

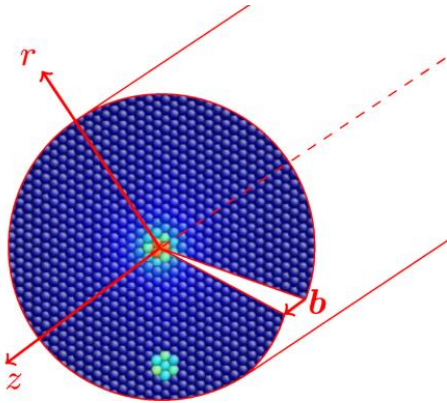
Machine-learning material properties with domain knowledge of the interatomic bond

Thomas Hammerschmidt,
Alvin Ladines, Aparna Subramanyam, Jan Jenke, Yury Lysogorskiy,
Ralf Drautz
ICAMS, Ruhr University Bochum, Germany

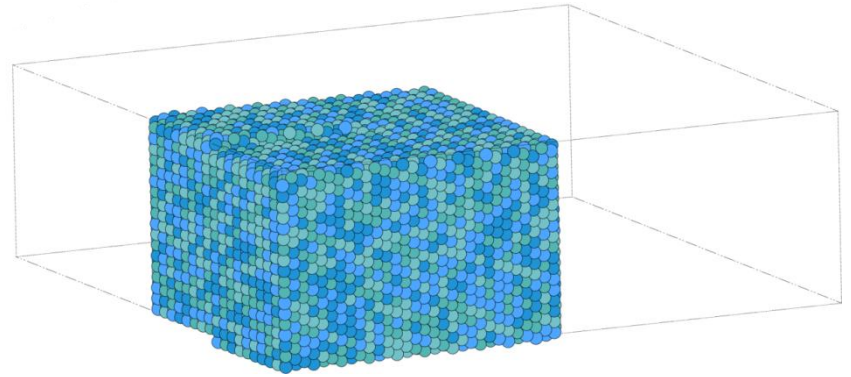
David Pettifor
MML, Oxford University

Atomistic simulations of structural materials

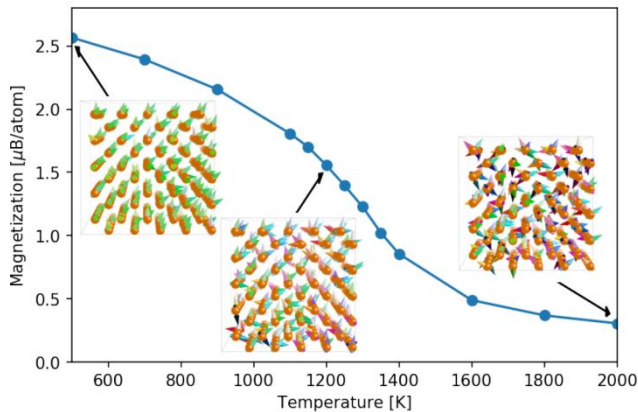
screw dislocation in bcc-Fe



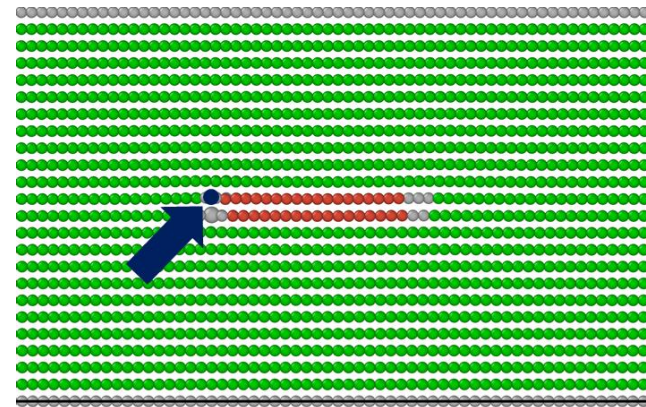
screw-dislocations in high-entropy alloys



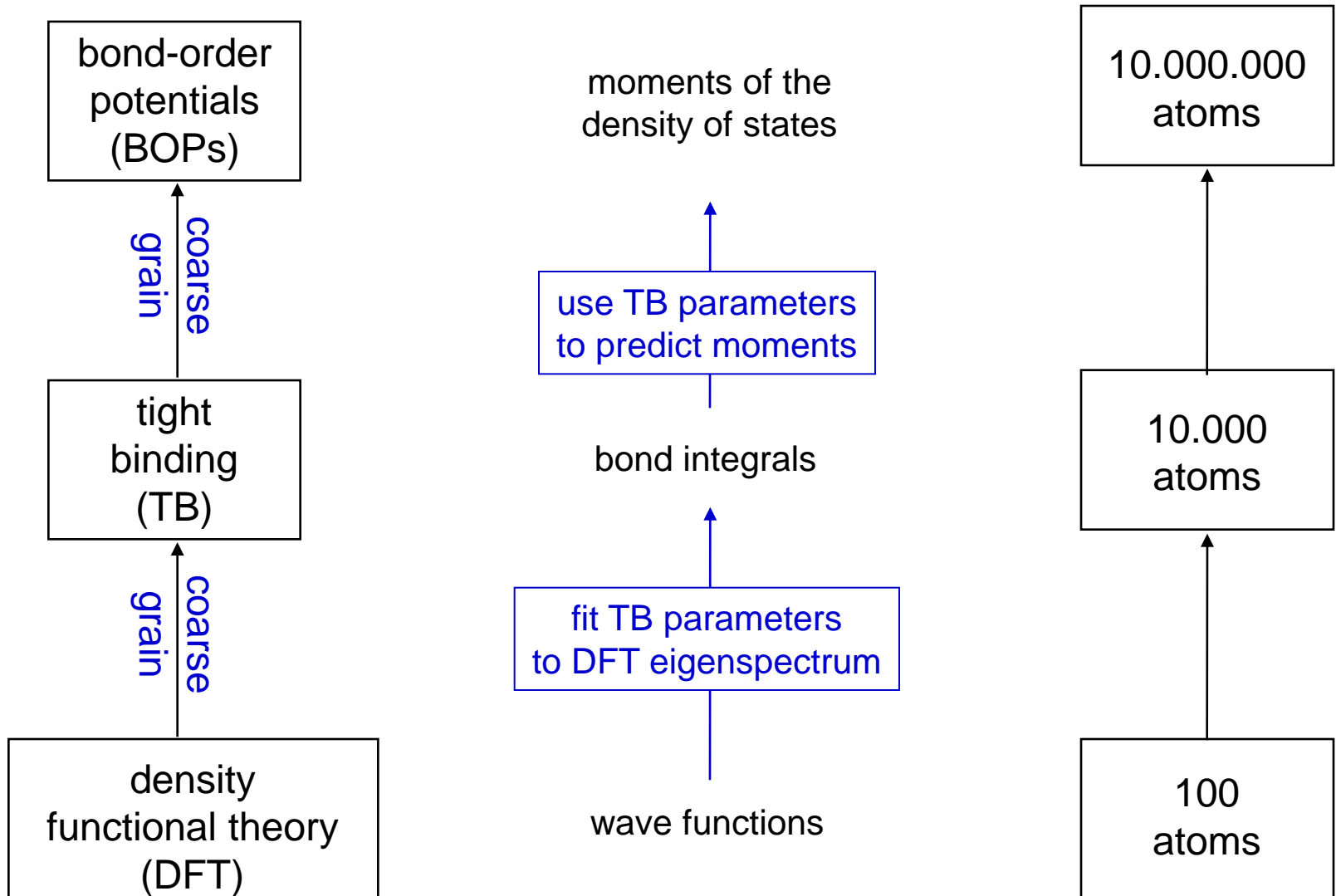
magnetic phase-transitions in Fe



edge dislocation in superalloys



Coarse-graining from DFT to TB to BOP



Tight-binding

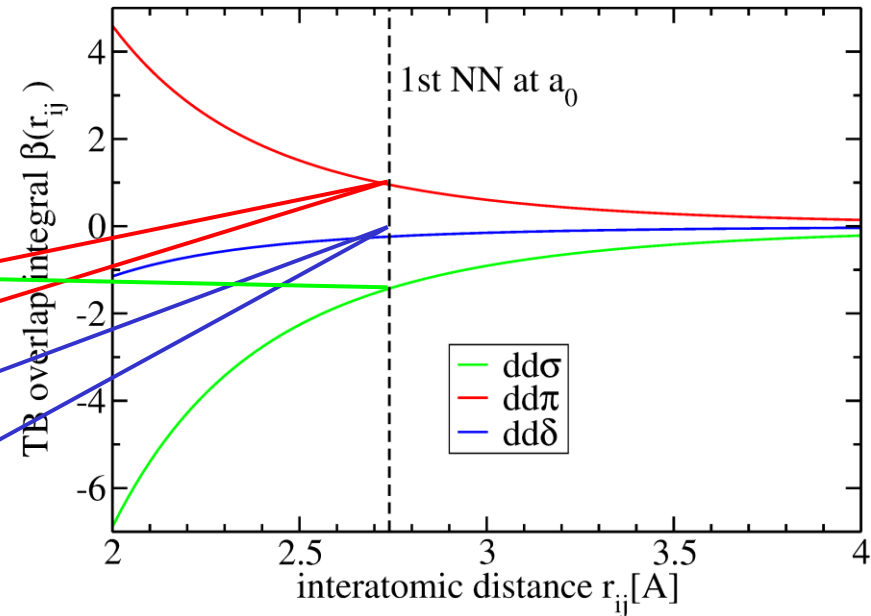
total energy in tight-binding bond model

Sutton et al, J Phys C 1988

$$U_B = U_{\text{bond}} + U_{\text{prom}} + U_{\text{ion}} + U_{\text{es}} + U_{\text{rep}} + U_{\text{mag}}$$

distance-dependent matrix-elements for pairs of interacting atoms

$$H_{ij}^{(b)} = \begin{array}{c} is \\ ip \\ id \end{array} \begin{array}{c} js \\ jp \\ jd \end{array} \begin{array}{c} \left(\begin{array}{ccc|ccc|ccc} \sigma & \sigma & 0 & 0 & \sigma & 0 & 0 & 0 & 0 \\ \sigma & \sigma & 0 & 0 & \sigma & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \pi & 0 & 0 & \pi & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi & 0 & 0 & \pi & 0 & 0 \\ \hline \sigma & \sigma & 0 & 0 & \sigma & 0 & 0 & 0 & 0 \\ 0 & 0 & \pi & 0 & 0 & \pi & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi & 0 & 0 & \pi & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta \end{array} \right) \end{array}$$



eigenspectrum from diagonalisation of system-wide Hamiltonian

Lanczos recursion algorithm

transformation of TB Hamiltonian to new basis

$$\{|u_0\rangle, |u_1\rangle, |u_2\rangle, \dots\}$$

start at atomic-like state $|u_0\rangle$ and generate $|u_{n+1}\rangle$ recursively using

$$b_1|u_1\rangle = (\hat{H} - a_0)|u_0\rangle \quad b_{n+1}|u_{n+1}\rangle = (\hat{H} - a_n)|u_n\rangle - b_n|u_{n-1}\rangle$$

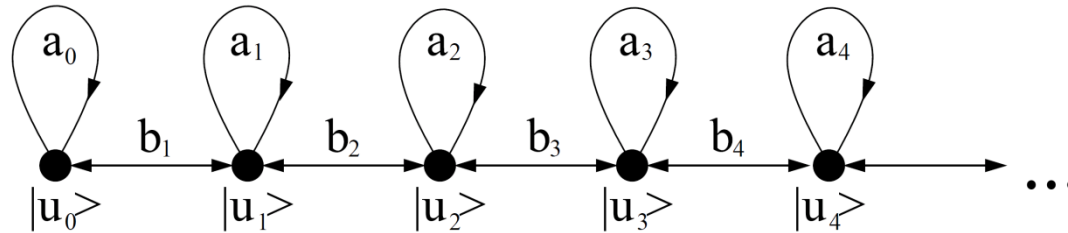
in this basis TB Hamiltonian takes tri-diagonal form

$$\langle u_n | \hat{H} | u_m \rangle = \begin{pmatrix} a_0 & b_1 & & & & \\ b_1 & a_1 & b_2 & & & \\ & b_2 & a_2 & b_3 & & \\ & & b_3 & a_3 & b_4 & \\ & & & b_4 & a_4 & \ddots \\ & & & & \ddots & \ddots & \ddots \\ & & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

Haydock, Comp Phys Comm 1980
Lanczos, J Res Natl Bur Stand 1950

Continued fraction

interpretation: 1d Lanczos chain with only nearest-neighbor matrix elements



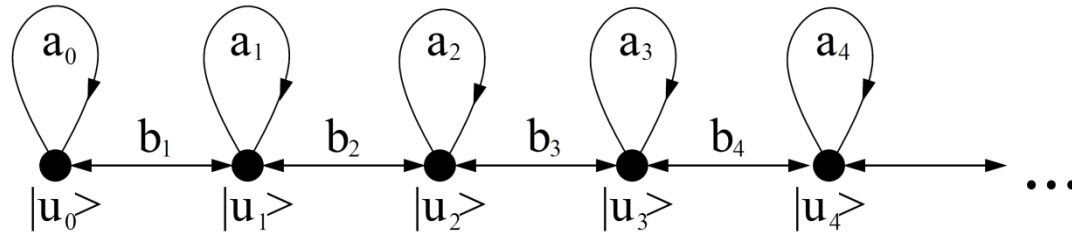
inversion of tridiagonal Hamiltonian by Greens functions yields continued fraction

$$n_{i\alpha}(E) = -\frac{1}{\pi} \text{Im} \frac{1}{E - a_0 - \frac{b_1^2}{E - a_1 - \frac{b_2^2}{E - a_2 - \frac{b_3^2}{\ddots}}}}$$

truncation after certain number of recursion levels \rightarrow linear-scaling BOP

Connection to moments

interpretation: 1d chain with only nearest-neighbor matrix elements



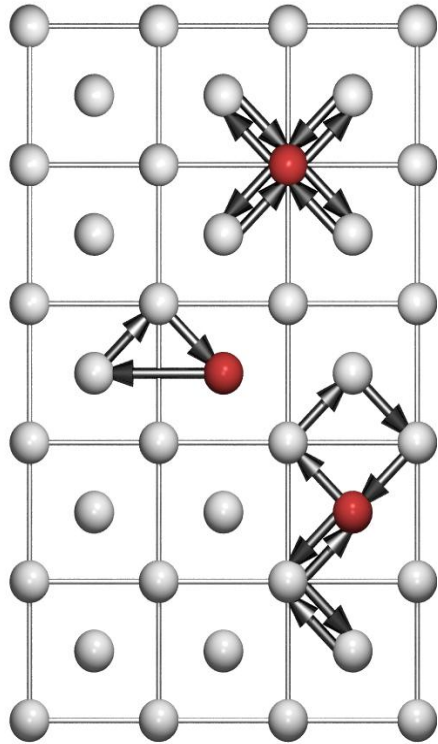
choose atomic orbital as starting orbital of Lanczos chain

$$\begin{aligned}\mu_{ilm}^{(n)} &= \langle \varphi_{ilm} | \hat{H}^n | \varphi_{ilm} \rangle = \langle u_0 | \hat{H}^n | u_0 \rangle \\ &= \sum_{i_1 \dots i_{n-1}} \langle u_0 | \hat{H} | u_{i_1} \rangle \langle u_{i_1} | \hat{H} | u_{i_2} \rangle \cdots \langle u_{i_{n-1}} | \hat{H} | u_0 \rangle\end{aligned}$$

direct relation of recursion coefficients to moments

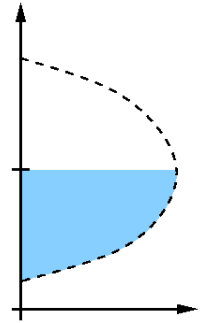
$$\begin{aligned}\mu_{ilm}^{(1)} &= a_0 \\ \mu_{ilm}^{(2)} &= a_0^2 + b_1^2 \\ \mu_{ilm}^{(3)} &= a_0^3 + 2a_0b_1^2 + a_1b_1^2\end{aligned}$$

Moments expansion of DOS



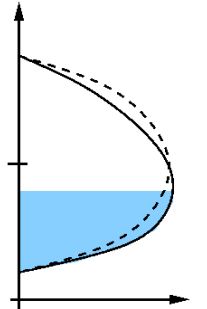
second moment (rms width)

$$\mu_i^{(2)} = \int \epsilon^2 n_i(\epsilon) d\epsilon = \sum_j \langle i | \hat{H} | j \rangle \langle j | \hat{H} | i \rangle$$



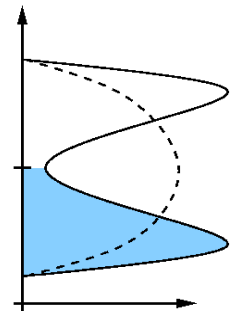
third moment (skewing)

$$\mu_i^{(3)} = \int \epsilon^3 n_i(\epsilon) d\epsilon = \sum_{jkl} \langle i | \hat{H} | j \rangle \langle j | \hat{H} | k \rangle \langle k | \hat{H} | i \rangle$$



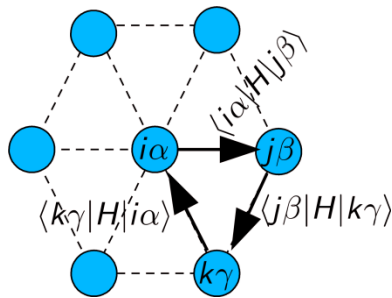
fourth moment (bimodality)

$$\mu_i^{(4)} = \int \epsilon^4 n_i(\epsilon) d\epsilon = \sum_{jkl} \langle i | \hat{H} | j \rangle \langle j | \hat{H} | k \rangle \langle k | \hat{H} | l \rangle \langle l | \hat{H} | i \rangle$$



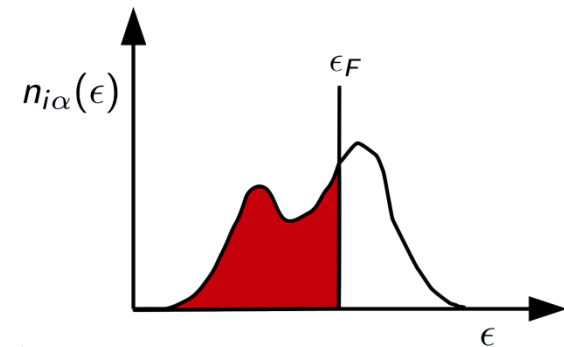
Moments link electronic and crystal structure

crystal structure



$$\mu_{i\alpha}^{(n)} = \langle i\alpha | H^n | i\alpha \rangle$$

electronic structure



$$\mu_{i\alpha}^{(n)} = \int \epsilon^n n_{i\alpha}(\epsilon) d\epsilon$$

bond energy

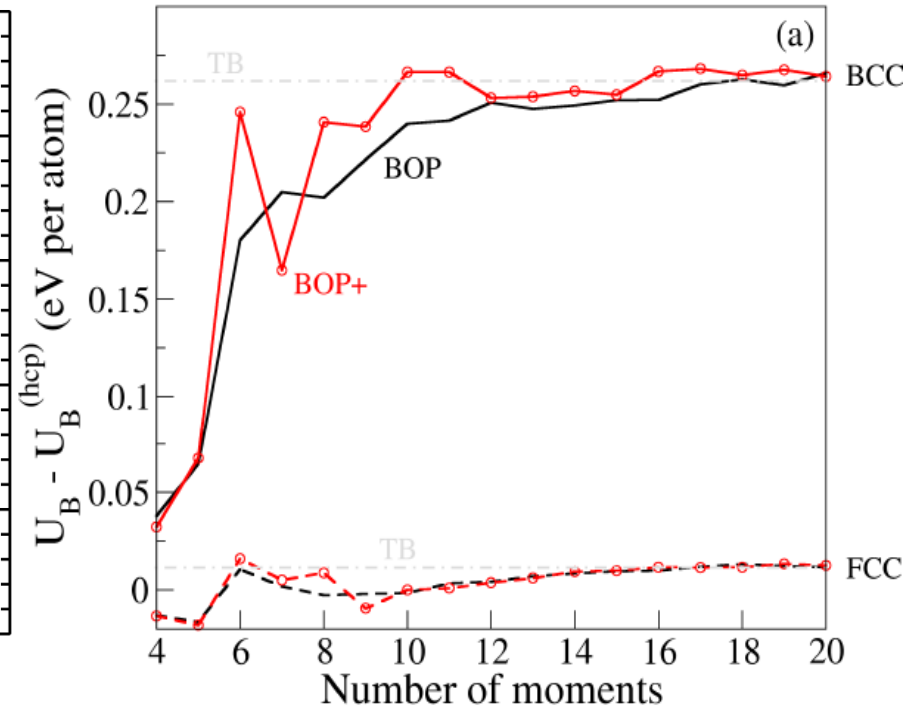
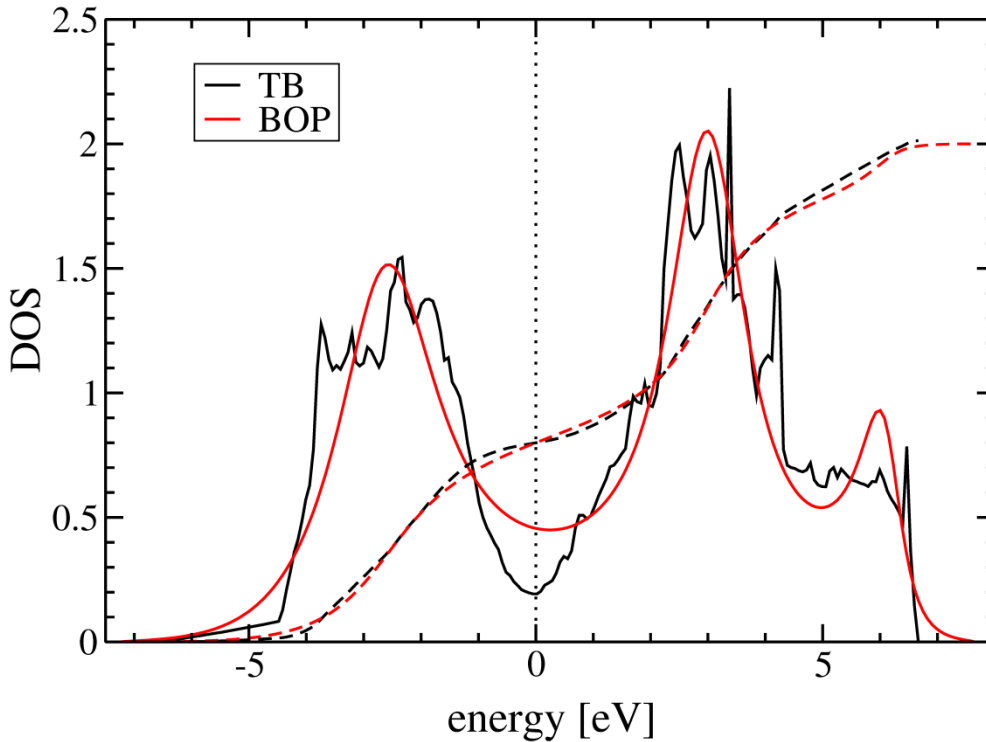
$$U_{bond} = \sum_{i\alpha} \int^{\epsilon_F} \epsilon n_{i\alpha}(\epsilon) d\epsilon$$

Cyrot-Lackmann,
Adv. Phys. 1967

Convergence with moments

bcc-Ta with 9 moments

magnetic bcc-Fe



rapid and robust convergence

TH, et al., Comp Phys Comm, 2019

Ford, et al., Mod Sim Mat Sci Eng, 2014

Moments based descriptors

normalisation of lowest moments

$$\mu_{i\alpha}^{(0)} = 1 \quad \mu_{i\alpha}^{(1)} = 0 \quad \mu_{i\alpha}^{(2)} = 1$$

rotation invariance

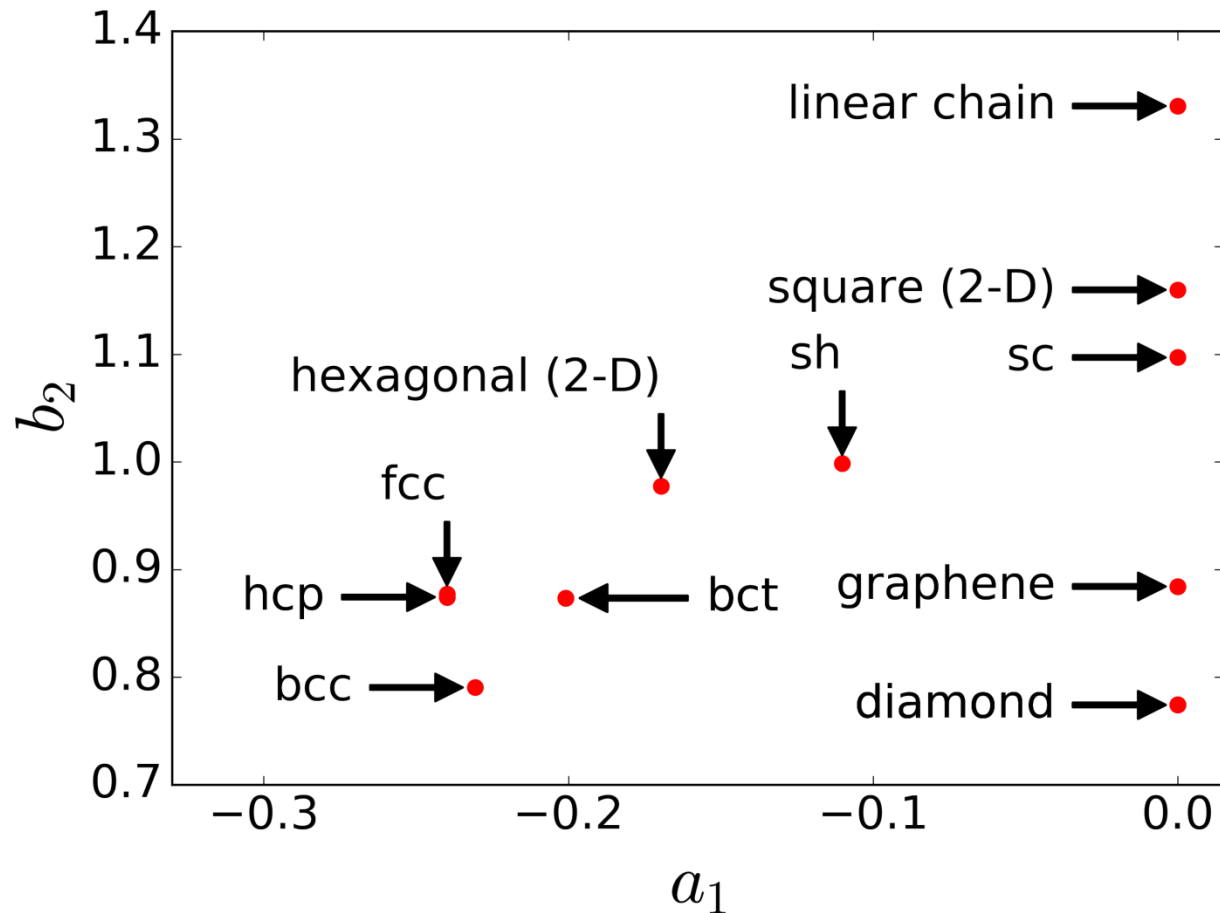
$$\mu_i^{(N)} = \frac{1}{N_{orb}} \sum_{\alpha} \mu_{i\alpha}^{(N)}$$

independent descriptors based on 3rd and 4th moment

$$\mathbf{a}_{i,1} = \mu_i^{(3)} \quad \mathbf{b}_{i,2} = \sqrt{\mu_i^{(4)} - \left(\mu_i^{(3)}\right)^2 + 1}$$

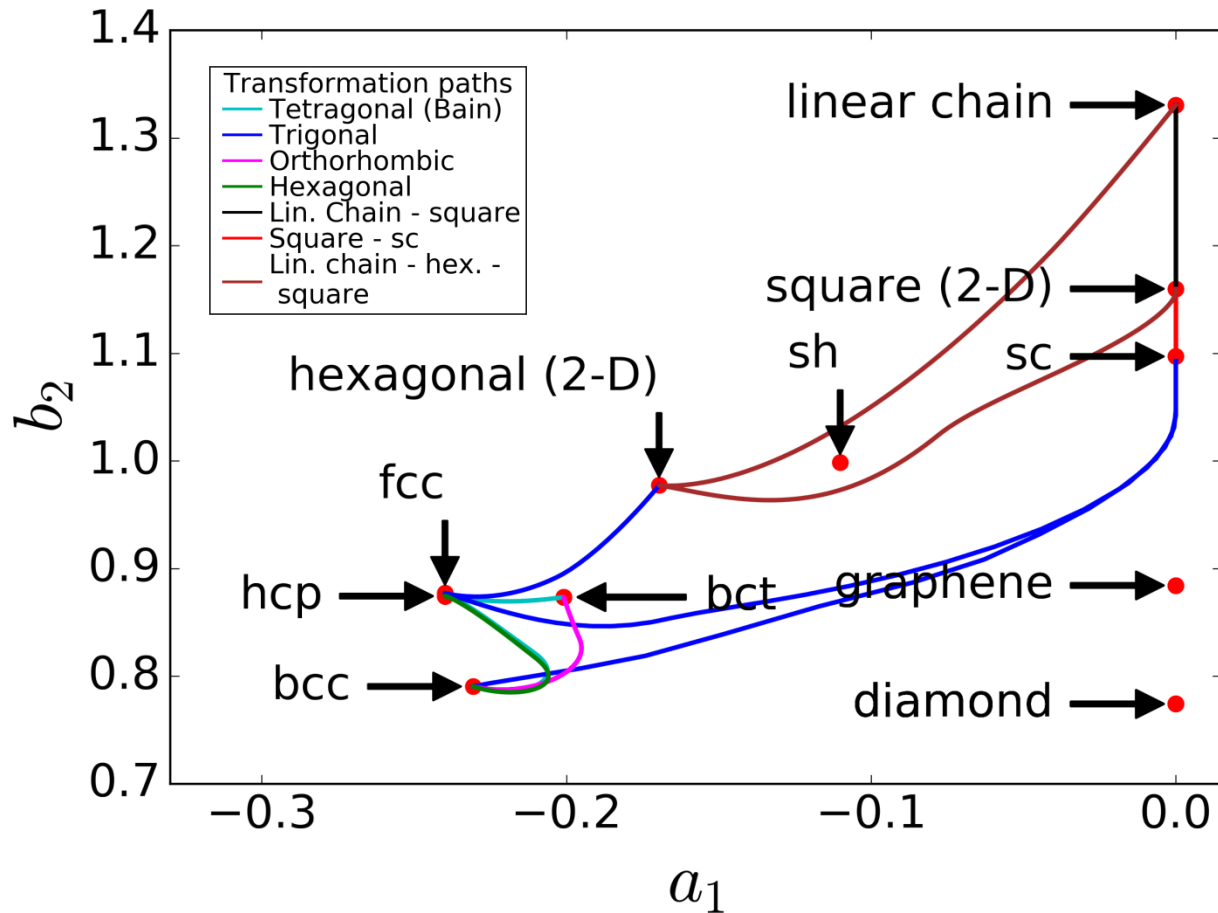
electronic-structure based descriptor of local atomic environment

Low-dimensional map of structural similarity



clear separation of typical 2D and 3D crystal structures

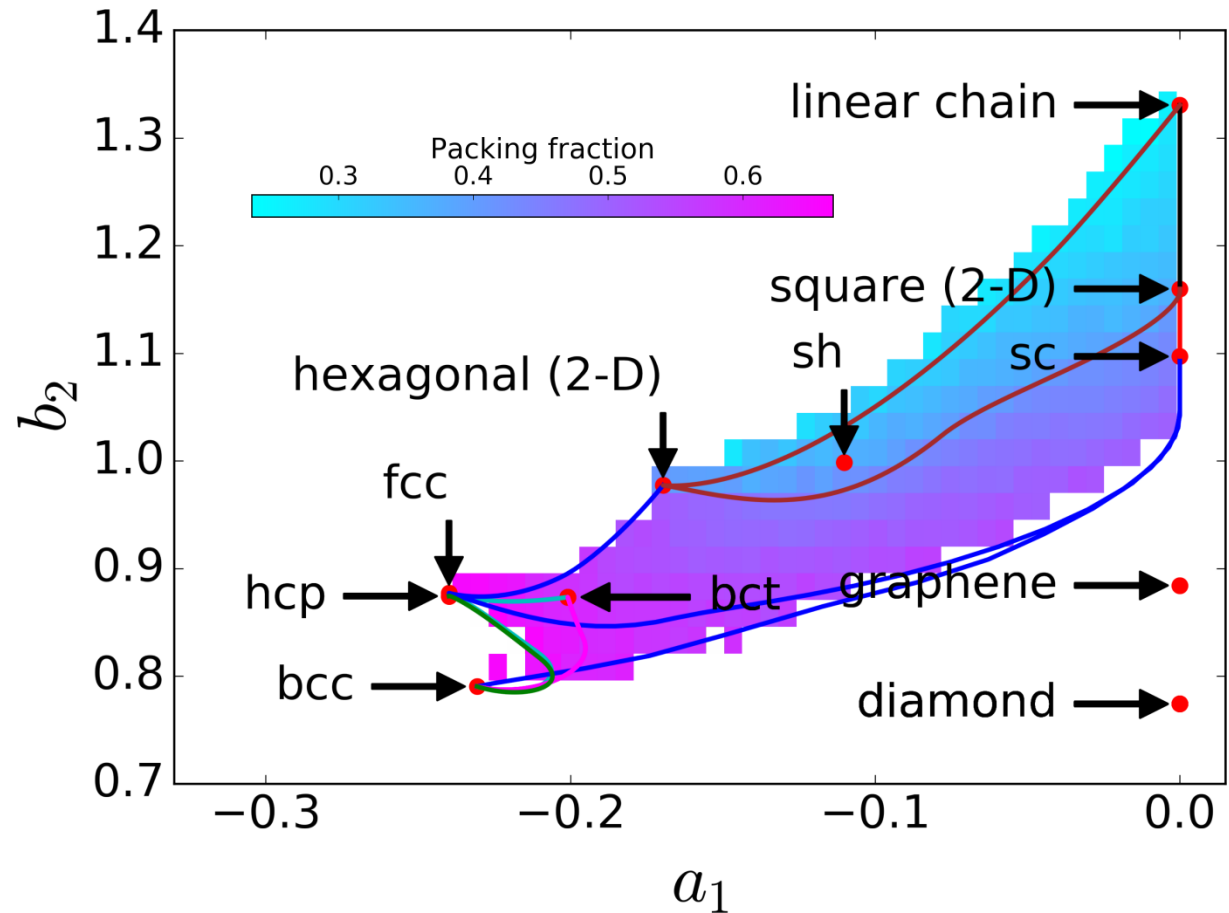
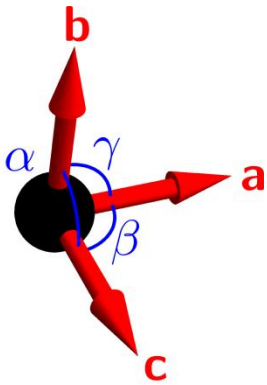
Transformation paths



transformation paths mark region of simple structures

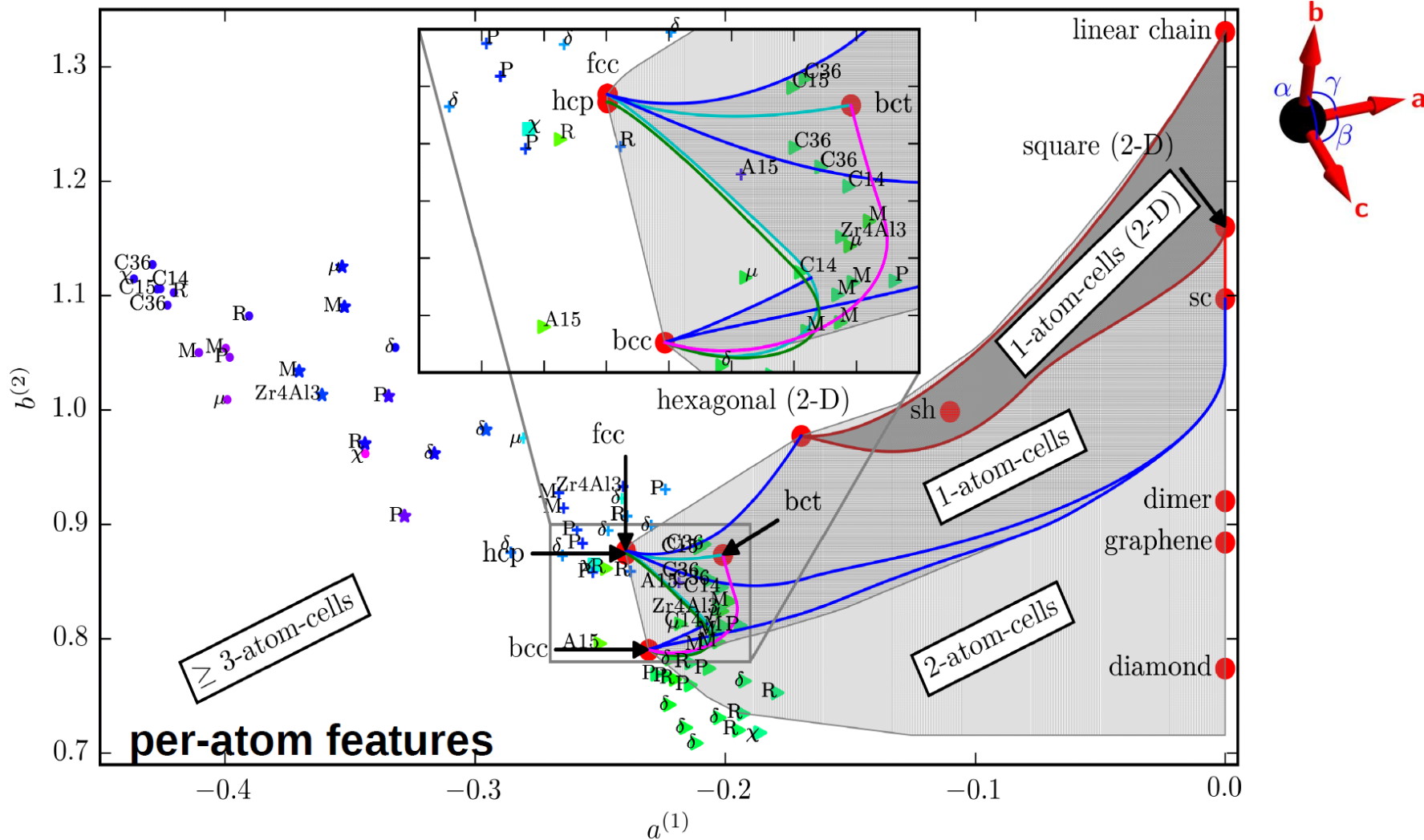
Random structures with one Wyckoff site

random unit cell
1 Wyckoff position
90000 structures



random structures (1 atom) fill region of simple structures

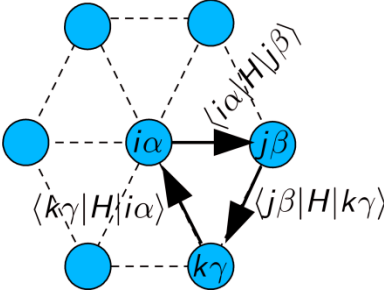
Map of local atomic environments



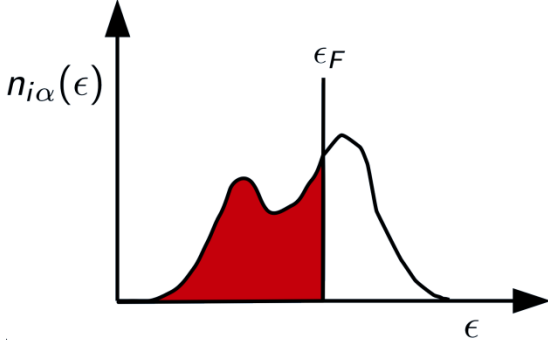
Jenke, Subramanyam, Densow, TH, Pettifor, Drautz, PRB 2018

Moments link electronic and crystal structure

crystal structure


$$\mu_{i\alpha}^{(n)} = \langle i\alpha | H^n | i\alpha \rangle$$

electronic structure


$$\mu_{i\alpha}^{(n)} = \int \epsilon^n n_{i\alpha}(\epsilon) d\epsilon$$

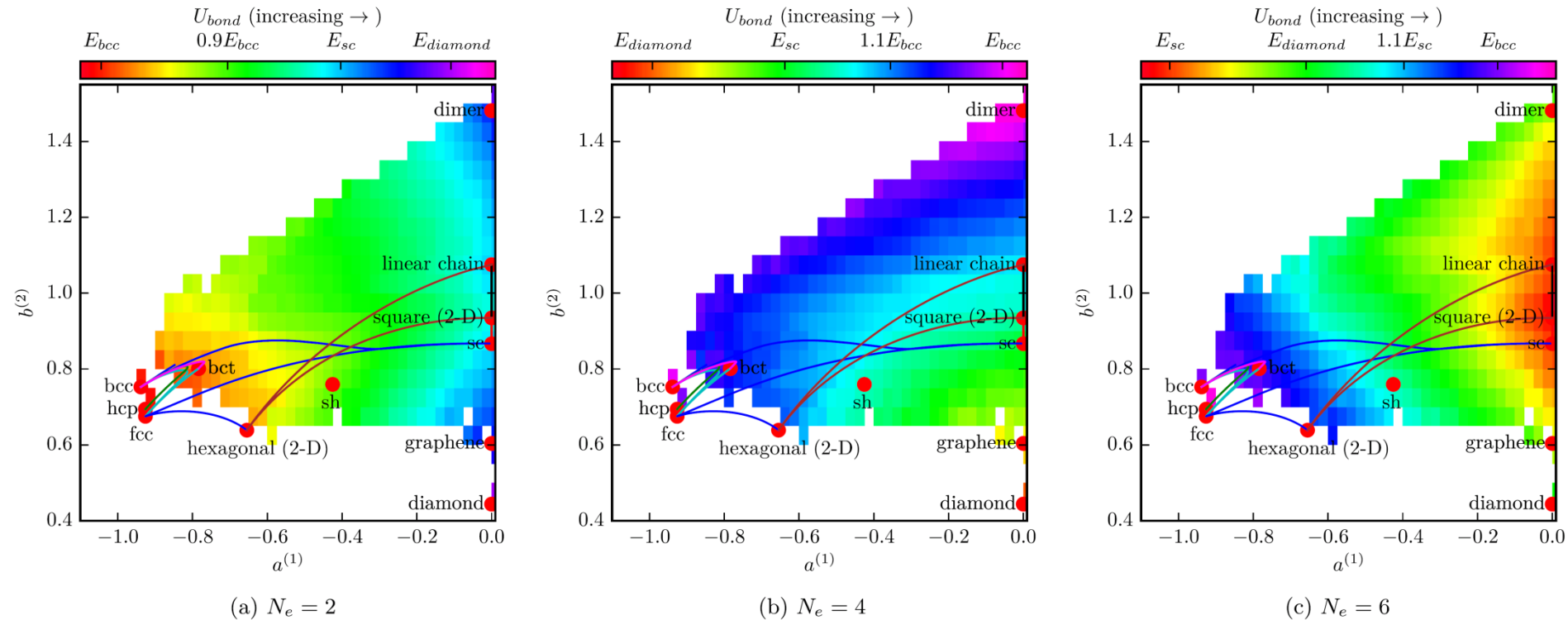
bond energy

$$U_{bond} = \sum_{i\alpha} \int^{\epsilon_F} \epsilon n_{i\alpha}(\epsilon) d\epsilon$$

Cyrot-Lackmann,
Adv. Phys. 1967

Influence of bandfilling on structural stability

canonical sp-valent TB with variation of valence-electron number N_e

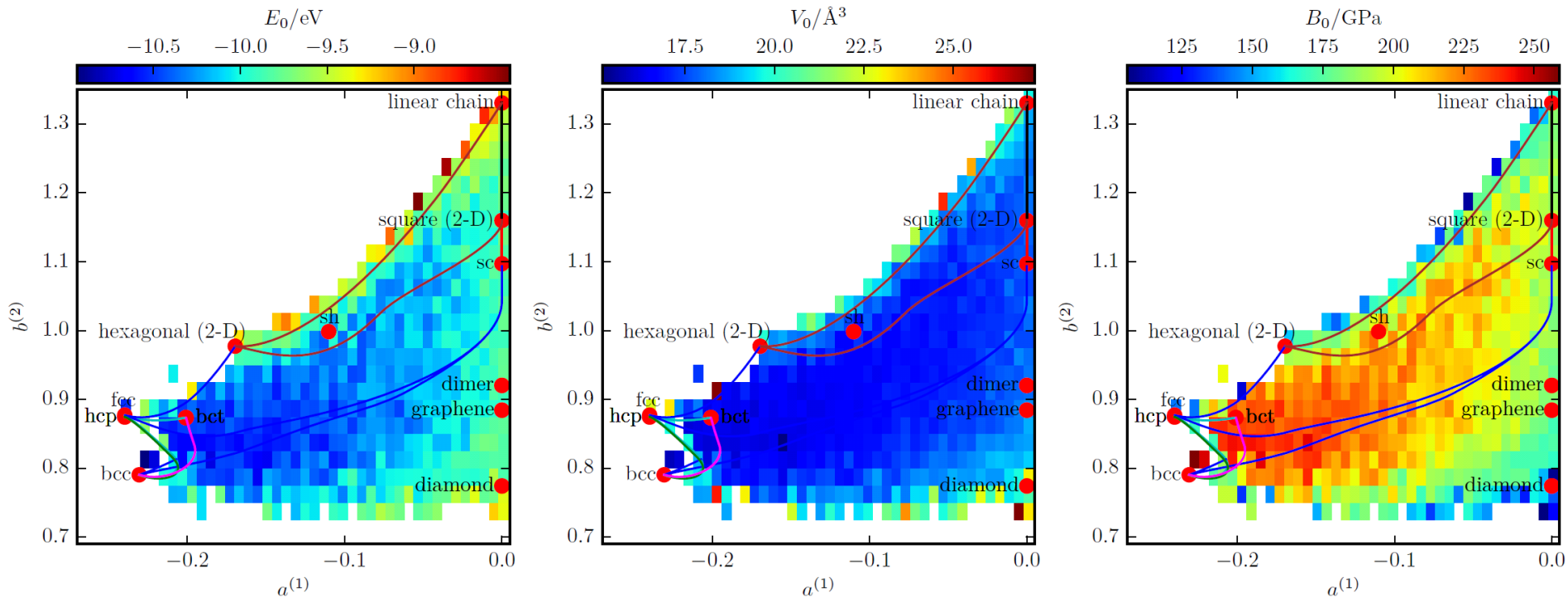


structural trend with band filling incorporated in descriptor

Jenke, Subramanyam, Densow, TH, Pettifor, Drautz, PRB 2018

Comparison to DFT calculations

Mo in 1-atom and 2-atom random structures (VASP with PBE, PAW)



smooth trend of E_0 , V_0 , B_0 \rightarrow domain knowledge of interatomic bond

Jenke, Subramanyam, Densow, TH, Pettifor, Drautz, PRB 2018

Application in machine learning

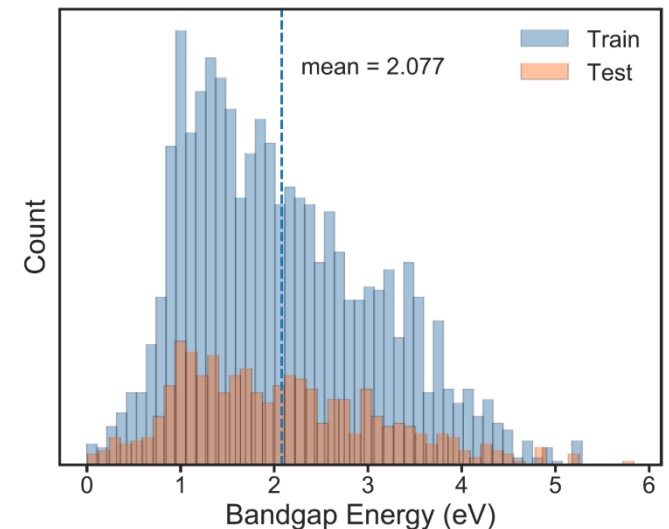
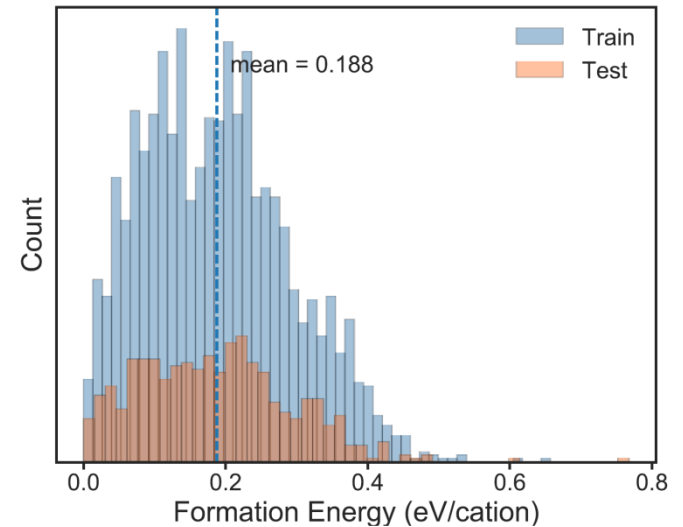


transparent conductors: $(\text{Al}_x\text{Ga}_y\text{In}_z)_2\text{O}_3$

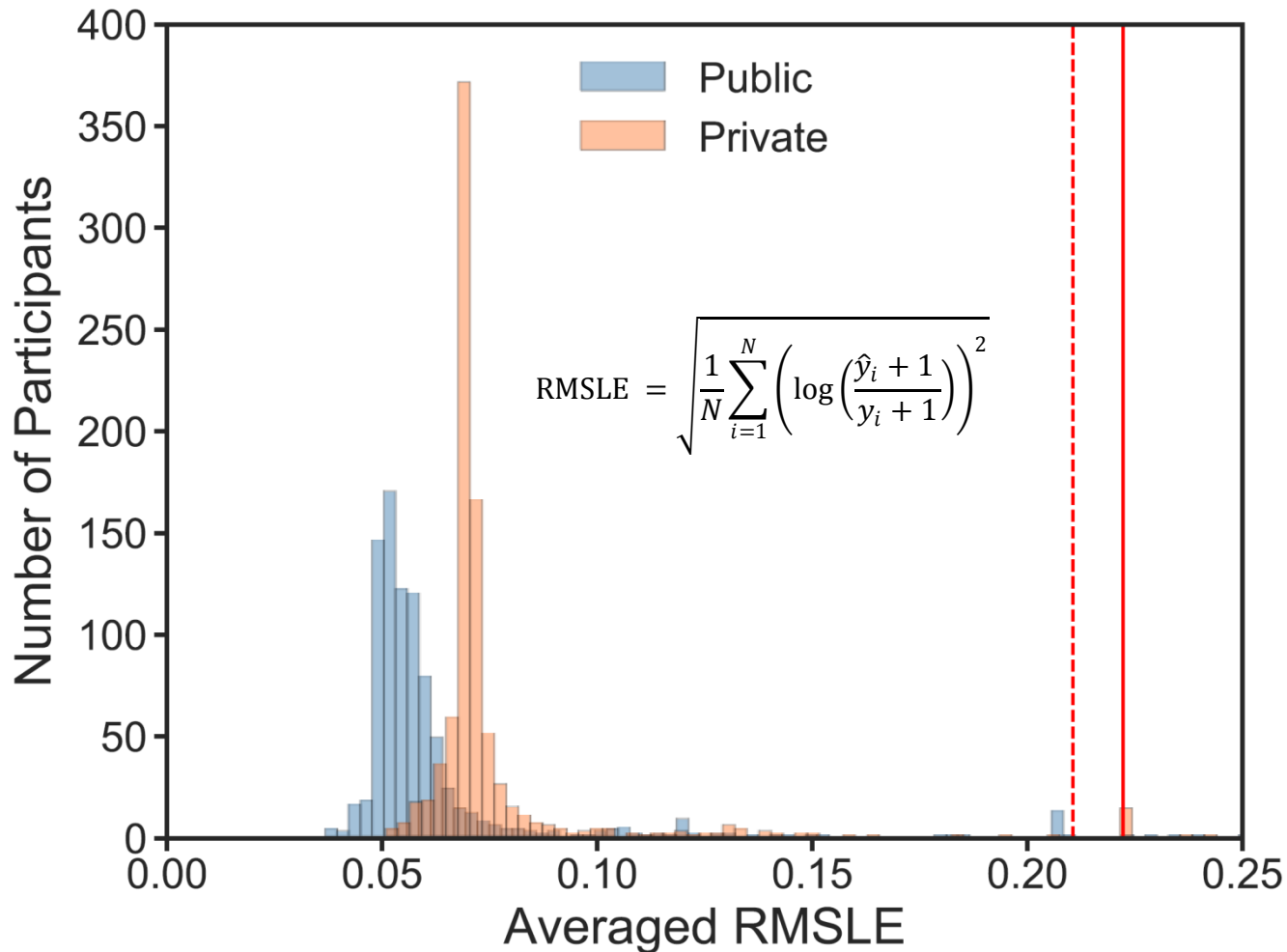
challenge: formation energy and bandgap

DFT: 2400 train (public), 600 test (secret)

~ 900 participants/teams



Final results of competition for test set



Sutton, Ghiringhelli, Yamamoto, Lysogorskiy, Blumenthal, TH, Golebiowski, Liu, Ziletti, Scheffler

npj Comp Mat 2019

Final results of competition for test set

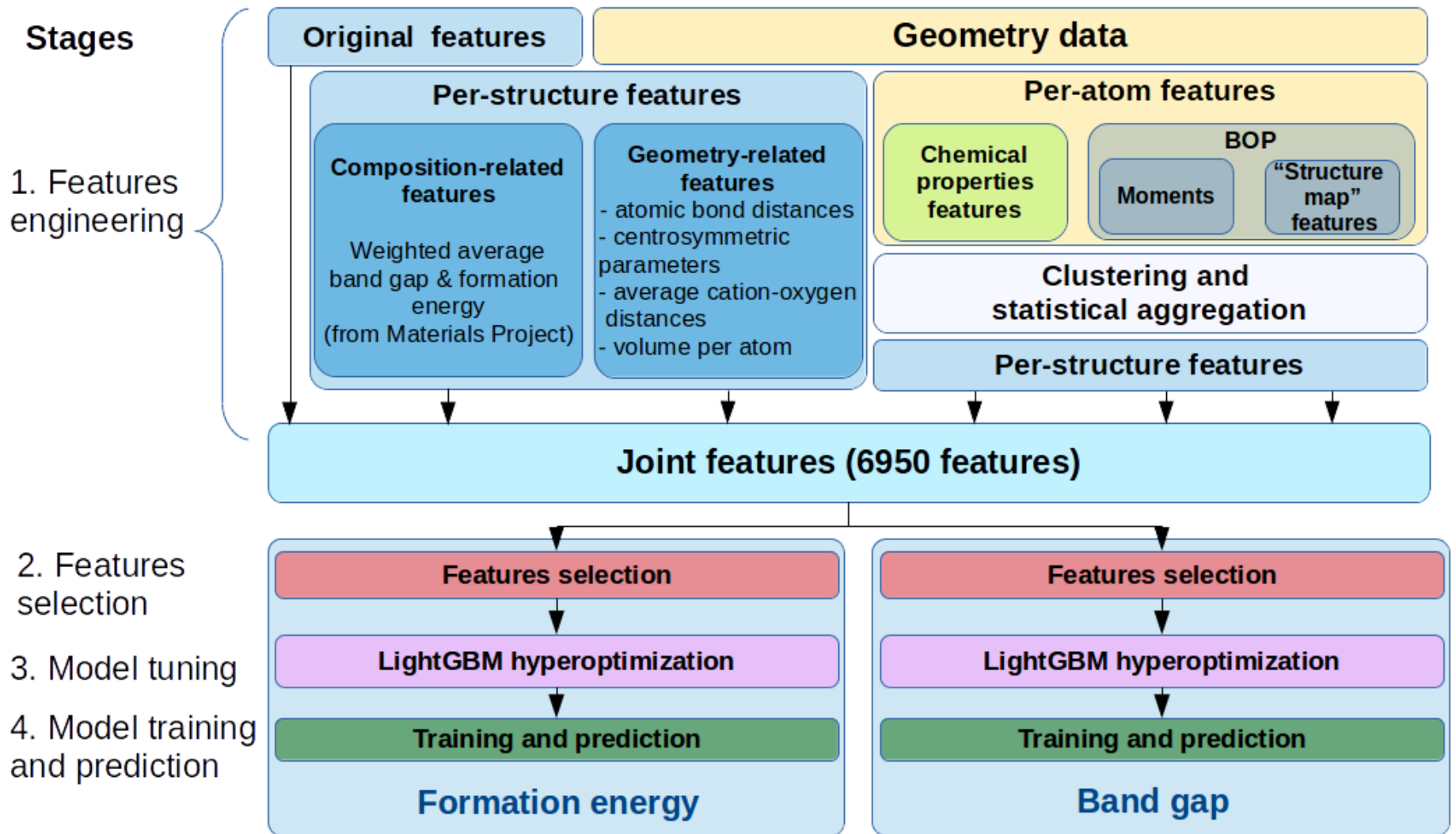


Ranking	ML representation + regression method	Band gap energy		Formation energy	
		Root mean square log error	Mean absolute error (meV)	Root mean square log error	Mean absolute error (meV/cation)
1st	<i>n</i> -gram+KRR	0.077	114	0.021	15
2nd	c/BOP+LGBM	0.081	93	0.022	15
3rd	SOAP+NN	0.081	98	0.021	13

Sutton, Ghiringhelli, Yamamoto, Lysogorskiy, Blumenthal, TH, Golebiowski, Liu, Ziletti, Scheffler

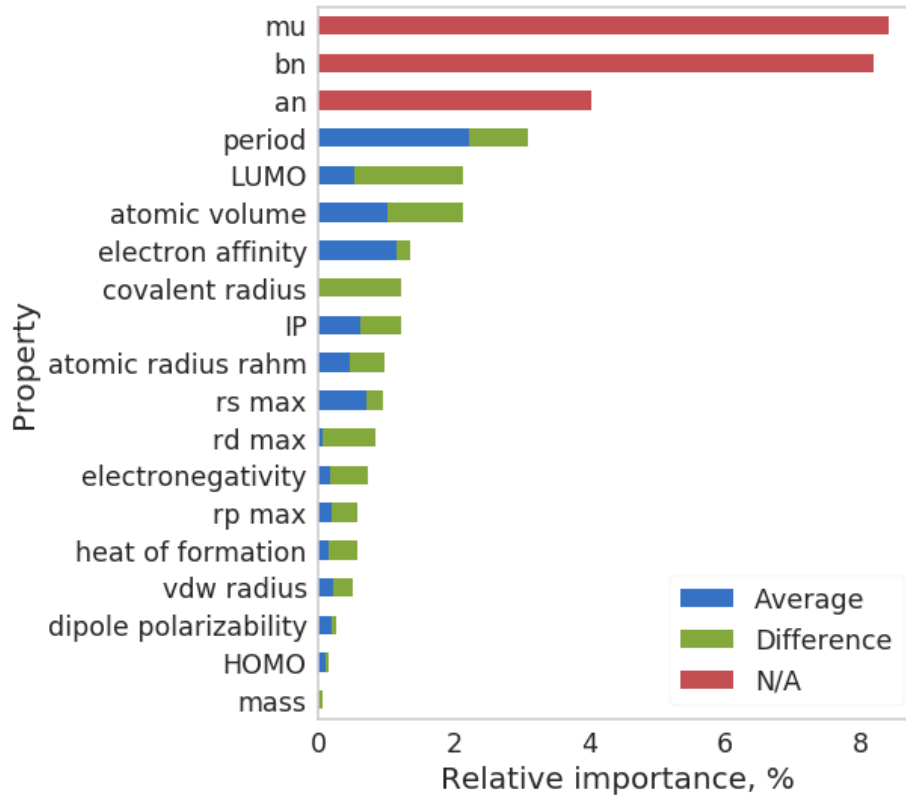
npj Comp Mat 2019

c/BOP+LGBM solution

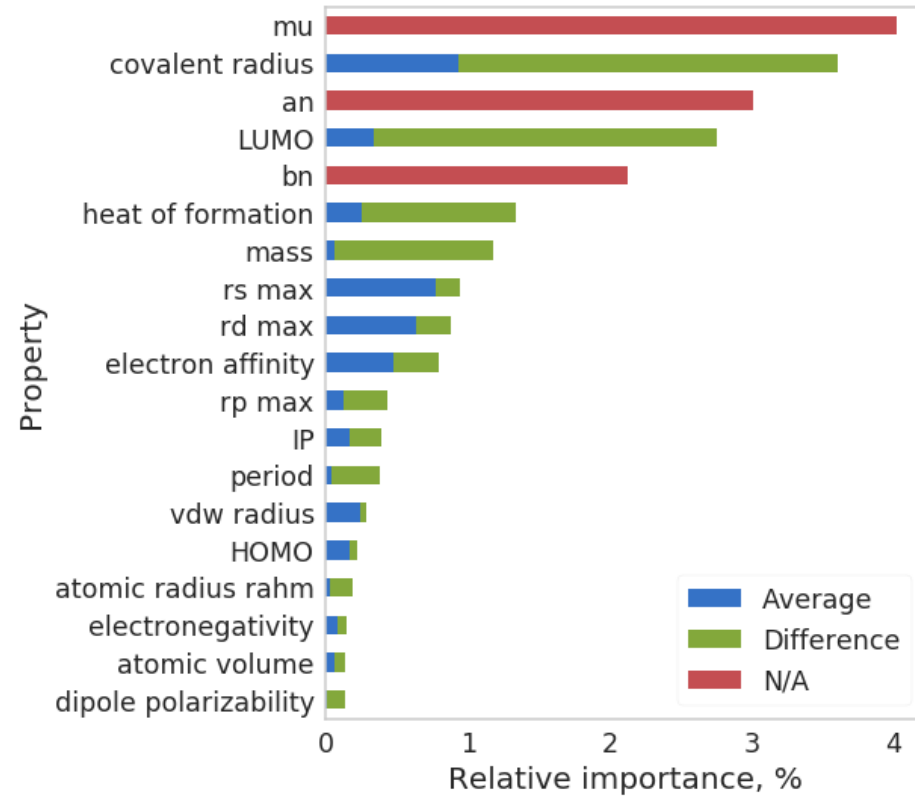


Relative feature importance

formation energy



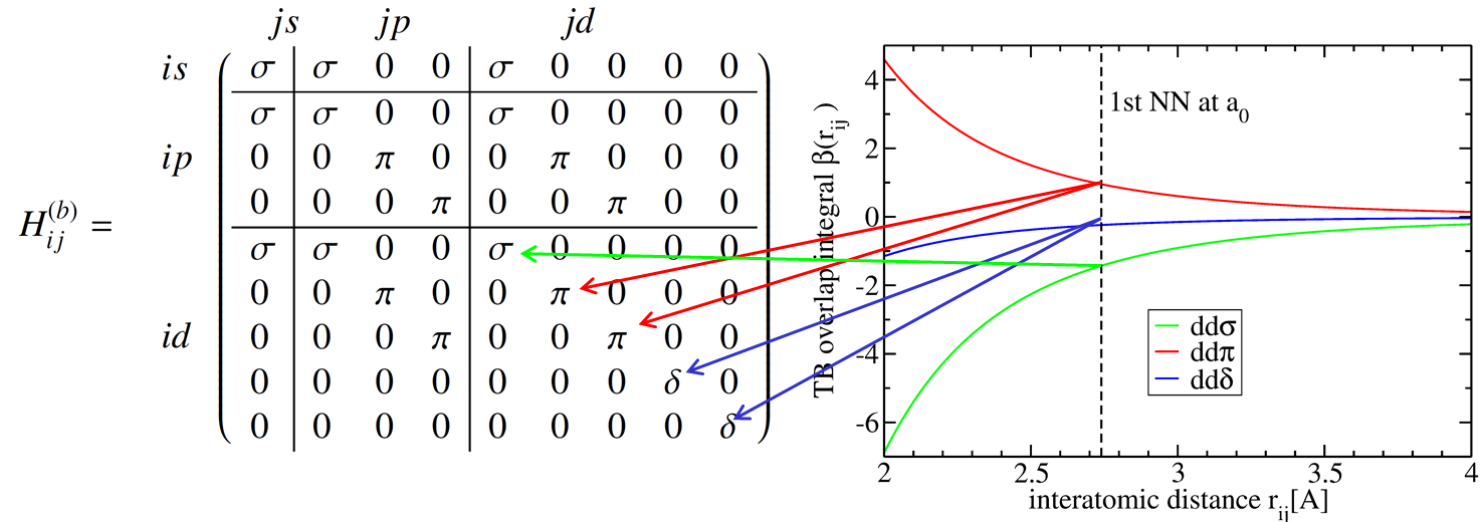
band gap



highest importance for electronic-structure based descriptors mu, an, bn

More chemistry by bond-specific Hamiltonians

description of chemistry: d-valent and sp-valent canonical TB Hamiltonians



$$\left. \begin{array}{l} dd\sigma \\ dd\pi \\ dd\delta \end{array} \right\} = \begin{array}{l} -6 \\ 4 \\ -1 \end{array} \beta(r) \quad \left. \begin{array}{l} ss\sigma \\ sp\sigma \\ pp\sigma \\ pp\pi \end{array} \right\} = \begin{array}{l} -1.00 \\ 1.31 \\ 2.31 \\ -0.76 \end{array} \beta(r) \quad \beta(r) = c/r^5$$

Downfold DFT eigenspectra to TB minimal basis

downfolding procedure:

Dimer DFT eigenstates ψ_n
for different distances R_{IJ}

Optimize projection

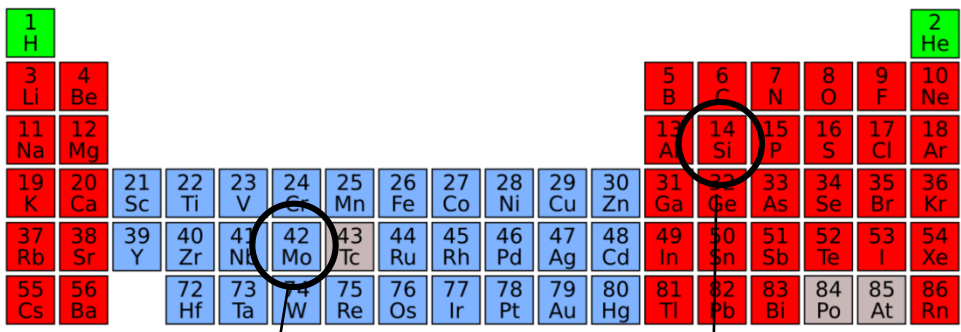
$$P = \frac{1}{N_e} \sum_n f_n \sum_{Jlm} \langle \psi_n | \varphi^{Jlm} \rangle \langle \varphi^{Jlm} | \psi_n \rangle$$
 on TB minimal basis

Construct TB matrix elements
 $H_{IJ} = \{ss\sigma, sd\sigma, dd\sigma, dd\pi, dd\delta, \dots\}$

Parameterize TB matrix elements

$$H_{IJ}(R_{IJ}) = \sum_i c_i \exp(-\lambda_i R_{IJ}^{n_i})$$

applied to:



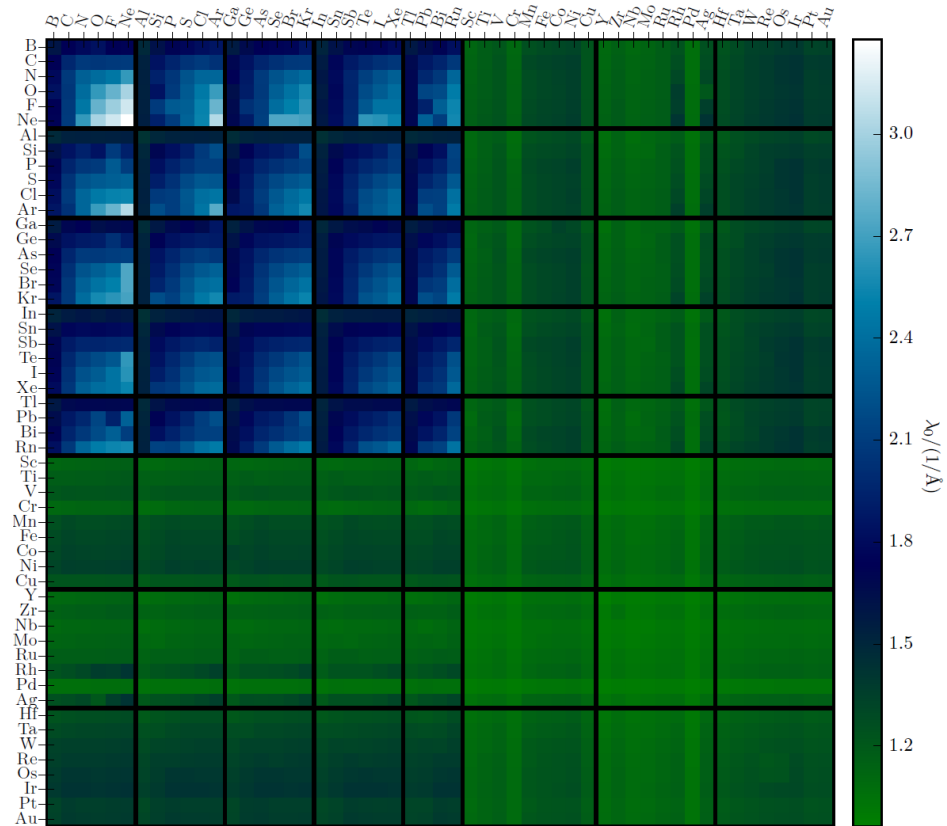
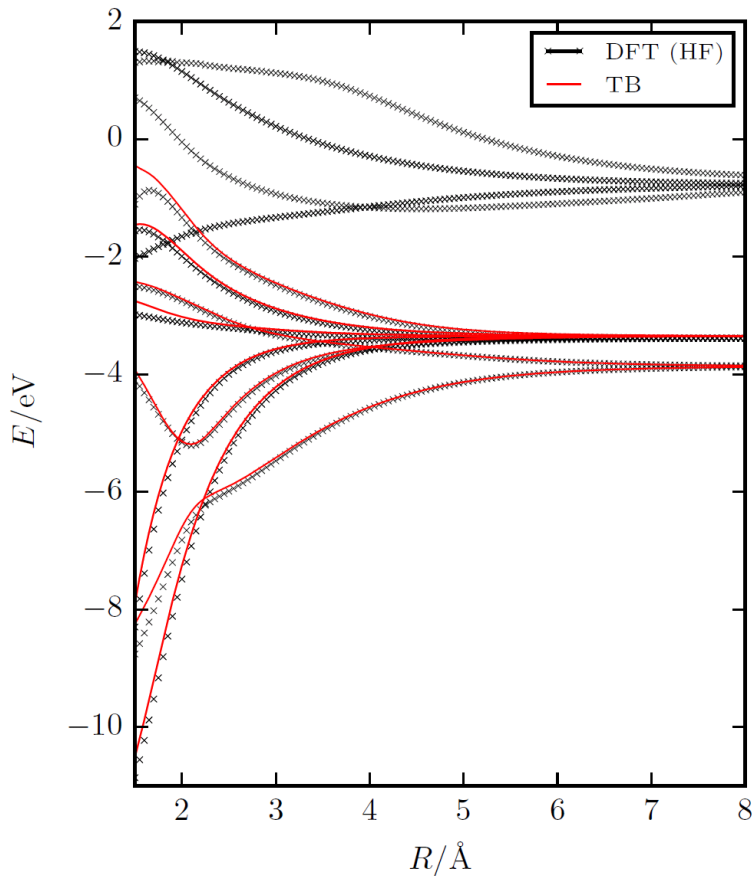
Madsen et al., PRB 2011
 Urban et al., PRB 2011

Jenke et al.
 (submitted)

Database of bond-specific H for dimers

Mo-Mo dimer

database of pairwise TB parameters

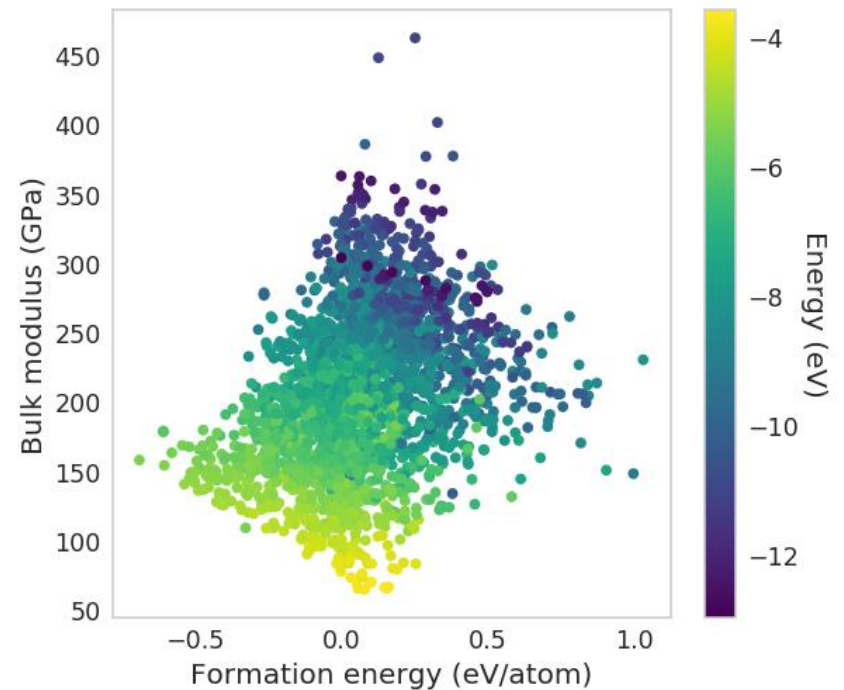
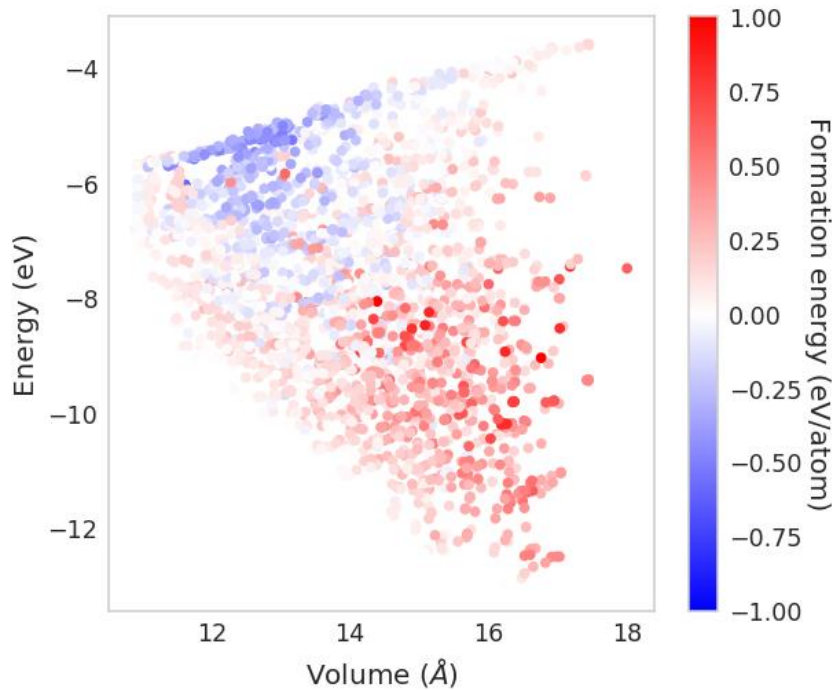


recover canonical TB models

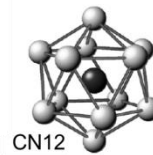
Jenke et al. (submitted)

Application to intermetallic phases

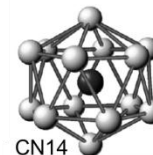
A15, C14, C15, C36, σ , χ , μ phases in ternary Ni-Al-Re and subsystems



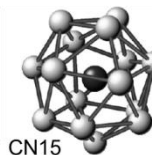
challenge: high structural similarity!



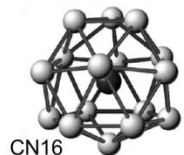
CN12



CN14

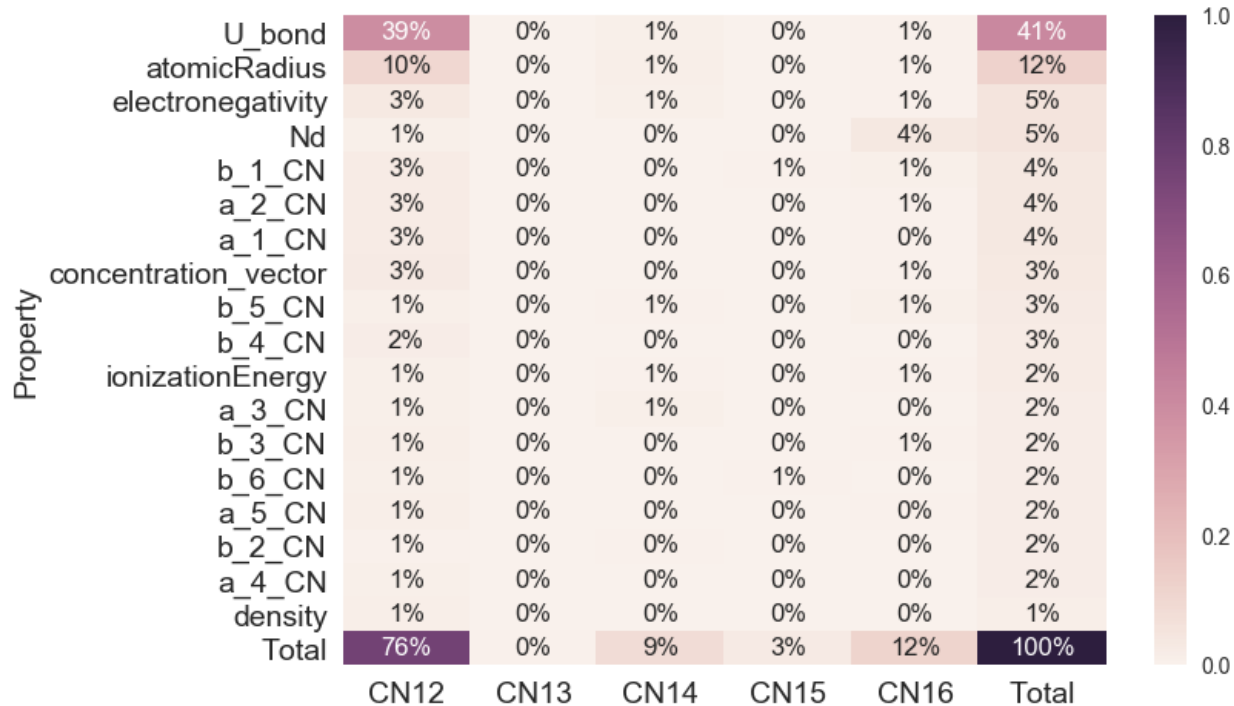
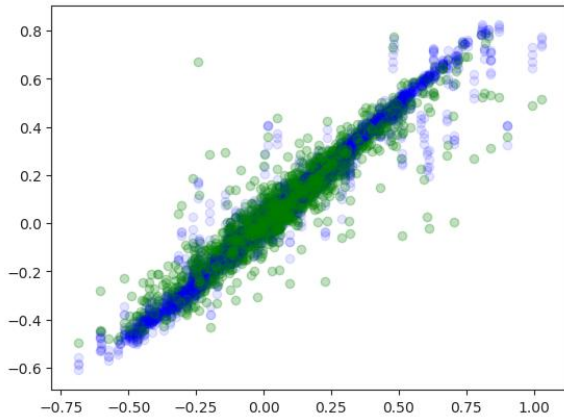


CN15



CN16

Prediction of formation energy (preliminary)



Kernel ridge regression

Train error (meV/at):

44 +/- 2

Test error (meV/at):

74 +/- 10

LGBM

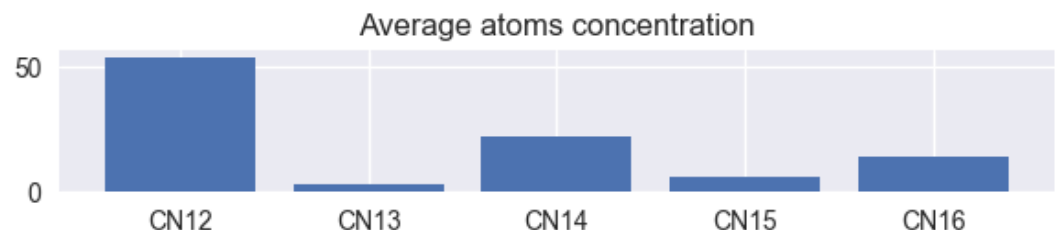
Train error (meV/at):

36.8 +/- 1.4

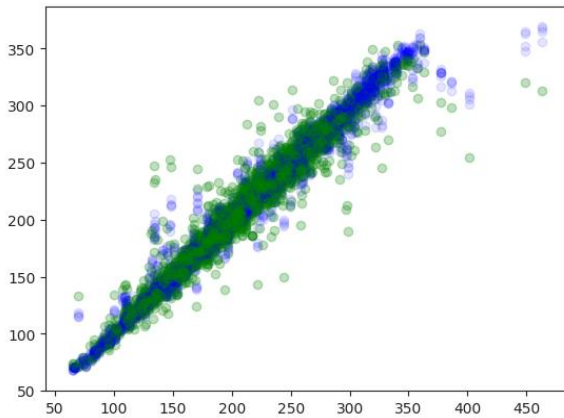
Test error (meV/at):

77 +/- 5

Maximum importance: bond energy, atomic radius



Prediction of bulk modulus (preliminary)



Kernel ridge regression

Train error (GPa):

13.10 +/- 0.38

Test error (GPa):

14.25 +/- 1.51

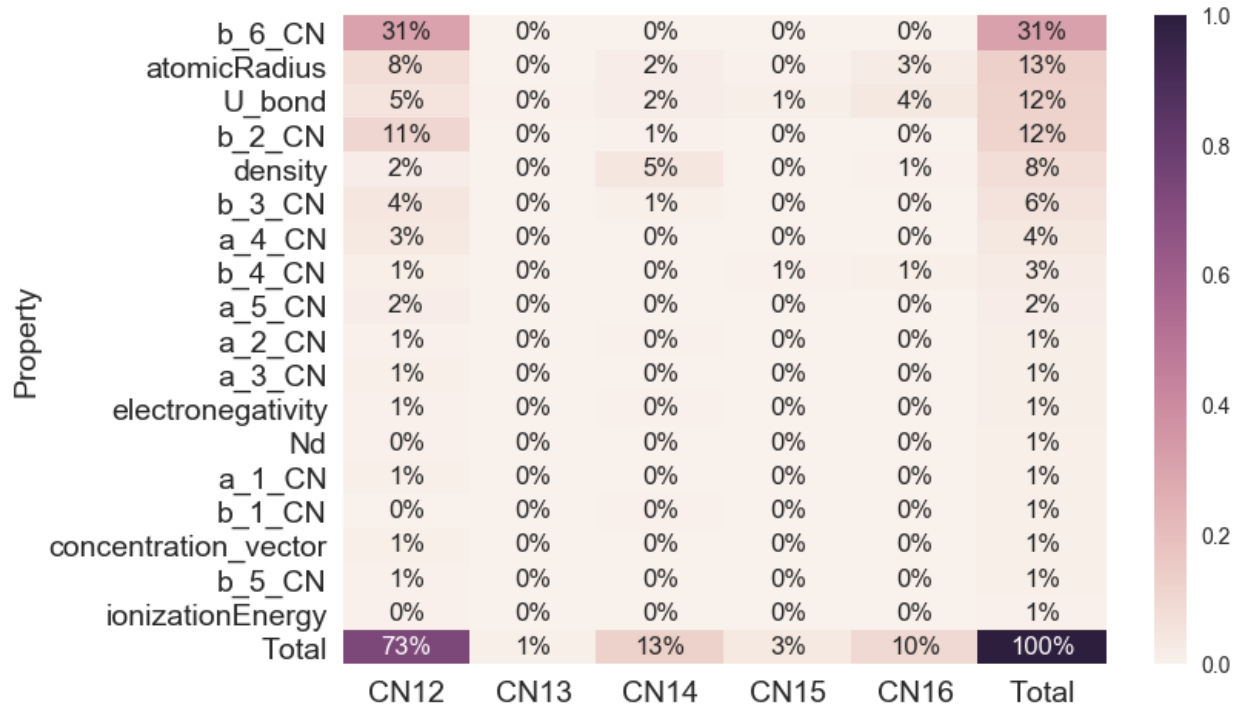
LGBM

Train error (GPa):

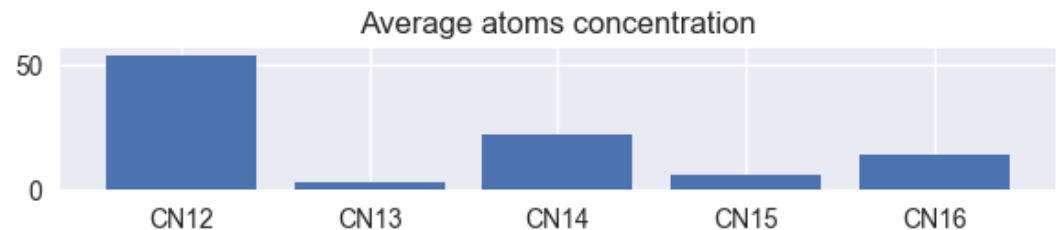
7.9 +/- 0.3

Test error (GPa):

14.9 +/- 1.7



Maximum importance: b6, atomic radius, bond energy, b2



Conclusion

electronic-structure based descriptors

- coarse-graining DFT → TB → BOP
- moments link atomic structure and electronic structure
- low-dimensional representation of local atomic environments
- complete and homogeneous samplings of atomic environments
- chemistry-aware descriptor includes valence-type and band-filling
- bond-specific pair-wise Hamiltonians from downfolding DFT eigenspectrum

applications

- smooth variation of formation energy, volume and bulk modulus
- formation energies and bandgaps of transparent oxides
- formation energies and bulk moduli of intermetallic phases
- high relative importance of descriptors with domain knowledge

funding: DFG, SFB/TR103, DAAD, IMPRS SurMat, EPSRC

