ESAIM: M2AN 54 (2020) 2295–2318
https://doi.org/10.1051/m2an/2020020

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

LOCALITY OF INTERATOMIC FORCES IN TIGHT BINDING MODELS FOR INSULATORS

Christoph Ortner[1], Jack Thomas[1,*] and Huajie Chen[2]

Abstract. The tight binding model is a minimalistic electronic structure model for predicting properties of materials and molecules. For insulators at zero Fermi-temperature we show that the potential energy surface of this model can be decomposed into exponentially localised site energy contributions, thus providing qualitatively sharp estimates on the interatomic interaction range which justifies a range of multi-scale models. For insulators at finite Fermi-temperature we obtain locality estimates that are uniform in the zero-temperature limit. A particular feature of all our results is that they depend only weakly on the point spectrum. Numerical tests confirm our analytical results. This work extends Chen and Ortner [*Multiscale Model. Simul.* **14** (2016) 232–264] and Chen *et al.* [*Arch. Ration. Mech. Anal.* **230** (2018) 701–733] to the case of zero Fermi-temperature as well as strengthening the results proved therein.

Mathematics Subject Classification. 74E15, 74G65.

Received August 17, 2019. Accepted March 24, 2020.

1. Introduction

A wide range of electronic, optical and magnetic properties of solids are determined by electronic structure. Computational methods, such as density functional theory, have been used successfully to model electronic structure and thus allowed the investigation and prediction of properties of materials [19,26]. The tight binding model is a simple quantum mechanical model lying, both in terms of computational cost and accuracy, between empirical interatomic potential methods and expensive *ab initio* calculations. Nevertheless, due to the underlying eigenvalue problem, a naive implementation of tight binding models requires $O(N^3)$ computational cost, where N denotes the number of particles in the simulation. A possible route to alleviate this cost are linear scaling algorithms [7,20,25,28], which rely on the "nearsightedness" of the density matrix.

If the quantities of interest in a simulation are mechanical properties, then it may be advantageous to entirely bypass the electronic structure model and replace it with an interatomic potential (IP). The recent transfer of machine learning technology into this domain has made it possible to "fit" high-accuracy IP models [4–6,34], which makes this approach particularly attractive. A starting assumption in most IP models for materials is

Keywords and phrases. Strong locality, tight binding, point defects.

[1] Mathematics Institute, University of Warwick, Zeeman Building, Coventry, UK.
[2] School of Mathematical Sciences, Beijing Normal University, Beijing, P.R. China.
*Corresponding author: J.Thomas.1@warwick.ac.uk

Article published by EDP Sciences

that the potential energy surface can be decomposed into site energies, *i.e.* contributions from individual atoms that depend only on a small neighbourhood.

A partial justification for this assumption was given in [9,11], for a linear tight binding model at finite Fermi-temperature. For nuclei positions $y = \{y_n\}$, it was shown that there is a decomposition of the total potential energy into site energies

$$G(y) = \sum_{\ell} G_\ell(y), \quad \text{where} \quad \left|\frac{\partial G_\ell(y)}{\partial y_n}\right| \lesssim e^{-\eta|y_\ell - y_n|}, \tag{1.1}$$

for some $\eta > 0$. Similar estimates are also shown for higher derivatives.

The exponent η in (1.1) measures the interatomic interaction range. Classical IPs are typically fairly short-ranged (using a cut-off on the order of 2–3 interatomic distances), which is only justified if η is not too small. Similarly, in QM/MM multi-scale schemes the exponent η determines the size of the QM region that must be imposed [9,10,15] as well as the interaction range of the coarse-grained MM model. We emphasize that results such as (1.1) do not follow from the classical near-sightedness of the density matrix. Indeed, in general, as pointed out in [15], exponential off-diagonal decay of the density matrix is not sufficient to validate multi-scale and hybrid models.

Unfortunately, one expects (and we make this precise in the present paper) that, in general,

$$\eta \sim \beta^{-1},$$

where β is the inverse Fermi-temperature. This means that, for moderate to low temperature regimes, the practical value of (1.1) is limited.

The main purpose of the present paper is to demonstrate that for insulators the presence of a spectral gap significantly improves the estimate. Specifically, we consider a linear tight-binding model at either zero or finite-Fermi temperature, with electrons in a grand-canonical ensemble. In this setting we prove that (1.1) holds with η independent of β, but instead η is linear in the spectral gap. Moreover, we demonstrate that "pollution" of the spectral gap by a point spectrum caused, for example, by local defects in the crystal, affects only the prefactors, but not the exponent η in (1.1). The present paper therefore significantly strengthens the estimates of [9,11] to the case of insulating multi-lattice materials in the presence of point defects as well as extending the results to zero Fermi-temperature models.

Related results, such as the near-sightedness of the density matrix (*e.g.* [1–3,18]), exist in the context of random Schrödinger operators. These papers use some similar techniques, such as resolvent calculus and low rank perturbation identities, however, unlike in the aforementioned papers, not only do we show that localisation still occurs for the perturbed Hamiltonian but also that this perturbation does not affect the exponents of the locality estimates. Moreover, our primary concern are potential energies and forces rather than the density matrix.

In addition to supporting the justification of interatomic potentials and QM/MM multi-scale models, our results also allow for an extension of the thermodynamic limit models for crystalline defects [9,12,17] to the zero-temperature case and an investigation of the (non-trivial) relationship between zero and finite-temperature models, which we will pursue in a forthcoming paper [30].

Outline

In Section 2, we state the main results of this paper. In order to do this, we introduce a simple two-centre linear tight binding model (Sect. 2.1) and show that, at both finite and zero Fermi-temperature, the total energy of the system can be decomposed into exponentially localised site energy contributions (Sect. 2.2). We then discuss how these results can be improved upon in the case of a point defect embedded into a reference configuration (Sect. 2.2), showing that the resulting point spectrum does not affect the exponent. In Section 3 we provide numerical tests confirming our analytical results. The main conclusions of this work are then discussed in Section 4 and all of the proofs are collected into Section 5.

Notation

The Frobenius norm will be denoted by $\|\cdot\|_F$ while $\|\cdot\|$ and $|\cdot|$ will denote the ℓ^2 and Euclidean norms, respectively. We write $b + A = \{b + a\colon a \in A\}$ and similarly for $A - b$. Moreover, for $A \subset \mathbb{C}$ and $b \in \mathbb{C}$, the distance between b and A is defined by $\operatorname{dist}(b, A) := \inf_A |b - \cdot|$. For a subset $A \subset \mathbb{R}$ and $\delta > 0$, we denote the ball of radius δ about A by $B_\delta(A) := \{r \in \mathbb{R}\colon \operatorname{dist}(r, A) < \delta\}$. For two subsets $A, B \subset \mathbb{R}$, the Hausdorff distance between A and B is denoted by $\operatorname{dist}(A, B) := \max\{\sup_{a\in A} \operatorname{dist}(a, B), \sup_{b\in B} \operatorname{dist}(b, A)\}$. For a finite set A, we denote by $\#A$ the cardinality of A. The set of strictly positive real numbers will be denoted by $\mathbb{R}_+ := \{r \in \mathbb{R}\colon r > 0\}$.

For an operator T, let us denote by $\sigma_{\text{disc}}(T)$, the discrete spectrum of T (that is, the set of isolated eigenvalues of T with finite multiplicity) and $\sigma_{\text{ess}}(T) := \sigma(T) \setminus \sigma_{\text{disc}}(T)$, the essential spectrum of T.

The symbol C will denote a generic positive constant that may change from one line to the next. In calculations, C will always be independent of Fermi-temperature. The dependencies of C will normally be clear from context or stated explicitly.

2. Results

2.1. Tight binding model

We consider a locally finite reference $\Lambda \subset \mathbb{R}^d$ and configuration $y\colon \Lambda \to \mathbb{R}^d$ satisfying the following uniform non-interpenetration condition:

(L). There exists $\mathfrak{m} > 0$ such that $|y(\ell) - y(k)| \geqslant \mathfrak{m}$ for all $\ell, k \in \Lambda$.

We consider N_{b} atomic orbitals per atom, indexed by $1 \leqslant a, b \leqslant N_{\text{b}}$. For a given admissible configuration y, we define the following two-centre tight binding Hamiltonian [13, 19, 35]:

(TB). For $\ell, k \in \Lambda$ and $1 \leqslant a, b \leqslant N_{\text{b}}$, we suppose that the Hamiltonian take the form

$$\mathcal{H}(y)^{ab}_{\ell k} = h^{ab}_{\ell k}(y(\ell) - y(k)) \tag{2.1}$$

where $h^{ab}_{\ell k}\colon \mathbb{R}^d \to \mathbb{R}$ are ν times continuously differentiable for some $\nu \geqslant 1$. Further, we assume that there exist $\boldsymbol{h} := (h_0, \dots, h_\nu)$, $\boldsymbol{\gamma} := (\gamma_0, \dots, \gamma_\nu) \in (\mathbb{R}_+)^{\nu+1}$ such that, for each $1 \leqslant j \leqslant \nu$,

$$\left|h^{ab}_{\ell k}(\xi)\right| \leqslant h_0\, e^{-\gamma_0|\xi|} \quad \text{and} \quad \left|\partial^\alpha h^{ab}_{\ell k}(\xi)\right| \leqslant h_j\, e^{-\gamma_j|\xi|} \quad \forall \xi \in \mathbb{R}^d \tag{2.2}$$

for all multi-indices $\alpha \in \mathbb{N}^d$ with $|\alpha| = j$. Finally, we suppose that $h^{ab}_{\ell k}(\xi) = h^{ba}_{k\ell}(-\xi)$ for all $\xi \in \mathbb{R}^d$ and $1 \leqslant a, b \leqslant N_{\text{b}}, \ell, k \in \Lambda$.

It is important to emphasise that the constants $\boldsymbol{h}, \boldsymbol{\gamma} \in (\mathbb{R}_+)^{\nu+1}$ in (2.2) are chosen to be independent of the atomic sites $\ell, k \in \Lambda$.

The condition in (2.2) with $j = 0$ is satisfied for most common linear tight binding models. In fact, in most tight binding models, a finite cut-off radius is used and so Hamiltonian entries are zero for atoms beyond a finite interaction range. For $j = 1$, (2.2) states that there are no long-range interactions. That is, the dependence of the Hamiltonian entry $\mathcal{H}(y)^{ab}_{\ell k}$ on site m decays exponentially to zero in $|y(\ell) - y(m)| + |y(m) - y(k)|$. In particular, we are assuming that the Coulomb interactions have been screened, which is typical in practical tight binding models [13, 27, 32].

Since the $h^{ab}_{\ell k}$ depend on the atomic sites, we allow for multi-lattice reference configurations with possibly multiple atomic species. While the assumptions are motivated by the lattice setting, we do not define exactly what we mean by Λ and thus the presentation is kept abstract and the mathematical results are more general.

Under (**TB**), $\mathcal{H}(y)$ is Hermitian and so the spectrum is real. Moreover, $\sigma(\mathcal{H}(y)) \subset [\underline{\sigma}, \overline{\sigma}]$ where $\underline{\sigma}, \overline{\sigma}$ only depend on $\mathfrak{m}, d, h_0, \gamma_0$ and are independent of system size and configuration y satisfying (**L**) with the constant $\mathfrak{m}$. A proof of this fact is an application of the Gershgorin circle theorem ([9], Lem. 4).

In practice, the Hamiltonian satisfies further symmetry properties [35] which are derived in Appendix A of [9].

While nuclei are treated as classical particles, we assume that electrons are described by a grand canonical potential model. That is, the Fermi-temperature, volume and chemical potential, μ, are fixed model parameters.

We first let Λ be a finite system. In this case, after diagonalising the Hamiltonian,

$$\mathcal{H}(y)\psi_s = \lambda_s \psi_s \quad \text{where} \quad \|\psi_s\| = 1, \tag{2.3}$$

(where the dependence of the eigenpair (λ_s, ψ_s) on y has been omitted) the potential energy surface is given by

$$G^\beta(y) \coloneqq \sum_s \mathfrak{g}^\beta(\lambda_s; \mu). \tag{2.4}$$

Here, β is the inverse Fermi-temperature given by $T = (k_{\mathrm{B}}\beta)^{-1}$ where k_{B} is the Boltzmann constant and T the Fermi-temperature. We consider $\beta < \infty$ (as defined in [11,16]) and $\beta = \infty$:

$$\mathfrak{g}^\beta(z;\mu) \coloneqq \frac{2}{\beta}\log(1 - f_\beta(z-\mu)) \quad \text{and} \quad \mathfrak{g}^\infty(z;\mu) \coloneqq 2(z-\mu)\chi_{(-\infty,\mu)}(z)$$

where $f_\beta \coloneqq (1+\exp(\beta\,\cdot))^{-1}$ is the Fermi-Dirac distribution which describes the occupation numbers for the electronic states. The factor of 2 accounts for the spin. The zero Fermi-temperature energy is simply given by the point-wise limit as $\beta \to \infty$.

Remark 2.1. Our analysis requires $G^\infty(y)$ to be a differentiable function of the configuration (that is, the derivative with respect to $[y(m)]_i$ exists for all $m \in \Lambda$ and $1 \leqslant i \leqslant d$) and so we will usually impose the condition that $\mu \notin \sigma(\mathcal{H}(y))$. Justification for considering this zero Fermi-temperature grand potential is given in [11] and a forthcoming paper [30] where we formulate the geometry relaxation problem for the grand potential (2.4) at zero Fermi-temperature as a variational problem and show that this is consistent with taking Fermi-temperature to zero.

2.2. Site energy decomposition

For a given configuration, $y\colon \Lambda \to \mathbb{R}^d$ with Λ finite, we can distribute the total energy of the system into site energy contributions. Since $\|\psi_s\| = 1$, we can decompose $G^\beta(y)$ into

$$G^\beta(y) = \sum_{\ell\in\Lambda} G^\beta_\ell(y) \quad \text{where} \quad G^\beta_\ell(y) \coloneqq \sum_s \mathfrak{g}^\beta(\lambda_s;\mu)\sum_a [\psi_s]^2_{\ell a} \tag{2.5}$$

where $[\psi_s]_{\ell a}$ denotes the $(\ell a)-$entry of ψ_s. Using resolvent calculus, the site energies defined in (2.5) can be extended to the case where Λ is infinite: By Lemma 5.1, $\mathfrak{g}^\beta(\,\cdot\,;\mu)$ extends to a holomorphic function defined on the set $\mathbb{C}\setminus\{\mu + ir\colon r\in\mathbb{R}, |r| \geqslant \pi\beta^{-1}\}$. Therefore, we may write

$$G^\beta_\ell(y) \coloneqq -\frac{1}{2\pi i}\sum_a \oint_{\mathscr{C}_\beta} \mathfrak{g}^\beta(z;\mu)(\mathcal{H}(y)-z)^{-1}_{\ell\ell,aa}\mathrm{d}z \tag{2.6}$$

where $\mathscr{C}_\beta$ is a simple closed contour contained within the region of holomorphicity of $\mathfrak{g}^\beta(z;\mu)$ and encircling the spectrum $\sigma(\mathcal{H}(y))$; see Figure 1. We may choose $\mathscr{C}_\beta$ such that

$$\operatorname{dist}\big(z, \sigma(\mathcal{H}(y)) \cup \{\mu + ir\colon r\in\mathbb{R}, |r| \geqslant \pi\beta^{-1}\}\big) \geqslant \frac{\pi}{2\beta} \quad \text{for all} \quad z \in \mathscr{C}_\beta. \tag{2.7}$$

Since the resolvent operator, $(\mathcal{H}(y)-z)^{-1}$, is a well defined bounded linear operator for all $z \in \mathbb{C}\setminus\sigma(\mathcal{H}(y))$, the definition in (2.6) is valid for countable Λ. Obtaining a site energy on the infinite domain can also be derived

by taking an appropriate sequence of finite domains Λ_R and considering the thermodynamic limit of the site energies along this sequence ([11], Lem. 3.1). This resolvent calculus approach for the tight binding model has been widely used [9, 11, 21, 36].

For finite Fermi-temperature, the site energies defined in (2.6) are exponentially localised [9, 11] in the sense of Proposition 2.2. The only difference between Proposition 2.2 and Lemma 7 of [9] or Lemma 2.1 of [11] is that we explicitly track the β-dependent constants in the estimates.

Proposition 2.2 (Finite Fermi-temperature locality for metals)**.**

(i) *Suppose $y\colon \Lambda \to \mathbb{R}^d$ and $\mathcal{H}(y)$ satisfy (**L**) and (**TB**), respectively. Then, for $1 \leqslant j \leqslant \nu$, there exist positive constants $C_j = C_j(\beta)$ and $\eta_j = \eta_j(\beta)$ such that*

$$\left| \frac{\partial^j G_\ell^\beta(y)}{\partial [y(m_1)]_{i_1} \dots \partial [y(m_j)]_{i_j}} \right| \leqslant C_j e^{-\eta_j \sum_{l=1}^j |y(\ell) - y(m_l)|} \tag{2.8}$$

for any $\ell, m_1, \dots, m_j \in \Lambda$ and $1 \leqslant i_1, \dots, i_j \leqslant d$.

(ii) *For all sufficiently large β, $C_j(\beta) = C\beta^\alpha$ where $C > 0$ depends only on $N_\mathrm{b}, \boldsymbol{h}, \mathfrak{m}$ and $\alpha > 0$ depends only on j and d. Further, $\eta_j(\beta) = c \min\{1, \beta^{-1}\}$ for some $c > 0$ depending only on $j, \boldsymbol{h}, \boldsymbol{\gamma}, \mathfrak{m}$ and d.*

Sketch of the proof. The β-dependence in the estimate (2.8) comes from the fact that the distance between the contour and the spectrum can, in general, only be bounded below by a constant multiple of β^{-1} as in (2.7). We summarise the main ideas in Section 5.2. □

For the case of insulators (where μ lies in a spectral gap), Proposition 2.2 can both be improved and extended to zero Fermi-temperature. In this case, the following constants are strictly positive

$$\mathsf{d}(y) \coloneqq \operatorname{dist}(\mu, \sigma(\mathcal{H}(y))) \quad \text{and} \tag{2.9}$$

$$\mathsf{g}(y) \coloneqq \inf(\sigma(\mathcal{H}(y)) \cap (\mu, +\infty)) - \sup(\sigma(\mathcal{H}(y)) \cap (-\infty, \mu)). \tag{2.10}$$

Using (2.10), and the fact that $z \mapsto 2(z - \mu)$ is analytic, means that the expression (2.6) holds for zero Fermi-temperature with a simple closed contour $\mathscr{C}_\infty$ encircling $\sigma(\mathcal{H}(y)) \cap (-\infty, \mu)$ and avoiding $\sigma(\mathcal{H}(y)) \cap (\mu, \infty)$; see Figure 1. Further, we may suppose that

$$\operatorname{dist}(z, \sigma(\mathcal{H}(y))) \geqslant \tfrac{1}{2}\mathsf{g}(y) \quad \forall z \in \mathscr{C}_\infty. \tag{2.11}$$

In this case, we obtain the following locality results that are uniform in Fermi-temperature:

Proposition 2.3 (Locality estimates for insulators)**.**

(i) *Suppose $y\colon \Lambda \to \mathbb{R}^d$ and $\mathcal{H}(y)$ satisfy (**L**) and (**TB**), respectively. Further, we assume that $\mu \notin \sigma(\mathcal{H}(y))$. Then, for $1 \leqslant j \leqslant \nu$, there exist positive constants C_j, η_j, such that*

$$\left| \frac{\partial^j G_\ell^\beta(y)}{\partial [y(m_1)]_{i_1} \dots \partial [y(m_j)]_{i_j}} \right| \leqslant C_j\, e^{-\eta_j \sum_{l=1}^j |y(\ell) - y(m_l)|} \tag{2.12}$$

for any $\ell, m_1, \dots, m_j \in \Lambda$ and $1 \leqslant i_1, \dots, i_j \leqslant d$.

(ii) *For $\beta < \infty$, the constants C_j and η_j depend on $\mathsf{d} = \mathsf{d}(y)$ from (2.9). For all sufficiently small d, $C_j(\mathsf{d}) = C\mathsf{d}^\alpha$ where $C > 0$ depends only on $N_\mathrm{b}, \boldsymbol{h}, \mathfrak{m}$ and $\alpha > 0$ depends only on j and d. Further, $\eta_j(\mathsf{d}) = c \min\{1, \mathsf{d}\}$ for some $c > 0$ depending only on $j, \boldsymbol{h}, \boldsymbol{\gamma}, \mathfrak{m}$ and d.*

(iii) *For $\beta = \infty$, the constants C_j and η_j depend on $\mathsf{g} = \mathsf{g}(y)$ from (2.10). For all sufficiently small g, $C_j(\mathsf{g}) = C\mathsf{g}^\alpha$ where $C > 0$ depends only on $N_\mathrm{b}, \boldsymbol{h}, \mathfrak{m}$ and $\alpha > 0$ depends only on j and d. Further, $\eta_j(\mathsf{g}) = c \min\{1, \mathsf{g}\}$ for some $c > 0$ depending only on $j, \boldsymbol{h}, \boldsymbol{\gamma}, \mathfrak{m}$ and d.*

Sketch of the proof. This result follows from the same arguments as in the proof of Proposition 2.2. In this case, the pre-factors and exponents are β-independent because the constants (2.9) and (2.10) are. Again, the main ideas are summarised in Section 5.2. □

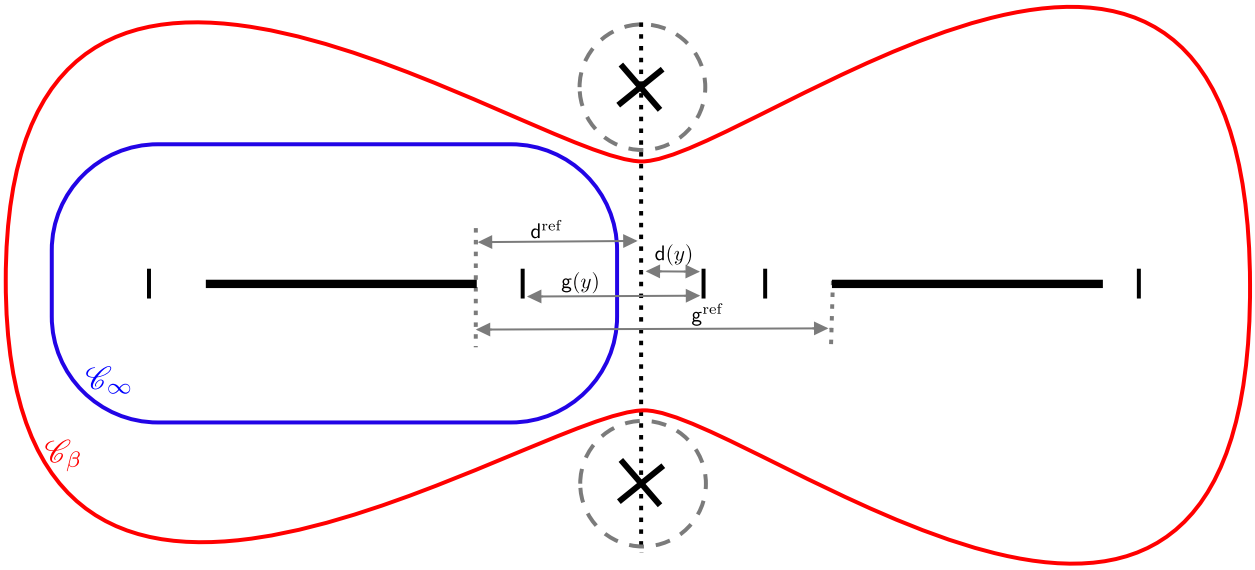


FIGURE 1. Cartoon depicting an approximation of $\sigma(\mathcal{H}(y))$ for $y \in \mathrm{Adm}(\Lambda)$ (in black on the real axis), see Lemma 2.5, and the contours $\mathscr{C}_\beta$ (in red) and $\mathscr{C}_\infty$ (in blue). The positive constants $\mathsf{d}(y), \mathsf{d}^{\rm ref}, \mathsf{g}(y)$ and $\mathsf{g}^{\rm ref}$ are also displayed. By (5.3), the finite Fermi-temperature contours avoid the balls of radius $\mathsf{b}\beta^{-1}$ (grey dashed circles) about $\mu \pm i\pi\beta^{-1}$ (shown with black crosses). The spectrum pictured in Figure 1 is qualitatively similar to that resulting from point defects in lattice structures.

2.3. Point defects

We suppose that a given reference configuration has a spectral gap and that the chemical potential is fixed within the gap. Then we show that point defect configurations introduce additional "defect states" into the system and also perturb the essential spectrum. The main result of this paper is that, within this setting, the locality results discussed in Section 2.2 are independent of discrete spectra inside the band gap in the sense of Theorems 2.7 and 2.9.

2.3.1. Reference configurations

In preparation for this result, we consider the Hamiltonian on a fixed reference configuration, $\Lambda^{\rm ref}$, given by $\big(\mathcal{H}^{\rm ref}\big)^{ab}_{\ell k} = h^{ab}_{\ell k}(\ell - k)$ for $\ell, k \in \Lambda^{\rm ref}$ and $1 \leqslant a, b \leqslant N_{\rm b}$. In order to keep the presentation abstract and the mathematical results general we will not explicitly define $\Lambda^{\rm ref}$, but we will always be thinking of a multi-lattice.

For the remainder of this paper, we make the following assumption:

(GAP). There exists a band gap in $\sigma(\mathcal{H}^{\rm ref})$ and the chemical potential, μ, lies in the interior of this band gap.

Under (**GAP**), we may introduce the following positive constants that will determine the interaction range in the improved locality estimates:

$$\mathsf{d}^{\rm ref} := \mathrm{dist}\big(\mu, \sigma(\mathcal{H}^{\rm ref})\big) \quad \text{and} \tag{2.13}$$

$$\mathsf{g}^{\rm ref} := \inf\big(\sigma(\mathcal{H}^{\rm ref}) \cap (\mu, +\infty)\big) - \sup\big(\sigma(\mathcal{H}^{\rm ref}) \cap (-\infty, \mu)\big). \tag{2.14}$$

Remark 2.4. In the case where $\Lambda^{\rm ref}$ is a multi-lattice formed by taking the union of finitely many copies of a Bravais lattice, $\mathbb{L}$, we may apply Bloch's theorem [24] to diagonalise $\mathcal{H}^{\rm ref}$ and write the spectrum as the union

of finitely many continuous energy bands:

$$\sigma(\mathcal{H}^{\mathrm{ref}}) = \bigcup_{\alpha} \{\varepsilon^{\alpha}(k) \colon k \in \mathrm{BZ}\} \tag{2.15}$$

where $\varepsilon^{\alpha} \colon \overline{\mathrm{BZ}} \to \mathbb{R}$ are continuous functions on the (closure of the) Brillouin zone.

2.3.2. Point defect reference configurations

From now on, we shall assume that Λ is a point defect reference configuration:

(P). Given a reference configuration $\Lambda^{\mathrm{ref}} \subset \mathbb{R}^d$, we suppose $\Lambda \subset \mathbb{R}^d$ is such that there exists a positive constant R_{def} with $\Lambda^{\mathrm{ref}} \setminus B_{R_{\mathrm{def}}} = \Lambda \setminus B_{R_{\mathrm{def}}}$ and $\#(B_{R_{\mathrm{def}}} \cap \Lambda) < \infty$.

We now introduce energy spaces of displacements which restricts the class of admissible configurations. Given $\ell \in \Lambda$ and $\rho \in \Lambda - \ell$, we define the finite difference $D_\rho u(\ell) \coloneqq u(\ell + \rho) - u(\ell)$. The full (infinite) finite difference stencil is then defined to be $Du(\ell) \coloneqq (D_\rho u(\ell))_{\rho \in \Lambda - \ell}$. For $\Upsilon > 0$, the ℓ^2_Υ semi-norm on the full interaction stencil is given by

$$\|Du\|_{\ell^2_\Upsilon} \coloneqq \left(\sum_{\ell \in \Lambda} \sum_{\rho \in \Lambda - \ell} e^{-2\Upsilon|\rho|} |D_\rho u(\ell)|^2 \right)^{1/2}.$$

All of the semi-norms $\|D \cdot\|_{\ell^2_\Upsilon}$ for $\Upsilon > 0$ are equivalent [12] and so we fix an exponent $\Upsilon > 0$ for the remainder of this paper and define the following function space of finite energy displacements:

$$\dot{\mathscr{W}}^{1,2}(\Lambda) \coloneqq \{u \colon \Lambda \to \mathbb{R}^d \colon \|Du\|_{\ell^2_\Upsilon} < \infty\}.$$

We hence define the space of admissible configurations by

$$\mathrm{Adm}(\Lambda) \coloneqq \left\{ y \in x + \dot{\mathscr{W}}^{1,2}(\Lambda) \colon y \text{ satisfies } (\mathbf{L}) \right\}$$

where $x \colon \Lambda \to \Lambda$ denotes the identity configuration.

For $y \in \mathrm{Adm}(\Lambda)$, the spectrum, $\sigma(\mathcal{H}(y))$, can be related to $\sigma(\mathcal{H}^{\mathrm{ref}})$:

Lemma 2.5 (Decomposition of the spectrum)**.** *Fix $y \in \mathrm{Adm}(\Lambda)$. Then, for all $\delta > 0$, there exists $R_\delta > 0$ such that $\#\big(\sigma(\mathcal{H}(y)) \setminus B_\delta(\sigma(\mathcal{H}^{\mathrm{ref}}))\big) \leqslant R_\delta$.*

Sketch of the proof. This result is slightly stronger than Weyl's theorem on the stability of the essential spectrum under compact perturbations. See Section 5.3 for a full proof. □

Remark 2.6. In the case that Λ^{ref} is a multi-lattice, $\sigma(\mathcal{H}^{\mathrm{ref}})$ is banded in the sense of (2.15). This means Lemma 2.5 states that point defects give rise to a finite number of "defect states" located away from the spectral bands as depicted in Figure 1.

2.3.3. Locality of site energies

Lemma 2.5, together with a locality result for the spectral projection (see Lem. 5.3) corresponding to the finitely many eigenvalues bounded away from the spectral bands, allows us to approximate $(\mathcal{H}(y) - z)^{-1}$ in terms of the reference resolvent, $(\mathcal{H}^{\mathrm{ref}} - z)^{-1}$. This means we can apply the existing locality estimates of Proposition 2.3 on the reference spectrum. The approximation does not affect the exponent in the estimates and only increases the constant pre-factor. We show that the pre-factor may be chosen to depend on the atomic sites and this converges exponentially to the corresponding pre-factor in the defect-free case, as we send the atomic sites away from the defect core. That is, away from the defect, the locality estimates resemble the corresponding estimates for the reference configuration.

Theorem 2.7 (Improved finite Fermi-temperature locality)**.**

(i) *Fix $y \in \mathrm{Adm}(\Lambda)$. Then, for $1 \leqslant j \leqslant \nu$, $\ell \in \Lambda$, $\boldsymbol{m} = (m_1, \dots, m_j) \in \Lambda^j$ and $1 \leqslant i_1, \dots, i_j \leqslant d$, there exists a positive constant $C_{\beta j} = C_{\beta j}(\ell, \boldsymbol{m})$ such that*

$$\left| \frac{\partial^j G_\ell^\beta(y)}{\partial [y(m_1)]_{i_1} \dots \partial [y(m_j)]_{i_j}} \right| \leqslant C_{\beta j} e^{-\eta_j^{\mathrm{ref}} \sum_{l=1}^j |y(\ell) - y(m_l)|} \tag{2.16}$$

where $\eta_j^{\mathrm{ref}} := c \min\{1, \mathsf{d}^{\mathrm{ref}}\}$ and $c > 0$ depends only on $j, \boldsymbol{h}, \boldsymbol{\gamma}, \mathfrak{m}$ and d.

(ii) *$C_{\beta j}(\ell, \boldsymbol{m})$ is uniformly bounded independently of ℓ and $\boldsymbol{m}$. Let $C_{\beta j}^{\mathrm{ref}} := C_{\beta j}$ when $\Lambda = \Lambda^{\mathrm{ref}}$ and $y = x$. If $\ell, m_1, \dots, m_j \in B_R(\xi)$ for some $R > 0$, then $C_{\beta j}(\ell, \boldsymbol{m}) \to C_{\beta j}^{\mathrm{ref}}$ as $|\xi| \to \infty$, with an exponential rate.*

(iii) *If $\mu \notin \sigma(\mathcal{H}(y))$, then $C_{\beta j}$ can be chosen to be β-independent. On the other hand, if $\mu \in \sigma(\mathcal{H}(y))$, then $C_{\beta j} = C\beta^{j-1}$ for some $C > 0$ depending only on $j, \boldsymbol{h}, \boldsymbol{\gamma}, N_{\mathrm{b}}, \mathfrak{m}$ and d.*

Remark 2.8. By (iii), if $\mu \in \sigma(\mathcal{H}(y))$, then the first derivatives of the site energies are uniformly bounded in β despite the fact that the zero Fermi-temperature site energies are not differentiable. In fact, the point-wise limit of the first derivatives as $\beta \to \infty$ exist.

We have the following analogous zero Fermi-temperature result:

Theorem 2.9 (Improved zero Fermi-temperature locality)**.**

(i) *Fix $y \in \mathrm{Adm}(\Lambda)$ with $\mu \notin \sigma(\mathcal{H}(y))$. Then, for $1 \leqslant j \leqslant \nu$, $\ell \in \Lambda, \boldsymbol{m} = (m_1, \dots, m_j) \in \Lambda^j$, and $1 \leqslant i_1, \dots, i_j \leqslant d$, there exists a positive constant $C_{\infty j}(\ell, \boldsymbol{m})$ such that*

$$\left| \frac{\partial^j G_\ell^\infty(y)}{\partial [y(m_1)]_{i_1} \dots \partial [y(m_j)]_{i_j}} \right| \leqslant C_{\infty j}(\ell, \boldsymbol{m})\, e^{-\eta_{\infty j}^{\mathrm{ref}} \sum_{l=1}^j |y(\ell) - y(m_l)|} \tag{2.17}$$

where $\eta_{\infty j}^{\mathrm{ref}} := c \min\{1, \mathsf{g}^{\mathrm{ref}}\}$ and $c > 0$ depends only on $j, \boldsymbol{h}, \boldsymbol{\gamma}, \mathfrak{m}$ and d.

(ii) *$C_{\infty j}(\ell, \boldsymbol{m})$ is uniformly bounded independently of ℓ and $\boldsymbol{m}$. Let $C_{\infty j}^{\mathrm{ref}} := C_{\infty j}$ when $\Lambda = \Lambda^{\mathrm{ref}}$ and $y = x$. If $\ell, m_1, \dots, m_j \in B_R(\xi)$ for some $R > 0$, then $C_{\infty j}(\ell, \boldsymbol{m}) \to C_{\infty j}^{\mathrm{ref}}$ as $|\xi| \to \infty$, with an exponential rate.*

Remark 2.10. In the case $j = 1$, Theorem 2.7 part (ii) takes the form

$$|C_{\beta 1}(\ell, m) - C_{\beta 1}^{\mathrm{ref}}| \lesssim e^{-\eta_1^{\mathrm{ref}}(|y(\ell)| + |y(m)| - |y(\ell) - y(m)|)}.$$

Similarly for Theorem 2.9 with $\beta = \infty$ and the exponent $\eta_{\infty 1}^{\mathrm{ref}}$. For higher derivatives, the relationship between ℓ and $\boldsymbol{m}$ is more complicated.

3. Numerical tests

In this section, we present numerical simulations to support our analytical results. We use a practical tight binding model, the NRL model [13, 27, 31], to test the force-locality in bulk carbon and silicon, both with and without an interstitial defect. Since we are unaware of established codes that compute site energies and their derivatives, we implemented these models in the `Julia` package `SKTB.jl` [8].

3.1. The NRL tight binding model

The NRL tight binding model, developed by Cohen, Mehl, and Papaconstantopoulos [13], is slightly more general than our formulation in Section 2. It is non-orthogonal, which means that the energy levels are now determined by the generalised eigenvalue problem

$$\mathcal{H}(y)\psi_s = \lambda_s \mathcal{M}(y)\psi_s \quad \text{where} \quad \psi_s^{\mathrm{T}} \mathcal{M}(y)\psi_s = 1, \tag{3.1}$$

which has an additional overlap matrix $\mathcal{M}(y)$ compared with (2.3). Furthermore, the NRL hamiltonian and overlap matrices are construct both from hopping elements as in (2.1) as well as on-site matrix elements as a function of the local environment. For carbon and silicon they are parameterised as follows (for other elements the parameterisation is similar):

To define the on-site terms, each atom ℓ is assigned a pseudo-atomic density

$$\rho_\ell := \sum_k e^{-\lambda^2 r_{\ell k}} f_{\rm c}(r_{\ell k}),$$

where the sum is over all of the atoms k within the cutoff $R_{\rm c}$ of atom ℓ, λ is a fitting parameter, $f_{\rm c}$ is a cutoff function

$$f_{\rm c}(r) = \frac{\theta(R_{\rm c} - r)}{1 + \exp\big((r - R_{\rm c})/l_c + L_c\big)},$$

with θ the step function, and the parameters $l_c = 0.5$, $L_c = 5.0$ for most elements. Although, in principle, the on-site terms should have off-diagonal elements, the NRL model follows traditional practice and only include the diagonal terms. Then, the on-site terms for each atomic site ℓ are given by

$$\mathcal{H}(y)_{\ell\ell}^{\upsilon\upsilon} := a_\upsilon + b_\upsilon \rho_\ell^{2/3} + c_\upsilon \rho_\ell^{4/3} + d_\upsilon \rho_\ell^2, \tag{3.2}$$

where $\upsilon = s, p$, or d is the index for angular-momentum-dependent atomic orbitals and (a_υ), (b_υ), (c_υ), (d_υ) are fitting parameters. The on-site elements for the overlap matrix are simply taken to be the identity matrix.

The off-diagonal NRL Hamiltonian entries follow the formalism of Slater and Koster who showed in [35] that all two-centre (spd) hopping integrals can be constructed from ten independent "bond integral" parameters $h_{\upsilon\upsilon'\mu}$, where

$$(\upsilon\upsilon'\mu) = ss\sigma,\ sp\sigma,\ pp\sigma,\ pp\pi,\ sd\sigma,\ pd\sigma,\ pd\pi,\ dd\sigma,\ dd\pi,\ \text{and}\ dd\delta.$$

The NRL bond integrals are given by

$$h_{\upsilon\upsilon'\mu}(r) := \big(e_{\upsilon\upsilon'\mu} + f_{\upsilon\upsilon'\mu} r + g_{\upsilon\upsilon'\mu} r^2\big) e^{-h_{\upsilon\upsilon'\mu} r} f_{\rm c}(r) \tag{3.3}$$

with fitting parameters $e_{\upsilon\upsilon'\mu}, f_{\upsilon\upsilon'\mu}, g_{\upsilon\upsilon'\mu}, h_{\upsilon\upsilon'\mu}$. The matrix elements $\mathcal{H}(y)_{\ell k}^{\upsilon\upsilon'}$ are constructed from the $h_{\upsilon\upsilon'\mu}(r)$ by a standard procedure [35].

The analogous bond integral parameterisation of the overlap matrix is given by

$$m_{\upsilon\upsilon'\mu}(r) := \big(\delta_{\upsilon\upsilon'} + p_{\upsilon\upsilon'\mu} r + q_{\upsilon\upsilon'\mu} r^2 + r_{\upsilon\upsilon'\mu} r^3\big) e^{-s_{\upsilon\upsilon'\mu} r} f_{\rm c}(r) \tag{3.4}$$

with the fitting parameters $(p_{\upsilon\upsilon'\mu}), (q_{\upsilon\upsilon'\mu}), (r_{\upsilon\upsilon'\mu}), (s_{\upsilon\upsilon'\mu})$ and $\delta_{\upsilon\upsilon'}$ the Kronecker delta function.

The fitting parameters in the foregoing expressions are determined by fitting to some high-symmetry first-principle calculations: In the NRL method, a database of eigenvalues (band structures) and total energies were constructed for several crystal structures at several volumes. Then the parameters are chosen such that the eigenvalues and energies in the database are reproduced. For practical simulations, the parameters for different elements can be found in [31].

3.2. Test systems

Our two test systems are diamond cubic bulk carbon and bulk silicon, which provide ideal test cases of our theory due to their clearly defined band gaps. Since carbon has a much larger band gap than silicon we will also be able to test how this affects locality of interaction.

For both elements, we simulate a supercell model (*i.e.* using periodic boundary conditions) consisting of $5 \times 5 \times 5$ diamond cubic unit cells, containing 1000 atoms in total. First, we use the NRL tight binding model

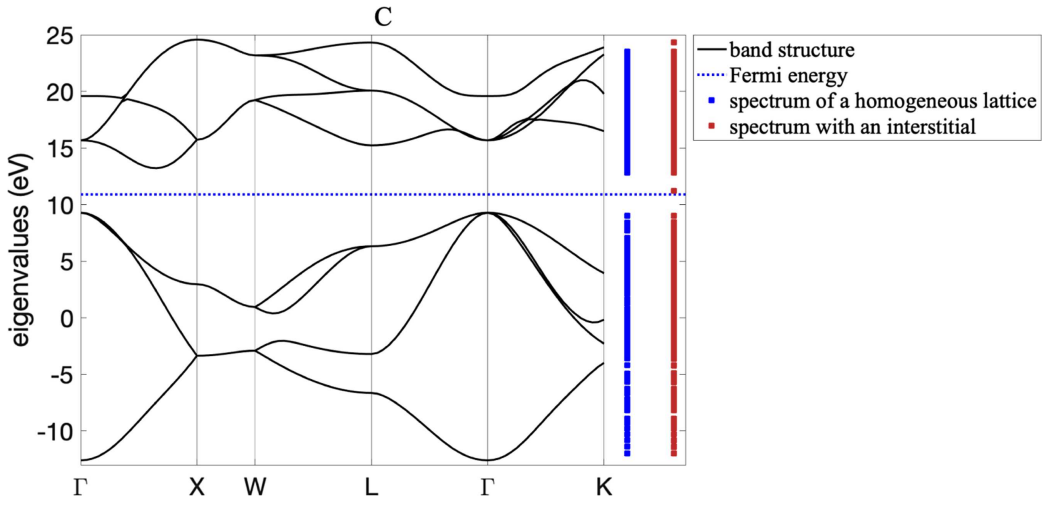


FIGURE 2. Band structure of C; spectrum of the homogeneous lattice (supercell approximation) and defective system.

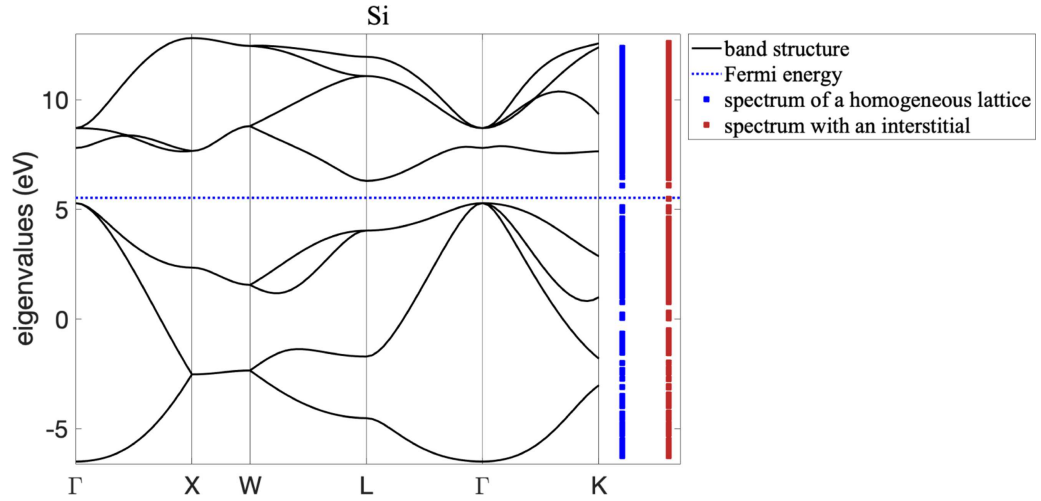
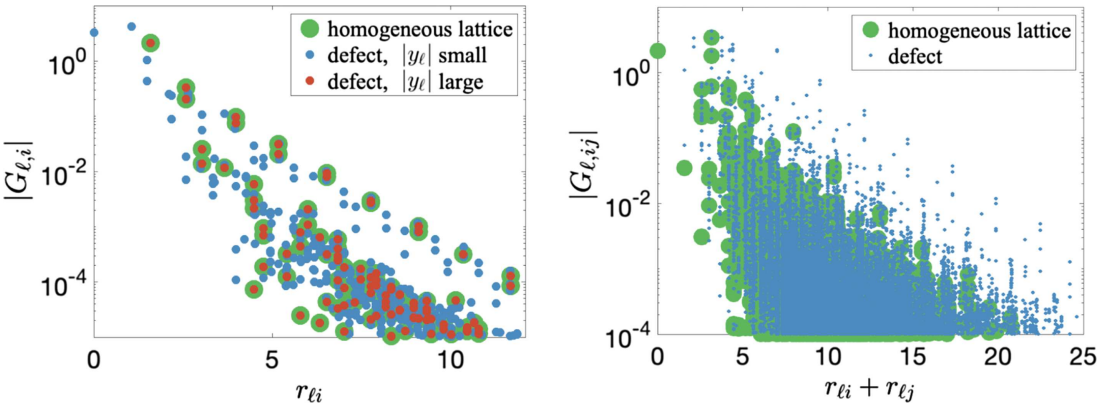


FIGURE 3. Band structure of Si, spectrum of the homogeneous lattice (supercell approximation) and defective system.

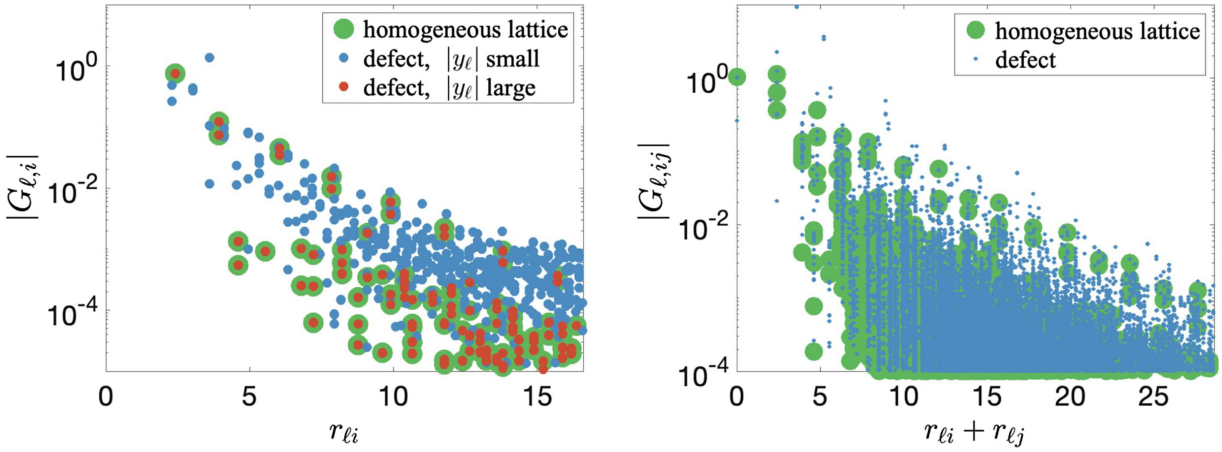
to relax the cells to their ground states (this only rescales the cells but does not change their shape). We then compute the band structures which are, respectively, shown in Figures 2 and 3. We verified our implementation by comparing the band structure for the silicon model against that published in [32]. The Fermi energy is chosen to be the mid point between the highest occupied state and the lowest unoccupied state of the homogeneous 1000-atom system. For both systems we observe clearly defined band gaps around the Fermi energy, approximately 0.98 eV for Si and 3.83 eV for C.

Next, we create a self-interstitial near the origin, and observe (Figs. 2 and 3) the expected pollution of the band gap in the defected system. By tweaking the position of the interstitial we are able to create configurations



(A) (B)

FIGURE 4. Carbon: locality of site energies in homogeneous lattice and defective system. (a) Decay of site energy derivatives. (b) Decay of site energy hessians.



(A) (B)

FIGURE 5. Silicon: locality of site energies in homogeneous lattice and defective system. (a) Decay of site energy derivatives. (b) Decay of site energy hessians.

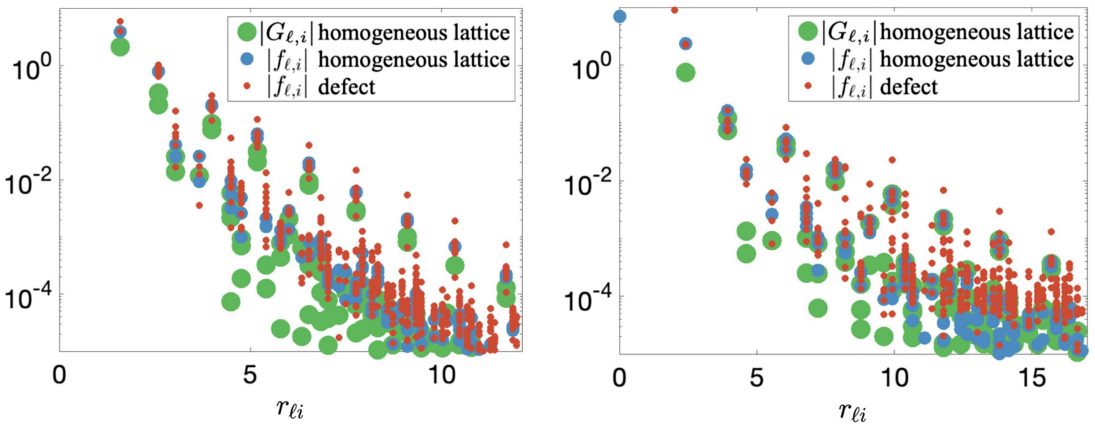
where an eigenvalue is arbitrarily close to the Fermi-energy in order to provide a challenging situation to confirm the result of Theorem 2.9.

3.3. Site energy locality

To test the locality of interatomic interaction we evaluate all first and second site energy derivatives $G_{\ell,j} = \partial_{R_j} G_\ell$ and $G_{\ell,ij} = \partial_{R_i}\partial_{R_j} G_\ell$ in both the homogeneous and defective system, and plot the data points

$$\big(r_{\ell j}, |G_{\ell,j}|\big) \quad \text{and} \quad \big(r_{\ell i} + r_{\ell j}, |G_{\ell,ij}|\big)$$

in Figures 4 and 5. For the homogeneous systems all sites are equivalent, hence we only plot the site energy derivatives for a single site. For the defective systems we plot the data points for the interstitial site itself ("$|y_\ell|$



(A) (B)

FIGURE 6. The decay of force derivatives in homogeneous lattice and defective system. (a) Carbon. (b) Silicon.

small") as well as for the site in the computational cell that has the largest distance to the interstitial atom ("$|y_\ell|$ large").

We clearly observe the exponential decay of interaction strength as predicted in Theorem 2.9. Moreover, we also observe that for sites ℓ far from the defect the site derivative decay perfectly matches that of the bulk system.

Two additional observations were unexpected for us: (1) the decay of site derivatives for "near-defect sites" does not exhibit the increased prefactor that we predicted; however we do see this increase in the second derivatives. (2) the decay of interaction in the silicon system is nearly identical (after rescaling by the lattice constants) to the carbon system even though silicon has a much smaller band gap.

These observations suggest that there are further effects leading to improved locality of interaction that our analysis does not fully capture. While a possible explanation is that the locality of the bond integral functions dominates the locality of the resolvents, this does not explain the excellent locality of the Si systems which have a fairly small band gap.

3.4. Force locality

Finally, we compare the decay of site energy derivatives to the decay of force derivatives. The reason for this additional test is that our definition of a site-energy is somewhat arbitrary. Indeed, there are infinitely many possible decompositions of total energy into site energies and each choice may lead to a different rate of decay of the interaction. Forces, on the other hand, are uniquely defined. Their locality is therefore "canonical" and provides a limit for the locality of site energies.

In Figure 6, we compare the decay of site energy derivatives and force derivatives. We evaluate the force derivatives $f_{\ell,j} = \partial_{R_j} f_\ell$, where the force is defined by the (negative) derivative of the total energy $f_\ell = -\partial_{R_\ell} G$, and plot the data points

$$\big(r_{\ell j}, |f_{\ell,j}|\big) \quad \text{and} \quad \big(r_{\ell i} + r_{\ell j}, |G_{\ell,ij}|\big)$$

in Figure 6. We observe that the site energy locality matches force locality very closely, which suggests that our choice of site energies leads to near-optimal locality of interaction.

4. Conclusions

We have extended the results of [9] to the zero Fermi-temperature case under the assumption that the chemical potential is not an eigenvalue of the Hamiltonian. We have described a site energy decomposition for a zero Fermi-temperature linear tight binding model and shown that the site contributions are exponentially localised. Most importantly, we have shown that the exponents in these estimates are independent of the discrete spectrum inside the band gap caused by point defects, and even the pre-factors converge to the pre-factors that would result from using the homogeneous site energy in the estimates, as the distance of a site to the defect increases. Our numerical results in Section 3 strongly support our analysis, but also point to possible further extensions in particular in the limit of small band gaps where our results may not yet be sharp.

The same analysis was also applied to the Helmholtz free energy in the canonical ensemble under the assumption that the Fermi level is fixed. In particular, this improves the locality results of [9] for insulators. Moreover, the analysis carries over to other quantities of interest as in [11].

The results of this paper allow us to formulate zero Fermi-temperature lattice relaxation as a variational problem on the energy space of displacements. In particular, for $y \in \mathrm{Adm}(\Lambda)$, we can define

$$\mathcal{G}^\infty(y) \coloneqq \sum_\ell (G^\infty_\ell(y) - G^\infty_\ell(x))$$

where $x\colon \Lambda \to \Lambda$ is the identity configuration. This *grand potential difference functional* is well defined if $\mu \notin \sigma(\mathcal{H}(y))$. This can be shown by applying results of [12] together with the locality estimates of this paper. We can then consider the zero Fermi-temperature limit problem:

$$\overline{y} \in \arg\min\{\mathcal{G}^\infty(y)\colon y \in \mathrm{Adm}(\Lambda),\ \mathrm{dist}(\mu, \sigma(\mathcal{H}(y))) > 0\} \tag{4.1}$$

where "arg min" denotes the set of local minimisers. The locality results presented in this paper allow us to show that the site energies and their derivatives converge exponentially quickly in the zero Fermi-temperature limit. This observation allows us to prove that (4.1) is indeed the limiting model of analogous finite Fermi-temperature models. Rigorous results are presented in a forthcoming paper [30].

5. Proofs of the main results

5.1. Definition of the site energies

Before we begin the proof of the locality estimates, we need to show that the definition of the finite Fermi-temperature site energy is valid. That is, we need $\mathfrak{g}^\beta(\,\cdot\,;\mu)$ to extend to a holomorphic function on some open neighbourhood of the spectrum and we need to consider appropriate contours $\mathscr{C}_\beta$, $\mathscr{C}_\infty$.

Lemma 5.1 (Analytic continuation of $\mathfrak{g}^\beta(z;\mu)$)**.** *Fix $\beta \in (0,\infty)$. Then, $z \mapsto \mathfrak{g}^\beta(z;\mu)$ can be analytically continued to the set $\mathbb{C} \setminus \{\mu + ir\colon r \in \mathbb{R}, |r| \geqslant \pi\beta^{-1}\}$.*

Proof. Extending $\mathfrak{g}^\beta(\,\cdot\,;\mu)$ into the complex plane amounts to choosing a branch cut of the complex logarithm. For each $n \in \mathbb{N}$, we define,

$$\begin{aligned} &\mathfrak{g}^\beta_n(z;\mu) \coloneqq \frac{2}{\beta}\big[\log|1 - f_\beta(z-\mu)| + i\mathrm{Arg}_n(1 - f_\beta(z-\mu))\big] \quad \text{where} \\ &\mathrm{Arg}_n(z) = \mathrm{Arg}(z) \pmod{2\pi} \quad \text{and} \quad \mathrm{Arg}_n(z) \in ((n-1)\pi, (n+1)\pi]. \end{aligned} \tag{5.1}$$

Choosing the principal branch of the complex logarithm, we get $\mathfrak{g}^\beta_0$ which agrees with $\mathfrak{g}^\beta$ on the real axis.

To simplify notation, and without loss of generality, we suppose $\mu = 0$.

It is clear that the Fermi-Dirac distribution has isolated singularities at $i(2k+1)\pi\beta^{-1}$ for all $k \in \mathbb{Z}$ (that is, when $e^{\beta z} = -1$) and is holomorphic away from these singularities. Therefore, $\mathfrak{g}^\beta_n(\,\cdot\,;0)$ is holomorphic on the

set that avoids the branch cut of the complex logarithm and the non-analyticity of $1 - f_\beta$. That is, $\mathfrak{g}_n^\beta(\,\cdot\,;0)$ is holomorphic on

$$\begin{cases} \{z \in \mathbb{C}\colon 1 - f_\beta(z) \notin (-\infty, 0]\} \setminus \left\{\frac{(2k+1)\pi i}{\beta}\right\}_{k\in\mathbb{Z}} & \text{for } n \text{ even}, \\ \{z \in \mathbb{C}\colon 1 - f_\beta(z) \notin [0, \infty)\} \setminus \left\{\frac{(2k+1)\pi i}{\beta}\right\}_{k\in\mathbb{Z}} & \text{for } n \text{ odd}. \end{cases}$$

Rewriting $1 - f_\beta$ we obtain,

$$1 - f_\beta(z) = \frac{e^{\beta z}}{1 + e^{\beta z}} = \frac{e^{\beta z}\big(1 + e^{\beta \overline{z}}\big)}{|1 + e^{\beta z}|^2} = \frac{e^{\beta \operatorname{Re}(z)}}{|1 + e^{\beta z}|^2}\Big(e^{i\beta \operatorname{Im}(z)} + e^{\beta \operatorname{Re}(z)}\Big). \tag{5.2}$$

The factor, $e^{\beta \operatorname{Re}(z)}|1 + e^{\beta z}|^{-2}$, is real and positive and so $1 - f_\beta(z)$ avoids the branch cut if and only if $h(z) \coloneqq e^{i\beta \operatorname{Im}(z)} + e^{\beta \operatorname{Re}(z)}$ does. Now, $h(z) \in (-\infty, 0]$ if and only if $\operatorname{Re}(z) \leqslant 0$ and $\beta \operatorname{Im}(z) = (2k+1)\pi$ for some $k \in \mathbb{Z}$. On the other hand $h(z) \in [0, \infty)$ if and only if $\beta \operatorname{Im}(z) = 2k\pi$ for some $k \in \mathbb{Z}$. We can therefore conclude that $\mathfrak{g}_0^\beta(\,\cdot\,;0)$ is holomorphic on

$$A_\beta^0 \coloneqq \Big\{z \in \mathbb{C}\colon \operatorname{Re}(z) > 0\Big\} \cup \Big\{z \in \mathbb{C}\colon \beta \operatorname{Im}(z) \in (-\pi, \pi)\Big\}$$

and that $\mathfrak{g}_n^\beta(\,\cdot\,;0)$ (for $n \neq 0$) is holomorphic on the set

$$A_\beta^n \coloneqq \Big\{z \in \mathbb{C}\colon \operatorname{Re}(z) < 0,\, \beta \operatorname{Im}(z) \in ((n-1)\pi, (n+1)\pi)\Big\}.$$

Since $A_\beta^n \cap A_\beta^{n+1} = \{z \in \mathbb{C}\colon \operatorname{Re}(z) < 0,\, \beta \operatorname{Im}(z) \in (n\pi, (n+1)\pi)\}$, and by (5.2), we have that

$$\operatorname{Arg}_n(1 - f_\beta(z)) = \operatorname{Arg}_{n+1}(1 - f_\beta(z)) \in (n\pi, (n+1)\pi]$$

for all $z \in A_\beta^n \cap A_\beta^{n+1}$. That is, $\mathfrak{g}_n^\beta(\,\cdot\,;0) = \mathfrak{g}_{n+1}^\beta(\,\cdot\,;0)$ on $A_\beta^n \cap A_\beta^{n+1}$.

We may therefore consider the analytic continuation of $\mathfrak{g}_n^\beta(\,\cdot\,;0)$ to $A_\beta^n \cup A_\beta^{n+1}$. We do this for each $n \in \mathbb{Z}$ and conclude since $\bigcup_{n\in\mathbb{Z}} A_\beta^n = \mathbb{C} \setminus \{ir\colon |r| \geqslant \pi\beta^{-1}\}$. □

From now on, we denote the analytic continuation by $z \mapsto \mathfrak{g}^\beta(z;\mu)$. We need conditions on the family of contours, $\{\mathscr{C}_\beta\}_\beta$, to ensure that $\mathfrak{g}^\beta(z;\mu)$ remains uniformly bounded for $z \in \mathscr{C}_\beta$ and $\beta > 0$. We suppose there exists some β-independent constant $0 < \mathsf{b} < \pi$ such that

$$\begin{gathered} \operatorname{dist}\big(z, \{\mu \pm i\pi\beta^{-1}\}\big) \geqslant \mathsf{b}\beta^{-1} \quad \forall z \in \mathscr{C}_\beta \quad \text{and} \\ \exists A \subset \mathbb{C} \text{ bounded s.t. } \mathscr{C}_\beta \subset A \end{gathered} \tag{5.3}$$

for all $\beta > 0$.

Lemma 5.2. *Fix $y \in \operatorname{Adm}(\Lambda)$. Suppose that $\{\mathscr{C}_\beta\}_\beta$ is a family of simple closed contours encircling $\sigma(\mathcal{H}(y))$ and satisfying (5.3). Then, for $\beta_0 > 0$,*

$$\sup_{\beta \geqslant \beta_0} \sup_{z \in \mathscr{C}_\beta} |\mathfrak{g}^\beta(z;\mu)| < \infty.$$

Proof. Since A is bounded, we can find a strip of width $r > 0$ about the real axis containing A. This means that for fixed $\beta > 0$, the number branches of the complex logarithm that we must consider, as in (5.1), in order to have extended $\mathfrak{g}^\beta(\,\cdot\,;\mu)$ to the whole of A is at most a constant multiple of $\frac{r\beta}{\pi}\cdot$ This means that

$$|\operatorname{Arg}_n(1 - f_\beta(z - \mu))| \leqslant (n+1)\pi \leqslant Cr\beta$$

for all n such that $A \cap \{z \in \mathbb{C}\colon \beta \operatorname{Im}(z) \in ((n-1)\pi, (n+1)\pi)\} \neq \emptyset$ and $z \in A_\beta^n$. Therefore, for $z \in A$ and $\beta > 0$, we have that $\operatorname{Im}\big(\mathfrak{g}^\beta(z;\mu)\big)$ is bounded on A independently of β.

Fix $\beta > \beta_0$. Now we show that, away from the singularities, $\operatorname{Re}\big(\mathfrak{g}^\beta(\,\cdot\,;\mu)\big)$ is uniformly bounded. We know that $\operatorname{Re}\big(\mathfrak{g}^\beta(z;\mu)\big) = 2\operatorname{Re}(z-\mu) - \frac{2}{\beta}\log\big|1+e^{\beta(z-\mu)}\big|$ and $2(z-\mu)$ is uniformly bounded on A. Moreover,

$$\tfrac{2}{\beta}\log\Big|1+e^{\beta(z-\mu)}\Big| \leqslant \tfrac{2}{\beta}\log\Big(1+\exp\Big(\beta \sup_{z\in A}|\operatorname{Re}(z)-\mu|\Big)\Big) \leqslant C$$

for some $C > 0$ depending only on A and β_0. Therefore, all that is left to show is that $|1+e^{\beta(z-\mu)}|$ is uniformly bounded below by a positive constant. If $\operatorname{Re}(z-\mu) < -c\beta^{-1}$ for some $c > 0$ then $|1+e^{\beta(z-\mu)}| \geqslant 1-e^{-c} > 0$ and if $\operatorname{Re}(z-\mu) > c\beta^{-1}$ then $|1+e^{\beta(z-\mu)}| \geqslant e^c - 1 > 0$. On the other hand, if $|\beta\operatorname{Im}(z-\mu) - r| \geqslant \theta$ for all $r \in \mathbb{R}$ such that $|r| \geqslant \pi$, then $|1+e^{\beta(z-\mu)}| \geqslant \tan(\theta) > 0$. □

5.2. Proof of Propositions 2.2 and 2.3: locality estimates

We now briefly sketch the main ideas in the proof of Propositions 2.2 and 2.3. We mainly do this so that we can track the β-dependent constants in the proof. A key ingredient is the following Combes–Thomas type estimate on the resolvent [14]:

Lemma 5.3 (Combes–Thomas). *Fix $y \in \operatorname{Adm}(\Lambda)$ and $z \in \mathbb{C}$ such that $\operatorname{dist}(z, \sigma(\mathcal{H}(y))) \geqslant \mathfrak{d}$ for some $\mathfrak{d} > 0$. Then,*

$$\Big|\Big[(\mathcal{H}(y)-z)^{-1}\Big]_{\ell k}^{ab}\Big| \leqslant \tfrac{2}{\mathfrak{d}} e^{-\gamma_{\mathrm{CT}}(\mathfrak{d})|y(\ell)-y(k)|},$$

where $\gamma_{\mathrm{CT}}(\mathfrak{d}) \coloneqq c\min\{1,\mathfrak{d}\}$ for some $c > 0$ depending only on $h_0, \gamma_0, \mathfrak{m}$ and d.

Proof. This follows the proof of Lemma 6 from [9] and the main ideas of [36]. The claimed $\mathfrak{d}$ dependence in the exponent can be obtained by replacing (34) in Lemma 6 of [9] with the following sharper estimate: there exists a $C > 0$ such that

$$\sup_{\ell\in\Lambda}\sum_{k\in\Lambda} h_0 e^{-\gamma_0|y(\ell)-y(k)|}\Big(e^{\gamma_{\mathrm{CT}}|y(\ell)-y(k)|}-1\Big) \leqslant C\gamma_{\mathrm{CT}} \tag{5.4}$$

for all $0 \leqslant \gamma_{\mathrm{CT}} \leqslant \frac{1}{2}\gamma_0$. To conclude, we note that in the proof of Lemma 6 from [9], $\gamma_{\mathrm{CT}} > 0$ must be chosen sufficiently small such that the right hand side of (5.4) is less than $\frac{1}{2}\mathfrak{d}$. □

To simplify notation, we shall write $r_{\ell k}(y) \coloneqq |y(\ell) - y(k)|$ for the distance between two atomic sites and $\mathscr{R}_z(y) \coloneqq (\mathcal{H}(y)-z)^{-1}$ for the resolvent operator corresponding to $\mathcal{H}(y)$. We will drop the argument (y) in $r_{\ell k}(y)$, the resolvent and Hamiltonian when the dependence on y is clear from context. Moreover, we shall use the following shorthand for derivatives of the Hamiltonian: for $\boldsymbol{m} = (m_1,\dots,m_j) \in \Lambda^j$ and $\boldsymbol{i} = (i_1,\dots,i_j) \in \big(\mathbb{R}^d\big)^j$, define

$$[\mathcal{H}_{,\boldsymbol{m}}]_{\boldsymbol{i}} \coloneqq \frac{\partial^j \mathcal{H}(y)}{\partial[y(m_1)]_{i_1}\dots\partial[y(m_j)]_{i_j}}.$$

Often, to simplify notation further, we will drop the Euclidean coordinate in this notation.

Before we prove the locality estimates of Propositions 2.2 and 2.3, we shall derive general bounds for the first and second derivatives of the resolvent. We choose $z \in \mathbb{C}$ such that

$$\operatorname{dist}(z, \sigma(\mathcal{H}(y))) \geqslant \mathfrak{d} \tag{5.5}$$

for some $\mathfrak{d} > 0$. Moreover, we let $\gamma_{\mathrm{CT}} = \gamma_{\mathrm{CT}}(\mathfrak{d}) > 0$ be the corresponding Combes–Thomas exponent from Lemma 5.3 and define $\mathfrak{d}_j \coloneqq \min\{\gamma_1,\dots,\gamma_j,\gamma_{\mathrm{CT}}\}$ for each j. By applying the Combes–Thomas estimate and

using the regularity of the Hamiltonian, we have

$$\begin{aligned}\left|\frac{\partial[\mathscr{R}_z(y)]^{aa}_{\ell\ell}}{\partial[y(m)]_i}\right| &= \left|\sum_{\substack{\ell_1,\ell_2\in\Lambda\\ 1\leqslant b,c\leqslant N_{\rm b}}} [\mathscr{R}_z]^{ab}_{\ell\ell_1}\big([\mathcal{H}_{,m}]_i\big)^{bc}_{\ell_1\ell_2}[\mathscr{R}_z]^{ca}_{\ell_2\ell}\right| \\ &\leqslant 4N_{\rm b}^2 h_1 \mathfrak{d}^{-2}\sum_{\ell_1,\ell_2\in\Lambda} e^{-\gamma_{\rm CT} r_{\ell_1\ell}} e^{-\gamma_1\left(r_{\ell_1 m}+r_{m\ell_2}\right)} e^{-\gamma_{\rm CT} r_{\ell_2\ell}} \\ &\leqslant C\mathfrak{d}^{-2}\left(\sum_{\ell_1\in\Lambda} e^{-\mathfrak{d}_1\left(r_{\ell\ell_1}+r_{\ell_1 m}\right)}\right)^2 \\ &\leqslant C\mathfrak{d}^{-2}\mathfrak{d}_1^{-2d} e^{-\mathfrak{d}_1 r_{\ell m}}.\end{aligned} \tag{5.6}$$

Similarly, for the second derivatives,

$$\frac{\partial^2[\mathscr{R}_z(y)]^{aa}_{\ell\ell}}{\partial y(m_1)\partial y(m_2)} = \Big[\mathscr{R}_z\mathcal{H}_{,m_1}\mathscr{R}_z\mathcal{H}_{,m_2}\mathscr{R}_z - \mathscr{R}_z\mathcal{H}_{,m_1m_2}\mathscr{R}_z + \mathscr{R}_z\mathcal{H}_{m_2}\mathscr{R}_z\mathcal{H}_{m_1}\mathscr{R}_z\Big]^{aa}_{\ell\ell}. \tag{5.7}$$

Each of the terms in (5.7) can be bounded separately:

$$\begin{aligned}&\left|\left[\mathscr{R}_z\mathcal{H}_{,m_1}\mathscr{R}_z\mathcal{H}_{,m_2}\mathscr{R}_z\right]^{aa}_{\ell\ell}\right| \\ &\quad\leqslant 8N_{\rm b}^4 h_1^2\mathfrak{d}^{-3}\sum_{\ell_1,\ell_2,\ell_3,\ell_4\in\Lambda} e^{-\gamma_{\rm CT}\left(r_{\ell\ell_1}+r_{\ell_2\ell_3}+r_{\ell_4\ell}\right)} e^{-\gamma_1\left(r_{\ell_1 m_1}+r_{m_1\ell_2}+r_{\ell_3 m_2}+r_{m_2\ell_4}\right)} \\ &\quad\leqslant C\mathfrak{d}^{-3}\mathfrak{d}_1^{-4d} e^{-\frac{1}{2}\mathfrak{d}_1\left(r_{\ell m_1}+r_{\ell m_2}\right)}; \quad \text{and} \\ &\left|\left[\mathscr{R}_z\mathcal{H}_{,m_1m_2}\mathscr{R}_z\right]^{aa}_{\ell\ell}\right| \\ &\quad\leqslant 4N_{\rm b}^2 h_2\mathfrak{d}^{-2}\sum_{\ell_1,\ell_2\in\Lambda} e^{-\gamma_{\rm CT}\left(r_{\ell\ell_1}+r_{\ell_2\ell}\right)} e^{-\gamma_2\left(r_{\ell_1 m_1}+r_{\ell_1 m_2}+r_{\ell_2 m_1}r_{\ell_2 m_2}\right)} \\ &\quad\leqslant C\mathfrak{d}^{-2}\left(\sum_{\ell_1\in\Lambda} e^{-\gamma_{\rm CT} r_{\ell\ell_1}} e^{-\gamma_2\left(r_{\ell_1 m_1}+r_{\ell_1 m_2}\right)}\right)^2 \leqslant C\mathfrak{d}^{-2}\mathfrak{d}_2^{-2d} e^{-\frac{1}{2}\mathfrak{d}_2\left(r_{\ell m_1}+r_{\ell m_2}\right)}.\end{aligned}$$

Therefore, we obtain the following bound:

$$\left|\frac{\partial^2[\mathscr{R}_z(y)]^{aa}_{\ell\ell}}{\partial y(m_1)\partial y(m_2)}\right| \leqslant C\mathfrak{d}^{-3}\max\left\{\mathfrak{d}_1^{-4d}, \mathfrak{d}\mathfrak{d}_2^{-2d}\right\} e^{-\frac{1}{2}\mathfrak{d}_2\left(r_{\ell m_1}+r_{\ell m_2}\right)}. \tag{5.8}$$

In particular, for $\mathfrak{d}$ sufficiently small, we have

$$\left|\frac{\partial[\mathscr{R}_z(y)]^{aa}_{\ell\ell}}{\partial y(m)}\right| \lesssim \mathfrak{d}^{-2(d+1)} e^{-\gamma_{\rm CT} r_{\ell m}} \quad \text{and} \quad \left|\frac{\partial^2[\mathscr{R}_z(y)]^{aa}_{\ell\ell}}{\partial y(m_1)\partial y(m_2)}\right| \lesssim \mathfrak{d}^{-(4d+3)} e^{-\frac{1}{2}\gamma_{\rm CT}\left(r_{\ell m_1}+r_{\ell m_2}\right)}. \tag{5.9}$$

It should be clear that, for higher derivatives, the same arguments can be made and similar estimates hold.

Proof of Proposition 2.2: Finite temperature locality for metals. We will only consider the case where $j \in \{1, 2\}$. For $j > 2$, similar arguments can be made but is omitted as the notation becomes tedious and no new ideas are used. Since for all $z \in \mathscr{C}_\beta$,

$$\operatorname{dist}\Big(z, \sigma(\mathcal{H}(y))\Big) \geqslant \frac{\pi}{2\beta},$$

we may use (5.6) and (5.8) with $\gamma_{\mathrm{CT}} = \gamma_{\mathrm{CT}}(\frac{\pi}{2\beta})$. First, we consider $j = 1$ and write:

$$\left|\frac{\partial G_\ell^\beta(y)}{\partial[y(m)]_i}\right| \leqslant \frac{1}{2\pi}\sum_a \left|\oint_{\mathscr{C}_\beta} \mathfrak{g}^\beta(z;\mu)\frac{\partial[\mathscr{R}_z(y)]_{\ell\ell}^{aa}}{\partial[y(m)]_i}\mathrm{d}z\right|$$
$$\leqslant CN_{\mathrm{b}}|\mathscr{C}_\beta| \max_{\mathscr{C}_\beta}\left|\mathfrak{g}^\beta(z;\mu)\right|\beta^2 \min\left\{\gamma_1, \gamma_{\mathrm{CT}}(\tfrac{\pi}{2\beta})\right\}^{-2d} e^{-\min\{\gamma_1,\gamma_{\mathrm{CT}}(\frac{\pi}{2\beta})\}\, r_{\ell m}}.$$

By (5.3) and Lemma 5.2, $\mathfrak{g}^\beta(\,\cdot\,;\mu)$ is uniformly bounded along $\mathscr{C}_\beta$ independently of β. We can thus conclude with $\eta_1 \coloneqq \min\{\gamma_1, \gamma_{\mathrm{CT}}(\frac{\pi}{2\beta})\}$.

Similarly, for $j = 2$, we may apply (5.8) together with

$$\left|\frac{\partial^2 G_\ell^\beta(y)}{\partial[y(m_1)]_{i_1}\partial[y(m_2)]_{i_2}}\right| \leqslant CN_{\mathrm{b}}|\mathscr{C}_\beta| \max_{1\leqslant a\leqslant N_{\mathrm{b}}}\sup_{z\in\mathscr{C}_\beta}\left|\frac{\partial^2[\mathscr{R}_z(y)]_{\ell\ell}^{aa}}{\partial y(m_1)\partial y(m_2)}\right|,$$

to conclude with $\eta_2 \coloneqq \frac{1}{2}\min\{\gamma_1, \gamma_2, \gamma_{\mathrm{CT}}(\frac{\pi}{2\beta})\}$. The fact that, for sufficiently large $\beta > 0$, the pre-factor is $C\beta^\alpha$ for some $\alpha = \alpha(j,d) > 0$ should be clear from (5.9) and Lemma 5.3. □

In the case of an insulator, the separation between the spectrum and the contour can be chosen to be β-independent and equal to $\mathsf{d}(y)$ as in (2.9). In the zero temperature case, this constant may be chosen to be $\frac{1}{2}\mathsf{g}(y)$ where $\mathsf{g}(y)$ is the constant from (2.10).

Proof of Proposition 2.3: locality estimates for insulators. The proof follows in the exact same way as Proposition 2.2 with $\gamma_{\mathrm{CT}} = \gamma_{\mathrm{CT}}(\mathsf{d}(y))$ for finite Fermi-temperature. In the case of zero Fermi-temperature, we use the proof of Proposition 2.2 with $\gamma_{\mathrm{CT}} = \gamma_{\mathrm{CT}}(\frac{1}{2}\mathsf{g}(y))$ and the fact that $|2(z-\mu)|$ is uniformly bounded along the contour $\mathscr{C}_\infty$. □

5.3. Decomposition of the spectrum

We need to show that the defective Hamiltonian can be written in terms of the reference Hamiltonian. However, we are considering a point defect reference configuration, Λ, for which $\Lambda \cap B_{\mathrm{def}} \neq \Lambda^{\mathrm{ref}} \cap B_{\mathrm{def}}$ in general. This means that the defective and reference Hamiltonians may be defined on different spaces. We shall extend the definitions to $\Lambda \cup \Lambda^{\mathrm{ref}}$: for $y \in \mathrm{Adm}(\Lambda)$ and $\ell, k \in \Lambda \cup \Lambda^{\mathrm{ref}}$, let us define

$$\widetilde{\mathcal{H}}(y)_{\ell k}^{ab} \coloneqq \begin{cases} \mathcal{H}(y)_{\ell k}^{ab} & \text{if } \ell, k \in \Lambda \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad [\widetilde{\mathcal{H}}^{\mathrm{ref}}]_{\ell k}^{ab} \coloneqq \begin{cases} [\mathcal{H}^{\mathrm{ref}}]_{\ell k}^{ab} & \text{if } \ell, k \in \Lambda^{\mathrm{ref}} \\ 0 & \text{otherwise.} \end{cases} \tag{5.10}$$

This only changes the spectrum by introducing additional zero eigenvalues. We shall shift the spectrum away from $\{0\}$ so that we can replace $\mathcal{H}(y)$ and $\mathcal{H}^{\mathrm{ref}}$ with $\widetilde{\mathcal{H}}(y)$ and $\widetilde{\mathcal{H}}^{\mathrm{ref}}$, respectively. This does not lead to any problems as we will now see: fix $\beta \in (0,\infty]$ and an appropriate contour $\mathscr{C}$. Choosing $z_0 \in \mathbb{C}$ such that the contour $\mathscr{C} + z_0$ does not encircle $\{0\}$, we have

$$\begin{aligned} G_\ell^\beta(y) &= -\frac{1}{2\pi i}\sum_a \oint_{\mathscr{C}} \mathfrak{g}^\beta(z;\mu)[\mathscr{R}_z(y)]_{\ell\ell}^{aa}\mathrm{d}z \\ &= -\frac{1}{2\pi i}\sum_a \oint_{\mathscr{C}+z_0} \mathfrak{g}^\beta(z-z_0;\mu)\Big[\big(\mathcal{H}(y)-(z-z_0)\big)^{-1}\Big]_{\ell\ell}^{aa}\mathrm{d}z \\ &= -\frac{1}{2\pi i}\sum_a \oint_{\mathscr{C}+z_0} \mathfrak{g}^\beta(z;\mu+z_0)\left[\Big((\mathcal{H}(y)+z_0)^{\Lambda\cup\Lambda^{\mathrm{ref}}}-z\Big)^{-1}\right]_{\ell\ell}^{aa}\mathrm{d}z \end{aligned}$$

where $(\mathcal{H}(y)+z_0)^{\Lambda\cup\Lambda^{\rm ref}}$ is the extension of $\mathcal{H}(y)+z_0$ to $\Lambda\cup\Lambda^{\rm ref}$ as in (5.10). Therefore, by considering $\mathcal{H}(y)+z_0$, we can shift the spectrum away from $\{0\}$ and this does not affect the site energies as long as we also shift the chemical potential and the contour by z_0.

We now show that the Hamiltonian may be decomposed into three terms: the reference Hamiltonian and two perturbations that are small in the sense of rank and Frobenius norm, respectively.

Lemma 5.4 (Decomposition of the Hamiltonian)**.** *Fix $y \in \mathrm{Adm}(\Lambda)$. For each $\delta > 0$ there exists $R_\delta > 0$ and operators $P_1(y), P_2(y)$ such that*

$$\widetilde{\mathcal{H}}(y) = \widetilde{\mathcal{H}}^{\rm ref} + P_1(y) + P_2(y), \tag{5.11}$$

$\|P_1(y)\|_F \leqslant \delta$ and $P_2(y)^{ab}_{\ell k} = 0$ for all $(\ell,k) \notin B_{R_\delta} \times B_{R_\delta}$.

Proof. To simplify notation, we let $u := y - x$. Firstly, since $u \in \dot{\mathscr{W}}^{1,2}(\Lambda)$ and y satisfies (**L**), there exists an accumulation parameter $0 < \mathfrak{m} < 1$ such that $|y(\ell) - y(k)| \geqslant \mathfrak{m}|\ell - k|$ for all $\ell, k \in \Lambda$ [29]. Moreover, since the semi-norm defined by

$$\|Du\|_{\ell^\infty} := \sup_{\ell\in\Lambda} \sup_{\rho\in\Lambda-\ell} \frac{|D_\rho u(\ell)|}{|\rho|}$$

is equivalent to $\|D\cdot\|_{\ell^2_\Upsilon}$ [12], we may choose $R > R_{\rm def}$ sufficiently large such that

$$|D_{k-\ell}u(\ell)| \leqslant \mathfrak{m}|\ell - k| \quad \forall\, \ell, k \in \Lambda \setminus B_R. \tag{5.12}$$

By applying Taylor's theorem we have: for all $\ell, k \in \Lambda \setminus B_R$ and atomic orbitals $1 \leqslant a, b \leqslant N_{\rm b}$,

$$\begin{aligned} \left|\left[\widetilde{\mathcal{H}}(y) - \widetilde{\mathcal{H}}^{\rm ref}\right]^{ab}_{\ell k}\right| &= \left|\left[\mathcal{H}(y) - \mathcal{H}(x)\right]^{ab}_{\ell k}\right| = \left|\nabla h^{ab}_{\ell k}(\xi)\cdot[u(k) - u(\ell)]\right| \\ &\leqslant h_1 e^{-\gamma_1|\xi|}|D_{k-\ell}u(\ell)| \end{aligned} \tag{5.13}$$

where $\xi = (1-\theta)(y(\ell) - y(k)) + \theta(\ell - k)$ for some $\theta = \theta(a,b,\ell,k) \in [0,1]$. Now, by (5.12), we necessarily have that $|\xi| \geqslant \frac{\sqrt{3}}{2}\mathfrak{m}|\ell - k|$. In particular, by (5.13), we obtain

$$\left|\left[\widetilde{\mathcal{H}}(y) - \widetilde{\mathcal{H}}^{\rm ref}\right]^{ab}_{\ell k}\right| \leqslant h_1\, e^{-\frac{\sqrt{3}}{2}\gamma_1\mathfrak{m}|\ell-k|}|D_{k-\ell}u(\ell)| \tag{5.14}$$

for all $\ell, k \in \Lambda \setminus B_R$.

The off-diagonal Hamiltonian entries decay exponentially and so we obtain: for $R' > 0$,

$$\begin{aligned} \sum_{1\leqslant a,b\leqslant N_{\rm b}} \sum_{\ell\in(\Lambda\cup\Lambda^{\rm ref})\cap B_R} \sum_{\substack{k\in\Lambda\setminus B_R \\ |\ell-k|>R'}} \left|\left[\widetilde{\mathcal{H}}(y) - \widetilde{\mathcal{H}}^{\rm ref}\right]^{ab}_{\ell k}\right|^2 &\leqslant C \sum_{\ell\in(\Lambda\cup\Lambda^{\rm ref})\cap B_R} \sum_{\substack{k\in\Lambda\setminus B_R \\ |\ell-k|>R'}} e^{-2\gamma_0\mathfrak{m}|\ell-k|} \\ &\leqslant C_R \int_{\substack{\boldsymbol{r}\in\mathbb{R}^d \\ |\boldsymbol{r}|>R'}} e^{-2\gamma_0\mathfrak{m}|\boldsymbol{r}|}\mathrm{d}\boldsymbol{r} \leqslant C_R\, p(R') e^{-2\gamma_0\mathfrak{m}R'} \end{aligned} \tag{5.15}$$

where $p(R')$ is a polynomial (of degree $d-1$) in R'. We let $P_1(y)$ be the operator (depending on R and R') defined by

$$P_1(y)^{ab}_{\ell k} = \begin{cases} \left[\mathcal{H}(y) - \mathcal{H}^{\rm ref}\right]^{ab}_{\ell k} & \text{if } \ell, k \in \Lambda \setminus B_R \\ \left[\widetilde{\mathcal{H}}(y) - \widetilde{\mathcal{H}}^{\rm ref}\right]^{ab}_{\ell k} & \text{if } |\ell - k| > R' \text{ and } (\ell \in B_R, k \notin B_R \text{ or } \textit{vice versa}) \\ 0 & \text{otherwise} \end{cases} \tag{5.16}$$

for $\ell, k \in \Lambda \cup \Lambda^{\mathrm{ref}}$. We then define $P_2(y)$ to be the finite rank operator such that (5.11) is satisfied. $P_2(y)$ is local in the sense that $P_2(y)^{ab}_{\ell k} = 0$ if $(\ell, k) \notin B_{R''} \times B_{R''}$ for some R'' (depending on R and R'). To conclude, we simply use (5.14) and (5.15) to bound $\|P_1(y)\|_F$:

$$\begin{aligned}\|P_1(y)\|_F^2 &= \sum_{1 \leqslant a,b \leqslant N_{\mathrm{b}}} \left(\sum_{\ell,k \in \Lambda \setminus B_R} \left|P_1(y)^{ab}_{\ell k}\right|^2 + 2 \sum_{\ell \in (\Lambda \cup \Lambda^{\mathrm{ref}}) \cap B_R} \sum_{\substack{k \in \Lambda \setminus B_R \\ |\ell - k| > R'}} \left|P_1(y)^{ab}_{\ell k}\right|^2 \right) \\ &\leqslant C \|Du\|^2_{\ell^2_\Upsilon(\Lambda \setminus B_R)} + C_R\, p(R') e^{-2\gamma_0 \mathfrak{m} R'}.\end{aligned}$$

Since this expression can be made arbitrarily small by choosing R and then R' sufficiently large, this completes the proof. □

We now use Lemma 5.4 to show that $\widetilde{\mathcal{H}}(y)$ is a compact perturbation of $\widetilde{\mathcal{H}}^{\mathrm{ref}}$ and thus prove that the spectrum can be decomposed as in Lemma 2.5:

Proof of Lemma 2.5: Decomposition of the spectrum. By Lemma 5.4, $\mathcal{H}(y) - \mathcal{H}^{\mathrm{ref}}$ is the $\|\cdot\|_{\ell^2 \to \ell^2}$−limit of a sequence of finite rank operators and is thus compact [33]. Indeed, for $\delta > 0$, we choose $P_1(y)$ and $P_2(y)$ as in Lemma 5.4 and write:

$$\|\mathcal{H}(y) - \mathcal{H}^{\mathrm{ref}} - P_2(y)\|_{\ell^2 \to \ell^2} = \|P_1(y)\|_{\ell^2 \to \ell^2} \leqslant \|P_1(y)\|_F \leqslant \delta.$$

Here, we have used the fact that the operator norm is bounded above by the Frobenius norm which is a simple application of the Cauchy–Schwarz inequality. Therefore by Weyl's theorem [23], $\sigma_{\mathrm{ess}}(\mathcal{H}(y)) = \sigma_{\mathrm{ess}}(\mathcal{H}^{\mathrm{ref}})$ and so,

$$\begin{aligned}\sigma(\mathcal{H}(y)) \setminus B_\delta(\sigma(\mathcal{H}^{\mathrm{ref}})) &\subset \sigma(\mathcal{H}(y)) \setminus B_\delta(\sigma_{\mathrm{ess}}(\mathcal{H}^{\mathrm{ref}})) \\ &= \sigma_{\mathrm{disc}}(\mathcal{H}(y)) \setminus B_\delta(\sigma_{\mathrm{ess}}(\mathcal{H}(y))).\end{aligned}$$

We can therefore conclude that $\sigma(\mathcal{H}(y)) \setminus B_\delta(\sigma(\mathcal{H}^{\mathrm{ref}}))$ is both compact and discrete and therefore finite. □

5.4. Proof of Theorems 2.7 and 2.9: Improved locality estimates

We will now prove general estimates for the resolvent operators which will be useful in the proof of the improved locality results. Firstly, we shift $\mathcal{H}(y)$ and $\mathcal{H}^{\mathrm{ref}}$ by the same constant multiple of the identity so that the spectrum of these operators is bounded below by a positive constant. We also fix $z \in \mathbb{C}$ in a bounded set such that

$$\mathrm{dist}\Big(z, \sigma(\widetilde{\mathcal{H}}(y))\Big) \geqslant \mathfrak{d} \quad \text{and} \quad \mathrm{dist}\Big(z, \sigma(\widetilde{\mathcal{H}}^{\mathrm{ref}})\Big) \geqslant \mathfrak{d}^{\mathrm{ref}} \tag{5.17}$$

for positive constants $\mathfrak{d}, \mathfrak{d}^{\mathrm{ref}}$. In the following, we use the notation, $\gamma_{\mathrm{CT}}(\mathfrak{d})$ and $\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})$, for the Combes–Thomas exponents from Lemma 5.3. Further, we fix $\delta > 0$ sufficiently small such that the finite rank perturbation $P_2(y)$, from Lemma 5.4 with the constant δ, is such that $\widetilde{\mathcal{H}}^{\mathrm{ref}} + P_2(y) - z$ is invertible. This can be done since (5.17) is satisfied and $P_1(y)$, also from Lemma 5.4 with the constant δ, only perturbs the spectrum by at most δ in the following sense [23]: $\mathrm{dist}(\sigma(\mathcal{H}(y)), \sigma(\mathcal{H}^{\mathrm{ref}} + P_2(y))) \leqslant \delta$.

Applying the Woodbury identity [22] with $(\widetilde{\mathcal{H}}^{\mathrm{ref}} + P_2(y) - z)^{-1}$ and $\mathscr{R}^{\mathrm{ref}}_z := (\widetilde{\mathcal{H}}^{\mathrm{ref}} - z)^{-1}$ yields

$$\Big(\widetilde{\mathcal{H}}^{\mathrm{ref}} + P_2(y) - z\Big)^{-1} = \mathscr{R}^{\mathrm{ref}}_z - \mathscr{R}^{\mathrm{ref}}_z (I + P_2(y)\mathscr{R}^{\mathrm{ref}}_z)^{-1} P_2(y) \mathscr{R}^{\mathrm{ref}}_z. \tag{5.18}$$

Here, $I + P_2(y)\mathscr{R}_z^{\rm ref} = (\widetilde{\mathcal{H}}^{\rm ref} + P_2(y) - z)\mathscr{R}_z^{\rm ref}$ is invertible since $\widetilde{\mathcal{H}}^{\rm ref} + P_2(y) - z$ and $\mathscr{R}_z^{\rm ref}$ are both invertible. For $k \notin B_{R_\delta}$, we have $\big[(I + P_2(y)\mathscr{R}_z^{\rm ref})^{-1} P_2(y)\big]_{\ell k}^{ab} = 0$. On the other hand, if $k \in B_{R_\delta} \cap \Lambda$, we have

$$
\begin{aligned}
&\left|\Big[\big(I + P_2(y)\mathscr{R}_z^{\rm ref}\big)^{-1} P_2(y)\Big]_{\ell k}^{ab}\right| \\
&\quad = \left|\left[(\widetilde{\mathcal{H}}^{\rm ref} - z)\left(\Big(\widetilde{\mathcal{H}}^{\rm ref} + P_2(y) - z\Big)^{-1} - \mathscr{R}_z^{\rm ref}\right)(\widetilde{\mathcal{H}}^{\rm ref} - z)\right]_{\ell k}^{ab}\right| \\
&\quad \leqslant C \max\Big\{\mathfrak{d}^{-1}, \big(\mathfrak{d}^{\rm ref}\big)^{-1}\Big\}\left[\sum_{\ell_1,\ell_2\in\Lambda} e^{-\eta(r_{\ell\ell_1}+r_{\ell_1\ell_2}+r_{\ell_2 k})} + |z|\sum_{\ell_1\in\Lambda} e^{-\eta(r_{\ell\ell_1}+r_{\ell_1 k})} + |z|^2 e^{-\eta r_{\ell k}}\right] \\
&\quad \leqslant c_{\mathfrak{d}\mathfrak{m}} e^{-\frac{1}{4}\eta r_{\ell k}} \leqslant c_{\mathfrak{d}\mathfrak{m}} e^{-\frac{1}{4}\eta(|\ell| - R_\delta)}
\end{aligned}
\tag{5.19}
$$

where $\eta \coloneqq \min\{\gamma_0, \mathfrak{m}\gamma_{\rm CT}(\mathfrak{d}), \gamma_{\rm CT}(\mathfrak{d}^{\rm ref})\}$. Here, we write $r_{\ell k}$ for $r_{\ell k}(x) = |\ell - k|$ where $x\colon \Lambda \to \Lambda$ is the identity configuration. It is important to note that, since we are considering z contained in a bounded set, the prefactor, $c_{\mathfrak{d}\mathfrak{m}}$, is bounded independently of z (in the bounded set).

Now we may bound the additional contribution in (5.18):

$$
\begin{aligned}
\left|\big[\mathscr{R}_z^{\rm ref}(I + P_2(y)\mathscr{R}_z^{\rm ref})^{-1} P_2(y)\mathscr{R}_z^{\rm ref}\big]_{\ell k}^{ab}\right| &\leqslant C c_{\mathfrak{d}\mathfrak{m}} \big(\mathfrak{d}^{\rm ref}\big)^{-2} \sum_{\ell_1\in\Lambda, \ell_2\in\Lambda\cap B_{R_\delta}} e^{-\gamma_{\rm CT}(\mathfrak{d}^{\rm ref})\left(r_{\ell\ell_1}+r_{\ell_2 k}\right)} e^{-\frac{1}{4}\eta|\ell_1|} \\
&\leqslant C c_{\mathfrak{d}\mathfrak{m}} \big(\mathfrak{d}^{\rm ref}\big)^{-2} e^{-\gamma_{\rm CT}(\mathfrak{d}^{\rm ref})(|\ell|+|k|)}.
\end{aligned}
\tag{5.20}
$$

Combining (5.20) and (5.18) results in the following improved Combes–Thomas type estimate:

$$
\begin{aligned}
&\left|\left[\Big(\widetilde{\mathcal{H}}^{\rm ref} + P_2(y) - z\Big)^{-1}\right]_{\ell k}^{ab}\right| \leqslant C_{\ell k} e^{-\gamma_{\rm CT}(\mathfrak{d}^{\rm ref}) r_{\ell k}} \quad \text{where,} \\
&\qquad C_{\ell k} = C\Big\{\big(\mathfrak{d}^{\rm ref}\big)^{-1} + c_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\rm ref}\big)^{-2} e^{-\gamma_{\rm CT}(\mathfrak{d}^{\rm ref})(|\ell|+|k|-r_{\ell k})}\Big\}.
\end{aligned}
\tag{5.21}
$$

Since, adding $P_1(y)$ only perturbs the spectrum by δ as in Lemma 2.5, the same estimates hold with exponent $\gamma_{\rm CT}(\mathfrak{d}^{\rm ref} - \delta)$ when $(\widetilde{\mathcal{H}}^{\rm ref} + P_2(y) - z)^{-1}$ is replaced by $\mathscr{R}_z(y)$.

Finally, we show that the estimate (5.21) implies improved estimates for derivatives of the resolvent. Moreover, we will show that the pre-factor decays away from the defect. To simplify notation further, we let $\mathfrak{d}_j^{\rm ref} \coloneqq \min\{\gamma_1, \dots, \gamma_j, \gamma_{\rm CT}(\mathfrak{d}^{\rm ref})\}$ for each j. In place of (5.6), we now have

$$
\begin{aligned}
\left|\frac{\partial[\mathscr{R}_z(y)]_{\ell\ell}^{aa}}{\partial[y(m)]_i}\right| &\leqslant C\left(\sum_{\ell_1\in\Lambda} C_{\ell\ell_1} e^{-\mathfrak{d}_1^{\rm ref}\left(r_{\ell\ell_1}+r_{\ell_1 m}\right)}\right)^2 \\
&\leqslant C\left\{\big(\mathfrak{d}^{\rm ref}\big)^{-2}\big(\mathfrak{d}_1^{\rm ref}\big)^{-2d} e^{-\mathfrak{d}_1^{\rm ref} r_{\ell m}} + c_{\mathfrak{d}\mathfrak{m}}^2\big(\mathfrak{d}^{\rm ref}\big)^{-4}\left(\sum_{\ell_1\in\Lambda} e^{-\mathfrak{d}_1^{\rm ref}\left(|\ell|+|\ell_1|+r_{\ell_1 m}\right)}\right)^2\right\} \\
&\leqslant C\big(\mathfrak{d}_1^{\rm ref}\big)^{-2d}\Big(\big(\mathfrak{d}^{\rm ref}\big)^{-2} + c_{\mathfrak{d}\mathfrak{m}}^2\big(\mathfrak{d}^{\rm ref}\big)^{-4} e^{-\mathfrak{d}_1^{\rm ref}(|\ell|+|m|-r_{\ell m})}\Big) e^{-\mathfrak{d}_1^{\rm ref} r_{\ell m}} \\
&=: C(\ell, m) e^{-\mathfrak{d}_1^{\rm ref} r_{\ell m}}.
\end{aligned}
\tag{5.22}
$$

The pre-factor, $C(\ell, m)$, in (5.22) converges to the corresponding pre-factor for the reference resolvent, *i.e.* to $C\big(\mathfrak{d}_1^{\rm ref}\big)^{-2d}\big(\mathfrak{d}^{\rm ref}\big)^{-2}$, as $|\ell| + |m| - r_{\ell m} \to \infty$.

We will now do the same calculation for the second order derivatives of the resolvent. In this case, (5.8) now takes the form

$$
\begin{aligned}
&\big|\big[\mathscr{R}_z\mathcal{H}_{,m_1}\mathscr{R}_z\mathcal{H}_{,m_2}\mathscr{R}_z\big]^{aa}_{\ell\ell}\big| \\
&\quad\leqslant \sum_{\ell_1,\ell_2,\ell_3,\ell_4\in\Lambda} C_{\ell\ell_1}C_{\ell_2\ell_3}C_{\ell_4\ell}e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+r_{\ell_2\ell_3}+r_{\ell_4\ell}\right)}e^{-\gamma_1\left(r_{\ell_1 m_1}+r_{m_1\ell_2}+r_{\ell_3 m_2}+r_{m_2\ell_4}\right)} \\
&\quad= C\sum_{\ell_1,\ell_2,\ell_3,\ell_4\in\Lambda}\Big\{\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-3}e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+r_{\ell_2\ell_3}+r_{\ell_4\ell}\right)} \\
&\qquad+ c_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-4}\Big(e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(|\ell|+|\ell_1|+r_{\ell_2\ell_3}+r_{\ell_4\ell}\right)}+e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+|\ell_2|+|\ell_3|+r_{\ell_4\ell}\right)} \\
&\qquad\qquad+e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+r_{\ell_2\ell_3}+|\ell_4|+|\ell|\right)}\Big) \\
&\qquad+ c^2_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-5}\Big(e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+|\ell_2|+|\ell_3|+|\ell_4|+|\ell|\right)}+e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(|\ell|+|\ell_1|+r_{\ell_2\ell_3}+|\ell_4|+|\ell|\right)} \\
&\qquad\qquad+e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(|\ell|+|\ell_1|+|\ell_2|+|\ell_3|+r_{\ell_4\ell}\right)}\Big) \\
&\qquad+ c^3_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-6}\Big(e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})(|\ell|+|\ell_1|+|\ell_2|+|\ell_3|+|\ell_4|+|\ell|)}\Big)\Big\}e^{-\gamma_1\left(r_{\ell_1 m_1}+r_{m_1\ell_2}+r_{\ell_3 m_2}+r_{m_2\ell_4}\right)} \\
&\quad\leqslant C\big(\mathfrak{d}_1^{\mathrm{ref}}\big)^{-4d}\Big\{\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-3} \\
&\qquad+ c_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-4}\Big(e^{-\frac{1}{2}\mathfrak{d}_1^{\mathrm{ref}}\left(|\ell|+|m_1|-r_{\ell m_1}\right)}+e^{-\frac{1}{2}\mathfrak{d}_1^{\mathrm{ref}}(|m_1|+|m_2|)}+e^{-\frac{1}{2}\mathfrak{d}_1^{\mathrm{ref}}\left(|\ell|+|m_2|-r_{\ell m_2}\right)}\Big) \\
&\qquad+ c^2_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-5}\Big(e^{-\mathfrak{d}_1^{\mathrm{ref}}\left(|\ell|+|m_2|-r_{\ell m_2}\right)}+e^{-\frac{1}{2}\mathfrak{d}_1^{\mathrm{ref}}\left(2|\ell|+|m_1|+|m_2|-r_{\ell m_1}-r_{\ell m_2}\right)} \\
&\qquad\qquad+e^{-\mathfrak{d}_1^{\mathrm{ref}}\left(|\ell|+|m_1|-r_{\ell m_1}\right)}\Big) \\
&\qquad+ c^3_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-6}e^{-\mathfrak{d}_1^{\mathrm{ref}}\left(2|\ell|+|m_1|+|m_2|-r_{\ell m_1}-r_{\ell m_2}\right)}\Big\}e^{-\frac{1}{2}\mathfrak{d}_1^{\mathrm{ref}}\left(r_{\ell m_1}+r_{\ell m_2}\right)} \\
&\quad=: C_1(\ell,m_1,m_2)e^{-\frac{1}{2}\mathfrak{d}_1^{\mathrm{ref}}\left(r_{\ell m_1}+r_{\ell m_2}\right)}.
\end{aligned}
\tag{5.23}
$$

Again, the pre-factor converges to the reference pre-factor, *i.e.* to $C\big(\mathfrak{d}_1^{\mathrm{ref}}\big)^{-4d}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-3}$, exponentially as $|\ell|+|m_1|-r_{\ell m_1}$ and $|\ell|+|m_2|-r_{\ell m_2}\to\infty$. Similarly,

$$
\begin{aligned}
&\big|\big[\mathscr{R}_z\mathcal{H}_{,m_1m_2}\mathscr{R}_z\big]^{aa}_{\ell\ell}\big|\leqslant C\sum_{\ell_1,\ell_2\in\Lambda}C_{\ell\ell_1}C_{\ell_2\ell}e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+r_{\ell_2\ell}\right)}e^{-\gamma_2\left(r_{\ell_1 m_1}+r_{\ell_1 m_2}+r_{\ell_2 m_1}r_{\ell_2 m_2}\right)} \\
&\quad= C\sum_{\ell_1,\ell_2\in\Lambda}\Big\{\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-2}e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+r_{\ell_2\ell}\right)} \\
&\qquad+ c_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-3}\Big(e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(|\ell|+|\ell_1|+r_{\ell\ell_2}\right)}+e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})\left(r_{\ell\ell_1}+|\ell_2|+|\ell|\right)}\Big) \\
&\qquad+ c^2_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-4}e^{-\gamma_{\mathrm{CT}}(\mathfrak{d}^{\mathrm{ref}})(|\ell|+|\ell_1|+|\ell_2|+|\ell|)}\Big\}e^{-\gamma_2\left(r_{\ell_1 m_1}+r_{\ell_1 m_2}+r_{\ell_2 m_1}r_{\ell_2 m_2}\right)} \\
&\quad\leqslant C\big(\mathfrak{d}_2^{\mathrm{ref}}\big)^{-2d}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-2}\Big[1+c_{\mathfrak{d}\mathfrak{m}}\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-2}e^{-\frac{1}{4}\mathfrak{d}_2^{\mathrm{ref}}\left(2|\ell|+|m_1|+|m_2|-r_{\ell m_1}-r_{\ell m_2}\right)}\Big]^2e^{-\frac{1}{2}\mathfrak{d}_2^{\mathrm{ref}}\left(r_{\ell m_1}+r_{\ell m_2}\right)} \\
&\quad=: C_2(\ell,m_1,m_2)e^{-\frac{1}{2}\mathfrak{d}_1^{\mathrm{ref}}\left(r_{\ell m_1}+r_{\ell m_2}\right)}.
\end{aligned}
\tag{5.24}
$$

Therefore, by using (5.7), we have

$$
\left|\frac{\partial^2[\mathscr{R}_z(y)]^{aa}_{\ell\ell}}{\partial y(m_1)\partial y(m_2)}\right|\leqslant\max\{C_1(\ell,m_1,m_2),C_1(\ell,m_1,m_2)\}e^{-\frac{1}{2}\mathfrak{d}_2^{\mathrm{ref}}\left(r_{\ell m_1}+r_{\ell m_2}\right)}
\tag{5.25}
$$

where the pre-factor converges to the pre-factor arising if $\mathscr{R}_z(y)$ is replaced with $\mathscr{R}_z^{\mathrm{ref}}$, *i.e.* it converges to $C\big(\mathfrak{d}^{\mathrm{ref}}\big)^{-3}\max\Big\{\big(\mathfrak{d}_1^{\mathrm{ref}}\big)^{-4d},\mathfrak{d}^{\mathrm{ref}}\big(\mathfrak{d}_2^{\mathrm{ref}}\big)^{-2d}\Big\}$, as we send ℓ, m_1 and m_2 away from the defect core together. Again, we omit the arguments for $j>2$.

Proof of Theorem 2.9: Improved zero temperature locality. We directly apply (5.22) and (5.25) with $\mathfrak{d}^{\mathrm{ref}}=\frac{1}{2}\mathfrak{g}^{\mathrm{ref}}$ and $\mathfrak{d}=\frac{1}{2}\mathfrak{g}$. We again use the fact that $|2(z-\mu)|$ is uniformly bounded along the contour $\mathscr{C}_\infty$. □

If $\mu\notin\sigma(\mathcal{H}(y))$, similar arguments can be made for the finite temperature case. However, if $\mu\in\sigma(\mathcal{H}(y))$, another contribution to the site energy must be considered:

Proof of Theorem 2.7: Improved finite temperature locality. In the case that $\mu\notin\sigma(\mathcal{H}(y))$, we can directly apply (5.22) and (5.25) with $\mathfrak{d}^{\mathrm{ref}}=\mathsf{d}^{\mathrm{ref}}$ and $\mathfrak{d}=\mathsf{d}(y)$. Here we again use the fact that the analytic continuation of $\mathfrak{g}^\beta(z;\mu)$ is uniformly bounded along $\mathscr{C}_\beta$.

In the case that $\mu\in\sigma(\mathcal{H}(y))$, we may split $\mathscr{C}_\beta$ into three simple closed contours $\mathscr{C}^-$, $\mathscr{C}^+$ and $\mathscr{C}_0$ such that $\mathscr{C}^-$ and $\mathscr{C}^+$ are contained in $\mathbb{C}\setminus(\mu+i\mathbb{R})$ and encircle $\sigma(\mathcal{H}(y))\cap[\underline{\sigma},\mu)$ and $\sigma(\mathcal{H}(y))\cap(\mu,\overline{\sigma}]$, respectively, and $\mathscr{C}_0$ encircles $\{\mu\}$ and avoids the rest of the spectrum. Now the finite temperature site energy is of the form:

$$\begin{aligned}G_\ell^\beta(y)=&-\frac{1}{2\pi i}\sum_a\oint_{\mathscr{C}^-}\mathfrak{g}^\beta(z;\mu)[\mathscr{R}_z(y)]_{\ell\ell}^{aa}\mathrm{d}z-\frac{1}{2\pi i}\sum_a\oint_{\mathscr{C}^+}\mathfrak{g}^\beta(z;\mu)[\mathscr{R}_z(y)]_{\ell\ell}^{aa}\mathrm{d}z\\&-\frac{1}{2\pi i}\sum_a\oint_{\mathscr{C}_0}\mathfrak{g}^\beta(z;\mu)[\mathscr{R}_z(y)]_{\ell\ell}^{aa}\mathrm{d}z.\end{aligned}\tag{5.26}$$

The first two expressions of (5.26) can be treated in the exact same way as in the case where $\mu\notin\sigma(\mathcal{H}(y))$. The additional term is

$$\sum_a\sum_{s=1}^{m(\mu)}\mathfrak{g}^\beta(\varepsilon_s(y);\mu)[\psi_s]_{\ell a}^2=\frac{2}{\beta}\log\Big(\frac{1}{2}\Big)\sum_a\sum_{s=1}^{m(\mu)}[\psi_s]_{\ell a}^2\tag{5.27}$$

where $m(\mu)$ is the multiplicity of μ as an eigenvalue of $\mathcal{H}(y)$, $\{\psi_s\}$ is basis for the eigenspace of μ and $\varepsilon_s(y)$ are the eigenvalues at μ written as functions of the configuration. We wish to show that (5.27) has the same locality properties as the first two terms of (5.26).

For $j=1$, we have

$$\begin{aligned}\frac{\partial}{\partial y(m)}\Big(\sum_s\mathfrak{g}^\beta(\varepsilon_s(y);\mu)[\psi_s]_{\ell a}^2\Big)&=\sum_s\Big(2f_\beta(\varepsilon_s(y)-\mu)\frac{\partial\varepsilon_s(y)}{\partial y(m)}[\psi_s]_{\ell a}^2+\mathfrak{g}^\beta(\varepsilon_s(y);\mu)\frac{\partial[\psi_s]_{\ell a}^2}{\partial y(m)}\Big)\\&=\sum_s\Big(\frac{\partial\varepsilon_s(y)}{\partial y(m)}[\psi_s]_{\ell a}^2-\frac{2}{\beta}\log(2)\frac{\partial[\psi_s]_{\ell a}^2}{\partial y(m)}\Big)\\&=\frac{\partial}{\partial y(m)}\Big(\sum_s\big(\varepsilon_s(y)-\mu-\tfrac{2}{\beta}\log(2)\big)[\psi_s]_{\ell a}^2\Big)\\&=-\frac{1}{2\pi i}\oint_{\mathscr{C}_0}\big(z-\mu-\tfrac{2}{\beta}\log(2)\big)\frac{\partial[\mathscr{R}_z(y)]_{\ell\ell}^{aa}}{\partial y(m)}\mathrm{d}z.\end{aligned}$$

Now, because $z\mapsto z-\mu-\frac{2}{\beta}\log(2)$ is analytic, there is no β-dependent restriction on the contour $\mathscr{C}_0$. This again allows us to apply the Woodbury identity and the Combes–Thomas type estimate on the reference resolvent to obtain improved locality results.

For $j=2$, we have

$$\frac{\partial^2}{\partial y(m_1)\partial y(m_2)}\Big(\sum_s\mathfrak{g}^\beta(\varepsilon_s(y);\mu)[\psi_s]_{\ell a}^2\Big)=-\frac{1}{2\pi i}\oint_{\mathscr{C}_0}\big(-\tfrac{1}{4}\beta(z-\mu)^2+z-\mu-\tfrac{2}{\beta}\log(2)\big)\frac{\partial^2[\mathscr{R}_z(y)]_{\ell\ell}^{aa}}{\partial y(m_1)\partial y(m_2)}\mathrm{d}z.$$

Again, we see that $z \mapsto -\frac{1}{4}\beta(z-\mu)^2 + z - \mu - \frac{2}{\beta}\log(2)$ is analytic and so we may use the improved resolvent estimates on a temperature independent contour. This results in improved locality estimates with temperature independent exponents but pre-factors of the form $C\beta$.

For higher derivatives, the same arguments can be made which gives rise to β-independent exponents but pre-factors that are of the form $C\beta^{j-1}$. □

Acknowledgements. JT and CO gratefully acknowledge the hospitality of Peking Normal University during the work on this project. CO is supported by EPSRC Grant EP/R043612/1 and Leverhulme Research Project Grant RPG-2017-191. JT is supported by EPSRC as part of the MASDOC DTC, Grant No. EP/HO23364/1. HC is supported by Thousand Talents Program for Young Professionals, and the Fundamental Research Funds for the Central Universities of China under grant 2017EYT22.

References

[1] M. Aizenman, Localization at weak disorder: some elementary bounds. *Rev. Math. Phys.* **06** (1994) 1163–1182.
[2] M. Aizenman and G.M. Graf, Localization bounds for an electron gas. *J. Phys. A: Math. Gen.* **31** (1998) 6783–6806.
[3] M. Aizenman and S. Molchanov, Localization at large disorder and at extreme energies: an elementary derivation. *Commun. Math. Phys.* **157** (1993) 245–278.
[4] A.P. Bartók, M.C. Payne, R. Kondor and G. Csányi, Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104** (2010) 136403.
[5] A.P. Bartók, J. Kermode, N. Bernstein and G. Csányi, Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8** (2018) 041048.
[6] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98** (2007) 146401.
[7] D.R. Bowler and T. Miyazaki, *O(N)* methods in electronic structure calculations. *Rep. Prog. Phys.* **75** (2012) 036503.
[8] H.C. Chen *et al.*, https://github.com/cortner/SKTB.jl.git.
[9] H. Chen and C. Ortner, QM/MM methods for crystalline defects. Part 1: locality of the tight binding model. *Multiscale Model. Simul.* **14** (2016) 232–264.
[10] H. Chen and C. Ortner, QM/MM methods for crystalline defects. Part 2: consistent energy and force-mixing. *Multiscale Model. Simul.* **15** (2017) 184–214.
[11] H. Chen, J. Lu and C. Ortner, Thermodynamic limit of crystal defects with finite temperature tight binding. *Arch. Ration. Mech. Anal.* **230** (2018) 701–733.
[12] H. Chen, F. Nazar and C. Ortner, Geometry equilibration of crystalline defects in quantum and atomic descriptions. *Math. Model. Methods Appl. Sci.* **29** (2019) 419–492.
[13] R. Cohen, M. Mehl and D. Papaconstantopoulos, Tight-binding total-energy method for transition and noble metals. *Phys. Rev. B* 50 (1994) 14694–14697.
[14] J.M. Combes and L. Thomas, Asymptotic behaviour of eigenfunctions for multiparticle Schrödinger operators. *Commun. Math. Phys.* **34** (1973) 251–270.
[15] G. Csányi, T. Albaret, G. Moras, M.C. Payne and A. De Vita, Multiscale hybrid simulation methods for material systems. *J. Phys. Condens. Matter* **17** (2005) 691–703.
[16] N. David Mermin, Thermal properties of the inhomogeneous electron gas. *Phys. Rev.* **137** (1965) 1441–1443.
[17] V. Ehrlacher, C. Ortner and A.V. Shapeev, Analysis of boundary conditions for crystal defect atomistic simulations. *Arch. Ration. Mech. Anal.* **222** (2016) 1217–1268.
[18] A. Figotin and A. Klein, Localization phenomenon in gaps of the spectrum of random lattice operators. *J. Stat. Phys.* **75** (1994) 997–1021.
[19] M. Finnis, Interatomic Forces in Condensed Matter. Oxford University Press (2003).
[20] S. Goedecker, Linear scaling electronic structure methods. *Rev. Mod. Phys.* 71 (1999) 1085–1123.
[21] S. Goedecker and M. Teter, Tight-binding electronic-structure calculations and tight-binding molecular dynamics with localized orbitals. *Phys. Rev. B* **51** (1995) 9455–9464.
[22] W.W. Hager, Updating the inverse of a matrix. *SIAM Rev.* **31** (1989) 221–239.
[23] T. Kato, Perturbation Theory for Linear Operators Springer, 2nd edition. Springer-Verlag Berlin Heidelberg (1995).
[24] C. Kittel, Introduction to Solid State Physics, eighth edition. Wiley (2004).
[25] W. Kohn, Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.* **76** (1996) 3168–3171.
[26] R.M. Martin, Electronic Structure: Basic Theory and Practical Methods. Cambridge University Press (2004).
[27] M. Mehl and D. Papaconstantopoulos, 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** (1996) 4519–4530.
[28] A.M.N. Niklasson, Density matrix methods in linear scaling electronic structure theory, chapter 16. In: Linear-Scaling Techniques in Computational Chemistry and Physics. Springer, Dordrecht (2011) 439–473.

[29] C. Ortner and A.V. Shapeev, Interpolants of lattice functions for the analysis of atomistic/continuum multiscale methods. Preprint arXiv:1204.3705v1 (2012).
[30] C. Ortner and J. Thomas, Point defects in tight binding models for insulators, in preparation.
[31] D. Papaconstantopoulos, Handbook of the Band Structure of Elemental Solids, from $Z = 1$ to $Z = 112$. Springer New York (2015).
[32] D. Papaconstantopoulos, M. Mehl, S. Erwin and M. Pederson, Tight-binding Hamiltonians for carbon and silicon. *Symp. R – Tight Binding Approach Comput. Mater. Sci.* **491** (1997) 221.
[33] M. Reed and B. Simon, Functional Analysis (Methods of Modern Mathematical Physics). Academic Press (1980).
[34] A.V. Shapeev, Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14** (2016) 1153–1173.
[35] J.C. Slater and G.F. Koster, Simplified LCAO method for the periodic potential problem. *Phys. Rev.* **94** (1954) 1498–1524.
[36] E. Weinan and J. Lu, Electronic structure of smoothly deformed crystals: Cauchy–Born rule for the nonlinear tight-binding model. *Commun. Pure Appl. Math.* **63** (2010) 1432–1468.