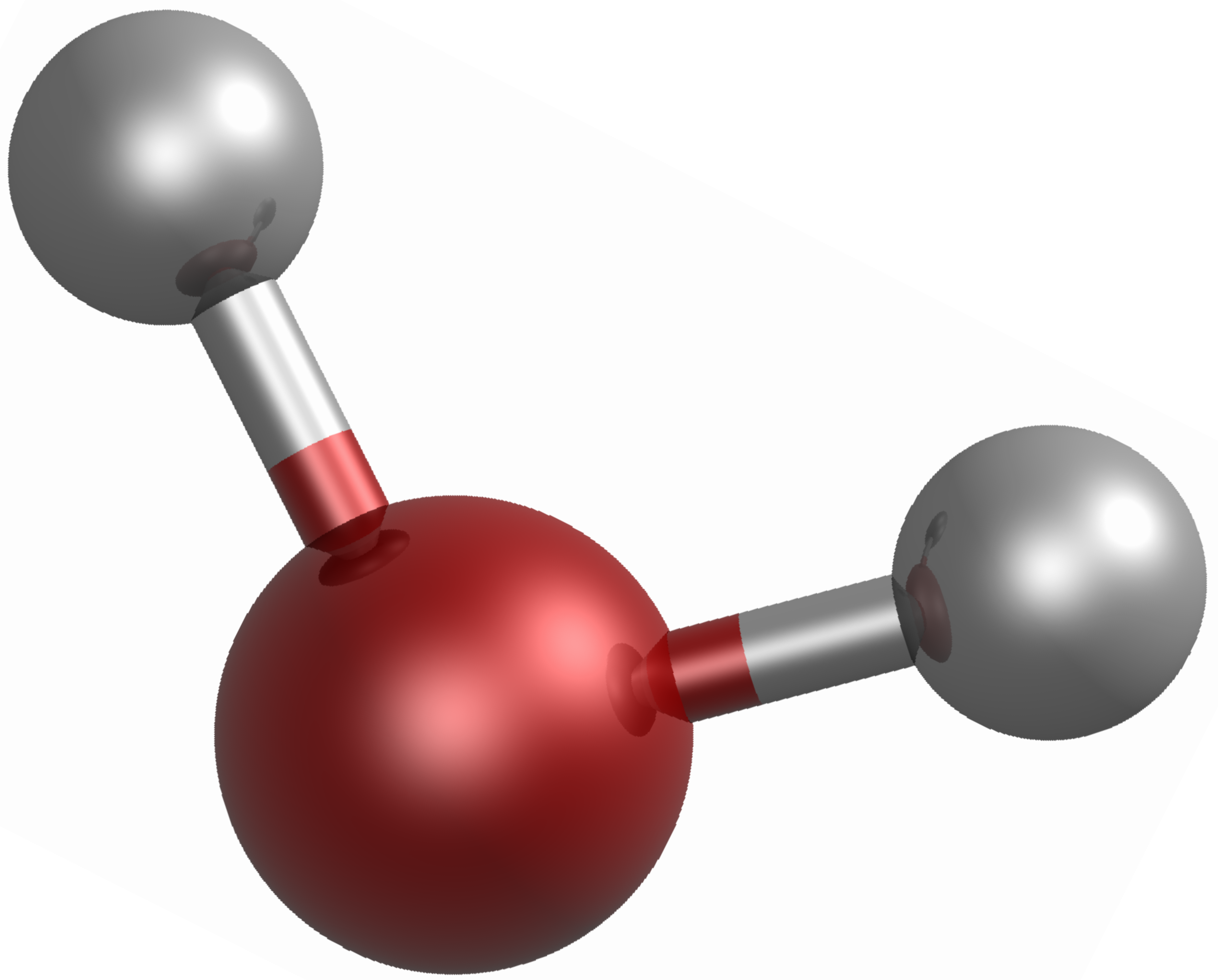
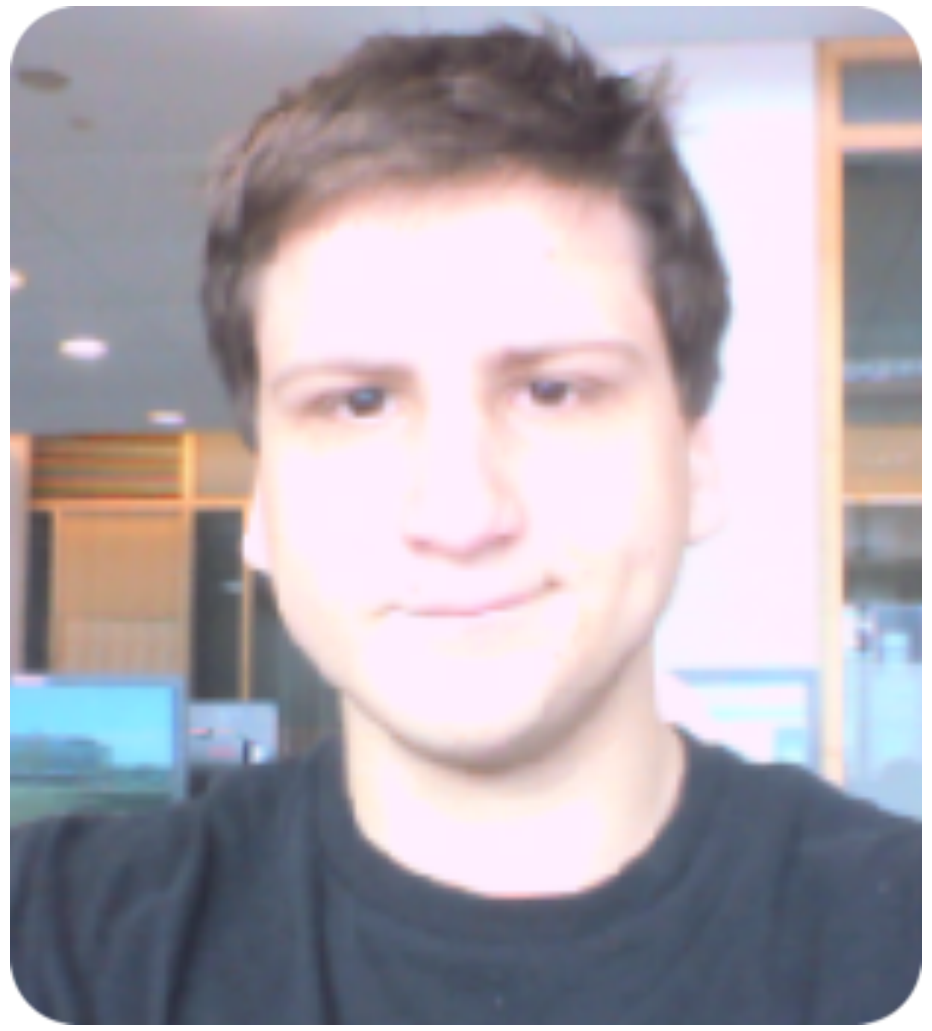
Atomistic Simulations of Liquid Water Next to Metal Surfaces



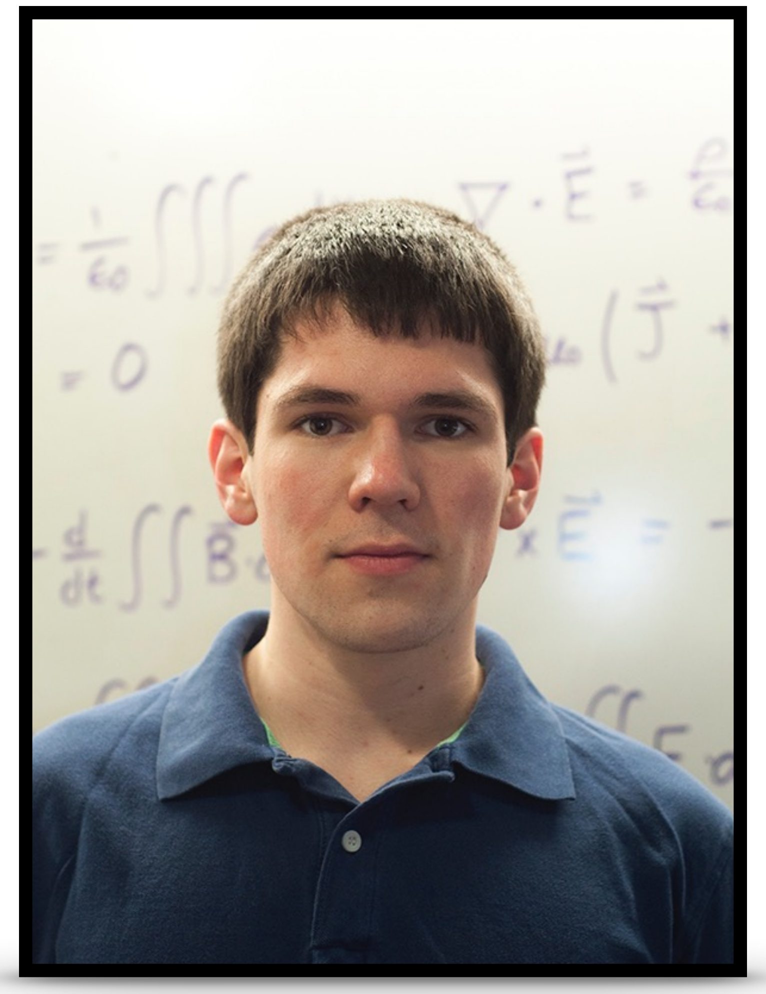
I. Tamblyn

http://faculty.uoit.ca/itamblyn



A. Domurad

A. Maharaj



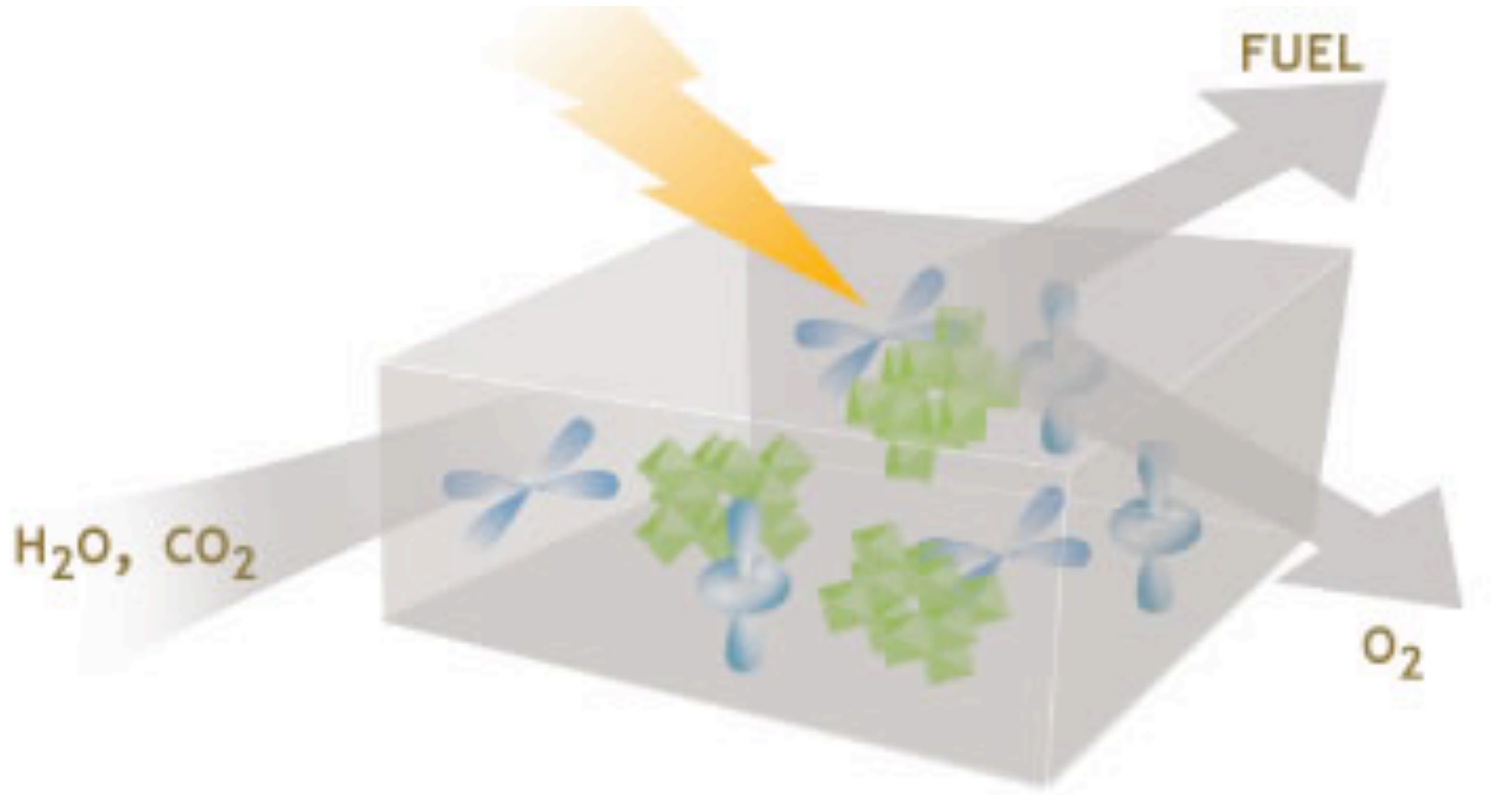
K. Mills

E. Selinger

D. Nemirovsky

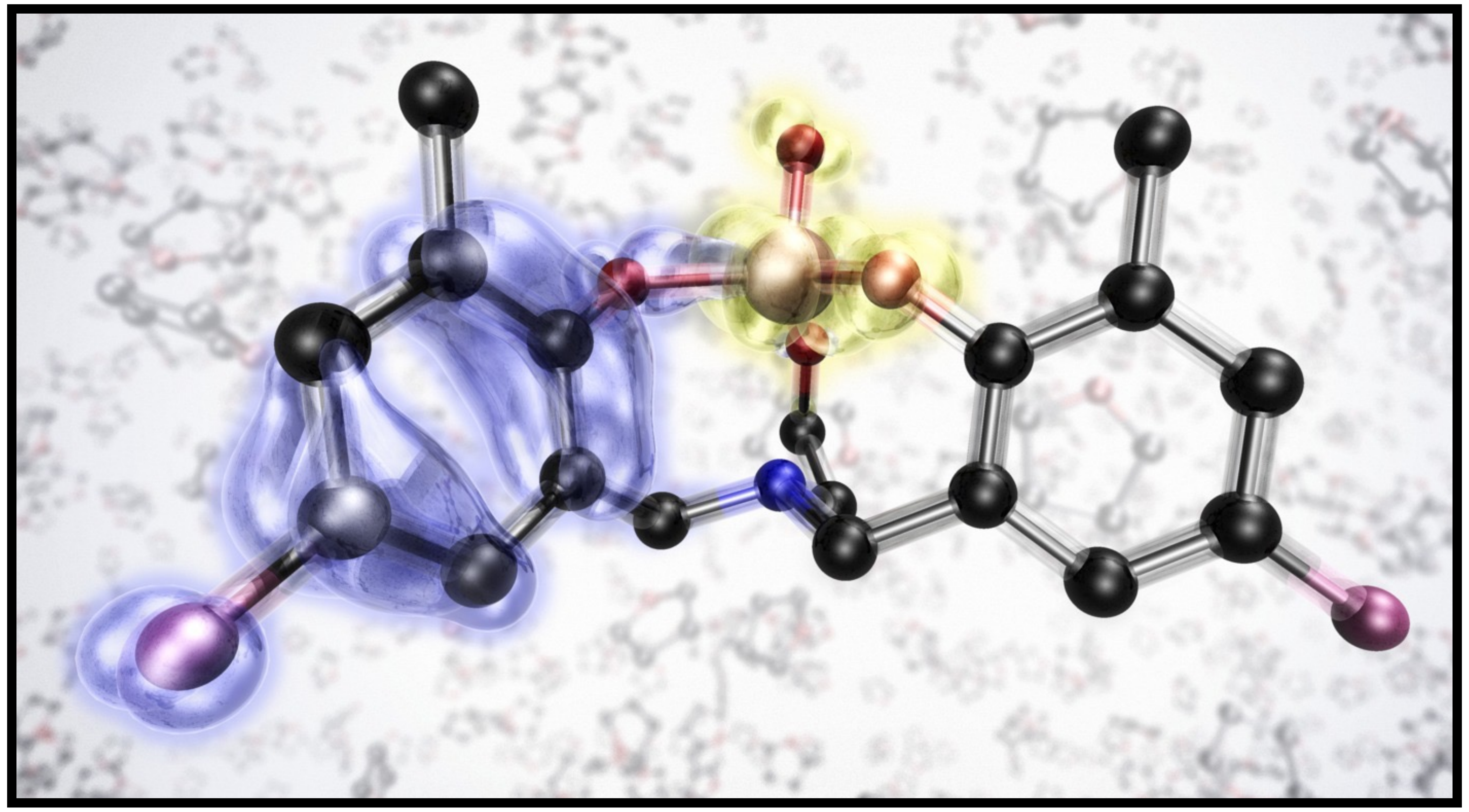
K. Ryczko

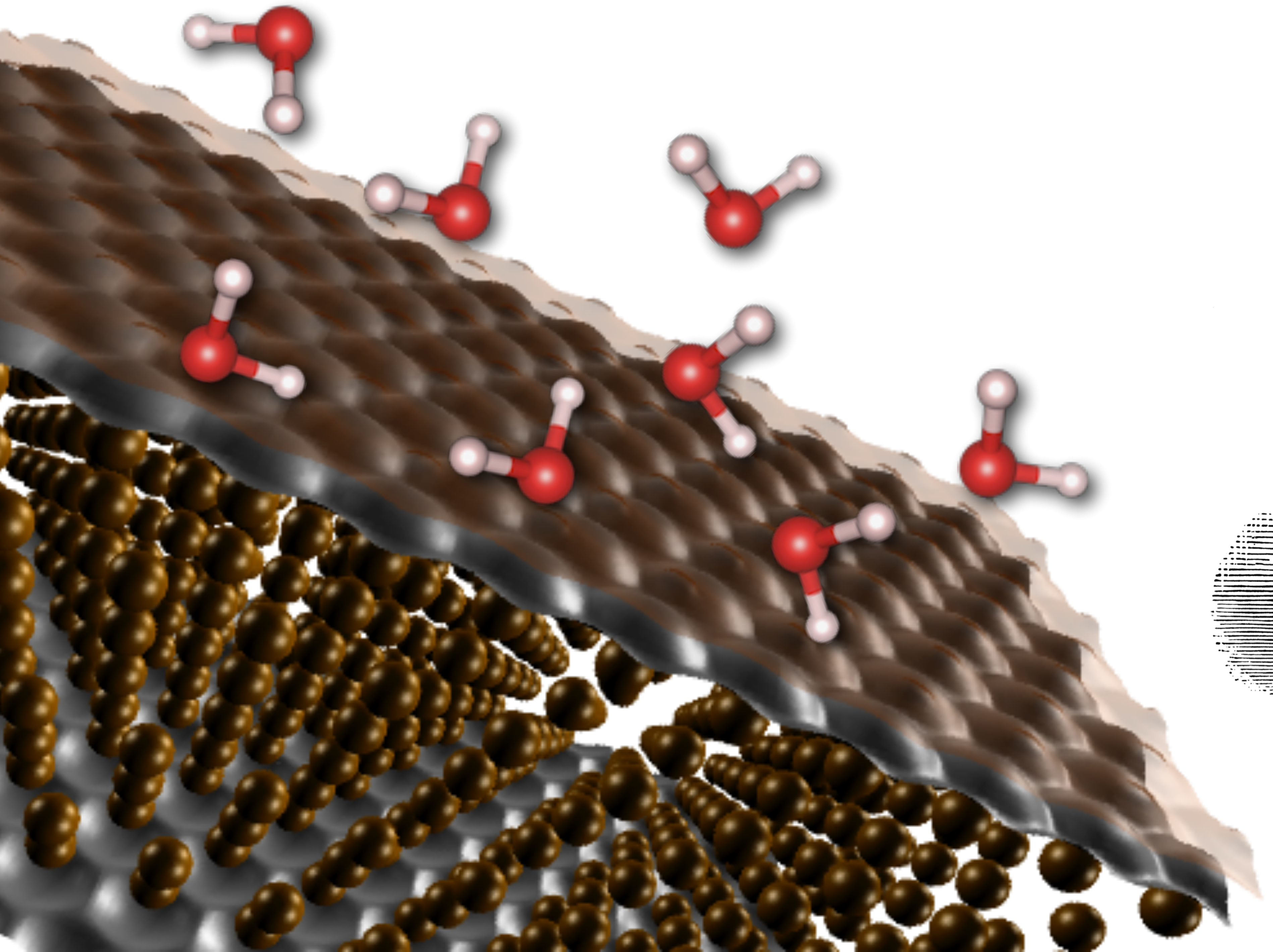
Artificial photosynthesis and water splitting

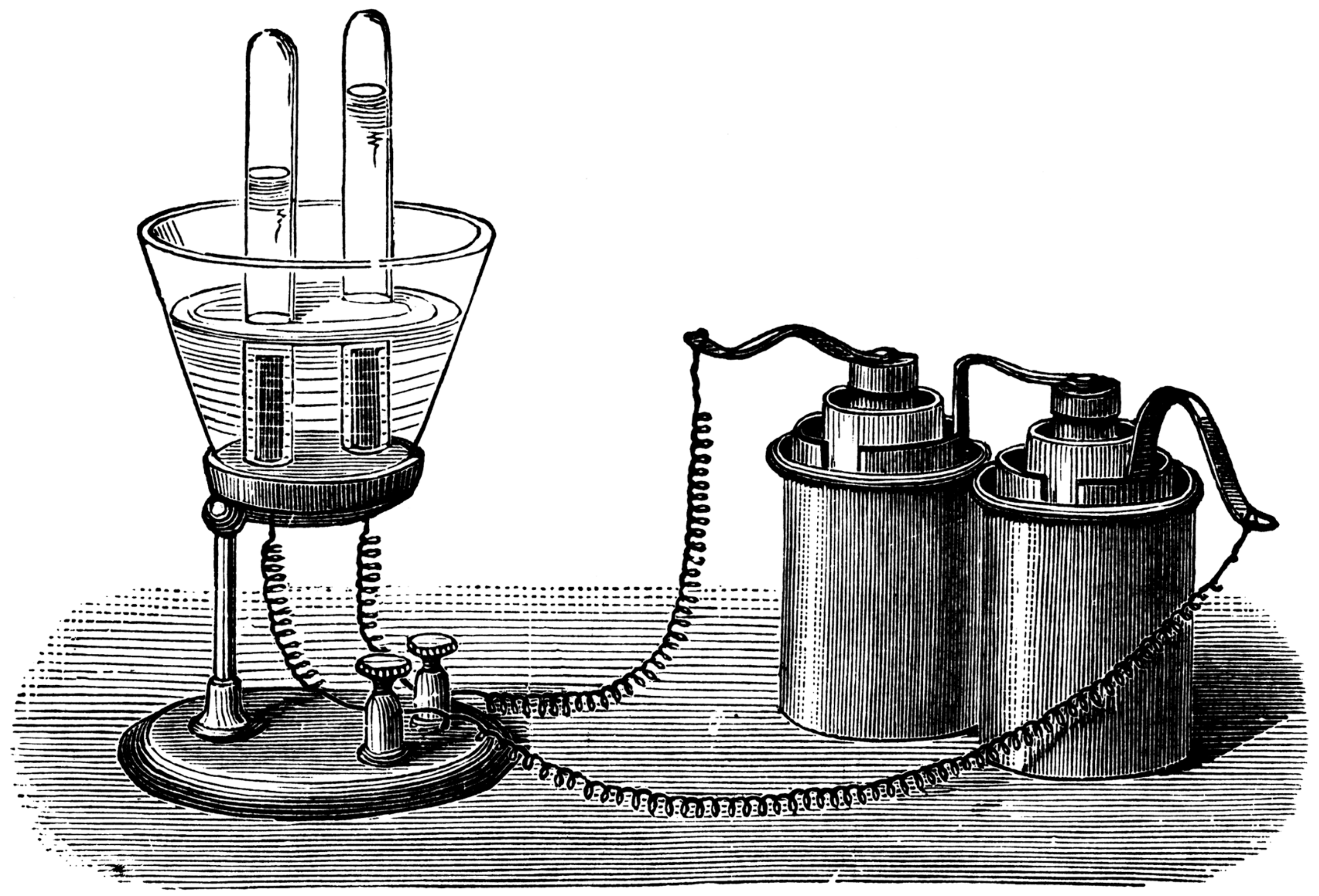


$CO_2 + H_2O + h\nu \rightarrow FUEL + O_2$

Solar PV







Methods

$$\boxed{i\hbar \frac{\partial}{\partial t}\Psi = \hat{H}\Psi}$$

$$\hat{H} = \underbrace{-\frac{\hbar}{2m_e}\sum_i \nabla_i^2}_{T_{\text{elec}}} + \underbrace{\sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|}}_{V_{\text{e-i}}} + \underbrace{\frac{1}{2}\sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}}_{V_{\text{e-e}}} \underbrace{- \frac{\hbar}{2}\sum_I \frac{\nabla_I^2}{M_I}}_{T_{\text{ion}}} + \underbrace{\frac{1}{2}\sum_{I \neq J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}}_{V_{\text{i-i}}}$$

Methods

Coarse graining/Finite element

$\updownarrow$

Force fields/Empirical methods

$\updownarrow$

Density functional theory

$\updownarrow$

Many body methods (*GW*, Quantum Chemistry, QMC)

$V_{\mathrm{ext}}(\mathbf{r})$ is uniquely determined by the ground state particle density $n_0(\mathbf{r})$

$$\begin{tikzcd}
V_{\mathrm{ext}}(\mathbf{r}) \arrow[d] & n_0(\mathbf{r}) \arrow[l, "\mathrm{HK}"'] \\
\Psi_i(\{\mathbf{r}\}) \arrow[r] & \Psi_0(\{\mathbf{r}\}) \arrow[u]
\end{tikzcd}$$

HK = Hohenberg-Kohn theorem

Methods

KS = Kohn-Sham theorem

Ground state density of original system can be described by the ground state density of a <u>non-interacting system</u>

$$\begin{tikzcd}
V_{\text{ext}}(\mathbf{r}) \arrow[d] & n_0(\mathbf{r}) \arrow[l, "\text{HK}"'] \arrow[r, leftrightarrow, "\text{KS}"] & n_0(\mathbf{r}) \arrow[r, "\text{HK}_0"] & V_{\text{eff}}(\mathbf{r}) \arrow[d] \\
\Psi_i(\{\mathbf{r}\}) \arrow[r] & \Psi_0(\{\mathbf{r}\}) \arrow[u] & \psi_{i=1,N_e}(\mathbf{r}) \arrow[u] & \psi_i(\mathbf{r}) \arrow[l]
\end{tikzcd}$$

$$E_{\mathrm{KS}}[n] = T_s[n] + \int d\mathbf{r} V_{\mathrm{ext}}(\mathbf{r}) n(\mathbf{r}) + E_{\mathrm{Hartree}}[n] + E_{\mathrm{ion-ion}} + \boxed{E_{\mathrm{xc}}[n]}$$

Exact form unknown

Successful approximations:

1. Local density approximation (LDA)

   → assume XC energy density ~ homogenous electron gas

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[n] = \int \mathrm{d}^3 r n(\mathbf{r}) \epsilon_{\mathrm{xc}}^{\mathrm{hom}}(n(\mathbf{r}))$$

2. Generalized gradient approximation (GGA)

   → include gradients of charge density

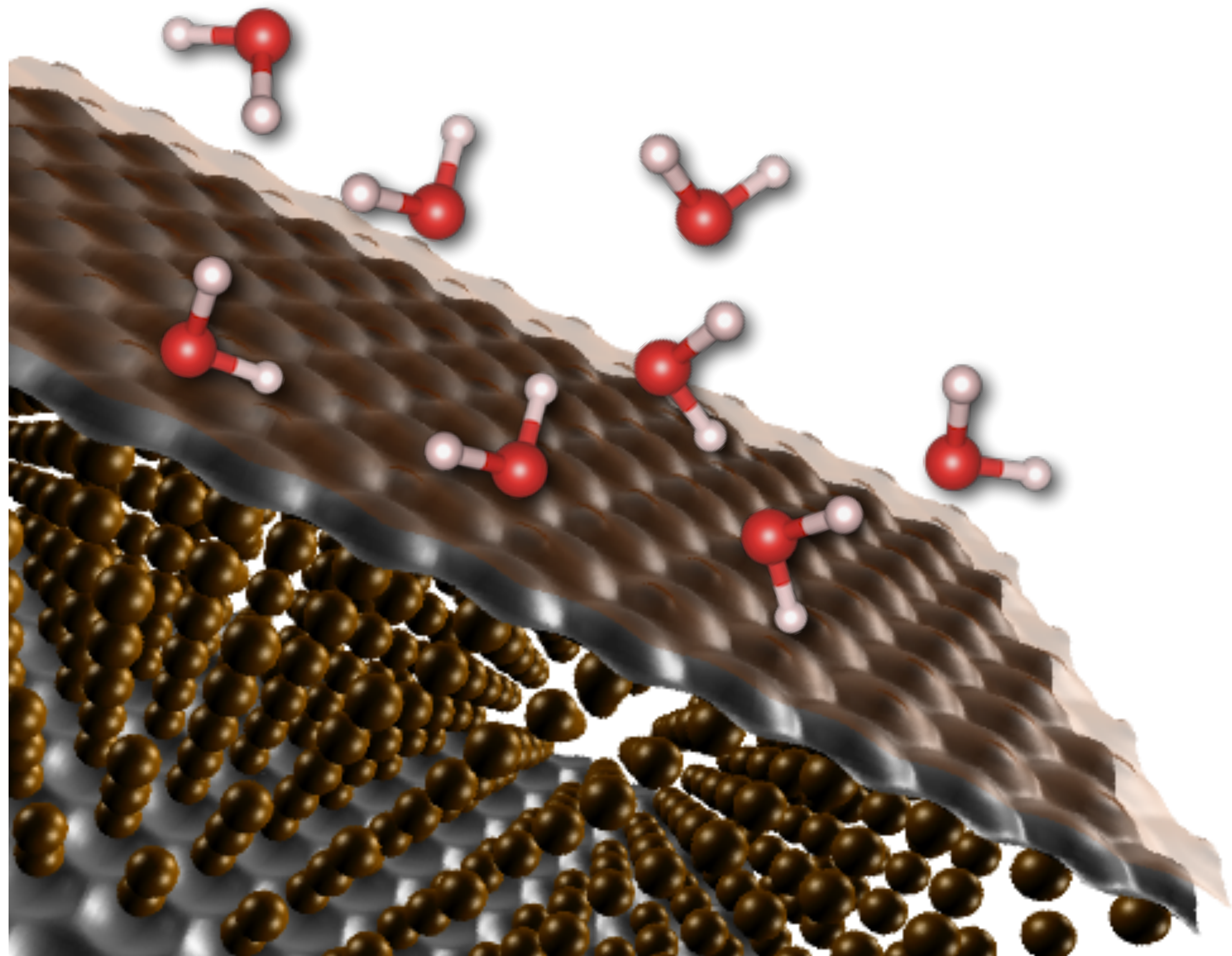
$$E_{\mathrm{xc}}^{\mathrm{GGA}}[n] = \int \mathrm{d}^3 r n(\mathbf{r}) \epsilon_{\mathrm{xc}}(n(\mathbf{r}), |\nabla n(\mathbf{r})|)$$

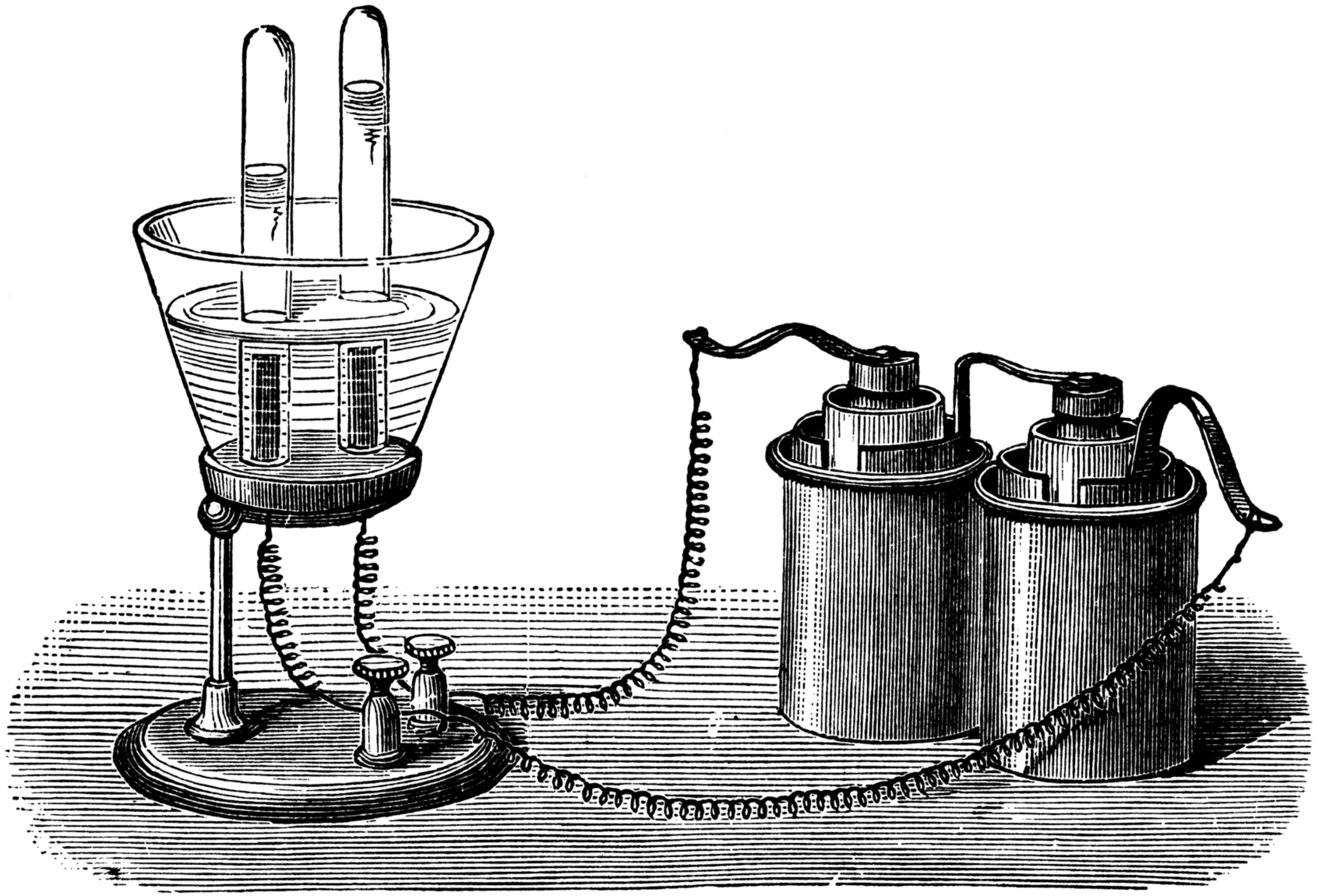
3. Hybrid functionals (B3LYP, HSE, etc)

   → Include fraction of exact exchange

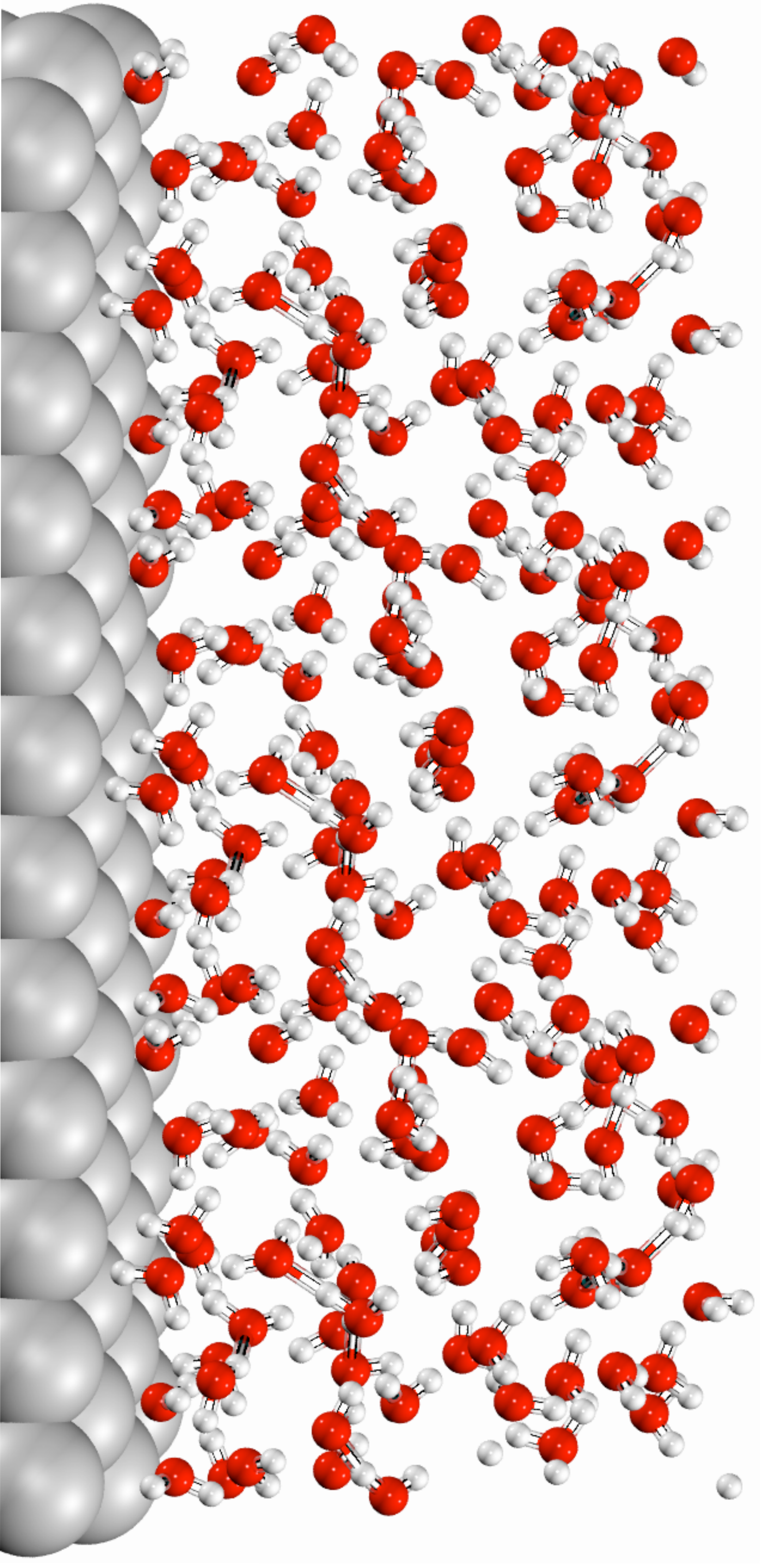
DFT provides a good starting point for higher or lower level calculations

First principles electrochemistry





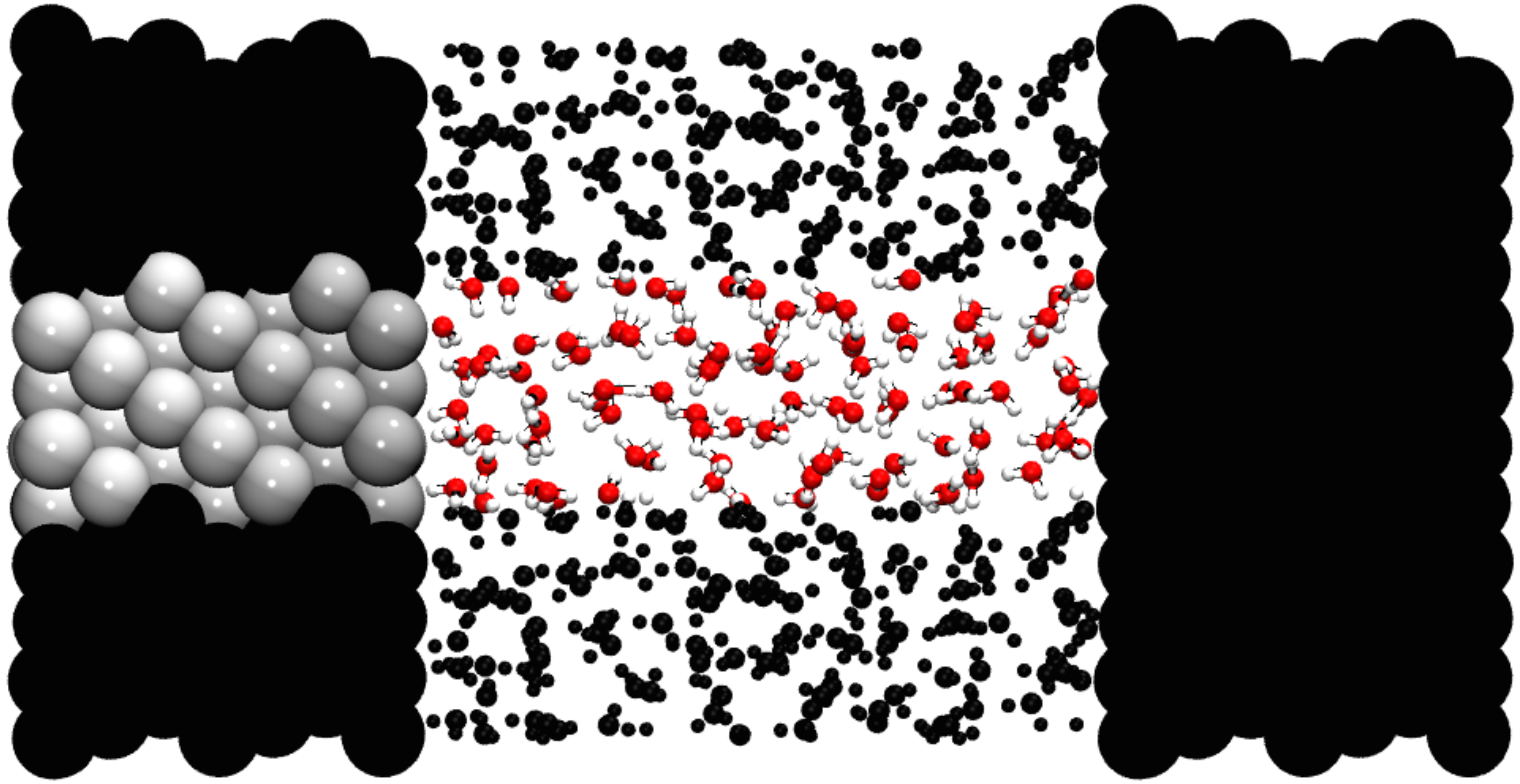
E.M Avery, 1895



A problem of …

- length-scales
- out-of-equilibrium dynamics
- stochastic, incoherent electron transfer
- boundary conditions
- level alignment
- mismatch of errors (reactants, intermediates, and products and therefore rate constants)
- non-local (e.g. vdW) interactions

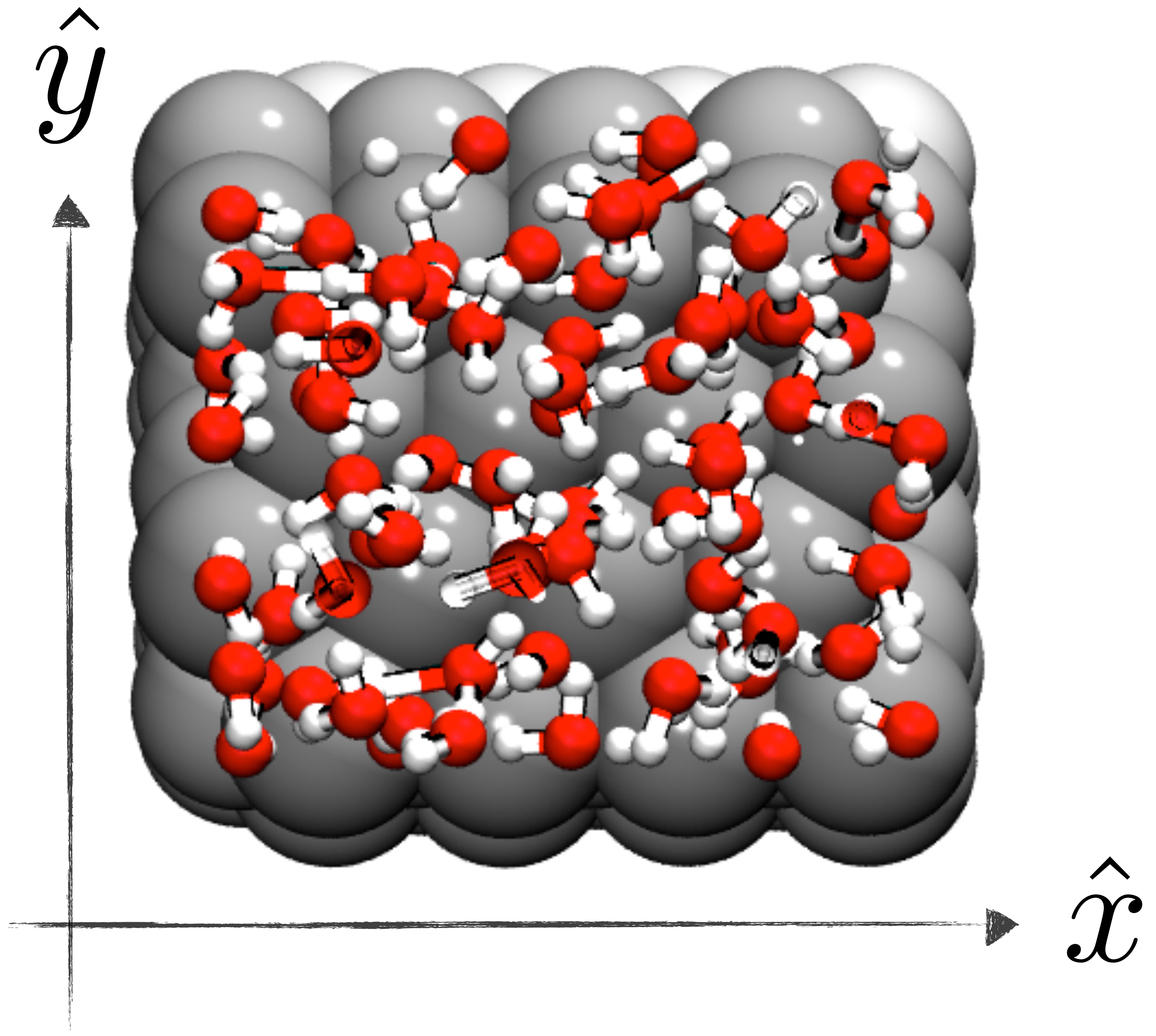
The details matter!



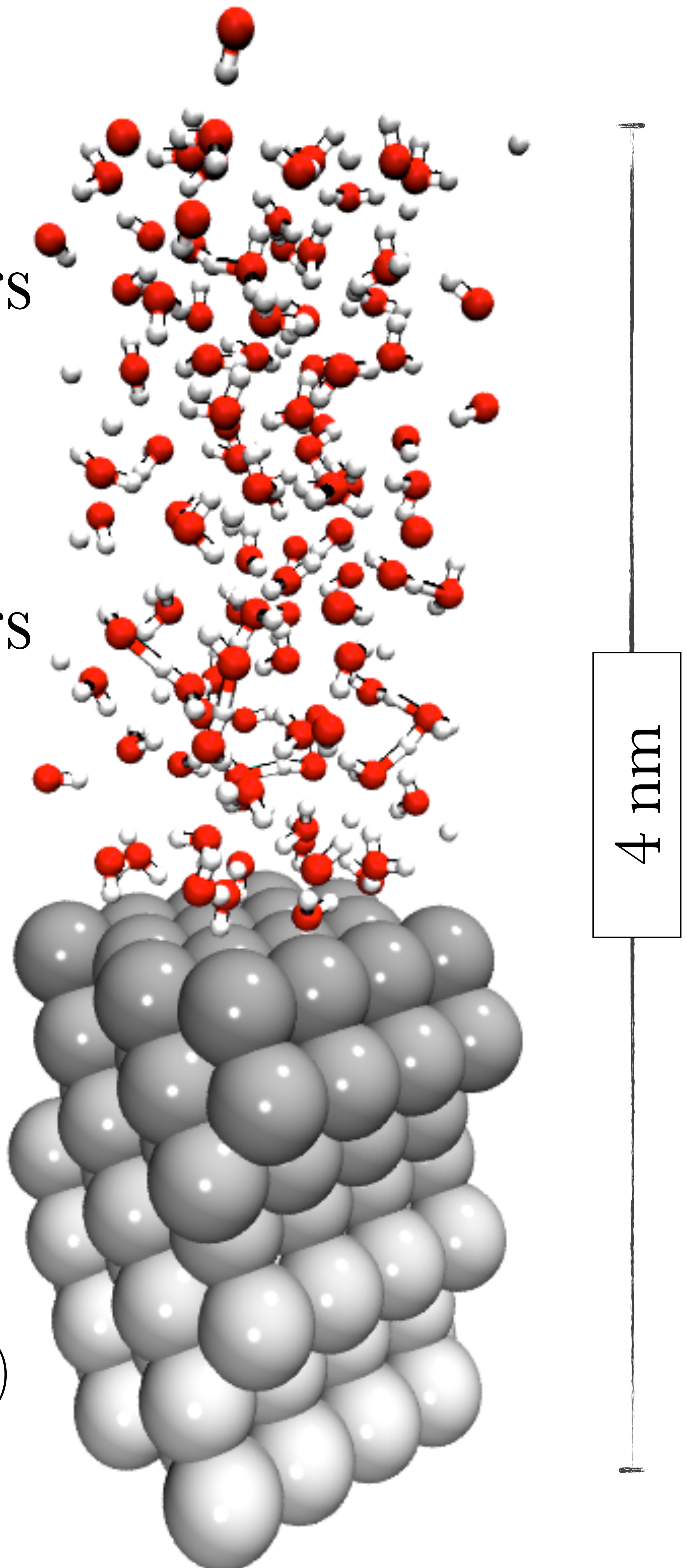
periodic boundary conditions

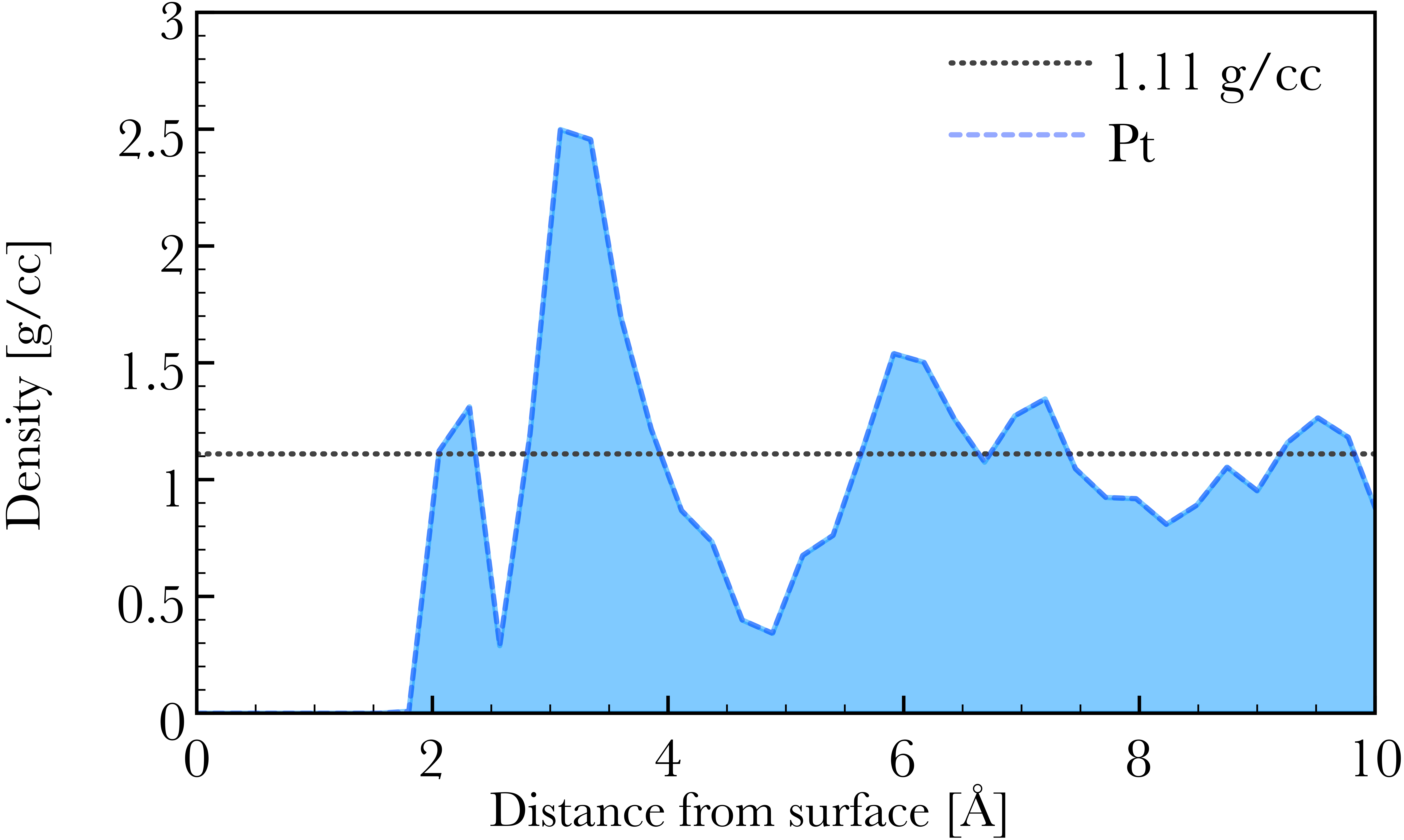
1 picosecond ~
125,000 CPU hours
(PBE)

1 picosecond ~
250,000 CPU hours
(vdW)

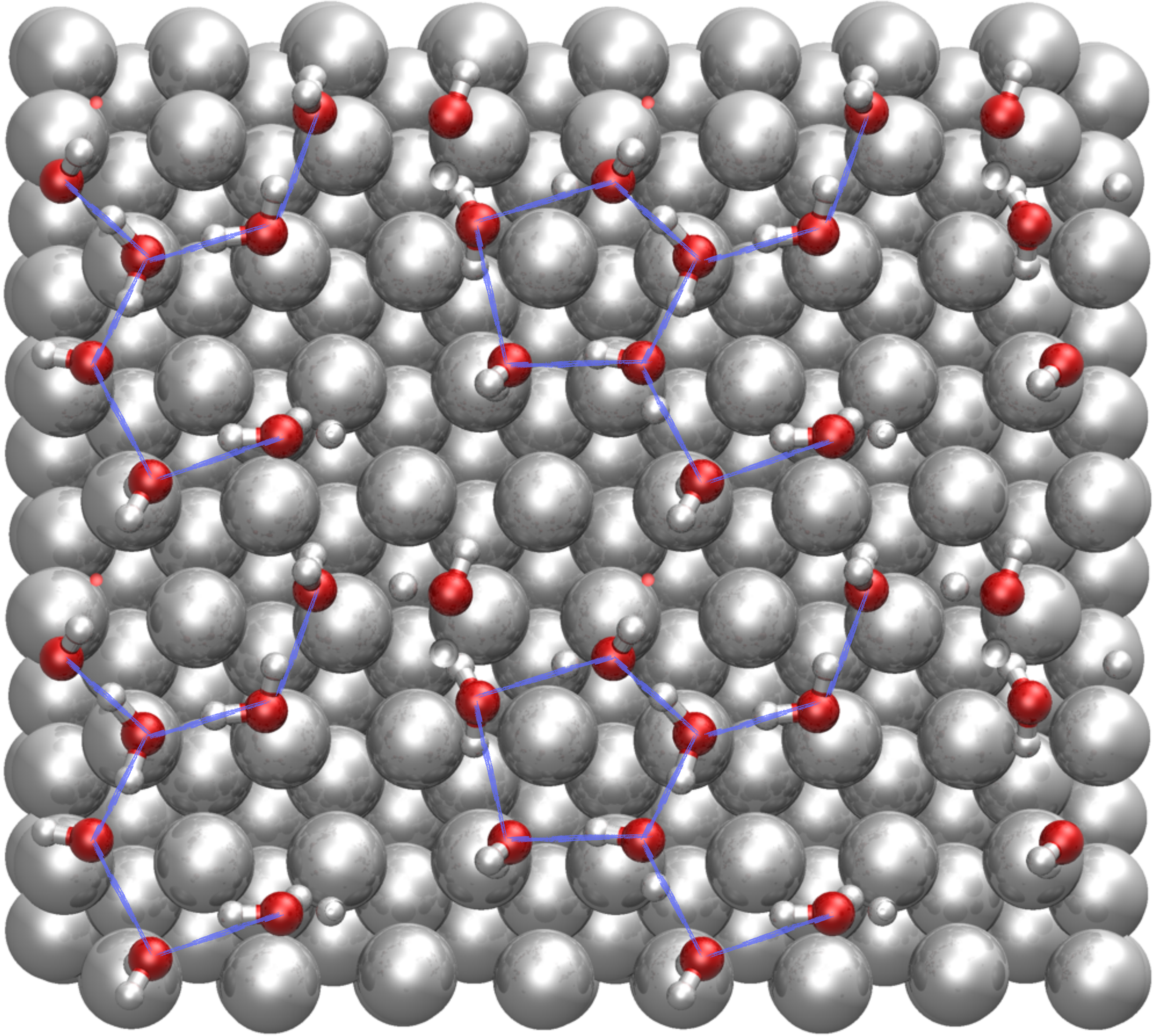


DFT (VASP)
Δt = 11 au (6+ ps simulation time)
2x2x1 *k*-points sampling
PAW potential, 37 Ry cutoff
T= 330-1000 K (Nose-Hover thermostat)

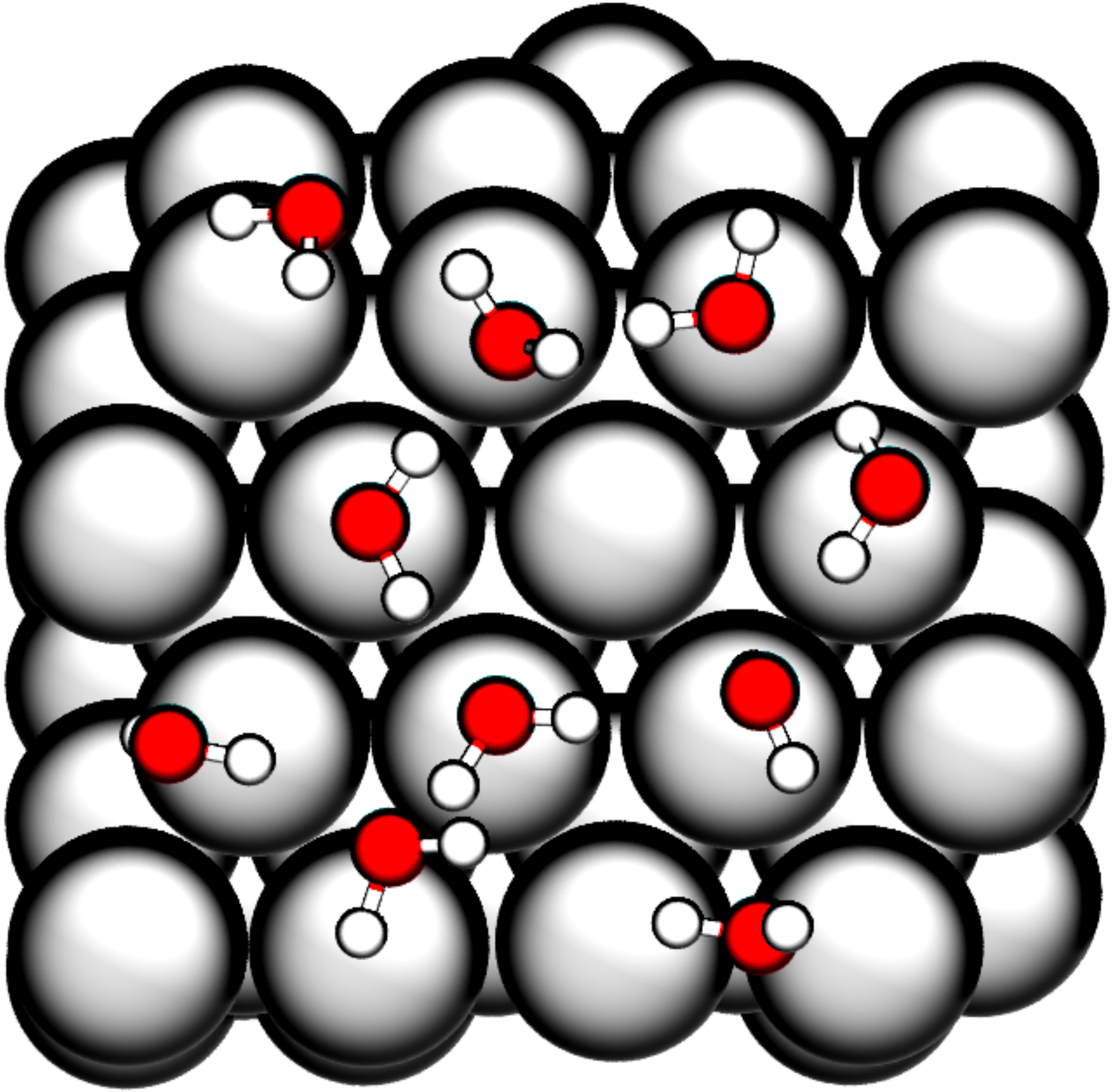


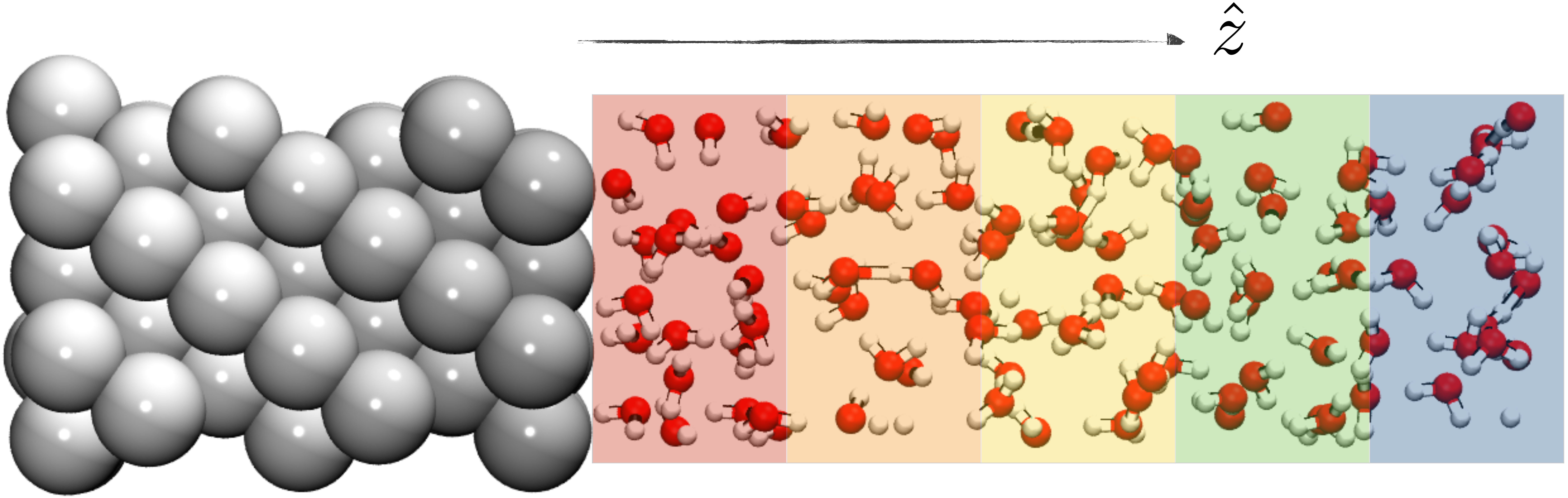


1.11 g/cc
Pt
Density [g/cc]
Distance from surface [Å]
0
0.5
1
1.5
2
2.5
3
0
2
4
6
8
10

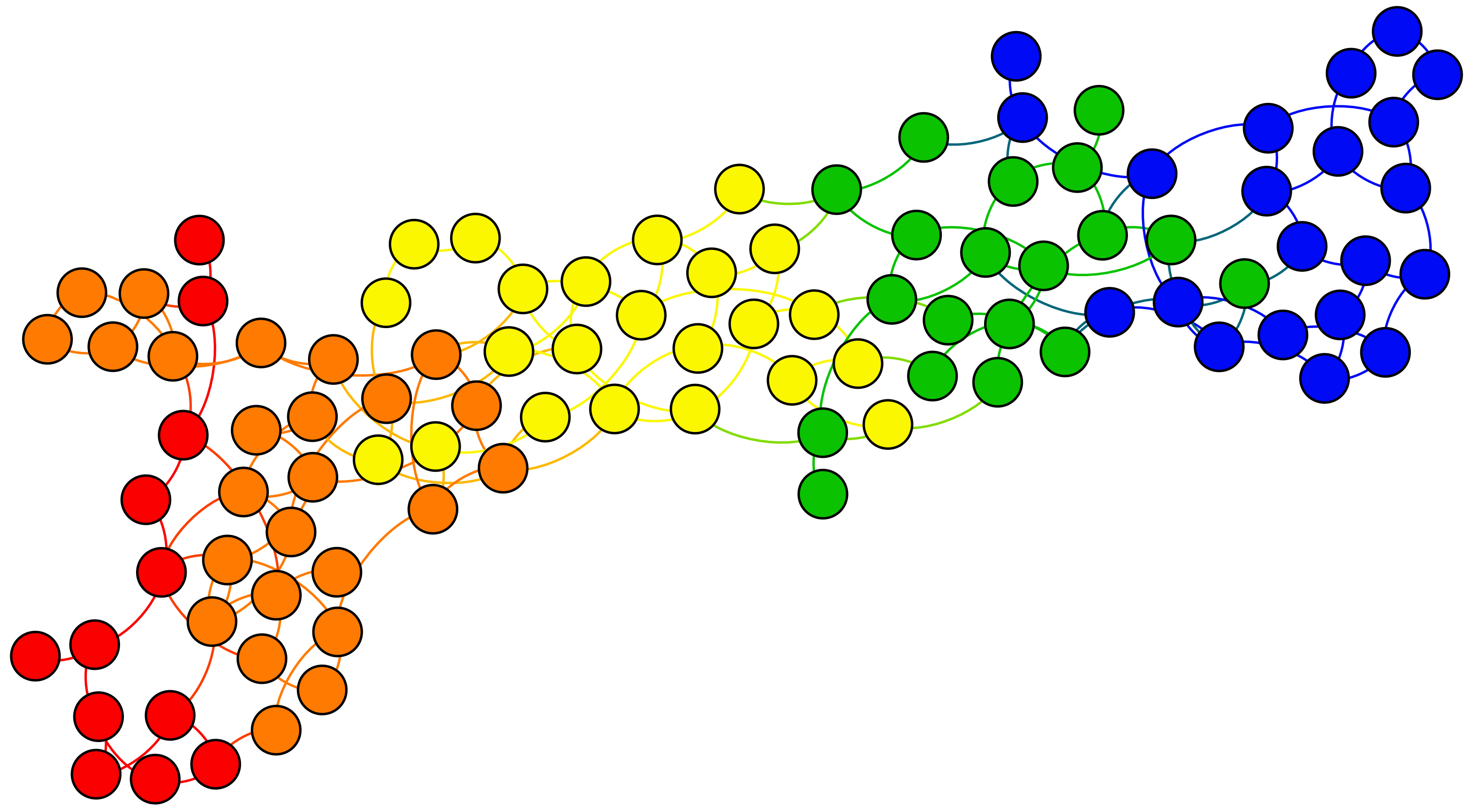


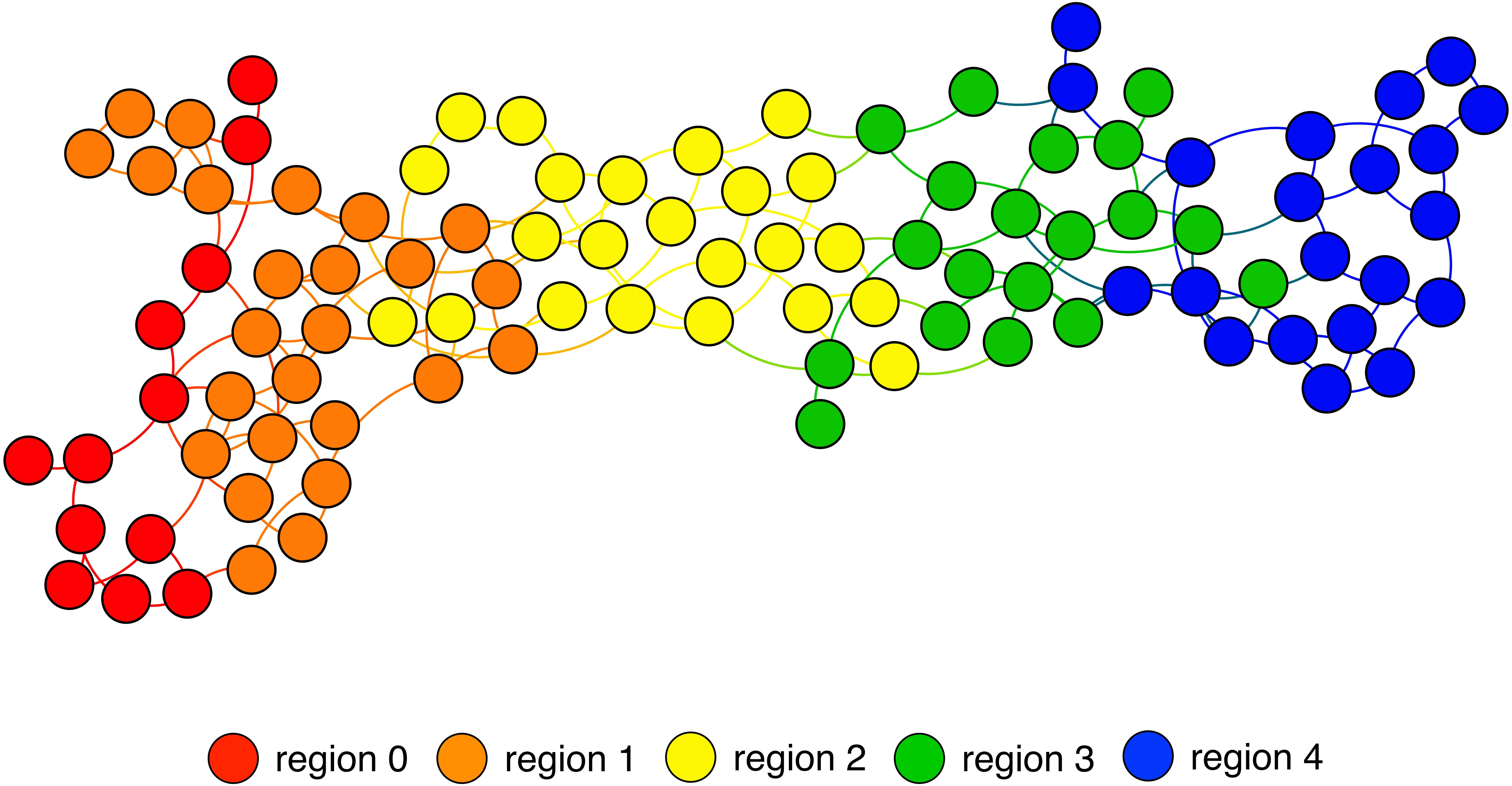
snapshot from MD

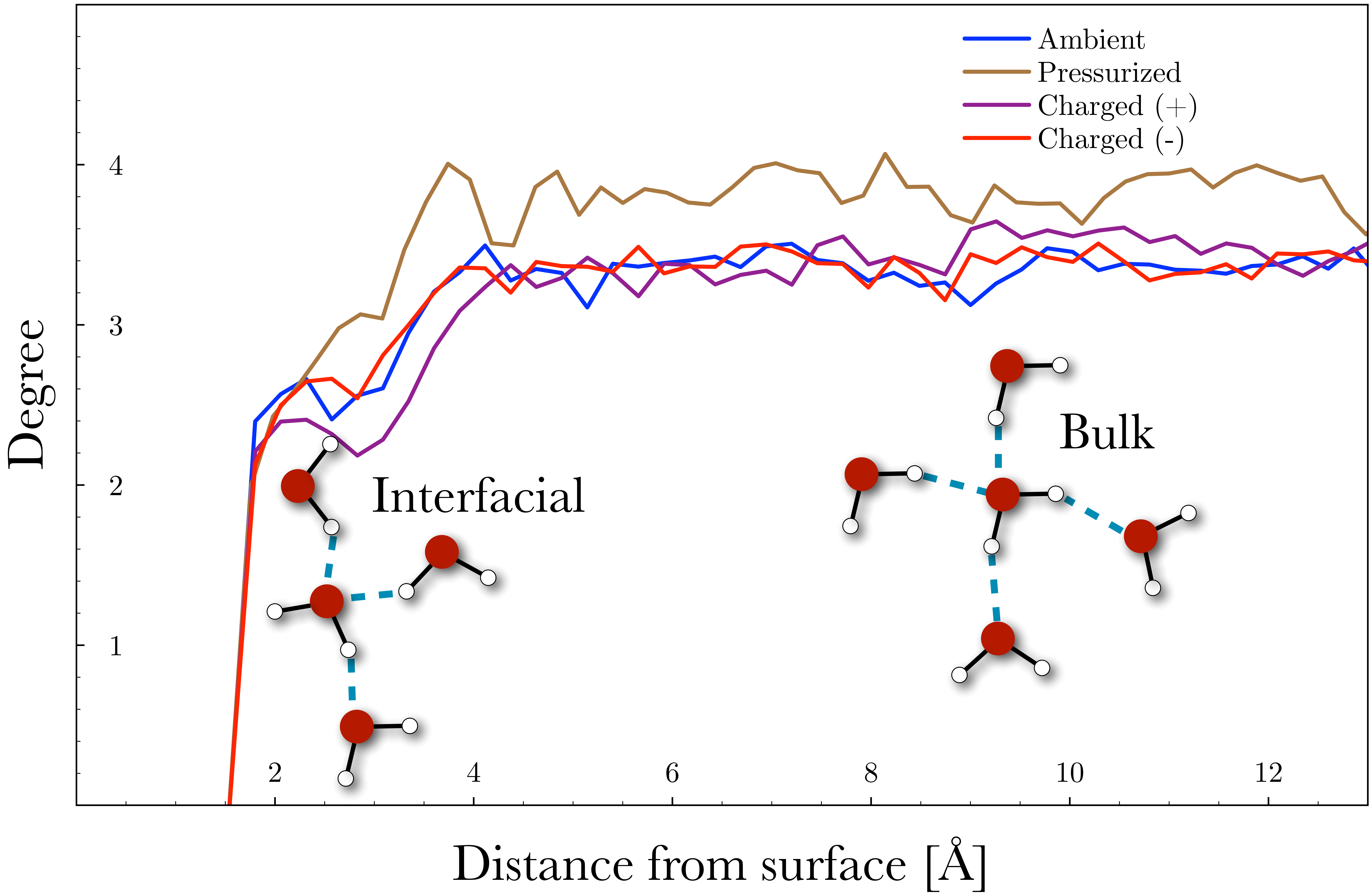




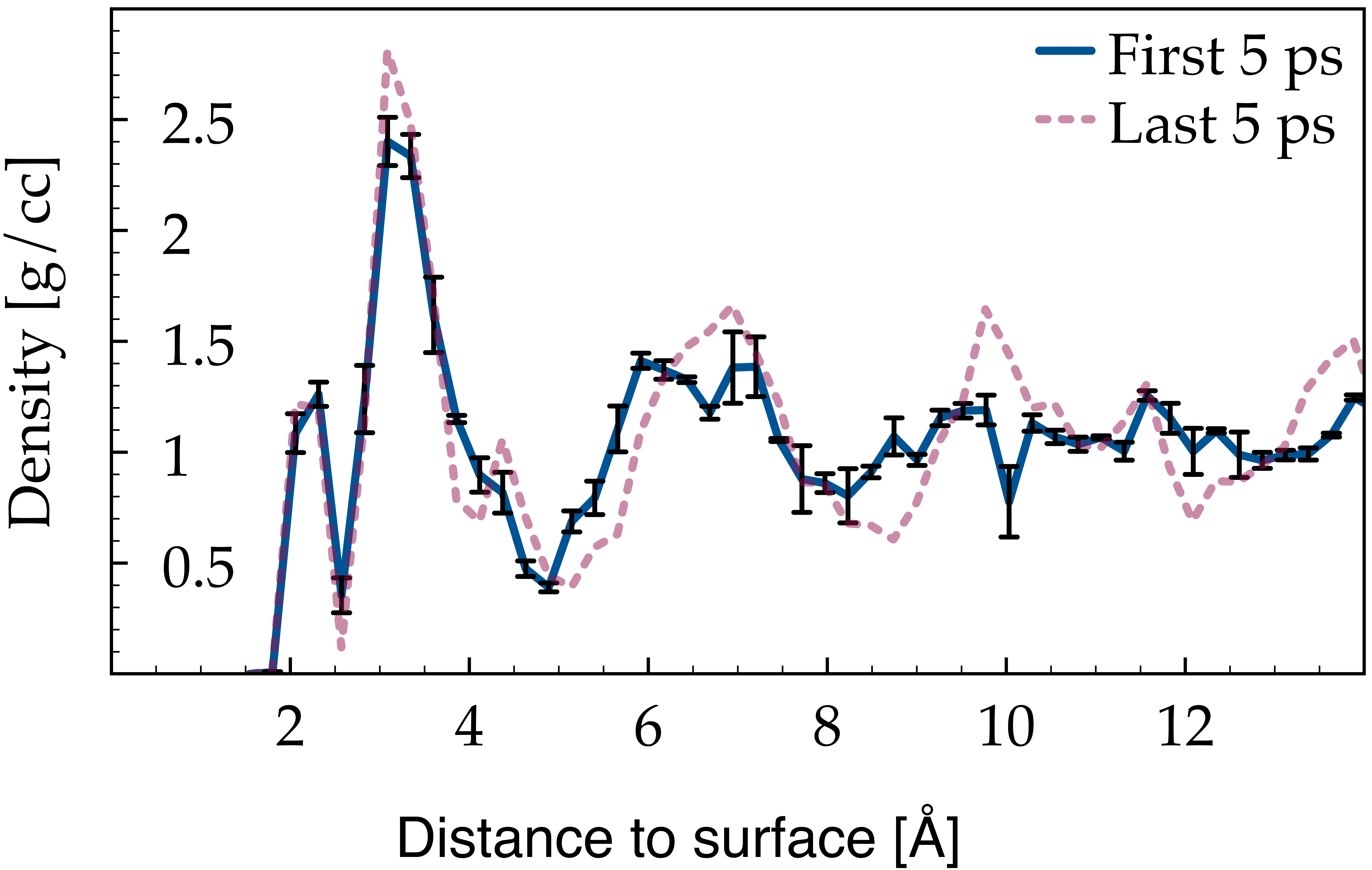
Hydrogen
bonding graph

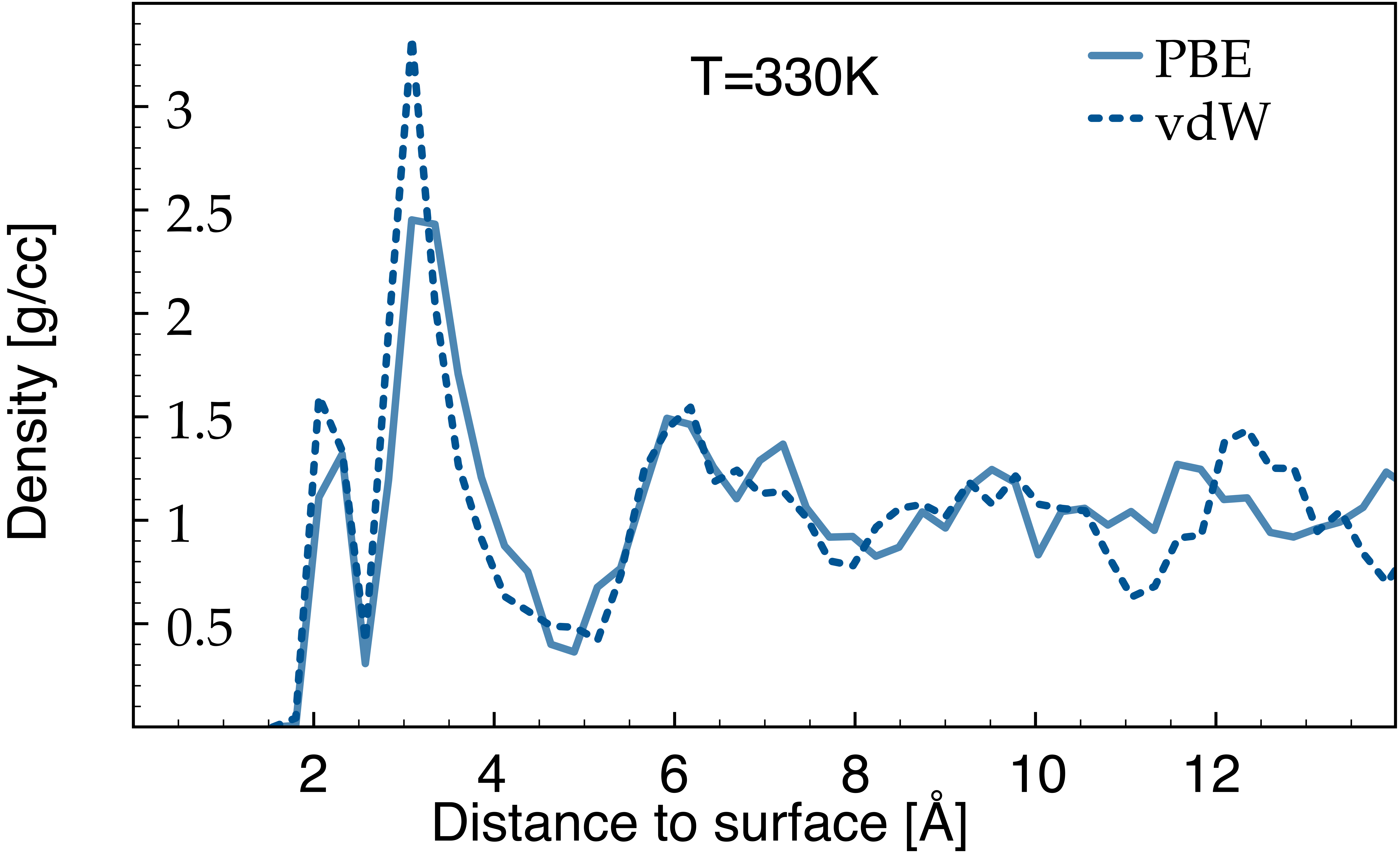



region 0
region 1
region 2
region 3
region 4

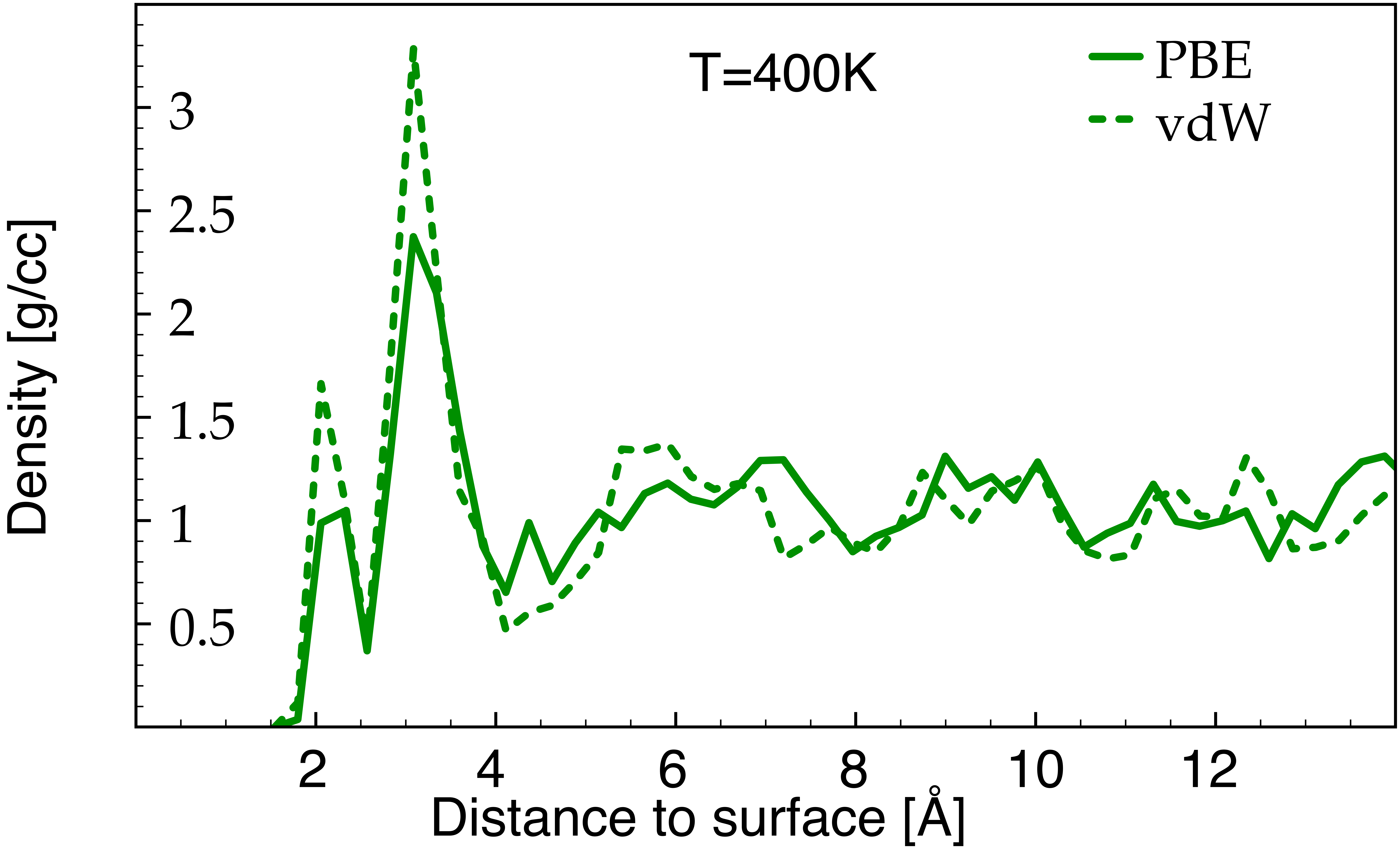


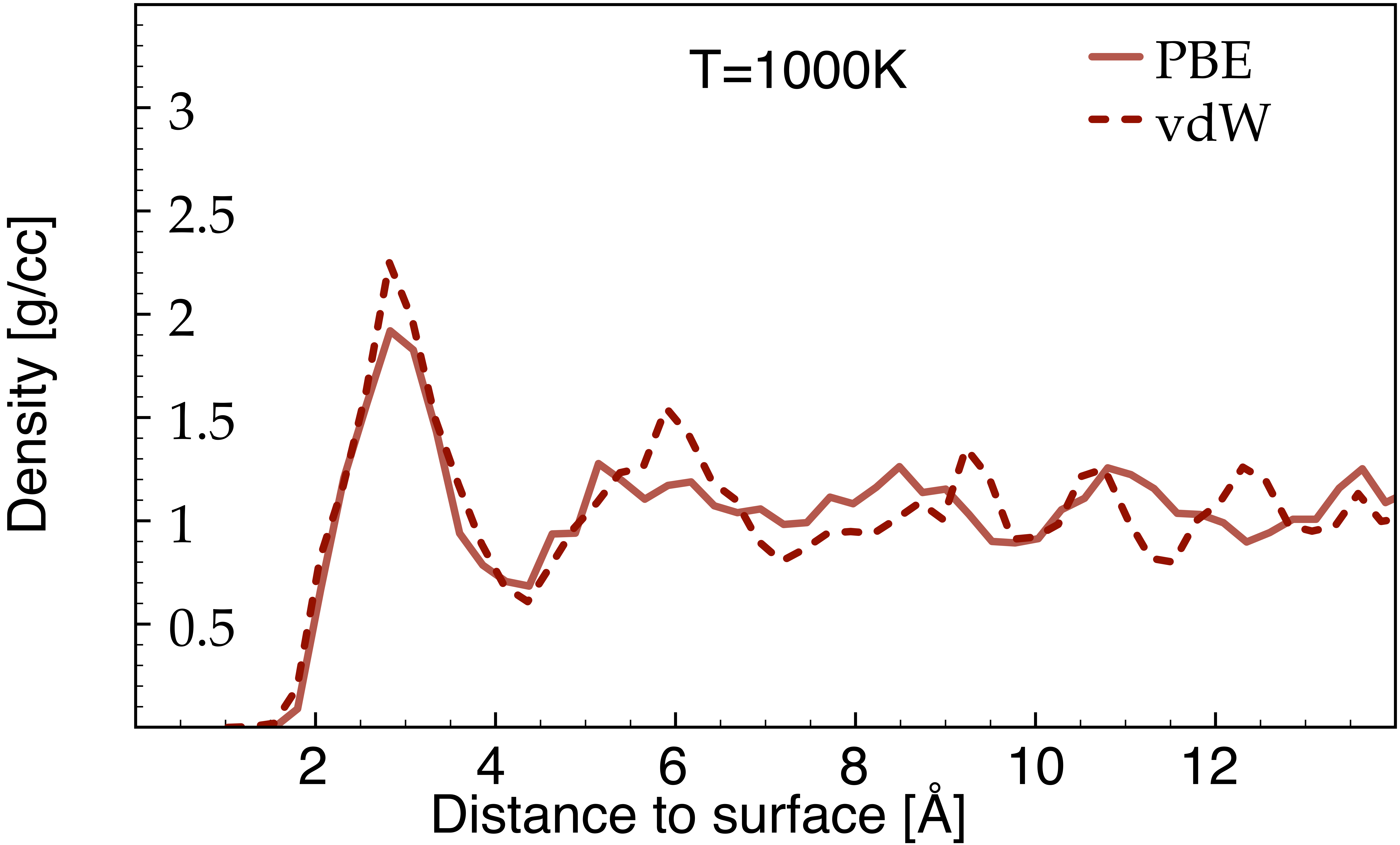
Ambient
Pressurized
Charged (+)
Charged (-)
Bulk
Interfacial
Degree
1
2
3
4
2
4
6
8
10
12
Distance from surface [Å]


First 5 ps
Last 5 ps
Density [g/cc]
0.5
1
1.5
2
2.5
2
4
6
8
10
12
Distance to surface [Å]

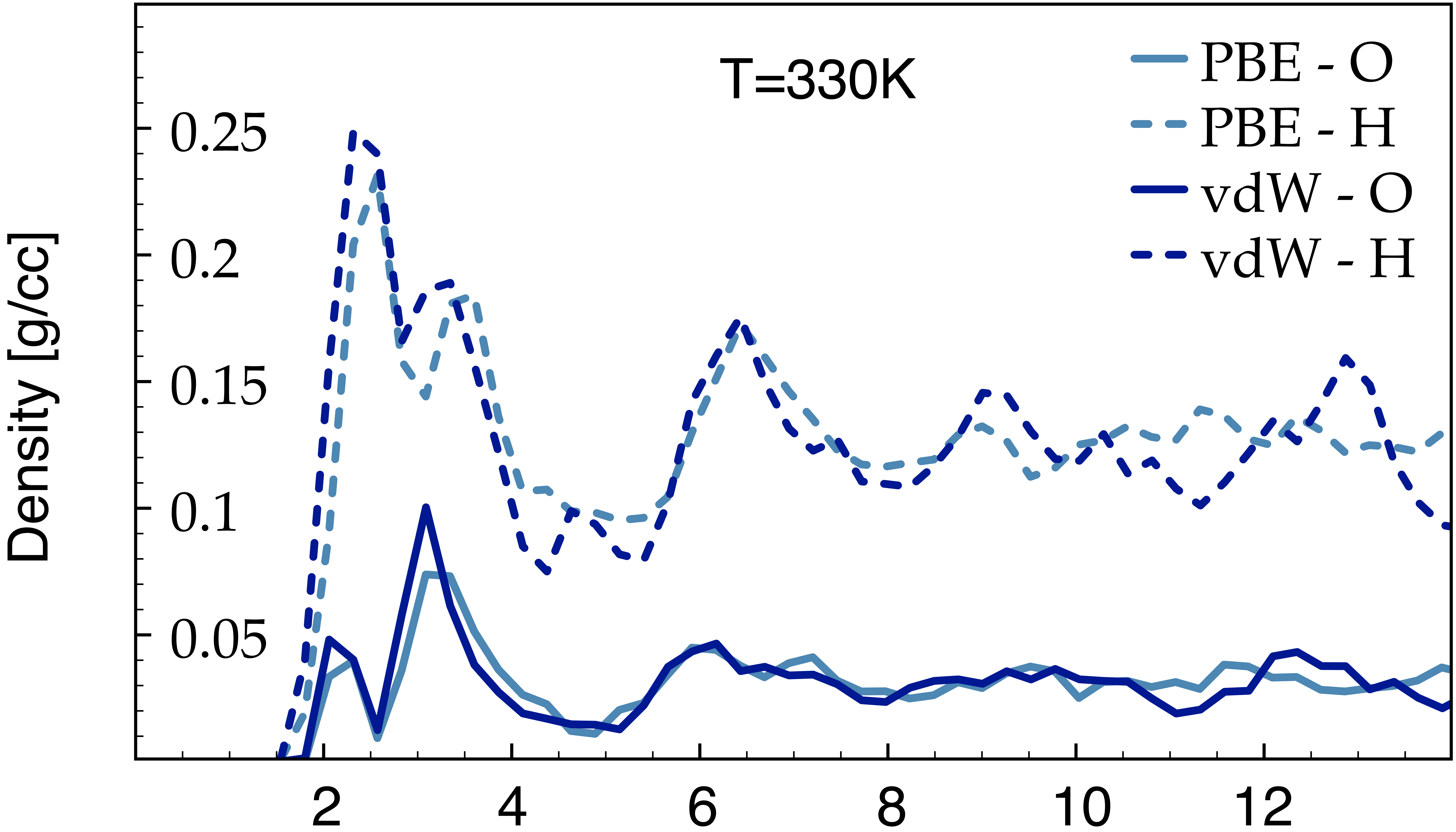


T=330K
PBE
vdW
Density [g/cc]
Distance to surface [Å]
0.5
1
1.5
2
2.5
3
2
4
6
8
10
12

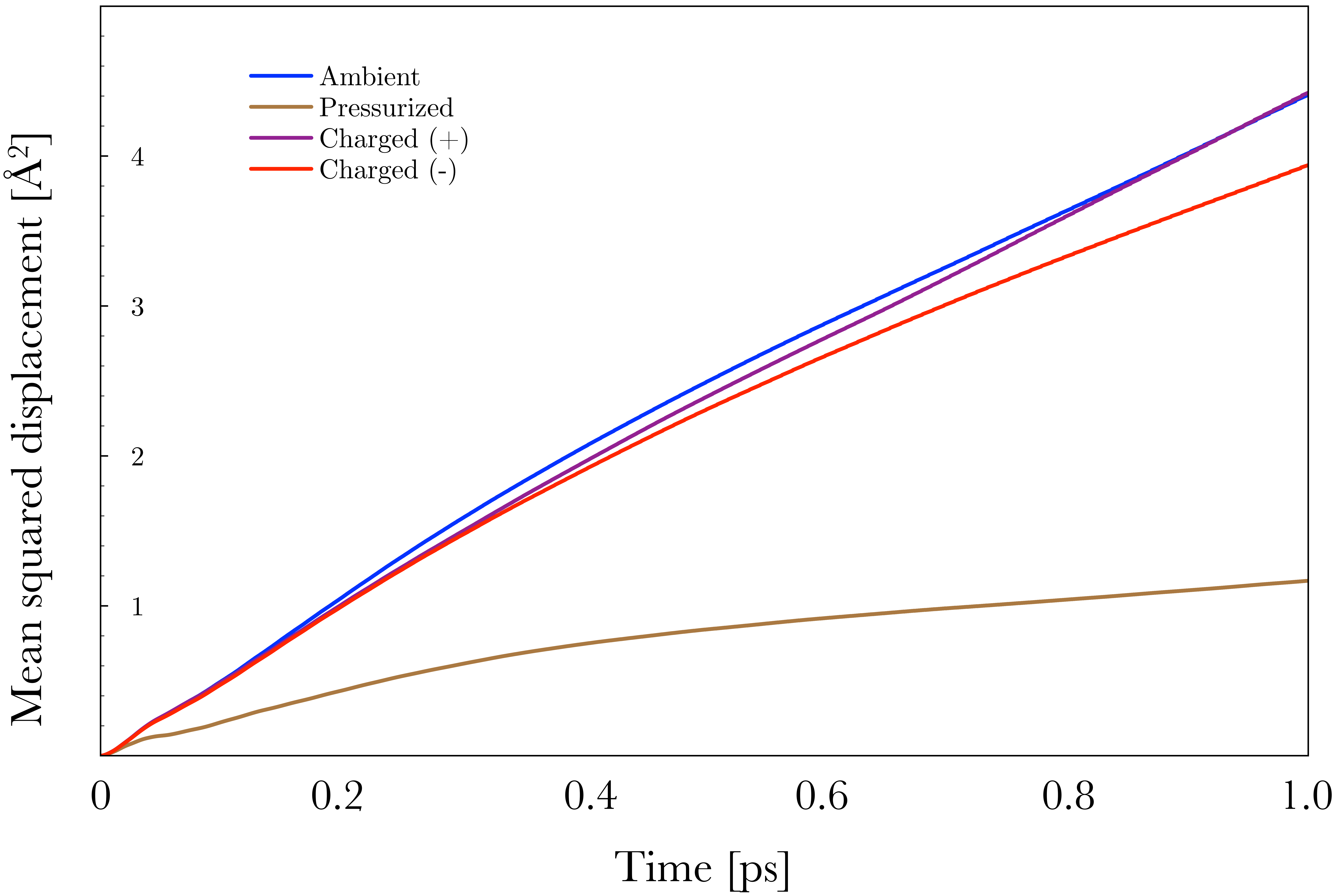

T=400K
PBE
vdW
Density [g/cc]
0.5
1
1.5
2
2.5
3
2
4
6
8
10
12
Distance to surface [Å]



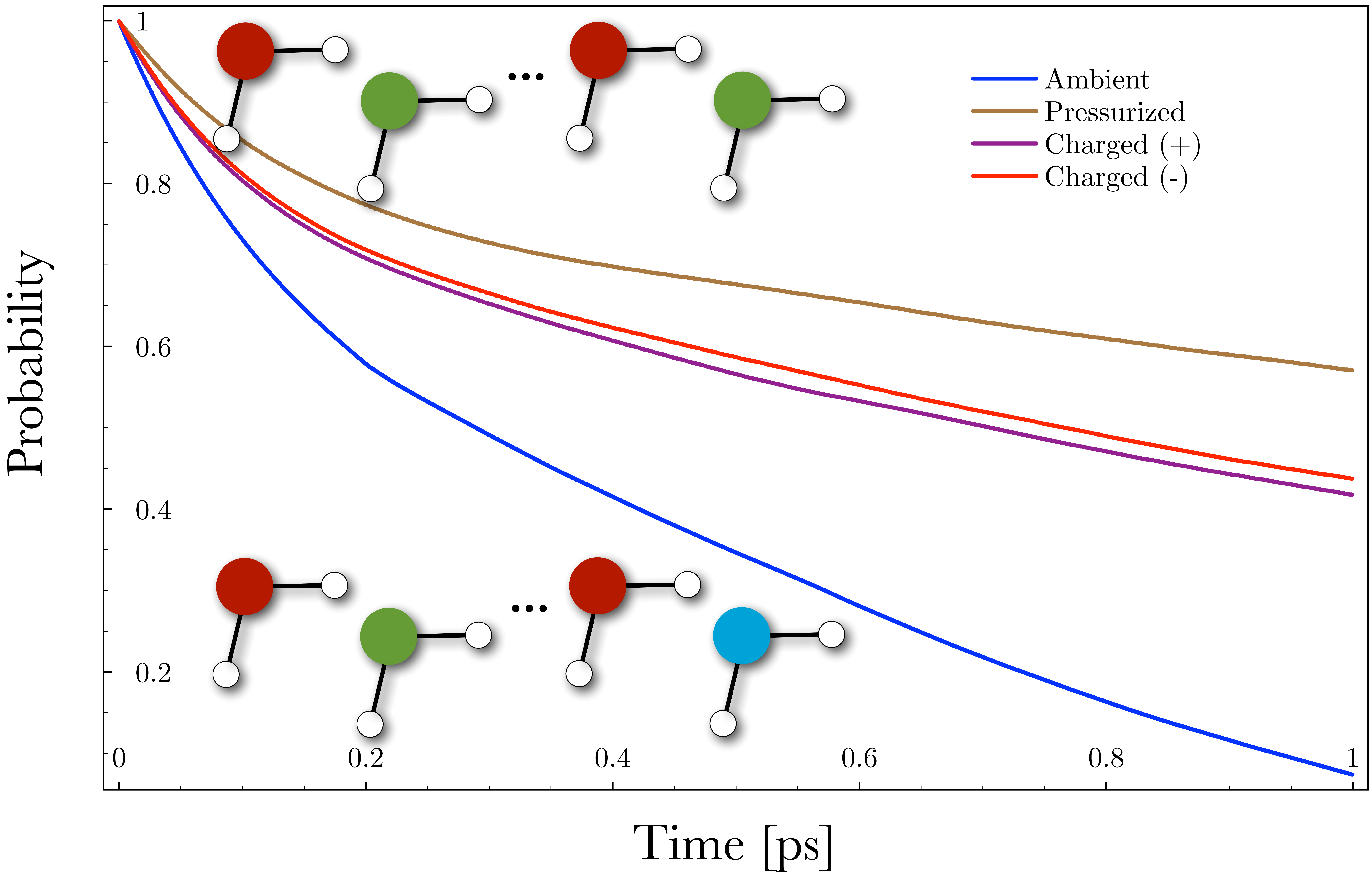
T=1000K
PBE
vdW
Density [g/cc]
3
2.5
2
1.5
1
0.5
2
4
6
8
10
12
Distance to surface [Å]

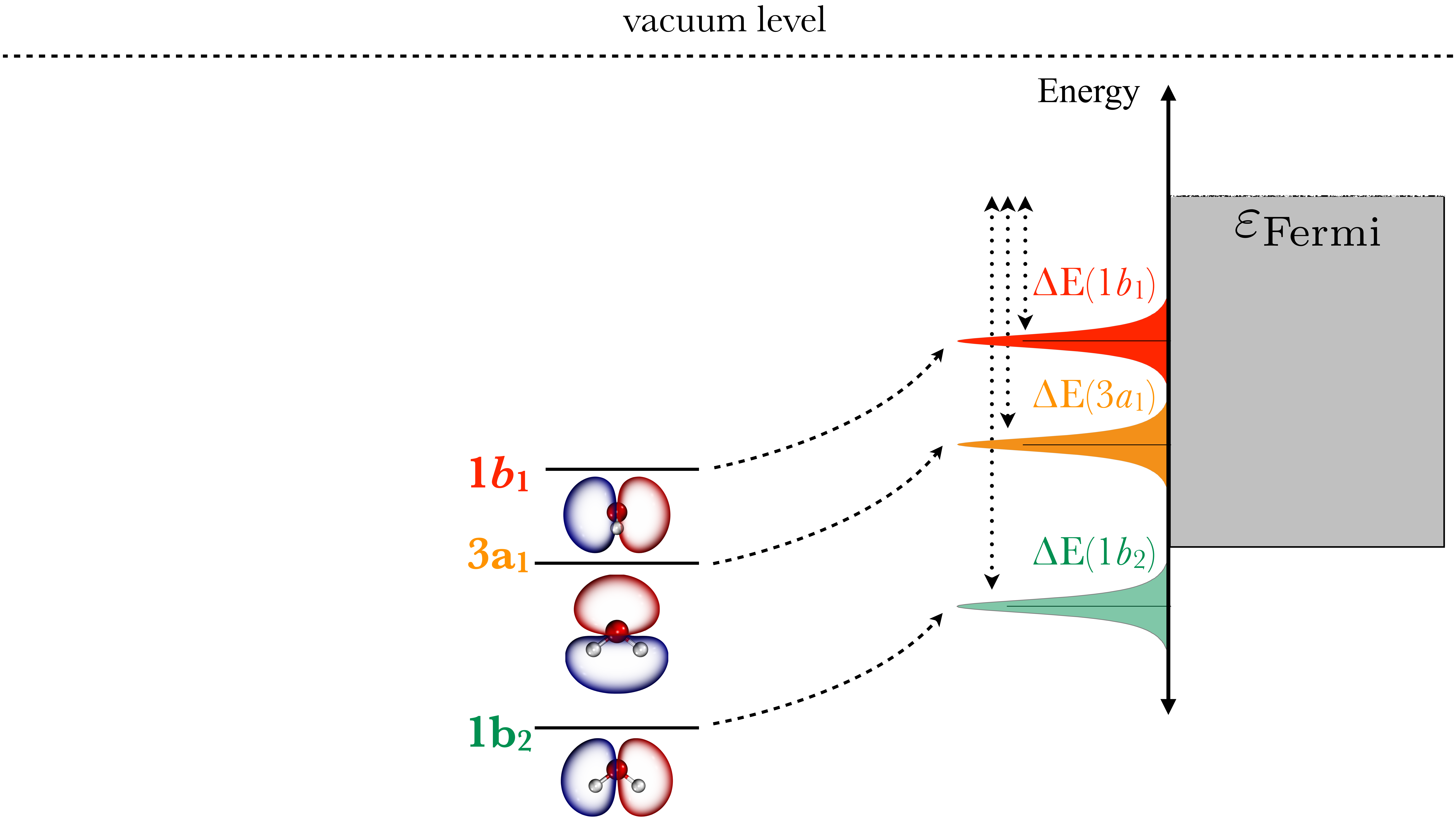


T=330K
PBE - O
PBE - H
vdW - O
vdW - H
Density [g/cc]
0.25
0.2
0.15
0.1
0.05
2
4
6
8
10
12



Ambient
Pressurized
Charged (+)
Charged (-)
Mean squared displacement [Å^2]
Time [ps]
0
0.2
0.4
0.6
0.8
1.0
1
2
3
4


Ambient
Pressurized
Charged (+)
Charged (-)
Probability
Time [ps]
0
0.2
0.4
0.6
0.8
1
0.2
0.4
0.6
0.8
1


vacuum level
Energy
$\varepsilon_{\mathrm{Fermi}}$
$\Delta E(1b_1)$
$\Delta E(3a_1)$
$\Delta E(1b_2)$
$1b_1$
$3a_1$
$1b_2$

Factors affecting level alignment

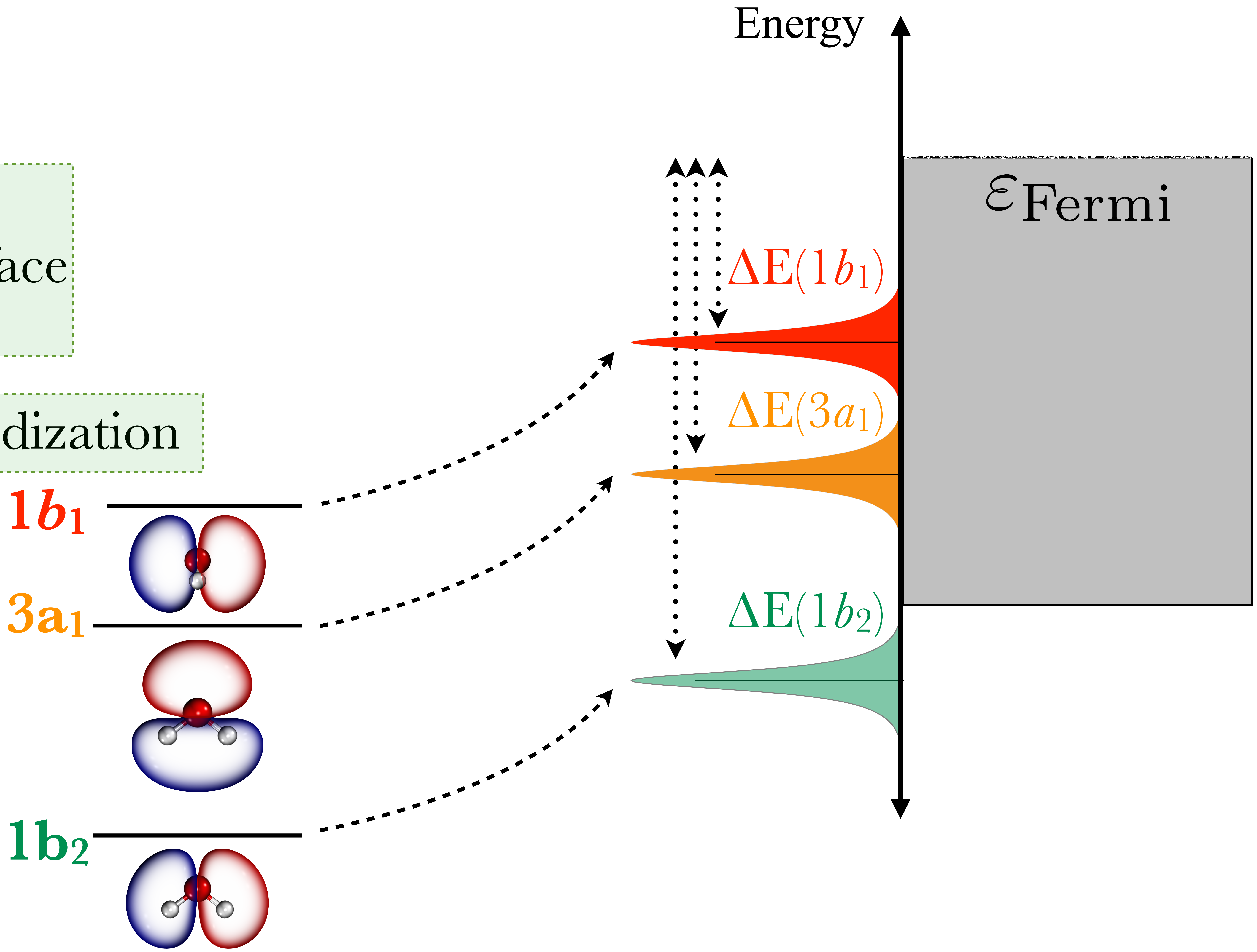
vacuum level

- molecular IE & EA, metal work function

- self consistent charge rearrangement and interface dipole formation

- broadening due to hybridization

- non-local correlation: static and dynamic polarization



JB Neaton, M. Hybertsen, S.G. Louie
Phys. Rev. Lett. 97, 216405 (2006)

interface
clean
Pt(111)
Density of states [states / eV]
25
20
15
10
5
0
-14
-12
-10
-8
-6
-4
-2
0
$\varepsilon - \varepsilon_{\mathrm{Fermi}}$ [eV]

Projected density of states [states / eV]
8
6
4
2
0
$1b_2$
$3a_1$
$1b_1$
-14
-12
-10
-8
-6
-4
-2
0
$\varepsilon - \varepsilon_{\mathrm{Fermi}}$ [eV]

8
6
4
2
0
Projected density of states [states / eV]
1b2
3a1
1b1
-14
-12
-10
-8
-6
-4
-2
0
ε − ε_Fermi [eV]

Perturbation to the DFT Hamiltonian

- Density functional theory

$$[-\nabla^2 + V_{\text{ext}} + V_{\text{Coul}} + V_{xc}]\psi_{nk}^{\text{KS}} = \epsilon_{\text{nk}}^{\text{KS}}\psi_{nk}^{\text{KS}}$$

Hohenberg, Kohn, and Sham 1965

Single particle energies, only E_{HOMO} has meaning

- Many-body approach, QP approximation

$$[-\nabla^2 + V_{\text{ext}} + V_{\text{Coul}} + \Sigma(E_{\text{nk}}^{\text{QP}})]\psi_{nk}^{\text{QP}} = \epsilon_{\text{nk}}^{\text{QP}}\,\psi_{nk}^{\text{QP}}$$

Quasi-particle spectra, *all* E correspond to observables

GW approximation to self-energy

$$\Sigma(r, r'; E) = \frac{i}{2\pi} \int d\omega G(r, r', E + \omega) W(r, r'; \omega) e^{i\delta\omega}$$

$$G(r, r', \omega) = \Sigma_{nk} \frac{\psi^{\dagger}_{nk}(r)\psi_{nk}(r')}{\omega - E_{nk} - i\delta_{nk}}$$

$$W(r, r', \omega) = \int d^3 r'' v(r, r'') \epsilon^{-1}(r'', r', \omega)$$

$$\mathrm{H} = \mathrm{H}_0 + (\mathrm{H}-\mathrm{H}_0)$$

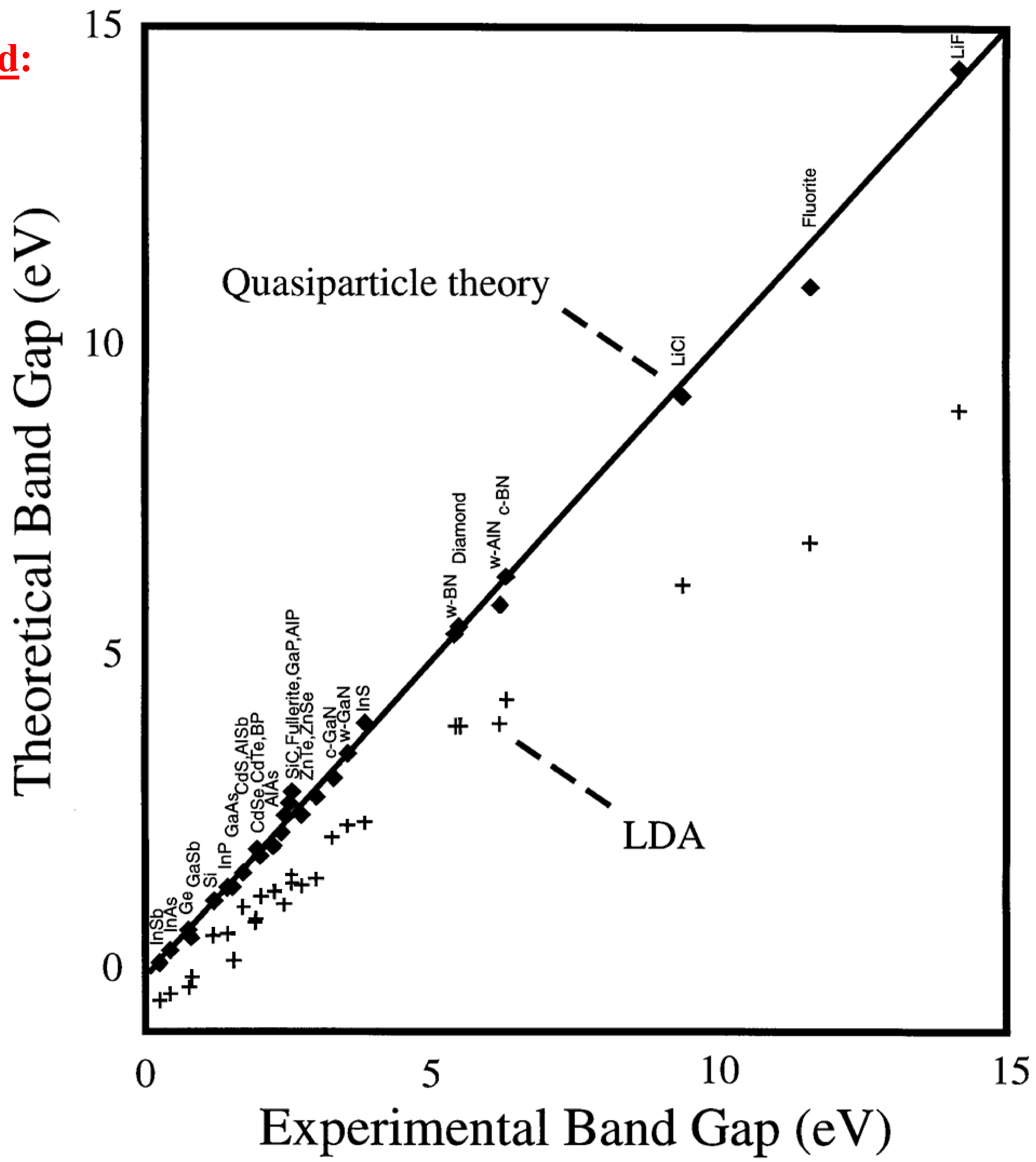
$$E_{QP} = E_{DFT} + \langle \psi_{DFT} | \Sigma - V_{xc} | \psi_{DFT} \rangle$$

Procedure

1. Calculate Ψ_{nk} and E_{nk}
2. Calculate dielectric matrix
3. Construct Σ operator
4. Compute QP spectra

Materials included:

InSb, InAs
Ge
GaSb
Si
InP
GaAs
CdS
AlSb, AlAs
CdSe, CdTe
BP
SiC
C_{60}
GaP
AlP
ZnTe, ZnSe
c-GaN, w-GaN
InS
w-BN, c-BN
diamond
w-AlN
LiCl
Fluorite
LiF



GW also gives good dispersion relations as compared to angle-resolved photoemission.

Compiled by
E. Shirley and
S. G. Louie

GW Details

- plasmon-pole model (Hybertsen & Louie [1])
- screened cutoff = 20 Ry (bare exchange 60 Ry)
- 4096 unoccupied states (~ 35 eV above $\varepsilon_{\mathrm{Fermi}}$, IE within 5%)
- extrapolation method = static remainder [2] approach
- 16,384 cores (32 TB RAM) **BlueGene/Q**
- "avoided" off diagonals through the right choice of basis

[1] M. S. Hybertsen and S. G. Louie, Phys. Rev. B 34, 5390 (1986)

[2] J. Deslippe, G. Samsonidze, M. Jain, M. L. Cohen, and S. G. Louie, Phys. Rev. B 87, 165124 (2013)

Evaluating $\hat{\Sigma}$

- the goal of a *GW* calculation is to obtain the quasi-particle wavefunctions and eigenvalues, i.e. the addition and removal energies of the system

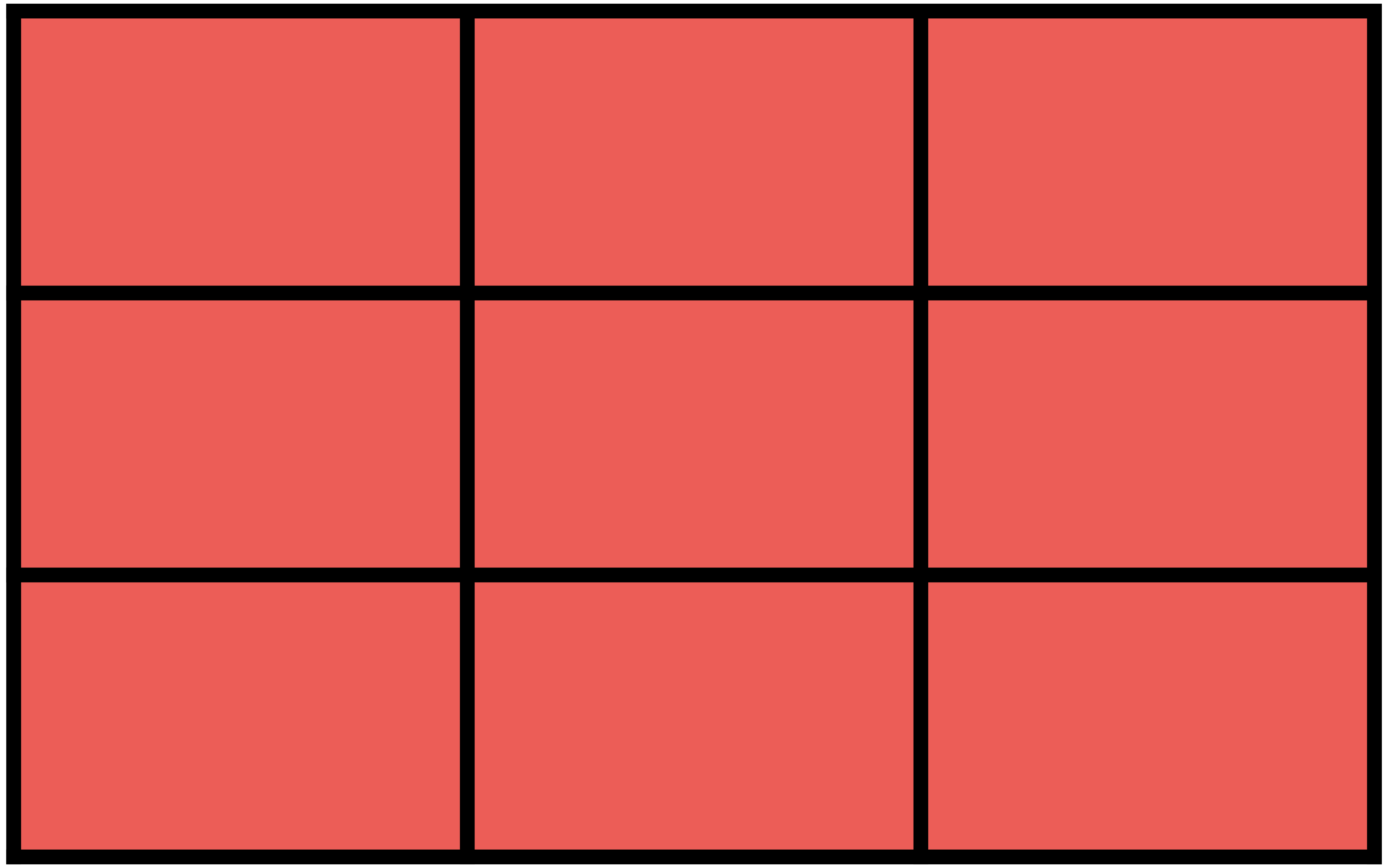
- KS-DFT is (typically) used as the mean field solution in a *GW* calculation

- in many *legacy* systems

$$|\Psi_{\mathrm{KS}}\rangle \approx |\Psi_{\mathrm{QP}}\rangle$$

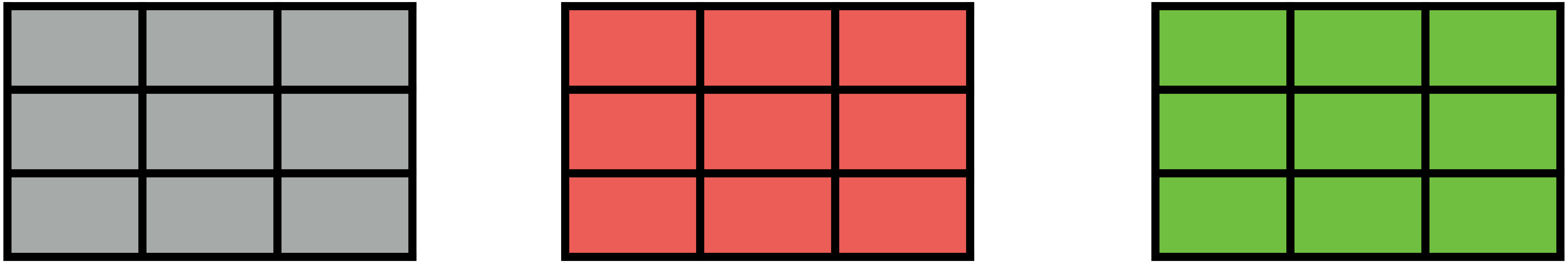
- the diagonal approximation has been used widely

$\hat{\Sigma}$ is an energy dependent operator

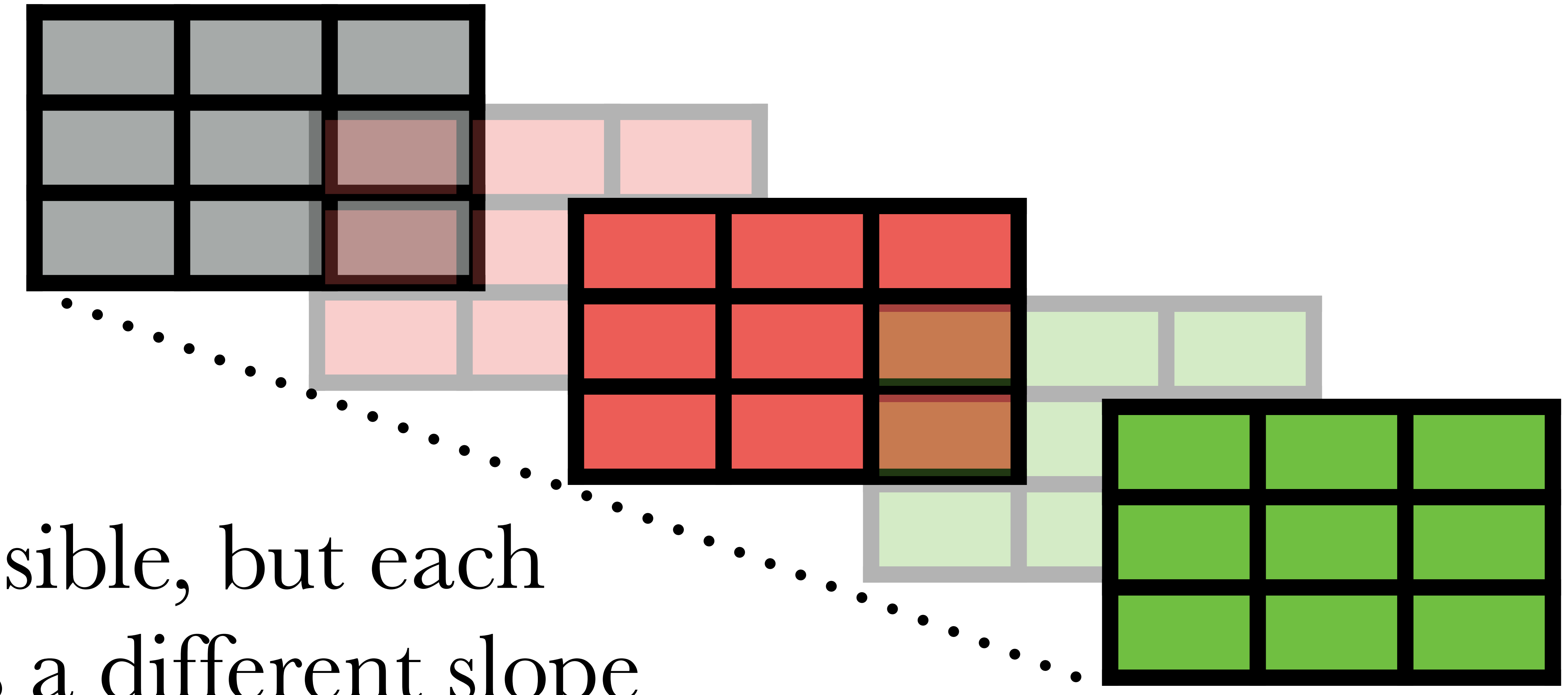


$\hat{\Sigma}(E = E_b)$

$\hat{\Sigma}$ is an energy dependent operator



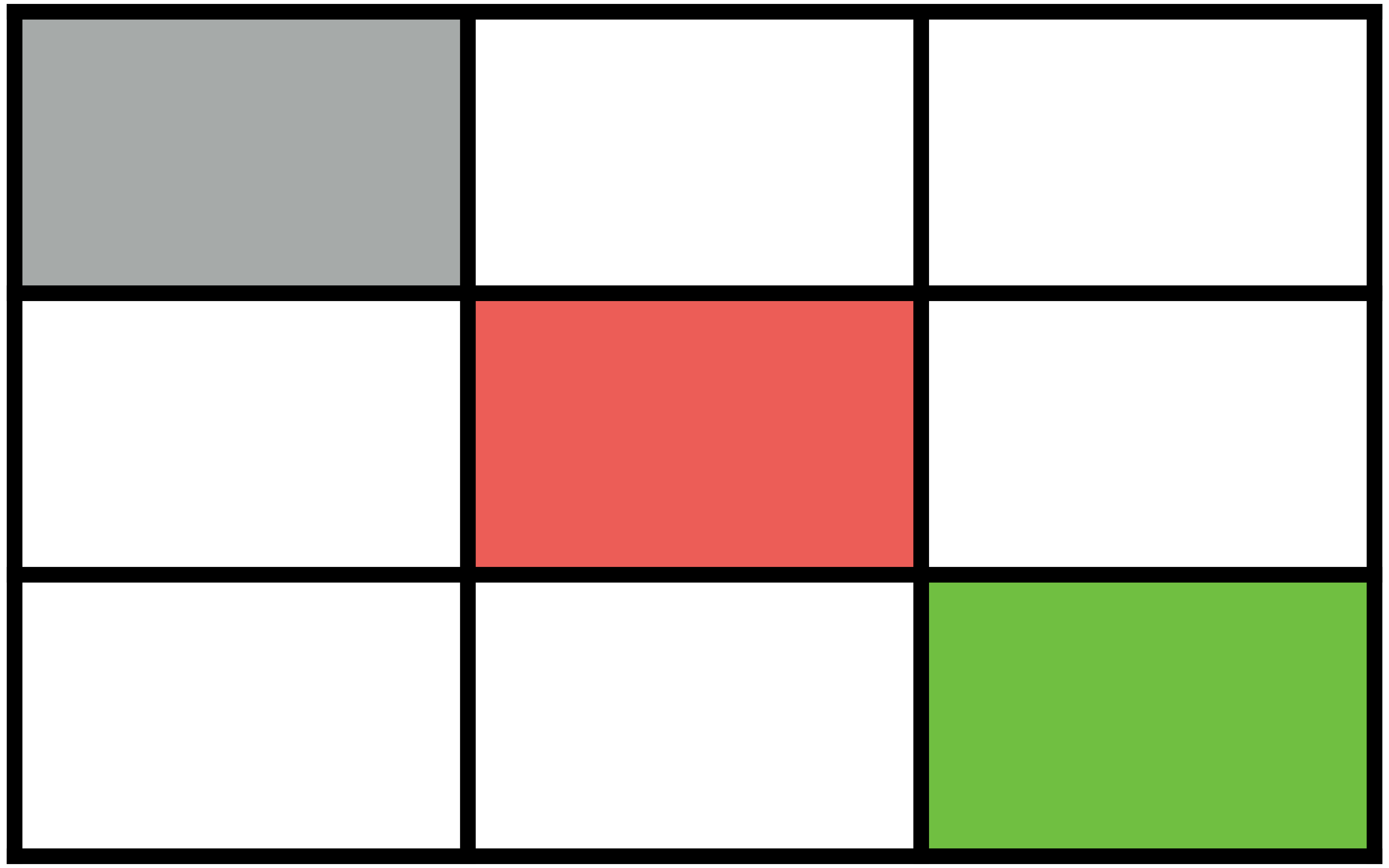
$\hat{\Sigma}(E = E_a)$ $\quad$ $\hat{\Sigma}(E = E_b)$ $\quad$ $\hat{\Sigma}(E = E_c)$

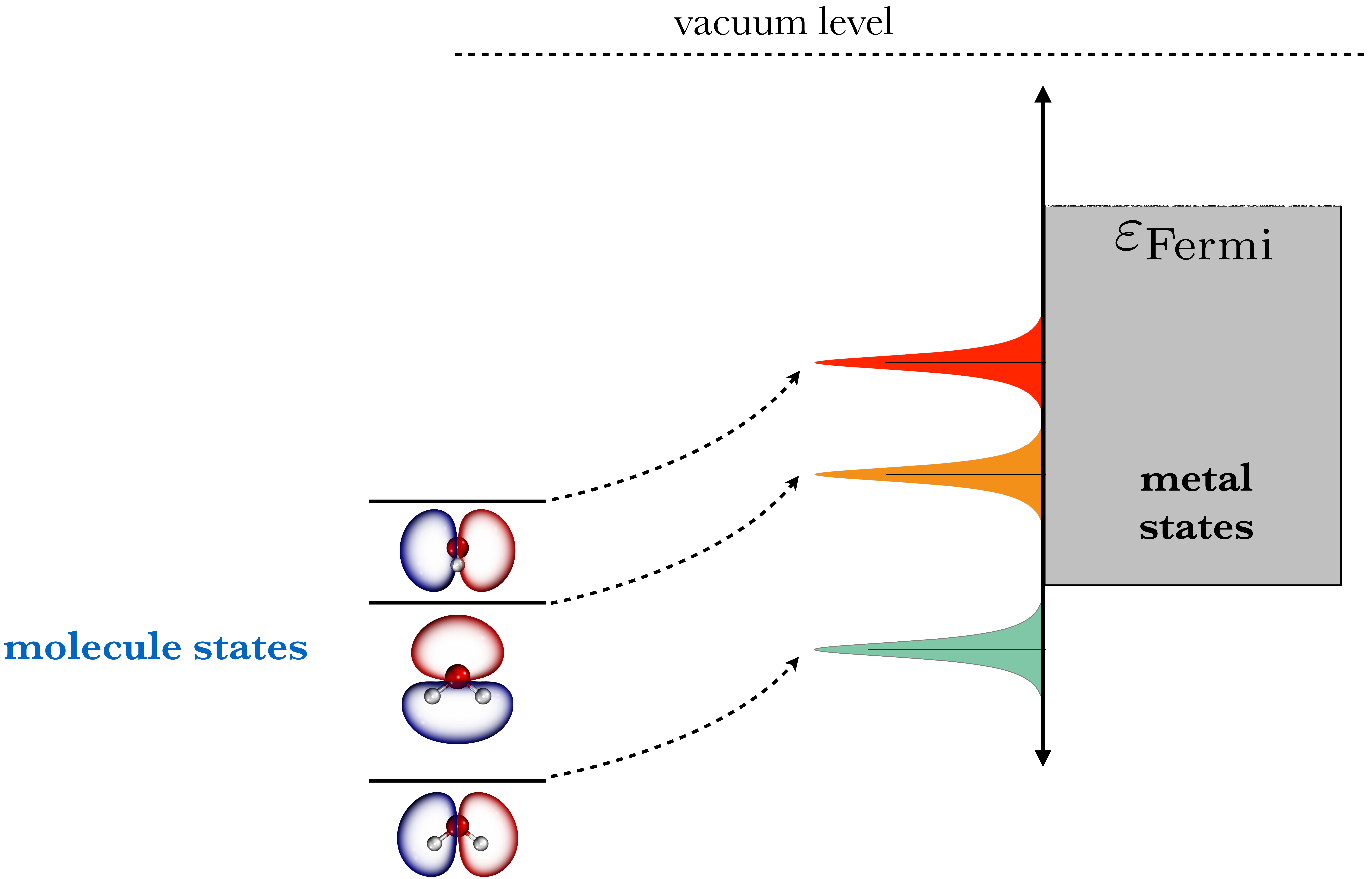


interpolation is possible, but each
matrix element has a different slope

$\hat{\Sigma}(E = E_{\mathrm{QP}})$

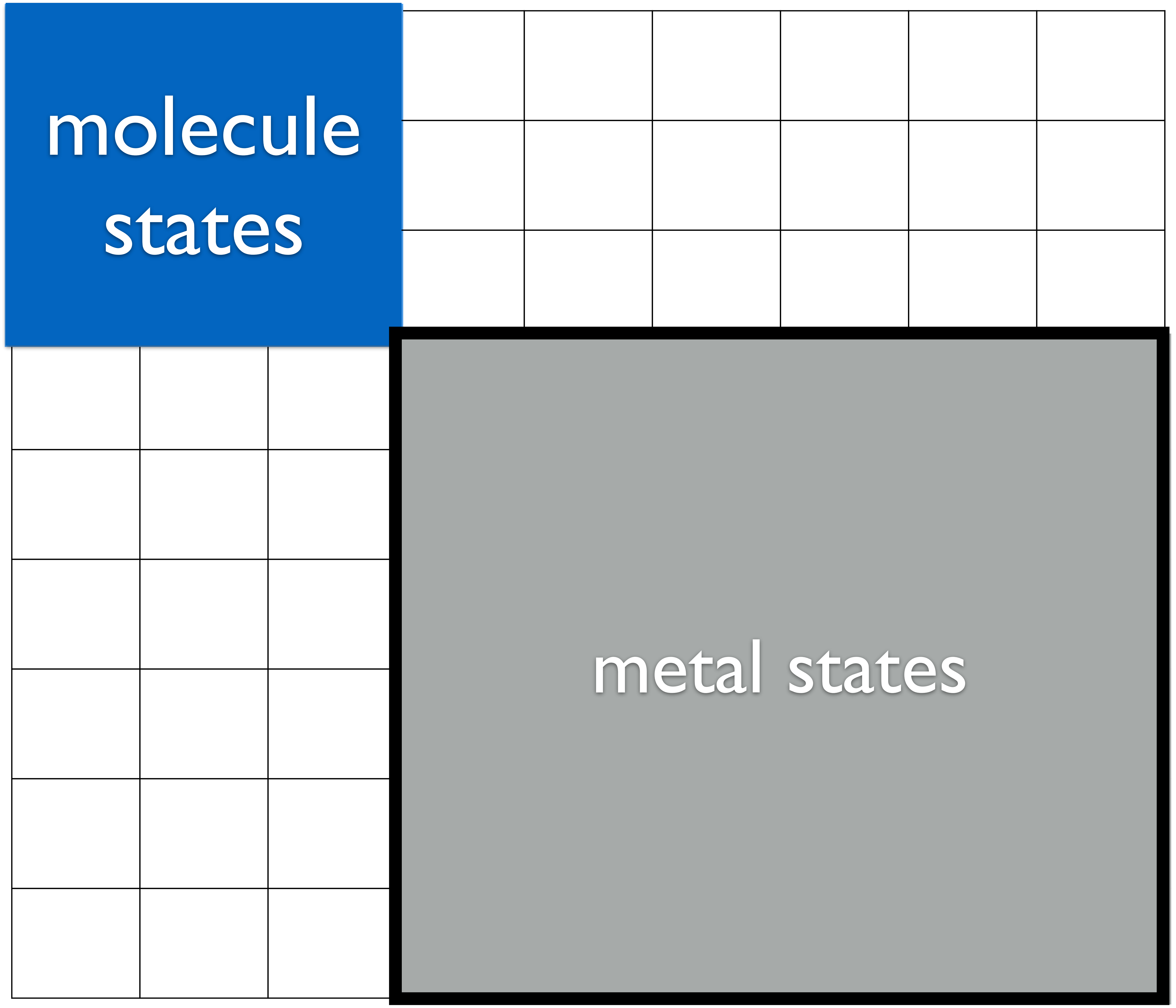
$$\text{If } |\Psi_{\mathrm{KS}}\rangle \approx |\Psi_{\mathrm{QP}}\rangle$$



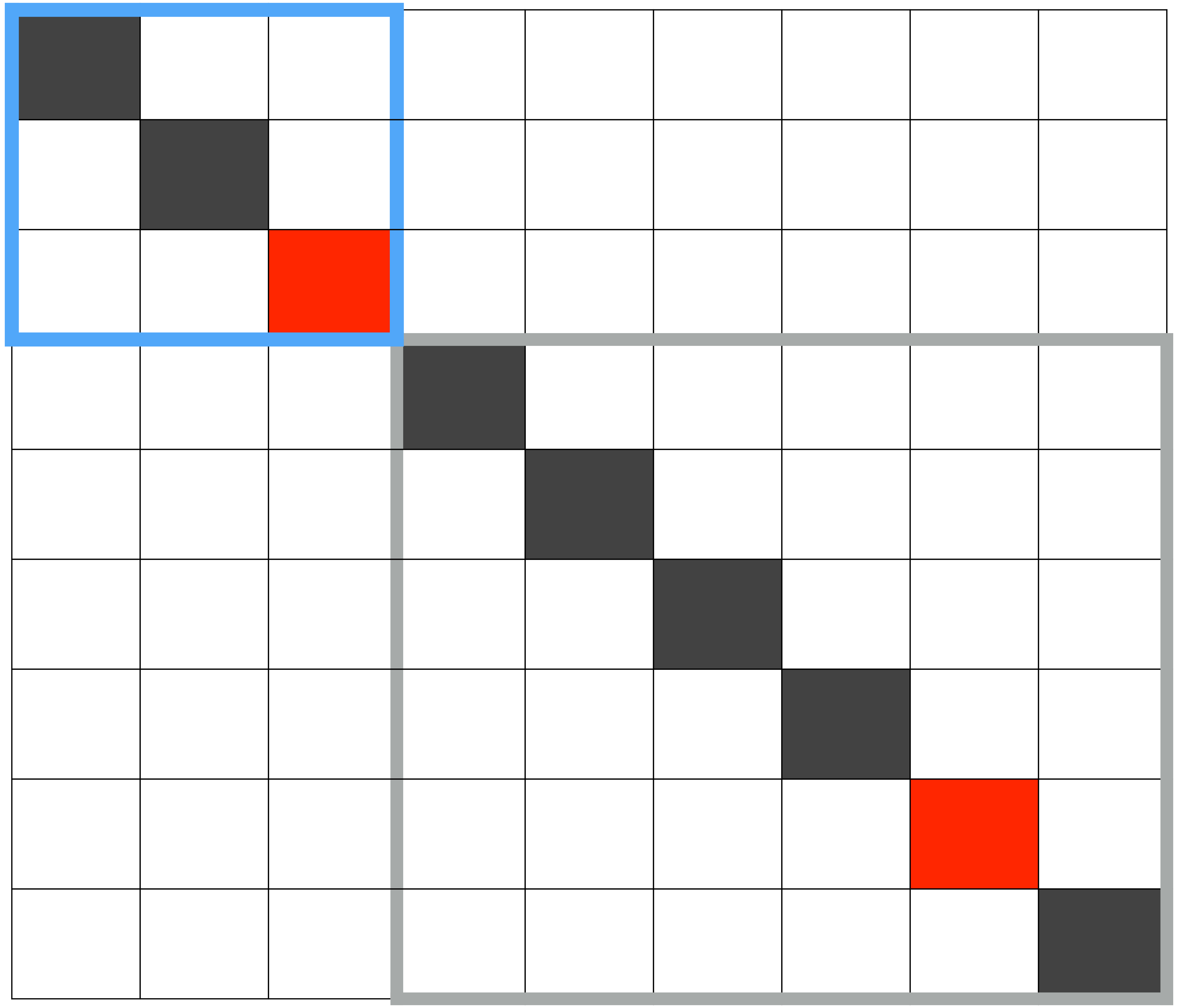


vacuum level
$\varepsilon_{\text{Fermi}}$
metal states
molecule states

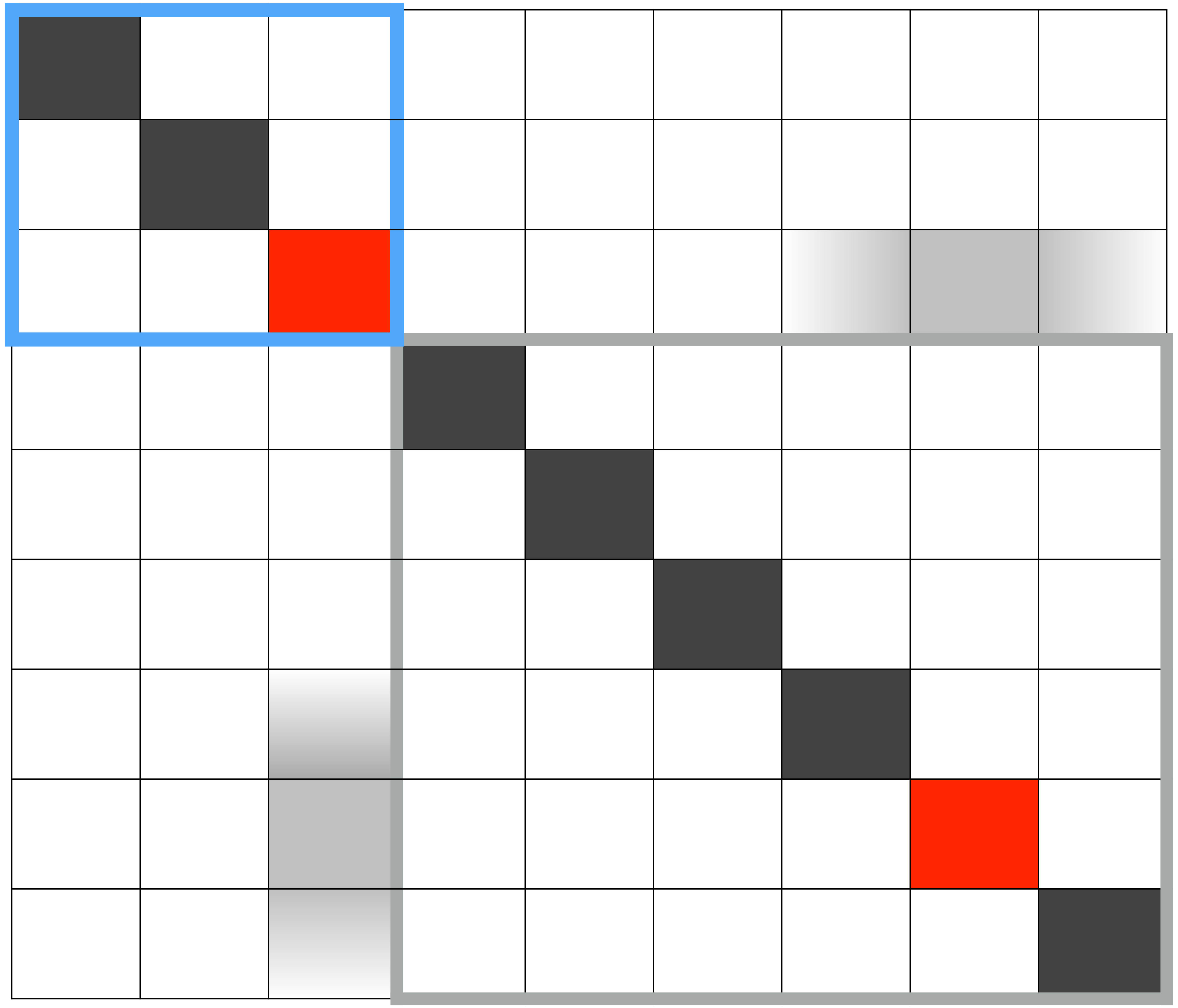
A new legacy



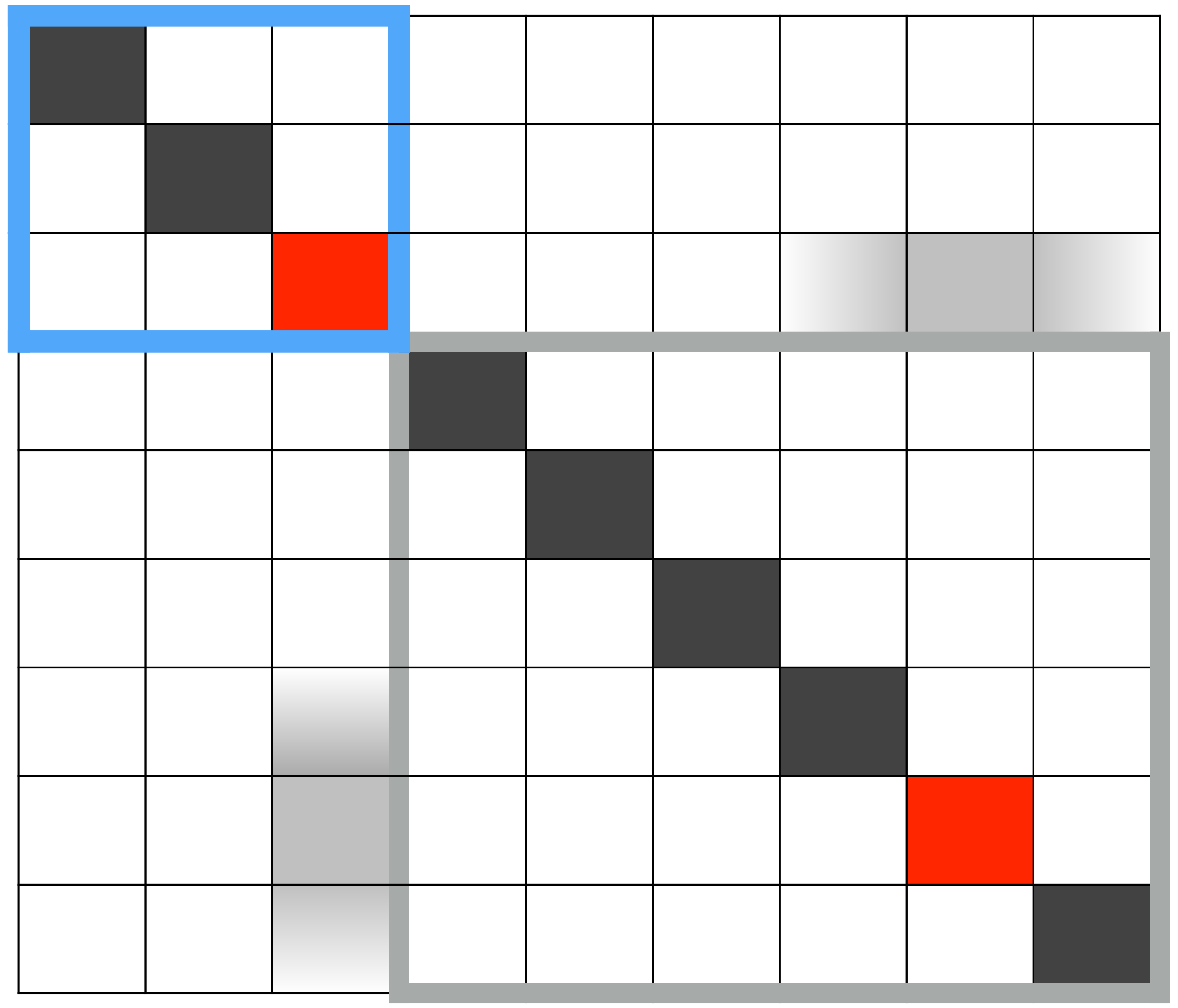
A new legacy



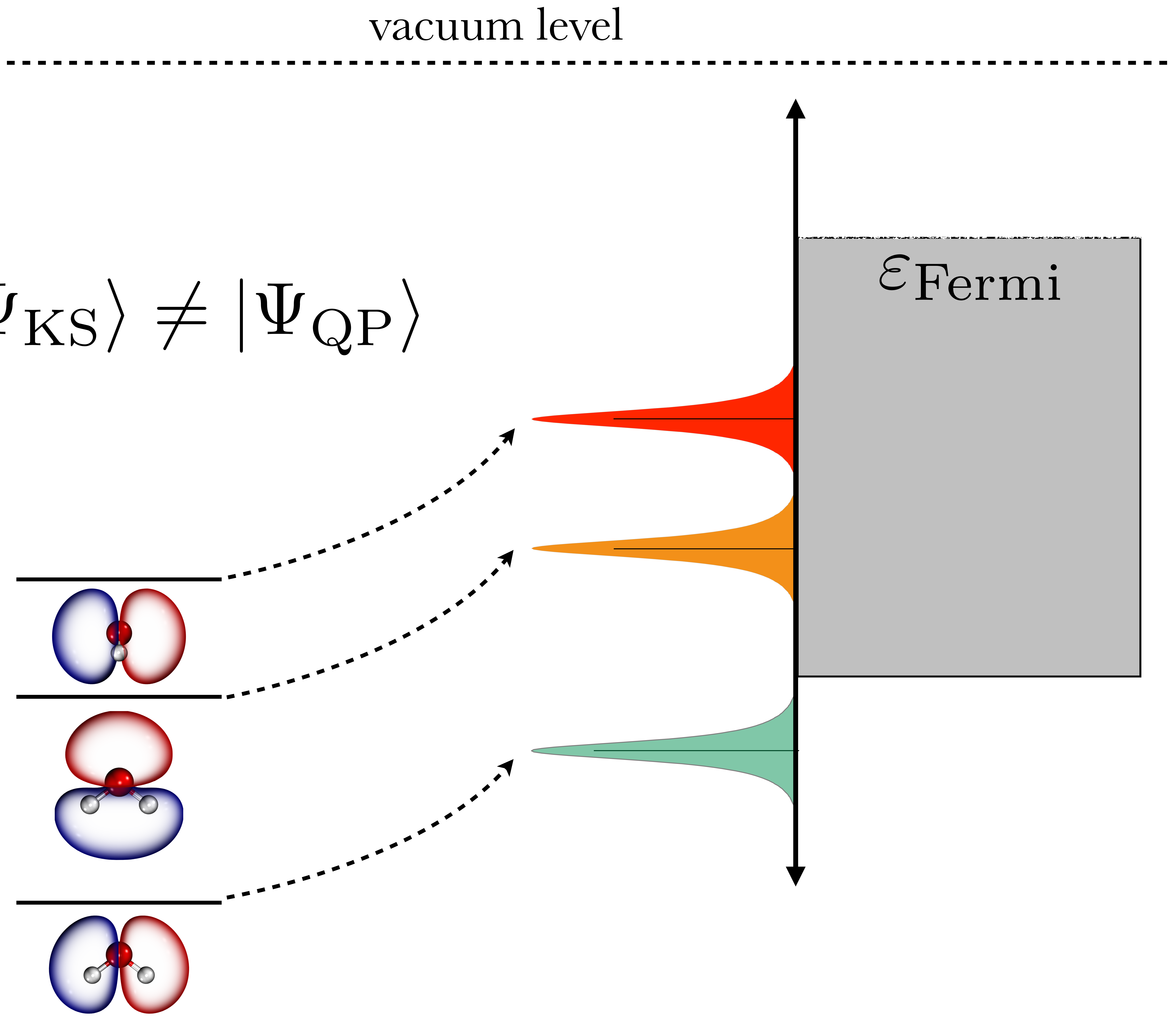
A new legacy




vacuum level
shallow MOs
DFT mixing
ε_{Fermi}



$$|\Psi_{\mathrm{KS}}\rangle \neq |\Psi_{\mathrm{QP}}\rangle$$



QP mixing

G.-M. Rignanese, *et al*., Phys. Rev. Lett. **86**, 2110 (2001)

R. Shaltaf *et al*., Phys. Rev. Lett. **100**, 186401 (2008)

I. Tamblyn *et al*., Phys. Rev. B 84, 201402 (2011)

When $|\Psi_{\mathrm{KS}}\rangle \neq |\Psi_{\mathrm{QP}}\rangle$, $\hat{\Sigma}$ is not diagonal

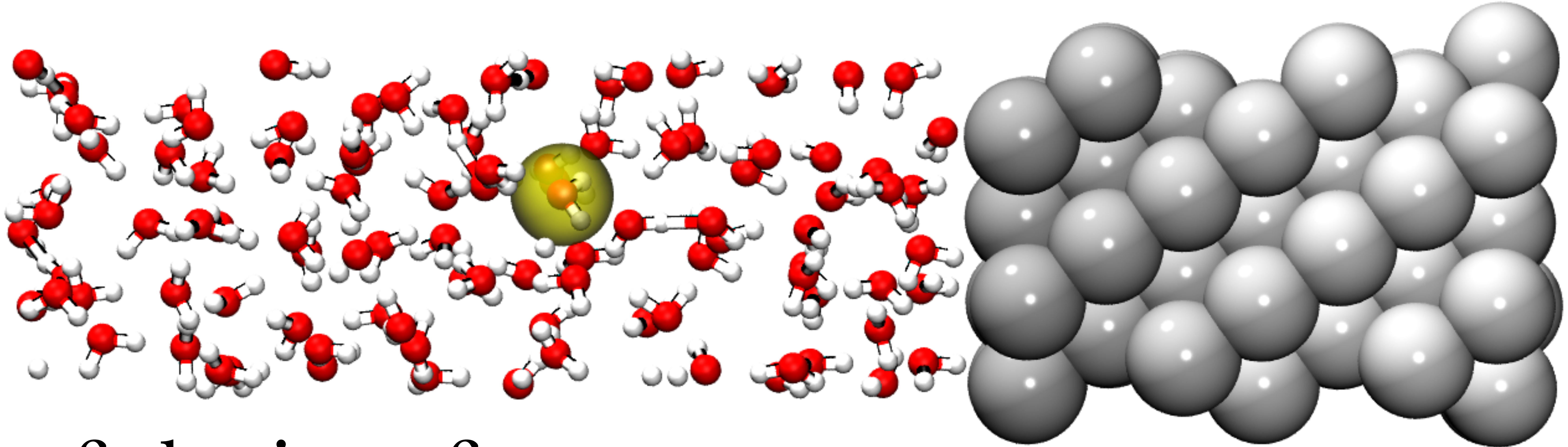
in the KS basis

It may be nearly diagonal in some other basis however…

Gas phase*, molecular orbitals $|\iota\rangle$

$\langle\iota|\hat{H}_K|\iota\rangle =$ (units of eV)

	$\mathbf{1b_2}$	$\mathbf{3a_1}$	$\mathbf{1b_1}$
$\mathbf{1b_2}$	-7.79	0.05	0.01
$\mathbf{3a_1}$	0.05	-4.71	0.01
$\mathbf{1b_1}$	0.01	0.01	-2.20



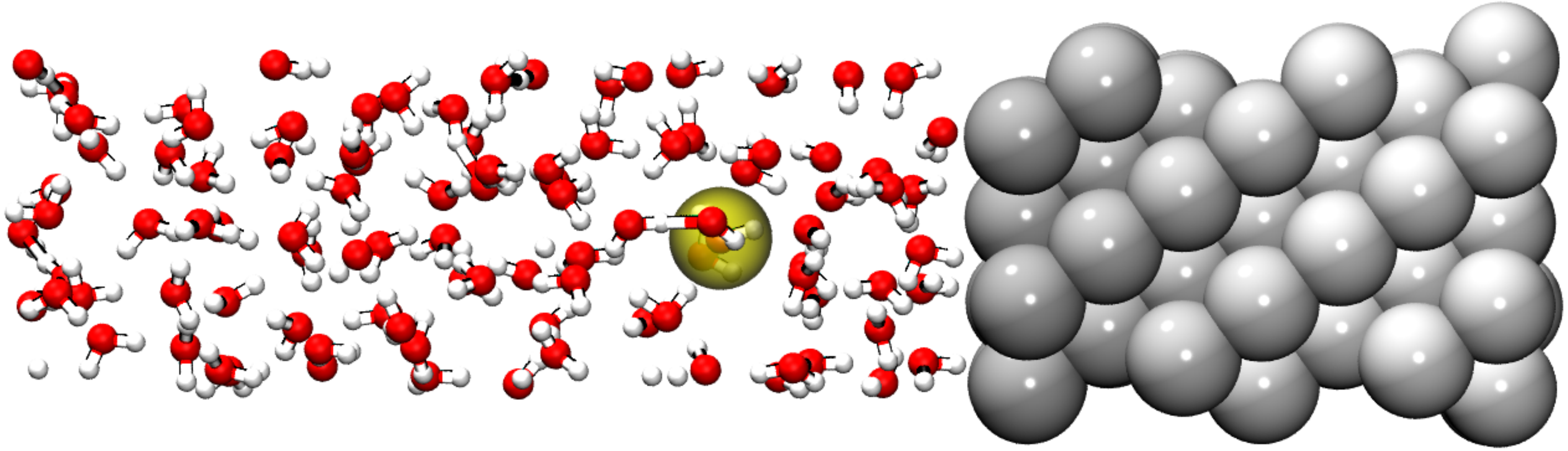
$\hat{H}_K$ = Hamiltonian of the interface

Gas phase*, molecular orbitals $|\iota\rangle$

$$\langle \iota | \hat{H}_K | \iota \rangle =$$

(units of eV)

	1*b*₂	**3*a*₁**	**1*b*₁**
1*b*₂	-7.61	0.04	0.05
3*a*₁	0.04	-3.64	0.17
1*b*₁	0.05	0.17	-1.70

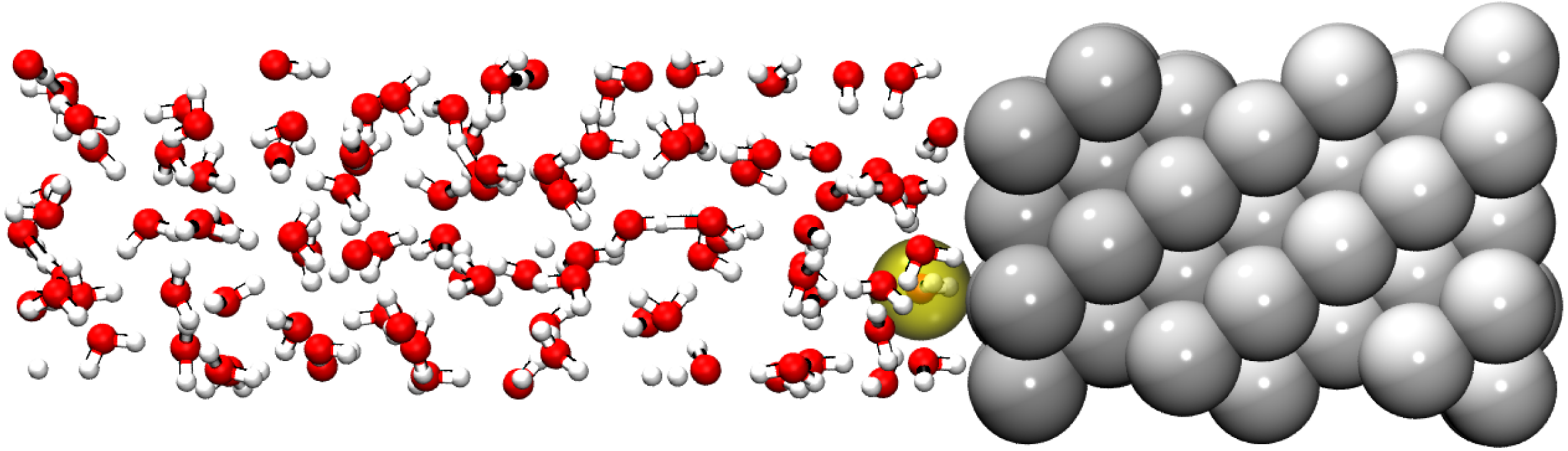


Gas phase*, molecular orbitals $|\iota\rangle$

$$\langle \iota | \hat{H}_K | \iota \rangle =$$

(units of eV)

$\mathbf{1b_2}$	$\mathbf{3a_1}$	$\mathbf{1b_1}$	
-8.09	0.01	0.01	$\mathbf{1b_2}$
0.01	-5.02	0.01	$\mathbf{3a_1}$
0.01	0.01	-3.18	$\mathbf{1b_1}$

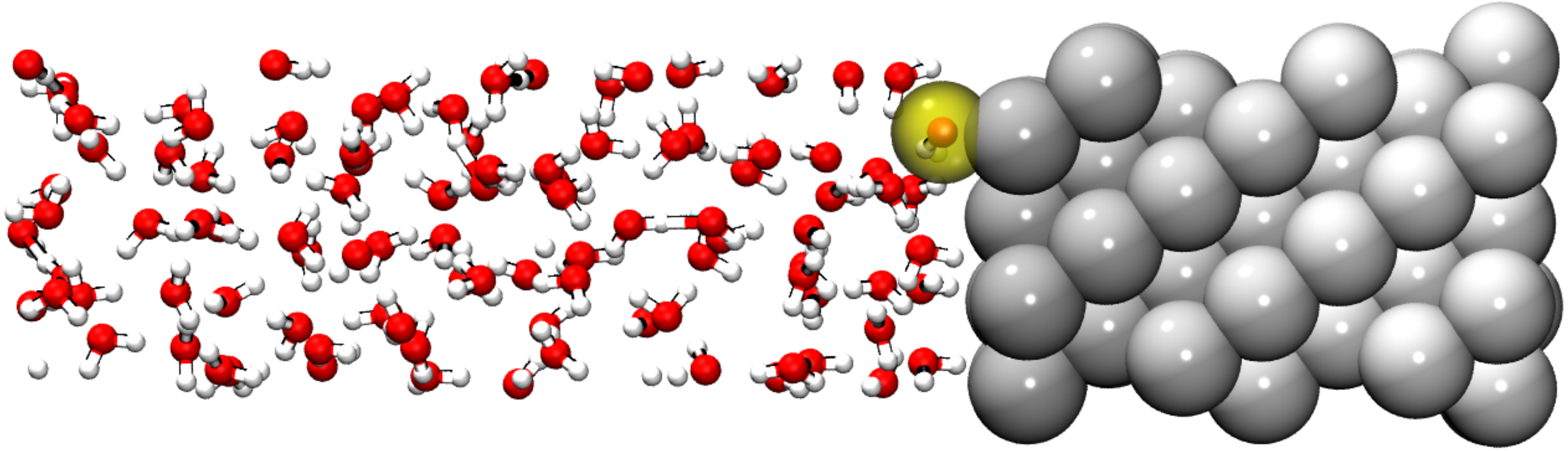


Gas phase*, molecular orbitals $|\iota\rangle$

$$\langle\iota|\hat{H}_K|\iota\rangle = \begin{array}{c} \\ \end{array}$$

(units of eV)

	$\mathbf{1}b_2$	$\mathbf{3}a_1$	$\mathbf{1}b_1$
$\mathbf{1}b_2$	-7.69	0.02	0.11
$\mathbf{3}a_1$	0.02	-4.92	0.35
$\mathbf{1}b_1$	0.11	0.35	-3.09



1^{st} year QM

For each molecule, compute $|\iota\rangle$

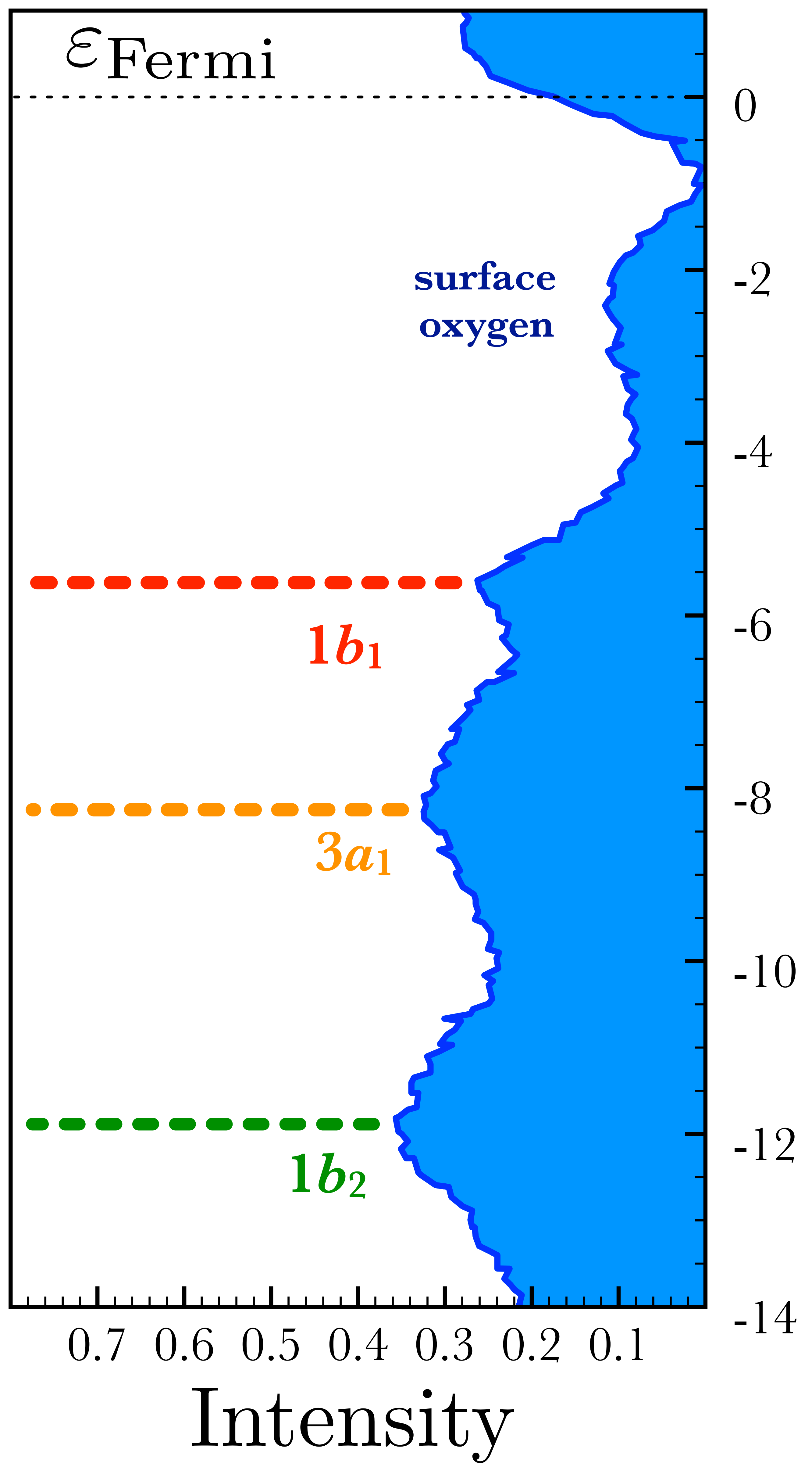
then projectors, $\langle\iota|\kappa\rangle$ for <u>many</u> KS states,

(I used ~ 3,000)

then the usual

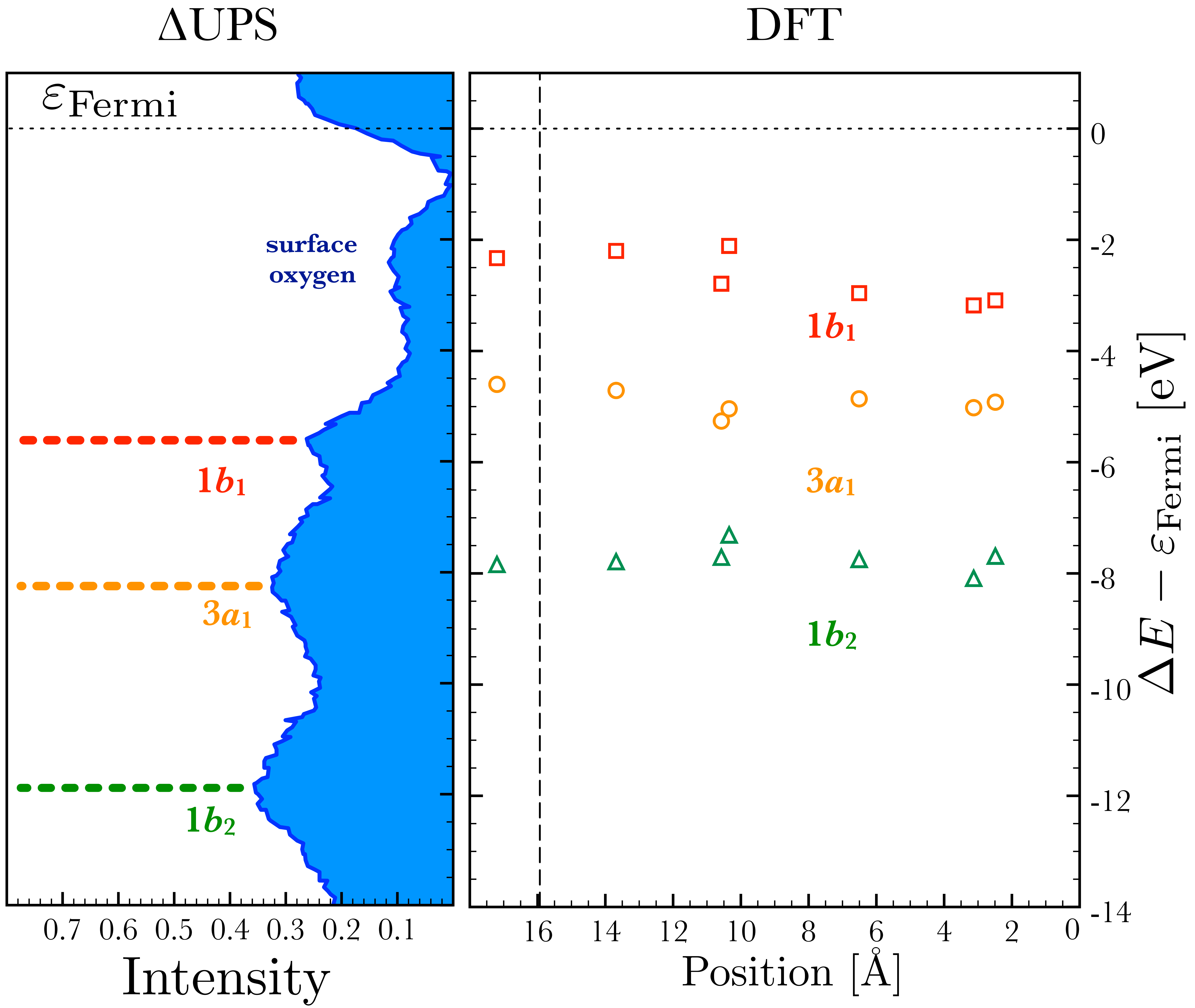
$$\langle\iota|\hat{O}|\iota\rangle = \langle\iota|\kappa\rangle\langle\kappa|\hat{O}|\kappa\rangle\langle\kappa|\iota\rangle$$

ΔUPS



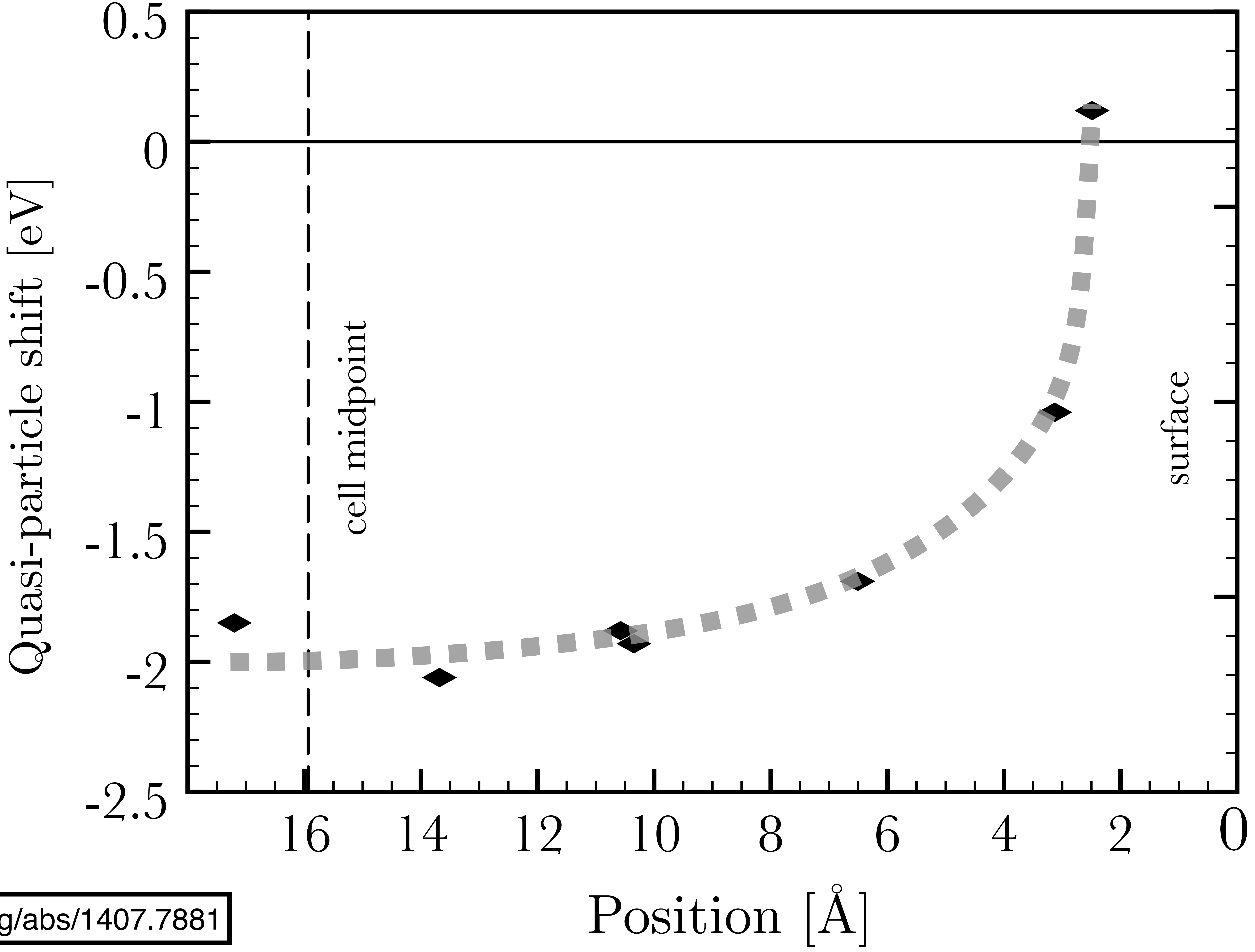
Experiment: Fisher et al., PRL, 44, 683 (1980)

SURFACE



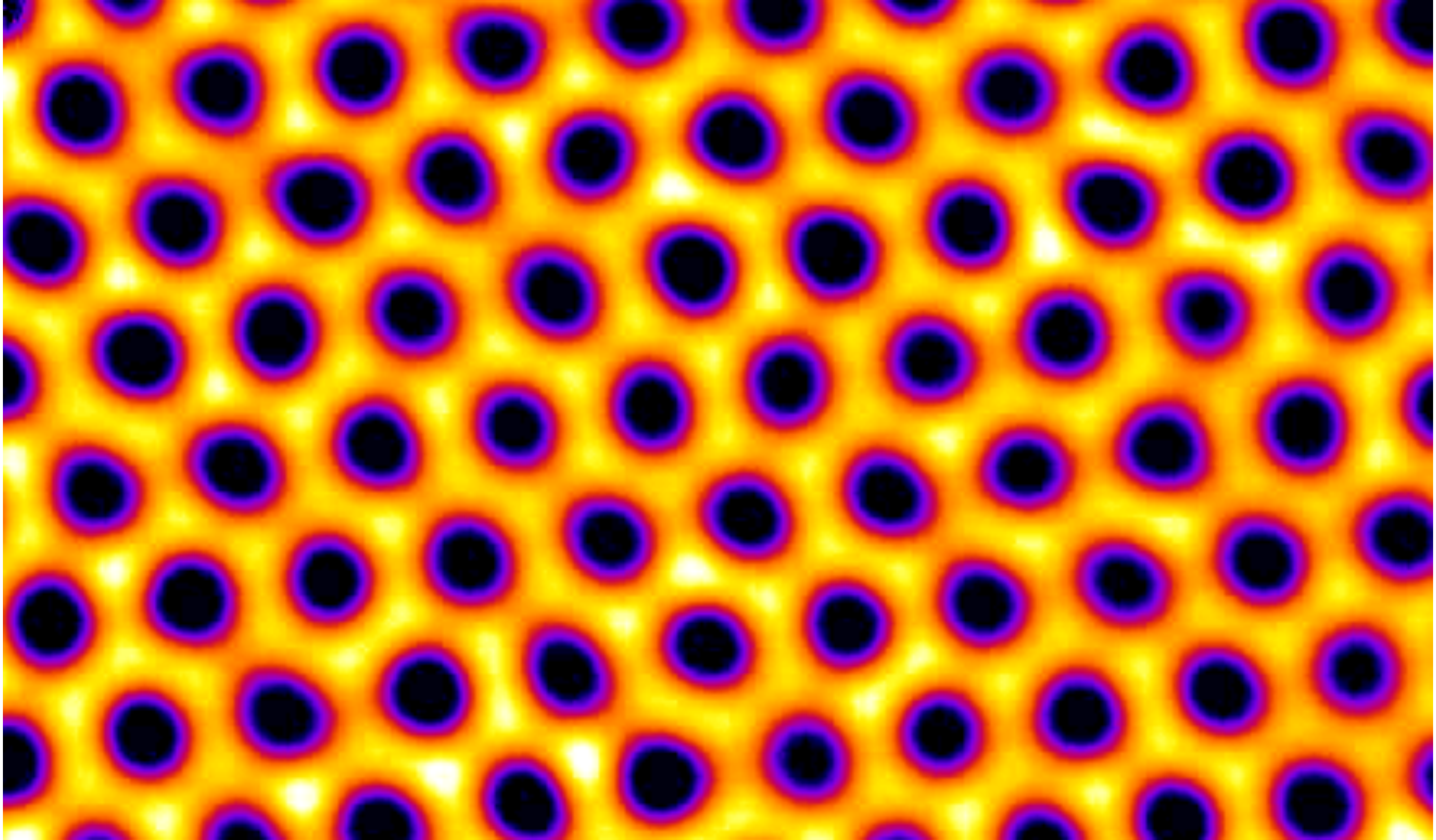
Experiment: Fisher et al., PRL, 44, 683 (1980)

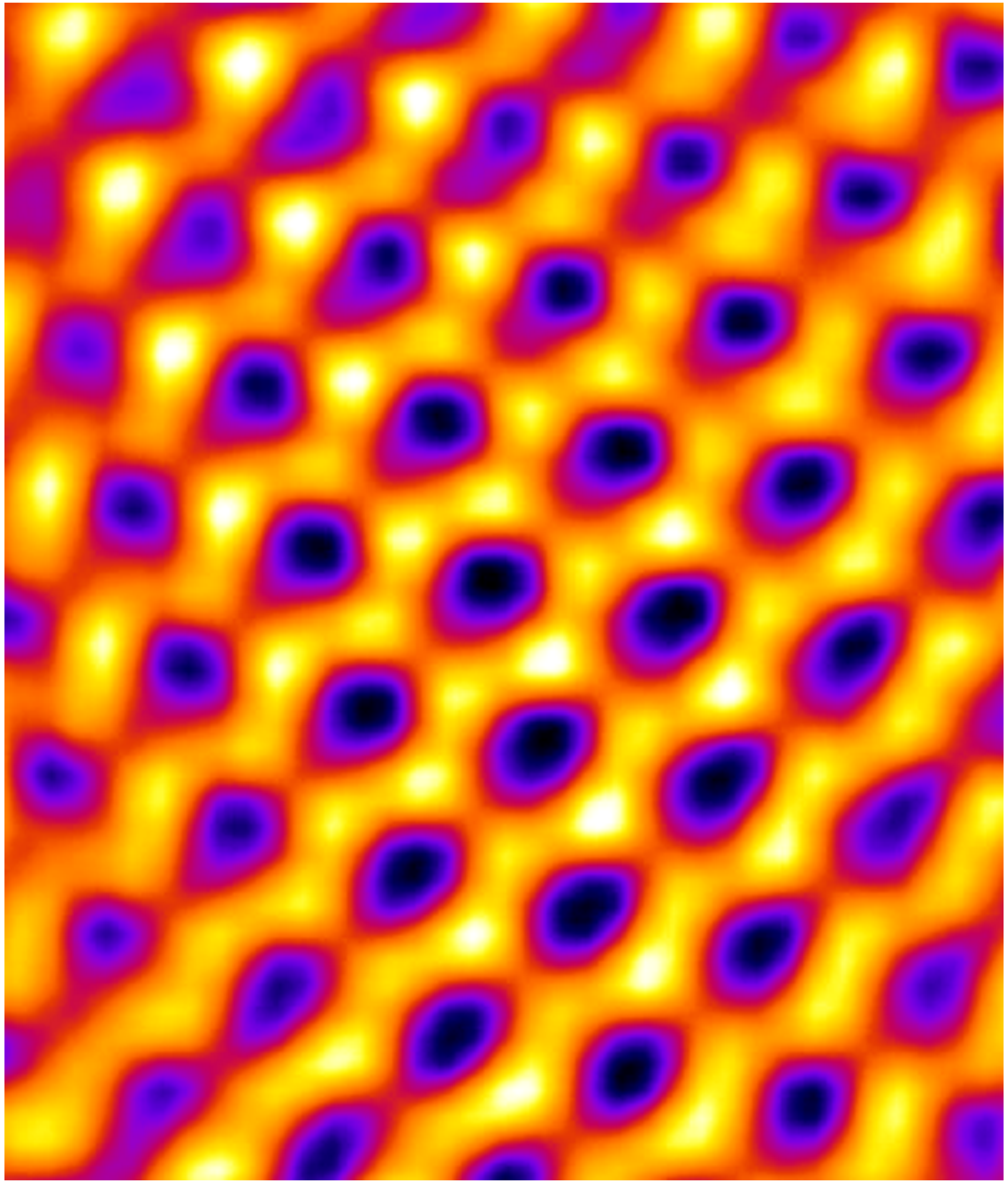
Experiment: Fisher et al., PRL, 44, 683 (1980)

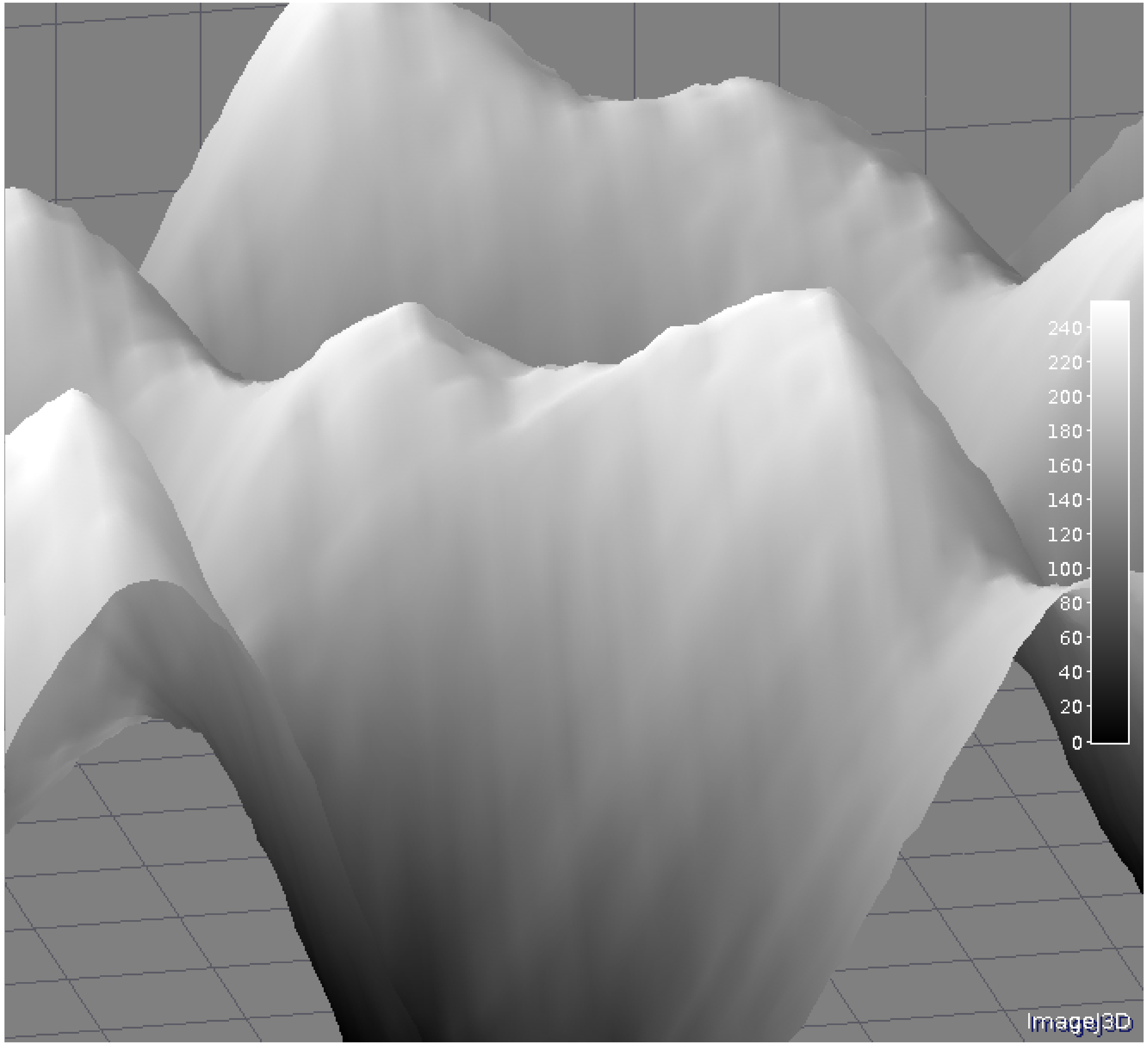


http://arxiv.org/abs/1407.7881

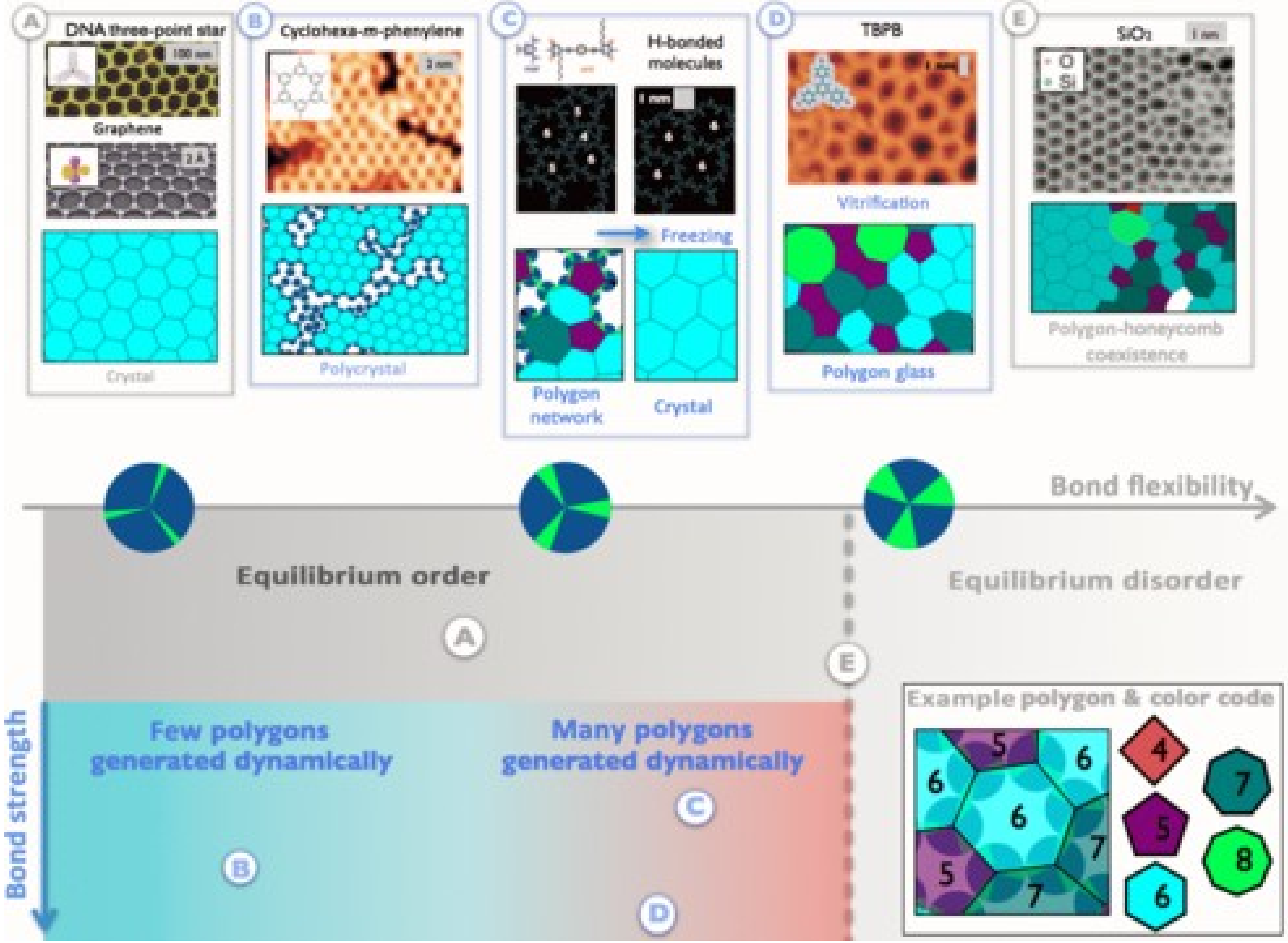
- GW significantly improves agreement with experiment for the water/Pt interface
- DFT is *qualitatively* incorrect for level alignment (and its spatial dependence)
- A judicious choice of basis is necessary to handle nanoscale systems
- Current focus is on <u>doping and defects</u>



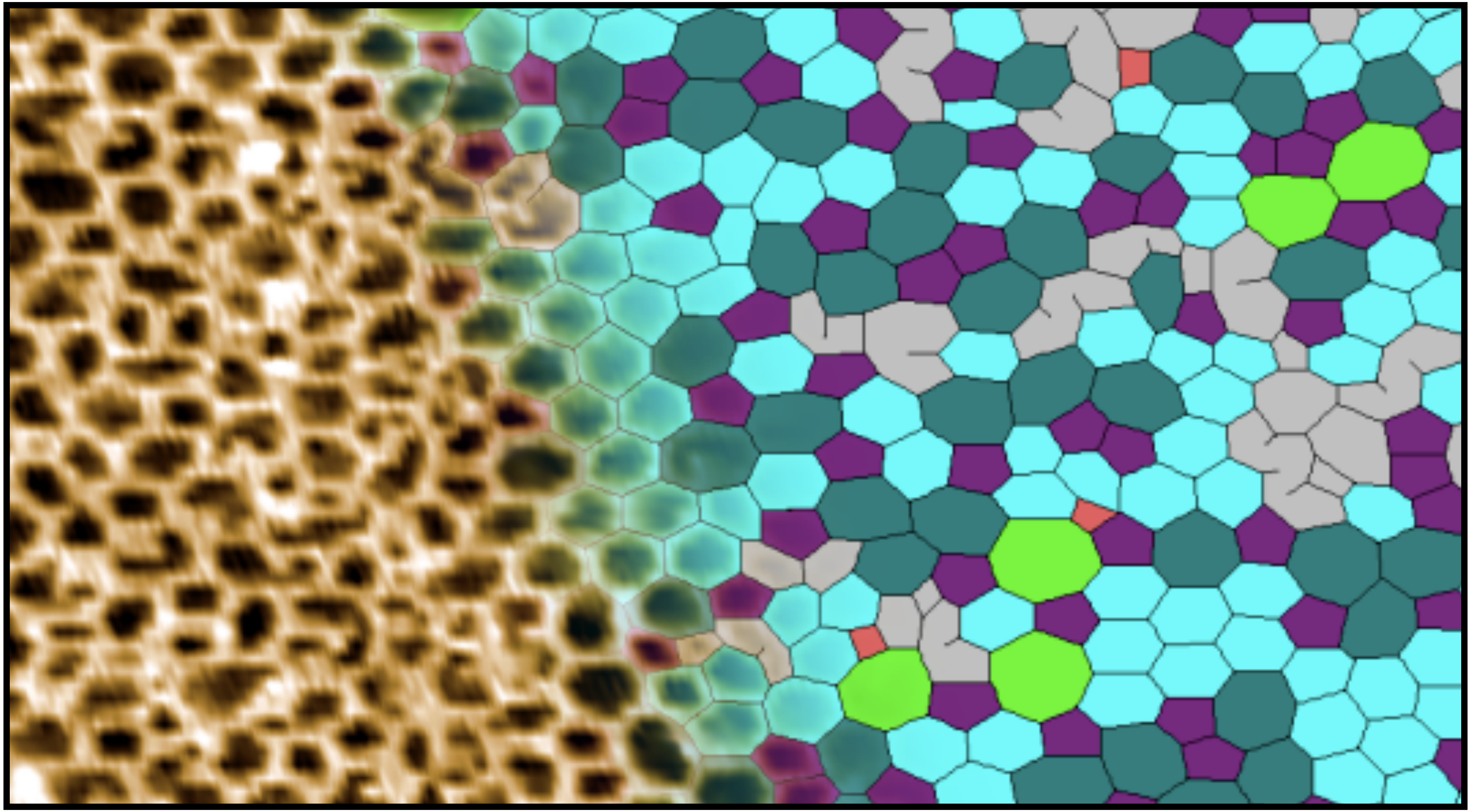




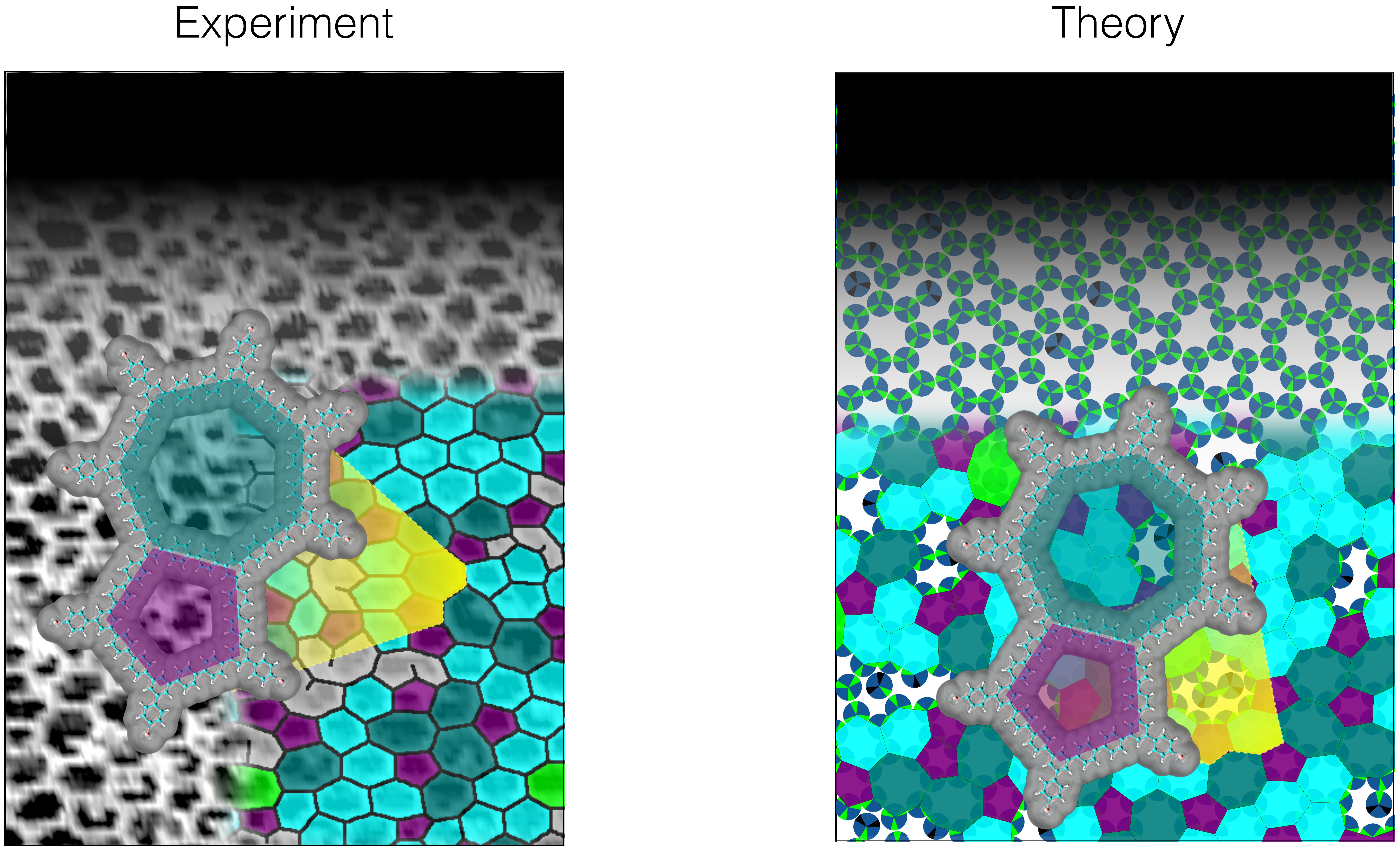
Aberration Corrected Graphene
britishcarbon.org/images.shtml



Whitelam, Tamblyn, Beton, and Garrahan, *Phys. Rev. Lett.* 114, 115702 (2015)
Whitelam, Tamblyn, *et al.*, *Phys. Rev. X*, 3, (2014)
Whitelam, Tamblyn, Beton, and Garrahan, *Phys. Rev. Lett.* 108, 035702 (2012)



Whitelam, Tamblyn, Beton, and Garrahan, *Phys. Rev. Lett.* 114, 115702 (2015)
Whitelam, Tamblyn, *et al.*, *Phys. Rev. X*, 3, (2014)
Whitelam, Tamblyn, Beton, and Garrahan, *Phys. Rev. Lett.* 108, 035702 (2012)

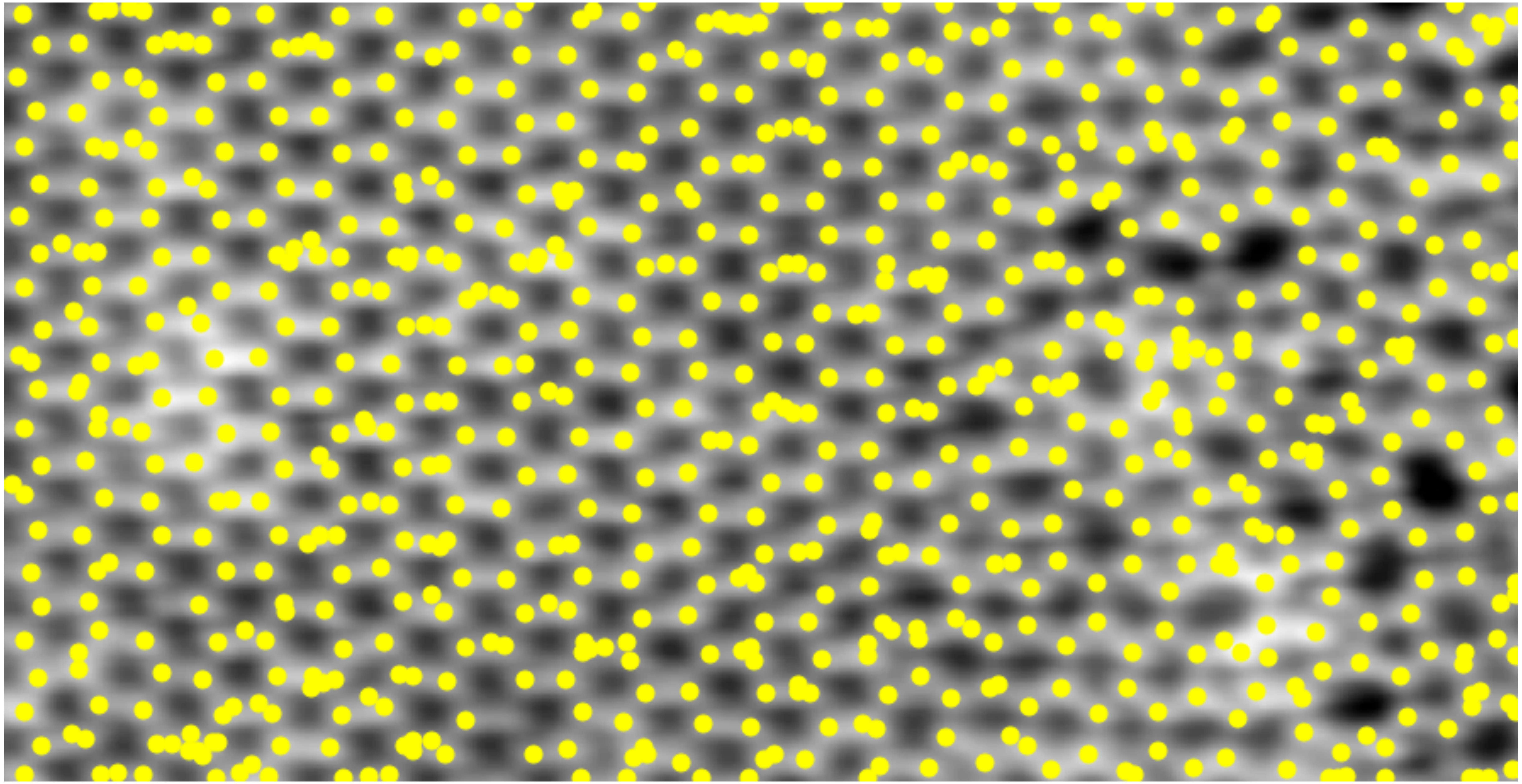

Experiment
Theory

Automate extraction

Stage 1: Initial Detections
Stage 2: Find and fix fractures and incorrect detections
Stage 3: Find and fix Merged detections

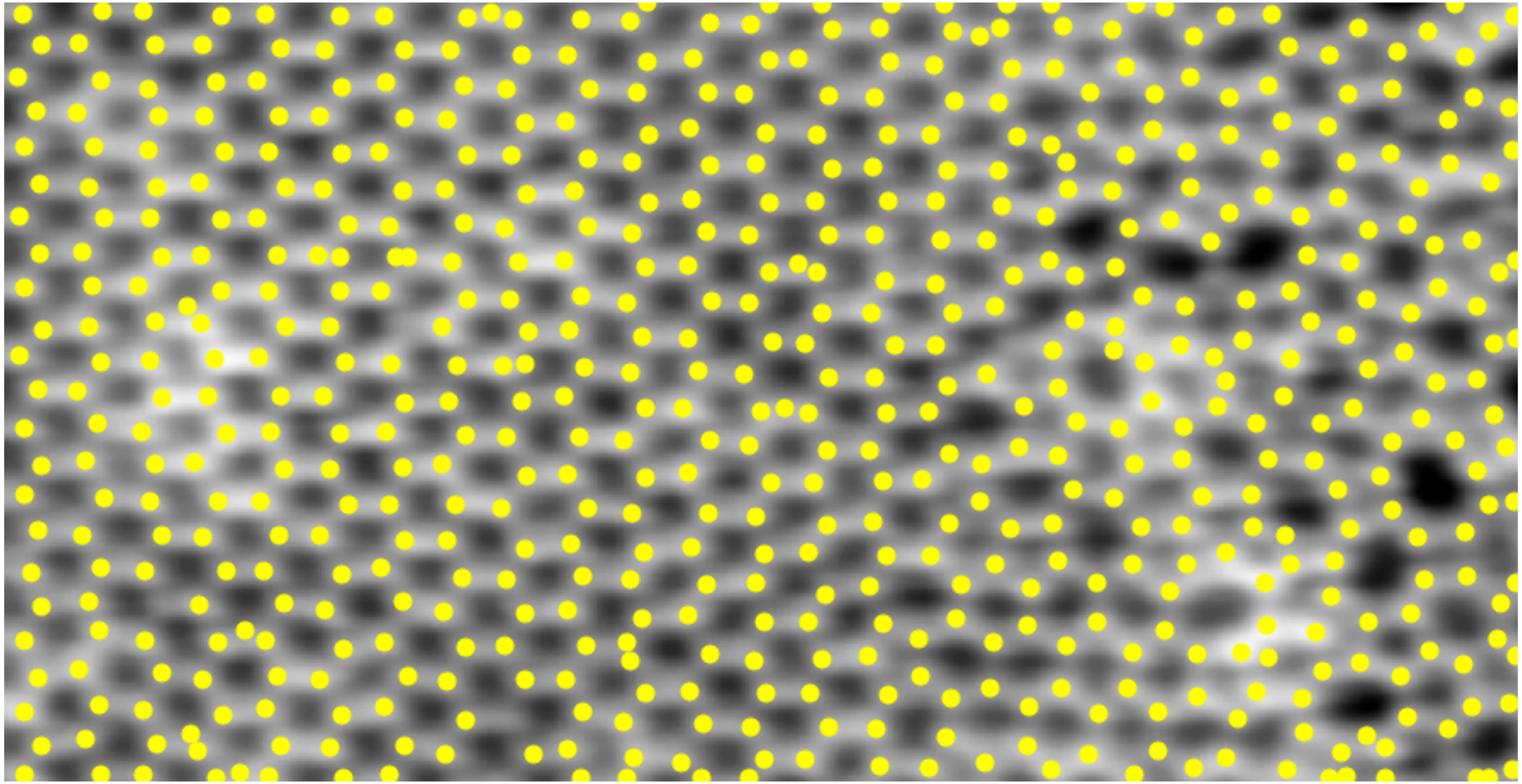
Stage 1 - Initial Detection

Find curvature of the image at each pixel.
Areas of negative curvature correspond to areas of atomic charge
Find local maxima
Let this be our initial set of potential atoms within the image.



Stage 2 - Fixing Fracture and Incorrect Detections

User defined length scale
Find set of pairs within one half of this distance.
Iterate through pairs and either remove a potential atom, if its blob of negative curvature is significantly smaller than its pair, or merge them together if the blobs are of similar size



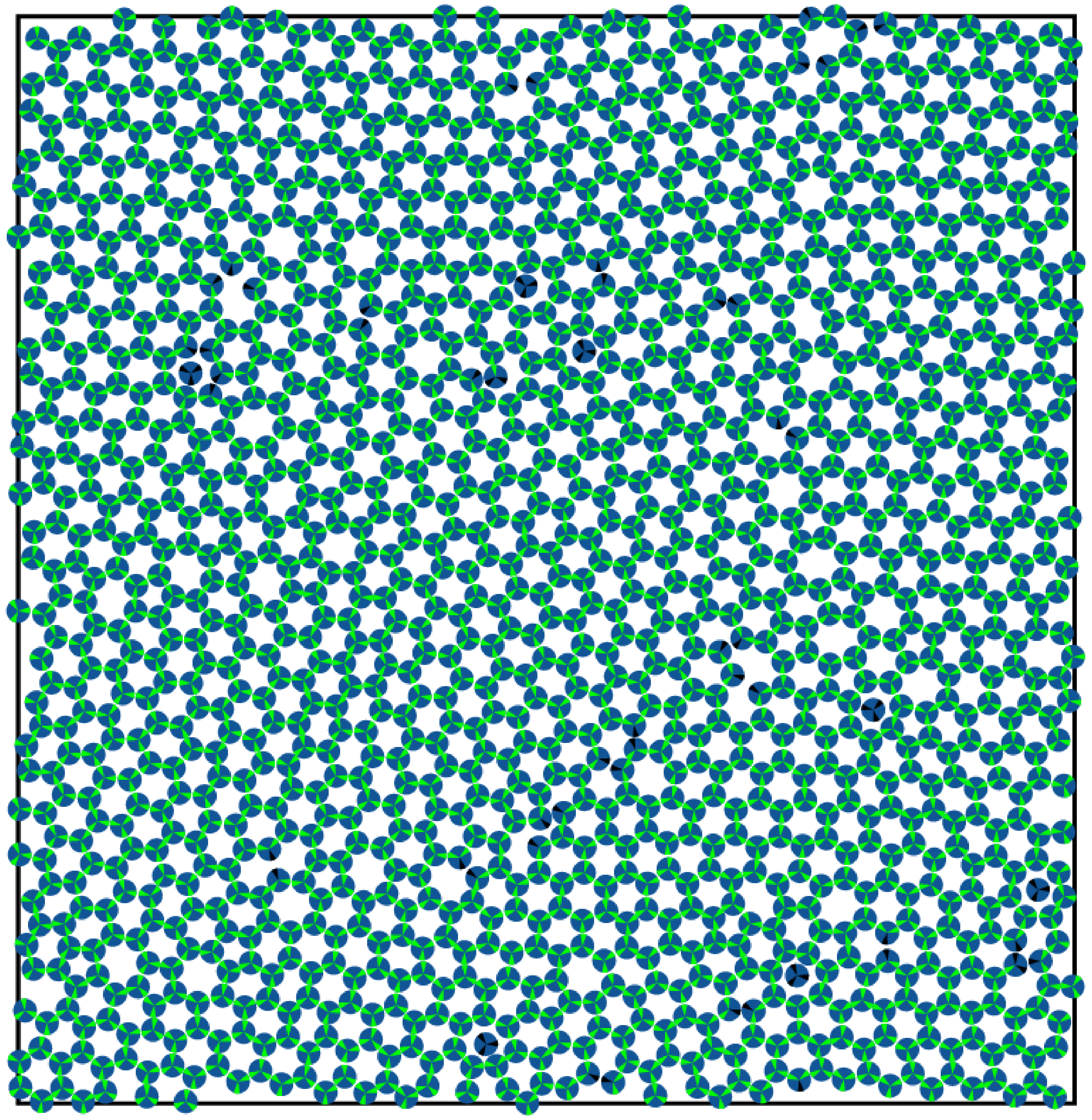
Stage 3 - Find and fix merged detections

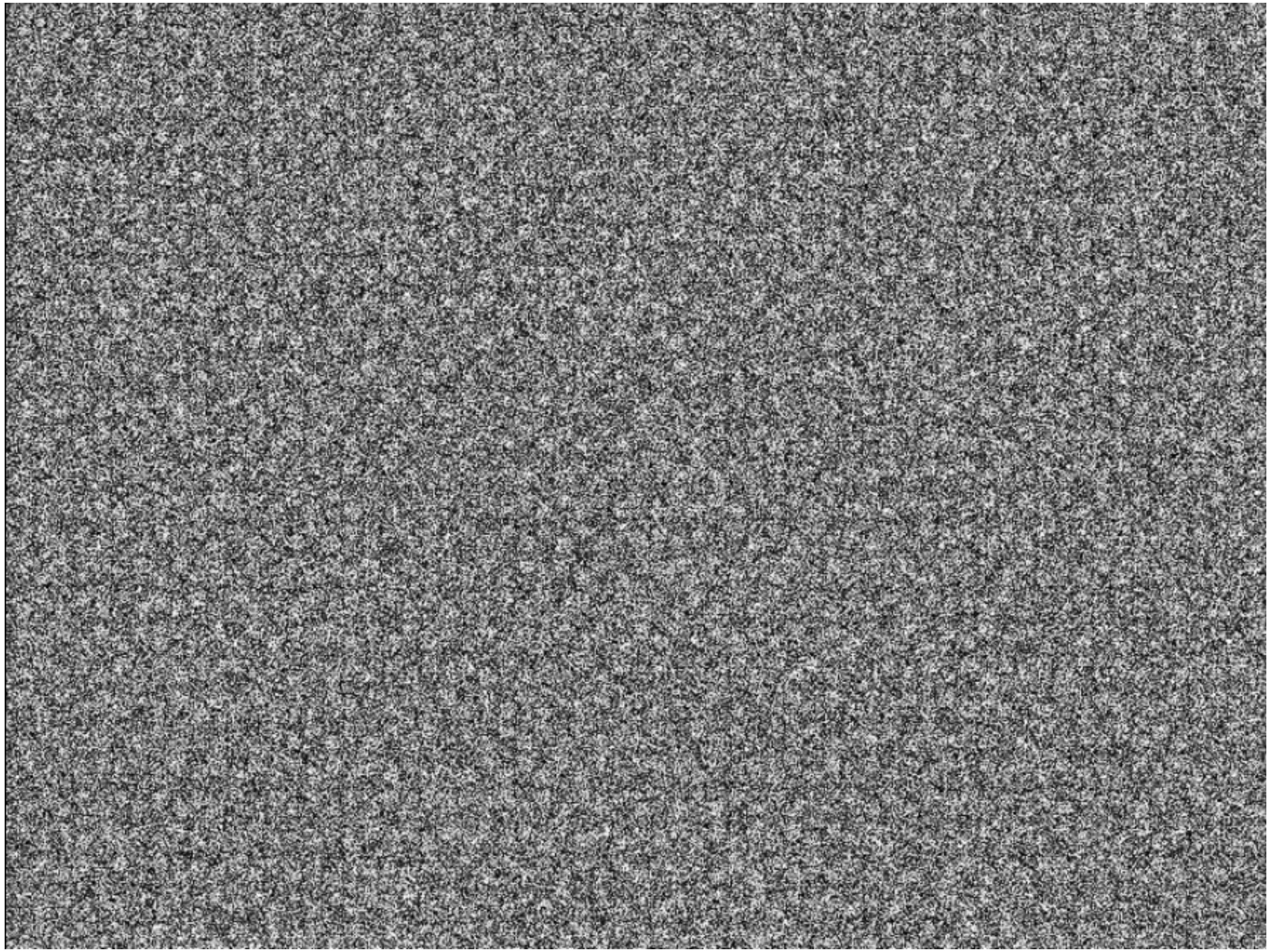
Fit ellipses to negative curvature regions that represent potential atoms

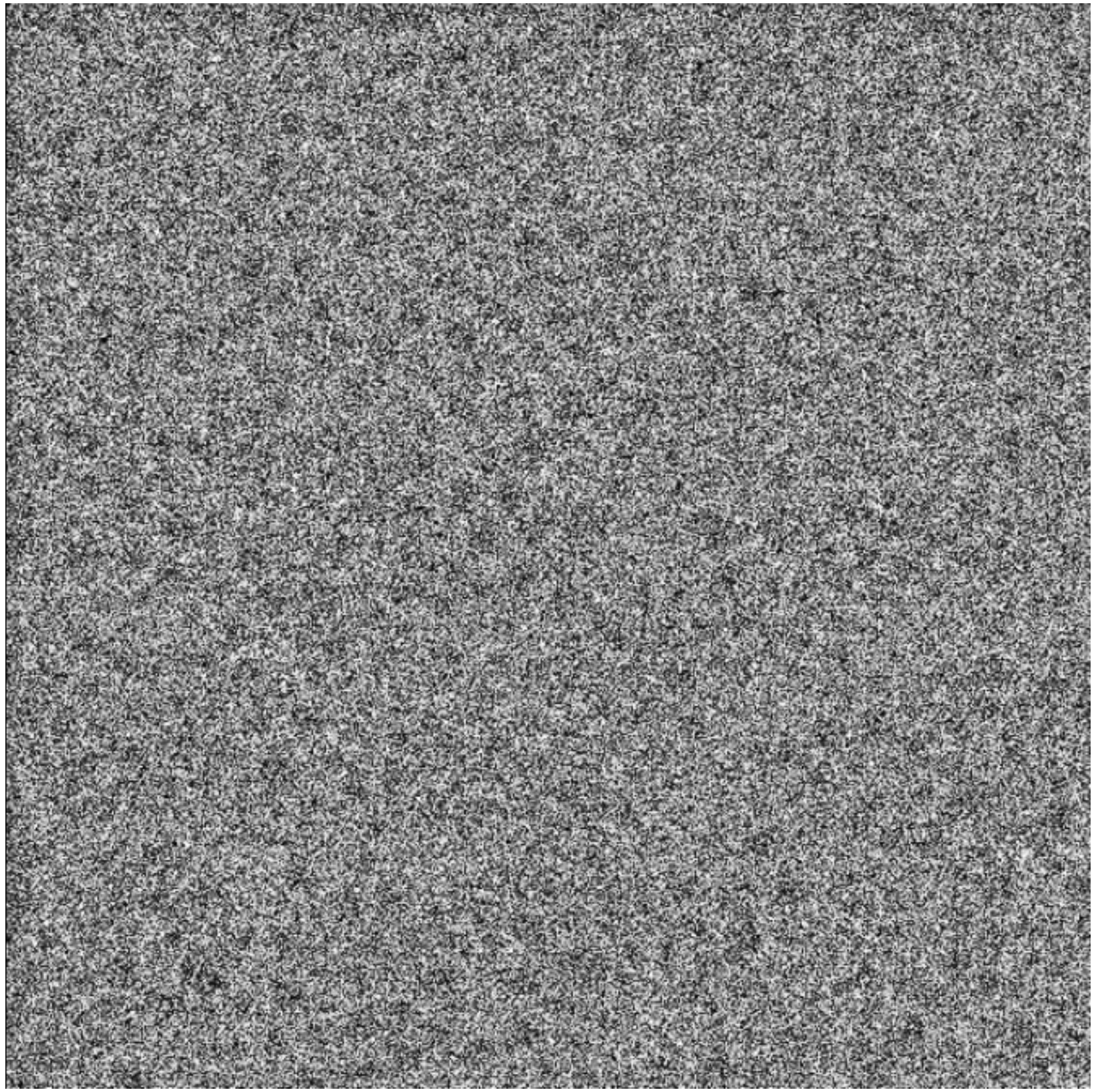
Test ellipses:

- eccentricity of the fitted ellipse
- size of the negative area
- distance between maxima of blob and center of ellipse

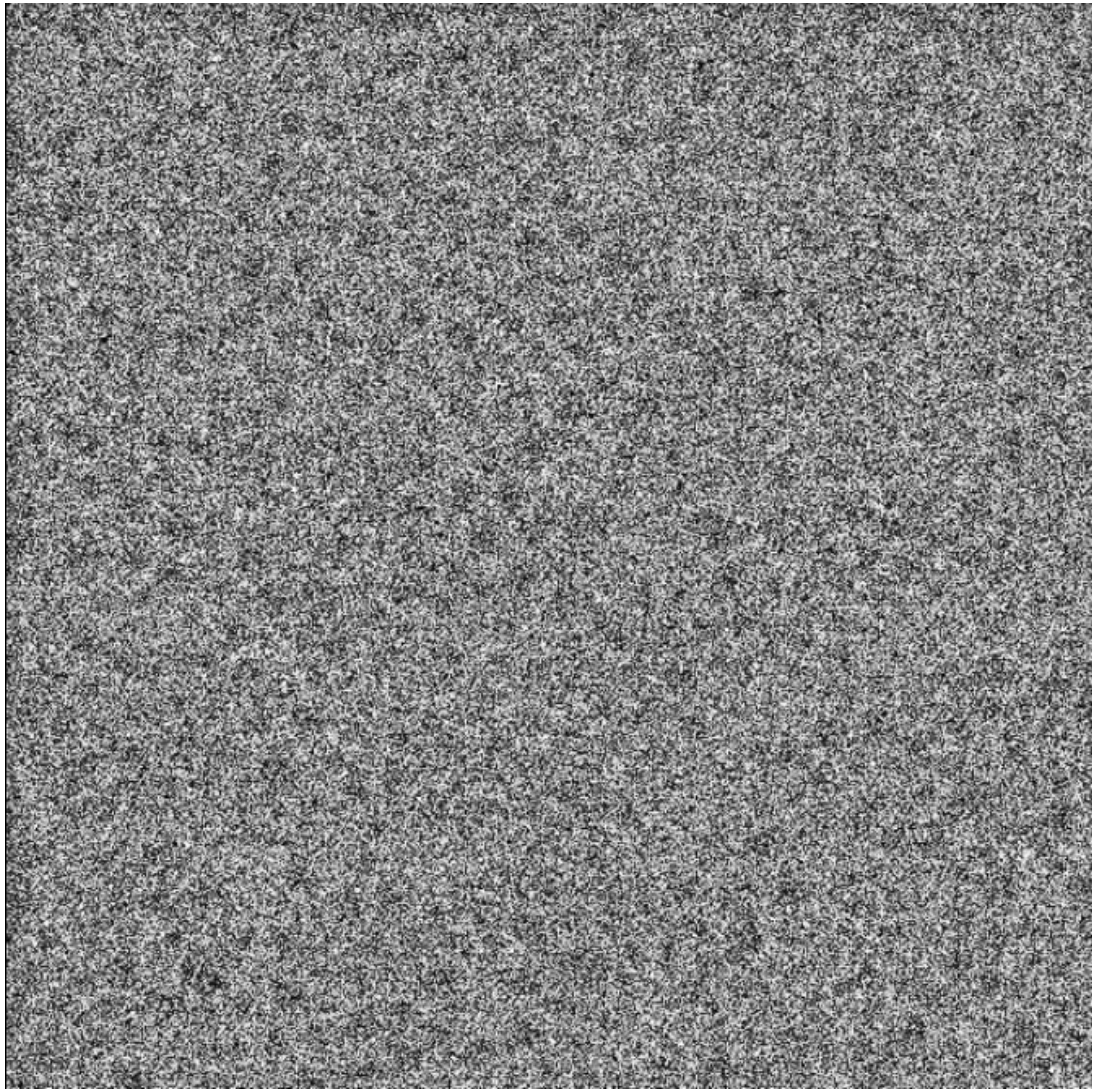
If potential atom is an outlier in all three of these criteria,split it

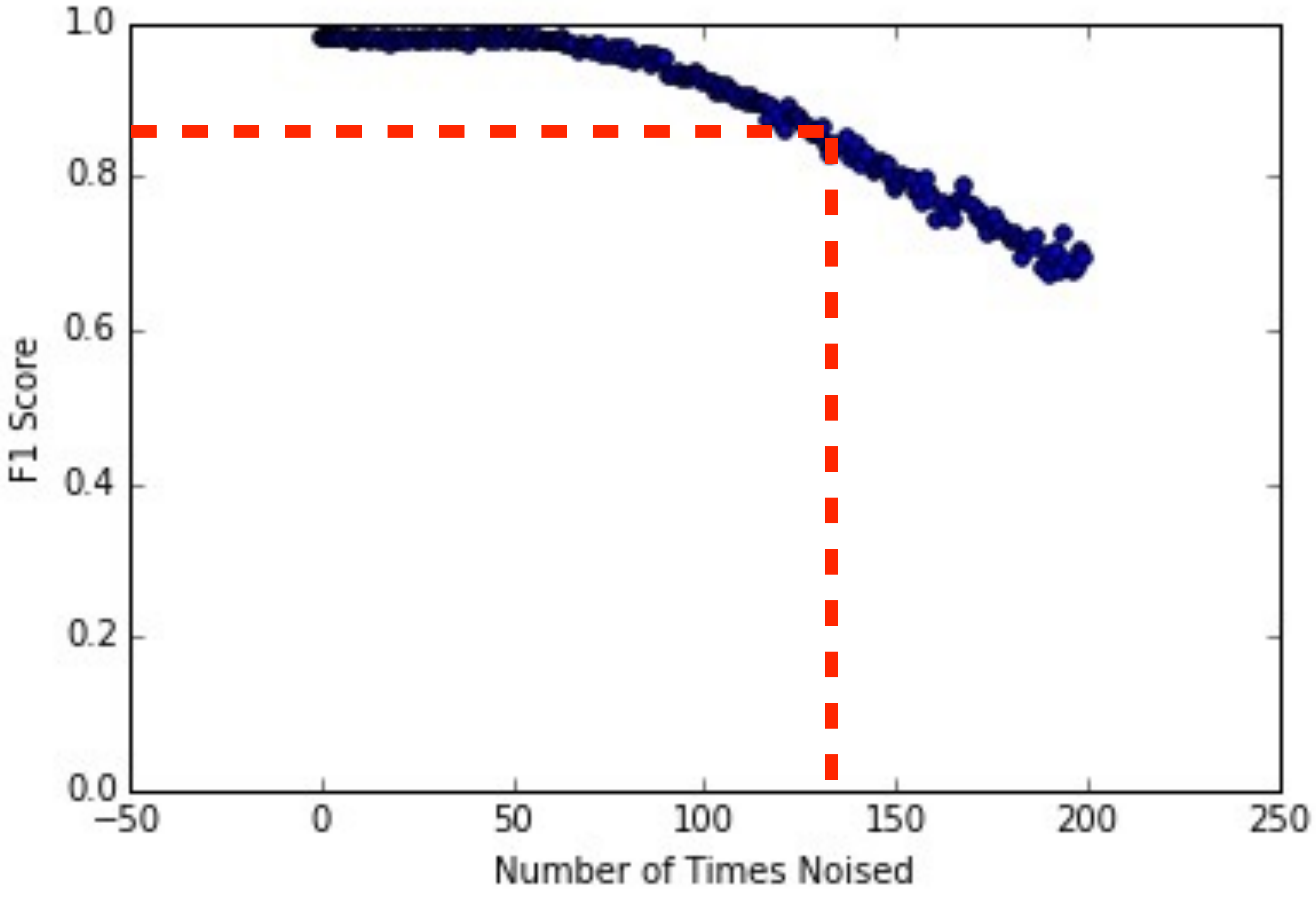






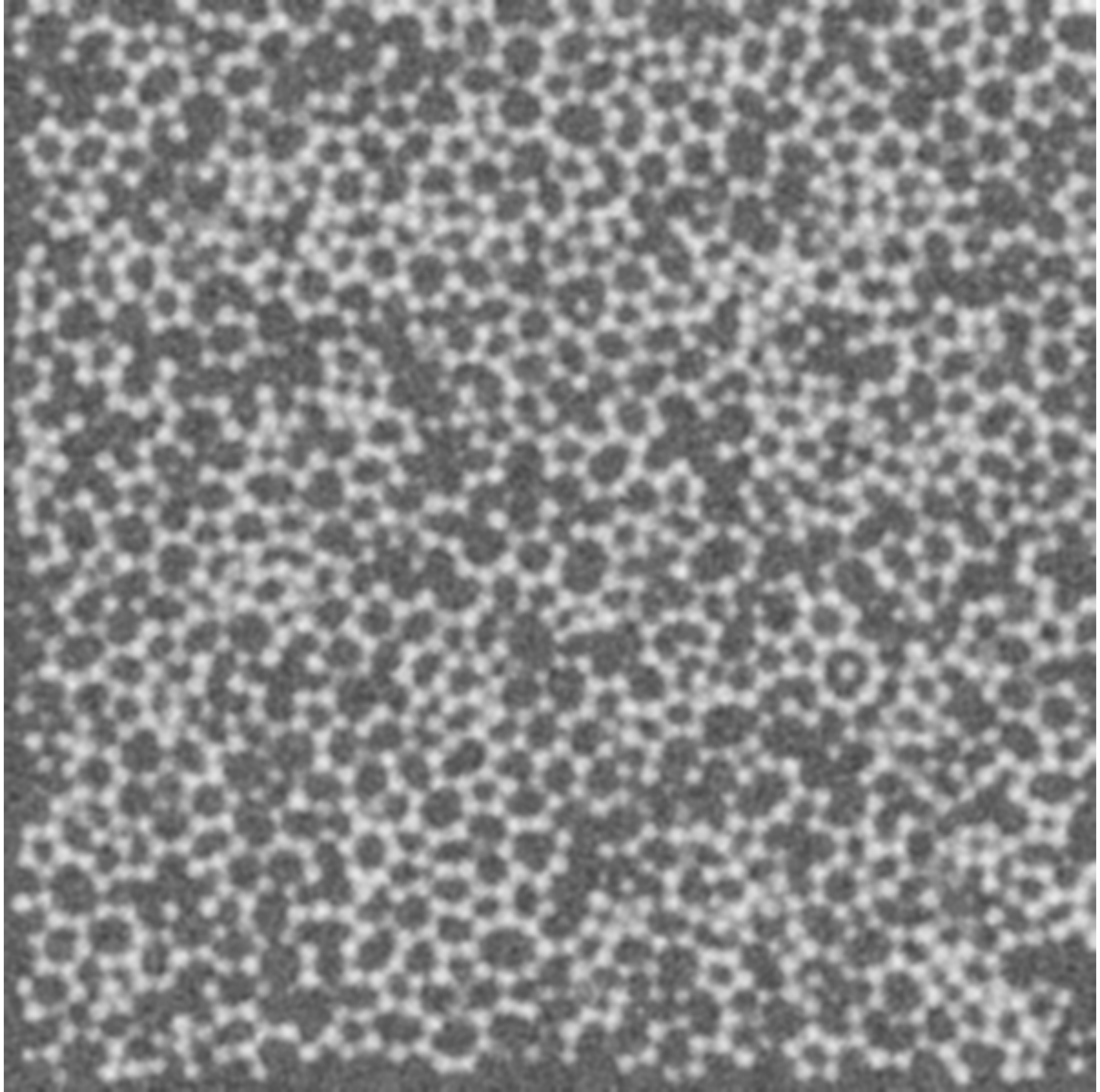
Performance degradation under simulated noise

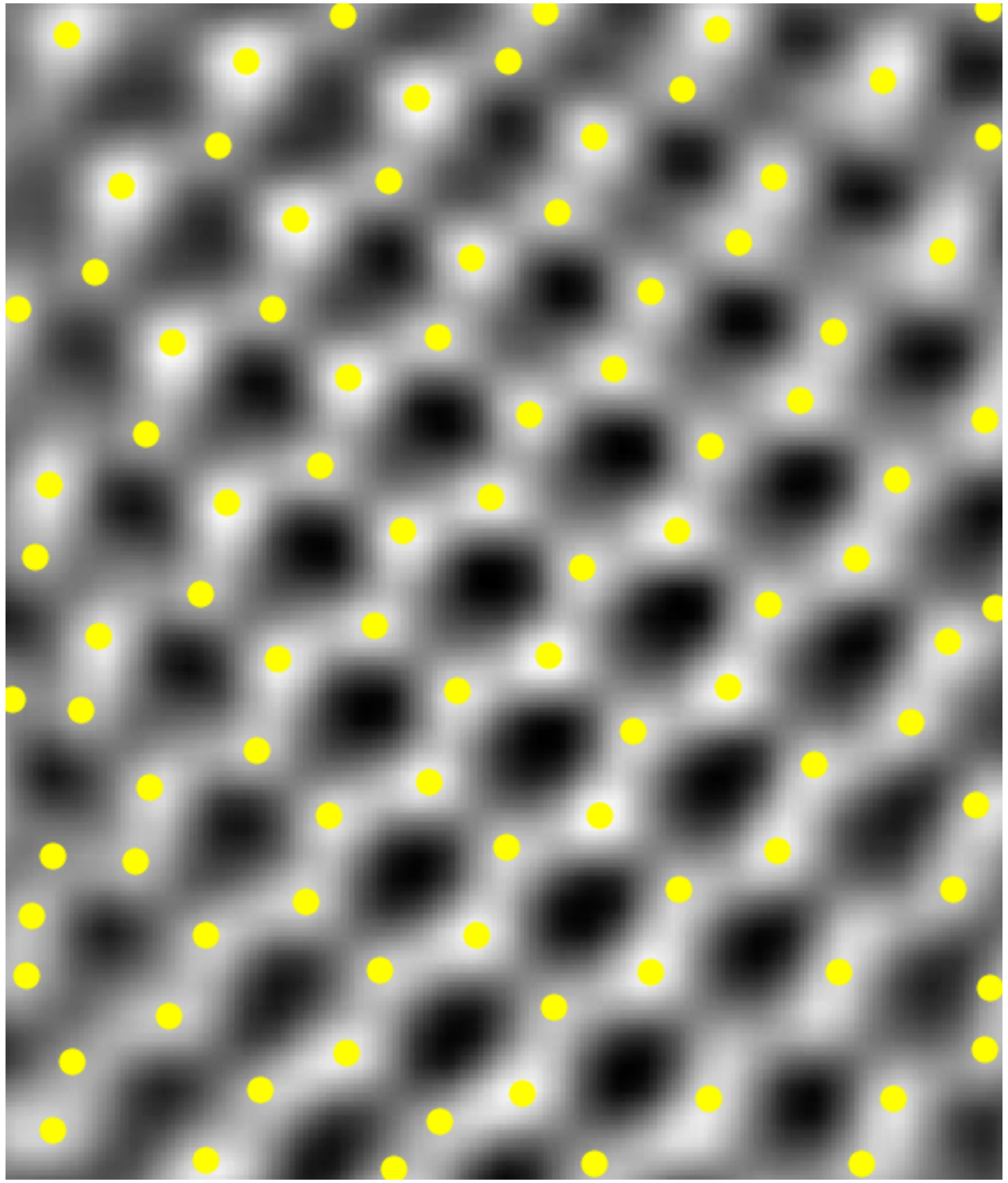




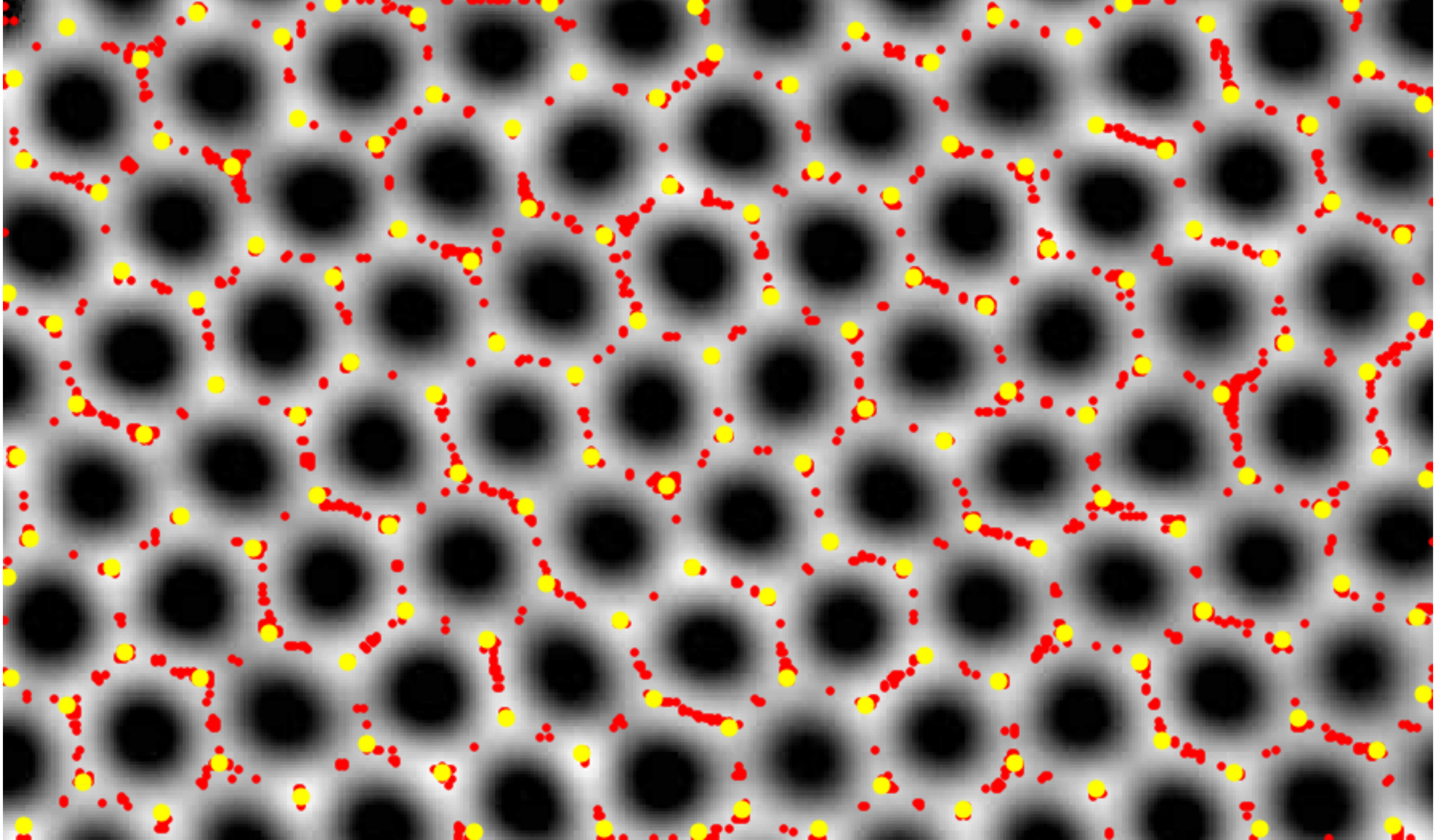
F1 = harmonic mean of *precision* and *recall*. *Precision* is the ratio of correct detection to all detections. *Recall* is the ratio of correct detections to the total number of true atoms.

Tune *Leucippus* for multiple phases and defects

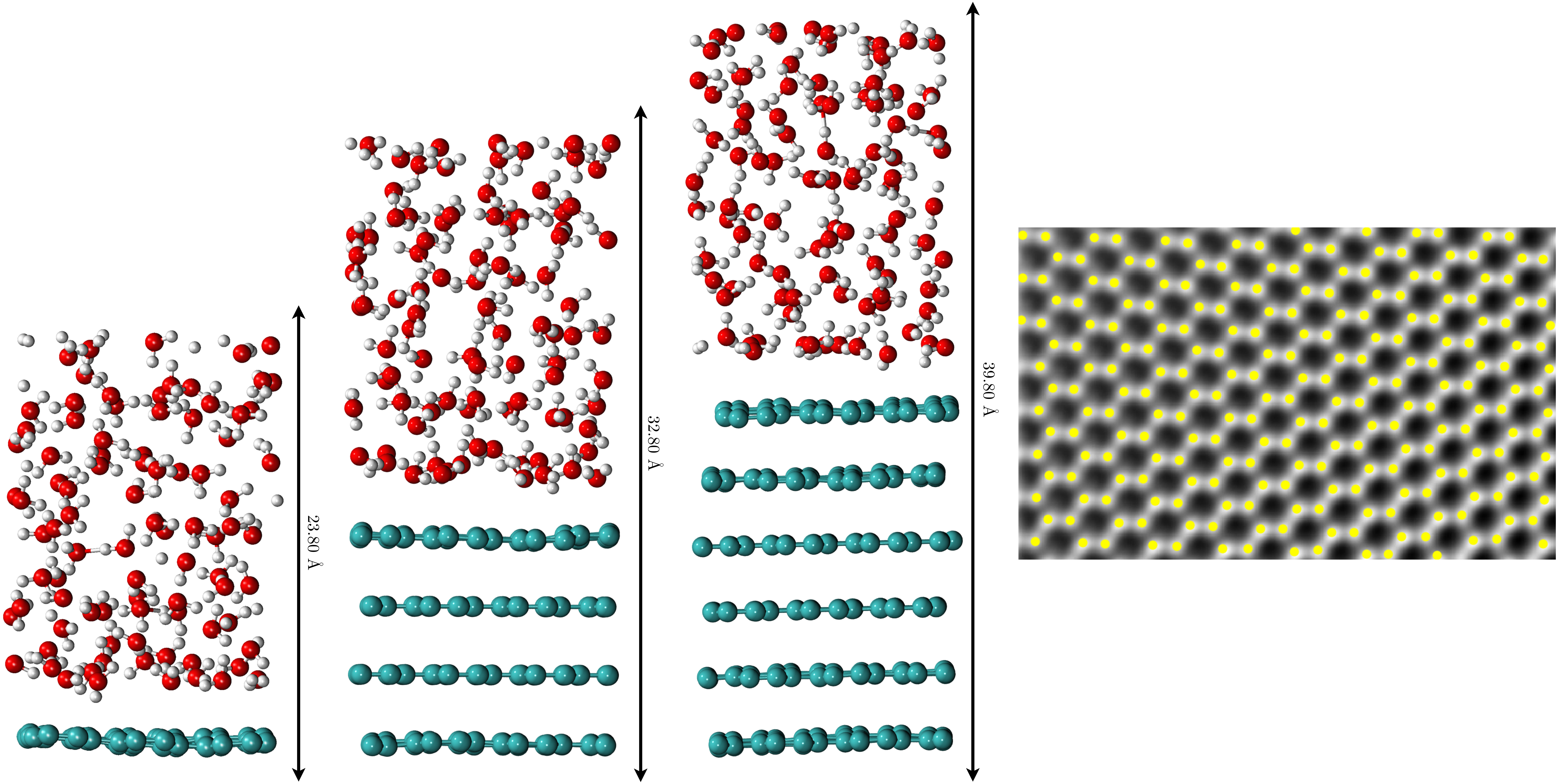




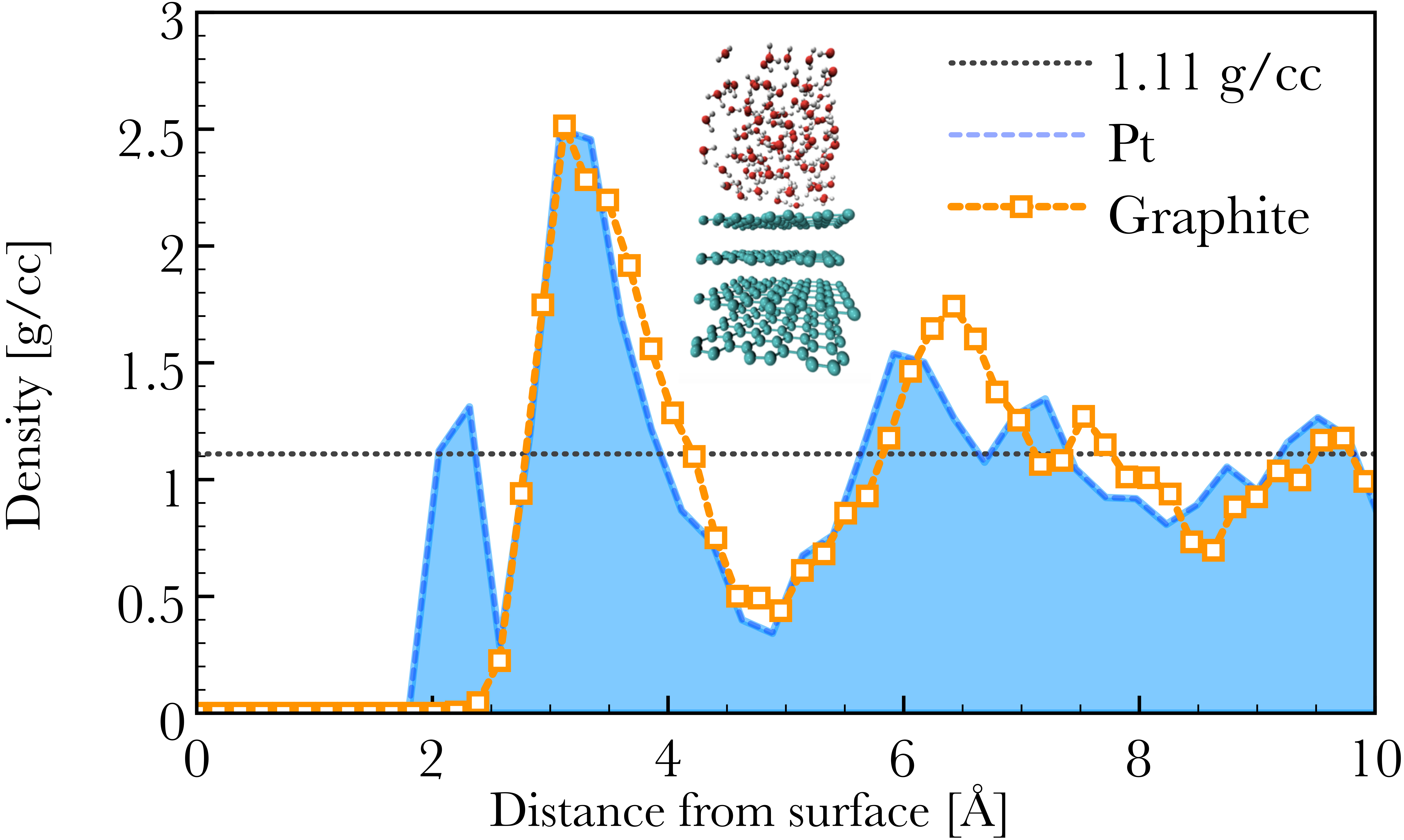
http://leucipp.us



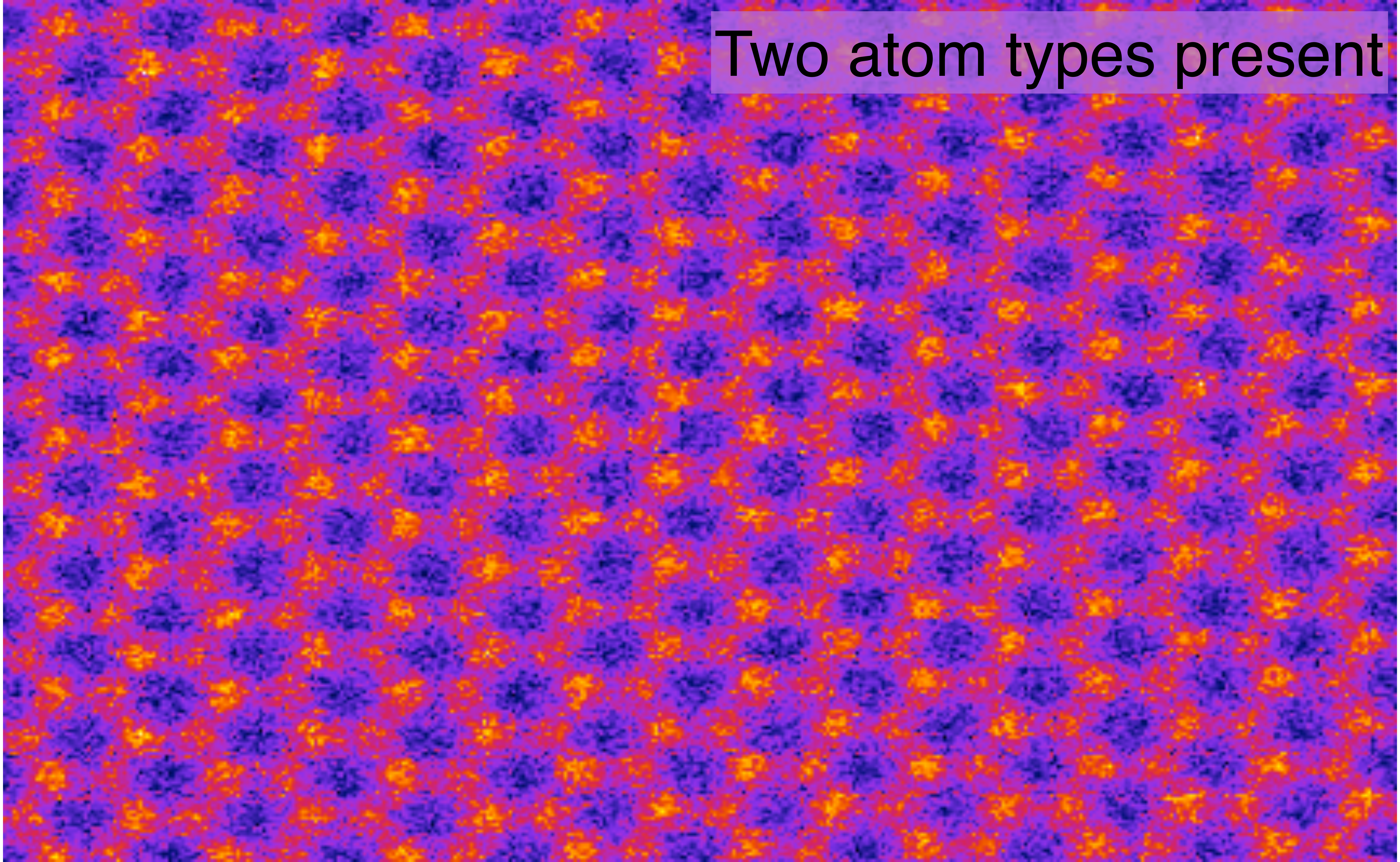

ImageJ
File
Edit
Image
Process
Analyze
Plugins
Window
Help
Dev
Stk
graphene1.png (200%)
439x258 pixels; RGB; 442K
Menu
ImageJ
graphene1.png (200%)
2
97%
20:09



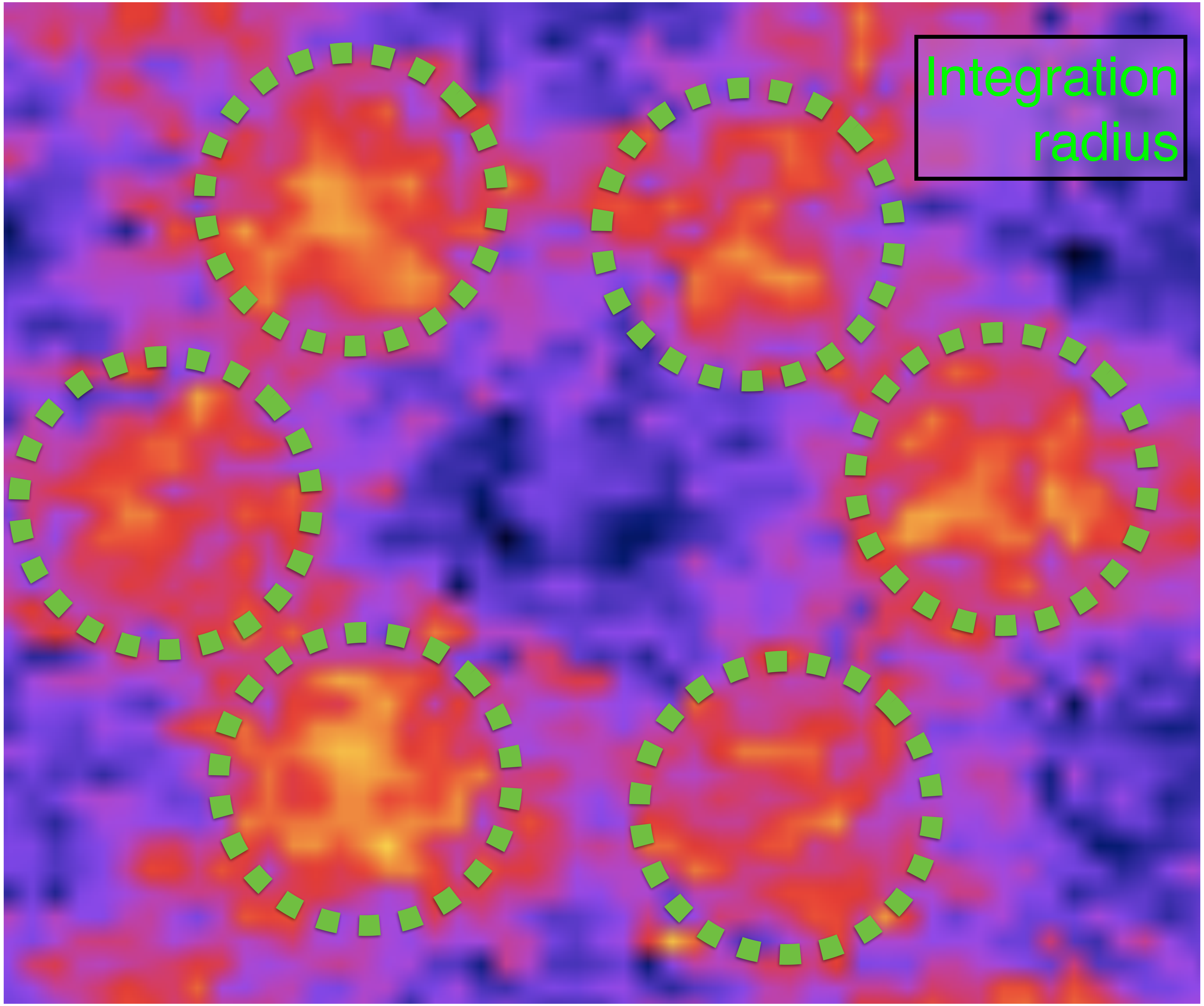
23.80 Å
32.80 Å
39.80 Å

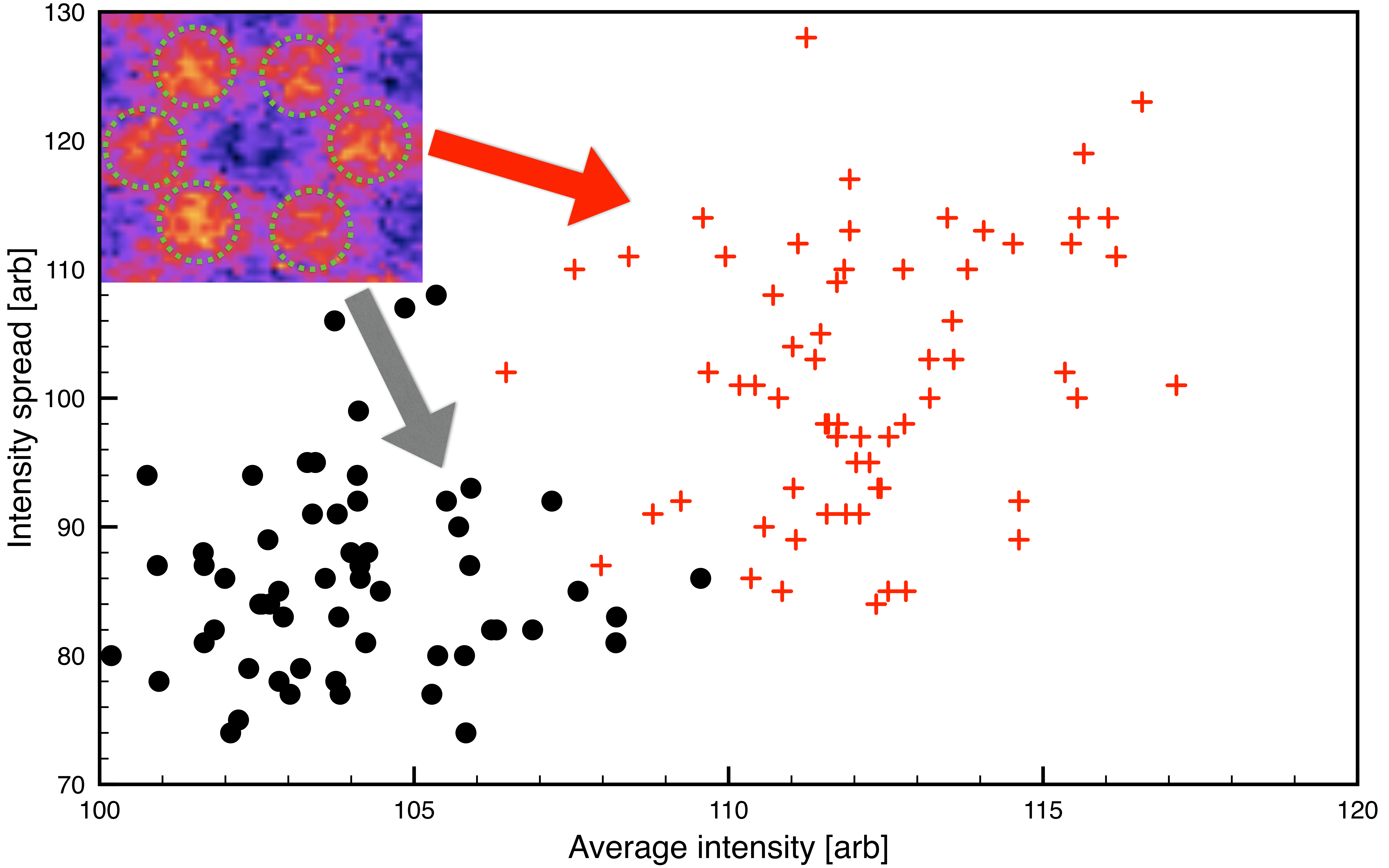


Density [g/cc]
Distance from surface [Å]
1.11 g/cc
Pt
Graphite
0
0.5
1
1.5
2
2.5
3
0
2
4
6
8
10

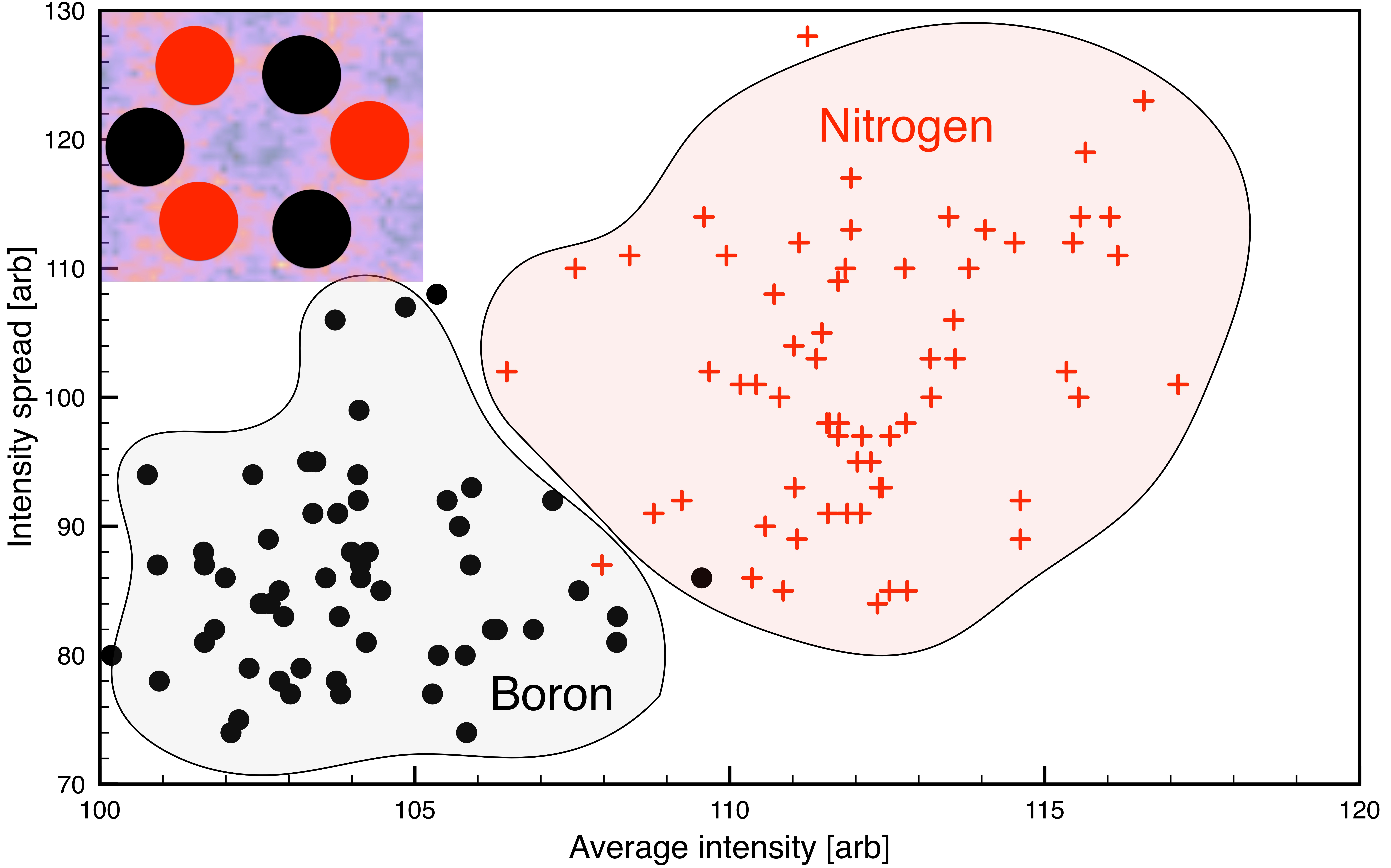

Two atom types present

Localized feature extraction



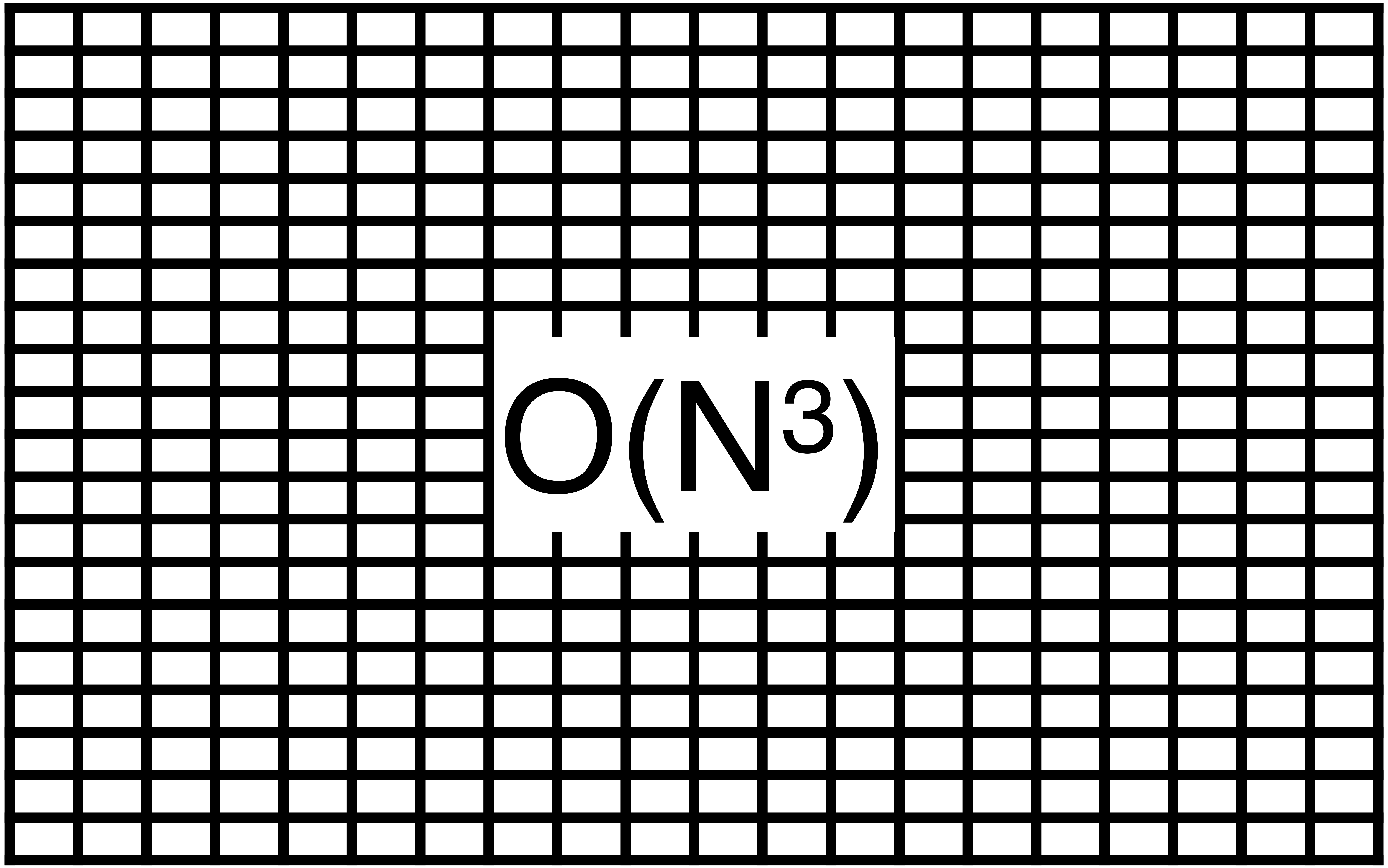


130
120
110
100
90
80
70
Intensity spread [arb]
100
105
110
115
120
Average intensity [arb]


Nitrogen
Boron
Intensity spread [arb]
Average intensity [arb]
130
120
110
100
90
80
70
100
105
110
115
120

Density functional theory maps the ground state electron density to physical observables of the system

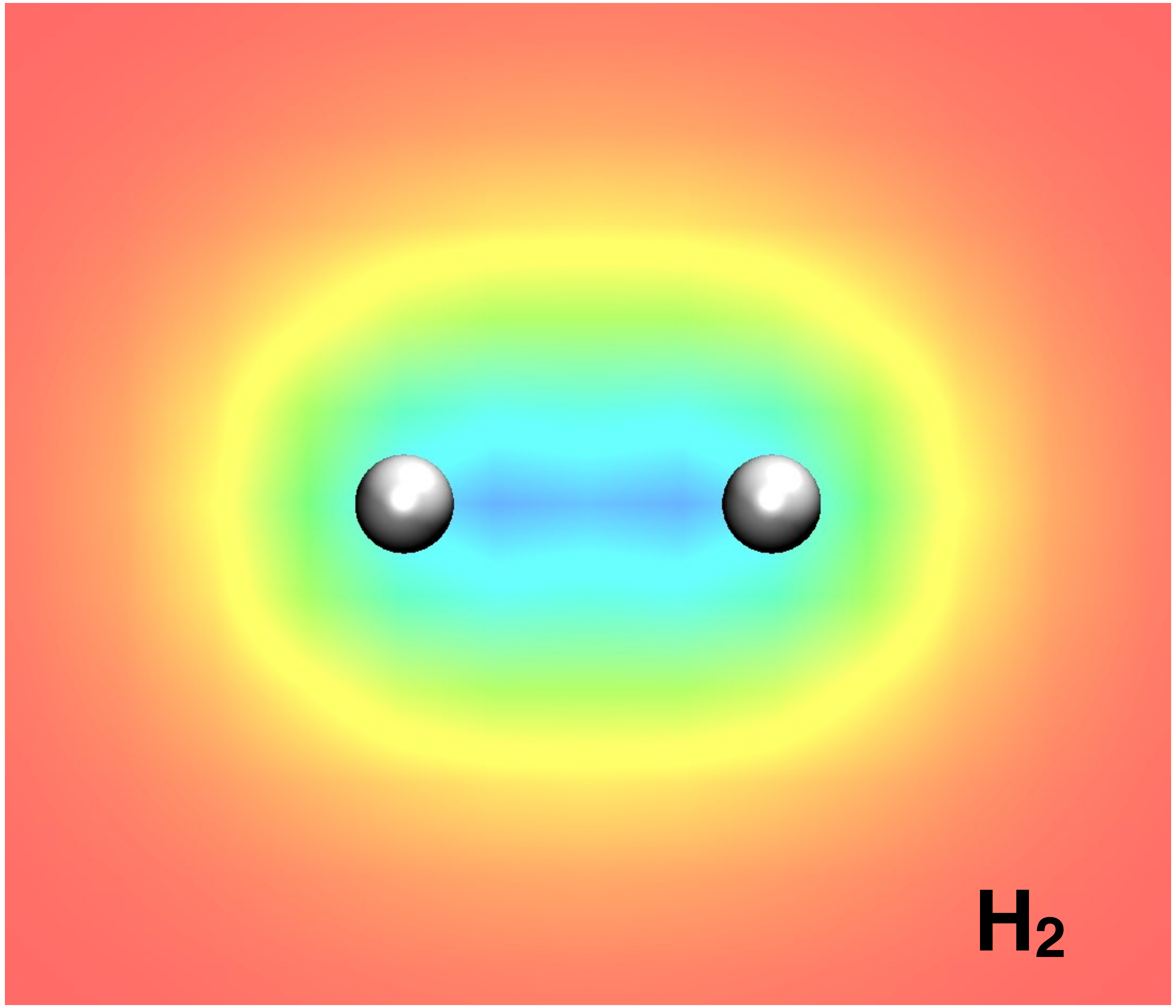
Only the density is needed, yet we ignore this and use wavefunctions



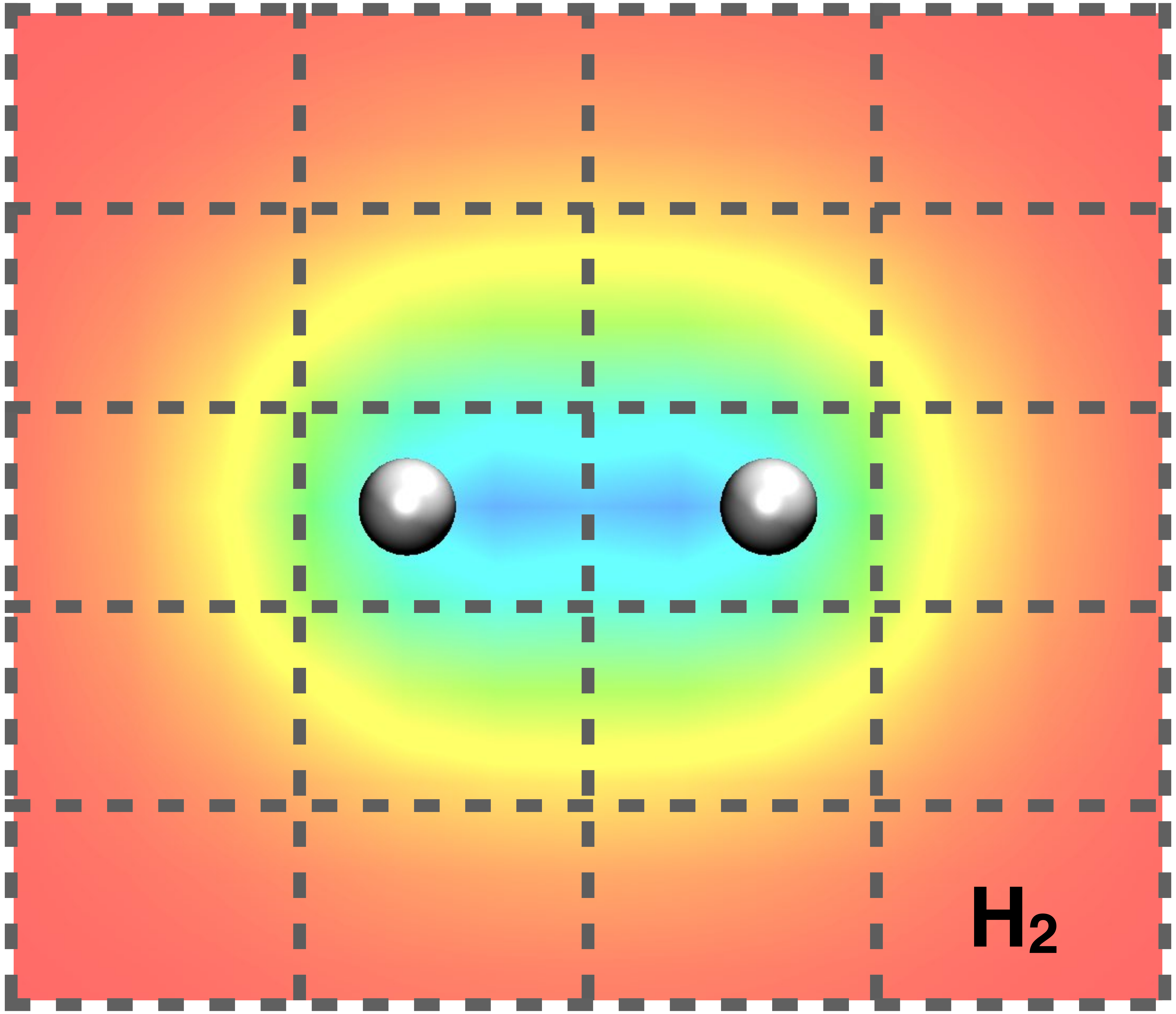

$V_{\text{ext}}(\mathbf{r}) \xleftarrow{\text{HK}} n_0(\mathbf{r})$
$\Psi_i(\{\mathbf{r}\}) \longrightarrow \Psi_0(\{\mathbf{r}\})$
O(N)
KS
$n_0(\mathbf{r}) \xrightarrow{\text{HK}_0} V_{\text{eff}}(\mathbf{r})$
$\psi_{i=1,N_e}(\mathbf{r}) \longleftarrow \psi_i(\mathbf{r})$
O(N³)

Density functional theory maps the ground state electron density to physical observables of the system

We can't read the map

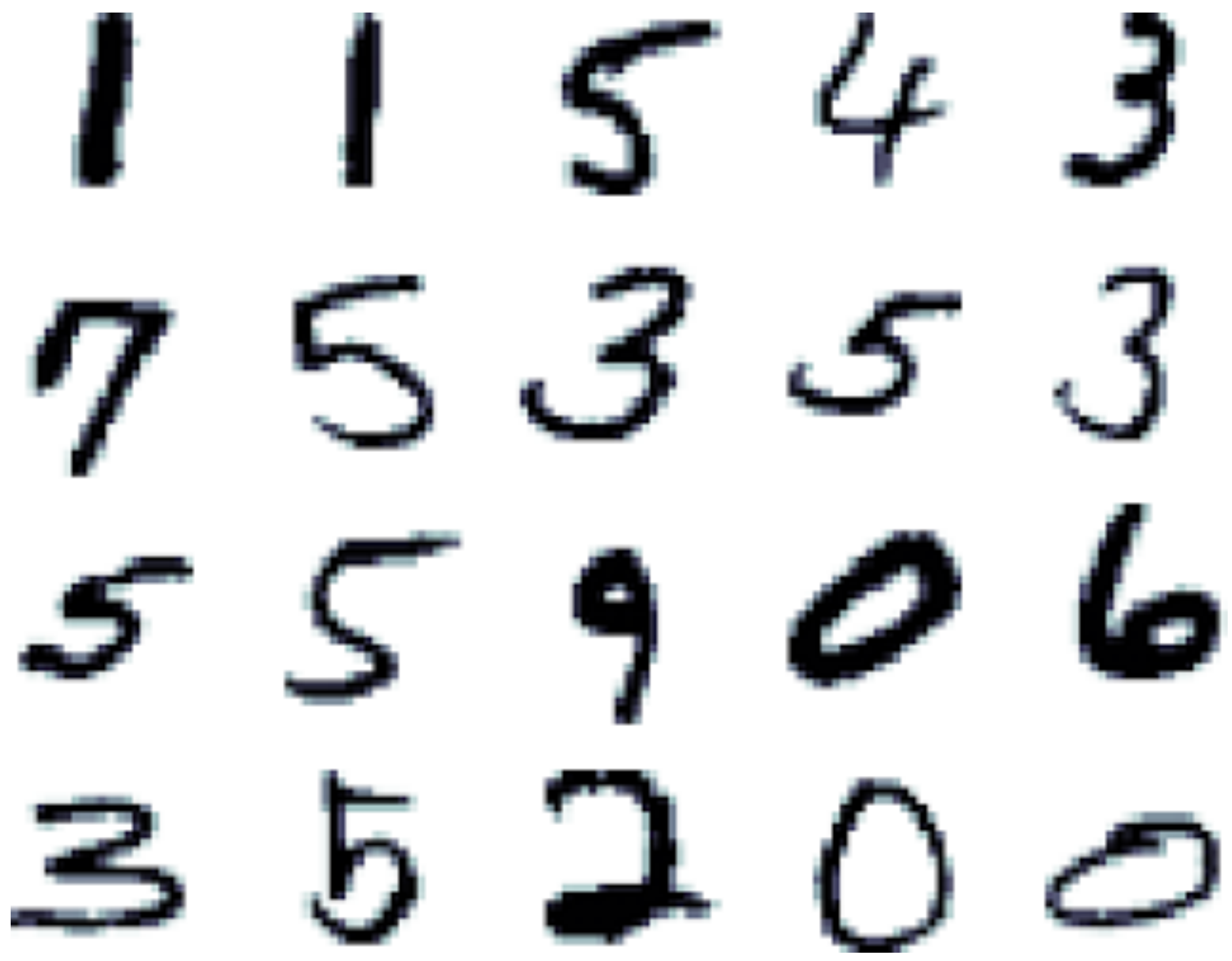

H_2

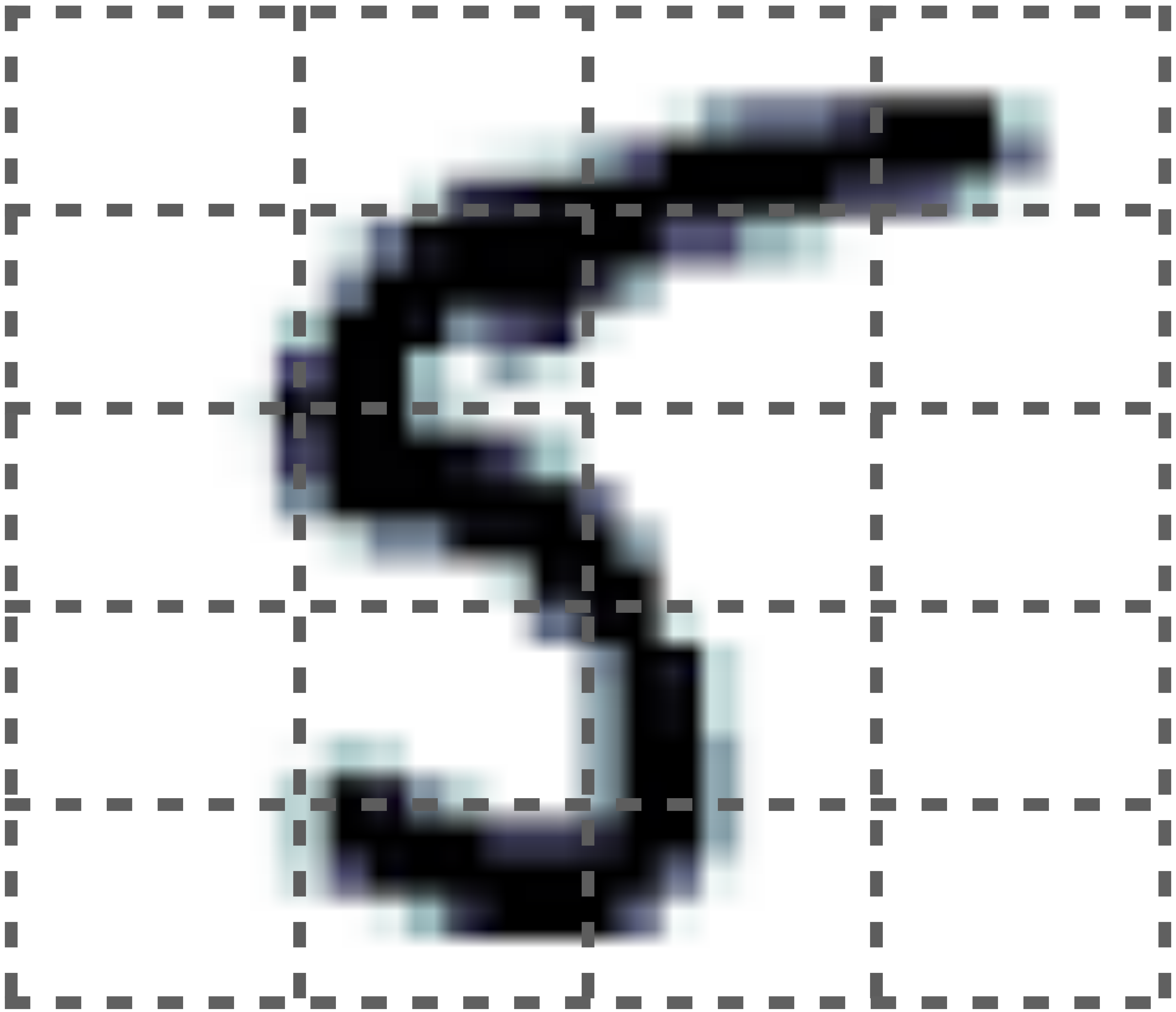
What wavelengths of light does it absorb?

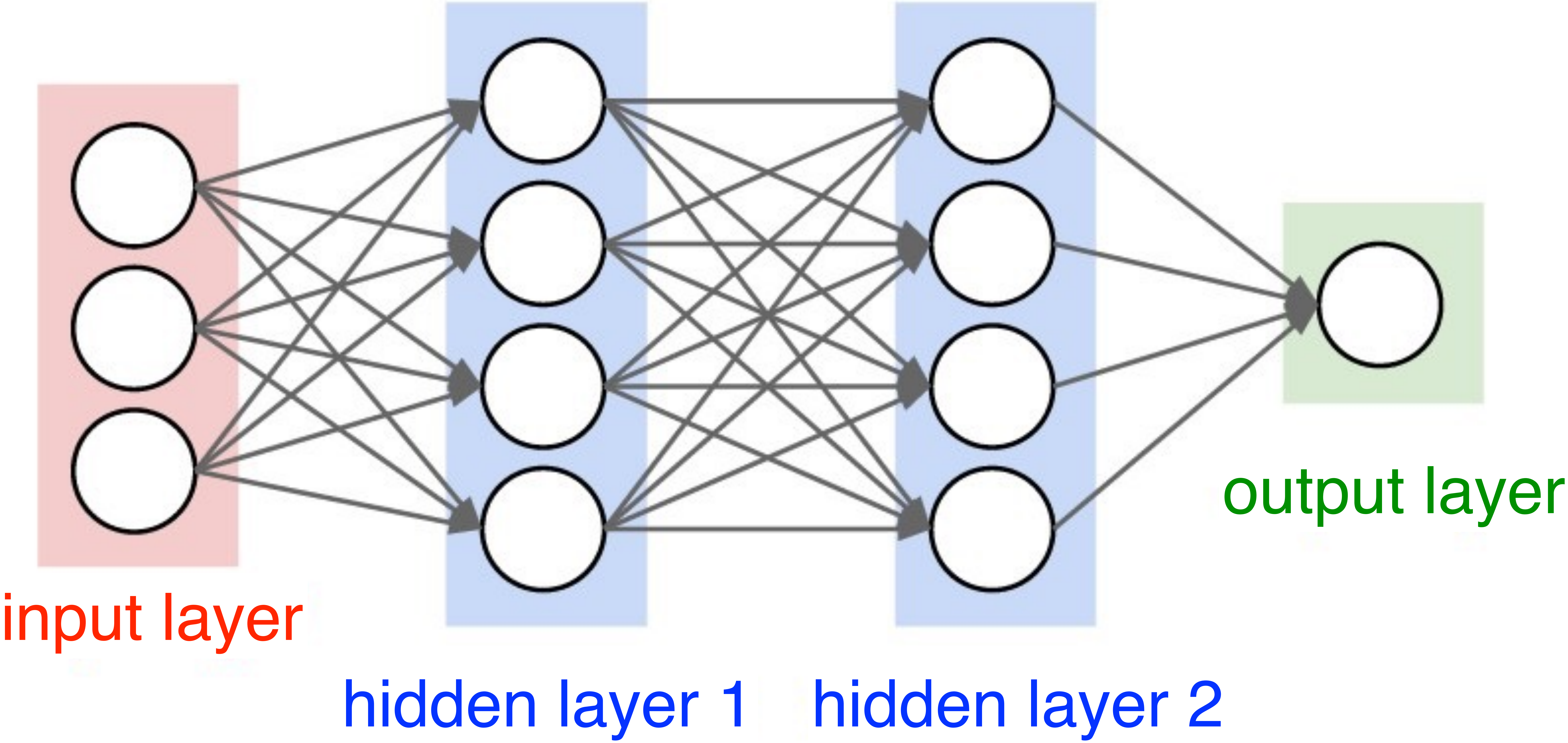


How much energy does it take to remove an electron (ionization energy)?

What are the forces?

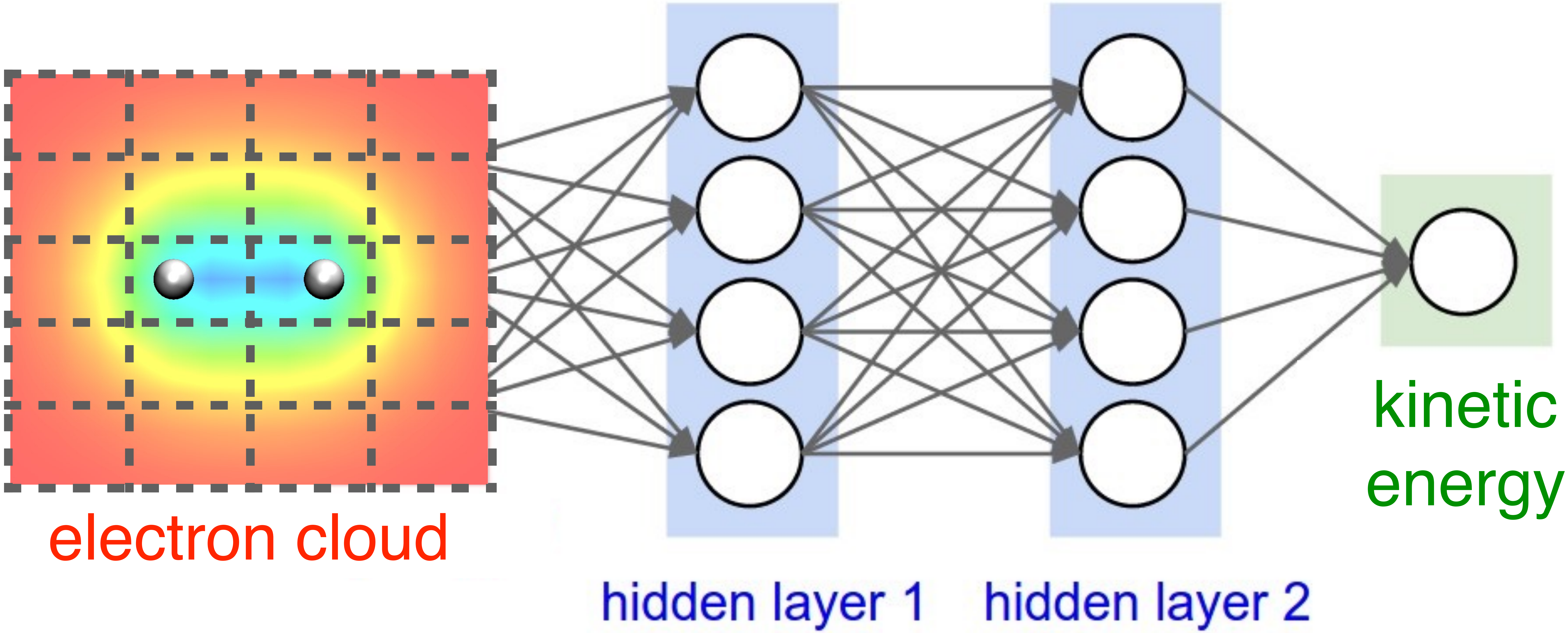




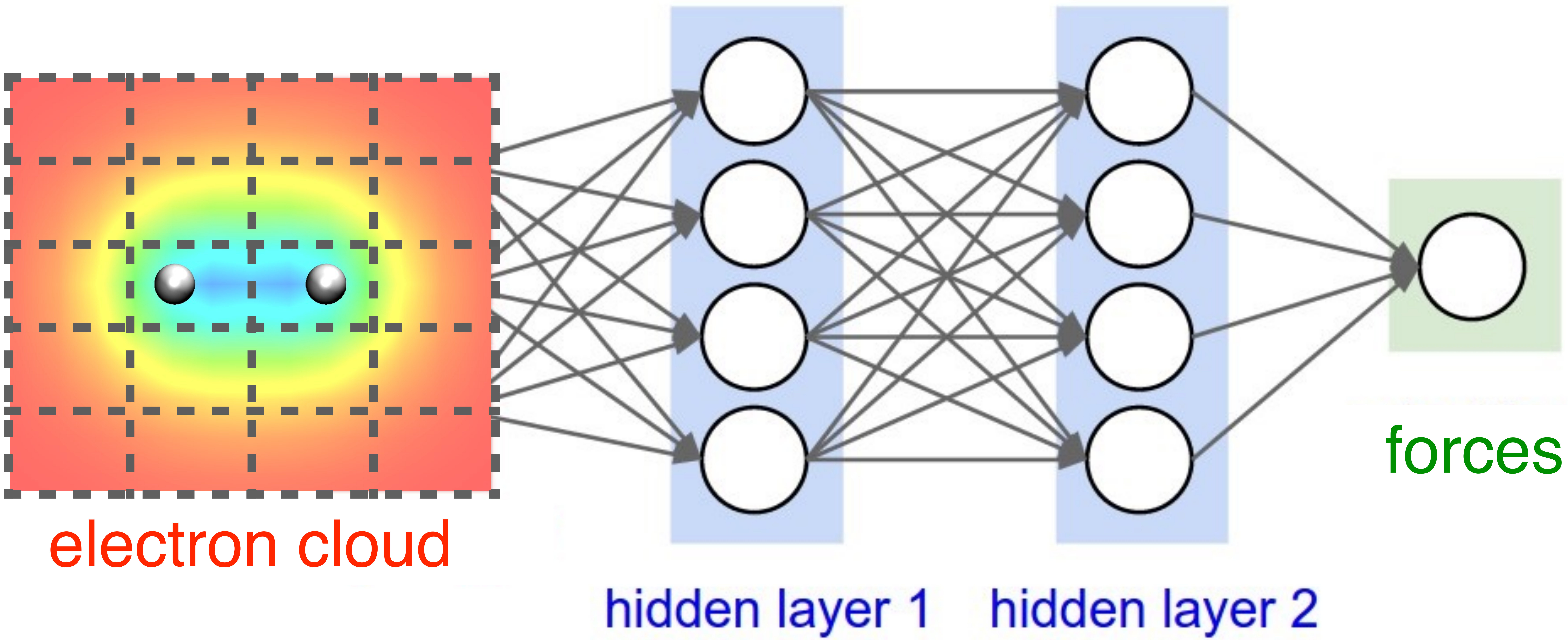


CS231n, Convolutional Neural Networks for Visual Recognition (http://cs231n.github.io)

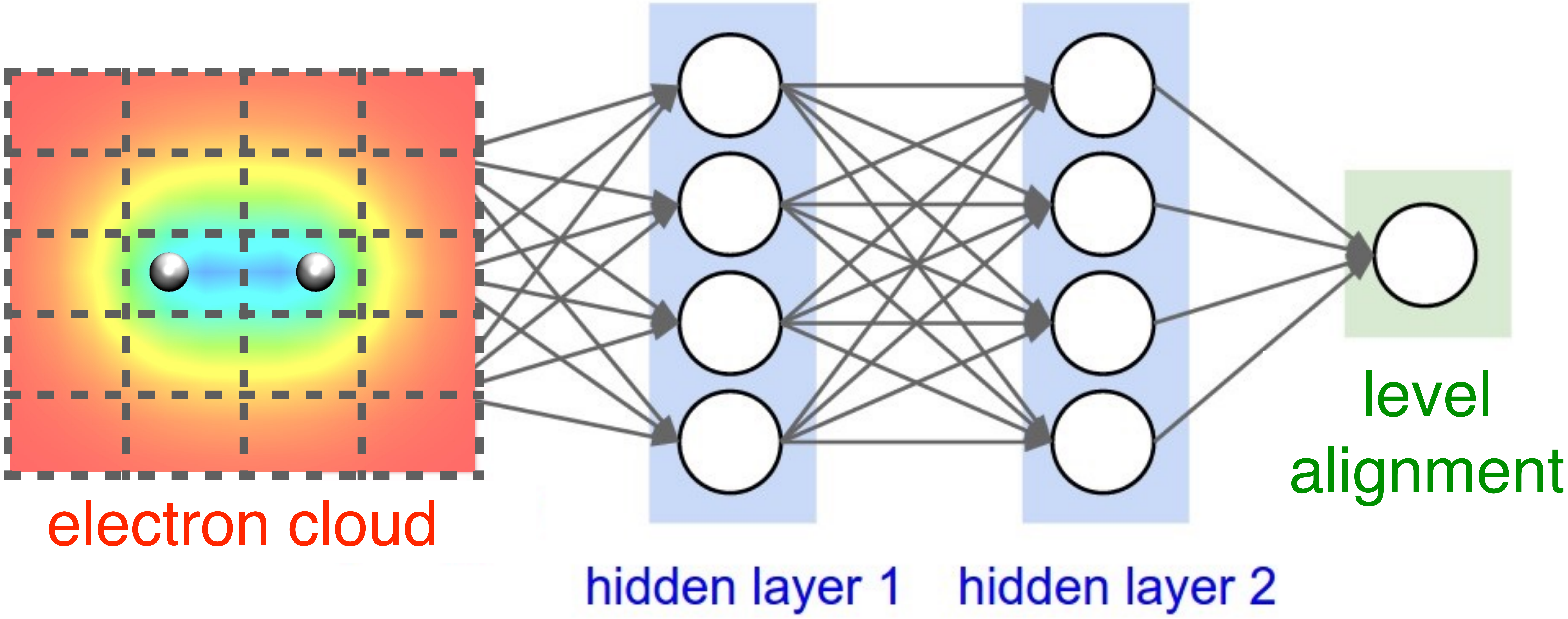
CS231n, Convolutional Neural Networks for Visual Recognition (http://cs231n.github.io/)



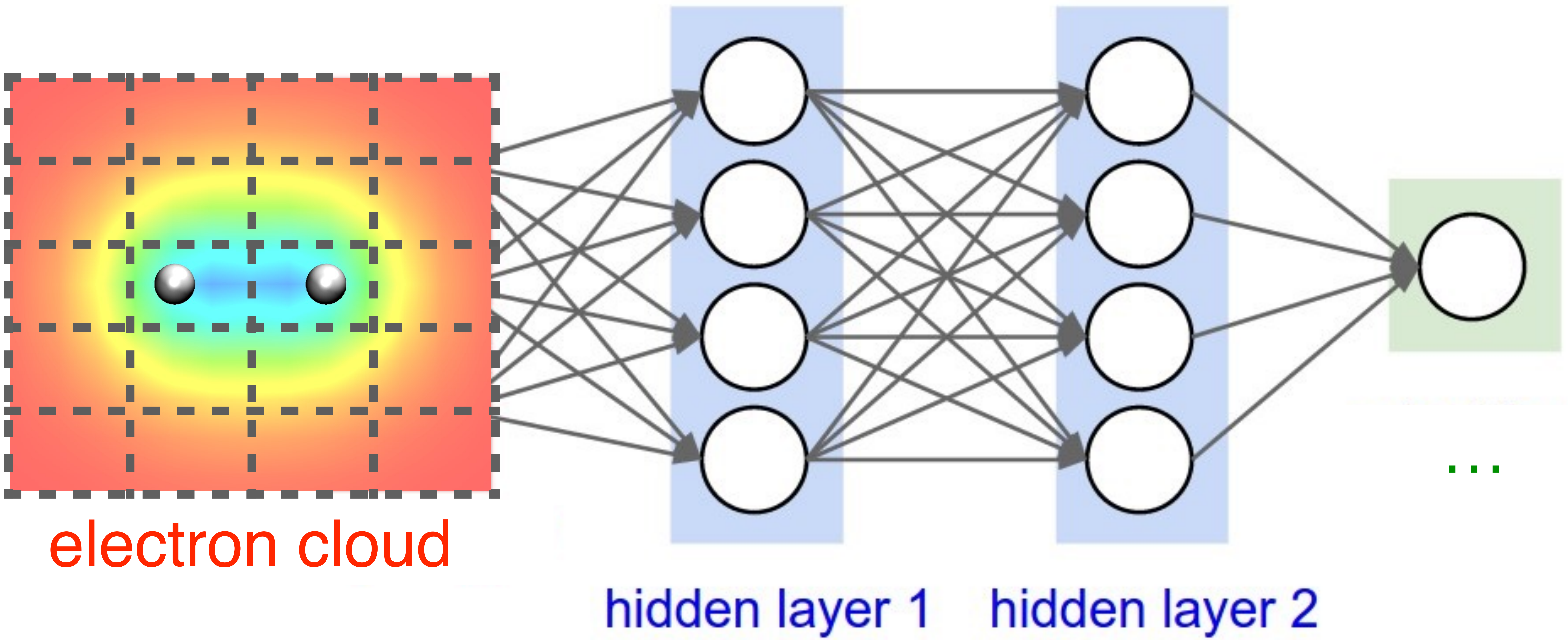
CS231n, Convolutional Neural Networks for Visual Recognition (http://cs231n.github.io/)



CS231n, Convolutional Neural Networks for Visual Recognition (http://cs231n.github.io/)



CS231n, Convolutional Neural Networks for Visual Recognition (http://cs231n.github.io/)



CS231n, Convolutional Neural Networks for Visual Recognition (http://cs231n.github.io/)


$V_{\mathrm{ext}}(\mathbf{r}) \xleftarrow{\mathrm{HK}} n_0(\mathbf{r})$
$\Psi_i(\{\mathbf{r}\}) \longrightarrow \Psi_0(\{\mathbf{r}\})$
O(N)
KS
$n_0(\mathbf{r}) \xrightarrow{\mathrm{HK}_0} V_{\mathrm{eff}}(\mathbf{r})$
$\psi_{i=1,N_e}(\mathbf{r}) \longleftarrow \psi_i(\mathbf{r})$
O(N^3)

CLEAN

Computational Laboratory for Energy And Nanoscience

- DFT+GW significantly improves agreement with experiment for the water/Pt interface
- "Good" & "Bad" surfaces look different at the interface
- A judicious choice of basis is necessary to handle nanoscale systems
- Large scale searching requires a fundamentally different approach

Acknowledgements

A. Willard
M. Morales
D. Prendergast
T. Ogitsu
BerkeleyGW Team

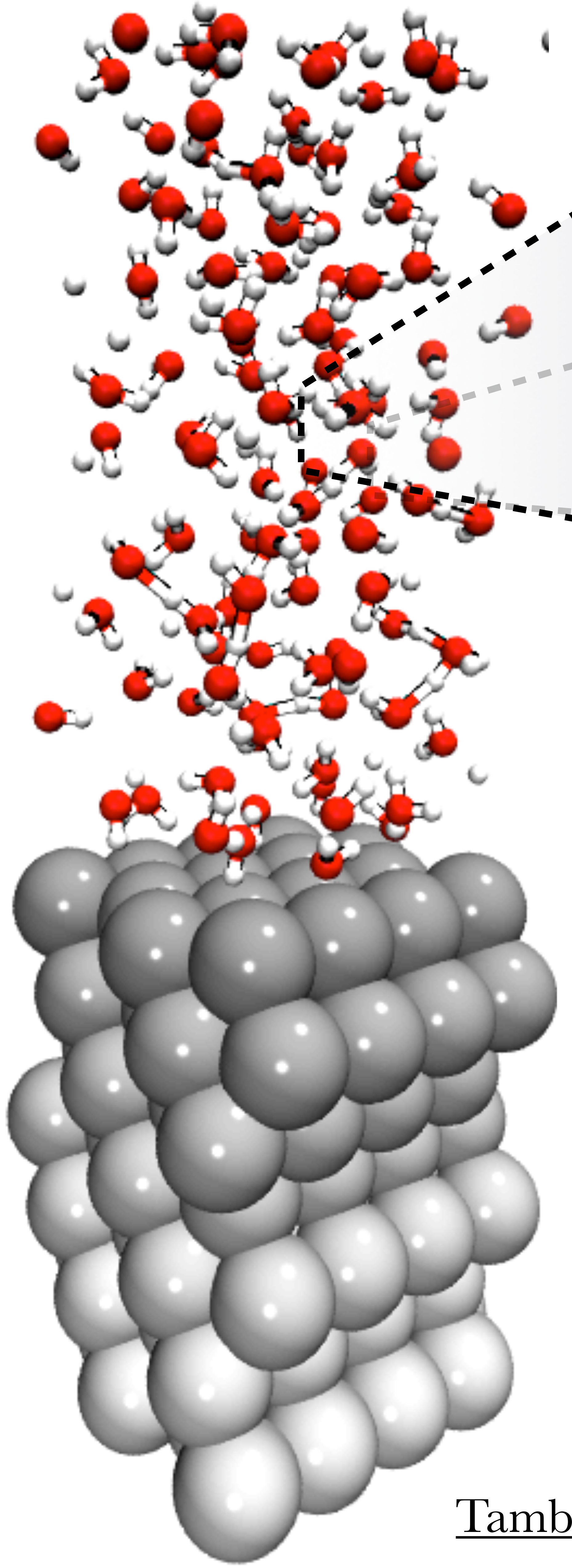
NSERC CRSNG

SOSCIP

berkeleygw.org

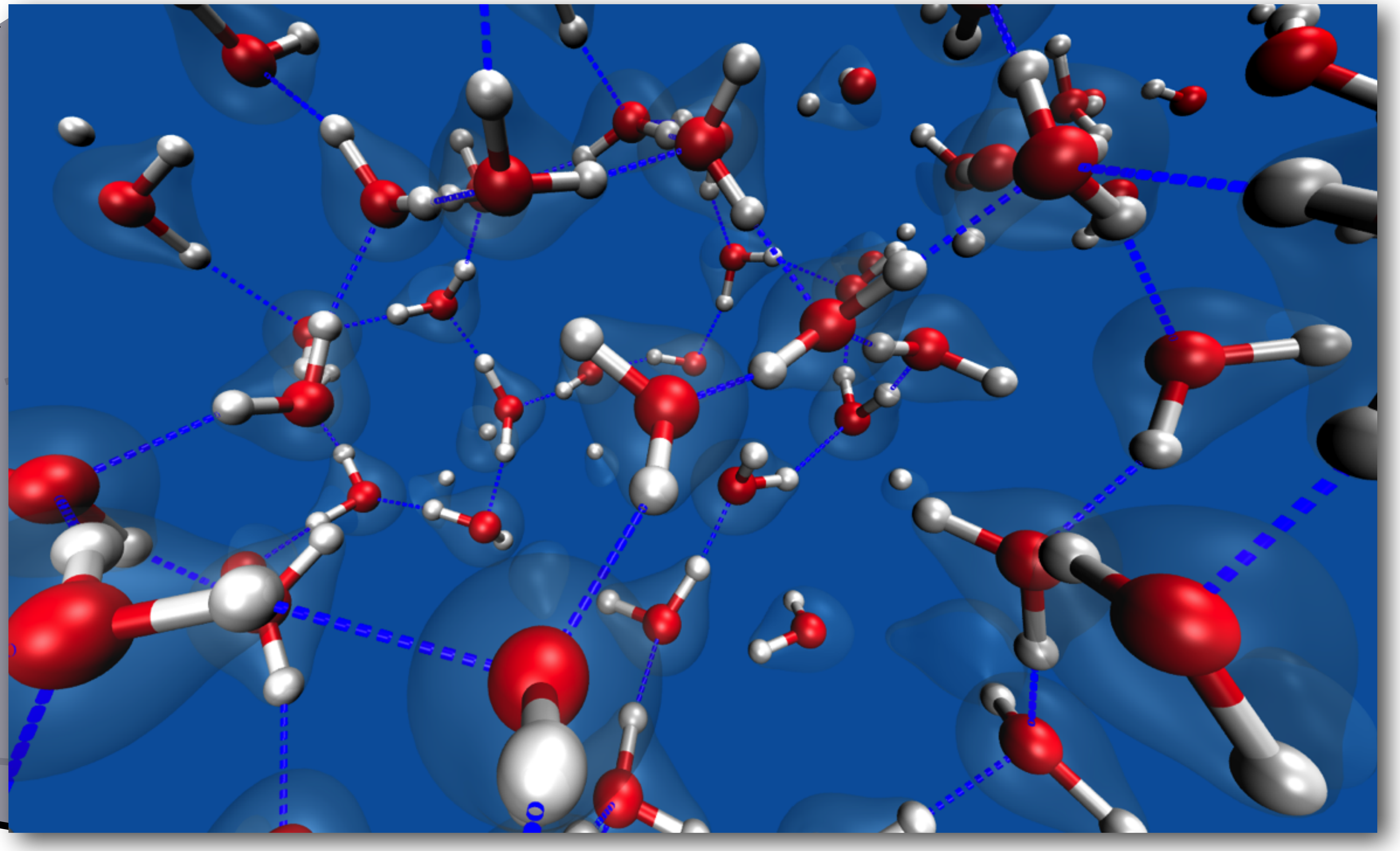
http://faculty.uoit.ca/itamblyn

liquid water

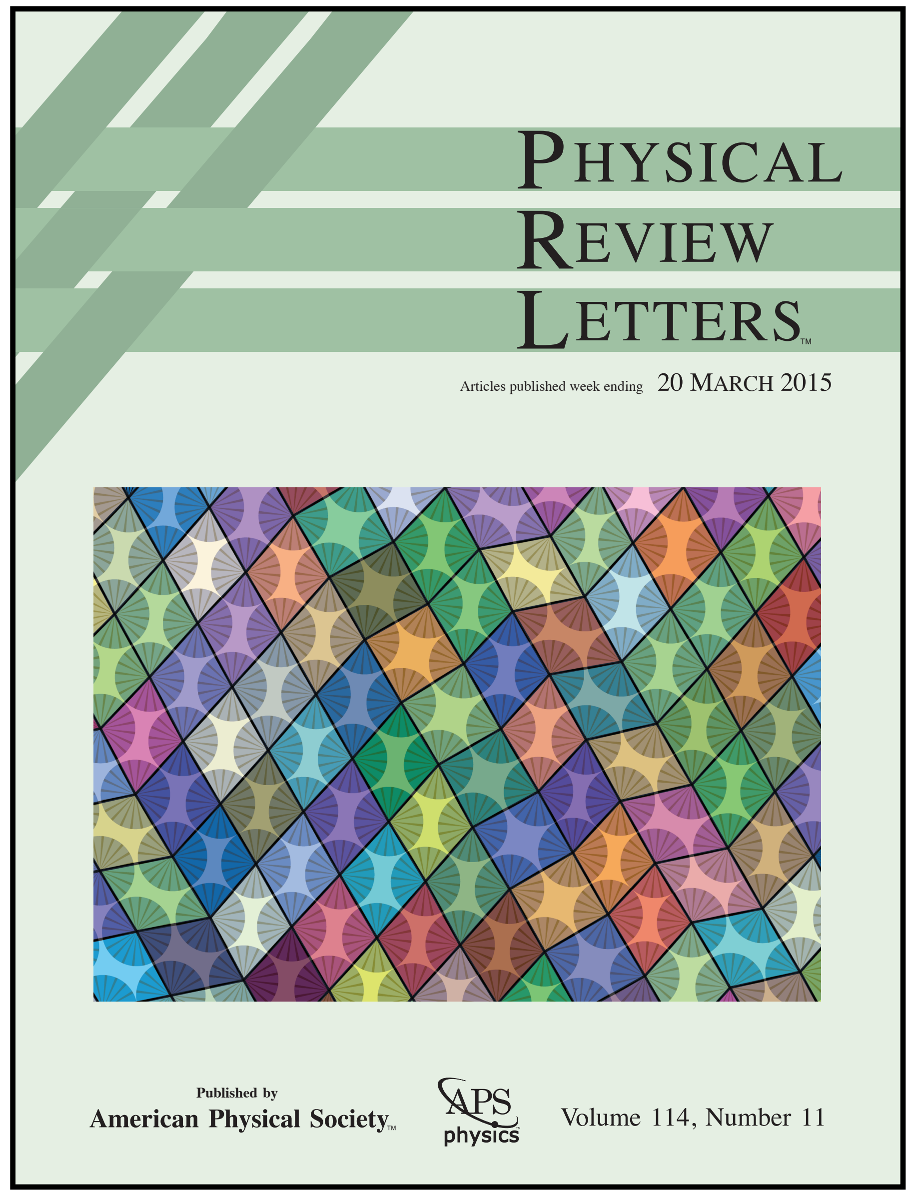


Platinum electrode

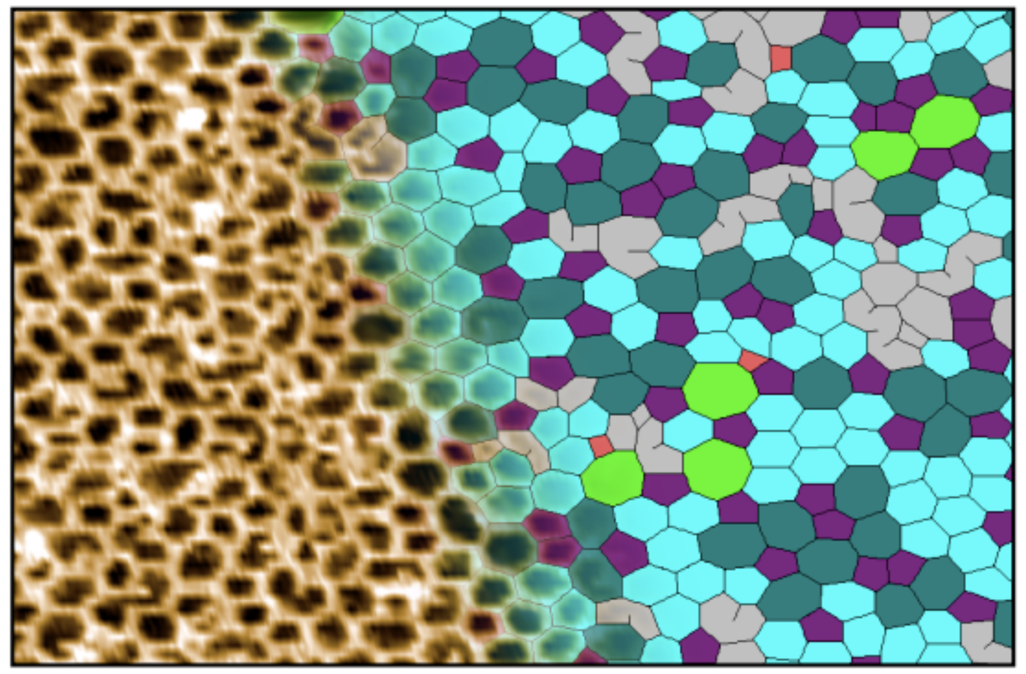
Tamblyn, *submitted*, (2015)



H-bonding

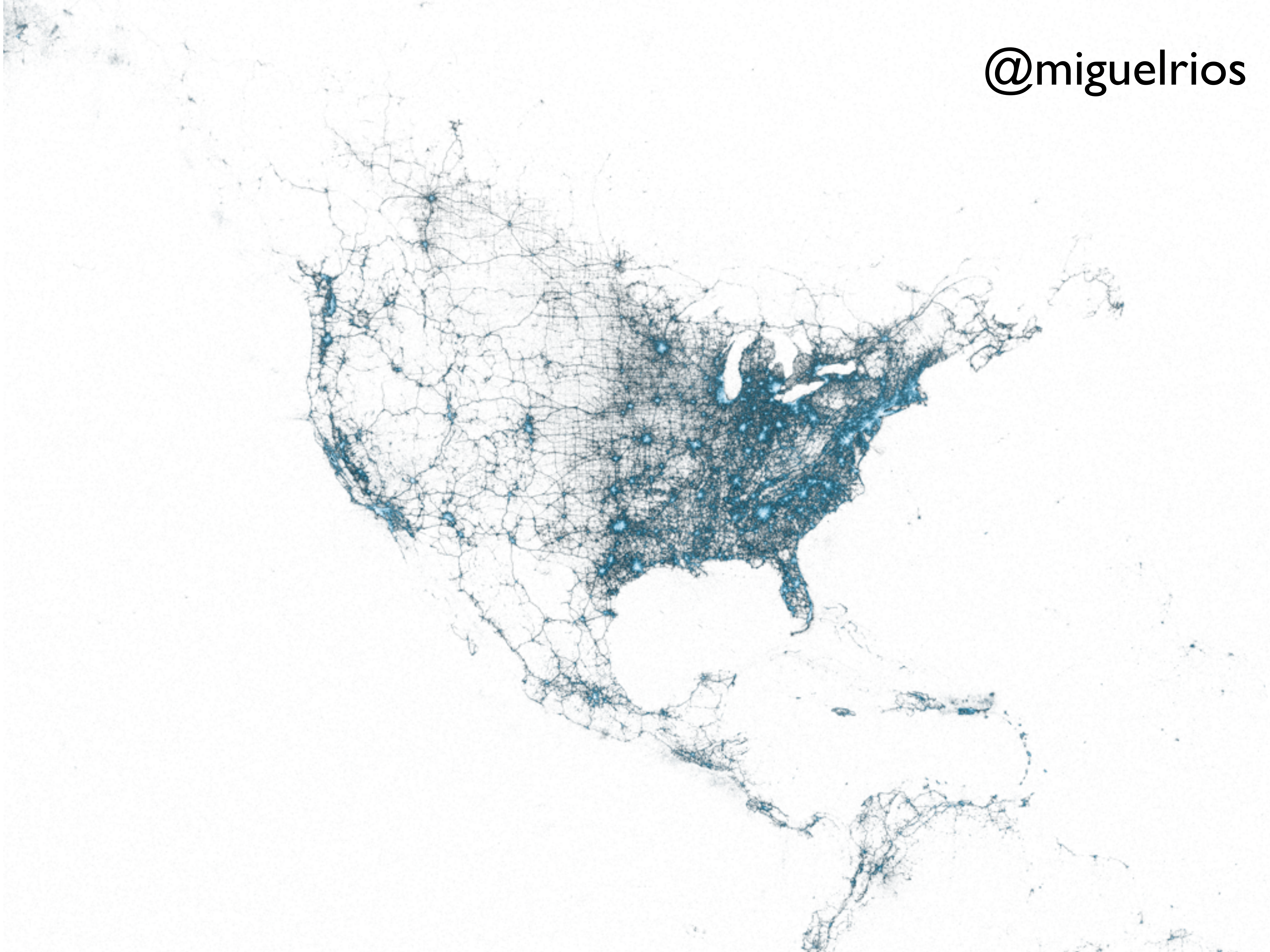


self assembled networks - OPV

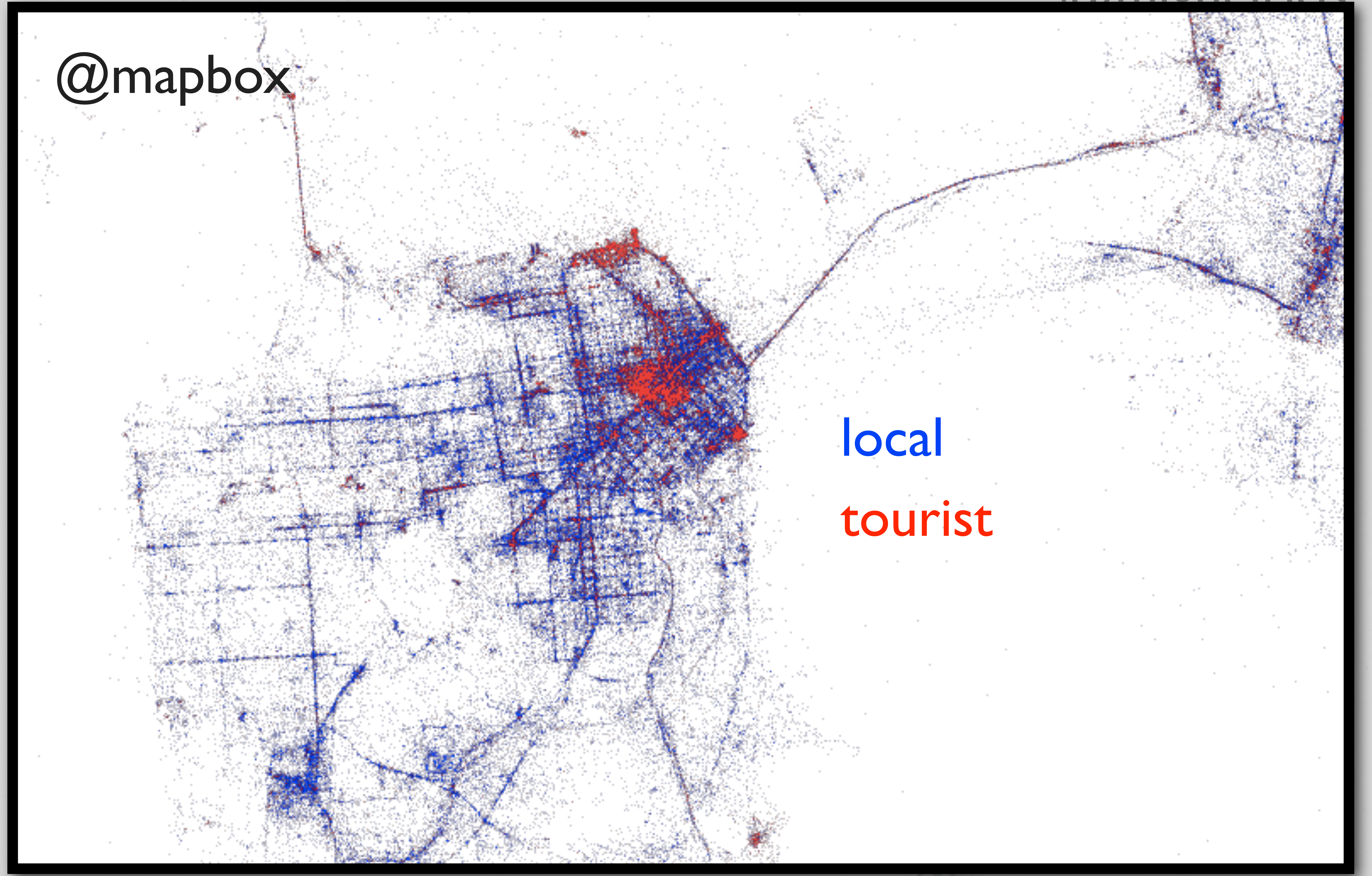


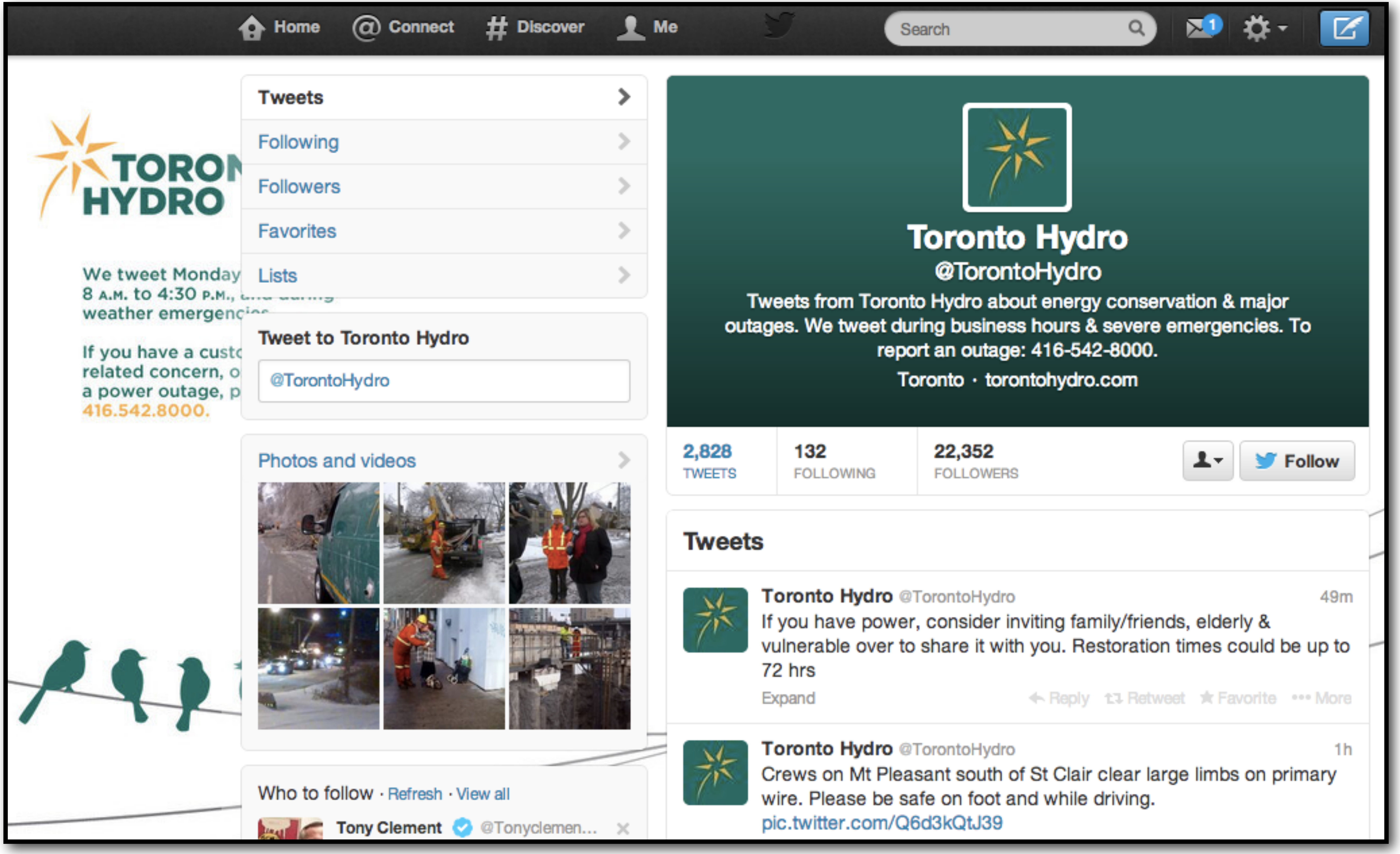
Whitelam, Tamblyn, Beton, and Garrahan, *Phys. Rev. Lett.* 114, 115702 (2015)
Whitelam, Tamblyn, *et al.*, *Phys. Rev. X*, 3, (2014)
Whitelam, Tamblyn, Beton, and Garrahan, *Phys. Rev. Lett.* 108, 035702 (2012)

@miguelrios



@miguelrios





December 2013 Power Outage


@jkrums
US AIRWAYS
2009

@UkrProgress
@StateDept

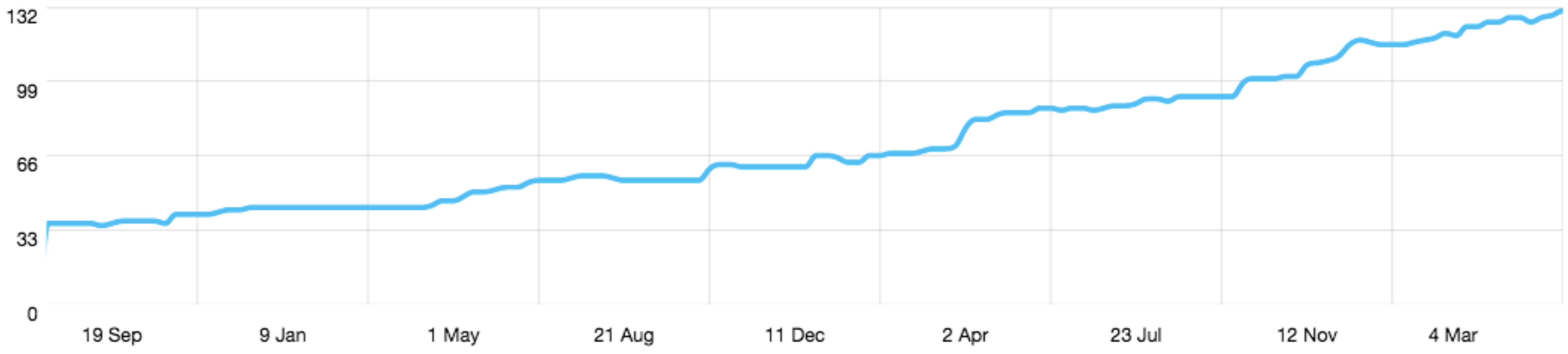
Прогресс для Украины
@UkrProgress

Follow

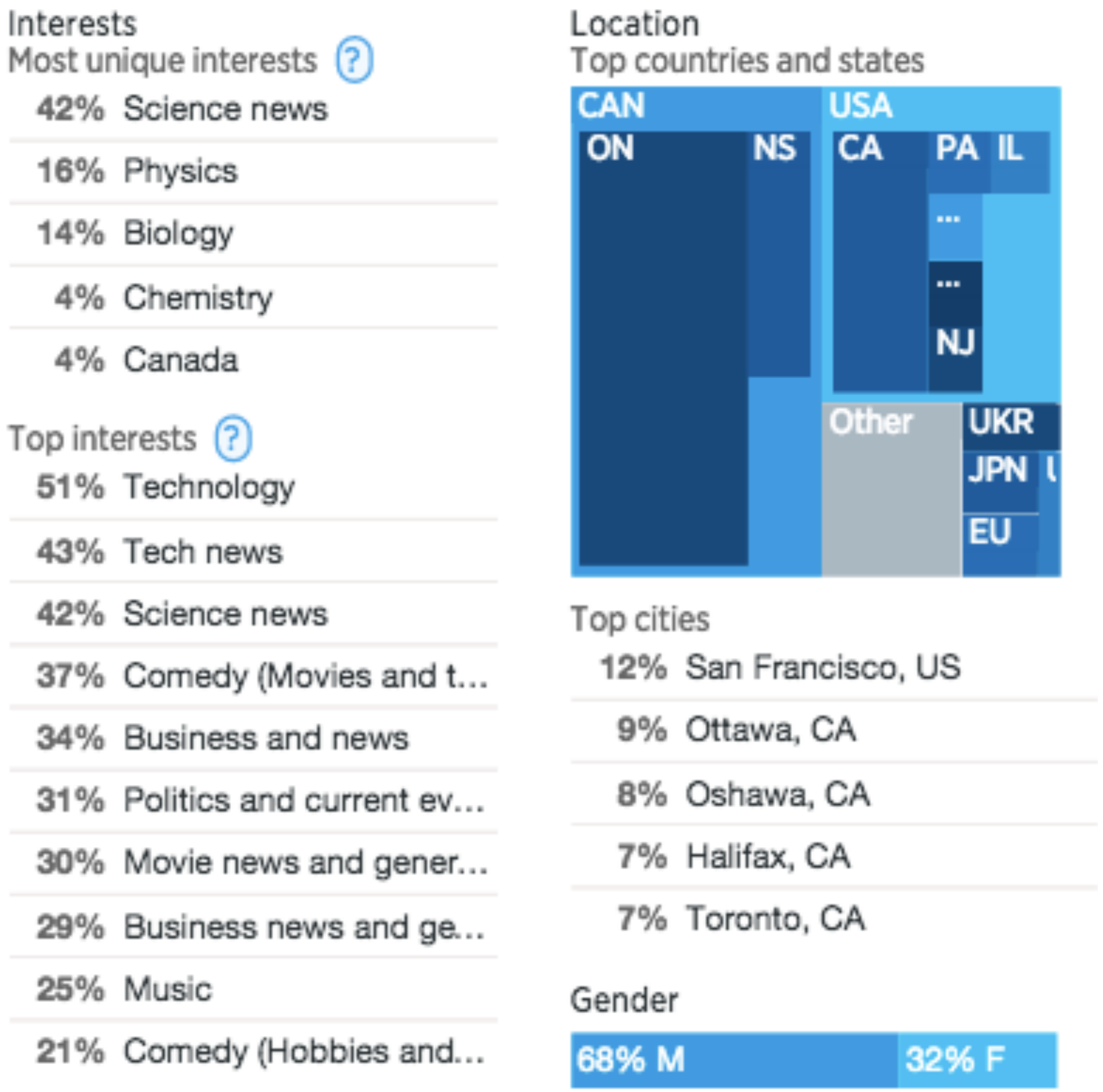
MT@Interpreter_Mag Поддерживаемые Россией сепаратисты заняли 28 городов и деревень с февраля bit.ly/1JqJ2Qp

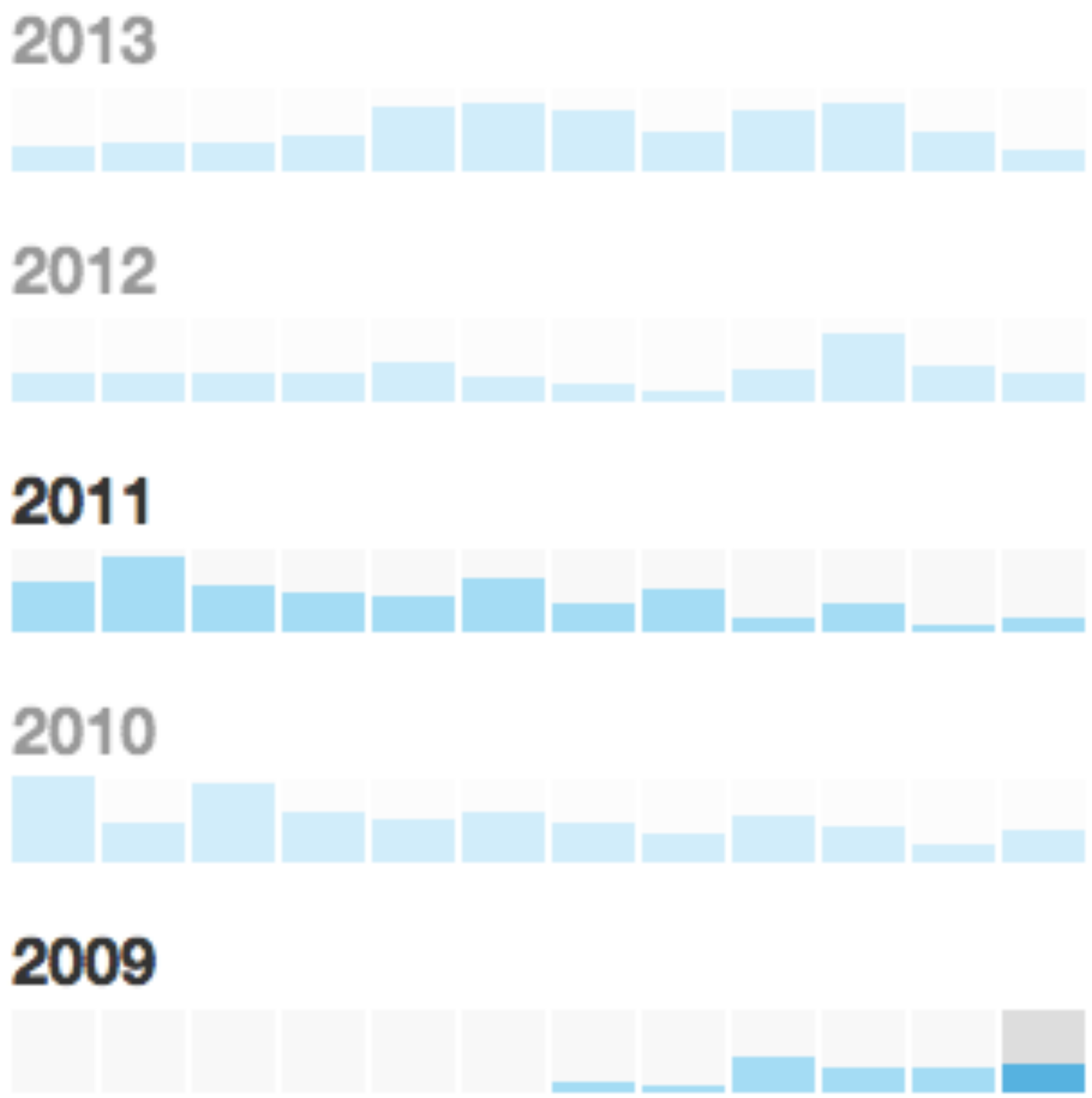
View translation

RETWEETS	FAVORITES
8	2



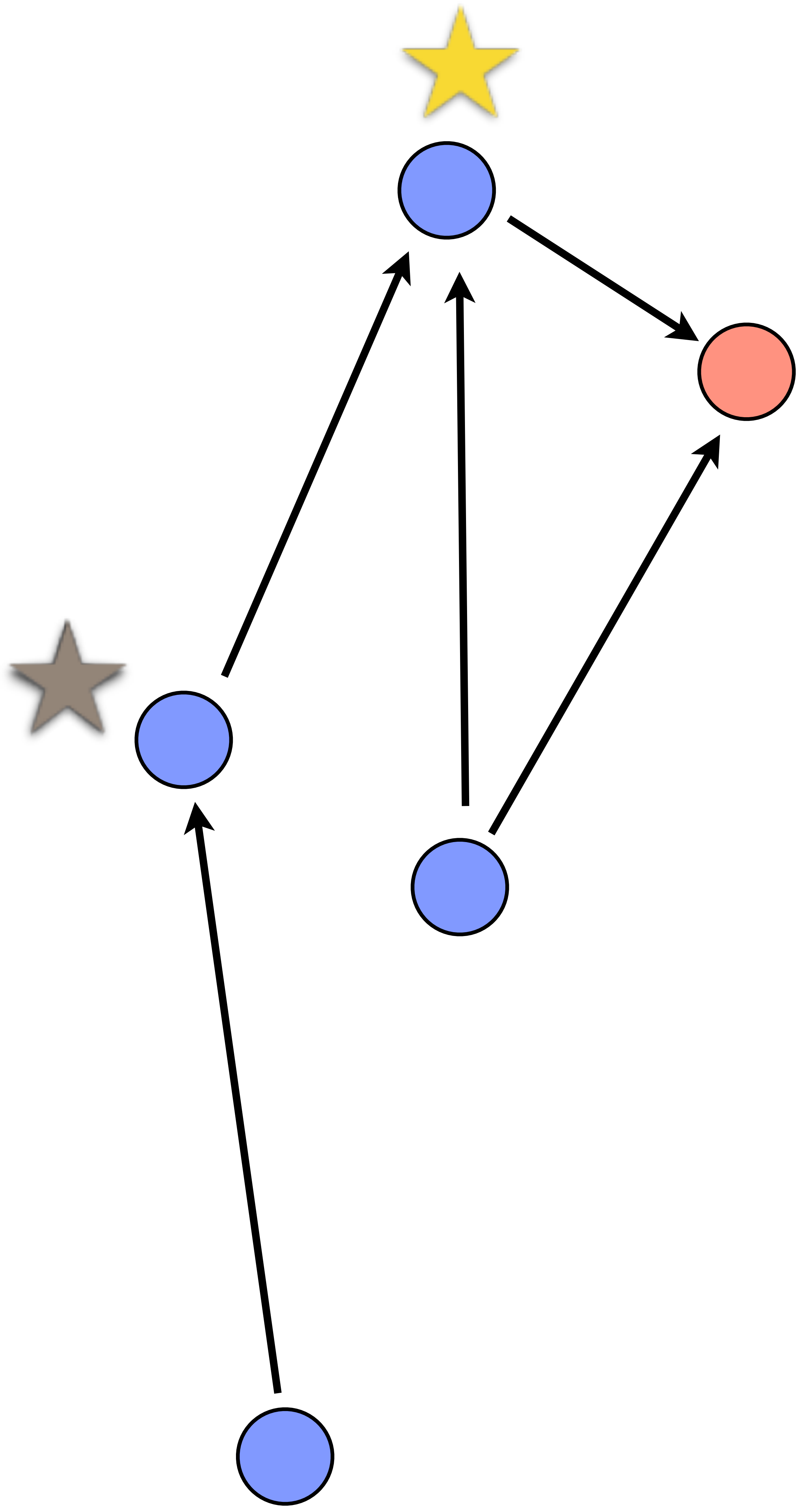
User behaviour changes over time





What events are possible?

- join
- tweet
- follow
- retweet / mention



- Upon follow: you can view and re-transmit information
- Your followers can do the same
- hashtags in the tweets, ‘#physics’ is an example

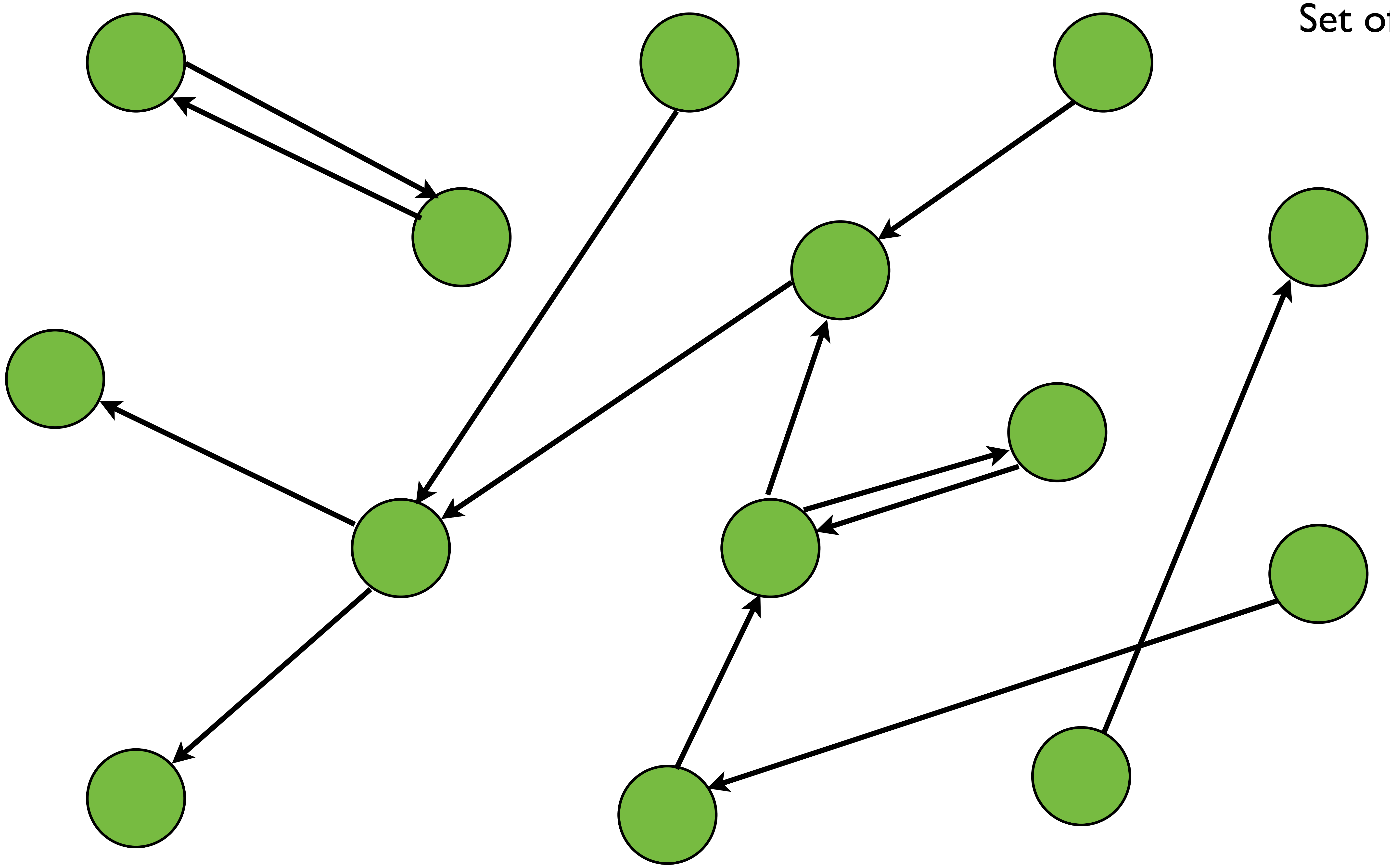
- Different online social networks have different rules

Introduction to networks/graphs

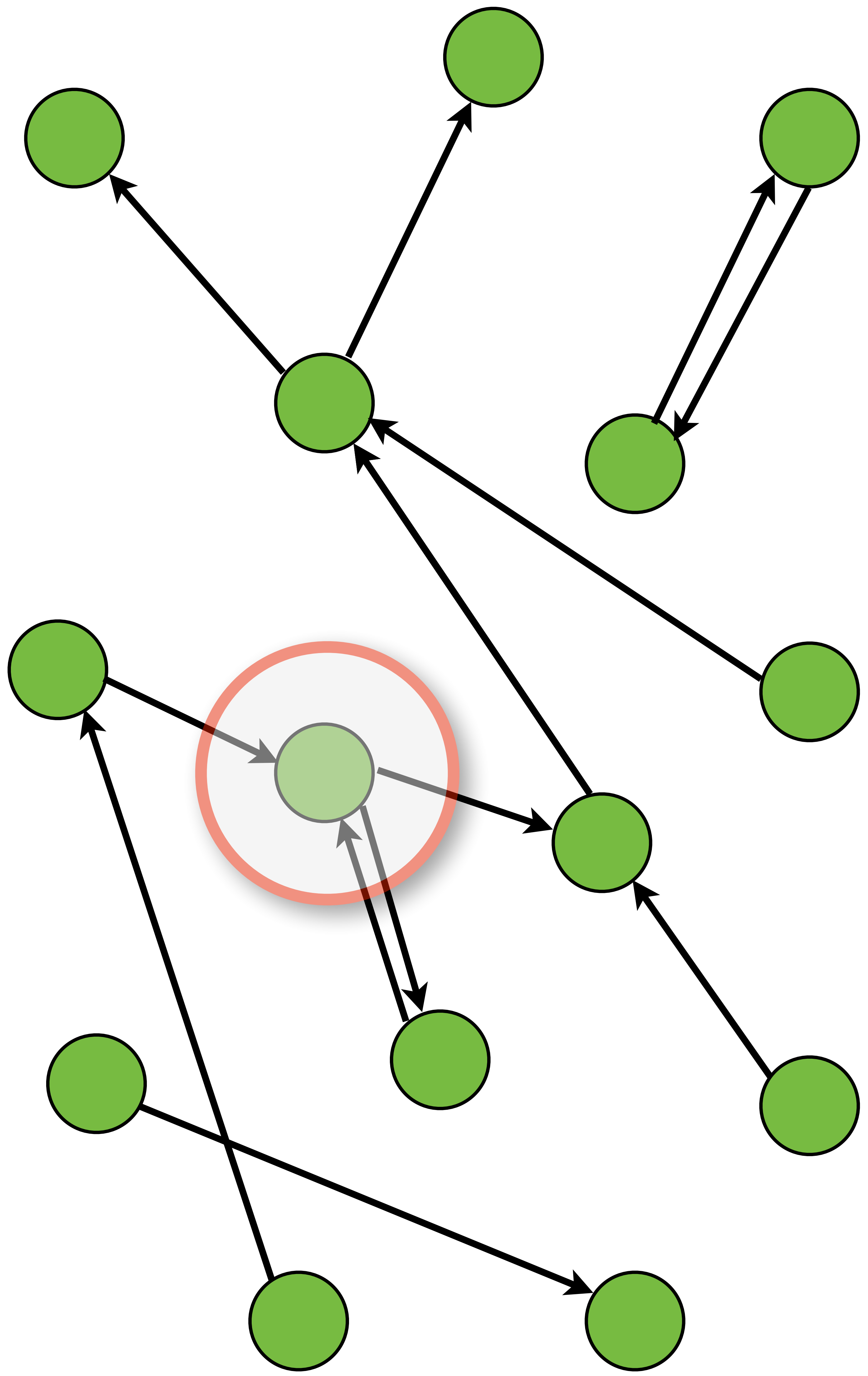
$$G(N, E) = \{N\} + \{E\}$$

Set of nodes -

Set of edges -



Degree of a node



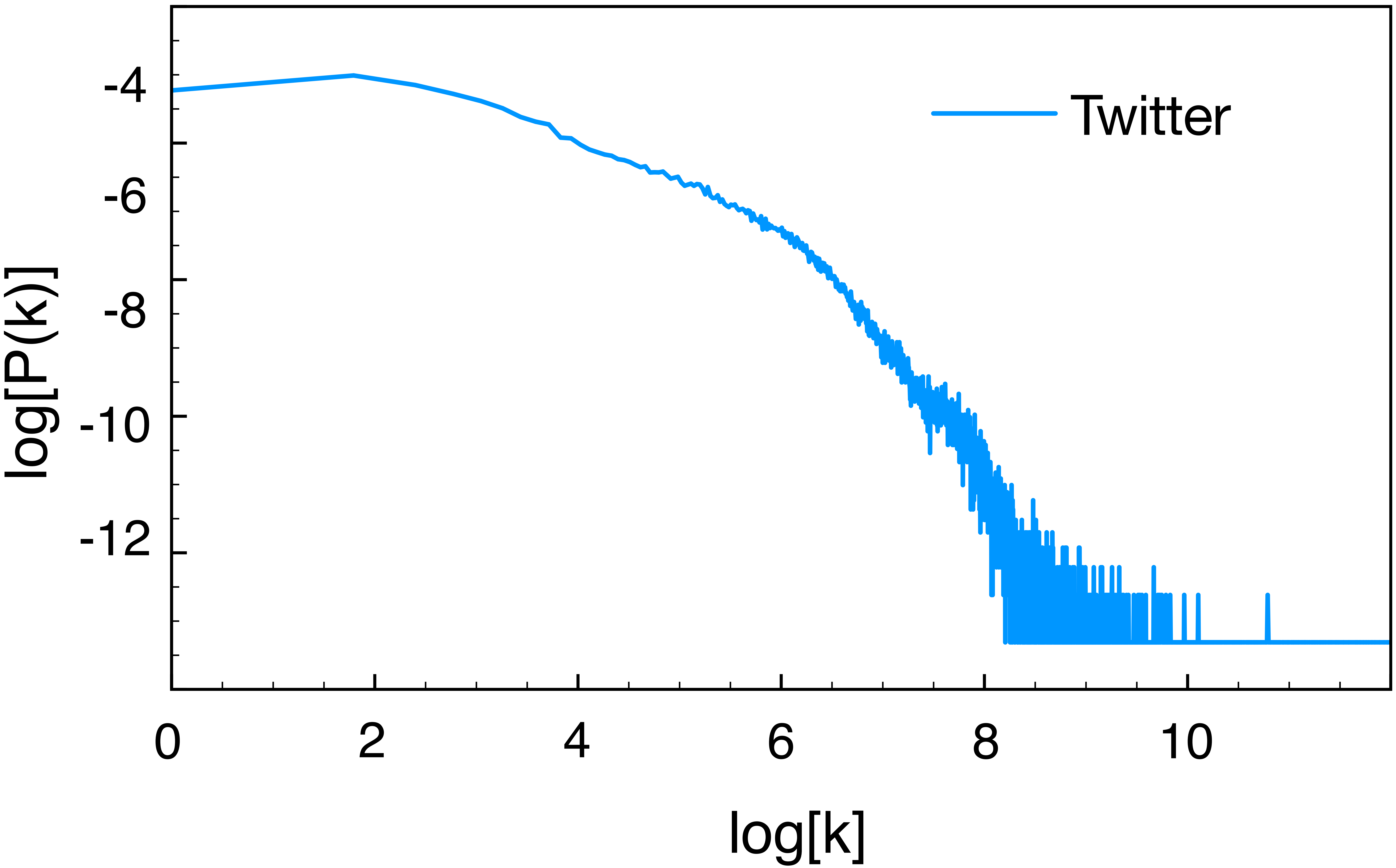
- How many connections does a selected node have?

  Example:

- In-degree: 2
- Out-degree: 2
- Cumulative degree: 4

- If we do this for all nodes we obtain a degree distribution.

Degree distribution



Our code #k@
http://hashkat.org

- Written mostly in C++, with some Python and Lua.
- Source ~ 7000 lines of code
- Repository consists of ~80 files
- Full documentation and tutorials are available online
- Obtain a random network with 5 M nodes in about an hour and 30 M (128 Gb RAM) nodes in about 42 hours.
- Free & Open Source (GPLv3)
- Agent based model

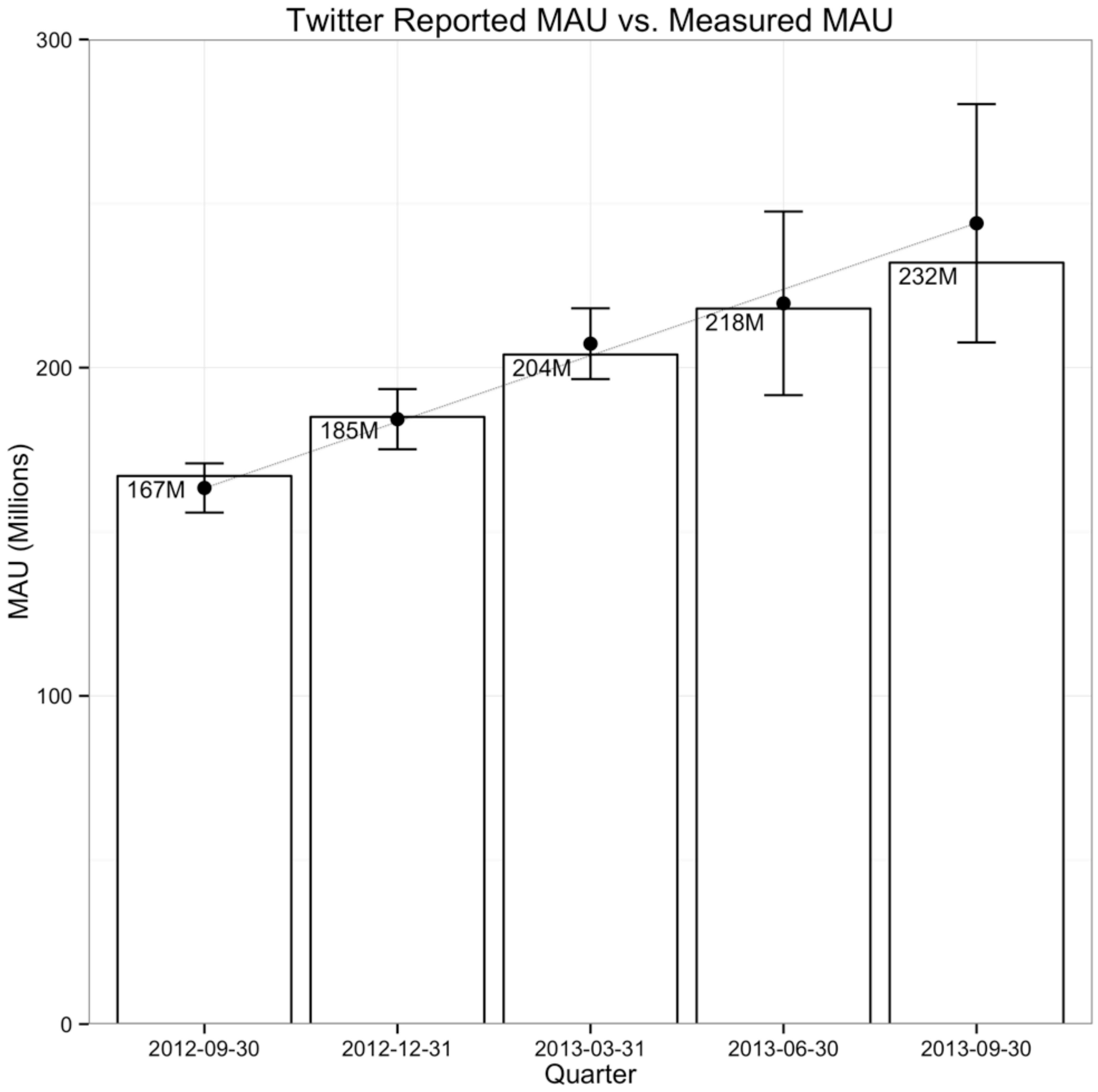
Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

- id = 1
- Agent type
- Number of tweets, retweets
- Region
- Creation time
- Language
- Ideology
- Humour preference
- List of chatty people
- Following and follower set

Experimental guidance



sampling error reflects collection method

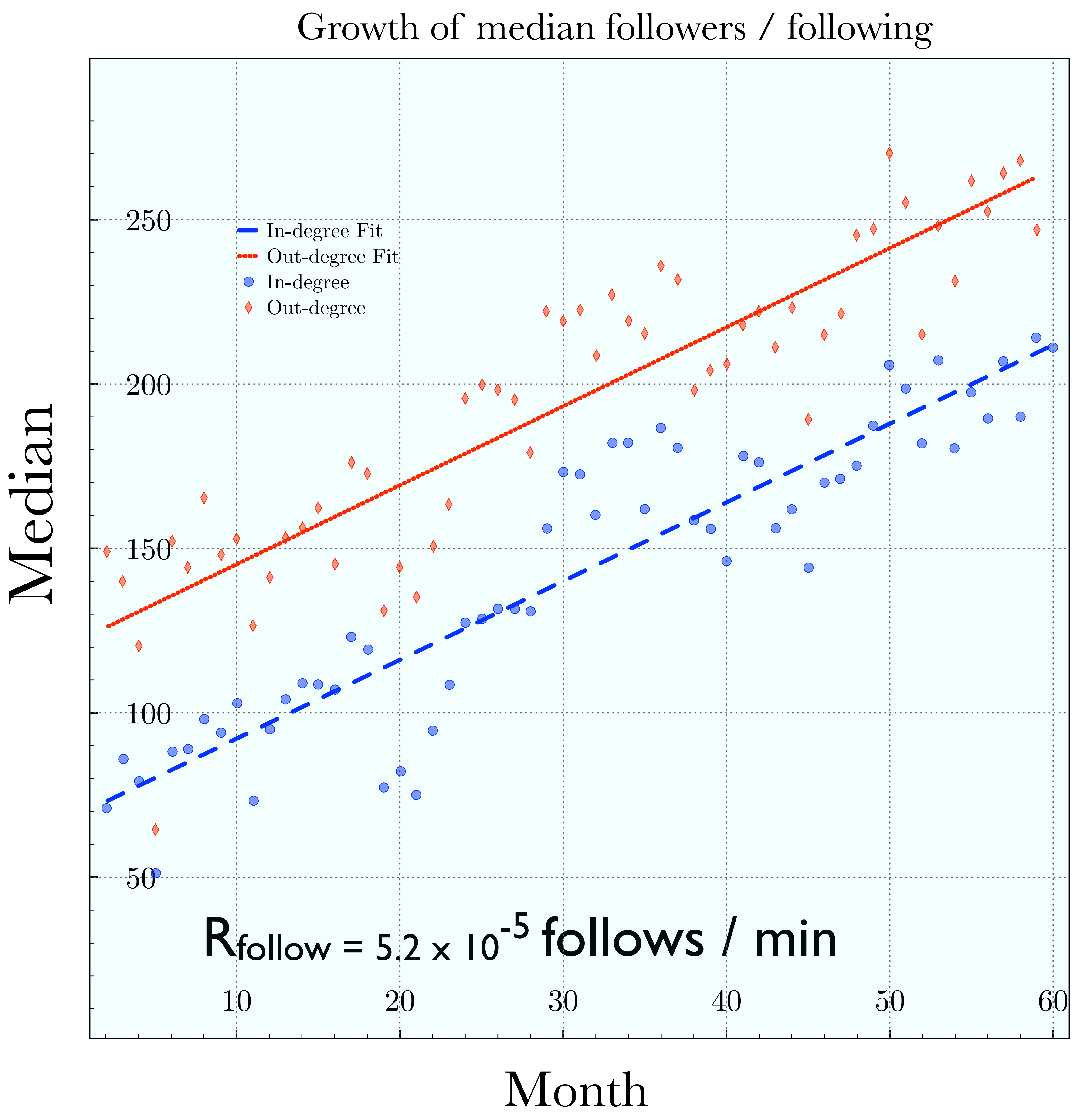
R_{add} = 123.7 users/min

Size of the graph (IPO)
- 300 million

Tweets / day (IPO)
- 500 million

K. White, private comm

Experimental guidance



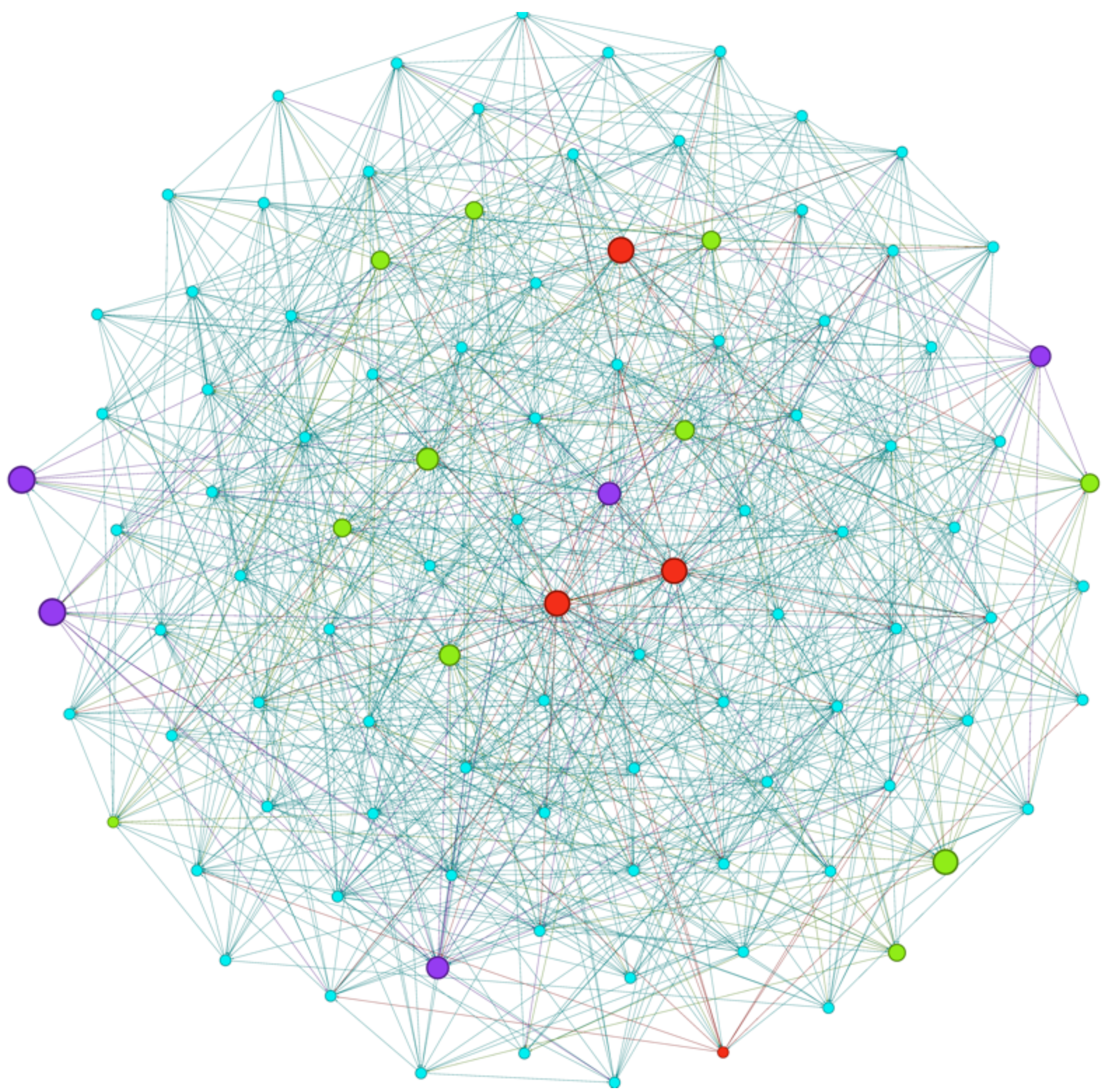
- Users on Twitter tend to follow more than being followed.
- The rates for followers and followings is roughly equal.
- The follow rate is much smaller.

K. White, private comm

Simulation run through

R_{add}	R_{follow}	R_{tweet}

Producing a random graph



- With probability p, place an edge between two nodes.

Total number of edges: $$E = \frac{N(N-1)}{2}$$

Total number of expected edges: $$E = p\left(\frac{N(N-1)}{2}\right)$$

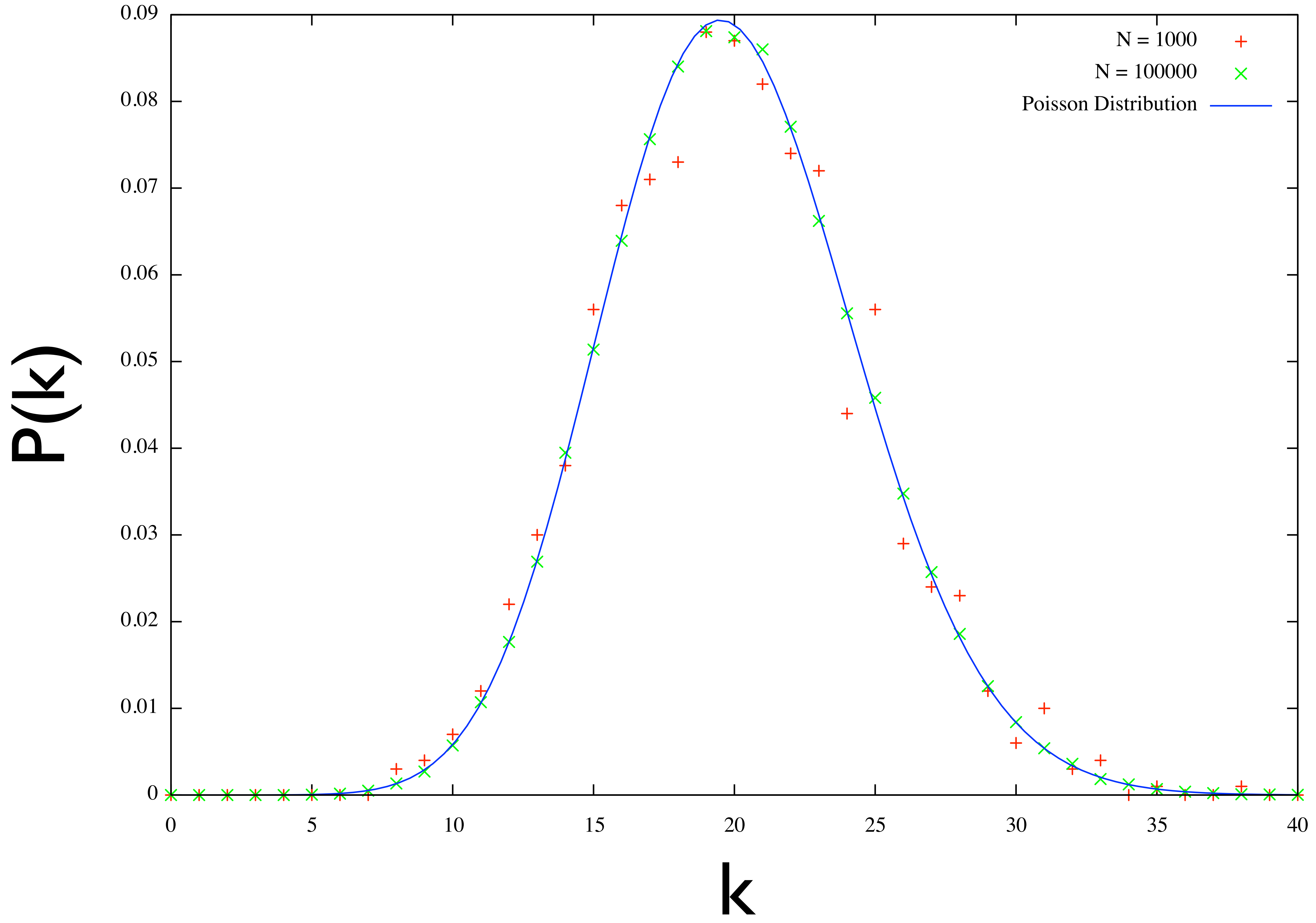
Example:

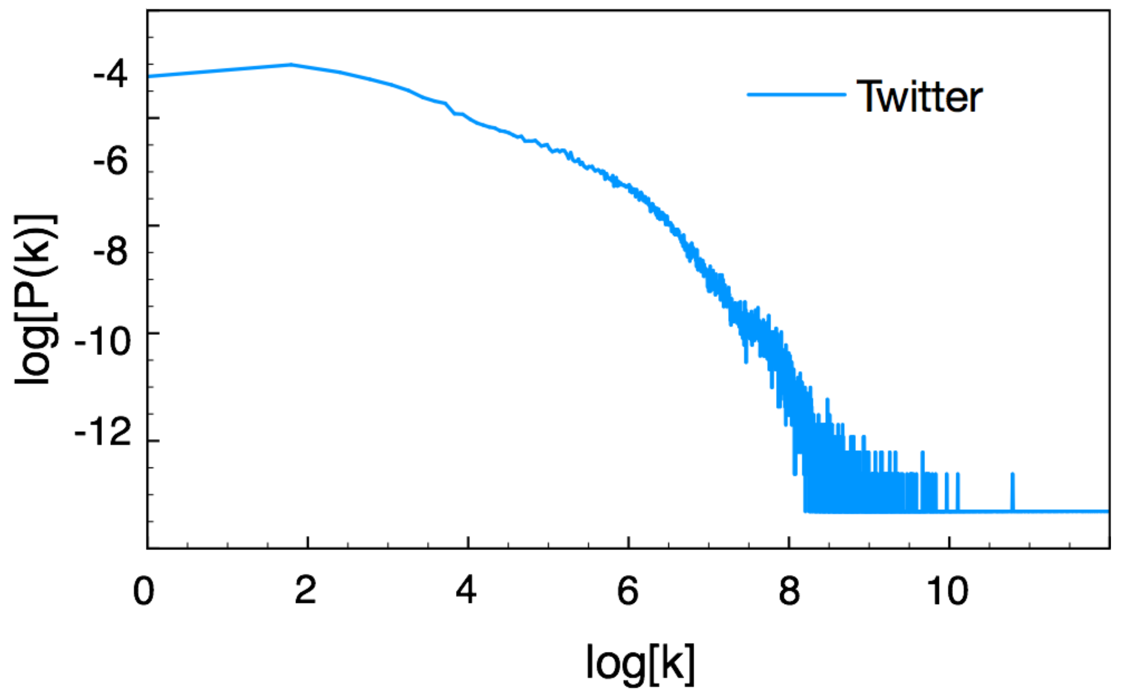
$$N = 10$$
$$E = (10 \cdot 9)/2 = 45$$
$$\text{if } p = 0.2$$
$$E(0.2) = 9$$

Degree distribution of a random graph

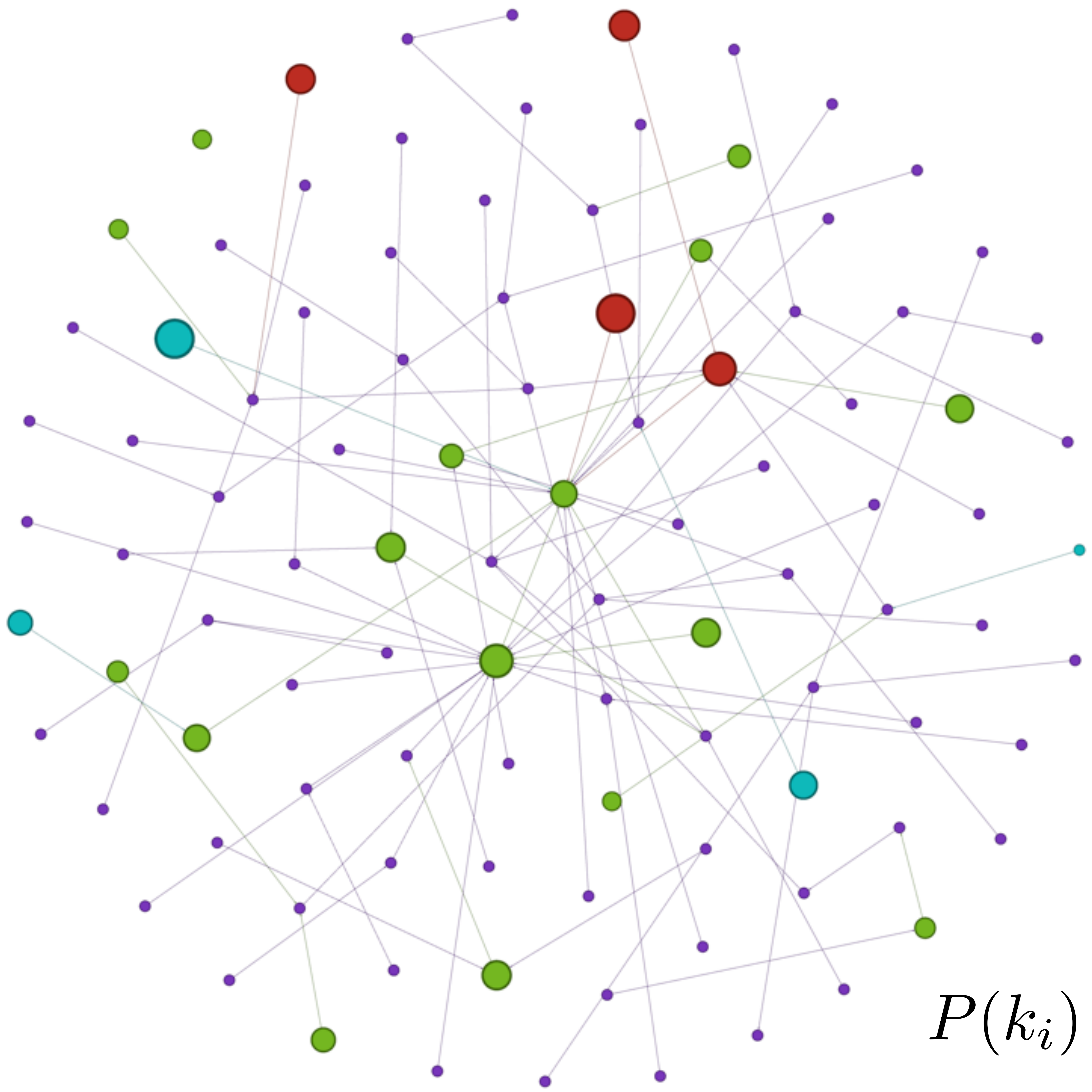
typo

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{\langle k \rangle^k e}{k!}$$





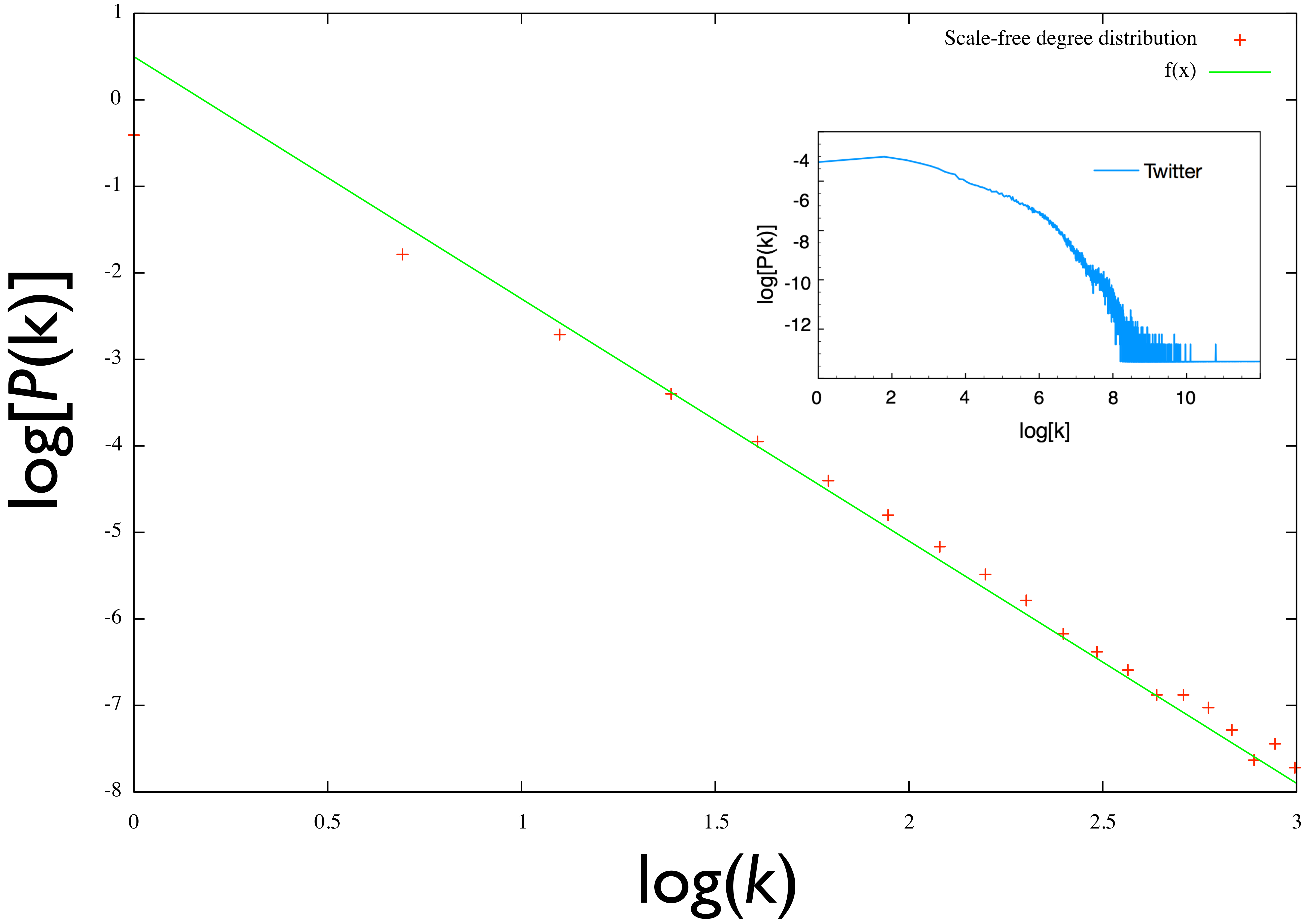
2-Following via preferential attachment



$$P(k_i) = \frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

Degree distribution of a scale-free network

$$P(k) \propto k^{-\gamma} \text{ where } \gamma \approx 3$$

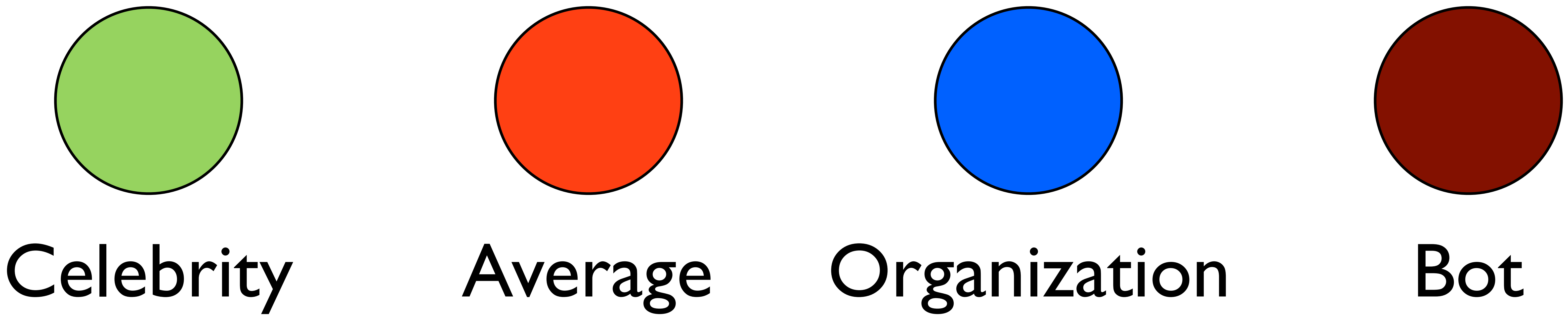


Following another entity

1. Random - follow randomly.
2. Preferential - follow based on the in-degrees of the entities.
3. Entity - follow based on their <u>exterior</u> titles.

3-Follow by entity type

- Follow an entity based on the exterior title.
- On Twitter, there are different types of users



@Pontifex

Celebrity

76 tweets,
+ 5 million
followers

@Canada

Organization

40,000 followers on 1st day

Average

Popularity and influence comes from activity *within* the network

Bot

Following another entity

1. Random - follow randomly.
2. Preferential - follow based on the degrees of the entities.
3. Entity - follow based on their exterior titles.
4. Preferential-entity - look to the exterior title first, and then follow based on the degrees of the specific entity types.
5. Message content - based on messages tweeted by entities

Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

- id = 1
- Agent type
- Number of tweets, retweets
- Region
- Creation time
- Language
- Ideology
- Humour preference
- Music preference
- List of chatty people
- Following and follower set

Network = | Agent 0 | Agent 1 | Agent 2 | Agent 3 | ... | Agent N |

Array of structs

- id = 1
- Agent type
- Number of tweets, retweets
- *Region*
- Creation time
- *Language*
- *Ideology*
- *Humour preference*
- *Music preference*
- List of chatty people
- Following and follower set

Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

- id = 1
- Average Joe
-
- *Canada*
-
- *English*
- *Politics* = ***Blue***
- *Funny*
- *Likes Jay-Z*
-
-

Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

- id = 3
- Average Joe
-
- *Canada*
-
- *English*
- *Politics* = ***Red***
- *Funny*
- *Likes ABBA*
-
-

Agent 1 Agent 3

Agent 1 and 2 share the same geographical location. They both appreciate funny things. They differ politically and musically, however.

Agent 1 can generate funny, *apolitical* tweets

Agent 3 may appreciate those tweets. She may follow Agent 1 and rebroadcast such content

Agent 1 Agent 3

Agent 1 and 2 share the same geographical location. They both appreciate funny things. They differ politically and musically, however.

Agent 1 can also generate politically charged tweets

Agent 3 will not rebroadcast these, as they are inconsistent with her political views

Agent 1

Agent 3

Agent 1 and 2 share the same geographical location. They both appreciate funny things. They differ politically and musically, however.

Agent 1

Agent 3

these agents can communicate because they *share the same language*

Agent 1 can also label tweets with #tags

A political tweet related to Canadian politics	#poli #cdn
…or music	#music
…or a joke	#funny

Network =

Agent 4	Agent 5	Agent 6	Agent 7	...	Agent N

Array of structs

- id = 7
- Average Joe
-
- Australia
-
- *English*
- *Politics* = ***Blue***
- *Funny*
- *Likes Jay-Z*
-
-

Agent 1 can also label tweets with #tags

#poli #cdn

…or #music

…or #funny

Agent 7 lives in a different region, thus she is interested only in #poli #aus even though both are Blue

Agent 1

can also label tweets with #tags

#poli #cdn

…or

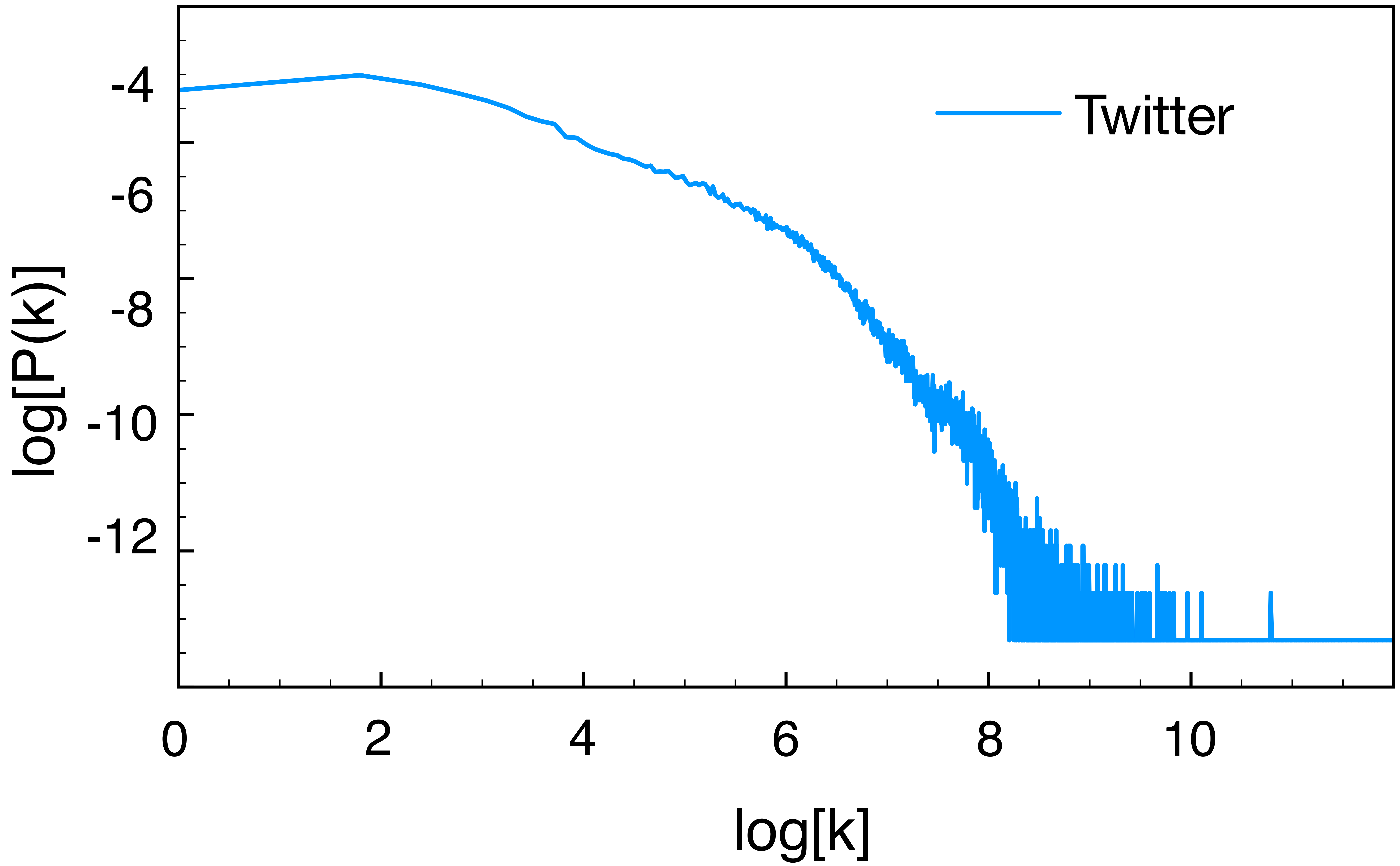
#music ✓

…or

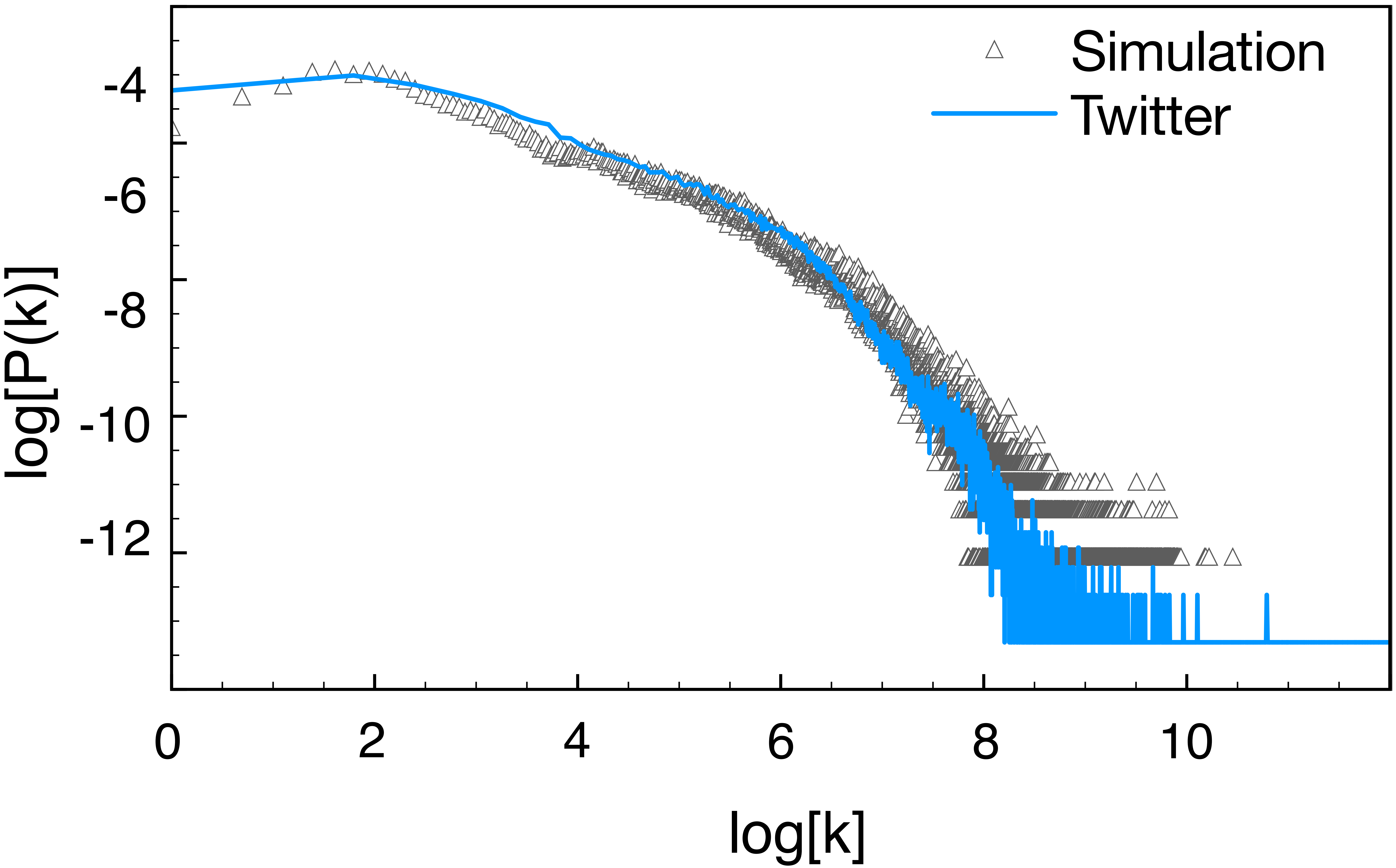
#funny ✓

Agent 7

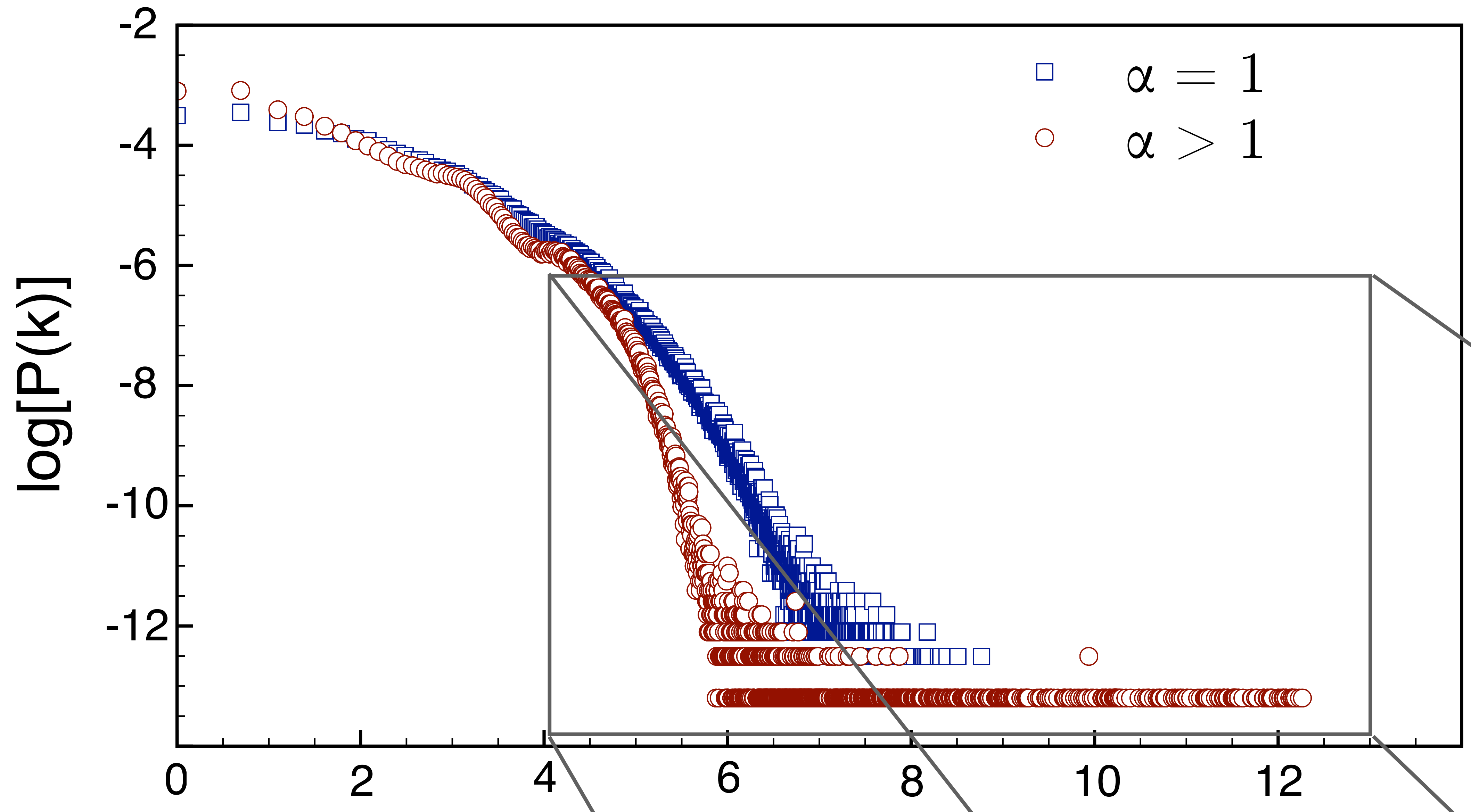
these agents agree on musical tastes and humour

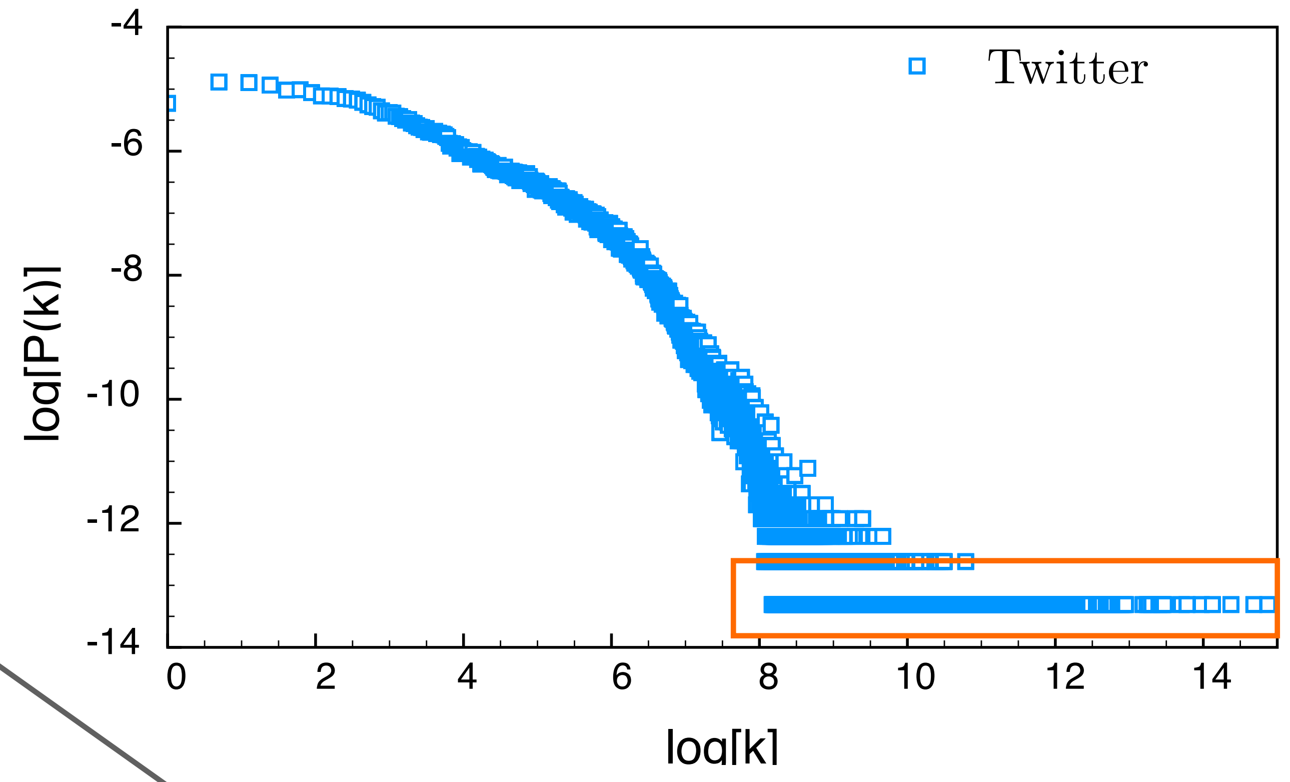


Twitter
-4
-6
-8
-10
-12
log[P(k)]
0
2
4
6
8
10
log[k]

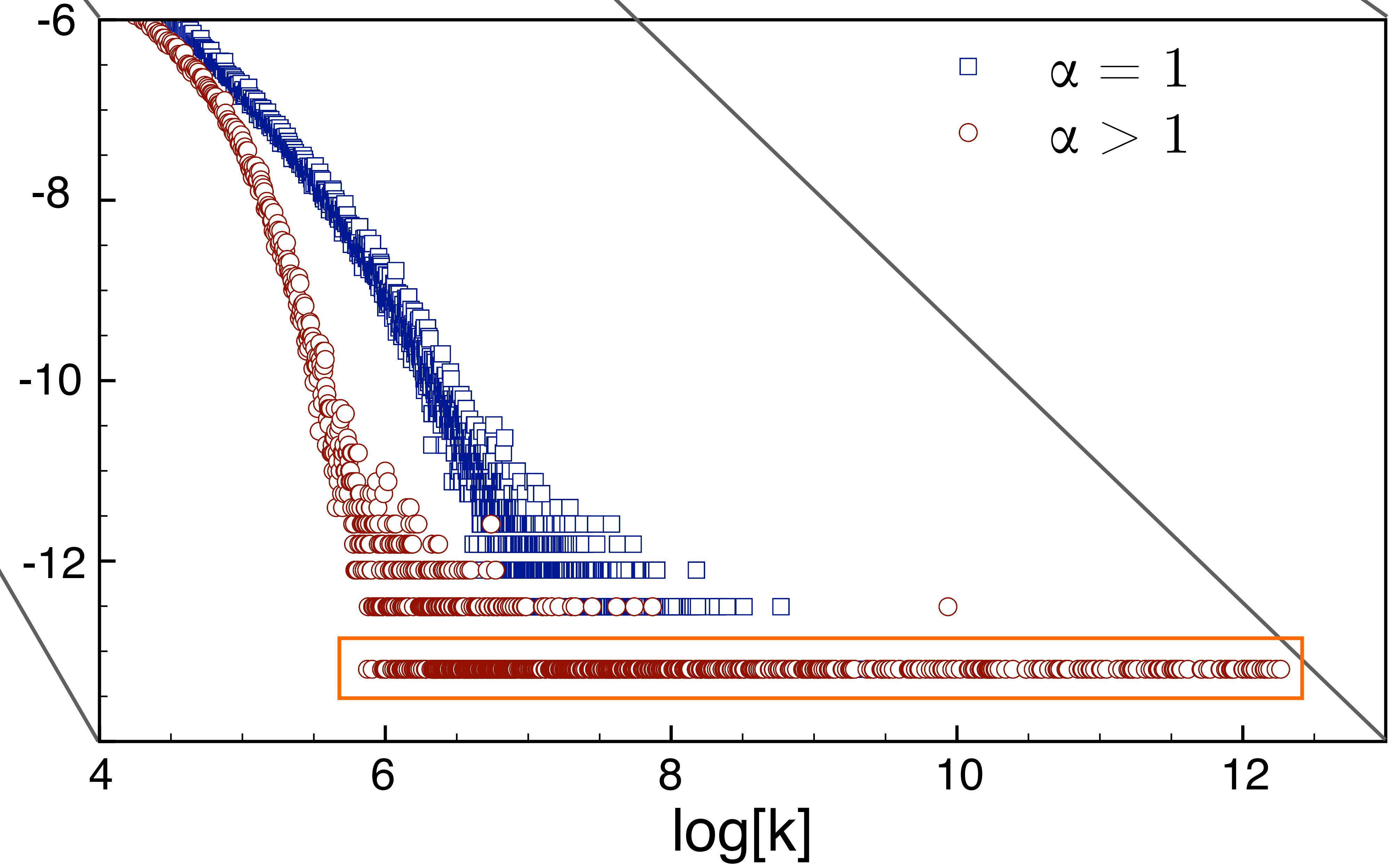


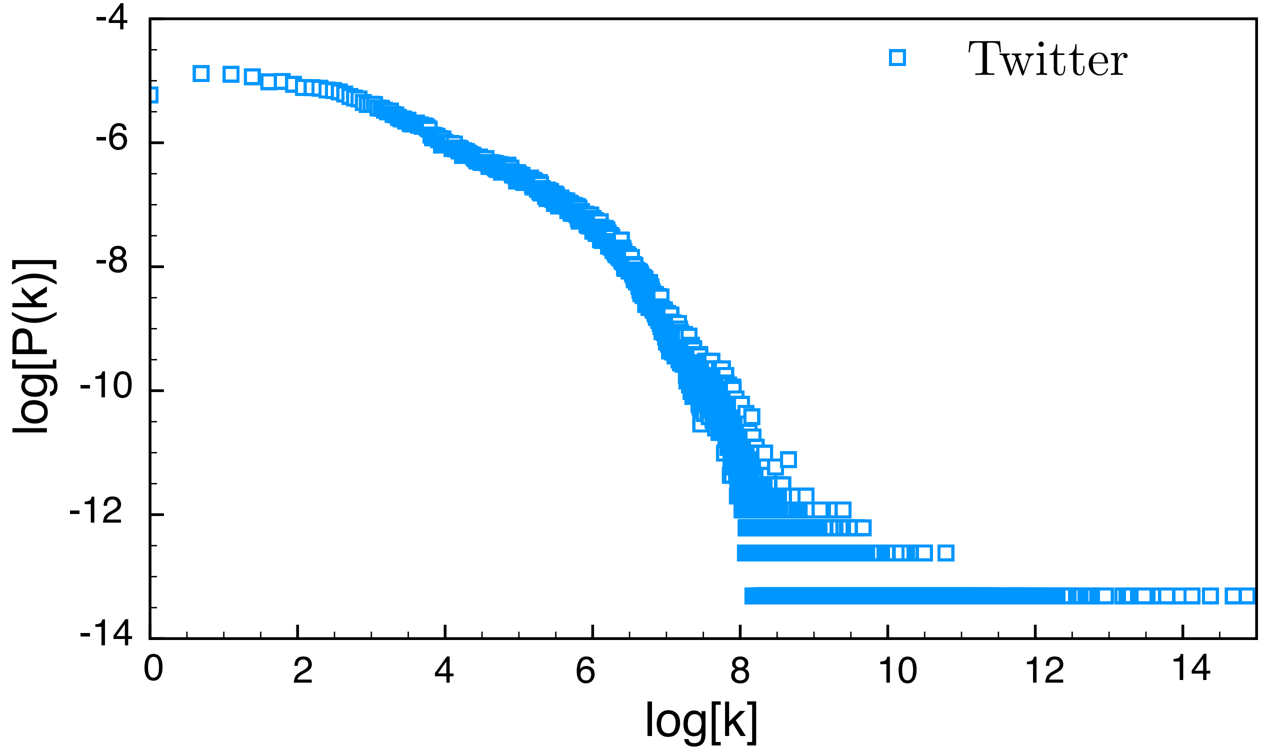
Simulation
Twitter
-4
-6
-8
-10
-12
0
2
4
6
8
10
log[P(k)]
log[k]

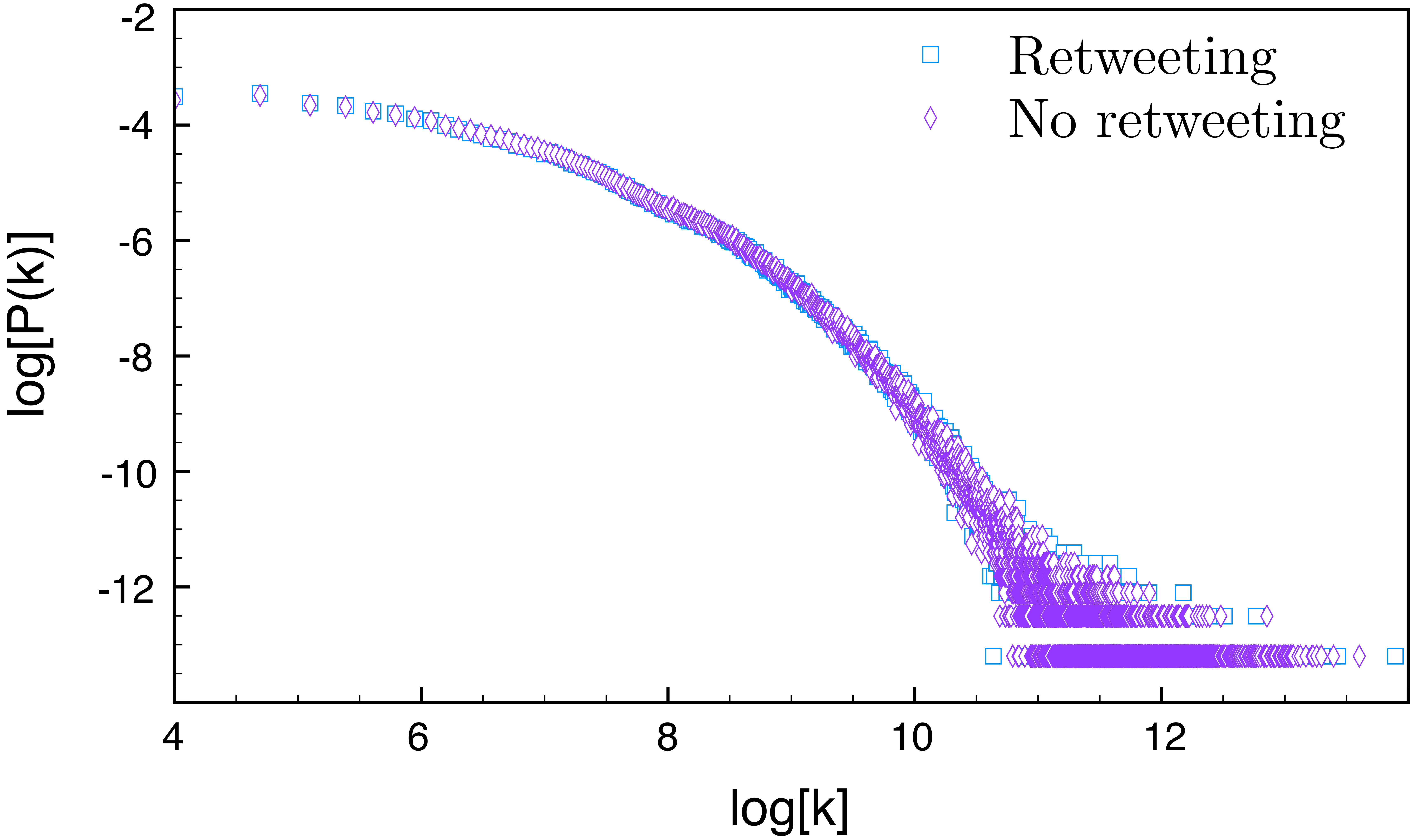




1) Gamma is the probability to follow node i with degree k, it is the mathematical form for preferential attachment.
2) I am zooming into the tail,




Twitter
log[P(k)]
log[k]


Retweeting
No retweeting
log[P(k)]
log[k]

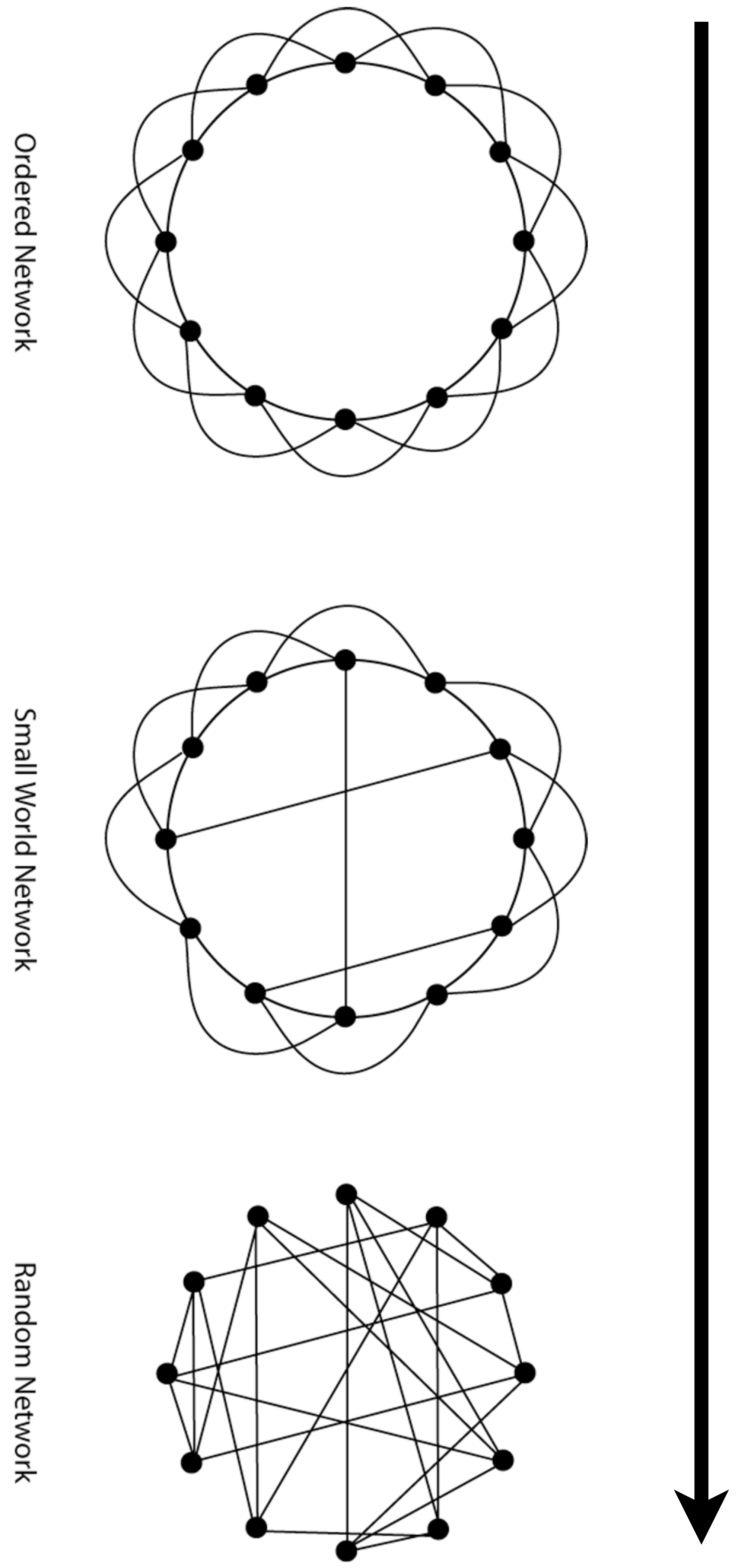
Conclusions & Acknowledgements

- Simulation accurately describes growth of Twitter
- Retweeting is fun, but not necessary
- Chattiness is in the eye of the beholder
- Models are very general: panspermia, religions, yawning, laughter, artistic styles, disease, political ideology - anything where information can be rebroadcast
- http://hashkat.org

NSERC
CRSNG

K. Ryczko
A. Domurad
K. White

Generating a small world network



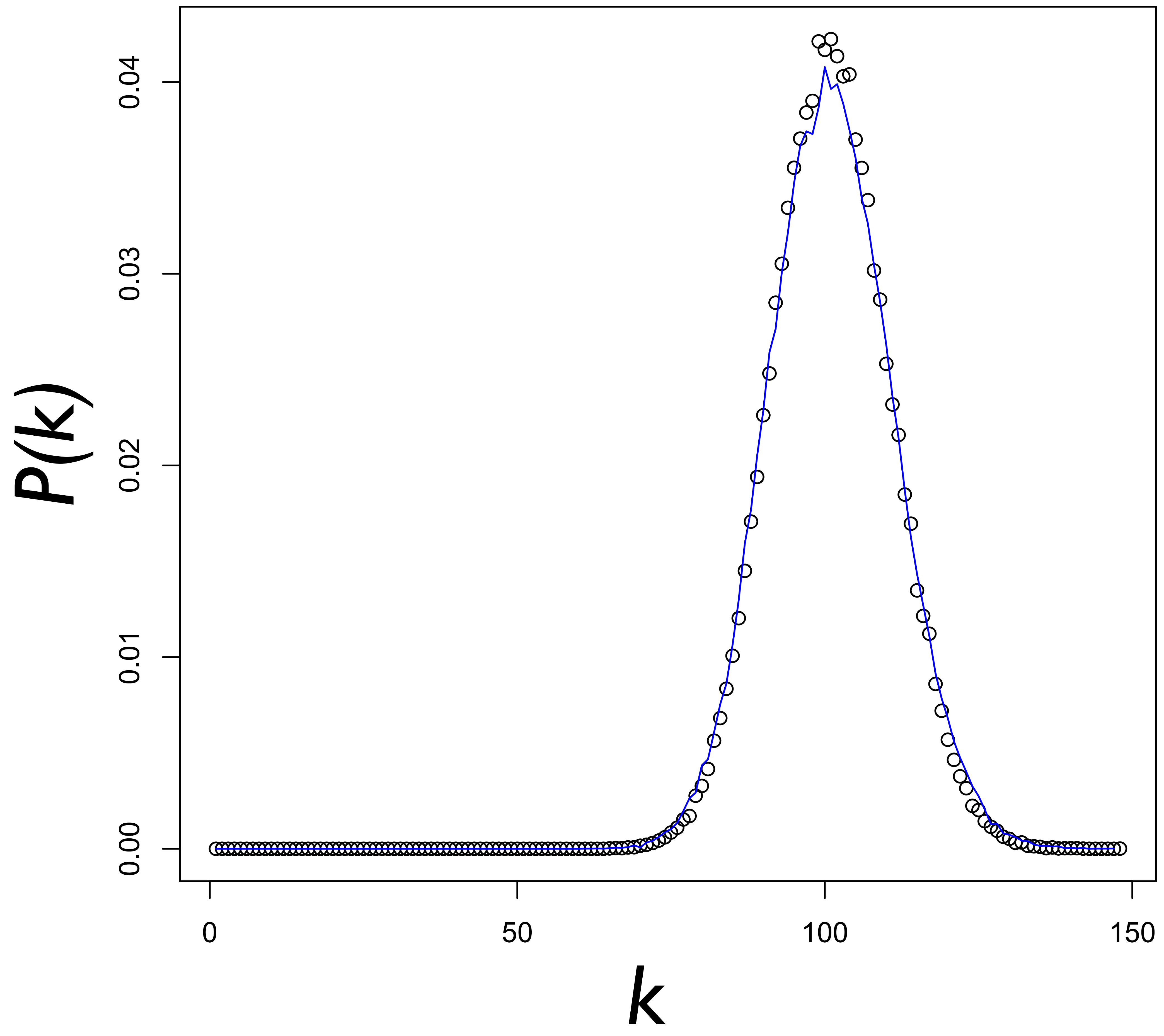
$p = 0$

$0 < p < 1$

$p = 1$

- Begin with a highly ordered network where every node has the same degree.
- With a probability p, reconnect the edges randomly.
- For some p less than 1 and greater than 0, you will generate a small world network.

Degree distribution of a small world network



- Very similar to a random graph.
- The physical interpretation is there is a threshold on the number of connections a node can have.