Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information for

Optimized multi-site local orbitals in the large-scale DFT program CONQUEST

Ayako Nakata,*[,a,b] David R. Bowler,[c,d,e] and Tsuyoshi Miyazaki**[,b,c]

[a]International Center for Young Scientists (ICYS), National Institute for Materials Science (NIMS)

[b]Computational Materials Science Unit (CMSU), National Institute for Materials Science (NIMS)

[c]Department of Physics & Astronomy, University College London

[d]WPI-MANA, National Institute for Materials Science (NIMS)

[e]London Centre for Nanotechnology, University College London

*NAKATA.Ayako@nims.go.jp

**MIYAZAKI.Tsuyoshi@nims.go.jp

Dependence of the total energy on the chemical potential in multi-site support functions

The PAO coefficients in the multi-site support functions contain the Fermi-Dirac function $f(\varepsilon)$ as shown in Eq. (6) in the main paper. The chemical potential μ in $f(\varepsilon)$ is set to be close to the local Fermi level, in order to eliminate the effects from unoccupied molecular orbitals (MOs) far from the Fermi level which usually have no physical meaning. Table S1 shows the difference in total energy of bulk Si for multi-site support functions with several μ values, relative to the primitive support function result. We use three values for μ: 0.000, 0.175, and the mean value of the energies of the highest occupied and the lowest unoccupied MOs of each subsystem. DZP primitive PAOs were used (the range of the s, p and d orbitals were (4.9, 4.4), (6.1, 5.0), and 6.1 bohr respectively). Non-self-consistent Harris-Foulkes energies [J. Harris, *Phys. Rev. B*, 1985, **31**, 1770.] are compared. As shown in the table, the dependence on μ is not large, especially when r_{MS} and r_{LD} are large. The dependence on μ will be eliminated when we optimize the PAO coefficients. The dependence on kT in $f(\varepsilon)$ is discussed in reference [15].

Table S1. Differences of the Harris-Foulkes energies [mhartree/atom] by multisite support functions (r_{LD}-r_{MS}) from that by primitive support functions.

μ	0.000	0.175	$(\varepsilon_{\mathrm{LUMO}}-\varepsilon_{\mathrm{HOMO}})/2$
(5.0-5.0)	18.0	5.7	13.6
(8.0-8.0)	2.4	1.0	1.4
(17.0-17.0)	0.1	0.1	0.1
Primitive DZP		-3959.1	

Brief note about the computational cost of CONQUEST

The computational cost for constructing the Hamiltonian (even for the whole system) is much smaller than that for the diagonalization calculations in CONQUEST, because all of the matrices in CONQUEST (including the Hamiltonian) are constructed as sparse matrices. Since the support functions are strictly zero beyond a cutoff region r_{SF}, we only need to calculate the matrix elements in a region of extent $2r_{\mathrm{SF}}$.

The charge density n at each grid point g is calculated as

$$n\left(\mathbf{r}_g\right) = \sum_{i\alpha, j\beta} \phi_{i\alpha}\left(\mathbf{r}_g\right) K_{i\alpha, j\beta} \phi_{j\beta}\left(\mathbf{r}_g\right)^*$$

where again only the element with "$(r_i - r_j) < 2r_{\mathrm{SF}}$" are taken into account. The Coulomb and exchange-correlation potentials are determined with the density $n(r)$. The Coulomb potential is calculated as

$$V^{Coulomb}(\mathbf{r}) = \int d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{\left|\mathbf{r} - \mathbf{r}'\right|}$$

by numerical integration with a fast Fourier transform (FFT). All other contributions to the Hamiltonian (the kinetic energy and the pseudopotential terms) are found via integration in local regions. Therefore, the construction of matrices is much less cost-consuming than the diagonalization. More details are given in references [2] and [9].