Open Catalyst Challenge 23

THE ITALIAN JOB

The team

ISTITUTO ITALIANO DI TECNOLOGIA

Luigi
Bonati

Simone
Perego

Pedro
Buigues

Pietro
Novelli

Riccardo
Grazzi

Massimiliano
Pontil

Atomistic Simulations

Machine Learning

Italian Institute of Technology, Genoa, Italy

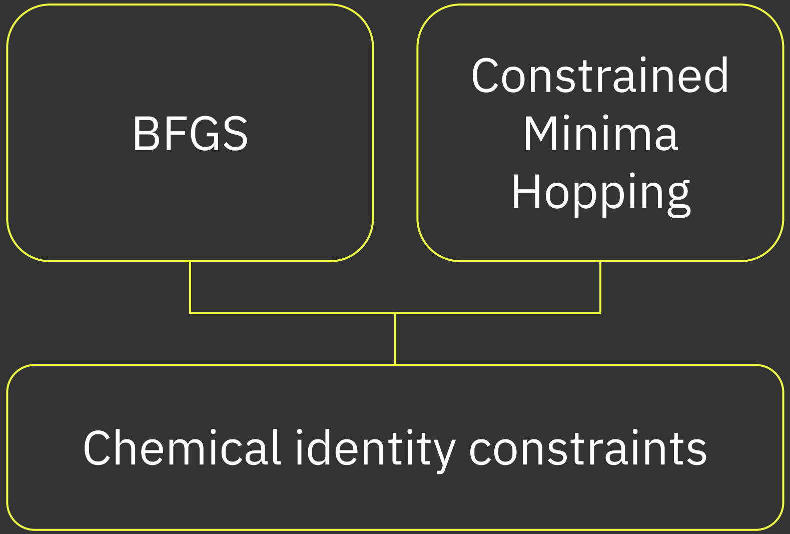
Outline of *the job*

1. Search candidate *configurations*
2. Improve *energy* estimation
3. Take the lowest minima

Outline of *the job*

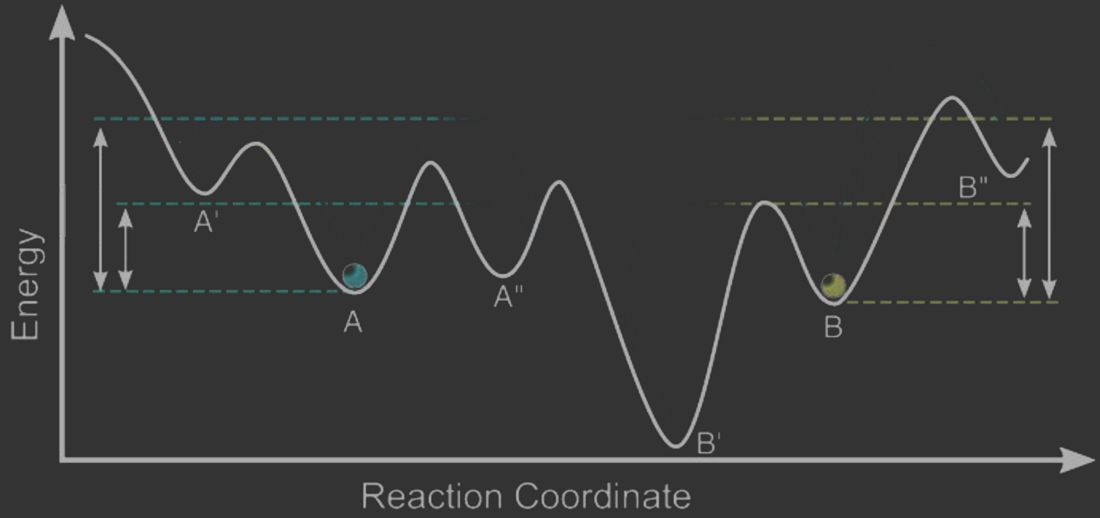
1. **Search candidate *configurations***
2. Improve *energy* estimation
3. Take the lowest minima

Local and global optimization schemes with different ML potentials [1]



[1] EquiformerV2, GemNet-OC and eSCN (OC20 2M)

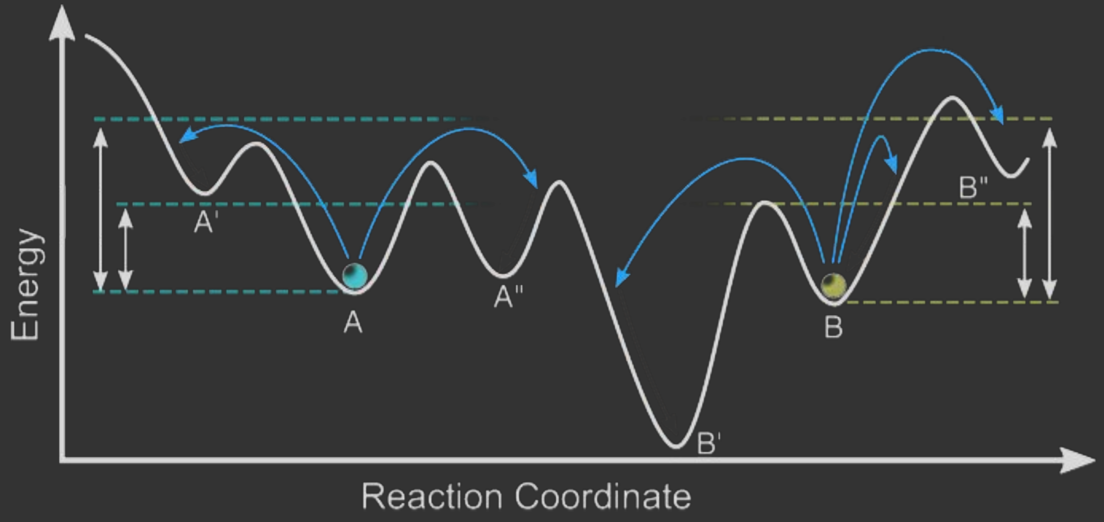
(Constrained) Minima Hopping



- (High temperature) molecular dynamics
- BFGS to minimize

Constraints are needed to avoid changing the molecular identity

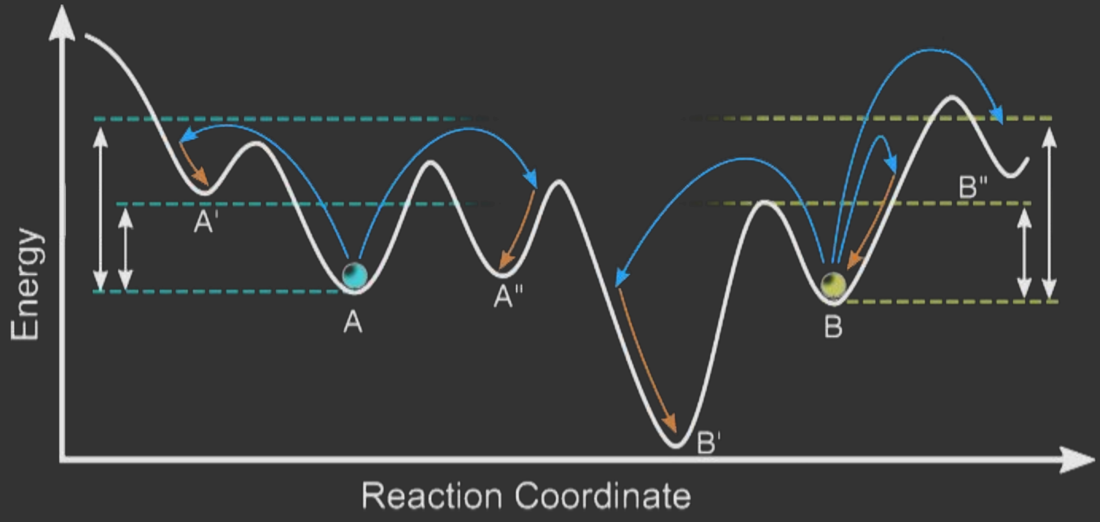
(Constrained) Minima Hopping



- (High temperature) molecular dynamics
- BFGS to minimize

Constraints are needed to avoid changing the molecular identity

(Constrained) Minima Hopping



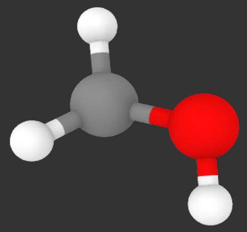
- (High temperature) molecular dynamics
- BFGS to minimize

Constraints are needed to avoid changing the molecular identity

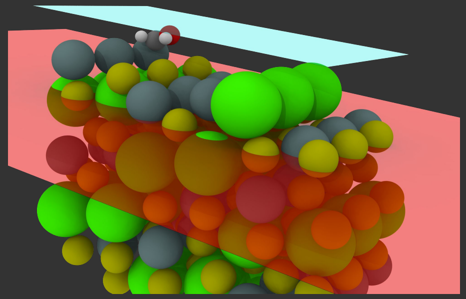
General constraints to preserve molecular identity

Steer optimization towards physically relevant configurations

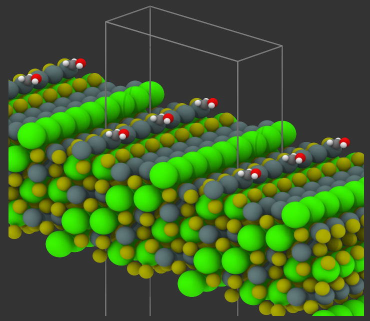
Dissociation

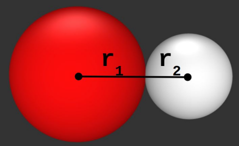


Desorption and Intercalation



Surface Reconstruction



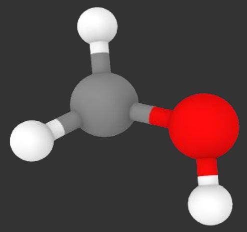


Generalized Hookean constraints based on contact matrix built from covalent radii

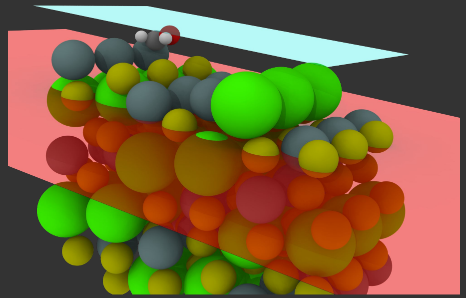
General constraints to preserve molecular identity

Steer optimization towards physically relevant configurations.

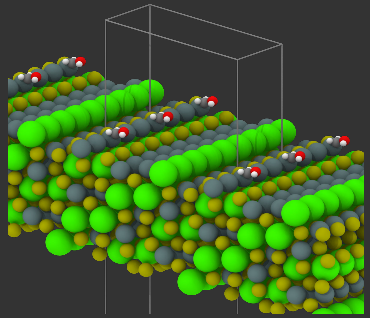
Dissociation



Desorption and Intercalation



Surface Reconstruction



Model	Anomalies	Dissociated	Desorbed	Intercalated	Reconstructed
Baseline	**19.3%**	0.3%	10.1%	2.3%	8.8%
Our approach	**12.4%**	0.0%	2.9%	2.0%	8.7 %

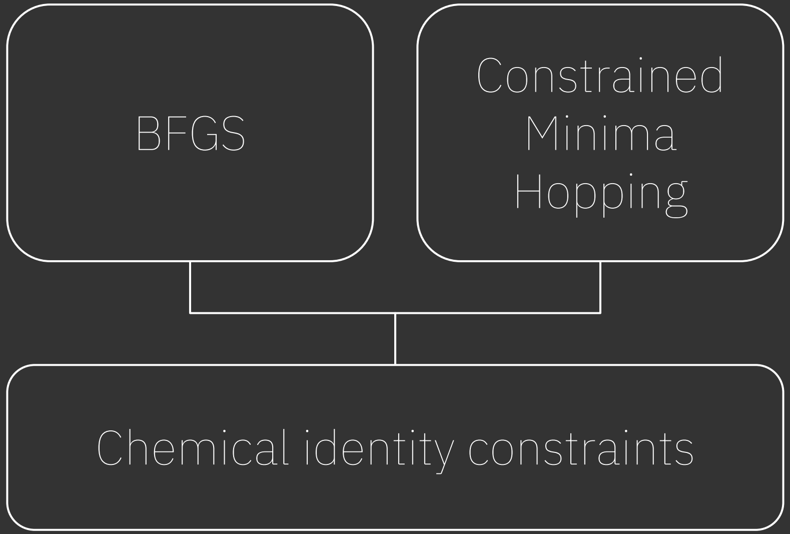
Outline of *the job*

1. Search candidate *configurations*
2. Improve *energy* estimation
3. Take the lowest minima

	# systems $E_{ML}(x) < E_{DFT} + 0.1$ eV
BFGS	79.5 %
BFGS + constraints	82.1 %
BFGS + Constraints + Minima Hopping	**87.2 %**

Results on the (balanced) validation dataset.

Local and global optimization schemes with different ML potentials [1]



[1] EquiformerV2, GemNet-OC and eSCN (OC20 2M)

Outline of *the job*

1. Search candidate configurations
2. **Improve energy estimation**
3. Take the lowest minima

Original idea:
Fine-tune a pre-trained ML model.
(discarded - too costly)

Outline of *the job*

1. Search candidate configurations
2. **Improve energy estimation**
3. Take the lowest minima

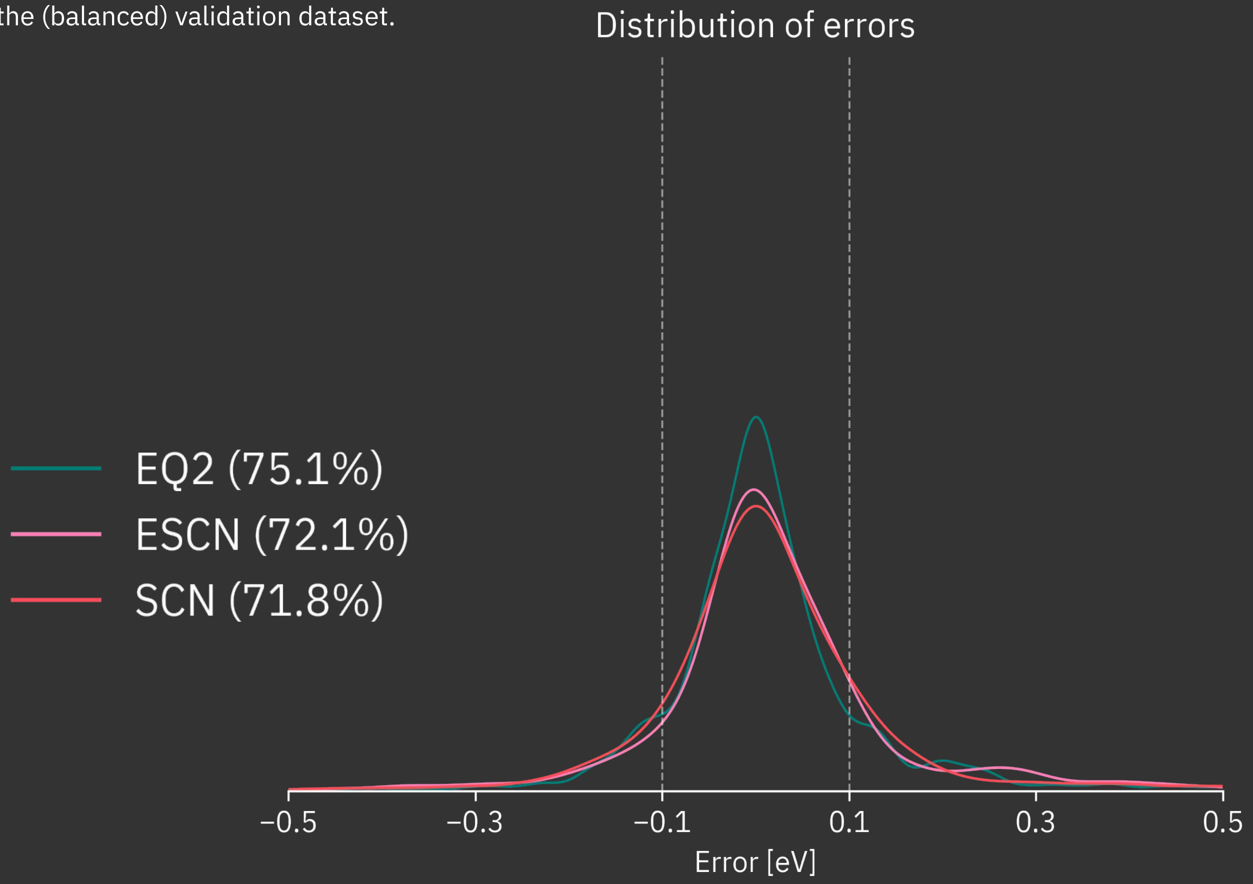
Our approach:
ML models are wrong in different ways

Ensembling

Uncertainty filtering

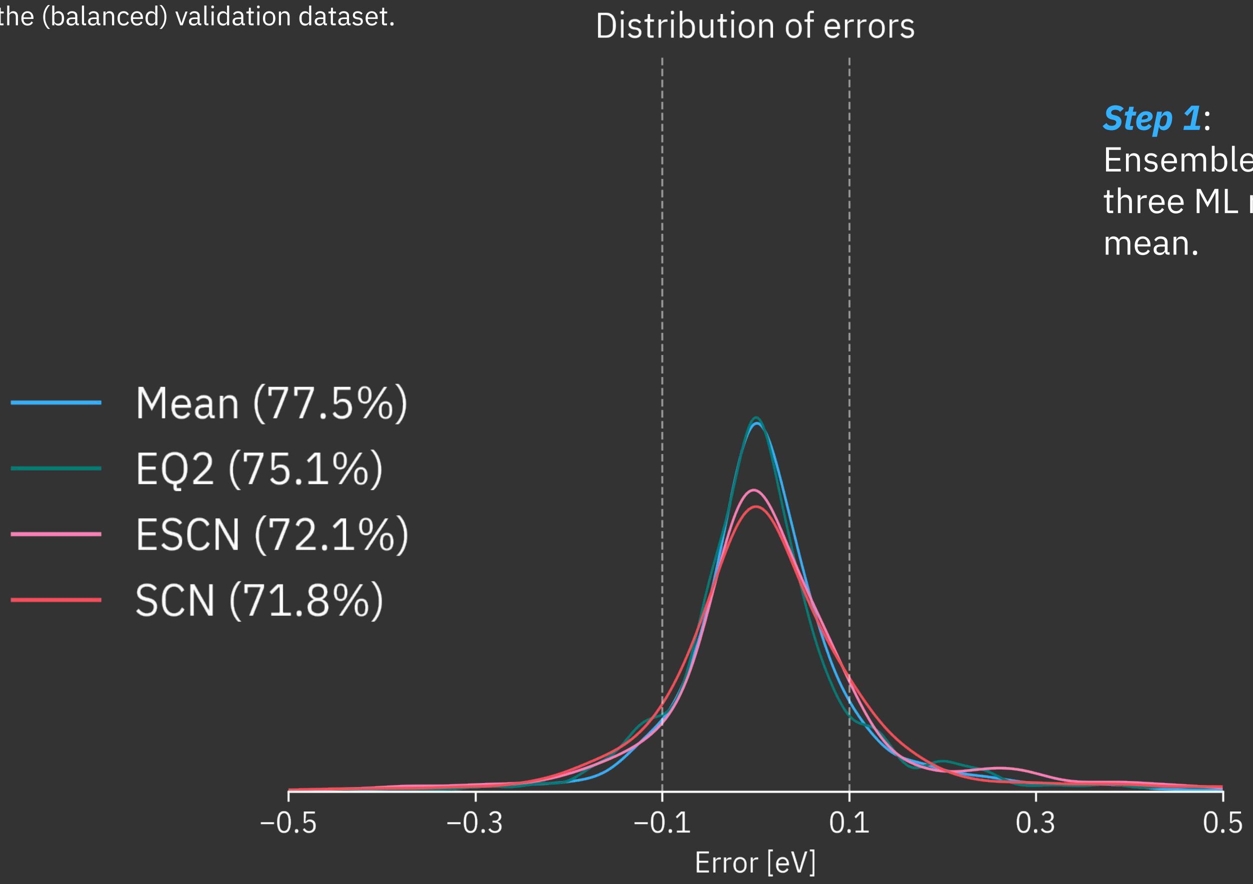
Improving energy estimation

Results on the (balanced) validation dataset.



Improving energy estimation via ensembling

Results on the (balanced) validation dataset.



Step 1:
Ensemble the prediction of the three ML models by taking their mean.

Improving energy estimation with filtering

Assumption: the spread of ML predictions is a proxy of their quality

Standard deviation
Range (*max* - *min*)

$$|E_{\mathrm{ML}} - E_{\mathrm{DFT}}|$$

Improving energy estimation with filtering

Assumption: the spread of ML predictions is a proxy of their quality

Standard deviation
Range (*max* - *min*)

$$|E_{\text{ML}} - E_{\text{DFT}}|$$

EQ2 (EquiformerV2)	1.332 eV	
eSCN	1.376 eV	→ Spread 0.115 eV
SCN	1.447 eV	

Improving energy estimation with filtering

Assumption: the spread of ML predictions is a proxy of their quality

Standard deviation
Range (*max* - *min*)

$$|E_{\mathrm{ML}} - E_{\mathrm{DFT}}|$$

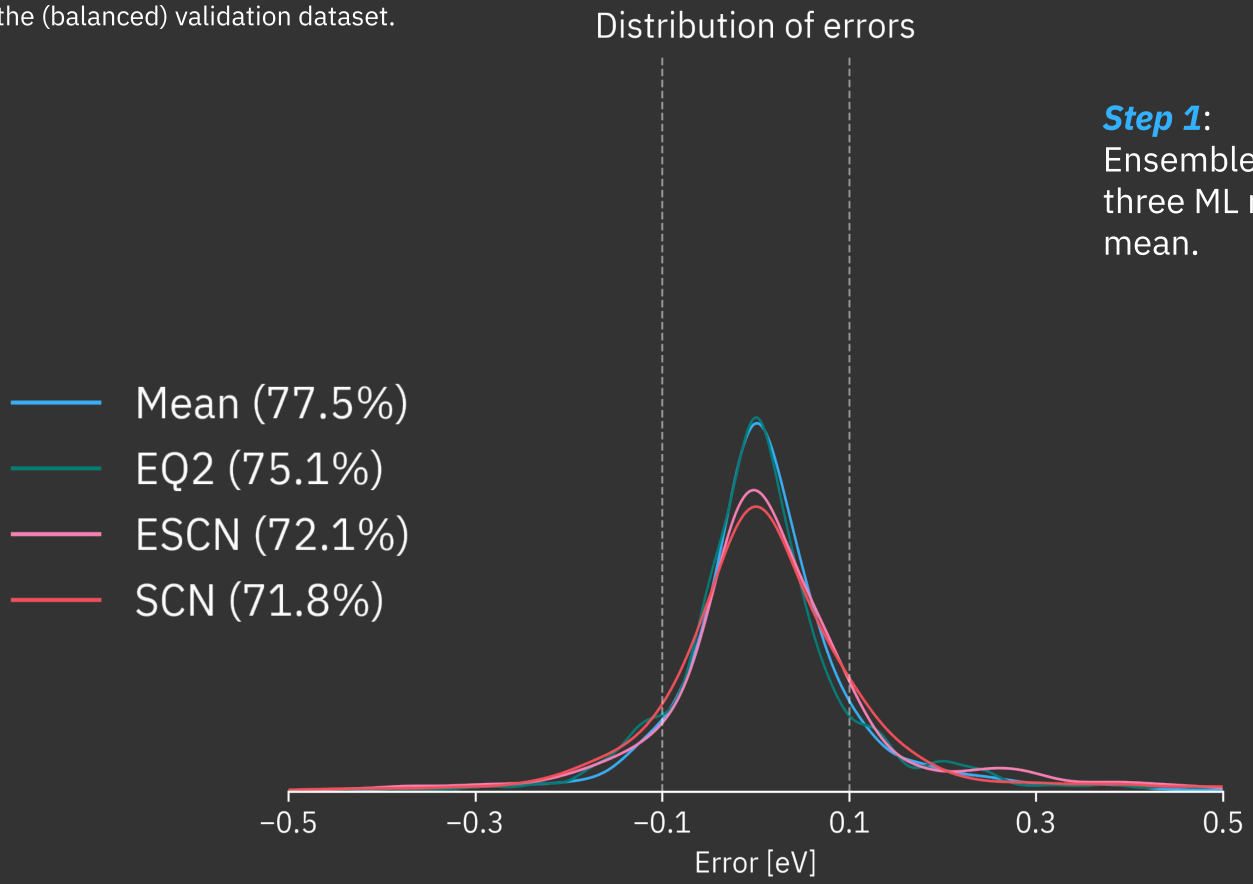
EQ2 (EquiformerV2)	1.332 eV
eSCN	1.376 eV
SCN	1.447 eV

→ Spread 0.115 eV

On validation, around **60 % correlation** between spread and DFT error.

Improving energy estimation via ensembling

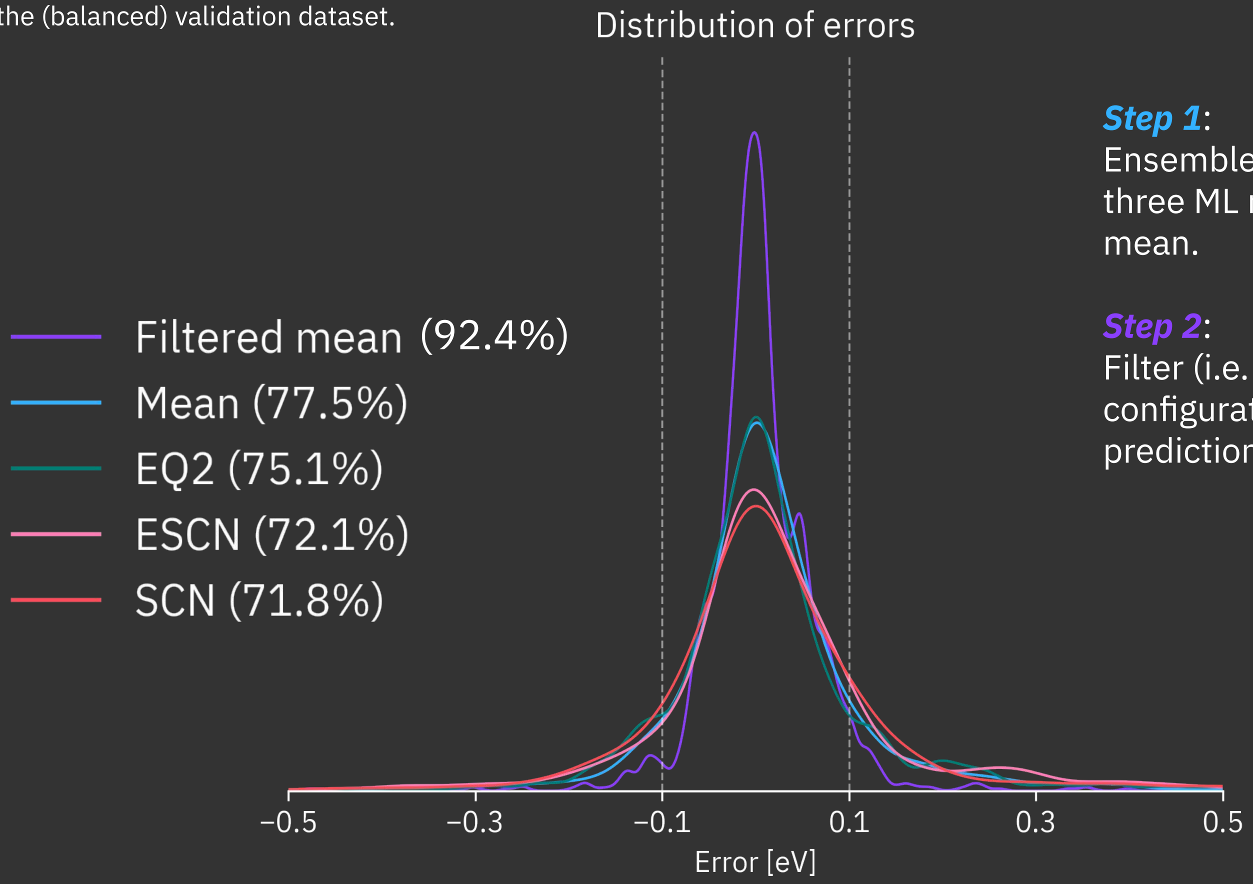
Results on the (balanced) validation dataset.



Step 1:
Ensemble the prediction of the three ML models by taking their mean.

Improving energy estimation with filtering

Results on the (balanced) validation dataset.



Step 1:
Ensemble the prediction of the three ML models by taking their mean.

Step 2:
Filter (i.e. ignore) the configurations with a ML prediction spread > 0.1 eV

Outline of *the job*

1. Search candidate configurations
2. Improve energy estimation
3. Take the lowest minima

Our approach:
ML models are wrong in different ways

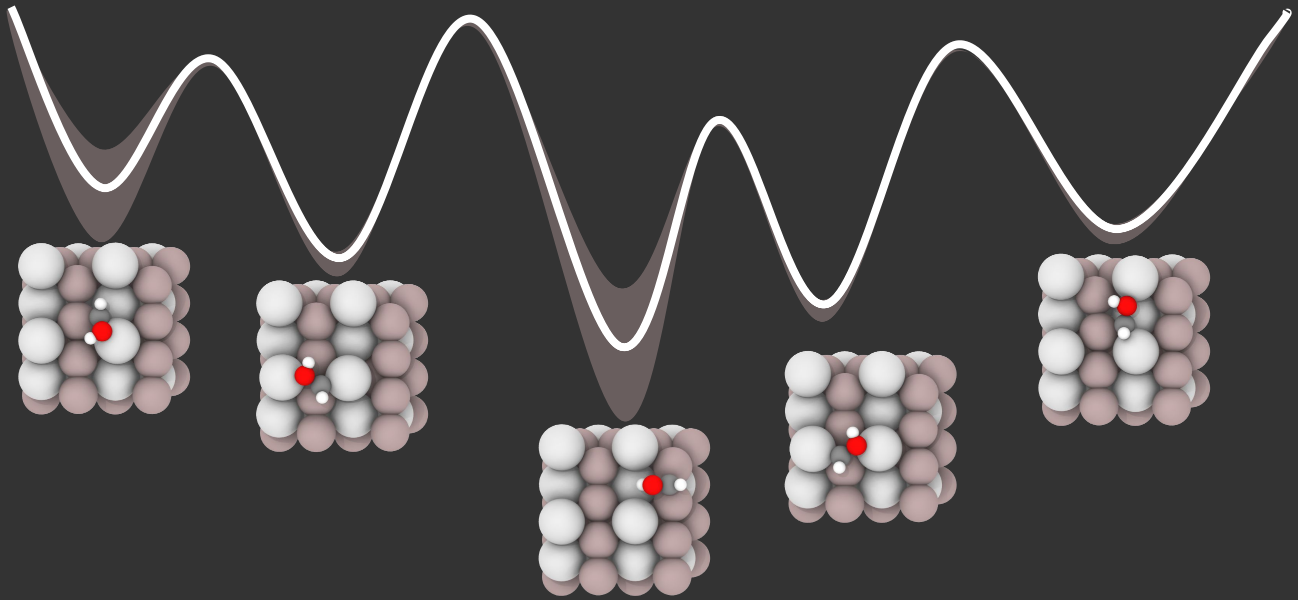
Ensembling

Uncertainty filtering

# configs.	$\lvert E_{ML} - E_{DFT} \rvert < 0.1$ eV
EquiformerV2	75.1 %
Ensembling	77.5 %
Ensembling +Filtering	**92.4 %**

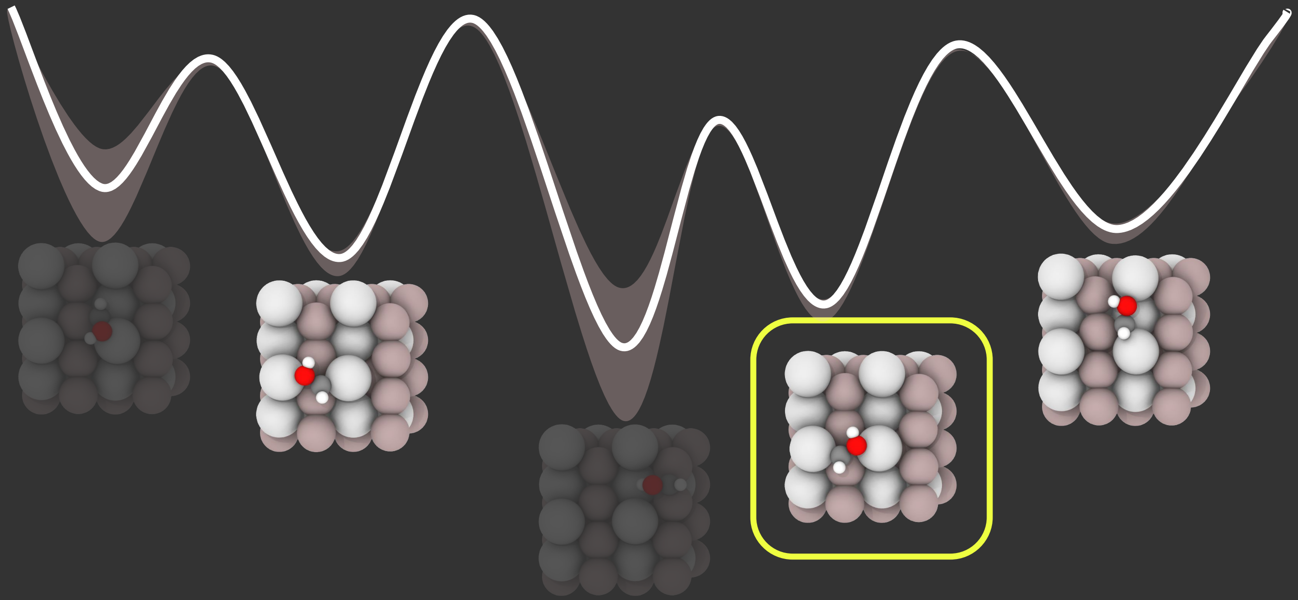
Outline of *the job*

1. Search candidate configurations
2. Improve energy estimation
3. **Take the lowest minima**



Outline of *the job*

1. Search candidate configurations
2. Improve energy estimation
3. **Take the lowest minima**



Thank you!