Hylleraas

Automated Determination of Hybrid Particle-Field Parameters by Machine Learning

May 29, 2020

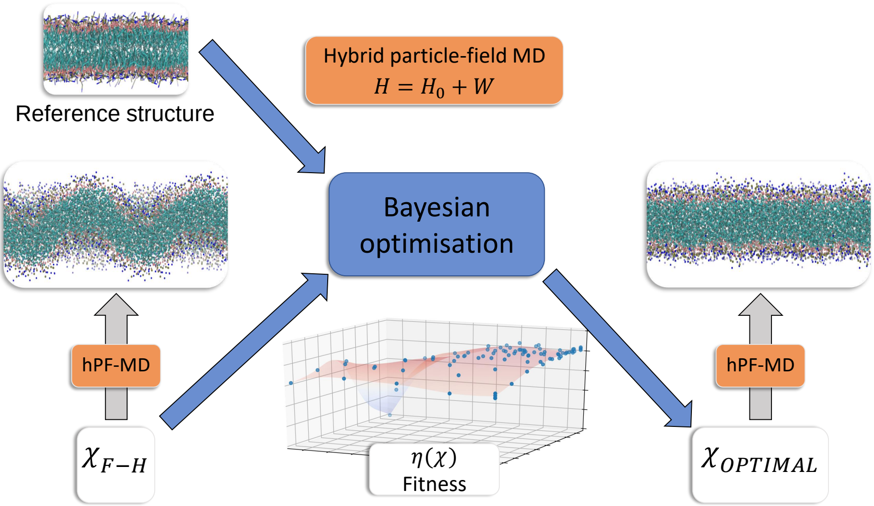
Morten Ledum

Norwegian Centre of Excellence

The Research Council of Norway

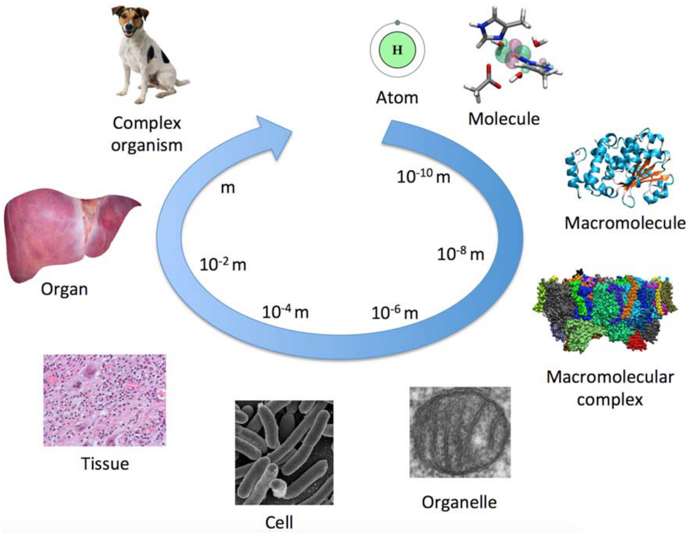
UiO : University of Oslo

Outline



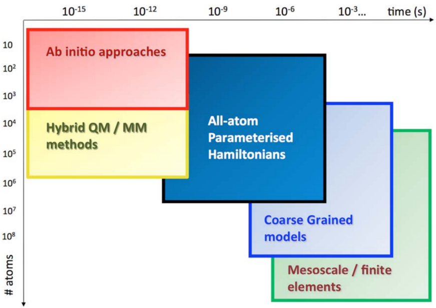
Reference structure
Hybrid particle-field MD
$H = H_0 + W$
Bayesian optimisation
hPF-MD
χ_{F-H}
$\eta(\chi)$
Fitness
hPF-MD
$\chi_{OPTIMAL}$

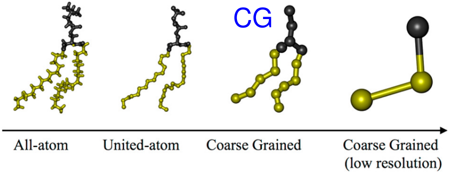
Outline



M. Cascella and S. Vanni, Chem. Modell. ***12****, 1–52 (2016)*

Coarse-graining





$$Z = \int d\Gamma \, e^{-\beta H(\Gamma)} \longrightarrow Z \approx \int d\Gamma_{CG} \, e^{-\beta H(\Gamma_{CG})}$$

M. Cascella and S. Vanni, Chem. Modell. ***12****, 1–52 (2016)*
T.A. Soares et al., J. Phys. Chem. Lett. ***8****, 3586–3594 (2017)*

Hylleraas

Hybrid particle-field

Hamiltonian

$$H(\{\mathbf{r}\}) = \sum_{m=1}^{N_{mol}} \underbrace{H_0(\{\mathbf{r}_m\})}_{\text{Intramolecular}} + \underbrace{W[\{\phi(\mathbf{r})\}]}_{\text{Intermolecular}}$$

External potential and forces

$$V_k(\mathbf{r}) = \frac{\delta W[\{\phi\}]}{\delta \phi_k(\mathbf{r})}; \quad and \quad \mathbf{F}_i = -\nabla_i V_k(\mathbf{r}_i)$$

Interaction energy $W[\{\phi(\mathbf{r})\}]$

$$W[\phi(\mathbf{r})] = \frac{1}{2\phi_0} \int d\mathbf{r} \left(\underbrace{\sum_{ij} \tilde{\chi}_{ij} \phi_i(\mathbf{r}) \phi_j(\mathbf{r})}_{\textit{Mixing}} + \underbrace{\frac{1}{\kappa} \left(\sum_j \phi_j(\mathbf{r}) - \phi_0 \right)^2}_{\textit{Compressibility}} \right),$$

Depends on a set of parameters

$\tilde{\chi}_{ij} < 0$ Particle types i and j prefer to mix

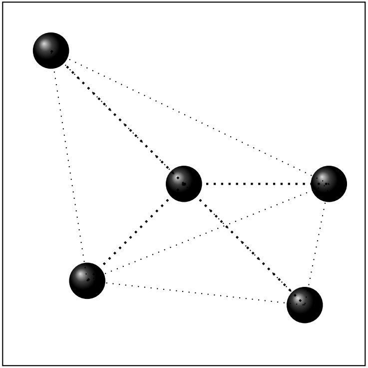
$\tilde{\chi}_{ij} > 0$ Particle types i and j prefer to avoid each other

and

$\kappa = 0$ Incompressible

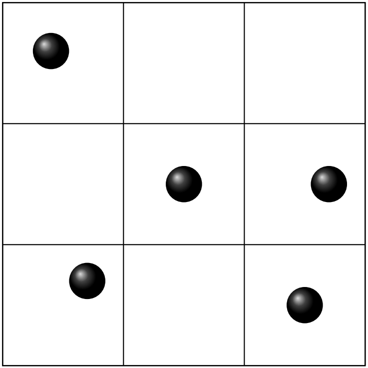
$\kappa > 0$ Compressible

Computing the density field



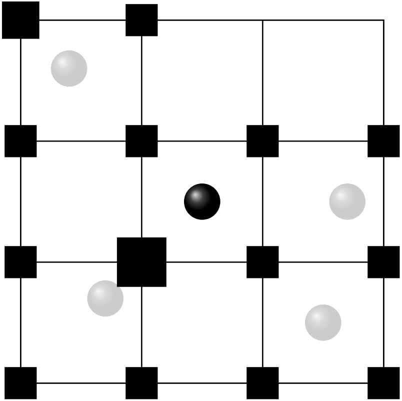
$$\mathrm{P}_{ij}\, \tilde{\nabla}(\mathbf{r}_{ij})$$

Computing the density field



$$\mathrm{P}_{i}\, V_k[\ \ (\mathbf{r}_i)]$$

Computing the density field

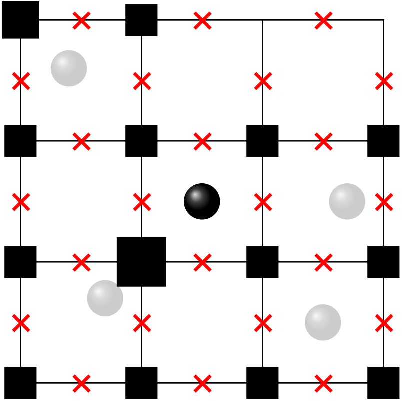


$$\mathrm{P}_{i}\, V_k[\ \ (\mathbf{r}_i)]$$

$$\mathbf{F}_i = \ \ r_{\ i} V_k[\ \ (\mathbf{r}_i)]$$

■ : *nml*

Computing the density field



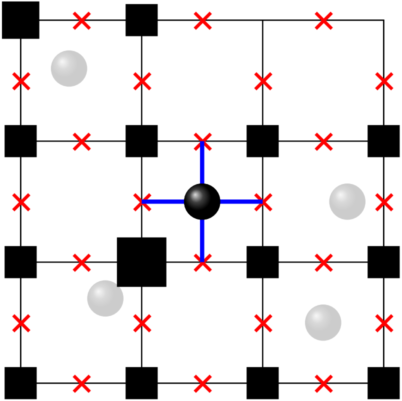
$$\mathrm{P}_{i}\, V_k[\ (\mathbf{r}_i)]$$

$$\mathbf{F}_i = \quad r_{\,i} V_k[\ (\mathbf{r}_i)]$$

■ : $_{nml}$

× : $r\, V_{nml}$

Computing the density field



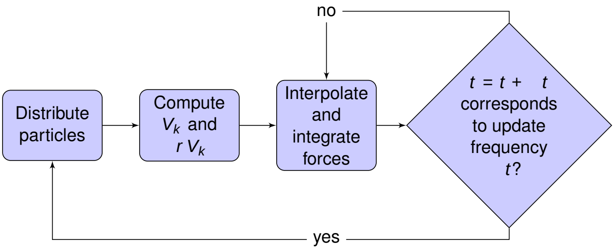
$$\text{P}_{i} V_k[\ (\mathbf{r}_i)]$$

$$\mathbf{F}_i = \quad r_{\ i} V_k[\ (\mathbf{r}_i)]$$

[black square] : $_{nml}$

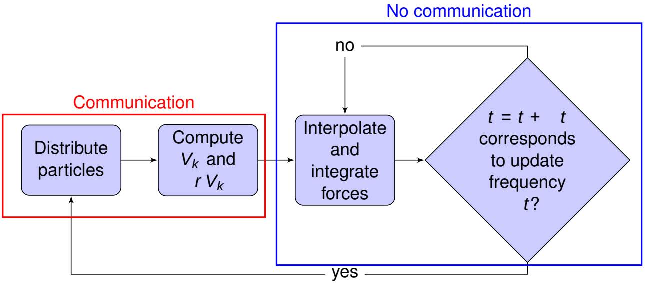
[red cross] : $r\ V_{nml}$

Implementation



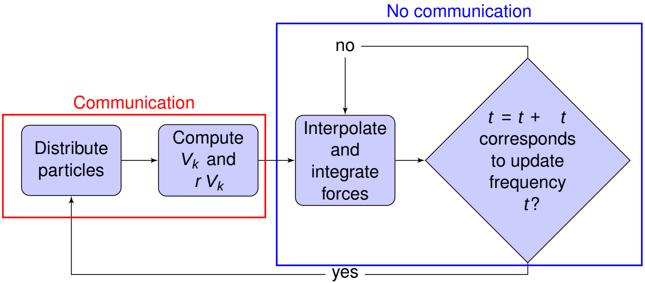
Hylleraas

Implementation

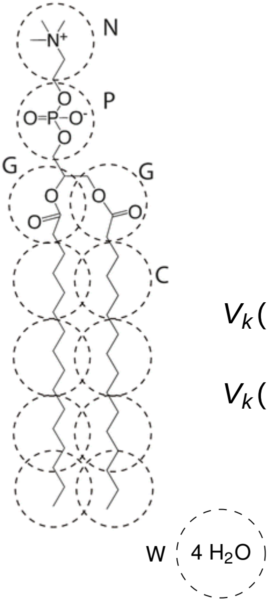


Implementation

OCCAM
Molecular Dynamics



Coarse grained model



Intramolecular Hamiltonian

$$H_0 = \sum \frac{m_i \dot{\mathbf{r}}_i^2}{2} + \sum \frac{k_r (r_{ij} \quad r_0)^2}{2} + \sum \frac{k \;(\cos(\;{}_{ijk}) \quad \cos(\;{}_0))^2}{2}$$

Interaction potential

$$V_k(\mathbf{r}) = \frac{W[f \quad g]}{{}_k(\mathbf{r})}$$

$$V_k(\mathbf{r}) = \frac{1}{{}_0} \sum \tilde{}_{ij} \;{}_j(\mathbf{r}) + \frac{1}{} \sum \;{}_j(\mathbf{r}) \qquad {}_0$$

Flory-Huggins ~ matrix

P	G	C	D	W	
-1.5	6.3	9.0	7.2	-8.1	N
	4.5	13.5	11.7	-3.6	P
		6.3	6.3	4.5	G
			0	13.5	C
				23.25	D

Calculated from

$$^{\text{F–H}}_{KK'} = \frac{z_{\text{CN}}}{2k_{\text{B}}T} \quad \frac{2u_{KK'} \qquad (u_{KK} + u_{K'K'})}{2} \quad ;$$

where $u_{KK'}$ is interpreted as the MARTINI model K–K^{0} '' parameter.

Introduction

Bayesian optimization

- Gaussian process
- Acquisition function

Lipid membranes

Summary

Hylleraas

Bayesian optimization

Constrained optimization scheme

$$\mathbf{x}_{\text{optimal}} = \arg\max_{\mathbf{x} \in X} \;(\mathbf{x})$$

Does not require derivatives, suitable for computationally expensive, noisy black-box functions.

Parameters $\mathbf{x} = (p_1; p_2; \ldots; p_n)$

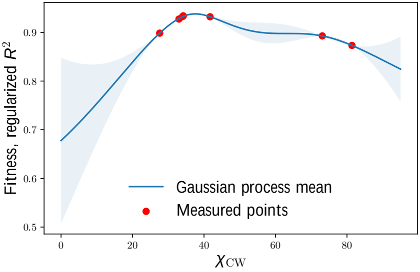
Parameter space X

Objective function $(\mathbf{x})$

Surrogate-based optimization

In general the parameter space X is high-dimensional and the objective function $(\mathbf{x})$ is unknown, non-convex, multimodal, and only accessible through noisy pointwise sampling.

! Place a **Gaussian process** function *prior* over the objective function and optimize it instead.



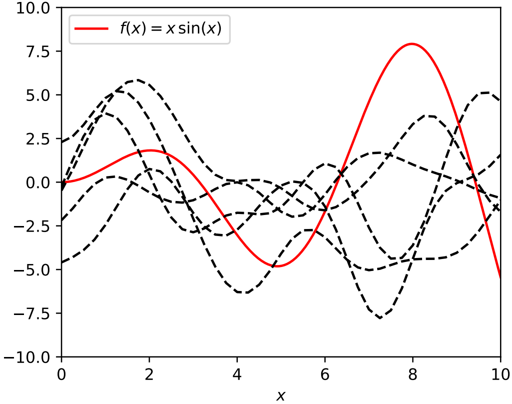
Gaussian process

A GP is a collection of random variables such that any linear combination of the variables induces a multivariate normal distribution.

Given data, a GP is a probability distribution over possible functions which fits the data points.

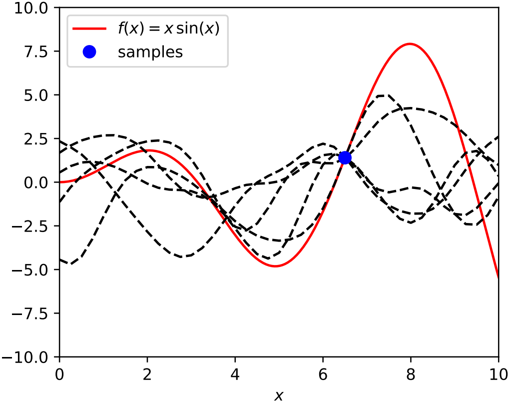
$$f(x) \quad GP(\;{}_0; \quad {}_0)$$

Conditioning

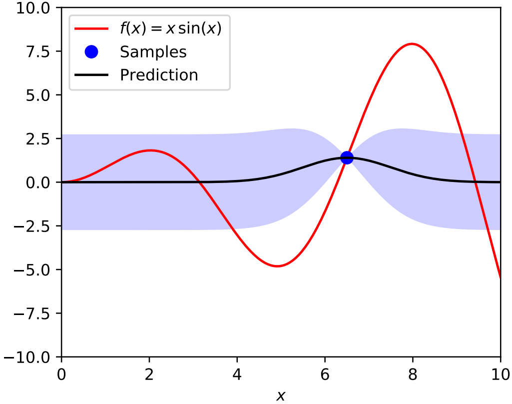


Hylleraas

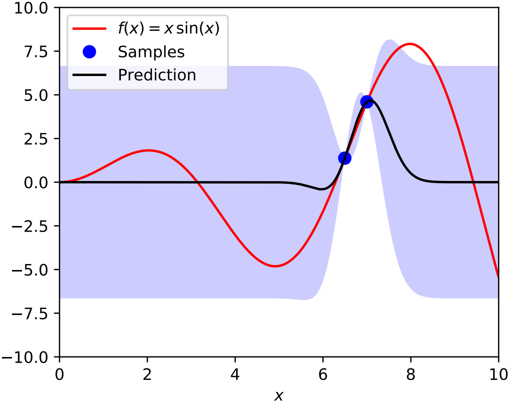
Conditioning



Conditioning

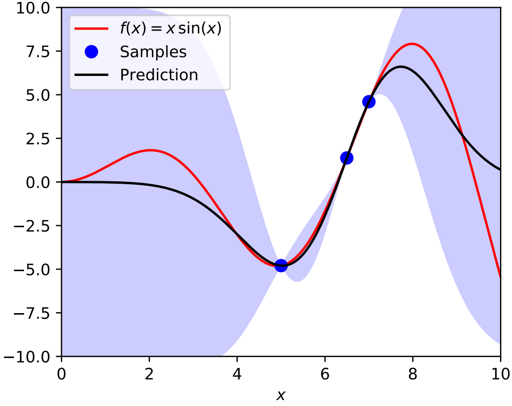


Conditioning

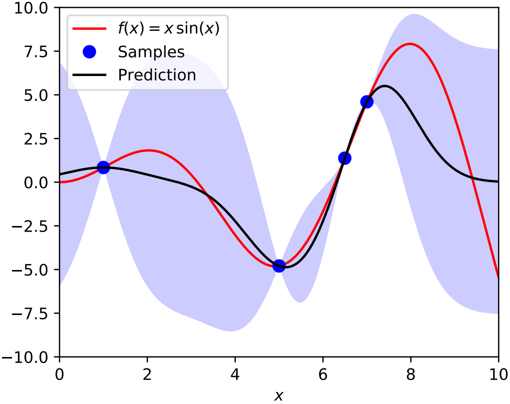


Hylleraas

Conditioning

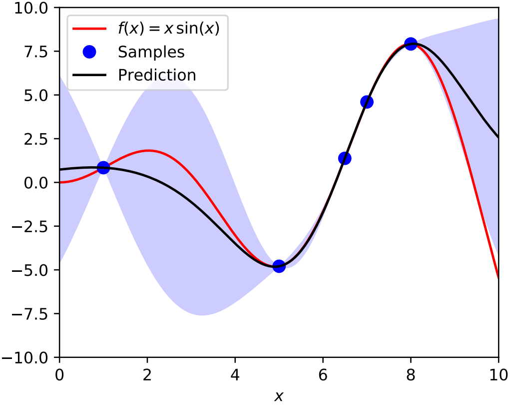


Conditioning

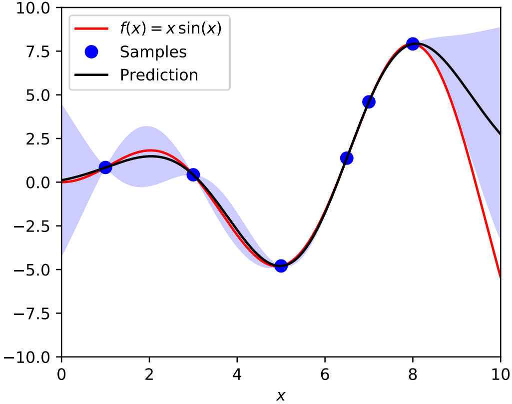


Hylleraas

Conditioning



Conditioning



Hylleraas

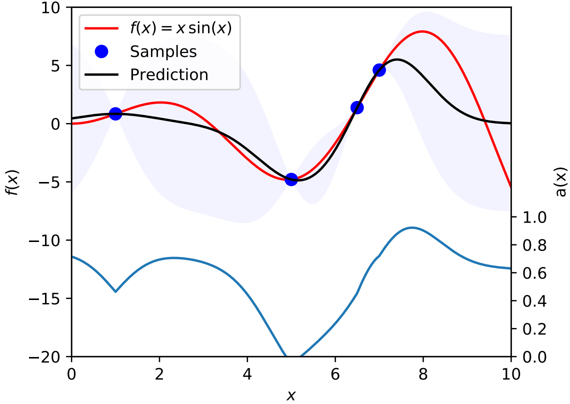
Acquisition function

In order to turn the GP into an optimization scheme, we pair it with an **acquisition function** to guide the sampling

$$
\begin{aligned}
a(\mathbf{x}) &= a(\quad(\mathbf{x});\quad(\mathbf{x})) \\
&= \quad(\mathbf{x}) + \qquad(\mathbf{x})
\end{aligned}
$$

Acquisition function

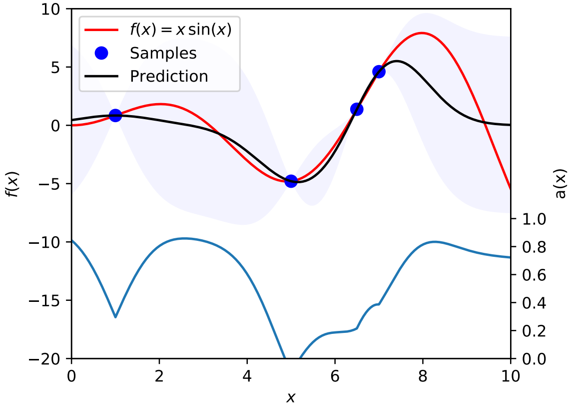
= 1:0



Hylleraas

Acquisition function

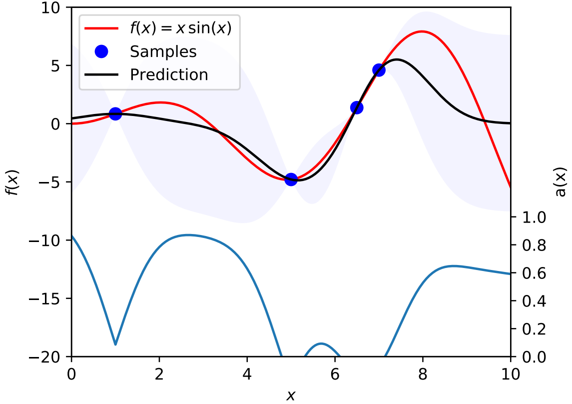
= 3:0



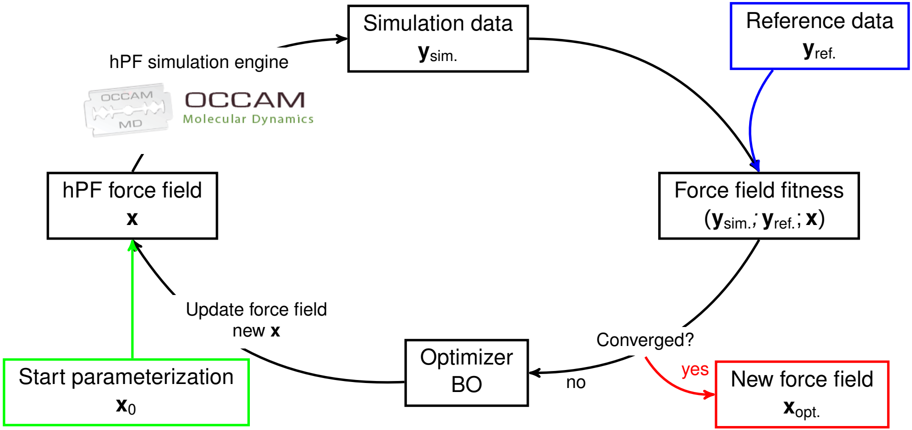
Hylleraas

Acquisition function

= 5:0

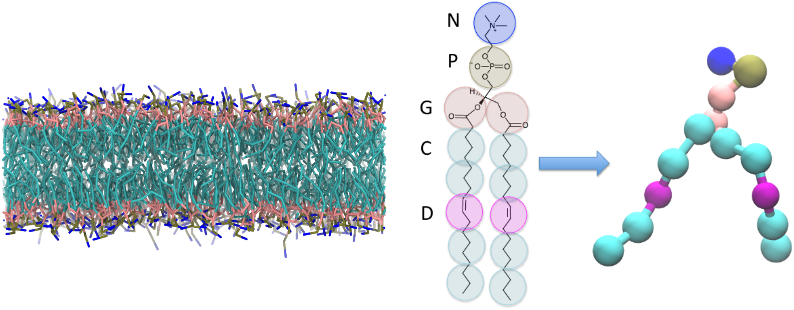


Bayesian optimization

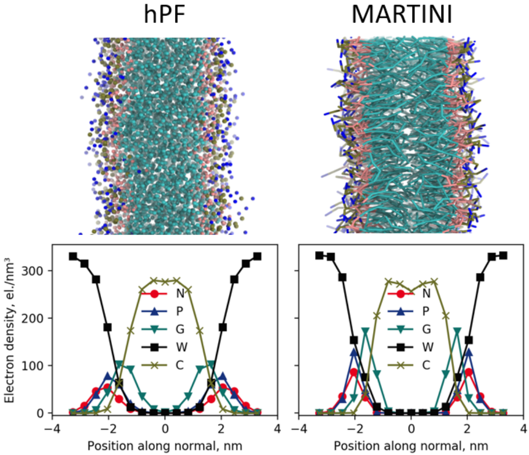


Lipid membrane test systems

Dipalmitoylphosphatidylcholine (DPPC)
Dimyristoylphosphatidylcholine (DMPC)
Distearoylphosphatidylcholine (DSPC)
Dioleoylphosphatidylcholine (DOPC)



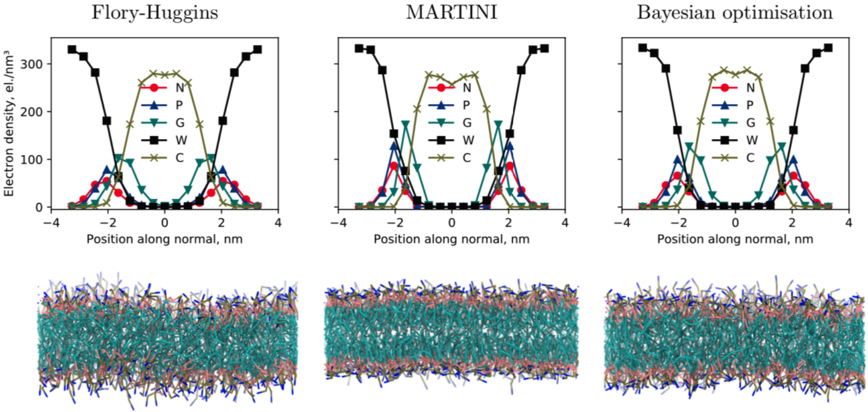
Fitness function



$$\left(\varphi;\tilde{\varphi}\right) = \frac{1}{nn_k}\sum_{k=1}^{n_k}\sum_{i=1}^{n}\left(\varphi_i^k - \hat{\varphi}_i^k\right)^2,$$

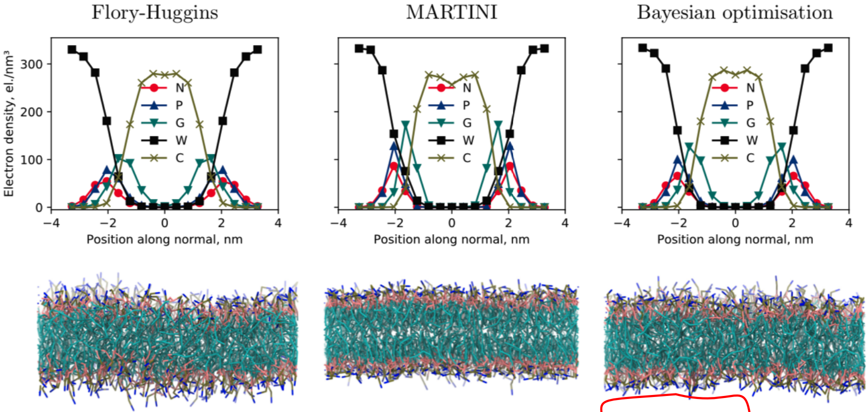
Hylleraas

Optimized DPPC parameters



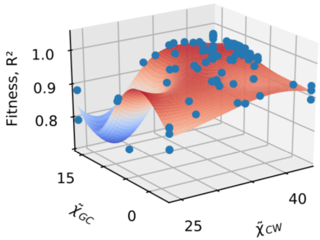
	N	P	G	C	W	average
BO (this work)	4.65	4.10	6.52	7.26	8.91	6.29 (1.71%)
F-H [21]	9.29	12.19	20.51	12.23	12.82	13.41 (4.10%)

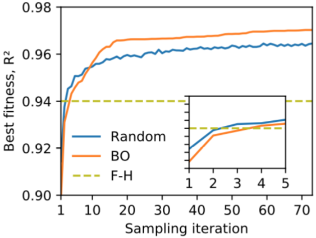
Optimized DPPC parameters

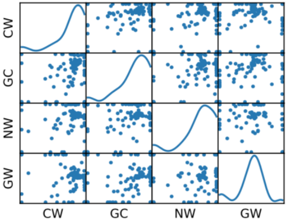


	N	P	G	C	W	average
BO (this work)	4.65	4.10	6.52	7.26	8.91	6.29 (1.71%)
F-H [21]	9.29	12.19	20.51	12.23	12.82	13.41 (4.10%)

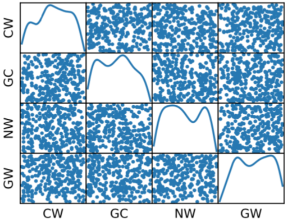
Sampling efficiency







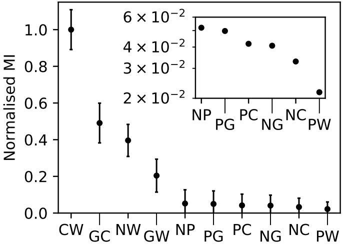
Bayesian optimization sampling

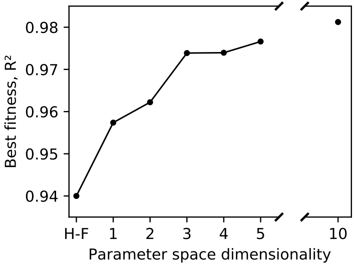


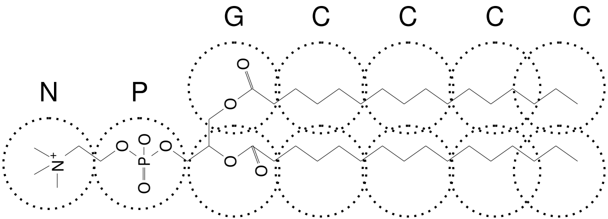
Random sampling

Feature importance

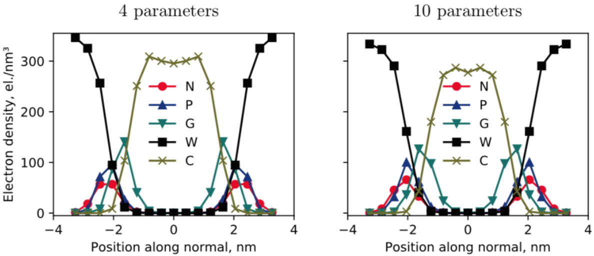
$$MI(X;Y) = \int_X \int_Y \mathrm{d}x\,\mathrm{d}y\, f_{X;Y}(x;y) \log \frac{f_{X;Y}(x;y)}{f_X(x) f_Y(x)};$$







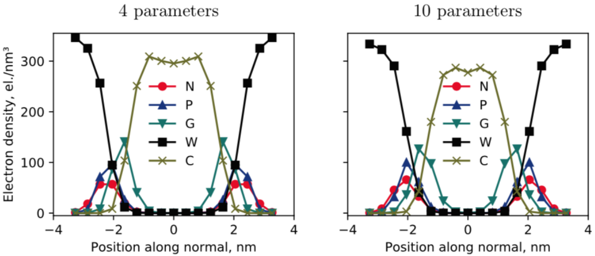
Feature importance



	parameter space dimensionality				
	10	4	3	2	1
C–W	42.24	43.68	43.63	42.09	38.16
G–C	10.47	14.00	15.33	14.69	6.30
N–W	-3.77	1.55	1.82	-8.10	-8.10
G–W	4.53	3.02	4.50	4.50	4.50
N–P	-9.34	-1.50	-1.50	-1.50	-1.50
P–G	8.04	4.50	4.50	4.50	4.50
N–G	1.97	6.30	6.30	6.30	6.30
P–C	14.72	13.50	13.50	13.50	13.50
P–W	-1.51	-3.60	-3.60	-3.60	-3.60
N–C	13.56	9.00	9.00	9.00	9.00
S_{p}	1.71%	1.96%	2.25%	2.29%	2.32%

Hylleraas

Feature importance



	parameter space dimensionality				
	10	4	3	2	1
C–W	42.24	43.68	43.63	42.09	38.16
G–C	10.47	14.00	15.33	14.69	6.30
N–W	-3.77	1.55	1.82	-8.10	-8.10
G–W	4.53	3.02	4.50	4.50	4.50
N–P	-9.34	-1.50	-1.50	-1.50	-1.50
P–G	8.04	4.50	4.50	4.50	4.50
N–G	1.97	6.30	6.30	6.30	6.30
P–C	14.72	13.50	13.50	13.50	13.50
P–W	-1.51	-3.60	-3.60	-3.60	-3.60
N–C	13.56	9.00	9.00	9.00	9.00
S_p	1.71%	1.96%	2.25%	2.29%	2.32%

Hylleraas

Transferability

Hylleraas

Transferability

Transferability

Transferability

Hylleraas

Summary

Our machine learning scheme systematically improves on interaction parameters used in the hPF literature

Makes possible systematic development of accurate and reproducible parameter sets without the need for human fine-tuning

Less important parameters may be identified on the fly and dropped from the optimization, thus drastically lowering the computational cost (with little impact on the resulting parameter set)

The optimized potentials show excellent transferability among chemically similar moieties

Acknowledgements

Sigbjørn Løland Bore[y]
Michele Cascella[y]
Antonio De Nicola[z]

[†] *University of Oslo, Norway*
[†] *Yamagata Univeristy, Japan*

Hylleraas

UiO

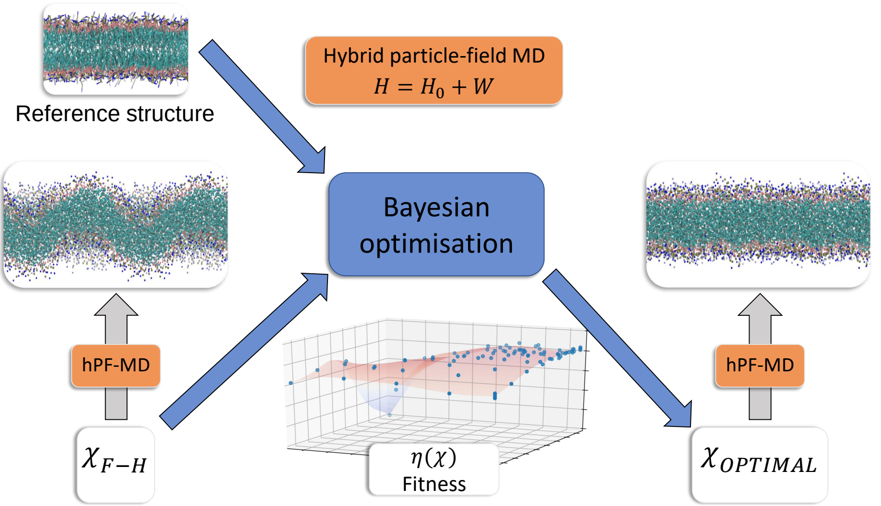
Norwegian Centre of Excellence
The Research Council of Norway

OCCAM Molecular Dynamics

notur

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Outline



Hylleraas