

SI Appendix

Deep Elastic Strain Engineering of Bandgap through Machine Learning

Zhe Shi ^{1,*}, Evgenii Tsymbalov ^{2,*}, Ming Dao ¹, Subra Suresh ³, Alexander Shapeev ², Ju Li ¹

¹Department of Materials Science and Engineering and Department of Nuclear Science Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²Skolkovo Institute of Science and Technology, Moscow 121205, Russia

³Nanyang Technological University, Singapore 639798, Republic of Singapore

**These authors contributed equally.*

[†] SSuresh@ntu.edu.sg, A.Shapeev@skoltech.ru, liju@mit.edu

SI Appendix, Note S1: Straining of diamond cubic crystals in real and reciprocal space

The straining can be best described by applying a 3×3 tensor transformation to the perfect silicon or diamond primitive cell to avoid the confounding effect of band folding in larger supercells which causes difficulty in identifying band structure information (1). Due to material anisotropy, we present all the related figure of merit results within 6D strain space instead of 3D space of principal strains. To avoid redundant computations, we ensured that each strain we applied to a crystal has a one-to-one correspondence to a distinct deformation case. The non-translational part of a homogeneous deformation of a crystal can be defined by a second-order deformation gradient tensor \mathbf{F} , which can be viewed as the Jacobi matrix linking deformed and undeformed lattice vectors. The relationship between the symmetric strain tensor $\boldsymbol{\varepsilon}$ and \mathbf{F} is given by $\boldsymbol{\varepsilon} = \frac{1}{2}(\mathbf{F} + \mathbf{F}^T) - \mathbf{I}$. Since the band structure does not change upon rotations of the crystal, we can eliminate the rotational degrees of freedom by adopting upper triangular \mathbf{F} to map out all deformation cases, as in SI Appendix, Figure S1a.

Under general 3D three-normal-strains deformation, the original O_h crystal point group of Si turns into a D_{2h} point group. The Brillouin zone for deformed Si in this case is shown in SI Appendix, Figure S1b. In general it is not anymore a regular truncated octahedron with equilateral hexagonal and square faces. The reciprocal space lattice vectors are adjusted by the inverse transpose of the deformation gradient tensor in real space, i.e. \mathbf{F}^{-T} , as a result of the deformation. The center of any type of Brillouin zone is labeled as Γ and we keep this tradition. In undeformed Si, the centers of the square and regular hexagonal surfaces on the Brillouin zone boundary are completely degenerate and labeled as X and L , respectively. For the simplicity of comparison, we follow the same spirit and still denote the ‘ X ’-type points as the centers of the tetragon surfaces and L -type points as the centers of the regular/non-regular hexagonal surfaces. The lines that connect the Γ point to the ‘ X ’-type points are labeled as ‘ Δ ’-type. This way, the six ‘ X ’- and ‘ L ’-type points, though non-degenerate, would keep the correct fractional coordinates of $\langle 0.5, 0, 0.5 \rangle$ - and $\langle 0.5, 0, 0 \rangle$ -type, and the \mathbf{k} -points along the Γ -‘ X ’ line would all have the $\langle \zeta, 0, \zeta \rangle$ -type coordinates. As the CBMs of our concern always appear on either the center of the Brillouin zone, center of the zone boundary surfaces, or the line connecting the zone center and surface center, our notations are sufficient.

SI Appendix, Note S2:

First-principles calculations

We used the Perdew-Burke-Ernzerhof (PBE) (2) exchange-correlation functional and the projector augmented wave method (PAW) (3) in our DFT simulations implemented in the Vienna Ab initio Simulation Package (4) with spin-orbit coupling incorporated. A plane wave basis set with an energy cutoff of 520 eV was adopted to expand the electronic wavefunctions. The Brillouin zone integration was conducted on a $13 \times 13 \times 13$ Monkhorst-Pack \mathbf{k} -point mesh ($6 \times 6 \times 6$ for GW calculations). Atomic coordinates in all the structures were relaxed until the maximum residual force was below $0.0005 \text{ eV \AA}^{-1}$. We focused on the strain range of $\{-5\% \leq \varepsilon_j \leq 10\%, j = 1 \dots 6\}$ for silicon and $\{-5\% \leq \varepsilon_j \leq 5\%, j = 1 \dots 6\}$ for diamond. The large strain values and corresponding strain energy density values are on the same order of magnitude compared to those achieved experimentally for bulk silicon and bulk diamond and these strains are all below theoretical failure strains, i.e. without phonon instability occurring.

Machine learning

Neural network (NN)

NN fitting is implemented within the Tensorflow (5) framework. To predict the bandgap we used deep NNs with four hidden layers with a (64 - 128 - 256 - 256) structure in the case of three-normal-strains strains ($\boldsymbol{\varepsilon}^{3D}$) and a (512 - 256 - 256 - 256) structure for the case with shear strains ($\boldsymbol{\varepsilon}^{6D}$), as shown in main text Figure 1a and SI Appendix, Figure S2. For the more complicated task of band energy prediction at a single \mathbf{k} -point, the architecture of (512 - 256 - 256 - 256) was used. The leaky rectified linear unit was chosen as an activation function. We used the Adam stochastic optimization method, the orthogonal weight initialization (6) and the dropout technique to prevent overfitting.

Tree-based ensemble algorithms

The algorithms were implemented in Scikit-learn (7). For our regression task, we used two types of ensembling on decision trees: the random forest regression (8) and the gradient boosting regression (9). The architecture is shown in SI Appendix, Figure S2. Hyper-parameters tuning was executed by using cross-validation on a training set to enhance the fitting process.

Data fusion

Data fusion represents the concept of combining different data sources in order to improve the model (10). We adopted this approach to further improve the learning outcome of E_g , namely the most technically important property for an electronic material. While the data fusion model prediction in Ref. (11) corresponds to a baseline value plus a correction, our data fusion approach is more advanced. More specifically, given E_g^{PBE} computed using an approximate baseline level of theory (PBE) at a particular query strain case, a related E_g^{GW} value corresponding to a more accurate and more demanding target level of theory (GW) can be estimated as a function of both E_g^{PBE} and $\boldsymbol{\varepsilon}$. Therefore, the E_g^{GW} consistent with the query

strain case is learned using exclusively ϵ and E_g^{PBE} as input, as illustrated in Figure 1a and b. The resulting data fusion model reduces the MAE in the prediction of bandgap by more than half for kernel-based ensemble methods and allow the bandgap predicted by NN be reach an extremely high accuracy of 8 meV, as shown in main text Figure 1b and SI Appendix, Table S1.

SI Appendix, Note S3:

Ranking of common Si crystal direction families for obtaining the same target bandgap through uniaxial compressive straining (from the most energy efficient strain direction to the least energy efficient strain direction):

$\langle 111 \rangle$, $\langle 332 \rangle$, $\langle 322 \rangle$, $\langle 221 \rangle$, $\langle 211 \rangle$, $\langle 321 \rangle$, $\langle 331 \rangle$, $\langle 320 \rangle$, $\langle 210 \rangle$, $\langle 311 \rangle$, $\langle 110 \rangle$, $\langle 310 \rangle$, $\langle 100 \rangle$

Ranking of common Si crystal direction families for obtaining the same target bandgap through uniaxial tensile straining (from the most energy efficient strain direction to the least energy efficient strain direction):

$\langle 111 \rangle$, $\langle 332 \rangle$, $\langle 221 \rangle$, $\langle 322 \rangle$, $\langle 331 \rangle$, $\langle 211 \rangle$, $\langle 311 \rangle$, $\langle 321 \rangle$, $\langle 110 \rangle$, $\langle 320 \rangle$, $\langle 210 \rangle$, $\langle 310 \rangle$, $\langle 100 \rangle$

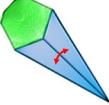
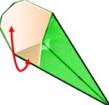
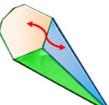
SI Appendix, Table S1:

Table S1: MAE and RMSE (in units of eV) for ML algorithms for bandgap prediction with or without the Δ -ML model. Here, the Lagrange polynomial of degree 8 is used. Relative error: norm of the difference between the true value and the prediction divided by the norm of the true value.

ML algorithms	GW		GW+PBE (Δ -ML)	
	MAE	RMSE	MAE	RMSE
Lagrange	0.0211	0.0274	0.0186	0.0241
GBR	0.0334	0.0521	0.0135	0.0209
RFR	0.0434	0.0596	0.0145	0.0215
NN	0.0099	0.0144	0.0080	0.0118
NN relative error	1.72%	2.78%	1.38%	2.05%

SI Appendix, Table S2:

Table S2: k -space CBM transitions. Each of 12 separating ridgelines of the iso-bandgap body tabulated. The constants k_1 and k_2 are approximately equal to 0.425 and 0.5, corresponding to points on Δ and L, respectively.

Type	Change of “carapace”	k -coordinate of CBM
‘ Δ ’-switching		$\Delta_1 \leftrightarrow \Delta_2$ $(0, k_1, k_1) \leftrightarrow (k_1, 0, k_1)$
		$\Delta_2 \leftrightarrow \Delta_3$ $(k_1, 0, k_1) \leftrightarrow (k_1, k_1, 0)$
		$\Delta_3 \leftrightarrow \Delta_1$ $(k_1, k_1, 0) \leftrightarrow (0, k_1, k_1)$
‘L’-switching		$L_1 \leftrightarrow L_2$ $(k_2, 0, 0) \leftrightarrow (0, k_2, 0)$
		$L_2 \leftrightarrow L_3$ $(0, k_2, 0) \leftrightarrow (0, 0, k_2)$
		$L_3 \leftrightarrow L_1$ $(0, 0, k_2) \leftrightarrow (k_2, 0, 0)$
‘L-to- Δ ’ transition		$L_1 \leftrightarrow \Delta_2$ $(k_2, 0, 0) \leftrightarrow (k_1, 0, k_1)$
		$L_1 \leftrightarrow \Delta_3$ $(k_2, 0, 0) \leftrightarrow (k_1, k_1, 0)$
		$L_2 \leftrightarrow \Delta_1$ $(0, k_2, 0) \leftrightarrow (0, k_1, k_1)$
		$L_2 \leftrightarrow \Delta_3$ $(0, k_2, 0) \leftrightarrow (k_1, k_1, 0)$
		$L_3 \leftrightarrow \Delta_1$ $(0, 0, k_2) \leftrightarrow (0, k_1, k_1)$
		$L_3 \leftrightarrow \Delta_2$ $(0, 0, k_2) \leftrightarrow (k_1, 0, k_1)$
Indirect-to-direct bandgap transition		$L_1 \leftrightarrow \Gamma$ $(k_2, 0, 0) \leftrightarrow (0, 0, 0)$
		$L_2 \leftrightarrow \Gamma$ $(0, k_2, 0) \leftrightarrow (0, 0, 0)$
		$L_3 \leftrightarrow \Gamma$ $(0, 0, k_2) \leftrightarrow (0, 0, 0)$

SI Appendix, Table S3:

Table S3: Si bandgap prediction errors, RMSE and MAE (in units of eV), for the incremental fitting scenario on reduced datasets. The error in both metrics is reduced for both ϵ^{3D} and ϵ^{6D} datasets after the incremental fitting.

	ϵ^{3D}		ϵ^{6D}	
	before	after	before	after
RMSE	0.0403	0.0069	0.0264	0.0253
MAE	0.0167	0.0052	0.0179	0.0167

SI Appendix, Figure S1:

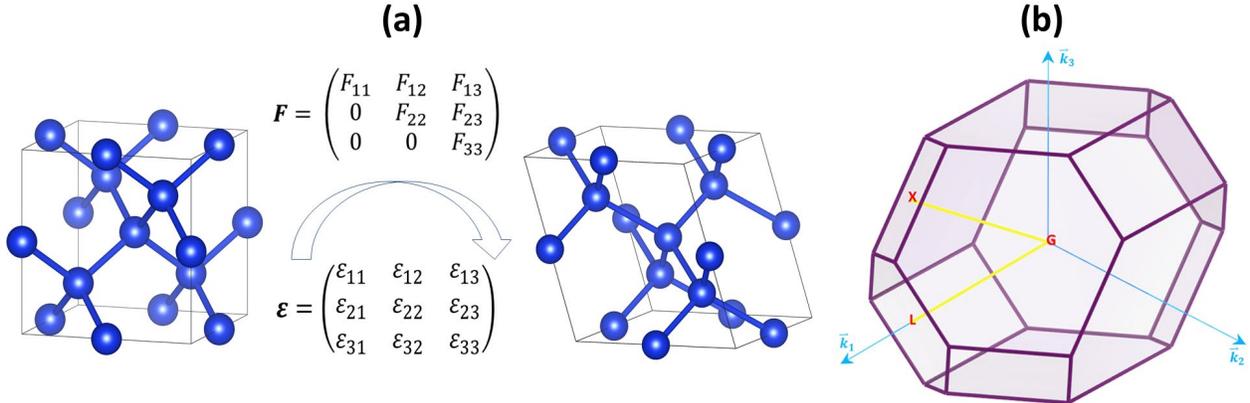


Figure S1: (a) ESE achieved by applying a reduced deformation gradient tensor to the undeformed diamond cubic lattice of Si or C in the real space. (b) Brillouin zone of diamond cubic crystal under three-normal-strains deformation. It is a tetradehedron with 8 hexagonal and 6 quadrilateral faces. The discussions based on Figure 5 of the main text incorporate the same labels and \mathbf{k} coordinates as here.

SI Appendix, Figure S2:

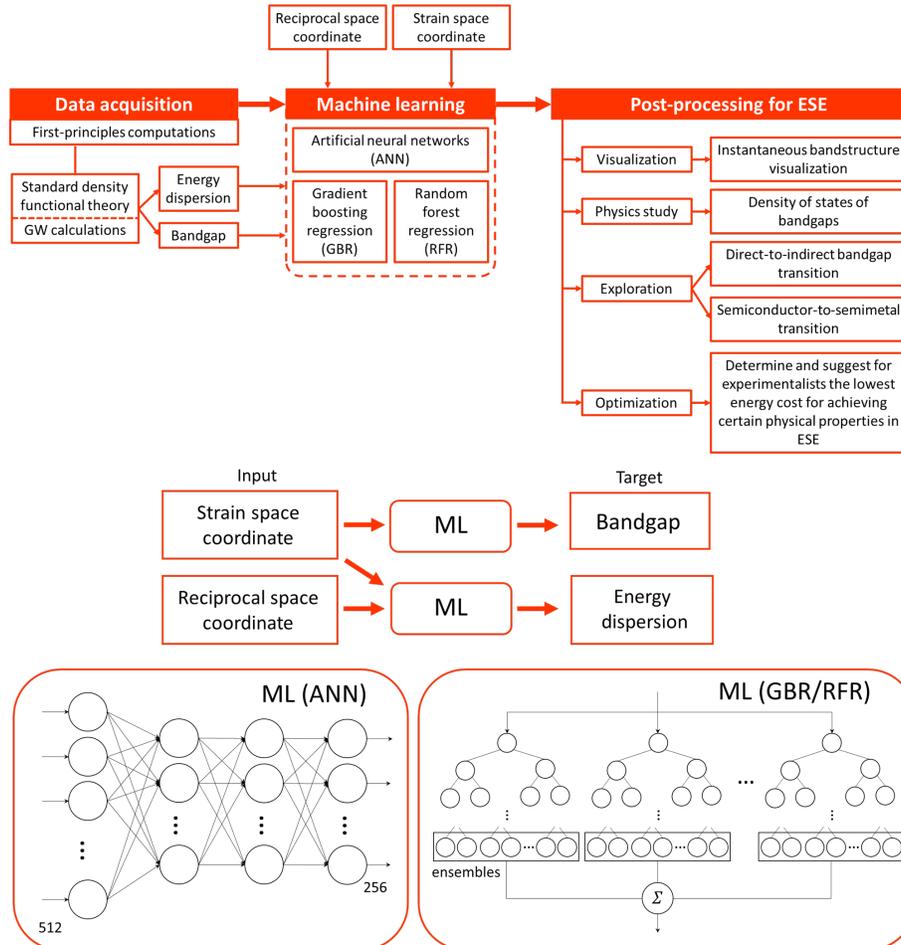


Figure S2: *Top*: Three major processing steps involved in the deep ESE of bandgap, including data acquisition through *ab initio* computations, ML and physical property exploration. *Middle*: Flowchart setting out the details of the ML process. Here, strain tensor and \mathbf{k} coordinate are used as the input whereas the bandgap and energy dispersion are the target for fitting. For the ML algorithms, our set-up supports ANN, GBR, RFR, and other kernel-based fitting methods. *Bottom*: Detailed architecture of the algorithms adopted.

SI Appendix, Figure S3:

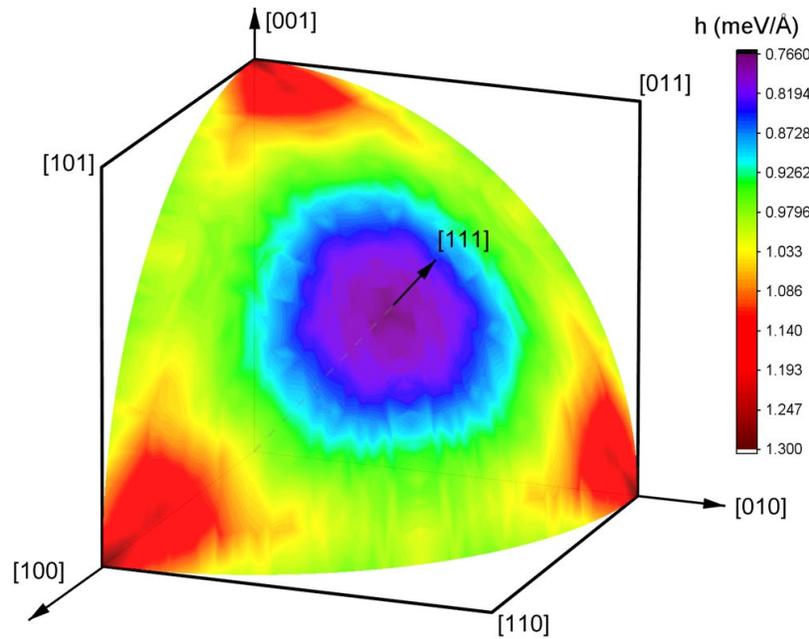


Figure S3: Color contour map of the elastic strain energy density (h) required to reach the same bandgap level of 0.6 eV through uniaxial compressive straining in Si.

References:

1. Umeno Y, Kushima A, Kitamura T, Gumbsch P, Li J (2005) Ab initio study of the surface properties and ideal strength of (100) silicon thin films. *Phys Rev B* 72(16):165431.
2. Perdew JP, Burke K, Ernzerhof M (1996) Generalized Gradient Approximation Made Simple. *Phys Rev Lett* 77(18):3865–3868.
3. Blöchl PE (1994) Projector augmented-wave method. *Phys Rev B* 50(24):17953–17979.
4. Kresse G, Furthmüller J (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput Mater Sci* 6(1):15–50.
5. Abadi M, et al. (2016) TensorFlow: A System for Large-scale Machine Learning.

Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, OSDI'16. (USENIX Association, Berkeley, CA, USA), pp 265–283.

6. Saxe AM, McClelland JL, Ganguli S (2013) Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *ArXiv13126120 Cond-Mat Q-Bio Stat*. Available at: <http://arxiv.org/abs/1312.6120> [Accessed October 8, 2018].
7. Pedregosa F, et al. (2011) Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12:2825–2830.
8. Breiman L (2001) Random Forests. *Mach Learn* 45(1):5–32.
9. Friedman JH (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat* 29(5):1189–1232.
10. Khaleghi B, Khamis A, Karray FO, Razavi SN (2013) Multisensor data fusion: A review of the state-of-the-art. *Inf Fusion* 14(1):28–44.
11. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA (2015) Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach. *J Chem Theory Comput* 11(5):2087–2096.