First-Principles and Machine Learning Study on NRR in Curvature-Tuned TM-Doped CNTs

Xiaolin Jiang[1], Hongchang xu[1], Lifu Zhang[2], Zhenpeng Hu[1*]

[1]School of Physics, Nankai University, Tianjin, 300071, China.

[2]School of Biomedical Engineering and Technology, Tianjin Medical University, Tianjin, 300070, China.

***Corresponding author E-mail(s): zphu@nankai.edu.cn**

1. SISSO algorithm principle

SISSO is a feature selection and descriptor identification method based on the compressed sensing framework. Its core idea is to alternately apply "Set Independence Screening" (SIS) and "Sparsification Operators" (SO) to select a small number of key features from a high-dimensional, highly correlated feature space, forming low-dimensional descriptors for accurate prediction of material properties.

The first step is the SIS process, where candidate features are ranked based on their correlation with the target property. The top M features are then selected for the next sparsification step, forming the sub-feature space S_{1D}. The theoretical guarantee of the SIS step is that, as long as the feature set is large enough and satisfies certain conditions, the truly relevant features will likely appear at the top of the correlation ranking. After the feature subspace is pre-screened by SIS, the SO step is used to solve the sparse linear combination problem, further determining the optimal low-dimensional feature combination. For example, in the case of a two-dimensional descriptor problem, the residual vector Δ_{2D} can be defined as:

$$\Delta_{2D} = P - d_{2D}c_{2D} \tag{2-5}$$

Where d_{2D} is the matrix composed of the two selected features, and

$c_{2D} = (d_{2D}^T d_{2D})^{-1} d_{2D}^T P$ is the least squares solution. In general, for a descriptor with dimension n, the residual is defined as:

$$\Delta_{nD} = P - d_{nD} c_{nD} \tag{2-6}$$

In practical applications, the process starts with a one-dimensional descriptor, and the dimensionality is gradually increased. The residual at each step is used as a reference for the next round of SIS screening, until the residual reaches the desired accuracy.

In addition, the construction of the feature space starts from the initial features provided by the user (Φ_0), and recursively uses predefined unary and binary mathematical operators to generate higher-order candidate feature sets (Φ_n):

$$\Phi_n = \bigcup_{i=1}^{n} \hat{H}^{(m)}[\phi_1, \phi_2], \quad \forall \phi_1, \phi_2 \in \Phi_{i-1} \tag{2-7}$$

Where $\hat{H}^{(m)}$ represents the operator set (such as +, −, ×, /, exp, log, | |, etc.). After several iterations, a vast candidate feature space can be obtained. The algorithm effectively performs selection in the high-dimensional space through these two steps, significantly reducing computational complexity, thereby enabling rapid convergence to the optimal low-dimensional descriptor combination.

2. SISSO parameter Settings

Table S1 ΔG(*NNH) training parameters

Parameters	Values
Ptype	1
ntask	1
desc_dim	3
task_weighting	1
scmt	false
nsamples	24
restart	0
nsf	8
ops	"(+)(-)(*)(/)(exp)(exp-)(^-1)(^2)(^3)(cbrt)(|-|)(^6)(sin)(cos)'
fcomplexity	7
funit	(1:1)(2:2)(3:3)(4:4)(5:5)(6:8)
fmax_min	1e-3

fmax_max	1e5
ns_sis	1000
method	'L0'
nmodel	10
metric	'RMSE'
fit_intercept	true

Table S2 fNRR training parameters

Parameters	Values
Ptype	1
ntask	1
desc_dim	3
task_weighting	1
scmt	false
nsamples	22
restart	0
nsf	8
ops	"(+)(-)(*)(/)(exp)(exp-)(^-1)(^2)(^3)(cbrt)(|-|)(^6)(sin)(cos)'
fcomplexity	7
funit	(1:1)(2:2)(3:3)(4:4)(5:5)(6:9)
fmax_min	1e-3
fmax_max	1e5
ns_sis	1000
method	'L0'
nmodel	10
metric	'RMSE'
fit_intercept	true

3. Result

Table S3 Binding energy of TM-N3CNT with different curvatures

System	（4，4）	（6，6）	（8，8）	（10，10）
Sc	-7.709	-7.535	-7.471	-7.420
Ti	-7.600	-7.904	-7.529	-7.521
V	-7.560	-7.879	-7.602	-7.996
Cr	-4.981	-5.044	-4.817	-4.844
Mn	-4.789	-4.775	-4.811	-4.812
Fe	-5.294	-5.634	-5.506	-5.362
Co	-5.801	-5.888	-5.942	-6.001
Ni	-5.790	-5.885	-5.858	-5.806
Cu	-3.995	-4.038	-4.034	-4.022
Zn	-1.834	-2.354	-2.454	-2.506
Mo	-7.909	-7.701	-7.509	-7.327
W	-6.809	-7.098	-7.301	-7.500

Table S4 Gibbs Free Energy for Adsorption of Multiple H, N_2, and Nitrogen Protonation on TM-N_3CNTs at Various Curvatures.

Element	(m, m)CNT	ΔG (eV)									
		*H	*NN	*NNH	*NN*H	*1N_2-NN	*1N_2-NNH	*2N_2*H	*2N_2-NN	*2N_2-NNH	*3N_2*H
Sc	(4, 4)	-0.479	-0.821	0.873	0.041	-0.580	0.966		-0.539	1.042	0.262

	(6, 6)	-0.621	-0.825	0.747	-0.068	-0.614	0.887	-0.017	-0.532	0.922	0.337
	(8, 8)	-0.599	-0.855	0.715	-0.123	-0.620	0.842	-0.090	-0.558	0.913	0.288
	(10, 10)	-0.656	-0.893	0.704	-0.221	-0.603	0.812	0.107	-0.582	0.892	0.281
Ti	(4, 4)	-0.929	-1.249	0.658	0.116	-0.719	0.707	0.188	-0.682	0.825	0.738
	(6, 6)	-0.882	-1.235	0.685	-0.038	-0.711	0.667	0.344	-0.764	0.814	0.740
	(8, 8)	-0.809	-1.246	0.683	-0.030	-0.714	0.644	0.189	-0.804	0.810	0.782
	(10, 10)	-0.868	-1.265	0.678	-0.067	-0.706	0.636	0.162	-0.792	0.770	0.783
V	(4, 4)		-1.127	0.650	0.007	-0.862	0.683	0.360	-0.670	0.759	-0.421
	(6, 6)	-0.655	-1.137	0.663	0.004	-0.872	0.686	0.347	-0.687	0.753	0.875
	(8, 8)	-0.684	-1.129	0.648	0.039	-0.872	0.658	2.157	-0.712	0.752	0.901
	(10, 10)	-0.690	-1.130	0.529	-0.189	-0.863	0.671	0.342	-0.695	0.747	0.909
Cr	(4, 4)		-0.960	0.643	-0.091	-0.813	0.595	0.167	-0.898	0.670	1.035
	(6, 6)	-0.419	-0.997	0.471	-0.173	-0.948	0.610	-0.035	-0.874	0.630	1.029
	(8, 8)	-0.473	-1.052	0.427	-0.175	-0.952	0.594	0.406	-0.884	0.613	1.050
	(10, 10)	-0.554	-1.139	0.400	-0.178	-1.752	1.375	0.784	-0.087	0.603	1.068
Mn	(4, 4)	-0.294	-1.044	0.722	0.166	-0.096	0.157	-0.467	-1.660	1.568	2.248
	(6, 6)	-0.318	-1.049	0.712	0.167	-0.822	0.797	0.176	-1.029	1.505	2.131
	(8, 8)	0.076	-1.053	0.731	0.147	-0.636	0.574	-0.016	-1.249	1.447	2.104
	(10, 10)	0.051	-1.022	0.692	0.118	-0.667	0.574	-0.010	-1.245	1.412	2.095
Fe	(4, 4)	-0.135	-1.296	0.955	0.042	-0.050	0.230	-0.869	-0.813	0.960	
	(6, 6)	-0.397	-1.136	0.745	-0.125	-0.927	0.883	-0.241	-0.339		
	(8, 8)	-0.398	-1.286	0.819	0.009	-0.805	0.854	-0.263	-0.448		
	(10, 10)	-0.147	-1.092	0.594	-0.202	-1.022	0.878	-0.242	-0.454	1.300	
Co	(4, 4)	-0.098	-1.222	1.007	-0.047	-0.154	0.962	0.120	-0.293	1.640	
	(6, 6)	-0.389	-1.479	0.929	-0.094	-0.351	1.055	0.199	-0.147		
	(8, 8)	-0.153	-1.231	0.897	-0.053	-0.381	1.020	0.146	-0.187		
	(10, 10)	-0.152	-1.212	0.862	-0.142	-0.436	1.050	0.165	-0.154		
Ni	(4, 4)	0.238	-0.920	0.760	0.111	-0.107	1.547				
	(6, 6)	0.120	-0.956	0.746	0.015	-0.197					
	(8, 8)	0.033	-0.996	0.776	-0.019	-0.207					
	(10, 10)	-0.033	-1.014	0.721	-0.051	-0.265					
Cu	(4, 4)	0.386	-1.135	1.682	1.477						
	(6, 6)	0.313	-1.220	1.665	1.432						
	(8, 8)	0.293	-1.277	1.659	1.392						
	(10, 10)	0.222	-1.311	1.806	1.449						
Zn	(4, 4)	-1.225	-0.576	0.965							
	(6, 6)	-1.066	-0.532	1.031							
	(8, 8)	-1.002	-0.523	1.055							
	(10, 10)	-0.995	-0.508	1.039							
Mo	(4, 4)	-0.847	-1.340	0.427	-0.415	-1.373	0.550	0.225	-1.190	0.470	0.415
	(6, 6)	-0.916	-1.338	0.280	-0.389	-1.359	0.459	0.221	-1.224	0.422	0.377
	(8, 8)	-0.962	-1.378	0.237	-0.402	-1.346	0.381	0.235	-1.254	0.400	0.384
	(10, 10)	-0.950	-1.367	0.210	-0.406	-1.334	0.356	0.244	-1.267	0.379	0.381
W	(4, 4)	-1.345	-1.530	0.196	-0.861	-1.599	0.375	-0.177	-1.461	0.309	0.285
	(6, 6)	-1.396	-1.538	0.057	-0.863	-1.594	0.2342	-0.160	-1.505	0.263	0.261
	(8, 8)	-1.406	-1.572	0.013	-0.885	-1.590	0.160	-0.130	-1.534	0.242	0.265
	(10, 10)	-1.406	-1.584	0.009	-0.870	-1.560	0.129	-0.124	-1.552	0.220	0.274

Table S5: Reaction Proportions of *2N_2*NNH and *3N_2*H.

(m,m) CNT	f_{NRR}						
	Sc	Ti	V	Cr	Mn	Mo	W
(4, 4)	0.000	0.220	0.000	0.995	1.000	0.310	0.414
(6, 6)	0.000	0.255	0.856	0.997	1.000	0.343	0.493
(8, 8)	0.000	0.399	0.896	0.998	1.000	0.442	0.584
(10, 10)	0.000	0.547	0.912	0.999	1.000	0.507	0.684

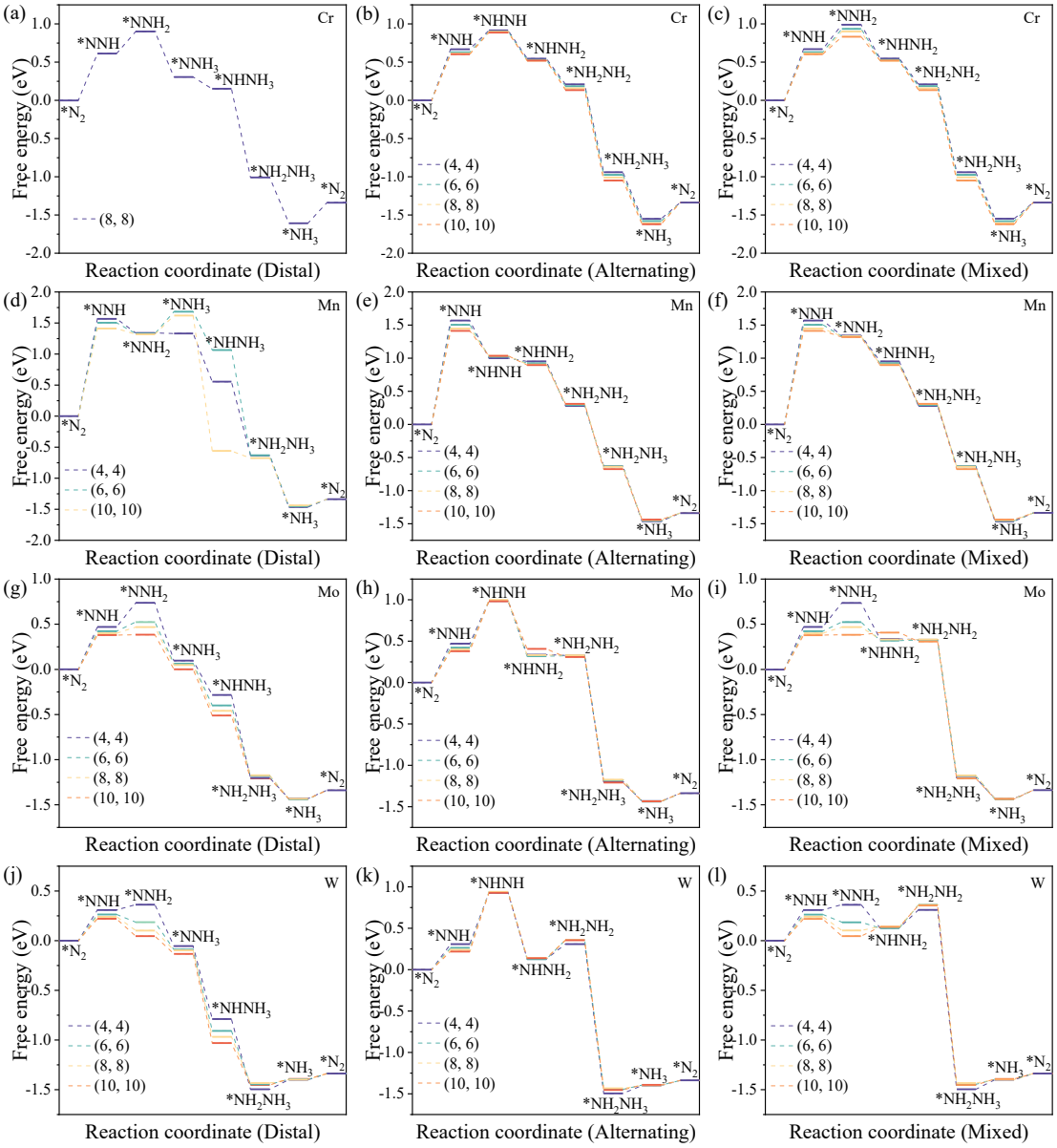

Fig.S1: Reaction Pathway Step Diagram of eNRR for Different Pathways.