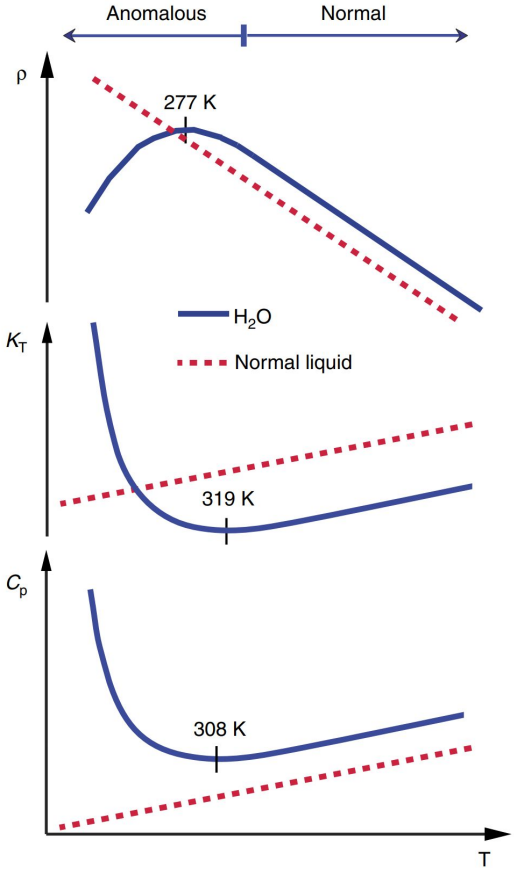
Exploring hidden structure in water using polarizable atom neural networks

HAMLET-Physics Conference, University of Copenhagen

2026-08-19

Fredrik Grote

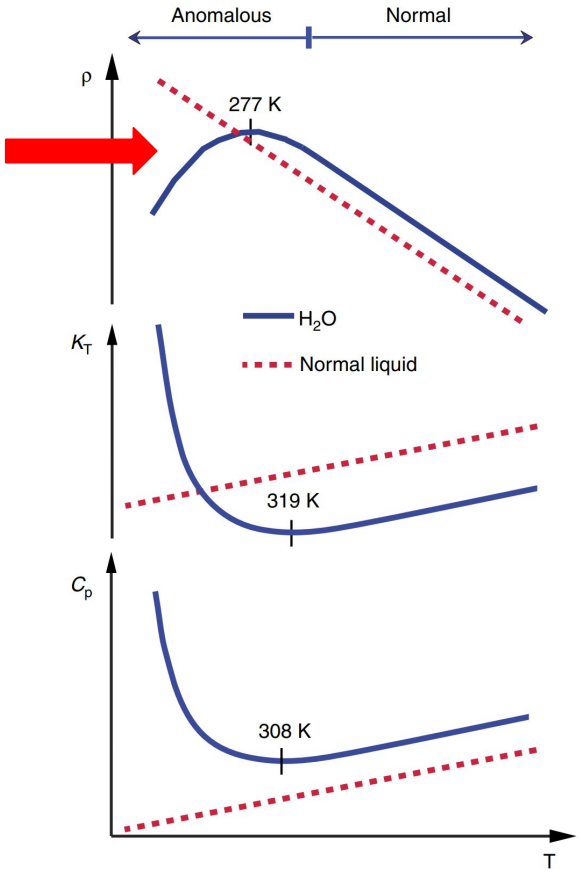
Water anomalies



R. J. Speedy and C. A. Angell, *J. Chem. Phys.*, 65, (1976)

Water anomalies

Density maximum

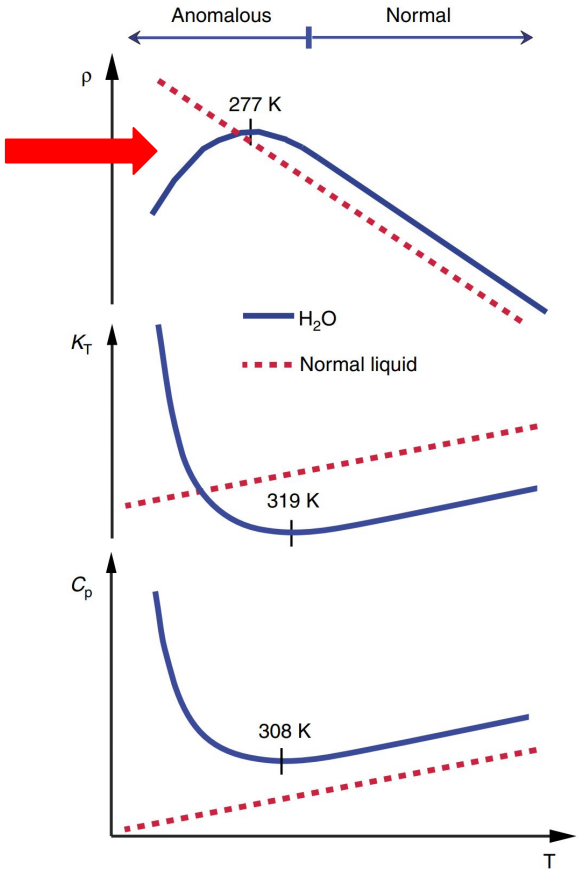


Density

R. J. Speedy and C. A. Angell, *J. Chem. Phys.*, 65, (1976)

Water anomalies

Density maximum

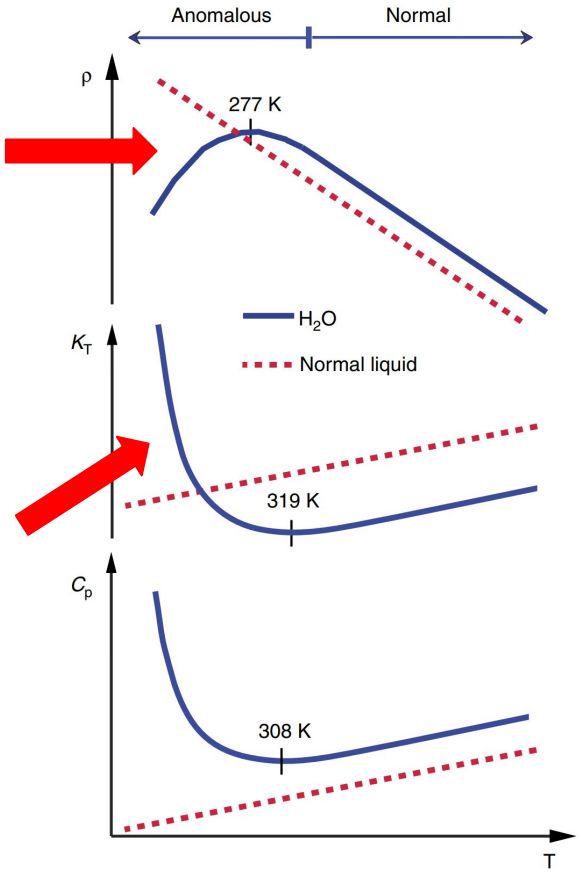


Density

R. J. Speedy and C. A. Angell, *J. Chem. Phys.*, 65, (1976)

Water anomalies

Density maximum



Density

Fluctuations increase as temperature decreases

Density fluctuations

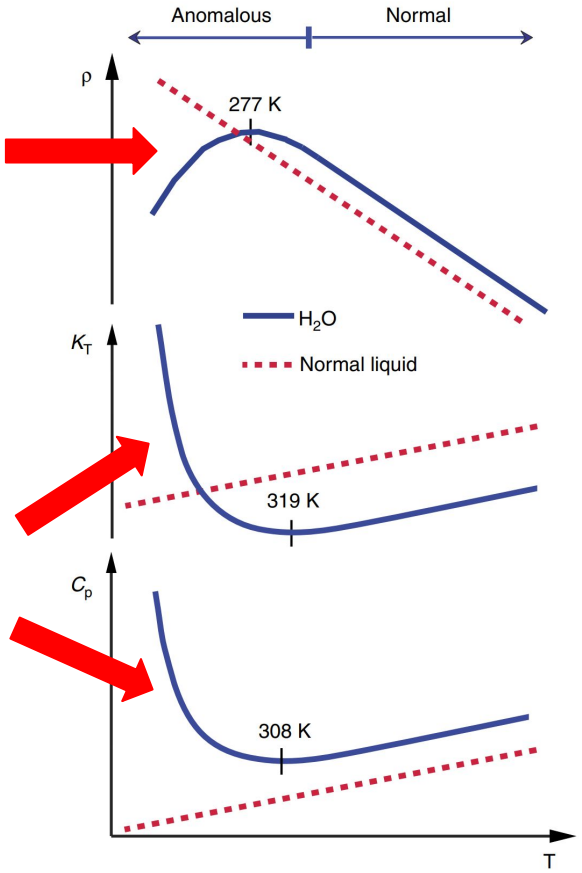
$$\kappa_T \propto \langle V^2 \rangle - \langle V \rangle^2$$

R. J. Speedy and C. A. Angell, *J. Chem. Phys.*, 65, (1976)

Water anomalies

Density maximum

Fluctuations increase as temperature decreases



Density

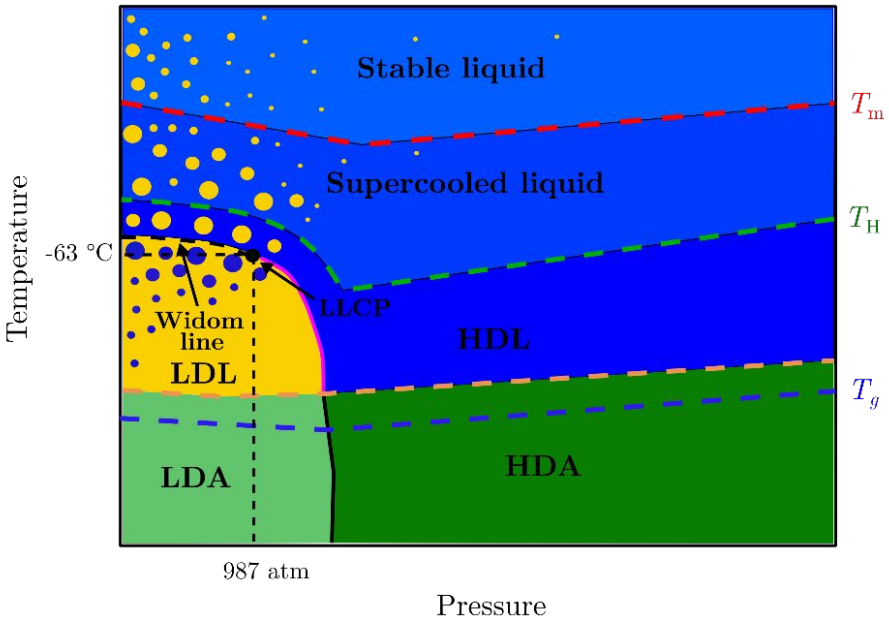
Density fluctuations $\kappa_T \propto \langle V^2 \rangle - \langle V \rangle^2$

Entropy fluctuations $C_p \propto \langle S^2 \rangle - \langle S \rangle^2$

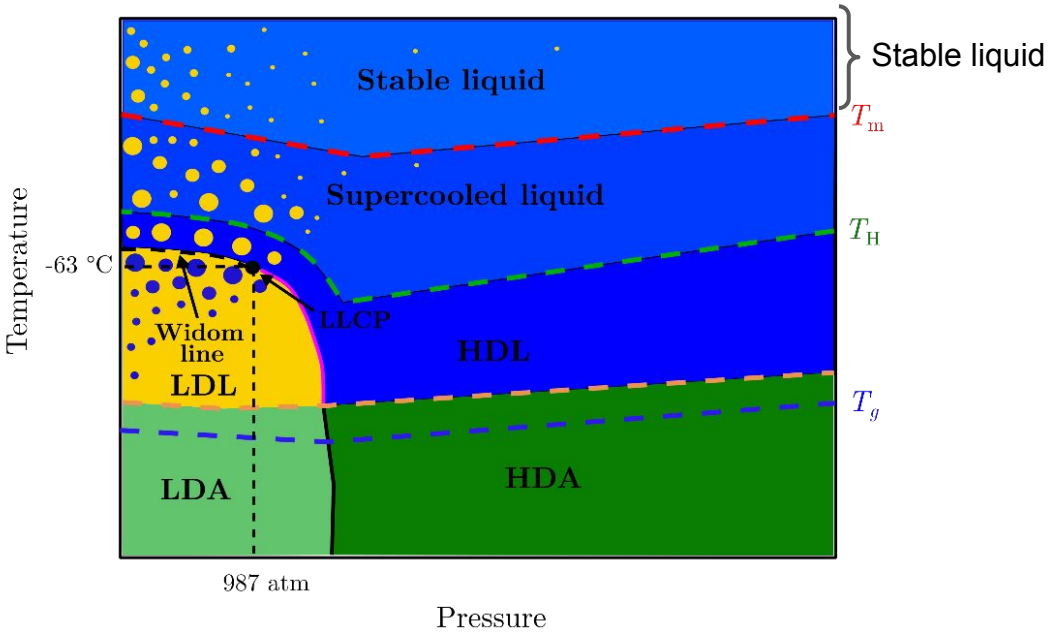
R. J. Speedy and C. A. Angell, *J. Chem. Phys.*, 65, (1976)

What is the origin of these *non-thermal* fluctuations?

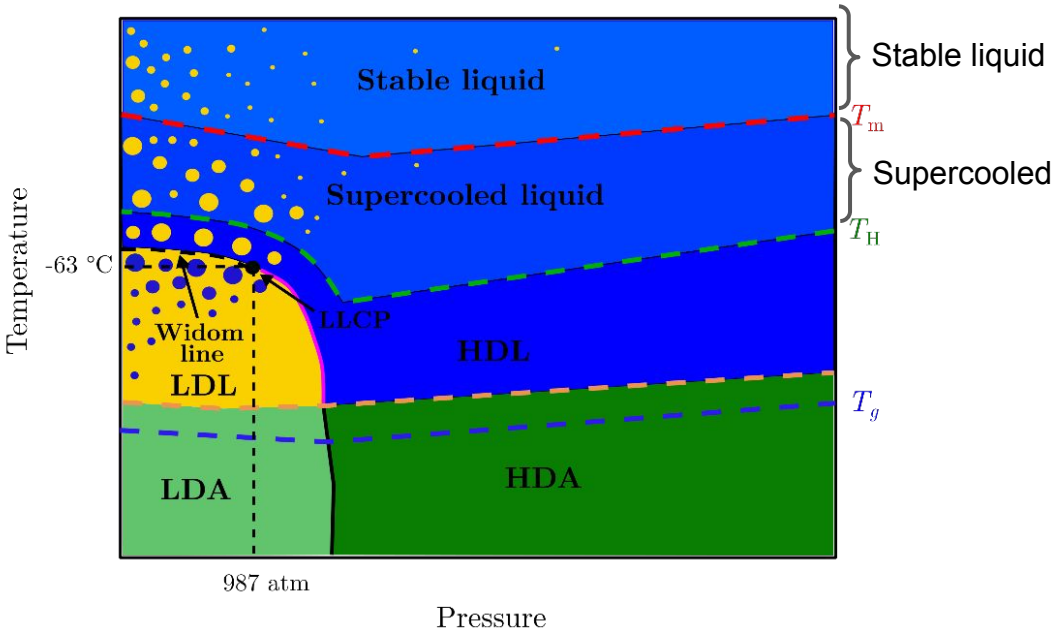
Liquid-liquid transition and liquid-liquid critical point



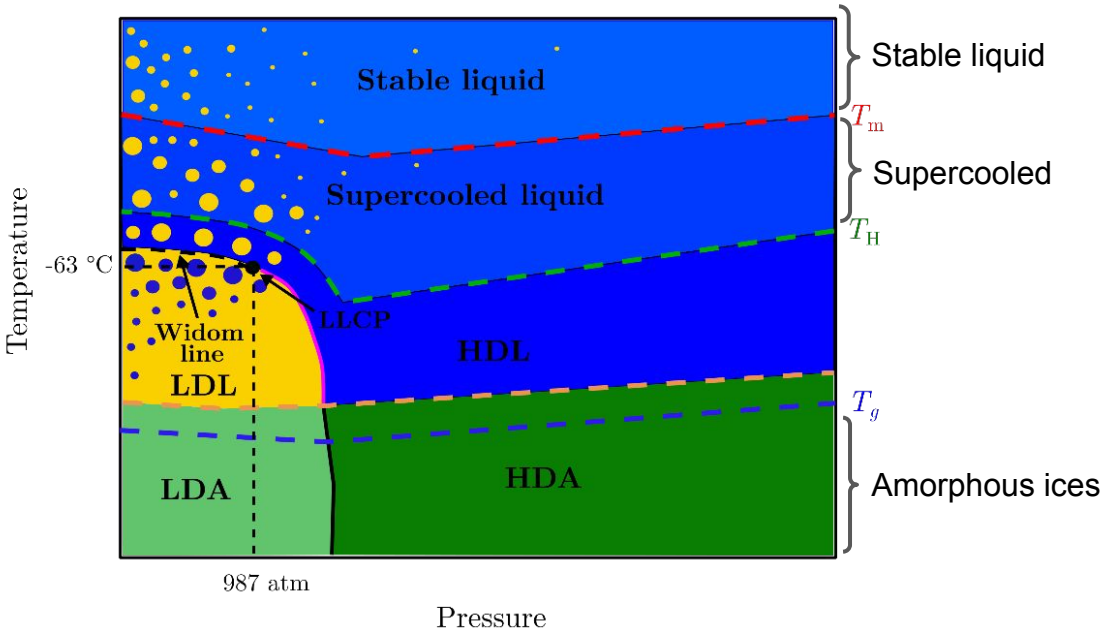
Liquid-liquid transition and liquid-liquid critical point



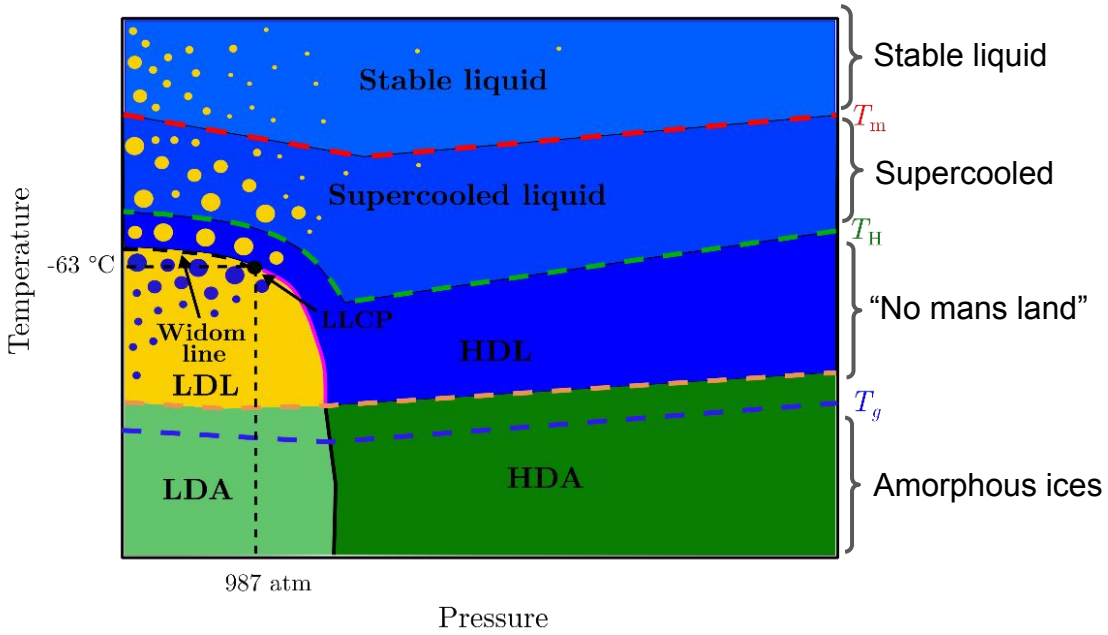
Liquid-liquid transition and liquid-liquid critical point



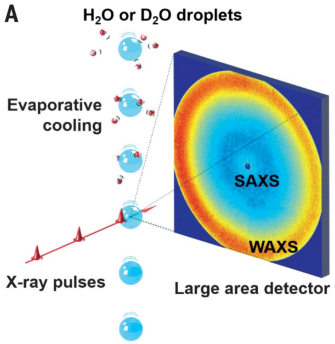
Liquid-liquid transition and liquid-liquid critical point

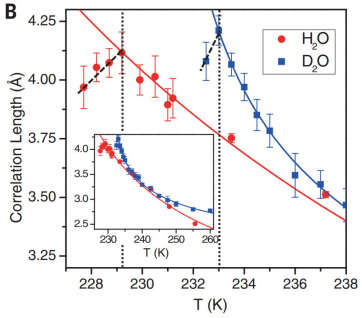


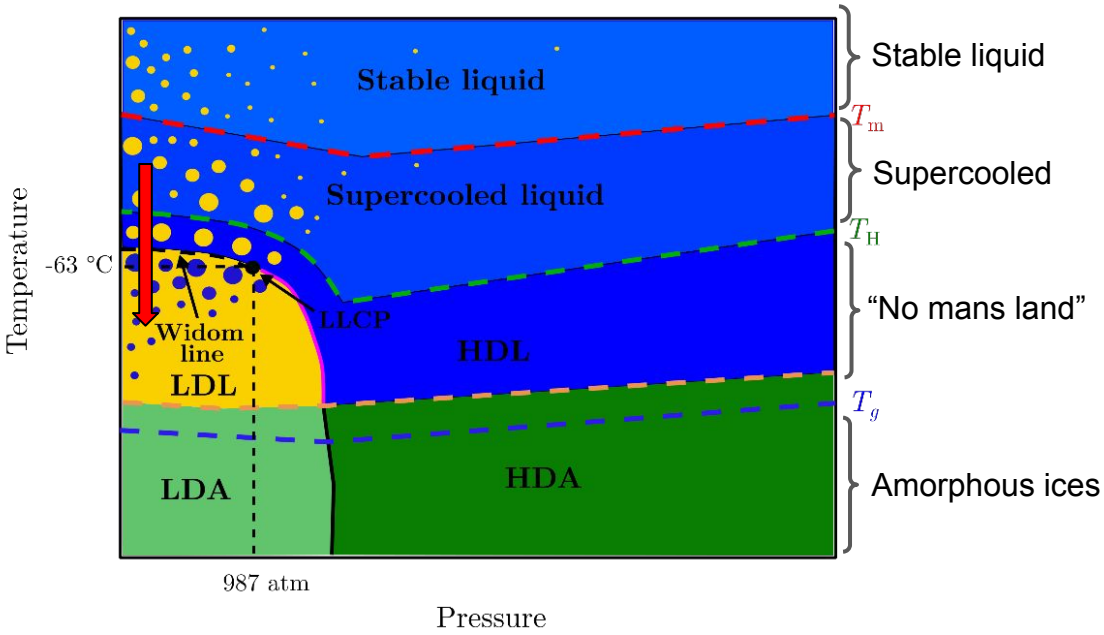
Liquid-liquid transition and liquid-liquid critical point



Liquid-liquid transition and liquid-liquid critical point

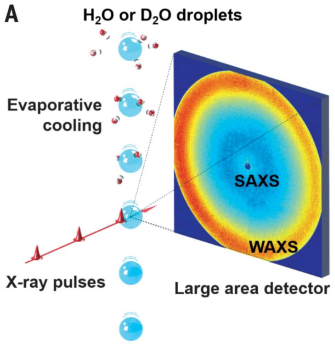


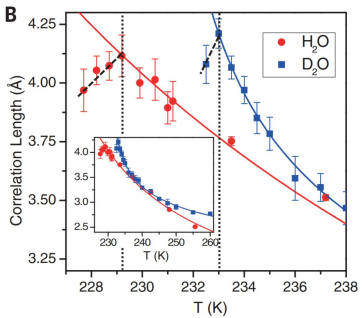


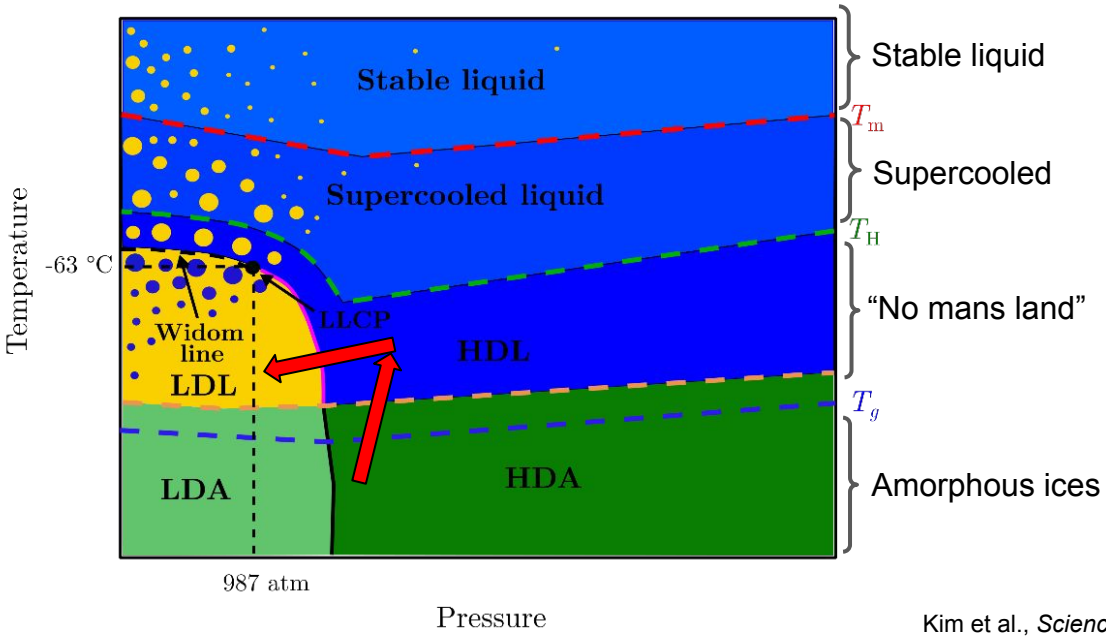


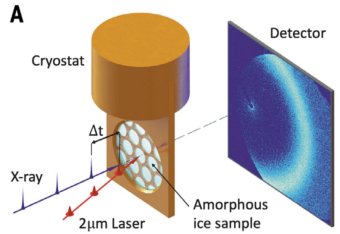
Kim et al., *Science* **358**, 1589-1593 (2017)

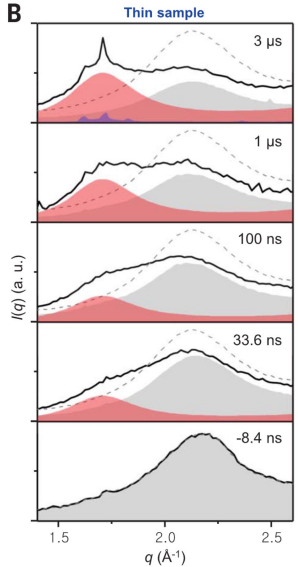
Liquid-liquid transition and liquid-liquid critical point







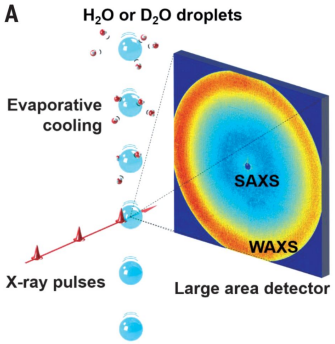


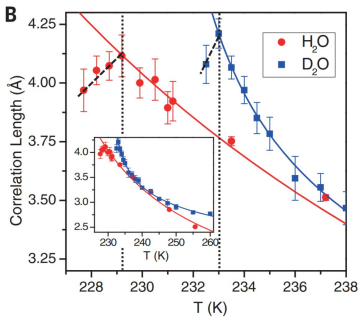


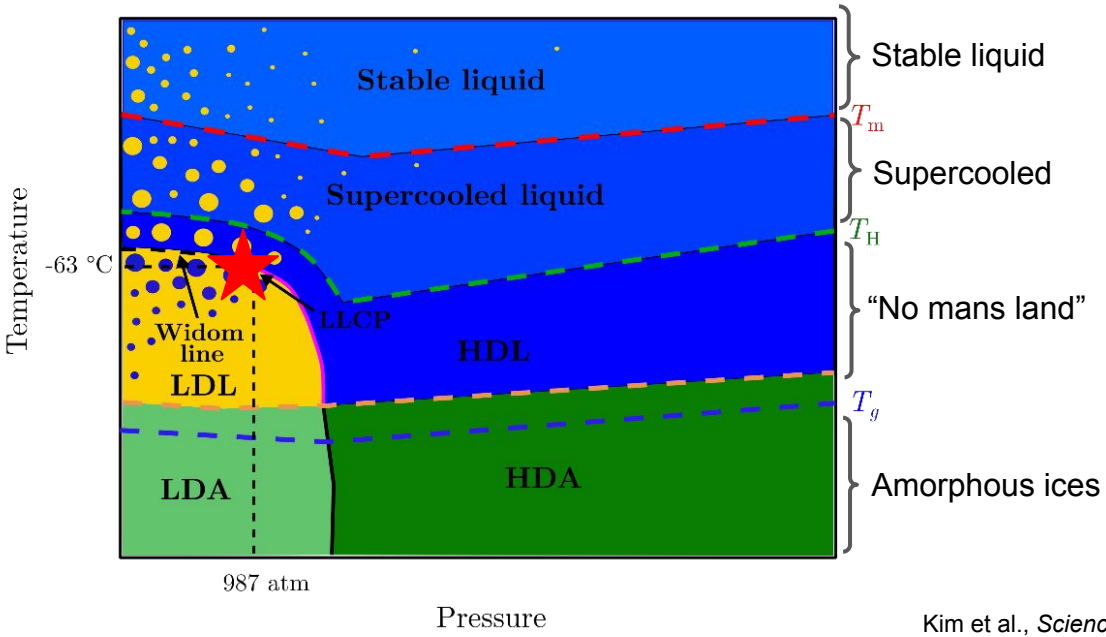
Kim et al., *Science* **370**, 978-982 (2020)

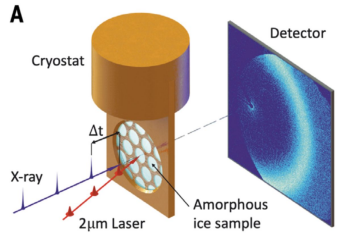
Kim et al., *Science* **358**, 1589-1593 (2017)

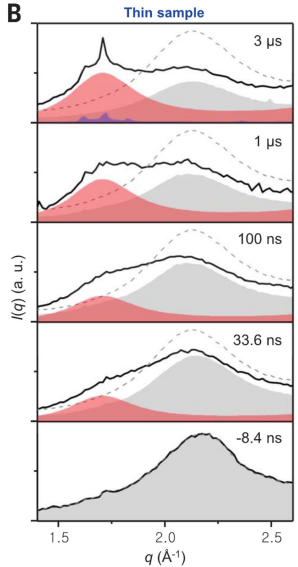
Liquid-liquid transition and liquid-liquid critical point









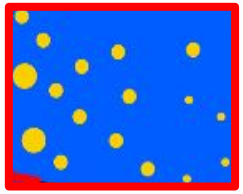


Kim et al., *Science* **358**, 1589-1593 (2017)

Kim et al., *Science* **370**, 978-982 (2020)

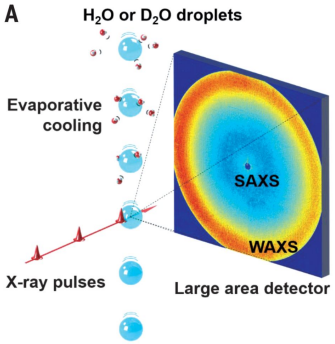
You et al., *Science* **391**, 1387-1391 (2026)

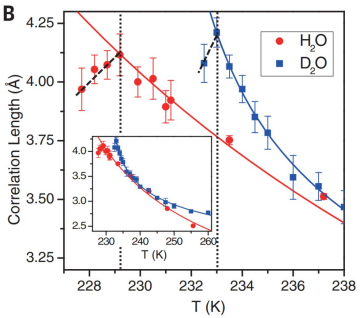
Liquid-liquid transition and liquid-liquid critical point

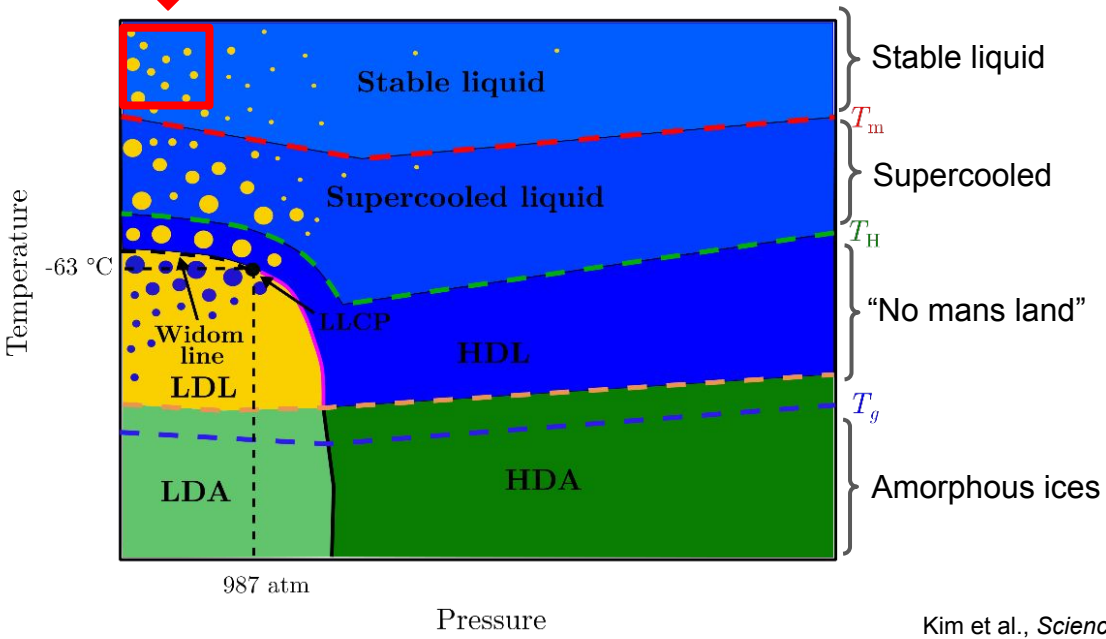


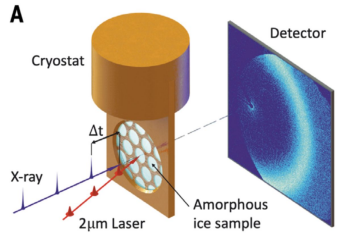
Ambient water is mainly **HDL**
with **LDL** fluctuations

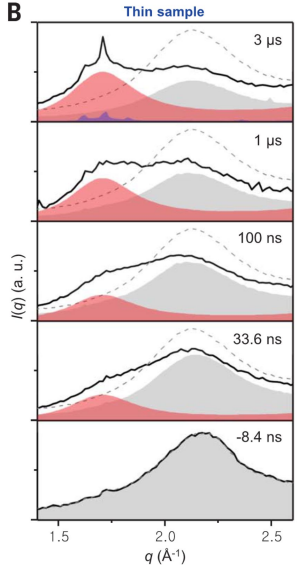
"Supercritical"











Kim et al., *Science* **358**, 1589-1593 (2017)

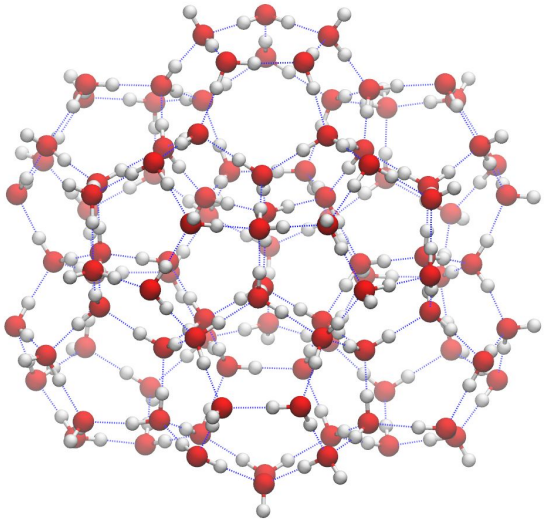
Kim et al., *Science* **370**, 978-982 (2020)

You et al., *Science* **391**, 1387-1391 (2026)

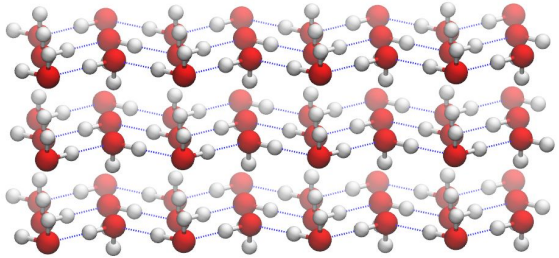
What is the structure of HDL and LDL?

A proposal for the structure of LDL and HDL

LDL - Fused dodecahedra



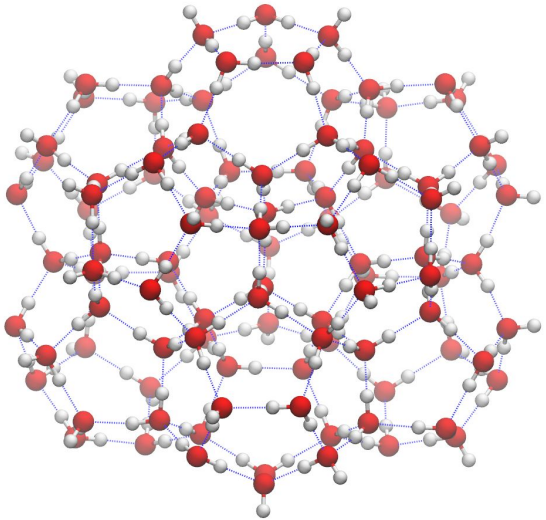
HDL - Chains



Camisasca et al., *J. Chem. Phys.* 151, (2019)

A proposal for the structure of LDL and HDL

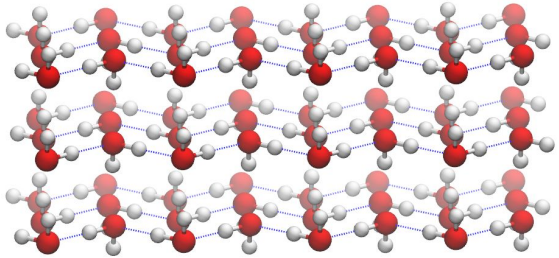
LDL - Fused dodecahedra



Maximize number of H-bonds

⇌

HDL - Chains



Librational motion

A competition between energy and entropy

Camisasca et al., *J. Chem. Phys.* 151, (2019)

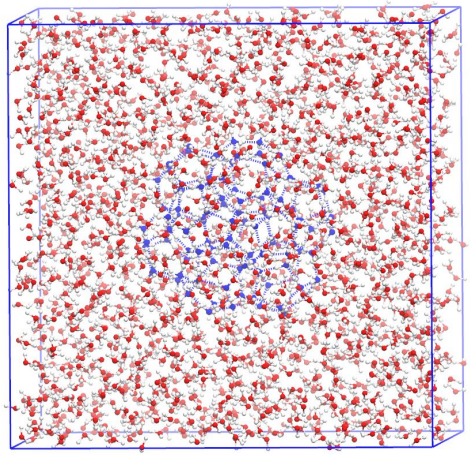
How can we study these template structures?

Molecular dynamics simulations and the FCM-GAP water model

Newtonian equation of motion

$$\mathbf{F}_i = m_i \mathbf{a}_i$$

$$\mathbf{F}_i = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n)$$

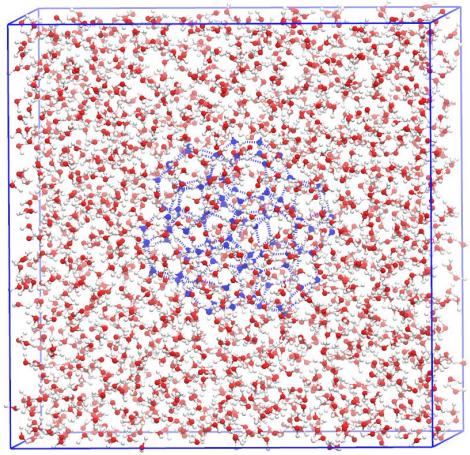


Molecular dynamics simulations and the FCM-GAP water model

Newtonian equation of motion

$$\mathbf{F}_i = m_i \mathbf{a}_i$$

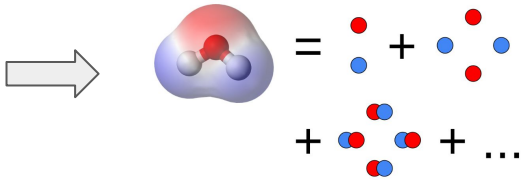
$$\mathbf{F}_i = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n)$$

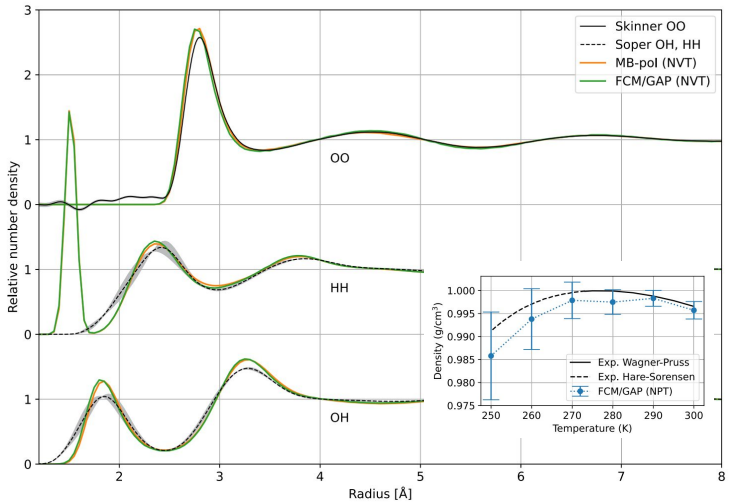


FCM-GAP water model

CCSD(T)

SAPT





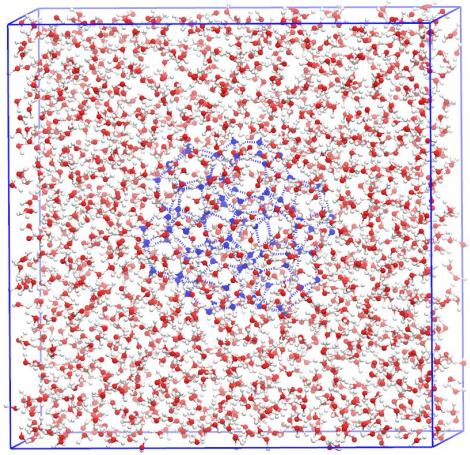
J. Öström and L. G. M. Pettersson, *J. Chem. Phys.* 163, (2025)

Molecular dynamics simulations and the FCM-GAP water model

Newtonian equation of motion

$$\mathbf{F}_i = m_i \mathbf{a}_i$$

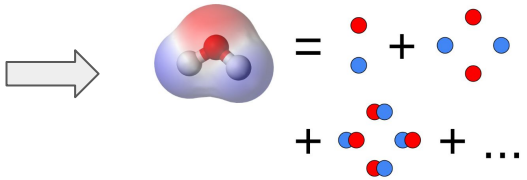
$$\mathbf{F}_i = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n)$$

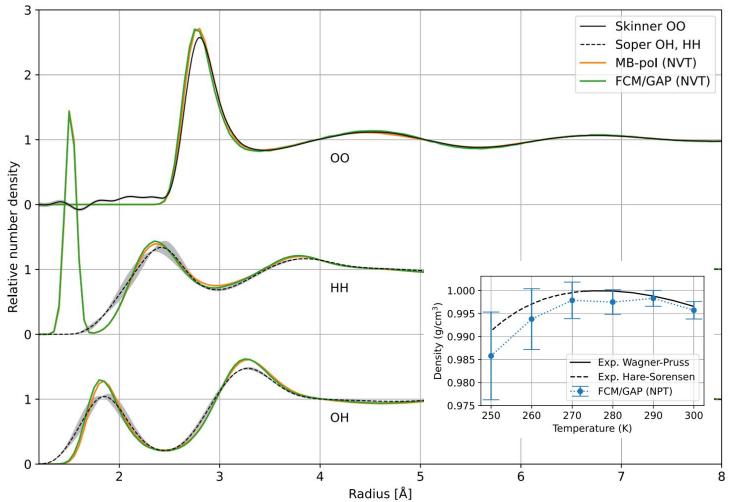


FCM-GAP water model

CCSD(T)

SAPT





High computational cost - Limited to system size ~100 molecules

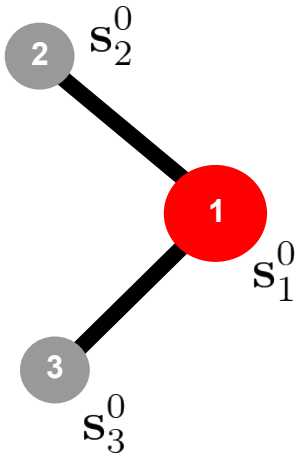
J. Öström and L. G. M. Pettersson, *J. Chem. Phys.* 163, (2025)

How can we extend the limits of the FCM-GAP water model?

Polarizable atom interaction neural network (PAINN)

Initialization

$$\mathbf{s}_i^0 \in \mathbb{R}^d$$



K. T. Schütt, O. T. Unke, M, Gastegger, *Proceedings of the 38 th International Conference on Machine Learning, PMLR,* 139, (2021).

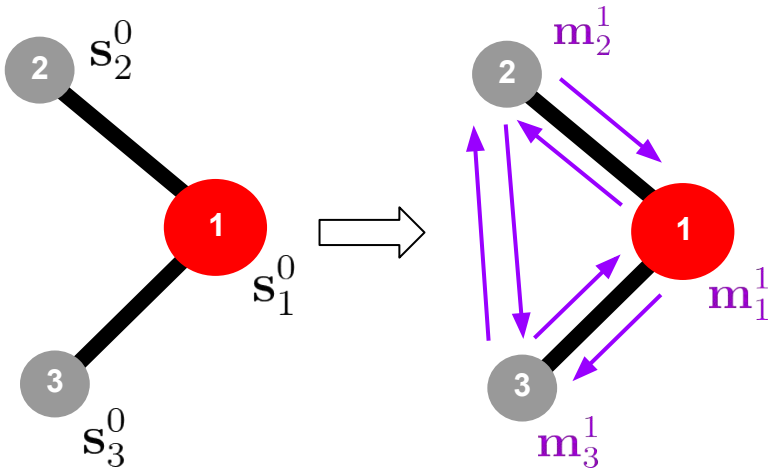
Polarizable atom interaction neural network (PAINN)

Initialization

$$\mathbf{s}_i^0 \in \mathbb{R}^d$$

Interaction

$$\mathbf{m}_i^{t+1} = \sum_{j \in N(i)} \mathbf{M}_t(\mathbf{s}_i^t, \mathbf{s}_j^t, \vec{r}_{ij})$$



K. T. Schütt, O. T. Unke, M, Gastegger, *Proceedings of the 38 th International Conference on Machine Learning, PMLR,* 139, (2021).

Polarizable atom interaction neural network (PAINN)

Initialization

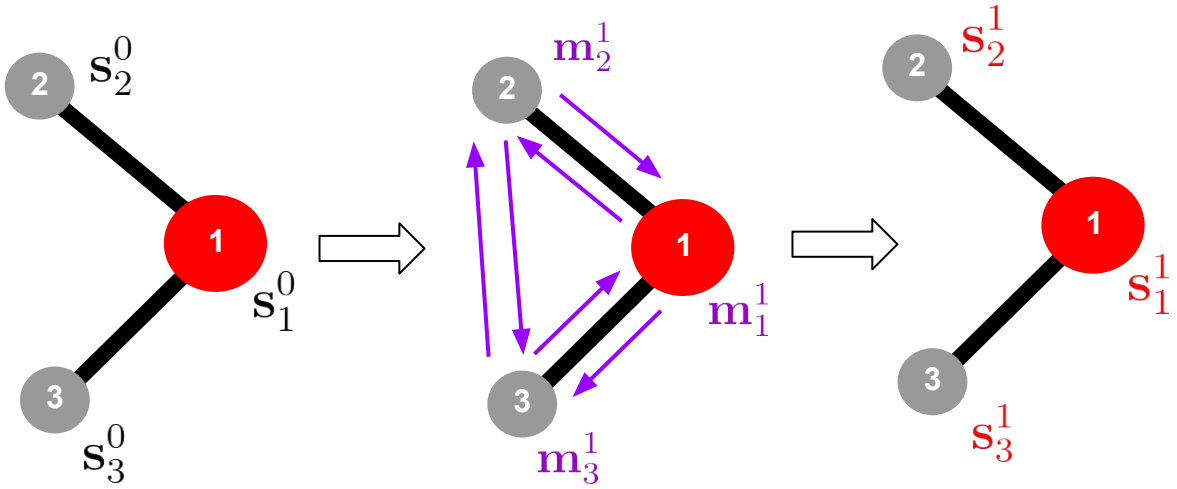
$$\mathbf{s}_i^0 \in \mathbb{R}^d$$

Interaction

$$\mathbf{m}_i^{t+1} = \sum_{j \in N(i)} \mathbf{M}_t(\mathbf{s}_i^t, \mathbf{s}_j^t, \vec{r}_{ij})$$

Update

$$\mathbf{s}_i^{t+1} = \mathbf{U}_t(\mathbf{s}_i^t, \mathbf{m}_i^{t+1})$$



K. T. Schütt, O. T. Unke, M, Gastegger, *Proceedings of the 38 th International Conference on Machine Learning, PMLR,* 139, (2021).

Polarizable atom interaction neural network (PAINN)

Initialization

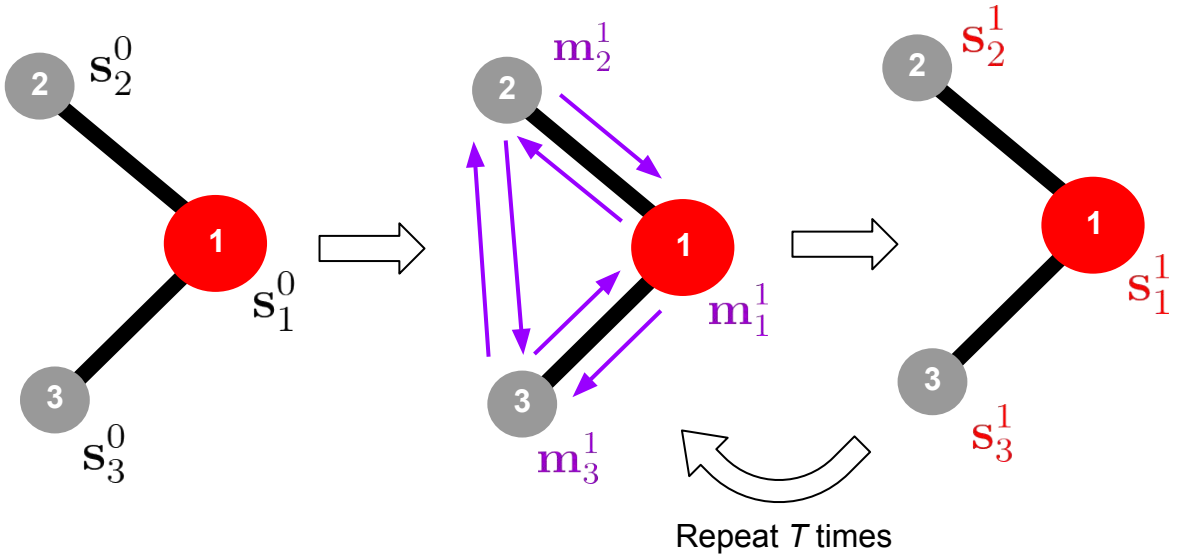
$$\mathbf{s}_i^0 \in \mathbb{R}^d$$

Interaction

$$\mathbf{m}_i^{t+1} = \sum_{j \in N(i)} \mathbf{M}_t(\mathbf{s}_i^t, \mathbf{s}_j^t, \vec{r}_{ij})$$

Update

$$\mathbf{s}_i^{t+1} = \mathbf{U}_t(\mathbf{s}_i^t, \mathbf{m}_i^{t+1})$$



K. T. Schütt, O. T. Unke, M, Gastegger, *Proceedings of the 38 th International Conference on Machine Learning, PMLR,* 139, (2021).

Polarizable atom interaction neural network (PAINN)

Initialization

$$\mathbf{s}_i^0 \in \mathbb{R}^d$$

Interaction

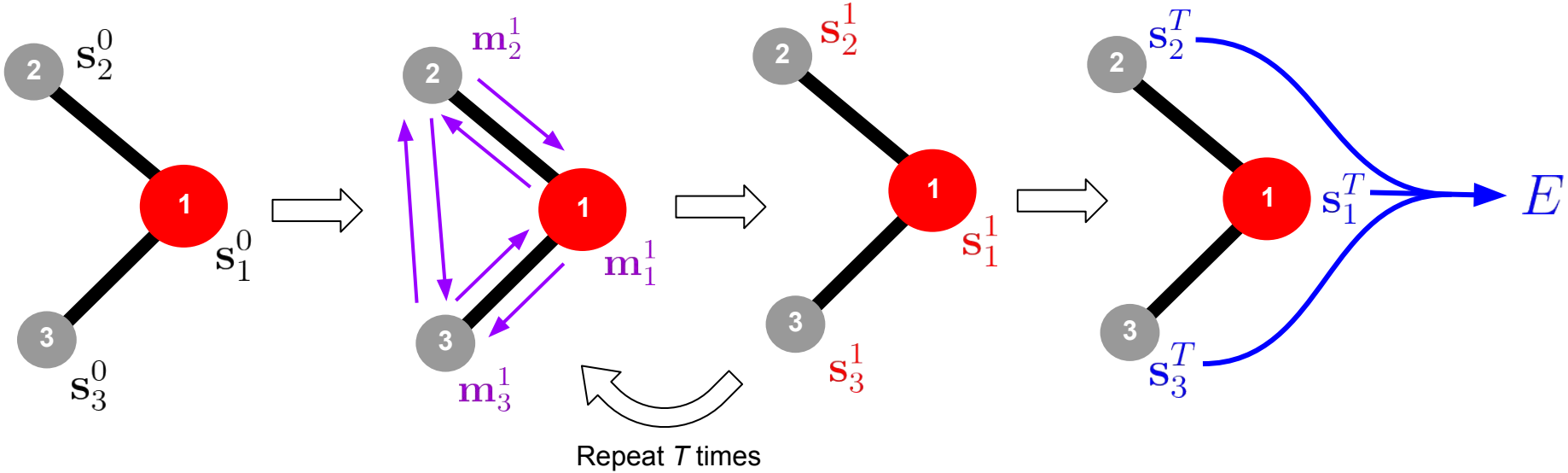
$$\mathbf{m}_i^{t+1} = \sum_{j \in N(i)} \mathbf{M}_t(\mathbf{s}_i^t, \mathbf{s}_j^t, \vec{r}_{ij})$$

Update

$$\mathbf{s}_i^{t+1} = \mathbf{U}_t(\mathbf{s}_i^t, \mathbf{m}_i^{t+1})$$

Output

$$E = \sum_i \epsilon(\mathbf{s}_i^T)$$



K. T. Schütt, O. T. Unke, M, Gastegger, *Proceedings of the 38 th International Conference on Machine Learning, PMLR,* 139, (2021).

Polarizable atom interaction neural network (PAINN)

Initialization

$$\mathbf{s}_i^0 \in \mathbb{R}^d$$

Interaction

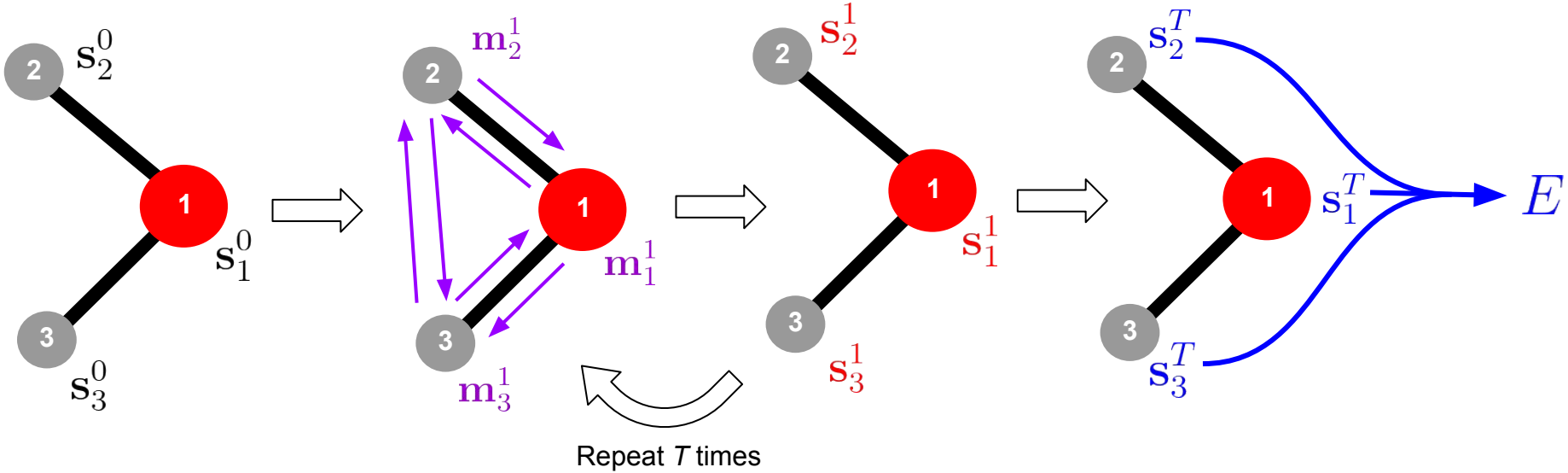
$$\mathbf{m}_i^{t+1} = \sum_{j \in N(i)} \mathbf{M}_t(\mathbf{s}_i^t, \mathbf{s}_j^t, \vec{r}_{ij})$$

Update

$$\mathbf{s}_i^{t+1} = \mathbf{U}_t(\mathbf{s}_i^t, \mathbf{m}_i^{t+1})$$

Output

$$E = \sum_i \epsilon(\mathbf{s}_i^T)$$

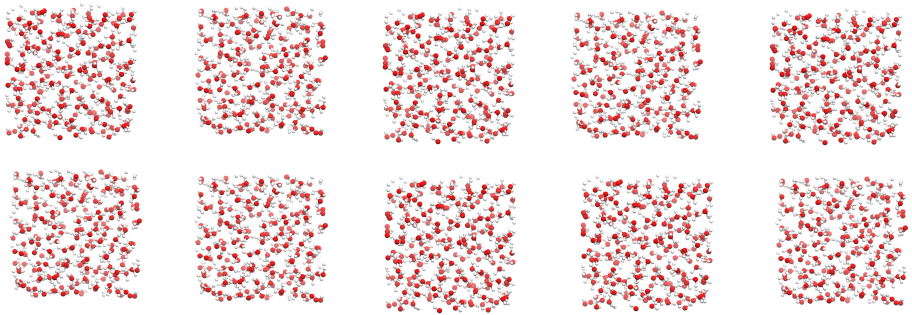


PAINN uses both *invariant* and *equivariant* features

K. T. Schütt, O. T. Unke, M, Gastegger, *Proceedings of the 38 th International Conference on Machine Learning, PMLR,* 139, (2021).

Data set

FCM-GAP: Potential energy, force and stress

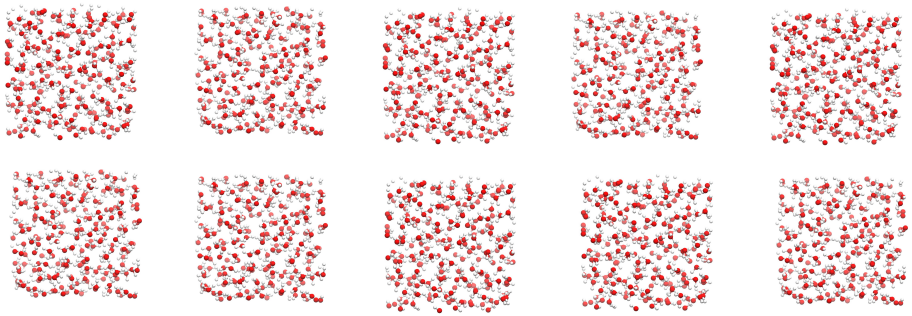


~2 000 structures:

- Bulk MD frames
- Fused dodecahedra
- Distorted geometries

Data set

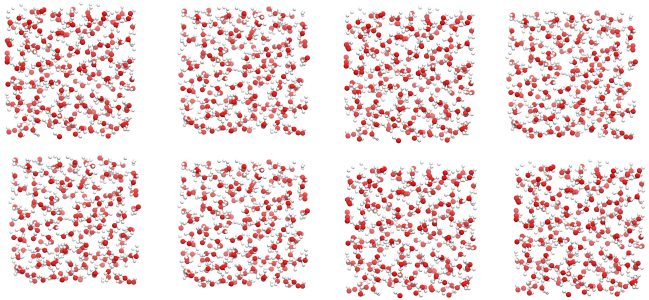
FCM-GAP: Potential energy, force and stress



~2 000 structures:

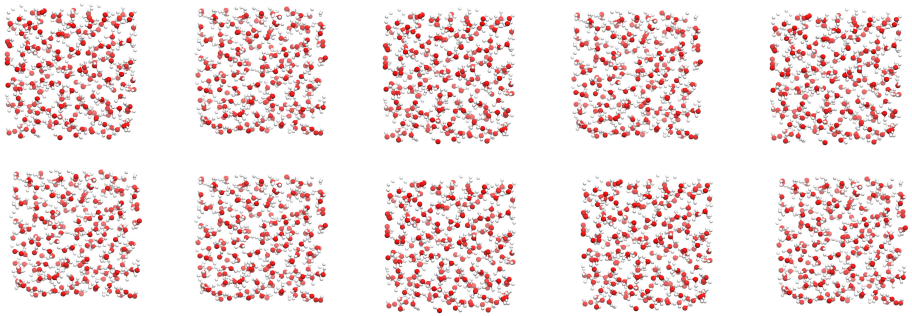
- Bulk MD frames
- Fused dodecahedra
- Distorted geometries

Training set: 80 %



Data set

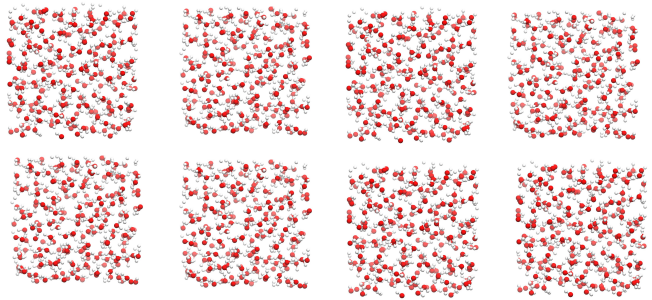
FCM-GAP: Potential energy, force and stress



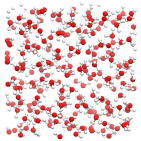
~2 000 structures:

- Bulk MD frames
- Fused dodecahedra
- Distorted geometries

Training set: 80 %

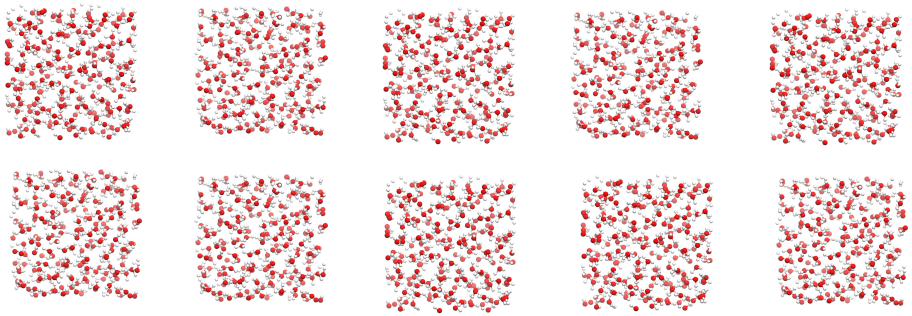


Validation set: 10 %



Data set

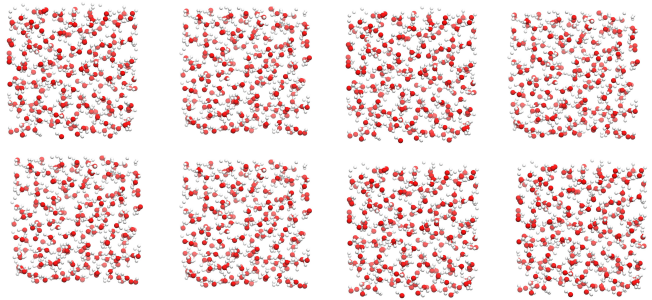
FCM-GAP: Potential energy, force and stress



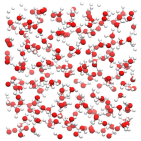
~2 000 structures:

- Bulk MD frames
- Fused dodecahedra
- Distorted geometries

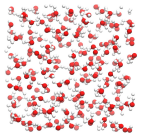
Training set: 80 %



Validation set: 10 %



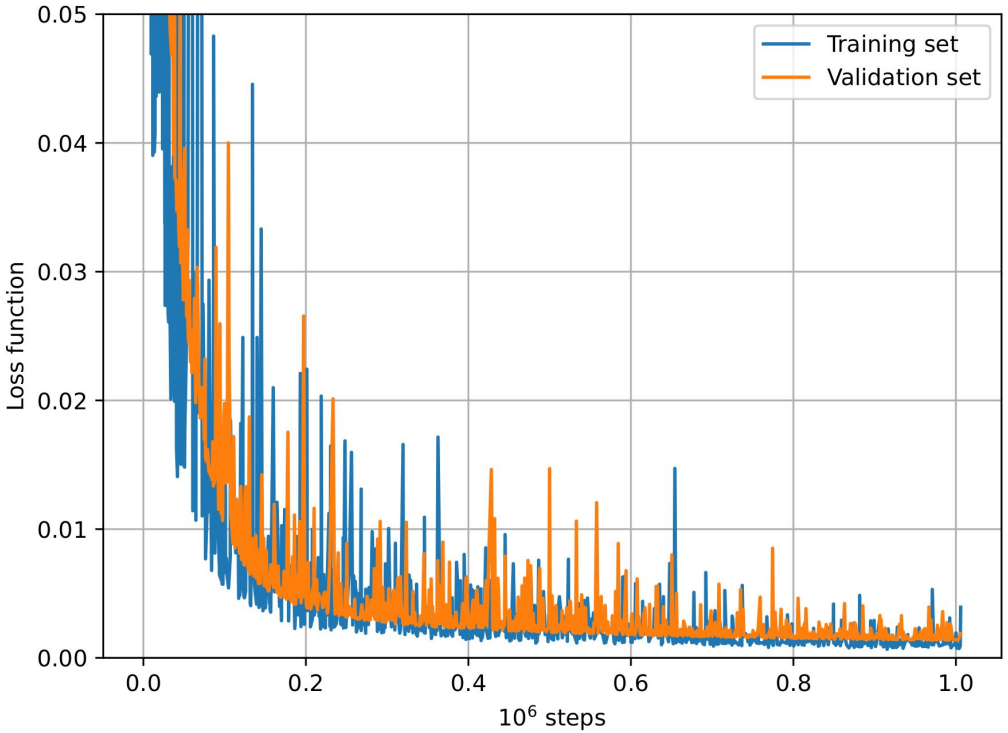
Test set: 10 %



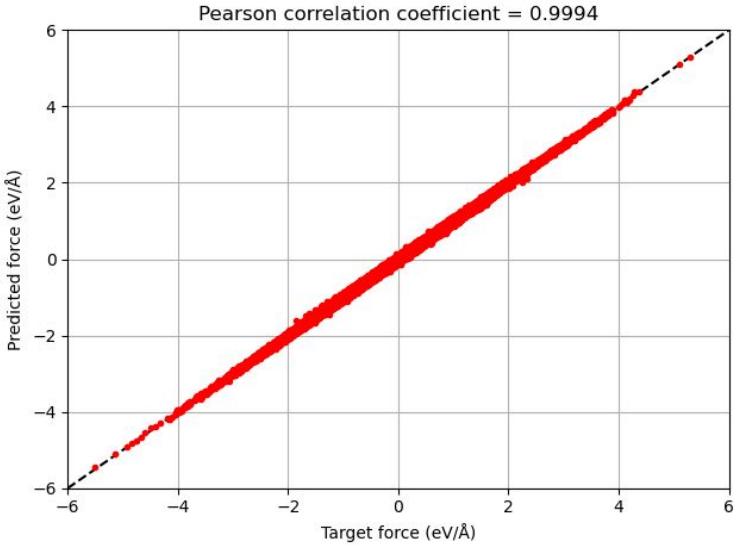
Training

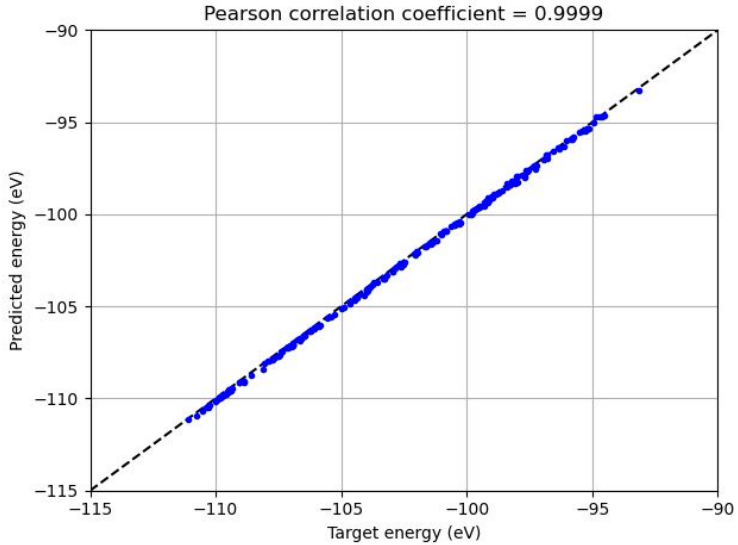
$$L = c_1 MSE_{\text{Force}} + c_2 MSE_{\text{Energy}} + c_3 MSE_{\text{Stress}}$$

$$c_1 = 0.90 \qquad c_2 = 0.01 \qquad c_3 = 0.09$$

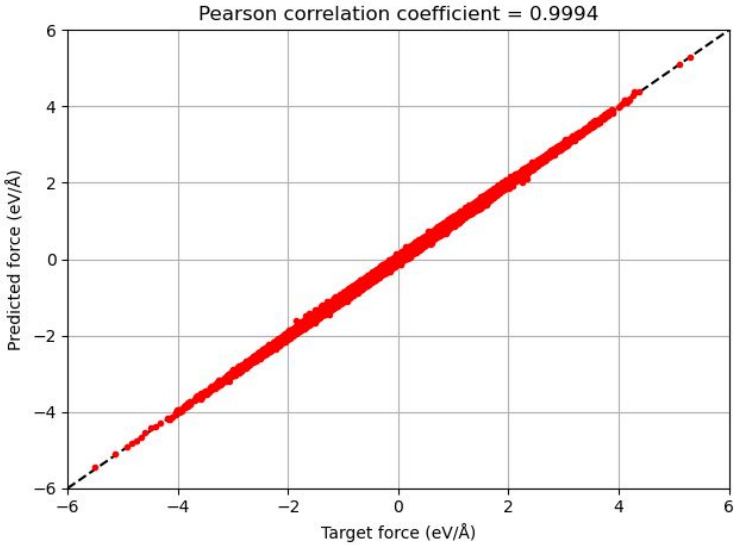


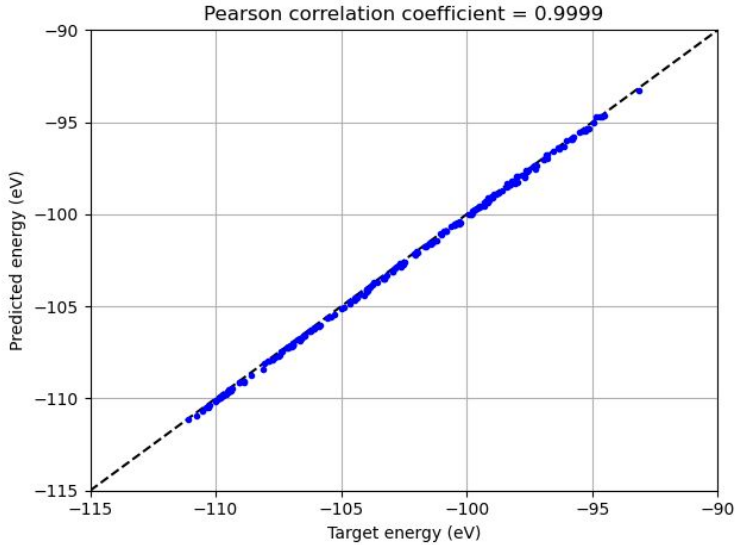
Accuracy of ML potential





Accuracy of ML potential





Metric	Train	Validation	Test
Energy MAE (eV)	0.116 ± 0.069	0.122 ± 0.066	0.121 ± 0.066
Energy ME (eV)	0.396	0.304	0.314
Force MAE (scalar) (eV/Å)	0.019 ± 0.016	0.019 ± 0.016	0.019 ± 0.016
Force ME (scalar) (eV/Å)	0.458	0.296	0.259
Force MAE (vector) (eV/Å)	0.038 ± 0.020	0.038 ± 0.020	0.038 ± 0.020
Force ME (vector) (eV/Å)	0.476	0.366	0.346

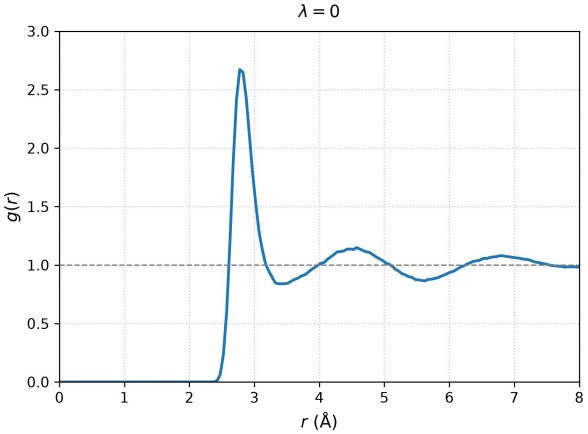
Are these errors small enough?

Force error analysis

$$\mathbf{F}_i = \mathbf{F}_i^{\text{FCM-GAP}} + \lambda \hat{\mathbf{R}}_i$$ ⟵ Random unit vector

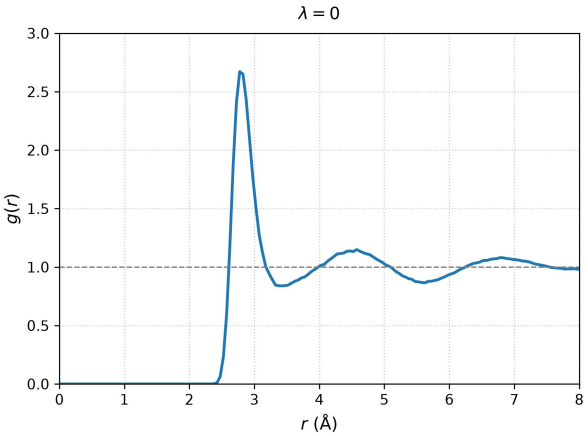
Force error analysis

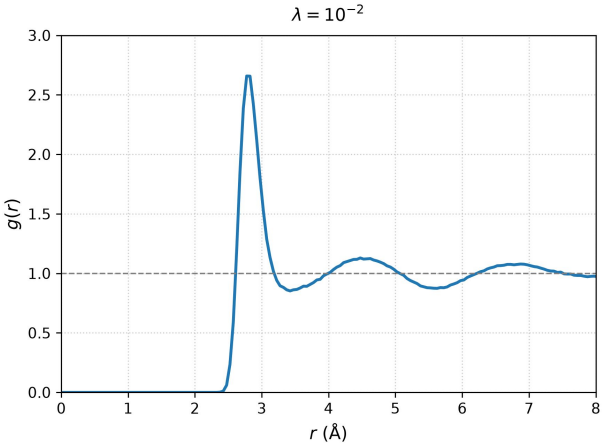
$$\mathbf{F}_i = \mathbf{F}_i^{\text{FCM-GAP}} + \lambda \hat{\mathbf{R}}_i$$ ⬅ Random unit vector



Force error analysis

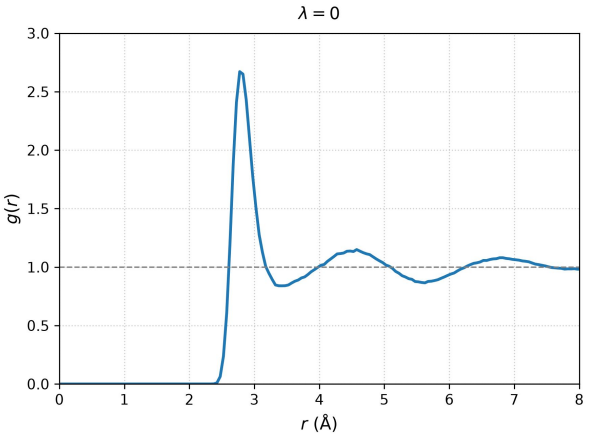
$$\mathbf{F}_i = \mathbf{F}_i^{\text{FCM-GAP}} + \lambda \hat{\mathbf{R}}_i$$ ⟵ Random unit vector

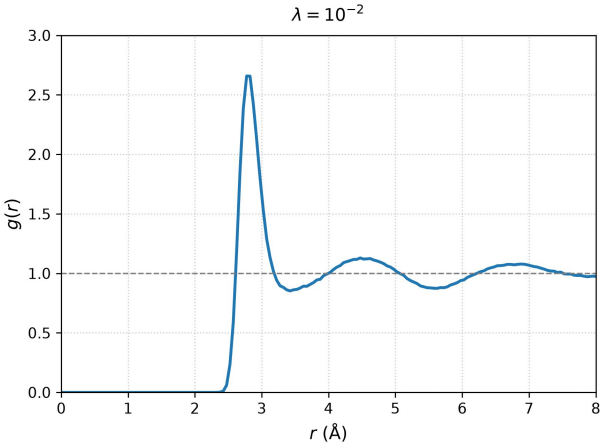


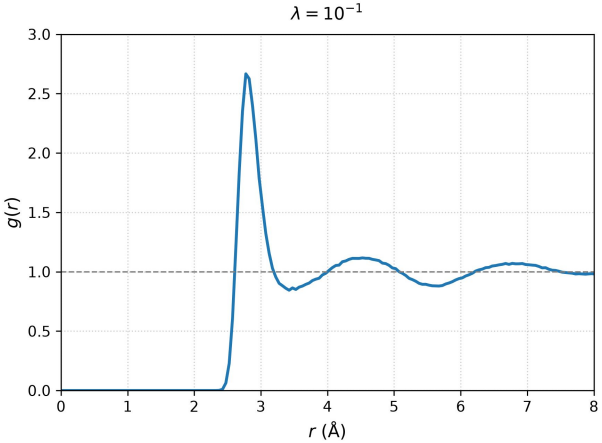


Force error analysis

$$\mathbf{F}_i = \mathbf{F}_i^{\text{FCM-GAP}} + \lambda \hat{\mathbf{R}}_i$$ ⬅ Random unit vector

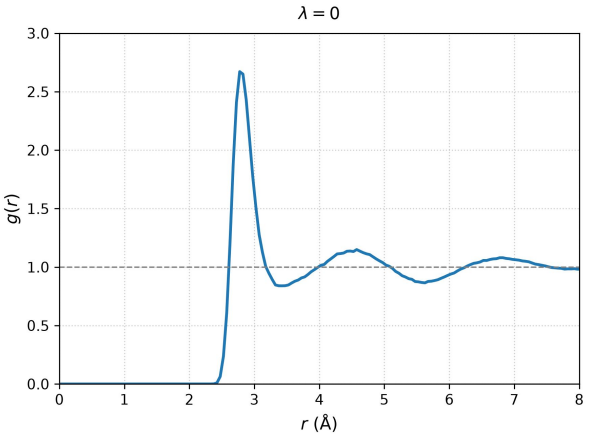


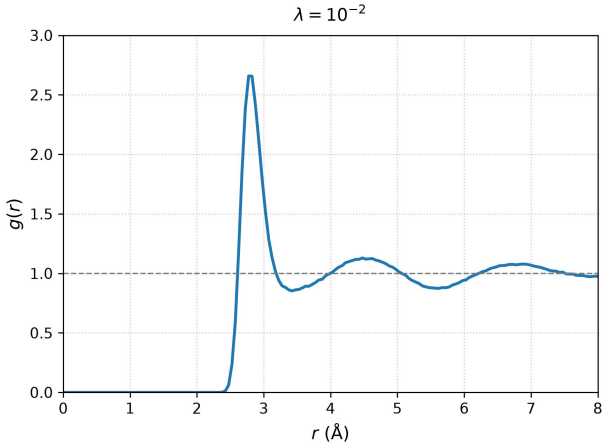


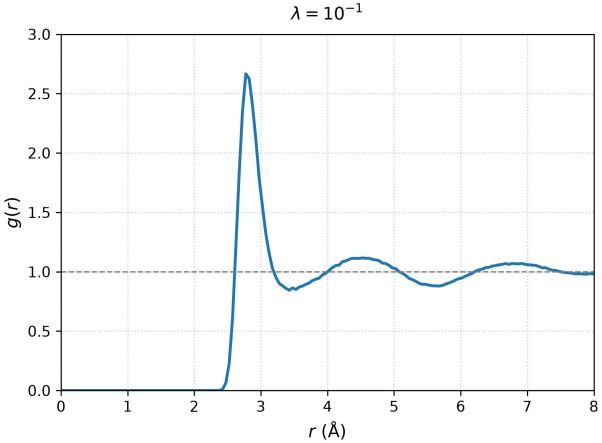


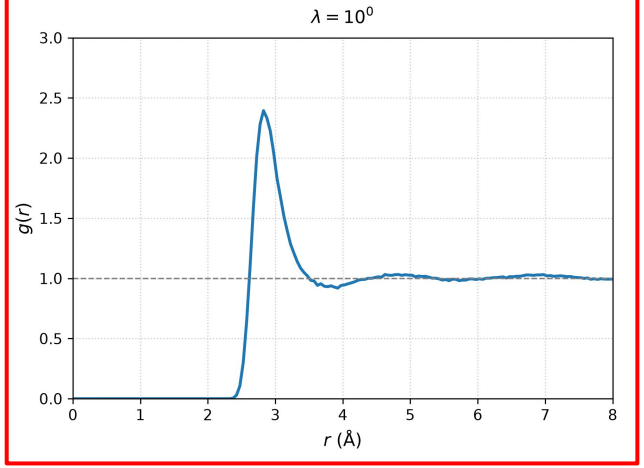
Force error analysis

$$\mathbf{F}_i = \mathbf{F}_i^{\text{FCM-GAP}} + \lambda \hat{\mathbf{R}}_i$$ ⟵ Random unit vector



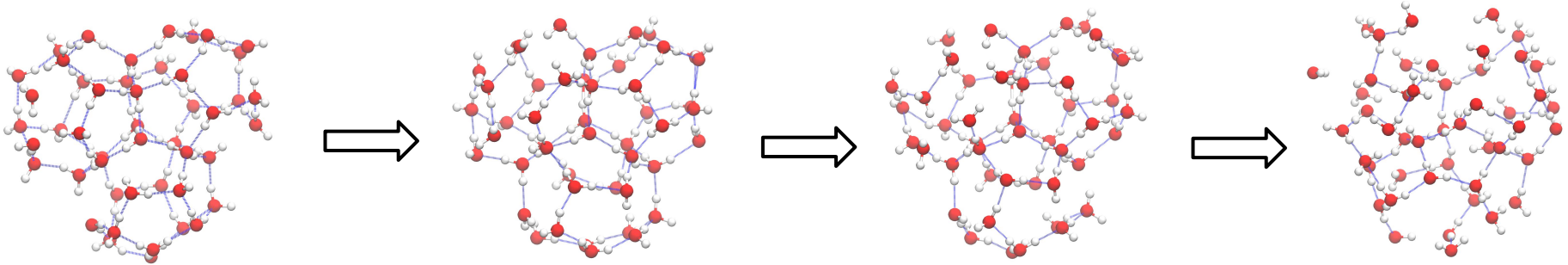






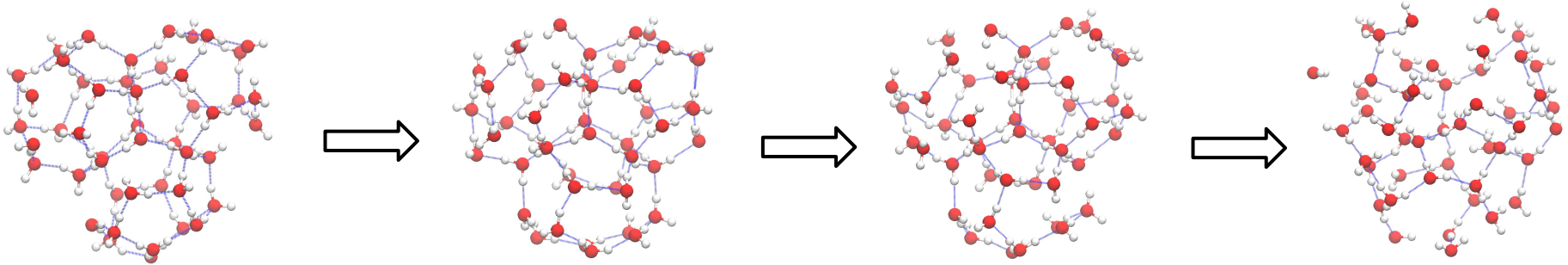
MD simulations of LDL motifs

Life-time of LDL motifs in ambient liquid



MD simulations of LDL motifs

Life-time of LDL motifs in ambient liquid



~ 10 picosecond time scale

Soft equilibration is needed

Conclusions

- The origin of water’s anomalies is a HDL-LDL transition and a liquid-liquid critical point

Conclusions

- The origin of water's anomalies is a HDL-LDL transition and a liquid-liquid critical point

- Ambient water is mainly HDL with LDL fluctuations - "supercritical"

Conclusions

- The origin of water's anomalies is a HDL-LDL transition and a liquid-liquid critical point

- Ambient water is mainly HDL with LDL fluctuations - "supercritical"

- We have trained PAINN models based on the FCM-GAP water model

Conclusions

- The origin of water's anomalies is a HDL-LDL transition and a liquid-liquid critical point
- Ambient water is mainly HDL with LDL fluctuations - "supercritical"
- We have trained PAINN models based on the FCM-GAP water model
- The model allows us to perform accurate simulations of LDL motifs in bulk water

Conclusions

- The origin of water's anomalies is a HDL-LDL transition and a liquid-liquid critical point
- Ambient water is mainly HDL with LDL fluctuations - "supercritical"
- We have trained PAINN models based on the FCM-GAP water model
- The model allows us to perform accurate simulations of LDL motifs in bulk water
- Fused dodecahedra have a life-time of ~10 ps

Conclusions

- The origin of water's anomalies is a HDL-LDL transition and a liquid-liquid critical point
- Ambient water is mainly HDL with LDL fluctuations - "supercritical"
- We have trained PAINN models based on the FCM-GAP water model
- The model allows us to perform accurate simulations of LDL motifs in bulk water
- Fused dodecahedra have a life-time of ~10 ps
- Future work will provide estimates for the abundance of different LDL motifs in ambient water

Acknowledgement

Principal investigator

Prof. Lars G. M. Pettersson

Machine learning expert

Dr. Mathias Ljungberg

Developer of FCM-GAP

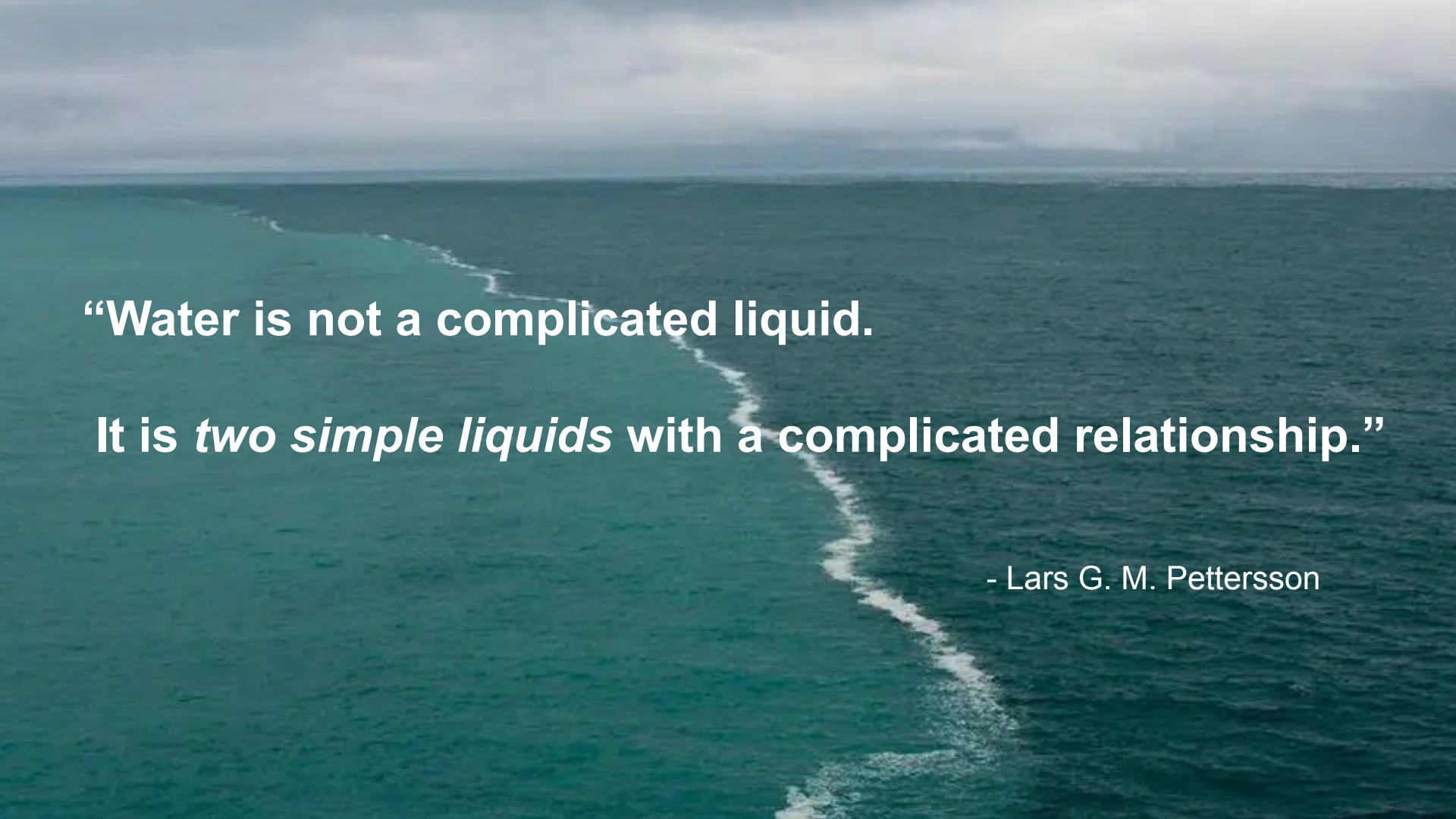
Dr. Jonatan Öström

erc

VETENSKAPSRÅDET
THE SWEDISH RESEARCH COUNCIL

NAISS

UNIVERSITET STOCKHOLMS
Stockholm University


“Water is not a complicated liquid.
It is *two simple liquids* with a complicated relationship.”
- Lars G. M. Pettersson