

# Comparison of the canonical transformation and energy functional formalisms for *ab initio* calculations of self-localized polarons

Yao Luo,<sup>1</sup> Benjamin K. Chang,<sup>1</sup> and Marco Bernardi<sup>1</sup>

<sup>1</sup>*Department of Applied Physics and Materials Science,  
California Institute of Technology, Pasadena, California 91125, USA*

In materials with strong electron-phonon (e-ph) interactions, charge carriers can distort the surrounding lattice and become trapped, forming self-localized (small) polarons. We recently developed an *ab initio* approach based on canonical transformations to efficiently compute the formation and energetics of small polarons [1]. A different approach based on a Landau-Pekar energy functional has been proposed in the recent literature [2, 3]. In this work, we analyze and compare these two methods in detail. We show that the small polaron energy is identical in the two formalisms when using the same polaron wave function. We also show that our canonical transformation formalism can predict polaron band structures and can properly treat zero- and finite-temperature lattice vibration effects, although at present using a fixed polaron wave function. Conversely, the energy functional approach can compute the polaron wave function, but as we show here it neglects lattice vibrations and cannot address polaron self-localization and thermal band narrowing. Taken together, this work relates two different methods developed recently to study polarons from first-principles, highlighting their merits and shortcomings and discussing them both in a unified formalism.

## I. INTRODUCTION

Self-localized (small) polarons are charge carriers that interact strongly with the lattice vibrations, becoming trapped as a result of their local lattice distortion [4]. Small polarons are key to understanding the physical properties of materials with strong or localized electron-phonon (e-ph) interactions, including alkali halides, organic molecular crystals, transition metal oxides, and some glasses [4]. Small polarons can be studied experimentally using spectroscopy, diffraction, and microscopy techniques sensitive to the local lattice distortion [5–7]. Their signatures are also found in transport properties: as small polarons move only in response to certain vibrations of the surrounding atoms, they are associated with a low charge carrier mobility ( $< 1 \text{ cm}^2/\text{Vs}$ ), which typically increases with temperature due to thermally activated small-polaron hopping [8, 9]. As a result, small polarons are detrimental in many technological applications where electrical transport limits device efficiency.

Theoretical treatments of small polarons span a wide range of analytic and numerical techniques [10, 11]. Focusing on first-principles approaches based on density functional theory (DFT), a key goal has been the development of parameter-free, quantitative predictions of the energetics and dynamics of small polarons. The conventional approach employs DFT calculations on supercells with excess charge or defects added explicitly [12–15]. Yet, recent work has developed a different family of first-principles methods aimed at computing polarons using only a unit cell of the material with *ab initio* e-ph calculations [1–3]. The goal of these approaches is two-fold: reducing computational cost by avoiding calculations on large supercells with many atoms, and formulating rigorous polaron calculations by combining many-body techniques with first-principles theories.

Within this recent body of work, we formulated an efficient approach based on the canonical transformation formalism [1]; this method can compute the small-polaron formation energy in a localized Wannier basis starting from a trial polaron wave function. Its computational cost is a minimal overhead to a DFT calculation on a unit cell, enabling investigations of small polarons in a wide range of materials with minimal computational effort [1]. A different approach, proposed by Sio et al. [2, 3], uses a Landau-Pekar-type energy functional to obtain coupled equations for the polaron wave function and its associated atomic displacements. Solving these equations on a fine reciprocal-space grid can provide the polaron formation energy and wave function.

Here we compare in detail these two methods, and explain how they address various aspects of polaron physics. We first discuss the canonical transformation approach, showing example calculations of the polaron energy and band structure in an ionic insulator (NaCl) and an organic semiconductor (naphthalene). We show that for small polarons the canonical transformation and energy functional formalisms give the same polaron energy when using the same polaron wave function. We also present a generalization of the canonical transformation formalism that can compute the polaron wave function and treat both small and delocalized polarons. This formulation allows us to relate the canonical transformation and energy functional approaches, and discuss differences in how they treat the polaron wave function, thermal effects, band narrowing and polaron self-localization. Of the two approaches, we show that only the canonical transformation framework can guarantee polaron self-localization (due to vanishing hopping) and correctly treat lattice vibrations and thermal effects. Taken together, our work advances the formulation of rigorous methods to study polarons with first-principles calculations on a unit cell of the material.

## II. CANONICAL TRANSFORMATION

Inspired by analytic treatments of small polarons [16, 17], we derive an effective polaron Hamiltonian by transforming the coordinates to the distorted lattice configuration induced by the localized charge carrier. This technique was introduced by Lee, Low, and Pines to study the large polaron problem [18], and it can be traced back to a method used by Tomonaga [19] to solve a meson problem. The transformation is closely related to the one that diagonalizes the charged harmonic oscillator (CHO) in an external electric field [17]. Therefore, we first briefly review the CHO treatment to set the stage for the polaron canonical transformation.

### A. Analogy with the charged harmonic oscillator

The Hamiltonian of a one-dimensional CHO is

$$H^{(\text{CHO})} = \frac{1}{2m}p^2 + \frac{1}{2}m\omega^2x^2 + eEx, \quad (1)$$

where  $x$  is the position and  $p$  the momentum of a particle with mass  $m$  and charge  $e$ . The oscillator frequency is  $\omega$ , and  $E$  is the external electric field that couples to the system (here and below, we set  $\hbar = 1$ ). The CHO Hamiltonian can be solved by completing the square and shifting to a new coordinate system:

$$x' = x - x_0, \quad (2)$$

$$H^{(\text{CHO})} = \frac{1}{2m}p^2 + \frac{1}{2}m\omega^2x'^2 - \frac{e^2E^2}{2m\omega^2}. \quad (3)$$

The particle then oscillates around the new equilibrium point,  $x_0 = -eE/m\omega^2$ , with the same frequency  $\omega$  as in the absence of the external field.

We study this transformation by introducing creation and annihilation operators for the harmonic oscillator,

$$\begin{aligned} x &= \frac{1}{\sqrt{2m\omega}}(b^\dagger + b) \\ p &= i\sqrt{\frac{m\omega}{2}}(b^\dagger - b). \end{aligned}$$

The CHO Hamiltonian becomes

$$H^{(\text{CHO})} = \omega(b^\dagger b + \frac{1}{2}) + \omega g(b + b^\dagger), \quad (4)$$

with the coupling constant  $g = eE/\sqrt{2m\omega^3}$ . As implied by Eq. (2), this Hamiltonian can be solved exactly by a translation to the new equilibrium position  $x_0$ , as can be achieved with the canonical transformation of operators  $\mathcal{O} \rightarrow \tilde{\mathcal{O}} = e^S \mathcal{O} e^{-S}$ , with the translation generated by

$$S^{(\text{CHO})} = ipx_0 = g(b^\dagger - b). \quad (5)$$

Using the Baker-Campbell-Hausdorff (BCH) formula

$$\tilde{\mathcal{O}} = e^S \mathcal{O} e^{-S} = \mathcal{O} + [S, \mathcal{O}] + \frac{1}{2!}[S, [S, \mathcal{O}]] + \dots, \quad (6)$$

we obtain the *transformed* annihilation operator and Hamiltonian (denoted by the tilde symbol):

$$\tilde{b} = b - g, \quad (7)$$

$$\tilde{H}^{(\text{CHO})} = \omega(b^\dagger b + \frac{1}{2}) - \omega g^2. \quad (8)$$

The second term in  $\tilde{H}^{(\text{CHO})}$  is the decrease in potential energy associated with stretching the oscillator spring:

$$-\omega g^2 = \frac{1}{2}m\omega^2x_0^2 + eEx_0. \quad (9)$$

The meaning of the operators  $b$  and  $b^\dagger$  after the transformation can be understood by analyzing the transformed position operator

$$\tilde{x} = x + [S^{(\text{CHO})}, x] = x + x_0 \quad (10)$$

where the  $x$  coordinate on the right-hand side now measures the position relative to  $x_0$ ; this way, the coordinates  $x$  and  $\tilde{x}$  in Eq. (9) correspond to  $x'$  and  $x$  in Eq. (1), respectively. Expressed in terms of creation and annihilation operators, this transformed coordinate becomes

$$\tilde{x} = \frac{1}{\sqrt{2m\omega}}(\tilde{b}^\dagger + \tilde{b}) = \frac{1}{\sqrt{2m\omega}}(b^\dagger + b) + x_0. \quad (11)$$

Therefore, by comparison it is clear that  $b^\dagger$  in  $\tilde{H}^{(\text{CHO})}$  creates oscillation quanta relative to the new equilibrium point  $x_0$ , as illustrated in Fig. 1(a)-(b).

The ground state wave function of the CHO, denoted here as  $|0\rangle$ , is centered at  $x_0$ , and is annihilated by the operator  $b$ , so that  $b|0\rangle = 0$ . Denote as  $|\tilde{0}\rangle$  the original ground state (before applying the electric field) which is annihilated by  $\tilde{b}$ . Then, since

$$b|\tilde{0}\rangle = (\tilde{b} + g)|\tilde{0}\rangle = g|\tilde{0}\rangle$$

we can see that  $|\tilde{0}\rangle$  is a coherent state of  $b$  with eigenvalue  $g$ , and therefore we have

$$|\tilde{0}\rangle = \exp[g(b^\dagger - b)]|0\rangle = \exp[S^{(\text{CHO})}]|0\rangle. \quad (12)$$

The exponential factor connecting  $|\tilde{0}\rangle$  and  $|0\rangle$  is the translation operator in Eq. (4). This result shows that the ground states before and after applying the electric field are related via a translation by  $x_0$  [17].

## B. Derivation of the polaron Hamiltonian

We now carry out a transformation analogous to the CHO case to obtain an effective polaron Hamiltonian. The starting point is the  $e$ -ph Hamiltonian in the electronic Wannier and phonon momentum basis [20],

$$H = \sum_{mn} \varepsilon_{mn} a_m^\dagger a_n + \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} \left( b_{\mathbf{Q}}^\dagger b_{\mathbf{Q}} + \frac{1}{2} \right) \quad (12)$$

$$+ \frac{1}{\sqrt{N_\Omega}} \sum_{mn} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} g_{\mathbf{Q}mn} \left( b_{\mathbf{Q}}^\dagger + b_{-\mathbf{Q}} \right) a_m^\dagger a_n,$$

where  $n = j_n \mathbf{R}_n$  is a collective index labelling the  $j_n$ -th Wannier function (WF) in the unit cell with origin at the Bravais lattice vector  $\mathbf{R}_n$ , and  $a_n = a_{j_n \mathbf{R}_n}$  is the corresponding electron annihilation operator;  $b_{\mathbf{Q}}$  is the phonon annihilation operator, where  $\mathbf{Q}$  is a collective label for the phonon mode  $\nu$  and momentum  $\mathbf{q}$ . The hopping strength and phonon energy are denoted as  $\varepsilon_{mn}$  and  $\omega_{\mathbf{Q}}$ , respectively, and  $N_\Omega$  is the number of unit cells in the crystal.

The  $e$ -ph coupling matrix elements in the Wannier basis, denoted as  $g_{\mathbf{Q}mn}$ , are unitless and do not include the phonon frequency factor, different from the standard convention [20]. They are obtained by transforming unitless  $e$ -ph matrix element in momentum space,  $\tilde{g}_{ij\nu}(\mathbf{k}, \mathbf{q}) = g_{ij\nu}(\mathbf{k}, \mathbf{q}) / (\hbar \omega_{\nu \mathbf{q}})$ , to the electron Wannier basis, where  $g_{ij\nu}(\mathbf{k}, \mathbf{q})$  are defined in Eq. (24) of Ref. [20], and  $i$  and  $j$  are band indices. Using the notation in Ref. [20], the explicit definition is  $g_{\mathbf{Q}mn} \equiv \tilde{g}_{mn\nu}(\mathbf{R}_e, \mathbf{q}) = \frac{1}{N_e} \sum_{\mathbf{k}, ij} e^{-i\mathbf{k} \cdot \mathbf{R}_e} \mathcal{U}_{mi}^\dagger(\mathbf{k} + \mathbf{q}) \tilde{g}_{ij\nu}(\mathbf{k}, \mathbf{q}) \mathcal{U}_{jn}(\mathbf{k})$ , where  $\mathcal{U}$  are unitary Wannier matrices [20]. Also recall that the  $e$ -ph coupling needs to satisfy the relation  $g_{\mathbf{Q}mn}^* = g_{-\mathbf{Q}nm}$  for the Hamiltonian to be Hermitian.

The  $e$ -ph interaction term in Eq. (12) has the same form as in  $H^{(\text{CHO})}$  in Eq. (3), but now the external field coupling to the ‘‘spring’’ of each phonon mode is controlled by the electronic configuration through the factor  $a_m^\dagger a_n$ . In analogy with Eq. (4), we stretch the spring of each phonon mode to a new equilibrium position using the canonical transformation  $\mathcal{O} \rightarrow \tilde{\mathcal{O}} = e^S \mathcal{O} e^{-S}$ , with the generator defined as

$$S = \sum_{mn} C_{mn} a_m^\dagger a_n, \quad (13)$$

$$C_{mn} = \frac{1}{\sqrt{N_\Omega}} \sum_{\mathbf{Q}} B_{\mathbf{Q}mn} (b_{\mathbf{Q}}^\dagger - b_{-\mathbf{Q}}). \quad (14)$$

Above, we introduced the undetermined distortion coefficients  $B_{\mathbf{Q}mn}$  which, analogous to the coupling  $g$  in the CHO example, quantify the stretching of the spring associated with each phonon mode, as we show below. To make the transformation unitary, we impose the conjugate relation  $B_{\mathbf{Q}mn}^* = B_{-\mathbf{Q}nm}$ , so that the operator  $S$  is anti-Hermitian.

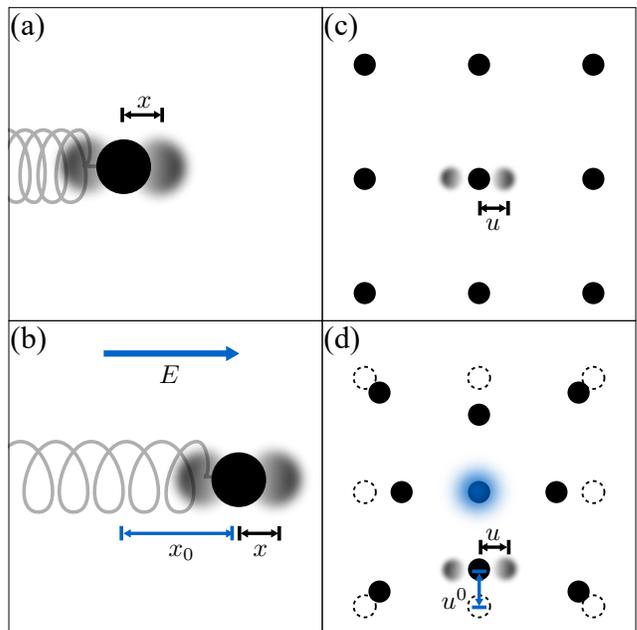


FIG. 1. Schematic of the charged harmonic oscillator, (a) without and (b) with the electric field applied. By analogy, panel (c) shows the unperturbed crystal lattice, and panel (d) the distorted lattice in the presence of the excess charge carrier. In panel (b), the spring is stretched by an amount of  $x_0$ ; analogously, in panel (d) the static lattice distortion is defined as  $u^0$ . The coordinates of the oscillation quanta relative to the new (distorted) equilibrium positions are denoted as  $x$  or  $u$ , in (b) and (d) respectively.

To obtain the transformed electron and phonon operators, we compute their commutators with  $S$ :

$$[S, a_m] = - \sum_n C_{mn} a_n,$$

$$[S, b_{\mathbf{Q}}] = - \frac{1}{\sqrt{N_\Omega}} \sum_{mn} B_{\mathbf{Q}mn} a_m^\dagger a_n.$$

Then using the BCH formula in Eq. (5), we have

$$\tilde{a}_m = \sum_n e_{mn}^{-C} a_n, \quad (15)$$

$$\tilde{b}_{\mathbf{Q}} = b_{\mathbf{Q}} - \frac{1}{\sqrt{N_\Omega}} \sum_{mn} B_{\mathbf{Q}mn} a_m^\dagger a_n, \quad (16)$$

where  $e_{mn}^{-C}$  is a shorthand notation for the phonon operator

$$e_{mn}^{-C} = \delta_{mn} - C_{mn} + \frac{1}{2!} \sum_i C_{mi} C_{in} - \dots \quad (17)$$

The physical intuition is that the distortion coefficients  $B_{\mathbf{Q}mn}$  quantify how the transformation stretches each phonon spring, as is seen by comparing Eq. (16) with the CHO case in Eq. (6).

It is instructive to examine the lattice displacements in real space [21]. Denoting as  $u_{cs\alpha}$  the displacement of the atom  $s$  in unit cell  $c$  along the Cartesian coordinate  $\alpha$ , the atomic positions in the transformed basis become:

$$\tilde{u}_{cs\alpha} = u_{cs\alpha} + u_{cs\alpha}^0 = \sum_{\mathbf{Q}} \frac{1}{\sqrt{2M_s\omega_{\mathbf{Q}}N_{\Omega}}} e_{\mathbf{Q}}^{s\alpha} e^{i\mathbf{Q}\cdot\mathbf{R}_c} \times \left[ \left( b_{\mathbf{Q}}^{\dagger} + b_{-\mathbf{Q}} \right) - \frac{2}{\sqrt{N_{\Omega}}} \sum_{mn} B_{-\mathbf{Q}mn} a_m^{\dagger} a_n \right], \quad (18)$$

in full analogy with Eq. (10) for the CHO (above,  $M_s$  is the mass of atom  $s$  and  $e_{\mathbf{Q}}^{s\alpha}$  is the phonon eigenvector). From Eq. (18), we see that the lattice distortion induced by the localized polaron gives new equilibrium atomic positions,

$$u_{cs\alpha}^0 = -\frac{1}{N_{\Omega}} \sum_{\mathbf{Q}} \sqrt{\frac{2}{M_s\omega_{\mathbf{Q}}}} e_{\mathbf{Q}}^{s\alpha} e^{i\mathbf{Q}\cdot\mathbf{R}_c} \sum_{mn} B_{-\mathbf{Q}mn} a_m^{\dagger} a_n. \quad (19)$$

The lattice distortion depends on the polaron electronic state, as is clear from the factors  $a_m^{\dagger} a_n$ . In addition, the distortion coefficients  $B_{\mathbf{Q}mn}$  can now be identified as generalized Fourier transforms of the lattice distortion. Figure 1(c)-(d) illustrate the lattice distortion to new equilibrium positions, displaced by  $\mathbf{u}_{cs}^0$  relative to the pristine lattice, together with the vibrational coordinates  $\mathbf{u}_{cs}$  relative to the distorted lattice.

After the transformation, the operators  $a_n^{\dagger}$  and  $b_{\mathbf{Q}}^{\dagger}$  create polarons and phonons in the distorted lattice, respectively, analogous to the CHO. The factor  $e_{mn}^{-C}$  in Eq. (15) is analogous to the translation operator  $\exp[-S^{(\text{CHO})}]$ , and thus it relaxes the distortion and sends each phonon to their corresponding vibrational mode of the undistorted lattice. Loosely speaking, Eq. (15) implies that annihilating a polaron from a given site amounts to first annihilating the electronic state from the distorted lattice and then removing the lattice distortion.

The transformed Hamiltonian is derived by replacing the electron and phonon operators with their transformed counterparts. We obtain:

$$\begin{aligned} \tilde{H} = & \sum_{mn} \tilde{\varepsilon}_{mn} a_m^{\dagger} a_n + \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} \left( b_{\mathbf{Q}}^{\dagger} b_{\mathbf{Q}} + \frac{1}{2} \right) \\ & + \frac{1}{\sqrt{N_{\Omega}}} \sum_{mn\mathbf{Q}} \omega_{\mathbf{Q}} (\tilde{g}_{\mathbf{Q}mn} - B_{\mathbf{Q}mn}) (b_{\mathbf{Q}}^{\dagger} + b_{-\mathbf{Q}}) a_m^{\dagger} a_n \\ & + \frac{1}{N_{\Omega}} \sum_{mnij\mathbf{Q}} \omega_{\mathbf{Q}} B_{-\mathbf{Q}ij} (B_{\mathbf{Q}mn} - 2\tilde{g}_{\mathbf{Q}mn}) a_i^{\dagger} a_j a_m^{\dagger} a_n, \end{aligned}$$

where the transformed hopping and e-ph coupling matrices  $\tilde{\varepsilon}_{mn}$  and  $\tilde{g}_{\mathbf{Q}mn}$ , denoted collectively as  $\tilde{M}_{mn}$ , are defined as

$$\tilde{M}_{mn} = \sum_{ij} e_{mi}^C M_{ij} e_{jn}^{-C}. \quad (20)$$

These transformed matrices encode the effect of a polaron dragging the lattice distortion. For example, the hopping term in the transformed Hamiltonian,  $\sum_{mn} a_m^{\dagger} \tilde{\varepsilon}_{mn} a_n = \sum_{mnij} a_m^{\dagger} (e_{mi}^C \varepsilon_{ij} e_{jn}^{-C}) a_n$ , shows that if a polaron hops to a nearby site, the lattice distortion is first removed at the original site by  $e_{jn}^{-C}$ , then the electron hops with amplitude  $\varepsilon_{ij}$ , and then the distortion is created at the new site by  $e_{mi}^C$ .

Note that the transformed matrices  $\tilde{M}$  in Eq. (20) still contain phonon operators  $b_{\mathbf{Q}}$  in the  $e^{\pm C}$  terms. Following Holstein [16], we take the thermal average of the phonon operators, obtaining averaged transformed matrices  $\langle \tilde{M} \rangle_{mn}$  as explained below. After collecting terms, and using the identity

$$a_i^{\dagger} a_j a_m^{\dagger} a_n = a_i^{\dagger} a_n \delta_{mj} - a_i^{\dagger} a_m^{\dagger} a_j a_n,$$

we obtain our effective polaron Hamiltonian:

$$\begin{aligned} \tilde{H} = & \sum_{mn} E_{mn} a_m^{\dagger} a_n + \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} \left( b_{\mathbf{Q}}^{\dagger} b_{\mathbf{Q}} + \frac{1}{2} \right) \\ & + \frac{1}{\sqrt{N_{\Omega}}} \sum_{mn\mathbf{Q}} \omega_{\mathbf{Q}} G_{\mathbf{Q}mn} (b_{\mathbf{Q}}^{\dagger} + b_{-\mathbf{Q}}) a_m^{\dagger} a_n \\ & - \frac{1}{N_{\Omega}} \sum_{ijmn\mathbf{Q}} V_{\mathbf{Q}ijmn} a_i^{\dagger} a_m^{\dagger} a_j a_n, \end{aligned} \quad (21)$$

where the polaron hopping strength  $E_{mn}$ , the residual polaron-phonon (pl-ph) coupling constant  $G_{\mathbf{Q}mn}$ , and the effective polaron-polaron (pl-pl) interaction  $V_{\mathbf{Q}ijmn}$  are defined respectively as

$$\begin{aligned} E_{mn} = & \langle \tilde{\varepsilon} \rangle_{mn} + \frac{1}{N_{\Omega}} \sum_{i\mathbf{Q}} \omega_{\mathbf{Q}} B_{-\mathbf{Q}mi} (B_{\mathbf{Q}in} - 2\langle \tilde{g}_{\mathbf{Q}} \rangle_{in}), \\ G_{\mathbf{Q}mn} = & \langle \tilde{g}_{\mathbf{Q}} \rangle_{mn} - B_{\mathbf{Q}mn}, \end{aligned} \quad (22)$$

$$V_{\mathbf{Q}ijmn} = \omega_{\mathbf{Q}} B_{-\mathbf{Q}ij} (B_{\mathbf{Q}mn} - 2\langle \tilde{g}_{\mathbf{Q}} \rangle_{mn}).$$

Due to the thermal averaging process, all these quantities are now c-numbers rather than phonon operators. Therefore the Hamiltonian in Eq. (21) reduces to an effective tight-binding model, which can be studied with standard approaches. In this work, we assume that the carrier concentration is low enough to neglect the pl-pl interaction  $V_{\mathbf{Q}ijmn}$ .

### C. Thermal average

The thermal average of the transformed matrices,  $\langle \tilde{M} \rangle_{mn}$ , appears above in the polaron Hamiltonian and needs to be evaluated. This thermal average admits an exact expression only in the Holstein model [17], in which all e-ph coupling constants  $g_{\mathbf{Q}mn}$  are zero unless

$m = n$ . However, in the general case, a closed-form expression for  $\langle \widetilde{M} \rangle_{mn}$  cannot be derived without assuming that the distortion coefficients commute with each other:  $[B_{\mathbf{Q}}, B_{\mathbf{Q}'}]_{mn} = 0$  for all pairs of  $\mathbf{Q}$  and  $\mathbf{Q}'$  [22]. There are two main approaches to calculate the thermal average in the general case, the first uses the Feynman disentangling of operators [23] and the second the BCH formula [22]. Here we follow the latter strategy and derive the expression for the thermally-averaged transformed matrices.

First, using the BCH formula in Eq. (5), we have

$$\langle \widetilde{M} \rangle_{mn} = \langle e^C M e^{-C} \rangle_{mn} = \langle M + \frac{1}{2!} [C, [C, M]] + \dots \rangle_{mn},$$

where the angle brackets  $\langle \dots \rangle$  indicate a thermal average over phonon states. In this expression, terms with an odd number of  $C_{mn}$  operators vanish because the thermal average of an odd number of  $b_{\mathbf{Q}}$  or  $b_{\mathbf{Q}}^\dagger$  is zero. Substituting the definition of  $C_{mn}$ , we get

$$\begin{aligned} \langle \widetilde{M} \rangle_{mn} &= \langle \sum_i \frac{1}{(2i)!} [C, [\dots [C, [C, M]] \dots]] \rangle_{mn} \\ &= \sum_i \frac{1}{(2i)! N_{\Omega}^i} \left[ \sum_{\mathbf{Q}_1} B_{\mathbf{Q}_1}, [\dots [\sum_{\mathbf{Q}_{2i}} B_{\mathbf{Q}_{2i}}, M] \dots]] \right] \times \\ &\quad \langle (b_{\mathbf{Q}_1}^\dagger - b_{-\mathbf{Q}_1}) \dots (b_{\mathbf{Q}_{2i}}^\dagger - b_{-\mathbf{Q}_{2i}}) \rangle_{mn}, \end{aligned}$$

where the phonon operator part can be factored out in the last equality because all permutations of  $(b_{\mathbf{Q}}^\dagger - b_{-\mathbf{Q}})$  give the same thermal average. Next, we apply the Wick theorem and use the well-known thermal averages  $\langle b_{\mathbf{Q}}^\dagger b_{\mathbf{Q}} \rangle = N_{\mathbf{Q}}$  and  $\langle b_{\mathbf{Q}} b_{\mathbf{Q}}^\dagger \rangle = N_{\mathbf{Q}} + 1$ , where  $N_{\mathbf{Q}}$  is the phonon thermal occupation, obtaining

$$\begin{aligned} \langle \widetilde{M} \rangle_{mn} &= \sum_i \frac{(-1)^i}{(2i)! N_{\Omega}^i} \sum_{\mathbf{Q}_1 \dots \mathbf{Q}_i} (2N_{\mathbf{Q}_1} + 1) \dots (2N_{\mathbf{Q}_i} + 1) \times \\ &\quad \sum_{\text{all pairings}} [B_{\mathbf{Q}_1}, [B_{-\mathbf{Q}_1}, [\dots [B_{\mathbf{Q}_i}, [B_{-\mathbf{Q}_i}, M]] \dots]]]_{mn}. \end{aligned}$$

Assuming all the distortion coefficients  $B_{\mathbf{Q}}$  commute with each other, the commutator factors are identical for every possible pairing. Under this assumption, each of the  $(2i)!/2^i i!$  possible pairings gives the same contribution, and thus

$$\begin{aligned} \langle \widetilde{M} \rangle_{mn} &= \sum_i \frac{(-1)^i}{N_{\Omega}^i i!} \sum_{\mathbf{Q}_1 \dots \mathbf{Q}_i} (N_{\mathbf{Q}_1} + \frac{1}{2}) \dots (N_{\mathbf{Q}_i} + \frac{1}{2}) \times \\ &\quad [B_{\mathbf{Q}_1}, [B_{-\mathbf{Q}_1}, [\dots [B_{\mathbf{Q}_i}, [B_{-\mathbf{Q}_i}, M]] \dots]]]_{mn}. \end{aligned}$$

Defining the linear operator  $\Lambda$  on  $M$  as

$$\sum_{\alpha\gamma} \Lambda_{mn, \alpha\gamma} M_{\alpha\gamma} = \frac{1}{N_{\Omega}} \sum_{\mathbf{Q}} (N_{\mathbf{Q}} + \frac{1}{2}) [B_{\mathbf{Q}}, [B_{-\mathbf{Q}}, M]]_{mn}, \quad (23)$$

we derive the final expression for the thermal averages

$$\langle \widetilde{M} \rangle_{mn} = \sum_{\alpha\gamma} e^{-\Lambda} M_{\alpha\gamma}. \quad (24)$$

In the special case where all the nonlocal distortion coefficients vanish, i.e. if  $B_{\mathbf{Q}mn}$  is nonzero only if  $m = n$ , the expressions in Eqs. (23) and (24) are exact [17, 22, 24]. In this case, using  $B_{\mathbf{Q}mn} = B_{\mathbf{Q}mm} \delta_{mn}$  in Eq. (23), we obtain:

$$\begin{aligned} \Lambda_{mn, \alpha\gamma} &= \lambda_{mn} \delta_{m\alpha} \delta_{n\gamma}, \\ \lambda_{mn}(T) &= \frac{1}{N_{\Omega}} \sum_{\mathbf{Q}} \left[ N_{\mathbf{Q}}(T) + \frac{1}{2} \right] |B_{\mathbf{Q}mm} - B_{\mathbf{Q}nn}|^2. \end{aligned} \quad (25)$$

With this definition, the thermal average of the transformed matrix becomes

$$\langle \widetilde{M} \rangle_{mn} = \exp(-\lambda_{mn}) M_{mn}. \quad (26)$$

Under these assumptions, the explicit expressions for the transformed polaron hopping and  $e$ -ph coupling are  $\langle \widetilde{\varepsilon} \rangle_{mn} = \exp(-\lambda_{mn}) \varepsilon_{mn}$  and  $\langle \widetilde{g} \rangle_{mn} = \exp(-\lambda_{mn}) g_{\mathbf{Q}mn}$ , respectively.

These expressions greatly simplify the evaluation of  $\langle \widetilde{\varepsilon} \rangle_{mn}$  and  $\langle \widetilde{g} \rangle_{mn}$ , which now involve only the exponential of a specific matrix element, rather than an exponential of an entire matrix as in Eq. (24). Below, we refer to  $e^{-\lambda_{mn}}$  as the *band narrowing factor* because the polaron hopping amplitude  $\langle \widetilde{\varepsilon} \rangle_{mn}$  is suppressed by  $e^{-\lambda_{mn}}$ .

#### D. Small polaron self-localization

Given a specific set of lattice distortion coefficients  $B_{\mathbf{Q}mn}$ , the effective polaron Hamiltonian can be obtained from Eq. (21) using the thermally averaged polaron hopping and pl-ph coupling in Eq. (22). We restrict the distortion coefficients to be local, and set them to

$$B_{\mathbf{Q}mn} = g_{\mathbf{Q}mn} \delta_{mn}. \quad (27)$$

Using this ansatz, the thermal average of the transformed matrix can be written as in Eq. (26), with the exponent  $\lambda_{mn}(T)$  given in Eq. (25), which depends on temperature  $T$  via the thermal phonon occupations  $N_{\mathbf{Q}}(T)$  and on the difference between the local  $e$ -ph coupling at the  $m$  and  $n$  WF sites:

$$\lambda_{mn}(T) = \frac{1}{N_{\Omega}} \sum_{\mathbf{Q}} \left( N_{\mathbf{Q}}(T) + \frac{1}{2} \right) |g_{\mathbf{Q}mm} - g_{\mathbf{Q}nn}|^2. \quad (28)$$

The diagonal part of  $\lambda_{mn}$  vanishes, and thus  $\exp(-\lambda_{mm}) = 1$  for all sites  $m$ . In ionic materials, usually the off-diagonal part of  $\exp(-\lambda_{mn})$  is orders of magnitude smaller than unity (typically of order  $10^{-2}$  to  $10^{-10}$  at 300 K), as we verify explicitly with numerical calculations here and in Ref. [1]. In this case, polaron hopping is negligible, and we have

$$\exp(-\lambda_{mn}) \approx \delta_{mn}. \quad (29)$$

Substituting Eqs. (26), (27) and (29) into Eq. (22), we obtain the key result for materials with negligible polaron hopping:

$$E_{mn} = \left( \varepsilon_{mm} - \frac{1}{N_{\Omega}} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |g_{\mathbf{Q}mm}|^2 \right) \delta_{mn}, \quad (30)$$

$$G_{\mathbf{Q}mn} = g_{\mathbf{Q}mn} \delta_{mn} - B_{\mathbf{Q}mn} = 0. \quad (31)$$

The first equation gives the on-site polaron energy  $E_{mm}$  as the sum of the electronic energy  $\varepsilon_{mm}$  of the WF describing the polaron wave function and the potential energy decrease due to the lattice distortion, analogous to the CHO case [compare the second terms in Eqs. (7) and (30)]. This equation further implies that the operator  $a_m^{\dagger}$  in the polaron Hamiltonian in Eq. (21) creates a self-localized polaron, as hopping to nearby sites is negligible due to the vanishing off-diagonal  $E_{mn}$  hopping amplitudes. The second equation implies that this small polaron is decoupled from all phonon modes as  $G_{\mathbf{Q}mn} = 0$ . Our previous work employed Eq. (30) to compute the polaron energy in various families of ionic materials. In these systems, the band narrowing factor can be approximated as  $e^{-\lambda_{mn}} \approx \delta_{mn}$ , which implies a polaron with a very large effective mass and an ideally flat polaron band.

In more weakly polar materials where polaron hopping is non-negligible, the canonical-transformation formalism is still valid, and as we show here it enables calculations of the polaron band structure. In this more general case, the off-diagonal elements of  $e^{-\lambda_{mn}}$  cannot be neglected, and the polaron Hamiltonian matrix  $E_{mn}$  can be obtained from Eq. (22), using the ansatz in Eq. (27) and the thermal averages in Eq. (26). We obtain:

$$E_{mn} = \left( \varepsilon_{mn} - \frac{2}{N_{\Omega}} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} g_{-\mathbf{Q}mm} g_{\mathbf{Q}mn} \right) e^{-\lambda_{mn}} + \frac{1}{N_{\Omega}} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |g_{\mathbf{Q}mm}|^2 \delta_{mn}, \quad (32)$$

where the on-site polaron energies  $E_{mm}$  are the same as in Eq. (30), but now we also compute the off-diagonal elements, namely the inter-site hopping amplitudes  $E_{mn}$  (with  $m \neq n$ ). Setting up a tight-binding model based on this polaron Hamiltonian matrix allows us to calculate the full polaron band structure.

### E. Small polaron formation

The formation of a self-localized charge carrier in a crystal depends on two competing energies: the kinetic energy increase resulting from localizing the electronic wave function, and the energy decrease from the lattice relaxation around the charge carrier. In our formalism, this competition is clearly seen in the polaron energy in Eq. (30), where  $\varepsilon_{mm}$  is the electronic contribution to the

polaron energy and the negative term proportional to  $\sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |g_{\mathbf{Q}mm}|^2$  is the energy decrease from the lattice relaxation.

Computing the on-site polaron energy  $E_{mm}$  allows us to predict whether a self-localized polaron will form in a material: if  $E_{mm}$  is lower than the conduction band minimum (CBM) for an electron carrier, or higher than the valence band maximum (VBM) for a hole carrier, then the self-localized polaron is energetically more favorable than a delocalized Bloch state. In this scenario, the electron or hole carrier forms a small polaron and becomes self-trapped by the lattice distortion. The polaron formation energy is computed as the difference between the polaron energy  $E_{mm}$  and the respective band edge; thus the formation energy for an electron polaron is

$$\Delta E_f = E_{mm} - \varepsilon_{\text{CBM}}, \quad (33)$$

and for a hole polaron

$$\Delta E_f = \varepsilon_{\text{VBM}} - E_{mm}, \quad (34)$$

where in both cases  $\Delta E_f < 0$  means that polaron formation is energetically favorable.

## III. NUMERICAL CALCULATIONS

### A. Workflow

In the canonical transformation formalism, calculations of small-polaron energies and wave functions are straightforward. In the first step, we generate maximally-localized WFs from the electronic band structure [25], and then calculate the  $e$ -ph matrix elements in the Wannier basis using the standard workflow [20]. The next step consists in verifying numerically the approximation of negligible hopping  $E_{mn} \approx E_{mm} \delta_{mn}$  (or equivalently  $e^{-\lambda_{mn}} \approx \delta_{mn}$ ), which typically works well for strongly polar materials with self-localized polarons. When this approximation holds, we evaluate the on-site polaron energies  $E_{mm}$  using Eq. (30), and then obtain the polaron formation energy using Eqs. (33)-(34). Note that computing  $E_{mm}$  with Eq. (30) is a simple post-processing of the  $e$ -ph calculations; it can be carried out with minimal computational cost (tens of CPU core-hours) using the PERTURBO code [20].

In materials where the off-diagonal elements of  $E_{mn}$  and  $e^{-\lambda_{mn}}$  are non-negligible, we evaluate the full effective polaron Hamiltonian matrix  $E_{mn}$  using Eq. (32), still with minimal computational cost. Starting from the Hamiltonian matrix  $E_{mn}$ , we calculate the polaron band structure using a standard tight-binding approach. Due to its simple workflow, the canonical-transformation method enables rapid calculations of small-polaron energies in a wide range of materials [1], and is particularly promising for high-throughput and data-driven studies of small polarons.

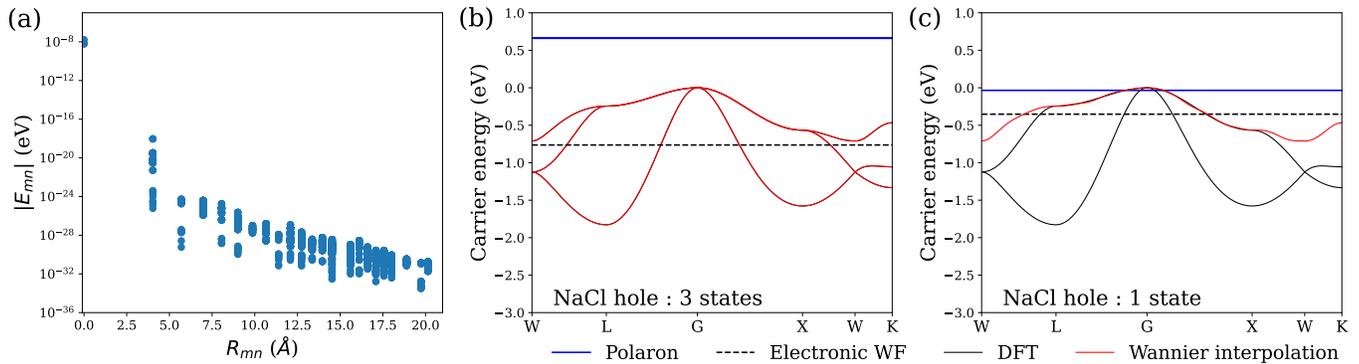


FIG. 2. (a) Computed polaron hopping amplitudes  $E_{mn}$  for hole carriers in NaCl, shown as a function of distance  $R_{mn}$  between the  $m$  and  $n$  WF sites at  $T = 0$  K. (b)-(c) Calculated polaron energy for holes in NaCl, obtained by wannierizing (b) three bands and (c) one band. Blue lines are polaron on-site energies  $E_{mm}$  and dashed black lines are electronic WF energies  $\varepsilon_{mm}$  [see Eq. (30)]. The solid black curves are the DFT band structure and the red curves the Wannier-interpolated bands, equal in number to the WFs used in the polaron calculation. The energy zero is set to the VBM.

## B. Computational details

We carry out numerical calculations on two paradigmatic systems with polarons, a simple ionic material (NaCl) and an organic crystal (naphthalene). For these case studies, we show calculations of polaron hopping amplitudes and polaron band structures, and discuss the choice of WFs to compute the polaron energy.

We carry out plane-wave DFT calculations on NaCl using the QUANTUM ESPRESSO code [26] with norm-conserving pseudopotentials [27, 28] and the Perdew-Burke-Ernzerhof generalized gradient approximation [29]. We use a kinetic energy cutoff of 100 Ry, an  $8 \times 8 \times 8$   $\mathbf{k}$ -point grid and relaxed lattice parameters in all DFT calculations. Density functional perturbation theory [30] is employed to compute phonon frequencies and eigenvectors on a coarse  $8 \times 8 \times 8$   $\mathbf{q}$ -point grid. The  $e$ -ph coupling matrix elements are first computed in the Bloch basis and then transformed to their Wannier basis counterparts,  $g_{\mathbf{Q}mn}$  defined above, using the PERTURBO code [20] with WFs generated from WANNIER90 [25]. The calculations on naphthalene follow the same workflow, using settings and numerical details provided in Ref. [31].

## IV. RESULTS

### A. Polaron hopping amplitude

The polaron energy in the canonical transformation method can be easily computed using Eq. (30). Yet, to use that formula one first needs to verify that the inter-site polaron hopping amplitude in Eq. (32) is negligible, so that  $E_{mn} = E_{mm}\delta_{mn}$ , as a result of a diagonal band-narrowing factor in Eq. (32),  $e^{-\lambda_{mn}} = \delta_{mn}$ . To that end, we compute  $\lambda_{mn}(T)$  by carrying out the numerical integration in Eq. (28), and then obtain  $E_{mn}$  using Eq. (32). Recall that for the diagonal entries with  $m = n$  we have  $e^{-\lambda_{mm}} = 1$  by definition. Therefore the key questions

are how  $E_{mn}$  decays with inter-site distance and whether its off-diagonal entries are small enough to approximate  $E_{mn} = E_{mm}\delta_{mn}$  as in our recent work [1].

Figure 2(a) shows the computed polaron hopping amplitudes  $E_{mn}$  for hole carriers in NaCl as a function of distance between the  $m$  and  $n$  WF sites. The results are given at  $T = 0$  K; as the hopping amplitudes decrease monotonically with temperature, these results are an upper bound to the finite temperature hopping values. We find that even at zero temperature the off-diagonal matrix elements of  $E_{mn}$  are smaller than  $10^{-6}$  eV, and they further decrease with inter-site distance and temperature. Therefore our approach predicts that in NaCl the hole polaron is self-localized and associated with a flat polaron band.

### B. Choice of Wannier functions

Due to the small value of the off-diagonal hopping energies  $E_{mn}$  for holes in NaCl, we can compute the hole polaron energy using Eq. (30). The first step in this calculation is the generation of WFs that accurately interpolate the band structure. However, the choice of WFs is not unique and is a subtle point in our canonical transformation approach. As the goal is to find the lowest-energy polaron state, one could test various choices of WFs, both by changing the WF generation parameters and by wannierizing a different number of bands [25]. Different WFs will lead to relatively small changes of polaron energy, within  $\sim 1$  eV based on our tests. Therefore, if one finds an electron polaron with energy lower than the CBM (or a hole polaron with energy higher than the VBM), then our method guarantees the existence of a self-localized polaron. Conversely, if a stable self-localized polaron is not found, but the polaron energy is within  $\sim 1$  eV of the band edge, it's still possible that a different choice of WFs will lead to a self-localized polaron.

We illustrate the role of different trial wave functions using NaCl as an example. We compare two calculations of the hole polaron energy in NaCl by generating WFs for three bands [Fig. 2(b)] or only one band [Fig. 2(c)], respectively. The calculation using three bands gives polaron energies above the VBM and thus correctly predicts that holes in NaCl form a self-localized small polaron, in agreement with experiments [32]. In the calculation using only one band, the polaron energy is just below the valence band edge, so a polaron is not predicted to form. However, the polaron energy is only 100 meV below the band edge, thus signaling the possible presence of a lower-energy polaron state, as confirmed in Fig. 2(b).

The choice of WFs influences both the on-site electronic energy  $\varepsilon_{mm}$  and the  $e$ -ph coupling  $g_{\mathbf{Q}mm}$  as both contribute to the polaron energy  $E_{mm}$  in Eq. (30). In the NaCl example, the three WFs used in Fig. 2(b) resemble the  $p$  orbitals of Cl and are more spatially localized than the WF used in Fig. 2(c). As a result, they possess a greater on-site hole energy  $|\varepsilon_{mm}|$  (i.e., a lower electronic energy  $\varepsilon_{mm}$  in Fig. 2(b)) and a greater overlap with the phonon perturbation, resulting in a stronger on-site coupling and larger potential energy decrease in the second term of Eq. (30), which leads to a more stable polaron state in Fig. 2(b). Note that the energy scale of these differences is only 0.5–1 eV. Therefore, whenever a self-localized polaron state is clearly unstable (say, by  $> 1$  eV), as we found in Ref. [1] for SrTiO<sub>3</sub>, the result can be trusted without comparing different WFs. Although one can generate and test many WFs, at present the use of a trial polaron wave function equal to a non-uniquely defined WF is a limitation of our method. It can be overcome by formulating a generalized canonical-transformation approach, as we show below in Sec. V C.

### C. Polaron band structure

We carry out a polaron band structure calculation on naphthalene, an organic semiconductor with non-negligible polaron hopping. Electron carriers in naphthalene possess a narrow band width of  $\sim 200$  meV, leading to pronounced polaron effects [31]. We wannierize the two lowest conduction bands, and calculate the polaron band structure using tight-binding with the energies  $E_{mn}$  in Eq. (32). In these calculations, we exclude phonons with energy lower than a cutoff  $\omega_c$  when computing  $\lambda_{mn}$  in Eq. (28) and  $E_{mn}$ . Excluding phonons with up to several times the electronic hopping energy  $|t|$  is physically justified in the canonical transformation when hopping is present [33]. As the charge carrier hops and the lattice rearranges, only phonons with frequency greater than the hopping energy can follow the charge carrier and make up its phonon cloud. Lower energy phonons contribute to the residual pl-ph interaction and can be treated as dynamical disorder [33]. Here, from WF analysis we obtain a maximum electron hopping energy  $|t| \approx 40$  meV.

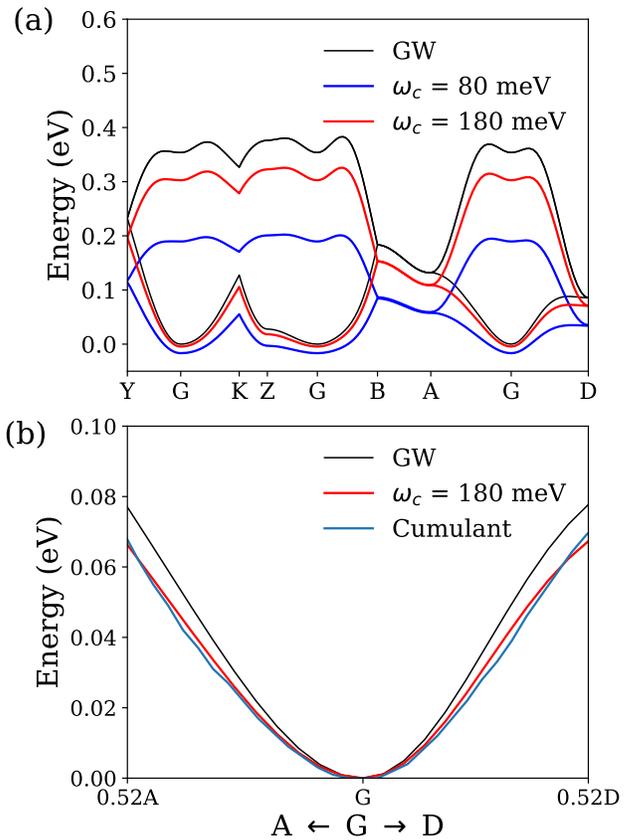


FIG. 3. (a) Band structure of naphthalene computed with the GW method (see Ref. [31]), shown together with the computed polaron band structure at 0 K for two different phonon cutoff energies,  $\omega_c = 80$  meV and 180 meV. (b) Polaron band structure from the canonical transformation (with  $\omega_c = 180$  meV) compared with polaron calculations using the cumulant method (Ref. [31]). The energy zero is the CBM.

Figure 3(a) shows the polaron band structure at 0 K for two different cutoffs,  $\omega_c = 80$  and 180 meV; the first cutoff is twice the hopping energy and the second is as high as possible but below a strongly coupled optical phonon. For both cutoffs, the minima of the polaron bands are below the conduction band, signaling the presence of a stable polaron. Even at 0 K, the polaron bands are narrower than the electronic GW bands due to the zero-point term in  $\lambda_{mn}$ . The band narrowing decreases for greater cutoff values, leaving an adjustable parameter in the theory. Guided by Holstein’s work [16], we argue that the optimal cutoff is the highest phonon energy that permits the inclusion of strongly coupled optical phonons (here this value is 180 meV). In Fig. 3(b) we compare the low-energy polaron dispersion in the canonical transformation with  $\omega_c = 180$  meV to our recent cumulant calculation in naphthalene [31]. For this optimal cutoff, the two methods give polaron band structures in quantitative agreement – both methods predict a polaron with dispersive bands and modest mass renormalization.

## V. METHOD COMPARISON

### A. Energy functional method

We briefly summarize the formalism of Sio et al. [2, 3] for first-principles polaron calculations. Their approach models the polaron as a *single* excess charge carrier, and calculates its electronic wave function  $\psi$  and associated lattice distortion  $u_\kappa^0$  by minimizing the polaron energy functional {see Eq. (23) in Ref. [3]}:

$$E_p[\psi, u_\kappa^0] = \frac{1}{2} \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_\kappa^0 u_{\kappa'}^0 \quad (35)$$

$$+ \int d\mathbf{r} \psi^*(\mathbf{r}) \left( H_{\text{KS}} + \sum_{\kappa} \frac{\partial V_{\text{KS}}}{\partial u_\kappa^0} u_\kappa^0 \right) \psi(\mathbf{r}),$$

where  $\kappa$  is a composite index for atoms and Cartesian coordinates (similar to Eq. (18),  $\kappa = c s \alpha$ ), and  $\Phi_{\kappa\kappa'}$  are interatomic force constants;  $H_{\text{KS}}$  and  $V_{\text{KS}}$  are, respectively, the Kohn-Sham (KS) Hamiltonian and KS potential at equilibrium without the excess charge carrier.

In Refs. [2, 3], the wave function  $\psi$  of the excess electron is written as a superposition of Bloch states (with band index  $i$  and crystal momentum  $\mathbf{k}$ ) or WFs (with composite index  $m = j_m \mathbf{R}_m$ , as above):

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{N_\Omega}} \sum_{i\mathbf{k}} A_{i\mathbf{k}} \psi_{i\mathbf{k}}(\mathbf{r}) = \sum_m A_m w_m(\mathbf{r}), \quad (36)$$

where  $A_{i\mathbf{k}}$  and  $A_m$  are expansion coefficients for the wave function in the Bloch and Wannier basis, respectively. The lattice distortion  $u_\kappa^0$  due to the excess charge carrier is expanded in the basis of phonon eigenvectors as

$$u_{c s \alpha}^0 = -\frac{2}{N_\Omega} \sum_{\mathbf{Q}} B_{\mathbf{Q}}^* \frac{1}{\sqrt{2M_s \omega_{\mathbf{Q}}}} e_{\mathbf{Q}}^{s\alpha} e^{i\mathbf{q} \cdot \mathbf{R}_c}, \quad (37)$$

where  $B_{\mathbf{Q}}$  are scalar lattice-distortion coefficients that are independent of electronic band or WF site. Minimizing the polaron energy functional with respect to  $\psi^*$  and  $u_\kappa^0$  gives a set of coupled polaron equations for the coefficients  $A_{i\mathbf{k}}$  and  $B_{\mathbf{Q}}$  {see Eqs. (37)-(38) in Ref. [3]} [34]:

$$\frac{2}{N_\Omega} \sum_{i'\mathbf{q}} \omega_{\mathbf{Q}} B_{\mathbf{Q}} g_{i'i\nu}^*(\mathbf{k}, \mathbf{q}) A_{i'\mathbf{k}+\mathbf{q}} = (\varepsilon_{i\mathbf{k}} - \varepsilon) A_{i\mathbf{k}} \quad (38)$$

$$B_{\mathbf{Q}} = \frac{1}{N_k} \sum_{i'\mathbf{k}+\mathbf{q}} A_{i'\mathbf{k}+\mathbf{q}}^* g_{i'i\nu}(\mathbf{k}, \mathbf{q}) A_{i\mathbf{k}} \quad (39)$$

which are solved self-consistently by first assuming a set of coefficients  $B_{\mathbf{Q}}$  and then solving for  $A_{i\mathbf{k}}$  in the first equation. The process is then repeated until convergence. The resulting polaron formation energy is [3]

$$\Delta E_f = \varepsilon - \varepsilon_{\text{CBM}} + \frac{1}{2} \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_\kappa^0 u_{\kappa'}^0$$

$$= \frac{1}{N_{\mathbf{k}}} \sum_{i\mathbf{k}} |A_{i\mathbf{k}}|^2 (\varepsilon_{i\mathbf{k}} - \varepsilon_{\text{CBM}}) - \frac{1}{N_\Omega} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |B_{\mathbf{Q}}|^2. \quad (40)$$

### B. Comparison I: Polaron energy

We now prove the equivalence of the small-polaron energy in the canonical transformation and energy functional methods: for a given small polaron wave function, the on-site polaron energy computed using Eq. (30) in our method is identical to the polaron energy in Refs. [2, 3].

In our canonical transformation formalism [1], we use a single WF as the trial small-polaron wave function. Suppose this WF is centered at site 0, then the polaron wave function is

$$|w_0\rangle = a_0^\dagger |0\rangle. \quad (41)$$

This wave function is defined in the canonical transformed Hamiltonian. For this state, we can set all the distortion coefficients to zero except  $B_{\mathbf{Q}00}$ , as can be seen from the expectation value of Eq. (16):  $\langle 0 | a_0 \sum_{mn} B_{\mathbf{Q}mn} a_m^\dagger a_n a_0^\dagger | 0 \rangle = B_{\mathbf{Q}00}$ . From Eqs. (27) and (30), the on-site energy of this polaron state is

$$E_{00} = \varepsilon_{00} - \frac{1}{N_\Omega} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |g_{\mathbf{Q}00}|^2, \quad (42)$$

where  $\varepsilon_{00}$  is the WF energy and  $g_{\mathbf{Q}00}$  the  $e$ -ph matrix element in Wannier basis at the polaron site. Both of these quantities depend on the WF choice, as discussed above. Without loss of generality, we assume that the WF with lowest on-site polaron energy  $E_{00}$  can be expanded in Bloch basis as

$$|w_0\rangle = \frac{1}{\sqrt{N_k}} \sum_{i\mathbf{k}} A_{i\mathbf{k}} |\psi_{i\mathbf{k}}\rangle. \quad (43)$$

Using the transformation between Wannier and Bloch basis [see Eq. (A3) in Appendix A], the polaron formation energy  $\Delta E_f = E_{00} - \varepsilon_{\text{CBM}}$  is

$$\Delta E_f = \frac{1}{N_k} \sum_{i\mathbf{k}} |A_{i\mathbf{k}}|^2 (\varepsilon_{i\mathbf{k}} - \varepsilon_{\text{CBM}}) - \frac{1}{N_\Omega} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |g_{\mathbf{Q}00}|^2, \quad (44)$$

where

$$g_{\mathbf{Q}00} = \frac{1}{N_k} \sum_{i'\mathbf{k}} A_{i'\mathbf{k}+\mathbf{q}}^* g_{i'i\nu}(\mathbf{k}, \mathbf{q}) A_{i\mathbf{k}}. \quad (45)$$

This polaron formation energy  $\Delta E_f$  and on-site  $e$ -ph coupling  $g_{\mathbf{Q}00}$  are exactly the same as, respectively, the formation energy and distortion coefficient  $B_{\mathbf{Q}}$  in Refs. [2, 3]. More precisely, the polaron energy  $E_{00}$  is identical to the eigenvalue  $\varepsilon$  of Eq. (38), the first polaron equation in Sio et al. [3], provided that we add to their eigenvalue the elastic energy associated with the polaron lattice distortion,  $\sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |B_{\mathbf{Q}}|^2 = 1/2 \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_\kappa^0 u_{\kappa'}^0$  (see Appendix A). This constant term is included in the canonical transformation polaron energy  $E_{00}$ , whereas in the energy functional method of Ref. [3] it is added to the eigenvalue after the calculation [see Eq. (40)].

Available numerical results confirm this equivalence for cases where the polaron is self-localized and well described by a single WF. For example, we recently computed the polaron formation energy for electrons in  $\text{Li}_2\text{O}_2$ , and obtained a value of -4.905 eV [1] that is nearly identical to the -4.87 eV value found in Ref. [3]. Note that in our canonical transformation method the polaron energy is computed straightforwardly using Eq. (30), with negligible computational cost even for large systems. By contrast, the energy functional method in Refs. [2, 3] requires solving an eigenvalue problem self-consistently and extrapolating the result to an infinite  $\mathbf{k}$ -point grid size, with significant computational cost. The key advantage of our approach is the use of WFs as a more natural, localized basis set to describe small polarons, which enables bypassing costly calculations in momentum space.

Comparing the two methods for a more delocalized polaron wave function is more challenging. On one hand, our approach can be generalized to take into account an arbitrary polaron wave function, giving a general canonical transformation formalism that can treat both small and large polarons (see below). On the other hand, the method of Refs. [2, 3] can already describe an arbitrary polaron wave function (as a superposition of WFs at multiple sites), but as we show below it has important limitations for addressing key polaron physics such as thermal effects and polaron localization.

### C. Comparison II: Polaron wave function

*Canonical transformation method.* In the canonical transformation formalism, it is not obvious how to determine the polaron wave function [1]. While Eq. (30) gives the polaron energy for a WF localized at site  $m$ , the choice of this WF is not unique. Different WFs may lead to different electronic energies  $\epsilon_{mm}$  and  $e$ -ph interactions  $g_{\mathbf{Q}mm}$ , and thus different polaron energies  $E_{mm}$ . The most stable polaron state corresponds to the wave function minimizing the polaron energy, but that wave function may be a nontrivial combination of WFs.

We present an extension to our canonical transformation formalism using a general polaron wave function written as a superposition of WFs,

$$|\psi\rangle = \sum_m A_m a_m^\dagger |0\rangle = \sum_m A_m |w_m\rangle \quad (46)$$

with normalization  $\sum_m |A_m|^2 = 1$ . The energy of this polaron state in this generalized canonical-transformation formalism is

$$\begin{aligned} E[A, B] &= \langle \psi | \tilde{H} | \psi \rangle = \sum_{mn} A_m^* A_n \langle 0 | a_m^\dagger \tilde{H} a_n | 0 \rangle \\ &= \frac{1}{N_\Omega} \sum_m |A_m|^2 \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |B_{\mathbf{Q}mm}|^2 \\ &+ \sum_{mn} A_m^* A_n e^{-\lambda_{mn}} \left( \epsilon_{mn} - \frac{2}{N_\Omega} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} B_{-\mathbf{Q}mm} g_{\mathbf{Q}mn} \right), \end{aligned} \quad (47)$$

where we ignored the off-diagonal matrix elements of  $B_{\mathbf{Q}mn}$ . This energy functional depends on the Wannier-basis coefficients  $A_m$  of the electronic wave function and on the lattice distortion coefficients  $B_{\mathbf{Q}mm}$  describing how the lattice responds to the charge carrier.

The resulting full wave function  $|\Phi\rangle$  of the combined electron-plus-phonon system has a general form with *entangled* electrons and phonons. In the canonical transformation framework, the full wave function in the original Hamiltonian is

$$\begin{aligned} |\Phi\rangle &= e^S |\psi\rangle \\ &= \sum_m A_m e^{\frac{1}{\sqrt{N_\Omega}} \sum_{\mathbf{Q}} B_{\mathbf{Q}mm} (b_{\mathbf{Q}}^\dagger - b_{-\mathbf{Q}})} a_m^\dagger |0\rangle \end{aligned} \quad (48)$$

where  $S$  is defined in Eq. (13) and we neglected the off-diagonal matrix elements of  $B_{\mathbf{Q}mn}$ .

Full minimization of the energy functional in Eq. (47) with respect to  $A_m$  and  $B_{\mathbf{Q}mm}$  is challenging and will be pursued in future work. For materials with strong  $e$ -ph interactions leading to self-localized small polarons, the ansatz introduced above of a polaron localized at a single site that is free of hopping (obtained respectively by setting  $A_n = \delta_{n0}$  and  $B_{\mathbf{Q}mn} = g_{\mathbf{Q}mm} \delta_{mn}$ ) is a good approximation to the stationary solution of the generalized canonical-transformation functional in Eq. (47). In that approximation, the polaron wave function reduces to the tensor product state

$$|\Phi\rangle = e^{\frac{1}{\sqrt{N_\Omega}} \sum_{\mathbf{Q}} g_{\mathbf{Q}00} (b_{\mathbf{Q}}^\dagger - b_{-\mathbf{Q}})} a_0^\dagger |0\rangle \quad (49)$$

defined uniquely by one distortion coefficient for each phonon mode,  $B_{\mathbf{Q}} = g_{\mathbf{Q}00}$ . For a more general polaron state, if one uses site-dependent distortion coefficients  $B_{\mathbf{Q}mm}$  together with an electronic wave function with amplitude at multiple sites, then the full wave function  $|\Phi\rangle$  entangles the electrons and phonons, and includes both the solution in Eq. (49) and the approach by Sio et al. [3] as subcases, as we show below.

*Energy functional method.* In the formalism of Refs. [2, 3], the full wave function is effectively a tensor product of the electron and phonon wave functions:

$$\begin{aligned} |\Phi\rangle &= \left( \sum_m A_m a_m^\dagger |0\rangle_e \right) \otimes \left( e^{\frac{1}{\sqrt{N_\Omega}} \sum_{\mathbf{Q}} B_{\mathbf{Q}} (b_{\mathbf{Q}}^\dagger - b_{-\mathbf{Q}})} |0\rangle_p \right) \\ &= \sum_m A_m e^{\frac{1}{\sqrt{N_\Omega}} \sum_{\mathbf{Q}} B_{\mathbf{Q}} (b_{\mathbf{Q}}^\dagger - b_{-\mathbf{Q}})} a_m^\dagger |0\rangle, \end{aligned} \quad (50)$$

where  $|0\rangle_e$  and  $|0\rangle_p$  are the electron and phonon vacuums, respectively, and  $|0\rangle = |0\rangle_p |0\rangle_e$  the combined vacuum.

We briefly outline the derivation of this result. The phonon state in Ref. [2, 3] is characterized only by the classical displacements  $u_\kappa^0$ . As discussed at the end of in Sec. II A, the ground state of a shifted harmonic oscillator centered at  $u^0 = \frac{B}{\sqrt{2m\omega}}$  can be described by the coherent state [35]

$$|u^0\rangle = e^{-\frac{B}{2}(b^\dagger - b)} |0\rangle_p. \quad (51)$$

We can extend this state to describe the entire distorted lattice, with the position of each atom  $\kappa$  displaced by  $u_\kappa^0$  as in Eq. (37):

$$|\{u_\kappa^0\}\rangle = e^{\frac{1}{\sqrt{N_\Omega}} \sum_{\mathbf{Q}} B_{\mathbf{Q}}[A](b_{\mathbf{Q}}^\dagger - b_{-\mathbf{Q}})} |0\rangle_p, \quad (52)$$

from which one obtains the full wave function in Eq. (50) for the approach in Refs. [2, 3]. Here, we noted explicitly that the distortion coefficients  $B_{\mathbf{Q}}[A]$  are functionals of the wave function coefficients  $A_m$  via the second polaron equation in Wannier form [see Eq. (A2)].

Therefore, while the electrons and phonons are disentangled as a tensor-product state in the energy-functional method, they are still coupled in a mean-field sense via the distortion coefficients  $B_{\mathbf{Q}}[A]$  from the second polaron equation, which describes the mean-field effect of the electrons and  $e$ -ph coupling on the lattice. Starting from our wave function in Eq. (48), if we make the canonical-transformation distortion coefficients  $B_{\mathbf{Q}mm}$  independent of site and electron orbital index  $m$ ,  $B_{\mathbf{Q}mm} \rightarrow B_{\mathbf{Q}}$ , we obtain the variational ansatz in Eq. (50) as a special case of the canonical transformation formalism. Similarly, the polaron energy functional in Refs. [2, 3] can be viewed a special case of the generalized canonical transformation approach given in Eq. (47).

#### D. Comparison III: Temperature dependence

Thermal effects are essential in polaron physics. For small-polarons, the polaron bands become progressively flatter as the temperature increases, until the charge carrier ultimately self-localizes. In the canonical transformation formalism, these effects are encoded in the temperature dependent band-narrowing factor,  $\exp[-\lambda_{mn}(T)]$ . Analysis of the exponent  $\lambda_{mn}(T)$  in Eq. (25) shows that the band-narrowing factor decreases with temperature due to an increase in the phonon occupations  $N_{\mathbf{Q}}$ , leading to a progressive flattening of the polaron bands, consistent with the picture discussed above. Even at  $T=0$  K, where the phonon occupations  $N_{\mathbf{Q}}$  vanish, the exponent  $\lambda_{mn}(T)$  can still be relatively large due to the zero-point motion of the lattice, resulting in a significant zero temperature band-narrowing factor:

$$e^{-\lambda_{mn}(T=0)} = e^{-\frac{1}{2N_\Omega} \sum_{\mathbf{Q}} |B_{\mathbf{Q}mm} - B_{\mathbf{Q}nn}|^2}. \quad (53)$$

This result shows that there is a finite zero-point polaron band renormalization due to the site and WF dependence of the distortion coefficients.

In the energy functional approach of Refs. [2, 3], these thermal and zero-point effects are missing entirely, an important limitation for a polaron theory. This point is clearly seen in their energy functional in Eq. (35), which includes only the static lattice distortion  $u_\kappa^0$  but no terms associated with lattice vibrations. It can be better understood by comparing their first polaron equation in Wannier form, Eq. (A1), with the corresponding term in

the generalized canonical-transformation functional, the third line in Eq. (47). This comparison shows that the approach by Sio et al. [2, 3] is equivalent to assuming  $e^{\lambda_{mn}} = 1$ , or equivalently  $\lambda_{mn} = 0$ , which neglects both the zero- and finite-temperature polaron band narrowing. Their lack of band narrowing and temperature dependence is a consequence of not including the lattice vibrations, which is equivalent to setting  $N_{\mathbf{Q}} + 1/2 = 0$  in Eq. (28) for the canonical transformation formalism. Note also that the energy functional method in Refs. [2, 3] describes the polaron as an isolated system (essentially, a localized defect), and thus a polaron band structure is missing altogether. This is why that method cannot be extended straightforwardly to include thermal effects on the polaron band structure and effective mass.

The effect of temperature is also critical for polaron dynamics. A polaron can hop from between different sites assisted by the thermal lattice vibrations, and the distortion gets transferred to the new site [4]. In many materials with polaron effects, as temperature increases charge transport transitions from a band-like mechanism to thermally-activated charge hopping. Due to its thermally activated nature, describing charge hopping requires distortion coefficients that depend on site and electronic state. In the canonical transformation formalism, the distortion coefficients  $B_{\mathbf{Q}mn}$  are associated with electronic hopping amplitudes between WF sites  $m$  and  $n$ , via terms proportional to  $B_{\mathbf{Q}mn} a_m^\dagger a_n$  that couple explicitly the electron and lattice dynamics. By contrast, the energy functional formalism, using site- and electronic state-independent distortion coefficients  $B_{\mathbf{Q}}$  [2, 3], couples the electron and lattice dynamics in a mean field way, as seen in the second polaron equation in Wannier basis, Eq. (A2).

#### E. Comparison IV: Polaron localization

In the canonical transformation method, polaron self-localization is easy to verify starting from the effective Hamiltonian. When the condition  $e^{-\lambda_{mn}(T)} \approx \delta_{mn}$  is satisfied, the Hamiltonian reduces to the diagonal matrix  $E_{mn}$  in Eq. (30), and thus the polaron is localized at a single site with a vanishing hopping amplitude. In this scenario, when the canonical transformation method predicts a polaron on-site energy lower than the band edge, we conclude that the formation of a self-localized polaron with a nearly flat polaron band is energetically favorable. In materials with non-negligible polaron hopping, one can use Eq. (32) in the canonical transformation approach to compute the temperature dependent polaron band structure.

To guarantee that these polaron band structure calculations are physically meaningful, the polaron Hamiltonian matrix  $E_{mn}$  in Eq. (32) needs to have the same translation symmetry as the lattice. This translational invariance is simple to show in our canonical transformation formalism. Recall that the WF index

$n$  is a composite index labeling both the site and WF,  $n = j_n \mathbf{R}_n$ . We translate the  $m$  and  $n$  WFs by a lattice vector  $\mathbf{R}$ , shifting them to new sites  $m' = j_m \mathbf{R}_m + \mathbf{R}$  and  $n' = j_n \mathbf{R}_n + \mathbf{R}$ . Using Eq. (A3), the translated  $e$ -ph coupling matrix elements become  $g_{\mathbf{Q}m'n'} = g_{\mathbf{Q}mn} e^{i\mathbf{q}\cdot\mathbf{R}}$ , and using this relation to evaluate the translated Hamiltonian matrix  $E_{m'n'}$  in Eq. (32), one obtains the translational invariance condition  $E_{m'n'} = E_{mn}$ .

The situation is different in the energy functional method [2, 3], where the polaron is described as an isolated system consisting of a single charge carrier plus a lattice distortion around it. The distorted lattice induces a local potential which explicitly breaks the translational symmetry. As a result, a polaron band structure cannot be defined at any temperature, and polaron self-localization is not deduced from a vanishing polaron hopping or bandwidth. Rather, in the energy functional method, polaron self-localization is inferred from the presence of a bound state in a static potential generated by the distorted lattice [2, 3]. This approach hides the complex physics of the polaron problem, with key temperature dependent inter-site hopping, and treats it as a simple quantum mechanical problem of a particle in a localized potential. Yet, in reality the attractive potential felt by the excess electron is neither static nor temperature independent, as it is determined by the zero-point and thermal motion of the lattice.

For small polarons, self-localization cannot be guaranteed by the presence of a bound state in the method of Refs. [2, 3], unless one can prove a negligible inter-site hopping and extend the method to finite temperatures. For materials with non-negligible polaron hopping, or where polaron localization varies significantly with temperature, the lack of translation symmetry, hopping amplitude, and thermal effects in the energy functional method of Refs. [2, 3] currently prevents quantitative comparisons with our canonical transformation approach.

## VI. CONCLUSION

This work analyzes and compares two methods that advance first-principles studies of polarons. Both methods can compute the polaron energy and lattice distortion with calculations that use only a unit cell of the material. These approaches leverage *ab initio*  $e$ -ph calculations and related software packages to carry them out efficiently on modern computer architectures. We have highlighted the proper treatment of thermal effects, translational invariance, and polaron self-localization in the canonical transformation framework.

We believe that more work is needed to bring our canonical transformation method to full fruition. We have shown that it can be extended to explicitly compute the polaron wave function and treat both small and more delocalized polarons. As it includes an explicit coupling of charge hopping and lattice distortion, the canon-

ical transformation method can also be extended, using linear-response theory, to study charge transport in the polaron hopping regime. Our analysis highlighted a common root for the canonical transformation and energy functional methods, suggesting that proper extensions of both approaches will enable exciting future developments in polaron physics.

## ACKNOWLEDGEMENTS

The authors thank Nien-En Lee for fruitful discussions. This work was supported by the Air Force Office of Scientific Research through the Young Investigator Program, Grant FA9550-18-1-0280. M.B. was partially supported by the Liquid Sunlight Alliance, which is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award No. DE-SC0021266. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

## Appendix A

To compare the polaron energy in the canonical transformation and energy functional methods, we use the identity [25]

$$A_{i\mathbf{k}} = \sum_{j_m \mathbf{R}_m} e^{-i\mathbf{k}\cdot\mathbf{R}_m} \mathcal{U}_{ij_m}^{\mathbf{k}} A_{j_m \mathbf{R}_m}$$

to replace  $A_{i\mathbf{k}}$  in Eqs. (38) and (39), where  $\mathcal{U}_{ij_m}^{\mathbf{k}}$  is the unitary matrix from the WF generation process [25]. The two polaron equations of Ref. [3], rewritten in the Wannier basis, become

$$\sum_{mn} A_m^* A_n \left[ \varepsilon_{mn} - \frac{2}{N_\Omega} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} B_{-\mathbf{Q}} g_{\mathbf{Q}mn} \right] = \varepsilon, \quad (\text{A1})$$

$$B_{\mathbf{Q}} = \sum_{mn} A_m^* g_{\mathbf{Q}mn} A_n, \quad (\text{A2})$$

where we have used the following definitions for the real-space electron hopping amplitude and  $e$ -ph coupling constants:

$$\varepsilon_{mn} = \frac{1}{N_\Omega} \sum_{i\mathbf{k}} e^{-i\mathbf{k}\cdot(\mathbf{R}_n - \mathbf{R}_m)} \mathcal{U}_{jm_i}^{\dagger\mathbf{k}} \varepsilon_{i\mathbf{k}} \mathcal{U}_{ij_n}^{\mathbf{k}}, \quad (\text{A3})$$

$$g_{\mathbf{Q}mn} = \frac{1}{N_\Omega} \sum_{i'\mathbf{k}} e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{R}_m} \mathcal{U}_{i'j_m}^{\dagger\mathbf{k}+\mathbf{q}} g_{i'iv}(\mathbf{k}, \mathbf{q}) \mathcal{U}_{ij_n}^{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}_n}.$$

The equality used above for the elastic energy,

$$\frac{1}{2} \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_{\kappa}^0 u_{\kappa'}^0 = \frac{1}{N_\Omega} \sum_{\mathbf{Q}} \omega_{\mathbf{Q}} |B_{\mathbf{Q}}|^2, \quad (\text{A4})$$

can be derived by substituting in the left-hand side the expression for  $u_{\kappa}^0$  in Eq. (37) and using the identities:

$$\begin{aligned}\Phi_{\kappa\kappa'} &= \frac{\sqrt{M_s M_{s'}}}{N_{\Omega}} \sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot(\mathbf{R}_{c'} - \mathbf{R}_c)} D_{s\alpha, s'\alpha'}(\mathbf{q}), \\ \sum_{s'\alpha'} D_{s\alpha, s'\alpha'}(\mathbf{q}) e_{\mathbf{Q}}^{s'\alpha'} &= \omega_{\mathbf{Q}}^2 e_{\mathbf{Q}}^{s\alpha}, \\ \sum_{s\alpha} e_{\nu\mathbf{q}}^{*s\alpha} e_{\nu'\mathbf{q}}^{s\alpha} &= \delta_{\nu\nu'}\end{aligned}$$

where  $D_{s\alpha, s'\alpha'}(\mathbf{q})$  is the dynamical matrix.

A subtle question is why the polaron energy includes this elastic energy term in the canonical transformation but not in the energy functional approach, where it needs to be added to the eigenvalue of the first polaron equation [see Eq. (40)]. A correction to the energy functional in Refs. [2, 3] allows us to properly include the elastic energy in the polaron equation eigenvalue. The functional in Eq. (35) suffers from an inaccuracy: when the excess charge vanishes, as can be obtained by setting the polaron wave function  $\psi = 0$ , the polaron energy does not vanish: incorrectly, it equals the elastic energy induced by the polaron. This unphysical behavior can be addressed by properly coupling the elastic energy with its source, the charge distribution  $|\psi(\mathbf{r})|^2$ , in the energy functional, changing the first term in Eq. (35) as

$$\frac{1}{2} \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_{\kappa}^0 u_{\kappa'}^0 \rightarrow \frac{1}{2} \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_{\kappa}^0 u_{\kappa'}^0 \int d\mathbf{r} |\psi(\mathbf{r})|^2. \quad (\text{A5})$$

This way,  $\psi = 0$  gives a polaron energy  $E_p = 0$ , and the revised energy functional becomes:

$$\begin{aligned}E_p[\psi, u_{\kappa}^0] &= \int d\mathbf{r} \psi^*(\mathbf{r}) \times \\ &\left( \frac{1}{2} \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_{\kappa}^0 u_{\kappa'}^0 + \sum_{\kappa} \frac{\partial V_{\text{KS}}}{\partial u_{\kappa}^0} u_{\kappa}^0 + H_{\text{KS}} \right) \psi(\mathbf{r}).\end{aligned} \quad (\text{A6})$$

Varying with respect to the polaron wave function, under the constraint of its normalization, gives a revised first polaron equation that properly includes the elastic energy in the eigenvalue:

$$\begin{aligned}\frac{\delta}{\delta\psi^*} \left[ E_p - \varepsilon \left( \int d\mathbf{r} |\psi(\mathbf{r})|^2 - 1 \right) \right] &= 0 \quad \rightarrow \\ \left( \frac{1}{2} \sum_{\kappa\kappa'} \Phi_{\kappa\kappa'} u_{\kappa}^0 u_{\kappa'}^0 + \sum_{\kappa} \frac{\partial V_{\text{KS}}}{\partial u_{\kappa}^0} u_{\kappa}^0 + H_{\text{KS}} \right) \psi(\mathbf{r}) &= \varepsilon \psi(\mathbf{r}).\end{aligned} \quad (\text{A7})$$

The solution of this equation is identical to its version without the elastic energy term proposed in Ref. [3]. However, now the polaron eigenvalue  $\varepsilon$  can be directly interpreted as the polaron energy since from Eqs. (A6) and (A7) one obtains  $E_p[\psi, u_{\kappa}^0] = \varepsilon$ . Therefore, the polaron formation energy is now given by

$$\Delta E_f = \varepsilon - \varepsilon_{\text{CBM}}, \quad (\text{A8})$$

consistent with Eq. (33) in the canonical transformation formalism, because the elastic energy has now been absorbed in the polaron eigenvalue  $\varepsilon$ .

- 
- [1] N.-E. Lee, H.-Y. Chen, J.-J. Zhou, and M. Bernardi, Facile ab initio approach for self-localized polarons from canonical transformations, *Phys. Rev. Mater.* **5**, 063805 (2021).
- [2] W. H. Sio, C. Verdi, S. Ponc e, and F. Giustino, Polarons from first principles, without supercells, *Phys. Rev. Lett.* **122**, 246403 (2019).
- [3] W. H. Sio, C. Verdi, S. Ponc e, and F. Giustino, Ab initio theory of polarons: Formalism and applications, *Phys. Rev. B* **99**, 235139 (2019).
- [4] D. Emin, Small polarons, *Phys. Today* **35**, 34 (1982).
- [5] E. Goovaerts and D. Schoemaker, Inelastic light scattering of the  $v_k$  centers in the alkali halides, *Phys. Stat. Sol. (b)* **88**, 615 (1978).
- [6] H. Sezen, H. Shang, F. Bebensee, C. Yang, M. Buchholz, A. Nefedov, S. Heissler, C. Carbogno, M. Scheffler, P. Rinke, and C. W oll, Evidence for photogenerated intermediate hole polarons in ZnO, *Nat. Commun.* **6**, 6901 (2015).
- [7] E. Pastor, J.-S. Park, L. Steier, S. Kim, M. Gr atzel, J. R. Durrant, A. Walsh, and A. A. Bakulin, In situ observation of picosecond polaron self-localisation in  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub> photoelectrochemical cells, *Nat. Commun.* **10**, 3962 (2019).
- [8] P. Nagels, M. Denayer, and J. Devreese, Electrical properties of single crystals of uranium dioxide, *Solid State Commun.* **1**, 35 (1963).
- [9] C. Crevecoeur and H. D. Wit, Electrical conductivity of Li doped MnO, *J. Phys. Chem. Solids* **31**, 783 (1970).
- [10] A. S. Alexandrov and J. T. Devreese, *Advances in Polaron Physics*, 1st ed. (Springer, 2010).
- [11] D. Emin, *Polarons* (Cambridge University Press, 2013).
- [12] J. B. Varley, A. Janotti, C. Franchini, and C. G. Van de Walle, Role of self-trapping in luminescence and  $p$ -type conductivity of wide-band-gap oxides, *Phys. Rev. B* **85**, 081109 (2012).
- [13] S. Kokott, S. V. Levchenko, P. Rinke, and M. Scheffler, First-principles supercell calculations of small polarons with proper account for long-range polarization effects, *New J. Phys.* **20**, 033023 (2018).
- [14] M. Reticcioli, U. Diebold, G. Kresse, and C. Franchini, Small Polarons in Transition Metal Oxides, in *Handbook of Materials Modeling* (Springer International Publishing, 2020) pp. 1–39.
- [15] N. Osterbacka, P. Erhart, S. Falletta, A. Pasquarello, and J. Wiktor, Small electron polarons in CsPbBr<sub>3</sub>: Competition between electron localization and delocalization, *Chem. Mater.* **32**, 8393 (2020).
- [16] T. Holstein, Studies of polaron motion: Part ii. the “small” polaron, *Ann. Phys. (N. Y.)* **8**, 343 (1959).
- [17] G. D. Mahan, *Many-Particle Physics*, 3rd ed. (Springer, 2000).
- [18] T. D. Lee, F. E. Low, and D. Pines, The motion of slow electrons in a polar crystal, *Phys. Rev.* **90**, 297 (1953).
- [19] S. Tomonaga, On the Effect of the Field Reactions on the Interaction of Mesotrons and Nuclear Particles. III, *Prog. Theor. Phys.* **2**, 6 (1947).
- [20] J.-J. Zhou, J. Park, I.-T. Lu, I. Maliyov, X. Tong, and M. Bernardi, PERTURBO: A software package for ab initio electron–phonon interactions, charge transport and ultrafast dynamics, *Comput. Phys. Commun.* **264**, 107970 (2021).
- [21] M. Bernardi, First-principles dynamics of electrons and phonons, *Eur. Phys. J. B* **89**, 239 (2016).
- [22] K. Hannewald, V. M. Stojanovi c, J. M. T. Schellekens, P. A. Bobbert, G. Kresse, and J. Hafner, Theory of polaron bandwidth narrowing in organic molecular crystals, *Phys. Rev. B* **69**, 075211 (2004).
- [23] R. W. Munn and R. Silbey, Theory of electronic transport in molecular crystals. ii. zeroth order states incorporating nonlocal linear electron–phonon coupling, *J. Chem. Phys.* **83**, 1843 (1985).
- [24] R. Silbey and R. W. Munn, General theory of electronic transport in molecular crystals. i. local linear electron-phonon coupling, *J. Chem. Phys.* **72**, 2763 (1980).
- [25] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, An updated version of wannier90: A tool for obtaining maximally-localised Wannier functions, *Comput. Phys. Commun.* **185**, 2309 (2014).
- [26] P. Giannozzi *et al.*, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, *J. Phys. Condens. Matter* **21**, 395502 (2009).
- [27] N. Troullier and J. L. Martins, Efficient pseudopotentials for plane-wave calculations, *Phys. Rev. B* **43**, 1993 (1991).
- [28] D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, *Phys. Rev. B* **88**, 085117 (2013).
- [29] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [30] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory, *Rev. Mod. Phys.* **73**, 515 (2001).
- [31] B. K. Chang, J.-J. Zhou, N.-E. Lee, and M. Bernardi, Intermediate polaronic charge transport in organic crystals from a many-body first-principles approach, *arXiv* 2106.09810 (2022).
- [32] T. G. Castner and W. K anzig, The electronic structure of V-centers, *J. Phys. Chem. Solids* **3**, 178 (1957).
- [33] J. H. Fetherolf, D. Gole z, and T. C. Berkelbach, A unification of the holstein polaron and dynamic disorder pictures of charge transport in organic crystals, *Phys. Rev. X* **10**, 021062 (2020).
- [34] Note that similar to above we factored out the phonon energy and defined the  $e$ -ph matrix elements as  $g_{jiv}(\mathbf{k}, \mathbf{q}) = \tilde{g}_{jiv}(\mathbf{k}, \mathbf{q})/(\hbar\omega_{\mathbf{Q}})$ , where  $\tilde{g}_{jiv}(\mathbf{k}, \mathbf{q})$  are the usual  $e$ -ph matrix elements in the Bloch basis [20].
- [35] T. Shi, E. Demler, and J. Ignacio Cirac, Variational study of fermionic and bosonic systems with non-Gaussian states: Theory and applications, *Ann. Phys.* **390**, 245 (2018).