Exchange-Correlation Functional with Uncertainty Quantification Capabilities for Density Functional Theory

Manuel Aldegunde

M.A.Aldegunde-Rodriguez@Warwick.ac.uk
Warwick Centre of Predictive Modelling (WCPM)
The University of Warwick

October 27, 2015

http://www2.warwick.ac.uk/wcpm/

Outline

Introduction

- Density Functional Theory (DFT) has become the most widespread technique to study materials in a quantum-mehanical framework
- It can provide a favourable trade-off between accuracy and computation
- It is commonly implemented using approximations to make it more manageable:

Introduction

- Density Functional Theory (DFT) has become the most widespread technique to study materials in a quantum-mehanical framework
- It can provide a favourable trade-off between accuracy and computation
- It is commonly implemented using approximations to make it more manageable:
  - simplify physics, e. g., Born-Oppenheimer approximation
  - numerical approximations, e.g., pseudopotentials, PAW, ...

Introduction

- Density Functional Theory (DFT) has become the most widespread technique to study materials in a quantum-mehanical framework
- It can provide a favourable trade-off between accuracy and computation
- It is commonly implemented using approximations to make it more manageable:
  - simplify physics, e. g., Born-Oppenheimer approximation
  - numerical approximations, e.g., pseudopotentials, PAW, ...
- One approximation is necessary:

Introduction

- Density Functional Theory (DFT) has become the most widespread technique to study materials in a quantum-mehanical framework
- It can provide a favourable trade-off between accuracy and computation
- It is commonly implemented using approximations to make it more manageable:
  - simplify physics, e. g., Born-Oppenheimer approximation
  - numerical approximations, e.g., pseudopotentials, PAW, ...
- One approximation is necessary:
  - Exchange-Correlation Functional (unknown in general)

Density Functional Theory

- Approximates the ground state energy of a material system with charge density n.
- Minimisation of the energy functional $E^{DFT}[n]$ for a given system

$$E^{DFT}[n] = \int n(\mathbf{r})v(\mathbf{r})\,d\mathbf{r} + T_0[n] + U[n] + E^{xc}[n]$$
$$= E^b[n] + E^{xc}[n] = E^b[n] + E^x[n] + E^c[n]$$

[1] R. Jones *et al.* (1989), R. Jones (2015)

Density Functional Theory

- Approximates the ground state energy of a material system with charge density n.
- Minimisation of the energy functional $E^{DFT}[n]$ for a given system

$$E^{DFT}[n] = \int n(\mathbf{r})v(\mathbf{r})\,d\mathbf{r} + T_0[n] + U[n] + E^{xc}[n]$$
$$= E^b[n] + E^{xc}[n] = E^b[n] + E^x[n] + E^c[n]$$

- $E^{xc}[n]$ not known... Need to specify an approximation

[1] R. Jones *et al.* (1989), R. Jones (2015)

Density Functional Theory: Kohn-Sham method

- Kohn and Sham (1965) introduced a methodology to solve the equations: most common formulation nowadays

Density Functional Theory: Kohn-Sham method

- Kohn and Sham (1965) introduced a methodology to solve the equations: most common formulation nowadays
- Solve a self-consistent problem using independent electrons in an effective potential:
  - $\left[-\frac{1}{2}\nabla^2 + v_{eff}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r})$
  - $v_{eff}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\mathbf{r}' + \frac{\delta E^{xc}[n]}{\delta n(\mathbf{r})}$
  - $n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2$

DFT: Exchange-correlation energies

- How to approximate $E^{xc}[n]$?
- We can write $E^{xc}[n] = \int n\varepsilon^{xc}(n;\mathbf{r})\,d\mathbf{r}$
  - $n\varepsilon^{xc}(n;\mathbf{r})$: XC energy density

[1] J. P. Perdew *et al.* (2001)

DFT: Exchange-correlation energies

- How to approximate $E^{xc}[n]$?
- We can write $E^{xc}[n] = \int n\varepsilon^{xc}(n;\mathbf{r})\,d\mathbf{r}$
  - $n\varepsilon^{xc}(n;\mathbf{r})$: XC energy density
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r})]$: LDA

[1] J. P. Perdew *et al.* (2001)

DFT: Exchange-correlation energies

- How to approximate $E^{xc}[n]$?
- We can write $E^{xc}[n] = \int n\varepsilon^{xc}(n;\mathbf{r})\,d\mathbf{r}$
  - $n\varepsilon^{xc}(n;\mathbf{r})$: XC energy density
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r})]$: LDA
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r}), \nabla n(\mathbf{r})]$: GGA

[1] J. P. Perdew *et al.* (2001)

DFT: Exchange-correlation energies

- How to approximate $E^{xc}[n]$?
- We can write $E^{xc}[n] = \int n\varepsilon^{xc}(n;\mathbf{r})\,d\mathbf{r}$
  - $n\varepsilon^{xc}(n;\mathbf{r})$: XC energy density
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r})]$: LDA
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r}), \nabla n(\mathbf{r})]$: GGA
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r})]$: meta-GGA
    $\left(\tau(\mathbf{r}) = 2\sum_i^{'} \frac{1}{2}\left|\nabla\psi_i(\mathbf{r})\right|^2\right)$

[1] J. P. Perdew *et al.* (2001)

DFT: Exchange-correlation energies

- How to approximate $E^{xc}[n]$?
- We can write $E^{xc}[n] = \int n\varepsilon^{xc}(n;\mathbf{r})\,d\mathbf{r}$
  - $n\varepsilon^{xc}(n;\mathbf{r})$: XC energy density
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r})]$: LDA
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r}), \nabla n(\mathbf{r})]$: GGA
  - $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r})]$: meta-GGA
    $\left(\tau(\mathbf{r}) = 2\sum_i{}' \frac{1}{2}\left|\nabla\psi_i(\mathbf{r})\right|^2\right)$
- We can add exact exchange: $E^x[n] = -\frac{1}{2}\sum_i \int \frac{\psi_i^*(\mathbf{r})\psi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}d\mathbf{r}'$
  - But still need to approximate the correlation energy...

[1] J. P. Perdew *et al.* (2001)

DFT: Exchange-correlation energies

- The double integral in the exact exchange makes it much more costly
- We chose the meta-GGA framework to build our approximation,

$$E^{xc}[n] = \int n\varepsilon^{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r}))\, d\mathbf{r}$$

- We will focus on exchange energy only,

$$E^{xc}[n] = E^{c}[n] + \int n\varepsilon^{x}(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r}))\, d\mathbf{r}$$

DFT: Exchange energy

- We can transform the dependence on $\nabla n(\mathbf{r})$ and $\tau(\mathbf{r})$ into two dimensionless parameters s and α:

$$s = \frac{|\nabla n|}{2(3\pi^2)^{1/3} n^{4/3}}; \quad \alpha = \frac{\tau - \tau^W}{\tau^{UEG}},$$

  - $\tau^W = |\nabla n|^2 / 8n$: Weizsäcker kinetic energy density
  - $\tau^{UEG} = \frac{3}{10}(3\pi^2)^{2/3} n^{5/3}$: UEG kinetic energy density
- Also, we can group all non-local contributions in the *exchange enhancement factor* $F^x(s, \alpha)$,

$$E^x[n] = \int n\varepsilon^x(n, \nabla n, \tau)\, d\mathbf{r} = \int n\varepsilon^x_{UEG}(n) F^x(s, \alpha)\, d\mathbf{r}$$

Linear model for the enhancement factor

- To specify a model for the exchange energy we just need to specify a model for the enhancement factor

Linear model for the enhancement factor

- To specify a model for the exchange energy we just need to specify a model for the enhancement factor
- We specify a linear model

$$F^x(s,\alpha) = \sum_i \xi_i^x \phi_i(s,\alpha) = (\boldsymbol{\xi}^x)^T \boldsymbol{\phi}(s,\alpha)$$

  - $\phi_i(s,\alpha)$: basis functions
  - ξ_i^x: linear model coefficients

Linear model for the exchange energy

- Given this model, the exchange energy becomes a linear model

$$E^x[n;\boldsymbol{\xi}^x] = \sum_{i=1}^{M} \xi_i^x \int n\varepsilon_{UEG}^x(n)\phi_i(s,\alpha)\,d\mathbf{r}$$
$$= \sum_{i=0}^{M-1} \xi_i^x E^x[n;\hat{\mathbf{e}}_i] = (\boldsymbol{\xi}^x)^T \mathbf{E}^x[n;\hat{\mathbf{e}}]$$

  - $E^x[n;\hat{\mathbf{e}}_i] = \int n\varepsilon_{UEG}^x(n)\phi_i(s,\alpha)\,d\mathbf{r}$ is the "basis exchange energy", which is obtained if we use $\boldsymbol{\xi}^x = \hat{\mathbf{e}}_i$, i.e., only basis ϕ_i with $\xi_i = 1$,

Linear model for the exchange energy

- How do we choose the basis functions $\phi_i(s, \alpha)$?
  - Follow selection in Wellendorff *et al.*
  - Physically based: inspired in previous non-empirical functionals (PBEsol, MS)
  - Complete basis: 2D Legendre Polynomials

[1] J. Wellendorff *et al.* (2014)

Linear model for the exchange energy

- How do we choose the basis functions $\phi_i(s, \alpha)$?
  - Follow selection in Wellendorff *et al.*
  - Physically based: inspired in previous non-empirical functionals (PBEsol, MS)
  - Complete basis: 2D Legendre Polynomials
- 2D Legendre polynomials: argument in [-1, 1]
- Rational transformations from s, α to the interval [-1, 1]

$$PBEsol \rightarrow t_s(s) = \frac{2s^2}{q + s^2} - 1$$

$$MS \rightarrow \quad t_\alpha(\alpha) = \frac{(1 - \alpha^2)^3}{1 + \alpha^3 + \alpha^6}$$

[1] J. Wellendorff *et al.* (2014)

Linear model for the exchange energy

- The final exchange enhancement model is

$$F^x(s,\alpha) = \sum_i^{M_s} \sum_j^{M_\alpha} \xi_{ij}^x P_i(t_s(s)) P_j(t_\alpha(\alpha))$$

- The final exchange energy model is therefore

$$E^x[n;\boldsymbol{\xi}^x] = \sum_i^{M_s} \sum_j^{M_\alpha} \xi_{ij}^x \int n\varepsilon_{UEG}^x(n) P_i(t_s(s)) P_j(t_\alpha(\alpha))\, d\mathbf{r}$$

Linear model for the exchange energy

- The final exchange enhancement model is

$$F^x(s,\alpha) = \sum_i^{M_s} \sum_j^{M_\alpha} \xi_{ij}^x P_i(t_s(s)) P_j(t_\alpha(\alpha))$$

- The final exchange energy model is therefore

$$E^x[n;\boldsymbol{\xi}^x] = \sum_i^{M_s} \sum_j^{M_\alpha} \xi_{ij}^x \int n\varepsilon_{UEG}^x(n) P_i(t_s(s)) P_j(t_\alpha(\alpha))\, d\mathbf{r}$$

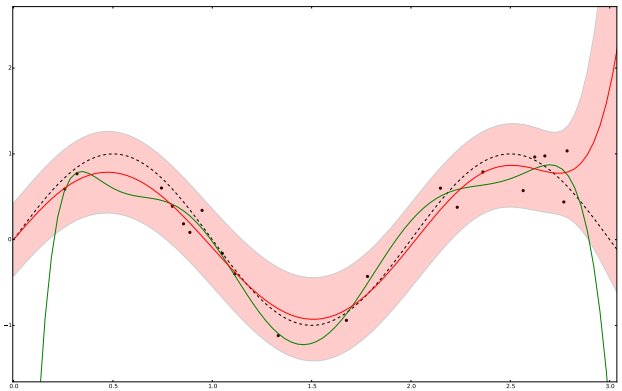
- How to obtain the coefficients?

Bayesian linear regression

- We want to account for uncertainty in the model
- Classical least squares fitting gives a point estimate, not appropriate
- A Bayesian model will give us probability distributions for the coefficients
  - Uncertainty from a limited data set for the regression
  - Uncertainty from an inadequate model

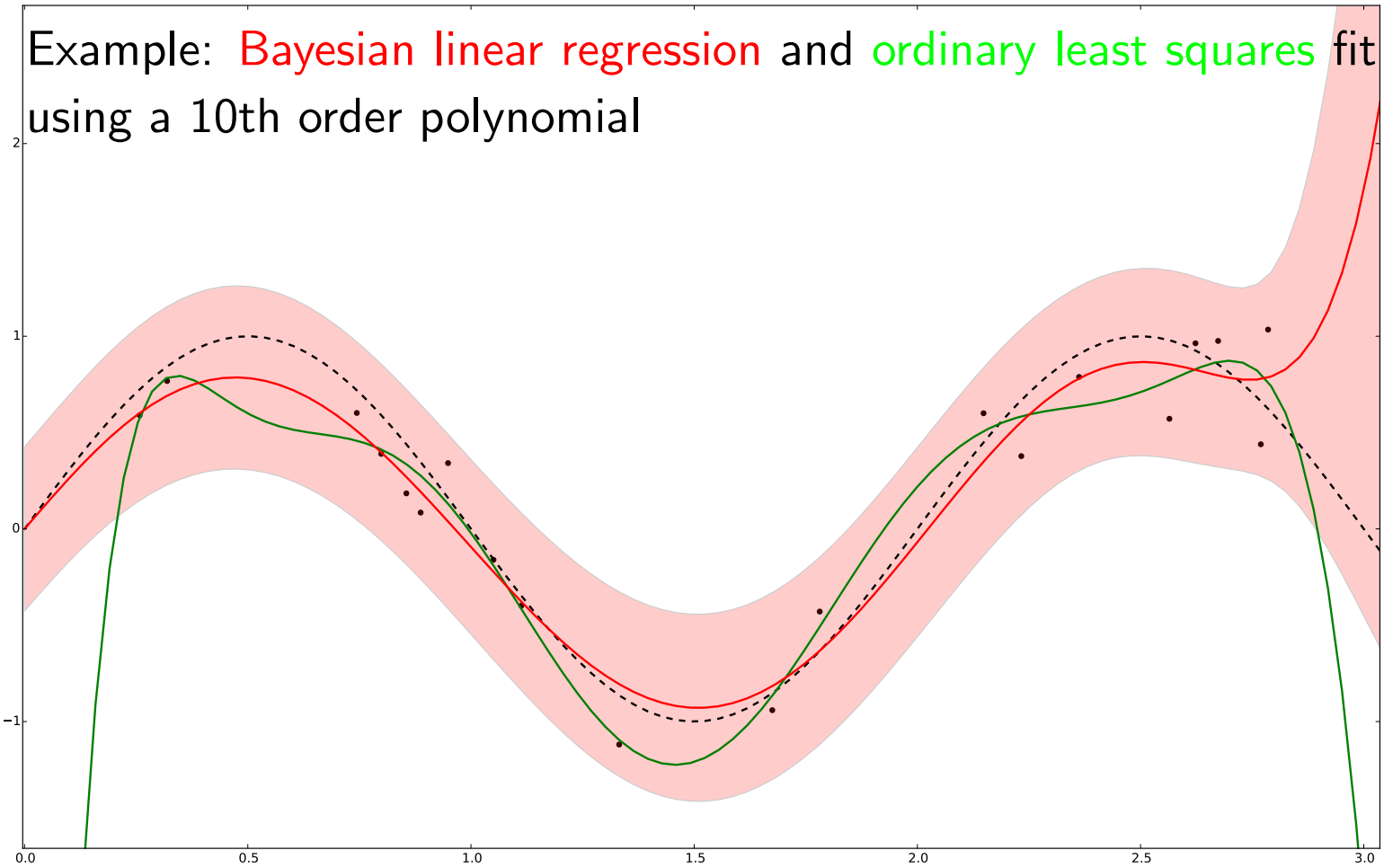
$$p(\boldsymbol{\xi}, \beta \mid \mathbf{t}) \propto \mathcal{L}(\mathbf{t} \mid \mathbf{x}, \boldsymbol{\xi}, \beta) p(\boldsymbol{\xi}, \beta)$$

Example: Bayesian linear regression and ordinary least squares fit using a 10th order polynomial



[1] C. Bishop (2006)

Bayesian linear regression



Some definitions

- $\mathbf{t} = (t_1, t_2, \ldots, t_N)^T$: given data (experimental, simulation, mix)
- $\mathbf{n} = (n_1, n_2, \ldots, n_N)^T$: input points (densities for DFT)
- $\mathcal{L}(\mathbf{t} \mid \mathbf{n}, \texttt{model})$: likelihood function
- $\mathcal{N}(x \mid \mu, v)$: normal distribution on x with mean μ and variance v
- $\mathcal{G}(x \mid \alpha, \beta)$: gamma distribution on x with parameters α and β
- $\mathcal{S}t(x \mid \mu, \lambda, \nu)$: Student t-distribution on x with parameters μ, λ and ν

Assumptions

- The observed data $\mathbf{t}$ follow on average our model and have a noise ε (includes model inaccuracy),

$$t_i = (\boldsymbol{\xi}^x)^T \mathbf{E}^x[n; \hat{\mathbf{e}}] + \varepsilon_i$$

Assumptions

- The observed data $\mathbf{t}$ follow on average our model and have a noise ε (includes model inaccuracy),

$$t_i = (\boldsymbol{\xi}^x)^T \mathbf{E}^x[n; \hat{\mathbf{e}}] + \varepsilon_i$$

- The noise is assumed Gaussian with precision $\beta = 1/v = 1/\sigma^2$ and uncorrelated, so that

$$t_i \sim \mathcal{N}(t \mid (\boldsymbol{\xi}^x)^T \mathbf{E}^x[n; \hat{\mathbf{e}}], \beta^{-1})$$

Assumptions

- The observed data $\mathbf{t}$ follow on average our model and have a noise ε (includes model inaccuracy),

$$t_i = (\boldsymbol{\xi}^x)^T \mathbf{E}^x[n; \hat{\mathbf{e}}] + \varepsilon_i$$

- The noise is assumed Gaussian with precision $\beta = 1/v = 1/\sigma^2$ and uncorrelated, so that

$$t_i \sim \mathcal{N}(t \mid (\boldsymbol{\xi}^x)^T \mathbf{E}^x[n; \hat{\mathbf{e}}], \beta^{-1})$$

- The likelihood function is, therefore,

$$\mathcal{L}(\mathbf{t} \mid \mathbf{n}, \boldsymbol{\xi}, \beta) = \prod_{i=1}^{N} \mathcal{N}(t_i \mid \boldsymbol{\xi}^T \mathbf{E}^x[n_i; \hat{\mathbf{e}}], \beta^{-1})$$

Assumptions

- The noise is assumed Gaussian with precision $\beta = 1/v = 1/\sigma^2$ and uncorrelated, so that

$$t_i \sim \mathcal{N}(t \mid (\boldsymbol{\xi}^x)^T \mathbf{E}^x[n; \hat{\mathbf{e}}], \beta^{-1})$$

- The likelihood function is, therefore,

$$\mathcal{L}(\mathbf{t} \mid \mathbf{n}, \boldsymbol{\xi}, \beta) = \prod_{i=1}^{N} \mathcal{N}(t_i \mid \boldsymbol{\xi}^T \mathbf{E}^x[n_i; \hat{\mathbf{e}}], \beta^{-1})$$

- We choose conjugate priors for $\boldsymbol{\xi}$ and β

Priors

- Incorporate prior beliefs into the model
- Depend on extra parameters: *hyperparameters*
- Conjugate priors keep the posterior propability distribution in the same family as the prior probability distribution
- Prior on $\boldsymbol{\xi}$: $p(\boldsymbol{\xi} \mid \beta, \mathbf{m}_0, \mathbf{S}_0) = \mathcal{N}(\boldsymbol{\xi} \mid \mathbf{m}_0, \beta^{-1}\mathbf{S}_0)$
- Prior on β: $p(\beta \mid a_0, b_0) = \mathcal{G}(\beta \mid a_0, b_0)$
- Joint prior: $p(\boldsymbol{\xi}, \beta) = p(\boldsymbol{\xi} \mid \beta)p(\beta) = \mathcal{N}(\boldsymbol{\xi} \mid \mathbf{m}_0, \beta^{-1}\mathbf{S}_0)\mathcal{G}(\beta \mid a_0, b_0)$

Posterior

- Probability of a set of parameters given the data
- Proportional to the prior distribution of parameters
- Proportional to the likelihood of the data

$$p(\boldsymbol{\xi}, \beta \mid \mathbf{t}) = \frac{\mathcal{L}(\mathbf{t} \mid \mathbf{x}, \boldsymbol{\xi}, \beta) p(\boldsymbol{\xi}, \beta)}{\int \mathcal{L}(\mathbf{t} \mid \mathbf{x}, \boldsymbol{\xi}, \beta) p(\boldsymbol{\xi}, \beta)\, d\boldsymbol{\xi}\, d\beta}$$
$$= \mathcal{N}(\boldsymbol{\xi} \mid \mathbf{m}_N, \beta^{-1}\mathbf{S}_N)\mathcal{G}(\beta \mid a_N, b_N)$$

Posterior

- The parameters of the posterior depend on those of the prior and the data,

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \mathbf{\Phi}^T\mathbf{\Phi}; \qquad \mathbf{m}_N = \mathbf{S}_N\left[\mathbf{S}_0^{-1}\mathbf{m}_0 + \mathbf{\Phi}^T\mathbf{t}\right]$$

$$a_N = a_0 + N/2$$

$$b_N = b_0 + \frac{1}{2}\left(\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 - \mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N + \mathbf{t}^T\mathbf{t}\right)$$

- $\mathbf{\Phi}$ is the *design matrix*

$$\mathbf{\Phi} = \begin{pmatrix} E^x[n_1^*;\hat{\mathbf{e}}_0] & \cdots & E^x[n_1^*;\hat{\mathbf{e}}_{M-1}] \\ \vdots & \ddots & \vdots \\ E^x[n_N^*;\hat{\mathbf{e}}_0] & \cdots & E^x[n_N^*;\hat{\mathbf{e}}_{M-1}] \end{pmatrix} = \begin{pmatrix} \mathbf{E}^x[n_1^*;\hat{\mathbf{e}}]^T \\ \vdots \\ \mathbf{E}^x[n_N^*;\hat{\mathbf{e}}]^T \end{pmatrix}$$

Posterior

- The parameters of the posterior depend on those of the prior and the data,
- $\mathbf{\Phi}$ is the *design matrix*

$$\mathbf{\Phi} = \begin{pmatrix} E^x[n_1^*; \hat{\mathbf{e}}_0] & \cdots & E^x[n_1^*; \hat{\mathbf{e}}_{M-1}] \\ \vdots & \ddots & \vdots \\ E^x[n_N^*; \hat{\mathbf{e}}_0] & \cdots & E^x[n_N^*; \hat{\mathbf{e}}_{M-1}] \end{pmatrix} = \begin{pmatrix} \mathbf{E}^x[n_1^*; \hat{\mathbf{e}}]^T \\ \vdots \\ \mathbf{E}^x[n_N^*; \hat{\mathbf{e}}]^T \end{pmatrix}$$

  - Rows: For a given data point, a vector with the "basis exchange energies"
  - Columns: For a given basis function, its value for every data point

Predictive distribution

- Once we have our probability distributions for model parameters, what can we say about new data points?

Predictive distribution

- Once we have our probability distributions for model parameters, what can we say about new data points?
- Prediction averaged over all possible parameters

Predictive distribution

- Once we have our probability distributions for model parameters, what can we say about new data points?
- Prediction averaged over all possible parameters
- Predictions given with probability distributions

$$
\begin{aligned}
p(\tilde{t} \mid \tilde{n}, \mathbf{t}) &= \int p(\tilde{t} \mid \tilde{n}, \boldsymbol{\xi}, \beta) p(\boldsymbol{\xi}, \beta \mid \mathbf{t}) \, d\boldsymbol{\xi} \, d\beta \\
&= \int \mathcal{N}(\tilde{t} \mid \boldsymbol{\xi}^T \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}], \beta^{-1}) \mathcal{N}(\boldsymbol{\xi} \mid \mathbf{m}_N, \beta^{-1} \mathbf{S}_N) \mathcal{G}(\beta \mid a_N, b_N) \, d\boldsymbol{\xi} \, d\beta \\
&= \mathcal{S}t(\tilde{t} \mid \mu, \lambda, \nu).
\end{aligned}
$$

Predictive distribution

- The Student t-distribution $\mathcal{S}t(\tilde{t} \mid \mu, \lambda, \nu)$ has parameters

$$\mu = \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}]^T \mathbf{m}_N$$
$$\lambda = \frac{a_N}{b_N} \left(1 + \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}]^T \mathbf{S}_N \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}]\right)^{-1}$$
$$\nu = 2a_N.$$

- Its mean, variance and mode are

$$\mathrm{E}[\tilde{t}] = \mu; \quad \nu > 1$$
$$\mathrm{cov}[\tilde{t}] = \frac{\nu}{\nu - 2}\lambda^{-1} = \frac{1 + \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}]^T \mathbf{S}_N \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}]}{\mathrm{mode}[\beta]}; \quad \nu > 2$$
$$\mathrm{mode}[\tilde{t}] = \mu$$

Predictive distribution

- Its mean, variance and mode are

$$\mathrm{E}[\tilde{t}] = \mu; \quad \nu > 1$$

$$\mathrm{cov}[\tilde{t}] = \frac{\nu}{\nu - 2}\lambda^{-1} = \frac{1 + \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}]^T \mathbf{S}_N \mathbf{E}^x[\tilde{n}; \hat{\mathbf{e}}]}{\mathrm{mode}[\beta]}; \quad \nu > 2$$

$$\mathrm{mode}[\tilde{t}] = \mu$$

  - The variance of the prediction depends on the data point

Predictive distribution

- If we make several predictions, they are correlated,

$$p(\tilde{\mathbf{t}} \mid \tilde{\mathbf{n}}, \mathbf{t}) = \int p(\tilde{\mathbf{t}} \mid \tilde{\mathbf{n}}, \boldsymbol{\xi}, \beta) p(\boldsymbol{\xi}, \beta \mid \mathbf{t})\, d\boldsymbol{\xi}\, d\beta = \mathcal{S}t(\tilde{\mathbf{t}} \mid \mu, \Lambda, \nu)$$

- The mean and covariance are

$$\mathrm{E}[\tilde{\mathbf{t}}] = \tilde{\mathbf{\Phi}} \mathbf{m}_N; \qquad \mathrm{cov}[\tilde{t}] = \frac{\mathbf{I} + \tilde{\mathbf{\Phi}} \mathbf{S}_N \tilde{\mathbf{\Phi}}^T}{\mathrm{mode}[\beta]}$$

  - $\tilde{\mathbf{\Phi}}$ is analogous to the design matrix where the rows are the points of the predictions

Hyperparameters: Evidence approximation

- One last point to be solved
- How do we obtain the hyperparameters?

Hyperparameters: Evidence approximation

- One last point to be solved
- How do we obtain the hyperparameters?
- We chose the evidence approximation: maximise the log of the marginal likelihood (*evidence function*)

$$\log p(\mathbf{t} \mid \mathbf{m}_0, \mathbf{S}_0, a_0, b_0) =$$

$$= log \int p(\mathbf{t} \mid \boldsymbol{\xi}, \beta, \mathbf{m}_0, \mathbf{S}_0, a_0, b_0) p(\xi, \beta \mid \mathbf{m}_0, \mathbf{S}_0, a_0, b_0) d\xi d\beta$$

$$= \mathcal{E}(\mathbf{m}_0, \mathbf{S}_0, a_0, b_0) = \frac{1}{2} \log \frac{|\mathbf{S}_N|}{|\mathbf{S}_0|} - \frac{N}{2} \log(2\pi) + \log \frac{\Gamma(a_N)}{\Gamma(a_0)} +$$

$$+ a_0 \log(b_0) - a_N \log(b_N)$$

Hyperparameters: Evidence approximation

- For a model with M parameters, $\mathbf{m}_0$, $\mathbf{S}_0^{-1}$, a_0 and b_0 have $\sim M^2$ parameters

[1] M. E. Tipping (2001), C. Bishop (2006)

Hyperparameters: Evidence approximation

- For a model with M parameters, $\mathbf{m}_0$, $\mathbf{S}_0^{-1}$, a_0 and b_0 have $\sim M^2$ parameters
- Using $\mathbf{m}_0 = 0$ and $\mathbf{S}_0^{-1} = \alpha\mathbb{I}$, we only have three hyperparameters and we can easily find a maximum of the evidence function (Bayesian ridge regression)

[1] M. E. Tipping (2001), C. Bishop (2006)

Hyperparameters: Evidence approximation

- For a model with M parameters, $\mathbf{m}_0$, $\mathbf{S}_0^{-1}$, a_0 and b_0 have $\sim M^2$ parameters
- Using $\mathbf{m}_0 = 0$ and $\mathbf{S}_0^{-1} = \alpha\mathbb{I}$, we only have three hyperparameters and we can easily find a maximum of the evidence function (Bayesian ridge regression)
- Using $\mathbf{S}_0^{-1} = diag(\alpha_0, \ldots, \alpha_{M-1})$, we have $M + 2$ hyperparameters (*Relevance Vector Machine*)
  - Induces sparsity (some of the α_i go to infinity)
  - Only keeps relevant terms: automatic model selection

[1] M. E. Tipping (2001), C. Bishop (2006)

Hyperparameters: Evidence approximation

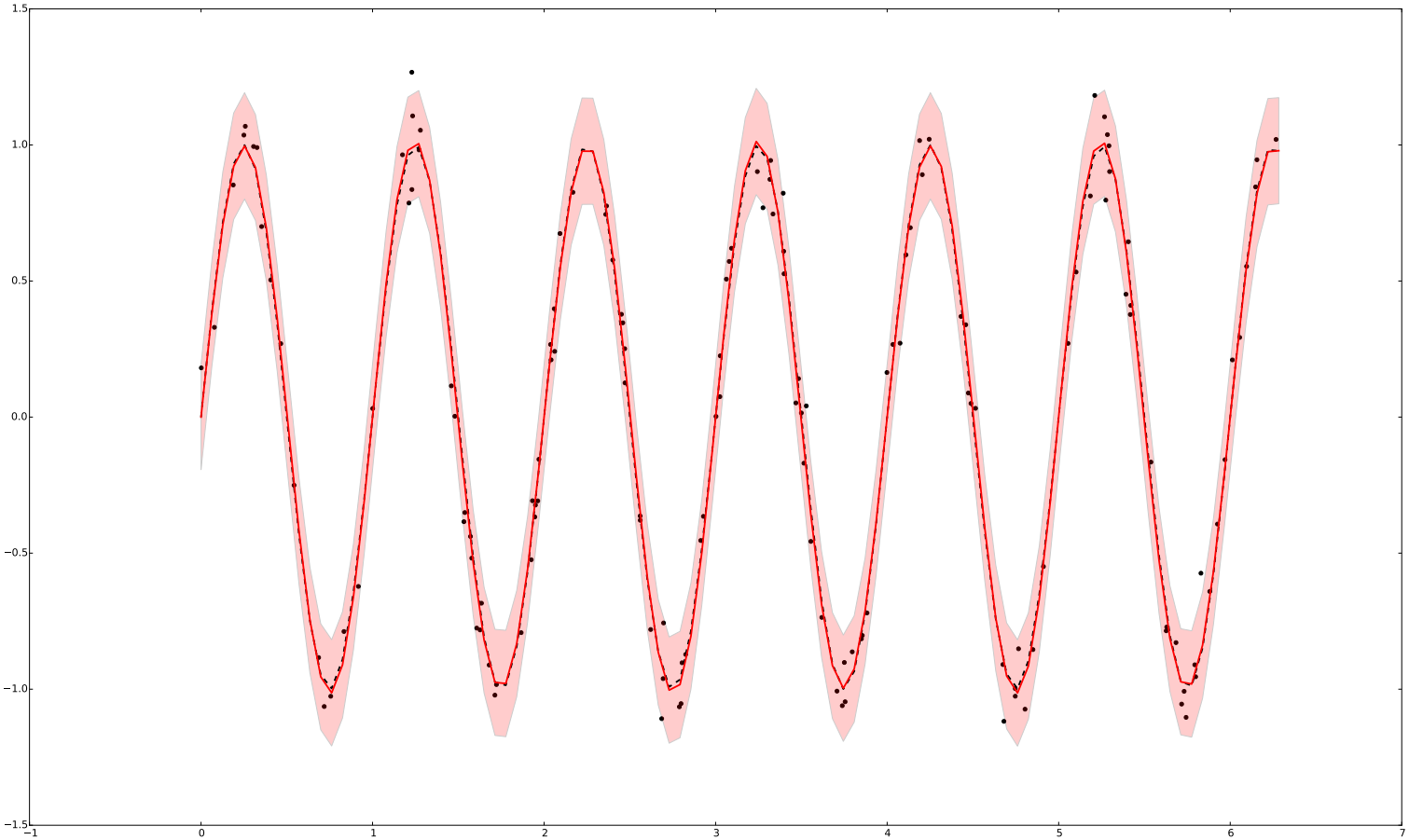
Algorithm 1 Hyperparameter optimisation for the RVM.

1: $\mathbf{S}_0^{-1} = diag(\alpha_0, \ldots, \alpha_{M-1})$, $\mathbf{m}_0 = 0$.
2: Initialize α_i from random numbers $r \in (0, 10^{10})$.
3: **repeat**
4: **repeat**
5: **for** all $i = 0, 1, \ldots, M-1$ **do**
6: Update α_i as $\alpha_i^{new} = \frac{1}{[\mathbf{S}_N]_{ii} + \frac{a_N}{b_N}[\mathbf{m}_N]_i^2}$.
7: Update $\mathbf{S}_N$, $\mathbf{m}_N$.
8: **end for**
9: **until** $\Delta\alpha_i < 10^{-5}\%$ or $\alpha_i > 10^{10}$
10: Update a_0, b_0 with a Newton iteration.
11: Update $\mathbf{S}_N$, $\mathbf{m}_N$.
12: **until** $\Delta\alpha, \Delta a_0, \Delta b_0 < 10^{-4}\%$

Relevance Vector Machine: Example

- Generate data from $f(x) = sin(2\pi x) + \varepsilon$
  - $\varepsilon \sim \mathcal{N}(\varepsilon \mid 0, 0.1^2)$
- Use 10 sine basis functions, $\sin(k\pi x)$; $\quad k = 0, 1, ..., 9$
- Fitted $\mathbf{m}_N = [0, 0.003, 1.006, -0.016, 0, 0, 0, 0, 0, 0]$, $mode[\sigma_N] = 0.097$
  - Only three coefficients remain

Relevance Vector Machine: Example



Training data

- The exchange energy cannot be measured
- Absolute energies cannot be measured
- How to train the model?
  - Easy to get from DFT simulations
  - Experimentally available

Training data

- The exchange energy cannot be measured
- Absolute energies cannot be measured
- How to train the model?
  - Easy to get from DFT simulations
  - Experimentally available
- Atomisation/cohesive energies
- Given a system $M = A_{n_A} B_{n_B} \ldots$,

$$E_{at} = \frac{1}{N} \left(\sum_I n_I E_I - E_M \right); \quad I = A, B, \ldots$$

Atomisation energies

- Decomposing energies into components...

$$E_{at} = \frac{1}{N}\left(\sum_I n_I(E_I^b + E_I^x + E_I^c) - (E_M^b + E_M^x + E_M^c)\right)$$
$$= E_{at}^b + E_{at}^x + E_{at}^c,$$

Atomisation energies

- Decomposing energies into components...
- Using our exchange energy model...

$$E_{at}^{x} = \boldsymbol{\xi}^{T} \frac{1}{N} \left[\sum_{I} n_I \mathbf{E}^{x}[n_i; \hat{\mathbf{e}}] - \mathbf{E}^{x}[n_M; \hat{\mathbf{e}}] \right]$$

Atomisation energies

- Decomposing energies into components...
- Using our exchange energy model...
- The design matrix becomes...

$$\mathbf{\Phi} = \begin{pmatrix} \frac{1}{N}\left(\sum_{I \in s_1} n_I \mathbf{E}^x[n_i; \hat{\mathbf{e}}] - \mathbf{E}^x[n_{s_1}; \hat{\mathbf{e}}]\right)^T \\ \frac{1}{N}\left(\sum_{I \in s_2} n_I \mathbf{E}^x[n_i; \hat{\mathbf{e}}] - \mathbf{E}^x[n_{s_2}; \hat{\mathbf{e}}]\right)^T \\ \vdots \\ \frac{1}{N}\left(\sum_{I \in s_N} n_I \mathbf{E}^x[n_i; \hat{\mathbf{e}}] - \mathbf{E}^x[n_{s_N}; \hat{\mathbf{e}}]\right)^T \end{pmatrix}$$

Atomisation energies

- What is my data vector $\mathbf{t}$?
- No access to exchange energy directly

Atomisation energies

- What is my data vector $\mathbf{t}$?
- No access to exchange energy directly
- But E_{at}^{b} and E_{at}^{c} do not depend on our model...

$$\mathbf{t} = \begin{pmatrix} E_{at}^{exp}(s_1) - E_{at}^{b}[n_1] - E_{at}^{c}[n_1] \\ E_{at}^{exp}(s_2) - E_{at}^{b}[n_2] - E_{at}^{c}[n_2] \\ \vdots \\ E_{at}^{exp}(s_N) - E_{at}^{b}[n_N] - E_{at}^{c}[n_N] \end{pmatrix}$$

Atomisation energies

- What is my data vector $\mathbf{t}$?
- No access to exchange energy directly
- But E_{at}^{b} and E_{at}^{c} do not depend on our model...

$$\mathbf{t} = \begin{pmatrix} E_{at}^{exp}(s_1) - E_{at}^{b}[n_1] - E_{at}^{c}[n_1] \\ E_{at}^{exp}(s_2) - E_{at}^{b}[n_2] - E_{at}^{c}[n_2] \\ \vdots \\ E_{at}^{exp}(s_N) - E_{at}^{b}[n_N] - E_{at}^{c}[n_N] \end{pmatrix}$$

- We have all we need for the regression

Indirect measurements

- What if we want to add other data?

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- What if we want to add other data?
- Linear on the energy?

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- What if we want to add other data?
- Linear on the energy?
  - We can use it directly within our model

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- What if we want to add other data?
- Linear on the energy?
- Not linear?

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- What if we want to add other data?
- Linear on the energy?
- Not linear?
  - Transform into impact on energy if possible
  - Non-analytically tractable posterior otherwise

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- What if we want to add other data?
- Linear on the energy?
- Not linear?
- Example: equilibrium volume (V_0), bulk modulus and pressure derivative (B_0, B_1) $\rightarrow$ $E(V)$ through equation of state

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- Example: equilibrium volume (V_0), bulk modulus and pressure derivative (B_0, B_1) $\rightarrow E(V)$ through equation of state

$$E(V) = a + b\frac{V_0^{1/3}}{V^{1/3}} + c\frac{V_0^{2/3}}{V^{2/3}} + d\frac{V_0}{V} = \gamma^T \phi(V)$$

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- Example: equilibrium volume (V_0), bulk modulus and pressure derivative (B_0, B_1) $\rightarrow E(V)$ through equation of state

$$E(V) = a + b\frac{V_0^{1/3}}{V^{1/3}} + c\frac{V_0^{2/3}}{V^{2/3}} + d\frac{V_0}{V} = \gamma^T \phi(V)$$

- where

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 0 \\ 18 & 10 & 4 & 0 \\ 108 & 50 & 16 & 0 \end{pmatrix} \gamma = \begin{pmatrix} -E_0 \\ 0 \\ 9V_0B_0 \\ 27V_0B_0B_1 \end{pmatrix}$$

[1] A. B. Alchagirov *et al.* (2001)

Indirect measurements

- Example: equilibrium volume (V_0), bulk modulus and pressure derivative (B_0, B_1) $\rightarrow E(V)$ through equation of state

$$E(V) = a + b\frac{V_0^{1/3}}{V^{1/3}} + c\frac{V_0^{2/3}}{V^{2/3}} + d\frac{V_0}{V} = \gamma^T \phi(V)$$

- From experimental V_0, B_0, B_1 we can also obtain cohesive energies of the strained material

[1] A. B. Alchagirov *et al.* (2001)

Data sets

- Elementary solids
- Molecules

Data sets

- Elementary solids
  - 20 cubic solids: 13 training + 7 testing
  - Extended using bulk properties (4 strains each)
- Molecules

Data sets

- Elementary solids
  - 20 cubic solids: 13 training + 7 testing
  - Extended using bulk properties (4 strains each)
- Molecules
  - G2/97 data set (small molecules): 120 training + 28 testing
  - Larger molecules from G3/99 only for testing

Model

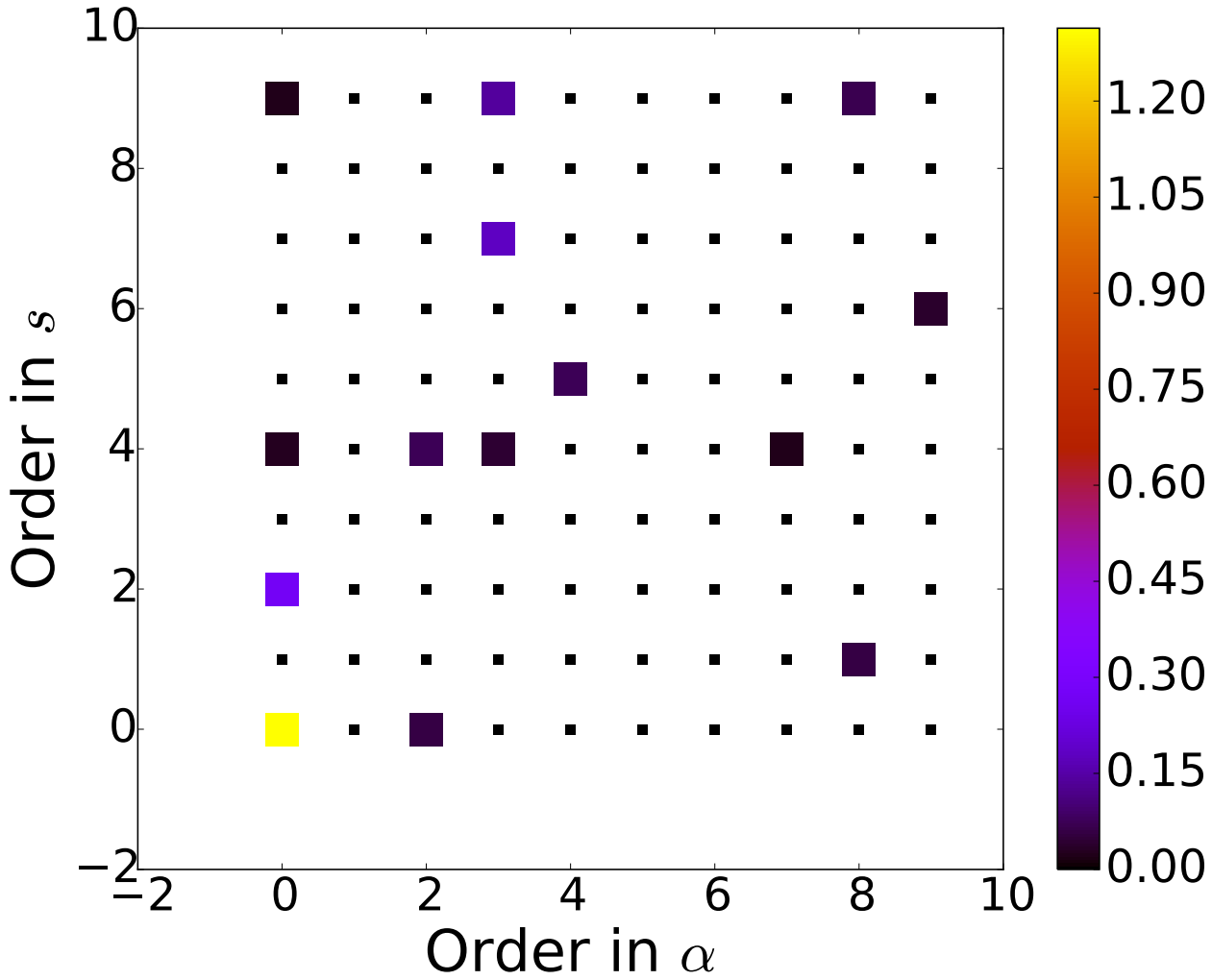
- Linear model with 10×10 terms

$$F^x(s,\alpha) = \sum_{i=0}^{9}\sum_{j=0}^{9} \xi_{ij}^x P_i(t_s(s)) P_j(t_\alpha(\alpha))$$

Model

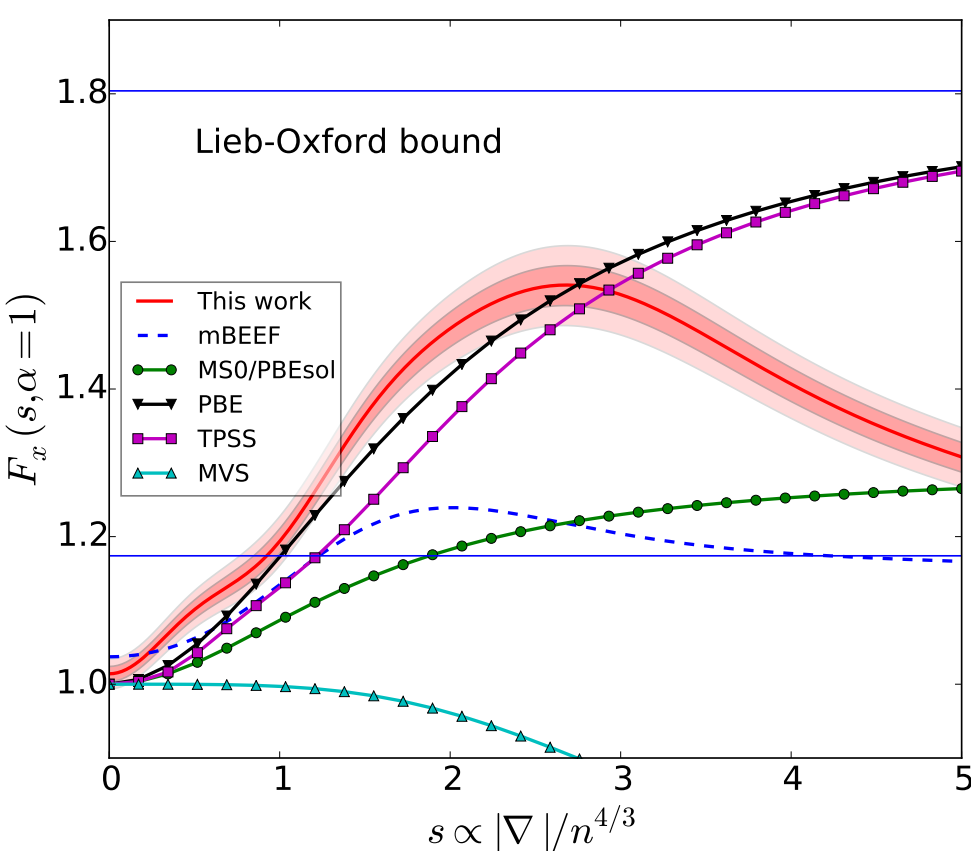
- Linear model with 10×10 terms
- DFT simulations run
  - Using PBE functional
  - Cut-off energy of 800 eV
  - Monkhorst-Pack mesh ($16 \times 16 \times 16$)
  - Relaxing with a maximum force criterion of 0.05 eV/Å

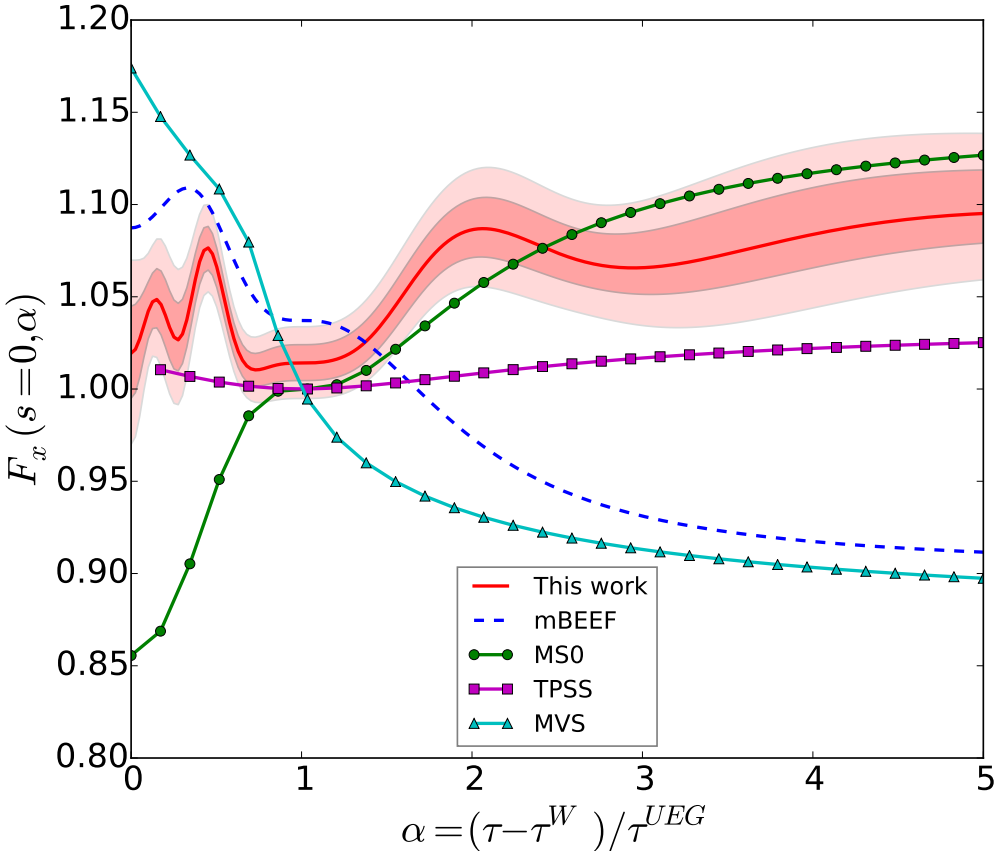
Results


10
8
6
4
2
0
−2
Order in s
−2
0
2
4
6
8
10
Order in α
1.20
1.05
0.90
0.75
0.60
0.45
0.30
0.15
0.00

Results

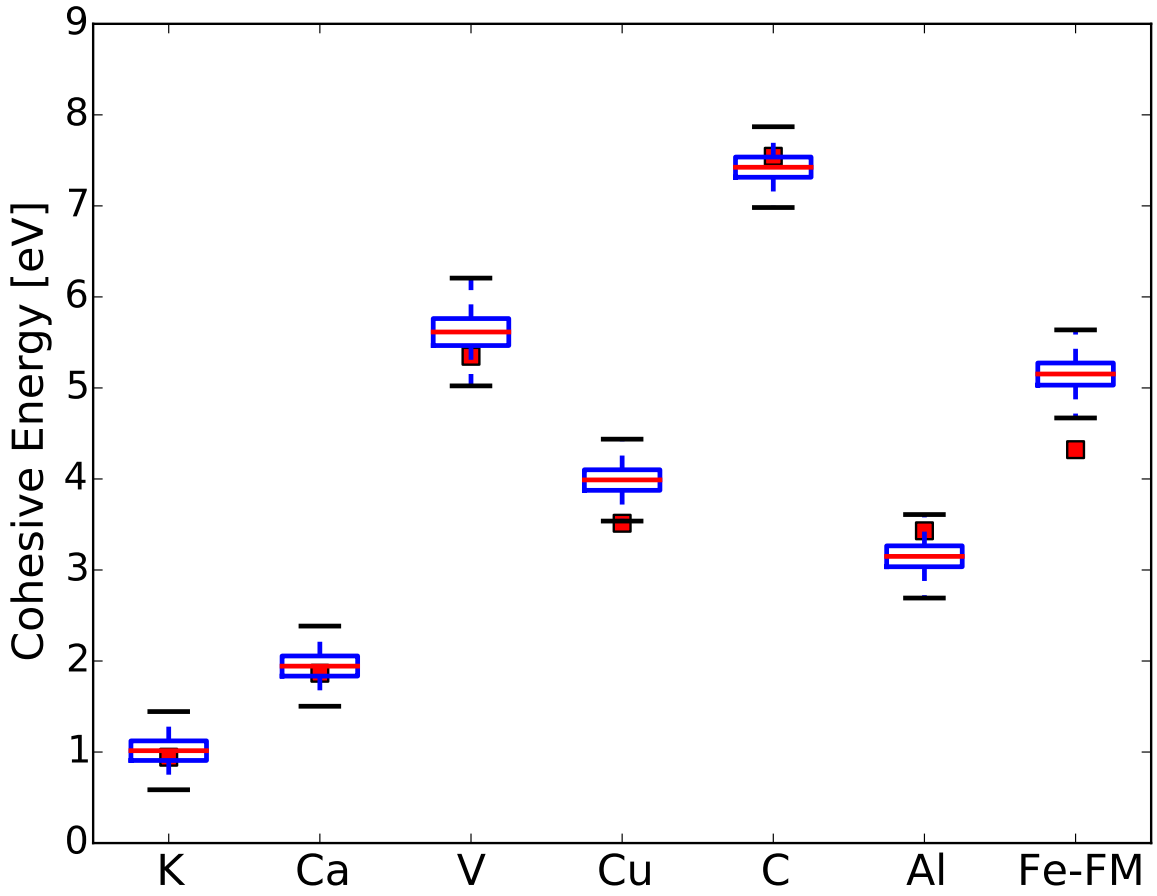
- Enhancement factor with 1 and 2 σ intervals





Atomisation energies

- Prediction of atomisation energies in the 7 test solids



Atomisation energies

- Prediction of atomisation energies in the 7 test solids
- Error in predictions for solids and molecules

XC functional	Error	G2/97-test	G2/97	EL20-test	EL20
This work	MAE	0.116	0.103	0.243	0.0975
	MARE	3.27	1.46	8.56	5.62
PBE	MAE		0.703		0.238
	MARE		5.09		6.88

Table: Mean absolute error (in eV) and mean absolute relative error (in %) of the predictions of atomisation energies using the average model for the training sets containing molecules (G2/97) and solids (EL20).

Atomisation energies

- Prediction of atomisation energies in the 7 test solids
- Error in predictions for solids and molecules

XC functional	Error	G2/97-test	G2/97	EL20-test	EL20
This work	MAE	0.116	0.103	0.243	0.0975
	MARE	3.27	1.46	8.56	5.62
PBE	MAE		0.703		0.238
	MARE		5.09		6.88

- G2/97 MAE better than TPSS (0.28 eV), BEEF-vdW (0.16 eV), B3LYP (0.14 eV) or PBE0 (0.21 eV)

Correlation functional

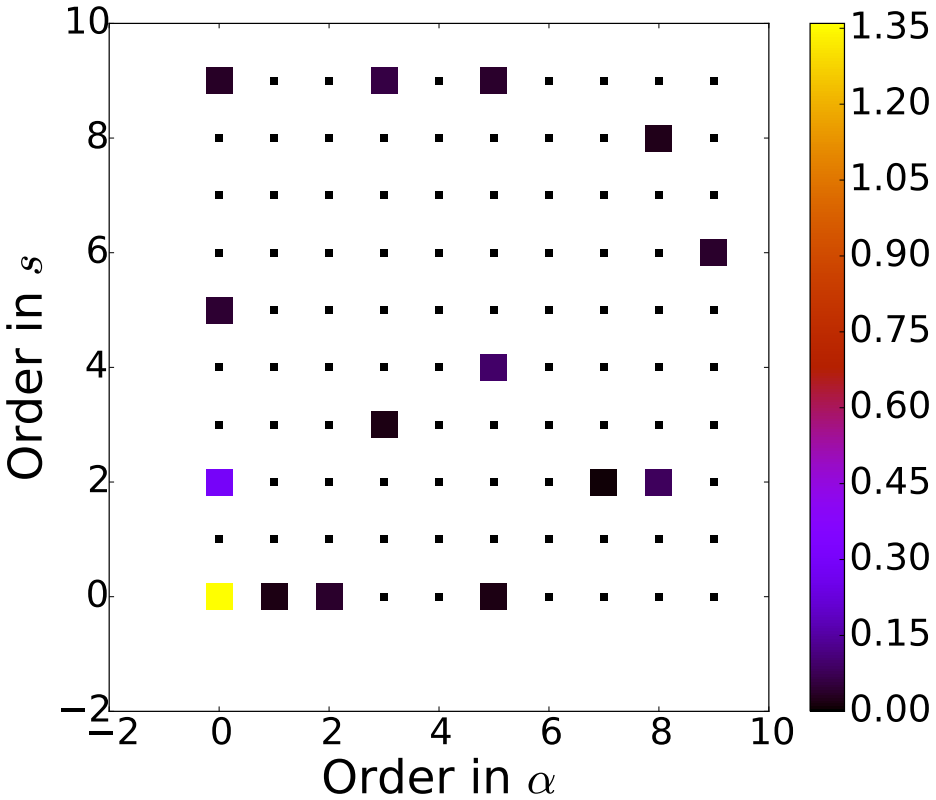
- We assumed a fixed correlation energy functional

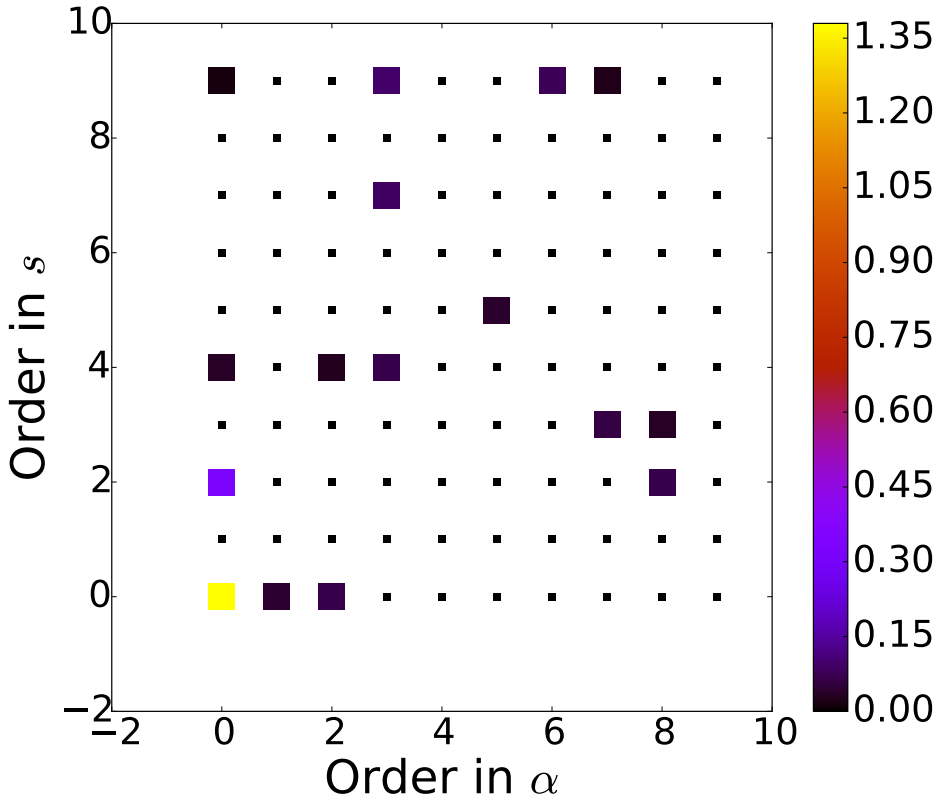
Correlation functional

- We assumed a fixed correlation energy functional
- What's the impact of our choice?

Correlation functional

- We assumed a fixed correlation energy functional
- What's the impact of our choice?
- Coefficients with PBEsol (left) and vPBE (right) correlations:





Correlation functional

- We assumed a fixed correlation energy functional
- What's the impact of our choice?
- Errors:

C functional	Error	G2/97-test	G2/97	EL20-test	EL20
PBE	MAE	0.116	0.103	0.243	0.0975
	MARE	3.27	1.46	8.56	5.62
PBEsol	MAE	0.116	0.108	0.204	0.172
	MARE	2.91	1.55	6.12	4.98
vPBE	MAE	0.110	0.107	0.226	0.184
	MARE	2.72	1.41	6.45	5.17
TPSS	MAE	0.108	0.104	0.227	0.190
	MARE	2.68	1.42	6.85	5.53

Test set

- SL20 test set including
  - 13 elemental solids
  - I-VII, II-VI, III-V and IV-IV compounds
  - 7 of them in the training set (elemental solids)

Uncertainty propagation from DFT

- Bulk properties are not obtained directly from DFT simulations
- How to propagate the uncertainty?
- We use a nested Monte Carlo approach
  - Sample model coefficients from the posterior distribution
  - Fit the EOS to the values from this X energy (Bayesian fit)
  - Sample coefficients from the fitting to calculate V_0, B_0, B_1

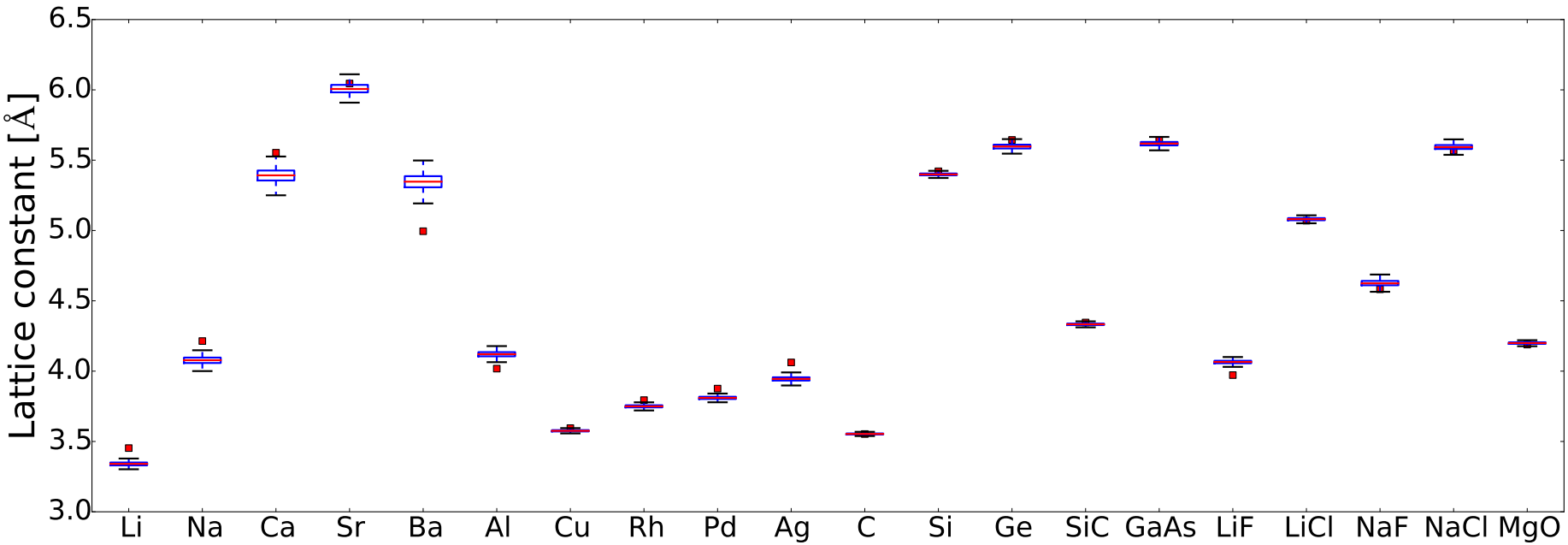
Uncertainty propagation from DFT

Algorithm 2 Calculation of uncertainty for V_0 and B_0.

Input: system s with unit cell $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$.
Input: N_1^{max}, N_2^{max}, the maximum iterations.
for 5 strains $0.95 \leq \sigma_i \leq 1.05$ **do**
    Strain unit cell by σ_i: $\mathbf{x}_\alpha \rightarrow \sigma_i \mathbf{x}_\alpha$, $\alpha = 1, 2, 3$.
    Self-consistent simulation of strained system.
    Keep the self-consistent electron density $n_i^* = n(\sigma_i)$.
end for
$N_1 = 0$
repeat
    Sample $\boldsymbol{\xi}_{N_1}$, β_{N_1}.
    Non self-consistent simulation with $\boldsymbol{\xi}_{N_1}$, β_{N_1} and n_i^*.
    $N_2 = 0$
    repeat
        Sample $\boldsymbol{\gamma}_{N_2}$.
        Calculate V_0, B_0.
    until $N_2 = N_2^{max}$
    $N_1 = N_1 + 1$
until $N_1 = N_1^{max}$
Collect statistics on calculated V_0, B_0.

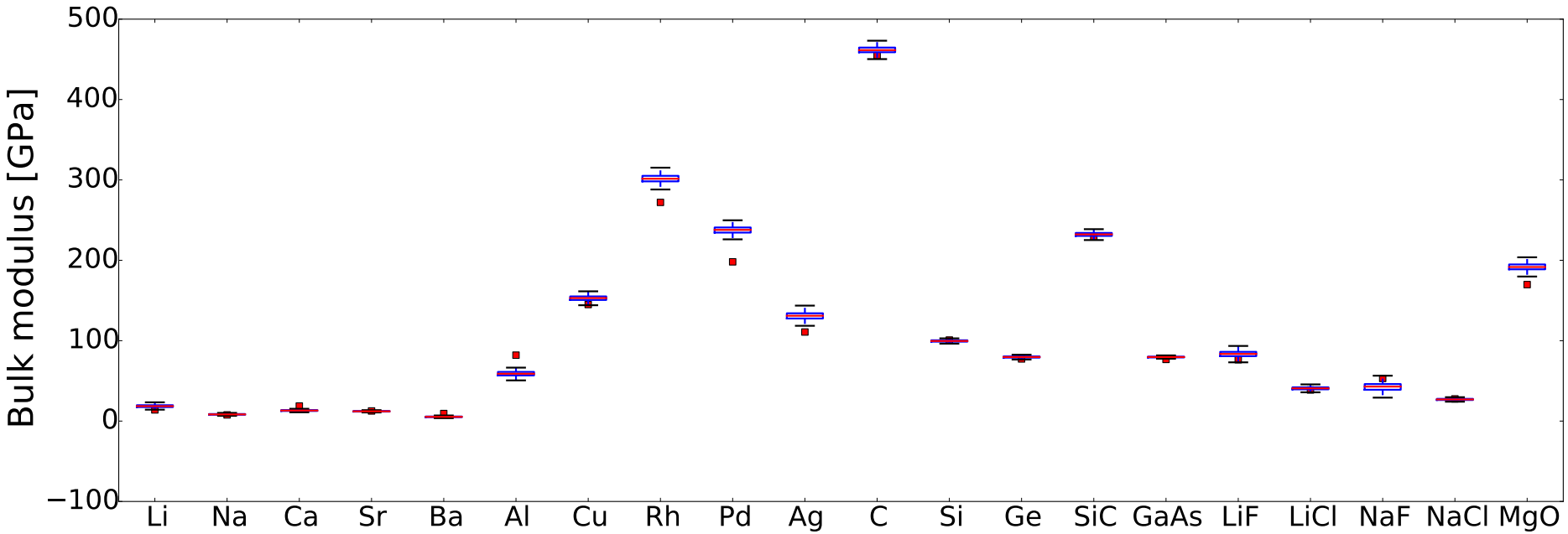
Results

- Equilibrium lattice constants for SL20 materials



Results

- Equilibrium bulk moduli for SL20 materials



Results

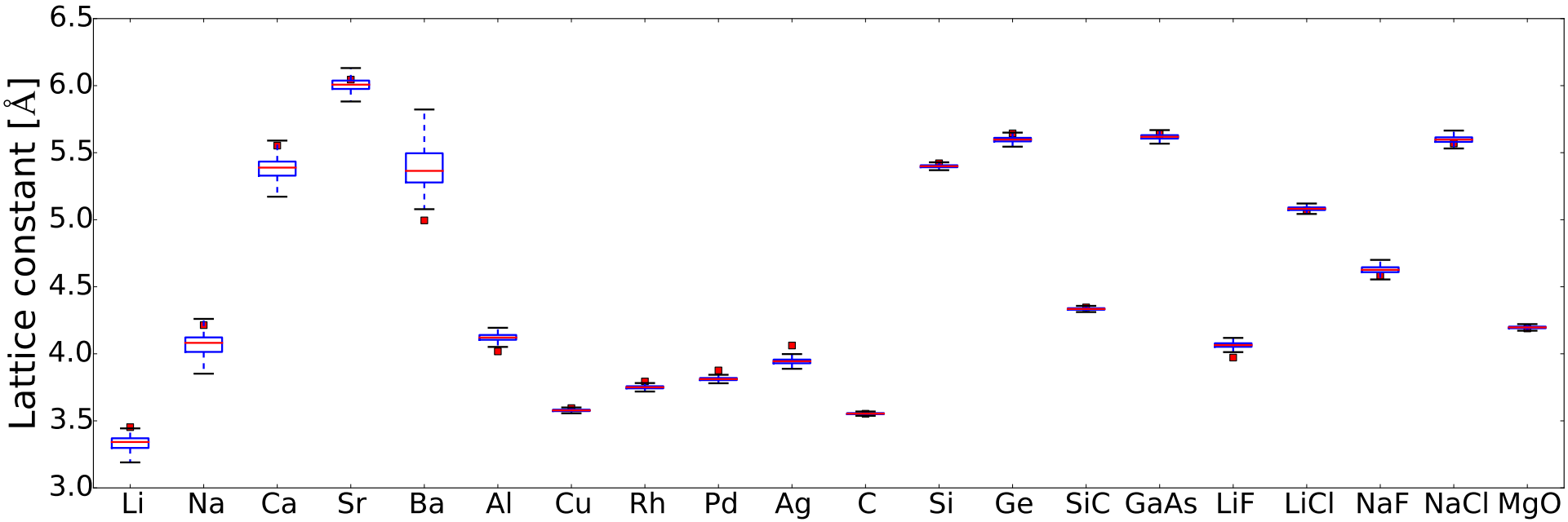
- What if the DFT results have another error sources?

Results

- What if the DFT results have another error sources?
- For example, assume a numerical error with Gaussian distribution and standard deviation 10 mV

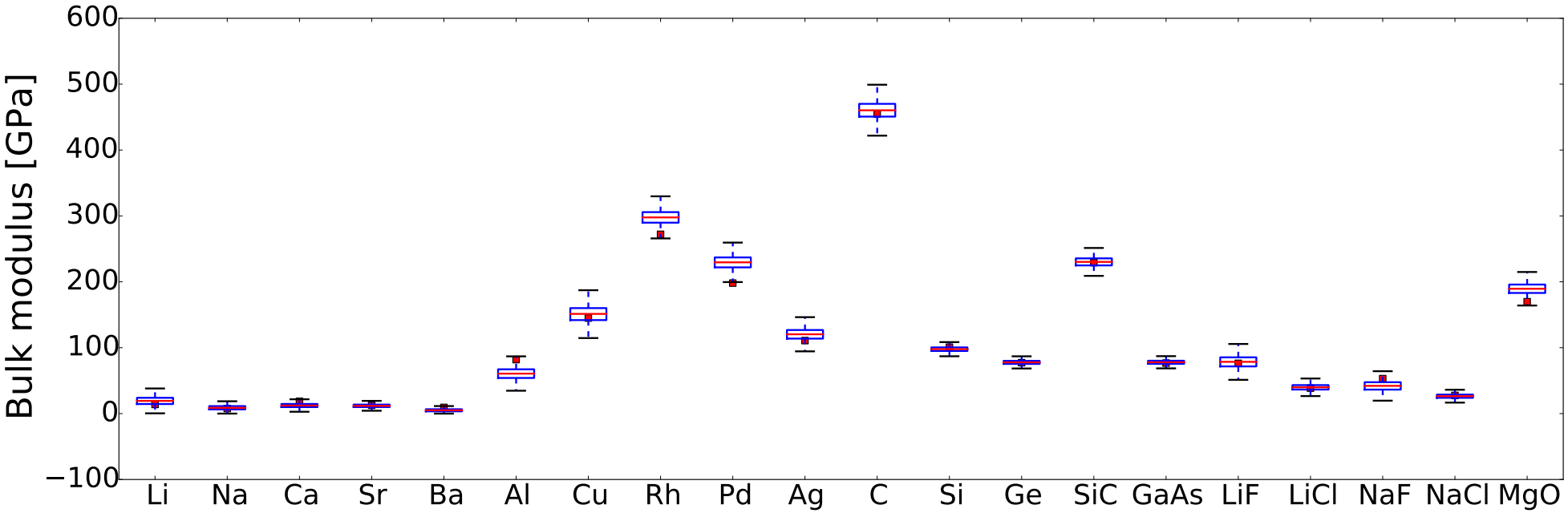
Results

- Equilibrium lattice constants for SL20 materials



Results

- Equilibrium bulk moduli for SL20 materials



Summary

- Bayesian framework to obtain an exchange energy functional
  - Use of a linear model
  - Coefficients of the model are random variables
  - Basis functions are fixed

Summary

- Bayesian framework to obtain an exchange energy functional
- Use of a relevance vector machine to find hyperparameters
  - Automatic selection of relevant basis functions (model selection)

Summary

- Bayesian framework to obtain an exchange energy functional
- Use of a relevance vector machine to find hyperparameters
- Obtained exchange energy from a simulation has an uncertainty
  - Limited data in the training (can be reduced to zero asymptotically with more data)
  - Limited model space, meta-GGA (cannot be reduced to zero asymptotically with more expansion terms)

Summary

- Bayesian framework to obtain an exchange energy functional
- Use of a relevance vector machine to find hyperparameters
- Obtained exchange energy from a simulation has an uncertainty
- This uncertainty can be propagated to other derived quantities
  - Bulk properties (shown)
  - Band diagrams, phonon properties, transport coefficients, enrgy barriers, etc.
  - Further models based on DFT results (e.g., cluster expansion for alloy modelling)

Thank you for your attention!

Acknowledgments

- EPSRC Strategic Package Project EP/L027682/1 for research at WCPM