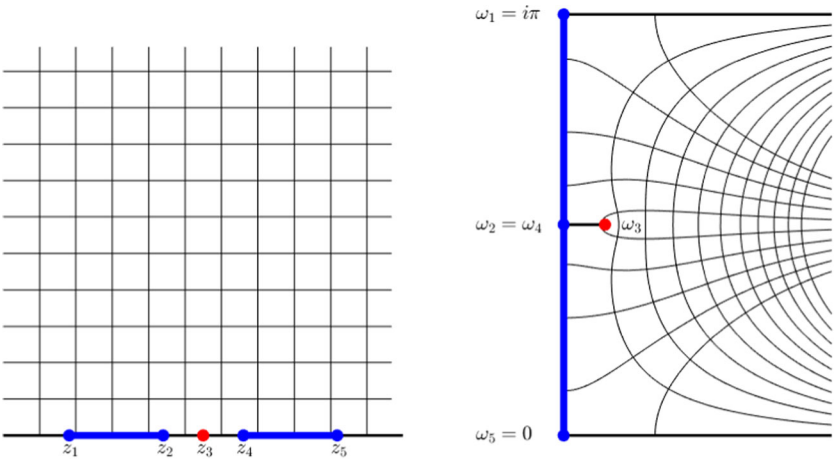


Fig. 4. The Schwarz–Christoffel mapping G_E with $E = [z_1, z_2] \cup [z_4, z_5]$ which maps the upper half plane (left) onto the infinite slit strip $\{\omega \in \mathbb{C}\colon \operatorname{Re}\omega > 0,\ \operatorname{Im}\omega \in (0, \pi)\}$ (right), is continuous on $\{z \in \mathbb{C}\colon \operatorname{Re}z \geqslant 0\}$ and maps the intervals $[z_1, z_2]$, $[z_4, z_5]$ to $[\omega_1, \omega_2]$, $[\omega_4, \omega_5] \subset i[0, \pi]$, respectively. We also plot the image of an 10×10 equi-spaced grid. A parameter problem is solved in order to obtain z_3 and thus ω_3 and $\omega_2 = \omega_4$ whereas the other constants are fixed. Here, we take $z_1 = -1, z_2 = -\varepsilon, z_4 = \varepsilon, z_5 = 1, \omega_1 = i\pi, \omega_5 = 0$ with $\varepsilon = 0.3$

4.1.6. Construction of the Green's function Now we restrict our attention to the particular case where $E \subset \mathbb{R}$ is a union of finitely many compact intervals of non-zero length.

It can be shown that the Green's function g_E satisfies the following Dirichlet problem on $\mathbb{C} \setminus E$ [75]:

$$\Delta g_E(z) = 0 \quad \text{on } \mathbb{C} \setminus E, \tag{4.11a}$$

$$g_E(z) \sim \log|z| \quad \text{as } |z| \to \infty, \tag{4.11b}$$

$$g_E(z) = 0 \quad \text{on } E. \tag{4.11c}$$

In fact, it can be shown that (4.11) admits a unique solution [75] and thus (4.11) is an alternative definition of the Green's function. Using this characterisation, it is possible to explicitly construct the Green's function g_E as follows. In the upper half plane, $g_E(z) = \operatorname{Re}(G_E(z))$ where $G_E\colon \{z \in \mathbb{C}\colon \operatorname{Im}(z) \geqslant 0\} \to \{z \in \mathbb{C}\colon \operatorname{Re}(z) \geqslant 0, \operatorname{Im}(z) \in [0, \pi]\}$ is a conformal mapping on $\{z\colon \operatorname{Im}(z) > 0\}$ such that $G_E(E) = i[0, \pi]$, $G_E(\min E) = i\pi$, and $G_E(\max E) = 0$. Using the symmetry of E with respect to the real axis, we may extend $\operatorname{Re}(G_E(z))$ to the whole complex plane via the Schwarz reflection principle. Then, one can easily verify that this analytic continuation satisfies (4.11). Since the image of G_E is a (generalised) polygon, $z \mapsto G_E(z)$ is an example of a Schwarz–Christoffel mapping [29]. See Figure 4 for the case $E = [-1, -\varepsilon] \cup [\varepsilon, 1]$.

We shall briefly discuss the construction of the Schwarz–Christoffel mapping G_E for $E = [-1, \varepsilon_-] \cup [\varepsilon_+, 1]$. We define the *pre-vertices* $z_1 = -1, z_2 = \varepsilon_-, z_4 = \varepsilon_+, z_5 = 1$ and wish to construct a conformal map G_E with $G_E(z_k) = \omega_k$ as in

Figure 4. For simplicity, we also define $z_0 := -\infty$ and $z_6 := \infty$ and observe that because the image is a polygon, $\arg G'_E(z)$ must be constant on each interval (z_{k-1}, z_k) and

$$\arg G'_E(z_k^+) - \arg G'_E(z_k^-) = (1 - \alpha_k)\pi, \tag{4.12}$$

where $z_k^- \in (z_{k-1}, z_k)$, $z_k^+ \in (z_k, z_{k+1})$, and $\alpha_k\pi$ is the interior angle of the infinite slit strip at vertex ω_k (that is, $\alpha_1 = \alpha_2 = \alpha_4 = \alpha_5 = \frac{1}{2}$ and $\alpha_3 = 2$). After defining $z^\alpha := |z|^\alpha e^{i\alpha \arg z}$ where $\arg z \in (-\pi, \pi]$, we can see that for $z \in (z_{k-1}, z_k)$, we have $\arg \prod_{j=k}^5 (z - z_j)^{\alpha_j - 1} = \sum_{j=k}^5 (\alpha_j - 1)\pi$ and so the jump in the argument of $z \mapsto \prod_{j=1}^5 (z - z_j)^{\alpha_j - 1}$ is $(1 - \alpha_k)\pi$ at z_k as in (4.12). Therefore, integrating this expression, we obtain

$$G_E(z) = A + B \int_1^z \frac{\zeta - z_3}{\sqrt{\zeta + 1}\sqrt{\zeta - \varepsilon_-}\sqrt{\zeta - \varepsilon_+}\sqrt{\zeta - 1}} \mathrm{d}\zeta. \tag{4.13}$$

Since $G_E(1) = A$, we take $A = 0$ (to ensure (4.11c) holds). Moreover, since the real part of the integral is $\sim \log|z|$ as $|z| \to \infty$, we apply (4.11b) to conclude $B = 1$. Finally, we can choose z_3 such that $\operatorname{Re} G_E(z) = 0$ for all $z \in E$; that is,

$$z_3 \in (\varepsilon_-, \varepsilon_+) : \qquad \int_{\varepsilon_-}^{\varepsilon_+} \frac{\zeta - z_3}{\sqrt{\zeta + 1}\sqrt{\zeta - \varepsilon_-}\sqrt{\varepsilon_+ - \zeta}\sqrt{1 - \zeta}} \mathrm{d}\zeta = 0. \tag{4.14}$$

For more details, see [37]. We use the Schwarz–Christoffel toolbox [29] in MATLAB to evaluate (4.13) and plot Figure 5.

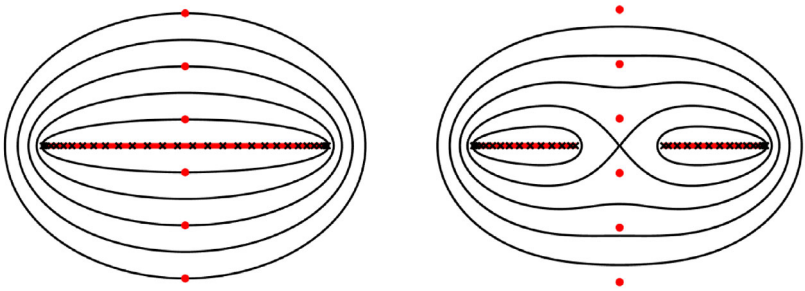
For the simple case $E := [-1, 1]$, by the same analysis, we can disregard z_2, z_3, z_4 and $\omega_2, \omega_3, \omega_4$ and integrate the corresponding expression to obtain the closed form $G_{[-1,1]}(z) = \log(z + \sqrt{z-1}\sqrt{z+1})$.

A similar analysis allows one to construct conformal maps from the upper half plane to the interior of any polygon. For further details, rigorous proofs and numerical considerations, see [31].

4.1.7. Interpolation nodes The only difficulty in obtaining (4.10) in practice is the fact that Fekete sets are difficult to compute. An alternative, based on the Schwarz–Christoffel mapping G_E, are *Fejér points*. For equally spaced points $\{\zeta_j\}_{j=1}^n$ on the interval $i[0, \pi]$, the n^{th} Fejér set is defined by $\{G_E^{-1}(\zeta_j)\}_{j=1}^n$. Fejér sets are also *asymptotically optimal* in the sense that (4.10) is satisfied where ℓ_n is now the node polynomial corresponding to n-point Fejér set.

Another approach is to use *Leja points* which are generated by the following algorithm: for fixed $z_1, \ldots, z_n$, the next interpolation node z_{n+1} is constructed by maximising $\prod_{j=1}^n |z_j - z|$ over all $z \in E$. Sets of this form are also asymptotically optimal [90] for any choice of $z_1 \in E$. Since we have fixed the previous nodes $z_1, \ldots, z_n$, the maximisation problem for constructing z_{n+1} is much simpler than that of (4.7).

More generally, if the normalised counting measure corresponding to a sequence of sets $\{z_j\}_{j=1}^n \subset E$ weak-$\star$ converges to the equilibrium measure ω_E, then the corresponding node polynomials satisfy (4.10).



(A) Metal: $E = [-1, 1]$. **(B)** Insulator: $E = [-1, -\varepsilon] \cup [\varepsilon, 1]$, $\varepsilon = 0.30$ (2 s.f.).

Fig. 5. Equi-potential curves $\mathscr{C}_{r_k} := \{z \in \mathbb{C} : e^{g_E(z)} = r_k\}$ for both metals (A) and insulators (B) where $\frac{1}{2}(r_k - r_k^{-1}) = \frac{k\pi}{\beta}$ for $k \in \{1, 2, 3, 4, 5\}$ and $\beta = 10$. In the case of metals (A), the equi-potential curves agree with Bernstein ellipses. We also plot the poles of $F^\beta(\cdot)$ which determine the maximal admissible integration contours: for (A), we can take contours $\mathscr{C}_r$ for all $r < r_1$ and, for (B), the contour $\mathscr{C}_{r_2}$ can be used for all positive Fermi-temperatures (we have chosen the gap carefully so that $\mathscr{C}_{r_2}$ self-intersects at μ). Shown in black crosses are 30 Fejér points in each case. To create these plots we consider an integral formula for the Green's function $z \mapsto g_E(z)$ [37] and use the Schwarz–Christoffel MATLAB toolbox [28,29] to approximate these integrals

For the simple case where $E = [-1, 1]$, many systems of zeros or maxima of sequences of orthogonal polynomials are asymptotically optimal in the sense of (4.10). In fact, since the equilibrium measure for $[-1, 1]$ is the arcsine measure [76]

$$\mathrm{d}\mu_{[-1,1]}(x) = \frac{1}{\pi} \frac{1}{\sqrt{1 - x^2}} \mathrm{d}x,$$

any sequence of sets with this limiting distribution is asymptotically optimal. An example of particular interest are the *Chebyshev points* $\{\cos \frac{j\pi}{n}\}_{0 \leqslant j \leqslant n}$ given by the $n+1$ extreme points of the Chebyshev polynomials defined by $T_n(\cos\theta) = \cos n\theta$.

4.1.8. Asymptotically optimal polynomial approximations Suppose that E is the union of finitely many compact intervals of non-zero length and $O : E \to \mathbb{C}$ extents to an analytic function in an open neighbourhood of E. On defining $\mathscr{C}_\gamma := \{z \in \mathbb{C} : g_E(z) = \gamma\}$, we denote by $\gamma^\star$ the maximal constant for which O is analytic on the interior of $\mathscr{C}_{\gamma^\star}$. We let $P_N^\star$ be the best $L^\infty(E)$-approximation to O in the space of polynomials of degree at most N and suppose that I_N is a polynomial interpolation operator in $N + 1$ points satisfying (4.11). Then, the Green's function g_E determines the asymptotic rate of approximation for not only polynomial interpolation, but also for best approximation:

$$\lim_{N\to\infty} \|O - P_N^\star\|_{L^\infty(E)}^{1/N} = \lim_{N\to\infty} \|O - I_N O\|_{L^\infty(E)}^{1/N} = e^{-\gamma^\star}. \tag{4.15}$$

For a proof that the asymptotic rate of best approximation is given by the Green's function see [76]. The result for polynomial interpolation uses the Hermite integral formula and (4.10), see (4.20) and (4.22), below.

4.2. Linear body-order approximation

In this section, we use the classical logarithmic potential theory from § 4.1.5 to prove the approximation error bounds for interpolation. However, we first show that polynomial approximations lead to body-order approximations:

Proof of Proposition 2.2. We first simplify the notation by absorbing the effective potential and two-centre terms into the three-centre summation:

$$\mathcal{H}(\boldsymbol{u})_{k_1k_2} = \sum_m \mathcal{H}_{k_1k_2m}, \quad \text{where}$$

$$\mathcal{H}_{k_1k_2m} := \begin{cases} \frac{1}{2}h(\boldsymbol{u}_{k_1k_2}) + \delta_{k_1k_2}v_{k_1}\mathrm{Id}_{N_\mathrm{b}}, & \text{if } m \in \{k_1, k_2\}, \\ t(\boldsymbol{u}_{k_1m}, \boldsymbol{u}_{k_2m}), & \text{if } m \notin \{k_1, k_2\}. \end{cases} \tag{4.16}$$

Now, supposing that $I_X O(z) = \sum_{j=0}^{|X|-1} c_j z^j$, we obtain

$$\begin{aligned} I_X O_\ell(\boldsymbol{u}) &= \mathrm{tr} \sum_{j=0}^{|X|-1} c_j \sum_{\ell_1,\dots,\ell_{j-1}} \mathcal{H}_{\ell\ell_1}\mathcal{H}_{\ell_1\ell_2}\dots\mathcal{H}_{\ell_{j-1}\ell} \\ &= \mathrm{tr} \sum_{j=0}^{|X|-1} c_j \sum_{\substack{\ell_1,\dots,\ell_{j-1} \\ m_1,\dots,m_j}} \mathcal{H}_{\ell\ell_1m_1}\mathcal{H}_{\ell_1\ell_2m_2}\dots\mathcal{H}_{\ell_{j-1}\ell m_j}. \end{aligned} \tag{4.17}$$

there the first two terms in the outer summation are c_0 and $c_1\mathcal{H}_{\ell\ell}$. Now, for a fixed body-order $(n+1)$, and $k_1 < \cdots < k_n$ with $k_l \neq \ell$, we construct $V_{nN}(\boldsymbol{u}_\ell; \boldsymbol{u}_{\ell k_1}, \dots, \boldsymbol{u}_{\ell k_n})$ by collecting all terms in (4.17) with $0 \leqslant j \leqslant |X| - 1$ and $\{\ell, \ell_1, \dots, \ell_{j-1}, m_1, \dots, m_j\} = \{\ell, k_1, \dots, k_n\}$. In particular, the maximal body-order in this expression is $2(|X| - 1)$ for three-centre models and $|X| - 1$ in the two-centre case.

More explicitly, using the notation (2.26), we have that

$$\begin{aligned} &V_{nN}(\boldsymbol{u}_\ell; \boldsymbol{u}_{\ell k_1}, \dots, \boldsymbol{u}_{\ell k_n}) \\ &\quad = \mathrm{tr} \sum_{j=0}^{|X|-1} c_j \sum_{\substack{\ell_1,\dots,\ell_{j-1},m_1,\dots,m_j \\ \{\ell,\ell_1,\dots,\ell_{j-1},m_1,\dots,m_j\}=\{\ell,k_1,\dots,k_n\}}} \mathcal{H}_{\ell\ell_1m_1}\mathcal{H}_{\ell_1\ell_2m_2}\dots\mathcal{H}_{\ell_{j-1}\ell m_j} \qquad (4.18) \\ &\quad = \mathrm{tr} \sum_{K\subseteq\{k_1,\dots,k_n\}} (-1)^{n-|K|} I_X O\big(\mathcal{H}\big|_{\ell;K}\big)_{\ell\ell}. \qquad (4.19) \end{aligned}$$

Here, we have applied an inclusion-exclusion principle to ensure that we are not only summing over sites in $\{k_1, \dots, k_n\}$ but we select at least one of each site in this set. Indeed, if we choose $\ell_1, \dots, \ell_{j-1}, m_1, \dots, m_j$ such that $\{\ell, \ell_1, \dots, \ell_{j-1}, m_1, \dots, m_j\} = \{\ell\} \cup K_0$, then the expression $\mathcal{H}_{\ell\ell_1m_1}\mathcal{H}_{\ell_1\ell_2m_2}\cdots\mathcal{H}_{\ell_{j-1}\ell m_j}$ appears in each term of (4.19) with $K \supseteq K_0$ exactly once (with a $\pm$ sign). Therefore, the number of times $\mathcal{H}_{\ell\ell_1m_1}\mathcal{H}_{\ell_1\ell_2m_2}\dots\mathcal{H}_{\ell_{j-1}\ell m_j}$ appears is exactly

$$\sum_{l=0}^{n-|K_0|} (-1)^{n-|K_0|-l} \binom{n-|K_0|}{l} = \begin{cases} 1 & \text{if } |K_0| = n, \\ 0 & \text{otherwise.} \end{cases}$$

That is, (4.19) only contains the terms in the summation (4.18). □

Proof of Theorem 2.3. We let $\ell_N(x) := \prod_j (x - x_j^N)$ be the node polynomial for $X_N := \{x_j^N\}_{j=0}^N$. Again, we fix the configuration $\boldsymbol{u}$ and consider $\mathcal{H} := \mathcal{H}(\boldsymbol{u})$.

Supposing that $\mathscr{C}$ is a simple closed positively oriented contour encircling $\sigma(\mathcal{H})$, we apply the Hermite integral formula (4.2) to obtain that

$$\begin{aligned}\left|O_\ell^\beta(\boldsymbol{u}) - I_{X_N} O_\ell^\beta(\boldsymbol{u})\right| &\leqslant \|O^\beta(\mathcal{H}) - I_{X_N} O^\beta(\mathcal{H})\|_{\ell^2\to\ell^2} = \sup_{z\in\sigma(\mathcal{H})} \left|O^\beta(z) - I_{X_N} O^\beta(z)\right| \\ &\leqslant \sup_{z\in\sigma(\mathcal{H})} \left|\frac{1}{2\pi i}\oint_{\mathscr{C}} \frac{\ell_N(z)}{\ell_N(\xi)} \frac{O^\beta(\xi)}{\xi - z} \mathrm{d}\xi\right| \leqslant C \sup_{z\in\sigma(\mathcal{H}),\, \xi\in\mathscr{C}} \left|\frac{\ell_N(z)}{\ell_N(\xi)}\right|,\end{aligned} \tag{4.20}$$

where

$$C := \frac{\mathrm{len}(\mathscr{C})}{2\pi} \frac{\max_{\xi\in\mathscr{C}} |O^\beta(\xi)|}{\mathrm{dist}\big(\mathscr{C}, \sigma(\mathcal{H})\big)}. \tag{4.21}$$

At this point we apply standard results of classical logarithmic potential theory (see, § 4.1.5 or [62]) and conclude by noting that if the interpolation points are asymptotically distributed according to the equilibrium distribution corresponding to $E := I_- \cup I_+$, then after applying (4.10), we have that

$$\lim_{N\to\infty} \left|\frac{\ell_N(z)}{\ell_N(\xi)}\right|^{\frac{1}{N}} = e^{g_E(z) - g_E(\xi)}. \tag{4.22}$$

Here, the equilibrium distribution and the Green's function $g_E(z)$ are concepts introduced in § 4.1.5 and § 4.1.6.

Therefore, by choosing the contour $\mathscr{C} := \{\xi \in \mathbb{C}\colon g_E(\xi) = \gamma\}$ for $0 < \gamma < g_E(\mu + i\pi\beta^{-1})$, the asymptotic exponents in the approximation error is γ.

The maximal asymptotic convergence rate is given by $g_E(\mu + i\pi\beta^{-1})$ since $\mathscr{C}$ must be contained in the region of analyticity of O^β and the first singularity of O^β is at $\mu + i\pi\beta^{-1}$ (for $O^\beta = F^\beta$ or G^β).

Examples of the equi-potential level sets $\mathscr{C}$ are given in Figure 5.

Using the Green's function results of § 4.1.6, $g_E(\mu + i\pi\beta^{-1}) = \mathrm{Re}\, G_E(\mu + i\pi\beta^{-1})$ where G_E is the integral (4.13). The asymptotic behaviour of this maximal asymptotic convergence rate for the separate $\beta \to \infty$ and $\mathsf{g} \to 0$ limits can be found in [37,81]. Here, we consider the $\beta^{-1} + \mathsf{g} \to 0$ limit where the gap remains symmetric about the chemical potential μ.

To simplify the notation we consider $I_- \cup I_+ = [-1, \varepsilon_-] \cup [\varepsilon_+, 1]$ where $\varepsilon_\pm = \mu \pm \frac{1}{2}\mathsf{g}$. By choosing to integrate (4.13) along the contour composed of the intervals $[1, \mu]$ and $[\mu, \mu + i\pi\beta^{-1}]$, we obtain

$$G_E(\mu + i\pi\beta^{-1}) = G_E(\mu) + \int_\mu^{\mu + i\pi\beta^{-1}} \frac{\zeta - z_3}{\sqrt{\zeta+1}\sqrt{\zeta-\varepsilon_-}\sqrt{\zeta-\varepsilon_+}\sqrt{\zeta-1}} \mathrm{d}\zeta. \tag{4.23}$$

Since $g_E(\mu) \sim \mathsf{g}$ as $\mathsf{g} \to 0$ [37], we only consider the remaining term in (4.23).

For $\zeta \in \mu + i[0, \pi\beta^{-1}]$, we have $c^{-1} \leqslant |\sqrt{\zeta \pm 1}| \leqslant c$, and so the integral in (4.23) has the same asymptotic behaviour as

$$\int_{\mu}^{\mu+i\pi\beta^{-1}} \frac{\zeta - z_3}{\sqrt{\zeta - \varepsilon_-}\sqrt{\zeta - \varepsilon_+}} \mathrm{d}\zeta = \mathsf{g} \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{i\pi\beta^{-1}}{\mathsf{g}}} \frac{\sqrt{\zeta}}{\sqrt{\zeta - 1}} \mathrm{d}\zeta$$
$$+(\varepsilon_- - z_3) \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{i\pi\beta^{-1}}{\mathsf{g}}} \frac{1}{\sqrt{\zeta}\sqrt{\zeta - 1}} \mathrm{d}\zeta, \tag{4.24}$$

where we have used the change of variables $\widetilde{\zeta} = \frac{\zeta - \varepsilon_-}{\varepsilon_+ - \varepsilon_-}$.

Since the integrands are uniformly bounded along the domain of integration, (4.24) is $\sim \beta^{-1}$ as $\beta \to \infty$.

The constant pre-factor in (4.21) is inversely proportional to the distance $\mathrm{dist}\big(\mathscr{C}, \sigma(\mathcal{H})\big)$ between the contour $\mathscr{C} = \{g_E = \gamma\}$ and the spectrum $\sigma(\mathcal{H})$. In particular, since g_E is uniformly Lipschitz with constant $L > 0$ on the compact region bounded by $\mathscr{C}$, we have: there exists $\lambda \in \sigma(\mathcal{H})$ and $\xi \in \mathscr{C}$ such that

$$\mathrm{dist}\big(\mathscr{C}, \sigma(\mathcal{H})\big) = |\xi - \lambda| \geqslant \frac{1}{L} |g_E(\xi) - g_E(\lambda)| = \frac{1}{L}\gamma.$$

Therefore, choosing γ to be a constant multiple of $g_E(\mu + i\pi\beta^{-1})$, we conclude that the constant pre-factor C satisfies $C \sim (\mathsf{g} + \beta^{-1})^{-1}$ as $\mathsf{g} + \beta^{-1} \to 0$.

To extend the body-order expansion results to derivatives (in particular, to forces), we write the quantities of interest using resolvent calculus, apply Lemma 2 to bound the derivatives of the resolvent, and use the Hermite integral formula (4.20) to conclude: for $\mathscr{C}_1, \mathscr{C}_2$ simple closed positively oriented contours encircling the spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ and $\mathscr{C}_1$, respectively, we have

$$\left| \frac{\partial O_\ell(\boldsymbol{u})}{\partial \boldsymbol{u}_m} - \frac{\partial I_{X_N} O_\ell(\boldsymbol{u})}{\partial \boldsymbol{u}_m} \right| = \frac{1}{2\pi} \left| \oint_{\mathscr{C}_1} \big(O(z) - I_{X_N} O(z)\big) \frac{\partial \big(\mathcal{H}(\boldsymbol{u}) - z\big)^{-1}_{\ell\ell}}{\partial \boldsymbol{u}_m} \mathrm{d}z \right|$$
$$= \frac{1}{4\pi^2} \left| \oint_{\mathscr{C}_1} \oint_{\mathscr{C}_2} \frac{\ell_N(z)}{\ell_N(\xi)} \frac{O(\xi)}{\xi - z} \frac{\partial \big(\mathcal{H}(\boldsymbol{u}) - z\big)^{-1}_{\ell\ell}}{\partial \boldsymbol{u}_m} \mathrm{d}\xi \mathrm{d}z \right|$$
$$\leqslant C e^{-\eta r_{\ell m}} \sup_{z \in \mathscr{C}_1, \xi \in \mathscr{C}_2} \left| \frac{\ell_N(z)}{\ell_N(\xi)} \right|. \tag{4.25}$$

We conclude by choosing appropriate contours $\mathscr{C}_l = \{g_E = \gamma_l\}$ for $l = 1, 2$ and applying (4.22). □

4.2.1. The role of the point spectrum To begin this section, we sketch the proof of Proposition 2.1.

Proof of Proposition 2.1. (i) Sup-norm perturbations. We suppose that $\sup_k \big[|\boldsymbol{r}_k - \boldsymbol{r}_k^{\mathrm{ref}}| + |v_k - v_k^{\mathrm{ref}}|\big] \leqslant \delta$ for $\delta > 0$ sufficiently small such that

$$\big|h(\boldsymbol{u}_{\ell k}) - h(\boldsymbol{u}_{\ell k}^{\mathrm{ref}})\big| = \big|\nabla h(\xi_{\ell k}) \cdot (\boldsymbol{r}_{\ell k} - \boldsymbol{r}_{\ell k}^{\mathrm{ref}})\big| \leqslant C\delta e^{-\frac{1}{2}\gamma_0 r_{\ell k}}, \quad \text{and}$$
$$\big|t(\boldsymbol{u}_{\ell m}, \boldsymbol{u}_{km}) - t(\boldsymbol{u}_{\ell m}^{\mathrm{ref}}, \boldsymbol{u}_{km}^{\mathrm{ref}})\big| = \big|\nabla_1 t(\xi_{\ell m}^{(1)}, \zeta_{km}^{(1)}) \cdot (\boldsymbol{r}_{\ell m} - \boldsymbol{r}_{\ell m}^{\mathrm{ref}})$$

$$+ \nabla_2 t(\xi_{\ell m}^{(2)}, \zeta_{km}^{(2)}) \cdot (\boldsymbol{r}_{km} - \boldsymbol{r}_{km}^{\mathrm{ref}})\big| \leqslant C\delta e^{-\frac{1}{2}\gamma_0(r_{\ell m}+r_{km})},$$

where, $\xi_{\ell k} \in [\boldsymbol{r}_{\ell k}, \boldsymbol{r}_{\ell k}^{\mathrm{ref}}]$, $\xi_{\ell m}^{(l)} \in [\boldsymbol{r}_{\ell m}, \boldsymbol{r}_{\ell m}^{\mathrm{ref}}]$, and $\zeta_{km}^{(l)} \in [\boldsymbol{r}_{km}, \boldsymbol{r}_{km}^{\mathrm{ref}}]$. Therefore, if $\psi \in \ell^2$, we have

$$\left\| \big(\mathcal{H}(\boldsymbol{u}) - \mathcal{H}(\boldsymbol{u}^{\mathrm{ref}})\big)\psi \right\|_{\ell^2}^2 \leqslant C\delta^2 \sum_{\ell k} e^{-\frac{1}{2}\gamma_0 r_{\ell k}} |\psi_k|^2 \leqslant C\delta^2 \|\psi\|_{\ell^2}^2. \tag{4.26}$$

Therefore, applying standard results from perturbation theory [56, p. 291], we obtain

$$\operatorname{dist}\Big(\sigma\big(\mathcal{H}(\boldsymbol{u})\big), \sigma\big(\mathcal{H}(\boldsymbol{u}^{\mathrm{ref}})\big)\Big) \leqslant \left\| \mathcal{H}(\boldsymbol{u}) - \mathcal{H}(\boldsymbol{u}^{\mathrm{ref}}) \right\|_{\ell^2 \to \ell^2} \leqslant C\delta.$$

(ii) Finite rank perturbations. The finite rank perturbation result has been presented in [70] in a slightly different setting. We sketch the main idea here for completeness.

Since the essential spectrum is stable under compact (in particular, finite rank) perturbations [56], the set

$$\sigma\big(\mathcal{H}(\boldsymbol{u})\big) \setminus B_\delta\big(\sigma\big(\mathcal{H}(\boldsymbol{u}^{\mathrm{ref}})\big)\big) \subseteq \sigma_{\mathrm{disc}}\big(\mathcal{H}(\boldsymbol{u})\big) \setminus B_\delta\big(\sigma_{\mathrm{ess}}\big(\mathcal{H}(\boldsymbol{u}^{\mathrm{ref}})\big)\big)$$

is both compact and discrete and therefore finite. □

Proof of Theorem 2.4. Suppose that $\mathscr{C}$ is a simple closed contour encircling the spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$ and (λ_s, ψ_s) are normalised eigenpairs corresponding to the finitely many eigenvalues outside $I_- \cup I_+$. Therefore, we have that

$$\begin{aligned} O_\ell^\beta(\boldsymbol{u}) - I_{X_N} O_\ell^\beta(\boldsymbol{u}) = \frac{1}{2\pi i} \oint_{\mathscr{C}} \big(O^\beta(z) - I_{X_N} O^\beta(z)\big) \operatorname{tr}\big[\big(z - \mathcal{H}(\boldsymbol{u})\big)^{-1}\big]_{\ell\ell} \mathrm{d}z \\ + \sum_s \big(O^\beta(\lambda_s) - I_{X_N} O^\beta(\lambda_s)\big) \big|[\psi_s]_\ell\big|^2. \end{aligned} \tag{4.27}$$

The first term of (4.27) may be treated in the same way as in the proof of Theorem 2.3. Moreover, derivatives of this term may be treated in the same way as in (4.25). It is therefore sufficient to bound the remaining term and its derivative.

Firstly, we note that the eigenvectors corresponding to isolated eigenvalues in the spectral gap have the following decay [17]: for $\mathscr{C}'$ a simple closed positively oriented contour (or system of contours) encircling the $\{\lambda_s\}$, we have that

$$\begin{aligned} \sum_s \big|[\psi_s]_\ell\big|^2 &= \frac{1}{2\pi} \left| \oint_{\mathscr{C}'} \big[\big(\mathcal{H}(\boldsymbol{u}) - z\big)^{-1}\big]_{\ell\ell} \mathrm{d}z \right| \\ &= \frac{1}{2\pi} \left| \oint_{\mathscr{C}'} \big[\big(\mathcal{H}(\boldsymbol{u}) - z\big)^{-1} - \big(\mathcal{H}(\boldsymbol{u}^{\mathrm{ref}}) - z\big)^{-1}\big]_{\ell\ell} \mathrm{d}z \right| \\ &\leqslant C e^{-\gamma_{\mathrm{CT}}[|\boldsymbol{r}_\ell| - R_{\mathrm{def}}]}, \end{aligned} \tag{4.28}$$

where γ_{CT} is the Combes-Thomas constant from Lemma 1 with $\mathfrak{d} = \operatorname{dist}\big(\mathscr{C}', \sigma(\mathcal{H}(\boldsymbol{u}))\big)$. The constant pre-factor in (4.28) depends on the distance between the

contour and the defect spectrum $\sigma\left(\mathcal{H}(\boldsymbol{u})\right)$. Similar estimates hold for the derivatives. For full details on the derivation of (4.28), see [17, (5.18)–(5.21)].

Therefore, combining (4.28) and the Hermite integral formula, we conclude as in the proof of Theorem 2.3. □

4.3. Non-linear body-order approximation

In this section, we prove Theorem 2.5 by applying the recursion method to reformulate the problem into a semi-infinite linear chain and replacing the far-field with vacuum.

4.3.1. Recursion method In that follows, we briefly introduce the recursion method [49,50], a reformulation of the Lanczos process [61], which generates a tri-diagonal (Jacobi) operator T [91] whose spectral measure is D_ℓ and the corresponding sequence of orthogonal polynomials [40]. This process provides the basis for constructing approximations to the LDOS giving rise to nonlinear approximation schemes satisfying Theorem 2.5.

Recall that D_ℓ is the LDOS satisfying (2.13). We start by defining $p_0 := 1, a_0 := \int x \mathrm{d}D_\ell(x)$ and $b_1 p_1(x) := x - a_0$ where b_1 is the normalising constant to ensure $\int p_1(x)^2 \mathrm{d}D_\ell(x) = 1$. Then, supposing we have defined $a_0, a_1, b_1, \ldots, a_n, b_n$ and the polynomials $p_0(x), \ldots, p_n(x)$, we set

$$b_{n+1} p_{n+1}(x) := (x - a_n) p_n(x) - b_n p_{n-1}(x), \qquad \text{with} \tag{4.29}$$

$$\int p_{n+1}(x)^2 \mathrm{d}D_\ell(x) = 1, \qquad a_{n+1} := \int x p_{n+1}(x)^2 \mathrm{d}D_\ell(x). \tag{4.30}$$

Then, $\{p_n\}$ is a sequence of orthogonal polynomials with respect to D_ℓ (i.e. $\int p_n p_m \mathrm{d}D_\ell = \delta_{nm}$) and we have that

$$T_N := \Big(\int x p_n p_m \mathrm{d}D_\ell \Big)_{0 \leqslant n,m \leqslant N} = \begin{pmatrix} a_0 & b_1 & & \\ b_1 & a_1 & \ddots & \\ & \ddots & \ddots & b_N \\ & & b_N & a_N \end{pmatrix} \tag{4.31}$$

(see Lemma D.1 for a proof). Moreover, we denote by T the infinite symmetric tridiagonal matrix on $\mathbb{N}_0$ with diagonal $(a_n)_{n \in \mathbb{N}_0}$ and off-diagonal $(b_n)_{n \in \mathbb{N}}$.

Remark 15. It will also prove convenient for us to renormalise the orthogonal polynomials by defining $P_n(x) := b_n p_n(x)$ and $b_0 := 1$; that is,

$$P_0(x) = 1, \quad P_1(x) = x - a_0, \quad \text{and} \tag{4.32}$$

$$P_{n+1}(x) = \frac{x - a_n}{b_n} P_n(x) - \frac{b_n}{b_{n-1}} P_{n-1}(x), \quad \text{for } n \geqslant 1 \tag{4.33}$$

$$b_{n+1}^2 = \int P_{n+1}(x)^2 \mathrm{d}D_\ell(x), \quad \text{and} \quad a_{n+1} = \frac{\int x P_{n+1}(x)^2 \mathrm{d}D_\ell(x),}{b_{n+1}^2}. \tag{4.34}$$

One advantage of this formulation is that it explicitly defines the coefficients $\{b_n\}$.

Therefore, if we have the first $2N+1$ moments $\mathcal{H}_{\ell\ell}, \dots, (\mathcal{H}^{2N+1})_{\ell\ell}$, it is possible to evaluate $Q_{2N+1}(\mathcal{H})_{\ell\ell}$ (that is, $\int Q_{2N+1} \mathrm{d}D_\ell$) for all polynomials Q_{2N+1} of degree at most $2N+1$, and thus compute T_N. In particular, for a fixed observable of interest O, we may write

$$\Theta_{2N+1}\big(\mathcal{H}_{\ell\ell}, \dots, [\mathcal{H}^{2N+1}]_{\ell\ell}\big) := O(T_N)_{00}. \tag{4.35}$$

Remark 16. In Appendix E we introduce more complex bond order potential (BOP) schemes based on the recursion method and show that they also satisfy Theorem 2.5.

4.3.2. Error estimates Equation (4.35) states that the nonlinear approximation scheme given by Θ_{2N+1} simply approximates the LDOS with the spectral measure of T_N corresponding to $\boldsymbol{e}_0 := (1, 0, \dots, 0)^{\mathrm{T}}$. We now show that $[(T_N)^n]_{00} = [T^n]_{00} = [\mathcal{H}^n]_{\ell\ell}$ for all $n \leqslant 2N+1$ and thus we may apply (2.14) to conclude.

By the orthogonality, we have $[T^0]_{ij} = \int p_i(x) x^0 p_j(x) \mathrm{d}D_\ell(x) = \delta_{ij}$. Therefore, assuming $[T^n]_{ij} = \int p_i(x) x^n p_j(x) \mathrm{d}D_\ell(x)$, we can conclude that

$$\begin{aligned}
[T^{n+1}]_{ij} &= \sum_k [T^n]_{ik} T_{kj} \\
&= \int p_i(x) x^n \big[b_j p_{j-1}(x) + a_j p_j(x) + b_{j+1} p_{j+1}(x)\big] \mathrm{d}D_\ell(x) \qquad (4.36) \\
&= \int p_i(x) x^{n+1} p_j(x) \mathrm{d}D_\ell(x). \qquad (4.37)
\end{aligned}$$

Here, we have applied (4.29) directly. In particular, if $i = j = 0$, we obtain $[T^n]_{00} = [\mathcal{H}^n]_{\ell\ell}$.

4.3.3. Analyticity To conclude the proof of Theorem 2.5, we show that Θ_{2N+1} as in (4.35) extents to an analytic function on some open set $U \subset \mathbb{C}^{2N+1}$. Throughout this section, we use the rescaled orthogonal polynomials $\{P_n\}$ from Remark 15.

For a polynomial $P(x) = \sum_{j=0}^m c_j x^j$, we use the notation $\mathcal{L}P(z_1, \dots, z_m) := c_0 + \sum_{j=1}^m c_j z_j$ for the linear function satisfying $P(x) = \mathcal{L}P(x, x^2, \dots, x^m)$. To extend the recurrence coefficients from (4.32), we start by defining

$$\begin{aligned}
&b_0 = 1, \quad a_0(z_1) := z_1, \quad P_1(x; z_1) := x - a_0(z_1) = x - z_1, \\
&b_1^2(z_1, z_2) := \mathcal{L}\big(x \mapsto P_1(x; z_1)^2\big)(z_1, z_2) = z_2 - z_1^2, \\
&a_1(z_1, z_2, z_3) := \frac{\mathcal{L}\big(x \mapsto x P_1(x; z_1)^2\big)(z_1, z_2, z_3)}{b_1^2(z_1, z_2)} = \frac{z_3 - 2z_1 z_2 + z_1^3}{z_2 - z_1^2}.
\end{aligned} \tag{4.38}$$

To simplify the notation, we write $z_{1:m}$ for the m-tuple $(z_1, \dots, z_m)$. Given $a_0(z_1), \dots, a_n(z_{1:2n+1})$ and $b_1(z_{1:2}), \dots, b_n(z_{1:2n})$, we define $P_{n+1}(x; z_{1:2n+1})$ to be the polynomial in x satisfying the same recursion as (4.32) but as a function of $z_{1:2n+1}$:

$$P_{n+1}(x; z_{1:2n+1}) = \frac{x - a_n(z_{1:2n+1})}{b_n(z_{1:2n})} P_n(x; z_{1:2n-1})$$

$$- \frac{b_n(z_{1:2n})}{b_{n-1}(z_{1:2n-2})} P_{n-1}(x; z_{1:2n-3}).$$

With this notation, we define

$$b_{n+1}^2(z_{1:2n+2}) := \mathcal{L}\big(x \mapsto P_{n+1}(x; z_{1:2n+1})^2\big)(z_{1:2n+2}), \tag{4.39}$$

$$a_{n+1}(z_{1:2n+3}) := \frac{\mathcal{L}\big(x \mapsto x P_{n+1}(x; z_{1:2n+1})^2\big)(z_{1:2n+3})}{b_{n+1}^2(z_{1:2n+2})}. \tag{4.40}$$

Since $P_{n+1}(x) = P_{n+1}(x; \mathcal{H}_{\ell\ell}, \ldots, [\mathcal{H}^{2n+1}]_{\ell\ell})$, we have extended the definition of the recursion coefficients (4.34) to functions of multiple complex variables.

We now show that $a_n(z_{1:2n+1})$ and $b_n^2(z_{1:2n})$ are rational functions. As a preliminary step, we show that both P_{n+1}^2 and $\frac{P_{n+1}P_n}{b_n}$ are polynomials in x with coefficients given by rational functions of a_n, b_n^2 and all previous recursion coefficients. This statement is clearly true for $n = 0$: $P_1^2 = (x - a_0)^2$ and $\frac{P_1 P_0}{b_0} = x - a_0$. Therefore, by induction and noting that

$$P_{n+1}^2 = \left(\frac{x - a_n}{b_n} P_n\right)^2 - 2(x - a_n)\frac{P_n P_{n-1}}{b_{n-1}} + \frac{b_n^2}{b_{n-1}^2} P_{n-1}^2 \tag{4.41}$$

$$\frac{P_{n+1}P_n}{b_n} = \frac{x - a_n}{b_n^2} P_n^2 - \frac{P_n P_{n-1}}{b_{n-1}}, \tag{4.42}$$

we can conclude. Therefore, by (4.38) and (4.39) and (4.40), we can apply another induction argument to conclude that $a_{n+1}(z_{1:2n+3})$ and $b_{n+1}^2(z_{1:2(n+1)})$ are rational functions.

We fix N and define the following complex valued tri-diagonal matrix

$$T_N(z) := \begin{pmatrix} a_0(z_1) & b_1^2(z_{1:2}) & & & \\ 1 & a_1(z_{1:3}) & b_2^2(z_{1:4}) & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & \ddots & b_N^2(z_{1:2N}) \\ & & & 1 & a_N(z_{1:2N+1}) \end{pmatrix}. \tag{4.43}$$

If $z_j = [\mathcal{H}^j]_{\ell\ell}$ for each $j = 1, \ldots, N$, (4.43) is similar to T_N from (4.31).

Now, on defining $U := \{z \in \mathbb{C}^{2N+1} : b_n^2(z_{1:2n}) \neq 0 \ \forall n = 1, \ldots, N\}$, the mapping $U \to \mathbb{C}^{(N+1)\times(N+1)}$ given by $z \mapsto T_N(z)$ is analytic. Therefore, for appropriately chosen contours $\mathscr{C}_z$ encircling $\sigma\big(T_N(z)\big)$, we have that

$$\Theta_{2N+1}(z) := O\big(T_N(z)\big)_{00} = -\frac{1}{2\pi i} \oint_{\mathscr{C}_z} O(\omega) \Big[\big(T_N(z) - \omega\big)^{-1}\Big]_{00} \mathrm{d}\omega. \tag{4.44}$$

In particular, Θ_{2N+1} is an analytic function on

$$\big\{z \in U : \ O \text{ analytic in an open neighbourhood of } \sigma\big(T_N(z)\big)\big\}.$$

Remark 17. Since $\mathbb{C}^{2N+1} \setminus U$ is the zero set for some (non-zero) polynomial P in $2N + 1$ variables, it has $(2N + 1)$-dimensional Lebesgue measure zero [45].

Remark 18. In Appendix D we show that the eigenvalues of $T_N(z)$ are distinct for z in some open neighbourhood, $U_0 \subset U$, of $\mathbb{R}^{2N+1}$, which leads to the following alternative proof. On U_0, the eigenvalues and corresponding left and right eigenvectors can be chosen to be analytic: there exist analytic functions $\varepsilon_j, \psi_j, \phi_j^\star$ for $j = 0, \dots, N$ such that

$$T_N(z)\psi_j(z) = \varepsilon_j(z)\psi_j(z), \quad \phi_j^\star(z)T_N(z) = \varepsilon_j(z)\phi_j^\star(z), \quad \text{and} \quad \phi_i^\star(z)\psi_j(z) = \delta_{ij}.$$

(More precisely, we apply [44, Theorem 2] to obtain analytic functions $\psi_j, \phi_j^\star$ of each variable $z_0, \dots, z_{2N+1}$ separately and then apply Hartog's theorem [60] to conclude that $\psi_j, \phi_j^\star$ are analytic as functions on $U \subset \mathbb{C}^{2N+1}$.) Therefore, the nonlinear method discussed in this section can also be written in the form

$$\Theta_{2N+1} = \sum_{j=0}^{N} [\psi_j]_0 [\phi_j^\star]_0 \cdot \big(O \circ \varepsilon_j\big), \tag{4.45}$$

which is an analytic function on $\{z \in U_0 \colon O \text{ analytic at } \varepsilon_j(z) \text{ for each } j\}$ (as it is a finite combination of analytic functions only involving products, compositions and sums).

4.4. Self-consistent tight binding models

We start with the following preliminary lemma:

Lemma 3. *Suppose that $T \colon \ell^2(\Lambda) \to \ell^2(\Lambda)$ is an invertible bounded linear operator with matrix entries $T_{\ell k}$ satisfying $\big|T_{\ell k}\big| \leqslant c_T e^{-\gamma_T r_{\ell k}}$ for some $c_T, \gamma_T > 0$.*

Then, there exists an invertible bounded linear operator $\overline{T} \colon \ell^\infty(\Lambda) \to \ell^\infty(\Lambda)$ extending $T \colon \ell^2(\Lambda) \to \ell^2(\Lambda)$ (that is, $\overline{T}\big|_{\ell^2(\Lambda)} = T$).

Proof. First, we denote the inverse of T and its matrix entries by $T^{-1} \colon \ell^2(\Lambda) \to \ell^2(\Lambda)$ and $T^{-1}_{\ell k}$, respectively. Then, applying the Combes-Thomas estimate to T yields the off-diagonal decay estimate $|T^{-1}_{\ell k}| \leqslant Ce^{-\gamma_{\mathrm{CT}} r_{\ell k}}$ for some $C, \gamma_{\mathrm{CT}} > 0$ [93].

Due to the off-diagonal decay properties of the matrix entries, the operators $\overline{T}, \overline{T}^{-1} \colon \ell^\infty(\Lambda) \to \ell^\infty(\Lambda)$ given by

$$[\overline{T}\phi]_\ell := \sum_{k \in \Lambda} T_{\ell k}\phi_k \qquad \text{and} \qquad [\overline{T}^{-1}\phi]_\ell := \sum_{k \in \Lambda} T^{-1}_{\ell k}\phi_k$$

are well defined bounded linear operators with norms $\sup_\ell \sum_{k\in\Lambda} |T_{\ell k}|$ and $\sup_\ell \sum_{k\in\Lambda} |T^{-1}_{\ell k}|$, respectively. To conclude, we note that

$$[\overline{T}\,\overline{T}^{-1}\phi]_\ell = \sum_k \sum_m T_{\ell k} T^{-1}_{km} \phi_m = \sum_m [TT^{-1}]_{\ell m}\phi_m = \phi_\ell \tag{4.46}$$

and so $\overline{T}^{-1}$ is the inverse of $\overline{T}$. Here, we have exchanged the summations over k and m by applying the dominated convergence theorem: $\big|\sum_k T_{\ell k} T^{-1}_{km}\phi_m\big| \leqslant Ce^{-\frac{1}{2}\min\{\gamma_T, \gamma_{\mathrm{CT}}\} r_{\ell m}} \|\phi\|_{\ell^\infty}$ is summable over $m \in \Lambda$. □

Throughout the following proofs, we denote by $B_r(\rho)$ the open ball of radius r about ρ with respect to the ℓ^∞-norm. Moreover, we briefly note that the stability operator can be written as the product $\mathscr{L}(\rho) := \mathscr{F}(\rho)\nabla w(\rho)$, where [93]

$$\mathscr{F}(\rho)_{\ell k} := \frac{1}{2\pi i}\oint_{\mathscr{C}} F^\beta(z)\Big[\big(\mathcal{H}(\boldsymbol{u}(\rho)) - z\big)^{-1}_{\ell k}\Big]^2 \mathrm{d}z, \tag{4.47}$$

where $\mathscr{C}$ is a simple closed contour encircling the spectrum $\sigma\big(\mathcal{H}(\boldsymbol{u}(\rho))\big)$.

Proof of Theorem 2.6. Since $\rho \mapsto F^\beta(\boldsymbol{u}(\rho))$ is C^2, and $\big(I - \mathscr{L}(\rho^\star)\big)^{-1}$ is a bounded linear operator, we necessarily have that $\big(I - \mathscr{L}(\rho)\big)^{-1}$ is a bounded linear operator for all $\rho \in B_r(\rho^\star)$ for some $r > 0$.

By applying Theorem 2.3, together with the assumption (**EP**), we obtain

$$\big|\big[\mathscr{L}(\rho) - \mathscr{L}_N(\rho)\big]_{\ell k}\big| \leqslant \sum_m \left|\left[\frac{\partial F^\beta_\ell(\boldsymbol{u})}{\partial v_m} - \frac{\partial I_{X_N} F^\beta_\ell(\boldsymbol{u})}{\partial v_m}\right]\frac{\partial w(\rho)_m}{\partial \rho_k}\right| \tag{4.48}$$

$$\leqslant C\Big[\sum_m e^{-\eta r_{\ell m}} e^{-\gamma_v r_{mk}}\Big] e^{-\frac{1}{2}\gamma_N N} \tag{4.49}$$

$$\leqslant C e^{-\frac{1}{2}\min\{\eta,\gamma_v\} r_{\ell k}} e^{-\frac{1}{2}\gamma_N N} \tag{4.50}$$

for all $\rho \in B_r(\rho^\star)$. As a direct consequence, we have $\|\mathscr{L}(\rho) - \mathscr{L}_N(\rho)\|_{\ell^2\to\ell^2} \leqslant C e^{-\frac{1}{2}\gamma_N N}$ and we may choose N sufficiently large such that $\|\mathscr{L}(\rho) - \mathscr{L}_N(\rho)\|_{\ell^2\to\ell^2} < \|(I - \mathscr{L}(\rho))^{-1}\|^{-1}_{\ell^2\to\ell^2}$. In particular, for such N, the operator $I - \mathscr{L}_N(\rho)\colon \ell^2 \to \ell^2$ is invertible with inverse bounded above in operator norm independently of N.

We now show that $I - \mathscr{L}_N(\rho)$ satisfies the assumptions of Lemma 3. Using (4.47) and (**EP**), together with the Combes-Thomas estimate (Lemma 1), we conclude that

$$\big|\mathscr{L}_N(\rho)_{\ell k}\big| \leqslant C \sup_{z\in\mathscr{C}} |I_{X_N} F^\beta(z)| \sum_{m\in\Lambda} e^{-2\gamma_{\mathrm{CT}} r_{\ell m}} e^{-\gamma_v r_{mk}} \leqslant C e^{-\frac{1}{2}\min\{2\gamma_{\mathrm{CT}},\gamma_v\} r_{\ell k}}$$

for all $\rho \in B_r(\rho^\star)$. In particular, $I - \mathscr{L}_N(\rho)$ extends to a invertible bounded linear operator $\ell^\infty \to \ell^\infty$ and thus its inverse $\big(I - \mathscr{L}_N(\rho)\big)^{-1}\colon \ell^\infty \to \ell^\infty$ is bounded.

Now, the mapping $\rho \mapsto \rho - I_{X_N} F^\beta\big(\boldsymbol{u}(\rho)\big)$ between $\ell^\infty \to \ell^\infty$ is continuously differentiable on $B_r(\rho^\star)$ and the derivative at $\rho^\star$ is invertible (i.e. $\big(I - \mathscr{L}_N(\rho^\star)\big)^{-1}\colon \ell^\infty \to \ell^\infty$ is a well defined bounded linear operator). Since the map $\rho \mapsto I_{X_N} F^\beta\big(\boldsymbol{u}(\rho)\big)$ is C^2, its derivative $\mathscr{L}_N$ is locally Lipschitz about $\rho^\star$ and so there exists $L > 0$ such that

$$\begin{aligned}&\big\|\big(I - \mathscr{L}_N(\rho^\star)\big)^{-1}\big(\mathscr{L}_N(\rho_1) - \mathscr{L}_N(\rho_2)\big)\big\|_{\ell^\infty\to\ell^\infty}\\ &\quad\leqslant L\|\rho_1 - \rho_2\|_{\ell^\infty} \qquad \text{for } \rho_1, \rho_2 \in B_r(\rho^\star).\end{aligned}$$

Moreover, by Theorem 2.3, we have that

$$\begin{aligned}&\big\|\big(I - \mathscr{L}_N(\rho^\star)\big)^{-1}\big(\rho^\star - I_{X_N} F^\beta(\boldsymbol{u}(\rho^\star))\big)\big\|_{\ell^\infty}\\ &\quad\leqslant C\big\|F^\beta(\boldsymbol{u}(\rho^\star)) - I_{X_N} F^\beta(\boldsymbol{u}(\rho^\star))\big\|_{\ell^\infty} =: b_N\end{aligned}$$

where $b_N \lesssim e^{-\gamma_N N}$. In particular, we may choose N sufficiently large such that $2b_N L < 1$ and $t_N^\star := \frac{1}{L}(1 - \sqrt{1 - 2b_N L}) < r$.

Thus, the Newton iteration with initial point $\rho^0 := \rho^\star$, defined by

$$\rho^{i+1} = \rho^i - \big(I - \mathscr{L}_N(\rho^i)\big)^{-1}\big(\rho^i - I_{X_N} F^\beta(\boldsymbol{u}(\rho^i))\big),$$

converges to a unique fixed point $\rho_N = I_{X_N} F^\beta(\boldsymbol{u}(\rho_N))$ in $B_{t_N^\star}(\rho^\star)$ [102,104]. That is, $\|\rho_N - \rho^\star\|_{\ell^\infty} \leqslant t_N^\star \leqslant 2b_N$. Here, we have used the fact that $1 - \sqrt{1-x} \leqslant x$ for all $0 \leqslant x \leqslant 1$.

Since $\rho_N \in B_r(\rho^\star)$, we have $I - \mathscr{L}_N(\rho_N)\colon \ell^2 \to \ell^2$ is invertible and thus Lemma 2.8 also holds. □

Proof of Proposition 2.9. We proceed in the same way as in the proof of Theorem 2.6. In particular, since ρ_N is stable, if $\|\rho^0 - \rho_N\|_{\ell^\infty}$ is sufficiently small, $(I - \mathscr{L}_N(\rho^0))^{-1}$ is a bounded linear operator on ℓ^2. Moreover, by the exact same argument as in the proof of Theorem 2.6, $I - \mathscr{L}_N(\rho^0)\colon \ell^\infty \to \ell^\infty$ defines an invertible bounded linear operator. Also, $I - \mathscr{L}_N(\rho)$ is Lipschitz in a neighbourhood about ρ^0 and

$$\begin{aligned}\big\|\big(I - \mathscr{L}_N(\rho^0)\big)^{-1}\big(\rho^0 - I_{X_N} F^\beta(\boldsymbol{u}(\rho^0))\big)\big\|_{\ell^\infty} &\leqslant C\big\|\rho^0 - \rho_N - \big(I_{X_N} F^\beta(\boldsymbol{u}(\rho^0)) \\ &\quad - I_{X_N} F^\beta(\boldsymbol{u}(\rho_N))\big)\big\|_{\ell^\infty} \\ &\leqslant C\|\rho^0 - \rho_N\|_{\ell^\infty}.\end{aligned}$$

Here, we have used that

$$\begin{aligned}&\big|I_{X_N} F_\ell^\beta(\boldsymbol{u}(\rho^0)) - I_{X_N} F_\ell^\beta(\boldsymbol{u}(\rho_N))\big| \\ &\quad= \frac{1}{2\pi}\Big|\oint_{\mathscr{C}} I_{X_N} F^\beta(z)\big[\mathscr{R}_z(\rho^0) - \mathscr{R}_z(\rho_N)\big]_{\ell\ell}\Big| \\ &\quad\leqslant C\sum_{k\in\Lambda} e^{-2\gamma_{\mathrm{CT}} r_{\ell k}}|v(\rho^0)_k - v(\rho_N)_k| \\ &\quad\leqslant C\sum_{k\in\Lambda} e^{-2\gamma_{\mathrm{CT}} r_{\ell k}}\Big|\int_0^1 \sum_{m\in\Lambda} \frac{\partial v(t\rho^0 + (1-t)\rho_N)_k}{\partial \rho_m}\big[\rho^0 - \rho_N\big]_m \mathrm{d}t\Big| \\ &\quad\leqslant C\sum_{m\in\Lambda} e^{-\frac{1}{2}\min\{2\gamma_{\mathrm{CT}},\gamma_v\} r_{\ell m}}\big|\big[\rho^0 - \rho_N\big]_m\big| \\ &\quad\leqslant C\|\rho^0 - \rho_N\|_{\ell^\infty}. \end{aligned} \tag{4.51}$$

Therefore, as long as $\|\rho^0 - \rho_N\|_{\ell^\infty}$ is sufficiently small, we may apply the Newton iteration starting from ρ^0 to conclude. □

Proof of Corollary 2.7. As a direct consequence of (4.51), we have that

$$\begin{aligned}\big|O_\ell^{\mathrm{sc}}(\boldsymbol{u}) - I_{X_N} O_\ell\big(\boldsymbol{u}(\rho_N)\big)\big| &\leqslant \big|O_\ell\big(\boldsymbol{u}(\rho^\star)\big) - I_{X_N} O_\ell\big(\boldsymbol{u}(\rho^\star)\big)\big| \\ &\quad + \big|I_{X_N} O_\ell\big(\boldsymbol{u}(\rho^\star)\big) - I_{X_N} O_\ell\big(\boldsymbol{u}(\rho_N)\big)\big| \\ &\leqslant C\big[e^{-\gamma_N N} + \|\rho_N - \rho^\star\|_{\ell^\infty}\big]\end{aligned}$$

$$\leqslant Ce^{-\gamma_N N}.$$

Here, we have applied the standard convergence result (Theorem 2.3) with fixed effective potential. □

Acknowledgements. We gratefully acknowledge stimulating discussions with Gábor Csányi, Simon Etter, and Jianfeng Lu.
JT is supported by EPSRC Grant EP/HO23364/1 (as part of the MASDOC DTC) and by EPSRC Grant EP/W522594/1. HC is supported by the Natural Science Foundation of China under grant 11971066. CO is supported by EPSRC Grant EP/R043612/1, Leverhulme Research Project Grant RPG-2017-191 and by the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number GR019381].

Appendix A. Notation

Here we summarise the key notation:

- Λ : finite or countable index set,
- $\boldsymbol{u}_\ell = (\boldsymbol{r}_\ell, v_\ell, Z_\ell)$: state of atom ℓ where $\boldsymbol{r}_\ell \in \mathbb{R}^d$ denotes the atomic position, v_ℓ the effective potential, and Z_ℓ the atomic species,
- $\boldsymbol{u} = \{\boldsymbol{u}_\ell\}_{\ell\in\Lambda}$: configuration,
- $\boldsymbol{r}_{\ell k} := \boldsymbol{r}_k - \boldsymbol{r}_\ell$ and $r_{\ell k} := |\boldsymbol{r}_{\ell k}|$: relative atomic positions,
- δ_{ij} : Kronecker delta ($\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 1$),
- Id_n : $n \times n$ identity matrix,
- $|\cdot|$: absolute value on $\mathbb{R}^d$ or $\mathbb{C}$,
- $|\cdot|$: Frobenius matrix norm on $\mathbb{R}^{n\times n}$,
- $\nabla h = (\nabla h_{ab})_{1\leqslant a,b\leqslant n}$: gradient of $h\colon \mathbb{R}^d \to \mathbb{R}^{n\times n}$,
- M^{T} : transpose of the matrix M,
- Tr : trace of an operator,
- $f \sim g$ as $x \to x_0 \in \mathbb{R}\cup\{\pm\infty\}$ or $\mathbb{C}\cup\{\infty\}$ if there exists an open neighbourhood N of x_0 and positive constants $c_1, c_2 > 0$ such that $c_1 g(x) \leqslant f(x) \leqslant c_2 g(x)$ for all $x \in N$,
- C : generic positive constant that may change from one line to the next,
- $f \lesssim g$: $f \leqslant Cg$ for some generic positive constant,
- $\mathbb{N}_0 = \{0, 1, 2, \dots\}$: Natural numbers including zero,

- $\delta(\cdot)$: Dirac delta, distribution satisfying $\int f(x)\mathrm{d}\delta(x) = f(0)$,
- $\|f\|_{L^\infty(X)} := \sup_{x\in X} |f(x)|$: sup-norm of f on X,
- $\mathrm{dist}(z, A) := \inf_{a\in A} |z - a|$: distance between $z \in \mathbb{C}$ and the set $A \subset \mathbb{C}$,
- $a + bS := \{a + bs : s \in S\}$,
- $[\psi]_\ell$: the ℓ^{th} entry of the vector ψ,
- $\|\psi\|_{\ell^2} := \left(\sum_k |[\psi]_k|^2\right)^{1/2}$: ℓ^2-norm of ψ,
- $\mathrm{tr}\, M := \sum_\ell M_{\ell\ell}$: trace of matrix M,
- $\|M\|_{\max} := \max_{\ell,k} |M_{\ell k}|$: max-norm of the matrix M,
- $\sigma(T)$: the spectrum of the operator T,
- $\sigma_{\mathrm{disc}}(T) \subset \sigma(T)$: isolated eigenvalues of finite multiplicity,
- $\sigma_{\mathrm{ess}}(T) := \sigma(T) \setminus \sigma_{\mathrm{disc}}(T)$: essential spectrum,
- $\|T\|_{X\to Y} := \sup_{x\in X, \|x\|_X=1} \|Tx\|_Y$: operator norm of $T : X \to Y$,
- ∇v : Jacobian of $v : \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda$,
- $[a, b] := \{(1 - t)a + tb : t \in [0, 1]\}$: closed interval between $a, b \in \mathbb{R}^d$ or $a, b \in \mathbb{C}$,
- $\int_a^b := \int_{[a,b]}$: integral over the interval $[a, b]$ for $a, b \in \mathbb{C}$,
- $\mathrm{len}(\mathscr{C})$: length of the simple closed contour $\mathscr{C}$,
- $\mathrm{supp}\, \nu$: support of the measure ν, set of all x for which every open neighbourhood of x has non-zero measure,
- $\mathrm{conv}\, A := \{ta + (1 - t)b : a, b \in A, t \in [0, 1]\}$: convex hull of A.

Appendix B. Locality: Truncation of the Atomic Environment

We have seen that analytic quantities of interest may be approximated by body-order approximations. However, each polynomial depends on the whole atomic configuration $\boldsymbol{u}$. In this section, we consider the truncation of the approximation schemes to a neighbourhood of the central site ℓ and prove the exponential convergence of the corresponding sparse representation.

B.1. Banded Approximation

One intuitive approach is to restrict the interaction range globally and consider the following *banded approximation*:

$$\widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u})_{km} := \begin{cases} h(\boldsymbol{u}_{km}) + \sum\limits_{\substack{m'\notin\{k,m\}:\\ r_{km'}, r_{mm'}\leqslant r_\mathrm{c}}} t(\boldsymbol{u}_{km'}, \boldsymbol{u}_{mm'}) + \delta_{km} v_k \mathrm{Id}_{N_\mathrm{b}} & \text{if } r_{km} \leqslant r_\mathrm{c} \\ 0 & \text{otherwise.} \end{cases} \tag{B.1}$$

Therefore, approximating $O_\ell(\boldsymbol{u})$ with a function depending on the first N moments $[\widetilde{\mathcal{H}}^{r_\mathrm{c}}]_{\ell\ell}$ (e.g. applying Theorem 2.4 or 2.5 to $\widetilde{\mathcal{H}}^{r_\mathrm{c}}$) results in an approximation scheme depending only on finitely many atomic sites in a neighbourhood of ℓ. This can be seen from the fact that

$$[(\widetilde{\mathcal{H}}^{r_\mathrm{c}})^n]_{\ell\ell} = \sum_{\substack{\ell_1,\dots,\ell_{n-1}\\ r_{\ell\ell_1}, r_{\ell_1\ell_2},\dots,r_{\ell_{n-1}\ell}\leqslant r_c}} \mathcal{H}_{\ell\ell_1}\mathcal{H}_{\ell_1\ell_2}\dots\mathcal{H}_{\ell_{n-1}\ell}. \tag{B.2}$$

Moreover, we obtain appropriate error estimates by combining Theorem 2.4 or 2.5 with the following estimate:

Proposition B.1. *Suppose* $\boldsymbol{u}$ *satisfies Definition 1. Fix* $0 < \beta \leqslant \infty$ *and suppose that, if* $\beta = \infty$, *then* $\mathsf{g}, \mathsf{g}^{\mathrm{def}} > 0$. *Then, we have*

$$\left| O_\ell(\boldsymbol{u}) - \operatorname{tr} O\big(\widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u})\big)_{\ell\ell} \right| \lesssim e^{-\frac{1}{2}\gamma_0 r_\mathrm{c}}.$$

Suppose $\gamma_N(r_\mathrm{c})$ *and* $\gamma_N^{\mathrm{def}}(r_\mathrm{c})$ *are the rates of approximation from Theorems 2.4 and 2.5 when applied to* $\widetilde{\mathcal{H}}^{r_\mathrm{c}}$. *Then* $\gamma_N(r_\mathrm{c}) \to \gamma_N$ *and* $\gamma_N^{\mathrm{def}}(r_\mathrm{c}) \to \gamma_N^{\mathrm{def}}$ *as* $r_\mathrm{c} \to \infty$, *with an exponential rate.*

Proof. We first note that

$$\big[\mathcal{H}(\boldsymbol{u}) - \widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u})\big]_{km} = \begin{cases} \mathcal{H}(\boldsymbol{u})_{km} & \text{if } r_{km} > r_\mathrm{c} \\ \sum_{\substack{m': \\ r_{km'} > r_\mathrm{c} \text{ or } r_{mm'} > r_\mathrm{c}}} t(\boldsymbol{u}_{km'}, \boldsymbol{u}_{mm'}) & \text{if } r_{km} \leqslant r_\mathrm{c}. \end{cases} \tag{B.3}$$

Therefore, applying (**TB**), we obtain

$$\big|[\mathcal{H}(\boldsymbol{u}) - \widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u})]_{km}\big| \lesssim e^{-\frac{1}{2}\gamma_0 r_\mathrm{c}} \sum_{m'} e^{-\frac{1}{2}\gamma_0 (r_{km'} + r_{mm'})} \lesssim e^{-\frac{1}{2}\gamma_0 r_\mathrm{c}} e^{-\frac{1}{4}\gamma_0 r_{km}}. \tag{B.4}$$

To conclude we choose a suitable contour $\mathscr{C}$ and apply the Combes-Thomas estimate (Lemma 1) together with (B.4):

$$\left| O_\ell(\boldsymbol{u}) - \operatorname{tr} O\big(\widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u})\big)_{\ell\ell} \right| \leqslant \left| \frac{\operatorname{tr}}{2\pi} \oint_{\mathscr{C}} O(z) \Big[(\mathcal{H}(\boldsymbol{u}) - z)^{-1} \big(\mathcal{H}(\boldsymbol{u}) \tag{B.5}$$

$$- \widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u})\big)(\widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u}) - z)^{-1} \Big]_{\ell\ell} \mathrm{d}z \Big| \tag{B.6}$$

$$\lesssim \max_{z \in \mathscr{C}} |O(z)| e^{-\frac{1}{2}\gamma_0 r_\mathrm{c}} \sum_{km} e^{-\gamma_{\mathrm{CT}}(r_{\ell k} + r_{m\ell})} e^{-\frac{1}{4}\gamma_0 r_{km}} \lesssim e^{-\frac{1}{2}\gamma_0 r_\mathrm{c}} \tag{B.7}$$

As a direct consequence of (B.4), we have also have $\|\mathcal{H}(\boldsymbol{u}) - \widetilde{\mathcal{H}}^{r_\mathrm{c}}(\boldsymbol{u})\|_{\ell^2 \to \ell^2} \lesssim e^{-\frac{1}{2}\gamma_0 r_\mathrm{c}}$ and so $\operatorname{dist}\big(\sigma(\mathcal{H}), \sigma(\widetilde{\mathcal{H}}^{r_\mathrm{c}})\big) \lesssim e^{-\frac{1}{2}\gamma_0 r_\mathrm{c}}$ [56]. This means that for sufficiently large r_c, we obtain the same rates of approximation when applying Theorems 2.4 and 2.5 to $\widetilde{\mathcal{H}}^{r_\mathrm{c}}$. □

B.2. Truncation

One downside of the banded approximation is that the truncation radius depends on the maximal polynomial degree (e.g. see (B.2)). In this section, we consider truncation schemes that only depend on finitely many atomic sites independent of the polynomial degree:

$$\widetilde{\mathcal{H}}^{r_\mathrm{c}} := \mathcal{H}\big|_{\ell;\, \Lambda \cap B_{r_\mathrm{c}}(\ell)} \tag{B.8}$$

where the restriction of the Hamiltonian has been introduced in (2.26).

On defining the quantities

$$I_{X_N} \widetilde{O}_\ell(\boldsymbol{u}) := \mathrm{tr}\big[I_{X_N} O\big(\widetilde{\mathcal{H}}^{r_c}\big)\big]_{\ell\ell}, \tag{B.9}$$

where the operators I_{X_N} are given by Theorem 2.3, we obtain a sparse representation of the N-body approximation depending only on finitely many atomic sites, independently of the maximal body-order N.

Proposition B.2. *Suppose $\boldsymbol{u}$ satisfies Definition 1. Fix $0 < \beta \leqslant \infty$ and suppose that, if $\beta = \infty$, then $\mathsf{g}, \mathsf{g}^{\mathrm{def}} > 0$. Then,*

$$\big|I_{X_N} O_\ell^\beta(\boldsymbol{u}) - I_{X_N} \widetilde{O}_\ell^\beta(\boldsymbol{u})\big| \lesssim e^{-\frac{1}{4}\min\{\gamma_{\mathrm{CT}}, \gamma_0\} r_c}$$

where $O^\beta = F^\beta$ or G^β and γ_{CT} is the constant from Lemma 1 applied to $\mathcal{H}(\boldsymbol{u})$.

Proof. Applying the Hermite integral formula (4.1) directly, we conclude that $I_{X_N} O^\beta(z)$ is bounded uniformly in N along a suitably chosen contour $\mathscr{C} := \{g_E = \gamma\}$ (examples of such contours are given in Figure 5). It is important to note that the contour $\mathscr{C}$ must be chosen to encircle both $\sigma(\mathcal{H})$ and $\sigma(\widetilde{\mathcal{H}}^{r_c})$.

In the following, we let γ_{CT} be the Combes-Thomas exponent from Lemma 1 corresponding to $\mathcal{H}$.

Similarly to (B.7), we obtain

$$\begin{aligned}
\Big|I_{X_N} O_\ell^\beta(\boldsymbol{u}) - I_{X_N} \widetilde{O}_\ell^\beta(\boldsymbol{u})\Big| &\lesssim \max_{z\in\mathscr{C}} |I_{X_N} O(z)| \sum_{km} e^{-\gamma_{\mathrm{CT}} r_{\ell k}} \big|\big[\mathcal{H}(\boldsymbol{u}) - \widetilde{\mathcal{H}}^{r_c}(\boldsymbol{u})\big]_{km}\big| \\
&\lesssim \sum_{\substack{k,m:\\ r_{\ell k}\geqslant r_c \text{ or } r_{\ell m}\geqslant r_c}} e^{-\gamma_{\mathrm{CT}} r_{\ell k}} e^{-\frac{1}{2}\gamma_0 r_{km}} \\
&\quad + \sum_{\substack{k,m:\\ r_{\ell k}, r_{\ell m} < r_c}} e^{-\gamma_{\mathrm{CT}} r_{\ell k}} \sum_{\substack{m':\\ r_{\ell m'}\geqslant r_c}} e^{-\gamma_0 (r_{km'} + r_{mm'})} \\
&\lesssim e^{-\frac{1}{2}\min\{\gamma_{\mathrm{CT}}, \frac{1}{2}\gamma_0\} r_c}.
\end{aligned} \tag{B.10}$$

This concludes the proof. □

The fact that the exponents of Proposition B.2 are independent of the defect states within the band gap is in the same spirit to the improved locality estimates of [17].

Remark 19. (Divide-and-conquer Methods) This truncation scheme is closely related to the divide-and-conquer method for solving the electronic structure problem [103]. In this context the system is split into many subsystems that are only related through a global choice of Fermi level. In our notation, this method consists of constructing N_{DAC} smaller Hamiltonians $\widetilde{\mathcal{H}}^{r_c,\ell_j}$ centred on the atoms ℓ_j (for $j = 1, \ldots, N_{\mathrm{DAC}}$) and approximating the quantities $O_\ell(\boldsymbol{u})$ for ℓ in a small neighbourhood of ℓ_j by calculating $\mathrm{tr}\, O\big(\widetilde{\mathcal{H}}^{r_c,\ell_j}\big)_{\ell\ell}$. That is, the eigenvalue problem for the whole system is approximated by solving N_{DAC} smaller eigenvalue problems

in parallel. In particular, this method leads to linear scaling algorithms [42]. Theorem B.2 then ensures that the error in this approximation decays exponentially with the distance between ℓ and the exterior of the subsystem centred on ℓ_j.

A similar error analysis in the context of divide-and-conquer methods in Kohn-Sham density functional theory can be found in [18].

Remark 20. (General truncation operators) It should be clear from the proof of Proposition B.2 that more general truncation operators may be used. Indeed, Proposition B.2 is satisfied for all truncation operators $\widetilde{\mathcal{H}}^{r_c} = \widetilde{\mathcal{H}}^{r_c}(\boldsymbol{u})$ satisfying the following conditions:

(T1) For every polynomial p, the quantity $p\big(\widetilde{\mathcal{H}}^{r_c}\big)_{\ell\ell}$ depends on at most finitely many atomic sites depending on r_c but not p,
(T2) For all $k, m \in \Lambda$, we have $[\widetilde{\mathcal{H}}^{r_c}]_{km} \to \mathcal{H}_{km}$ as $r_c \to \infty$,
(T3) There exists $c_0 > 0$ such that for all $\gamma, r_c > 0$,

$$\sum_{km} e^{-\gamma r_{\ell k}} \left|\left[\mathcal{H} - \widetilde{\mathcal{H}}^{r_c}\right]_{km}\right| \leqslant C e^{-c_0 \min\{\gamma_0, \gamma\} r_c}$$

for some $C > 0$ depending on γ but not on r_c.

Due to the exponential weighting of the summation, (T3) states that $\widetilde{\mathcal{H}}^{r_c}$ captures the behaviour of the Hamiltonian in a small neighbourhood of the site ℓ. Moreover, when making the approximation $I_{X_N} O\big(\mathcal{H}\big)_{\ell\ell} \approx I_{X_N} O\big(\widetilde{\mathcal{H}}^{r_c}\big)_{\ell\ell}$, the number of atomic sites involved is finite by (T1).

Remark 21. (Non-linear schemes) One may be tempted to approximate the Hamiltonian with the truncation, $\widetilde{\mathcal{H}}^{r_c}$, and then apply the nonlinear scheme of Theorem 2.5. In doing so, we obtain the following error estimates:

$$\left|O_\ell(\boldsymbol{u}) - \Theta_N\big([\widetilde{\mathcal{H}}^{r_c}]_{\ell\ell}, \ldots, [(\widetilde{\mathcal{H}}^{r_c})^N]_{\ell\ell}\big)\right| \tag{B.11}$$

$$\leqslant \left|O(\mathcal{H})_{\ell\ell} - O\big(\widetilde{\mathcal{H}}^{r_c}\big)_{\ell\ell}\right| + \left|O\big(\widetilde{\mathcal{H}}^{r_c}\big)_{\ell\ell} - \Theta_N\big([\widetilde{\mathcal{H}}^{r_c}]_{\ell\ell}, \ldots, [(\widetilde{\mathcal{H}}^{r_c})^N]_{\ell\ell}\big)\right| \tag{B.12}$$

$$\lesssim e^{-\frac{1}{4}\min\{\gamma_0, \gamma_{\mathrm{CT}}\} r_c} + e^{-\widetilde{\gamma}_N(r_c) N}. \tag{B.13}$$

A problem with this analysis is that the constant $\widetilde{\gamma}_N(r_c)$ in (B.13) arises by applying Theorem 2.5 to $\widetilde{\mathcal{H}}^{r_c}$ rather than the original system $\mathcal{H}$. In particular, this means that $\widetilde{\gamma}_N(r_c)$ depends on the spectral properties of $\widetilde{\mathcal{H}}^{r_c}$ rather than $\mathcal{H}$. Since spectral pollution is known to occur when applying naive truncation schemes [64], the choice of $\widetilde{\mathcal{H}}^{r_c}$ is important for the analysis. In particular, it is not clear that $\widetilde{\gamma}_N(r_c) \to \gamma_N$ in general. This is in contrast the the result of Proposition B.1.

Appendix C. Convergence of Derivatives in the Nonlinear Approximation Scheme

As mentioned in Remark 10, the results of this section depend on the "regularity" properties of D_ℓ:

Definition 2. (Regular n^{th}-root Asymptotic Behaviour) For a unit measure ν with compact support $E := \operatorname{supp}\nu \subset \mathbb{R}$, we say ν is regular and write $\nu \in \mathbf{Reg}$ if the corresponding sequence of orthonormal polynomials $\{p_n(\cdot\,;\nu)\}$ satisfy

$$\lim_{n\to\infty} |p_n(z;\nu)|^{\frac{1}{n}} = e^{g_E(z)}$$

locally uniformly on $\mathbb{C} \setminus \operatorname{conv}(E)$.

Remark 22. The regularity condition says that the n^{th}-root asymptotic behaviour of $|p_n(z;\nu)|$ is minimal: in general, we have [85, Theorem 1.1.4]

$$e^{g_E(z)} \leqslant \liminf_{n\to\infty} |p_n(z;\nu)|^{\frac{1}{n}} \leqslant \limsup_{n\to\infty} |p_n(z;\nu)|^{\frac{1}{n}} \leqslant e^{g_\nu(z)}$$

where $g_\nu \geqslant g_E$ is the *minimal carrier Green's function of* ν [85].

Under the regularity condition of Definition 2, we obtain results analogous to (2.21):

Theorem C.1. *Suppose that* $\boldsymbol{u}$ *satisfies Definition 1 and* $\ell \in \Lambda$ *is such that* $D_\ell \in$ **Reg**. *Then, with the notation of Theorem 2.5, we in addition have*

$$\left| \frac{\partial}{\partial \boldsymbol{u}_m} \Big(O_\ell^\beta(\boldsymbol{u}) - \Theta_N\big(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \dots, [\mathcal{H}^N]_{\ell\ell}\big)\Big) \right| \lesssim e^{-\frac{1}{2}\gamma_N N} e^{-\eta\, r_{\ell m}}.$$

More generally, if the regularity assumption is not satisfied, it may still be the case that Theorem C.1 holds but with reduced locality exponent η. To formulate this result, we require the notion of *minimal carrier capacity*:

Definition 3. (Minimal carrier capacity) For arbitrary Borel sets C, the *capacity of* C is defined as

$$\operatorname{cap}(C) := \sup\{\operatorname{cap}(K)\colon K \subset C, \text{compact}\},$$

where $\operatorname{cap}(K)$ is defined as in § 4.1.5.

For a unit measure ν with compact support $E := \operatorname{supp}\nu \subset \mathbb{R}$, the set of *carriers of* ν and the *minimal carrier capacity* are defined as

$$\Gamma(\nu) := \{C \subset \mathbb{C}\colon C \text{ Borel and } \nu(\mathbb{C}\setminus C) = 0\}, \quad \text{and} \tag{C.1}$$

$$c_\nu := \inf\{\operatorname{cap}(C)\colon C \in \Gamma(\nu), C \text{ bounded}\} \leqslant \operatorname{cap}(E), \tag{C.2}$$

respectively.

Under these definitions, we have the following [85, p. 8–10]:

Remark 23. For a unit measure ν with compact support $E := \operatorname{supp}\nu \subset \mathbb{R}$, we have

(i) The set of *minimal carriers* $\Gamma_0(\nu) := \{C \in \Gamma(\nu)\colon \operatorname{cap}(C) = c_\nu, C \subset E\}$ is nonempty,

(ii) If $c_\nu > 0$, then there exists a *minimial carrier equilibrium distribution* ω_ν, a (uniquely defined) unit measure with $\operatorname{supp}\omega_\nu \subset E$ satisfying

$$g_\nu(z) = -\int \log\frac{1}{|z-t|}\mathrm{d}\omega_\nu(t) - \log c_\nu,$$

(iii) $g_\nu \equiv g_E$ if and only if $c_\nu = \mathrm{cap}(E)$,

(iv) In particular, if $c_\nu = \mathrm{cap}(E)$, then $\nu \in \mathbf{Reg}$ (although the converse is false [85, Example 1.5.4]),

(v) Suppose $c_\nu > 0$. Then, on defining ν_n to be the discrete unit measure giving equal weight to each of the zeros of $p_n(\,\cdot\,;\nu)$, the condition that

$$\nu_n \rightharpoonup^\star \omega_E,$$

where ω_E is the equilibrium distribution for E, is equivalent to $\nu \in \mathbf{Reg}$ [85, Thm. 3.1.4]. In particular, this justifies (4.10).

We therefore arrive at the corresponding result for $\ell \in \Lambda$ for which the corresponding LDOS has positive minimal carrier capacity:

Proposition C.2. *Suppose that* $\boldsymbol{u}$ *satisfies Definition* 1 *and* $\ell \in \Lambda$ *such that* $c_{D_\ell} > 0$. *Then, with the notation of Theorem* 2.5, *we in addition have*

$$\left| \frac{\partial}{\partial \boldsymbol{u}_m} \Big(O_\ell^\beta(\boldsymbol{u}) - \Theta_N\big(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \ldots, [\mathcal{H}^N]_{\ell\ell}\big)\Big) \right| \lesssim e^{-\frac{1}{2}\gamma_N N} e^{-\eta_\ell r_{\ell m}}$$

where $\eta_\ell > 0$,

$$\eta_\ell \to \eta \quad \text{as } c_{D_\ell} \to \mathrm{cap}(\mathrm{supp}\, D_\ell),$$

and $\eta > 0$ *is the constant from Theorem* C.1.

The proofs of Theorem C.1 and Proposition C.2 follow from the following estimates on the derivatives of the recursion coefficients $\{a_n, b_n\}$, and the locality of the tridiagonal operators T_N, together with the asymptotic upper bounds (i.e. Definition 2 or Remark 22).

Lemma C.3. *Suppose* $\boldsymbol{u}$ *satisfies Definition* 1. *Then, for a simple closed positively oriented contour* $\mathscr{C}'$ *encircling the spectrum* $\sigma\big(\mathcal{H}(\boldsymbol{u})\big)$, *there exists* $\eta = \eta(\mathscr{C}') > 0$ *such that*

$$\left|\frac{\partial b_n}{\partial \boldsymbol{u}_m}\right| \leqslant C \|p_n\|^2_{L^\infty(\mathscr{C}')} e^{-\eta r_{\ell m}} \quad \text{and} \quad \left|\frac{\partial a_n}{\partial \boldsymbol{u}_m}\right| \leqslant C \sum_{l=0}^{n} \|p_l\|^2_{L^\infty(\mathscr{C}')} e^{-\eta r_{\ell m}} \tag{C.3}$$

where $\eta \sim \mathfrak{d}$ *as* $\mathfrak{d} \to 0$ *where* $\mathfrak{d} := \mathrm{dist}\big(\mathscr{C}', \sigma\big(\mathcal{H}(\boldsymbol{u}^{\mathrm{ref}})\big)\big)$.

In the following, we denote by T_∞ the infinite symmetric matrix on $\mathbb{N}_0$ with diagonal $(a_n)_{n\in\mathbb{N}_0}$ and off-diagonal $(b_n)_{n\in\mathbb{N}}$.

Lemma C.4. *Fix* $N \in \mathbb{N} \cup \{\infty\}$. *Suppose that* $z \in \mathbb{C}$ *with* $\mathfrak{d}_N := \mathrm{dist}\big(z, \sigma(T_N)\big) > 0$. *Then, for each* $i, j \in \mathbb{N}_0$, *we have*

$$\left|(T_N - z)^{-1}_{ij}\right| \leqslant C e^{-\gamma_{|i-j|,N}|i-j|}.$$

(i) For each $r \in \mathbb{N}$, *we have* $\gamma_{r,N} \sim \mathfrak{d}_N$ *as* $\mathfrak{d}_N \to 0$.

(ii) We have $\lim_{r\to\infty} \gamma_{r,\infty} = \lim_{N\to\infty} \gamma_{N,N} = g_{\sigma(T_\infty)}(z)$ *where* $g_{\sigma(T_\infty)}$ *is the Green's function for the set* $\sigma(T_\infty)$ *as defined in* (4.10).

Remark 24. The fact that $g_{\sigma(T_\infty)}$ does not depend on the discrete eigenvalues of T_∞ means that asymptotically the locality estimates do not depend on defect states in the band gap arising due to perturbations satisfying Proposition 2.1, for example. Indeed, this has been shown more generally for operators with off-diagonal decay [17]. We show an alternative proof using logarithmic potential theory.

We will assume Lemmas C.4 and C.3 for now and return to their proofs below.

We first add on a constant multiple of the identity, cI, to the operators $\{T_N\}$ so that the spectra are contained in an interval bounded away from $\{0\}$. Moreover, we translate the integrand by the same constant: $\widetilde{O}(z) := O(z-c)$. Then, we extend T_N to an operator on $\ell^2(\mathbb{N}_0)$ by defining $[T_N\psi]_i = \sum_{j=0}^N [T_N]_{ij}\psi_j$ for $0 \leqslant i \leqslant N$ and $[T_N\psi]_i = 0$ otherwise. We therefore choose a simple closed contour (or system of contours) $\mathscr{C}$ encircling $\bigcup_N \sigma(T_N)$ so that

$$\begin{aligned}
&\frac{\partial\big[O_\ell(\boldsymbol{u}) - O(T_N)_{00}\big]}{\partial \boldsymbol{u}_m} \\
&\quad= \frac{1}{2\pi i}\oint_{\mathscr{C}} \widetilde{O}(z)\frac{\partial}{\partial \boldsymbol{u}_m}\Big[(T_\infty - z)^{-1}_{0,N+1} b_{N+1}(T_N - z)^{-1}_{N0}\Big]\mathrm{d}z \\
&\quad= \frac{1}{2\pi i}\oint_{\mathscr{C}} \widetilde{O}(z)\Big[\Big[(T_\infty - z)^{-1}\frac{\partial T_\infty}{\partial \boldsymbol{u}_m}(T_\infty - z)^{-1}\Big]_{0,N+1} b_{N+1}(T_N - z)^{-1}_{N0} \\
&\qquad + (T_\infty - z)^{-1}_{0,N+1}\frac{\partial b_{N+1}}{\partial \boldsymbol{u}_m}(T_N - z)^{-1}_{N0} \\
&\qquad + (T_\infty - z)^{-1}_{0,N+1} b_{N+1}\Big[(T_N - z)^{-1}\frac{\partial T_N}{\partial \boldsymbol{u}_m}(T_N - z)^{-1}\Big]_{N0}\Big]\mathrm{d}z. \qquad \text{(C.4)}
\end{aligned}$$

Therefore, applying Lemma C.3, a simple calculation reveals that

$$\begin{aligned}
&\left|\frac{\partial\big[O_\ell(\boldsymbol{u}) - O(T_N)_{00}\big]}{\partial \boldsymbol{u}_m}\right| \\
&\quad\leqslant C\sum_{n=0}^{\infty}\Big[\Big|\frac{\partial a_n}{\partial \boldsymbol{u}_m}\Big| + \Big|\frac{\partial b_n}{\partial \boldsymbol{u}_m}\Big|\Big]e^{-\min\{\gamma_{n,N},\gamma_{n,\infty}\}n}e^{-\min\{\gamma_{N,N},\gamma_{N+1,\infty}\}N} \\
&\quad\leqslant C\Big[\sum_{n=0}^{\infty}\sum_{l=0}^{n}\|p_l\|^2_{L^\infty(\mathscr{C}')}e^{-\min\{\gamma_{n,N},\gamma_{n,\infty}\}n}\Big]e^{-\min\{\gamma_{N,N},\gamma_{N+1,\infty}\}N}e^{-\eta r_{\ell m}}, \qquad \text{(C.5)}
\end{aligned}$$

where $\gamma_{r,N} = \gamma_{r,N}(\mathscr{C})$ is the constant from Lemma C.4. We therefore may conclude by choosing $\mathscr{C}' := \{g_E = \gamma\}$ if $D_\ell \in \mathbf{Reg}$ and $\mathscr{C}' := \{g_{D_\ell} = \gamma\}$ otherwise for some constant $\gamma > 0$ sufficiently small such that the summation in the square brackets converges.

Proof of Lemma C.3. The proof follows from the following identities:

$$\frac{\partial(b_n^2)}{\partial \boldsymbol{u}_m} = \oint_{\mathscr{C}} b_n^2 p_n(z)^2\Big[(\mathcal{H} - z)^{-1}\frac{\partial \mathcal{H}(\boldsymbol{u})}{\partial \boldsymbol{u}_m}(\mathcal{H} - z)^{-1}\Big]_{\ell\ell}\frac{\mathrm{d}z}{2\pi i} \quad \text{and} \qquad \text{(C.6)}$$

$$\frac{\partial a_n}{\partial \boldsymbol{u}_m} = \oint_{\mathscr{C}} \Big((z-a_n)p_n(z)^2 + \sum_{k=0}^{n-1} (-1)^{n-k}(2z-3a_k)p_k(z)^2 \Big) \Big[(\mathcal{H}-z)^{-1} \frac{\partial \mathcal{H}(\boldsymbol{u})}{\partial \boldsymbol{u}_m} (\mathcal{H}-z)^{-1} \Big]_{\ell\ell} \frac{\mathrm{d}z}{2\pi i}. \tag{C.7}$$

To do this, it will be convenient to renormalise the orthogonal polynomials as in Remark 15 (that is, we consider $P_n(x) := b_n p_n(x)$). Moreover, we define $b_{-1} := 1$. Using the shorthand $\partial := \frac{\partial}{\partial \boldsymbol{u}_m}$, we therefore obtain: $\partial b_{-1} = \partial b_0 = 0$, $\partial P_{-1}(x) = \partial P_0(x) = 0$, and

$$\partial P_{n+1}(x) = \frac{x-a_n}{b_n}\partial P_n(x) - \frac{b_n}{b_{n-1}}\partial P_{n-1}(x) - \partial\Big(\frac{a_n}{b_n}\Big)P_n(x) - \partial\Big(\frac{b_n}{b_{n-1}}\Big)P_{n-1}(x) \tag{C.8}$$

for all $n \geqslant 0$.

By noting $\partial P_1(x) = -\partial a_0$ and applying (C.8), we can see that ∂P_n is a polynomial of degree $n-1$ for all $n \geqslant 0$. Therefore, since P_n is orthogonal to all polynomials of degree $n-1$, we have

$$\partial(b_n^2) = 2\int P_n(x)\partial P_n(x)\mathrm{d}D_\ell + \oint_{\mathscr{C}} P_n(z)^2 \Big[(\mathcal{H}-z)^{-1} \frac{\partial \mathcal{H}}{\partial \boldsymbol{u}_m} (\mathcal{H}-z)^{-1} \Big]_{\ell\ell} \frac{\mathrm{d}z}{2\pi i} \tag{C.9}$$

$$= \oint_{\mathscr{C}} P_n(z)^2 \Big[(\mathcal{H}-z)^{-1} \frac{\partial \mathcal{H}}{\partial \boldsymbol{u}_m} (\mathcal{H}-z)^{-1} \Big]_{\ell\ell} \frac{\mathrm{d}z}{2\pi i} \tag{C.10}$$

which concludes the proof of (C.7).

To prove a similar formula for the derivatives of a_n, we first state a useful identity which will be proved after the conclusion of the proof of (C.7):

$$x\partial P_n(x) = \sum_{k=0}^{n} c_{nk} P_k(x), \quad \text{where} \quad c_{nn} = \sum_{k=0}^{n-1} \Big(a_k \frac{\partial b_k}{b_k} - \partial a_k \Big). \tag{C.11}$$

Therefore, we have that

$$\begin{aligned} \partial a_n &= \frac{1}{b_n^2}\Big(\oint_{\mathscr{C}} z P_n(z)^2 \Big[(\mathcal{H}-z)^{-1} \frac{\partial \mathcal{H}}{\partial \boldsymbol{u}_m} (\mathcal{H}-z)^{-1} \Big]_{\ell\ell} \frac{\mathrm{d}z}{2\pi i} \\ &\quad + 2\int x P_n(x)\partial P_n(x)\mathrm{d}D_\ell(x) \Big) - a_n \frac{\partial(b_n^2)}{b_n^2} \\ &= \frac{1}{b_n^2} \oint_{\mathscr{C}} (z-a_n) P_n(z)^2 \Big[(\mathcal{H}-z)^{-1} \frac{\partial \mathcal{H}}{\partial \boldsymbol{u}_m} (\mathcal{H}-z)^{-1} \Big]_{\ell\ell} \frac{\mathrm{d}z}{2\pi i} \\ &\quad + \sum_{k=0}^{n-1} \Big(a_k \frac{\partial(b_k^2)}{b_k^2} - 2\partial a_k \Big). \end{aligned} \tag{C.12}$$

Applying (C.11) for $k \leqslant n-1$, we can see that ∂a_n can be written as

$$\partial a_n = \oint_{\mathscr{C}} \Big((z-a_n)p_n(z)^2$$

$$+\sum_{k=0}^{n-1}(d_{1,k}z+d_{0,k})p_k(z)^2\Big)\Big[(\mathcal{H}-z)^{-1}\frac{\partial\mathcal{H}}{\partial \boldsymbol{u}_m}(\mathcal{H}-z)^{-1}\Big]_{\ell\ell}\frac{\mathrm{d}z}{2\pi i}.$$

for some coefficients $d_{1,k}$, $d_{0,k}$. Using (C.11) and assuming the result for $k \leqslant n-1$, we have

$$d_{1,k}z+d_{0,k} = a_k - 2(z-a_k) - 2\sum_{l=k+1}^{n-1}(-1)^{l-k}(2z-3a_k) \tag{C.13}$$

$$= -2z+3a_k-(-1)^k\big((-1)^{k+1}+(-1)^{n-1}\big)(2z-3a_k) \tag{C.14}$$

$$= (-1)^{n-k}(2z-3a_k). \tag{C.15}$$

for all $k \leqslant n-1$. □

Proof of (C.11). We have that

$$x\partial P_n(x) = \frac{x}{b_{n-1}}x\partial P_{n-1}(x) - \partial\Big(\frac{a_{n-1}}{b_{n-1}}\Big)xP_{n-1}(x) + \text{l.o.t.} \tag{C.16}$$

$$= \frac{1}{b_{n-1}}c_{n-1,n-1}xP_{n-1}(x) - \partial\Big(\frac{a_{n-1}}{b_{n-1}}\Big)b_{n-1}P_n(x) + \text{l.o.t.} \tag{C.17}$$

$$= c_{n-1,n-1}P_n(x) - \partial\Big(\frac{a_{n-1}}{b_{n-1}}\Big)b_{n-1}P_n(x) + \text{l.o.t,} \tag{C.18}$$

where l.o.t. ("lower order term") denotes a polynomial of degree strictly less than n that changes from one line to the next. That is, since $c_{11} = -\partial a_0 = \partial\big(\frac{a_0}{b_0}\big)b_0$, we apply an inductive argument to conclude that

$$c_{nn} = c_{n-1,n-1} - \partial\Big(\frac{a_{n-1}}{b_{n-1}}\Big)b_{n-1} = -\sum_{k=0}^{n-1}\partial\Big(\frac{a_k}{b_k}\Big)b_k = \sum_{k=0}^{n-1}\Big(a_k\frac{\partial b_k}{b_k} - \partial a_k\Big)$$

$$= \sum_{k=0}^{n-1}\Big(\frac{a_k}{2}\frac{\partial(b_k^2)}{b_k^2} - \partial a_k\Big).$$

□

Proof of Lemma C.4. The first statement is the Combes-Thomas resolvent estimate (Lemma 1) for tridiagonal operators (which, in particular, satisfy the off-diagonal decay assumptions of Lemma 1).

To obtain the asymptotic estimates of *(ii)*, we apply a different approach based on the banded structure of the operators. Since T_N is tri-diagonal, $[(T_N)^n]_{ij} = 0$ if $|i-j| > n$. Therefore, for any polynomial P of degree at most $|i-j|-1$, we have [9]

$$|(T_N-z)^{-1}_{ij}| = \Big|\big[(T_N-z)^{-1} - P(T_N)\big]_{ij}\Big| \leqslant \Big\|\frac{1}{\cdot - z} - P\Big\|_{L^\infty(\sigma(T_N))}. \tag{C.19}$$

We may apply the results of logarithmic potential theory (see (4.15)), to conclude. Here, it is important that $|\sigma(T_\infty)\setminus\sigma(T_N)|$ remains bounded independently of N so that, asymptotically, (C.19) has exponential decay with exponent $g_{\sigma(T_\infty)}$.

The proof that $|\sigma(T_\infty)\setminus\sigma(T_N)|$ is uniformly bounded can easily be shown when considering the sequence of orthogonal polynomials generated by T_∞. A full proof is given in parts *(ii)* and *(iv)* of Lemma D.1. □

Appendix D. Quadrature Method

The quadrature method as outlined in this section was introduced in [69] to approximate the LDOS. For a comparison of various nonlinear approximation schemes, see [51] and [41]. The former is a practical comparison of quadrature and BOP methods, while the later also discusses the maximum entropy method [67].

We now give an alternative proof of Theorem 2.5 by introducing the quadrature method [69].

Recall that D_ℓ is the local density of states (LDOS) satisfying (2.13) and $\{p_n\}$ is the corresponding sequence of orthogonal polynomials generated via the recursion method:

$$[p_n(\mathcal{H})p_m(\mathcal{H})]_{\ell\ell} = \int p_n(x)p_m(x)\mathrm{d}D_\ell(x) = \delta_{nm}$$

(see the proof of Lemma D.1, below).

We use the set of zeros of p_{N+1}, denoted by $X_N = \{\varepsilon_0, \ldots, \varepsilon_N\}$, as the basis for the following quadrature rule:

$$O(\mathcal{H})_{\ell\ell} = \int O(x)\mathrm{d}D_\ell(x) \approx \int I_{X_N}O(x)\mathrm{d}D_\ell(x) = \sum_{j=0}^{N} w_j O(\varepsilon_j), \qquad \text{where}$$

$$w_j = \int \ell_j(x)\mathrm{d}D_\ell(x) = \ell_j(\mathcal{H})_{\ell\ell}, \quad \text{and} \quad \ell_j(x) = \prod_{i\neq j}\frac{x-\varepsilon_i}{\varepsilon_j-\varepsilon_i}. \tag{D.1}$$

Here, ℓ_j is the polynomial of degree N with $\ell_j(\varepsilon_i) = \delta_{ij}$.

The following lemma highlights the fundamental properties of Gauss quadrature and allows us to show that the approximation scheme given by

$$\Theta^{\mathrm{q}}_{2N+1}\left(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \ldots, [\mathcal{H}^{2N+1}]_{\ell\ell}\right) := \sum_{j=0}^{N} w_j O(\varepsilon_j). \tag{D.2}$$

satisfies Theorem 2.5.

Lemma D.1. *Suppose that $\{p_n\}$ is the sequence of polynomials generated by the recursion method (4.29), X_N is the set of zeros of p_{N+1}, and $\{w_j\}$ are the weights satisfying $\int I_{X_N}O(x)\mathrm{d}D_\ell(x) = \sum_{j=0}^{N} w_j O(\varepsilon_j)$. Then,*

(i) $\{p_n\}$ is orthonormal with respect to D_ℓ: $\int p_n(x)p_m(x)\mathrm{d}D_\ell(x) = [p_n(\mathcal{H})p_m(\mathcal{H})]_{\ell\ell} = \delta_{nm}$,
(ii) $X_N = \sigma(T_N)$ where T_N is given by (4.31),
(iii) $X_N \subset \mathbb{R}$ is a set of $N+1$ distinct points,
(iv) If $[a,b] \cap \operatorname{supp} D_\ell = \emptyset$, then the number of points in $X_N \cap [a,b]$ is at most one,
(v) If P_{2N+1} is a polynomial of degree at most $2N+1$, then $P_{2N+1}(\mathcal{H})_{\ell\ell} = \sum_{j=0}^{N} w_j P_{2N+1}(\varepsilon_j)$,
(vi) The weights $\{w_j\}$ are positive and sum to one.

Proof. The idea behind the proofs are standard in the theory of Gauss quadrature (e.g. see [40]) but, for the convenience of the reader, they are collected together in D.3. □

Remark 25. The quadrature rule discussed in this section can be seen as the exact integral with respect to the following approximate LDOS

$$D_\ell^{2N+1,\mathrm{q}} := \sum_{j=0}^{N} w_j \delta(\cdot - \varepsilon_j).$$

This measure has unit mass by Lemma D.1 *(vi)*, and, by Lemma D.1 *(v)*, the first $2N+1$ moments of $D_\ell^{2N+1,\mathrm{q}}$ are given by $[\mathcal{H}^n]_{\ell\ell}$ for $n = 1, \dots, 2N+1$.

In the following two sections we prove error estimates and show that the functional form is analytic on an open set containing $\big([\mathcal{H}]_{\ell\ell}, \dots, [\mathcal{H}^{2N+1}]_{\ell\ell}\big)$.

D.1. Error Estimates.

Applying Remark 25, together with (2.14), we have: for every polynomial P_{2N+1} of degree at most $2N+1$,

$$\left| O_\ell(\boldsymbol{r}) - \sum_{j=0}^{N} w_j O(\varepsilon_j) \right| \leqslant 2 \| O - P_{2N+1} \|_{L^\infty(\sigma(\mathcal{H}) \cup X_N)}. \tag{D.3}$$

Now, since $\sigma\big(\mathcal{H}\big) \subset I_- \cup \{\lambda_j\} \cup I_+$ where $\{\lambda_j\}$ is a finite set, we may apply part *(iv)* of Lemma D.1 to conclude that the number of points in $X_N \setminus \big(I_- \cup I_+\big)$ is bounded independently of N. Accordingly, we may apply (4.15) with $E = I_- \cup I_+$, to obtain the following asymptotic bound

$$\lim_{N\to\infty} \left| O_\ell(\boldsymbol{r}) - \sum_{j=0}^{N} w_j O(\varepsilon_j) \right|^{1/(2N+1)} \leqslant e^{-\gamma^\star}$$
$$\text{where} \quad O \text{ is analytic on } \{z \colon g_E(z) < \gamma^\star\}.$$

In particular, we obtain the stated asymptotic behaviour.

Remark 26. (Spectral pollution) While $\sigma(\mathcal{H}) \subset \liminf_{N\to\infty} \sigma(T_N)$, we do not claim that the sequence $\sigma(T_N)$ is free from spurious eigenvalues. That is, there may exist sequences $\lambda_N \in \sigma(T_N)$ such that $\lambda_N \to \lambda$ along a subsequence and $\lambda \notin \sigma(\mathcal{H})$. Indeed, there exist measures supported on a union of disjoint intervals $[a,b] \cup [c,d]$ for which the corresponding sequences of orthogonal polynomials suffer from spurious eigenvalues at every point of the gap (b,c) [24,85]. In this paper, we only require the much milder property that the number of eigenvalues in the gap remains uniformly bounded in the limit $N \to \infty$.

For a more general discussion of spectral pollution, see [13,64].

D.2. Analyticity.

To conclude the proof of Theorem 2.5, we show that $\Theta^{\mathrm{q}}_{2N+1}$ as defined in (D.2) is analytic in a neighbourhood of $(\mathcal{H}_{\ell\ell}, [\mathcal{H}^2]_{\ell\ell}, \dots, [\mathcal{H}^{2N+1}]_{\ell\ell})$. Recall that in (4.43) we have extended the definition of T_N to an analytic function on $U := \{\boldsymbol{z} \in \mathbb{C}^{2N+1} : b_n^2(\boldsymbol{z}_{1:2n}) \neq 0 \ \forall n = 1, \dots, N\}$.

We define $X_N(\boldsymbol{z})$ to be the set of eigenvalues of $T_N(\boldsymbol{z})$. Since $X_N = X_N\big(\mathcal{H}_{\ell\ell}, \dots, [\mathcal{H}^{2N+1}]_{\ell\ell}\big)$ is a set of $N+1$ distinct points (Lemma D.1 *(iii)*), there exists a continuous choice of eigenvalues $X_N(\boldsymbol{z}) = \{\varepsilon_0(\boldsymbol{z}), \dots, \varepsilon_N(\boldsymbol{z})\}$ such that $X_N(\boldsymbol{z})$ is a set of $N+1$ distinct points in a neighbourhood, U_0, of $(\mathcal{H}_{\ell\ell}, \dots, [\mathcal{H}^{2N+1}]_{\ell\ell}) \in U$ and each ε_j is analytic on U_0 [56,96]. With this in hand, we define $\Theta^{\mathrm{q}}_{2N+1} : U_0 \to \mathbb{C}$ by

$$\Theta^{\mathrm{q}}_{2N+1}(\boldsymbol{z}) := \mathcal{L}\big(x \mapsto I_{X(\boldsymbol{z})} O(x)\big)(\boldsymbol{z}_{1:N}) = \sum_{j=0}^{N} \mathcal{L}\Big(x \mapsto \prod_{i \neq j} \frac{x - \varepsilon_i}{\varepsilon_j - \varepsilon_i}\Big) \cdot O \circ \varepsilon_j, \tag{D.4}$$

which is analytic on $\{\boldsymbol{z} \in U_0 : O \text{ analytic at } \varepsilon_j(\boldsymbol{z}) \ \forall j = 0, \dots, N\}$.

D.3. Proof of Lemma D.1

Proof of (i). First note that $\int p_0 p_1 \mathrm{d}D_\ell = 0$. We assume that $p_0, \dots, p_n$ are mutually orthogonal with respect to D_ℓ, and note that,

$$\begin{aligned} b_1 = b_1 \int p_1^2 \mathrm{d}D_\ell &= \int (x - a_0) p_1(x) \mathrm{d}D_\ell(x) = \int x p_0(x) p_1(x) \mathrm{d}D_\ell(x), \qquad \text{and} \\ b_n = b_n \int p_n^2 \mathrm{d}D_\ell &= \int \big((x - a_{n-1}) p_{n-1}(x) p_n(x) - b_{n-1} p_{n-2}(x) p_n(x)\big) \mathrm{d}D_\ell(x) \\ &= \int x p_{n-1}(x) p_n(x) \mathrm{d}D_\ell(x) \qquad \text{for } n \geqslant 2. \end{aligned} \tag{D.5}$$

Therefore, we conclude by noting

$$b_{n+1} \int p_{n+1} p_j \mathrm{d}D_\ell = \int \big((x - a_n) p_n(x) p_j(x) - b_n p_{n-1}(x) p_j(x)\big) \mathrm{d}D_\ell(x) \tag{D.6}$$

$$= \begin{cases} \int x p_n(x)^2 \mathrm{d}D_\ell - a_n & \text{if } j = n, \\ \int x p_n(x) p_{n-1}(x) \mathrm{d}D_\ell - b_n & \text{if } j = n-1 \\ 0 & \text{if } j \leqslant n-2, \end{cases} \tag{D.7}$$

and applying (D.5). Equation (D.5) also justifies the tri-diagonal structure (4.31).

Proof of (ii). We may rewrite the recurrence relation (4.29) as $x\,\boldsymbol{p}(x) = T_N \boldsymbol{p}(x) + b_{N+1} p_{N+1}(x) \boldsymbol{e}_N$ where $\boldsymbol{p}(x) := \big(1, p_1(x), \dots, p_N(x)\big)^T$, $[\boldsymbol{e}_N]_j = \delta_{jN}$, and T_N is the tri-diagonal matrix (4.31). In particular, each $\varepsilon_j \in X_N$ is an eigenvalue of T_N (with eigenvector $\boldsymbol{p}(\varepsilon_j)$).

Proof of (iii). Since T_N is symmetric, the spectrum is real. Now, for each $\varepsilon_j \in X_N = \sigma(T_N)$, the matrix $(T_N - \varepsilon_j)_{\neg N \neg 0}$ formed by removing the N^{th} row and 0^{th} column is lower-triangular with diagonal $(b_1, \dots, b_N)$. Since each $b_i > 0$, $(T_N - \varepsilon_j)_{\neg N \neg 0}$ has full rank and thus ε_j is a simple eigenvalue of T_N.

Proof of (iv). Suppose that (after possibly relabelling) $\varepsilon_0, \varepsilon_1 \in X_N \cap [a, b]$. After defining $R(x) := \prod_{j=2}^{N}(x - \varepsilon_j)$, a polynomial of degree $N - 1$, and noting $(x - \varepsilon_0)(x - \varepsilon_1) > 0$ on supp D_ℓ, we obtain

$$\int p_{N+1}(x) R(x) \mathrm{d}D_\ell(x) = \int R(x)^2 (x - \varepsilon_0)(x - \varepsilon_1) \mathrm{d}D_\ell(x) > 0,$$

contradicting part *(i)*.

Proof of (v). We may write $P_{2N+1} = p_{N+1} q_N + r_N$ where q_N, r_N are polynomials of degree at most N and note that $[p_{N+1}(\mathcal{H}) q_N(\mathcal{H})]_{\ell\ell} = 0$ by *(i)* and $P_{2N+1}(\varepsilon_j) = r_N(\varepsilon_j)$ since X is the set of zeros of p_{N+1}. Therefore,

$$\int P_{2N+1}(x) \mathrm{d}D_\ell(x) = \int \Big[p_{N+1}(x) q_N(x) + r_N(x) \Big] \mathrm{d}D_\ell(x) \tag{D.8}$$

$$= \int r_N(x) \mathrm{d}D_\ell(x) = \int I_X r_N(x) \mathrm{d}D_\ell(x) \tag{D.9}$$

$$= \sum_{j=0}^{N} w_j r_N(\varepsilon_j) = \sum_{j=0}^{N} w_j P_{2N+1}(\varepsilon_j). \tag{D.10}$$

In (D.9) we have used the fact that polynomial interpolation in $N+1$ distinct points is exact for polynomials of degree at most N.

Proof of (vi). $\ell_j(x)^2$ is a polynomial of degree $2N$ and so, by *(v)*, we have

$$0 \leqslant \int \ell_j(x)^2 \mathrm{d}D_\ell(x) = \sum_{i=0}^{N} w_i \ell_j(\varepsilon_i)^2 = w_j.$$

Moreover, $\sum_{j=0}^{N} \ell_j(x)$ is a polynomial of degree N equal to one on X_N (a set of $N+1$ distinct points) and so $\sum_{j=0}^{N} \ell_j(x) \equiv 1$. Finally, $\sum_{j=0}^{N} w_j = \int \big(\sum_{j=0}^{N} \ell_j(x) \big) \mathrm{d}D_\ell(x) = 1$.

Appendix E. Numerical Bond-Order Potentials (BOP)

In mathematical terms, the idea behind BOP methods is to replace the local density of states (LDOS) with an approximation using only the information from the truncated tri-diagonal matrix T_N (and possibly additional hyper-parameters). Since the first N coefficients contain the same information as the first $2N + 1$ moments $\mathcal{H}_{\ell\ell}, \dots, [\mathcal{H}^{2N+1}]_{\ell\ell}$, this approach is closely related to the method of moments [22].

Equivalently, the resolvent $[(z-\mathcal{H})^{-1}]_{\ell\ell}$, which can be written conveniently as the continued fraction expansion

$$[(z-\mathcal{H})^{-1}]_{\ell\ell} = \cfrac{1}{z-a_0-\cfrac{b_1^2}{z-a_1-\cfrac{b_2^2}{z-a_2-\ddots}}}, \tag{E.1}$$

is replaced with an approximation G_ℓ^N only involving the coefficients from T_N. For example, for fixed *terminator* t_∞, we may define

$$G_\ell^N(z) := \cfrac{1}{z-a_0-\cfrac{b_1^2}{z-a_1-\cfrac{b_2^2}{\ddots\ -\cfrac{b_N^2}{z-a_N-t_\infty(z)}}}}. \tag{E.2}$$

Truncating (E.1) to level N, which is equivalent to replacing the far-field of the linear chain with vacuum and choosing $t_\infty = 0$, results in a rational approximation to the resolvent and thus a discrete approximation to the LDOS. We have seen that truncation of the continued fraction in this way leads to an approximation scheme satisfying Theorem 2.5.

Alternatively, the far-field may be replaced with a constant linear chain with $a_{N+j} = a_\infty$ and $b_{N+j} = b_\infty$ for all $j \geqslant 1$ leading to the square root terminator $t_\infty(z) = \frac{b_\infty^2}{z-a_\infty-t_\infty(z)}$ [38,49,97].

More generally, one may choose any "approximate" local density of states $\widetilde{D}_\ell$ and construct a corresponding terminator that encodes the information from $\widetilde{D}_\ell$ [52,65]. For example, $\widetilde{D}_\ell(x) := \frac{1}{b_\infty \pi}\sqrt{1-\left(\frac{x-a_\infty}{2b_\infty}\right)^2}$ results in the square root terminator. While we are unaware of any rigorous results, there is numerical evidence [52] to suggest that the error in the approximation scheme is related to the smoothness of the difference $D_\ell - \widetilde{D}_\ell$.

Equivalently, we may choose any bounded symmetric tri-diagonal (Jacobi) operator $\widetilde{T}_N$ with diagonal $a_0, a_1, \ldots, a_N, \widetilde{a}_{N+1}, \ldots$ and off-diagonal $b_1, \ldots, b_N$, $\widetilde{b}_{N+1}, \ldots$. That is, we may evaluate the recursion method exactly to level N and append the far-field boundary condition $\{\widetilde{a}_n, \widetilde{b}_n\}_{n \geqslant N+1}$ to the semi-infinite linear chain. This approach also includes the case $t_\infty = 0$ as in § 4.3 by choosing $\widetilde{a}_n = \widetilde{b}_n = 0$ for all n.

With this in hand, we define

$$O_\ell^{2N+1,\mathrm{BOP}}(\boldsymbol{u}) := O(\widetilde{T}_N)_{00} = \int O \,\mathrm{d}\widetilde{D}_\ell^{2N+1,\mathrm{BOP}} \tag{E.3}$$

where $\widetilde{D}_\ell^{2N+1,\mathrm{BOP}}$ is the appropriate spectral measure corresponding to $\widetilde{T}_N$.

E.1. Error estimates

Since $[(\widetilde{T}_N)^n]_{00} = [(T_N)^n]_{00} = [(T_\infty)^n]_{00}$ is independent of the far-field coefficients $\{\widetilde{a}_j, \widetilde{b}_j\}$ for all $n \leqslant 2N+1$, we can immediately see that the first $2N+1$ moments of $\widetilde{D}_\ell^{2N+1,\mathrm{BOP}}$ agree with those of D_ℓ. In particular, we may immediately apply (2.14) to obtain error estimates that depend on $\mathrm{supp}\big(D_\ell - \widetilde{D}_\ell^{2N+1,\mathrm{BOP}}\big)$.

Therefore, as long as the far-field boundary condition is chosen so that there are only finitely many discrete eigenvalues in the band gap independent of N, the more complicated BOP schemes converge at least as quickly as the $t_\infty = 0$ case. Intuitively, if the far-field boundary condition is chosen to capture the behaviour of the LDOS (e.g. the type and location of band-edge singularities), then the integration against the signed measure $D_\ell - \widetilde{D}_\ell^{2N+1,\mathrm{BOP}}$ as in (2.14) may lead to improved error estimates. A rigorous error analysis to this affect is left for future work.

E.2. Analyticity

Since $\widetilde{T}_N$ is bounded and symmetric, the spectrum $\sigma(\widetilde{T}_N)$ is contained in a bounded interval of the real line. In particular, we can apply the same arguments as in (4.44) to conclude that (E.3) defines a nonlinear approximation scheme given by an analytic function on an open subset of $\mathbb{C}^{2N+1}$.

Appendix F. Kernel Polynomial Method & Analytic Bond Order Potentials

We first introduce the Kernel Polynomial Method (KPM) for approximating the LDOS [82,83,98]. In this section, we scale the spectrum and assume that $\sigma(\mathcal{H}) \subset [-1, 1]$.

For a sequence of *kernels* $K_N(x, y)$, we define the approximate quantities of interest

$$O_\ell^N := \int K_N \star O \,\mathrm{d}D_\ell =: \iint K_N(x, y) O(y) \,\mathrm{d}y \,\mathrm{d}D_\ell(x). \tag{F.1}$$

Under the choice $K_N(x, y) := \frac{2}{\pi}\sqrt{1-y^2}\sum_{n=0}^N U_n(x)U_n(y)$ (where U_n denotes the n^{th} Chebyshev polynomial of the second kind), we arrive at a projection method similar to that discussed in § 4.1.4: if $O(x) = \sum_{m=0}^\infty c_m U_m(x)$, then

$$K_N \star O(x) = \sum_{m,n} c_m U_n(x) \frac{2}{\pi} \int_{-1}^1 U_n(y) U_m(y) \sqrt{1-y^2} \,\mathrm{d}y = \sum_{n=0}^N c_n U_n(x). \tag{F.2}$$

Equivalently, we may consider the corresponding approximate LDOS

$$O_\ell^N = \int O(x) \mathrm{d}D_\ell^N(x) \quad \text{where} \quad D_\ell^N(x) = \frac{2}{\pi}\sqrt{1-x^2} \sum_{n=0}^N U_n(\mathcal{H})_{\ell\ell} U_n(x).$$

However, truncation of the Chebyshev series in this way leads to artificial oscillations in the approximate LDOS known as Gibbs oscillations [43]. Moreover, without *damping* these oscillations, the approximate LDOS need not be positive. However, on defining

$$K_N^{\text{Fejer}}(x,y) := \frac{1}{N}\sum_{n=1}^{N} K_n(x,y) = \frac{2}{\pi}\sqrt{1-y^2}\sum_{n=0}^{N}\left(1-\tfrac{n}{N}\right)U_n(x)U_n(y), \quad \text{(F.3)}$$

we obtain a positive approximate LDOS [98] where the *damping coefficients* $d_n := (1-\frac{n}{N})$ reduce the effect of Gibbs ringing. In practice, one may instead choose the *Jackson kernel* [47].

The problem with the above analysis in practice is that the damping factors that we have introduced mean that more moments $[\mathcal{H}^n]_{\ell\ell}$ are required in order to obtain good approximations to the LDOS. Instead, analytic BOP methods [74,79] compute the first N rows of the tridiagonal operator T_∞, thus obtaining the first $2N+1$ moments exactly. Then, a far-field boundary condition (such as a constant infinite linear chain) is appended to form a corresponding Jacobi operator $\widetilde{T}_N$ as in Appendix E. Now, since higher order moments of $\widetilde{T}_N$ can be efficiently computed, we may evaluate the following approximate LDOS

$$D_\ell^{N,M}(x) := \frac{2}{\pi}\sqrt{1-x^2}\sum_{n=0}^{M} d_n U_n(\widetilde{T}_N)_{00}U_n(x) \quad \text{(F.4)}$$

where d_n are damping coefficients and $M > 2N+1$. The damping is chosen so that the lower order moments which are computed exactly and are more important for the reconstruction of the LDOS are only slightly damped. With this choice of kernel, the approximate quantities of interest take the form

$$O_\ell^{N,M} = \sum_{n=0}^{2N+1} d_n c_n U_n(\mathcal{H})_{\ell\ell} + \sum_{n=2N+2}^{M} d_n c_n U_n(\widetilde{T}_N)_{00}.$$

Efficient implementation of analytic BOP methods can be carried out using the BOPfox program [47].

References

1. Aupetit, B.: A Primer on Spectral Theory. Springer, Berlin (1991)
2. Bachmayr, M., Csanyi, G., Drautz, R., Dusson, G., Etter, S., van der Oord, C., Ortner, C.: *Atomic cluster expansion: Completeness, efficiency and stability*, ArXiv e-prints arXiv:1911.03550 (2019).
3. Bak, J., Newman, D.J.: Complex Analysis. Springer, Berlin (2010)
4. Bartók, A.P., Kermode, J., Bernstein, N., Csányi, G.: Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**, 041048, 2018
5. Bartók, A.P., Payne, M.C., Kondor, R., Csányi, G.: Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403, 2010

6. BASKES, M.I.: Modified embedded-atom potentials for cubic materials and impurities. *Phys. Rev. B: Condens. Matter* **46**, 2727–2742, 1992
7. BAZANT, M.Z., KAXIRAS, E., JUSTO, J.F.: Environment-dependent interatomic potential for bulk silicon. *Phys. Rev. B Condens. Matter* **56**, 8542–8552, 1997
8. BEHLER, J., PARRINELLO, M.: Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401, 2007
9. BENZI, M., BOITO, P., RAZOUK, N.: Decay properties of spectral projectors with applications to electronic structure. *SIAM Rev.* **55**, 3–64, 2013
10. BISWAS, R., HAMANN, D.R.: New classical models for silicon structural energies. *Phys. Rev. B* **36**, 6434–6445, 1987
11. BRAAMS, B.J., BOWMAN, J.M.: Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **28**, 577–606, 2009
12. CANCÈS, É., KEMLIN, G., LEVITT, A.: Convergence analysis of direct minimization and self-consistent iterations. *SIAM J. Matrix Anal. Appl.* **42**, 243–274, 2021
13. CANCÈS, E., EHRLACHER, V., MADAY, Y.: Periodic schrödinger operators with local defects and spectral pollution. *SIAM J. Numer. Anal.* **50**, 3016–3035, 2012
14. CHEN, H., LU, J., ORTNER, C.: Thermodynamic limit of crystal defects with finite temperature tight binding. *Arch. Ration. Mech. Anal.* **230**, 701–733, 2018
15. CHEN, H., NAZAR, F.Q., ORTNER, C.: Geometry equilibration of crystalline defects in quantum and atomistic descriptions. *Math. Models Methods Appl. Sci.* **29**, 419–492, 2019
16. CHEN, H., ORTNER, C.: QM/MM methods for crystalline defects. Part 1: locality of the tight binding model. *Multiscale Model. Simul.* **14**, 232–264, 2016
17. CHEN, H., ORTNER, C., THOMAS, J.: Locality of interatomic forces in tight binding models for insulators. *ESAIM Math. Model. Numer. Anal.* **54**, 2295–2318, 2020
18. CHEN, J., LU, J.: Analysis of the divide-and-conquer method for electronic structure calculations. *Math. Comput.* **85**, 2919–2938, 2016
19. CHUPIN, M., DUPUY, M.-S., LEGENDRE, G., SÉRÉ, É.: Convergence analysis of adaptive DIIS algorithms with application to electronic ground state calculations, ArXiv e-prints arXiv:2002.12850 (2020).
20. COHEN, R.E., MEHL, M.J., PAPACONSTANTOPOULOS, D.A.: Tight-binding total-energy method for transition and noble metals. *Phys. Rev. B* **50**, 14694–14697, 1994
21. COMBES, J., THOMAS, L.: Asymptotic behavior of eigenfunctions for multiparticle Schrödinger operators. *Commun. Math. Phys.* **34**, 251–270, 1973
22. CYROT-LACKMANN, F.: On the electronic structure of liquid transitional metals. *Adv. Phys.* **16**, 393–400, 1967
23. DAW, M.S., BASKES, M.I.: Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B Condens. Matter* **29**, 6443–6453, 1984
24. DENISOV, S.A., SIMON, B.: Zeros of orthogonal polynomials on the real line. *J. Approx. Theory* **121**, 357–364, 2003
25. DRAUTZ, R.: Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104, 2019
26. DRAUTZ, R.: From electrons to interatomic potentials for materials simulations. In: PAVARINI, E., KOCH, E. (eds.) Topology, Entanglement, and Strong Correlations. Forschungszentrum Jülich GmbH, Institute for Advanced Simulation, Berlin (2020)
27. DRAUTZ, R., FÄHNLE, M., SANCHEZ, J.M.: General relations between many-body potentials and cluster expansions in multicomponent systems. *J. Phys. Condens. Matter* **16**, 3843–3852, 2004
28. DRISCOLL, T.A.: Algorithm 756: a MATLAB toolbox for Schwarz–Christoffel mapping. *ACM T. Math. Softw.* **22**, 168–186, 1996
29. DRISCOLL, T.A.: *Schwarz–Christoffel toolbox*. https://github.com/tobydriscoll/sc-toolbox (2007)
30. DRISCOLL, T.A., TOH, K.-C., TREFETHEN, L.N.: From potential theory to matrix iterations in six steps. *SIAM Rev.* **40**, 547–578, 1998

31. Driscoll, T.A., Trefethen, L.N.: Schwarz–Christoffel Mapping. Cambridge University Press, Cambridge (2002)
32. Ehrlacher, V., Ortner, C., Shapeev, A.V.: Analysis of boundary conditions for crystal defect atomistic simulations. *Arch. Ration. Mech. Anal.* **222**, 1217–1268, 2016
33. Elstner, M., Seifert, G.: Density functional tight binding. *Philos. Trans. R. Soc. A* **372**, 20120483, 2014
34. Embree, M., Trefethen, L.N.: Green's functions for multiply connected domains via conformal mapping. *SIAM Rev.* **41**, 745–761, 1999
35. Ercolessi, F.: Tight-binding molecular dynamics and tight-binding justification of classical potentials, lecture notes, 2005
36. Ercolessi, F., Adams, J.B.: Interatomic potentials from first-principles calculations: the force-matching method. *EPL* **26**, 583, 1994
37. Etter, S.: Polynomial and Rational Approximation for Electronic Structure Calculations, Ph.D. thesis. University of Warwick, UK (2019)
38. Finnis, M.: Interatomic Forces in Condensed Matter. Oxford University Press, Oxford (2003)
39. Finnis, M.: Bond-order potentials through the ages. *Prog. Mater Sci.* **52**, 133–153, 2007
40. Freud, G.: Orthogonal Polynomials. Elsevier, Amsterdam (2014)
41. Glanville, S., Paxton, A.T., Finnis, M.W.: A comparison of methods for calculating tight-binding bond energies. *J. Phys. F Metal Phys.* **18**, 693–718, 1988
42. Goedecker, S.: Linear scaling electronic structure methods. *Rev. Mod. Phys.* **71**, 1085–1123, 1999
43. Grafakos, L.: Classical Fourier Analysis. Springer, Berlin (2016)
44. Greenbaum, A., Li, R.-C., Overton, M.L.: First-order perturbation theory for eigenvalues and eigenvectors. *SIAM Rev.* **62**, 463–482, 2020
45. Gunning, R.C., Rossi, H.: Analytic Functions of Several Complex Variables. Prentice-Hall, New York (1965)
46. Haliciogli, T., Pamuk, H.O., Erkoc, S.: Interatomic potentials with multi-body interactions. *Phys. Status Solidi (b)* **149**, 81–92, 1988
47. Hammerschmidt, T., Seiser, B., Ford, M., Ladines, A., Schreiber, S., Wang, N., Jenke, J., Lysogorskiy, Y., Teijeiro, C., Mrovec, M., Cak, M., Margine, E., Pettifor, D., Drautz, R.: BOPfox program for tight-binding and analytic bond-order potential calculations. *Comput. Phys. Commun.* **235**, 221–233, 2019
48. Hartree, D.R.: The wave mechanics of an atom with a non-Coulomb central field. Part I: theory and methods. *Math. Proc. Camb. Philos. Soc.* **24**, 89–110, 1928
49. Haydock, R., Heine, V., Kelly, M.J.: Electronic structure based on the local atomic environment for tight-binding bands. *J. Phys. C Solid State Phys.* **5**, 2845–2858, 1972
50. Haydock, R., Heine, V., Kelly, M.J.: Electronic structure based on the local atomic environment for tight-binding bands, II. *J. Phys. C Solid State Phys.* **8**, 2591–2605, 1975
51. Haydock, R., Nex, C.M.M.: Comparison of quadrature and termination for estimating the density of states within the recursion method. *J. Phys. C Solid State Phys.* **17**, 4783–4789, 1984
52. Haydock, R., Nex, C.M.M.: A general terminator for the recursion method. *J. Phys. C Solid State Phys.* **18**, 2235–2248, 1985
53. Herbst, M.F., Levitt, A.: Black-box inhomogeneous preconditioning for self-consistent field iterations in density functional theory. *J. Phys. Condens. Matter* **33**, 085503, 2020
54. Hohenberg, P., Kohn, W.: Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871, 1964
55. Horsfield, A.P., Bratkovsky, A.M., Fearn, M., Pettifor, D.G., Aoki, M.: Bond-order potentials: theory and implementation. *Phys. Rev. B* **53**, 12694–12712, 1996
56. Kato, T.: Perturbation Theory for Linear Operators, 2nd edn. Springer, Berlin (1995)
57. Kittel, C.: Introduction to Solid State Physics, 8th edn. Wiley, New York (2004)
58. Kohn, W., Sham, L.J.: Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138, 1965

59. KOSKINEN, P., MÄKINEN, V.: Density-functional tight-binding for beginners. *Comput. Mater. Sci.* **47**, 237–253, 2009
60. KRANTZ, S.: Function Theory of Several Complex Variables. American Mathematical Society, Providence (2001)
61. LANCZOS, C.: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 1950
62. LEVIN, E., SAFF, E.B.: Potential theoretic tools in polynomial and rational approximation. In *Harmonic Analysis and Rational Approximation. Lecture Notes in Control and Information Science*, pp. 71–94. Springer, Berlin (2006)
63. LEVITT, A.: Screening in the finite-temperature reduced Hartree–Fock model. *Arch. Rat. Mech. Anal.* **238**, 901–927, 2020
64. LEWIN, M., SÉRÉ, É.: Spectral pollution and how to avoid it. *Proc. Lond. Math. Soc.* **100**, 864–900, 2009
65. LUCHINI, M.U., NEX, C.M.M.: A new procedure for appending terminators in the recursion method. *J. Phys. C Solid State Phys.* **20**, 3125–3130, 1987
66. LYSOGORSKIY, Y., VAN DER OORD, C., BOCHKAREV, A., MENON, S., RINALDI, M., HAMMERSCHMIDT, T., MROVEC, M., THOMPSON, A., CSANYI, G., ORTNER, C., DRAUTZ, R.: Performant implementation of the atomic cluster expansion (PACE): application to copper and silicon, ArXiv e-prints arXiv:2103.00814 (to appear in NPJ Computational Materials) (2021)
67. MEAD, L.R., PAPANICOLAOU, N.: Maximum entropy in the problem of moments. *J. Math. Phys.* **25**, 2404–2417, 1984
68. MEHL, M.J., PAPACONSTANTOPOULOS, D.A.: Applications of a tight-binding total-energy method for transition and noble metals: elastic constants, vacancies, and surfaces of monatomic metals. *Phys. Rev. B* **54**, 4519–4530, 1996
69. NEX, C.M.M.: Estimation of integrals with respect to a density of states. *J. Phys. A Math. Gen.* **11**, 653–663, 1978
70. ORTNER, C., THOMAS, J.: Point defects in tight binding models for insulators. *Math. Models Methods Appl. Sci.* **30**, 2753–2797, 2020
71. PAPACONSTANTOPOULOS, D.A.: Handbook of the Band Structure of Elemental Solids, From Z = 1 To Z = 112. Springer, New York (2015)
72. PAPACONSTANTOPOULOS, D.A., MEHL, M.J., ERWIN, S.C., PEDERSON, M.R.: Tight-binding Hamiltonians for carbon and silicon. *Symposium R - Tight Binding Approach to Comput. Mater. Sci.* **491**, 221, 1997
73. PARR, R.G., WEITAO, Y.: Density-Functional Theory of Atoms and Molecules. Oxford University Press, Oxford (1994)
74. PETTIFOR, D.: New many-body potential for the bond order. *Phys. Rev. Lett.* **63**, 2480–2483, 1989
75. RANSFORD, T.: Potential Theory in the Complex Plane. Cambridge University Press, Cambridge (1995)
76. SAFF, E.B.: Logarithmic potential theory with applications to approximation theory. *Surveys in Approximation Theory* **5**, 165–200, 2010
77. SAFF, E.B., TOTIK, V.: Logarithmic Potentials with External Fields. Springer, Berlin (1997)
78. SEIFERT, G., JOSWIG, J.-O.: Density-functional tight binding-an approximate density-functional theory method. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 456–465, 2012
79. SEISER, B., PETTIFOR, D.G., DRAUTZ, R.: Analytic bond-order potential expansion of recursion-based methods. *Phys. Rev. B* **87**, 094105, 2013
80. SHAPEEV, A.: Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173, 2016
81. SHEN, J., STRANG, G., WATHEN, A.J.: The potential theory of several intervals and its applications. *Appl. Math. Opt.* **44**, 67–85, 2001
82. SILVER, R., RODER, H.: Densities of states of mega-dimensional Hamiltonian matrices. *Int. J. Mod. Phys. C* **05**, 735–753, 1994

83. Silver, R., Roeder, H., Voter, A., Kress, J.: Kernel polynomial approximations for densities of states and spectral functions. *J. Comput. Phys.* **124**, 115–130, 1996
84. Slater, J.C., Koster, G.F.: Simplified LCAO method for the periodic potential problem. *Phys. Rev.* **94**, 1498–1524, 1954
85. Stahl, H., Totik, V.: General Orthogonal Polynomials, Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge (1992)
86. Stillinger, F.H., Weber, T.A.: Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B Condens. Matter* **31**, 5262–5271, 1985
87. Suryanarayana, P., Bhattacharya, K., Ortiz, M.: Coarse-graining Kohn–Sham density functional theory. *J. Mech. Phys. Solids* **61**, 38–60, 2013
88. Suryanarayana, P., Pratapa, P.P., Sharma, A., Pask, J.E.: SQDFT: spectral quadrature method for large-scale parallel O(N) Kohn–Sham calculations at high temperature. *Comput. Phys. Commun.* **224**, 288–298, 2018
89. Sutton, A.P.: Electronic Structure of Materials. Oxford University Press, Oxford (1993)
90. Taylor, R., Totik, V.: Lebesgue constants for Leja points. *IMA J. Num. Anal.* **30**, 462–486, 2008
91. Teschl, G.: Jacobi operators and completely integrable nonlinear lattices, vol. 72. Mathematical Surveys and Monographs, Providence (2000)
92. Teschl, G.: Mathematical Methods in Quantum Mechanics. The American Mathematical Society, Providence (2014)
93. Thomas, J.: Locality of interatomic interactions in self-consistent tight binding models. *J. Nonlinear Sci.* **30**, 3293–3319, 2020
94. Thomas, L.H.: The calculation of atomic fields. *Math. Proc. Camb. Philos. Soc.* **23**, 542–548, 1927
95. Trefethen, L.N.: Approximation Theory and Approximation Practice, Extended. SIAM, Philadelphia (2019)
96. Tsing, N.-K., Fan, M.K., Verriest, E.I.: On analyticity of functions involving eigenvalues. *Linear Algebra Appl.* **207**, 159–180, 1994
97. Turchi, P., Ducastelle, F., Treglia, G.: Band gaps and asymptotic behaviour of continued fraction coefficients. *J. Phys. C Solid State Phys.* **15**, 2891–2924, 1982
98. Voter, A.F., Kress, J.D., Silver, R.N.: Linear-scaling tight binding from a truncated-moment approach. *Phys. Rev. B* **53**, 12733–12741, 1996
99. Weinan, E., Lu, J.: Electronic structure of smoothly deformed crystals: Cauchy–Born rule for the nonlinear tight-binding model. *Commun. Pure Appl. Math.* **63**, 1432–1468, 2010
100. Weinan, E., Lu, J.: The Kohn–Sham equation for deformed crystals. *Mem. Am. Math. Soc.* **221**, 1, 2012
101. Woods, N.D., Payne, M.C., Hasnip, P.J.: Computing the self-consistent field in Kohn–Sham density functional theory. *J. Phys. Condens. Matter* **31**, 453001, 2019
102. Yamamoto, T.: A convergence theorem for Newton's method in Banach spaces. *Jpn. J. Appl. Math.* **3**, 37–52, 1986
103. Yang, W.: Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **66**, 1438–1441, 1991
104. Zhengda, H.: A note on the Kantorovich theorem for Newton iteration. *J. Comput. Appl. Math.* **47**, 211–217, 1993
105. Zhu, L., Amsler, M., Fuhrer, T., Schaefer, B., Faraji, S., Rostami, S., Ghasemi, S.A., Sadeghi, A., Grauzinyte, M., Wolverton, C., Goedecker, S.: A fingerprint based metric for measuring similarities of crystalline structures. *J. Chem. Phys.* **144**, 034203, 2016
106. Zuo, Y., Chen, C., Li, X., Deng, Z., Chen, Y., Behler, J., Csányi, G., Shapeev, A.V., Thompson, A.P., Wood, M.A., Ong, S.P.: Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745, 2020

JACK THOMAS
Mathematics Institute, Zeeman Building,
University of Warwick,
Coventry
UK.
e-mail: j.thomas.1@warwickgrad.net

and

HUAJIE CHEN
School of Mathematical Sciences,
Beijing Normal University,
Beijing
China.
e-mail: chen.huajie@bnu.edu.cn

and

CHRISTOPH ORTNER
Department of Mathematics,
University of British Columbia,
Vancouver
Canada.
e-mail: ortner@math.ubc.ca

(*Received June 30, 2021 / Accepted June 27, 2022*)
Published online August 6, 2022