Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supplementary of The Catalysis Deep Neural Networks (Cat-DNNs) in Singlet Fission Property Prediction

Shuqian Ye, Jiechun Liang, Xi Zhu*
School of Science and Engineering (SSE), Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS)
The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)
14-15F, Tower G2, Xinghe World, Rd Yabao, Longgang District, Shenzhen, Guangdong, China, 518172
Email: zhuxi@cuhk.edu.cn

S1. Singlet Fission

For a generalized excited state with occupied orbitals and unoccupied orbitals, we assume an electron-hole pair is created at sites r_e and r_h with spins σ_e and σ_h, respectively. The excited state wavefunction Φ can be expanded as the summation of ground state Φ_0 over sites:

$$\Phi = \sum_{r_e, r_h} C(r_e, r_h) \sum_{\sigma_e, \sigma_h} a^-(\sigma_e) a^+(\sigma_h) \Phi_0(r) \#(S1)$$

where $C(r_e, r_h) = N^{-1} \exp(i(k_e \cdot r_e + k_h \cdot r_h))$ is the probability amplitude coefficient, $a^-(\sigma_e)$ and $a^+(\sigma_h)$ are annihilation and creation operators of electrons and holes. The energy eigenvalue for this wavefunction is obtained through Schrodinger equation: $E = \frac{\langle \Phi | H | \Phi \rangle}{\langle \Phi | \Phi \rangle}$. Expand the energy function about $k = 0$ and apply the effective mass approximation for both charge carriers, the expression of eigenvalue E reads:

$$E = E_0 + E_g + \frac{p_e^2}{2m_e^*} + \frac{p_h^2}{2m_h^*} - \frac{1}{N^2} \sum_{r_e, r_h} CE_{eh} \#(S2)$$

E_0 is the ground state energy, E_g is the optical band gap for bulk materials, and p_i & m_i^* ($i = e, h$) denote the momenta and effective mass for electrons and holes, respectively. The last term E_{eh} represents the interaction between the excited electron hole pair. With two-center approximations, the interaction can be described with two types of energy integrals related to spin S: coulomb integral J_{eh} as well as exchange integral K_{eh}[1, 2]:

$$E_{eh} = J_{eh} - (1 - S) K_{eh} \#(S3)$$

$$J_{eh} = \iint |\Phi_0(r_1 - r_e)|^2 \frac{\kappa e^2}{|r_1 - r_2|} |\Phi_0(r_2 - r_h)|^2 \, d^3 r_1 d^3 r_2 \#(S4)$$

$$K_{eh} = \iint |\Phi_0(r_1 - r_e)| |\Phi_0(r_2 - r_e)| \frac{\kappa e^2}{|r_1 - r_2|} |\Phi_0(r_1 - r_h)| |\Phi_0(r_2 - r_h)| \, d^3 r_1 d^3 r_2 \#(S5)$$

Here we change the variable $r_1' = r_1 - r_e$, $r_2' = r_2 - r_h$, considering that $\frac{|r_1' - r_2'|}{|r_e - r_h|} \approx 1$ and using Haken's approximation applied for Wannier excitons that

$\int |\Phi_0(r_1')|^2 d^3r_1' = \int |\Phi_0(r_2')|^2 d^3r_2' \approx \frac{1}{\sqrt{\varepsilon}}$ where ε is the static dielectric constant of the solid[3], J_{eh} can be simplified as: $J_{eh} \approx \frac{\kappa e^2}{\varepsilon|r_e - r_h|}$. Unlike J_{eh}, there is a lack of analytical simplification for the exchange term K_{eh}. However, from a qualitative analysis, it makes sense that the exchange should also have a negative scaling with respect to the distance between electron and holes. For Wannier excitons we can approximate K_{eh} using similar structure of J_{eh} by adding a ratio $\alpha(r_e,r_h)$: $K_{eh} \approx \frac{\kappa e^2}{\alpha\varepsilon|r_e - r_h|_2}$. Combining these interaction terms and applying hydrogen atom model for electrons and holes by using center of mass R, relative coordinate r_{eh} and reduced mass μ, the last three terms in Eq (S2) can be considered as energy operator of an electron- hole pair. Since the ground state energy as well as bandgap for bulk materials can be treated as a constant, we can reformulate the wavefunction and Hamiltonian applying hydrogen model:

$$\Phi(r_e,r_h) = \Phi(R,r_{eh}) = \frac{1}{\sqrt{V}}\exp\left(i\frac{P\cdot R}{\hbar}\right)\phi_n(r_{eh}) \#(S6)$$

$$H = E_0 + E_g + \frac{P^2}{2M} - \frac{\hbar^2\nabla^2}{2\mu} - \frac{\kappa e^2}{\varepsilon'(S)r_{eh}} \#(S7)$$

where $\varepsilon'(S) = \varepsilon[1 - \frac{(1-S)}{\alpha}]^{-1}$ follows from Eq (S3). The formulation of wavefunction and Hamiltonian naturally separates the variable, hence the we can solve the spin related energy eigenvalue of hydrogenic energy state for electron-hole pairs from separated Schrodinger equation:

$$\left[-\frac{\hbar^2\nabla^2}{2\mu} - \frac{\kappa e^2}{\varepsilon'(S)r_{eh}}\right]\phi_n(r_{eh}) = E_n(S)\phi_n(r_{eh}) \#(S8)$$

Giving eigenvalue as:

$$E_n(S) = -\frac{\mu e^4\kappa^2}{2\hbar^2\varepsilon'(S)^2 n^2}\frac{1}{} \#(S9)$$

S2. Physical error detail

For orbit-related properties, the physical error should consider 3 parts: the mean absolute error, i.e., the L1 distance from prediction value to ground-truth value; the L1 distance between of

predicted band gap $\hat{\varepsilon}_{gap}$ and $\hat{\varepsilon}_{HOMO} - \hat{\varepsilon}_{LUMO}$; the punishment for non-physical reliable prediction, i.e., the predicted $\hat{\varepsilon}_{LUMO}$ is larger than $\hat{\varepsilon}_{HOMO}$. The punishment is the L1 distance from the average band gap of all molecules $\bar{\varepsilon}_{gap}$ to the ground truth band gap ε_{gap} of this molecule. The physical error can be written as:

$$err(\varepsilon_{gap}) = \left|\hat{\varepsilon}_{gap} - \varepsilon_{gap}\right| + \left|\hat{\varepsilon}_{gap} - \left(\hat{\varepsilon}_{LUMO} - \hat{\varepsilon}_{HOMO}\right)\right| + \frac{1}{2}\left(1 - sgn\left(\hat{\varepsilon}_{LUMO} - \hat{\varepsilon}_{HOMO}\right)\right)\left|\varepsilon_{gap} - \bar{\varepsilon}_{gap}\right|$$

$$err(\varepsilon_{HOMO}) = \left|\hat{\varepsilon}_{HOMO} - \varepsilon_{HOMO}\right| + \left|\hat{\varepsilon}_{gap} - \left(\hat{\varepsilon}_{LUMO} - \hat{\varepsilon}_{HOMO}\right)\right| + \frac{1}{2}\left(1 - sgn\left(\hat{\varepsilon}_{LUMO} - \hat{\varepsilon}_{HOMO}\right)\right)\left|\right.$$

$$err(\varepsilon_{LUMO}) = \left|\hat{\varepsilon}_{LUMO} - \varepsilon_{LUMO}\right| + \left|\hat{\varepsilon}_{gap} - \left(\hat{\varepsilon}_{LUMO} - \hat{\varepsilon}_{HOMO}\right)\right| + \frac{1}{2}\left(1 - sgn\left(\hat{\varepsilon}_{LUMO} - \hat{\varepsilon}_{HOMO}\right)\right)\left|\right.$$

Similarly, the physical error for excited energy properties can be defined. The punishment is given to the prediction that predicted singlet energy is lower than triplet energy.

$$err(E_f) = \left|\hat{E}_f - E_f\right| + \left|\hat{E}_f - \left(\hat{E}_S - 2\hat{E}_T\right)\right| + \left|\hat{\Delta E} - \left(\hat{E}_S - \hat{E}_T\right)\right| + \frac{1}{2}\left(1 - sgn\left(\hat{E}_S - \hat{E}_T\right)\right)\left|E_f - \bar{E}_f\right|$$

$$err(E_S) = \left|\hat{E}_S - E_S\right| + \left|\hat{E}_f - \left(\hat{E}_S - 2\hat{E}_T\right)\right| + \left|\hat{\Delta E} - \left(\hat{E}_S - \hat{E}_T\right)\right| + \frac{1}{2}\left(1 - sgn\left(\hat{E}_S - \hat{E}_T\right)\right)\left|E_S - \bar{E}_S\right|$$

$$err(E_{fT}) = \left|\hat{E}_T - E_T\right| + \left|\hat{E}_f - \left(\hat{E}_S - 2\hat{E}_T\right)\right| + \left|\hat{\Delta E} - \left(\hat{E}_S - \hat{E}_T\right)\right| + \frac{1}{2}\left(1 - sgn\left(\hat{E}_S - \hat{E}_T\right)\right)\left|E_T - \bar{E}_T\right|$$

$$err(\Delta E) = \left|\hat{\Delta E} - \Delta E\right| + \left|\hat{E}_f - \left(\hat{E}_S - 2\hat{E}_T\right)\right| + \left|\hat{\Delta E} - \left(\hat{E}_S - \hat{E}_T\right)\right| + \frac{1}{2}\left(1 - sgn\left(\hat{E}_S - \hat{E}_T\right)\right)\left|\Delta E - \bar{\Delta E}\right|$$

Based on the energy constraint mentioned above, we can define the physical error of the error as the L1 distance between predicted energy and ground truth value plus the punishment if the predicted values violate the physical constraint.

Physical Error function for internal energy U:

$$err(U) = \left|\hat{U} - U\right| + \frac{1}{2}\left(1 - sgn\left(\hat{H} - \hat{U}\right)\right)\left|U - \bar{U}\right|$$

Physical Error function for internal energy at 0K U_0:

$$err(U_0) = \left|\hat{U}_0 - U_0\right| + \frac{1}{2}\left(1 - sgn\left(\hat{U} - \hat{U}_0\right)\right)\left|U_0 - \bar{U}_0\right|$$

Physical Error function for free energy G:

$$err(G) = |\hat{G} - G| + \frac{1}{2}(1 - sgn(\hat{H} - \hat{G}))|G - \bar{G}|$$

S3. Ratio of non-physical reliable prediction.

Table S1 shows the percentage to non-physical reliable prediction, i.e., the prediction violates the physical correlations in the paper. "Y" represents the model with catalyst, while "N" represents models without catalyst. The column "U" represents the prediction violates the physical correlation for internal energy U introduced in the paper, the column "orbit" represents the prediction violates the physical correlation for orbital energy, the column "excited" represents the prediction violates the physical correlation for excited-state energy, and the column "all" represents the prediction violates any one or more physical correlations we introduced.

Table S1: The ratio of non-physical reliable prediction from various model. (Better is in **bold**).

Model	Catalyst	U	U_0	G	orbit	excited	all
GCN[4]	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	46.79%	43.75%	0.01%	**0.00%**	0.23%	90.55%
GAT[5]	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	44.92%	45.17%	**0.00%**	**0.00%**	0.15%	89.93%
GIN[6]	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	46.98%	41.41%	0.08%	**0.00%**	0.13%	88.33%
ChebyNet[7]	Y	**0.00%**	**0.00%**	**0.00%**	0.16%	**0.00%**	**0.16%**
	N	22.31%	68.33%	9.84%	**0.00%**	**0.00%**	87.68%
GGNN[8]	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	45.59%	44.27%	0.49%	**0.00%**	0.17%	89.80%
MPNN[9]	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	44.36%	42.04%	0.01%	**0.00%**	0.22%	86.30%
RGCN[10]	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	42.45%	42.78%	0.01%	**0.00%**	0.26%	85.22%
SchNet[11]	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	55.26%	35.35%	5.53%	**0.00%**	0.22%	84.10%

S4. Deep learning details

Besides we used above, we apply the catalyst to 8 GNN models, including graph convolutional networks[4] (GCN), graph attention network[5] (GAT), graph isomorphism net[6] (GIN), ChebyNet[7], gated graph neural network[8] (GGNN), message passing neural network[9] (MPNN), relational graph convolutional network[10] (RGCN), and SchNet[11]. The result is shown in Table S2.

The code is available at https://github.com/yippp/CatDNNs.

Table S2: The physical error for various properties with/without catalyst. In column "Catalyst", "N" represents no catalyst endorsed, while "Y" represents catalyst endorsed. Better is in **bold**. All Properties with unit eV.

Model	Catalyst	U	U_0	H	G	ε_{gap}	ε_{HOMO}	ε_{LUMO}	E_S	E_T	ΔE	E_f
GCN	Y	**0.072**	**0.072**	**0.072**	**0.074**	**0.002**	0.001	**0.002**	**0.069**	**0.057**	**0.045**	**0.079**
	N	2.850	0.118	0.088	0.087	0.002	**0.001**	0.002	0.076	0.069	0.054	0.092
GAT	Y	**0.066**	**0.066**	**0.066**	**0.068**	0.003	**0.002**	0.003	0.091	**0.065**	0.064	0.100
	N	2.751	0.124	0.086	0.088	**0.002**	0.002	**0.002**	**0.081**	0.070	**0.057**	**0.095**
GIN	Y	**0.068**	**0.070**	**0.068**	**0.071**	**0.003**	**0.002**	**0.003**	0.090	**0.070**	**0.062**	0.103
	N	2.839	0.118	0.087	0.085	0.003	0.002	0.002	**0.087**	0.074	0.062	**0.100**
ChebyNet	Y	**0.098**	**0.109**	0.098	0.104	0.012	0.011	0.010	**0.278**	**0.169**	**0.212**	**0.269**
	N	1.525	0.163	**0.086**	**0.102**	**0.010**	**0.010**	**0.009**	0.418	0.313	0.387	0.410
GGNN	Y	**0.065**	**0.065**	**0.065**	**0.066**	**0.004**	**0.003**	**0.004**	**0.105**	**0.075**	**0.075**	**0.117**
	N	2.813	0.116	0.080	0.088	0.004	0.004	0.004	0.129	0.098	0.099	0.146
MPNN	Y	**0.072**	**0.076**	**0.072**	**0.074**	**0.002**	**0.002**	**0.002**	**0.054**	**0.026**	**0.050**	**0.060**
	N	2.614	0.119	0.091	0.094	0.002	0.002	0.002	0.061	0.035	0.060	0.074
RGCN	Y	**0.058**	**0.059**	**0.057**	**0.061**	0.003	0.003	0.003	0.079	0.040	0.068	0.082
	N	2.490	0.113	0.081	0.080	**0.001**	**0.001**	**0.001**	**0.035**	**0.025**	**0.036**	**0.047**
SchNet	Y	**0.075**	**0.079**	**0.075**	**0.081**	0.003	0.003	0.003	0.086	0.053	0.081	0.111
	N	3.235	0.125	0.075	0.084	**0.001**	**0.001**	**0.001**	**0.031**	**0.025**	**0.032**	**0.043**

We use machine learning framework PyTorch[12] 1.8, geometric deep learning extension library PyTorch Geometric[13] 1.7, and wrapper PyTorch Lightning[14] to build all models. All the experiments are run on the server with graphic card NVIDIA GeForce RTX 2080 Ti (11 GB graphic memory). The operating system is Ubuntu 18.04 LTS, with graphic card driver version 440.82, and CUDA version 10.2.

We use 90% of the data in QM-symex[15] for training, and the remaining 10% for testing. The dataset is split by random shuffle to ensure the training and testing set distribution is the same. The DFT calculation level of QM-symex is B3LYP/6-31G with Symm=VeryLoose. The number of transition states is 10, which means Nstates=10 in Gaussian09.

We use multitask version of Adam optimizer[16] based on stochastic gradient descent to optimize the parameters of the DNN model. The main idea of multitask Adam optimizer[17] is that for each predicted property, maintain a group of parameters, and switch to the corresponding group of

parameters when optimizing one of the properties. During optimization, the loss for only one property is back-forwarding during each optimization iteration.

For all models, the initial learning rate is set to 0.001, the batch size is set to 32 so that all models can be trained using single graphic card. The learning will multiply 0.8 every 20 batches, and the total training procedure containing 300 epochs.

The detail hyperparameters for each model is listed below:

ChebyNet:

Node_hidden_dim	128
Polynomial_order	5
Num_step_prop	6
Num_step_set2set	6

GAT:

Node_hidden_dim	128
Num_step_prop	6
Num_step_set2set	6

GCN:

Node_hidden_dim	128
Num_step_prop	6
Num_step_set2set	6

GGNN:

Node_hidden_dim	64
Num_step_prop	3

GIN:

Node_hidden_dim	128
Num_step_prop	6
Num_step_set2set	6

MPNN:

Node_hidden_dim	48
Edge_hidden_dim	48
Num_step_message_passsing	3
Num_step_set2set	6

RGCN:

Node_hidden_dim	128
Num_step_prop	6
Num_step_set2set	6

SchNet:

Hidden_channels	128
Num_filters	128
Num_interactions	6
Num_gaussians	50
Cutoff	10.0
Readout	mean

S5. Extension to machine learning model

This catalyst concept also works well on some machine learning methods. To prove this idea, we use Least Absolute Shrinkage and Selection Operator[18] (Lasso) to perform a prediction task on polarizability (α) and electron mass (m_e) and the catalyst in this task is PBE band gap database. The feature used for training includes crystal features from C2DB database[19] and element features. All the used features are presented in Table S3. Figure S1 denotes the process of how the training is improved through catalyst. To show the difference, we use raw features as training data to predict α and m_e directly at first. The coefficient of determination (R^2) of α and m_e predictions are 0.84 and 0.51, the detailed data in Table S4. The catalysis process here is defined as adding a PBE bandgap (E_{gap}) prediction between raw features and α, m_e. Since there are sure to be a negative correlation between α and E_g^2, and a positive correlation between m_e and E_g, we use raw features to predict E_{gap} firstly. The E_{gap} is predicted with R^2=0.95, which is a satisfying result. Next the k and x are fitted with correlation $\alpha \sim k\frac{1}{E_{gap}^2} + p$ and $m_e \sim xE_g + q$ where k, x are coefficients and p, q are minor features that contribute little to the prediction. The result R^2 are increased to 0.91 and 0.64 comparing to previous prediction. This simple attempt strengthens the effect of the physical-endorsed catalyst in the machine learning process.

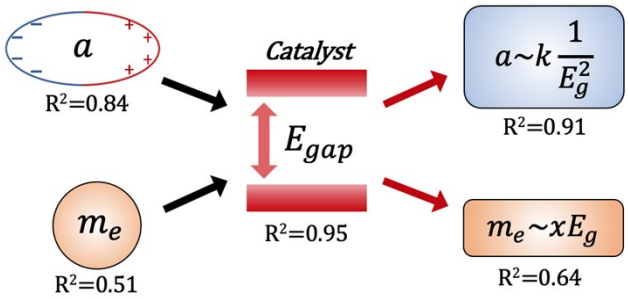

Figure S1: A sketch of catalyst-improved machine learning on polarizability and electron mass prediction. The catalyst in this case is band gap energy.

Table S3: Used features for polarizability and electron mass prediction.

Crystal feature	Element feature
Basis Vector	Ionicity
Heat of Formation	Period Number
Total Energy	Specific Heat
Area	Density
Mass	Curie Point
Charge	Average Covalent Radius
Volume	Electron Affinity
PBE Bandgap	Boiling Point
	Atom Radius
	Absolute Boiling Point
	Absolute Melting Point
	Atom Mass
	Number of Atoms
	Critical Temperature
	Critical Pressure
	Heat of Fusion
	Heat of Vaporization
	Melting Point
	Thermal Conductivity
	Brinell Hardness
	Vickers Hardness
	Bulk Modulus
	Shear Modulus
	Young Modulus
	Electron Negativity
	Number of Outer Electron
	Total Nuclear Charge
	State

Table S4: The prediction performance. (R^2 is higher better, MAE is lower better, best is in **bold**).

Property	R^2	MAE
E_{gap}	0.95	0.21 eV
m_e	0.51	0.19 m_e
m_e (w/ catalyst)	**0.65**	**0.17** m_e
a	0.84	1.08 Ang
a (w/ catalyst)	**0.91**	**0.84 Ang**

S6. Singlet fission prediction result

In Table S5, Y in column “Catalyst” represents the model with catalyst, while N represents models without catalyst. True positive (TP) is the number of predictions that the molecule actually has fission, and the prediction is fission and physical-reliable in all physical correlations. True negative (TN) is the number of predictions that the molecule actually does not have fission,

and the prediction is non-fission and physical-reliable in all physical correlations. False positive (FP) is the number of predictions that the molecule actually does not have fission, and the prediction is fission and physical-reliable in all physical correlations. False negative (FN) is the number of predictions that the molecule actually has fission, and the prediction is non-fission and physical-reliable in all physical correlations. Column "valid" represents the number of valid predictions, which is physical-reliable, i.e., the sum-up of TP, TN, FP, and FN. Note that the total number of samples for testing is 15030.

Precision in Figure 4 is to represent how precise the prediction in all "fission" predictions, i.e., the percentage of true positive (physical reliable) samples in all "fission" predictions (no matter whether the prediction follows physical correlations).

Accuracy in Table S5 measures how accurate all predictions, i.e., the ratio of TP and TN in all predictions.

Table S5: The detailed prediction results for singlet fission. (Better in **bold**)

Model	Catalyst	TP	TN	FP	FN	Valid	Accuracy
GAT	Y	3252	10738	474	566	15030	**93.1%**
	N	253	1302	15	11	1581	10.4%
GCN	Y	3405	10693	519	413	15030	**93.8%**
	N	273	1452	12	17	1754	11.5%
GGNN	Y	3407	10886	326	411	15030	**95.1%**
	N	507	1934	27	44	2512	16.2%
GIN	Y	3390	10595	617	428	15030	**93.1%**
	N	250	1354	14	21	1639	10.7%
MPNN	Y	3639	11055	157	179	15030	97.8%
	N	350	1613	27	24	2014	13.1%
ChebyNet	Y	3425	10769	443	393	15030	**94.4%**
	N	272	1152	11	15	1450	9.5%
RGCN	Y	3555	10949	263	263	15030	**96.5%**
	N	339	1891	14	11	2255	14.8%
SchNet	Y	1642	11100	112	2176	15030	**84.8%**
	N	109	572	2	4	687	4.5%

S7. Results on QM9

The results like QM-symex shown in Section S3 and S6 are shown below.

Table S6: The ratio of non-physical reliable prediction from various models in QM9. (Better is in **bold**).

Model	Catalyst	U	U_0	G	orbit	excited	all
GCN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	78.11%	32.89%	23.86%	**0.00%**	0.02%	92.40%
GAT	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	63.19%	46.33%	46.20%	**0.00%**	**0.00%**	93.67%
GIN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	98.37%	51.40%	87.60%	**0.00%**	**0.00%**	99.40%
ChebyNet	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	70.24%	59.43%	32.17%	**0.00%**	**0.00%**	92.85%
GGNN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	0.42%	5.88%	0.04%	**0.00%**	0.01%	6.27%
MPNN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	75.62%	0.95%	34.32%	**0.00%**	**0.00%**	80.59%
RGCN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	54.97%	59.67%	54.53%	**0.00%**	0.02%	90.32%
SchNet	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	94.08%	19.44%	57.18%	**0.00%**	0.06%	98.77%

Table S7: The detailed prediction results for singlet fission in QM9. (Better in **bold**)

Model	Catalyst	TP	TN	FP	FN	Valid	Accuracy
GAT	Y	39	13136	8	14	13197	**99.8%**
	N	0	1002	0	1	1003	7.6%
GCN	Y	47	13135	9	6	13197	**99.9%**
	N	0	836	0	0	836	6.3%
GGNN	Y	42	13136	8	11	13197	**99.9%**
	N	0	79	0	0	79	0.6%
GIN	Y	37	13136	8	16	13197	**99.8%**
	N	0	943	0	0	943	7.2%
MPNN	Y	44	13136	8	9	13197	**99.9%**
	N	31	12321	7	11	12370	93.6%
ChebyNet	Y	43	13140	4	10	13197	**99.9%**
	N	0	2562	0	0	2562	19.4%
RGCN	Y	45	13137	7	8	13197	**99.9%**
	N	0	1275	2	1	1278	9.7%
SchNet	Y	47	13107	37	6	13197	**99.7%**
	N	0	162	0	0	162	1.2%

S8. Generalization experiment results

The total number of molecules with different structures is 4906 in the Perkinson *et al.* screened SF

molecules dataset[20]. The molecules contain elements H, C, O, N, S, F, Cl, Br, B, P, Si, and the molecule size can reach 150. Since the molecule is larger than QM-symex (no more than 90 atoms), the elements are more various, and the training set (4404 molecules) is much smaller, the prediction difficulty in predicting SF in this dataset is much higher compared with QM-symex and QM9. We use the method as described in QM-symex[15] to calculate all properties needed in the catalyst. The models and training process have the same configuration as in Section S4, except the batch size is set to 16. Since the molecules are much larger, the graphic memory is not enough for some models if the batch size is set to 32. The prediction result for Perkinson *et al.*'s dataset using various models is shown in Table S8, S9, and S10.

Table S8: The physical error for various properties with/without catalyst. In column "Catalyst", "N" represents no catalyst endorsed, while "Y" represents catalyst endorsed. Better is in **bold**.

Model	Catalyst	U	U_0	H	G	ε_{gap}	ε_{HOMO}	ε_{LUMO}	E_S	E_T	ΔE	E_f
		eV	eV	eV	eV	eV	eV	eV	eV	eV	eV	eV
GCN	Y	**79.842**	**79.842**	**79.842**	**79.848**	**0.255**	**0.254**	**0.259**	**0.105**	**0.077**	**0.067**	**0.109**
	N	994.778	697.599	68.457	968.204	0.255	0.254	0.259	0.126	0.102	0.099	0.138
GAT	Y	**70.791**	**70.791**	**70.791**	**70.791**	0.255	0.256	0.260	**0.146**	**0.096**	**0.097**	**0.142**
	N	705.417	832.032	94.017	803.487	**0.254**	**0.253**	**0.258**	0.208	0.164	0.162	0.217
GIN	Y	**97.013**	**97.013**	**97.013**	**97.013**	**0.254**	**0.254**	**0.259**	**0.148**	**0.108**	**0.098**	**0.159**
	N	906.400	816.018	86.746	890.006	0.255	0.255	0.259	0.203	0.157	0.157	0.216
ChebyNet	Y	**14.813**	**14.152**	**14.815**	**13.716**	**0.255**	0.255	0.259	**0.211**	**0.157**	**0.142**	**0.220**
	N	1,081.176	736.652	21.086	786.228	0.255	**0.254**	**0.259**	0.313	0.251	0.233	0.319
GGNN	Y	**245.362**	**245.362**	**245.362**	**245.366**	**0.254**	**0.255**	0.259	**0.215**	**0.148**	**0.133**	**0.213**
	N	826.467	1,025.906	232.697	1,046.386	0.260	0.260	**0.265**	0.262	0.207	0.188	0.274
MPNN	Y	**39.136**	**39.131**	**39.136**	**39.145**	0.255	0.255	0.259	**0.186**	**0.138**	**0.126**	**0.204**
	N	595.265	909.041	33.943	902.823	**0.254**	**0.254**	**0.258**	0.248	0.205	0.181	0.270
RGCN	Y	**59.300**	**59.303**	**59.300**	**59.260**	0.256	0.255	0.260	**0.090**	**0.057**	**0.067**	**0.097**
	N	626.588	1,044.292	70.394	755.135	**0.255**	**0.254**	**0.259**	0.155	0.128	0.120	0.167
SchNet	Y	**93.346**	**93.344**	**93.346**	**93.440**	0.255	**0.254**	0.259	**0.129**	**0.090**	**0.093**	**0.135**
	N	773.359	821.539	73.966	764.310	**0.254**	0.254	**0.259**	0.171	0.134	0.144	0.182

Table S9: The ratio of non-physical reliable prediction. (Better is in **bold**).

Model	Catalyst	U	U0	G	orbit	excited	all
ChebyNet	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	45.42%	44.02%	49.00%	**0.00%**	**0.00%**	83.07%
GAT	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	61.95%	59.56%	**58.96%**	**0.00%**	**0.00%**	96.02%
GCN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	42.03%	58.17%	58.96%	**0.00%**	**0.00%**	90.44%
GGNN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	30.48%	36.45%	35.86%	**0.00%**	**0.00%**	62.75%
GIN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	55.98%	32.87%	41.63%	**0.00%**	**0.00%**	84.26%
MPNN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	44.42%	63.15%	56.57%	**0.00%**	**0.00%**	93.43%
RGCN	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	17.73%	53.19%	26.49%	**0.00%**	**0.00%**	69.32%
SchNet	Y	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
	N	56.97%	60.96%	55.58%	**0.00%**	**0.00%**	91.43%

Table S10: The detailed prediction results for singlet fission. (Better in **bold**)

Model	Catalyst	TP	TN	FP	FN	Valid	Accuracy
GAT	Y	16	475	5	6	502	**97.8%**
	N	0	18	0	2	20	3.6%
GCN	Y	15	477	3	7	502	**98.0%**
	N	5	42	0	1	48	9.4%
GGNN	Y	0	480	0	22	502	**95.6%**
	N	0	184	0	3	187	36.7%
GIN	Y	7	478	2	15	502	**96.6%**
	N	0	78	0	1	79	15.5%
MPNN	Y	5	480	0	17	502	**96.6%**
	N	0	33	0	0	33	6.6%
ChebyNet	Y	17	477	3	5	502	**98.4%**
	N	3	80	1	1	85	16.5%
RGCN	Y	17	477	3	5	502	**98.4%**
	N	4	148	1	1	154	30.3%
SchNet	Y	3	480	0	19	502	**96.2%**
	N	3	40	0	0	43	8.6%

S9. Detailed discussion

We could give some analysis for Cat-DNNs scheme from the deep learning aspect. Traditional MTL methods will not consider the correlation of the predicted properties explicitly. Researchers hope

to find the correlation among properties on the DNN model itself, even though the DNN with one output layer with "squashing" activation can approximate any Borel measurable function[21]. However, limited by the cragged loss curve and stochastic gradient descent optimization method, more complex DNN architectures are needed to obtain a better approximation for different tasks.

On quantum-mechanic tasks like molecular properties prediction, the correlation of some properties is easy to be theoretically determined as introduced above. The endorsed catalysts reduce the search space of the DNN models, so that the loss can be easier optimized to a better minima. For example, to predict energy properties U, U_0, H, G, the traditional models need to find out the correlation between energy and molecular structure 4 times, some of the model parameters are duplicated, and the fitting ability is limited. Added the catalysts, the correlation of energy and molecular could be only considered once and the saved parameters can be used to fit the difference of various properties, like the gap between U_0 and U, which is much easier than fitting the correlation of molecular 3D structure to energy property once more.

Reference

1. J. Michl and E. W. Thulstrup, *Tetrahedron*, 1976, **32**, 205-209.
2. J. Singh and K. Shimakawa, *Advances in amorphous semiconductors*, CRC Press, 2003.
3. J. Singh, *Excitation energy transfer processes in condensed matter: theory and applications*, Springer Science & Business Media, 2013.
4. T. N. Kipf and M. Welling, *arXiv preprint arXiv:1609.02907*, 2016.
5. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio and Y. Bengio, *arXiv preprint arXiv:1710.10903*, 2017.
6. K. Xu, W. Hu, J. Leskovec and S. Jegelka, *arXiv preprint arXiv:1810.00826*, 2018.
7. M. Defferrard, X. Bresson and P. Vandergheynst, *arXiv preprint arXiv:1606.09375*, 2016.
8. Y. Li, D. Tarlow, M. Brockschmidt and R. Zemel, *arXiv preprint arXiv:1511.05493*, 2015.
9. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, 2017.
10. M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov and M. Welling, 2018.
11. K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *The Journal of Chemical Physics*, 2018, **148**, 241722.
12. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein and L. Antiga, *arXiv preprint arXiv:1912.01703*, 2019.
13. M. Fey and J. E. Lenssen, *arXiv preprint arXiv:1903.02428*, 2019.
14. W. Falcon, et al., *GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning*, 2019,

3.

15. J. Liang, S. Ye, T. Dai, Z. Zha, Y. Gao and X. Zhu, *Scientific Data*, 2020, **7**, 1-6.

16. D. P. Kingma and J. Ba, *arXiv preprint arXiv:1412.6980*, 2014.

17. S. Ye, J. Liang, R. Liu and X. Zhu, *The Journal of Physical Chemistry A*, 2020, **124**, 6945-6953.

18. R. Tibshirani, *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, **58**, 267-288.

19. S. Haastrup, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen and A. C. Riis-Jensen, *2D Materials*, 2018, **5**, 042002.

20. C. F. Perkinson, D. P. Tabor, M. Einzinger, D. Sheberla, H. Utzat, T.-A. Lin, D. N. Congreve, M. G. Bawendi, A. Aspuru-Guzik and M. A. Baldo, *The Journal of chemical physics*, 2019, **151**, 121102.

21. Y. B. Ian Goodfellow, Aaron Courville, *Deep Learning*, 2016.