Accepted to JCTC (2023)

Benchmark datasets for crystalline organic semiconductors (BMCOS)

Andriy Zhugayevych

Max Planck Institute for Polymer Research, Polymer Theory Department

November 5, 2023

Acknowledgments

Wenbo Sun, Tammo van der Heide, Carlos Lien-Medrano, Thomas Frauenheim (BCCMS), Sergei Tretiak (LANL)

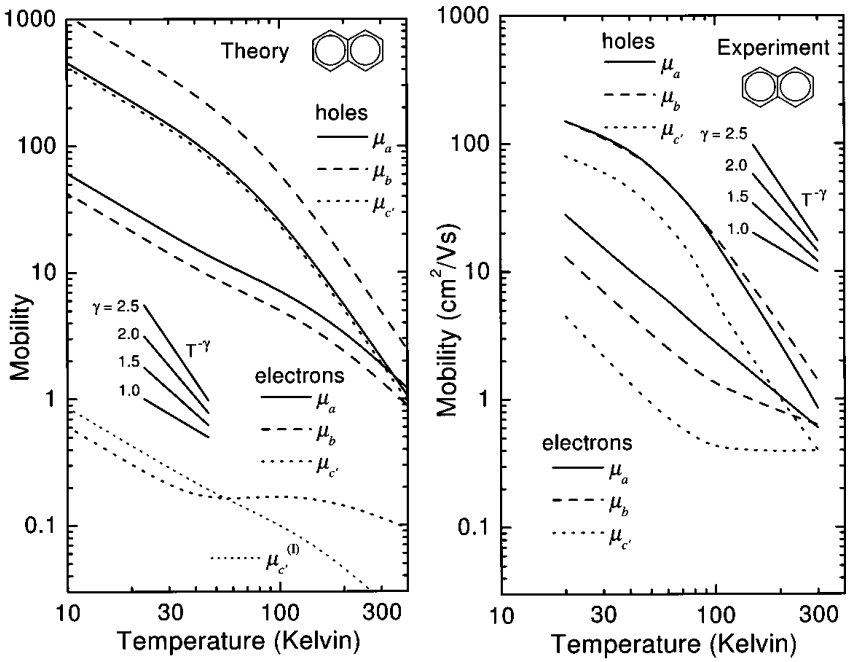
Denis Andrienko's group

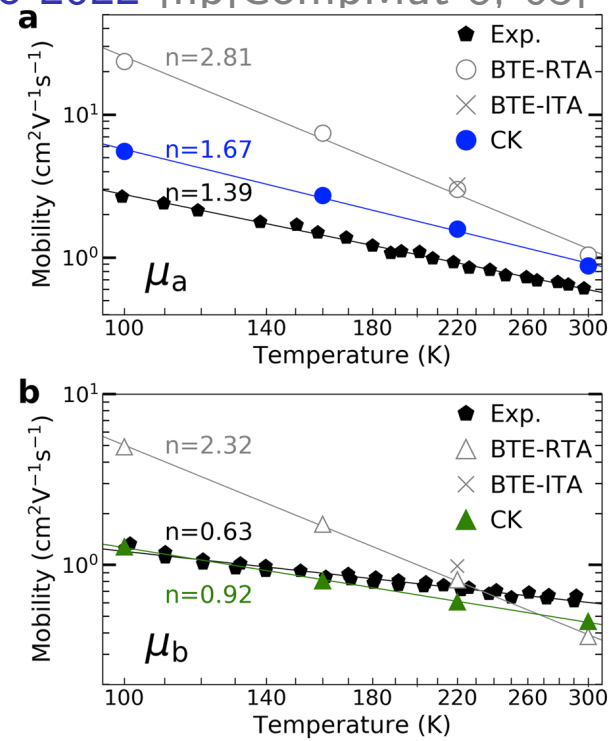
Outline

- Motivation
- Benchmark datasets
- Benchmarking DFT-D and semiempirical methods

Charge carrier mobility in naphthalene (20 πe/cell)

Advance in modeling from 2004 [APL 85, 1535] to 2022 [npjCompMat 8, 63]*





Depends on accuracy of

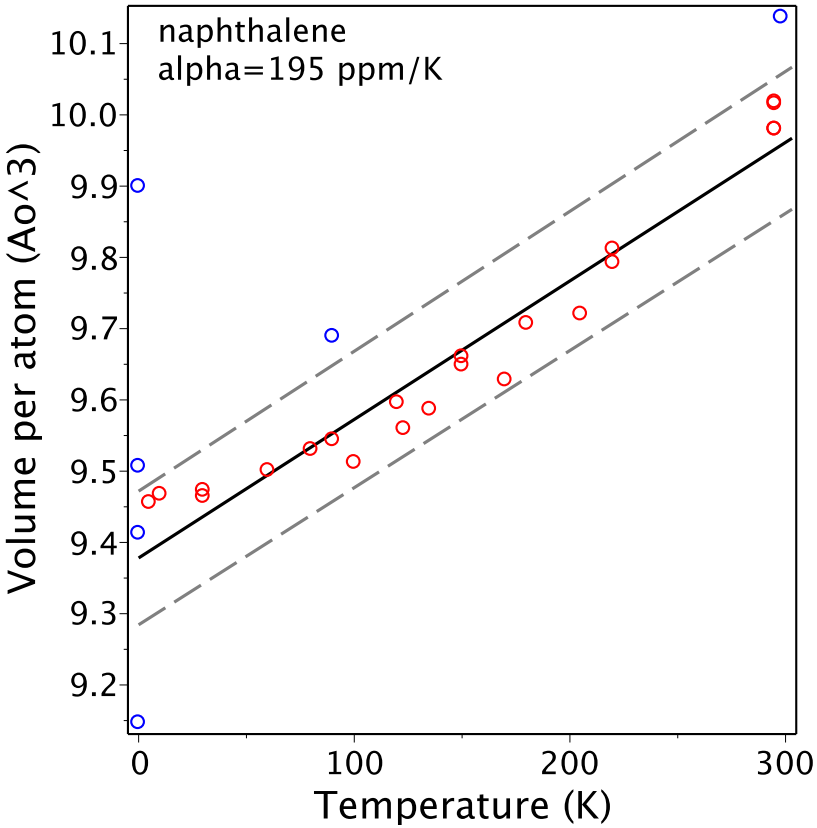
- geometry and force constants
- electronic and vibronic couplings
- e-p Hamiltonian and its solution

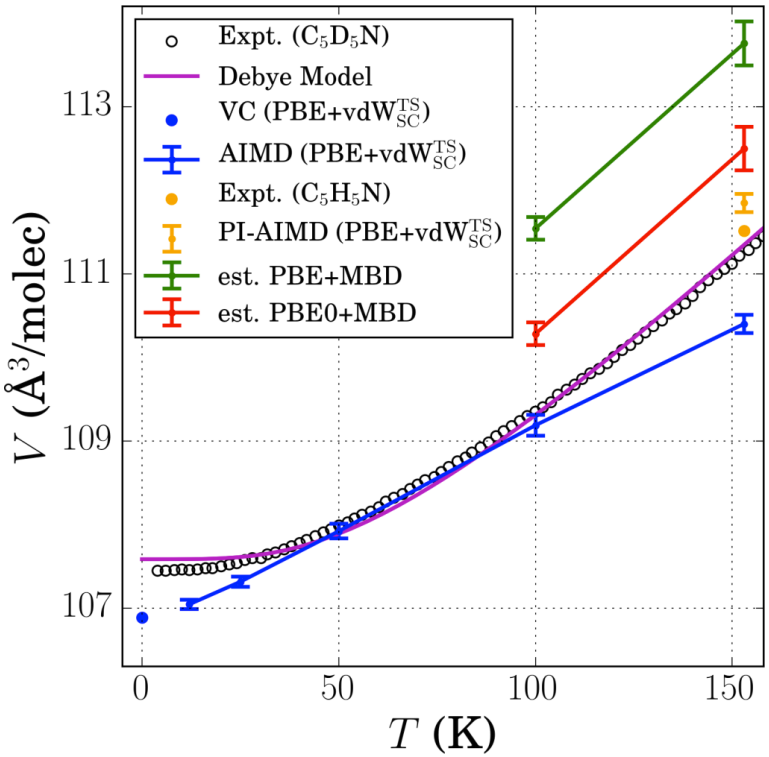
How accurate is crystal geometry and force constants?

---

* GGA+GW, unit cell is taken from experiment at different temperatures

Measuring/predicting crystal structure of naphthalene





PRM 2, 055603 (2018)

- 1% – experimental uncertainty at fixed temperature
- 1% – quantum effects at low temperatures
- 5% – thermal expansion at 300 K

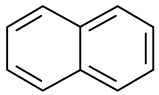
This sets the target accuracy for theoretical methods to 1%, that is smaller than variation of DFT-D results with functional

$\Longrightarrow$ *need methods benchmarking* $\Longrightarrow$ *need benchmark dataset*

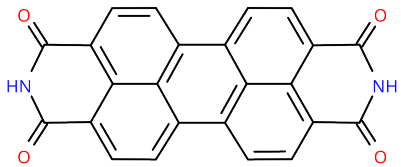
What accuracy of crystal geometry is needed

Depends on goals and feasibility

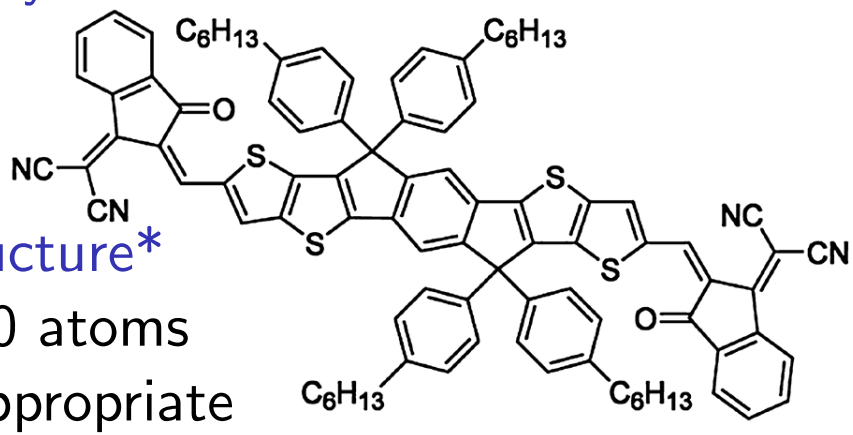
- Topology and morphology
  - generic force fields



- To rank polymorphs by energy, to match experim. uncertainty
  - DFT-D
  - advanced force fields



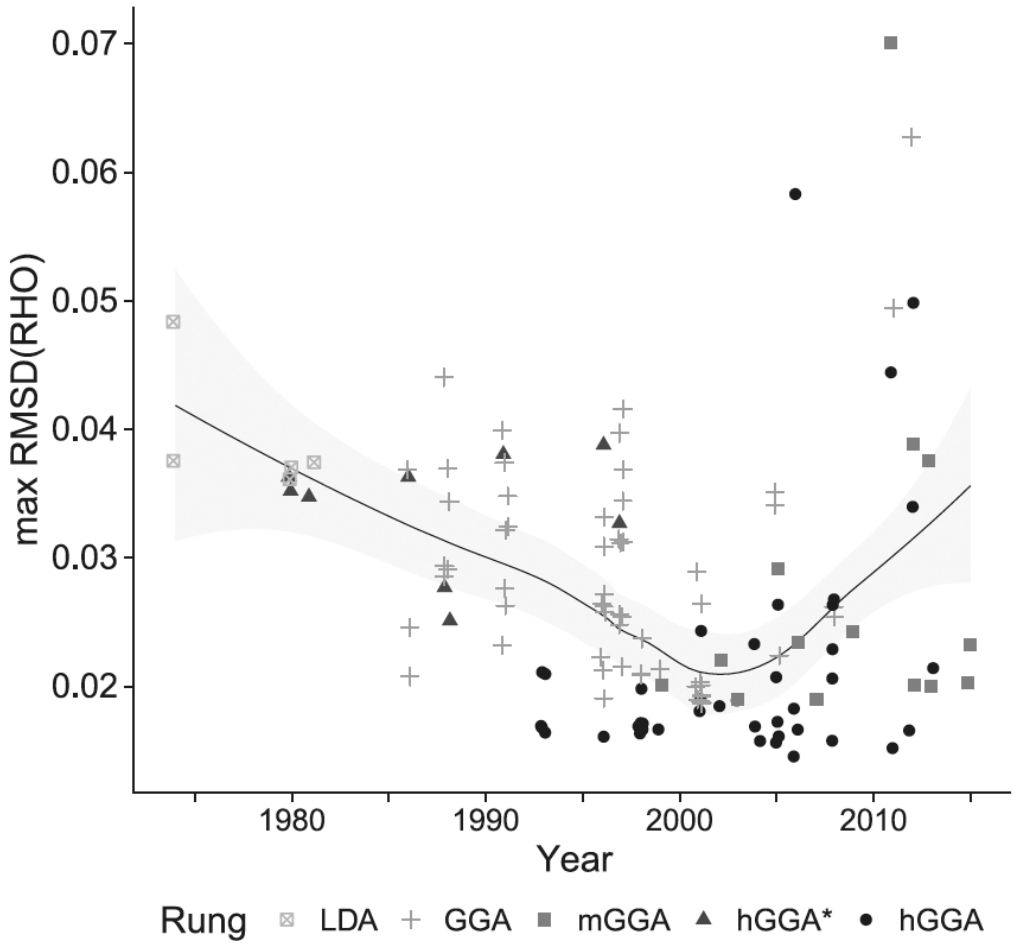
- To not distort electronic structure*
  - DFT-D – good but $<$ 1000 atoms
  - DFTB – looks the most appropriate
  - well-parameterized force fields – how to parameterize



Need a set of well-benchmarked methods to cover all scales

One method to obtain geometry, another method to get electronic structure.
DFTB=Density Functional based Tight Binding

Why we need separate dataset for organic semiconductors?



Science 355, aah5975 (2017)

Another example: CAM-B3LYP is one of the best for organic semiconductors but inaccurate for inorganic ones

Modeling of organic semiconductors: open problems

The table is for molecular systems, but there are also polymers, frameworks, hybrid

	single molecule	cluster, solution	single crystal	bulk glass	films ...
predict topology	✓✓	✓	?	CG	
optimize geometry	✓✓	✓	DFT-D	?	
electronic structure	✓✓	✓	ECG	localized	
excited states	✓✓	✓	?	localized	
electronic transfer	NAMD	NAMD	?	hopping	
reactions ...					

ECG=Electronic Coarse Graining, NAMD=NonAdiabatic Molecular Dynamics

Goals

- Create benchmark datasets for organic semiconductors
- Benchmark DFT-D for crystals up to 1000 atoms
- Benchmark DFTB for larger systems
- Benchmark generic force fields
- Benchmark electronic structure methods
- Create library of electronic structure prototypes
- Search for novel electronic structure architectures
- What is maximum possible charge carrier mobility for organic semiconductors?

Existing databases relevant to organic semiconductors

- Databases of crystalline organic semiconductors:
  large OMDB, OCELOT and small [Yavuz2016]
  but not designed for benchmarking
- Benchmark datasets for molecular crystals:
  CPOSS, [Reilly2016, Brandenburg2016, Dolgonos2019]
  but no extended π-conjugated molecules
- In principle, one can use Cambridge Structural Database
  but ambiguity in selection and preprocessing of raw CIFs

BMCOS: several datasets are needed

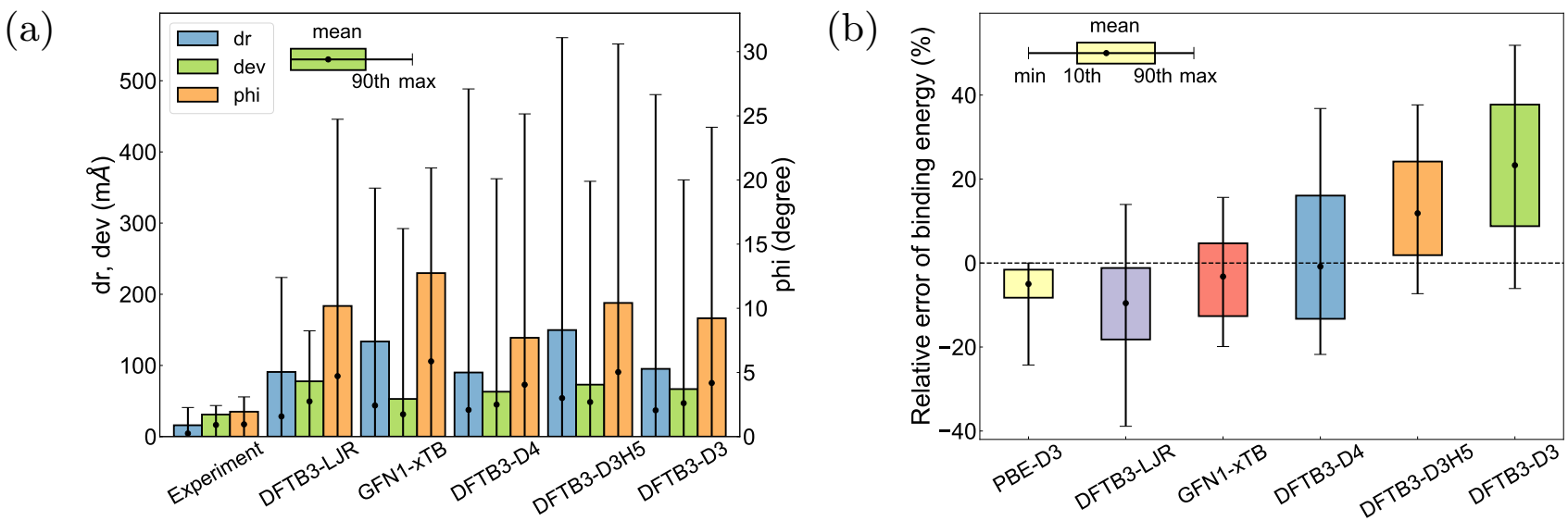
cmsos.github.io/bmcos/

- Main polymorph of single-conformer molecules (BMCOS1)
- Polymorphs
- Multi-conformer molecules with side chains
- Polymers
- Metal-organics

BMCOS1 is ready with 67 systems: acenes, oligothiophenes, PAHs, azaacenes, thienoacenes, imides, quinones, indigos, and others including TCNQs, stilbene, TTF, C60, see picture at cmsos.github.io/bmcos/BMCOS1_alt.html

Nongeometrical parameters available in BMCOS1

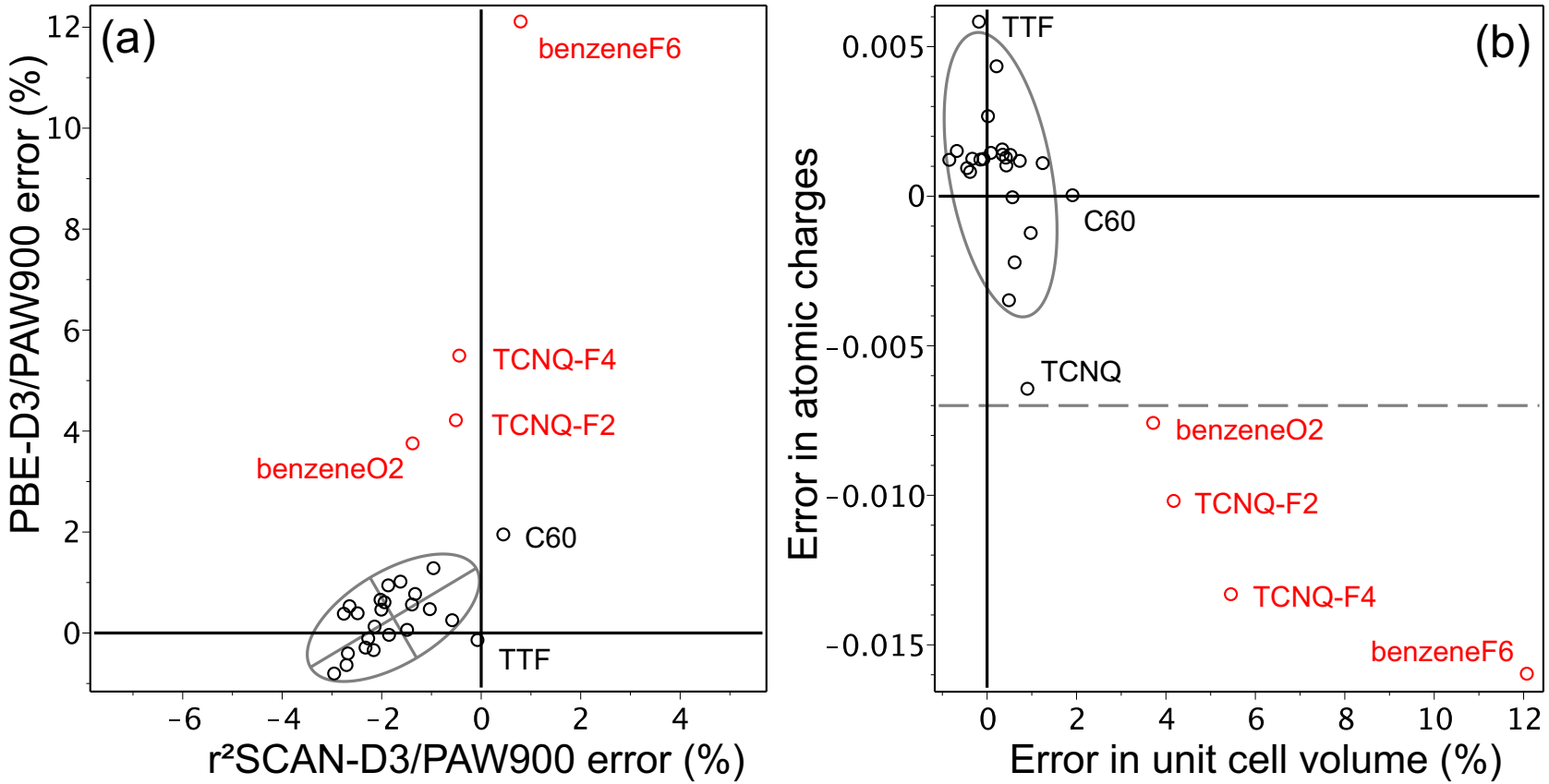
- Binding energy `Eb` (eV)
- Bulk modulus `K` (GPa) and its derivative `K'`, `MolMod/EOSfit`
- Elasticity tensor `C` (GPa) and derived quantities [Sewell2003]
- Vibrational frequencies at Γ-point and derived quantities $F(T)$
- Other parameters can be added later



Methodology of methods benchmarking

- Experimental data from Cambridge Structural Database
- DFT-D tested on a subset of 28 systems for which 0 K structure can be extrapolated
- Benchmarked functionals: PBE-D3 and r^{2}SCAN-D3
- Selectively benchmarked: PBE-MBD and vdW-DF2
- DFTB and xTB benchmarked against DFT-D
- force fields – *we are open for collaboration*

Results: PBE-D3 and r^{2}SCAN-D3



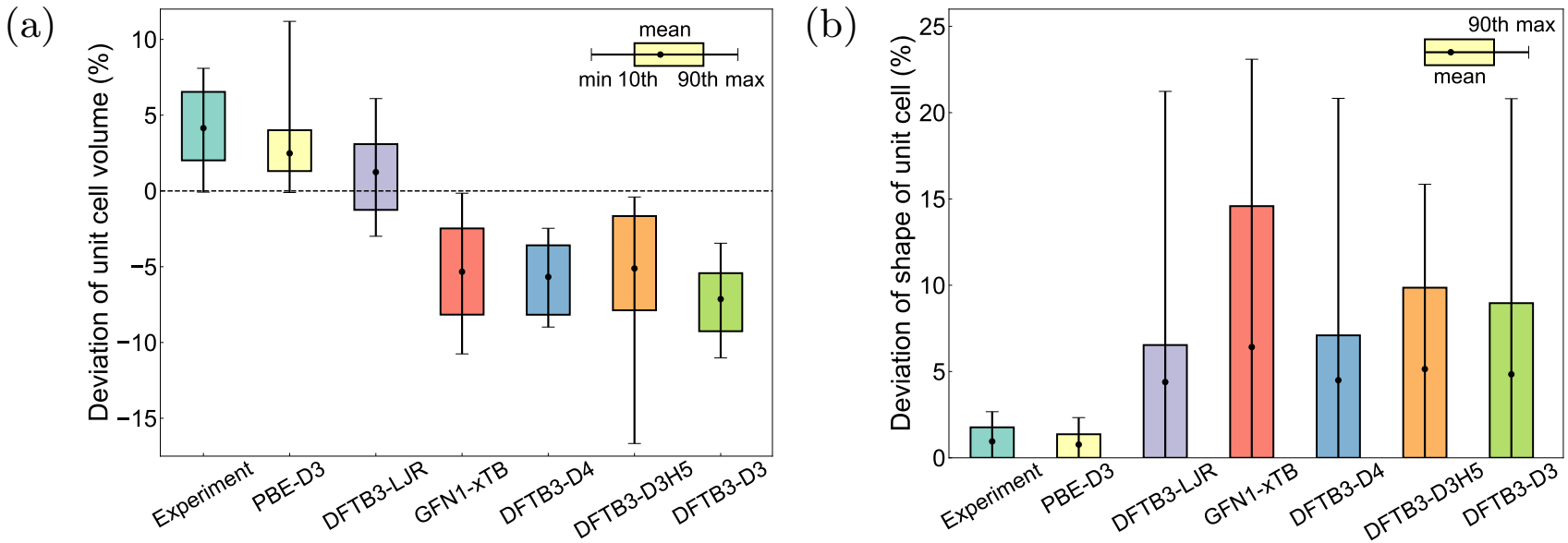
PBE-D3 – Optimal trade-off between accuracy and cost

- Inaccurate for large atomic charges, but might be corrected

r^{2}SCAN-D3 – Robust but computationally demanding

- Underestimates unit cell volume by 2%
- Slower than PBE-D3 by factor of 2-10

Results: DFTB and xTB vs PBE-D3



Tolerable mean deviation but unacceptable max error

Conclusions

- Benchmark dataset of crystalline organic semiconductors has been proposed
- PBE-D3 is an inexpensive "almost fault-free" method with accuracy comparable to experimental uncertainty
- r^2SCAN-D3 is robust but computationally demanding
- Known DFTB/xTB parameterizations give tolerable mean deviation but unacceptable max error

Appendix

Geometry optimization of molecular crystals

Must be fast, accurate, and produce usable output (will be used many times)

Common mistake: *spend minutes for geometry optimization and then hours to transform the output into a usable form*

- Preprocessing
  - clean up initial geometry (remove disorder, add hydrogens)
- Geometry optimization
  - Conserve symmetry
  - Large dynamic range of forces, PES is often flat
  - Unit cell optimization is nontrivial, especially angles
- Postprocessing
  - Symmetrize, connect molecules, reorder atoms for consistency
  - Keep required info (entire cell, molecules partitioning)