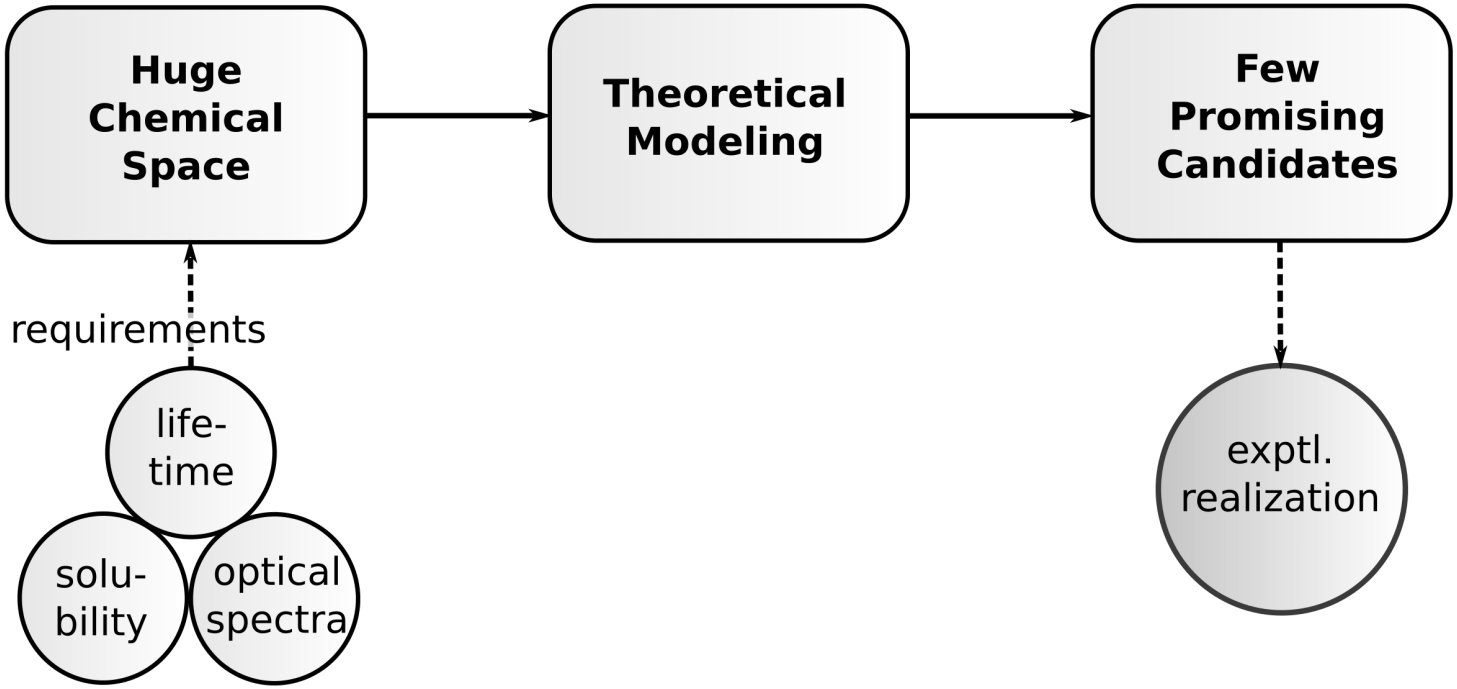
Towards the design of molecular materials

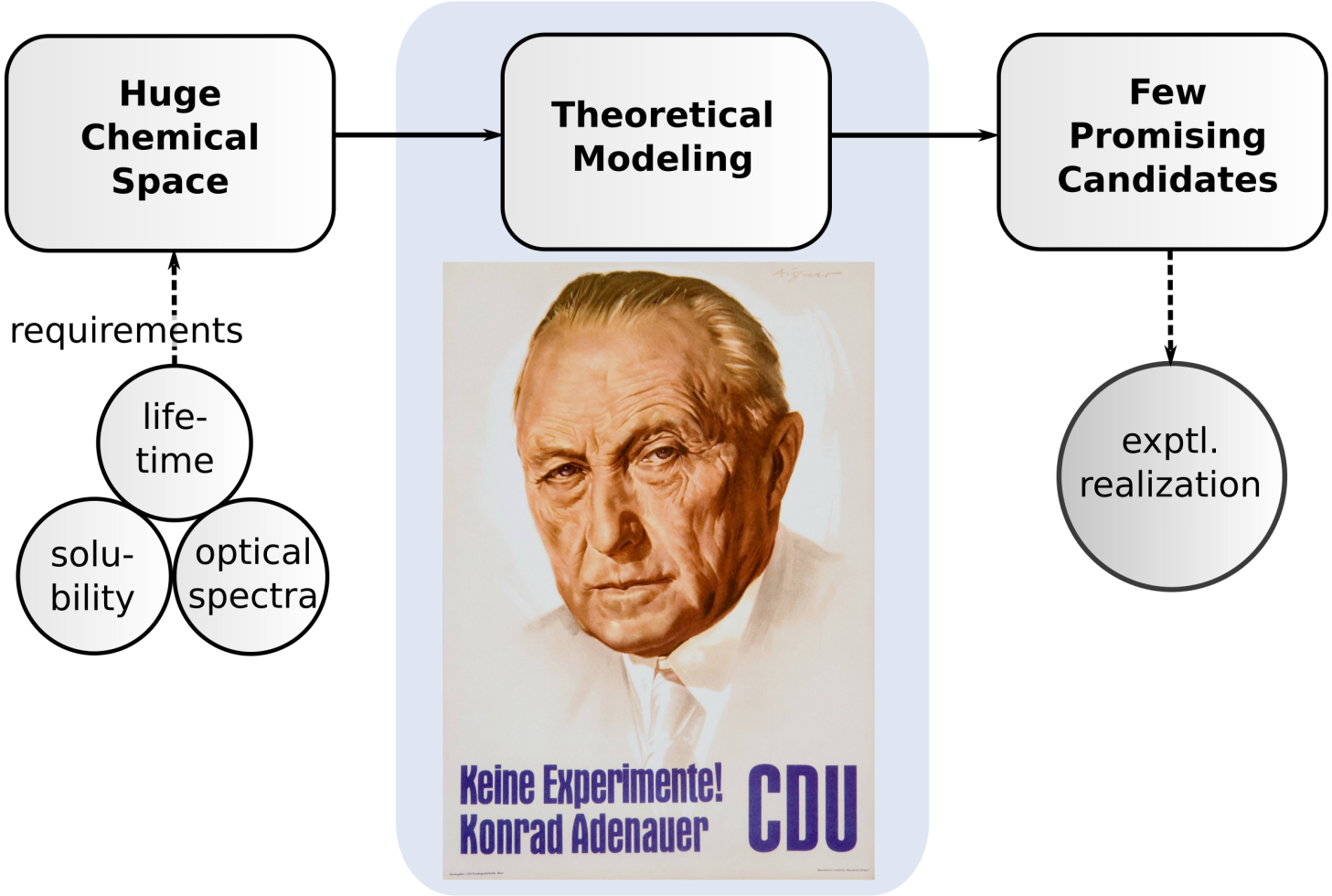
Gerit Brandenburg <j.g.brandenburg@gmx.de> | September 16, 2019

TALK AT THEORY OF CONDENSED MATTER SEMINAR, CAMBRIDGE UNIVERSITY, UK

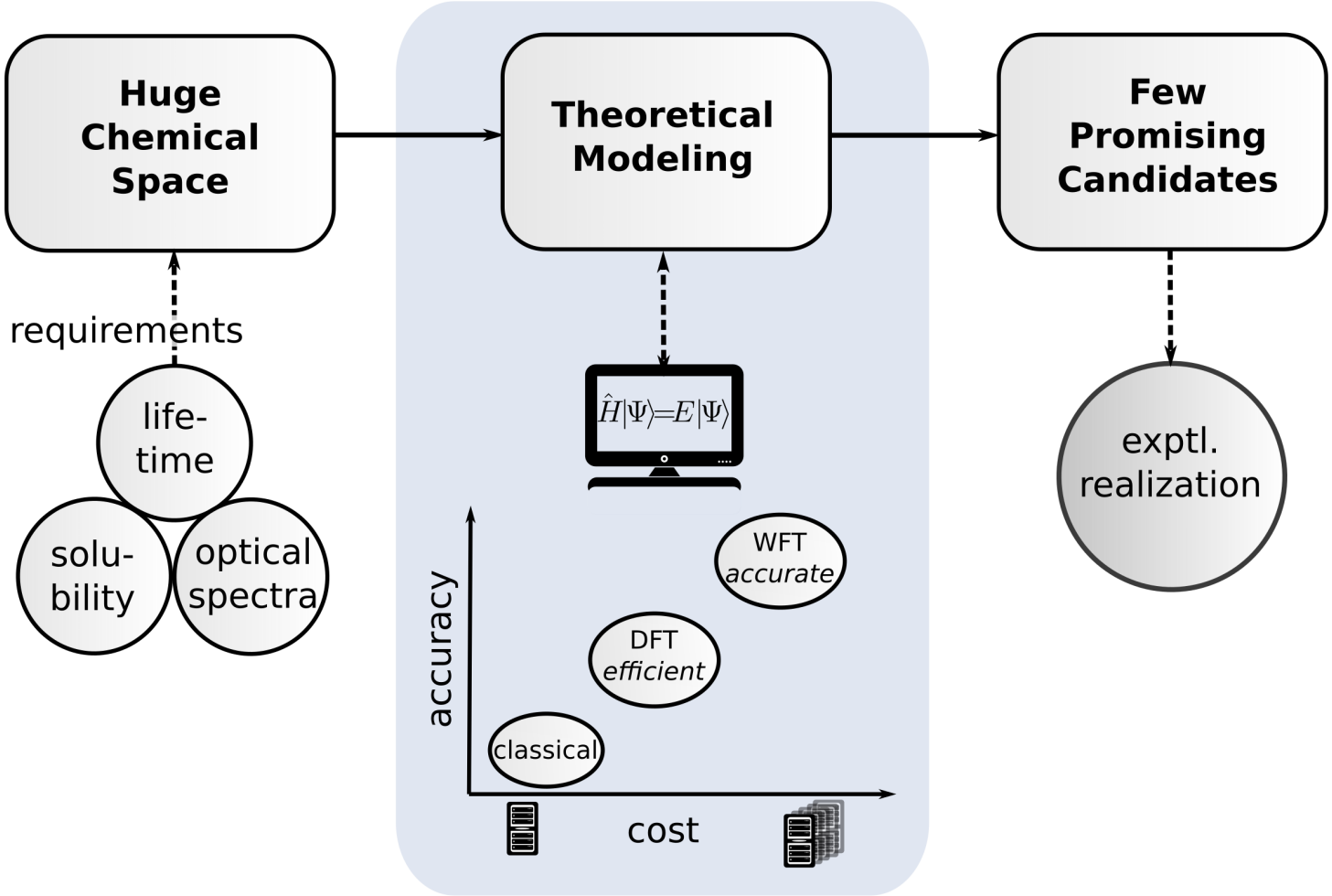
Material discovery can employ computational models



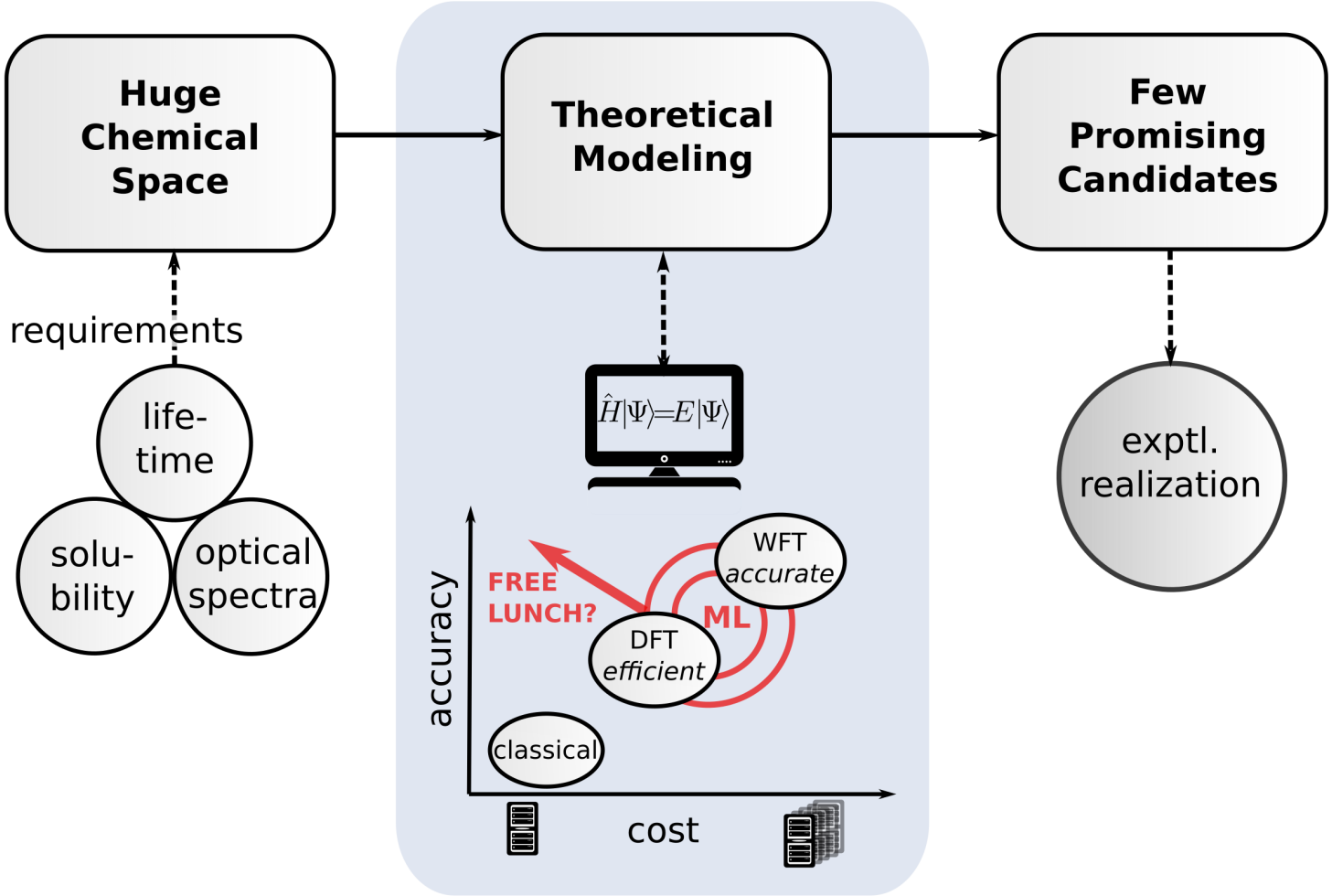
Material discovery can employ computational models



Material discovery can employ computational models

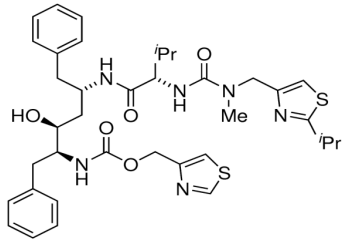


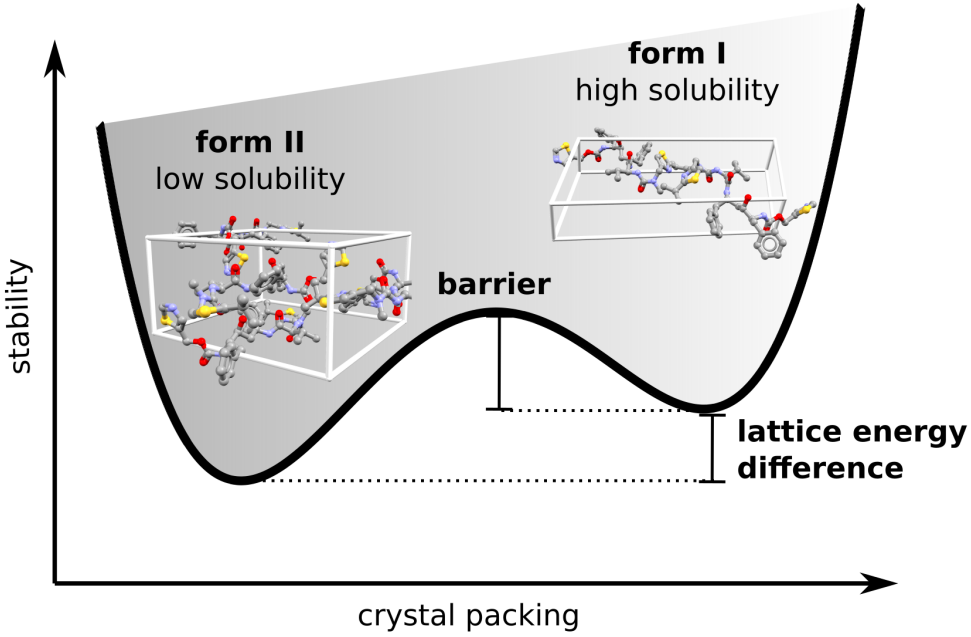
Material discovery can employ computational models



Many properties depend on the polymorphic form of a crystal

API: ritonavir
- HIV antiviral drug
- industrially produced in form I
- form II found 2 years after market launch
- reformulation required [1]





- tools to predict possible polymorphs would be valuable[2]

  → Simulation methods aim at complementing experimental screening

[1] J. Bauer, et al., *J. Pharm. Res.* **18**, 859-866 (2001).

[2] S. L. Price, JGB, *Molecular Crystal Structure Prediction*; Elsevier Australia ISBN: 9780128098356 (2017).

Predict most stable crystal polymorphs based on the molecular diagram

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

The inability to predict something as simple as how a molecule would crystallize is one of the continuing scandals in the physical sciences.[3–5]

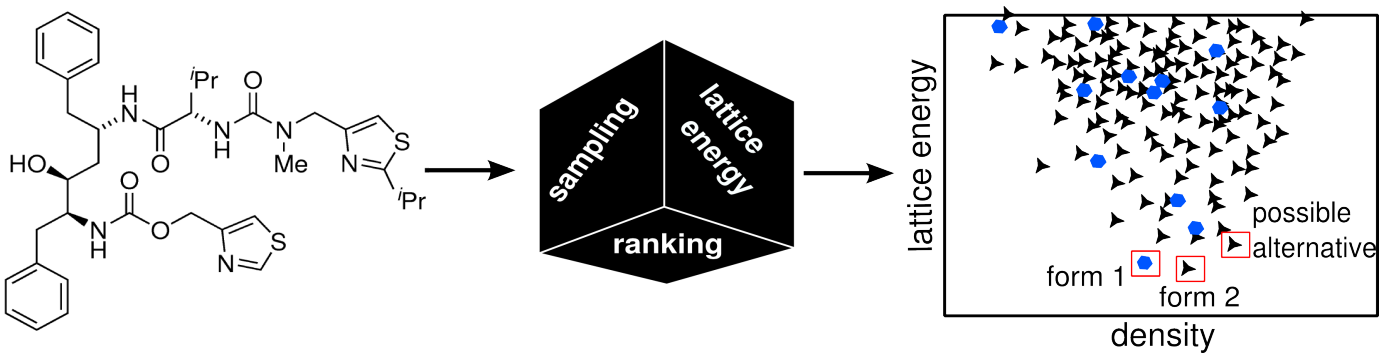
Task

- molecule is chosen due to its chemical/physical/biological properties
- based on the molecular diagram only, the most stable crystal structures should be predicted
- predict properties of interest for the most promising candidates

[3] A. Gavezzotti, *Acc. Chem. Res.* **27**, 309-314 (1994). [4] J. Maddox, *Nature* **335**, 201-201 (1988).

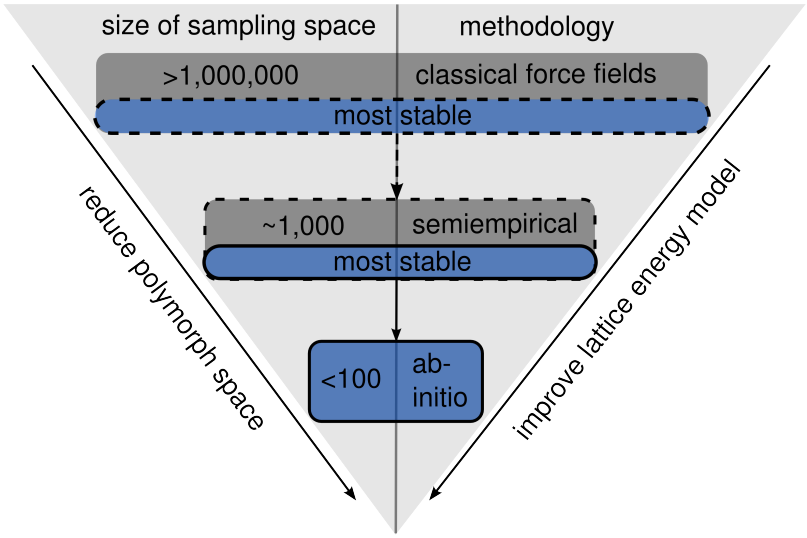
[5] K. N. Houk, F. Liu, *Acc. Chem. Res.* **50**, 539 (2017).

Sampling and energetic ranking for crystal structure prediction



Layers of complexity:

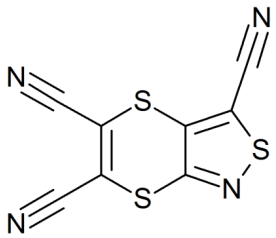
- molecular conformational space
- space groups (varying orientation, conformation, etc.)
- polymorph ranking according to free energy[6,7]



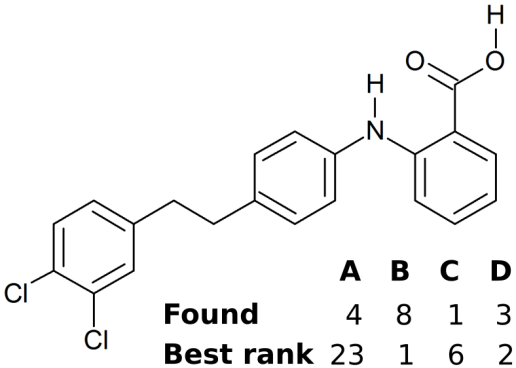
[6] S. Price, *Chem. Soc. Rev.* **43**, 2098 (2014)

[7] S. L. Price, JGB, *Molecular Crystal Structure Prediction*; Elsevier Australia, 336-363 (2017).

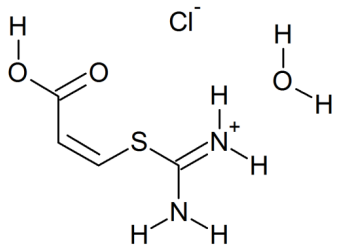
Results of the 6th blind test for organic crystal structure prediction



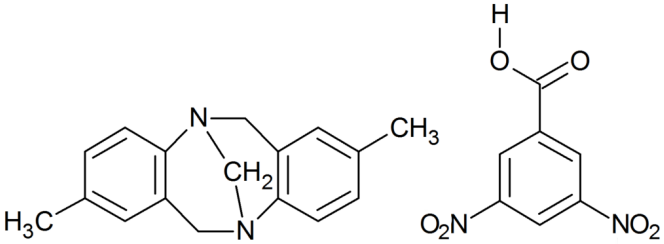
Found	12
Best rank	1



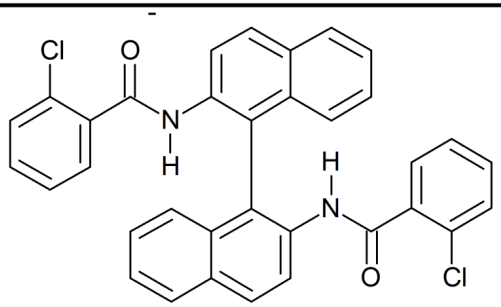
	A	B	C	D	E
Found	4	8	1	3	0
Best rank	23	1	6	2	



Found	1
Best rank	2



Found	5
Best rank	1



Found	3
Best rank	1

- No group had a perfect sampling and ranking of structures[8,9]

[8] A. Reilly, et al. *Acta Cryst. B*, **72**, 439 (2016). [9] J. G. Brandenburg, S. Grimme, *Acta Cryst. B*, **72**, 502 (2016).

CSP bind test highlights two remaining challenges

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

1. Ranking stage

Crystal polymorphs are often separated by just a few kJ/mol, exceeding the accuracy of standard density functional approximations (DFAs).
$\Longrightarrow$ **Are many-body methods feasible?**

2. Sampling stage

Dealing with a vast search space, in particular for molecules with increased flexibility, one has to cover too many structures at DFT accuracy.
$\Longrightarrow$ **Can we improve modern DFAs & employ ML techniques?**

Outline of talk

Methods with 'benchmark quality' for noncovalent interactions

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

CCSD(T)

- + 'gold standard' of quantum chemistry
- + linear scaling variants
- - pair-screening thresholds
- - slow basis set convergence
- - non-trivial to parallelize (shared memory: 100 GB/core)

DMC

- + exact projection to ground state
- + N^3 scaling, [-] huge prefactor
- + fast basis set convergence
- + scaleable to HPCs
- - stochastic error
- - uncertainties from FN and non-local pseudopotentials

No benchmark quality: MP2, RPA, CCSD, CCSD(T)/cc-pVTZ, VMC

Quantum Monte-Carlo in a nutshell: A scalable high-level method

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Fixed-node diffusion Monte-Carlo

1) Enforce nodal surface of Fermions

$$\Gamma = \{\mathbf{R}; |\Psi_T\rangle = 0\}$$

2) Walkers in configuration space

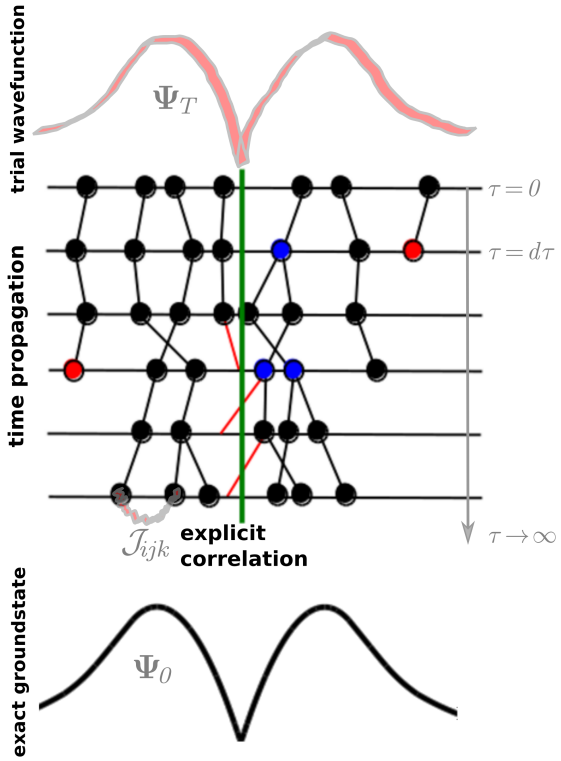
$$|\Psi_T(\mathbf{R},\tau)\rangle = \text{hist}\left[\sum \delta(\mathbf{R} - \mathbf{R_i}(\tau))\right]$$

3) Diffusion in imaginary time

$$\partial_\tau |\Psi_T(\mathbf{R},\tau)\rangle = \left[\frac{1}{2}\nabla^2_{\mathbf{R}} - (V - E_T)\right] |\Psi_T(\mathbf{R},\tau)\rangle$$

4) Projection to exact ground state

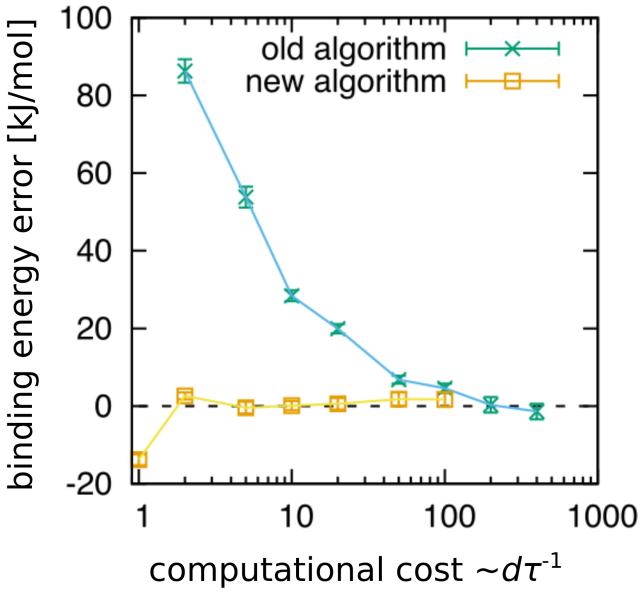
$$|\Psi_0(\mathbf{R})\rangle = \lim_{\tau\to\infty} \exp\left[-\tau(\hat{H} - E_T)\right] |\Psi_T(\mathbf{R},\tau)\rangle$$

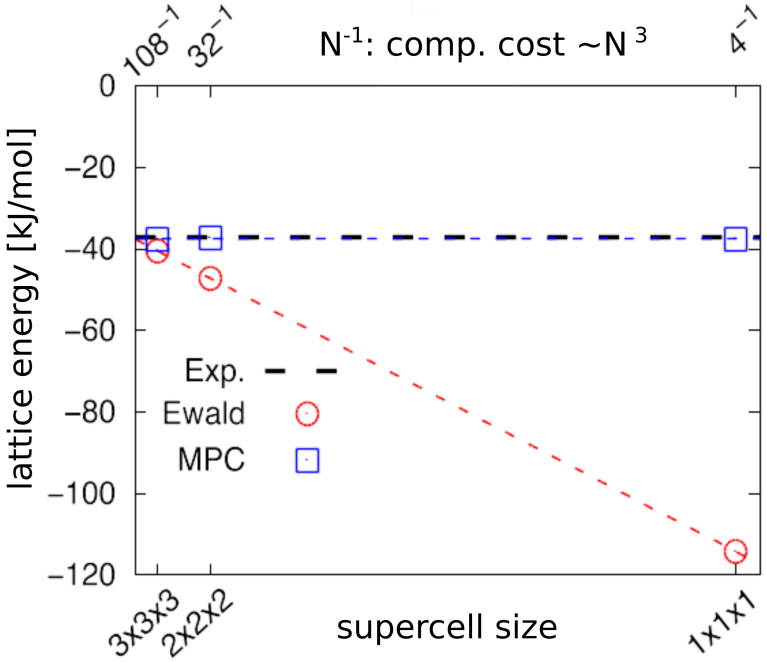


- exact within Γ and $d\tau \to 0$ and $\tau \to \infty$[15]
- new DMC algorithms lead to substantial speed up[16]

[15] *Chem. Rev.* **116**, 5188 (2016) [16] A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, D. Alfé, *Phys. Rev. B* **93**, 241118(R) (2016).

New DMC algorithm leads to substantial speed up



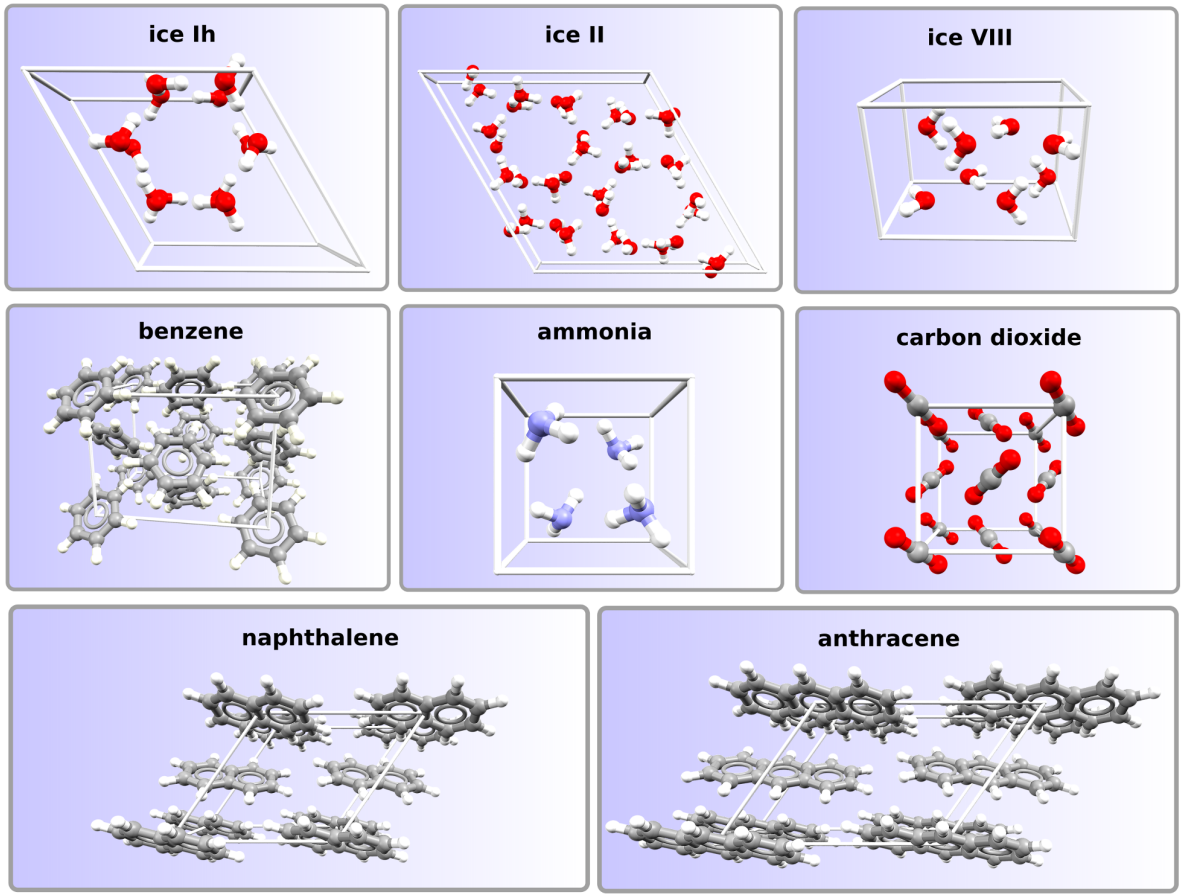


- new size-consistent implementation reduces $\Delta\tau$ error drastically[13]
- Model periodic Coulomb for finite size correction[14]

[16] A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, D. Alfé, *Phys. Rev. B* **93**, 241118(R) (2016).

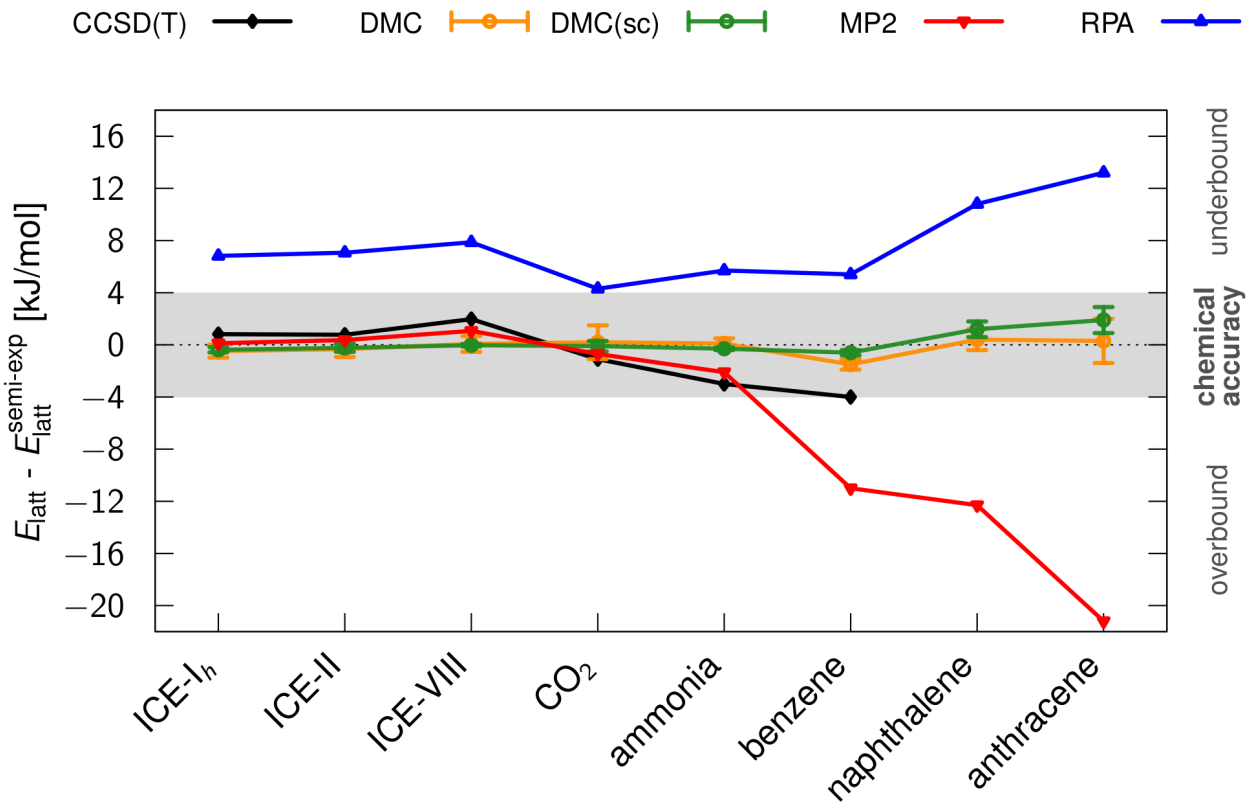
[17] L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, A. J. Williamson, *Phys. Rev. B* **53**, 1814 (1996).

Molecular crystals as challenging test



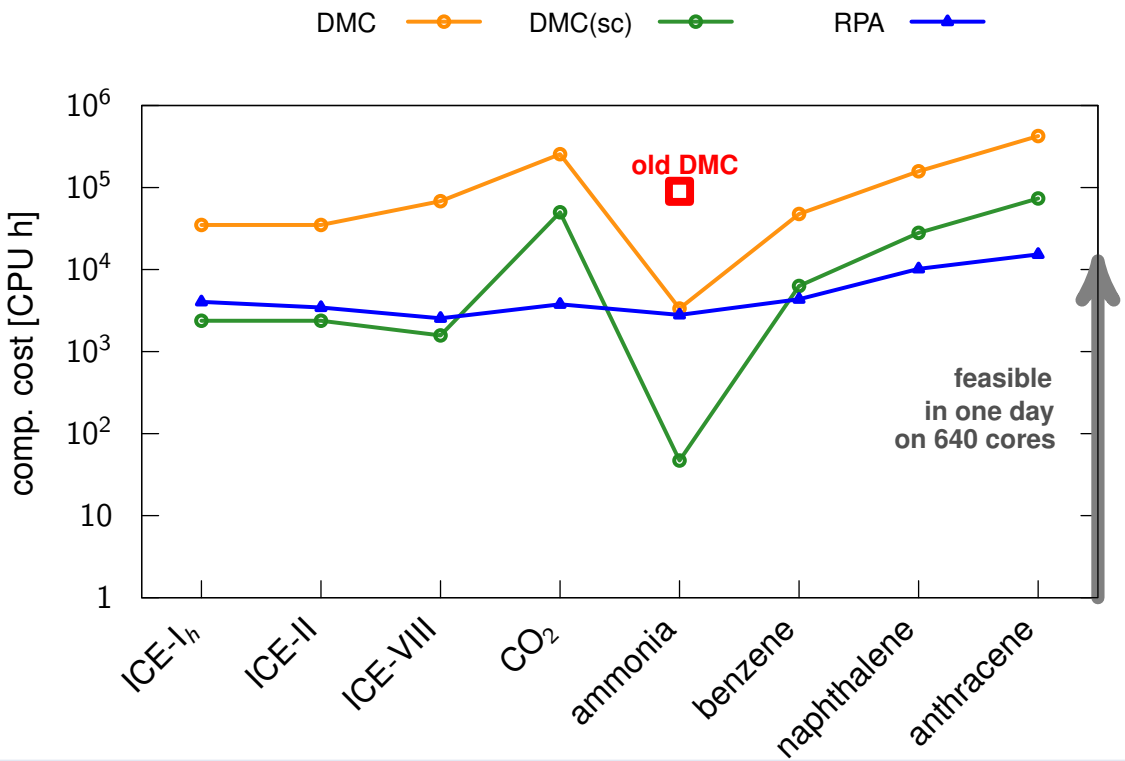
- strong H-bonds, vdW of saturated and unsaturated molecules
- problematic for all readily applicable methods (DFA-DISP, MP2)

QMC delivers (sub-) chemical accuracy for all tested systems



- excellent agreement with experiment and CCSD(T)
- uncertainty in H_{sub}^{exp} probably larger than DMC errors

QMC is feasible within one day on standard computer cluster



- up to three orders of magnitude speed-up compared to best DMC practice two years ago
- significance will extend to all classes of systems[18]

[18] A. Zen, JGB, J. Klimeš, A. Tkatchenko, D. Alfè, A. Michaelides, *Proc. Natl. Acad. Sci. U.S.A* , **115**, 1724 (2018).

Outline of talk

Multilevel methodologies: Finding the right compromise

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

	task/property	example method
accurate QM	single-point energy	many-body WFT (DMC[18])
cheap QM	optimization	DFT[19–20] (HSE-3c[21])
very cheap QM	optimization/Hessians conformations	semi-empirical[22,23]
force field	dynamics conformational sampling	transferable or molecule specific (QM derived) FF

[19] JGB, J. E. Bates, J. Sun, J. P. Perdew *Phys. Rev. B*, **94**, 115144 (2016).

[20] JGB, C. Bannwarth, A. Hansen, S. Grimme, *JCP*, **148**, 064104 (2018). [21] JGB, E. Caldeweyher, S. Grimme, *Phys. Chem. Chem. Phys.* **18**, 15519 (2016).

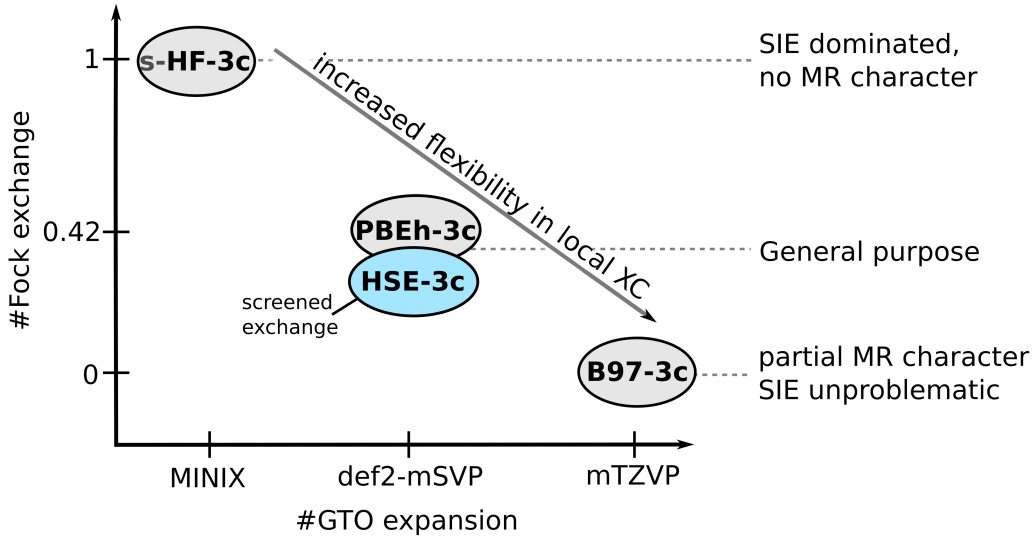
[22] JGB, S. Grimme, *JPCL*, **5**, 1785 (2014). [23] M. Mortazavi, JGB, R. J. Maurer, A. Tkatchenko, *JPCL*, **9**, 399 (2018).

Full hierarchy of 'low-cost' methods

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

HSE-3c construction principles

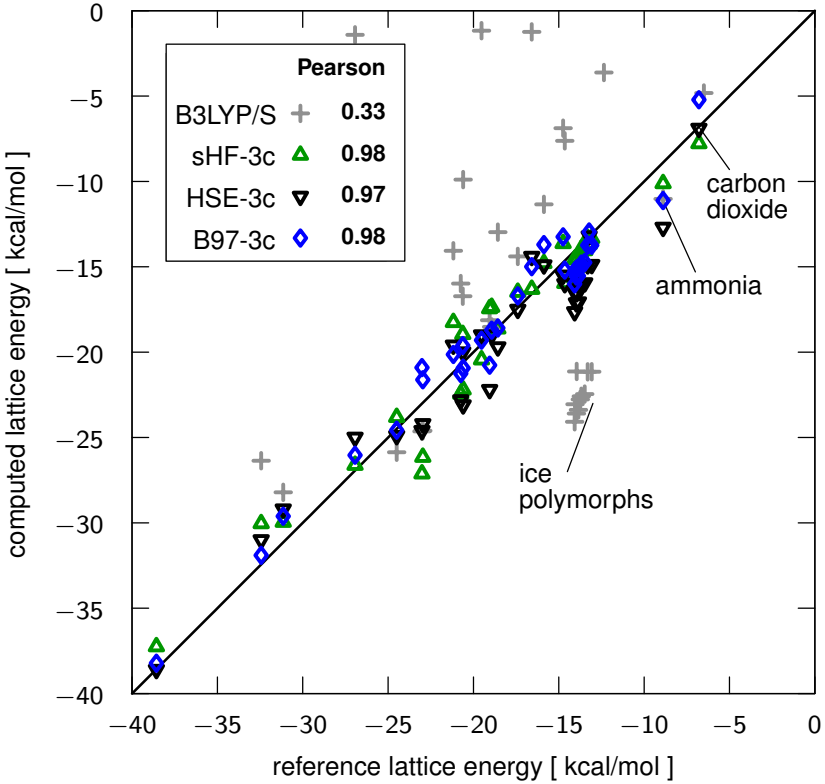
- fast for optimizations and frequencies → compact orbital expansion
- avoid most self-interaction error → use Fock exchange
- numerically robust → screening of Fock exchange
- combine with semi-classical correction potentials (3c)

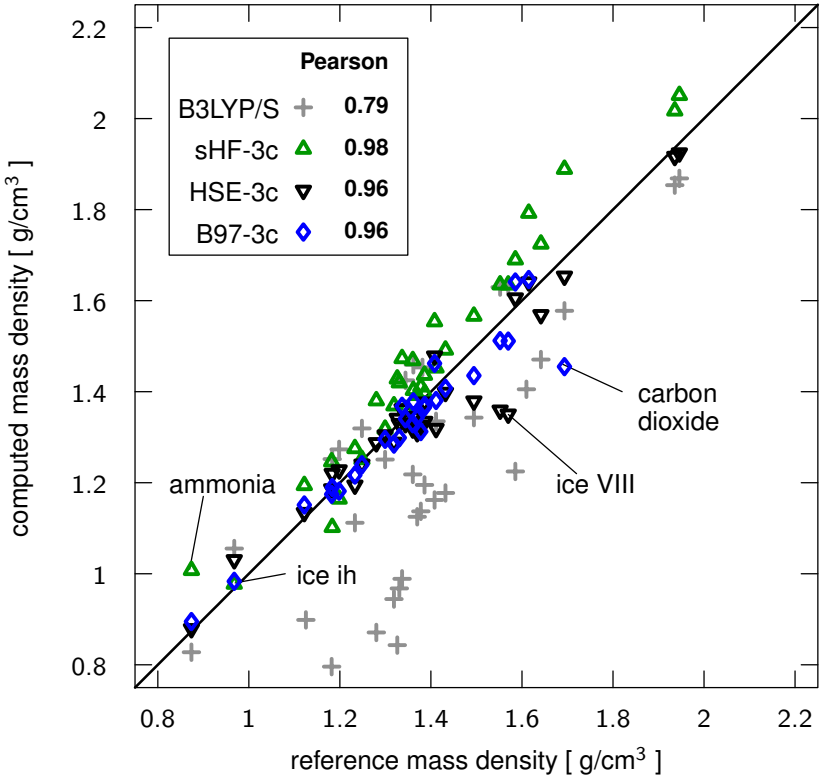


[24] E. Caldeweyher, JGB, *J. Phys.: Condens. Matter* **30**, 213001 (2018) *[Psi-k Highlight Jan. 2019]*

Good results on molecular crystals

X23[25,26] and ICE10[27] benchmark sets:

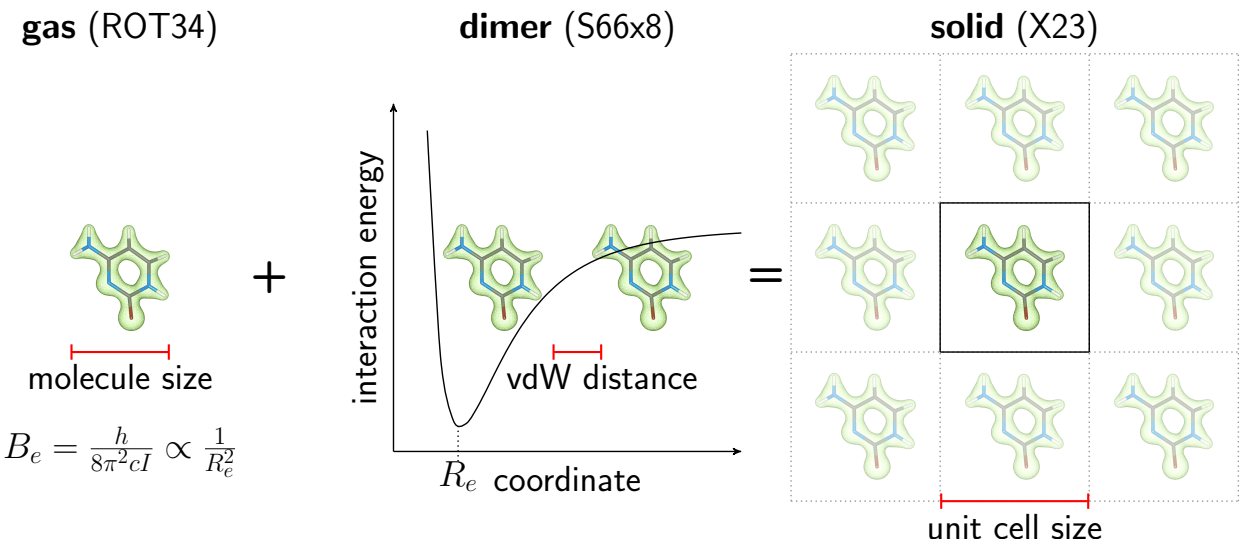




- consistent structures and interaction energies by HSE-3c

[25] E. Johnson, A. Otero-de-la-Roza, *J. Chem. Phys.* **137**, 054103 (2012), [26] A. Reilly, A. Tkatchenko, *JCP* **139**, 024705 (2013)
[27] JGB, T. Maas, S. Grimme, *J. Chem. Phys.* **142**, 124104 (2015)

Consistency for structures achieved



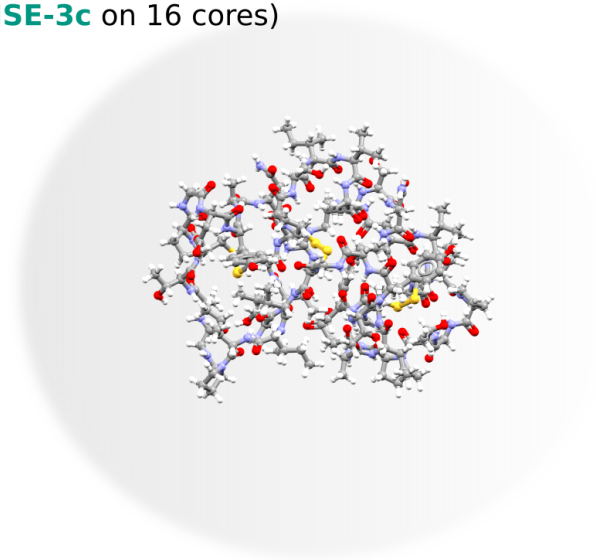
	ROT34 org. mol.	**S66x8** non-covalent	**X23** molecular crystal
TPSS/large basis	1.9	14.6	27.9
TPSS-D3/large basis	1.3	1.3	1.0
HF-3c (very cheap QM)	1.5	-1.2	-5.7
DFTB3-D3 (very cheap QM)	1.2	-2.9	-12.6
HSE-3c (cheap QM)	**0.2**	**1.3**	**0.7**

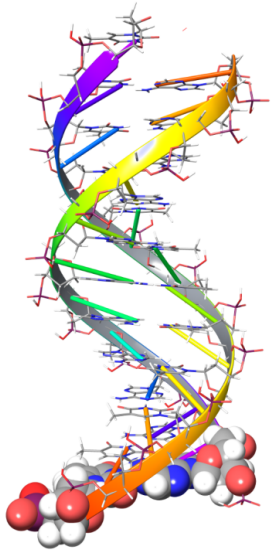
mean of relative deviation in %

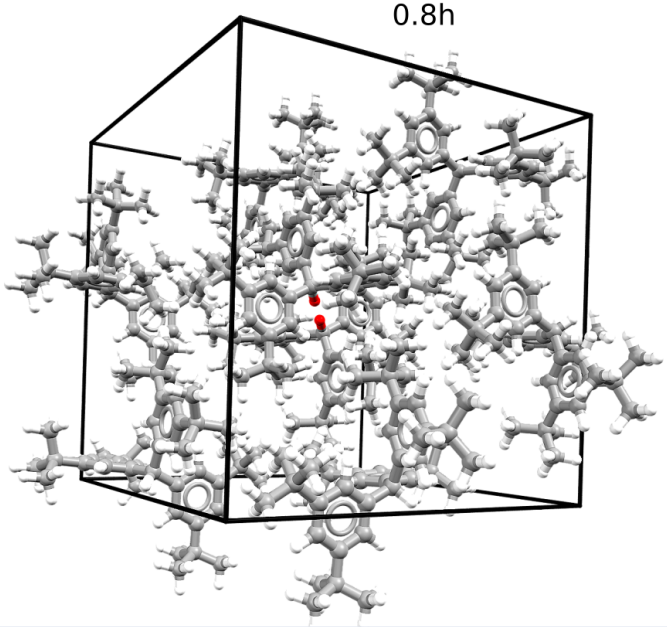
Fast electronic structure with ab-initio accuracy

	gas phase protein	DNA helix	molecular crystal
dim:	0D	1D	3D
#atoms/unit:	647 (647)	726 (66)	856 (37)
wall time:	5h	0.5h	0.8h

(**HSE-3c** on 16 cores)





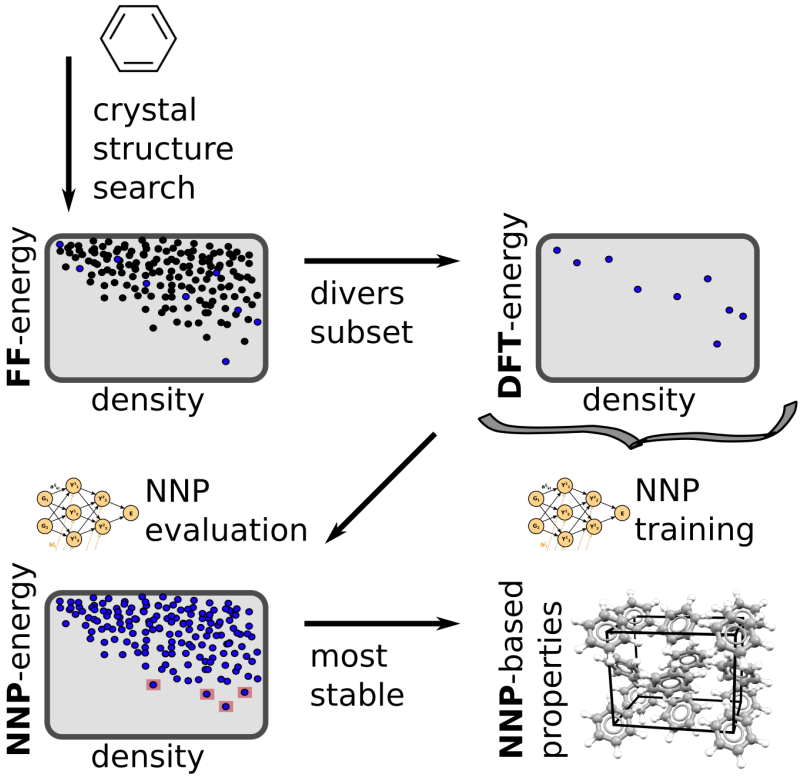


- fast computer code CRYSTAL17[28] with cost-efficient methods
- enabling routine electronic structure calculation of large systems

[28] R. Dovesi, A. Erba, R. Orlando, C. Zicovich-Wilson, B. Civalleri, L, Maschio, et al., *WIREs Comput. Mol. Sci.* **8**, 1360 (2018).

Employing ML to enhance CSP workflow

1. Generate putative structures with crude FF stability
2. Construct a divers subset of structures
3. Compute DFT stability on subset
4. Train a high-dimensional NNP on DFT references
5. Recompute stability of all structures employing the NNP



[29] V. L. Deringer, D. M. Proserpio, G. Csányi, and C. J. Pickard, *Faraday Discuss.* **211**, 45 (2018).

[30] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, *Phys. Rev. B* **99**, 064114 (2019).

[31] JGB and J. Behler, *in preparation*.

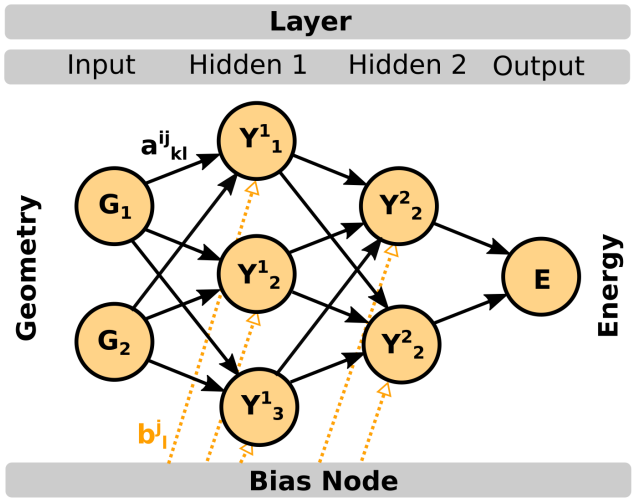
High-dimensional neural network for benzene crystal landscape

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Energy decomposed to atomic contributions

$$E_{\text{tot}} = \sum_{i}^{\text{atoms}} E_i \left[Z_i, G(\{R_j\}) \right]$$

- symmetry functions as local geometry descriptors G[32−34]
- element specific feed-forward neural network (2 hidden layers, 25 nodes each)
- training based on electronic energies and nuclear forces



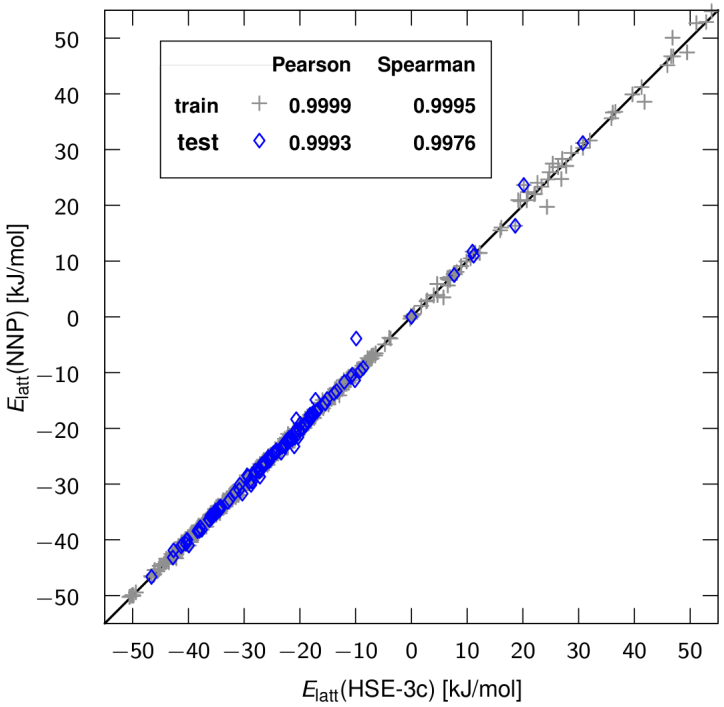
[32] J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).

[33] J. Behler, *J. Chem. Phys.* **134**, 074106 (2011).

[34] J. Behler, *Angew. Chem. Int. Ed.* **56**, 12828 (2017).

Stable training on 10% of generated putative polymorphs

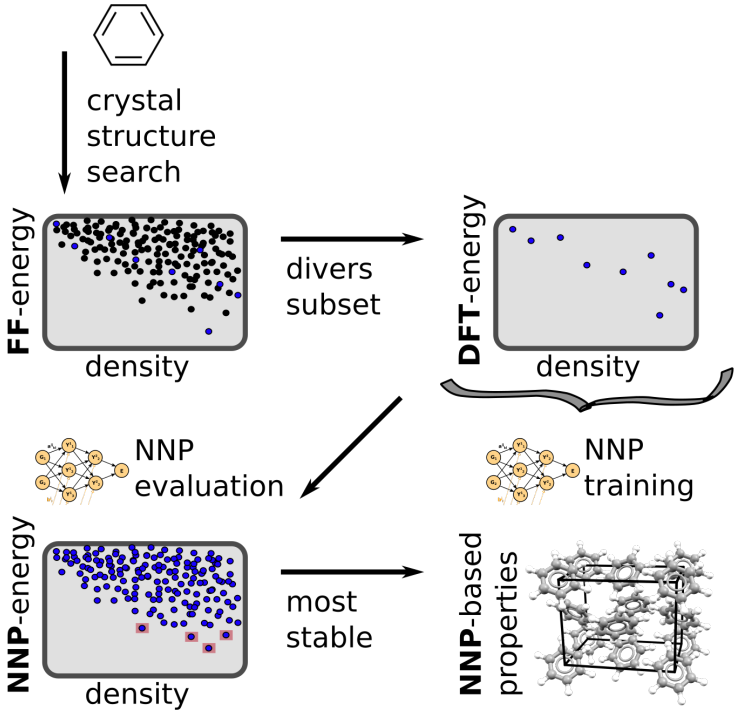
	training	testing
#structures	1 077	127
#data points	39 849	4 699
range E [kJ/mol]	[-50:147]	[-55:120]
range ρ [g/cm^3]	[0.1:1.1]	[0.1:1.1]
RMSD(E) [kJ/mol]	**0.53**	**0.93**
RMSD(F) [a.u.]	0.001	0.001

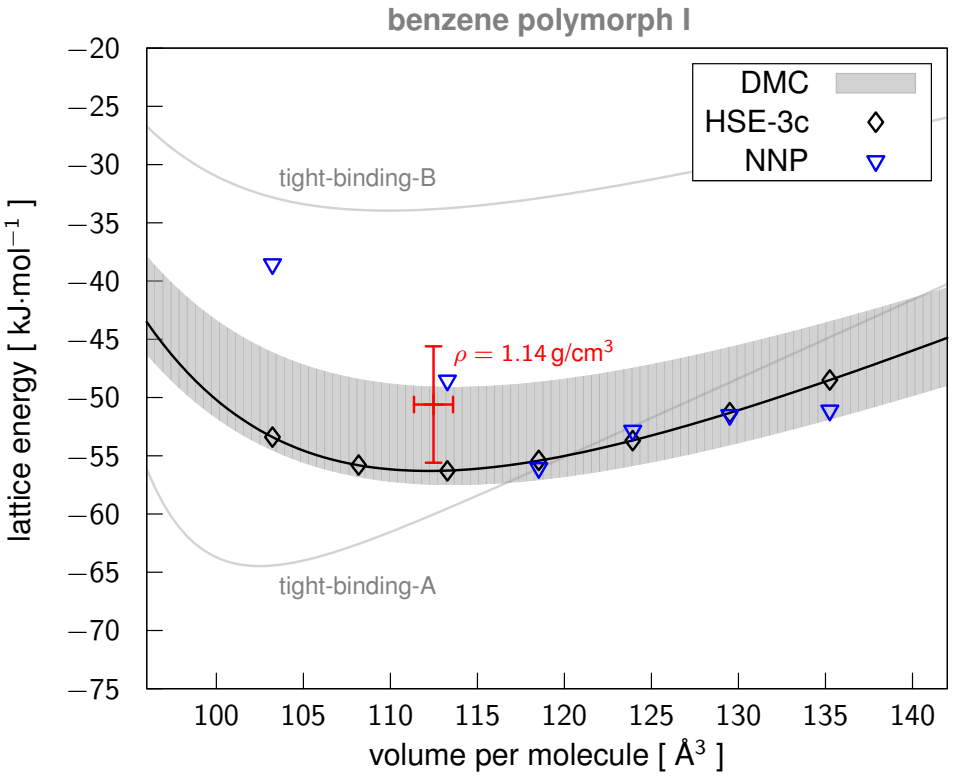


- neural network delivers sub-chemical accuracy
- CSP cost reduced by one order of magnitude [32]

[32] JGB and J. Behler, *in preparation*.

Employing ML to enhance CSP workflow



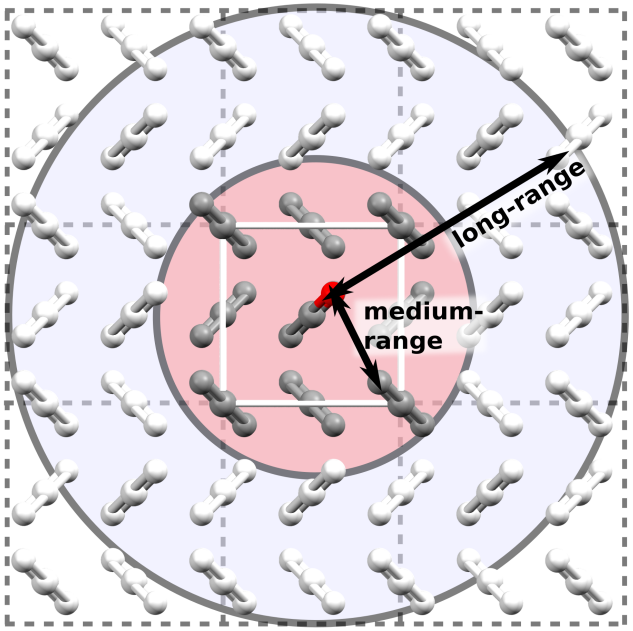


- EOS within 1 kJ/mol for ρ <1.1 g/cm^3, large errors for ρ >1.2 g/cm^3
- too few data points to generate globally smooth EOS

[32] JGB and J. Behler, *in preparation*.

Outline of talk

Learn many-body exchange-correlation to improve DFAs



A) long-range
physics based interactions

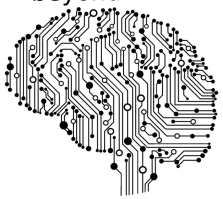
(i) electrostatics by DFT

(ii) induction by DFT

(iii) semi-classical vdW

B) medium-range
flexible artificial AI

(iv) many-body effects
beyond DFT-vdW



- DFT-vdW is capable to describe long-range interactions very accurately
- ML with local descriptor to learn medium-range many-body effects

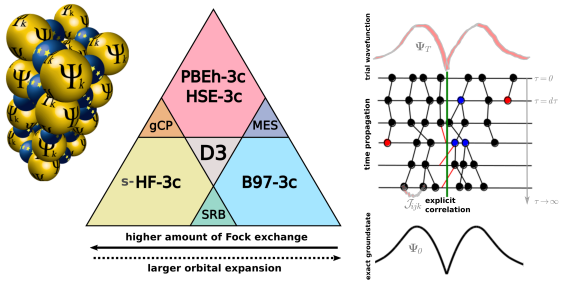
Remaining challenges in crystal structure prediction

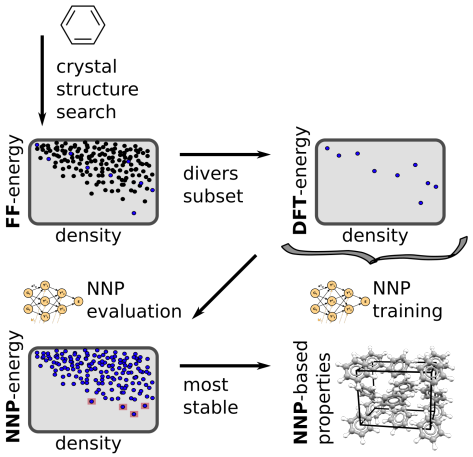
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

- Thermodynamics of crystals, i.e. full PT phase diagram
- Kinetics of crystallization
- Reducing the human and CPU time for prediction

Summary

- DMC delivers (sub-) chemical accuracy with three orders of magnitude speed up
- hierarchy of simplified DFAs has been developed
- enhancing CSP by combining simplified DFA & ML





Acknowledgements

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

PhD / Postdocs


Andrea Zen


Martin Fitzner


Benjamin Ramberger


Theodoros Tsatsoulis


Eike Caldeweyher


Luca Iuzzolino

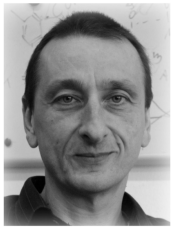
Principal investigator

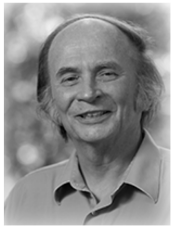

Michaelides (UCL)


Alfè (UCL)


Tkatchenko (Luxembourg)


Klimes (Prague)


Grimme (Bonn)


Perdew (Temple)


Kresse (Vienna)


Behler (Göttingen)


Rebecca Sure (BASF)


Price (UCL)


Grüneis (Vienna)


Civalleri (Turin)


Maschio (Turin)


Antreas Afantitis (NovaMechanics)

Alexander von Humboldt Stiftung/Foundation

THE THOMAS YOUNG CENTRE TYC

erc

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

MATERIALS AND MOLECULAR MODELLING HUB

CSCS

Acknowledgements

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

PhD / Postdocs


Andrea Zen


Martin Fitzner


Benjamin Ramberger


Theodoros Tsatsoulis


Eike Caldeweyher


Luca Iuzzolino

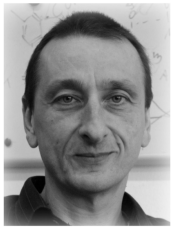
Principal investigator

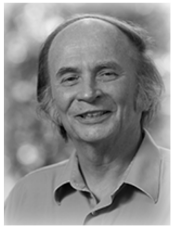

Michaelides (UCL)


Alfè (UCL)


Tkatchenko (Luxembourg)


Klimes (Prague)


Grimme (Bonn)


Perdew (Temple)


Kresse (Vienna)


Behler (Göttingen)


Rebecca Sure (BASF)


Price (UCL)


Grüneis (Vienna)


Civalleri (Turin)


Maschio (Turin)


Antreas Afantitis (NovaMechanics)

Thanks

Alexander von Humboldt Stiftung/Foundation

THOMAS YOUNG CENTRE

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

MATERIALS AND MOLECULAR MODELLING HUB

CSCS

Key references

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

- Crystal structure prediction:

  S. L. Price, JGB, *Molecular Crystal Structure Prediction*, G. DiLabio, A. Otero-de-la-Roza, Eds., Elsevier Australia, 336-363 (2017).

  L. Iuzzolino, P. McCabeb, S. L. Price, JGB, *Faraday Discuss.* **211**, 275 (2018).

- QMC for molecular crystals:

  A. Zen, JGB, J. Klimeš, A. Tkatchenko, D. Alfè, A. Michaelides, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1724 (2018).

- DFT development

  S. Grimme, A. Hansen, JGB, C. Bannwarth, *Chem. Rev.* **116**, 5105 (2016).

  E. Caldeweyher, JGB, *J. Phys.: Condens. Matter* **30**, 213001 (2018).

- website: **gerit-brandenburg.de**