SUPPLEMENTARY INFORMATION

Identifying key factors of peroxymonosulfate activation on single-atom M–N–C catalysts: A combined density functional theory and machine learning study

Yun Sun[1], Jiachun Cao[1,2], Qianyu Li[2], Didi Li[1], Zhimin Ao[2]*

[1]*Institute of Environmental Health and Pollution Control, School of Environmental Science and Engineering, Guangdong University of Technology, Guangzhou 510006, PR China*

[2]*Advanced Interdisciplinary Institute of Environment and Ecology, Beijing Normal University, Zhuhai 519087, PR China*

Corresponding author's email address: zhimin.ao@bnu.edu.cn

Pages (9)

Text (1)

Text S1 Empirical Risk Minimization for few-shot learning (FSL)

Tables (6)

Table S1 The formation energy (E_f) of M@N_2C_2, M@N_3C_1 and M@N_4.

Table S2 The adsorption energy (E_{ad}), O-O bond length ($l_{O\text{-}O}$), and electron transfer (q) for the four adsorption configurations of PMS on Fe@N_4.

Table S3 The d-band center (ε_d), the upper-edge of the d-band (εdW) and adsorption energy (E_{ad}).

Table S4 The DFT calculated activation energy barrier (E_{bar}) and reaction energy (ΔE).

Table S5 The input data electronegativity (E), d-electron count (ϑ_d), group number (g), radius (r_M) and the number of nitrogen atoms (N_N).

Table S6 The mutual information (MI) and Pearson correlation coefficient (P) of the features.

Figures (4)

Fig. S1 Partial density of states (PDOS) of d orbitals for $M@N_2C_2$. The d-band centers (ε_d) are labeled for 3d metals, from Sc to Zn (green); 4d metals, from Y to Cd (purple); and 5d metals, from Hf to Au (yellow). The Fermi level (E_F) is set to 0 eV.

Fig. S2 Partial density of states (PDOS) of d orbitals for $M@N_3C_1$. The d-band centers (ε_d) are labeled for 3d metals, from Sc to Zn (green); 4d metals, from Y to Cd (purple); and 5d metals, from Hf to Au (yellow). The Fermi level (E_F) is set to 0 eV.

Fig. S3 The selected 27 catalysts for training (yellow) and 34 catalysts for predicting (blue).

Fig. S4 Partial density of states (PDOS) for (a) $Fe@N_2C_2$, (b) $Fe@N_3C_1$, and (c) $Fe@N_4$.

Text S1 Empirical Risk Minimization for few-shot learning (FSL)

The core issue of FSL based on error decomposition in supervised machine learning is empirical risk minimization. Given a hypothesis *h*, we want to minimize its expected risk *R*, which is the loss measured with respect to $p(x,y)$. Specifically,

$$R(h) = \int l(h(x),y)dp(x,y) = E[l(h(x),y)].$$

As $p(x,y)$ is unknown, the empirical risk (which is the average of sample losses over the training set D_{train} of *I* samples)

$$R_I(h) = \frac{1}{I}\sum_{i=1}^{I} l(h(x),y),$$

is usually used as a proxy for $R(h)$, leading to empirical risk minimization (with possibly some regularizers). For illustration, let

- $\hat{h} = \arg min_h R(h)$ be the function that minimizes the expected risk;
- $h^* = \arg min_{h \in H} R(h)$ be the function in *H* that minimizes the expected risk;
- $h_I = \arg min_{h \in H} R_I(h)$ be the function in *H* that minimizes the empirical risk.

As $\hat{h}$ is unknown, one has to approximate it by some $h \in H$. h^* is the best approximation for $\hat{h}$ in *H*, while h_I is the best hypothesis in *H* obtained by empirical risk minimization. For simplicity, we assume that $\hat{h}$, h^* and h_I are unique. The total error can be decomposed as:

$$E\left[R(h_I) - R(\hat{h})\right] = \underbrace{E\left[R(h^*) - R(\hat{h})\right]}_{\varepsilon_{app}(H)} + \underbrace{E\left[R(h_I) - R(h^*)\right]}_{\varepsilon_{est}(H,I)},$$

where the expectation is with respect to the random choice of D_{train}. The approximation error $\varepsilon_{app}(H)$ measures how close the functions in *H* can approximate the optimal hypothesis $\hat{h}$, and the estimation error $\varepsilon_{est}(H,I)$ measures the effect of

minimizing the empirical risk $R_I(h)$ instead of the expected risk $R(h)$ within *H*.

As shown, the total error is affected by *H* (hypothesis space) and *I* (number of examples in D_{train}). In other words, learning to reduce the total error can be attempted from the perspectives of (i) data, which provides D_{train}; (ii) model, which determines *H*; and (iii) algorithm, which searches for the optimal $h_I \in H$ that fits D_{train}.[1]

Table S1 The formation energy (E_f) of M@N_2C_2, M@N_3C_1 and M@N_4.

No.	Metal	Base	E_f
1	Sc	N_2C_2	-3.061
2	Ti	N_2C_2	-2.223
3	V	N_2C_2	-1.998
4	Cr	N_2C_2	-2.454
5	Mn	N_2C_2	-3.599
6	Fe	N_2C_2	-2.704
7	Co	N_2C_2	-5.238
8	Ni	N_2C_2	-4.695
9	Cu	N_2C_2	-4.348
10	Zn	N_2C_2	-3.806
11	Y	N_2C_2	-1.279
12	Zr	N_2C_2	-0.070
13	Nb	N_2C_2	0.479
14	Mo	N_2C_2	0.073
15	Tc	N_2C_2	0.728
16	Ru	N_2C_2	-1.829
17	Rh	N_2C_2	-3.800
18	Pd	N_2C_2	-3.550
19	Ag	N_2C_2	-2.373
20	Cd	N_2C_2	-2.069
21	Hf	N_2C_2	-0.519
22	Ta	N_2C_2	0.641
23	W	N_2C_2	0.964
24	Re	N_2C_2	0.482
25	Os	N_2C_2	-1.453
26	Ir	N_2C_2	-3.603
27	Pt	N_2C_2	-3.074
28	Au	N_2C_2	-2.129
29	Sc	N_3C_1	-2.982
30	Ti	N_3C_1	-1.771
31	V	N_3C_1	-1.380
32	Cr	N_3C_1	-2.235
33	Mn	N_3C_1	-3.072
34	Fe	N_3C_1	-3.022
35	Co	N_3C_1	-4.808
36	Ni	N_3C_1	-4.113
37	Cu	N_3C_1	-3.367
38	Zn	N_3C_1	-3.318
39	Y	N_3C_1	-1.200
40	Zr	N_3C_1	0.286
41	Nb	N_3C_1	1.061
42	Mo	N_3C_1	0.814
43	Tc	N_3C_1	1.609
44	Ru	N_3C_1	-1.346
45	Rh	N_3C_1	-3.551
46	Pd	N_3C_1	-3.096
47	Ag	N_3C_1	-1.377
48	Cd	N_3C_1	-1.472
49	Hf	N_3C_1	-0.153
50	Ta	N_3C_1	1.210
51	W	N_3C_1	1.701
52	Re	N_3C_1	1.379
53	Os	N_3C_1	-0.783
54	Ir	N_3C_1	-3.344
55	Pt	N_3C_1	-2.627

No.	Metal	Base	E_f
56	Au	N_3C_1	-1.149
57	Sc	N_4	-2.712
58	Ti	N_4	-1.108
59	V	N_4	-0.616
60	Cr	N_4	-1.808
61	Mn	N_4	-2.247
62	Fe	N_4	-2.260
63	Co	N_4	-4.042
64	Ni	N_4	-3.021
65	Cu	N_4	-2.301
66	Zn	N_4	-3.008
67	Y	N_4	-1.163
68	Zr	N_4	0.937
69	Nb	N_4	2.079
70	Mo	N_4	2.061
71	Tc	N_4	2.445
72	Ru	N_4	-0.767
73	Rh	N_4	-2.867
74	Pd	N_4	-2.016
75	Ag	N_4	0.106
76	Cd	N_4	-1.192
77	Hf	N_4	0.492
78	Ta	N_4	2.188
79	W	N_4	3.053
80	Re	N_4	2.400
81	Os	N_4	-0.203
82	Ir	N_4	-2.677
83	Pt	N_4	-1.588
84	Au	N_4	0.625

*The 23 catalysts with positive formation energy were marked in gray, and excluded from subsequent calculations.

Table S2. The adsorption energy (E_{ad}), O-O bond length ($l_{O\text{-}O}$), and electron transfer (q) for the four adsorption configurations of PMS on Fe@N_4.

	E_{ad} (eV)	$l_{O\text{-}O}$ (Å)	q (e)
1	-2.20	1.47	0.19
2	-1.86	1.46	0.26
3	-1.87	1.47	0.26
4	-2.16	1.50	0.12

Table S3. The d-band center (ε_d), the upper-edge of the d-band (ε_d^W) and adsorption energy (E_{ad}).

No.	Metal	Base	ε_d (eV)	ε_d^W (eV)	E_{ad} (eV)
1	Sc	N_2C_2	0.07	7.86	-5.69
2	Ti	N_2C_2	-0.32	6.31	-5.03
3	V	N_2C_2	-0.87	5.47	-4.37
4	Cr	N_2C_2	-0.27	6.23	-3.56
5	Mn	N_2C_2	-1.61	4.64	-2.73
6	Fe	N_2C_2	-1.55	4.18	-3.06
7	Co	N_2C_2	-0.81	4.50	-2.23
8	Ni	N_2C_2	-1.21	3.58	-1.82
9	Cu	N_2C_2	-4.24	0.35	-1.94
10	Zn	N_2C_2	-6.70	-3.23	-2.26
11	Y	N_2C_2	-0.67	8.08	-7.79
12	Zr	N_2C_2	-0.67	6.86	-7.20
13	Ru	N_2C_2	-2.27	4.67	-4.35
14	Rh	N_2C_2	-2.00	3.82	-3.16
15	Pd	N_2C_2	-2.23	4.04	-1.91
16	Ag	N_2C_2	-5.99	-0.40	-1.98
17	Cd	N_2C_2	-9.53	-4.93	-3.79
18	Hf	N_2C_2	-0.74	6.89	-7.63
19	Os	N_2C_2	-2.31	4.76	-4.72
20	Ir	N_2C_2	-2.10	3.89	-3.72
21	Pt	N_2C_2	-2.98	2.73	-2.07
22	Au	N_2C_2	-6.08	0.09	-2.14
23	Sc	N_3C_1	0.83	8.23	-5.45
24	Ti	N_3C_1	-0.50	6.29	-4.81
25	V	N_3C_1	-0.81	5.71	-4.00
26	Cr	N_3C_1	-0.26	6.47	-2.96
27	Mn	N_3C_1	-1.22	5.49	-2.59
28	Fe	N_3C_1	-1.27	4.61	-2.37
29	Co	N_3C_1	-0.67	4.79	-2.23
30	Ni	N_3C_1	-1.66	3.26	-1.96
31	Cu	N_3C_1	-4.01	0.61	-2.03
32	Zn	N_3C_1	-6.56	-3.04	-2.56
33	Y	N_3C_1	-0.68	8.54	-7.80
34	Ru	N_3C_1	-1.67	4.73	-3.35
35	Rh	N_3C_1	-1.78	4.27	-2.26
36	Pd	N_3C_1	-2.94	3.42	-1.96
37	Ag	N_3C_1	-6.15	-0.50	-1.97
38	Cd	N_3C_1	-9.04	-4.53	-4.32
39	Hf	N_3C_1	-0.70	7.22	-6.87
40	Os	N_3C_1	-1.69	4.76	-3.82
41	Ir	N_3C_1	-1.87	4.30	-2.52
42	Pt	N_3C_1	-3.36	3.07	-2.05
43	Au	N_3C_1	-6.28	0.00	-2.15
44	Sc	N_4	0.16	7.51	-5.14
45	Ti	N_4	-0.49	6.39	-4.23
46	V	N_4	-0.01	7.04	-4.12
47	Cr	N_4	-0.10	6.97	-2.50
48	Mn	N_4	-1.13	5.65	-2.36
49	Fe	N_4	-0.39	5.28	-2.20
50	Co	N_4	-1.02	4.58	-1.99

No.	Metal	Base	ε_d (eV)	(eV)	E_{ad} (eV)
51	Ni	N_4	-2.00	3.02	-1.85
52	Cu	N_4	-3.39	1.19	-2.01
53	Zn	N_4	-6.65	-3.00	-2.46
54	Y	N_4	-0.24	8.92	-7.33
55	Ru	N_4	-1.53	5.17	-2.53
56	Rh	N_4	-2.10	4.20	-1.93
57	Pd	N_4	-3.27	3.16	-1.93
58	Cd	N_4	-8.56	-4.03	-4.18
59	Os	N_4	-1.51	5.22	-2.68
60	Ir	N_4	-2.16	4.27	-1.98
61	Pt	N_4	-3.35	3.63	-1.98

Table S4. The DFT calculated activation energy barrier (E_{bar}) and reaction energy (ΔE).

No.	Metal	Base	E_{bar}	ΔE
1	Sc	N_2C_2	2.36	-41.55
2	Ti	N_2C_2	4.20	-41.20
3	V	N_2C_2	2.25	-28.67
4	Cr	N_2C_2	1.78	-38.61
5	Co	N_2C_2	7.81	-9.29
6	Y	N_2C_2	0.75	-3.54
7	Zr	N_2C_2	4.91	-32.66
8	Hf	N_2C_2	0.47	-34.96
9	Ir	N_2C_2	1.64	-4.35
10	V	N_3C_1	1.29	-52.66
11	Ni	N_3C_1	9.62	-0.12
12	Cu	N_3C_1	17.45	0.00
13	Zn	N_3C_1	9.44	-4.11
14	Ru	N_3C_1	2.95	-16.30
15	Rh	N_3C_1	6.86	-8.22
16	Ag	N_3C_1	12.17	-18.92
17	Cd	N_3C_1	4.95	-0.95
18	Ir	N_3C_1	4.99	-23.35
19	Ti	N_4	0.84	-43.75
20	Mn	N_4	1.98	-33.01
21	Fe	N_4	3.09	-22.69
22	Co	N_4	7.45	-13.70
23	Ni	N_4	14.16	-1.38
24	Cu	N_4	14.68	-3.53
25	Zn	N_4	4.91	-7.01
26	Ru	N_4	4.77	-22.30
27	Os	N_4	4.38	-24.98

Table S5. The input data electronegativity (E), d-electron count (ϑ_d), group number (g), radius (r_M) and the number of nitrogen atoms (N_N).

No.	Metal	Base	E	ϑ_d	g	r_M	N_N
1	Sc	N_2C_2	1.36	1	3	1.7	2
2	Ti	N_2C_2	1.54	2	4	1.6	2
3	V	N_2C_2	1.63	3	5	1.53	2
4	Cr	N_2C_2	1.66	5	6	1.39	2
5	Co	N_2C_2	1.88	7	8	1.26	2
6	Y	N_2C_2	1.22	1	3	1.9	2
7	Zr	N_2C_2	1.33	2	4	1.75	2
8	Hf	N_2C_2	1.30	2	4	1.75	2
9	Ir	N_2C_2	2.20	7	8	1.41	2
10	V	N_3C_1	1.63	3	5	1.53	3
11	Ni	N_3C_1	1.91	8	8	1.24	3
12	Cu	N_3C_1	1.90	10	1	1.32	3
13	Zn	N_3C_1	1.65	10	2	1.22	3
14	Ru	N_3C_1	2.20	7	8	1.46	3
15	Rh	N_3C_1	2.28	8	8	1.42	3
16	Ag	N_3C_1	1.93	10	1	1.45	3
17	Cd	N_3C_1	1.69	10	2	1.44	3
18	Ir	N_3C_1	2.20	7	8	1.41	3
19	Ti	N_4	1.54	2	4	1.6	4
20	Mn	N_4	1.55	5	7	1.39	4
21	Fe	N_4	1.83	6	8	1.32	4
22	Co	N_4	1.88	7	8	1.26	4
23	Ni	N_4	1.91	8	8	1.24	4
24	Cu	N_4	1.90	10	1	1.32	4
25	Zn	N_4	1.65	10	2	1.22	4
26	Ru	N_4	2.20	7	8	1.46	4
27	Os	N_4	2.20	6	8	1.44	4

Table S6. The mutual information (MI) and Pearson correlation coefficient (P) of the features.

	E	ϑ_d	g	r_M	N_N
MI	0.33	0.35	0.53	0.28	0.1
P	0.34	0.71	0.68	-0.57	0.30

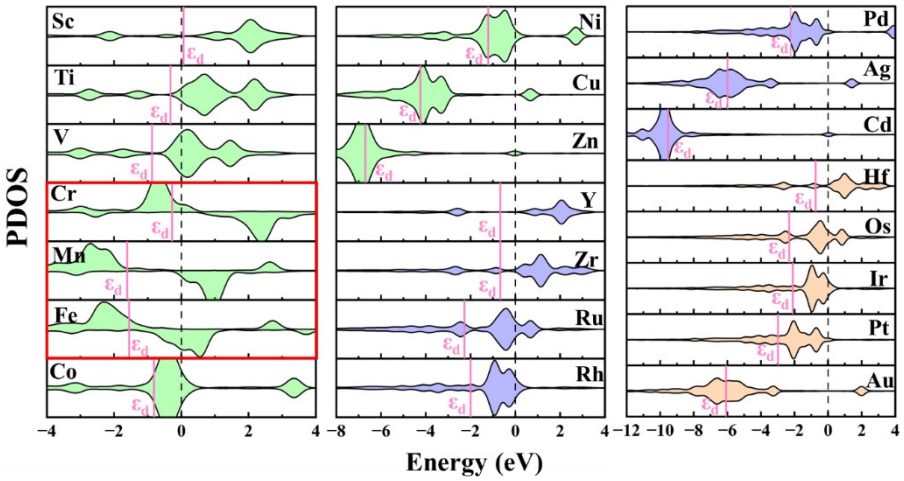

Fig. S1. Partial density of states (PDOS) of d orbitals for M@N_2C_2. The d-band centers (ε_d) are labeled for 3d metals, from Sc to Zn (green); 4d metals, from Y to Cd (purple); and 5d metals, from Hf to Au (yellow). The Fermi level (E_F) is set to 0 eV.

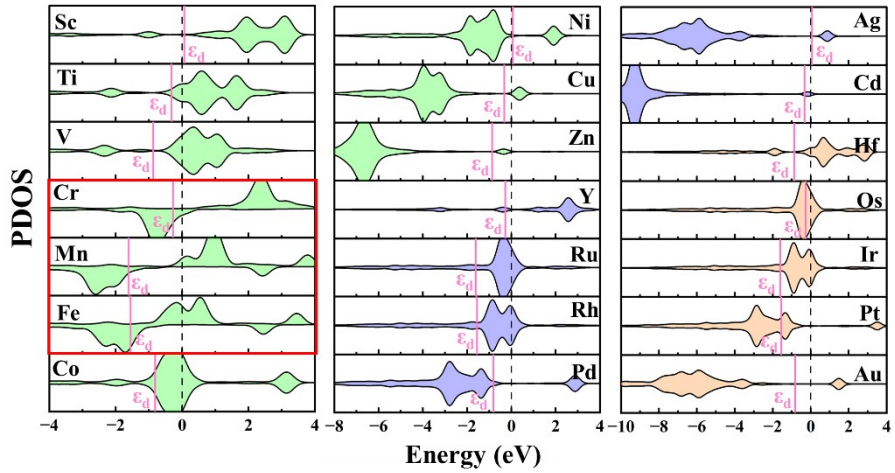

Fig. S2. Partial density of states (PDOS) of d orbitals for M@N_3C_1. The d-band centers (ε_d) are labeled for 3d metals, from Sc to Zn (green); 4d metals, from Y to Cd (purple); and 5d metals, from Hf to Au (yellow). The Fermi level (E_F) is set to 0 eV.

	3	4	5	6	7	8	9	10	11	12
$M@N_2C_2$	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
		Hf	Ta	W	Re	Os	Ir	Pt	Au	
$M@N_3C_1$	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
		Hf	Ta	W	Re	Os	Ir	Pt	Au	
$M@N_4$	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
		Hf	Ta	W	Re	Os	Ir	Pt	Au	

Fig. S3. The selected 27 catalysts for training (yellow) and 34 catalysts for predicting (blue).

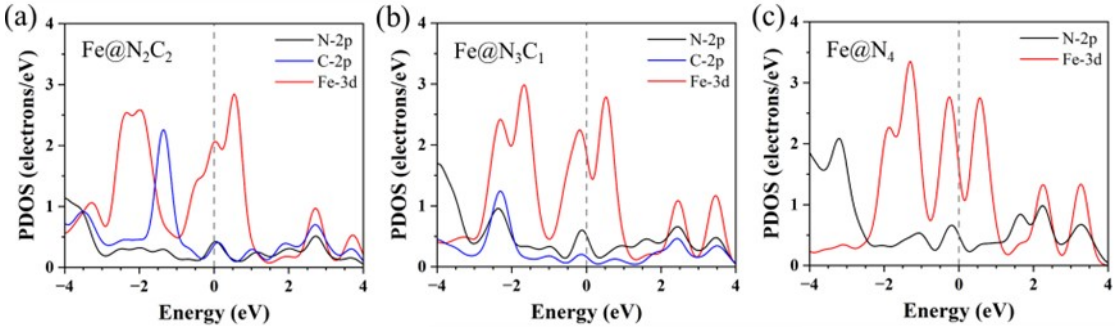

Fig. S4 Partial density of states (PDOS) for (a) $Fe@N_2C_2$, (b) $Fe@N_3C_1$, and (c) $Fe@N_4$.

Reference

1. Y. Q. Wang, Q. M. Yao, J. T. Kwok and L. M. Ni, Acm Computing Surveys, 2020, 53, 1-34.