Supplementary Information for "De novo exploration and self-guided learning of potential-energy surfaces"

Noam Bernstein[1], Gábor Csányi[2] & Volker L. Deringer[2,3]

[1]Center for Materials Physics and Technology, U.S. Naval Research Laboratory, Washington, DC 20375, United States
[2]Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom
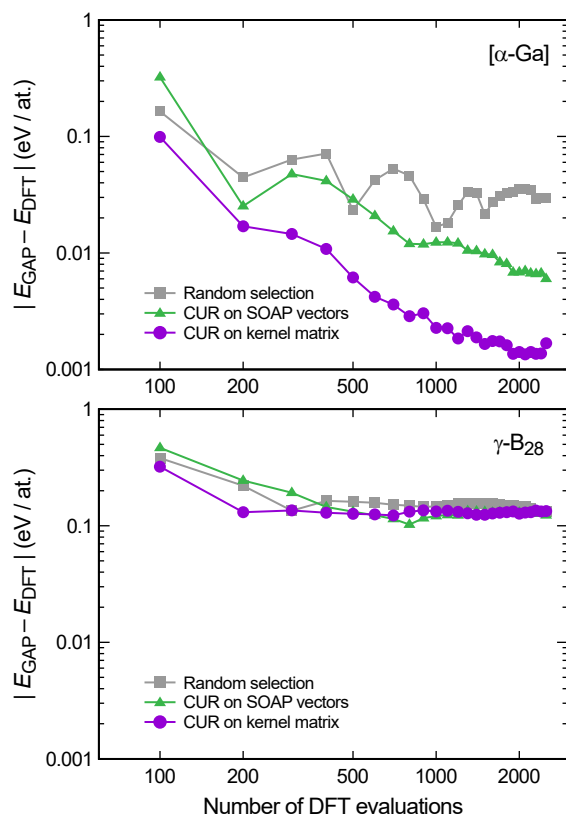[3]*Present address:* Department of Chemistry, University of Oxford, Oxford OX1 3QR, United Kingdom

FIG. S1. Evolution of energy errors for different boron allotropes and with different selection methods (*as a supplement to Fig. 2*). We show the errors for other, known boron crystal structures (α-Ga, *top*, and γ-B_{28}, *bottom*) as a function of the number of total DFT evaluations, and again for different diverse structure selection methods. Purple curves indicate leverage-score CUR on kernel similarity matrix, compared to random selection (grey) and CUR on SOAP vectors (green). For α-Ga, in line with our results for α-B_{12} in the main text, the random selection produces the least accurate potentials, while CUR on SOAP vectors (ignoring the non-linear exponentiation producing the similarity kernel) is better, but not as accurate as CUR on similarity kernels. The structure of γ-B_{28} contains more atoms in the unit cell than our search space encompasses (up to 24 atoms/cell), and its description therefore improved during the first iterations but not very accurate with the final potential (irrespective of the method used for selecting structures), as expected.
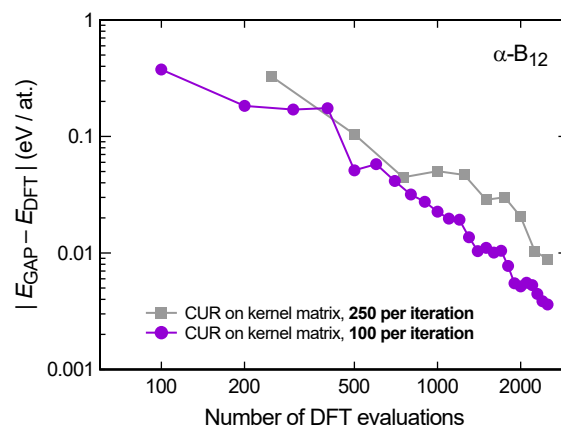

FIG. S2. Comparison of different N per iteration (*as a supplement to Fig. 2*). The overall performance is similar, but in this case using 100 DFT-evaluated structures per iteration produces a somewhat lower error for the final potential at the same (overall) DFT computation cost.
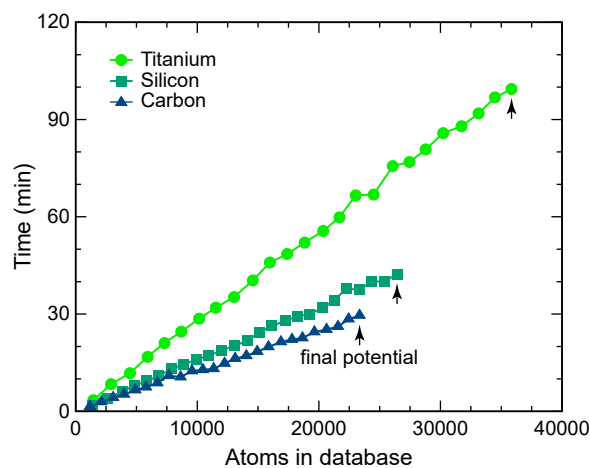

FIG. S3. Computational time for fitting potentials (*as a supplement to Fig. 3*). The cost increases linearly with iteration (due to the linear increase in database size), constituting a significant fraction of the total time per iteration by the end of the process. The time required to fit the final potentials is below one hour (carbon, silicon) and below two hours (titanium), respectively.
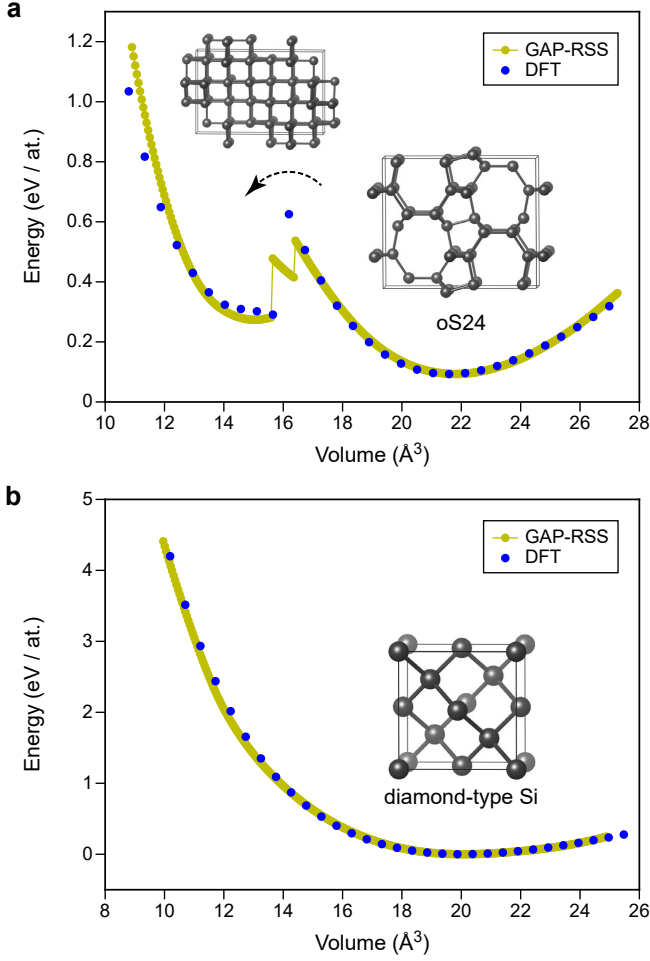
FIG. S4. Silicon allotropes at high pressure (*as a supplement to Fig. 3e*). We here show energy–volume curves as in the main text, but with smaller increments and extending down to much smaller unit-cell volumes (that is, higher-pressure regions), down to 50% of the equilibrium volume. In the case of the open-framework silicon structure oS24 (Ref. 65; panel **a**), a collapse is observed at ≈ 16 Å^3, both in the DFT reference computation and in the GAP-RSS prediction. We also performed a test for diamond-type silicon, compressing it to similarly small volumes (and much more strongly than shown in Fig. 3e). Over the full range of volumes studied, good agreement is observed between the DFT reference and the GAP-RSS prediction (panel **b**), which indicates a robust "learning" of repulsion at very small interatomic distances.

TABLE S1. Additional information regarding the selected structures shown as insets in Fig. 5: external pressure applied for this relaxation trajectory, p_{ext}; energy relative to the respective ground state, ΔE; maximum DFT-computed force component F_i in this structure.

		p_{ext}	ΔE	max $\{F_i\}$
		(GPa)	(eV / at.)	(eV / Å)
(i)	C (graphite)	0.111	+0.15	0.302
(ii)	C (buckled)	0.088	+0.23	1.678
(iii)	Si (dist. **lon**)	0.048	+0.32	0.742
(iv)	Si (**unj**)	0.000	+0.06	0.022
(v)	Si (sp^3 network)	0.124	+0.29	0.962
(vi)	Si (simple hex.)	0.208	+0.25	0.653
(vii)	Ti (ω phase)	0.185	+0.03	0.000
(viii)	Ti (hcp)	0.324	+0.05	0.238
(ix)	Si (RSS intermed.)	0.172	+0.59	0.857

TABLE S2. Lattice parameters and atomic coordinates for silicon structure (v), as added to the GAP-RSS reference database in the final iteration (Fig. 5).

$a = b = c = 5.5869$ Å		
$\alpha = \beta = \gamma = 109.08°$		
Si -1.68605982	-0.41624606	3.95802416
Si 0.15435366	4.30188720	0.38981490
Si 4.01815977	-0.40190360	1.54106451
Si -2.28421007	2.67006652	2.34218566
Si 3.26751299	2.67507753	-0.00486561
Si 1.46783885	-1.91088147	3.46795123
Si 1.26796258	1.77652573	3.00303591
Si 0.48155278	0.67469729	1.14050708

TABLE S3. As Table S2 but for silicon structure (ix).

$a = b = c = 6.5595$ Å		
$\alpha = 84.76°$ $\beta = 130.49°$ $\gamma = 116.23°$		
Si -1.20744491	4.11837921	3.06869747
Si 1.15061078	2.70140772	0.21706451
Si -2.74075733	-1.74587757	4.91801644
Si -2.97936509	2.43953892	3.67216632
Si -1.80907975	3.75555505	5.55352463
Si -2.12808832	-0.09478641	3.03665790
Si 1.86693516	0.25171580	3.77895234
Si -0.51394068	1.68239975	3.54412519
Si 3.12840615	4.18317367	0.57109224
Si 2.03701947	0.79695955	1.44756468