Graph Neural Network Techniques for Molecular Prediction

Domain Adaptation applied to organic semiconductors.

Neel Misciasci

M.Sc. Mathematics in Science and Engineering

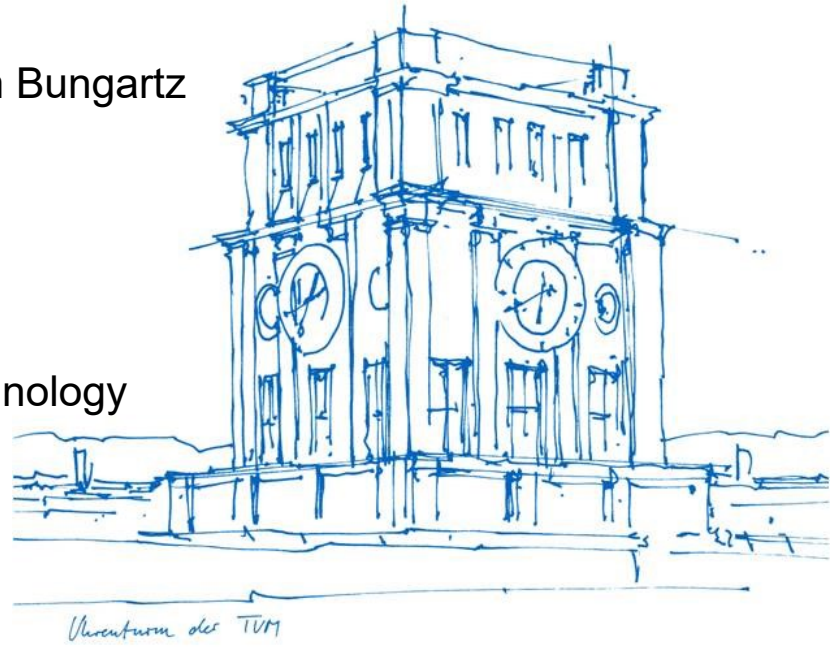
Supervisor: Prof. Dr. rer. nat. habil. Hans-Joachim Bungartz

Advisor: Kouroudis Ioannis, M.Sc.

Technische Universität München

TUM School of Computation, Information and Technology

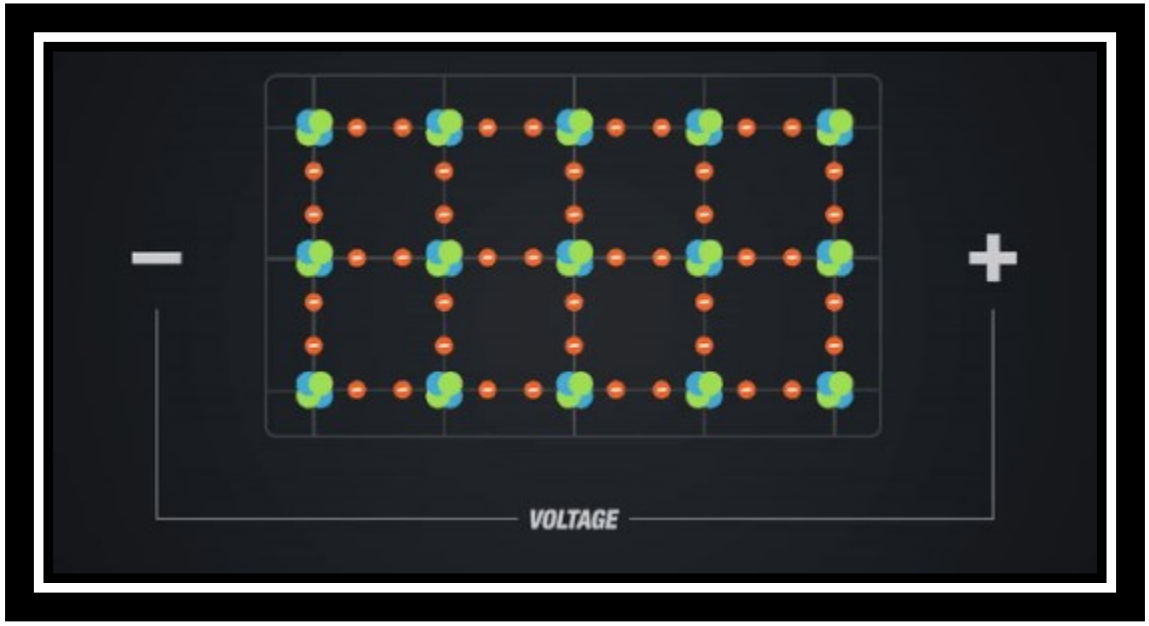
Munich, June 21st 2023



Semiconductors control the flow of current

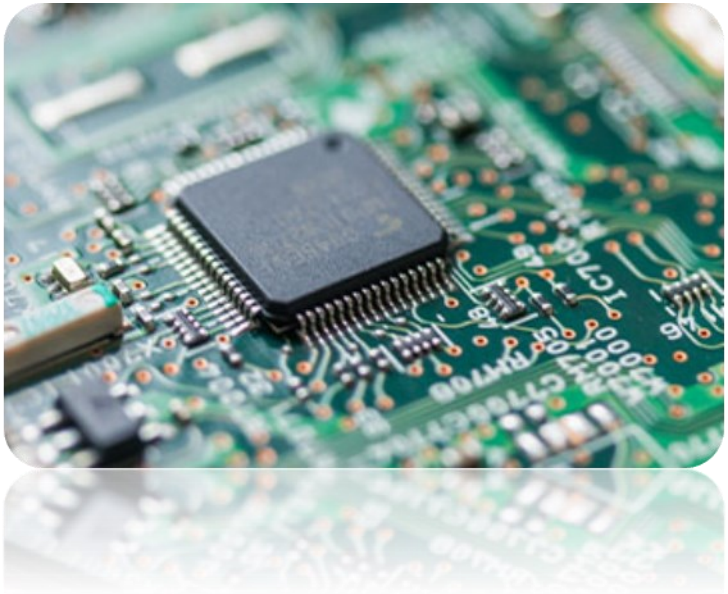
A semiconductor is the building material of ALL electronic devices.

Is neither an insulator nor a conductor.



Inorganic semiconductors are rigid

Semiconductors = Silicon



- Rigid
- Crystalline structure
- Relatively expensive

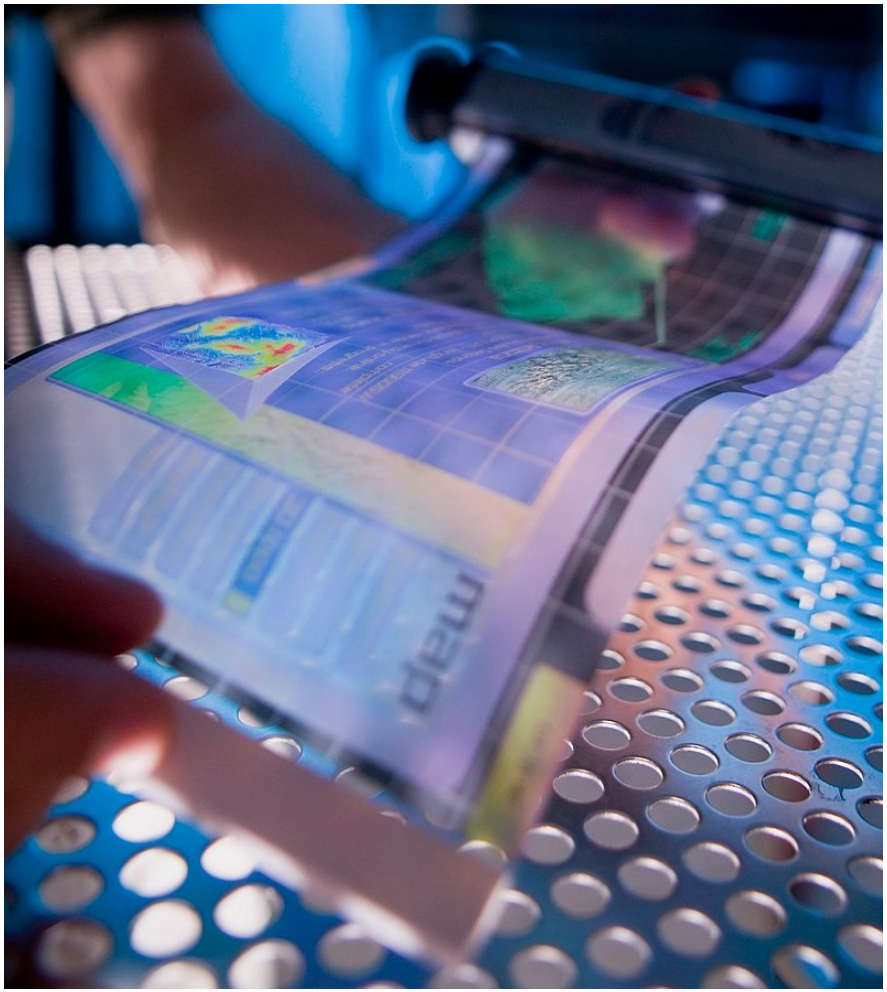
Organic Semiconductors (OS) are flexible

Semiconductors = Organic material

- Thin
- Flexible
- Stretchable
- Cheap to produce

OS have innovative applications

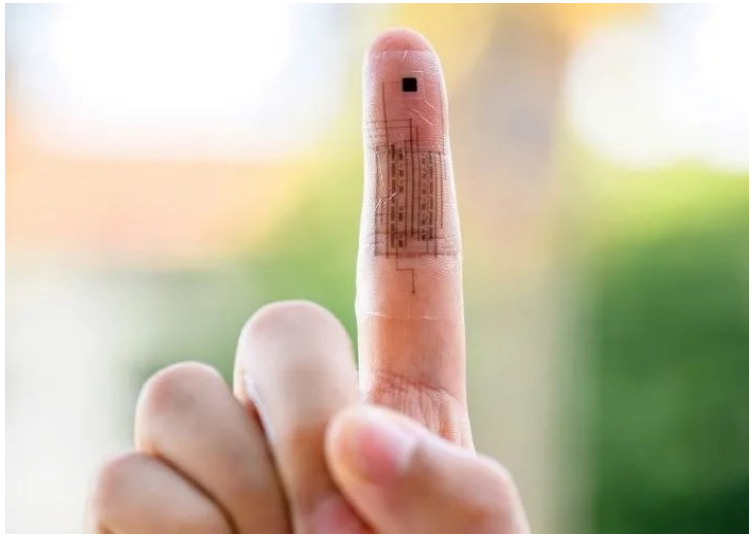
Flexible displays



OS have innovative applications

Next-gen wearable electronics

- Electronic skin for health monitoring.



Source: [1]

- Prosthetic limbs that mimic the sense of touch.

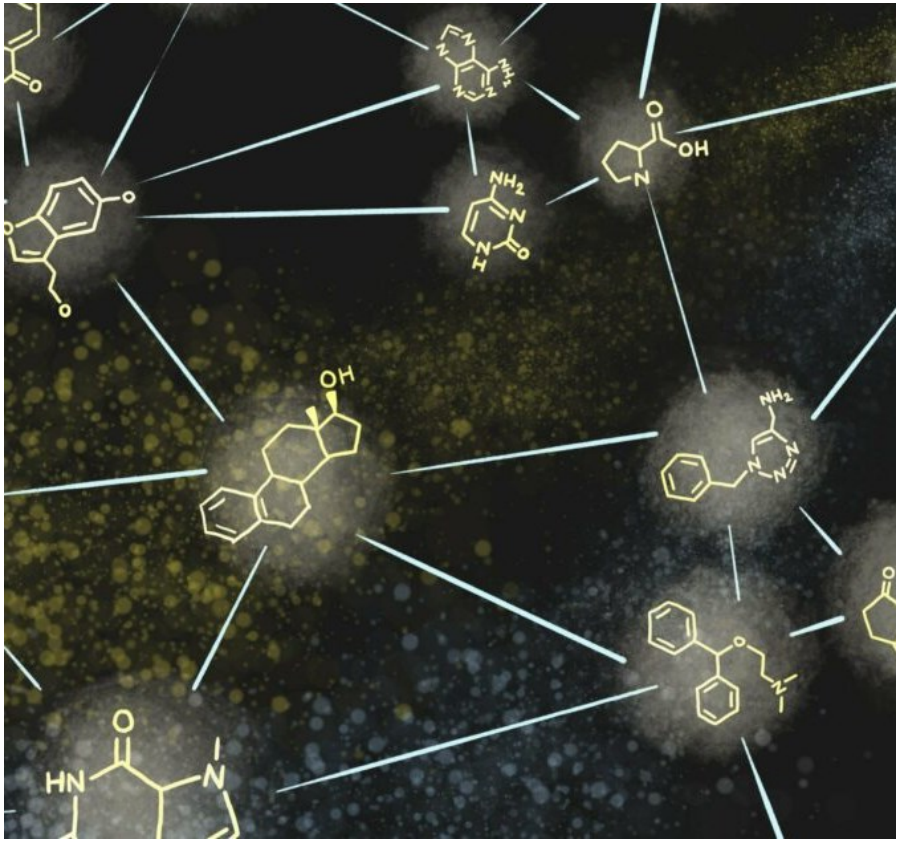
Source: [2]

These technologies are already available!

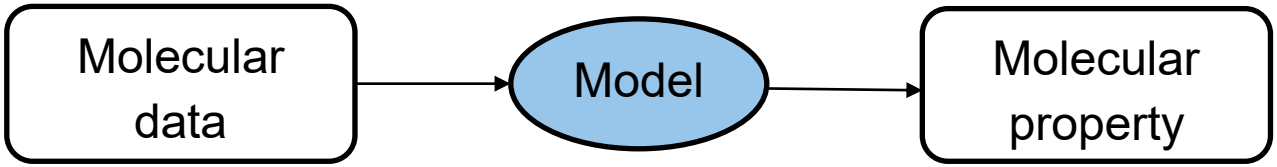
However, they need to be more efficient.

We want to find an organic material with good molecular properties.

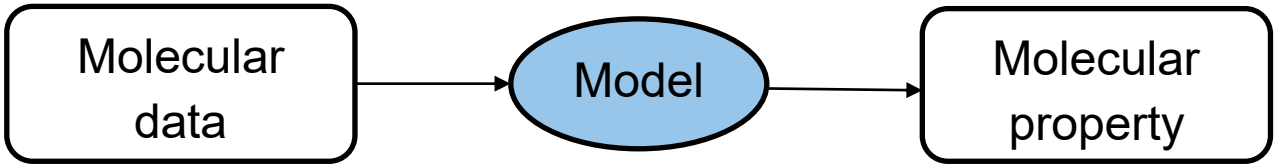
Good = Transfer Electric Charge



How do scientists address this problem?



How do scientists address this problem?

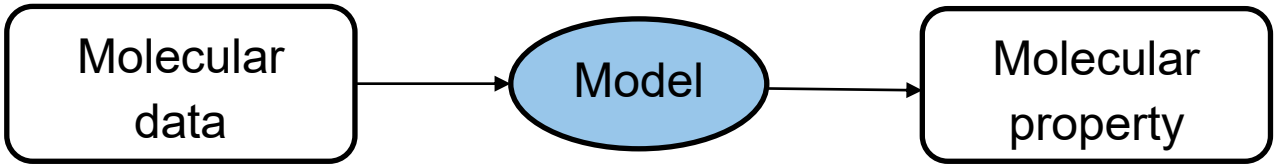


Simulations

Reproduce physics of material.

- Accurate.
- Computationally expensive.

How do scientists address this problem?



Simulations

Reproduce physics of material.

- Accurate.
- Computationally expensive.

Artificial Intelligence (I)

Predict molecular properties given a training dataset.

- Faster results.
- Need ground truth from simulation.

We take it a step further:

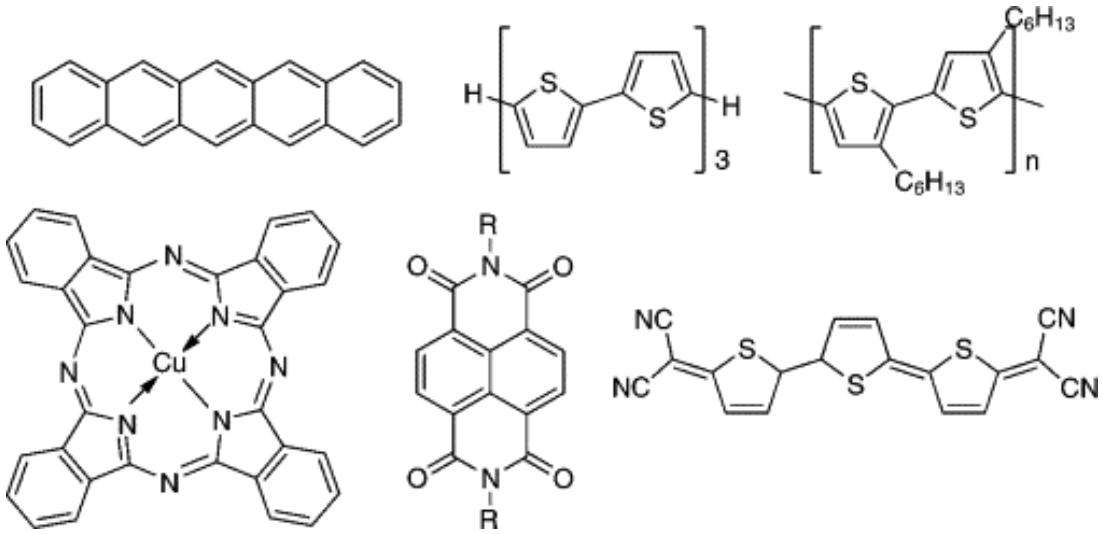
Artificial Intelligence (II)

Transfer the knowledge to new materials.

- Faster results. 
- Don’t need ground truth. 

Why do we want transfer learning?

- It accelerates discovery of innovative materials.
- It allows more creativity in trying novel combinations of materials.
- It enables cost-effective research.



How do we implement transfer learning?

* Source: [3]

How do we implement transfer learning?

1. We study a known Machine Learning technique: Domain Adaptation.

* Source: [3]

How do we implement transfer learning?

1. We study a known Machine Learning technique: Domain Adaptation.

2. We employ this method to predict the properties of new organic semiconductors.

* Source: [3]

How do we implement transfer learning?

1. We study a known Machine Learning technique: Domain Adaptation.

2. We employ this method to predict the properties of new organic semiconductors.

3. We show that this technique outperforms the state of the art*.

* Source: [3]

Table of Contents

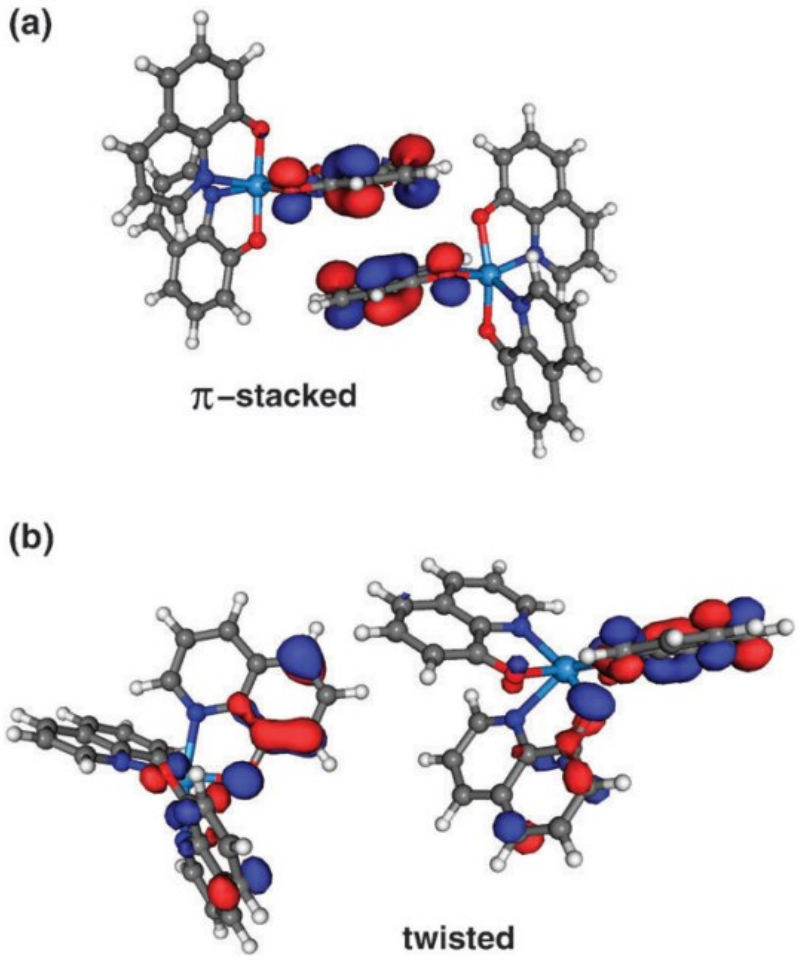
Charge Transfer Integral: the molecular property

Charge Transfer Integral:

Is a quantity that characterizes the electronic coupling between two molecules.

It is a complex property to evaluate.

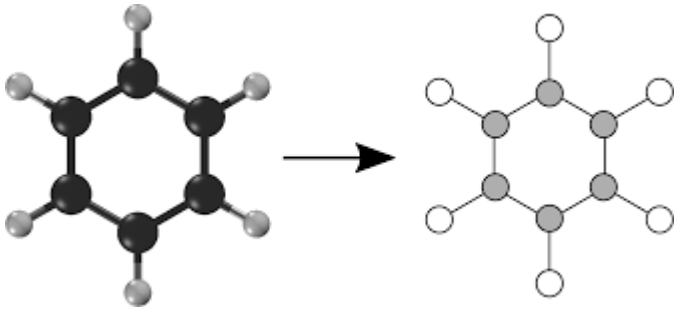
Figure: the orientation of the molecules affects the charge transfer integral.



Source: [4]

Graph Neural Networks for molecular data

Molecules can be encoded as graphs.



Graph Neural Networks (GNN): Local Message Passing Mechanism for Graph-structured data.

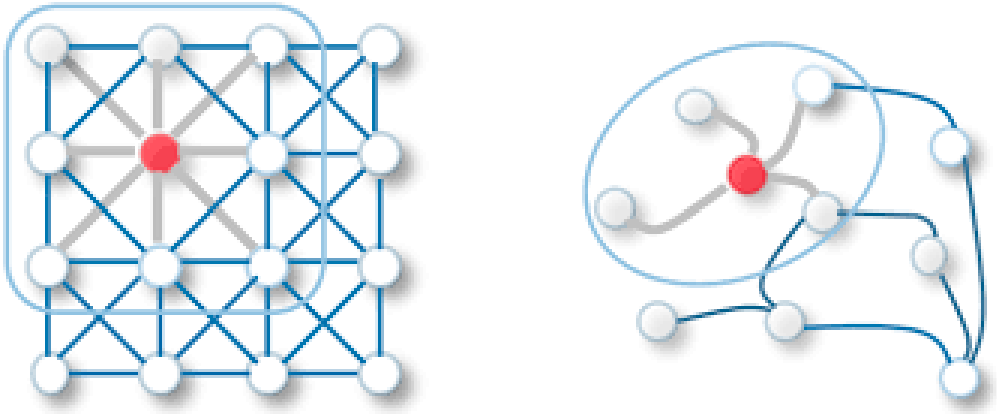


Figure: the receptive field of an image and graph convolution.

Domain Adaptation (DA): training

Train simultaneously two networks.

- Label Predictor: predict charge transfer integral.
- Domain Discriminator: predict the type of molecule.

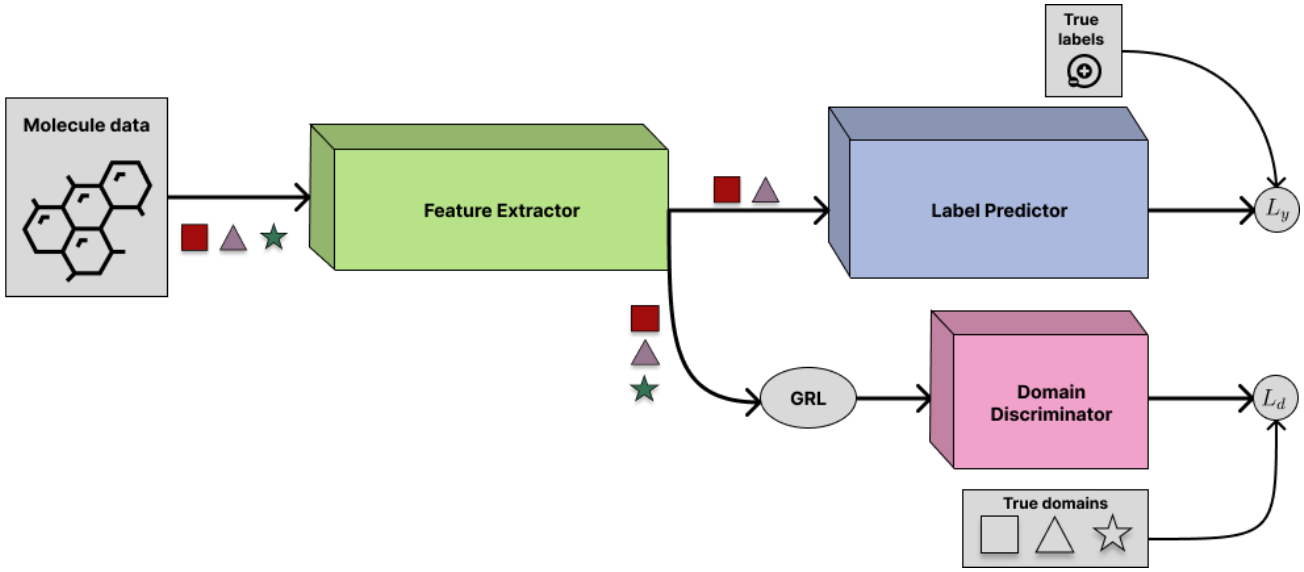


Figure: DA network during training time.

Domain Adaptation (DA): training

Train simultaneously two networks.

- Label Predictor: predict charge transfer integral.
- Domain Discriminator: predict the type of molecule.

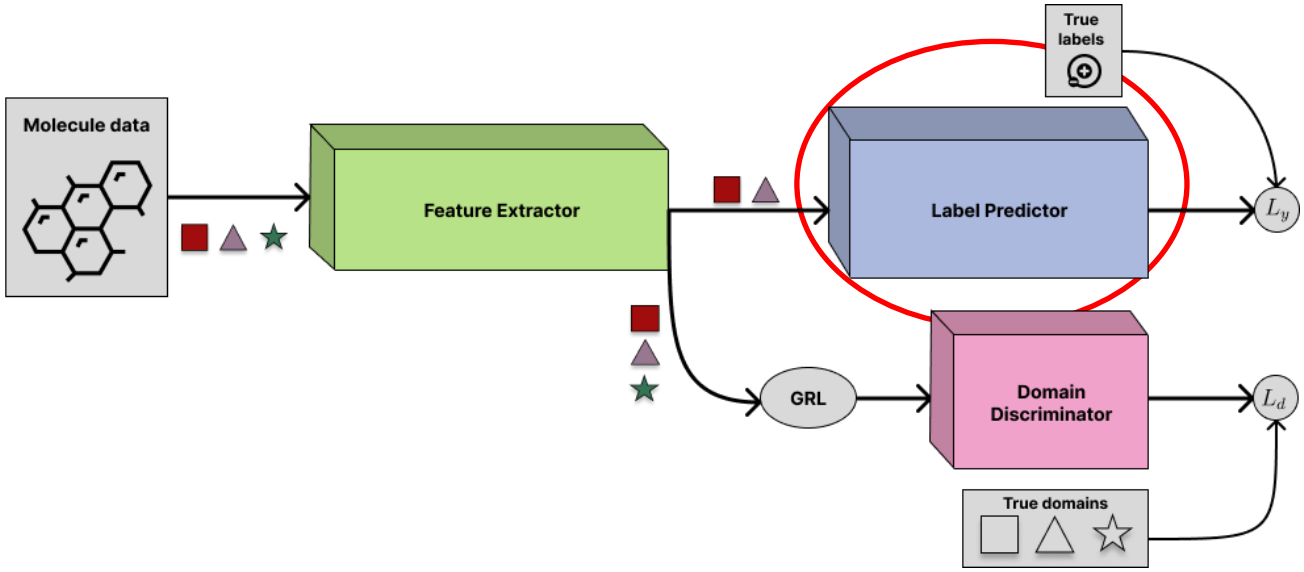


Figure: DA network during training time.

Domain Adaptation (DA): training

Train simultaneously two networks.

- Label Predictor: predict charge transfer integral.
- Domain Discriminator: predict the type of molecule.

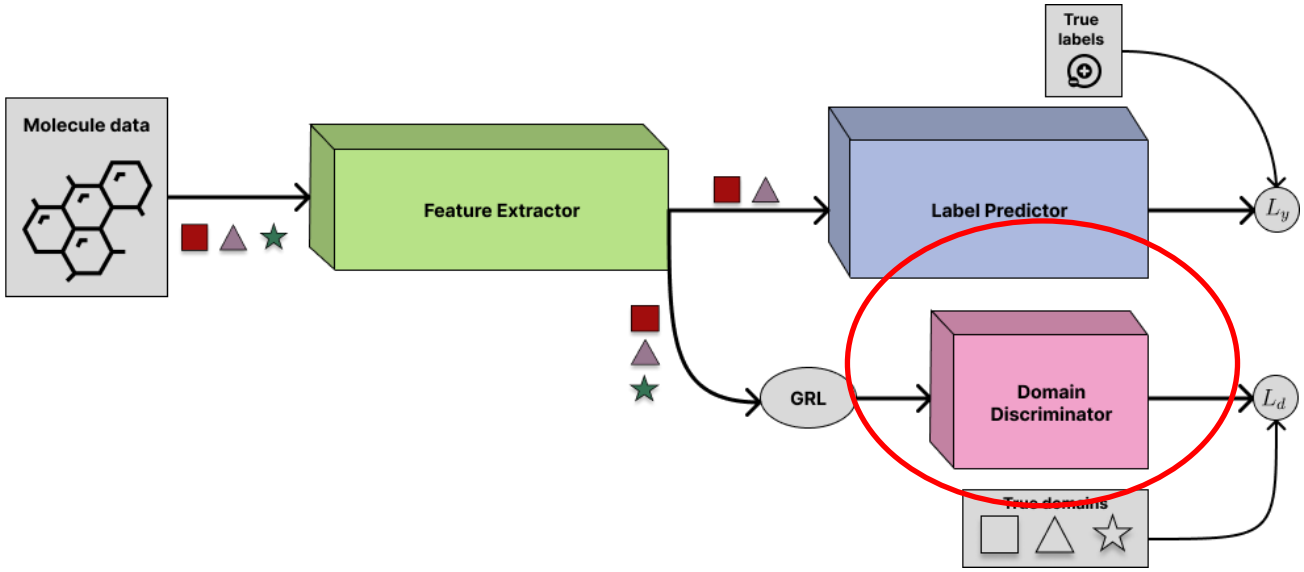


Figure: DA network during training time.

Domain Adaptation (DA): inference

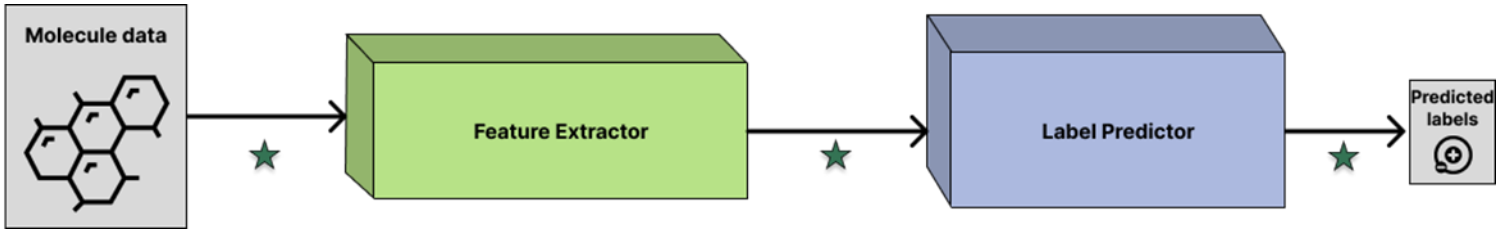


Figure: DA network during inference time.

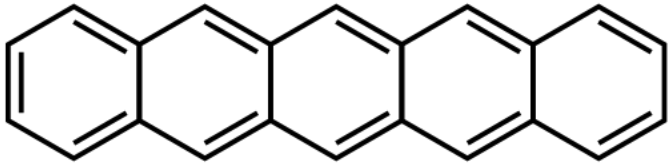
Dataset: Features and molecules

Node features:

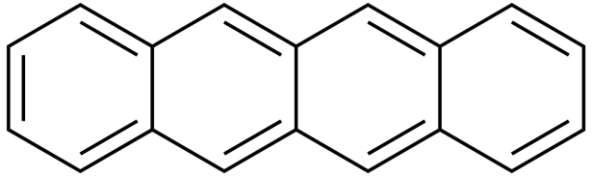
- Atomic number
- Electronegativity
- Molecule belonging (one-hot encoding).
- Atom type (one-hot encoding).

Edge features:

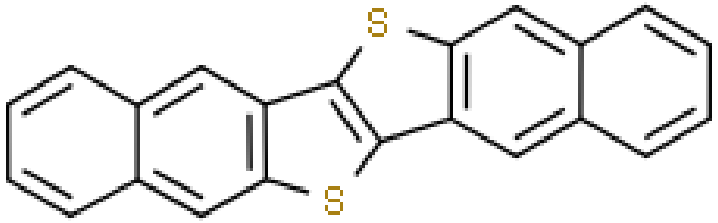
- Distance between atoms.
- Coulomb Matrix (CM).



Pentacene



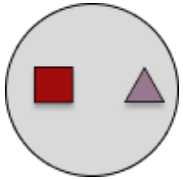
Tetracene



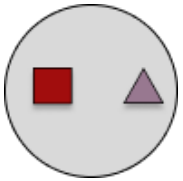
DNTT

Experiments can be in-domain or out-of-domain

In-domain: the test set contains molecules of the same type as the labeled training set.

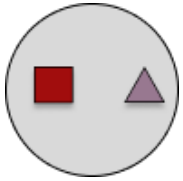


Labeled training set



Test set

Out-of-domain: the test set contains molecules of different type than the labeled training set.



Labeled training set

Test set

Results: In-domain

DA GNN performance does not degrade on in-domain predictions.

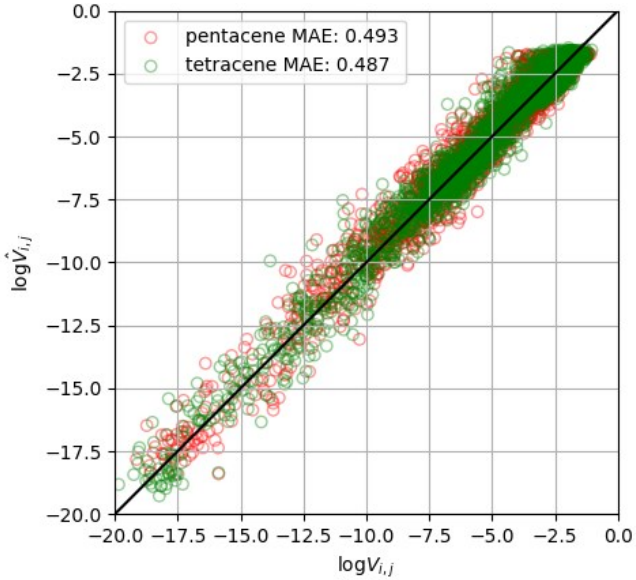


Fig: vanilla GNN results on experiment 2.

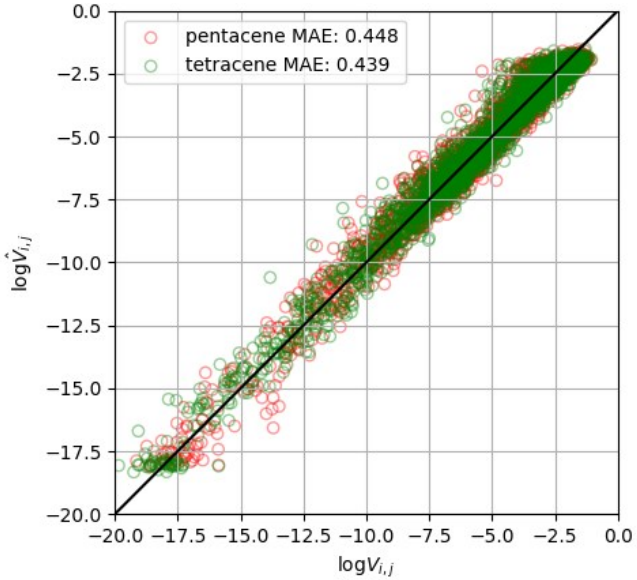


Fig: DA GNN results on experiment 2.

Results: Out-of-domain

DA GNN is able to generalize on different types of domains.

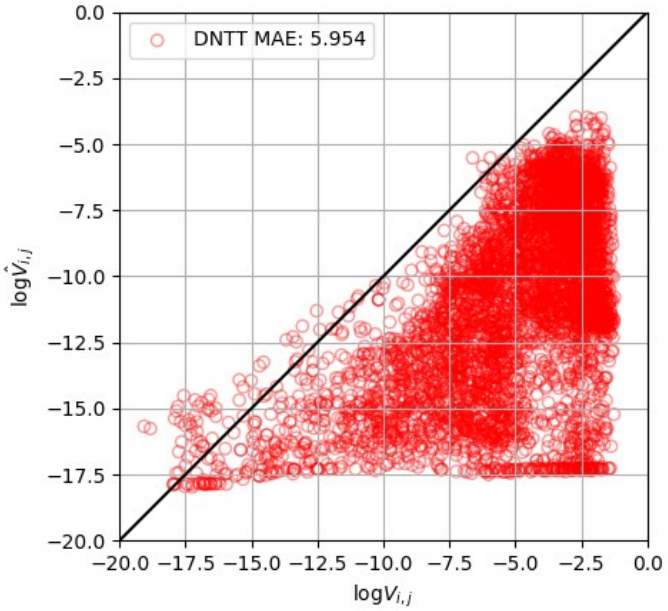


Fig: results of vanilla GNN on experiment 4.

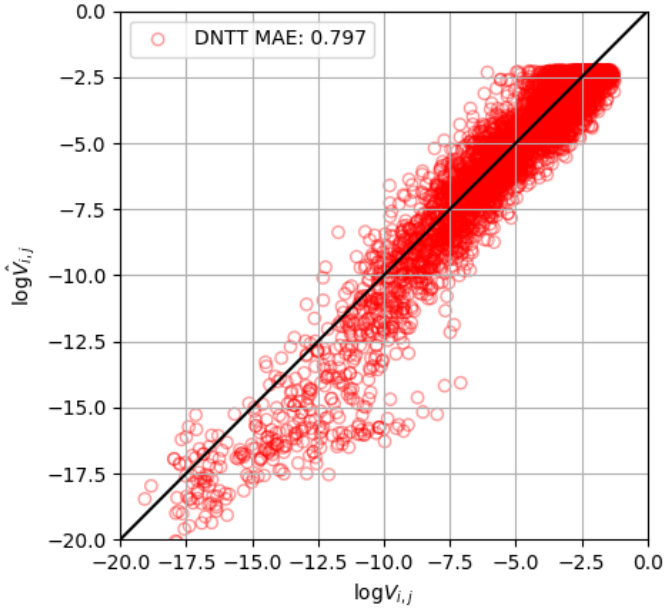


Fig: results of DA GNN on experiment 4.

Results: summary

In-domain experiments: vanilla and DA GNN perform similarly.
-> Domain Adaptation can replace vanilla on regular prediction tasks.

Out-of-domain experiments: DA GNN outperforms vanilla.
-> DA GNN can be an effective transfer learning technique.

Potential weaknesses:

- **In-domain experiments**: With multiple domains, DA GNN is similar, but slightly less accurate than vanilla. How will DA GNN perform when many domains are involved?
- **Out-of-domain experiments:** If the domains are very similar, vanilla can also predict decently.

Conclusion

We have showed that:

Conclusion

We have showed that:

1. Organic Semiconductors have interesting applications that can enhance our lives.

Conclusion

We have showed that:

1. Organic Semiconductors have interesting applications that can enhance our lives.

2. Transfer learning could accelerate the development of innovative materials (no ground truth needed).

Conclusion

We have showed that:

1. Organic Semiconductors have interesting applications that can enhance our lives.

2. Transfer learning could accelerate the development of innovative materials (no ground truth needed).

3. Domain Adaptation is an effective transfer learning technique. We proved it in our case study.

Conclusion

We have showed that:

1. Organic Semiconductors have interesting applications that can enhance our lives.
2. Transfer learning could accelerate the development of innovative materials (no ground truth needed).
3. Domain Adaptation is an effective transfer learning technique. We proved it in our case study.
4. There are potential pitfalls in the method. We have to be mindful about them.

Improvements

- Increase accuracy: DimeNet [5].
- Test DA on benchmark datasets (QM9) [6].

Sources

[1]: Soft ‘e-skin’ that talks to the brain | Stanford News

[2]: Roh, H., Cunin, C., Samal, S., & Gumyusenge, A. (2022). Towards organic electronics that learn at the body-machine interface: A materials journey. MRS Communications, 12(5), 565-577. https://doi.org/10.1557/s43579-022-00269-3

[3]: Rinderle, M., Kaiser, W., Mattoni, A., & Gagliardi, A. (2020). Machine-Learned Charge Transfer Integrals for Multiscale Simulations in Organic Thin Films. *The Journal of Physical Chemistry C*, *124*(32), 17733–17743. doi:10.1021/acs.jpcc.0c04355

[4]: Baumeier, B., Kirkpatrick, J., & Andrienko, D. (2010). Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies. *Phys. Chem. Chem. Phys.*, *12*, 11103–11113. doi:10.1039/C002337J

[5]: Klicpera, J., Groß, J., & Günnemann, S. (2020). Directional Message Passing for Molecular Graphs. CoRR, abs/2003.03123. Retrieved from https://arxiv.org/abs/2003.03123

Sources:

[6]: Ramakrishnan, R., Dral, P. O., Rupp, M., & von Lilienfeld, O. A. (2014). Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, *1*(1), 140022. doi:10.1038/sdata.2014.22

[7]: Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural Message Passing for Quantum Chemistry. arXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/1704.01212

[8]: Ganin, Y., & Lempitsky, V. (2015). Unsupervised Domain Adaptation by Backpropagation. arXiv [Stat.ML]. Retrieved from http://arxiv.org/abs/1409.7495

Thank you!

Neel Misciasci

München, June 21st 2023

Questions?

Neel Misciasci

München, June 21st 2023

Appendix: Table of results

Experiment number	Test set	Type of test	vanilla GNN MAE	DA GNN MAE
1	penta_5k	In-domain	0.466 ± 0.030	**0.457 ± 0.001**
2	penta_2p5k_tetra_2p5k	In-domain	**0.444 ± 0.011**	0.618 ± 0.110
3	tetra_5k	Out-of-domain	**1.447 ± 0.488**	1.778 ± 0.499
4	DNTT_5k	Out-of-domain	>2e25	**1.88 ± 1.07**

Table: Final results.

GNN: from graph to feature vector

Update the embedding h_v^{t+1} of node v at time $t+1$:

- Message passing phase

$$h_v^{t+1} = U_t \left(h_v^t, \bigoplus_{u \in \mathcal{N}(v)} M_t \left(h_v^t, h_u^t \right) \right)$$

- Readout phase (or *pooling* phase)

$$\hat{y} = R(\{h^T v \mid v \in G\})$$

GNN: the message passing phase broken down

1. Compute and aggregate the messages in m_v^{t+1} .
2. Update the node embedding in h_v^{t+1}.

$$m_v^{t+1} = \sum_{w \in N(v)} M_t \left(h_v^t, h_w^t, e_{vw} \right)$$

$$h_v^{t+1} = U_t \left(h_v^t, m_v^{t+1} \right)$$

MPNN for prediction of molecular properties

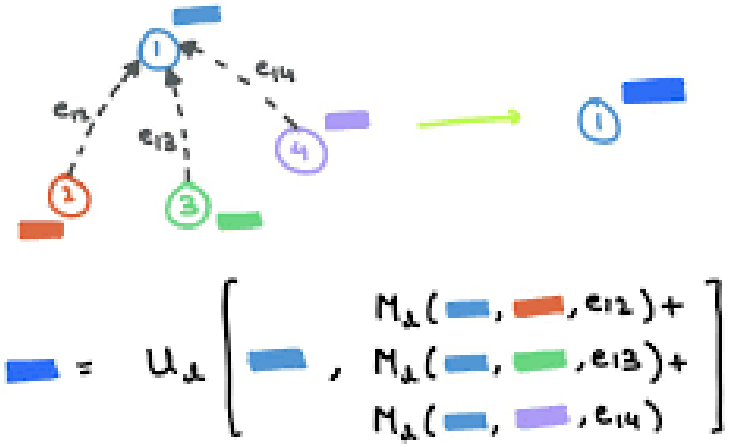
$$m_v^{t+1} = \sum_{u \in \mathcal{N}(v)} h_u^t NN_\Theta(e_{v,u})$$

$$h_v^{t+1} = \Theta h_v^t + m_v^{t+1}$$

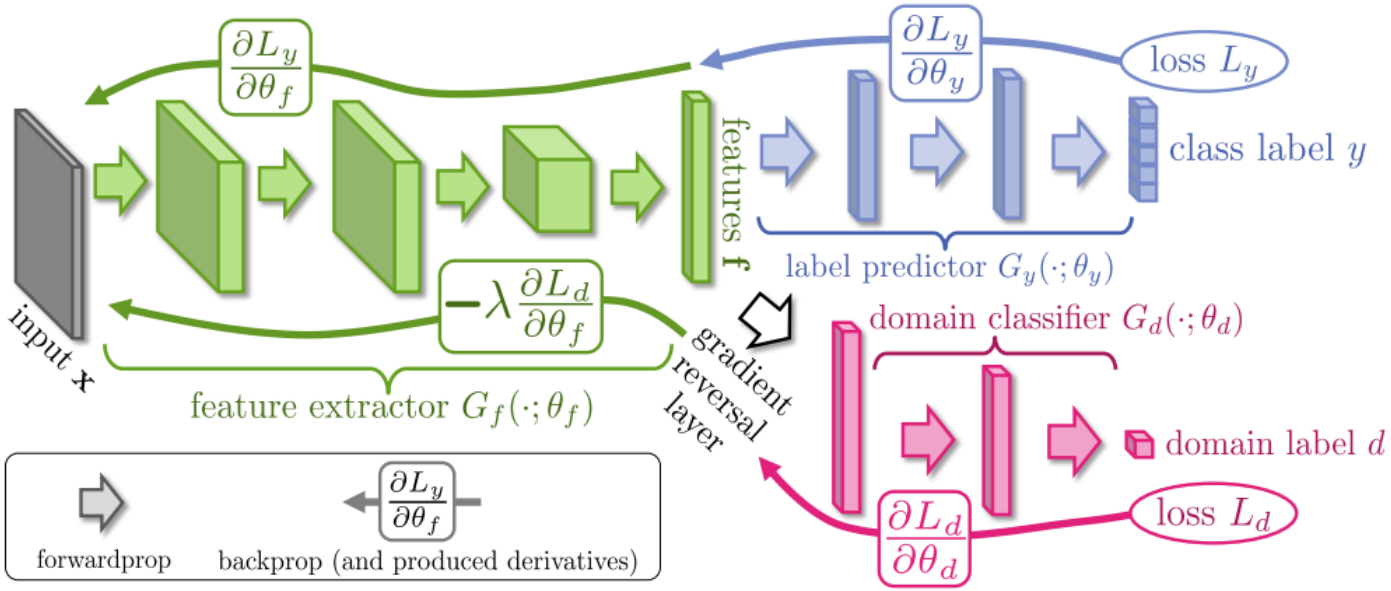
where $h_v^{t+1} \in \mathbb{R}^{F_{out}}$, $h_v^t \in \mathbb{R}^{F_{in}}$, are the input and output node embedding for node v at time t and $t+1$, $\Theta \in \mathbb{R}^{F_{out} \times F_{in}}$ and $NN_\Theta : D_{in} \rightarrow \mathbb{R}^{F_{in} \times F_{out}}$ are respectively a learnable matrix and a neural network mapping to a matrix.
The dimensions F_{out}, F_{in}, D_{in} represent respectively the dimensions of the incoming hidden message, the updated hidden message, and the edge features dimensions.

Source: [7]

Intuition behind the message passing-scheme



Domain Adaptation: a closer look



Source: [8]

Domain Adaptation: a closer look

We want the features from the Feature Extractor to be:

1. Discriminative for the task at hand.

2. Domain-invariant.

We achieve this by optimizing in two directions.

Domain Adaptation: a mathematical perspective

G_f is the Feature Extractor (maps input x to D dimensional feature).

G_y is the Label Predictor (maps feature $\boldsymbol{f}$ to target output).

G_d is the Domain Discriminator (maps feature $\boldsymbol{f}$ to domain label).

$$\mathbf{f} = G_f(\mathbf{x}; \theta_f)$$
$$\hat{y} = G_y(\mathbf{f}; , \theta_y)$$
$$\hat{d} = G_d(\mathbf{f}; \theta_d)$$

Domain Adaptation: a mathematical perspective

We want to **find the min-max saddle point** of the following functional, which combines the losses of Label Predictor L_y and of Domain Discriminator L_d:

$$E(\theta_f, \theta_y, \theta_d) = \sum_{\substack{i=1..N \\ d_i=0}} L_y\left(G_y\left(G_f\left(\mathbf{x}_i; \theta_f\right); \theta_y\right), y_i\right) - \lambda \sum_{i=1..N} L_d\left(G_d\left(G_f\left(\mathbf{x}_i; \theta_f\right); \theta_d\right), y_i\right)$$

$$= \sum_{\substack{i=1..N \\ d_i=0}} L_y^i\left(\theta_f, \theta_y\right) - \lambda \sum_{i=1\ldots N} L_d^i\left(\theta_f, \theta_d\right)$$

I.e. find the points:

$$\left(\hat{\theta}_f, \hat{\theta}_y\right) = \arg\min_{\theta_f, \theta_y} E\left(\theta_f, \theta_y, \hat{\theta}_d\right)$$

$$\hat{\theta}_d = \arg\max_{\theta_d} E\left(\hat{\theta}_f, \hat{\theta}_y, \theta_d\right).$$

Domain Adaptation implementation

We cannot directly use a Stochastic Gradient Descent (SGD) solver for a minmax problem.

Inject a Gradient Reversal Layer R_λ:

$$R_\lambda(\mathbf{x}) = \mathbf{x}$$
$$\frac{dR_\lambda}{d\mathbf{x}} = -\lambda\mathbf{I}$$

Domain Adaptation implementation

Thanks to the GRL we can follow a SGD procedure to solve the minmax problem. The updates become:

$$\theta_f \longleftarrow \theta_f - \mu \left(\frac{\partial L_y^i}{\partial \theta_f} - \lambda \frac{\partial L_d^i}{\partial \theta_f} \right)$$

$$\theta_y \longleftarrow \theta_y - \mu \frac{\partial L_y^i}{\partial \theta_y}$$

$$\theta_d \longleftarrow \theta_d - \mu \frac{\partial L_d^i}{\partial \theta_d}$$

This is equivalent to **minimizing** the functional with the injected GRL:

$$\tilde{E}\left(\theta_f, \theta_y, \theta_d\right) = \sum_{\substack{i=1....N \\ d_i=0}} L_y\left(G_y\left(G_f\left(\mathbf{x}_i; \theta_f\right); \theta_y\right), y_i\right) +$$

$$\sum_{i=1..N} L_d\left(G_d\left(R_\lambda\left(G_f\left(\mathbf{x}_i; \theta_f\right)\right); \theta_d\right), y_i\right)$$

The Coulomb Matrix (CM) encodes the Coulombian interaction between atoms

The CM is a symmetric matrix of size $N_{atoms} \times N_{atoms}$ where N_{atoms} is the number of atoms in the molecule.

$$C_{ij} = \begin{cases} \frac{Z_i \cdot Z_j}{|\mathbf{r}_i - \mathbf{r}_j|} & \text{if } i \neq j \\ \frac{1}{2} Z_i^{2.4} & \text{if } i = j \end{cases}$$

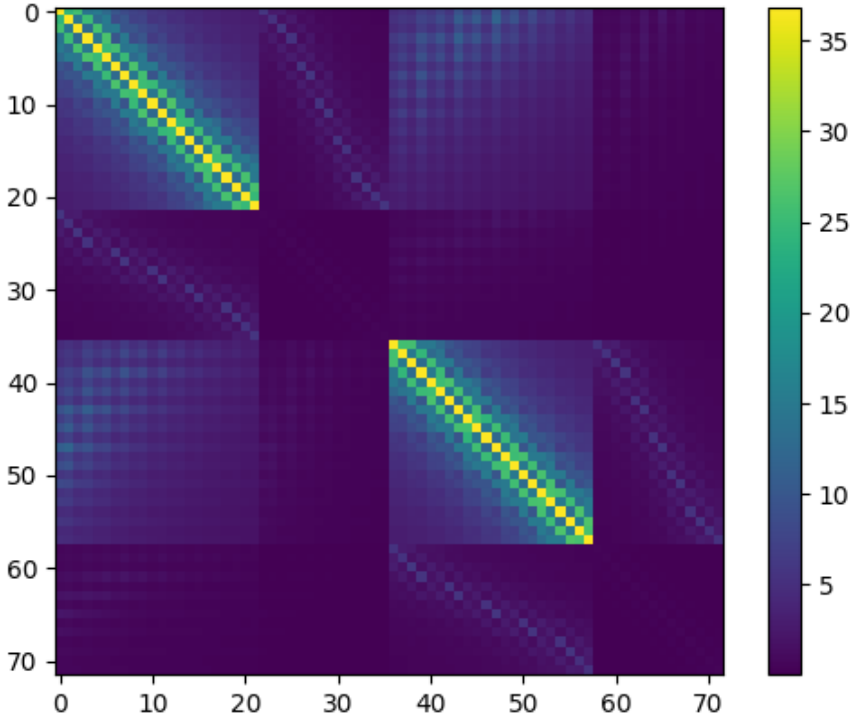
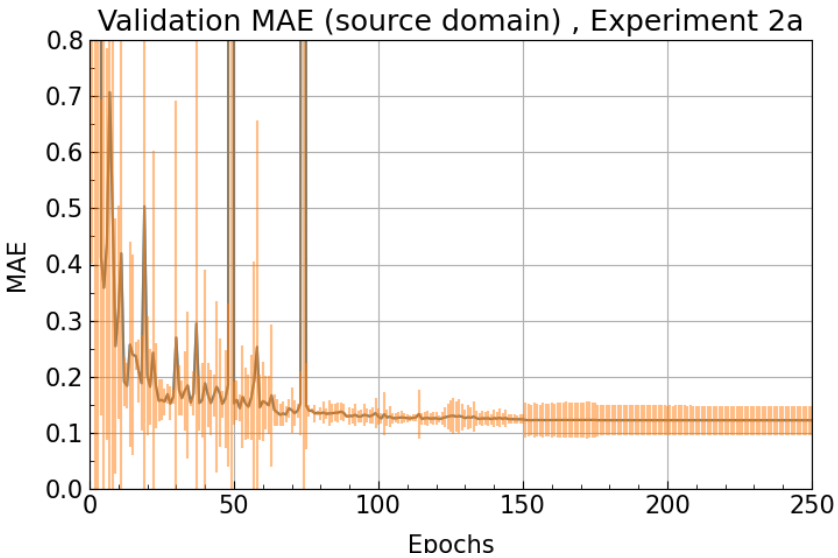
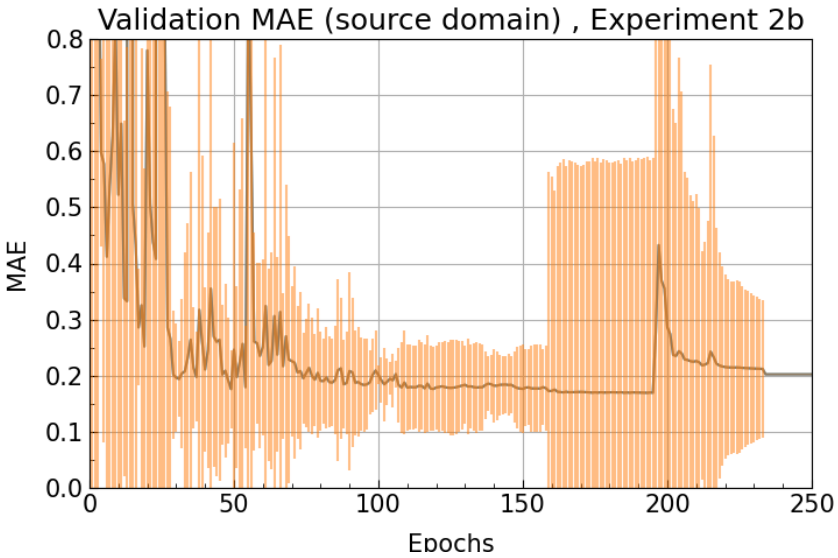


Figure: CM of a pentacene dimer.

Validation MAE

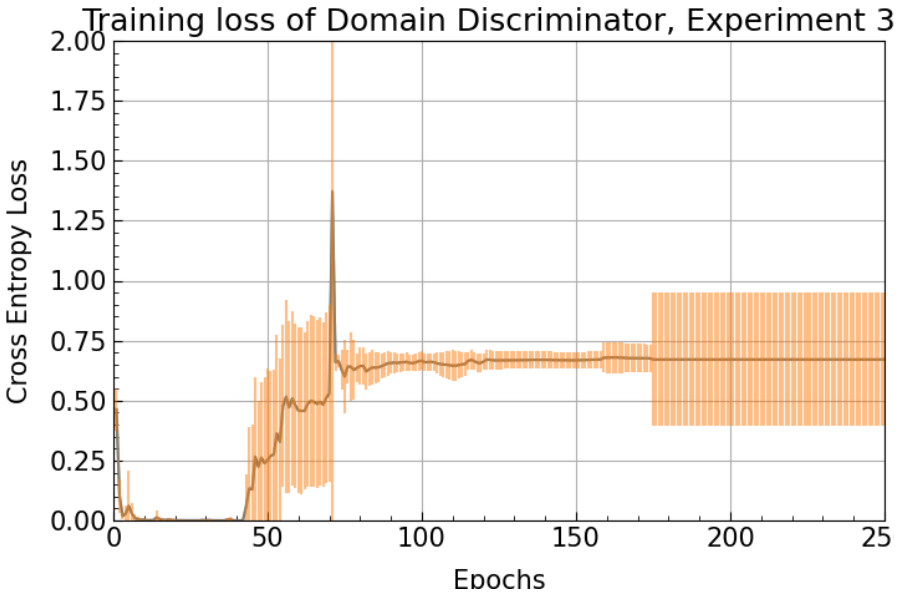


Vanilla GNN

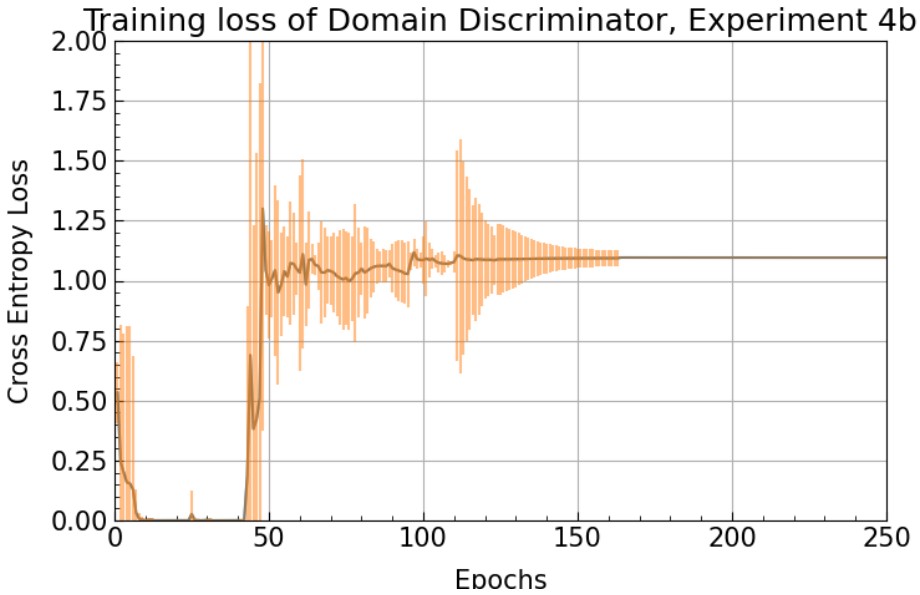


DA GNN

Domain Discriminator Loss



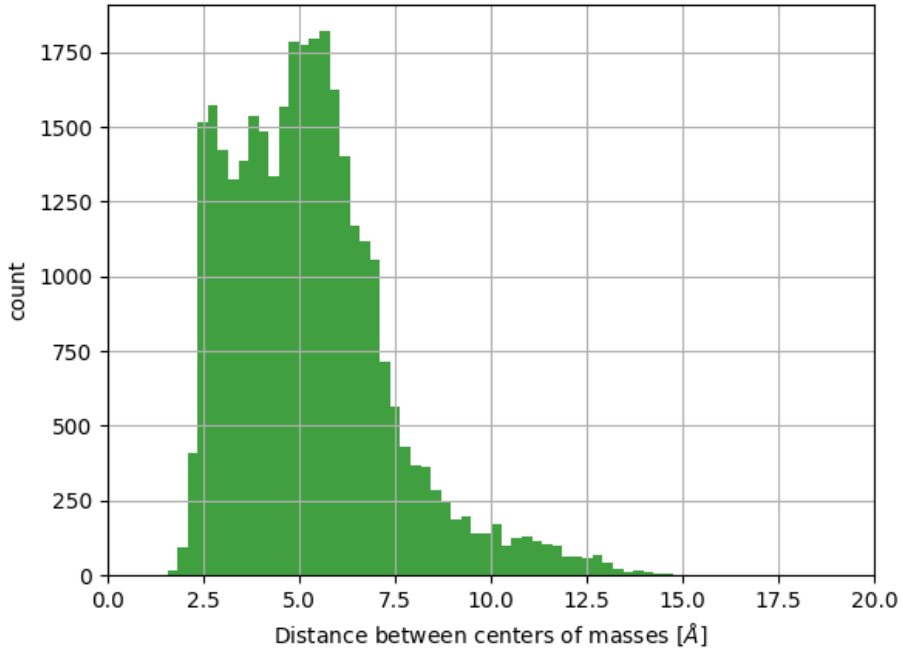
DA GNN

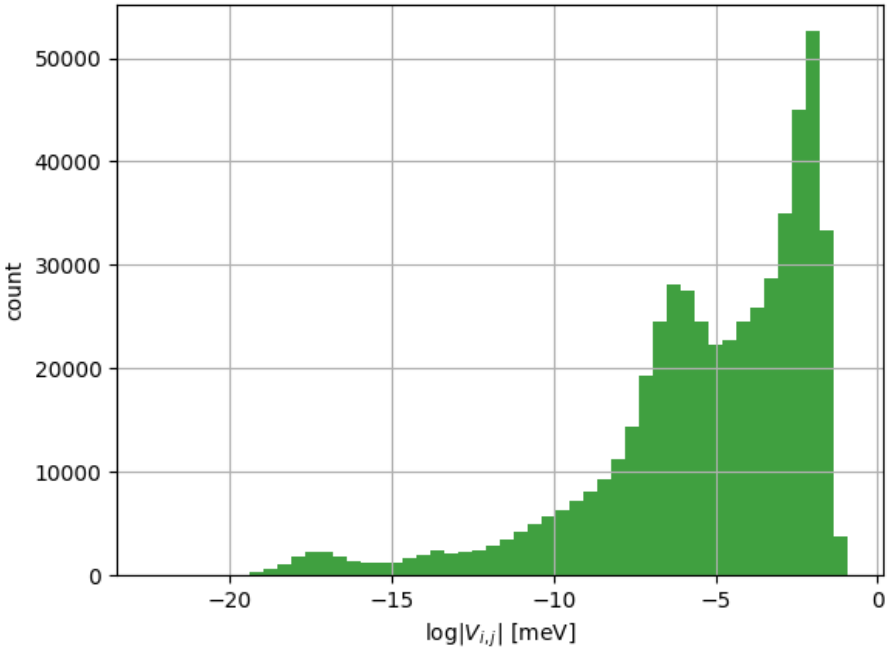


DA GNN

The GRL is activated slowly and steadily from epochs 22 to 55.
Result: increase in loss.

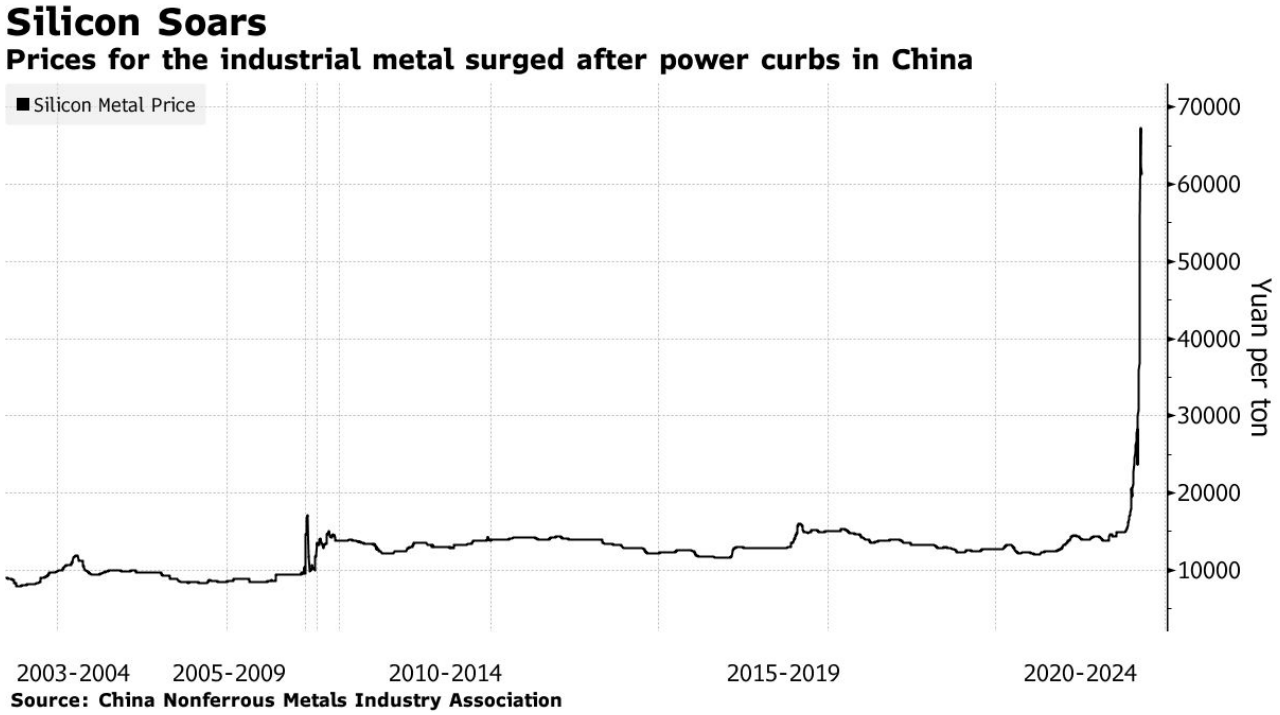
Distribution of targets and of center of masses





Molecule represented: pentacene.

The prices of silicon are rising:



Markets

Silicon's 300% Surge Throws Another Price Shock at the World

Source: Bloomberg