Entropy as a tool for crystal discovery

Pablo Piaggi (EPFL and USI, Switzerland)

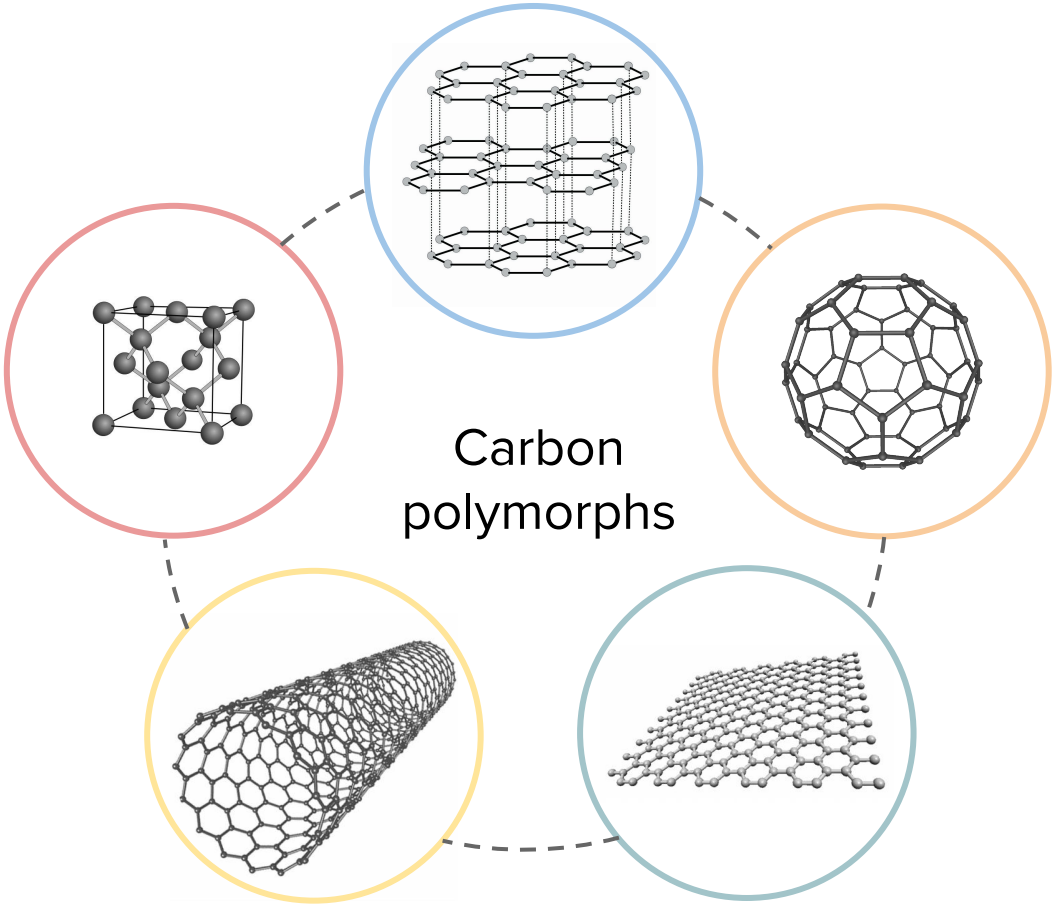
Workshop on Crystal Structure Prediction, ICTP, Trieste January 14-18, 2019

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MARVEL

NATIONAL CENTRE OF COMPETENCE IN RESEARCH

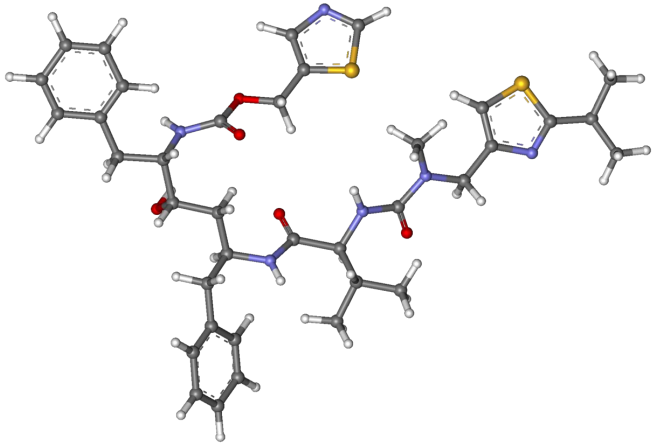
Some substances have more than one crystal structure



Polymorphism is particularly important for the pharmaceutical industry

- Molecules used as drugs exhibit rich polymorphism
- Each polymorph can be patented separately
- Polymorphs have different solubilities/bioavailability

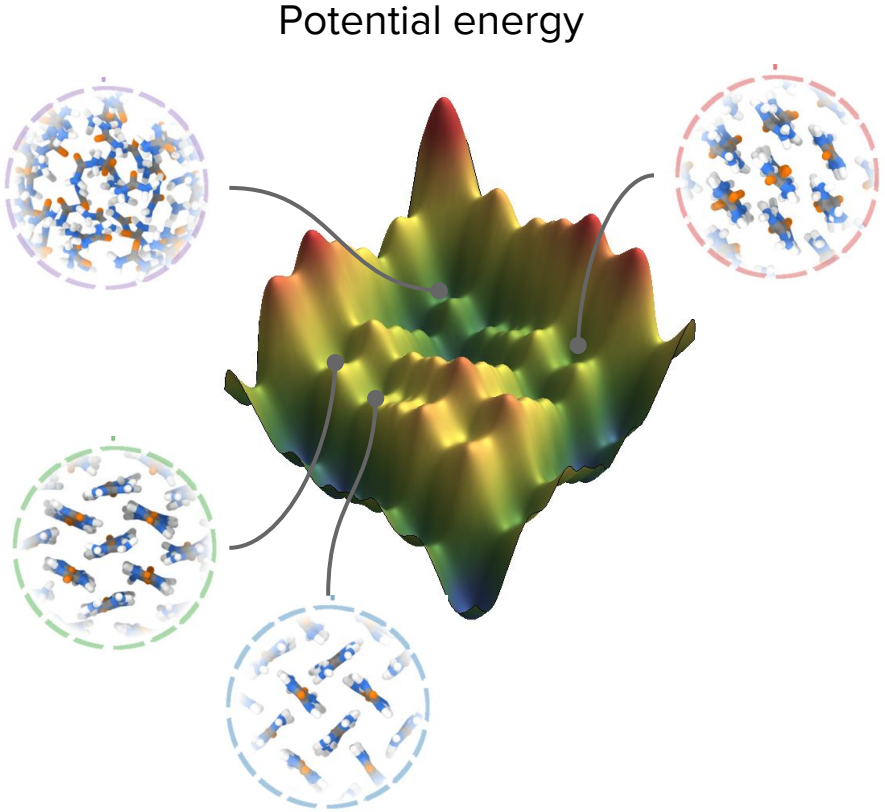
The case of ritonavir



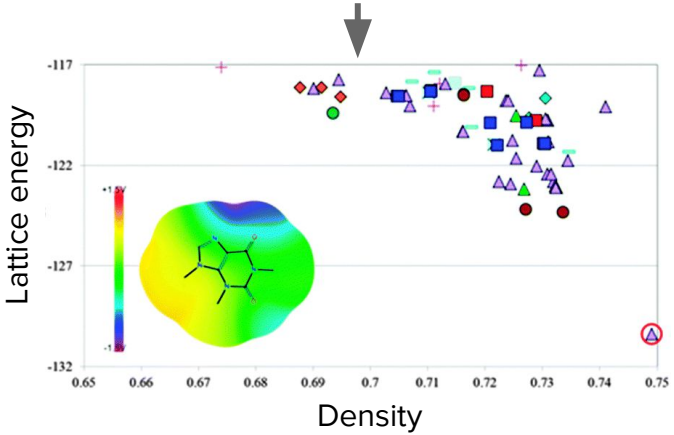
- Medication to treat HIV/AIDS
- During development form I was found
- Once in the market, the more stable form II appeared and its bioavailability is much lower
- The company lost US$ 250 million

J. Bauer et al., Pharmaceutical research 18 (2001)

Search for potential energy minima



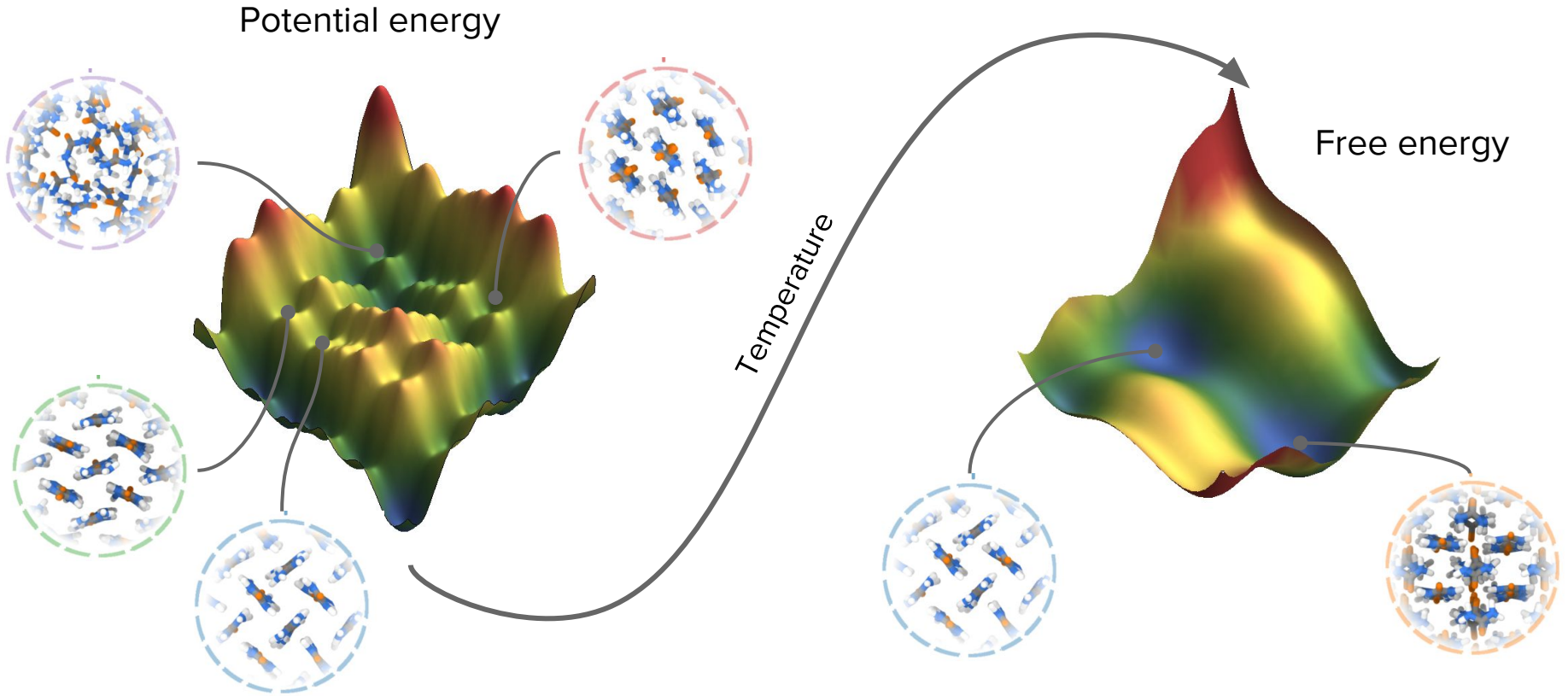
- Current methods search at 0 K
- Random search, evolutionary algorithms
- Many minima are found



Are these minima relevant at finite T?

S.L. Price, Chemical Society Reviews 43 (2014)

The effects of temperature



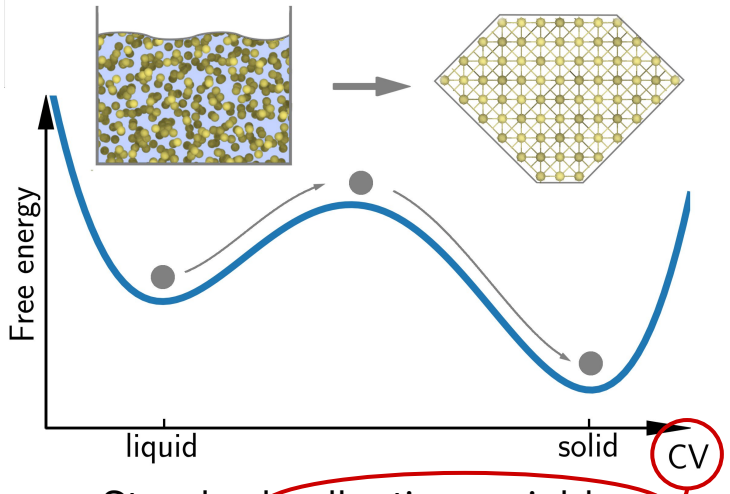
Is it possible to predict the crystal structure of a substance (directly) at finite temperature?

Search for free energy minima using metadynamics

$$V(\mathbf{s},t) = \sum_{k=1}^{n} W e^{-||\mathbf{s}-\mathbf{s}_k||^2} \exp\left[-\frac{1}{\gamma-1}\beta V_{k-1}(\mathbf{s}_k)\right].$$

$$V(\mathbf{s},t) = -\left(1-\frac{1}{\gamma}\right)F(\mathbf{s}) + c(t),$$

A Laio, and M Parrinello, PNAS 99, (2002)
A Barducci, G Bussi, and Parrinello, Physical Review Letters 100, (2008)

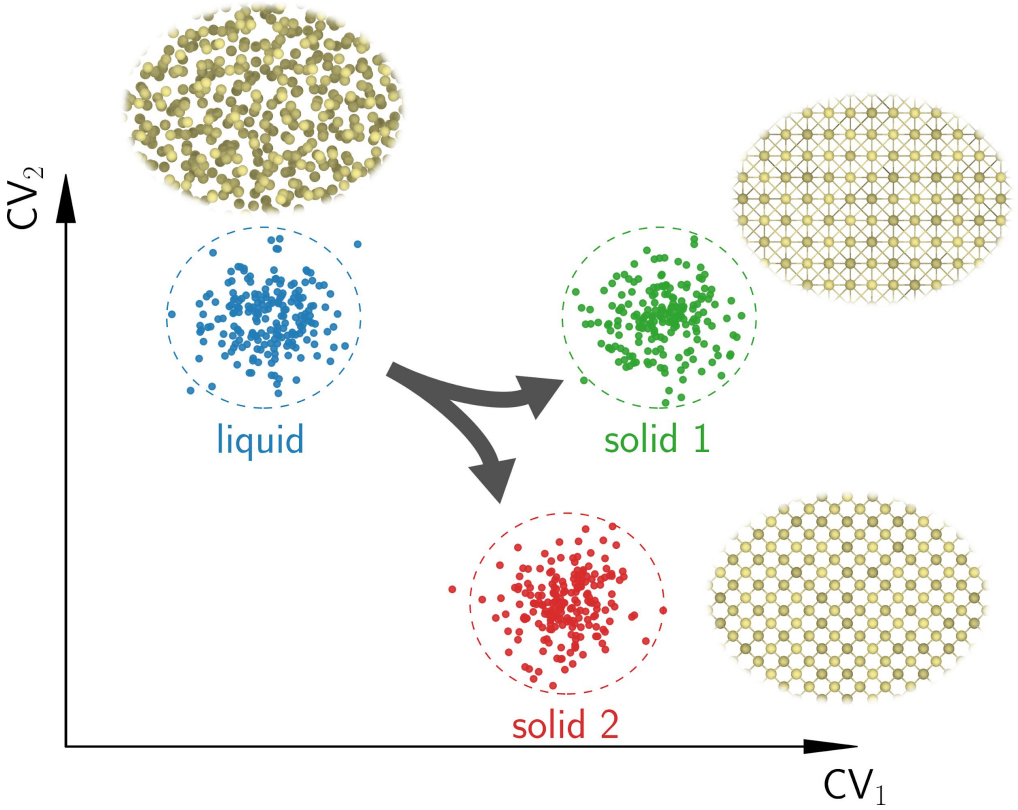


- Standard collective variables assume the final crystal structure from the start
- Steinhardt parameters, structure factor peaks
- Not useful for crystal discovery

H Niu, P Piaggi, M Invernizzi, and M Parrinello, PNAS 115, (2018)

The quest for a structure agnostic CV

Can we find a CV that does not assume the final structure from the start?



Crystallization as a trade off between enthalpy and entropy

In first order phase transitions there is a trade off between enthalpy and entropy

$$\Delta G = \Delta H - T\Delta S$$

ΔG	ΔH	$T\Delta S$
↓	↓	↓
Free energy	Enthalpy	Entropy

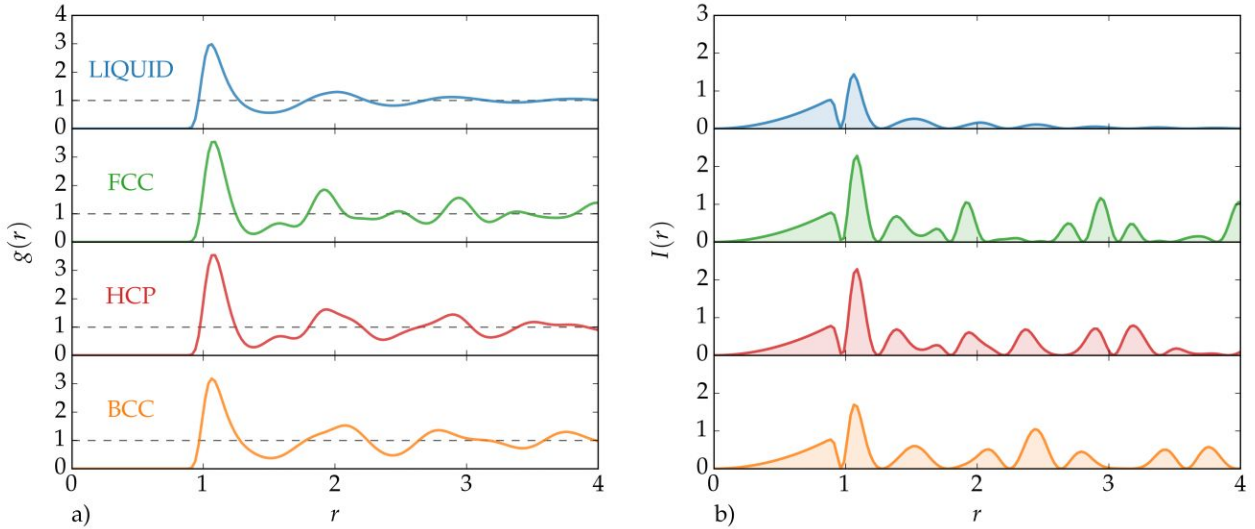
P. M. Piaggi, O. Valsson, and M. Parrinello, Physical Review Letters 119, 015701 (2017)

Approximate expression for the entropy

Entropy expansion in multibody correlation functions

$$S = S_1 + S_2 + S_3 + \cdots$$

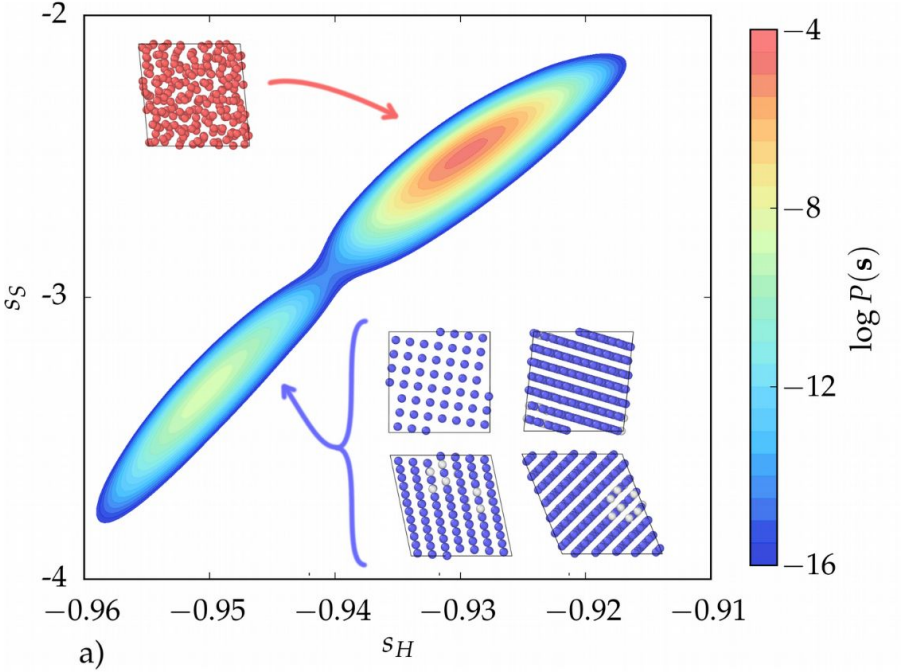
$$S_2 = -2\pi\rho k_B \int_0^\infty [g(r)\ln g(r) - g(r) + 1]\, r^2 dr$$

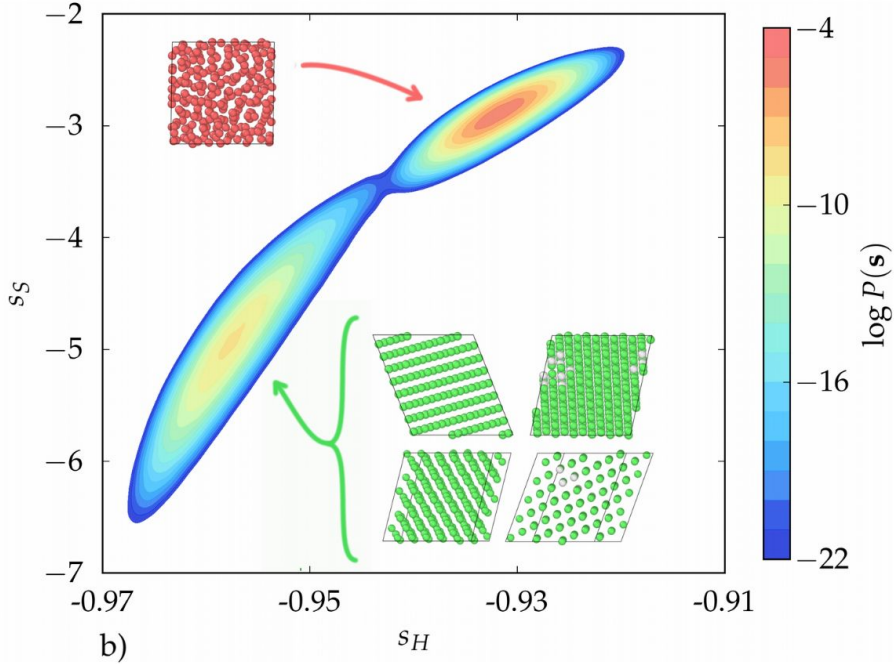


See for instance, A. Baranyai and D. J. Evans, Physical Review A 40, 3817 (1989)

Enhancing enthalpy and entropy fluctuations

Two examples: Na and Al



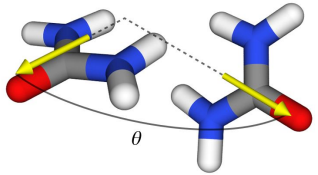


P. M. Piaggi, O. Valsson, and M. Parrinello, Physical Review Letters 119, 015701 (2017)

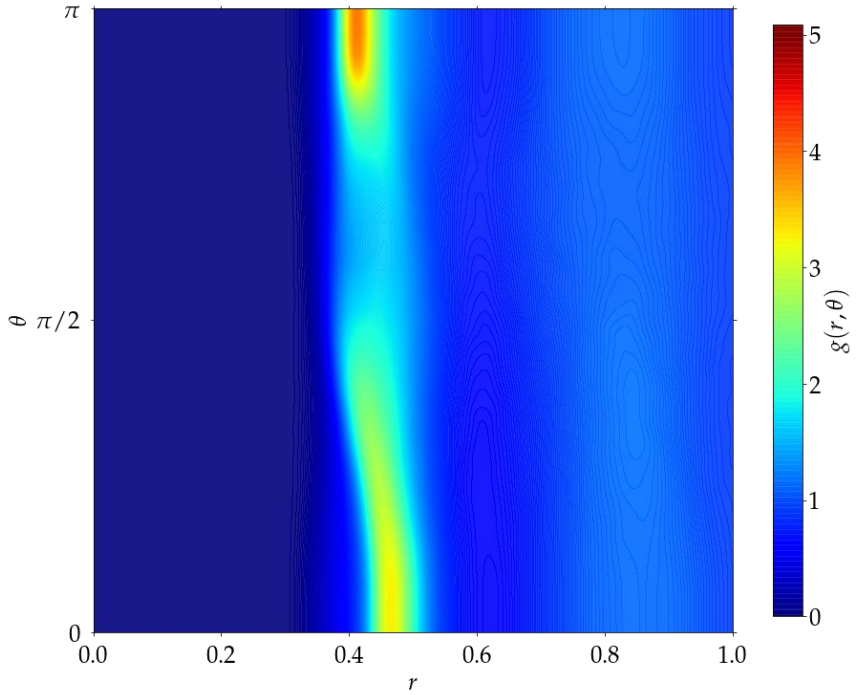
From atoms to molecules ...

g(r,θ) is a natural way to describe molecular crystals

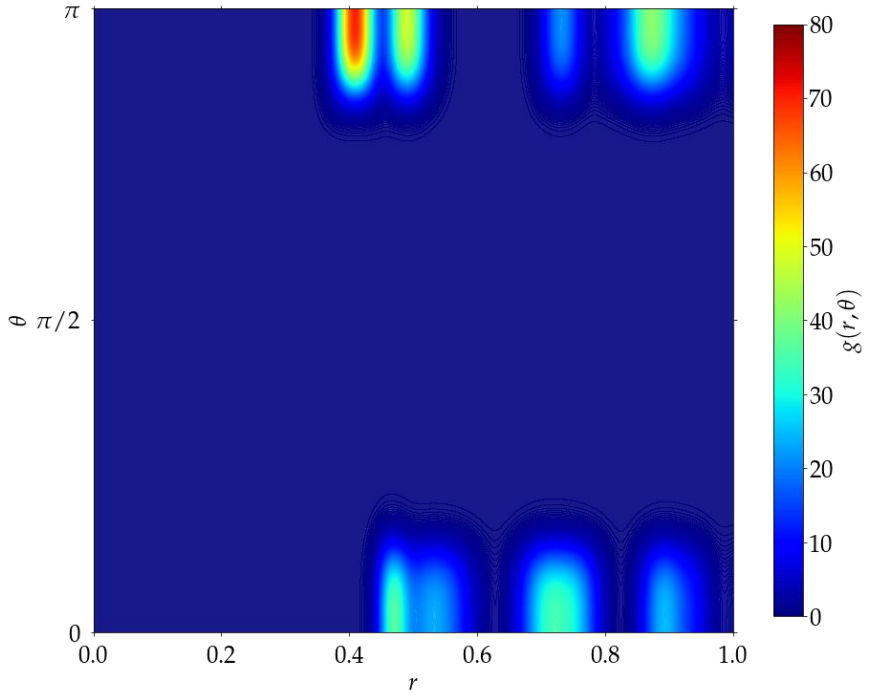
Examples of g(r,θ) - the case of Urea



Liquid



Solid



We define a corresponding pair entropy

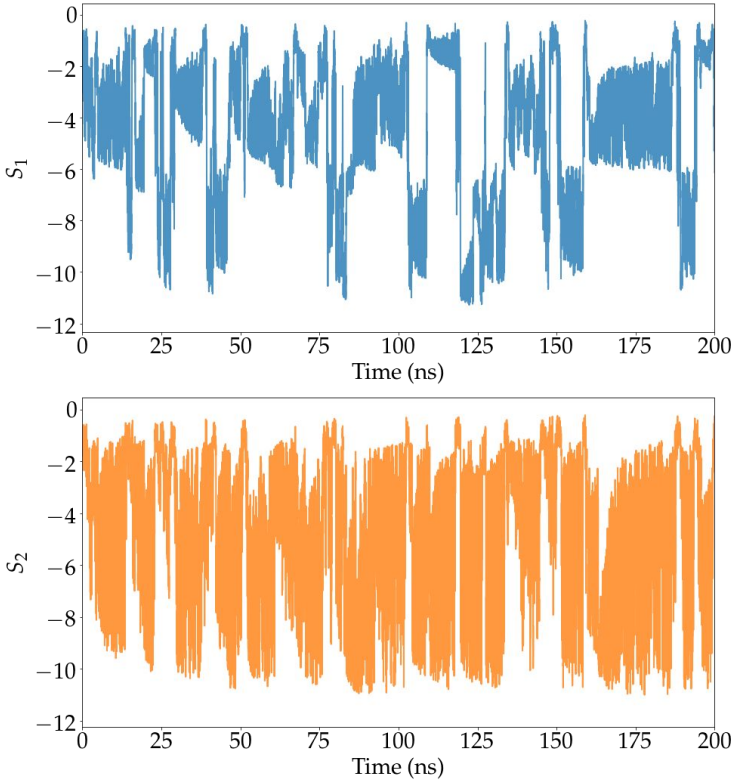
$$S_2 = -2\pi\rho k_B \int_0^\infty [g(r)\ln g(r) - g(r) + 1]\, r^2 dr$$

$$S_2 = -\pi\rho k_B \int_0^\infty \int_0^\pi [g(r,\theta)\ln g(r,\theta) - g(r,\theta) + 1]\, r^2 \sin\theta\, dr d\theta$$

P. M. Piaggi and M. Parrinello, PNAS 115 (41), 10251 (2018)

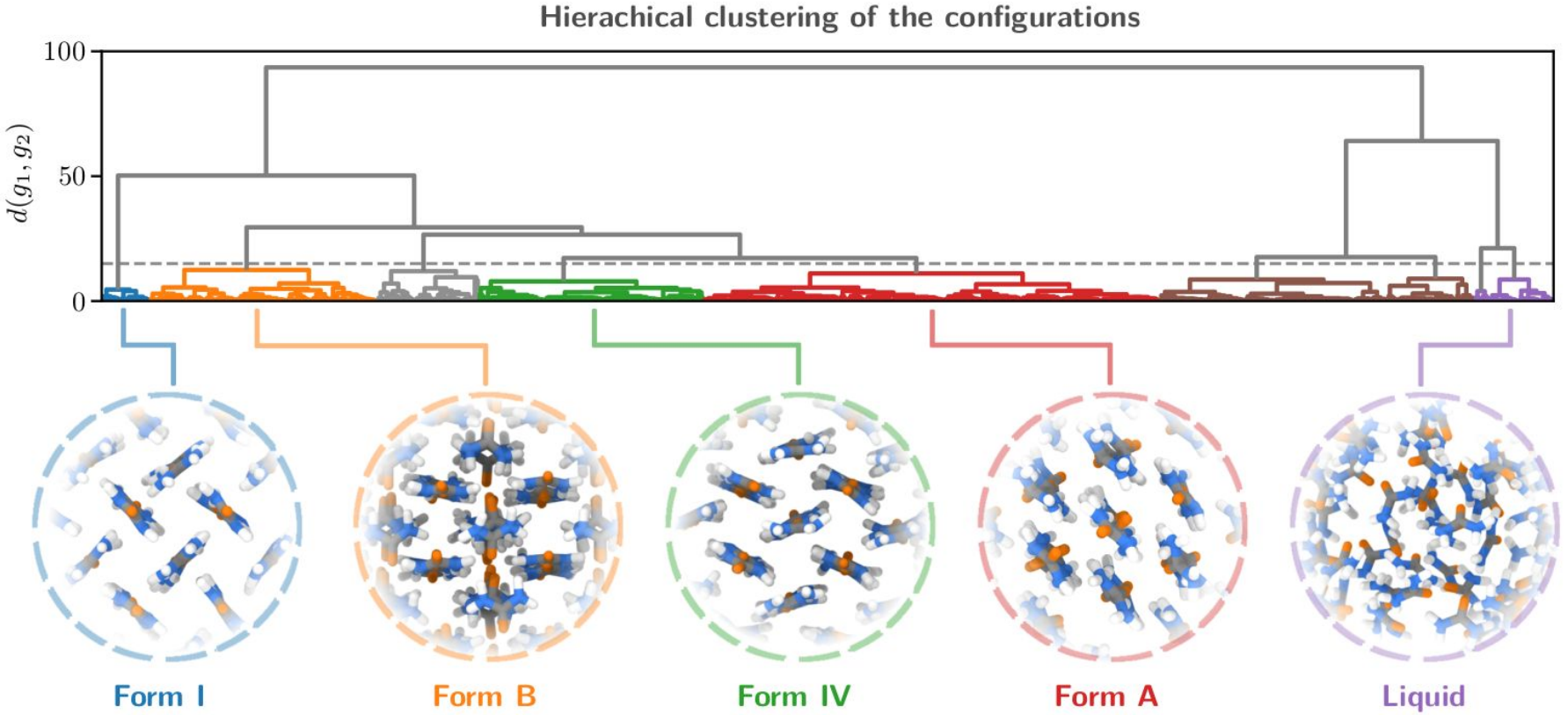
Good exploration - boon or bane?

Urea at finite temperature



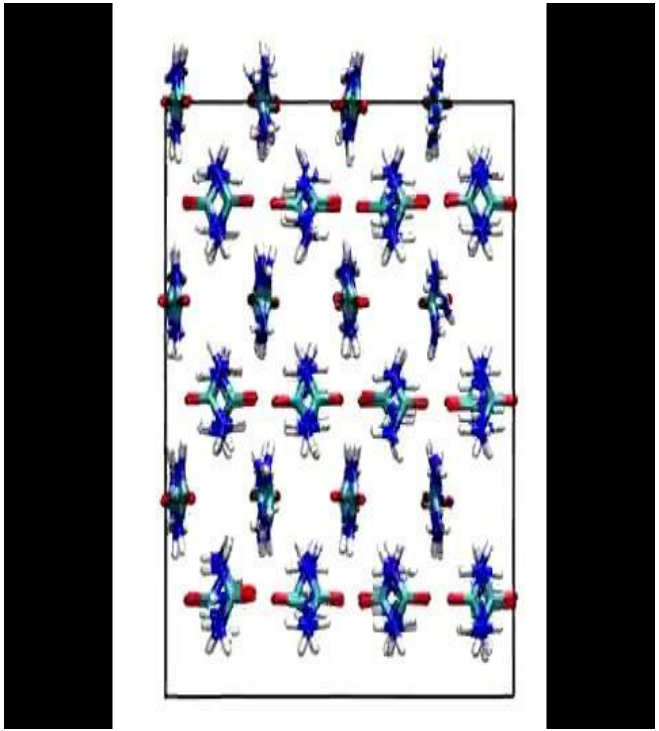
P. M. Piaggi and M. Parrinello, PNAS 115 (41), 10251 (2018)

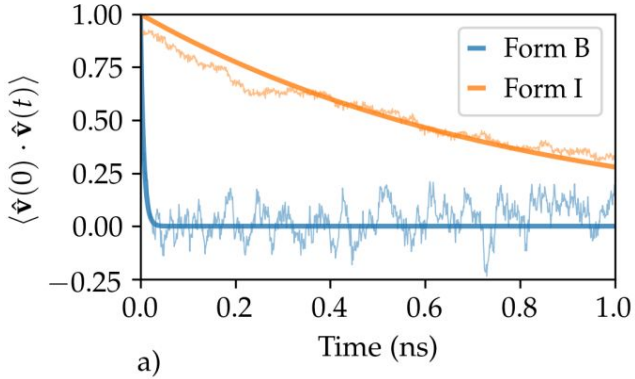
Clustering to understand complex data



P. M. Piaggi and M. Parrinello, PNAS 115 (41), 10251 (2018)

Urea form B is stabilized by entropy





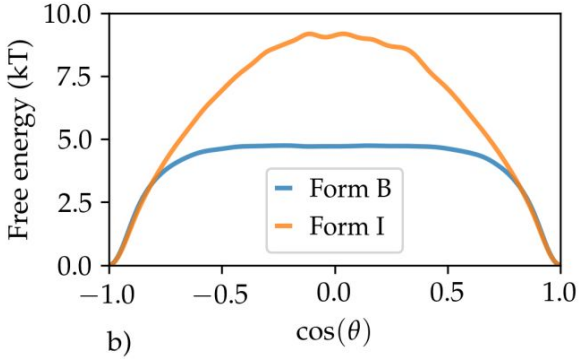
Time autocorrelation function

Free energy

$$G(\theta) = -k_B T \log p(\theta) \sin\theta$$

Entropy

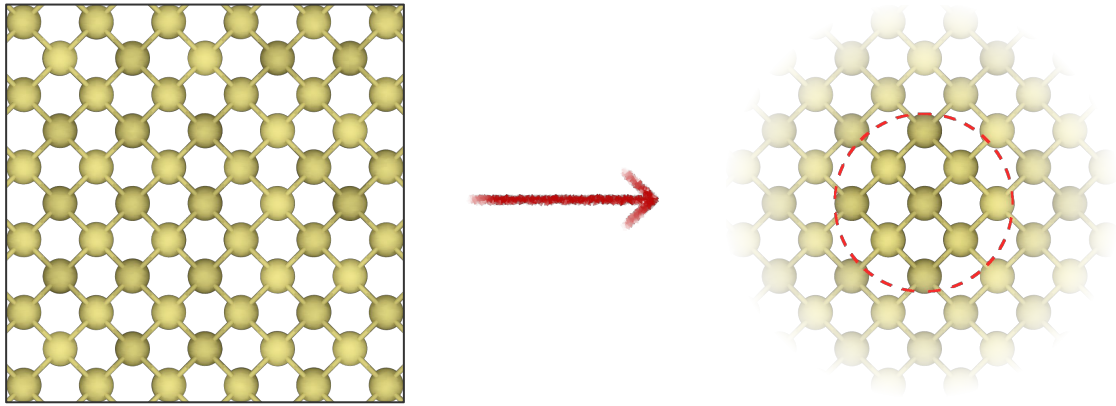
$$k_B T \int p(\theta) \log p(\theta) \sin\theta d\theta$$



P. M. Piaggi and M. Parrinello, PNAS 115 (41), 10251 (2018)

From global to local ...

From global to local



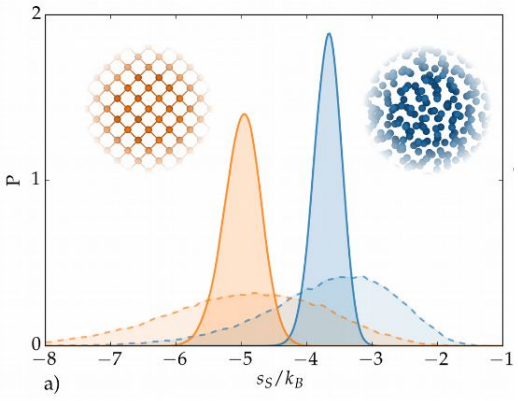
- Projection onto each atom

$$s_S^i = -2\pi\rho k_B \int_0^{r_m} \left[g_m^i(r) \ln g_m^i(r) - g_m^i(r) + 1 \right] r^2 dr,$$

- Average over first neighbors

P. M. Piaggi and M. Parrinello, Journal of Chemical Physics 147, 114112 (2017)

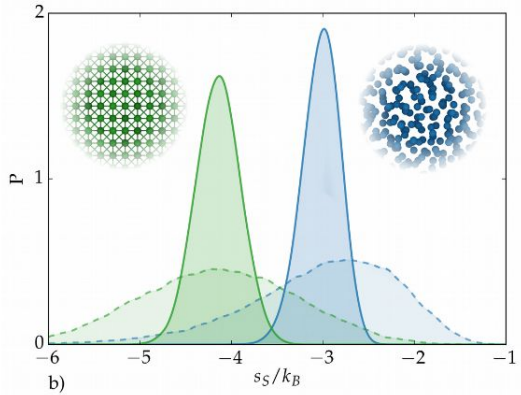
A fingerprint for local crystalline order



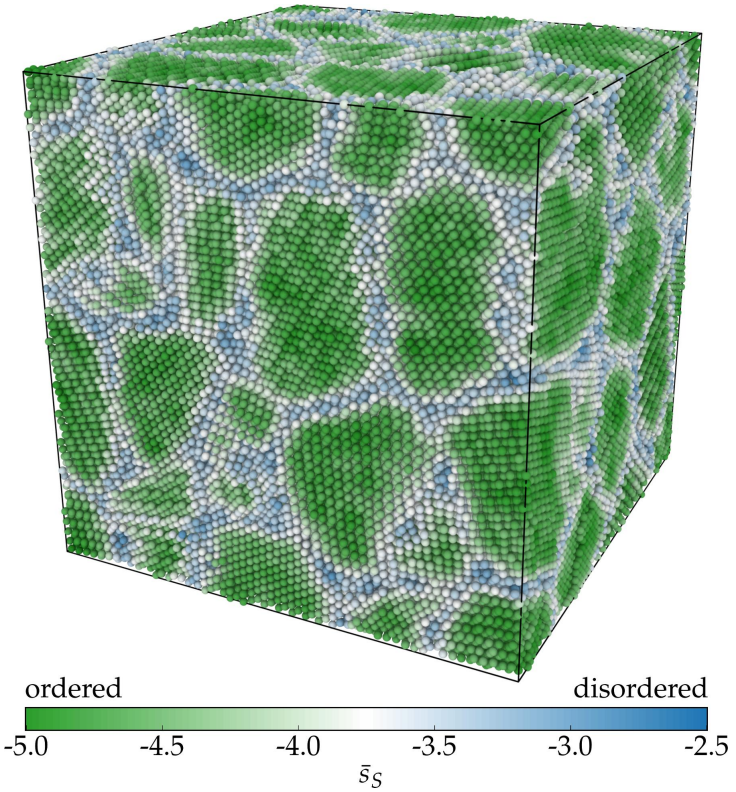
bcc

Fingerprint distributions

fcc

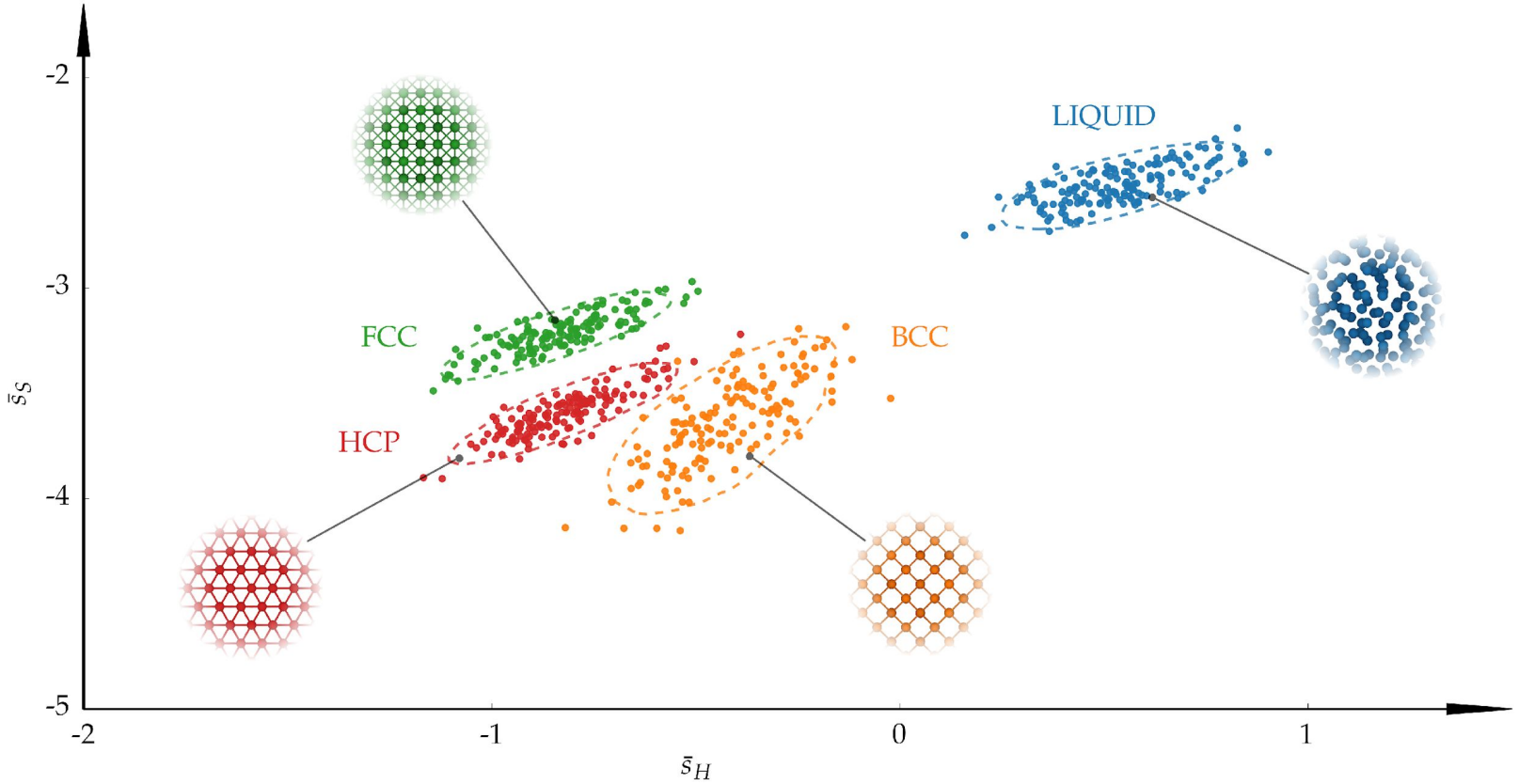


Application



P. M. Piaggi and M. Parrinello, Journal of Chemical Physics 147, 114112 (2017)

Distinguish between polymorphs

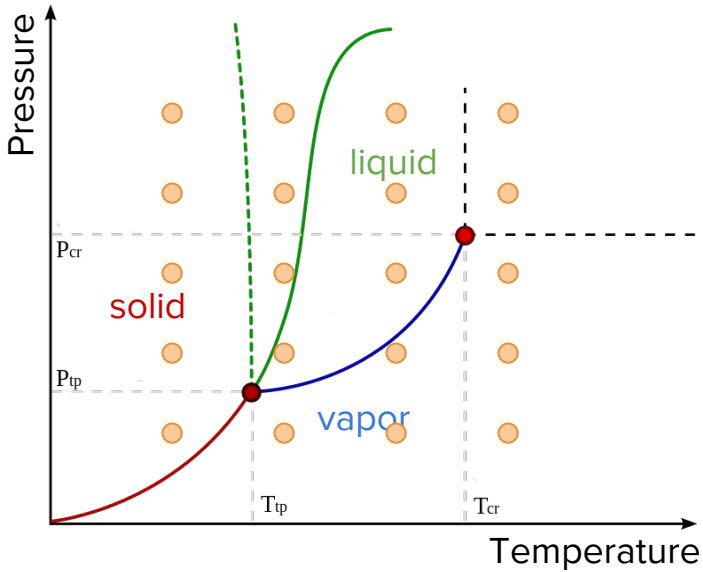


P. M. Piaggi and M. Parrinello, Journal of Chemical Physics 147, 114112 (2017)

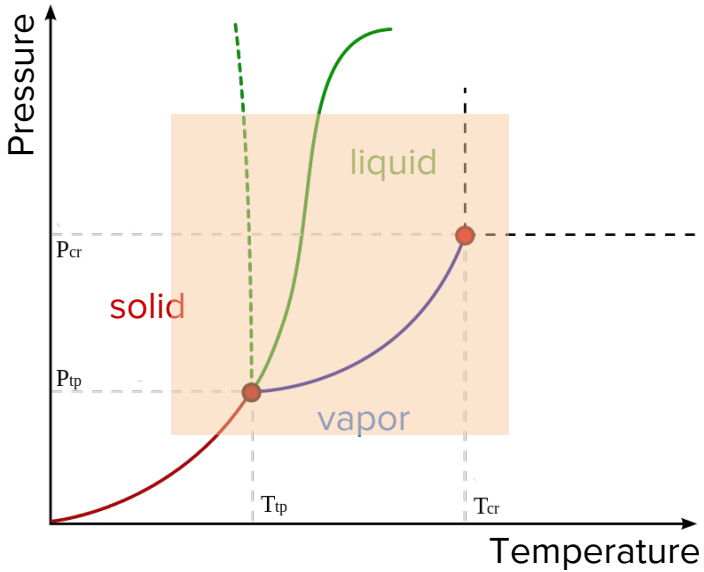
Multithermal-multibaric simulations from a variational principle

The idea

Isothermal-isobaric vs **multithermal-multibaric**



N simulations



1 simulation

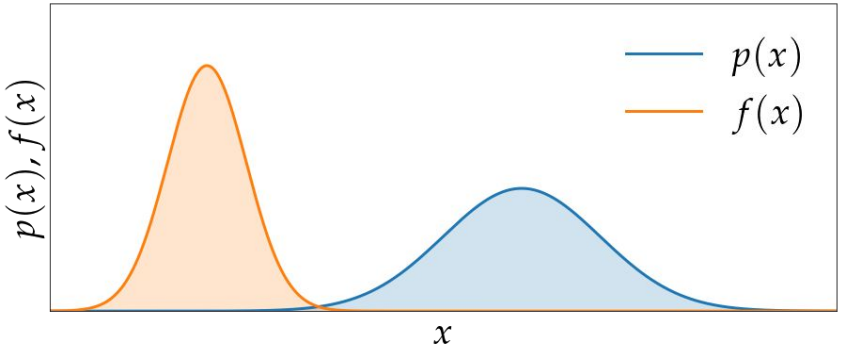
P. M. Piaggi and M. Parrinello, arXiv:1811.08253 (2018)

How?

Importance sampling

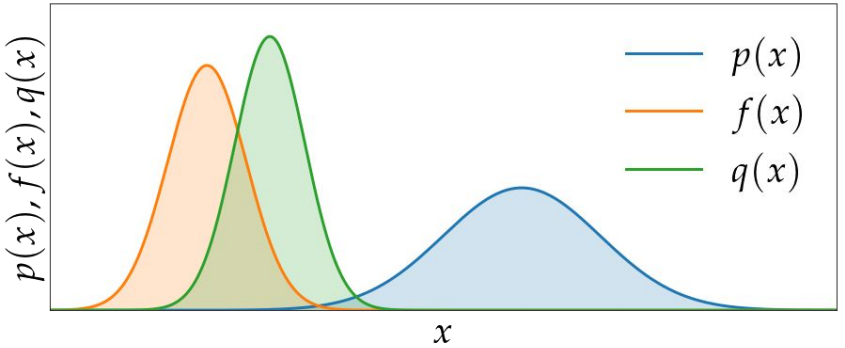
We would like to calculate:

$$\langle f \rangle_p = \int f(x)p(x)dx$$



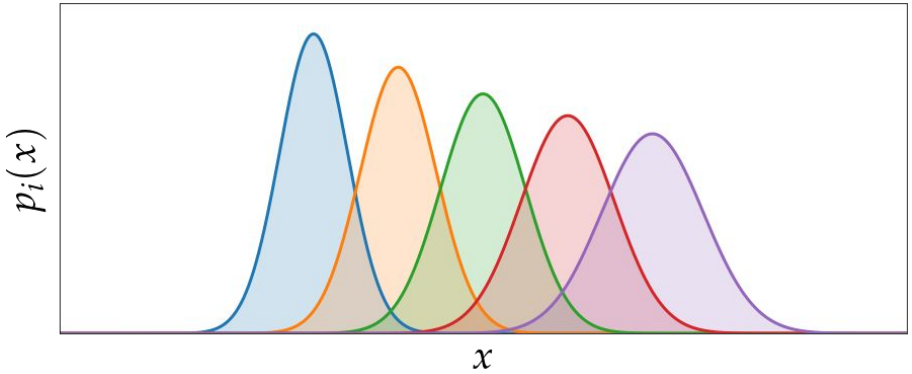
Use a different distribution:

$$\langle f \rangle_p = \int \frac{f(x)p(x)}{q(x)} q(x)dx = \left\langle \frac{fp}{q} \right\rangle_q$$

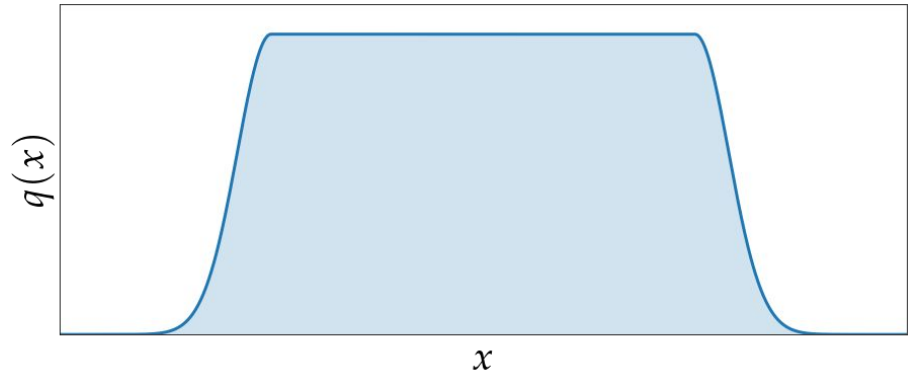


Sample several distributions simultaneously

Find a q(x) useful to sample several distributions $p_i(x)$



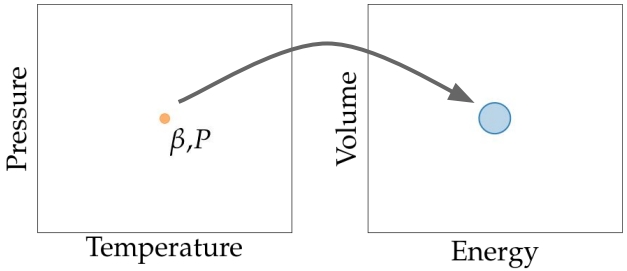
To have a small variance, q(x) must be large everywhere $p_i(x)$ are large



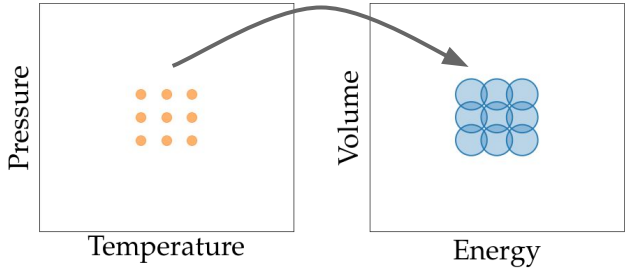
Recipe: q(x) and all the $p_i(x)$ should have **good overlap**

Multithermal-multibaric simulations

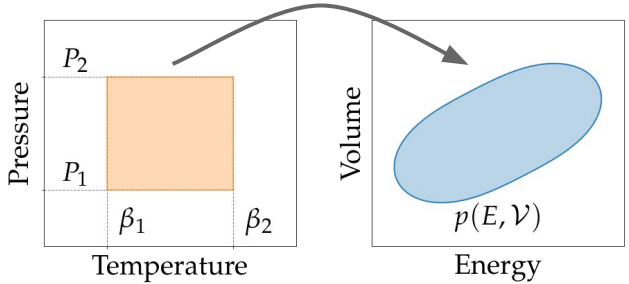
Isothermal-isobaric simulation



Several Isothermal-isobaric simulations



Multithermal-multibaric simulation



Find distribution that encompasses all the isothermal-isobaric distributions in the desired T-P range. But how?

Variationally enhanced sampling

Introduce a bias potential V(**s**) - **s** are the collective variables

Convex functional of the bias potential:

$$\Omega[V] = \frac{1}{\beta} \log \frac{\int d\mathbf{s}\, e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s}\, e^{-\beta F(\mathbf{s})}} + \int d\mathbf{s}\, p(\mathbf{s}) V(\mathbf{s})$$

Made stationary by,

$$V(\mathbf{s}) = -F(\mathbf{s}) - \frac{1}{\beta} \log p(\mathbf{s})$$

Then,

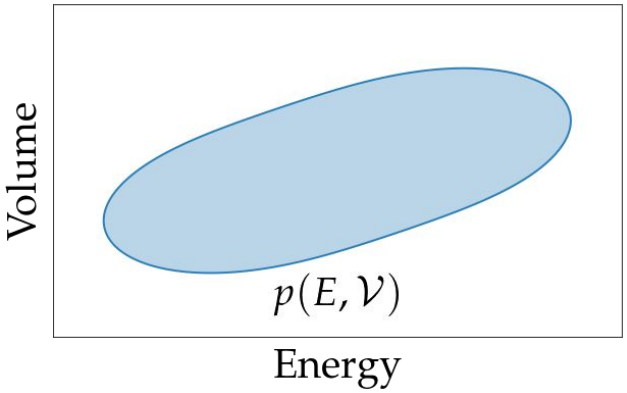
$$p(\mathbf{s}) = \frac{e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s}\, e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}} = P_V(\mathbf{s})$$

Therefore, once that Ω[V] is minimized, the distribution of CVs is p(**s**)

O. Valsson and M. Parrinello, Physical Review Letters 113 (9), 090601 (2014)

Multithermal-multibaric sampling with VES

- Use potential energy E and volume as CVs
- Choose some basis set for the bias
- Use a 2D uniform p(s). Region not known beforehand. Determine it self-consistently.

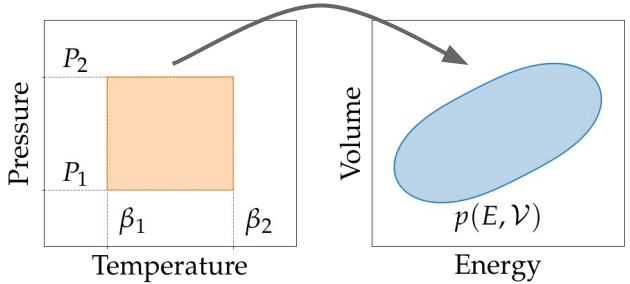


Rigorous link between free energies

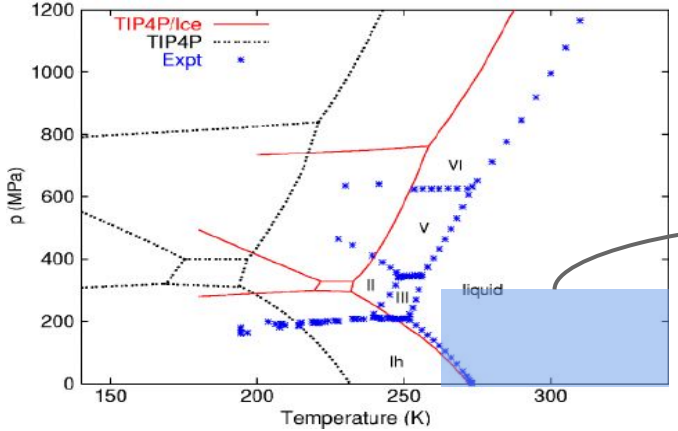
$$\beta' F_{\beta',P'}(E,\mathcal{V}) = \beta F_{\beta,P}(E,\mathcal{V}) + (\beta' - \beta)E + (\beta' P' - \beta P)\mathcal{V} + C'',$$

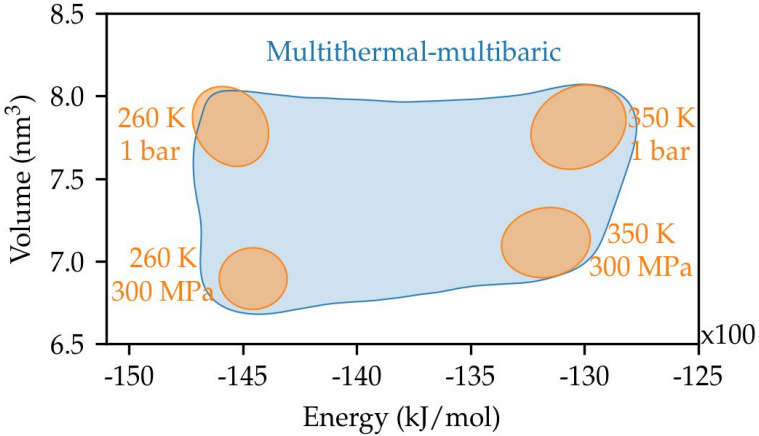
Definition of p(E,V)

$$p(E,\mathcal{V}) = \begin{cases} 1/\Omega_{E,\mathcal{V}} & \text{if there is at least one } \beta', P' \text{ such that } \beta' F_{\beta',P'}(E,\mathcal{V}) < \epsilon \text{ with } \beta_1 > \beta' > \beta_2 \text{ and } P_1 > P' > P_2 \\ 0 & \text{otherwise} \end{cases}$$

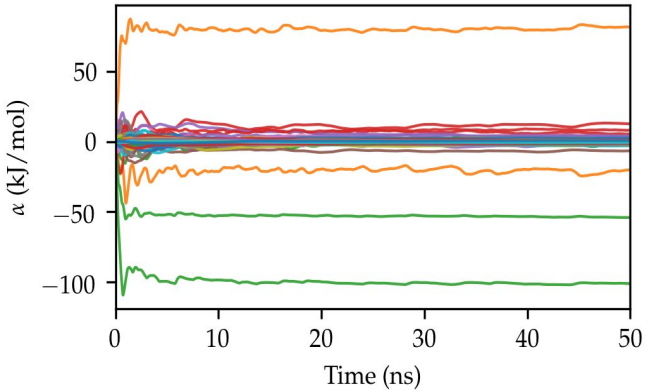


Density anomaly in TIP4P/Ice water





Variational coefficients (~100)

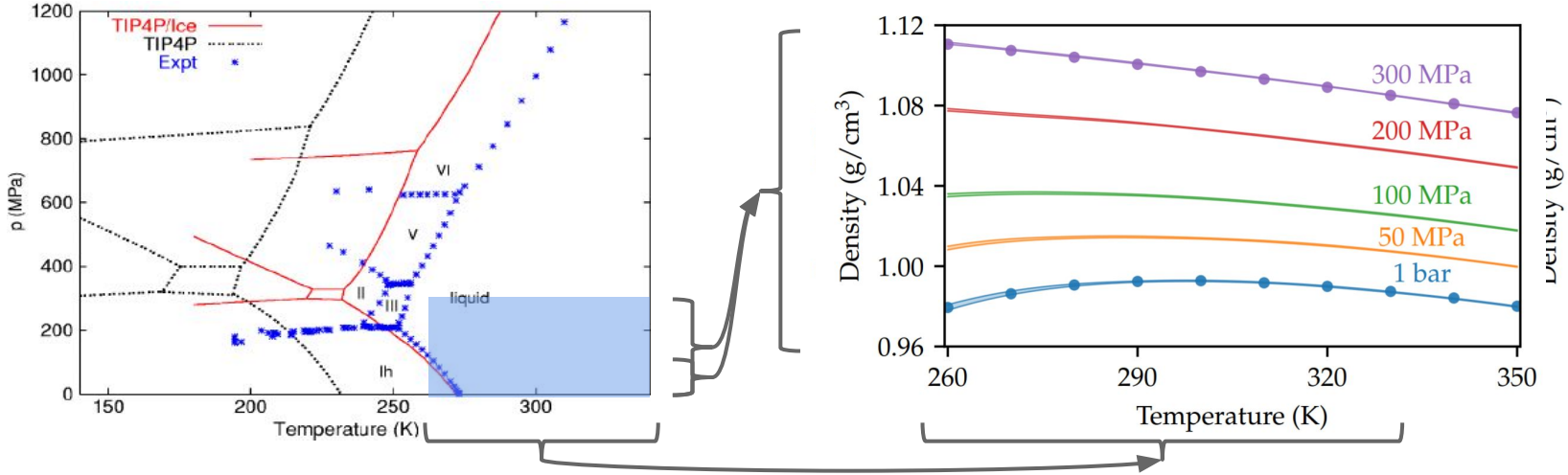


Reweight from biased ensemble at β and P to isothermal-isobaric ensemble at β' and P'

$$\langle O(\mathbf{R},\mathcal{V})\rangle_{\beta',P'} = \frac{\langle O(\mathbf{R},\mathcal{V})w(\mathbf{R},\mathcal{V})\rangle_{\beta,P,V}}{\langle w(\mathbf{R},\mathcal{V})\rangle_{\beta,P,V}}$$

$$w(\mathbf{R},\mathcal{V}) = e^{(\beta-\beta')E(\mathbf{R})+(\beta P-\beta' P')\mathcal{V}}e^{\beta V(E)}$$

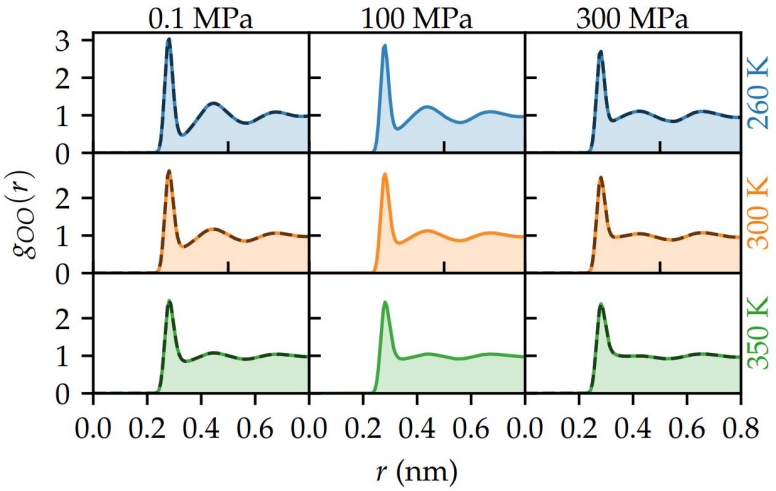
Density anomaly for all T and P



Excellent agreement with individual isothermal-isobaric simulations!

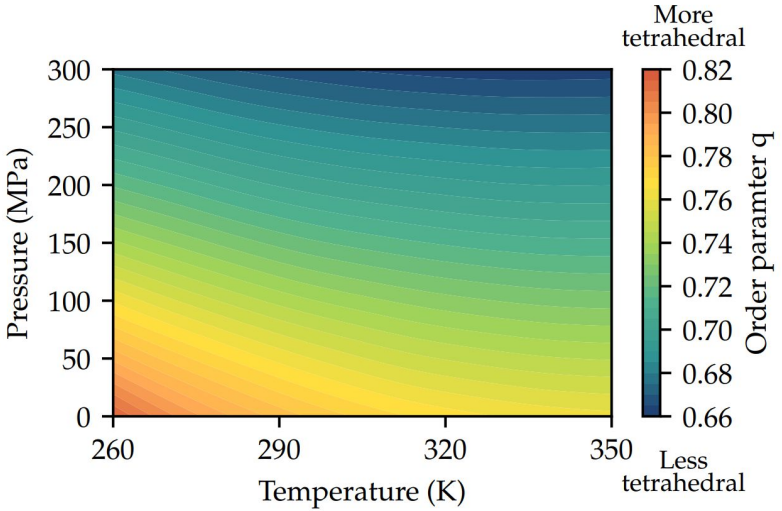
Also other static physical quantities

Radial distribution function



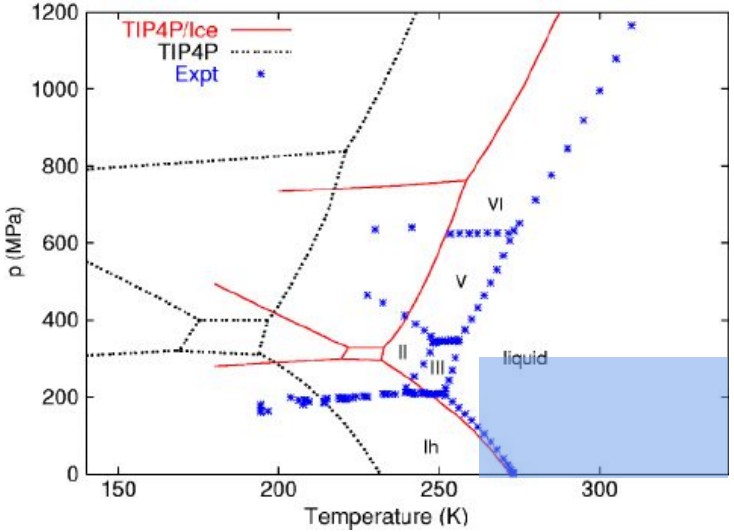
water becomes less structured as the temperature and pressure increase

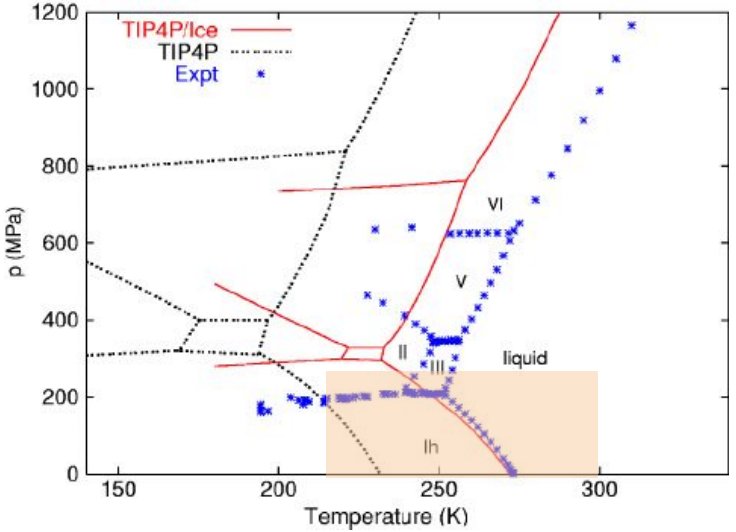
Tetrahedral order parameter



Also specific heat ...

What if there are **phase transitions** in the chosen regions of the phase diagram?

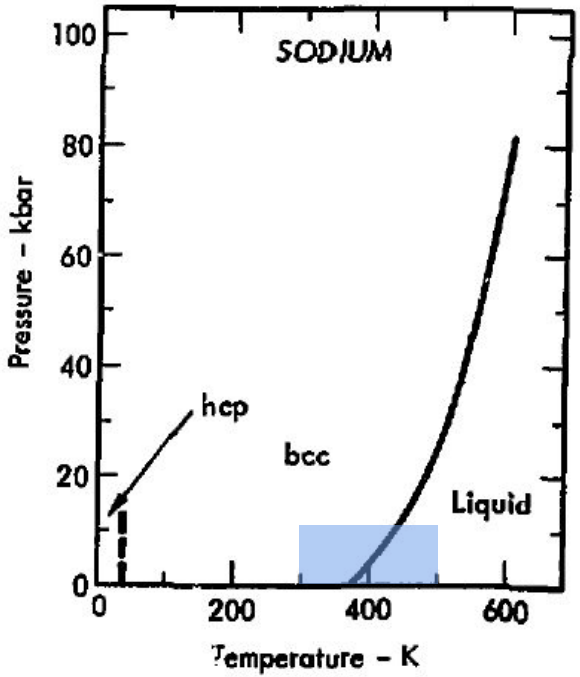


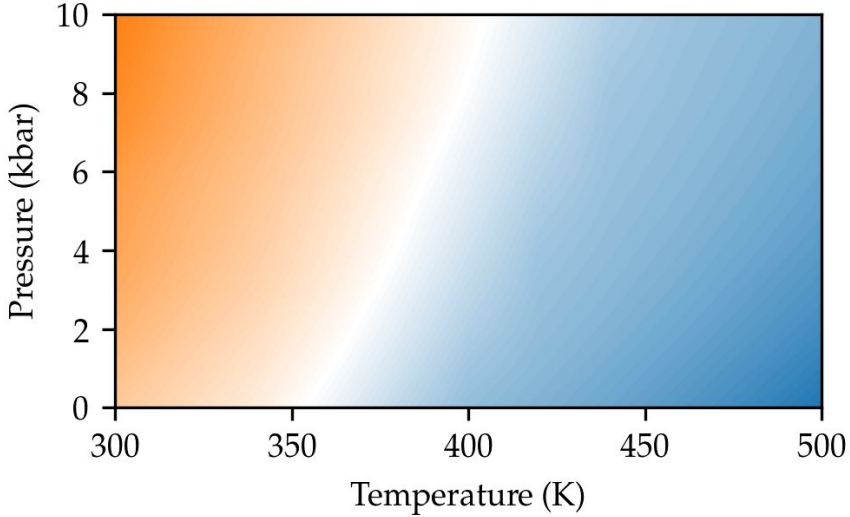


Solid-liquid transition

Combination with metadynamics

Example of **Sodium**





Y. Yang, H. Niu, M. Parrinello, Journal of Physical Chemistry Letters 9 (22), 6426 (2018)

Conclusions

Entropy-inspired CV

- The pair entropy is a collective variable based on the g(r) and it doesn't require any information about the final structure
- It has proven to be effective in predicting crystals structures in many systems from metals, to ionic crystals, to molecular crystals
- Useful to find structures at finite temperature, e.g. high entropy structures
- Pair entropy fingerprint to characterize order-disorder environments

Multithermal-multibaric

- I presented a method for performing multithermal-multibaric simulations
- The temperature and pressure interval is given as input and the relevant region of energy and pressure is determined automatically
- Once that the algorithm has converged, the simulation can be used to calculate all static physical quantities
- Can be used both in Lammps and Gromacs and is fully integrated in Plumed

Thank you for your attention!
Questions?

Acknowledgments

- NCCR MARVEL for funding
- The organizers for inviting me
- Prof. Parrinello
- Collaborators: Omar Valsson, Sergio Perez-Conesa