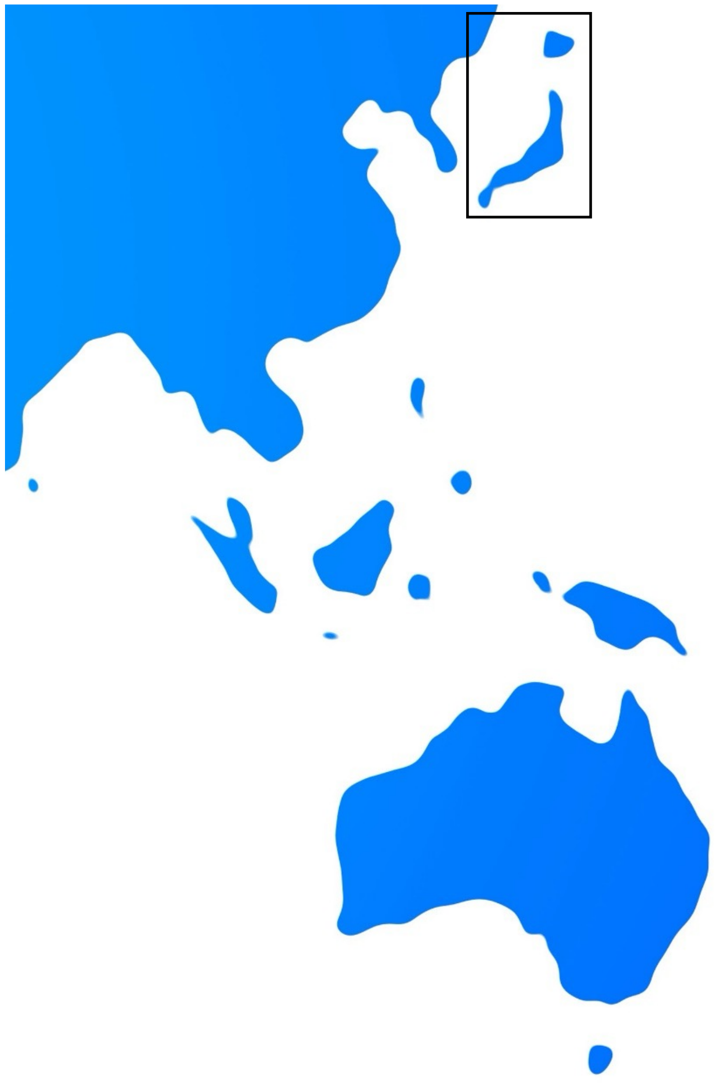
Applications of data science and machine learning to organic materials

Daniel Packwood

iCeMS KYOTO UNIVERSITY WPI Research Center

KYOTO UNIVERSITY KYOTO JAPAN FOUNDED 1897

Self-introduction

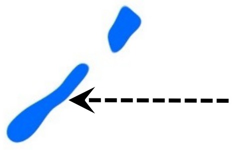


Senior lecturer and PI at iCeMS, Kyoto University (2016 - 2023)

Associate Professor (2023 -)

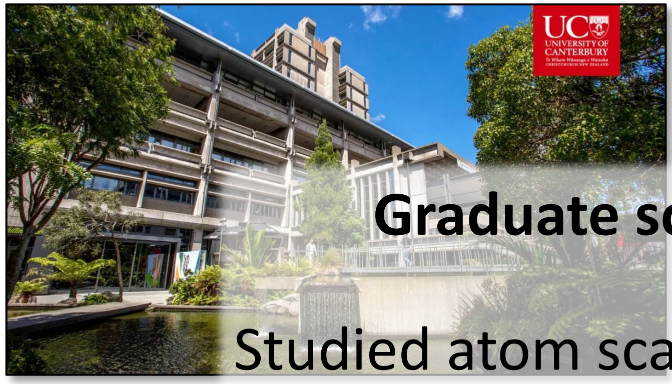
Assistant Professor, Mathematical Sciences Unit, AIMR, Tohoku University (2012 – 2016)

JSPS Postdoc, Quantum Chemistry Lab, Dept. of Chemistry, Graduate School of Science, Kyoto University (2010-2012)



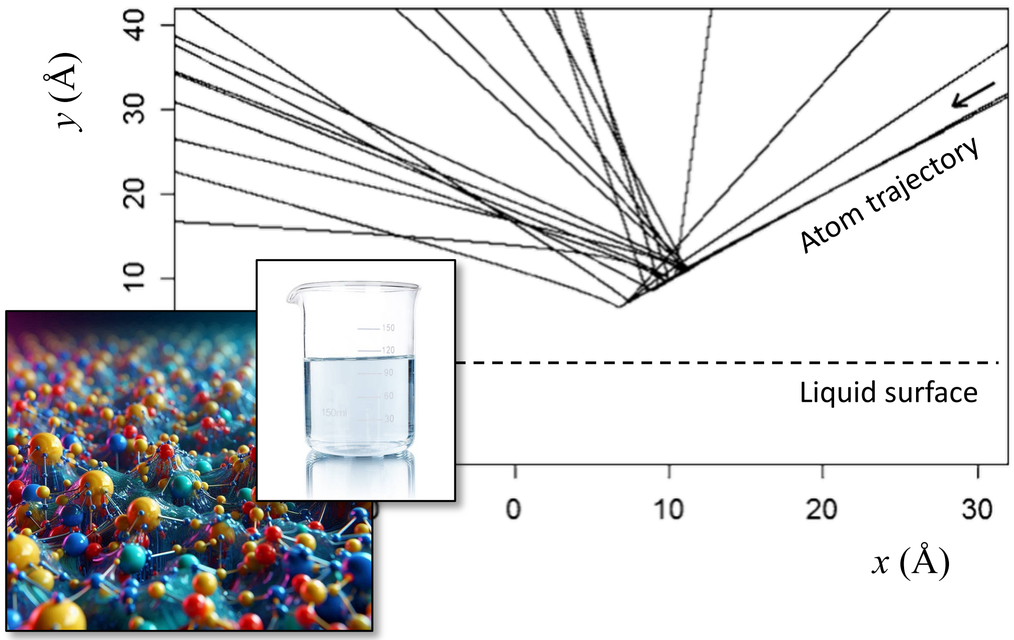
University of Canterbury (PhD 2010)
Major: Chemistry, Minor: Statistics

Why did I come to Japan?



Graduate school research (late 2000s)

Studied atom scattering from liquid surfaces using **non-equilibrium statistical mechanics** and **stochastic differential equations**.



Two names frequently appeared during my study:

Kiyoshi Ito

- Kyoto University mathematician
- **Stochastic differential equations pioneer**

Ryogo Kubo

- Tokyo University physicist
- **Non-equilibrium statistical mechanics pioneer**

Realization: my field was pioneered in Japan!

Why did I come to Japan?

HOPE MEETINGS
with Nobel Laureates

第2回HOPEミーティング開催概要

日時：2009年9月27日（日）～10月1日（木）
会場： ザ・プリンス箱根
テーマ：Art in Science
対象分野： 化学及び関連分野（物理学、生物学等）
主催： （独）日本学術振興会

https://www.jsps.go.jp/hope/gaiyou2.html

- Gathering of around 50 students from the Asia-Pacific region. Activities with Japanese graduate students and lectures from Japanese Nobel laureates.
- High level of research from the students impressed me, left me with a strong impression.

→ Go to Japan for postdoctoral research!

Career in Japan

Tanimura laboratory, Department of Chemistry, Grad. School of Science, Kyoto University

2010

Postdoc

2012

Mathematical Sciences Unit, Advanced Institute for Materials Research, Tohoku University

Assistant Professor

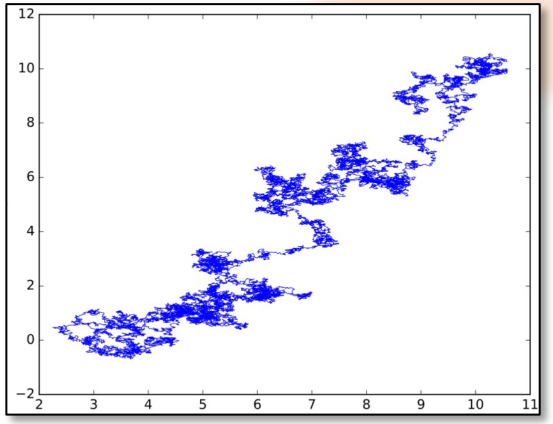
Phase 1: Stochastic processes

Institute for Integrated Cell-Material Sciences (iCeMS), Kyoto University

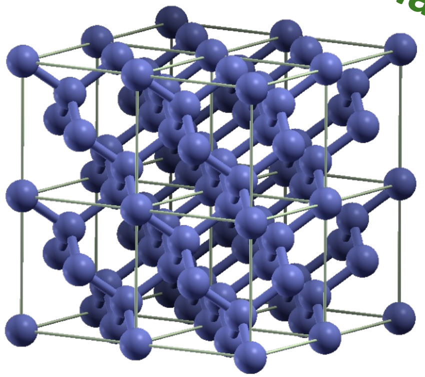
2016

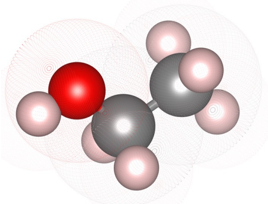
Lecturer

Associate Professor



Phase 2: Computational chemistry





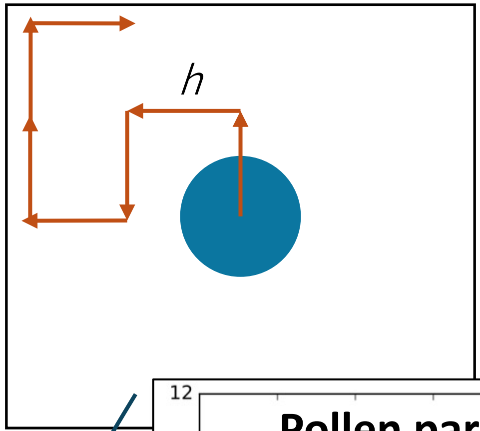
My previously research…

I studied **stochastic processes** (random walks) in physics and chemistry.

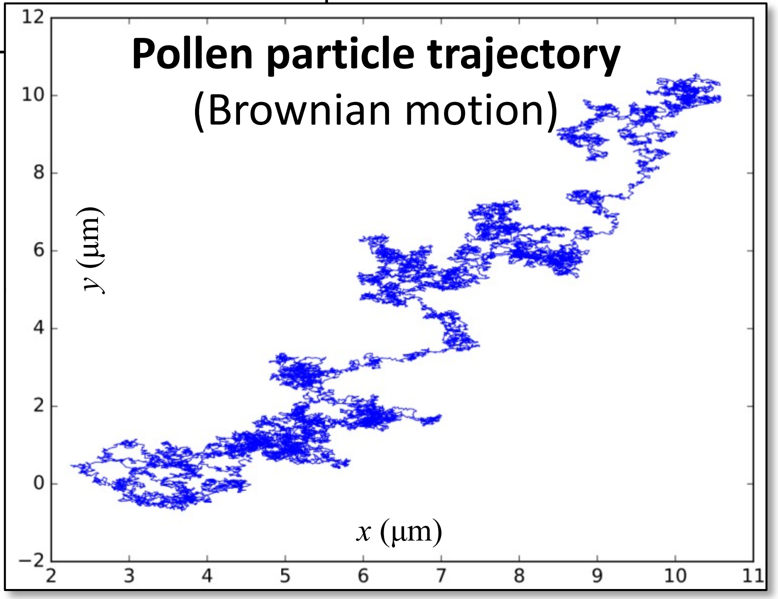
The most famous type of random walk in these fields is **Brownian motion**:

Pollen particles on water
Random motion due to collisions
with water molecules

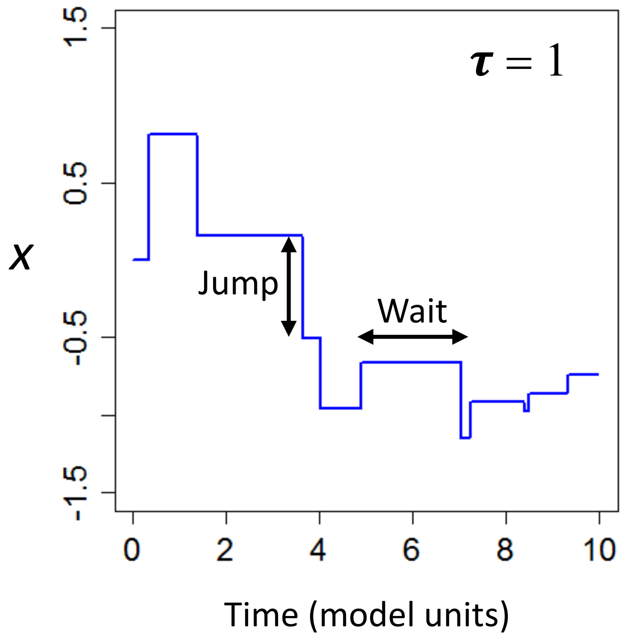
Model motion as
a random walk

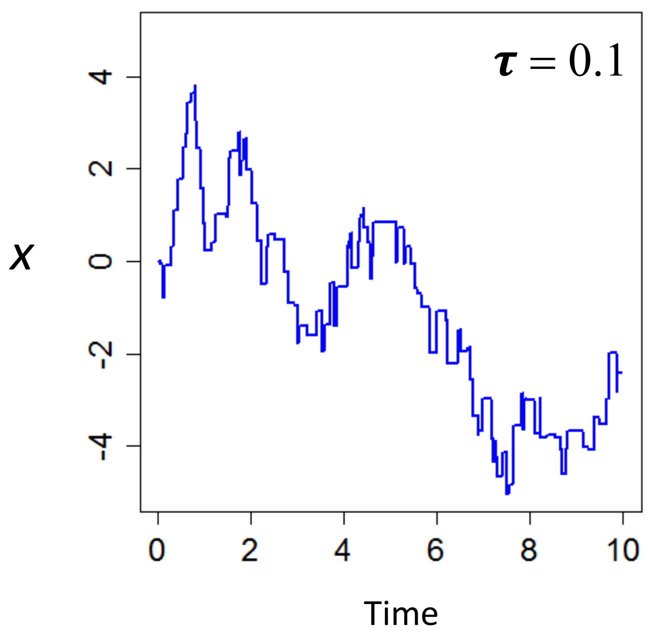


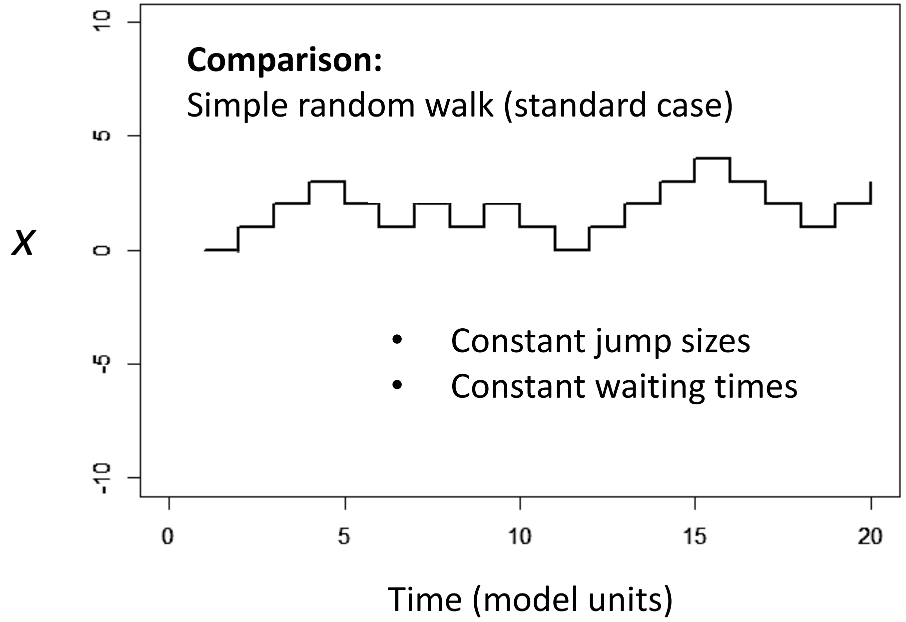
Take limit
$h \to 0$



Postdoc research (2010 – 2012)
Continuous-time random walk



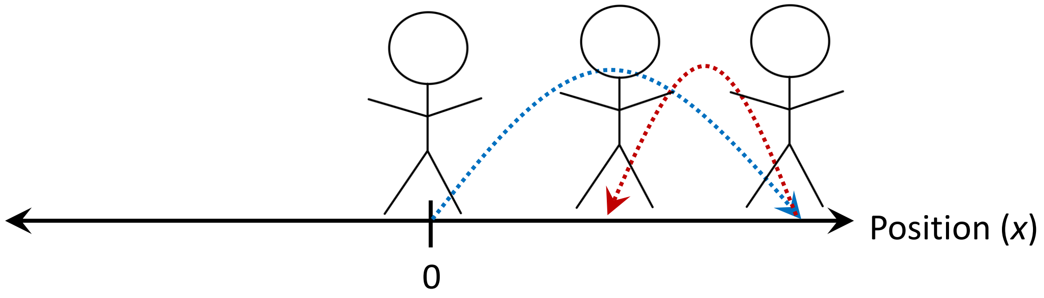




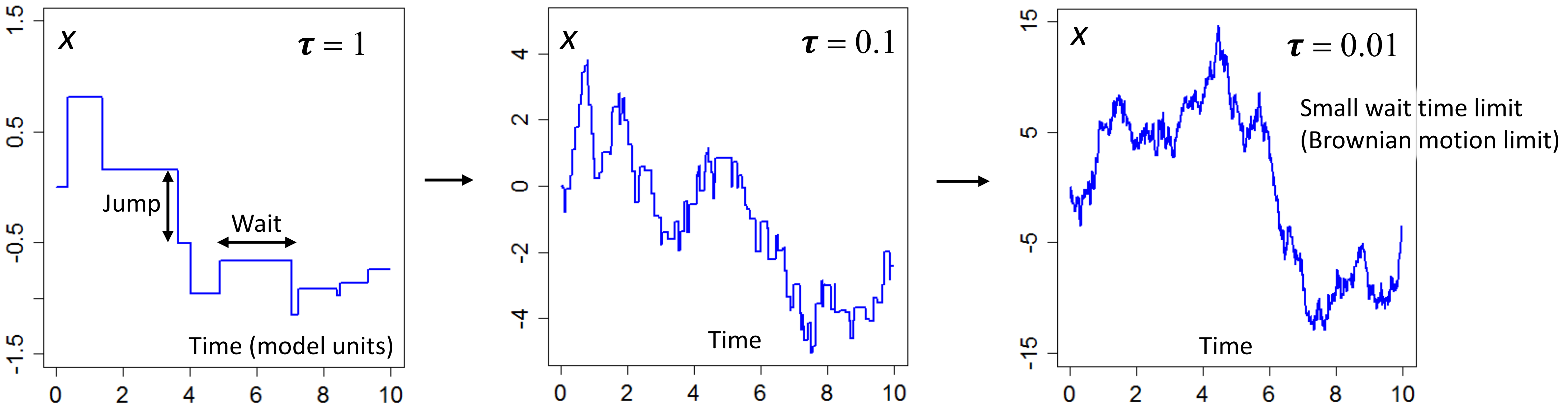
Consider a person hopping in one dimension.
Let *x* denote their position.

In the **continuous-time random walk**,

- the waiting times between jumps are random (average waiting time = τ), and
- the jump size is random



Typical results



1. Derivation of conditions for convergence to Brownian motion as wait times go to zero.

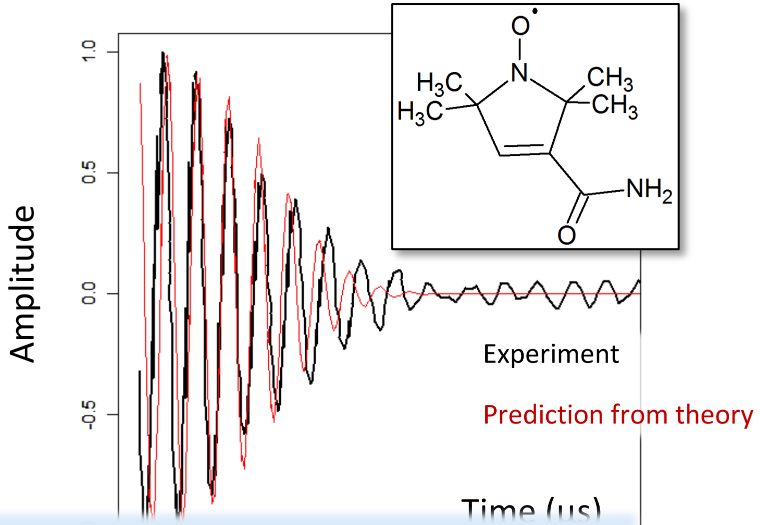
(the maximum jump size M must become $(3\tau)^{1/2}$)

J. Phys. A.: Math. Theor. **43**, 2010, 464001
arXiv: 1105,6283 (2011)
Phys. Rev. E. **84**, 2011, 61111

2. Derivation of the linear response function (important in spectroscopy)

Phys. Rev. E. **86**, 2012, 11130

$$F(t) = e^{-t/\tau} e^{i\omega_0 t} \exp\left(\frac{1}{\tau M}\int_0^t \frac{\sin r/\tau}{r} dr\right)$$



Fun, but a very narrow topic…

Career in Japan

Tanimura laboratory, Department of Chemistry, Grad. School of Science, Kyoto University

2010

Postdoc

2012

Mathematical Sciences Unit, Advanced Institute for Materials Research, Tohoku University

Assistant Professor

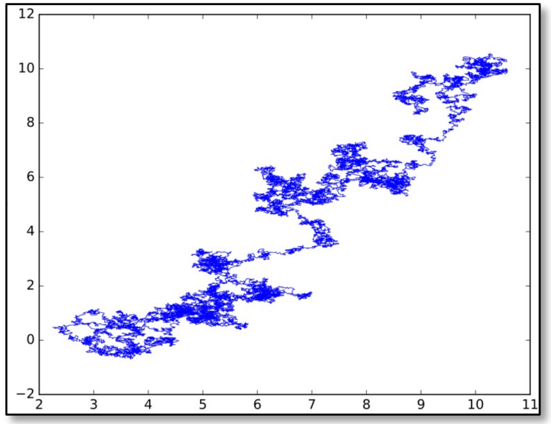
2016

Institute for Integrated Cell-Material Sciences (iCeMS), Kyoto University

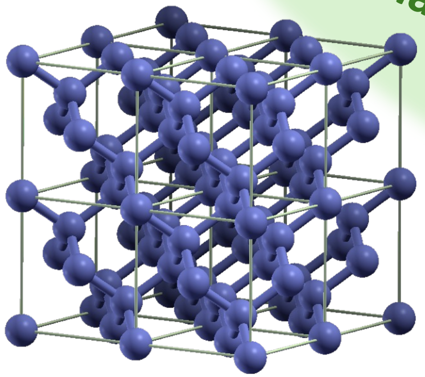
Lecturer

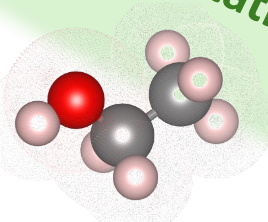
Associate Professor

Phase 1: Stochastic processes



Phase 2: Computational chemistry





Chance to shift to computational research:

Emergence of data science in computational materials science

Around 2012 – 2014, several papers appeared showing how machine learning could be applied in computational materials science.

PRL **108**, 058301 (2012) PHYSICAL REVIEW LETTERS week ending 3 FEBRUARY 2012

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

Matthias Rupp,[1,2] Alexandre Tkatchenko,[3,2] Klaus-Robert Müller,[1,2] and O. Anatole von Lilienfeld[4,2,*]
[1]*Machine Learning Group, Technical University of Berlin, Franklinstr 28/29, 10587 Berlin, Germany*
[2]*Institute of Pure and Applied Mathematics, University of California Los Angeles, Los Angeles, California 90095, USA*
[3]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany*
[4]*Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, USA*
(Received 15 June 2011; published 31 January 2012)

Phys. Rev. Lett. **108**, 2012, 058301

PHYSICAL REVIEW B **89**, 054303 (2014)

Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids

Atsuto Seko,[1,*] Tomoya Maekawa,[1] Koji Tsuda,[2,3] and Isao Tanaka[1,4,5]
[1]*Department of Materials Science and Engineering, Kyoto University, Kyoto 606-8501, Japan*
[2]*Minato Discrete Structure Manipulation System Project, ERATO, Japan Science and Technology Agency, Sapporo 060-0814, Japan*
...*Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan*
...*for Elements Strategy Initiative for Structure Materials (ESISM), Kyoto University, Kyoto 606-8501, Japan*
[5]*Nanostructures Research Laboratory, Japan Fine Ceramics Center, Nagoya 456-8587, Japan*
(Received 6 October 2013; revised manuscript received 29 December 2013; published 18 February 2014)

Phys. Rev. B. **89**, 2014, 054303

Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies

Katja Hansen,[*,†] Grégoire Montavon,[‡] Franziska Biegler,[‡] Siamac Fazli,[‡] Matthias Rupp,[¶] Matthias Scheffler,[†] O. Anatole von Lilienfeld,[§] Alexandre Tkatchenko,[†] and Klaus-Robert Müller[*,‡,||]
[†]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany
[‡]Machine Learning Group, TU Berlin, Germany
[¶]Institute of Pharmaceutical Sciences, ETH Zurich, Switzerland
[§]Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, Illinois, United States
[||]Department of Brain and Cognitive Engineering, Korea University, Korea

J. Chem. Theory Comput. **9**, 2013, 3404

Materials for a Sustainable Energy Future
Long program at the Institute for Pure and Applied Mathematics, UCLA, Sep – Dec 2013

Moreover, in 2013 I got to spend time with some of the authors during a stay at University of California, Los Angeles (UCLA)

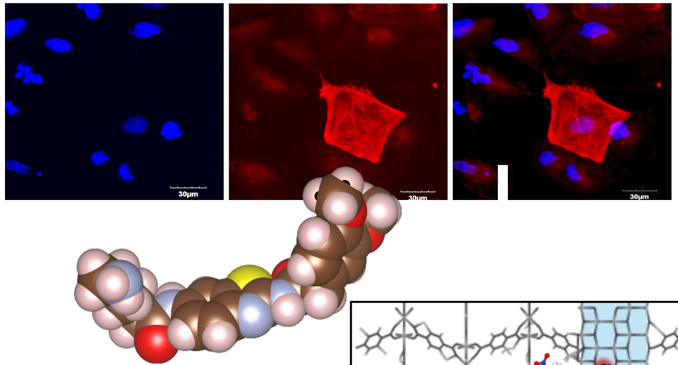
Machine learning and data science gave mathematicians a way to enter the computational materials science field!

… we just had to learn how to do density functional theory calculations.

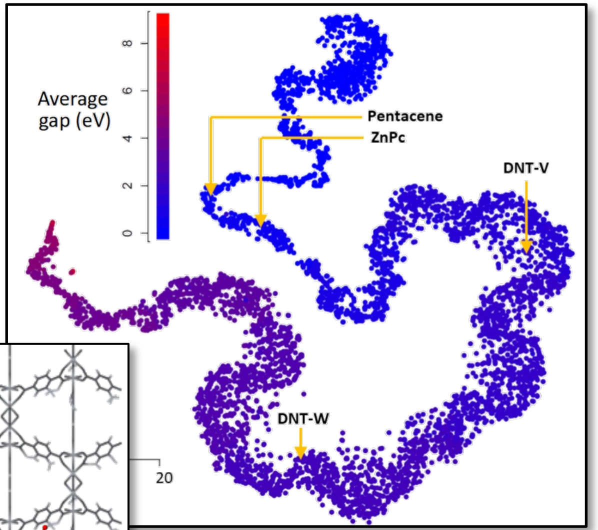
Research in the Packwood group

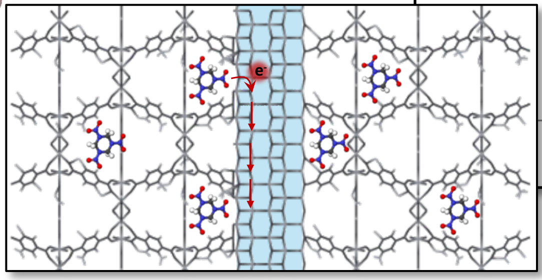
organic materials x simulation x data science

Computational design of bioactive compound for iPS cell control (*J. Chem. Inf. Model.* 2025)

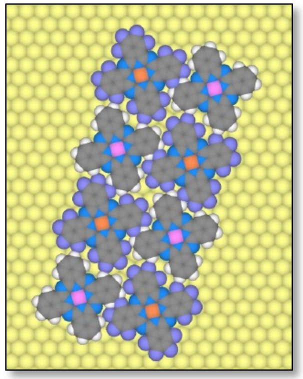


Visualization of organic semiconductor database (*Adv. Theory Simul.* 2023)

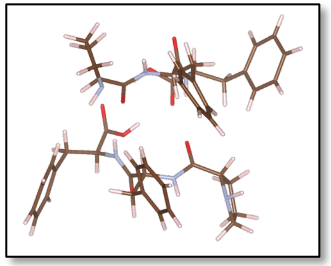




Disease detection potential in a MOF-semiconductor sensor (*Adv. Theory Simul.* 2025)



Simulation of metal-complex self-assembly (*Adv. Physics Res.* 2022)



Game theory analysis of peptide dimer dynamics (in process)

Chemical biology

Small molecule aggregation (2021 – now)

Bioactive molecule discovery (2021 – now)

Materials science

Coordination polymers / metal-organic frameworks (2018 – now)

Organic semiconductor (2021 – now)

On-surface molecular self-assembly (2015 – 2022)

2015 → **2025**

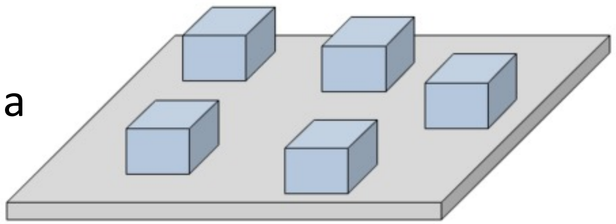
Lecture topics

Simulation of on-surface molecular self-assembly

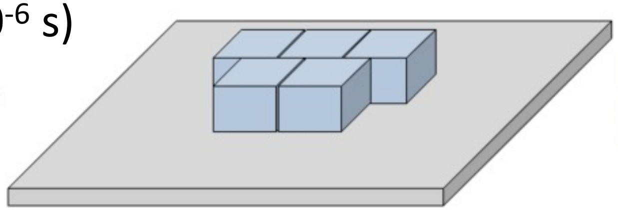
Machine learning for organic photovoltaic materials

On-Surface Molecular Self-Assembly

Molecules adsorbed to a surface



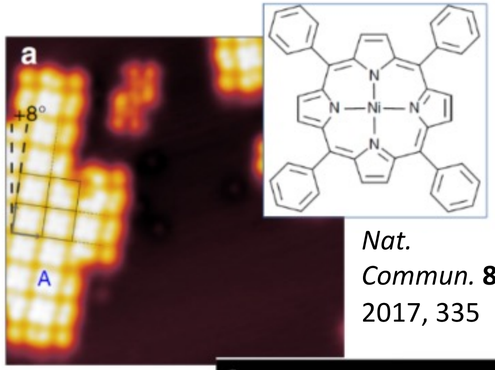
Time (< 10^{-6} s)



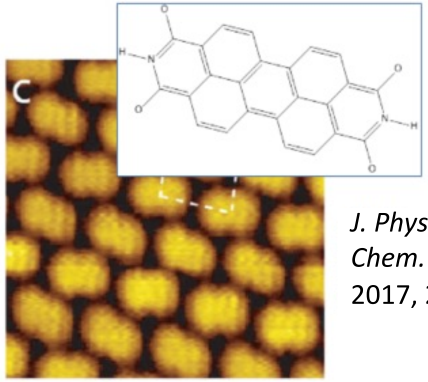
Molecular assembly (forms spontaneously)

Many experimental reports…

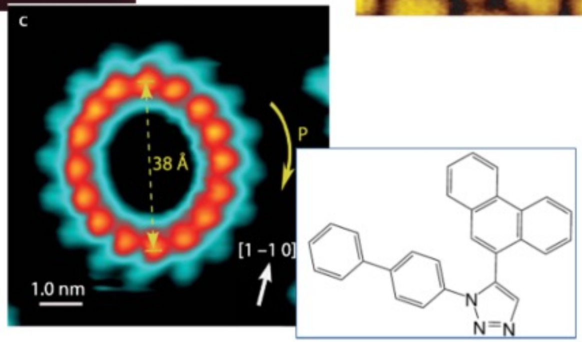
(Scanning tunneling microscopy images)



Nat. Commun. **8**, 2017, 335

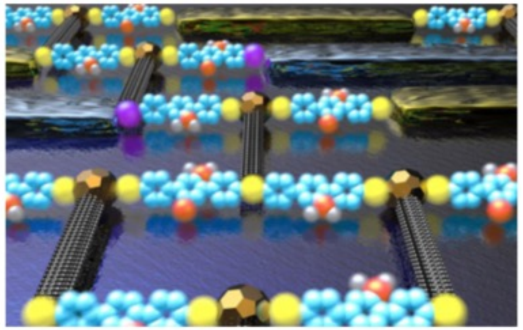


J. Phys. Chem. C. **121**, 2017, 20986



ACS Nano **11**, 2017, 8302

… and many potential applications



Molecular electronics

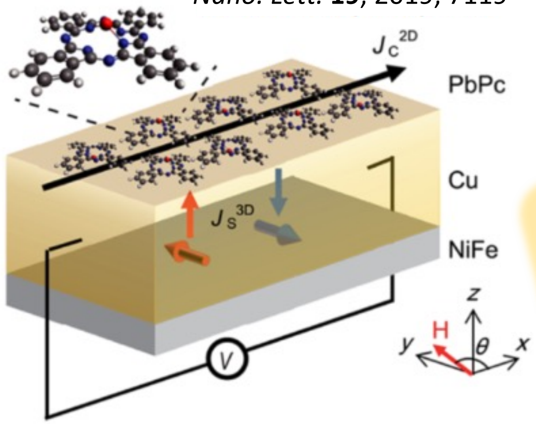
http://asdn.net/asdn/electronics/molecular_electronics.php

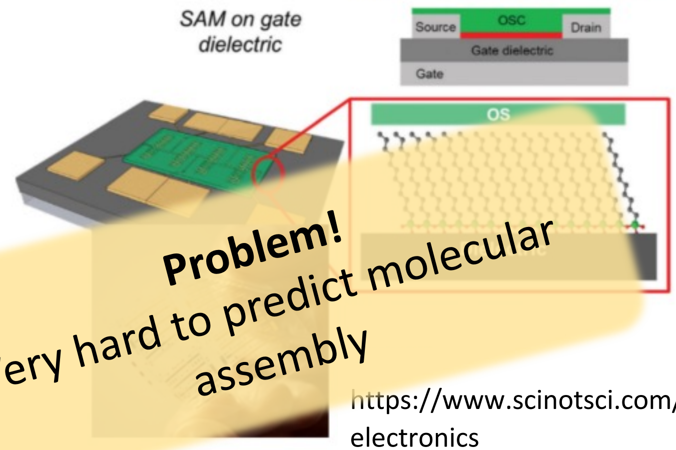
Organic electronics

Chem. Soc. Rev. **46**, 2017, 40

Spintronics

Nano. Lett. **19**, 2019, 7119



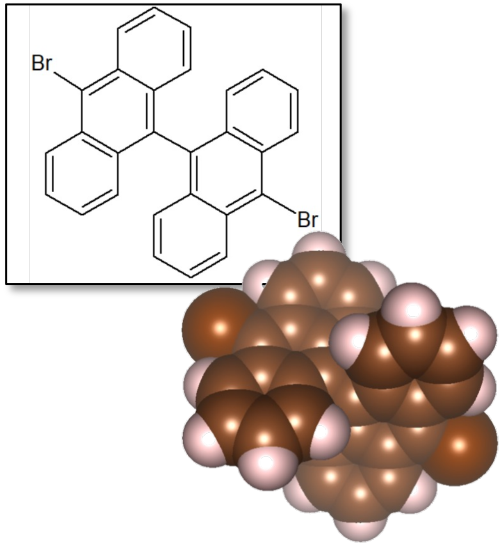


Problem!

Very hard to predict molecular assembly

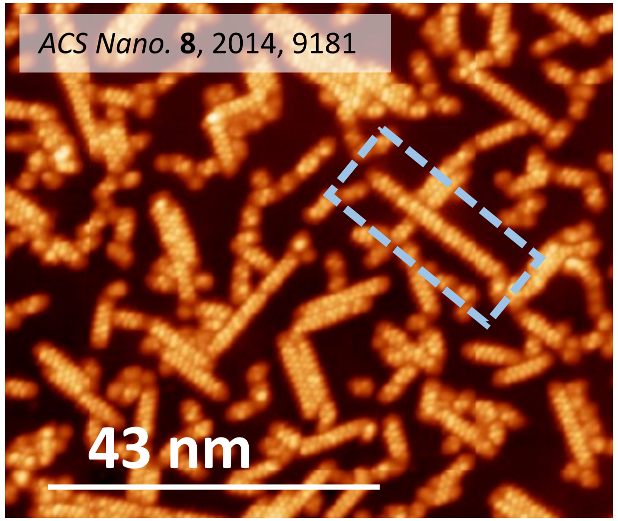
https://www.scinotsci.com/electronics

How I came to this topic?



Dibromo-bianthracene molecule

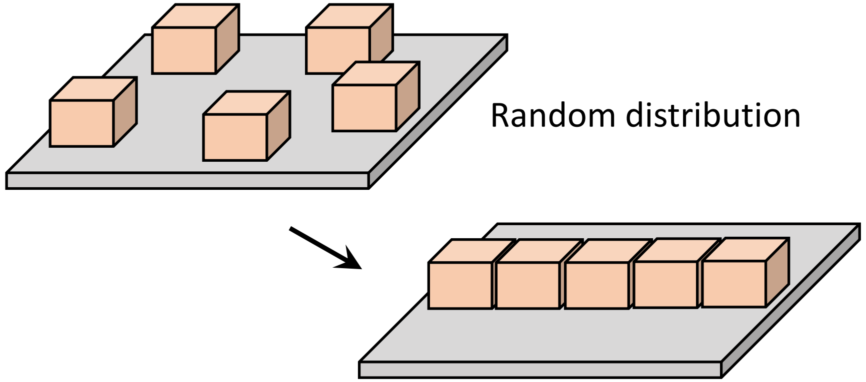
Deposition on Cu(111)



Self-assembly of molecular chains

Dr. Patrick Han

Prof. Taro Hitosugi
(now at Tokyo Univ)

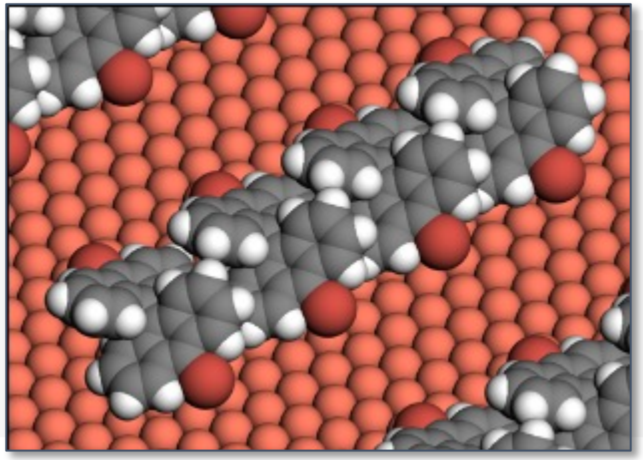


About 10 years ago, colleagues at Tohoku University were studying molecular self-assembly using scanning tunneling microscopy (STM).

They wanted to know whether the chains could be predicted computationally.

Collaborating together, we created a computational method to predict chain formation.

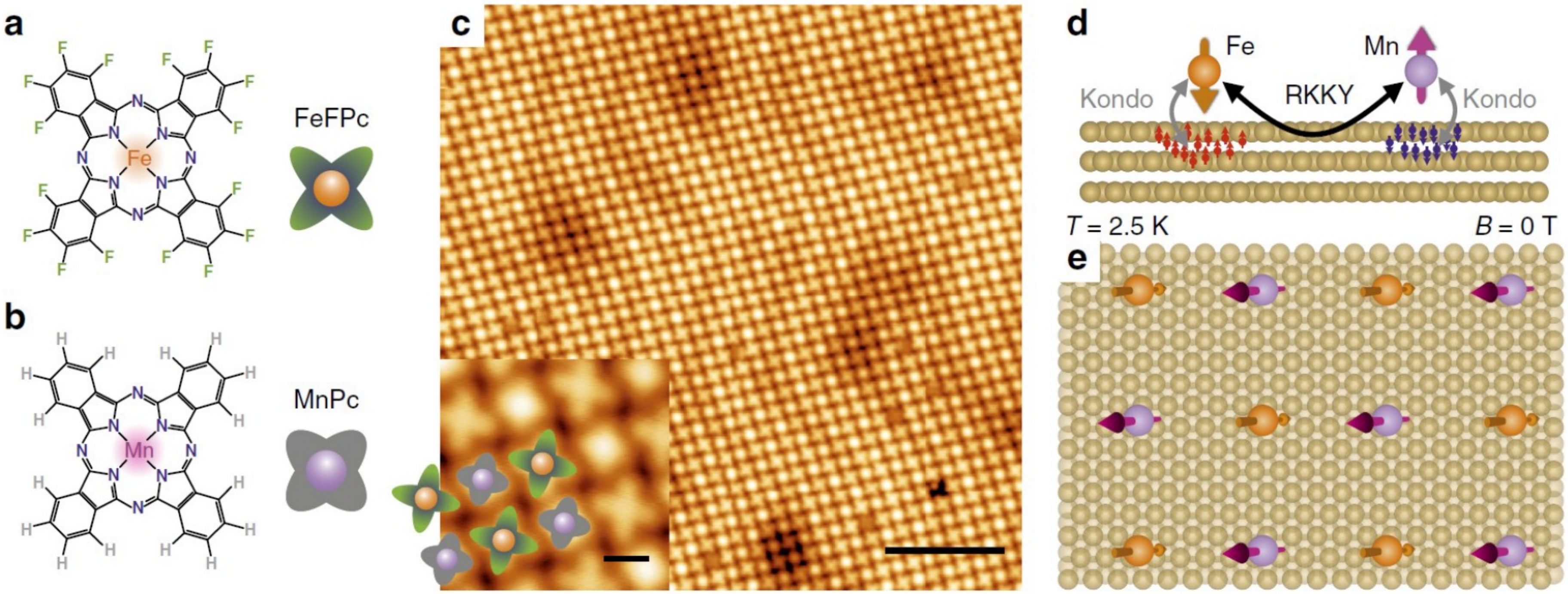
Nat. Commun. **8**, 2017, 14463



Today, I will discuss an updated version of this method applied to a different type of system…

Self-assembly of FeFPc and MnPc on Au(111)

Girovsky *et al. Nat. Commun.* **8**, 2017, 15388

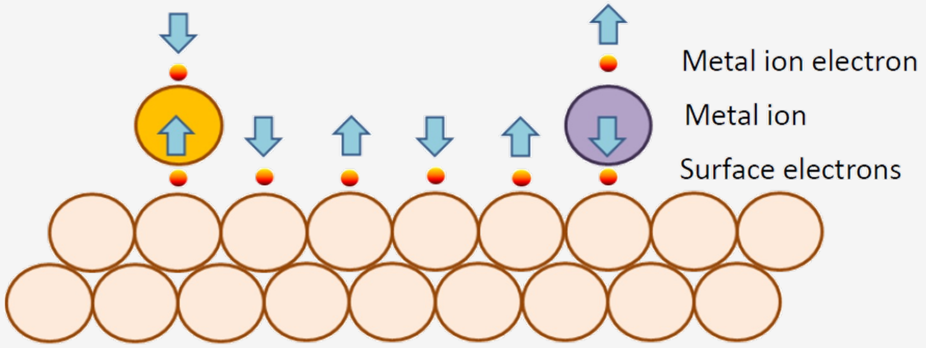


STM image

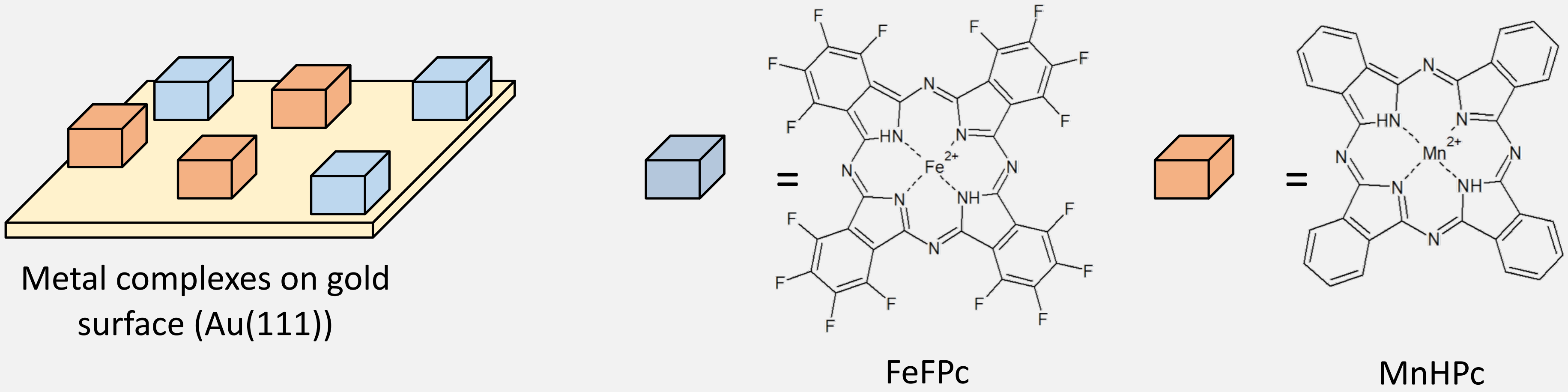
Metal ion spins antiferromagnetically coupled

The antiferromagnetic coupling is due to the Ruderman-Kittel-Kasuya-Yoshida (RKKY) interaction

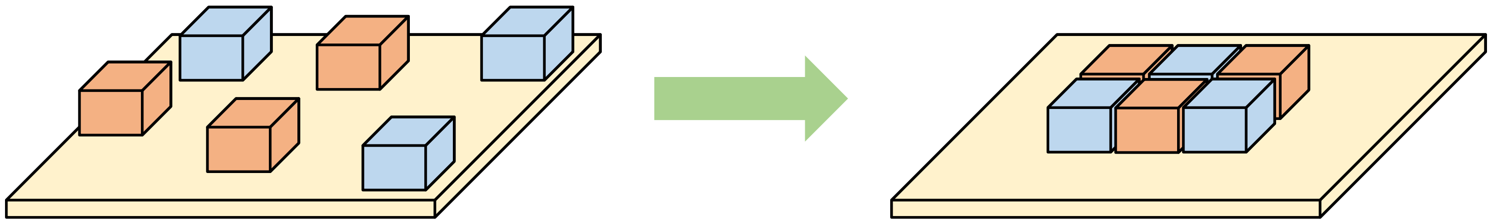
Can we reproduce this result with computation?



Q1. How to build a simplified model for the system?

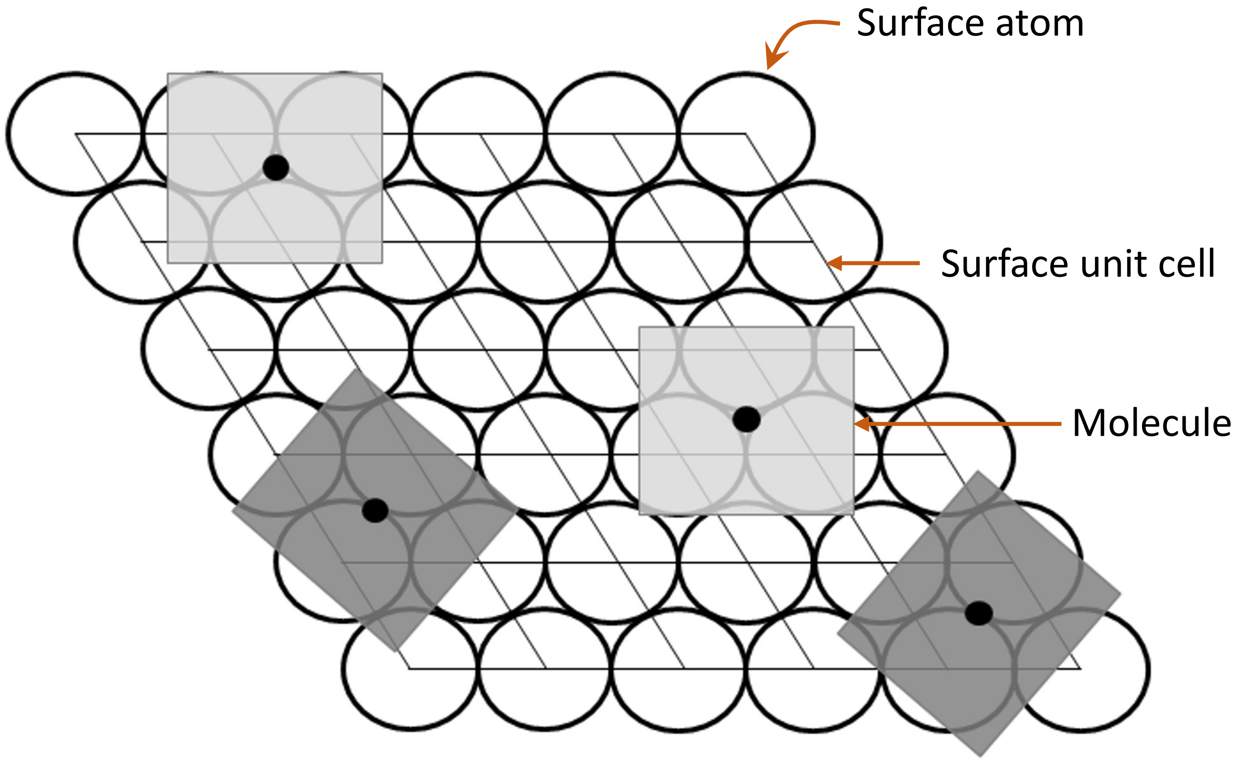


Q2. How to obtain predictions about self-assembly from the model?



Model assumptions

- Perfectly crystalline surface
- Two types of molecules on the surface; n_1 of the first type, n_2 of the second type.
- Finite number of adsorption sites (places where the molecules can sit)
- Finite number of molecule orientations.
- Rigid molecules. All molecules of the same type have the same conformation.
- Two-body energy function



$$E = \sum_i u_i + \sum_{i \neq j} v_{ij}$$

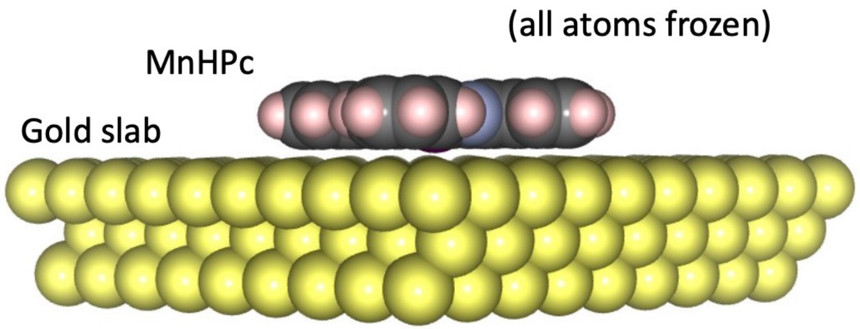
u_i = molecule *i*-surface interaction energy

v_{ij} = molecule *i*-molecule *j* interaction energy

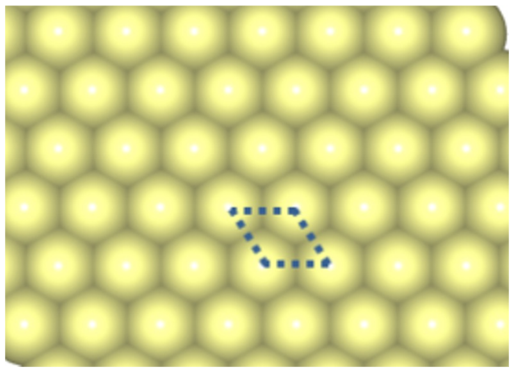
How do we assign the energy parameters (u_i, v_{ij})?

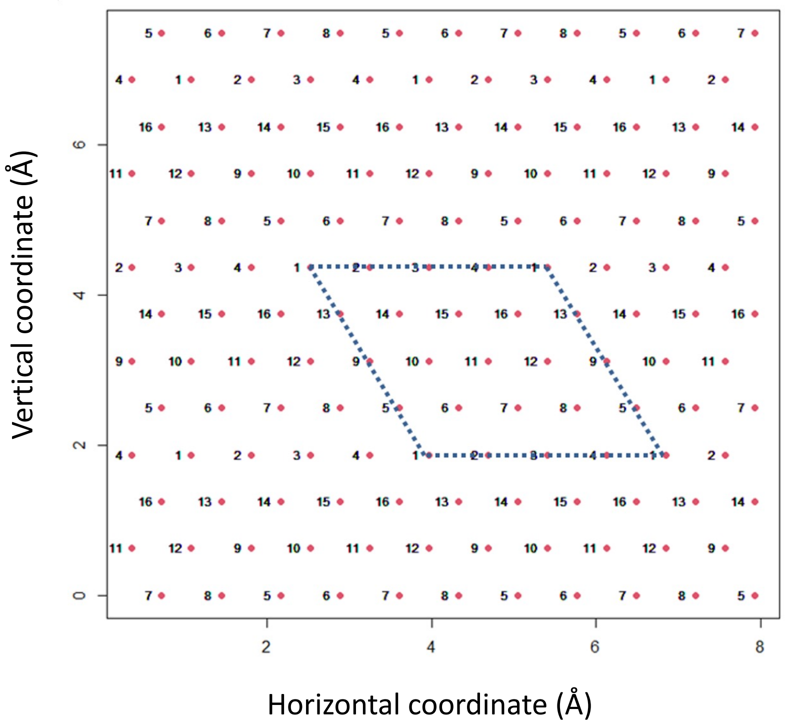
Molecule-surface interaction energies – assign using DFT*

i. Obtain the optimal molecule-surface adsorption height

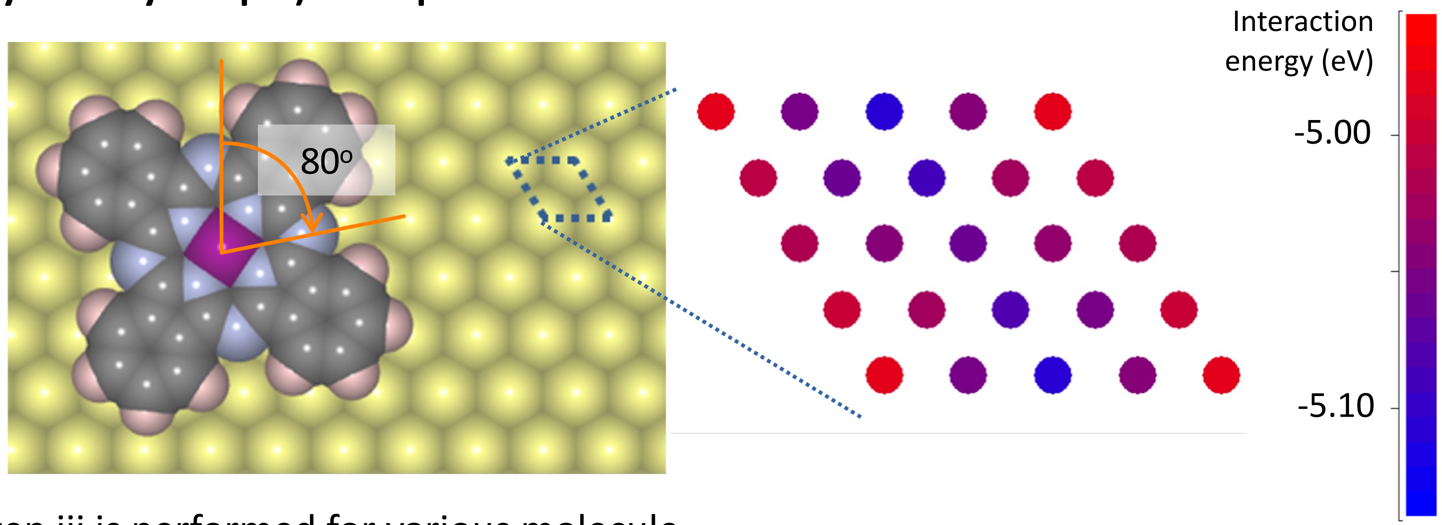


ii. Discretize surface
(around 40,000 points)





iii. Compute interaction energy at each (symmetry unique) adsorption site



Step iii is performed for various molecule orientations (0º, 40º, 80º, …, 320º)

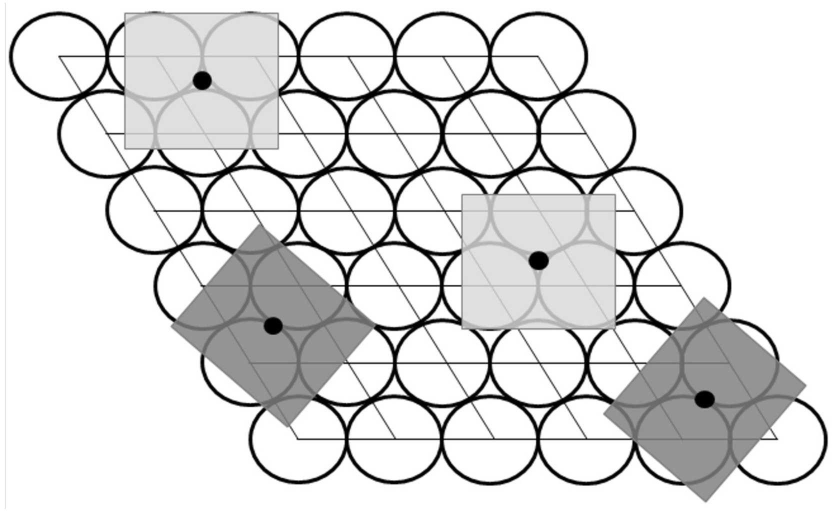
→ ***Surface-molecule interaction parameters assigned!***

$$E = \sum_{i} u_i + \sum_{i \neq j} v_{ij}$$

* DFT as implemented in FHI-aims, with the PBE exchange-correlation functional, TS vdW corrections, and "light" basis set defaults.

Molecule-molecule interaction energies

$$E = \sum u_i + \sum v_{ij}$$



Problem!

For a model with 40,000 adsorption sites, around 10^{11} unique molecule-molecule interactions are possible.

We cannot calculate the energy of each one individually using DFT.

Our solution

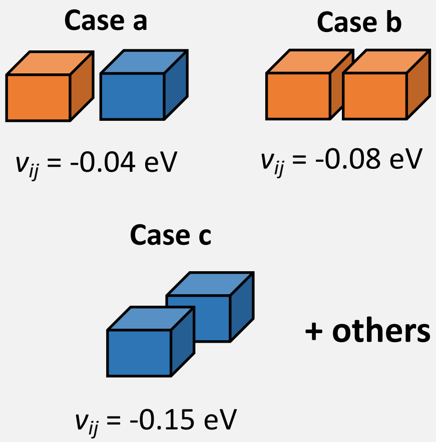
Take a small sample of around 5000 intermolecular interactions, calculate their energies with DFT.

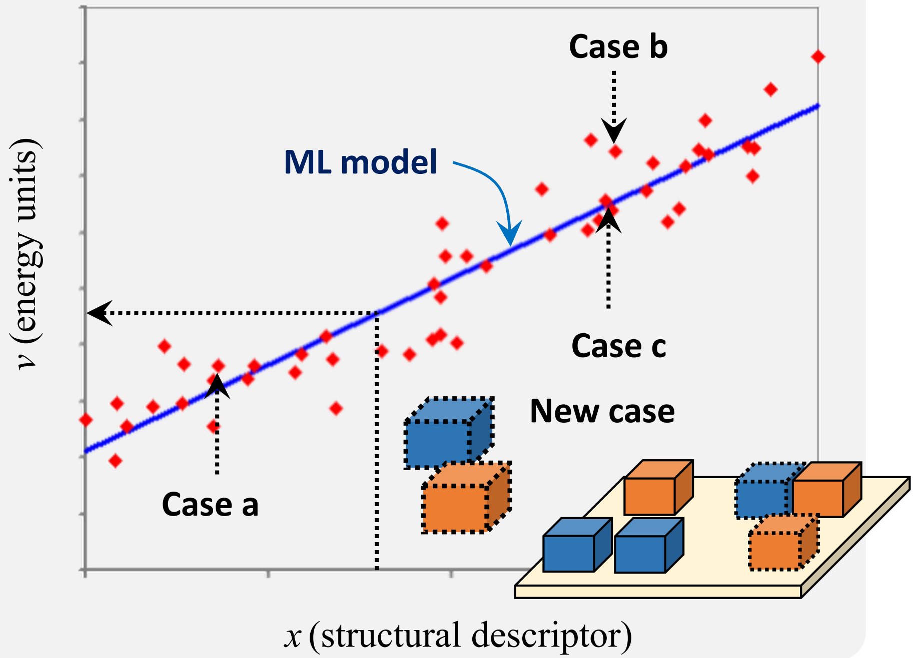
Use the sample data to build a **machine-learning (ML) model**.

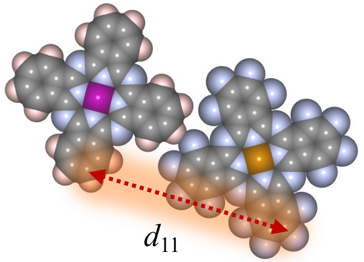
Use the model to assign the energies of the remaining cases.

<u>**Sample data**</u>

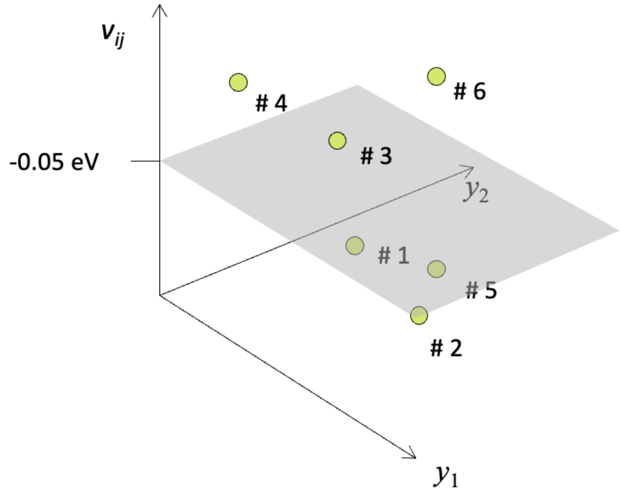
(calculated from DFT)

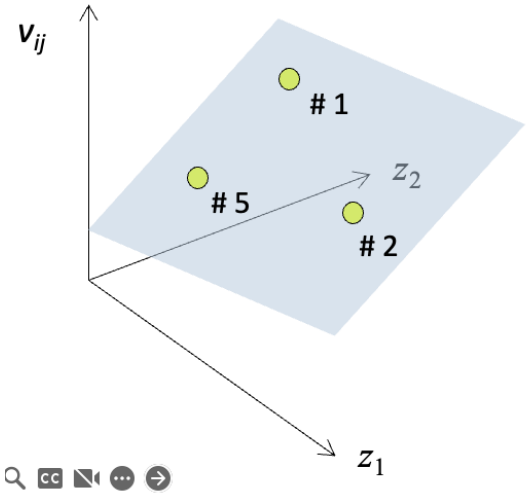






Input case (interaction energy = v_{ij})





Coulomb-type descriptors
$x_1 = 1/d_{11}, x_2 = 1/d_{12}, \ldots$

Support vector machine
Transform data so that it can be separated by a linear plane
$(x_1, x_2, \ldots) \Rightarrow (y_1, y_2, \ldots)$

$v_{ij} < -0.05$ eV

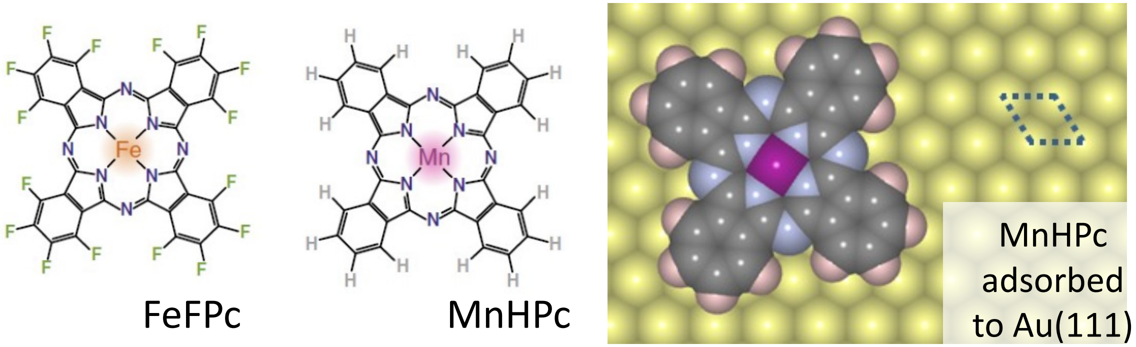
$-0.05 \text{ eV} \leq v_{ij} \leq 0$

Kernel ridge regression
Transform data so that it lies on a linear plane
$(y_1, y_2, \ldots) \Rightarrow (z_1, z_2, \ldots)$

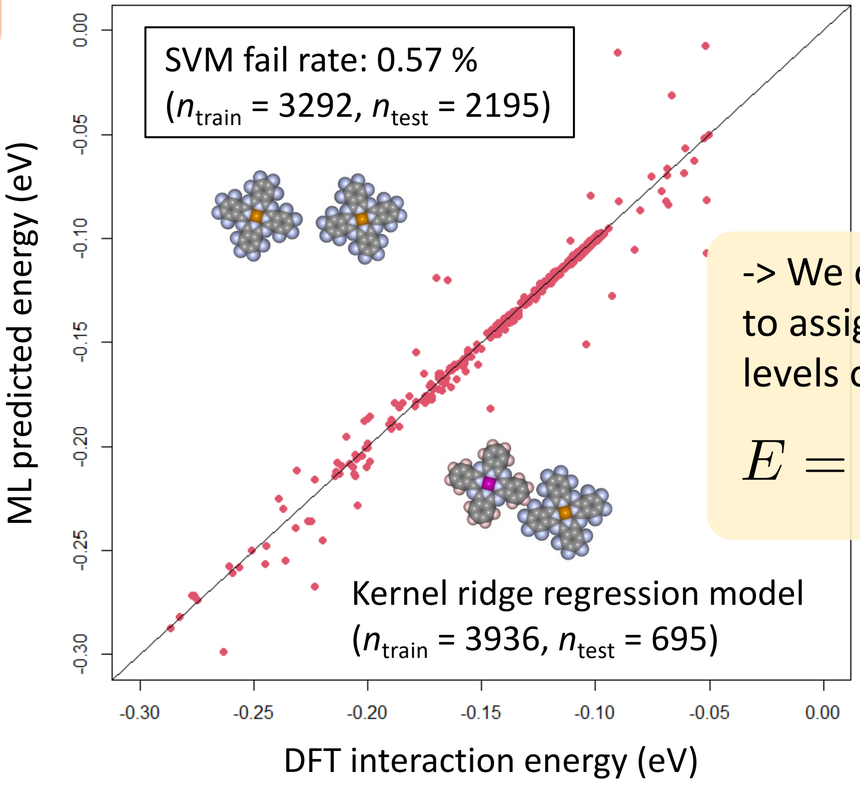
v_{ij} prediction

$v_{ij} = 0$

System under consideration:
Phthalocyanine molecules adsorbed to gold(111)



Performance on test data

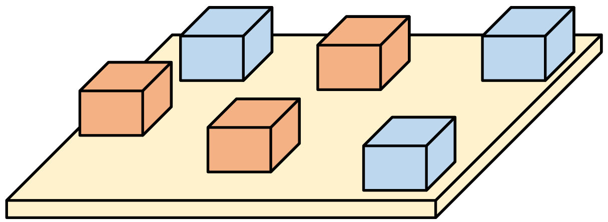


-> We can use the ML model to assign the v_{ij} with DFT-levels of accuracy:

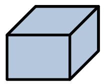
$$E = \sum u_i + \sum v_{ij}$$

DFT details: FHI-aims code, PBE xc functional, TS vdW corrections, 'light' basis set defaults.

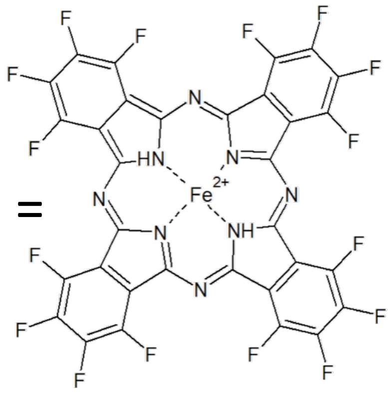
Q1. How to build a simplified model for the system?



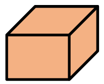
Metal complexes on gold surface (Au(111))



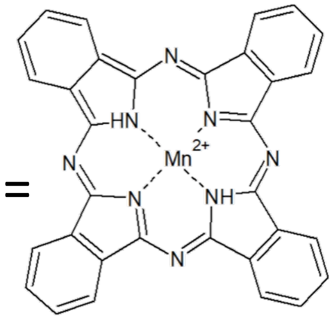
=



FeFPc

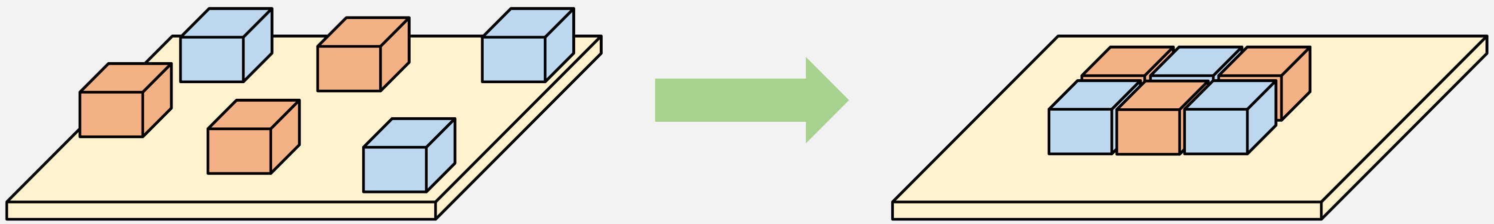


=



MnHPc

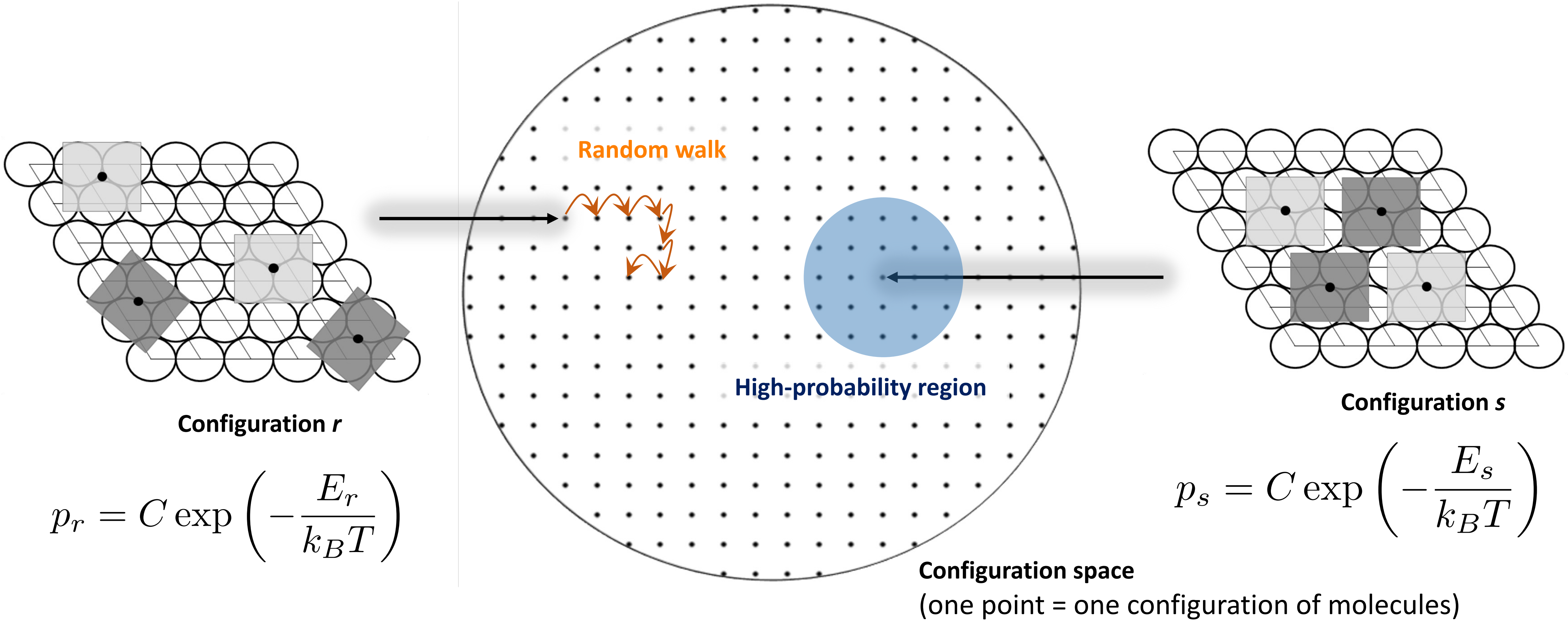
Q2. How to obtain predictions from the model?



We need to identify molecule configurations with high formation probability *p*:

$$p = C \exp\left(-\frac{E}{k_B T}\right)$$

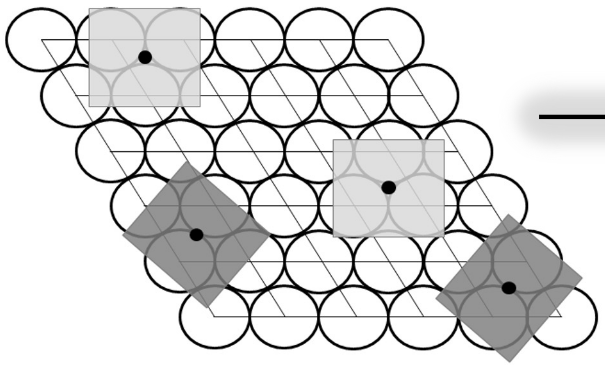
Markov chain Monte Carlo (MCMC)



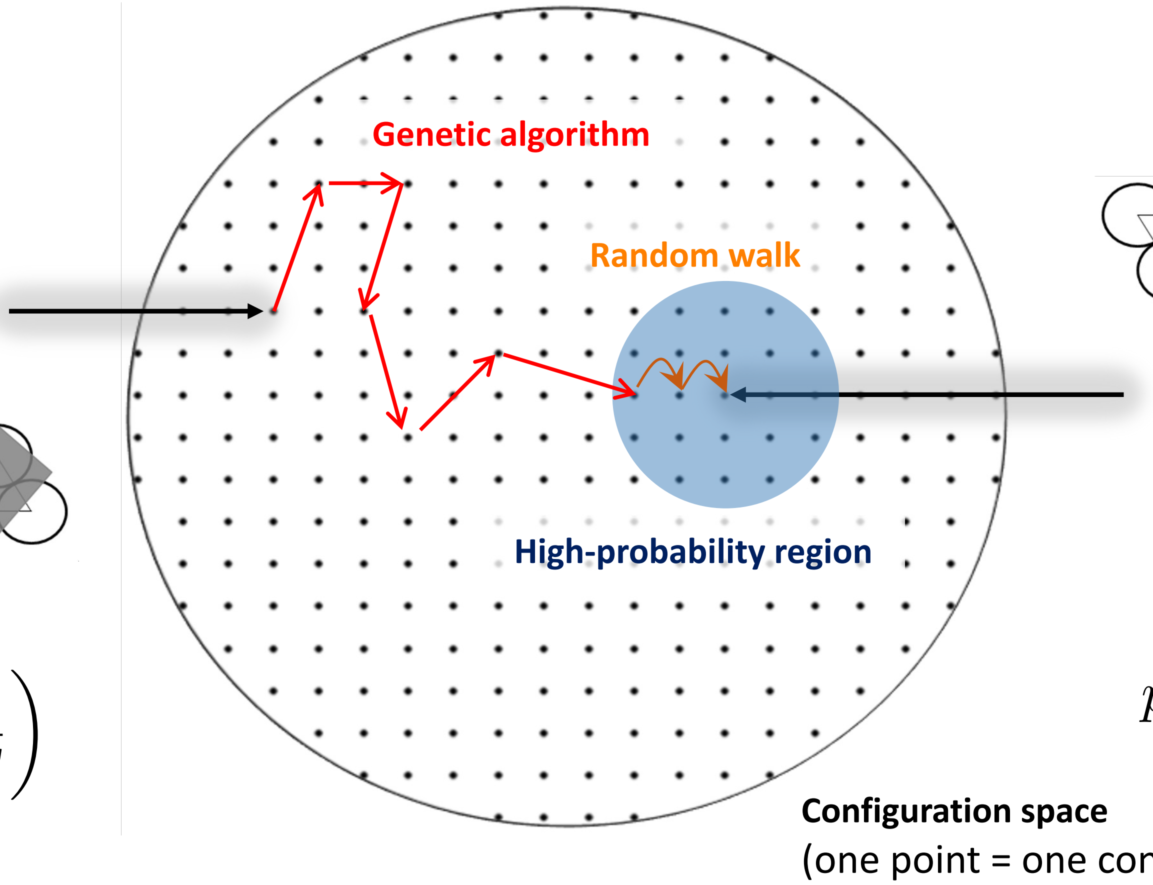
In MCMC, we simulate a random walk over the configuration space. It is simulated in such a way that, after a long length of time, the number of visits to configuration *s* is proportional to p_s (the formation probability).

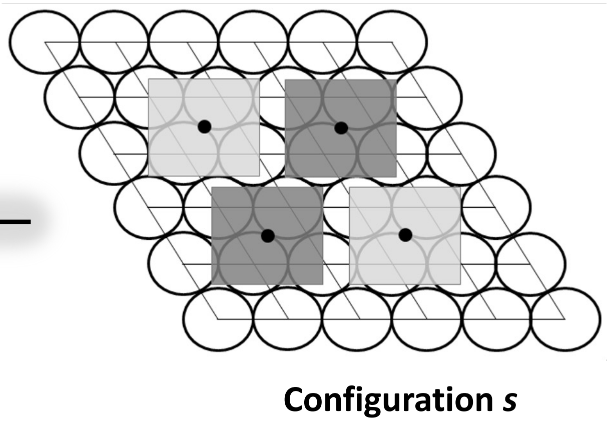
Problem: MCMC can be very inefficient. The configuration space is huge. The high-probability region (where the molecules are closely packed) is very small. Long simulation times are usually required to reach it.

Our solution: genetic algorithm + Markov chain Monte Carlo



$$p_r = C \exp\left(-\frac{E_r}{k_B T}\right)$$



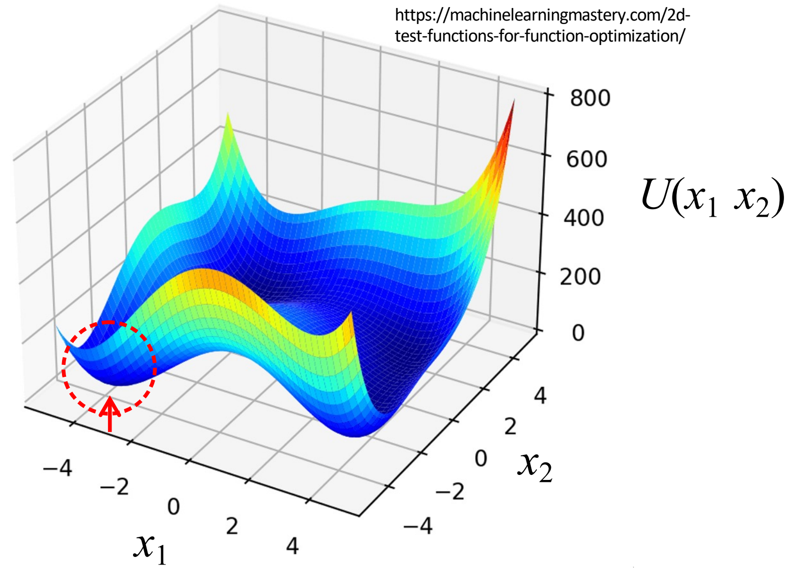


$$p_s = C \exp\left(-\frac{E_s}{k_B T}\right)$$

The genetic algorithm makes large steps in the configuration space, quickly bringing us to the high-probability region (coarse search)

Markov chain Monte Carlo (random walk) makes short steps, thoroughly exploring the high-probability region. (fine search)

Genetic algorithm (general concept)



Task: Find the values of $x_1, x_2, \ldots, x_n$ which minimize the objective function $U(x_1, x_2, \ldots, x_n)$

Initial setup: *N* vectors of random numbers (x_{ij})

$$\mathbf{v}_1 = (x_{11}, x_{12}, \ldots, x_{1n}) \qquad \mathbf{v}_2 = (x_{21}, x_{22}, \ldots, x_{2n}) \qquad \ldots \qquad \mathbf{v}_N = (x_{N1}, x_{N2}, \ldots, x_{Nn})$$

These vectors are called *chromosomes*. The elements x_{ij} are called *genes*.

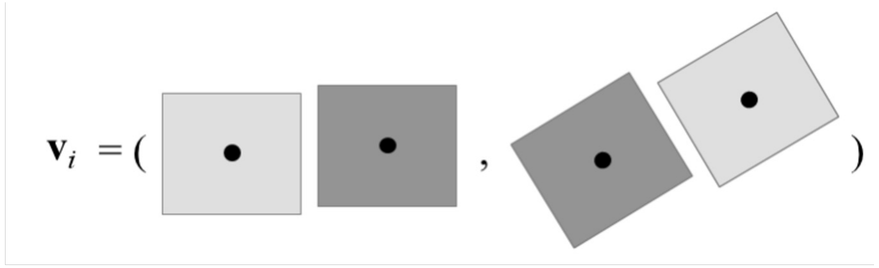
Algorithm:

Population of chromosomes	Calculate objective function	Calculate fitness	Create *k* new chromosomes by random mixing	Make new population with *N* – *k* high-fitness chromosomes from original population and the *k* new ones	Mutate all chromosomes by adding small random numbers
$\mathbf{v}_1$	U_1	f_1		$\mathbf{v}_1$	$\mathbf{v}_1'$
			$\mathbf{v}_{12} =$ $(x_{11}, x_{22}, \ldots, x_{2n})$		
$\mathbf{v}_2$	U_2	f_2		$\mathbf{v}_{12}$	$\mathbf{v}_{12}' = (x_{11} + w_1, x_{22} + w_2, \ldots, x_{2n} + w_n)$
⋮	⋮	⋮		⋮	⋮
$\mathbf{v}_N$	U_N	f_N		$\mathbf{v}_N$	$\mathbf{v}_N'$

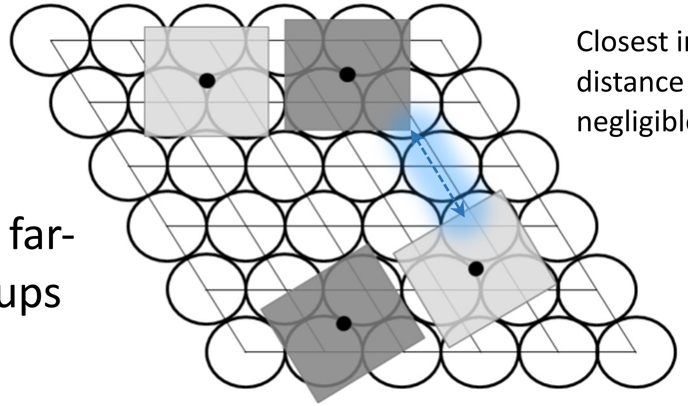
As algorithm is iterated, $\min(U_1, U_2, \ldots, U_N)$ converges to the global minimum.

Genetic algorithm implementation

Chromosomes: Vector of molecular clusters



Clusters correspond to far-separated groups of molecules



Closest interatomic distance > r_{cut}, negligible interaction

Objective function: a free energy function

$$U\left(\mathbf{v}_i\right) = E\left(\mathbf{v}_i\right) - JS\left(\mathbf{v}_i\right) \qquad S\left(\mathbf{v}_i\right) = k_B \ln W\left(\mathbf{v}_i\right)$$

Energy

Pseudo-temperature

Configuration entropy

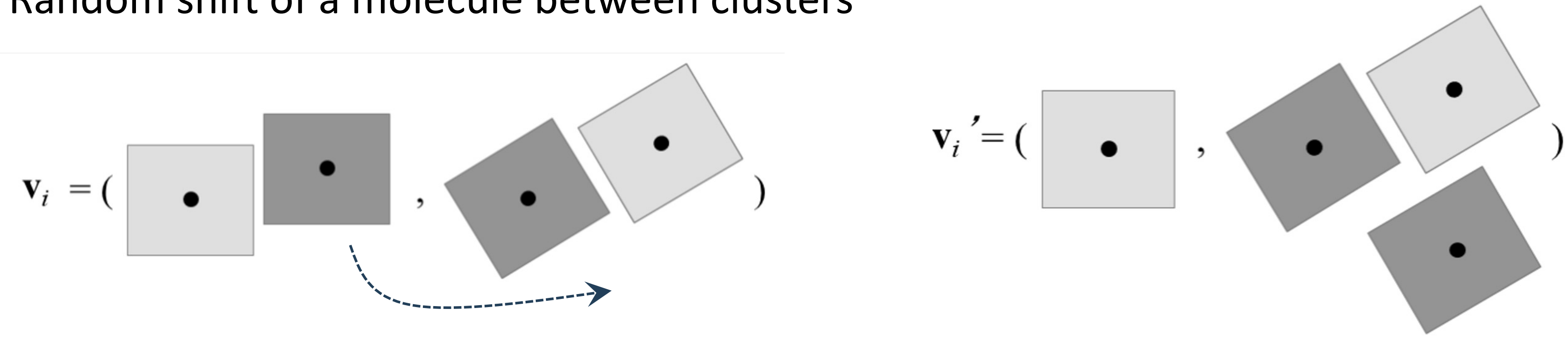
Number of unique ways molecular clusters can be placed on the surface (approximate formula: *R. Soc. Open. Soc.* **3**, 2016, 150681)

Chromosome mixing:

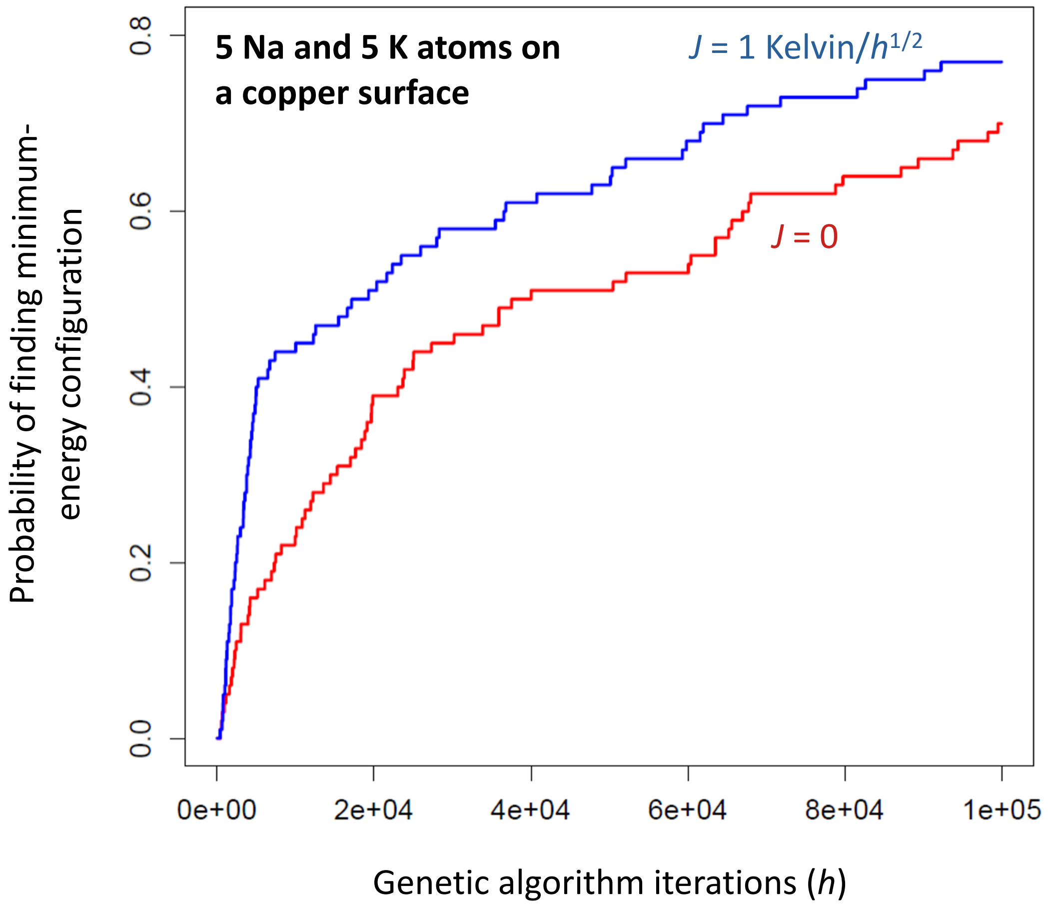
As before, but with conditions (to ensure that number of molecules n_1, n_2, is constant)

Chromosome mutations:

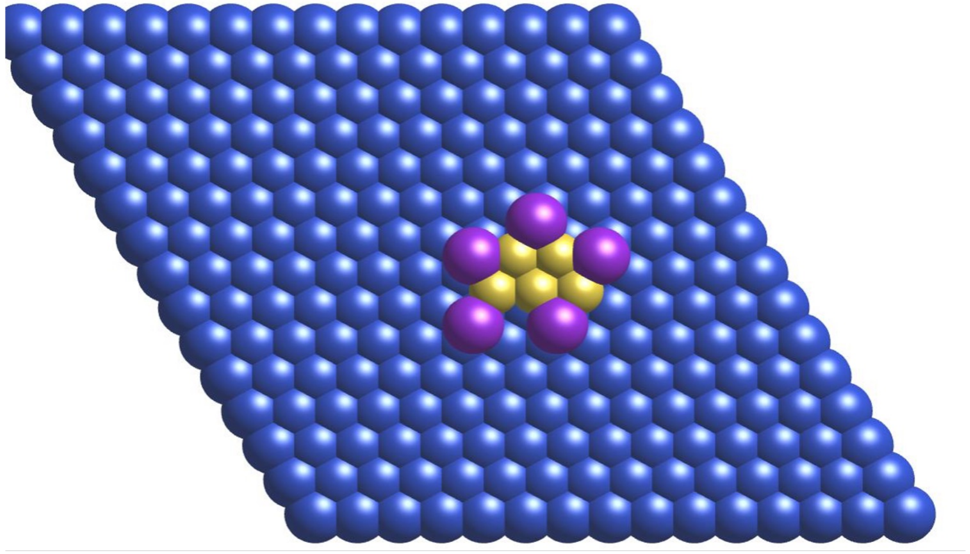
Random shift of a molecule between clusters



Special feature of our genetic algorithm: simulated annealing



Minimum-energy configuration
(yellow = Na atom, purple = K atom)

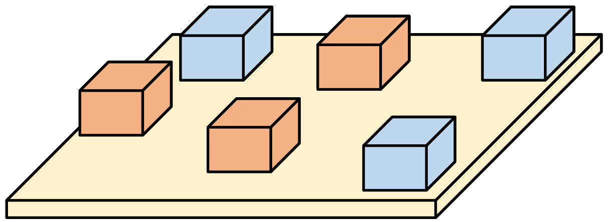


We gradually decrease the pseudo-temperature J as the algorithm proceeds. This helps the algorithm move out of local energy minima, improving performance.

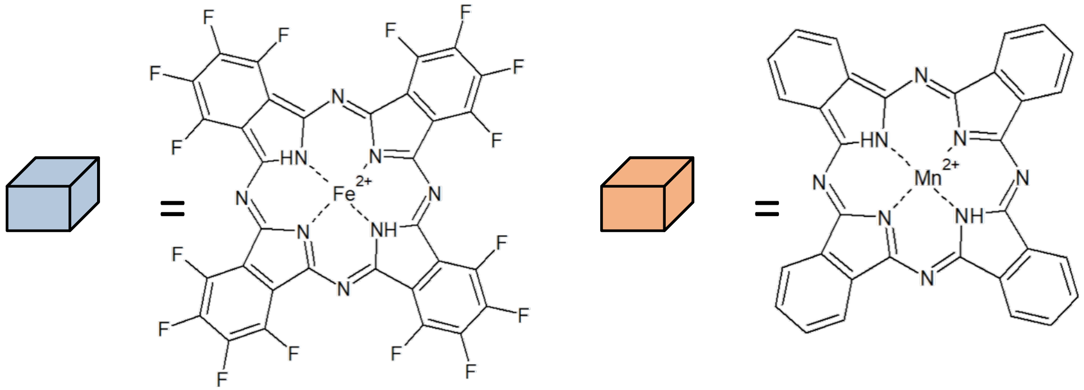
Note: J is just a parameter to control the algorithm, and not the true surface temperature.

$$U(\mathbf{v}_i) = E(\mathbf{v}_i) - JS(\mathbf{v}_i)$$

Q1. How to build a simplified model for the system?

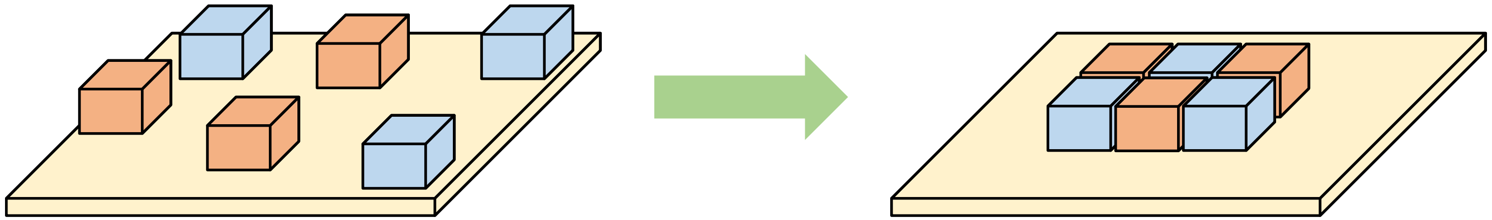


Metal complexes on gold surface (Au(111))



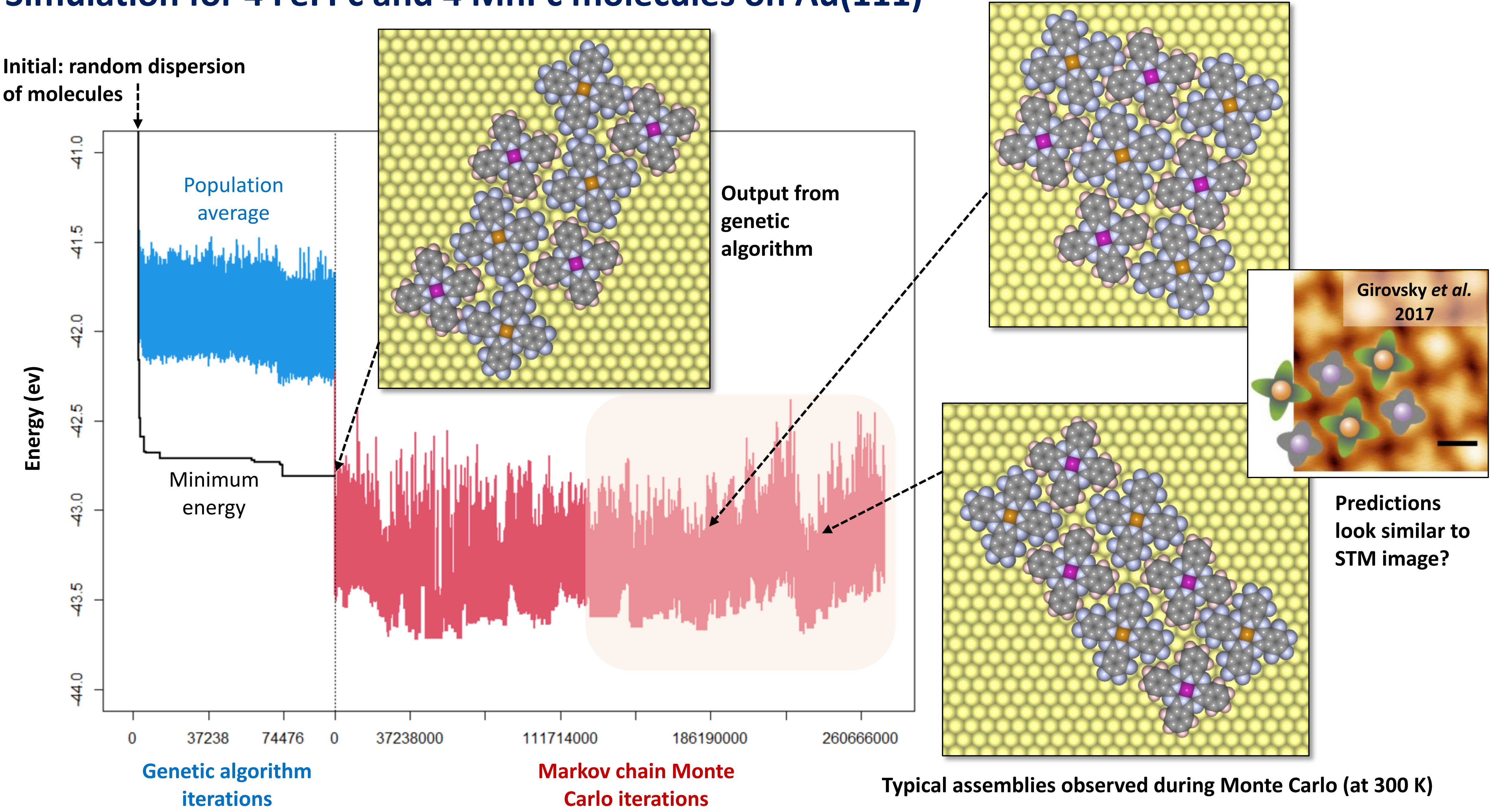
FeFPc MnHPc

Q2. How to obtain predictions from the model?

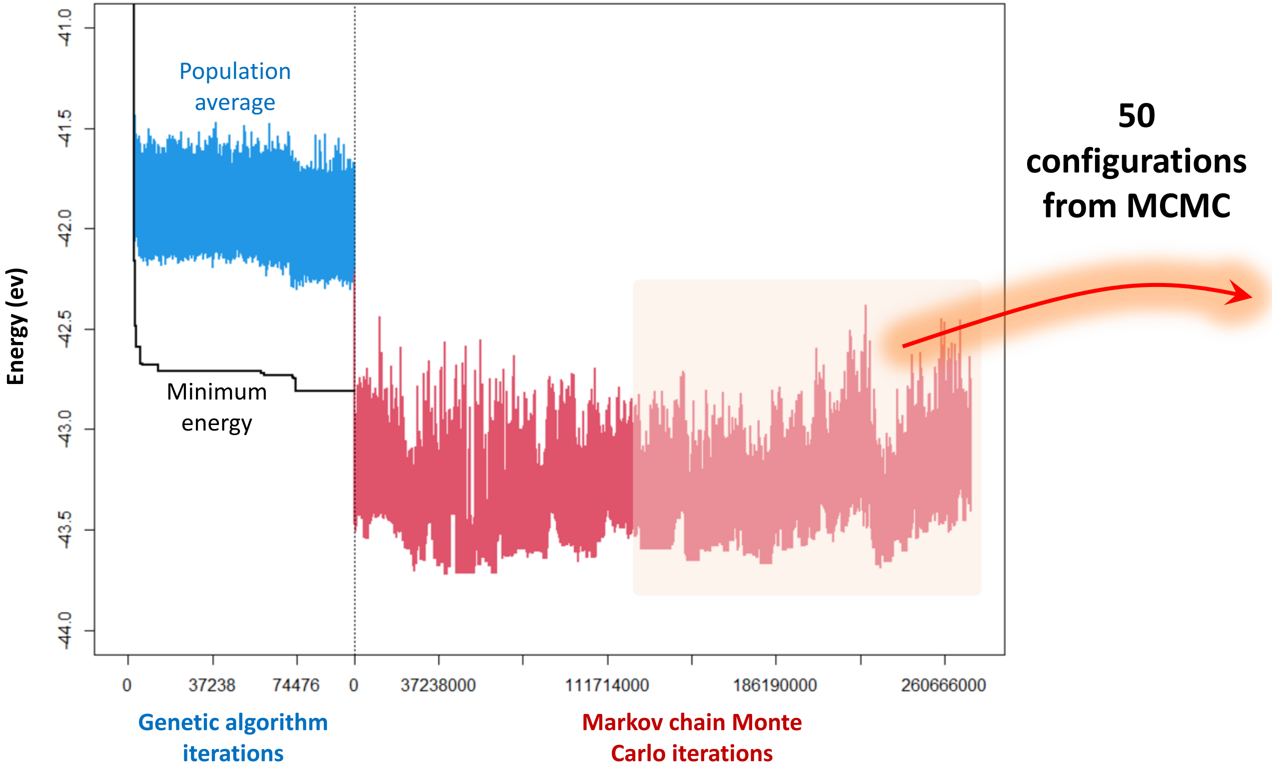


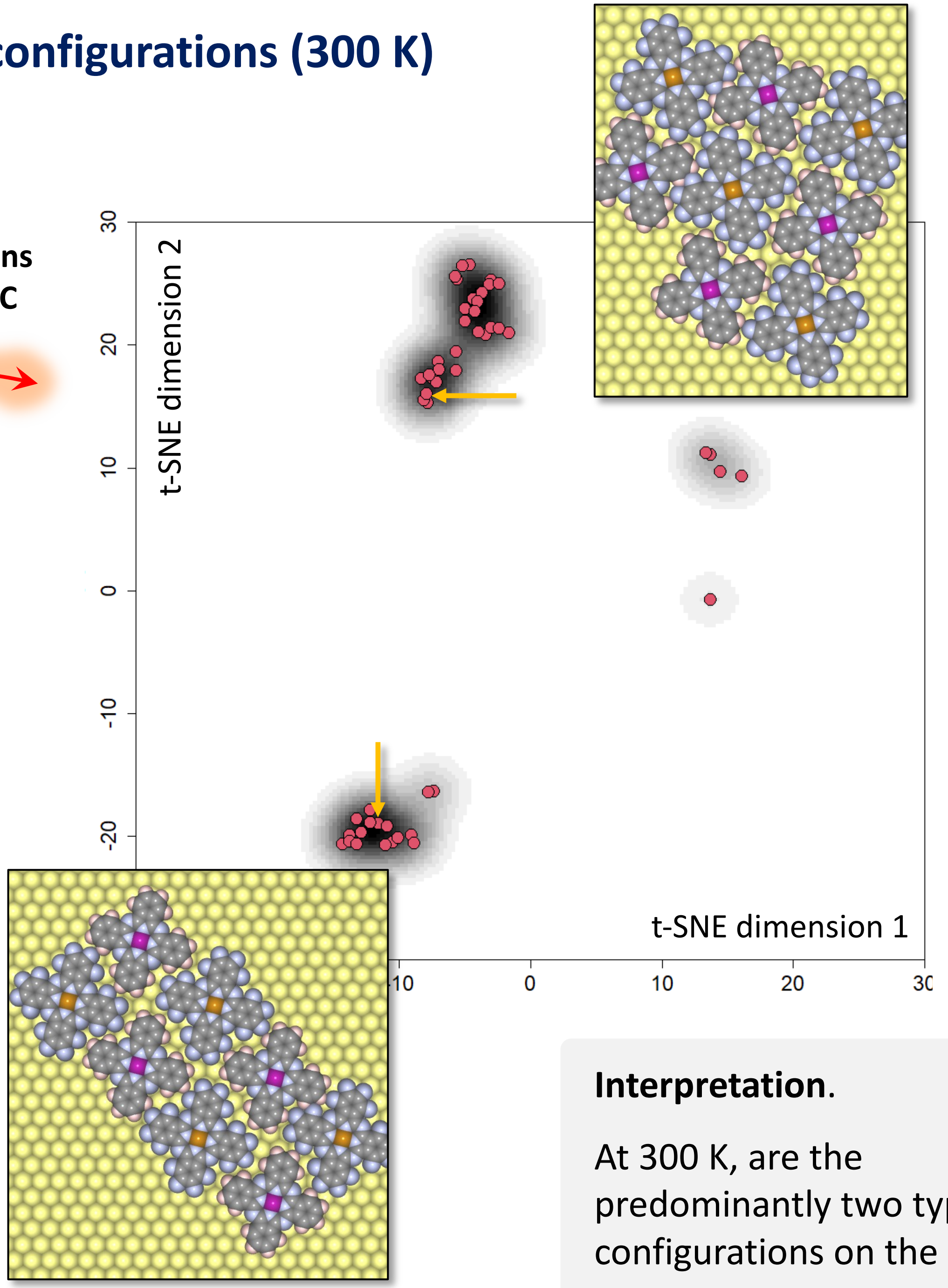
Results…

Simulation for 4 FeFPc and 4 MnPc molecules on Au(111)



Visualization of equilibrium distribution of molecular configurations (300 K)

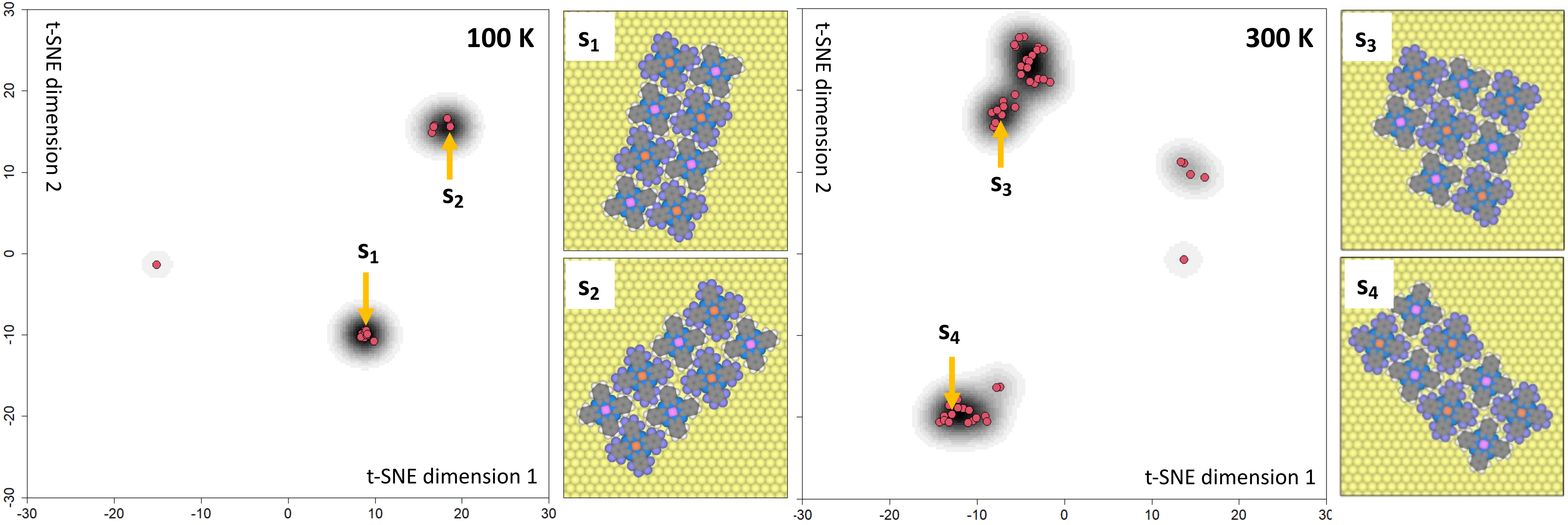




- MCMC gives us a sample of configurations that appear at thermodynamic equilibrium.
- We apply t-distributed stochastic neighborhood embedding (tSNE) to visualize the MCMC sample in 2D. The configurations (red points) are arranged according to their structures.
- The configurations sampled by MCMC are mostly located in two clusters. Within these clusters, the configurations are mostly identical.

Interpretation.

At 300 K, are the predominantly two types of configurations on the surface.

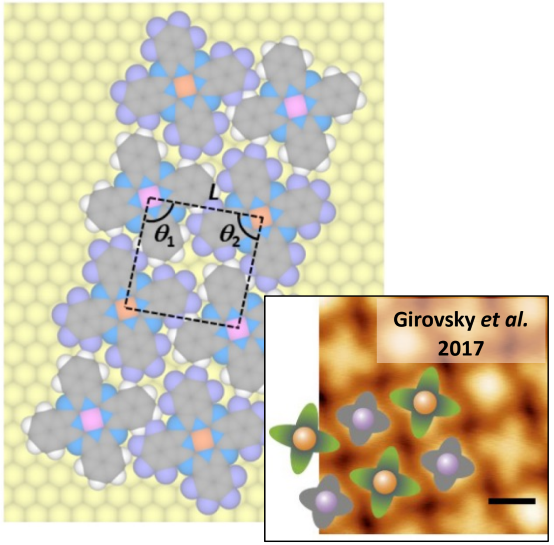


To compare with experiment, we consider simulations performed at low temperature (100 K).

Two types of configurations are also seen at 100 K, although different from the ones at 300 K.

Configuration s_1 achieves quantitative agreement with cryogenic STM images. Another phase also reported experimentally (perhaps s_2?) .

	Calculation (mean ± st err)	**Experiment (mean ± st err)**
L	14.14 ± 0.07	14.05 ± 0.08
θ_1	95.27 ± 1.86	95.84 ± 1.26
θ_2	84.73 ± 1.87	84.84 ± 0.79



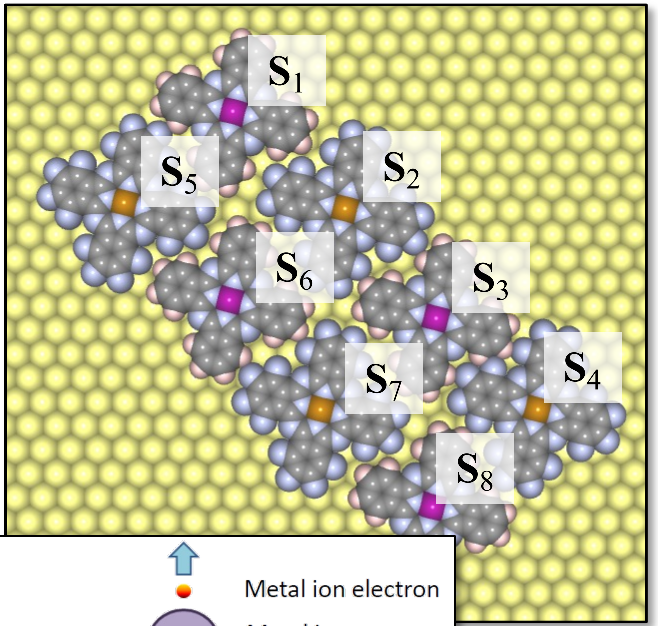
Predicting magnetic properties

The spin directions of the Mn / Fe ions are very difficult to predict from first-principles. We therefore use a classical Ising model:

$$H = \sum_{i \neq j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$$

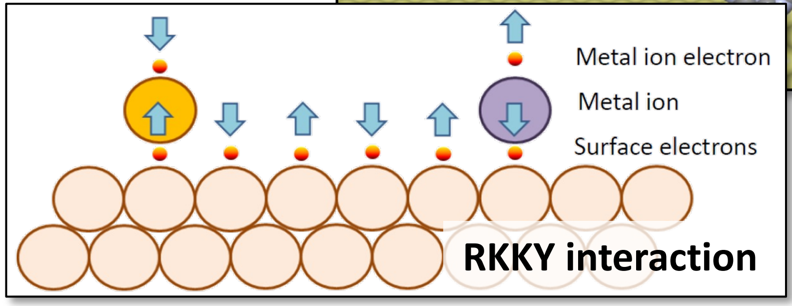
$\mathbf{S}_i$ = spin vector for spin *i*

J_{ij} = Exchange constant



For the case of RKKY interactions between spins*:

$$J_{ij} = Qg(2k_F r_{ij}) \qquad g\left(x\right) = \left(\frac{2}{x}\right)^4 \left(\sin x - x\cos x\right)$$



Q = System-dependent constant

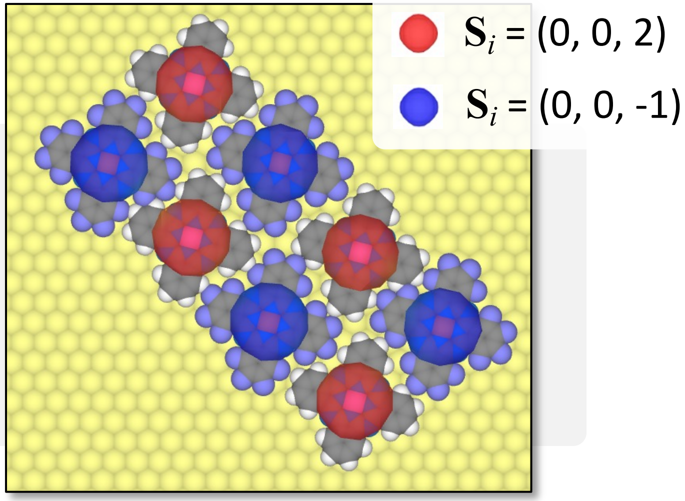
r_{ij} = Distance between spins *i* and *j*

k_F = Fermi vector for surface (about 0.18 Å^{-1} for Au(111)*)

Result:

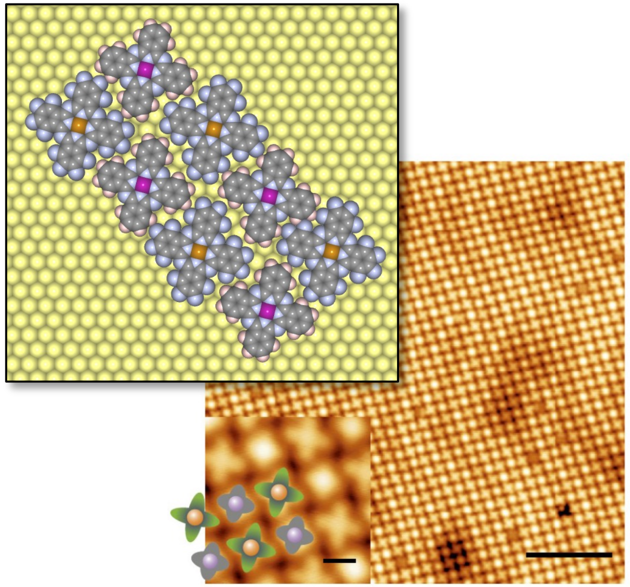
This model correctly predicts an antiferromagnetic ground state for the molecular assembly!

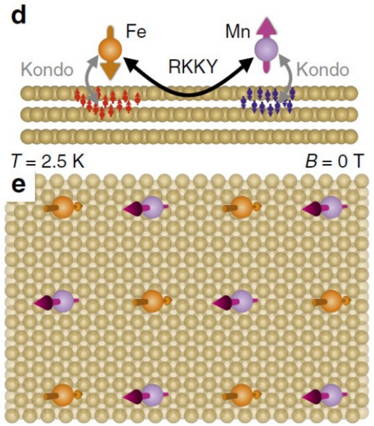
(a direct calculation of the spin orientations using density functional theory would be preferred for studying the magnetic ordering in more detail)



*Patterson and Bailey. *Solid-State Physics.* 2018. Springer

** Reinert *et al. Phys. Rev. B.* **63**, 2001, 115415





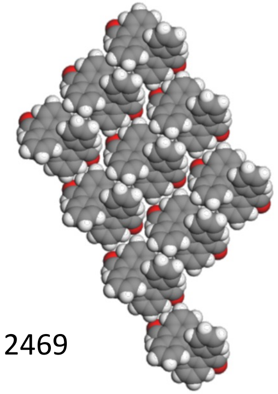
Final comments for part 1

- We succeeded to create a method that can predict how molecules assemble on metal surfaces.
- It combines a machine-learned interaction potential with genetic algorithms and Monte Carlo sampling.

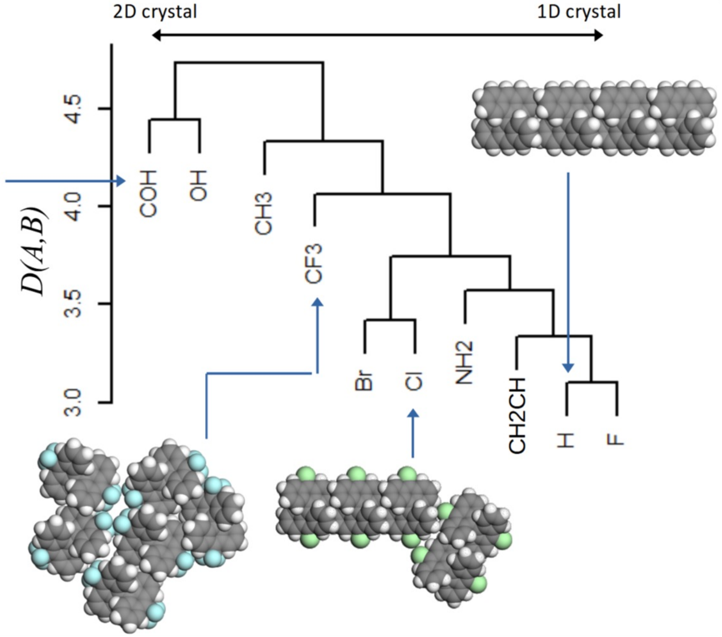
References: *Nat. Commun.* **8**, 2017, 14463;
Adv. Phys. Res. **1**, 2022, 2200019 <- this one is better!

We have done some other things with this method:

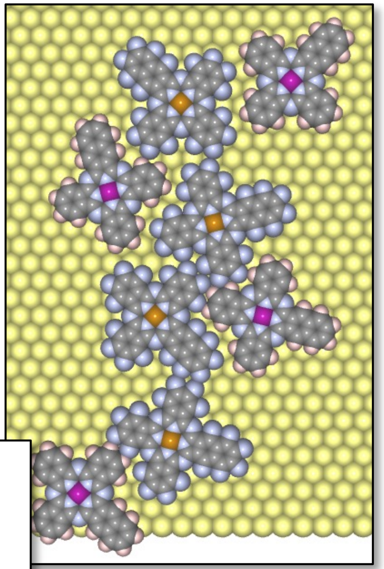
- Discovery of connection between single-molecule properties and molecular assembly shape.

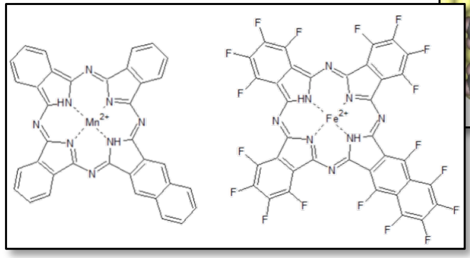


Nat. Commun. **9**, 2018, 2469



- Prediction of disordered molecular assemblies using asymmetric molecules





Adv. Phys. Res. **1**, 2022, 2200019

Lecture topics

Simulation of on-surface molecular self-assembly

Machine learning for organic photovoltaic materials

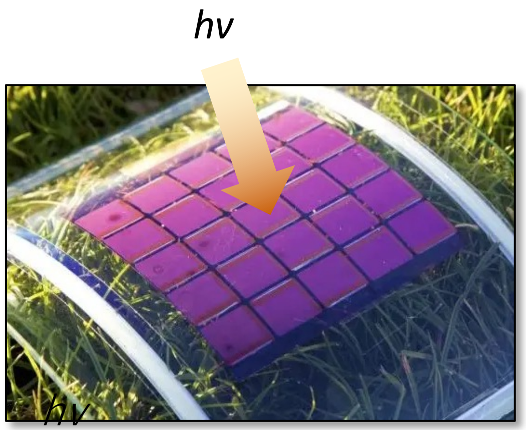
April 2021

iCeMS-MacDiarmid Institute Online Workshop

iCeMS
KYOTO UNIVERSITY
WPI Research Center

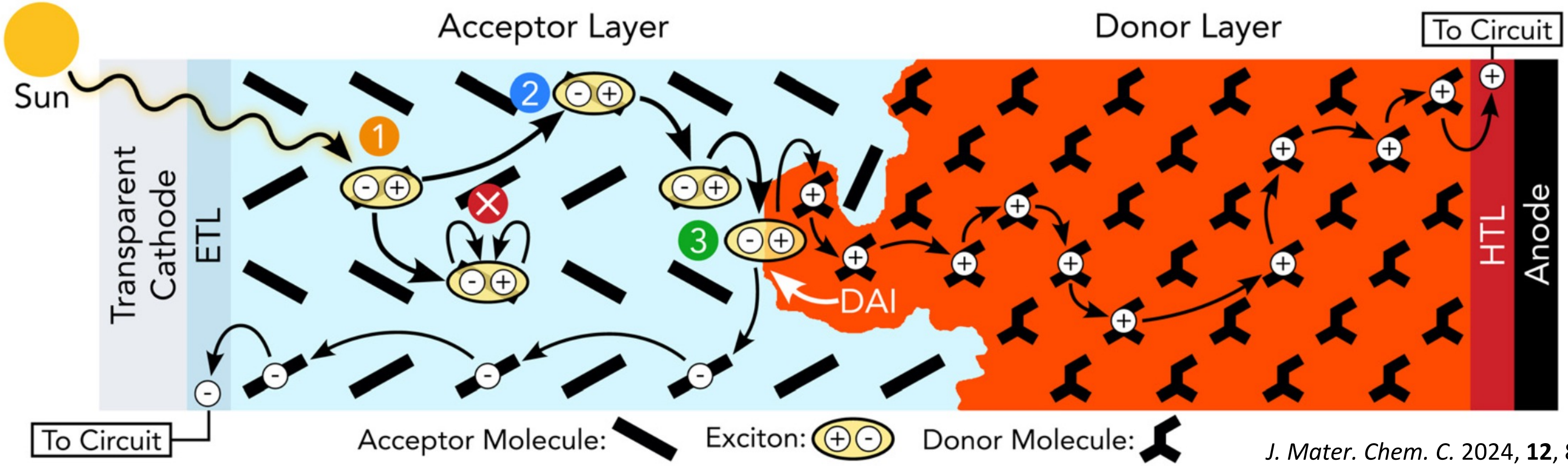
Te Mana
Tangata Whakawhanake
MacDiarmid Institute
Advanced Materials & Nanotechnology

Organic solar cells and exciton diffusion (Hodgkiss group)



Organic solar cell

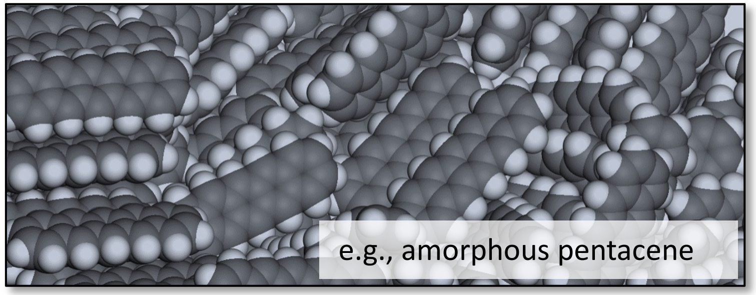
(www.solarreviews.com/blog/organic-solar-cells/)



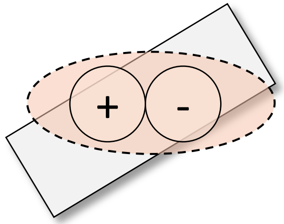
J. Mater. Chem. C. 2024, **12**, 8747

Acceptor layer is an organic semiconductor
(crystal or amorphous solid of organic molecules)

Donor layer is another semiconductor (can be organic or inorganic)

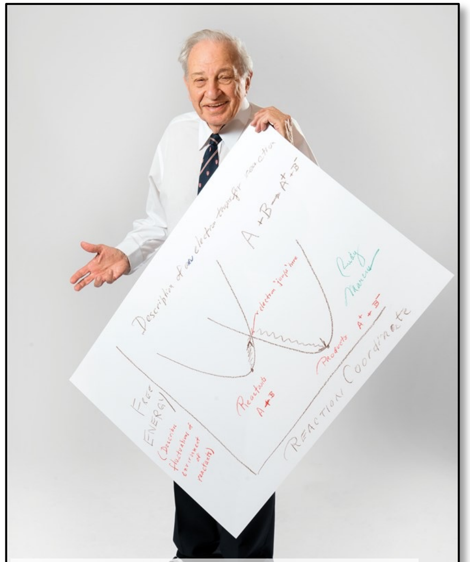


Light absorption by the acceptor layer results in **Frenkel excitons**
– tightly-bound electron-hole pairs localized to single molecules.



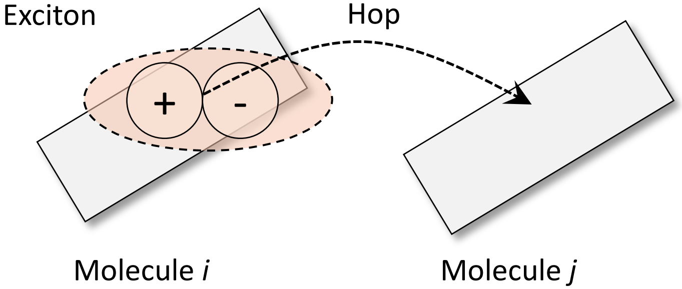
(Frenkel) exciton

Excitons hop between molecules to the donor-acceptor interface (DAI), where they dissociate into electron and hole pairs and generate electricity. However, if the exciton moves too slowly, it will die by electron-hole recombination.



https://mediatheque.lindau-nobel.org/laureates/marcus/cv

Exciton hopping



About 60 years ago, Rudolf Marcus showed that the rate of exciton hopping between molecules is approximately

Exciton coupling

Reorganization energy

$$k_{ij} = \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \frac{v_{ij}^2}{\hbar} \exp\left(-\frac{\lambda}{4k_B T}\right)$$

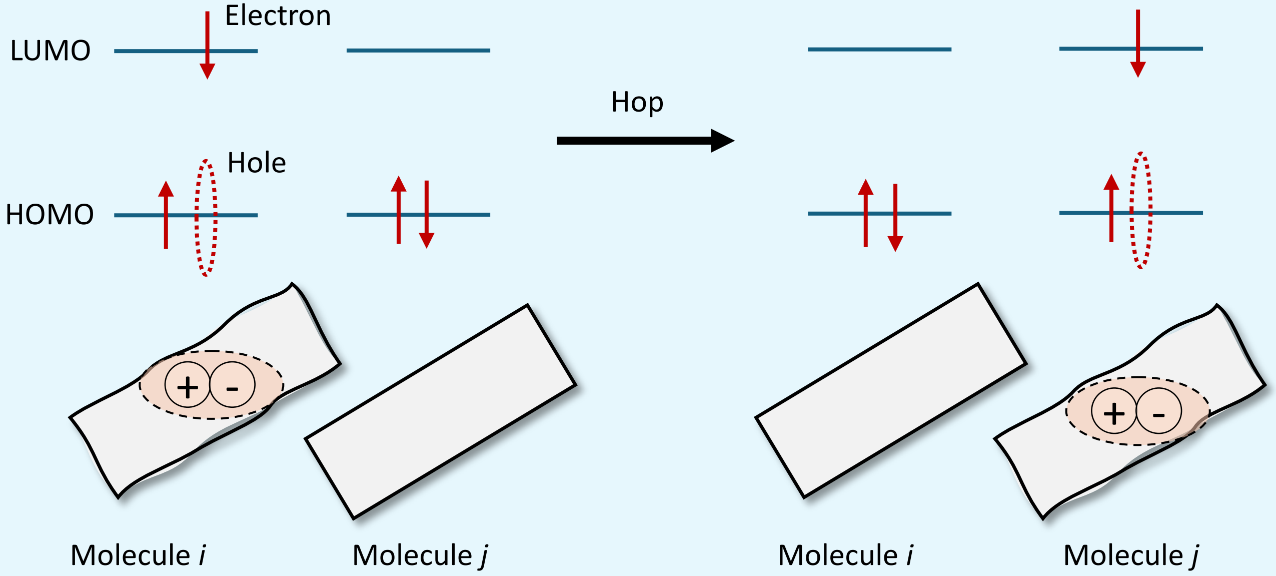
Temperature

$1/k_{ij}$ is the (average) time taken for the exciton to hop from molecule *i* to molecule *j*.

Important parameter 1.
Reorganization energy (λ)

The exciton distorts the shape (conformation) of the molecule.

The reorganization energy measures the energy required to change the shapes of the molecules as the exciton shifts.

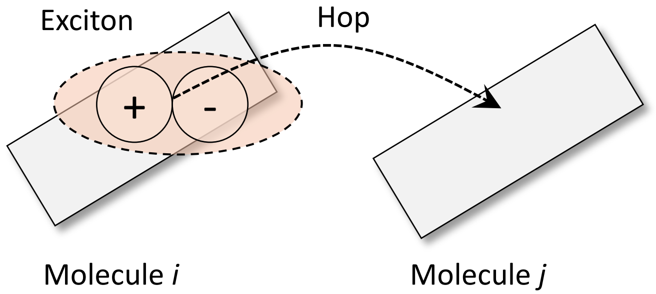


Important parameter 2.

Exciton coupling

Marcus equation:

$$k_{ij} = \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \frac{\overbrace{v_{ij}^2}^{\text{Exciton coupling}}}{\hbar} \exp\left(-\frac{\overbrace{\lambda}^{\text{Reorganization energy}}}{4k_B T}\right)$$



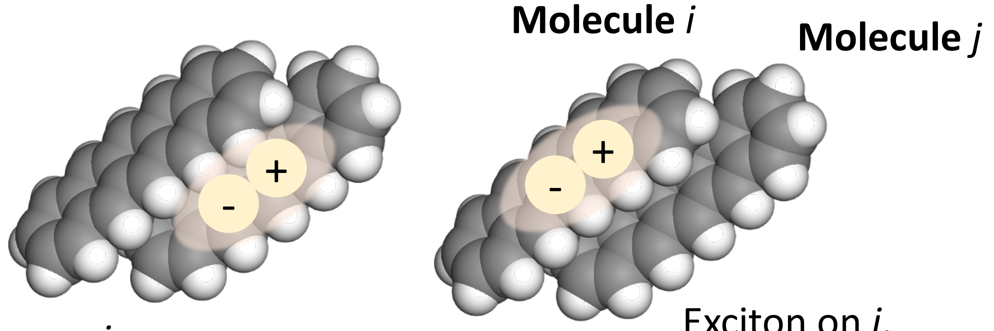
The exciton coupling v_{ij} is a quantum mechanical parameter. Its interpretation is somewhat abstract:

- Start with $|\psi_{ij}\rangle$. This is the wavefunction for the two molecules together, with the exciton on molecule *i* (initial state, before the hop).

- Electron-electron interactions between molecules perturbs $|\psi_{ij}\rangle$. The perturbed wavefunction is represented as $H_e|\psi_{ij}\rangle$.

- Now consider $\langle\psi_{ji}|$. This is the final wavefunction, with the exciton on molecule *j* (final state, after the hop).

- Finally, $\langle\psi_{ji}|H_e|\psi_{ij}\rangle$ tells us the overlap between the perturbed initial wavefunction and the final wavefunction.



Exciton on *j*, *i* in ground state.

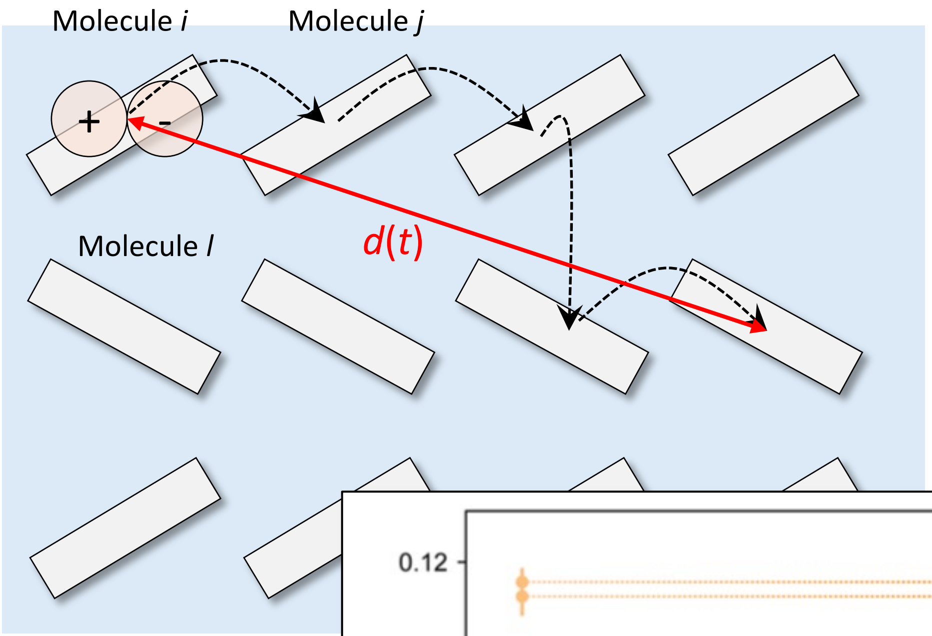
Exciton on *i*, *j* in ground state.

$$v_{ij} = \langle\psi_{ji}|H_e|\psi_{ij}\rangle$$

Electron-electron interaction operator

Both of these parameters can be calculated from first-principles (time-dependent DFT / TDDFT). However, these calculations are very time consuming!

Kinetic Monte Carlo (kMC) simulation of exciton diffusion



Once the hopping rates are calculated, the hopping process can be simulated using the kinetic Monte Carlo (kMC) method.

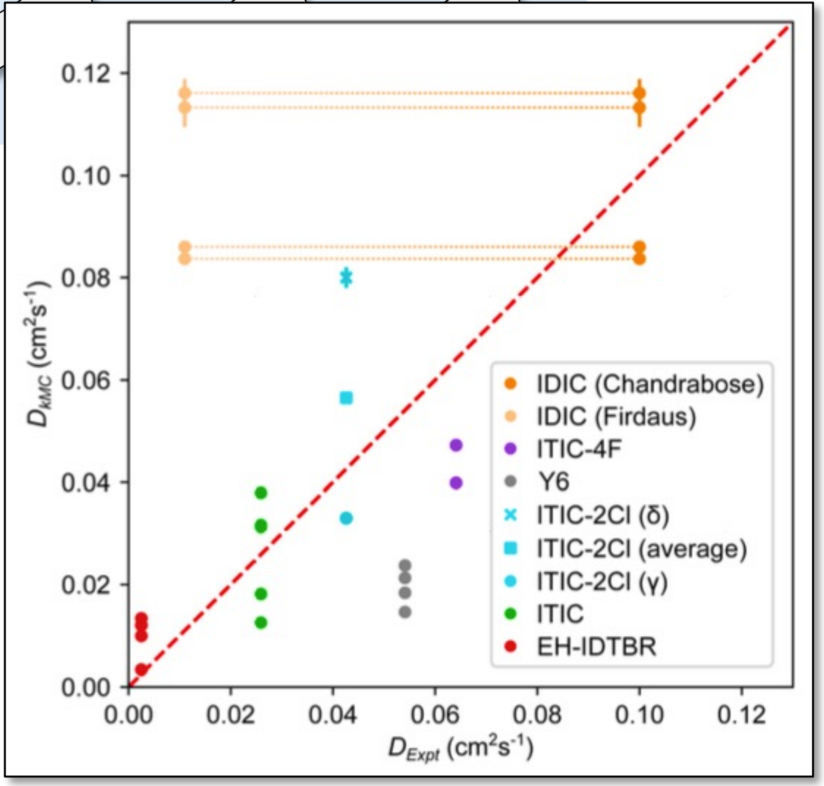
Roughly, the exciton hops to its neighbor *j* with probability proportional to k_{ij}. This results in a random walk-type motion through the crystal*.

Exciton diffusion coefficient can be estimated as:

Mean-square displacement

$$D = \lim_{t\to\infty} \frac{1}{6} \frac{\overline{d(t)^2}}{t}$$

Calculations of *D* compared to spectroscopic measurements (crystalline materials)



J. Mater. Chem. C. **24**, 2024, 8747

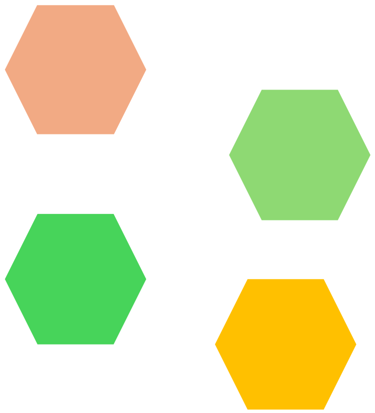
kMC simulations yield reasonable predictions of diffusion coefficients.

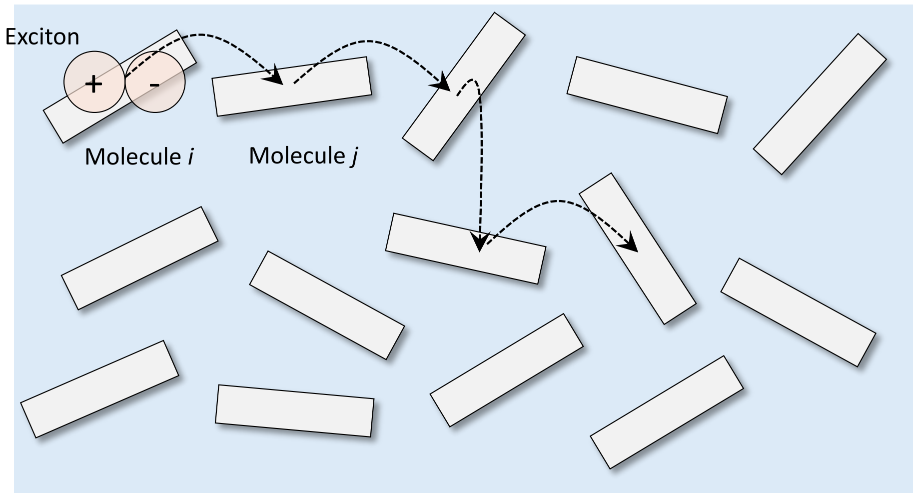
Could we discover new organic materials with large diffusion coefficients by high-throughput kMC simulations?

* Other details: *Phys. Chem. Chem. Phys.* **21**, 2019, 25023

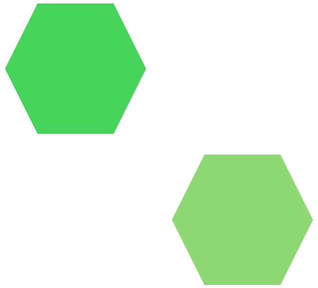
Kinetic Monte Carlo (kMC) for materials discovery?

Candidate organic semiconductors





Candidates with high exciton diffusion coefficients



Looks feasible – kMC can be done quite quickly on a laptop (minutes per candidate).

Problem: kMC has a big computational overhead!

For each candidate, couplines and reorganization energies need to be pre-computed from TDDFT before kMC can start.

Can we use machine learning to predict these parameters quickly?

Exciton coupling

Reorganization energy

$$k_{ij} = \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \frac{v_{ij}^2}{\hbar} \exp\left(-\frac{\lambda}{4 k_B T}\right)$$

2021 - 2022

First attempt at machine learning for exciton couplings

Dr. Chayanit Wechwithayakhlung


Data science

Prof. Justin Hodgkiss

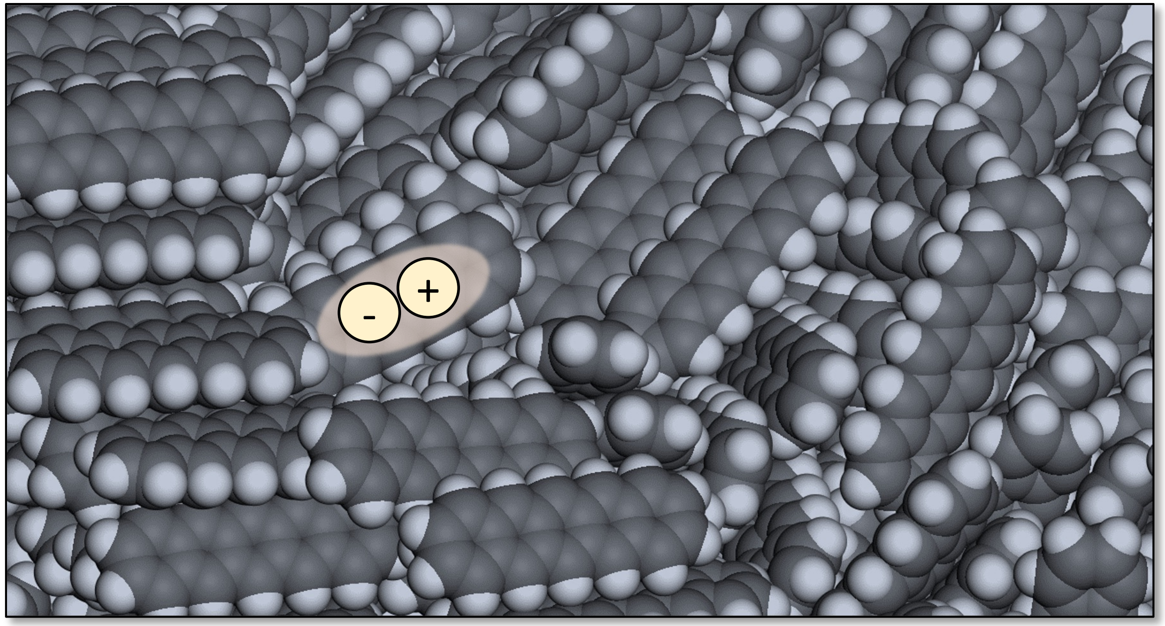
Dr. Paul Hume

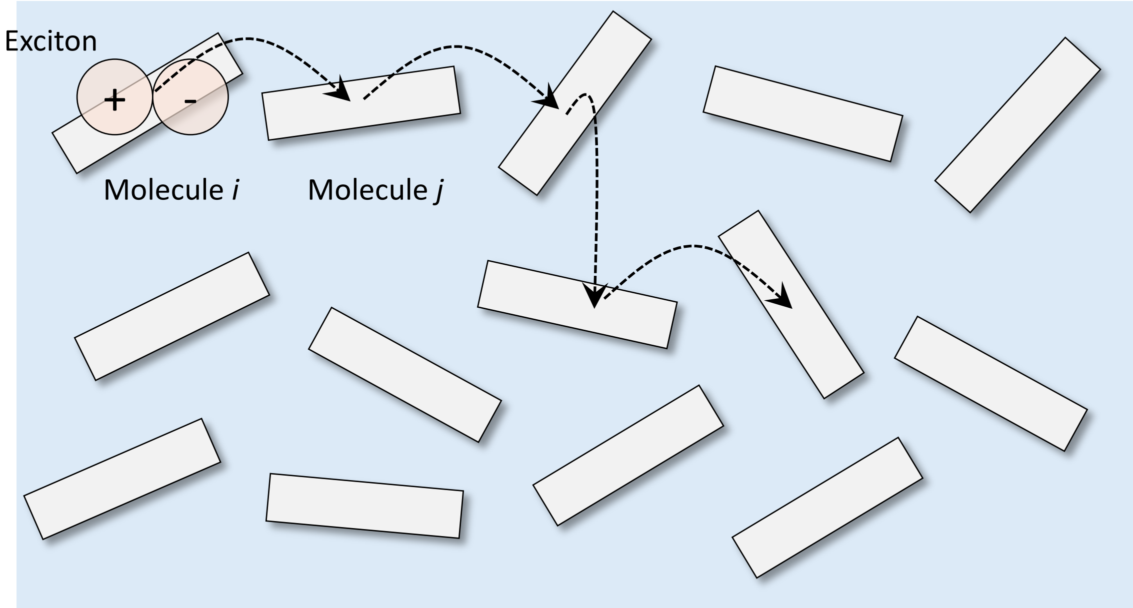
Dr. Geoff Weal



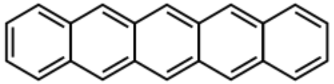
Ultrafast spectroscopy
Excited state quantum chemistry

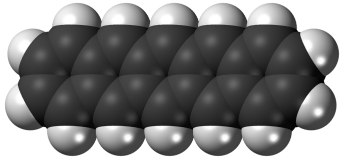
First target - exciton diffusion in amorphous pentacene





Pentacene molecule:

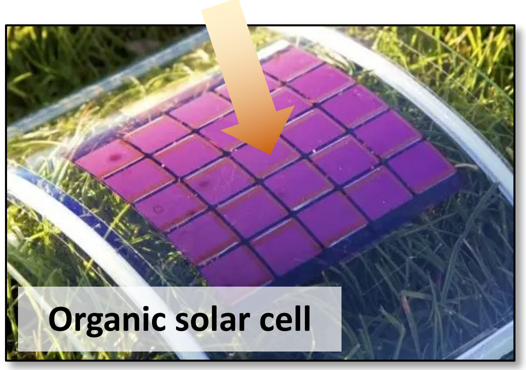




Exciton coupling

$$k_{ij} = \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \frac{v_{ij}^2}{\hbar} \exp\left(-\frac{\lambda}{4k_B T}\right)$$

hv



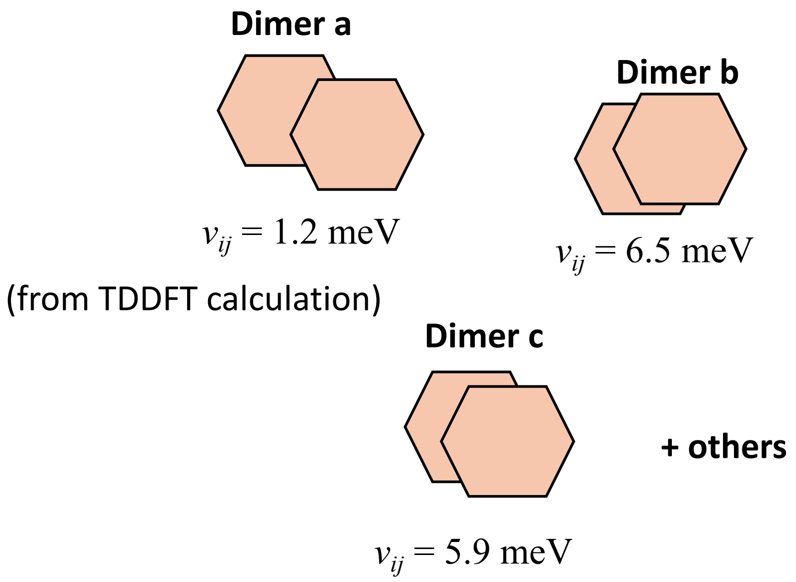
(www.solarreviews.com/blog/organic-solar-cells/)

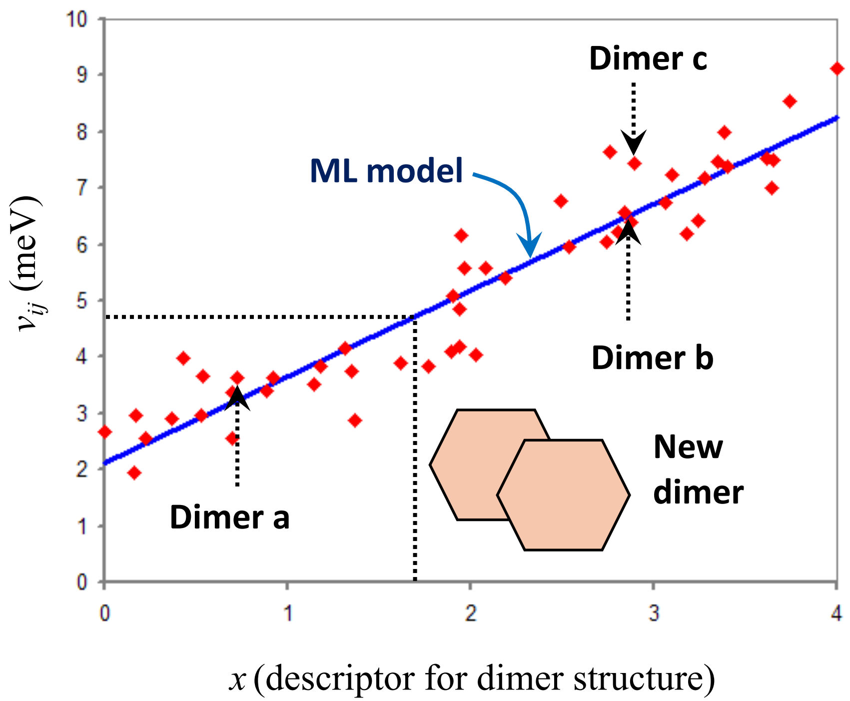
Amorphous pentacene is an organic semiconductor which can be used in organic solar cells.

Simple and relevant molecule, good starting point.

Machine learning concept

Training data

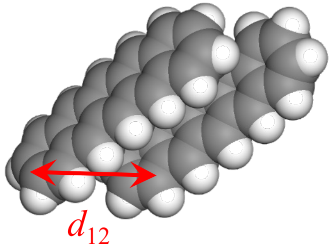




Again, we aim to fit a model which predicts coupling from dimer structure (x).

With the fitted model, couplings for new dimers can be quickly predicted.

ML model for exciton couplings



Dimer (coupling = v_{ij})

Pair of molecules within a 5 Å interatomic cut-off

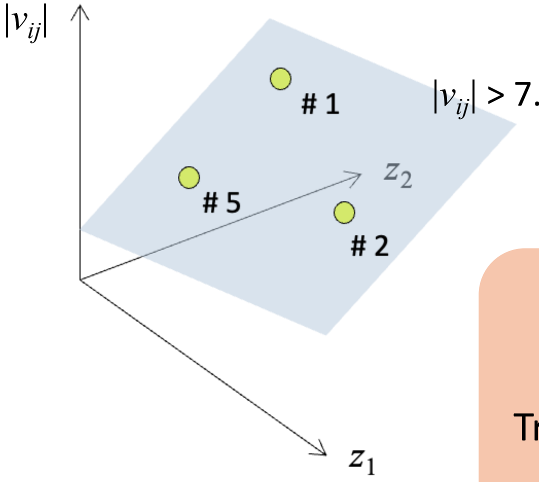
Coulomb-type descriptors

$x_1 = 1/d_{11}, x_2 = 1/d_{12}, \ldots$

Sorting
PCA

Support vector machine

Classify coupling as strong or weak



$|v_{ij}|$ > 7.5 meV

$|v_{ij}|$ < 7.5 meV

Kernel ridge regression 1

Trained on strong cases only

Kernel ridge regression 2

Trained on weak cases only

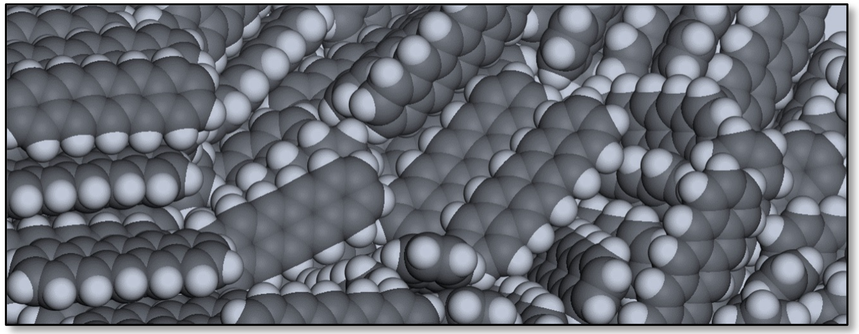
v_{ij} prediction

v_{ij} prediction

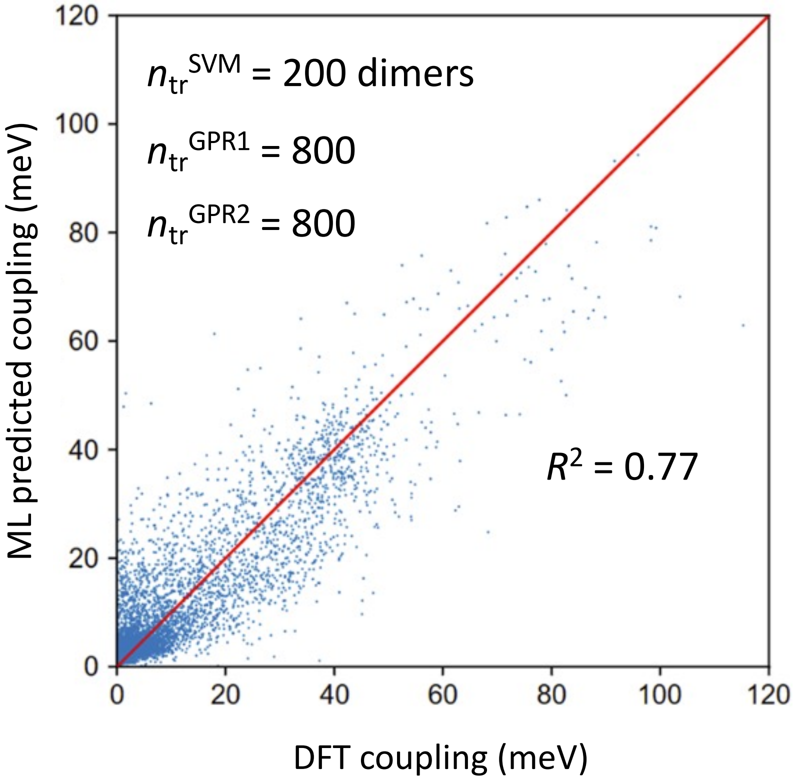
DFT settings: CAM-B3LYP xc functional with 6-3111+G(2d,p) basis set as implemented in Gaussian.

System under consideration:

Amorphous pentacene (generated by **Dr. Yu Kaneko (DAICEL)**)



Performance on 4127 dimers



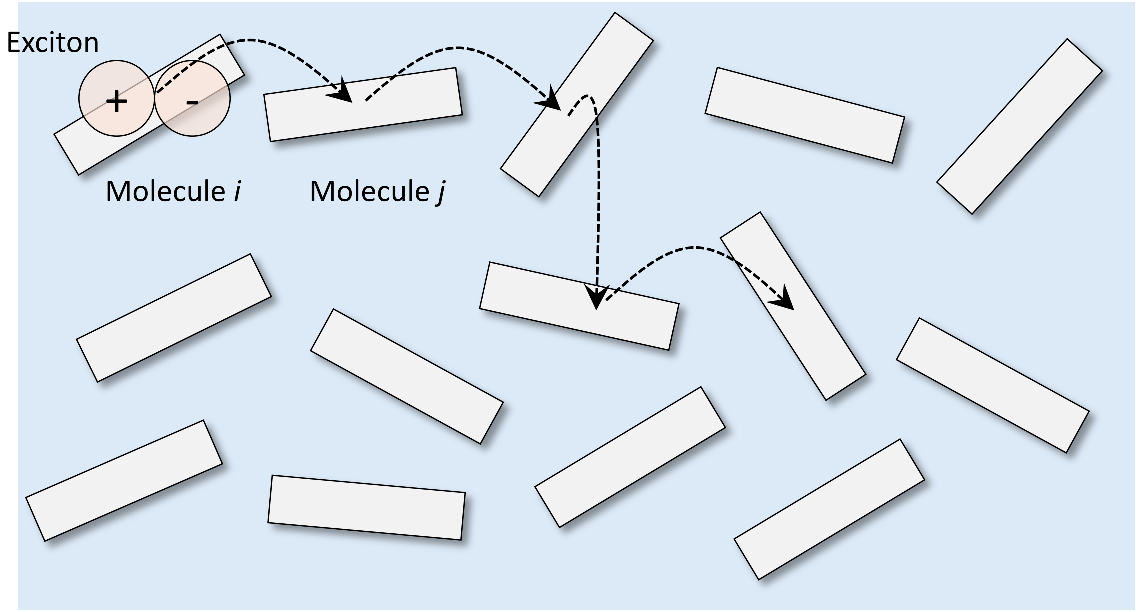
Average prediction time around 7 ms on an office workstation*

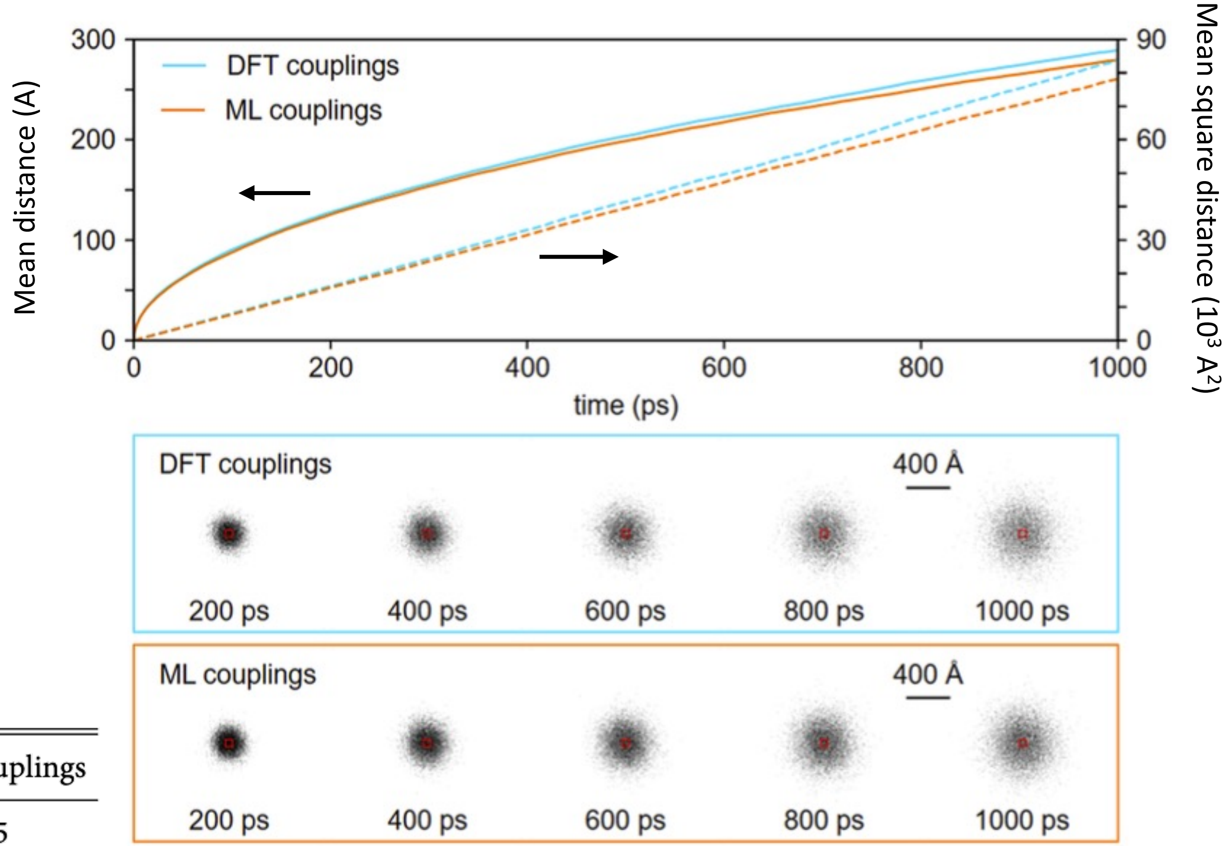
(cf. 7.4 hours for TDDFT on supercomputer**)

* Single 3.50 GHz Intel Xeon E5-1620 core

** 64 cores (128 threads), 2.0 – 3.35 GHz AMC Epyc 7702 processors.

Kinetic Monte Carlo simulation in amorphous pentacene





(Statistics from 10^4 trials)

	Ab initio couplings	Model-predicted couplings
Diffusion coefficient ($\times 10^{-3}$ cm^2 s^{-1})	1.630 ± 0.011	1.547 ± 0.005
Diffusion tensor eigenvalues ($\times 10^{-3}$ cm^2 s^{-1})		
Major	1.815 ± 0.014	1.686 ± 0.017
Middle	1.551 ± 0.016	1.492 ± 0.007
Minor	1.525 ± 0.012	1.462 ± 0.014

(Experimental diffusion coefficient for multicrystalline pentacene: 0.5 x 10^{-3} cm^2 s^{-1})

Good start, but a serious problem remains:

Coupling model restricted to only one type of molecule (pentacene). Cannot use this for virtual screening!

Withwechayakhlung *et al. J. Chem. Phys.* **158**, 2023, 204106

2023 - 2024

General exciton coupling model

Dr. Geoff Weal

Data science

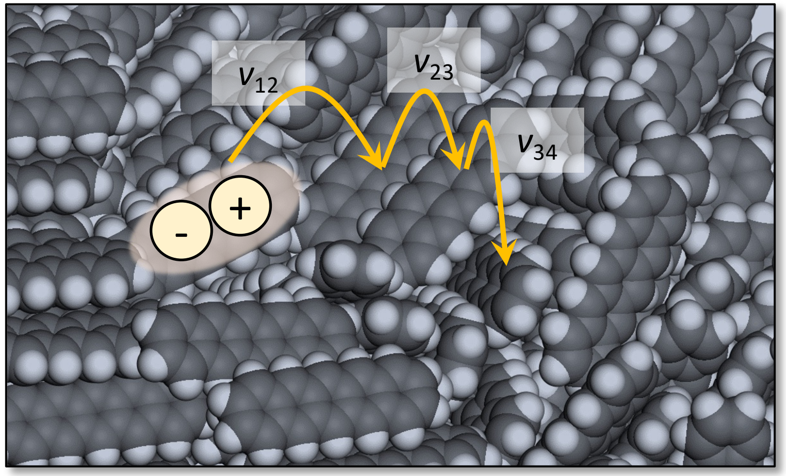
Prof. Justin Hodgkiss

Dr. Paul Hume

Ultrafast spectroscopy
Excited state quantum chemistry

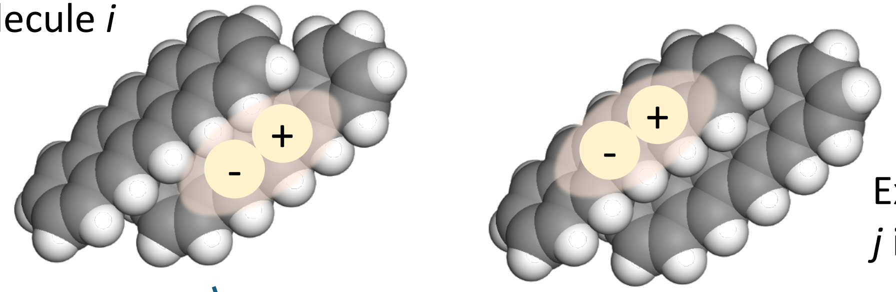
JSPS Postdoctoral Fellowship

Towards a new coupling model that works for all types of molecules?



Molecule i

Molecule j



Excitation on j,
i in ground state.

Excitation on i,
j in ground state.

$$v_{ij} = \langle \psi_{ji} | H_e | \psi_{ij} \rangle$$

Longuet-Higgins showed that v_{ij} can be approximated as a sum over simple, Coulomb-like terms*:

$$v_{ij} \approx \sum_{\alpha \in i, \beta \in j} \frac{Q^i_\alpha Q^j_\beta}{\mathbf{r}_\alpha - \mathbf{r}_\beta}$$

$Q_\alpha{}^i$ is the **atomic transition charge** (ATC) for atom α.

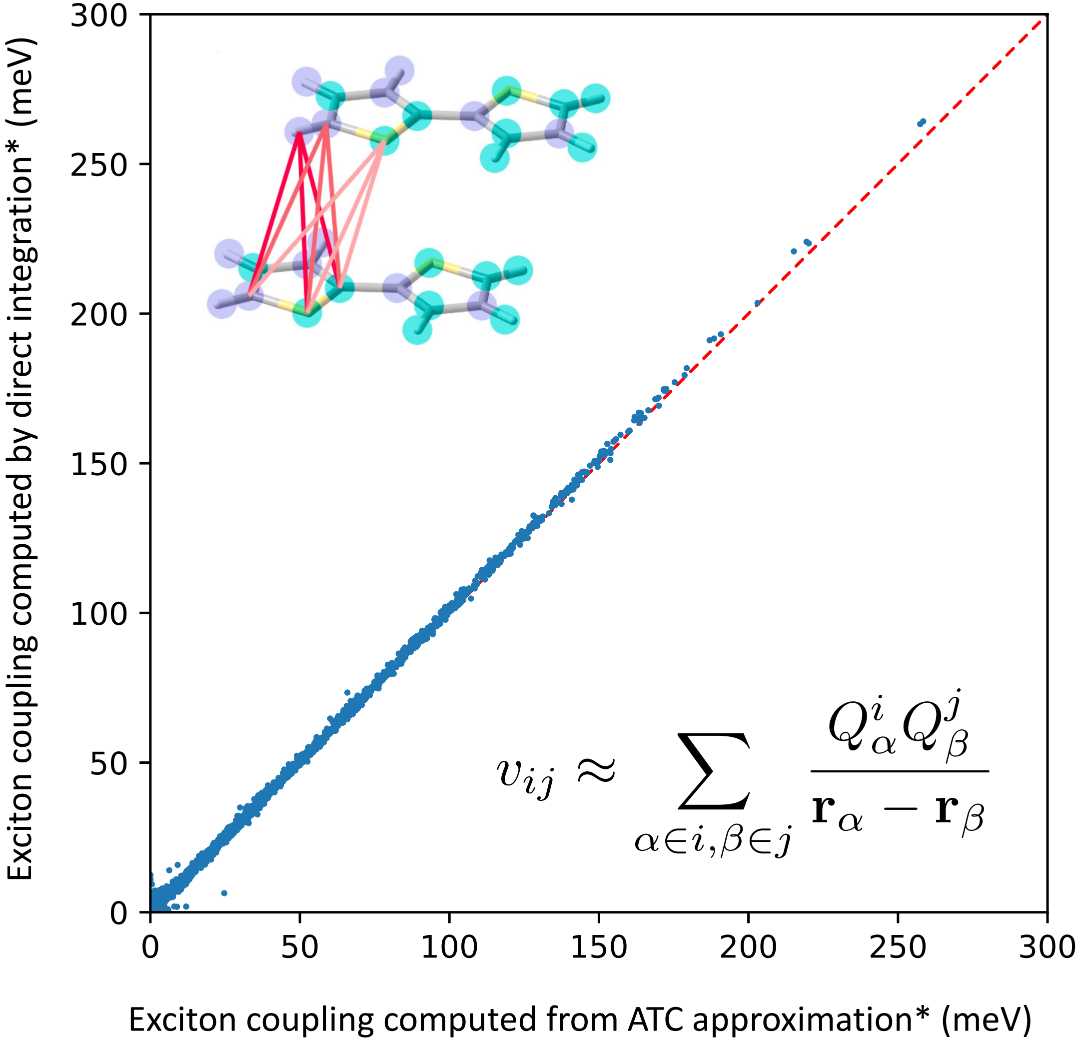
It measures the change in the electron density on atom α when molecule i transitions from the ground to the excited state ($\rho_{\mathrm{tr}}{}^i(\mathbf{r})$ is called the transition density)

$$Q^i_\alpha = \int_{\mathbf{r}\text{ on atom }\alpha} \rho^i_{\mathrm{tr}}(\mathbf{r})\, d\mathbf{r}$$

This is known as the **ATC approximation**. It has never been tested extensively, so we put it to work…

*Longuet-Higgins. *Proc. Roy. Soc.* **235**, 1956, 537

Towards a new coupling model that works for all types of molecules?



We tested the ATC approximation for molecular dimers extracted from 1989 organic crystals in the CCDC database (Cambridge Crystal Data Center)

Good agreement, with only small deviations for strong coupling cases (ATC approximation neglects some close-range quantum effects)

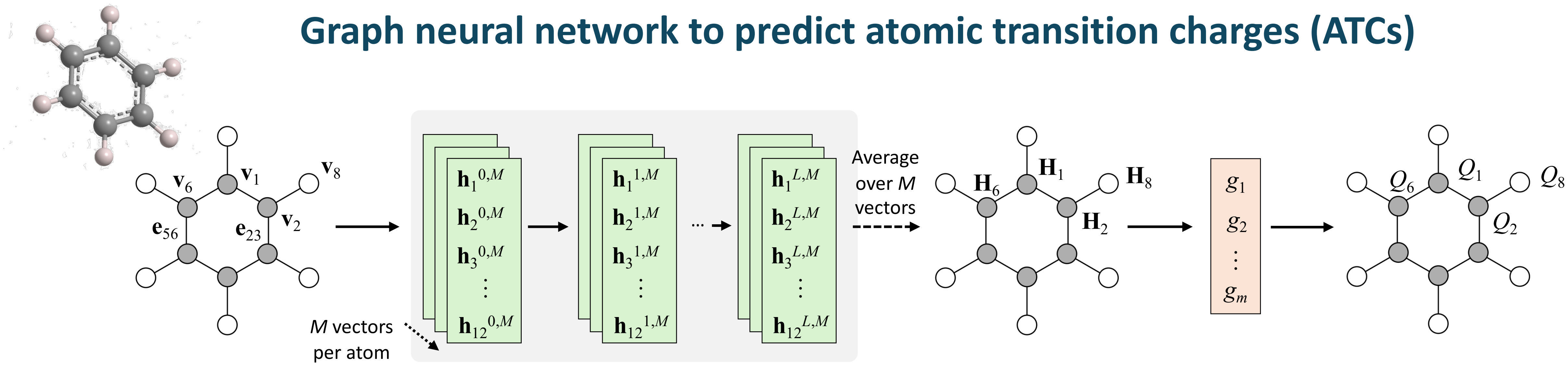
=> Instead of creating an ML model for couplings directly, let's **make an ML model for atomic transition charges** (Q_α).

We could then compute v_{ij} for any type of molecule using the ATC approximation.

* TDDFT with ω-B97XD xcf with 6-31+G(d,p) basis set as implemented in Gaussian 16.

Weal *et al. J. Chem. Phys.* **163**, 2025, 024125

Graph neural network to predict atomic transition charges (ATCs)



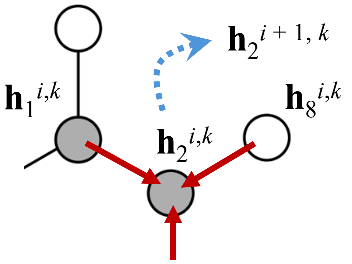
Molecular graph

Initial atom descriptors $\mathbf{v}_i$

Initial bond descriptors $\mathbf{e}_{ij}$

Iterative feature embedding

Message parsing between atoms



Molecular graph

Final atom descriptors

Graph readout

Atom readout

ATC predictions

→ Neural networks / non-linear transformations

Implementation

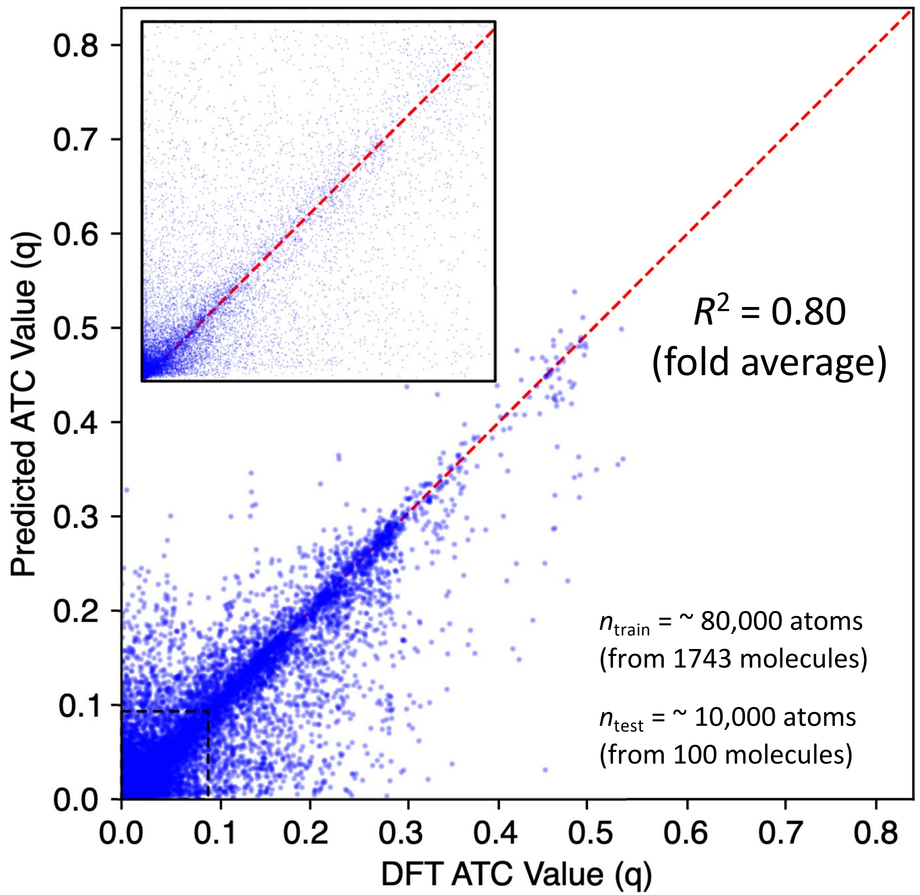
SOAP (Smooth Overlap of Atomic Positions) descriptors used for $\mathbf{v}_i$.
An integer-valued descriptor corresponding to hybridization state also incorporated (sp, sp2, sp3).

Integer-valued descriptor for bond type (single, double, etc) used for $\mathbf{e}_{ij}$.

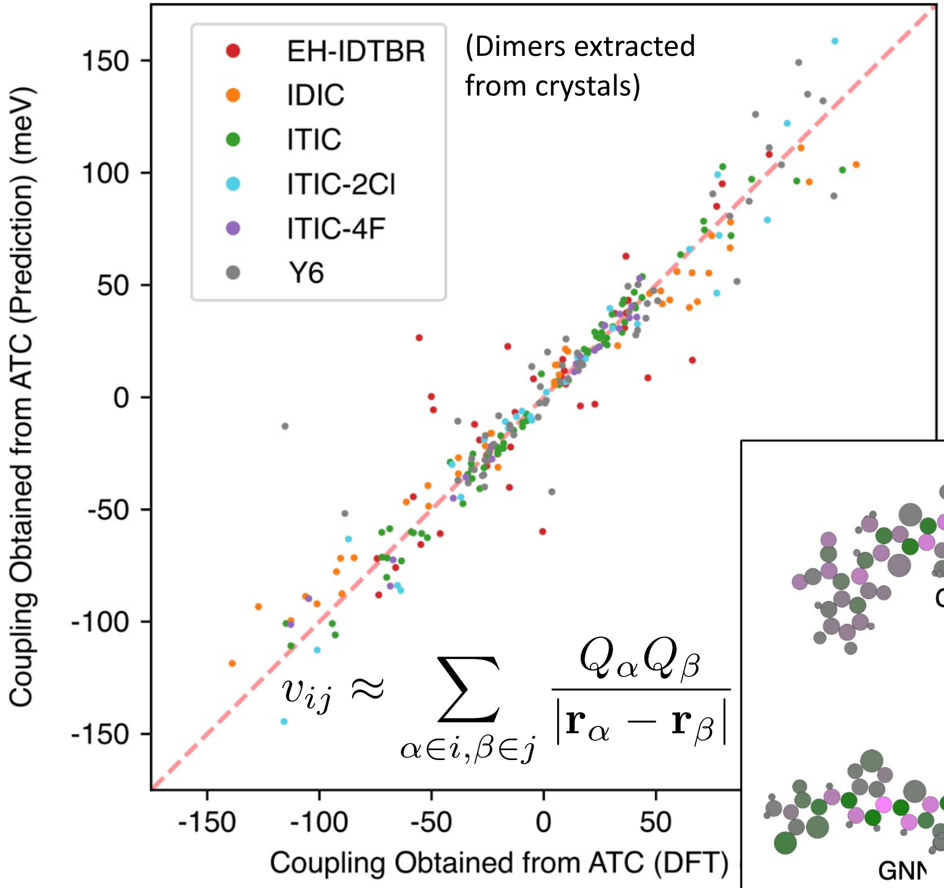
Other settings and training procedure followed Han *et al. Phys. Chem. Chem. Phys.* **24**, 2022, 26870.

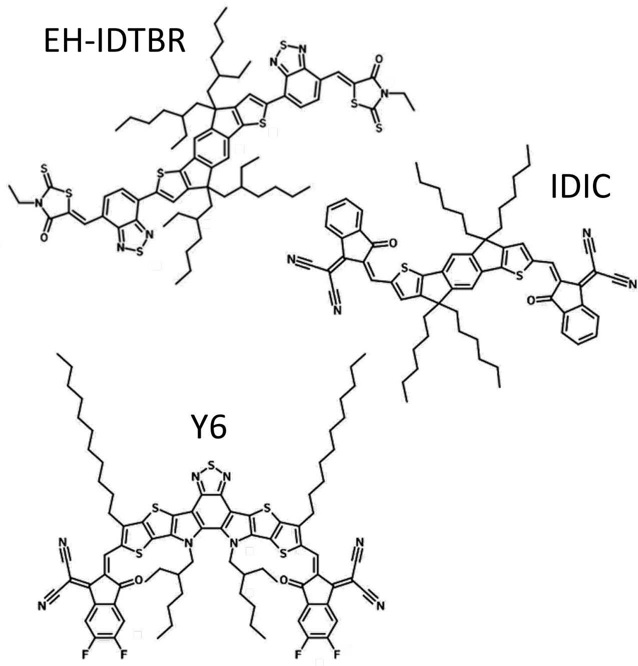
Graph neural network performance

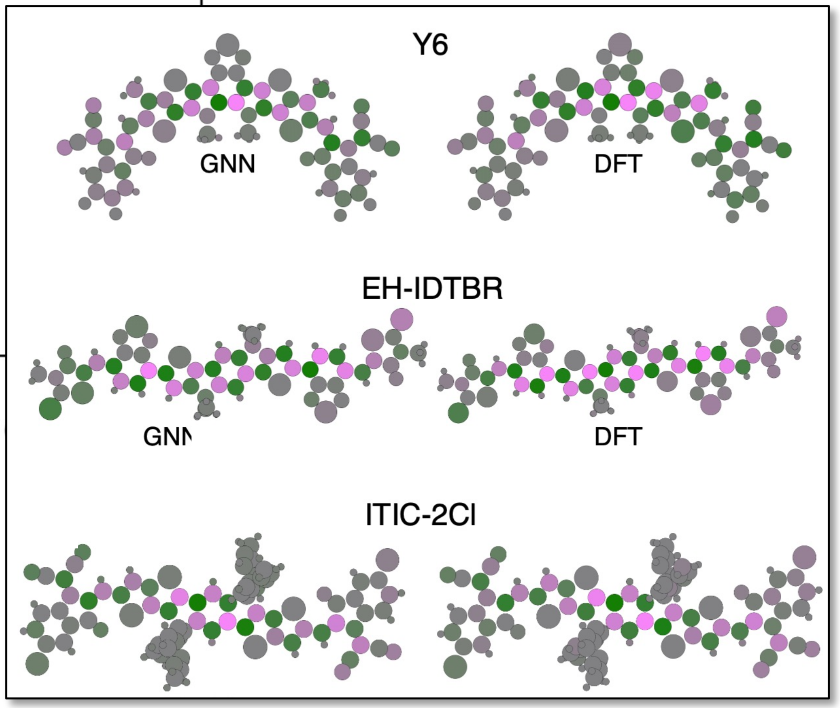
Performance on test set (10xCV fold)



Exciton coupling predictions for important organic photovoltaic materials (FREAs)





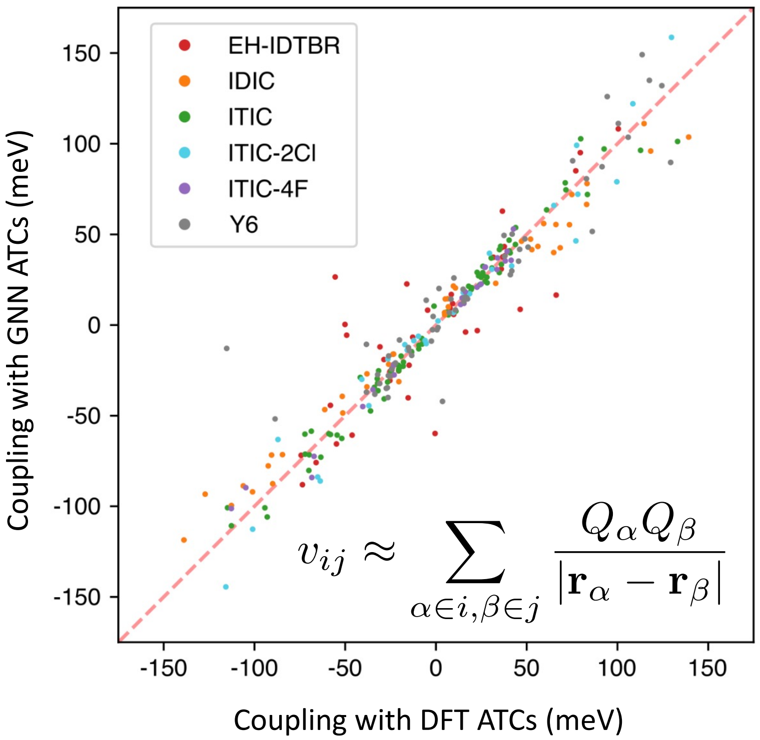


- Reasonable accuracy when predicting ATCs.
- **Generality achieved** - couplings predicted with good accuracy for a range of important organic photovoltaic materials!

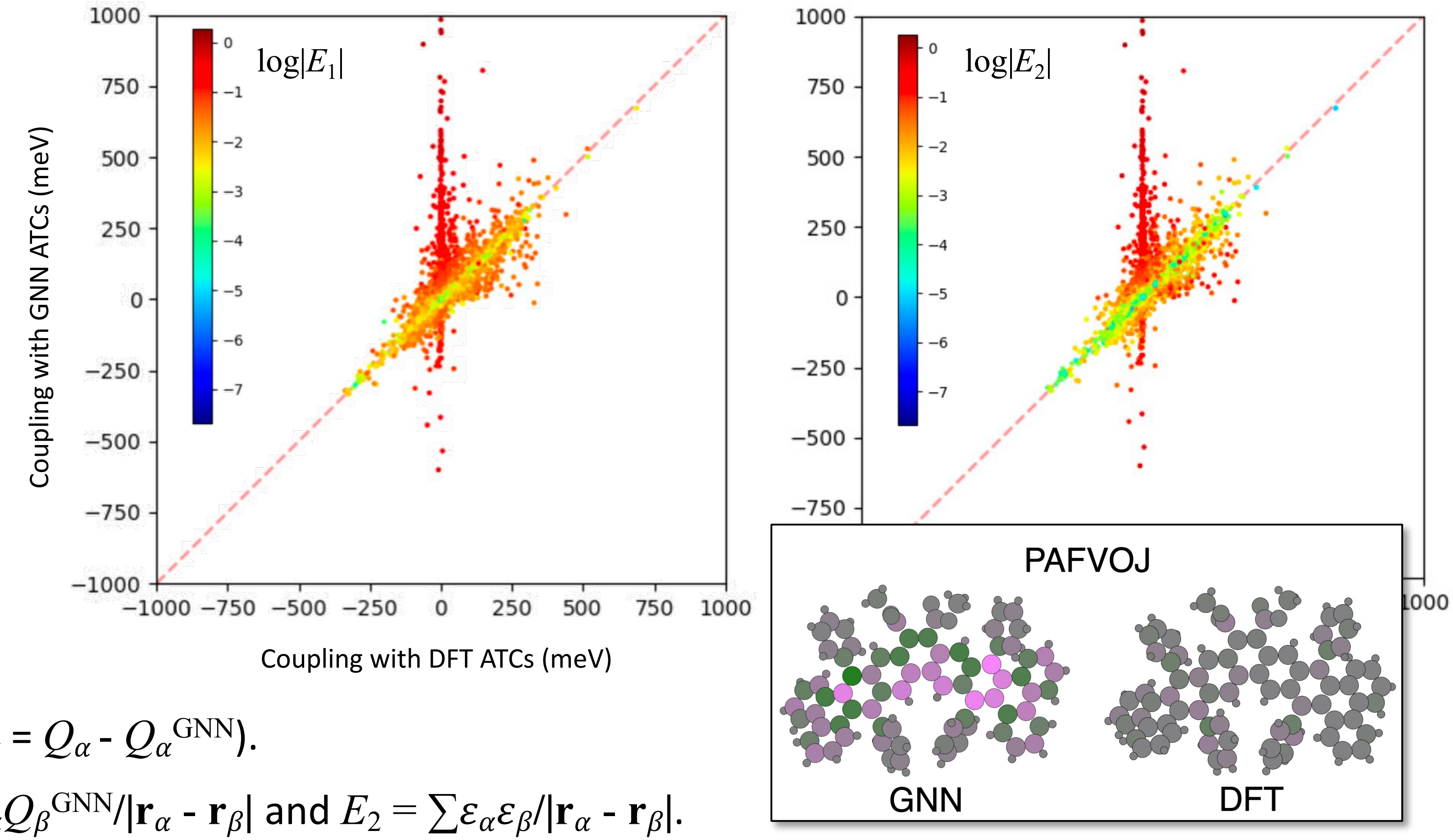
Weal *et al. J. Chem. Phys.* **163**, 2025, 024125

How general? Error propagation limits accuracy for weakly coupled cases

Coupling predictions for important organic photovoltaic materials (FREAs)



Couplings for dimers extracted from 1000 organic crystals



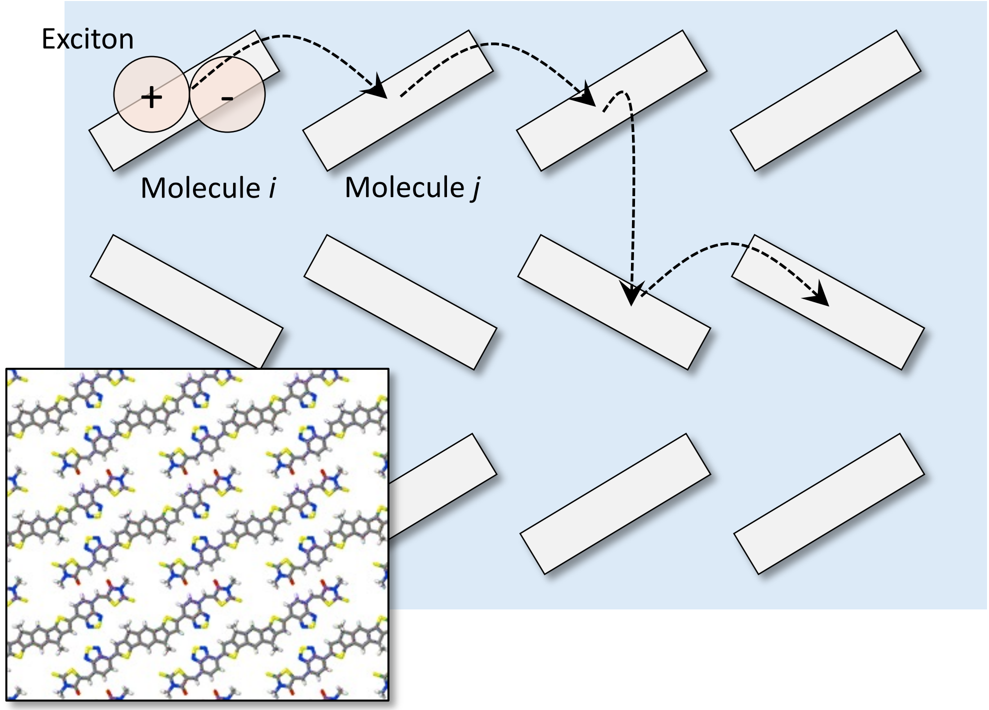
Let ε_α be the error of the Q_α prediction ($\varepsilon_\alpha = Q_\alpha - Q_\alpha^{\mathrm{GNN}}$).

Then $v_{ij} = v_{ij}^{\mathrm{GNN}} + E_1 + E_2$, where $E_1 = \sum \varepsilon_\alpha Q_\beta^{\mathrm{GNN}}/|\mathbf{r}_\alpha - \mathbf{r}_\beta|$ and $E_2 = \sum \varepsilon_\alpha \varepsilon_\beta/|\mathbf{r}_\alpha - \mathbf{r}_\beta|$.

First- and second-order errors can be large for dimers where coupling is in fact weak ($|v_{ij}|$ less than about 75 meV). For other cases, the method seems reliable.

Weal *et al. J. Chem. Phys.* **163**, 2025, 024125

Kinetic Monte Carlo simulations using GNN-predicted ATCs

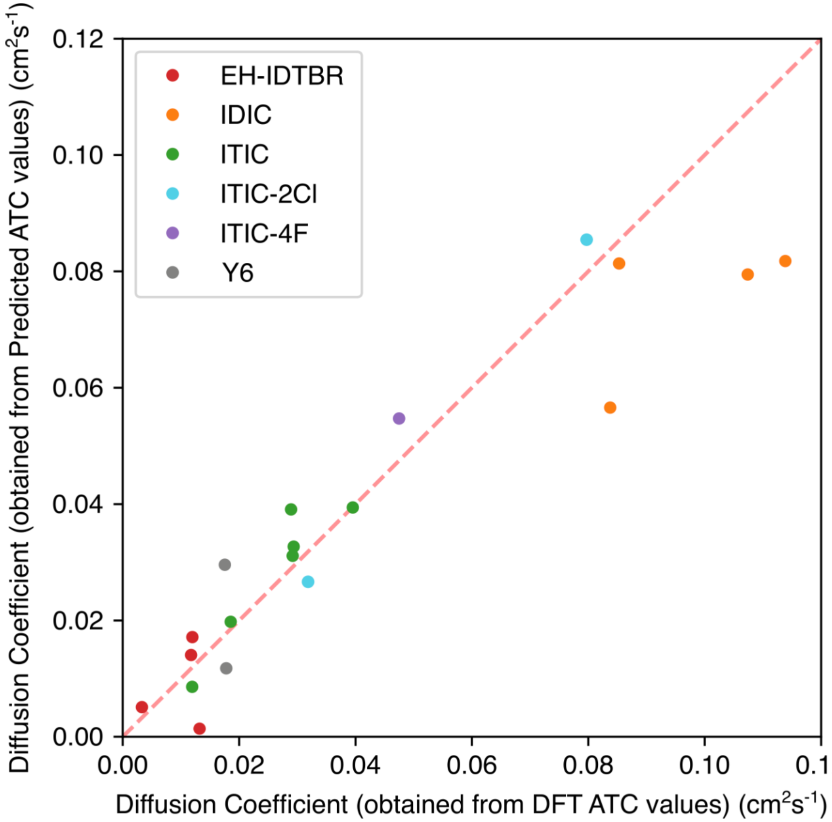


Marcus hopping rates:

GNN-predicted exciton coupling

Reorganization energy (TDDFT-calculated)

$$k_{ij} = \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \frac{v_{ij}^2}{\hbar} \exp\left(-\frac{\lambda}{4k_B T}\right)$$

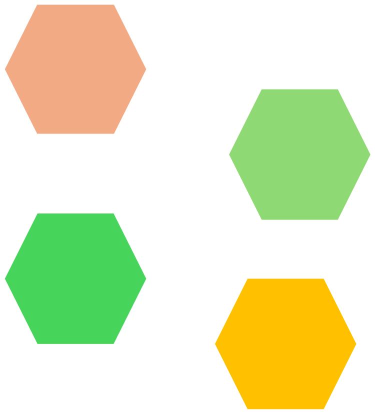


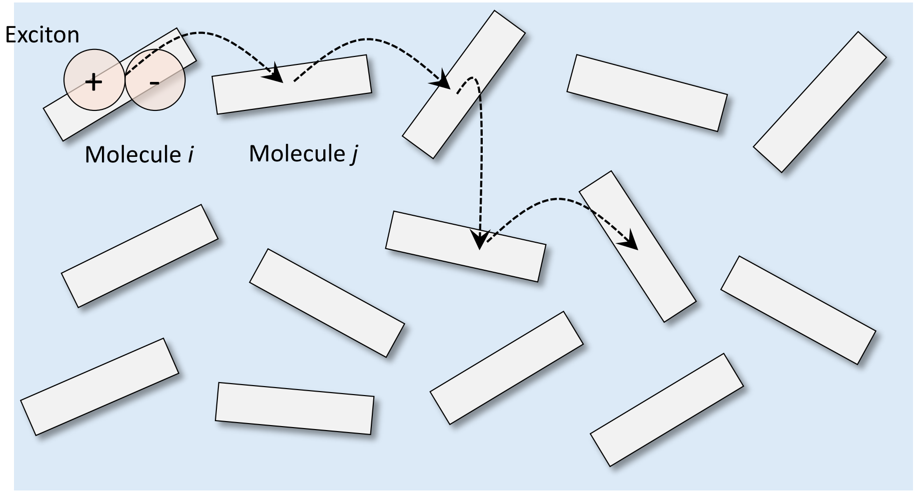
Good predictions of diffusion coefficient obtained for multiple materials!

Weal *et al. J. Chem. Phys.* **163**, 2025, 024125

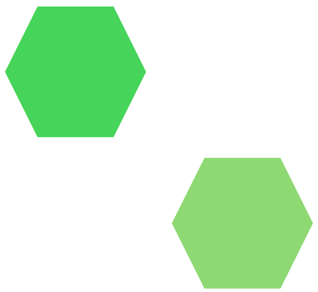
Closer to high-throughput kinetic Monte Carlo (kMC)?

Candidate organic semiconductors





Candidates with high exciton diffusion coefficients



Computational overhead remains!

We succeeded to create a general ML scheme to predict couplings.

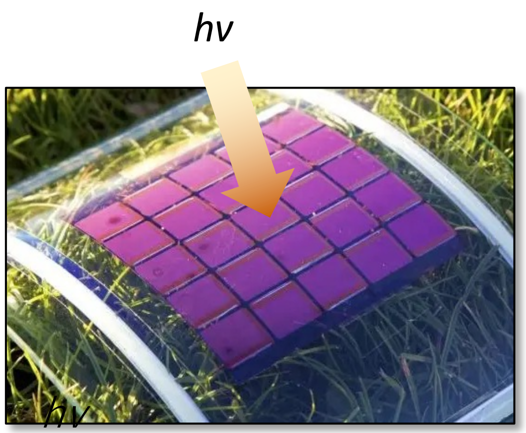
But reorganization energies still require expensive time-dependent DFT calculations. We still have work to do!

GNN-predicted exciton coupling

Reorganization energy (TDDFT-calculated)

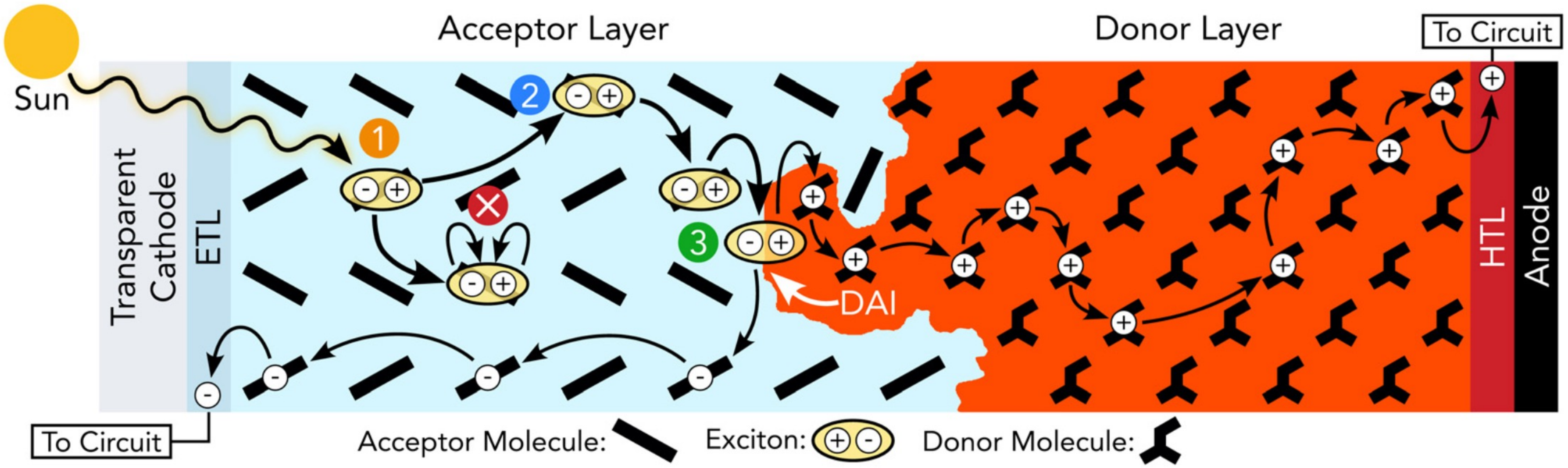
$$k_{ij} = \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \frac{v_{ij}^2}{\hbar} \exp\left(-\frac{\lambda}{4 k_B T}\right)$$

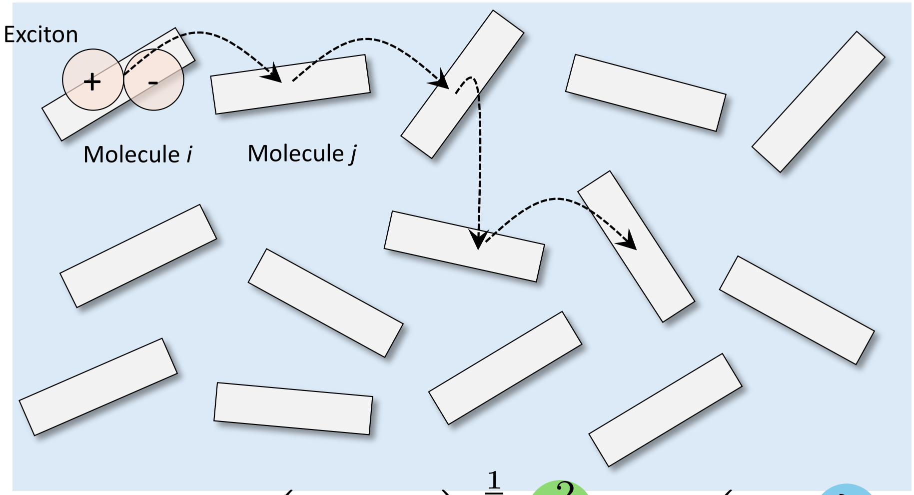
Summary of part 2



Organic solar cell

(www.solarreviews.com/blog/organic-solar-cells/)





$$k_{ij} = \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \frac{v_{ij}^2}{\hbar} \exp\left(-\frac{\lambda}{4k_B T}\right)$$

- Kinetic Monte Carlo simulations *might* be used for screening organic semiconducting materials. However, the computational times required for exciton hopping rates need to be significantly reduced first.
- We created a new method for quickly computing exciton coupling parameters. It combines the atomic transition charge approximation and a graph neural network.
- The method generalizes widely across different molecule types.

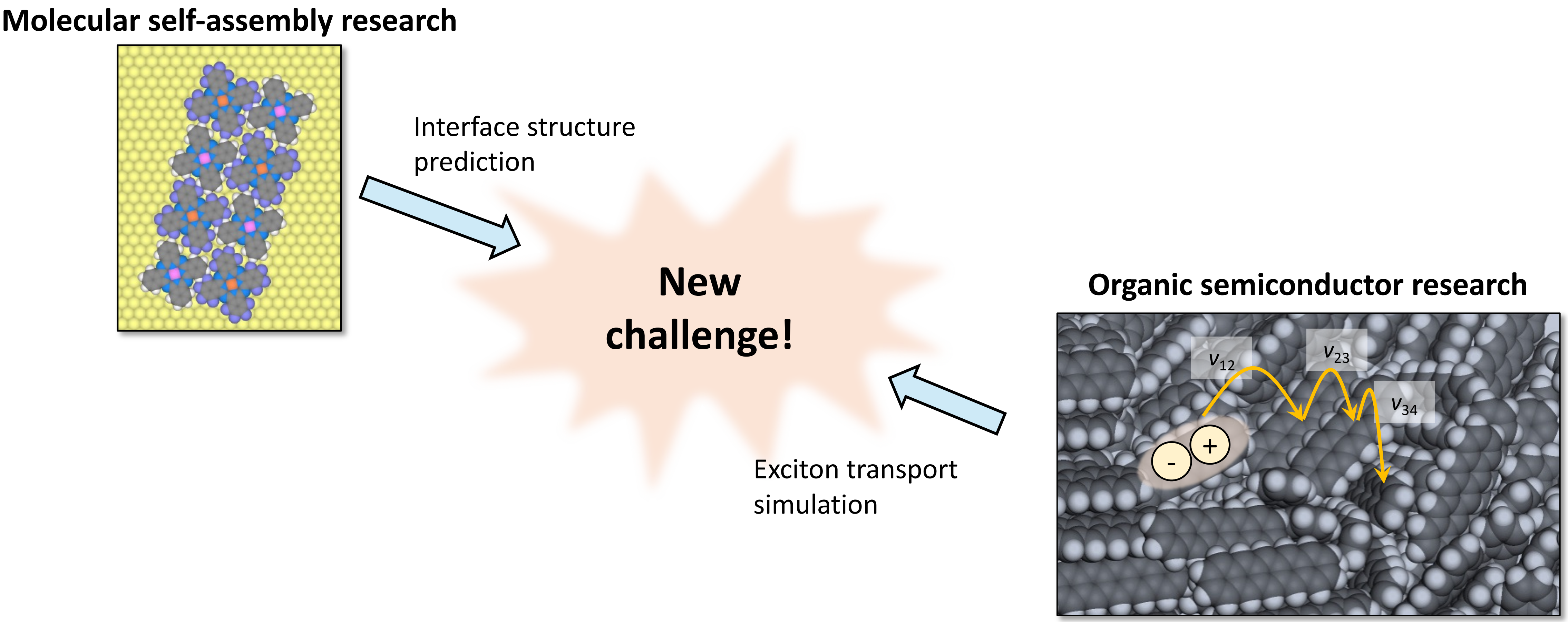
References: Wechwithayakhlung *et al. J. Chem. Phys.* **158**, 2023, 204106
Weal *et al. J. Mater. Chem. C.* **24**, 2024, 8748
Weal *et al. J. Chem. Phys.* **163**, 2025, 024125

Lecture topics

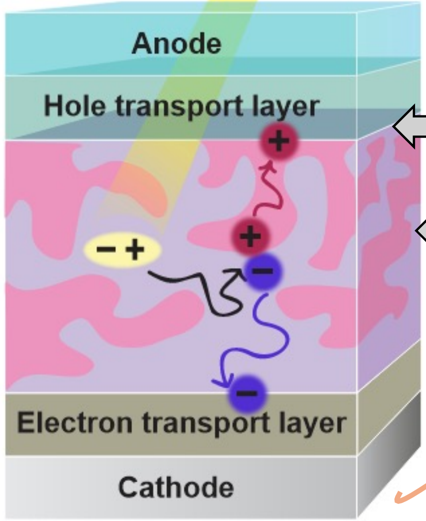
Simulation of on-surface molecular self-assembly

Machine learning for organic photovoltaic materials

What’s next?


Molecular self-assembly research
Interface structure prediction
New challenge!
Organic semiconductor research
V_{12}
V_{23}
V_{34}
+
-
Exciton transport simulation

e-ASIA JRP Project:
Interface Materials Informatics

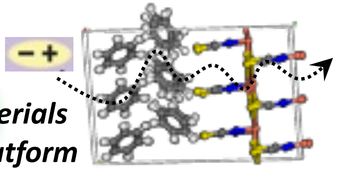


This work:
Interface region

Previous work:
Active organic layer

Thailand team
Inorganic semiconductors, experiment

Goal
Interface Materials Informatics Platform



Japan team
Interface modeling, machine learning

NZ team
Organic semiconductors, charge and energy simulations

iCeMS
KYOTO UNIVERSITY
WPI Research Center

VISTEC
VIDYASIRIMEDHI
INSTITUTE OF SCIENCE AND TECHNOLOGY

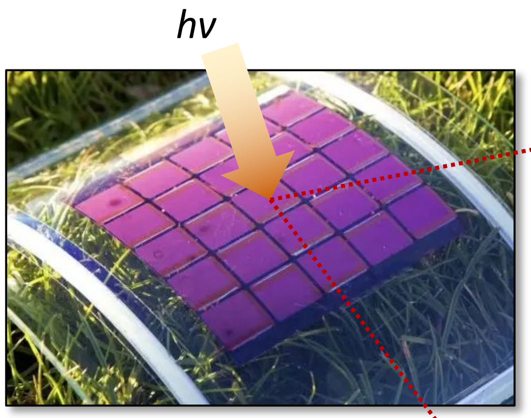
Pichaya Pattanasattayavong
(inorganic semiconductors)

Justin Hodgkiss
(organic semiconductors)

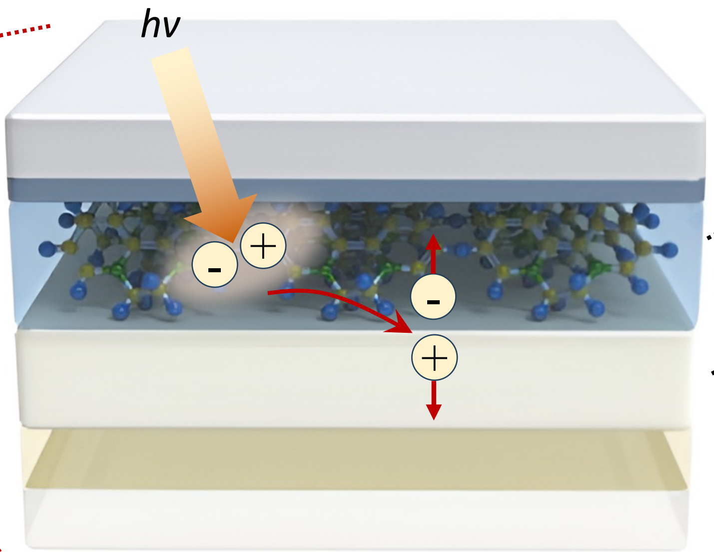
The MacDiarmid Institute
for Advanced Materials and Nanotechnology

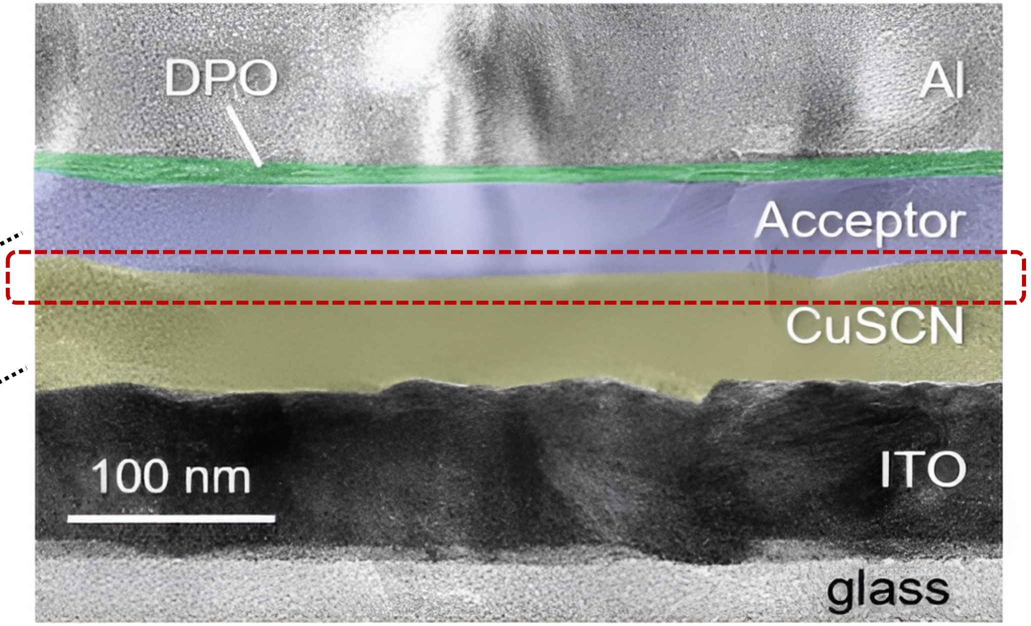
Started April 2025
(Japan side)

Project focus: organic-inorganic interface in organic solar cells

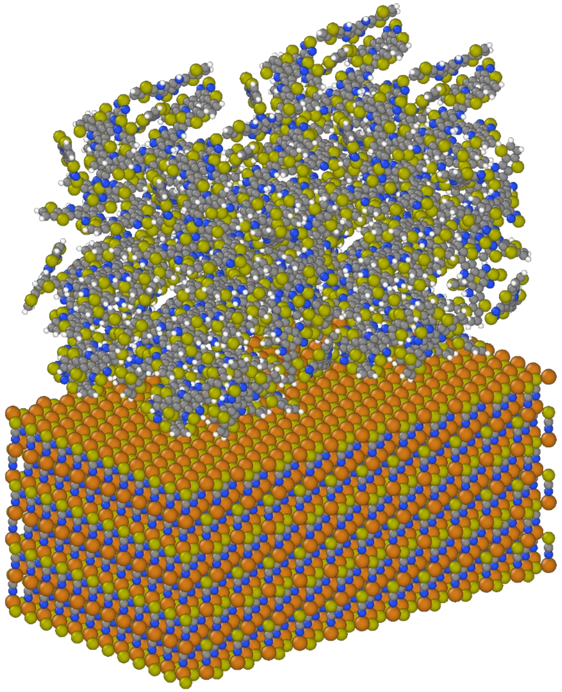


(www.solarreviews.com/blog/organic-solar-cells/)



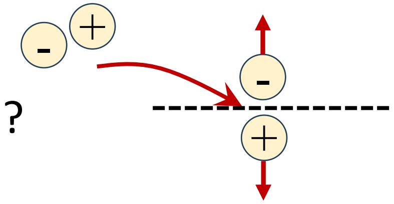


(https://discovery.kaust.edu.sa/en/article/6432/guiding-the-way-to-improved-solar-cell-performance/)



Can we predict the atomic-level structure of the interface?

Can we simulate charge separation at the interface?



Can we use these simulations to guide experimental solar cell fabrication?

Can experiment guide simulations?

(https://www.solar.fau.de/research/devices/)

Group members (2025)

Dr. Priya Dey
Chio Hayashi
Dr. Bandon Meza
Dr. Maryam Nurhuda
Dr. Shreya Rastogi
Dr. James Scott

Previous members

Dr. Geoff Weal
Dr. Chayanit Wechwithayakhlung

Collaborators (on-surface self-assembly)

Dr. Patrick Han
Prof. Taro Hitosugi

Collaborators (organic photovoltaics)

Dr. Yu Kaneko (DAICEL)
Dr. Paul Hume (MacDiarmid Institute)
Dr. Joshua Sutton (MacDiarmid Institute)
Prof. Justin Hodgkiss (MacDiarmid Institute)

2024

iCeMS KYOTO UNIVERSITY WPI Research Center

2022

Chayanit Wechwithayakhlung

Funding (on-surface self-assembly, organic photovoltaics)

JST PRESTO (2014 – 2018)
JSPS Kakenki Kiban C (21K05003)
JSPS Kakenhi Shingakujyutsu Koubo (19H04574)
JSPS Kakenhi Wakate (18K14126)
JSPS Kakenhi Shingakujyutsu Koubo (16H00879)
JSPS Bilateral Projects / MBIE CATALYST
JSPS Postdoctoral Fellowship for Foreign Researchers