Linear-scaling DFT+U with full local orbital optimization

David D. O'Regan,[1,2,*] Nicholas D. M. Hine,[1,3] Mike C. Payne,[1] and Arash A. Mostofi[3]

[1] *Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

[2] *Theory and Simulation of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

[3] *The Thomas Young Centre and the Department of Materials, Imperial College London, London SW7 2AZ, United Kingdom*

(Dated: 13 February 2012)

We present an approach to the DFT+U method (Density Functional Theory + Hubbard model) within which the computational effort for calculation of ground state energies and forces scales linearly with system size. We employ a formulation of the Hubbard model using nonorthogonal projector functions to define the localized subspaces, and apply it to a local-orbital DFT method including *in situ* orbital optimization. The resulting approach thus combines linear-scaling and systematic variational convergence. We demonstrate the scaling of the method by applying it to nickel oxide nano-clusters with sizes exceeding 7,000 atoms.

PACS numbers: 71.15.Mb, 71.15.Ap, 31.15.aq

I. INTRODUCTION

The success of the Kohn-Sham formulation of density functional theory (DFT)[1,2] is largely owed to its capability of accurately and reliably reproducing the ground-state properties of quantum-mechanical systems. Two factors that limit the applicability of DFT are the computational expense of treating large systems, and the difficulties encountered in simulating so-called "strongly-correlated" systems. The realistic study of many technologically and biologically important structures requires the explicit treatment of very large system sizes, yet the asymptotic scaling of conventional DFT algorithms is cubic in the number of atoms, so that the feasible size limit for routine calculations is typically around 1,000 atoms even on powerful high-performance computing architectures. Only by using DFT codes for which the effort increases linearly with system size, recently reviewed in Ref. 3 and examples of which include ONETEP[4–6], OPENMX[7,8], and CONQUEST[9,10], we may routinely bring first-principles simulation to bear on pertinent technological, environmental and medical problems. Furthermore, for many functional materials, most typically comprising open-shell first-row transition metal or lanthanoid ions, DFT with local or semi-local functionals performs very poorly, failing to obtain qualitative agreement with experimental observations in the most severe cases. Many methods have been developed to overcome this deficiency, and here we focus on the DFT+U[11] technique due to its widespread adoption and its amenability to linear-scaling implementation.

In this article, we present a computational methodology to tackle the obstacles of large system size and strong correlation effects simultaneously. Working in the framework of a linear-scaling implementation of DFT in order to tackle the issue of system size, we fully detail a DFT+U implementation, including self-consistent total-energies and forces, and demonstrate computational scaling tests on a strongly interacting oxide system of over 7,000 atoms. Previous linear-scaling or otherwise large-scale implementations of DFT+U, examples including Refs. 12–14, have relied on a basis of fixed user-defined or numerically pre-solved atomic orbitals. A noteworthy advantage of our approach is that we allow for the optimization of the orbitals representing the Kohn-Sham density-matrix *in situ*, that is during the process of total-energy minimization, with respect to an underlying, systematic plane-wave basis[4,15]. In this manner, we move beyond the fixed-orbital approximation to linear-scaling DFT and DFT+U. Using this approach, truly first-principles simulations may be carried out on systems comprising both strong electronic interactions and large spatial disorder, examples including layered transition-metal and lanthanoid oxide structures, catalytic surfaces, molecular magnets and organometallic biomolecules.

The article is organized as follows. We describe the DFT+U technique and its generalization to the nonorthogonal case in Section II, after which we introduce linear-scaling DFT and define the notation and sparse matrix algebra for linear-scaling DFT+U in Section III. Minimization of the DFT+U total-energy with respect to the density-matrix is detailed in Section IV. The method is applied to nickel oxide clusters exceeding 7,000 atoms in Section V, and linear-scaling performance is demonstrated. Following some concluding remarks, we detail the method used to preserve the density-matrix idempotency and normalization in Appendices A and B, and to compute the DFT+U ionic forces in Appendix C.

II. DFT+U METHOD FOR NONORTHOGONAL PROJECTORS

The use of approximations such as the local spin-density approximation (LSDA)[16], for the exchange-correlation (XC) functional in Kohn-Sham DFT, is appropriate and highly successful in systems where the magnitude of each electron's kinetic energy t is large

compared with the Coulomb interaction U acting on it. In such systems, usually comprising elements whose $3d$ or $4f$ atomic-like states are either completely empty or filled, the LSDA typically provides a good qualitative description of both the ground-state density and the insulating gap. In strongly correlated systems such as Mott-Hubbard insulators[17], however, these states are localized, partially occupied, and do not fall in the regime of $U \ll t$. In such cases, the LSDA may thus perform very poorly unless it is corrected. The DFT+U method[11] reintroduces the explicit Coulomb interaction terms, and thus the appropriate derivative discontinuity with respect to electronic occupation number, to the approximate XC functional.

In the DFT + Hubbard U method (DFT+U)[11], a number of spatially localized subspaces, sites labeled I, wherein the $U \ll t$ regime is not expected to hold, are selected for supplementation with explicit Coulomb correlations beyond the LSDA level, retaining the bare, inexpensive XC functional for the remainder of the system. The strongly interacting subspaces are spanned by sets of localized orbitals, termed the Hubbard projectors $\{\varphi_m^{(I)}\}$. The selection of Hubbard projectors is a topic of interest in itself and possible choices include localized Wannier functions built from the Kohn-Sham eigenfunctions according to maximal localization[18,19], energy downfolding[20] or maximal Coulomb repulsion[21] criteria, or indeed a total-energy minimization criterion in combination with a self-consistency scheme, as we have proposed in a related article, Ref. 22. In the description of our linear-scaling method, here we assume only that the projectors are confined to a spatial region, real-valued, and expressed in the same underlying, systematic basis as the orbitals representing the Kohn-Sham density-matrix (in the present case a truncated set of plane waves).

In the tensorial representation, developed in order to maintain the tensorial invariance of subspace occupancies, moments, ionic forces and the total-energy[23], localized Hubbard projector duals are defined by[24]

$$O^{(I)}_{mm'} = \langle \varphi^{(I)}_m | \varphi^{(I)}_{m'} \rangle; \quad |\varphi^{(I)m}\rangle = |\varphi^{(I)}_{m'}\rangle O^{(I)m'm}, \qquad (1)$$

where, by definition, $O^{(I)}_{mm''} O^{(I)m''m'} = \delta_m{}^{m'}$, such that an individual metric tensor $O^{(I)}$ is generated and used for each subspace. The occupancy matrix is then most conveniently expressed as a mixed tensor (specifically a tensor with one contravariant index and one covariant index), following Refs. 23 and 25, so that its trace is a tensorial invariant, as per

$$n^{(I)(\sigma)m}{}_{m'} = \langle \varphi^{(I)m} | \hat{\rho}^{(\sigma)} | \varphi^{(I)}_{m'} \rangle. \qquad (2)$$

Here, $\hat{\rho}^{(\sigma)}$ is the single-particle density-matrix for electrons of spin σ, formally defined by

$$\hat{\rho}^{(\sigma)} = \sum_i |\psi_i^{(\sigma)}\rangle f_i^{(\sigma)} \langle \psi_i^{(\sigma)}|, \qquad (3)$$

where $f_i^{(\sigma)}$ is the occupancy of the Kohn-Sham orbital $|\psi_i^{(\sigma)}\rangle$. Using this definition, we can cast the rotationally-invariant DFT+U functional of Refs. 26 and 27 into a more general, tensorially invariant form – that is invariant under arbitrary linear combinations of the Hubbard projectors for a given site, following Ref. 23. Specifically, we use a tensorially invariant generalization of the widely-used, simplified DFT+U functional of Ref. 28, where the energy functional is given by $E_{DFT+U} = E_{DFT} + E_U$, with

$$E_U = \sum_{I\sigma} \frac{U^{(I)}}{2} \left(n^{(I)(\sigma)m}{}_{m} - n^{(I)(\sigma)m'}{}_{m} n^{(I)(\sigma)m}{}_{m'} \right), \qquad (4)$$

and $U^{(I)}$ is the screened subspace-averaged Coulomb repulsion. The DFT+U penalty functional approximately emulates the exact exchange-correlation functional by introducing a derivative discontinuity in the total-energy with respect to the occupancy matrix, in effect approximately enforcing the unrestricted Hartree-Fock approximation within the subspaces I.

III. LINEAR-SCALING DFT+U

We now describe the steps necessary to perform DFT+U calculations with linear-scaling expense. We have previously demonstrated the features of projector self-consistency[22] and tensorial invariance[23], in our implementation of the method in the ONETEP code[4,5], however the method described in this article is applicable to linear-scaling DFT methods[3] generally. The method is rigorously general to the case of nonorthogonality of both the local orbitals and the Hubbard projectors.

A. Framework and notation

Linear-scaling DFT revolves around the optimization not of the eigenstates of the Kohn-Sham Hamiltonian, but of the density-matrix of Eqn. 3 expressed in terms of a set of nonorthogonal local orbitals, $\{|\phi_\alpha\rangle\}$ (known in the ONETEP code as Nonorthogonal Generalized Wannier Functions, NGWFs[15]), that is

$$\hat{\rho}^{(\sigma)} = |\phi_\alpha\rangle K^{(\sigma)\alpha\beta} \langle \phi_\beta|, \qquad (5)$$

where the tensor K is known as the *density kernel* and is generally non-diagonal. The exponential spatial localization of the density matrix for insulating materials[29],

$$\rho^{(\sigma)}(\mathbf{r}, \mathbf{r}') = \langle \mathbf{r} | \hat{\rho}^{(\sigma)} | \mathbf{r}' \rangle \sim \exp\left(-\gamma |\mathbf{r} - \mathbf{r}'|\right), \qquad (6)$$

must be exploited to achieve linear-scaling, by strictly limiting the spatial extent of the local orbitals and truncating the density kernel to an appropriate length-scale.

The contravariant duals of the local orbitals are denoted by $\{|\phi^\alpha\rangle\}$, and the contravariant metric on the

orbitals is the inverse of the covariant metric S, so that

$$S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle, \quad S^{\alpha\beta} = \langle\phi^\alpha|\phi^\beta\rangle = \left(S^{-1}\right)^{\alpha\beta}. \tag{7}$$

We emphasize that the metric S on the local orbitals, generally a large matrix containing information about the entire system, and the individual metric $O^{(I)}$ on each DFT+U subspace, a small matrix (usually 5×5 for $3d$-type subspaces) and a localized quantity, are distinct even in cases where the Hubbard projectors are selected as a proper subset of the local orbitals.

The occupation matrix for each subspace is expressed in terms of local orbital matrix elements by inserting the expansion of the density-matrix, Eq. 5, into the natural occupancy representation of Eq. 2. Making use of the transformation rules for Hubbard projectors given by Eq. 1, we find that

$$n^{(I)(\sigma)m}_{\quad\quad m'} = O^{(I)mm''}\langle\varphi^{(I)}_{m''}|\phi_\alpha\rangle K^{(\sigma)\alpha\beta}\langle\phi_\beta|\varphi^{(I)}_{m'}\rangle. \tag{8}$$

We denote the overlap between Hubbard projectors $\{|\varphi^{(I)}_m\rangle\}$ and local orbitals $\{|\phi_\alpha\rangle\}$ by

$$V^{(I)}_{\beta m} = \langle\phi_\beta|\varphi^{(I)}_m\rangle; \quad W^{(I)}_{m\alpha} = V^{(I)\dagger}_{\alpha m} = \langle\varphi^{(I)}_m|\phi_\alpha\rangle, \tag{9}$$

which may be very sparse matrices, particularly for a low density of subspaces. The DFT+U correction to the total-energy, given by Eqn. 4, is then computed with linear-scaling cost using the sparse matrix trace

$$E_U = \sum_{I,\sigma}\frac{U^{(I)}}{2}\mathrm{Tr}\big[OWKV\big(1-OWKV\big)\big]^{(I)(\sigma)}. \tag{10}$$

We have assumed throughout, for notational clarity, that identical Hubbard projectors are used for each spin channel, although the generalization to σ-dependent Hubbard projectors, and thus σ-dependent O, V, and W matrices, is straightforward.

B. Efficient use of matrix sparsity

The DFT+U functional of Eqn. 10, does not depend on the inter-site occupancy matrices generated using Hubbard projectors for different subspaces, although the generalization to inter-site occupancies, DFT+U+V, has been introduced in Ref. 30. In DFT+U, these non-local occupancies do not contribute to E_U and thus should not be computed unnecessarily. On the other hand, it is undesirable from the point of view of both ease of implementation and computational efficiency to explicitly store separate $V^{(I)}$, $W^{(I)}$ and $O^{(I)}$ matrices for each site, thereby necessitating individual matrix products for each site before explicit summation in, for example, Eqn. 10.

Our solution is to embed these small transformation matrices into large, though very sparse, V, W and O matrices for the entire system, where they then fit seamlessly into the hierarchical, parallelized, sparse algebra routines found in a contemporary linear-scaling DFT code[5,6]. The overlap O matrix is block-diagonal in either its covariant or contravariant form, the dimension of each block being the number of projectors spanning the subspace on the site in question, typically 5(7) for a subspace of $3d(4f)$ orbital symmetry. The Hubbard interaction parameters are also placed into a sparse matrix U for the entire system, of the same sparsity of O, although, in practice, diagonal in the simplest case of a scalar parameter on each site. The V matrix has the row sparsity of the orbital overlap matrix S, depending on the orbital cutoff radii, and the column sparsity of O; W is its transpose.

Let us take as an example the computation of the occupancy matrix given by Eqn. 8. We henceforth suppress the spin index for notational simplicity, on the understanding that the density kernel, its derivatives and derivatives with respect to it are generally spin-dependent. Working from left to right, temporarily placing a site index before each projector index to clarify to which subspace it belongs, we first consider the product

$$\begin{aligned}(OW)^{(I)m'}_{\quad\quad\beta} &= \sum_J O^{(I)m'(J)m''}W_{(J)m''\beta} \\ &= O^{(I)m'(I)m''}W_{(I)m''\beta},\end{aligned} \tag{11}$$

which is a matrix with the same sparsity pattern as W due to the block-sparsity of the O matrix. Next, taking the product with the density kernel on the right, as per

$$(OWK)^{(I)m'\alpha} = (OW)^{(I)m'}_{\quad\quad\beta}K^{\beta\alpha}, \tag{12}$$

we see that this matrix has the sparsity of WK, dense in the row index when no density-kernel truncation is applied. When kernel truncation is enforced, however, the number of values which α can take is reduced and the effort needed for the sum over β is diminished.

On the final step, where we compute

$$n^{(I)m'}_{\quad\quad(J)m} = (OWK)^{(I)m'\alpha}V_{\alpha(J)m}, \tag{13}$$

we accumulate extraneous information on the inter-subspace non-locality of the density-matrix. Were we to compute this matrix in full and then consider its square, we would find that

$$\sum_K n^{(I)m'}_{\quad\quad(K)m}n^{(K)m}_{\quad\quad(I)m'} \neq n^{(I)m'}_{\quad\quad(I)m}n^{(I)m}_{\quad\quad(I)m'}, \tag{14}$$

the former being generated in the full matrix product, while only the latter is required in Eqn. 4. This problem is resolved by always truncating the occupancy matrix

$$n^{(I)m'}_{\quad\quad(I)m} = (OWK)^{(I)m'\alpha}V_{\alpha(I)m} \tag{15}$$

to the block-diagonal sparsity pattern as O *in advance* of computing such products, eliminating the off-site occupancies. In practice, the unnecessary elements are never actually computed, and no wasted effort is incurred, since the sparse algebra system computes only elements in the sparsity pattern of the product matrix[6].

Matrix sparsity thus plays an important role in the construction of our linear-scaling DFT+U method, as it permits DFT+U calculations involving a large number of subspaces to be carried out efficiently. We hereafter suppress the site index, both to clarify the notation and to reflect the fact that the matrix operations are implemented *in practice* in terms of calls to sparse algebra subroutines which take the matrices V, W, O and U as arguments, and not their site-indexed counterparts.

IV. OPTIMIZATION OF THE DENSITY-MATRIX

In order to minimize the total-energy with respect to the density-matrix with linear-scaling cost, while affording it the variational freedom of a systematically improvable basis, it is performed in two nested conjugate gradients minimization loops. We first describe the inner loop, in Section IV A, a methodology common to many contemporary linear-scaling codes, where the energy is minimized with respect to the density kernel for a fixed set of local orbitals.

In the outer loop, the local orbitals $\{|\phi_\alpha\rangle\}$ which span the Hilbert space available to the density-matrix are optimized in order to minimize the total-energy. This technique is used to obviate the choice of a fixed local orbital basis, and numerous variations have been previously described[4,15,31–37]. The orbitals are assumed to be truncated to some region, in order to allow for linear-scaling cost, and refined with respect to the underlying basis, for a fixed density kernel $K^{\alpha\beta}$, in a manner which is furthermore compatible with a linear-scaling method for optimizing local orbitals for unoccupied states recently proposed in Ref. 38. We return to discuss the orbital optimization technique in Section IV B.

Recent success with the DFT+U method and its generalization to inter-site interactions, DFT+U+V[30], encourages us to think of DFT+U as a true method for first-principles energetics[39–42]. We have therefore implemented the DFT+U forces terms, as well as the total-energy minimization scheme, in the ONETEP code of which the capability of accurately optimizing geometries has been previously demonstrated[43]. We describe the required methodology in Appendix C.

A. Kernel Optimization

Minimization of the energy with respect to the density kernel is typically carried out, in practice, using a generalization of the Li-Nunes-Vanderbilt (LNV) technique[44], which simultaneously drives the density-matrix to idempotency while it evolves towards commutativity with its corresponding Kohn-Sham Hamiltonian. In this section however, for clarity, we assume that the energy may be straightforwardly minimized with respect to the density kernel. We return to the adaptations to the density kernel optimization method required for idempotency preservation in Appendix A.

The DFT+U contribution to the Hamiltonian is thus simply given by the derivative of the DFT+U energy term of Eqn. 4 with respect to an arbitrary density kernel, that is

$$H^U_{\beta\alpha} = \frac{U}{2}\Big\{\frac{\partial n^m_{\ m}}{\partial K^{\alpha\beta}} - \frac{\partial n^{m'}_{\ m}}{\partial K^{\alpha\beta}} n^m_{\ m'} - n^{m'}_{\ m}\frac{\partial n^m_{\ m'}}{\partial K^{\alpha\beta}}\Big\}. \quad (16)$$

In order to simplify this derivative, we begin by noting that the partial derivative of the occupation matrix with respect to the density kernel is given by

$$\begin{aligned}\frac{\partial n^{m'}_{\ m}}{\partial K^{\alpha\beta}} &= \frac{\partial}{\partial K^{\alpha\beta}}\left[O^{m'm''}W_{m''\gamma}K^{\gamma\delta}V_{\delta m}\right] \quad (17)\\ &= O^{m'm''}W_{m''\gamma}\delta^\gamma_\alpha\delta^\delta_\beta V_{\delta m}\\ &= O^{m'm''}W_{m''\alpha}V_{\beta m}.\end{aligned}$$

The trace of this derivative over Hubbard projectors gives the covariant, local-orbital representation of the sum of projections over subspaces, which is a Hermitian tensor by construction, given by

$$\frac{\partial n^m_{\ m}}{\partial K^{\alpha\beta}} = V_{\beta m}O^{mm''}W_{m''\alpha} \equiv P_{\beta\alpha}. \quad (18)$$

It follows that the products of the occupancy matrix and its derivative, each always computed in terms of the Hubbard projector indices, in practice, since there they have the block-diagonal sparsity pattern of O, are given by

$$\frac{\partial n^{m'}_{\ m''}}{\partial K^{\alpha\beta}} n^{m''}_{\ m} = (OW)^{m'}_{\ \alpha}(PKV)_{\beta m} \quad \text{and} \quad (19)$$

$$n^{m'}_{\ m''}\frac{\partial n^{m''}_{\ m}}{\partial K^{\alpha\beta}} = (OWKP)^{m'}_{\ \alpha}V_{\beta m'}. \quad (20)$$

As a result, the DFT+U term in the covariant Hamiltonian, denoted by H^U, may be succinctly expressed as

$$H^U_{\beta\alpha} = \frac{\partial E_U}{\partial K^{\alpha\beta}} = \frac{U}{2}(P - 2PKP)_{\beta\alpha}. \quad (21)$$

The DFT+U contribution to the total-energy is efficiently computed, correspondingly, using the trace

$$E_U = \frac{U}{2}(PK - PKPK)_\alpha^{\ \alpha}. \quad (22)$$

The DFT+U Hamiltonian and total-energy terms are added to their uncorrected DFT counterparts, giving $H_{\alpha\beta} = H^{DFT}_{\alpha\beta} + H^U_{\alpha\beta}$ and $E = E_{DFT} + E_U$, respectively, and similarly for the independent-particle, or "band-structure" energy $E^{IP} = E^{IP}_{DFT} + E^{IP}_U$, where $E_{DFT} \neq E^{IP}_{DFT} = H^{DFT}_{\alpha\beta}K^{\beta\alpha}$ and $E_U \neq E^{IP}_U = H^U_{\alpha\beta}K^{\beta\alpha}$.

For a refinement of the auxiliary density kernel $K^{\alpha\beta}$, any update to it must also be a contravariantly transforming tensor, as noted in Refs. 25 and 45. In order to provide such a search direction, it is necessary that we pre- and post-multiply the covariant gradient of Eqn. 21

with the contravariant metric tensor on the orbitals, that is their inverse overlap matrix evaluated at the point at which the gradient itself is computed, to give

$$G^{\alpha\beta} = \left(S^{-1}\right)^{\alpha\gamma} H_{\gamma\delta} \left(S^{-1}\right)^{\delta\beta}. \tag{23}$$

The inner product of two second-order tensors, X and Y, is defined with respect to the metric S on the local orbitals, so that

$$\langle X|Y\rangle_S = X^{\alpha\beta}Y_{\beta\alpha} = X^{\alpha\beta}S_{\beta\gamma}Y^{\gamma\delta}S_{\delta\alpha}. \tag{24}$$

This allows us to define the search direction norm $\|G\|_S = \langle G|G\rangle_S^{1/2}$, and the conjugacy condition $\langle G_{i+1}|G_i\rangle_S = 0$, and thus to minimize the total-energy by iteratively updating the density kernel according to

$$K_{i+1}^{\alpha\beta} \to K_i^{\alpha\beta} + \lambda_{i+1} G_{i+1}^{\alpha\beta}, \tag{25}$$

where the optimal step lengths $\{\lambda_i\}$ are computed using an appropriate non-linear conjugate gradients algorithm.

B. Orbital optimization

We now consider the optimization of the local orbitals, specifically the DFT+U contribution to total-energy variations with respect to the expansion coefficients of the orbitals in the underlying variational basis. We again assume that the energy may be directly minimized in this section, for simplicity, returning to the alterations necessary for idempotency preservation under local-orbital optimization in Appendix B. This procedure occurs in the outer of the two energy minimization loops in the ONETEP code used here, however the results of this section apply to any technique which optimizes its representation functions for minimal energy, such as those described in Refs. 4, 15, 31–37.

It is clear from Eqn. 22 that the derivative of the total-energy with respect to the expansion of the local orbitals on the grid (or, in general, the basis) may explicitly depend only on the matrix elements of the projection P, defined in Eqn. 18, so that

$$\frac{\partial E_U}{\partial \phi_\alpha(\mathbf{r})} = \frac{\partial E_U}{\partial P_{\beta\gamma}} \frac{\partial P_{\beta\gamma}}{\partial \phi_\alpha(\mathbf{r})}. \tag{26}$$

Since this derivative involves the expansion of the Hubbard projections on the grid, it incurs changes beyond simple linear mixing of the orbitals. Evaluating this, we first take the action of the DFT+U Hamiltonian contribution on the subspace projections, that is

$$\begin{aligned}\frac{\partial E_U}{\partial P_{\beta\gamma}} &= \frac{U}{2}\frac{\partial}{\partial P_{\beta\gamma}}\left[(PK - PKPK)_\alpha{}^\alpha\right] \\ &= K^{\gamma\delta}H^U_{\delta\epsilon}P^{\epsilon\beta}; \quad P^{\alpha\beta} = \left(P^{-1}\right)^{\alpha\beta}.\end{aligned} \tag{27}$$

The Hubbard projection operators depend explicitly on the covariant orbitals which overlap with their corresponding Hubbard projectors (and Hubbard projector duals) and this dependence may be expressed as

$$\begin{aligned}\frac{\partial P_{\beta\gamma}}{\partial \phi_\alpha(\mathbf{r})} &= \frac{\partial}{\partial \phi_\alpha(\mathbf{r})}\left[V_{\beta m}O^{mm'}W_{m'\gamma}\right] \\ &= \delta^\alpha_\beta \varphi_m(\mathbf{r})\, O^{mm'}W_{m'\gamma} + V_{\beta m}O^{mm'}\varphi_{m'}(\mathbf{r})\,\delta^\alpha_\gamma.\end{aligned} \tag{28}$$

Combining this result with Eqs. 26 and 27, we may compute the DFT+U term in the local-orbital gradient,

$$\begin{aligned}\frac{\partial E_U}{\partial \phi_\alpha(\mathbf{r})} &= 2K^{\alpha\delta}H^U_{\delta\epsilon}P^{\epsilon\beta}V_{\beta m}O^{mm'}\varphi_{m'}(\mathbf{r}) \\ &= 2K^{\alpha\delta}V_{\delta m''}H^{m''m'}\varphi_{m'}(\mathbf{r}),\end{aligned} \tag{29}$$

where formally, though never explicitly in practice,

$$H_U^{mm'} = O^{mm''}W_{m''\alpha}S^{\alpha\beta}H^U_{\beta\gamma}S^{\gamma\delta}V_{\delta m'''}O^{m'''m'}. \tag{30}$$

Due to the subspace-localized nature of the DFT+U correction in the tensorial representation[23], only those local orbitals $|\phi_\delta\rangle$ in Eqn. 29 which explicitly overlap with the Hubbard projectors $|\varphi_{m'}\rangle$ contribute and thus require summation over.

Since, crucially, we require a covariantly transforming orbital update in order to improve upon those functions, to preserve their tensorial character, the above contravariant gradient must be multiplied with the covariant metric tensor in order to provide the necessary covariant DFT+U orbital search direction term, given by

$$g^U_\alpha(\mathbf{r}) = 2S_{\alpha\beta}K^{\beta\delta}V_{\delta m''}H^{m''m'}\varphi_{m'}(\mathbf{r}). \tag{31}$$

This contribution may then be combined with the uncorrected DFT search direction, giving the total $|g_\alpha\rangle = |g_\alpha^{DFT}\rangle + |g_\alpha^U\rangle$. The inner product and norm of first-order tensors, or orbitals, $|x\rangle$ and $|y\rangle$, are defined such that

$$\langle x|y\rangle_S = \langle x^\alpha|y_\alpha\rangle = \langle x_\alpha|S^{\alpha\beta}|y_\beta\rangle; \tag{32}$$

$$\|x\|_S = \langle x|x\rangle_S^{1/2}, \tag{33}$$

and, using these definitions, the total-energy may be minimized by iteratively updating the orbitals, where the $\{\mu^i\}$ are computed using one of many available non-linear conjugate gradients algorithms, according to

$$|\phi_\alpha^{i+1}\rangle \to |\phi_\alpha^i\rangle + \mu^{i+1}|g_\alpha^{i+1}\rangle. \tag{34}$$

At each such orbital update step, in practice, we carry out a complete re-optimization of the density kernel, according to the procedure of Sub-section IV A.

V. APPLICATION TO NICKEL OXIDE NANO-CLUSTERS

We performed scaling tests on NiO nano-clusters of varying size, comparing the computational effort required for DFT+U and uncorrected DFT calculations. An antiferromagnetic insulator, NiO is a well known example where LSDA-type approximations[16] fail to qualitatively

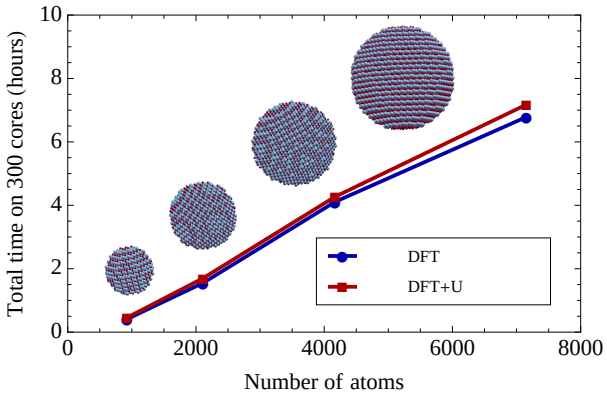

FIG. 1. (Color online) Scaling of the energy minimization algorithm for NiO nanoclusters of increasing size. Timings are for three density kernel optimization steps and one orbital optimization step, comparing DFT and DFT+U calculations. Simulations were performed on 300 Intel Westmere 2.67 Ghz cores connected using quad data rate Infiniband.

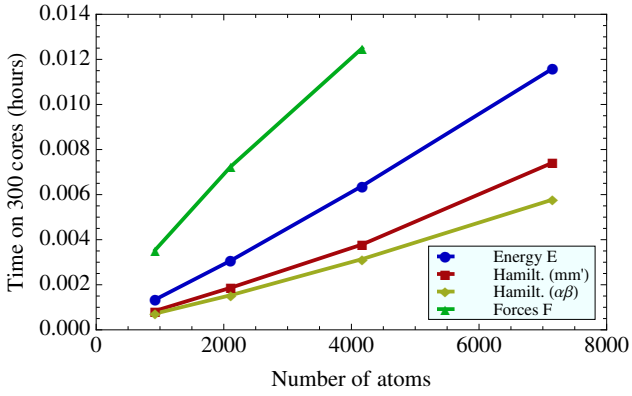

FIG. 2. (Color online) Computational time spent in subroutines associated with the DFT+U functionality in the tests shown in Fig. 1. Specifically timings shown are for computing the DFT+U energy of Eqn. 4, the Hamiltonian matrix of Eqn. 21 in its Hubbard projector and local orbital representations, and the ionic forces given by Eqn. C6.

reproduce the correct insulating gap and local magnetic moments (of between 1.64 μ_B and 1.9 μ_B[28]) due to a poor description of $3d$ orbital localization. The gap, of approximately 4 eV, is of predominantly Mott-Hubbard type since it persists above the Néel temperature[46], albeit with a significant charge-transfer component[47]. It is thus successfully recovered by a number of methods, which either include many-body Coulomb correlation effects explicitly, such as LDA+DMFT[48], or introduce an appropriate derivative discontinuity with respect to occupancy at the single-particle level, examples including unrestricted Hartree-Fock[46] and the self-interaction corrected local density approximation[49]. The correct description of the physics of NiO was an early success for DFT+U, the method of interest here, and this has been repeated using numerous functional forms[11,28,30,50–52].

The method described in this article has previously been successfully applied to bulk NiO[23]. For a demonstration of computational scaling, we have chosen spherical nano-clusters of NiO with even numbers of nickel ions, so that an open-shell singlet multiplicity, analogous to the bulk antiferromagnetic ground state, could be tentatively assumed. We may expect that a transition to a ferrimagnetic or ferromagnetic state occurs below some critical cluster size, as it has been predicted for very small iron oxide clusters of interest for data-storage technology[53,54].

Run-time parameters included a 500 eV equivalent plane-wave cutoff energy, a spin polarized density kernel cutoff at 25 a_0, the LSDA exchange-correlation functional[16], nine local orbitals (NGWFs) for each nickel ion and four each for oxygen, all with 7.5 a_0 cutoff radii, and norm-conserving pseudopotentials[55]. Atomic Hubbard projectors of hydrogenic form were used. Since calculations on nano-clusters of varying sizes are expected to exhibit differing convergence behavior, the energy minimization algorithm was simply run for a fixed number of iterations. One orbital optimization step and three density kernel steps, with three penalty-functional idempotency corrections iterations at each of the latter, were allowed. Orbital overlap matrix inversion was carried out using a sparse matrix implementation of Hotelling's algorithm[56] and a cubic supercell of length three times the diameter of each nano-cluster was used, up to a maximum supercell length of approximately 300 a_0.

A. Scaling of computational effort for DFT+U

Algorithmic timing data for ONETEP energy minimization of NiO nano-clusters, containing up to $7,153$ atoms across 300 Intel Westmere 2.67 Ghz cores, is shown in Fig. 1. A reasonable linear fit was obtained for the timing; with a slightly negative fitted time intercept at $450-500$ atoms indicating a very efficient initialization of the pre-requisite data in these calculations. The NiO nano-clusters in question do not represent a favorable case for the DFT+U method, since approximately half of the ions host correlated subspaces. Nonetheless, we observed a very small increase in computational time when the DFT+U functionality was invoked, at approximately $5-6\%$, and preservation of linear-scaling performance.

Timings for generating the DFT+U Hamiltonian and its contribution to the total-energy and forces, for those calculations which fell within memory resources, are depicted in Fig. 2. This indicates that no direct DFT+U functionality appreciably deviates from linear-scaling behavior. We note, in particular, that the total time spent in these DFT+U specific subroutines makes up only a small fraction of the increase in cost incurred by DFT+U, at less than 1% of the total computational time.

In order to understand where the dominant contribution to the DFT+U cost originates, since it is not di-

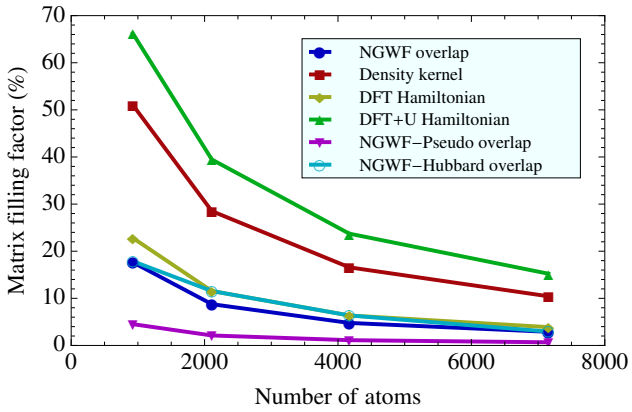

FIG. 3. (Color online) Nano-cluster size dependence of filling factors of principal matrices affecting the computational cost of linear-scaling DFT and DFT+U. These, namely, are the overlap between orbitals, the density kernel, the Hamiltonian matrices conventional for DFT and for DFT+U, and the overlap matrices between the local orbitals and the non-local pseudopotential projectors and Hubbard projectors.

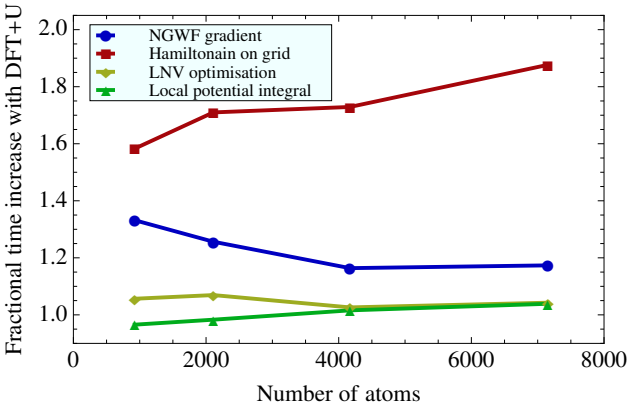

FIG. 4. (Color online) Fractional change in time expended on energy-minimization operations when the DFT+U functionality is activated. Namely, these are calculation of the local orbital (NGWF) gradient, expansion of the Hamiltonian in the basis, kernel optimization using the LNV method, and calculation of the matrix elements of the local potential.

rectly in the DFT+U subroutines themselves, we direct the reader to Fig. 3, where the system dependent sparse matrix filling is quantified. In a conventional DFT calculation, the sparsity of the Hamiltonian matrix is dominated by the orbital representation of the non-local pseudopotential, proportional to the product of the overlap matrix between the orbitals and the non-local pseudopotential projectors with its transpose. In essence, pairs of orbitals which overlap with a common non-local projector contribute to the energy, and the same holds for the Hubbard projectors of DFT+U. While non-local pseudopotential projectors tend to have cutoff radii not in excess of approximately 2 a_0, Hubbard projectors of $3d$ symmetry may require greater cutoff radii (for hydrogenic orbitals of effective nuclear charge in the typical range for transition-metals, $Z = \{8, 9, 10, 11\}$, the normalization spillage at 2 a_0 is $\{9.3, 4.6, 2.1, 0.9\}\%$). In our calculations the projector cutoff radius is set equal to that of the local orbitals, 7.5 a_0.

This increased Hamiltonian filling has consequences additionally for the calculation of energy gradients, as indicated in Fig. 4, which shows the fractional change in time spent in carrying out certain energy minimization operations. Most notably, it takes close to twice as much effort to calculate its expansion on the *psinc* grid due to the inclusion of Hubbard projectors in the Hamiltonian. The dominant part of the overall expense of the calculations is from operations on large matrices, however, so that grid expansion of the Hamiltonian is not significant for large systems. The incurred increase in the filling of the Hamiltonian matrix in DFT+U over DFT, and also in the expense of computing its products with quantities such as the density kernel and its expansion on the underlying, systematic plane-wave basis, is thus largely responsible for the observed, albeit moderate, increase in computational expense indirectly introduced by DFT+U, from 1% to $5-6$%.

VI. CONCLUDING REMARKS

We have detailed a linear-scaling implementation of the DFT+U method for treating strongly-correlated systems from first-principles. The formalism is generally appropriate to methods which minimize the energy with respect to the single-particle density-matrix, and allows for the optimization of both nonorthogonal Hubbard projectors[22] and ionic positions.

The preservation of linear-scaling performance on metal-oxide nano-clusters in excess of $7,000$ atoms is demonstrated. For systems of this type, with a high density of correlated sites, the increase in computational prefactor remains rather modest. The DFT+U functionality, furthermore, incurs negligible cost in large systems comprising only a small number of Hubbard subspaces.

Ground state calculations employing our method have previously been demonstrated on both bulk and molecular strongly interacting systems[22,23], with further examples on large-scale systems such as dilute magnetic semiconductor (Ga,Mn)As[57] and disordered VO_2[58], using an extension of the method to DFT+DMFT, forthcoming. Further examples of candidate systems include organometallic molecules, such as metalloproteins and molecular magnets, where the method is particularly efficient for a low density of strongly interacting subspaces, and solids such as magnetic heterostructures, defective and doped oxides or catalytic interfaces with oxide surfaces. We envisage that the technique described may aid in bringing linear-scaling DFT to bear on more challenging systems than those to which it is has been typically applied to date.

ACKNOWLEDGMENTS

All figures are reproduced with permission from Ref. 59. We are grateful to Emilio Artacho and Adrian Ionescu for helpful discussions. D.D.O'R acknowledges the support of EPSRC and the National University of Ireland. N.D.M.H and A.A.M. acknowledge the support of EPSRC (Grant EP/G05567X/1), and A.A.M further acknowledges support from RCUK. M.C.P. acknowledges the support of EPSRC (Grants EP/G055904/1, EP/F032773/1). Calculations were performed on the Cambridge HPCS Darwin computer.

Appendix A: Preservation of density-matrix purity under kernel optimization

In the LNV method[44], the Kohn-Sham density kernel is related to the auxiliary density kernel via one iteration of the McWeeny purification transform, that is

$$K^{\alpha\beta} = \left(3LSL - 2LSLSL\right)^{\alpha\beta}. \tag{A1}$$

In our treatment of DFT+U, we go a step further and provide the more general expressions needed for the HSMP[60] variant of the LNV method, in which the density kernel $\tilde{K}$ is expressed as a purified and normalized auxiliary density kernel, explicitly

$$\begin{aligned}\tilde{K}^{\alpha\beta} &= \frac{N}{S_{\gamma\delta}K^{\delta\gamma}}K^{\alpha\beta} \\ &= \frac{N\left(3LSL - 2LSLSL\right)^{\alpha\beta}}{S_{\gamma\delta}\left(3LSL - 2LSLSL\right)^{\delta\gamma}},\end{aligned} \tag{A2}$$

where N is the correct occupancy for the spin channel in question. The kernel renormalization introduces terms in the gradient proportional to an effective chemical potential, projecting out any first-order changes to the electron number, driving the density kernel $\tilde{K}$ towards both normalization and idempotency as the energy is minimized.

To locate the derivative of the DFT+U energy term with respect to the auxiliary density kernel, stressing that it is computed strictly using the purified and renormalized density kernel, we make use of the chain-rule for matrix derivatives to write

$$\frac{\partial E_U}{\partial L^{\alpha\beta}} = \frac{\partial E_U}{\partial K^{\gamma\delta}}\frac{\partial K^{\gamma\delta}}{\partial L^{\alpha\beta}}, \tag{A3}$$

where we carefully note that the partial derivative with respect to a doubly contravariant tensor is a doubly covariant tensor with indices once permuted to allow for the complex case. It may be readily shown, using Eqn. A1, that the latter term is given by

$$\begin{aligned}\frac{\partial K^{\gamma\delta}}{\partial L^{\alpha\beta}} &= 3\left(\delta^{\gamma}_{\alpha}S_{\beta\epsilon}L^{\epsilon\delta} + L^{\gamma\epsilon}S_{\epsilon\alpha}\delta^{\delta}_{\beta}\right) \\ &\quad - 2\begin{pmatrix}\delta^{\gamma}_{\alpha}S_{\beta\epsilon}L^{\epsilon\zeta}S_{\zeta\eta}L^{\eta\delta} \\ +L^{\gamma\epsilon}S_{\epsilon\alpha}S_{\beta\zeta}L^{\zeta\delta} + L^{\gamma\epsilon}S_{\epsilon\zeta}L^{\zeta\eta}S_{\eta\alpha}\delta^{\delta}_{\beta}\end{pmatrix}.\end{aligned} \tag{A4}$$

The derivative of the DFT+U energy with respect to the purified density kernel K may be broken into products of derivatives, and rearranged as follows

$$\begin{aligned}\frac{\partial E_U}{\partial K^{\gamma\delta}} &= \frac{\partial}{\partial \tilde{K}^{\zeta\epsilon}}\left[E_U\left(\tilde{K}\right)\right]\frac{\partial \tilde{K}^{\zeta\epsilon}}{\partial K^{\gamma\delta}} \\ &= H^U_{\epsilon\zeta}\frac{N}{\left(S_{\eta\theta}K^{\theta\eta}\right)}\left\{\delta^{\zeta}_{\gamma}\delta^{\epsilon}_{\delta} - \frac{K^{\zeta\epsilon}}{\left(S_{\iota\kappa}K^{\kappa\iota}\right)}S_{\delta\gamma}\right\}.\end{aligned} \tag{A5}$$

We may next write the gradient with respect to the density kernel in terms of a preconditioned contribution to the Hamiltonian, denoted by

$$\tilde{H}^U_{\delta\gamma} = H^U_{\delta\gamma} - \mu^U S_{\delta\gamma}; \quad \mu^U = \frac{H^U_{\epsilon\zeta}K^{\zeta\epsilon}}{\left(S_{\iota\kappa}K^{\kappa\iota}\right)}, \tag{A6}$$

where μ^U is identified as the DFT+U correction to the chemical potential, since

$$\begin{aligned}\frac{\partial E_U}{\partial K^{\gamma\delta}} &= \frac{N}{\left(S_{\eta\theta}K^{\theta\eta}\right)}\left\{H^U_{\delta\gamma} - \frac{H^U_{\epsilon\zeta}K^{\zeta\epsilon}}{\left(S_{\iota\kappa}K^{\kappa\iota}\right)}S_{\delta\gamma}\right\} \\ &= \frac{N}{\left(S_{\eta\theta}K^{\theta\eta}\right)}\tilde{H}^U_{\delta\gamma}.\end{aligned} \tag{A7}$$

It is worth noting that, just as the DFT+U independent-particle energy correction, $E^{IP}_U = H^U_{\alpha\beta}\tilde{K}^{\beta\alpha}$, does not equal the energy term E_U, so the energy correction entering into the computation of μ^U is not identical to the DFT+U correction to the total-energy.

The required DFT+U energy gradient is provided by the product of the preconditioned term in the Hamiltonian and the derivative of the density kernel with respect to its auxiliary counterpart, given by

$$\frac{\partial E_U}{\partial L^{\alpha\beta}} = \frac{N}{\left(S_{\iota\kappa}K^{\kappa\iota}\right)}\tilde{H}^U_{\delta\gamma}\frac{\partial K^{\gamma\delta}}{\partial L^{\alpha\beta}}. \tag{A8}$$

Finally, combining Eqs. A4 and A8, and recalling Eqn. 23, we arrive at the DFT+U contribution to the contravariant density kernel gradient,

$$\begin{aligned}G^{\alpha\beta}_U &= \left(S^{-1}\right)^{\alpha\gamma}\frac{\partial E_U}{\partial L^{\delta\gamma}}\left(S^{-1}\right)^{\delta\beta} \\ &= \frac{N}{S_{\gamma\delta}\left(3LSL - 2LSLSL\right)^{\delta\gamma}} \\ &\quad \times \begin{Bmatrix} 3\left(S^{-1}\tilde{H}L + L\tilde{H}S^{-1}\right) \\ -2L\tilde{H}L \\ -2\left(S^{-1}\tilde{H}LSL + LSL\tilde{H}S^{-1}\right)\end{Bmatrix}^{\alpha\beta}.\end{aligned} \tag{A9}$$

Appendix B: Preservation of density-matrix purity under orbital optimization

As in the case of the density kernel gradient, the orbital gradient is calculated using the purified and renormalized density kernel, and so it contains a preconditioning term

which drives the trace of the density-matrix to the correct occupancy of the system. The energy derivative with respect to covariant orbitals may be decomposed as

$$\frac{\partial E_U}{\partial \phi_\alpha(\mathbf{r})} = \frac{\partial E_U}{\partial \tilde{K}^{\beta\gamma}}\left(\frac{\partial \tilde{K}^{\beta\gamma}}{\partial K^{\delta\epsilon}}\frac{\partial K^{\delta\epsilon}}{\partial S_{\zeta\eta}} + \frac{\partial \tilde{K}^{\beta\gamma}}{\partial S_{\zeta\eta}}\right)\frac{\partial S_{\zeta\eta}}{\partial \phi_\alpha(\mathbf{r})} + \frac{\partial E_U}{\partial P_{\beta\gamma}}\frac{\partial P_{\beta\gamma}}{\partial \phi_\alpha(\mathbf{r})}. \tag{B1}$$

The terms contained in parentheses in Eqn. B1 may be evaluated, respectively yielding

$$\frac{\partial \tilde{K}^{\beta\gamma}}{\partial S_{\delta\epsilon}} = \frac{N}{(S_{\theta\iota}K^{\iota\theta})}\left\{-\frac{K^{\beta\gamma}}{(S_{\kappa\lambda}K^{\lambda\kappa})}K^{\epsilon\delta}\right\}, \tag{B2}$$

$$\frac{\partial \tilde{K}^{\beta\gamma}}{\partial K^{\delta\epsilon}} = \frac{N}{(S_{\theta\iota}K^{\iota\theta})}\left\{\delta^\beta_\delta\delta^\gamma_\epsilon - \frac{K^{\beta\gamma}}{(S_{\kappa\lambda}K^{\lambda\kappa})}S_{\epsilon\delta}\right\}, \quad \text{and}$$

$$\frac{\partial K^{\delta\epsilon}}{\partial S_{\zeta\eta}} = 3L^{\delta\zeta}L^{\eta\epsilon} - 2L^{\delta\zeta}(LSL)^{\eta\epsilon} - 2(LSL)^{\delta\zeta}L^{\eta\epsilon}.$$

The covariant metric explicitly depends only on the covariant orbitals, so that

$$\frac{\partial S_{\zeta\eta}}{\partial \phi_\alpha(\mathbf{r})} = \delta^\alpha_\zeta\phi_\eta(\mathbf{r}) + \delta^\alpha_\eta\phi_\zeta(\mathbf{r}). \tag{B3}$$

Contraction of the DFT+U term in the Hamiltonian and the terms in Eqn. B2 provides a tensor $\tilde{Q}$ which represents a contribution to the local orbital gradient purely due to mixing among the orbitals, given by

$$\tilde{Q}^{\eta\zeta} = H^U_{\gamma\beta}\left(\frac{\partial \tilde{K}^{\beta\gamma}}{\partial K^{\delta\epsilon}}\frac{\partial K^{\delta\epsilon}}{\partial S_{\zeta\eta}} + \frac{\partial \tilde{K}^{\beta\gamma}}{\partial S_{\zeta\eta}}\right) \tag{B4}$$

$$= \frac{N}{(S_{\theta\iota}K^{\iota\theta})}\left\{\begin{array}{c} 3L\tilde{H}L - 2L\tilde{H}LSL - 2LSL\tilde{H}L \\ -\mu^U K \end{array}\right\}^{\eta\zeta}.$$

To conclude, by combining Eqs. 29, B1, and B4, the contravariant gradient of the DFT+U energy with respect to the orbitals is given by

$$\frac{\partial E_U}{\partial \phi_\alpha(\mathbf{r})} = 2\left(\tilde{K}^{\alpha\zeta}V_{\zeta m''}\tilde{H}^{m''m'}\varphi_{m'} + \tilde{Q}^{\alpha\zeta}\phi_\zeta\right)(\mathbf{r}), \tag{B5}$$

which is then transformed to the required covariant form, in the same manner as per Eq. 31, giving

$$g^U_\alpha(\mathbf{r}) = 2S_{\alpha\beta}\tilde{K}^{\beta\zeta}V_{\zeta m''}\tilde{H}^{m''m'}\varphi_{m'}(\mathbf{r}) + 2S_{\alpha\beta}\tilde{Q}^{\beta\zeta}\phi_\zeta(\mathbf{r}). \tag{B6}$$

Appendix C: Ionic forces

We may assume that the ground-state density is located for a given ionic configuration before the forces are computed, so that the total-energy is variationally minimized with respect to both the orbital expansion coefficients and the matrix elements of the density kernel. The DFT+U correction then contributes to the ionic forces only via the spatial dependence of the Hubbard projection operators, that is for the ion labeled j,

$$\mathbf{F}^j_U = -\frac{\partial E_U}{\partial \mathbf{R}_j} = -\frac{\partial E_U}{\partial P_{\alpha\beta}}\frac{\partial P_{\alpha\beta}}{\partial \mathbf{R}_j}. \tag{C1}$$

The lattermost derivative may be expressed in terms of gradients of the covariant projectors and contravariant subspace metric tensors, specifically

$$\frac{\partial P_{\alpha\beta}}{\partial \mathbf{R}_j} = \sum_I \frac{\partial P_{\alpha\beta}}{\partial \varphi^{(I)}_m(\mathbf{r})}\frac{\partial \varphi^{(I)}_m(\mathbf{r})}{\partial \mathbf{R}_j} + \frac{\partial P_{\alpha\beta}}{\partial O^{mm'}}\frac{\partial O^{mm'}}{\partial \mathbf{R}_j}, \tag{C2}$$

where, in the tensorial representation[23],

$$\frac{\partial O^{mm'}}{\partial \mathbf{R}_j} = -O^{mm''}\frac{\partial O_{m''m'''}}{\partial \mathbf{R}_j}O^{m'''m'}, \tag{C3}$$

vanishes in the conventional case that the Hubbard projectors are rigidly translated with their host ions.

We note, next, that the Hubbard projectors are usually considered to be associated with one atomic site only, and so the subspace index I need only run over subspaces centered on ion j. We may thus suppress the summation symbol in Eqn. C1, for notational clarity, since the generalization to multiple subspaces per ion is straightforward. Denoting the spatial derivative of the Hubbard projectors by the three-component vector

$$\mathbf{X}^{(j)}_{\alpha m} = \sum_{I\cap j}\langle\phi_\alpha|\nabla|\varphi^{(I)}_m\rangle \tag{C4}$$

$$= \sum_{I\cap j}\int d\mathbf{r}\ \phi_\alpha(\mathbf{r})\left[\int d\mathbf{G}\ (-i\mathbf{G})\,e^{-i\mathbf{G}.\mathbf{r}}\varphi^{(I)}_m(\mathbf{G})\right],$$

the remaining terms in Eqn. C2 may be expressed as

$$\begin{aligned}\frac{\partial P_{\alpha\beta}}{\partial \mathbf{R}_j} &= \int\left(\frac{\partial}{\partial \varphi^{(j)}_m(\mathbf{r})}\left[V_{\alpha m'}O^{m'm''}W_{m''\beta}\right]\right) \\ &\quad\times\frac{\partial}{\partial \mathbf{R}}\sum_{I\cap j}\left[\int d\mathbf{G}\ e^{-i\mathbf{G}.\mathbf{r}}\varphi^{(I)}_m(\mathbf{G})\right]d\mathbf{r} \\ &= \int\left(\phi_\alpha O^{mm''}W_{m''\beta} + V_{\alpha m'}O^{m'm}\phi_\beta\right)(\mathbf{r}) \\ &\quad\times\sum_{I\cap j}\left[\int d\mathbf{G}\ (-i\mathbf{G})\,e^{-i\mathbf{G}.\mathbf{r}}\varphi^{(I)}_m(\mathbf{G})\right]d\mathbf{r} \\ &= \mathbf{X}^{(j)}_{\alpha m}O^{mm''}W_{m''\beta} + V_{\alpha m'}O^{m'm}\mathbf{X}^{(j)\dagger}_{m\beta}.\end{aligned} \tag{C5}$$

Combining the latter result with Eqs. 27 and C1, we conclude that the tensorially consistent DFT+U contribution to the ionic forces is succinctly given by the sparse matrix trace, noting the resemblance to Eqn. 29,

$$\mathbf{F}^j_U = -2\tilde{K}^{\beta\alpha}V_{\alpha m}H^{m'm}\mathbf{X}^{(j)\dagger}_{m\beta}. \tag{C6}$$

This may be used, for example, to perform DFT+U corrected ionic geometry optimization, molecular dynamics, or calculations of vibrational spectra on large systems.

* david.oregan@epfl.ch

[1] P. Hohenberg and W. Kohn, Phys. Rev., **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev., **140**, A1133 (1965).
[3] D. R. Bowler and T. Miyazaki, (2012), Rep. Prog. Phys. (in press), arXiv:1108.5976.
[4] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys., **122**, 084119 (2005).
[5] N. D. M. Hine, P. D. Haynes, A. A. Mostofi, C.-K. Skylaris, and M. C. Payne, Comp. Phys. Comms., **180**, 1041 (2009).
[6] N. D. M. Hine, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys., **133**, 114111 (2010).
[7] T. Ozaki and H. Kino, Phys. Rev. B, **72**, 045121 (2005).
[8] T. Ozaki, Phys. Rev. B, **82**, 075131 (2010).
[9] D. R. Bowler, R. Choudhury, M. J. Gillan, and T. Miyazaki, Phys. Status Solidi B, **243**, 989 (2006).
[10] D. R. Bowler and T. Miyazaki, J. Phys.: Condens. Matter, **22**, 074207 (2010).
[11] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B, **44**, 943 (1991); V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyżyk, and G. A. Sawatzky, *ibid.*, **48**, 16929 (1993).
[12] M. Wierzbowska, D. Sánchez-Portal, and S. Sanvito, Phys. Rev. B, **70**, 235209 (2004).
[13] M. J. Han, T. Ozaki, and J. Yu, Phys. Rev. B, **73**, 045110 (2006).
[14] S. Sanna, B. Hourahine, U. Gerstmann, and T. Frauenheim, Phys. Rev. B, **76**, 155128 (2007).
[15] C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez, and M. C. Payne, Phys. Rev. B, **66**, 035119 (2002).
[16] J. P. Perdew and A. Zunger, Phys. Rev. B, **23**, 5048 (1981).
[17] N. F. Mott, Proc. Phys. Soc., **A 62**, 416 (1949).
[18] W. Ku, H. Rosner, W. E. Pickett, and R. T. Scalettar, Phys. Rev. Lett., **89**, 167204 (2002).
[19] V. V. Mazurenko, S. L. Skornyakov, A. V. Kozhevnikov, F. Mila, and V. I. Anisimov, Phys. Rev. B, **75**, 224408 (2007).
[20] A. Yamasaki, M. Feldbacher, Y.-F. Yang, O. K. Andersen, and K. Held, Phys. Rev. Lett., **96**, 166401 (2006).
[21] T. Miyake and F. Aryasetiawan, Phys. Rev. B, **77**, 085122 (2008).
[22] D. D. O'Regan, N. D. M. Hine, M. C. Payne, and A. A. Mostofi, Phys. Rev. B, **82**, 081102 (2010).
[23] D. D. O'Regan, M. C. Payne, and A. A. Mostofi, Phys. Rev. B, **83**, 245124 (2011).
[24] We employ the Einstein convention, where pairs of identical indices are summed over unless in parentheses.
[25] E. Artacho and L. Miláns del Bosch, Phys. Rev. A, **43**, 5770 (1991).
[26] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys.: Condens. Matter, **9**, 767 (1997).
[27] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, Phys. Rev. B, **52**, R5467 (1995).
[28] M. Cococcioni and S. de Gironcoli, Phys. Rev. B, **71**, 035105 (2005).
[29] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, Phys. Rev. Lett., **98**, 046402 (2007).
[30] V. L. Campo Jr. and M. Cococcioni, J. Phys.: Condens. Matter, **22**, 055602 (2010).
[31] T. Ozaki, Phys. Rev. B, **67**, 155108 (2003).
[32] T. Ozaki and H. Kino, Phys. Rev. B, **69**, 195113 (2004).
[33] T. Ozaki and H. Kino, J. Chem. Phys., **121**, 10879 (2004).
[34] F. Mauri, G. Galli, and R. Car, Phys. Rev. B, **47**, 9973 (1993).
[35] F. Mauri and G. Galli, Phys. Rev. B, **50**, 4316 (1994).
[36] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B, **52**, 1640 (1995).
[37] P. Ordejón, D. A. Drabold, R. M. Martin, and M. P. Grumbach, Phys. Rev. B, **51**, 1456 (1995).
[38] L. E. Ratcliff, N. D. M. Hine, and P. D. Haynes, Phys. Rev. B, **84**, 165131 (2011).
[39] D. A. Scherlis, M. Cococcioni, P. Sit, and N. Marzari, J. Phys. Chem. B, **111**, 7384 (2007).
[40] H. J. Kulik and N. Marzari, J. Chem. Phys., **129**, 134314 (2008); J. Chem. Phys, **133**, 114103 (2010); J. Chem. Phys., **134**, 094103 (2011).
[41] H. Hsu, K. Umemoto, M. Cococcioni, and R. Wentzcovitch, Phys. Rev. B, **79**, 125124 (2009).
[42] B. Himmetoglu, R. M. Wentzcovitch, and M. Cococcioni, Phys. Rev. B, **84**, 115108 (2011).
[43] N. D. M. Hine, M. Robinson, P. D. Haynes, C.-K. Skylaris, M. C. Payne, and A. A. Mostofi, Phys. Rev. B, **83**, 195102 (2011).
[44] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B, **47**, 10891 (1993); R. W. Nunes and D. Vanderbilt, **50**, 17611 (1994); M. S. Daw, **47**, 10895 (1993).
[45] C. A. White, P. Maslen, M. S. Lee, and M. Head-Gordon, Chem. Phys. Lett., **276**, 133 (1997).
[46] M. D. Towler, N. L. Allan, N. M. Harrison, V. R. Saunders, W. C. Mackrodt, and E. Aprà, Phys. Rev. B, **50**, 5041 (1994).
[47] G. A. Sawatzky and J. W. Allen, Phys. Rev. Lett., **53**, 2339 (1984).
[48] X. Ren, I. Leonov, G. Keller, M. Kollar, I. Nekrasov, and D. Vollhardt, Phys. Rev. B, **74**, 195114 (2006).
[49] A. Svane and O. Gunnarsson, Phys. Rev. Lett., **65**, 1148 (1990).
[50] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B, **57**, 1505 (1998).
[51] O. Bengone, M. Alouani, P. Blöchl, and J. Hugel, Phys. Rev. B, **62**, 16392 (2000).
[52] W. E. Pickett, S. C. Erwin, and E. C. Ethridge, Phys. Rev. B, **58**, 1201 (1998).
[53] S. López, A. H. Romero, J. Mejía-López, J. Mazo-Zuluaga, and J. Restrepo, Phys. Rev. B, **80**, 085107 (2009).
[54] K. Palotás, A. N. Andriotis, and A. Lappas, Phys. Rev. B, **81**, 075403 (2010).
[55] Norm-conserving pseudopotentials, relativistically corrected and with a non-linear core correction for nickel, were generated using the Opium code available at `http://opium.sourceforge.net`.
[56] T. Ozaki, Phys. Rev. B, **64**, 195110 (2001).
[57] D. D. O'Regan, M. C. Payne, and A. A. Mostofi, In preparation (2012).
[58] C. Weber, D. D. O'Regan, N. D. M. Hine, M. C. Payne, G. Kotliar, and P. B. Littlewood, (2012) arXiv:1202.1423.
[59] D. D. O'Regan, *Optimised Projections for the Ab Initio Simulation of Large and Strongly Correlated Systems*, 1st ed., Springer Theses, Vol. XVI (Springer, Berlin, Heidelberg, 2012) p. 225.
[60] P. D. Haynes, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, J. Phys.: Condens. Matter, **20**, 294207 (2008).