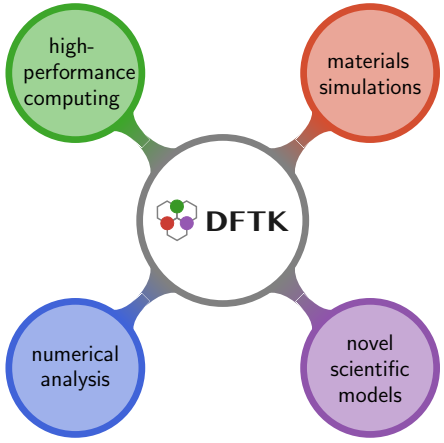
# Errors and uncertainty quantification in electronic-structure theory

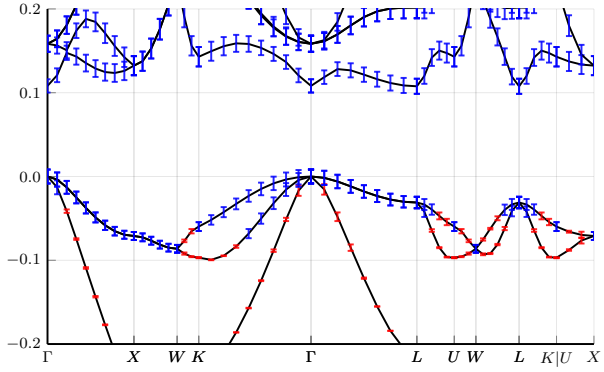
Michael F. Herbst

*Applied and Computational Mathematics, RWTH Aachen University
https://michael-herbst.com

8th June 2021

Slides: https://michael-herbst.com/talks/2021.06.08_mit_dft_uq.pdf



# About this talk

- A talk of questions, not of answers
- A biased summary
- Stimulate discussion

$\Rightarrow$ Please interrupt me if . . .

- . . . you have a question
- . . . you have a suggestion
- . . . you have no clue why one would possibly do this

# Contents

# Contents

# Societal challenges of 21st century

- Renewable energy
- Green chemistry and catalysts
- Drug design
- Transportation
- Data storage and communication

$\Rightarrow$ Need for novel materials

$\Rightarrow$ High-throughput computational screening

# High-throughput screening: Approach

- Given: Design space of novel materials
- Aim: Optimise wrt. target descriptors:
  1. Solve physical model for material
  2. Computed derived properties
- Computational approaches / fidelities:
  - Empirical models (Statistical surrogates, coarse-graining)
  - Density-functional theory (DFT)
  - Post-DFT methods (GW, . . . )
- Statistical learning:
  - Still needs (lots of) high-level samples

# Typical scale

- One DFT calculation: $\mathcal{O}$(hours) to $\mathcal{O}$(days)
- E.g. Open Catalyst Project[1]
  - $1.3$ million DFT calculations
  - $> 250$ million DFT energy evaluations
  - Workflow success rate: $\simeq 50\%$[2]

$\Rightarrow$ Reliability needs to be improved!

$\Rightarrow$ Need for careful understanding of errors

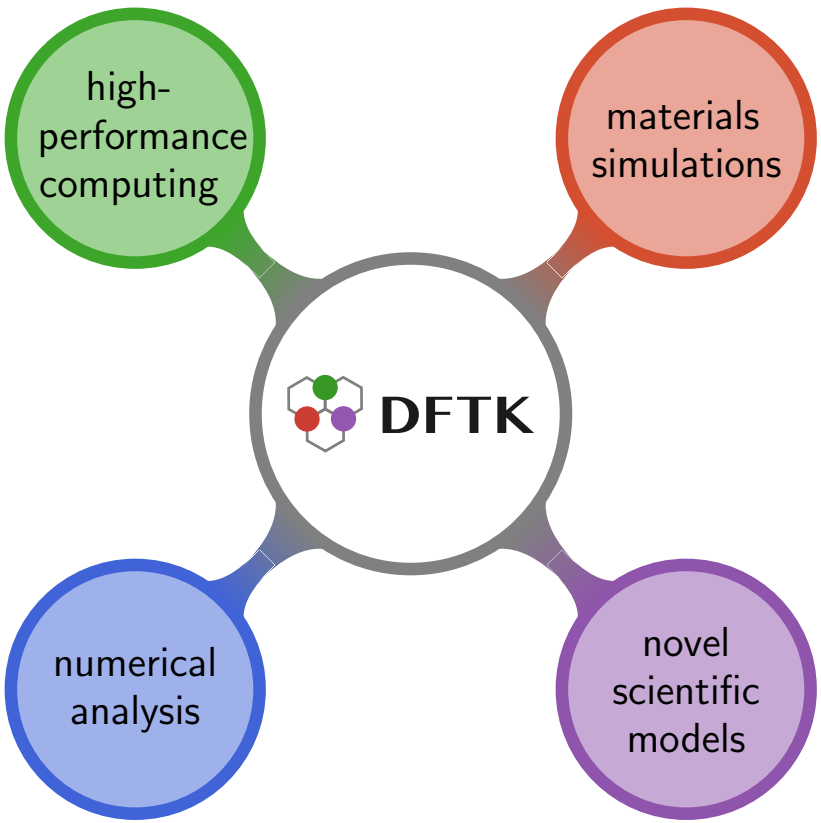
$\Rightarrow$ How much effort is really needed?

- Multidisciplinary research problem

[1] L. Chanussot *et. al.* The Open Catalyst 2020 (OC20) Dataset, 2020, arXiv 2010.09990.

[2] Z. Ulissi, private communication in ARPAE differentiate group seminar, Dec 2020.

# Density-functional toolkit (DFTK)[1]




- `https://dftk.org`
- 2 years of development
- Pure julia code

- Supports mathematical developments *and* scale-up to regime relevant to applications
- Low entrance barrier: Only 6k lines of code!

- International and interdisciplinary user base:
  - Analysis, mathematical physics, applications, ...

[1] M. F. Herbst, A. Levitt and E. Cancès. JuliaCon Proc., **3**, 69 (2021).

# Contents

# The essence of density-functional theory

$$\gamma_0 = \operatorname*{arg\,min}_{\gamma \in \mathcal{P}_N} \mathcal{E}_{\mathsf{DFT}}(\gamma)$$

- Energy functional

$$\mathcal{E}_{\mathsf{DFT}}(\gamma) = \operatorname{tr}_{L^2}\left(-\frac{1}{2}\Delta\gamma\right) + \int \rho_\gamma(\underline{\boldsymbol{r}}) V_{\mathsf{ext}}(\underline{\boldsymbol{r}})\,\mathrm{d}\underline{\boldsymbol{r}} + \frac{1}{2}\int \rho_\gamma(\underline{\boldsymbol{r}}) v_C(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') \rho_\gamma(\underline{\boldsymbol{r}}')\,\mathrm{d}\underline{\boldsymbol{r}}\,\mathrm{d}\underline{\boldsymbol{r}}' + E_{\mathsf{xc}}(\rho_\gamma)$$

- Density matrix $\gamma \in \mathcal{P}_N$, $0 \leq \gamma \leq 1$, $\operatorname{tr}_{L^2}(\gamma) = N$
- Density $\rho_\gamma(\underline{\boldsymbol{r}}) = \gamma(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}})$
- Coulomb kernel $v_C(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') = \|\underline{\boldsymbol{r}} - \underline{\boldsymbol{r}}'\|^{-1}$
- Exchange-correlation energy $E_{\mathsf{xc}}(\rho_\gamma)$
- External potential $V_{\mathsf{ext}}$

# Euler-Lagrange equation

- Coupled set of non-linear elliptic partial differential equations:

$$\Big(-\frac{1}{2}\Delta + \mathcal{V}(\rho)\Big)\psi_i = \varepsilon_i \psi_i, \quad \int \psi_i^* \psi_j = \delta_{ij}$$

$$\rho = \sum_{i=1}^{\infty} f\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i|^2, \quad \text{with } \varepsilon_F \text{ such that } \int \rho = N$$

- Density-dependent potential

$$\mathcal{V}(\rho) = V_{\text{ext}} + \int (v_C \rho) + V_{\text{xc}}(\rho)$$

with Coulomb kernel $v_C(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') = \|\underline{\boldsymbol{r}} - \underline{\boldsymbol{r}}'\|^{-1}$

- Fermi-Dirac function $f(x) = 1/(1+e^x)$
- Temperature $T$, electron count $N$
- Exchange-correlation potential $V_{\text{xc}} = \frac{\mathrm{d}E_{\text{xc}}(\rho)}{\mathrm{d}\rho}$

# Self-consistent field (SCF) as a fixed-point problem

- Density-dependent potential

$$\mathcal{V}(\rho) = V_{\text{ext}} + \int (v_C \rho) + V_{\text{xc}}(\rho)$$

- Potential-to-density map $F$

$$F(V) = \sum_{i=1}^{\infty} f\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i|^2$$

with $(\varepsilon_i, \psi_i)$ eigenpairs of $-\frac{1}{2}\Delta + V$.

$\Rightarrow$ SCF solves $\rho = F(\mathcal{V}(\rho))$

- Numerically: Damped fixed-point scheme

$$\rho_{n+1} = \rho_n + \alpha P^{-1} \left[F(\mathcal{V}(\rho_n)) - \rho_n\right]$$

with preconditioner ("mixing") $P$

# DFT properties and descriptors

- Assume fixed-point problem solved: $\rho^* = \arg\min_\rho \mathcal{E}_{\mathsf{DFT}}(\rho)$
- DFT properties:
  - Derivatives $\dfrac{\mathrm{d}\mathcal{E}^\lambda_{\mathsf{DFT}}(\rho^*)}{\mathrm{d}\lambda}$
  - $\lambda$: Physical system-dependent parameter of $V_{\mathsf{ext}}$
  - Examples: Forces, stresses, dipole moment, phonons, ...
  - Total derivative: Might require derivative of $\rho(\lambda)$
- Descriptors to compute for screening:
  - DFT quantities: $\mathcal{E}_{\mathsf{DFT}}$, $\varepsilon_i$, ... ($\simeq$ stability)
  - DFT properties ($\simeq$ usefulness)

# DFT takeaways

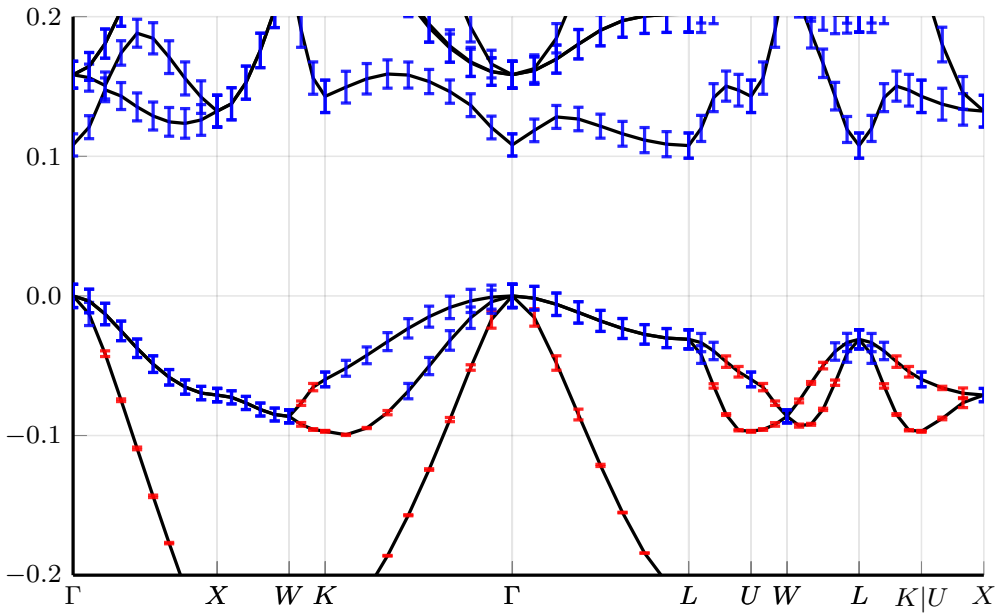
- Nested iterative problem:
  - Fixed point search, with diagonalisation for each iterate $\rho_n$
- Multi-step procedure:
  - (Geometry optimisation)
  - Fixed-point problem
  - Properties
- Numerous modelling parameters:
  - Model parameters: $E_{\mathsf{xc}}$?
  - Numerical parameters: Discretisation , $T$, ...
  - Algorithmic parameters: Tolerances, preconditioners, ...
  - Arithmetic parameters: Floating-point type, ...

# Contents

# Perspective of error estimation in DFT

- Nowadays: Parameter selection by trial and error
- Sources of error in DFT:
  - Model error: $\leftrightarrow E_{\text{xc}}$
  - Numerical error: $\leftrightarrow$ Discretisation , $T$, ...
  - Algorithmic error: $\leftrightarrow$ Tolerances, preconditioners, ...
  - Arithmetic error: $\leftrightarrow$ Floating-point type, ...
- Robust error estimates:
  - Total error known $\Rightarrow$ Error bars
  - Error-guided adaptive numerics $\Rightarrow$ Efficiency gain
- Vision: Robust automatic selection of numerical parameters

# A posteriori error analysis: First results with DFTK [1]




- Reduced model: Non-self-consistent Kohn-Sham
- Estimation of full numerical error
- Discretisation + algorithm + arithmetic
- Extensions ongoing work . . .

[1] M. F. Herbst, A. Levitt and E. Cancès. Faraday Discuss. **224**, 227 (2020).

# Error categories

- Numerical error:
  - Discretisation + algorithm + arithmetic
  - Promising *a posteriori* error estimates emerging
  - But: Usually the smaller contribution to the total error
  - Reason: Calculations "over-converged"

  $\Rightarrow$ Source for extra performance

- Model error ($E_{\mathrm{xc}}$):
  - Plethora of fidelity options (more in a sec)
  - *A posteriori* error estimates very challenging

  $\Rightarrow$ Regime of UQ / statistical methods

  $\Rightarrow$ Answers: How much numerical accuracy needed?

# Contents

# DFT model classes

- "Jacob's ladder" hierarchy for $E_{\text{xc}}$:
  - LDA, GGA, meta-GGA, Hybrid, RPA-like, ...
  - Each "step" defines parametrised model $\mathcal{E}^{\alpha}_{\text{DFT}}$
  - Hundreds of members differing by $\alpha$
  - Found by fitting and/or from physics
  - Parameter space increases $\simeq$ models get more accurate
- Additional correction terms (+U, dispersion, counterpoise, ...)

- DFT is a *non-variational* approximation to exact physics

$\Rightarrow$ No strictly guaranteed ranking

# Quantifying model uncertainty

- Error measures: Mean error, mean absolute error, relative error
- Conclusions depend on error measure[1]
- Best: Problem-specific error measure (e.g. $\Delta$-test)
- But: Hardly any specific ones developed

- Predominant generic approach: Regression-based methods[2]

$\Rightarrow$ Not predictive

[1] B. Civalleri, D. Presti, R. Dovesi, A. Savin, *On choosing the best density functional approximation*, in: Chemical Modelling: Applications and Theory, 168 (2012).

[2] K. Lejaeghere. *The uncertainty pyramid for electronic-structure methods*, 41. Elsevier (2020).

# The need for predictive UQ

- Screening decisions: Comparison with best case
- I.e. both best case *and* design space are simulated
- Then descriptors compared

- Model uncertainty usually neglected

$\Rightarrow$ Errors in differences can be large

$\Rightarrow$ Without UQ: Comparison of simulations can be misleading[1]

[1] G. Houchins, D. Krishnamurthy and V. Viswanathan. MRS Bull., **44**, 204 (2019).

# BEEF: Predictive UQ for the model error[1]

- BEEF: Bayesian Error Estimation Functional
- Based on ideas from Bayesian statistics
- BEEF approach:
  - Given some observations of a quantity $q$ (e.g. experiment)
  - Find "best" choice for $\alpha$ on a "step" by fit
  - Additionally: Distribution $P(\alpha)$ constructed
  - Aim: Ensemble spread of $\mathcal{E}_{\mathrm{DFT}}^{\alpha}$ represents model error
- Challenges for construction:
  - No model class is exact
  - Approximation known to not contain exact result

[1] R. Christensen, T. Bligaard and K. W. Jacobsen. *Bayesian error estimation in density functional theory*, 77-91. Elsevier (2020).

# BEEF: Key idea[1]

- Consider $N$ observations $q_i$, model predictions $q_i(\alpha)$, least-squares cost $C(\alpha) = \sum_i (q_i - q_i(\alpha))^2$
- Assume we found $\alpha^* = \arg\min_\alpha C(\alpha)$
- For a model $\mathcal{E}^\alpha_{\mathrm{DFT}}$ define
  - Error $\Delta q_i = q_i - q_i(\alpha)$
  - Model deviation $\delta q_i(\alpha) = q_i(\alpha) - q_i(\alpha^*)$
- Goal: Spread on $P(\alpha)$ should be model deviation (on average)

$$\sum_i \int [\delta q_i(\alpha)]^2 P(\alpha)\, \mathrm{d}\alpha = \sum_i (\Delta q_i)$$

- By maximum entropy principle (introduce least bias):

$$P(\alpha) \propto \exp\left(-\frac{NC(\alpha)}{2C(\alpha^*)}\right).$$

[1] V. Petzold, T. Bligaard and K. W. Jacobsen. Top. Catal., **55**, 402 (2012).

# BEEF: Usage in practice

- $q$ typically an energy difference quantity
  - E.g. Formation energy: $\mathcal{E}_{\text{DFT}}^{\text{compound}} - \mathcal{E}_{\text{DFT}}^{\text{atoms}}$
- Offline: Determine $\alpha^*$
  - Pick one DFT model class
  - Fit against data $q_i$
- Solve DFT: $\rho^* = \arg\min_\rho \mathcal{E}_{\text{DFT}}^{\alpha^*}(\rho)$, $\mathcal{E}^* = \mathcal{E}_{\text{DFT}}^{\alpha^*}(\rho^*)$
- Sample $\tilde{\alpha}$ from $P$:
  - One-shot computation: $\mathcal{E}_{\text{DFT}}^{\tilde{\alpha}}(\rho^*)$ (*same* $\rho^*$, different $\alpha$)
  - Resulting energy distribution: Uncertainty estimate
- Propagate distribution forward to quantity of interest $q'$

# BEEF: Issues and open questions

- BEEF-based approaches termed *uncertainty quantification*
- Issues:
  - Sensitivity of model parameters $\neq$ error to experiment
  - Experimental error ignored (just regularisation)
  - Self-consistency not treated
  - Limitation to energies (and derivatives)
- Open questions:
  - Rigid mathematical justification for this framework?
  - Usually $q' = q$, but not always: Is error propagation valid?
  - What if numerics is not perfect? Experimental error?

# Multi-fidelity approaches

- Natural approach for design-space searches:
  - Far from optimum: Full accuracy not needed
- Possible axes of fidelity:
  - Discretisation parameters (some related, some not)
  - DFT models
- Initial developments[1,2]
  - Two *a priori* defined fidelity layers
  - Co-kriging, Gaussian process regression approaches
  - Two models *or* two accuracies

[1] G. Pilania, J. Gubernatis and T. Lookman. Comput. Mater. Sci., **129**, 156 (2017).

[2] R. Batra, G. Pilania, B. P. Uberuaga et al. ACS Appl. Mater. Interfaces, **11**, 24906 (2019).

# Multi-fidelity: Open questions and challenges

- Which approach?
  - Kriging, Bayesian optimisation, ...
- Design space $\leftrightarrow$ feature space for learning
- Experimental design and sampling strategy?
  - ...in feature space
  - ...in design space
- Adaptive definition of numerical fidelity layers?
- DFT models: No strict ordering of improvement
  - At best empirically for selected classes of materials
- Selection of good features

# Opportunities for UQ: Uses of DFT data

- ✓ Design-space searches
- Potential fitting
  - For molecular dynamics
  - Propagation of DFT error / fitting error?
- Geometry optimisation
  - Sensitivity of DFT model?

# Summary

https://michael-herbst.com/talks/2021.06.08_mit_dft_uq.pdf

- Density-functional theory:
  - Key method for high-throughput materials discovery
  - Nested iterative procedures
  - Sizable number of numerical parameters: Reliability issue
- Error estimation in DFT:
  - *A posteriori* error estimation: Numerical error
  - Model error estimation $\Rightarrow$ UQ research needed
- Perspectives and efficiency gains by better UQ:
  - Error balancing model $\leftrightarrow$ numerics
  - Multi-fidelity design-space searches
  - Error propagation beyond energy quantities

# Questions?

https://michael-herbst.com/talks/2021.06.08_mit_dft_uq.pdf

 mfherbst

 https://michael-herbst.com/blog

 herbst@acom.rwth-aachen.de