Uncertainty-aware Molecular Dynamics from Bayesian Active Learning for Phase Transformations and Thermal Transport in SiC

Yu Xie,[1,*] Jonathan Vandermause,[1,2] Senja Ramakers,[3,4]
Nakib H. Protik,[1] Anders Johansson,[1] and Boris Kozinsky[1,5,†]

[1] *John A. Paulson School of Engineering and Applied Sciences, Harvard University*
[2] *Department of Physics, Harvard University*
[3] *Corporate Sector Research and Advance Engineering, Robert Bosch GmbH, Renningen*
[4] *Interdisciplinary Centre for Advanced Materials Simulation, Ruhr-Universität Bochum*
[5] *Robert Bosch LLC, Research and Technology Center*

Supplementary Materials

* xiey@g.harvard.edu
† bkoz@seas.harvard.edu

SUPPLEMENTARY METHODS

Supplementary Method 1: Select Uncertainty Threshold by Training Set Uncertainty

In the active learning, the Bayesian force field predicts uncertainty to quantify the confidence of the model of the current frame. Then the uncertainty is compared to an acquisition threshold. If the uncertainty is higher than the threshold then a DFT call is made to compute the "ground truth" energy (e), forces (f) and stress (s) of the current configuration. The uncertainties of total potential energy, forces and stress are computed from the square root of the variance of the full GP:

$$V_{pp}(\rho) = k_{pp} - \mathbf{k}_{pF}^{\top}(\mathbf{K}_{FF} + \mathbf{\Lambda})^{-1}\mathbf{k}_{Fp}, \quad p = e, f, s \tag{1}$$

which depend on the signal variance of the kernel function and the noise parameters σ_n. Therefore, choosing an absolute (constant) threshold throughout the active learning ignores the dependence and may determine a frame as "familiar/unfamiliar" inaccurately. In Ref.[1, 2], a multiple of $\sigma_n$is used as the threshold, such that the threshold changes while the noise parameter is optimized to a different value. In Ref.[3], a simple form of the raw uncertainty of sparse GP is used

$$V(\rho) = \frac{1}{\sigma^2}(k_{\varepsilon\varepsilon} - \mathbf{k}_{\varepsilon S}K_{SS}^{-1}\mathbf{k}_{S\varepsilon}) \tag{2}$$

where the normalization by the signal variance σ make $V(\rho)$ independent of the signal variance and noise. Then an absolute (constant) threshold can be used in this special case.

In this work, we introduce a more general way to determine the threshold. Particularly, we compute the average uncertainty $\bar{V}_{\text{train}}$ of all the training data, and then use a multiple of $\bar{V}_{\text{train}}$ as a threshold. In the active learning of SiC, we call the DFT when $\sqrt{\bar{V}} > 1.5\sqrt{\bar{V}_{\text{train}}}$. The relative uncertainty plotted in the Fig.2a in main text is $\sqrt{\bar{V}/\bar{V}_{\text{train}}}$.

SUPPLEMENTARY TABLES

Supplementary Table 1: On-the-fly Training Settings

Parameter	Value	Description
cutoff (Å)	5.0	Cutoff radius of the local environment for ACE descriptors
$\sigma^{(\mathrm{B1})}$	78.2991	Signal variance of kernel of B1 descriptors
$\sigma^{(\mathrm{B2})}$	3.5271	Signal variance of kernel of B2 descriptors
$\sigma_{\mathrm{n}}^{(\mathrm{energy})}$	0.1086	Noise parameter for total energy
$\sigma_{\mathrm{n}}^{(\mathrm{forces})}$	0.0895	Noise parameter for forces
$\sigma_{\mathrm{n}}^{(\mathrm{stress})}$	0.0021	Noise parameter for stress
$n_{\mathrm{max}}^{(\mathrm{B1})}$	11	Maximal index of radial basis function of ACE B1 descriptors
$n_{\mathrm{max}}^{(\mathrm{B2})}$	8	Maximal index of radial basis function of ACE B2 descriptors
$l_{\mathrm{max}}^{(\mathrm{B2})}$	4	Maximal index of spherical harmonics function of ACE B2 descriptors

Table S1: For the details of the parameters of the kernels, we direct the reader to Ref.[3]. For the details of the parameters of the ACE descriptors, we direct the reader to Ref.[4]

Supplementary Table 2: Training Set Collected from Bayesian Active Learning

System	Pressure	Temperature	τ_{sim}	τ_{wall}	N_{nodes}	N_{atoms}	N_{struc}	N_{envs}	N_{sparse}	N_{labels}
3C	0-480	2000	0.85	20.94	1	64	110	7040	550	21890
2H	0-480	2000	0.85	34.94	1	72	125	9000	625	27875
4H	0-480	2000	0.85	36.87	1	72	131	9432	655	29213
6H	0-480	2000	0.85	74.04	1	108	122	13176	610	40382
3C	0-800	300	0.85	9.78	1	64	72	4608	360	14328
RS	200-0	2000	0.55	17.22	1	64	120	7680	600	23880
Total	-	-	-	193.79	-	-	680	50936	3400	157568

Table S2: Summary of on-the-fly training for the SiC system: the simulation temperature T (K); the pressure (GPa) range of the simulation; the total simulation time τ_{sim} (ns); the total wall time of the simulation τ_{wall} (hr); the number of CPU nodes used for the training (with 32 CPUs per node); the number of atoms in the simulation N_{atoms}; the total number of structures added to the training set N_{struc}; the total number of local environments in all training structures N_{envs}; the total number of local environments added to the sparse set N_{sparse}; and the total number of training labels N_{labels}.

We note that since each on-the-fly training only uses a single CPU node with 32 cores, in practice we run these on-the-fly training simulations simultaneously on a cluster. The "System" in the table specifies the initial polytype/phase of the simulation. The simulations starting with 3C/2H/4H/6H are compressive with increasing pressures. The simulation starting with RS is decompressive with decreasing pressures. The pressure range in the table also specifies the compression/decompression. For example, "0-480" means a compressive simulation from 0 to 480 GPa, and "200-0" means a decompressive simulation from 200 to 0 GPa.

Supplementary Table 3: Elasticity

Polytype	Method	C_{11}	C_{33}	C_{12}	C_{13}	C_{44}	C_{66}	B_V	G_V
	EXP	390[a], 395[e], 371[c]		142[a], 123[e], 146[c]		150[b], 256[a], 236[e], 111[c]			
	DFT	384.3	384.3	127.9	127.9	239.9	239.9	213.6	195.4
3C	FLARE BFF	373.1	373.1	128.0	128.0	230.3	230.3	209.7	187.2
	Tersoff	426	426	109.9	109.9	252.7	252.7	215.3	214.8
	MEAM	402.2	402.2	115.6	115.6	215.1	215.1	211.1	186.4
	EXP								
	DFT	494.3	533.5	102.1	50.7	151.3	196.1	214.3	187.6
2H	FLARE BFF	482.8	527.2	101.0	52.7	158.1	190.9	211.7	187.2
	Tersoff	502.2	541.6	84.4	42.3	188.8	208.9	209.3	215.9
	MEAM	467.4	497.9	98.2	67.7	167.2	184.6	211.1	189.1
	EXP	501[d]	553[d]	111[d]	52[d]	163[d]			
	DFT	487.8	533.4	105	51.9	157.9	191.4	214	188.1
4H	FLARE BFF	476.0	514.9	104.8	55.0	158.4	185.6	210.7	184.0
	Tersoff	506.6	547.7	86.6	44.9	189.1	210	212.7	216.9
	MEAM	467.2	497.9	98.4	67.7	167.2	184.4	211.1	189.1
	EXP	501[d]	553[d]	112[d]	52[d]	163[d]			
	DFT	485.2	534.1	106	52.3	160.1	189.6	213	188.2
6H	FLARE BFF	472.2	515.3	106.1	55.3	158.4	183.1	210.3	182.8
	Tersoff	507.2	548.6	87	45.3	189.1	210.1	213.1	217
	MEAM	467.2	497.9	98.4	67.7	167.2	184.4	211.1	189

[a] Ref. [5] [b] Ref. [6] [c] Ref. [7] [d] Ref. [8] [e] Ref. [9] (Unit: GPa)

Table S3: Elastic tensor ($C_{11} \sim C_{66}$), bulk modulus (B_V) and shear modulus (G_V) from experiments (EXP), DFT[10], FLARE BFF and empirical potentials (Tersoff, MEAM).

We note that since FLARE BFF is trained on DFT data but not the experiments, we expect it to give predictions closer to DFT than experiments, while as shown in Table S3, we still find good match for the four polytypes between FLARE BFF and experimental results.

Supplementary Table 4: Performance Comparison

Model	Parameters	Speed (potential)	Speed (variance)	Force MAE
2+3 body	2-body: $r_{\text{cut}} = 5$Å, $n_{\text{grid}} = 32$ 3-body: $r_{\text{cut}} = 5$Å, $n_{\text{grid}} = 16^3$	0.20525	$O(N_{\text{ranks}} t^{(2+3)}_{\text{potential}})$	136.5
ACE-based	$r_{\text{cut}} = 5$Å, $n_{\text{max}} = 8$, $l_{\text{max}} = 3$	0.21123	$O(t^{(\text{ace})}_{\text{potential}})$	68.4

Table S4: Performance test with MGP potentials of SiC. Speed unit: ms $\cdot$ CPU/(atom $\cdot$ timestep). Force MAE unit: meV/Å.

2+3 body based mapped GP is an approximation, whose accuracy depends on the grid resolution of the spline interpolation, and the ranks from PCA (for variance). ACE based mapped GP is exact, which only differs from the original Sparse GP model by floating point errors. To reduce computational cost, we chose to use the $\xi = 2$ potential and $\xi = 1$ variance, where the $\xi = 1$ variance is a good approximation of the $\xi = 2$ variance, as shown in the Supplementary Information Figure S1.

The comparison of the performance between 2+3 body based MGP force field and the ACE based MGP is shown in the Table 4. The computational cost of the mapped variance of the 2+3 body based model is dominated by the PCA decomposed 3-body variance $O(N_{\text{ranks}} t^{(2+3)}_{\text{potential}})$. The computational cost of the mapped force field is $t^{(\text{ace})}_{\text{potential}} \sim O(n^{\xi}_{\text{descriptor}})$, and that of the mapped variance of ACE based model is $O(n^{2\xi}_{\text{descriptor}})$, where the ξ is the power of the kernel. Since we use $\xi = 1$ for variance and $\xi = 2$ for the force field, the cost of variance is comparable to the force field $O(t^{(\text{ace})}_{\text{potential}})$.

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Variance Correlations

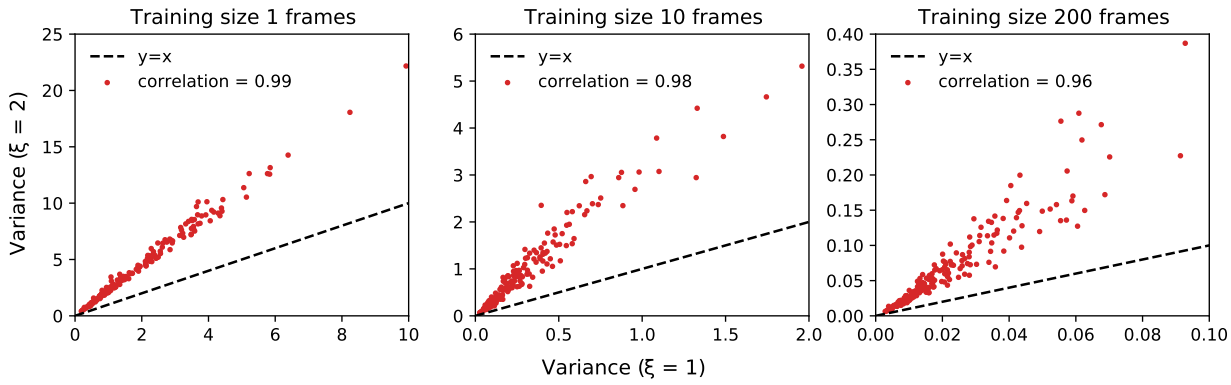

Figure S1: Variance comparison between different powers.

For the computational efficiency, we map the variance with power $\xi = 1$ of the normalized inner product kernel of sparse GP for prediction in LAMMPS. While the energy, forces and stress are predicted from power $\xi = 2$ kernel, the variance with $\xi = 1$ is an approximation of the variance from $\xi = 2$ kernel. We test the correlation between the two variances for the same testing configurations with different training set sizes. As shown in Fig.S1, the two variances have a high correlation above 90%, even though the values present a constant ratio. From Supplementary Method, we compare the uncertainty of testing data with the average training set uncertainty in the active learning. Therefore, the constant ratio does not make much difference as long as the approximate uncertainty preserves the order of high/low uncertainties.

Supplementary Figure 2: Polytype Stability

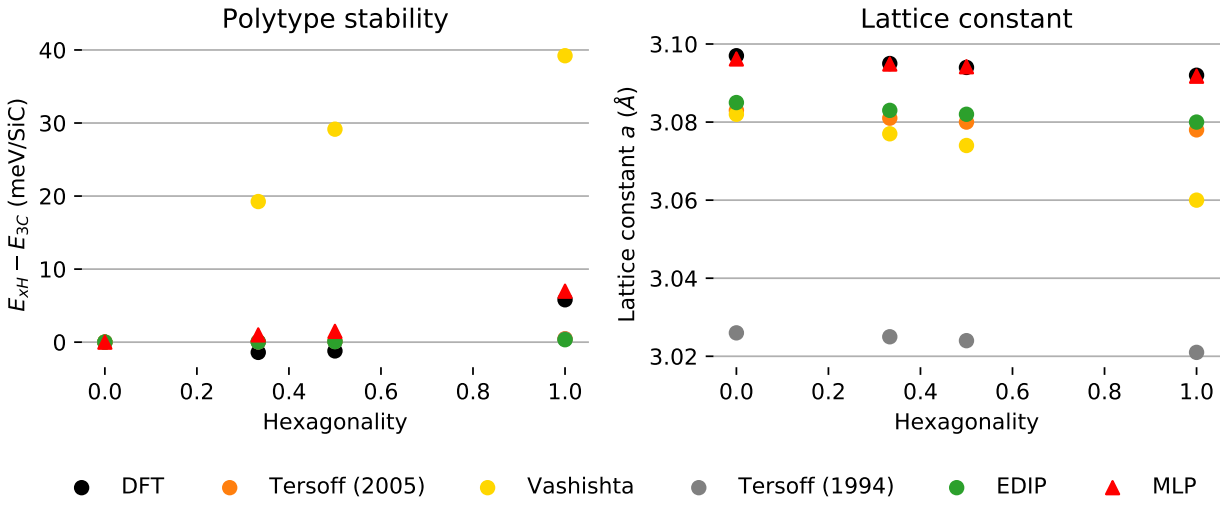

Figure S2: Polytype stability and lattice constant of 3C (hexagonality=0), 6H (hexagonality=1/3), 4H (hexagonality=1/2) and 2H (hexagonality=1), predicted from DFT, FLARE BFF and empirical potentials including Tersoff[11, 12], Vashishta[13] and EDIP[14].

We compare FLARE BFF with DFT and empirical potentials on the polytype stability and lattice constant at zero pressure. Specifically, we compute ground state energy per Si-C pair and compare the energy differences between hexagonal polytypes (2H, 4H and 6H) with the cubic one (3C). While the DFT energy differences between 4H, 6H and 3C are as small as $\sim$ 1 meV, 2H is significantly more unstable than the other polytypes by $\sim$ 6 meV. As shown in Fig.S2a, empirical methods such as Tersoff[11, 12] and EDIP[14] do not distinguish the energy differences between different polytypes, and Vashishta[13] potential predicts very large differences which over-stabilizes the cubic type, while the FLARE BFF reaches around 1 meV error with DFT and gives the correct polytype stability. The lattice constants of different polytypes are also compared in Fig.S2b, where the discrepancy between FLARE BFF and DFT is around 0.002Å. And all methods manage to predict the decreasing trend while the hexagonality increases.

Supplementary Figure 3: Phonon Dispersion from Models with Different Cutoffs

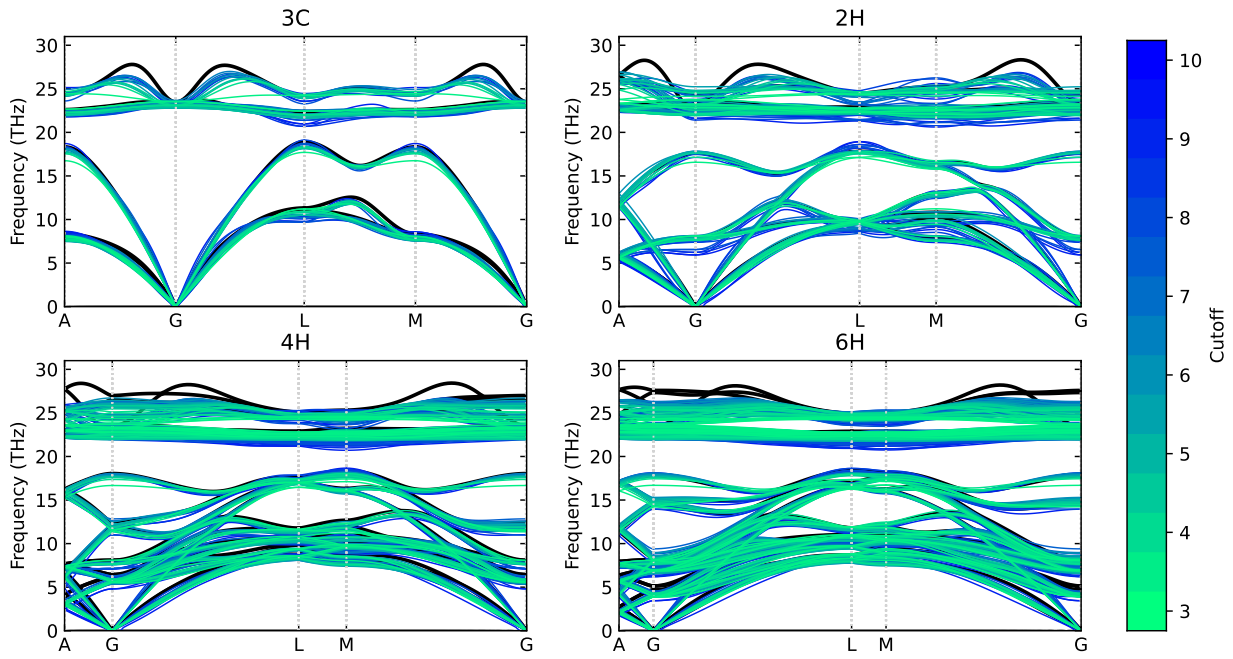

Figure S3

As shown in Fig.5 of the main text, the FLARE BFF and DFT phonon dispersions have an excellent agreement in the acoustic and transverse optical (TO) branches, but show mismatch in the longitudinal optical (LO) branches. Considering the LO-TO splitting, even though we do not include non-analytical correction, it is still likely that the polarization makes a non-trivial contribution and is not captured due to the near-sightedness of our machine learning model. Therefore, we tested how the model cutoff affects the phonon. We trained the model with cutoff from 3 to 10Å, and the phonons predicted by the models are plotted from light to dark colors. As shown in Fig.S3, all models have an excellent agreement with DFT in acoustic and TO branches, even with a cutoff as short as 3Å. While in the LO branches, the models with smaller cutoffs predict significantly flatter bands than the ones with larger cutoffs. Though with the largest cutoff our FLARE BFF still does not predict the LO branches perfectly, this result shows an evidence that one source of the mismatch in the LO branches is the long-range effect. Further investigation is needed such as introducing long-range interactions to the FLARE BFF, and will be left to our future work.

Supplementary Figure 4: Non Analytic Correction on the Thermal Conductivity

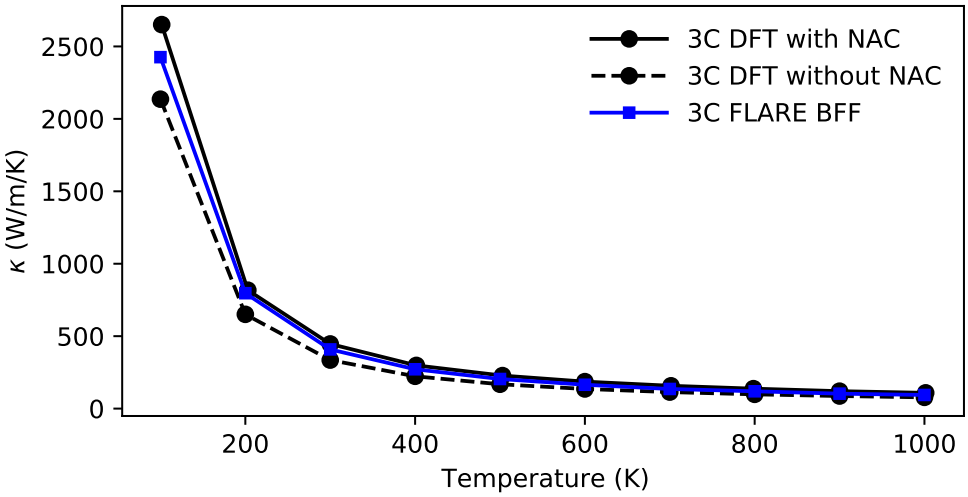

Figure S4: Thermal conductivity without and with NAC [15] at different temperatures.

We calculated the thermal conductivity from Boltzmann transport equation with the non analytical correction (NAC) term from Born effective charge, and compared it to that without NAC term at temperatures from 100 to 1000 K. Fig.S4 indicates that the exclusion of the non analytical correction term underestimates the thermal conductivity, but the underestimation is not significant. Since FLARE BFF is trained on the DFT data without long-range electrostatic interactions, and the force constants are computed without NAC term, it underestimates thermal conductivity compared to the DFT with NAC as expected.

Supplementary Figure 5: Accuracy Comparison of 2+3-body and ACE-based Models

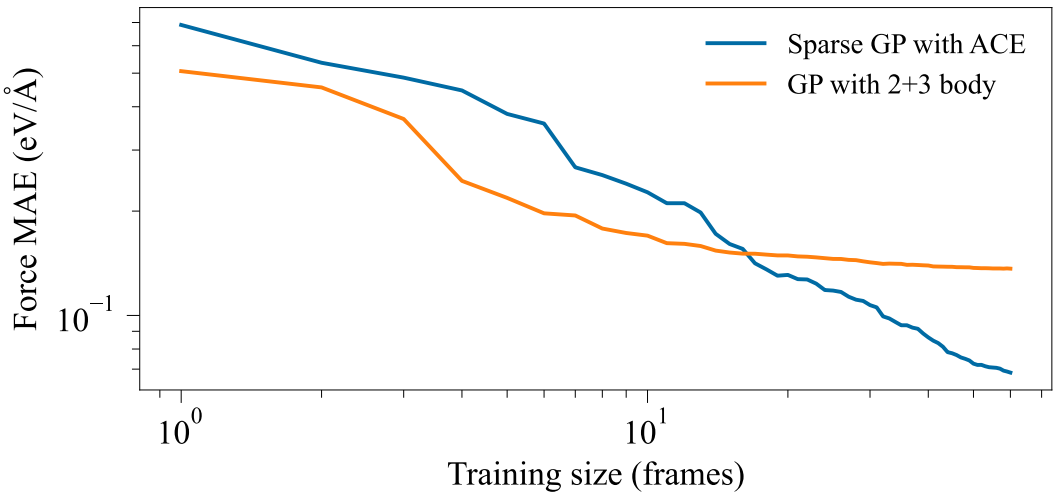

Figure S5: Comparison of accuracy between ACE-based and 2+3 body based Gaussian process force field on the bulk SiC dataset.

A comparison of accuracy between the GP model with 2+3 body descriptors and the one with ACE descriptors is made with the SiC bulk dataset, as shown in Fig.S5.

SUPPLEMENTARY REFERENCES

[1] Vandermause, J. *et al.* On-the-fly active learning of interpretable bayesian force fields for atomistic rare events. *npj Comput. Mater.* **6**, 1–11 (2020).

[2] Xie, Y., Vandermause, J., Sun, L., Cepellotti, A. & Kozinsky, B. Bayesian force fields from active learning for simulation of inter-dimensional transformation of stanene. *npj Computational Materials* **7**, 1–10 (2021).

[3] Vandermause, J., Xie, Y., Lim, J. S., Owen, C. J. & Kozinsky, B. Active learning of reactive bayesian force fields: Application to heterogeneous hydrogen-platinum catalysis dynamics (2021). 2106.01949.

[4] Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B* **99** (2019). URL https://link.aps.org/doi/10.1103/PhysRevB.99.014104.

[5] Feldman, D. W., Parker, J. H., Choyke, W. J. & Patrick, L. Phonon Dispersion Curves by Raman Scattering in SiC, Polytypes 3 C , 4 H , 6 H , 1 5 R , and 2 1 R. *Physical Review* **173**, 787–793 (1968). URL https://link.aps.org/doi/10.1103/PhysRev.173.787.

[6] Harrison, W. & Freeman, W. Electronic structure and the properties of solids: the physics of the chemical bond. *Journal of Molecular Structure* **71**, 355 (1981). URL https://linkinghub.elsevier.com/retrieve/pii/0022286081851368.

[7] Pestka, K. A., Maynard, J. D., Gao, D. & Carraro, C. Measurement of the Elastic Constants of a Columnar SiC Thin Film. *Physical Review Letters* **100** (2008). URL https://link.aps.org/doi/10.1103/PhysRevLett.100.055503.

[8] Kamitani, K. *et al.* The elastic constants of silicon carbide: A Brillouin-scattering study of 4h and 6h SiC single crystals. *Journal of Applied Physics* **82**, 3152–3154 (1997). URL http://aip.scitation.org/doi/10.1063/1.366100.

[9] Djemia, P., Roussigné, Y., Dirras, G. F. & Jackson, K. M. Elastic properties of β-SiC films by Brillouin light scattering. *Journal of Applied Physics* **95**, 2324–2330 (2004). URL http://aip.scitation.org/doi/10.1063/1.1642281.

[10] Ramakers, S. *et al.* Effects of thermal, elastic, and surface properties on the stability of sic

polytypes. *Phys. Rev. B* **106**, 075201 (2022). URL https://link.aps.org/doi/10.1103/PhysRevB.106.075201.

[11] Erhart, P. & Albe, K. Analytical potential for atomistic simulations of silicon, carbon, and silicon carbide. *Physical Review B* **71** (2005). URL https://link.aps.org/doi/10.1103/PhysRevB.71.035211.

[12] Tersoff, J. Chemical order in amorphous silicon carbide. *Physical Review B* **49**, 16349 (1994). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.49.16349.

[13] Vashishta, P., Kalia, R. K., Nakano, A. & Rino, J. P. Interaction potential for silicon carbide: A molecular dynamics study of elastic constants and vibrational density of states for crystalline and amorphous silicon carbide. *Journal of Applied Physics* **101**, 103515 (2007). URL http://aip.scitation.org/doi/10.1063/1.2724570.

[14] Jiang, C., Morgan, D. & Szlufarska, I. Carbon tri-interstitial defect: A model for the D II center. *Physical Review B* **86** (2012). URL https://link.aps.org/doi/10.1103/PhysRevB.86.144118.

[15] Protik, N. H. *et al.* Phonon thermal transport in 2h, 4h and 6h silicon carbide from first principles. *Materials Today Physics* **1**, 31–38 (2017). URL https://linkinghub.elsevier.com/retrieve/pii/S2542529317300743.