New Carbon Materials, 2023, 38(5): 887 - 897

Supporting Information

Optimizing the growth of vertically aligned carbon nanotubes by literature mining and high-throughput experiments

Zhang-Dan Gao[1,2,†], Zhong-Hai Ji[1,2,†], Lili Zhang[1,*], Dai-Ming Tang[3,*], Meng-Ke Zou[1,2], Rui-Hong Xie[1,2], Shao-Kang Liu[1,2], Chang Liu[1,*]

[1]Shenyang National Laboratory for Materials Science, Institute of Metal Research (IMR), Chinese Academy of Sciences, Shenyang 110016, China;

[2]School of Materials Science and Engineering, University of Science and Technology of China, Hefei 230026, China;

[3]Research Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki 305-0044, Japan

NEW CARBON MATERIALS

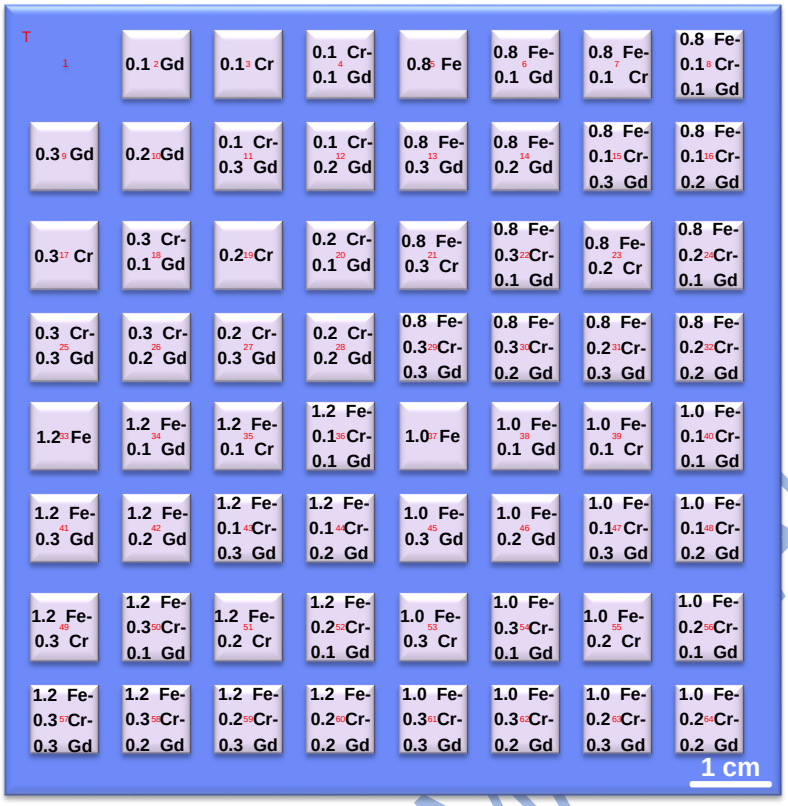

Figure S1. Schematic diagram of the nominal thickness of the Fe-Cr-Gd ternary catalyst library obtained using the combined masking method. The values of t_{Cr} and t_{Gd} are from 0 to 0.3 nm and the values of t_{Fe} are from 0 to 1.2 nm. The red numbers from 1 to 64 represent the marks on the silicon wafer.

Table S1. Statistical classification of manual data mining results.

Input parameter					
Annealing temperature	Annealing atmosphere	Primary catalyst	Co-catalyst	Support layer	Annealing time
/	H_2 Ar C_2H_2 ⋯	Fe Co	Gd	Al_2O_3 SiO_2	/
Growth temperature	Carbon source	Growth promoter	Carrier gas		Growth time
/	C_2H_2 C_2H_4 C_2H_5OH ⋯	Ar H_2	O_2 H_2O ⋯		/
Output parameter					
Height		I_G/I_D value			

NEW CARBON MATERIALS

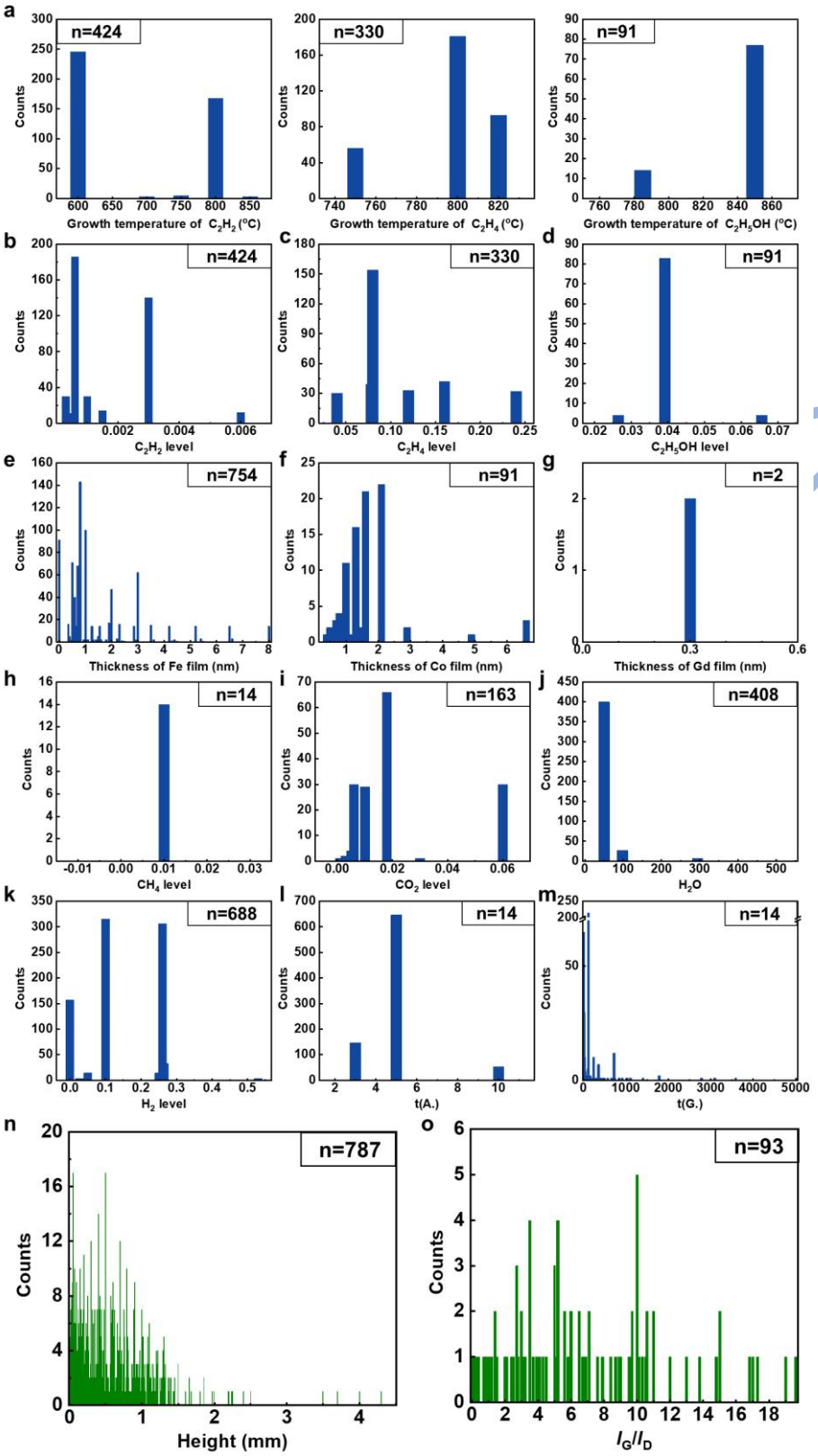

Figure S2. Distribution of 13 growth parameters (a-m) and 2 structural data (n-o) in the literature investigation. (a) Demonstrates the growth temperature distribution of C_2H_2, C_2H_4, C_2H_5OH. n denotes the number of recipes.

Table S2. Optimized hyperparameters for random forest regression (RFR) and support vector regression (SVR) on the training data for different output parameters. A grid search method was used to obtain the listed values for optimizing the structure of the machine learning models.

Output parameters	Hyperparameter		
	RFR	SVR	
	Number of trees	Penalty parameter C)	Gamma (γ)
Height	50	200	500
I_G/I_D	70	20	1

Table S3. Average k-fold cross-validation (CV) scores for prediction height (10-fold) and I_G/I_D values (5-fold) using 3 different machine learning models.

Output parameters	CV score		
	LR	RFR	SVR
Height	0.30	0.73	0.32
I_G/I_D	0.44	0.66	0.26

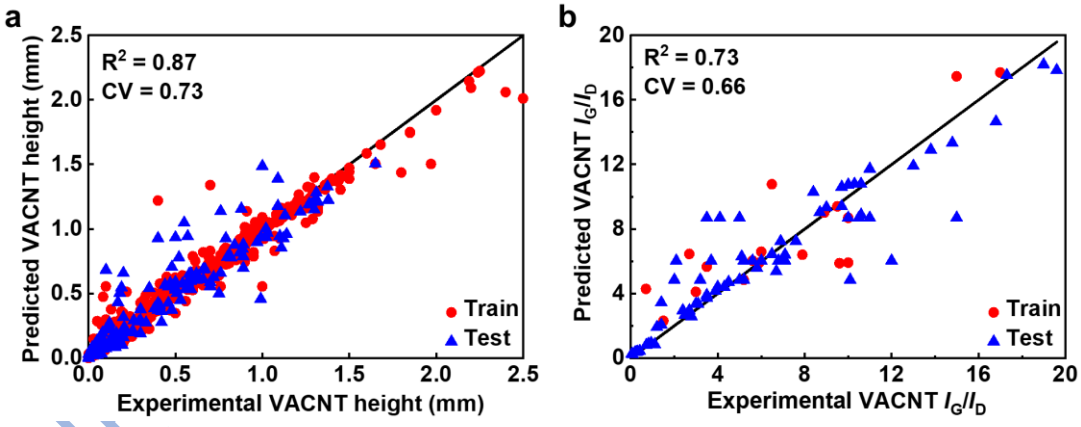

Figure S3. Linear fit between literature statistics and RFR predicted values for performance assessment of the model. (a) Height. (b) I_G/I_D values. Red circles and blue triangles represent the training and test set, respectively.

NEW CARBON MATERIALS

Table S4. Range and step size of growth parameters for determining the growth window for VACNTs.

Growth parameters	Range	Step
Thickness of Fe film /nm	0-9.45	0.15
Growth temperature /°C	750-820	20
Carbon source concentration /%	0-0.201	0.067

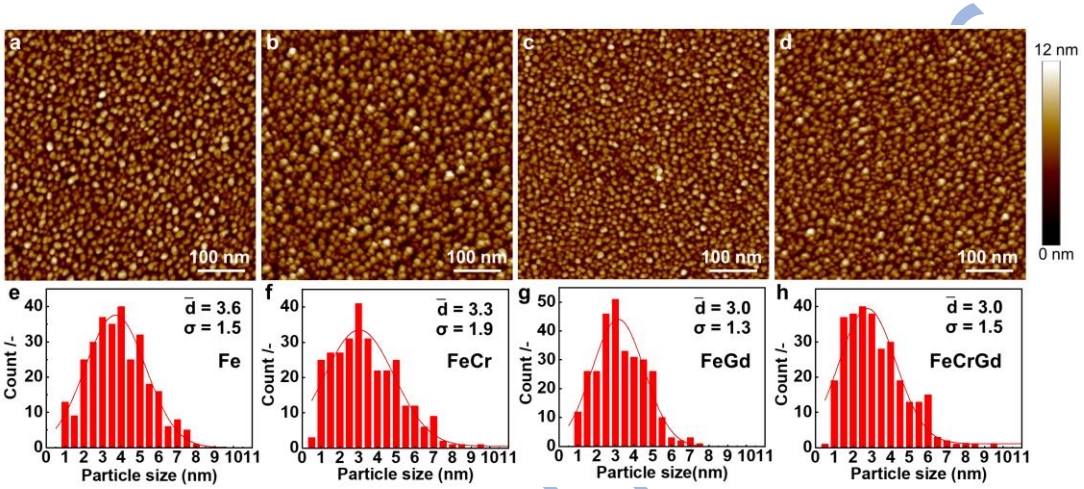

Figure S4. Average nanoparticle size and their size distribution of different catalysts on Al_2O_3 support layer after annealing under H_2 and Ar at 800°C for 10 min. (a) 1.2 nm Fe. (b) 1.1 nm Fe/0.1 nm Cr. (c) 1.1 nm Fe/0.1 nm Gd. (d) 1.0 nm Fe/0.1 nm Cr/0.1 nm Gd.

Table S5. Optimized hyperparameters for random forest regression (RFR) and support vector regression (SVR) on the training data for different output parameters based on high-throughput growth experiments. A grid search method was used to obtain the listed values for optimizing the structure of the machine learning models.

Output parameters	Hyperparameter		
	RFR	SVR	
	Number of trees	Penalty parameter C)	Gamma (γ)
Height	70	30	1
I_G/I_D	60	200	1

NEW CARBON

Table S6. Average 10-fold cross-validation (CV) scores for prediction Height and I_G/I_D values using 3 different machine learning models based on high-throughput growth experiments.

Output parameters	CV score		
	LR	RFR	SVR
Height	0.27	0.76	0.34
I_G/I_D	0.36	0.73	0.18

Table S7. Regression model validation and performance evaluation based on high-throughput growth experiments.

Output Parameters	Regression model	Performance evaluation metrics		
		RMSE	MAE	Coefficient of Determination (R^2)
Height	LR	0.18	0.14	0.29
	RFR	0.09	0.06	0.83
	SVR	0.12	0.09	0.70
I_G/I_D	LR	1.09	0.79	0.32
	RFR	0.61	0.42	0.77
	SVR	0.80	0.49	0.67

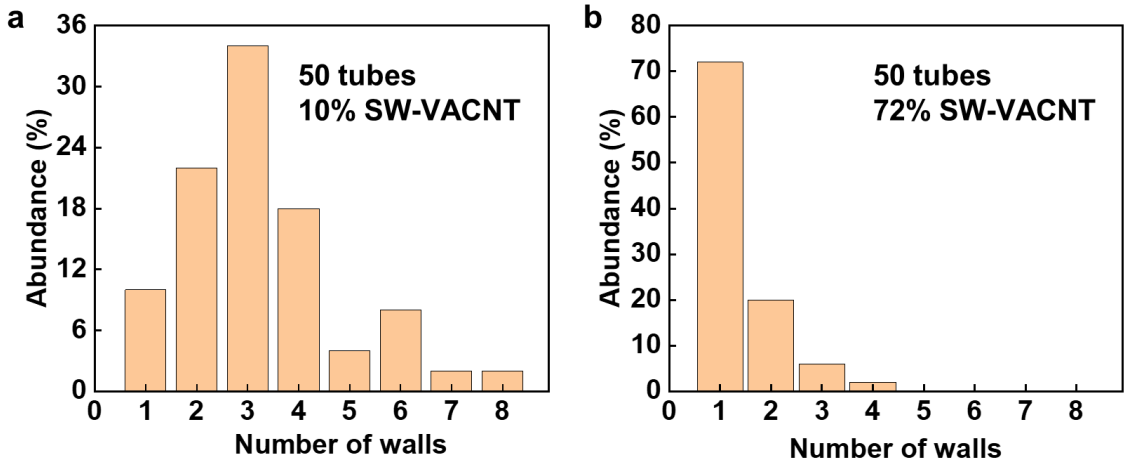

Figure S5. Wall number distributions of VACNT arrays grown from (a) 1.8 nm Fe catalyst and (b) 1.5 nm Fe/0.2 nm Gd catalyst.

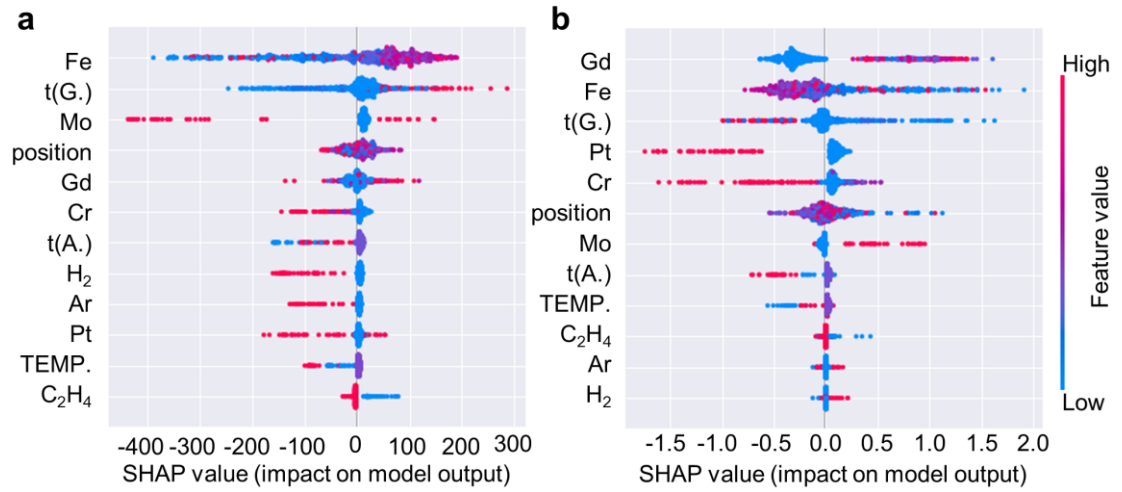

Figure S6. SHAP model for ranking the importance of growth parameters. The output parameters are height (a) and I_G/I_D value (b) of VACNTs. t(G.) and t(A.) denote the growth time and annealing time, respectively. Position indicates the coordinates of the sample in the marked silicon wafer.

NEW CARBON MATERIALS