Research Article

Nano Research
2021, 14(12): 4610–4615

ISSN 1998-0124 CN 11-5974/O4
https://doi.org/10.1007/s12274-021-3387-y

High-throughput screening and machine learning for the efficient growth of high-quality single-wall carbon nanotubes

Zhong-Hai Ji[1,2], Lili Zhang[1], Dai-Ming Tang[3] (✉), Chien-Ming Chen[4], Torbjörn E. M. Nordling[4,5], Zheng-De Zhang[6], Cui-Lan Ren[6], Bo Da[7], Xin Li[1,2], Shu-Yu Guo[1], Chang Liu[1] (✉), and Hui-Ming Cheng[1,8]

[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] *International Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan*
[4] *Department of Mechanical Engineering, "National Cheng Kung University", No. 1, University Road, Tainan City 701*
[5] *Department of Applied Physics and Electronics, Umeå University, 90187 Umeå, Sweden*
[6] *Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China*
[7] *Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science (NIMS), Ibaraki 305-0047, Japan*
[8] *Tsinghua-Berkeley Shenzhen Institute (TBSI), Tsinghua University, Shenzhen 518055, China*

Received: 9 November 2020 / **Revised:** 23 January 2021 / **Accepted:** 5 February 2021

ABSTRACT

It has been a great challenge to optimize the growth conditions toward structure-controlled growth of single-wall carbon nanotubes (SWCNTs). Here, a high-throughput method combined with machine learning is reported that efficiently screens the growth conditions for the synthesis of high-quality SWCNTs. Patterned cobalt (Co) nanoparticles were deposited on a numerically marked silicon wafer as catalysts, and parameters of temperature, reduction time and carbon precursor were optimized. The crystallinity of the SWCNTs was characterized by Raman spectroscopy where the featured G/D peak intensity (I_G/I_D) was extracted automatically and mapped to the growth parameters to build a database. 1,280 data were collected to train machine learning models. Random forest regression (RFR) showed high precision in predicting the growth conditions for high-quality SWCNTs, as validated by further chemical vapor deposition (CVD) growth. This method shows great potential in structure-controlled growth of SWCNTs.

KEYWORDS

single-wall carbon nanotube, high throughput, machine learning, optimization, chemical vapor deposition

1 Introduction

Structure-controlled synthesis is essential to achieve the excellent physical and chemical properties of single-wall carbon nanotubes (SWCNTs) and to realize their practical applications [1]. This requires optimizing the growth conditions to controllably synthesize SWCNTs with targeted structures and properties. For the typical chemical vapor deposition (CVD) synthesis of SWCNTs, there are at least 12 growth parameters: composition and size of catalyst particles (2), composition and flow rate of carbon precursors (2), type and flow rate of carrier gases (2), promotor gases and flow rate (2), pretreatment and growth temperatures (2), pretreatment and growth time (2). If 10 conditions are tried for each variable, there will be a huge parameter space with 10^{12} possible combinations. The traditional trial-and-error process to find the optimum growth conditions is extremely time-consuming, because of the complex, non-linear, cross-related, high dimensional parameter space. Theoretical calculations and simulations have been used to understand the growth mechanism [2–4], but it is impractical to model the whole growth process of SWCNTs, with various reactions coupled with multiple times and length scales.

In recent years, big data-based information technology, especially machine learning, has become a new scientific paradigm and a powerful tool to tackle complex problems, and has been applied to accelerate the development of new materials [5]. To optimize the growth parameters of CNTs, combinatorial methods [6–9] and autonomous growth systems [10, 11] have been reported. Noda et al. reported the fabrication of catalyst libraries with continuously varying compositions to search for the optimum growth condition of SWCNT arrays [12–14]. Mirkin et al. combined polymer pen lithography and ink spray-coating to prepare nanoscale Au-Cu libraries and to screen the active compositions for SWCNT growth [15]. On the other hand, machine learning has been used to model CNT growth [16–18]. Nasibulin et al. trained an artificial neural network (ANN) to fit the growth conditions to the yield, diameter, and the intensity ratio of the graphitic G-band to disorder-related D-band (I_G/I_D) of produced SWCNT films [18]. Recently, Maruyama et al. reported an attractive concept of "robot scientist", where growth experiments could be conducted autonomously so that the structure could be analyzed and then fed back to machine

Address correspondence to Dai-Ming Tang, tang.daiming@nims.go.jp; Chang Liu, cliu@imr.ac.cn

learning algorithm for further optimization [11, 16, 19]. By combining microscale pillar arrays, *in-situ* Raman characterization, and automated control, over 100 experiments were conducted in a single day. Experimental conditions such as growth temperature and flow rate of gases were effectively screened for selective growth of single- and multi-wall CNTs [11] or maximizing CNT growth rates [16, 19].

In this study, we have developed a high-throughput method coupled with machine learning to optimize the growth of high-quality SWCNTs. Numerically indexed catalyst patterns were discretely deposited on silicon wafers to grow SWCNTs from nanoparticles (NPs) with different sizes. By Raman mapping and automatic I_G/I_D extraction, a database of 1,280 was generated, one order larger compared to the datasets in previous reports [11, 16, 18]. This larger database overcomes the data scale barrier and enables accurate machine learning and prediction of the correlation between the growth conditions and the quality of SWCNTs.

2 Results and discussion

Within the huge parameter space, apparently, it is impractical to search for an optimized condition using an exhaustive approach and it is necessary to start from the most influential variables. In a typical CVD growth process of SWCNTs, carbon precursors are catalytically decomposed on the surface of catalyst NPs. Carbon atoms are then dissolved and diffused in the catalyst NPs. After becoming oversaturated, SWCNTs are nucleated and grown from the NPs as growth seeds [20]. In this study, SWCNTs were grown using Co as the catalyst and ethanol as carbon precursor, where the size of the Co catalyst NPs, growth temperature, reduction time, and flow rate of carrier gas through ethanol were selected as variables for optimizing the quality of SWCNTs. The workflow is shown in Fig. 1, including (a) high throughput CVD growth from catalyst patterns with a gradient in thickness, (b) Raman characterization of SWCNTs to build a database, (c) modeling of the growth parameters and SWCNT crystalline quality by using machine learning, and (d) validation of the predicted optimum growth conditions by CVD growth on common Si wafers.

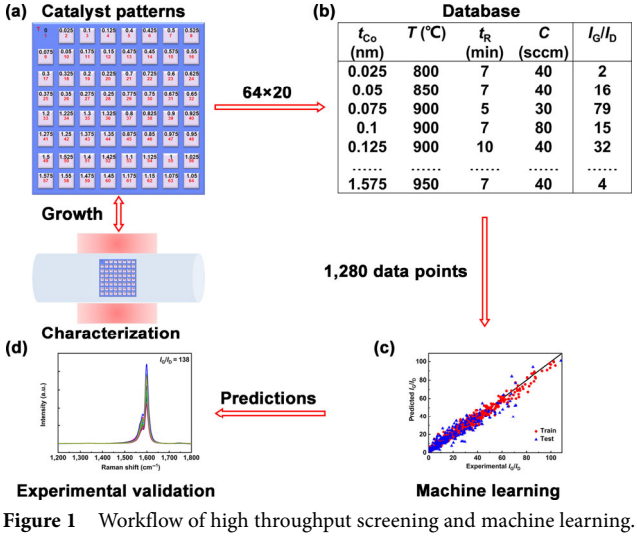

Figure 1 Workflow of high throughput screening and machine learning. (a) High throughput CVD growth of SWCNTs from numerically indexed catalyst patterns. The red numbers from 1 to 64 represent the marks on the silicon wafer. The black numbers ranging from 0 to 1.575 represent the deposited thicknesses of Co film (nm). (b) Construction of a database linking the growth parameters to the quality of the SWCNTs characterized by Raman (I_G/I_D) measurements. (c) Machine learning model with a plot of the experimental and predicted I_G/I_D. (d) Experimental validation of the predicted optimum growth conditions.

2.1 High-throughput growth

A quaternary combinatorial masking method [21] was used to prepare the discrete, numerically indexed Co catalyst patterns. A total of 64 catalyst patterns were generated in an array on a numerically marked silicon wafer by ion-beam sputtering deposition, and their nominal thicknesses were in the range of 0–1.575 nm with a step of 0.025 nm (Fig. S1 in the Electronic Supplementary Material (ESM)). After annealing under hydrogen atmosphere, the Co films were transformed to discrete NPs. Atomic force microscopy (AFM) images show that the sizes of catalyst NPs have a Gaussian distribution (Fig. S2 in the ESM), showing average sizes of 1.1, 2.2 and 4.2 nm for the Co films with nominal thicknesses of 0.025, 0.200 and 1.000 nm, respectively. During CVD growth, the growth temperature, reduction time and flow rate through ethanol were changed as variables. A scanning electron microscopy (SEM) image (Fig. S3(a) in the ESM) shows that the grown SWCNTs are confined within the catalyst patterns with their digital identification clearly resolved, so that the characterization results could be related with the growth conditions to construct a database. In a higher magnification SEM image (Fig. S3(b) in the ESM), we can see that the nanotubes form a uniform network. Transmission electron microscopy (TEM) images (Figs. S3(c)–S3(e) in the ESM) confirm that the SWCNTs possess high crystallinity with long and straight walls.

2.2 Raman characterization

Raman spectroscopy was used to characterize the quality of the grown SWCNTs by calculating I_G/I_D [22]. Figure 2(a) shows a map of G-band intensity, where the brightness indicates the density of the SWCNTs. The Raman mapping patterns are consistent with the SEM observations (Fig. S3(a) in the ESM). Three representative results were further analyzed, with the G-D Raman spectra shown in Figs. 2(b)–2(d) and corresponding SEM images in Figs. 2(e)–2(g). When the catalyst film thickness increased from 0.025 to 0.2 nm, the density of SWCNTs increased, and the average I_G/I_D was increased from ~ 8 to ~ 102 (Figs. 2(b) and 2(c)). When the Co film thickness was further increased to 1 nm (Fig. 2(d)), the SWCNT density decreased, and nanotubes were mainly observed along the edges. To avoid such "edge effects", average I_G/I_D from the inner area (0.6 mm × 0.6 mm) was mapped to the growth condition and used for further machine learning.

The Raman spectra in the radial breathing mode (RBM) region corresponding to Figs. 2(b)–2(d) are shown in Fig. S4 in the ESM. When the thicknesses of catalyst films are 0.025, 0.2, and 1 nm, the numbers of RBM peaks under the 532 nm laser are 26, 125, and 18, respectively. Therefore, there is an optimum Co catalyst thickness for the growth efficiency that is closely related to the number of observed RBM peaks. In addition, the positions of the RBM peaks down shifted to lower wavenumbers (ω_{RBM}), indicating an increase of tube diameter ($d \approx 248/\omega_{RBM}$) with the increase of catalyst film thickness. More detailed statistical analysis on the G-D and RBM peaks for a high throughput experiment (Fig. 2(a)) is shown in Fig. S5 in the ESM, where the I_G/I_D, RBM peak number, mean diameter and percentage of semiconducting SWCNTs (s-SWCNTs) are plotted against the thickness of Co film. The number of RBM peaks and I_G/I_D values have similar trends with the increase of thickness of Co film, indicating a higher yield of SWCNTs in the high-quality samples. With the increase of Co catalyst thickness, a general trend of increased tube

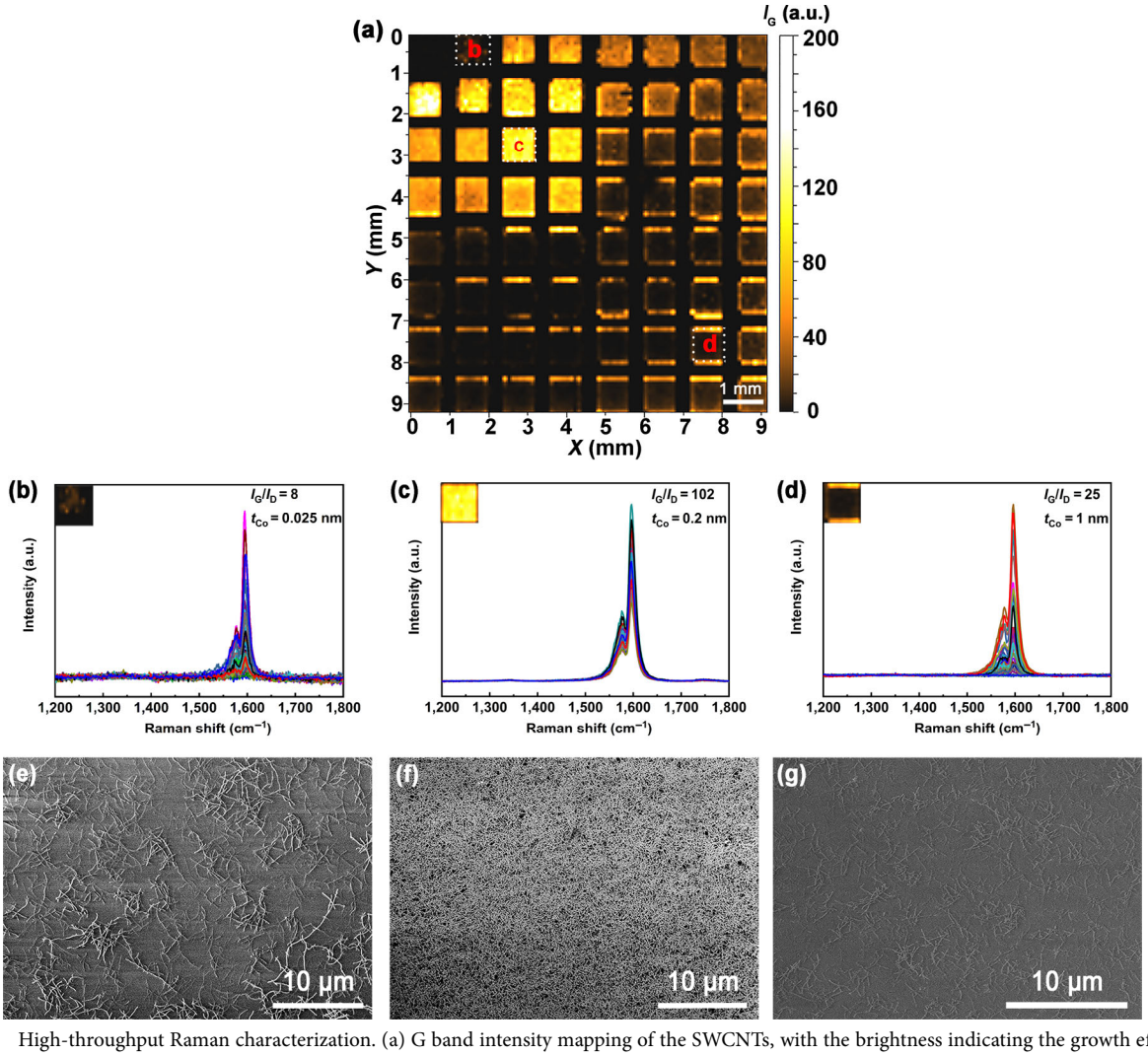

Figure 2 High-throughput Raman characterization. (a) G band intensity mapping of the SWCNTs, with the brightness indicating the growth efficiency. (b)–(d) G and D bands of the Raman spectra of the SWCNTs grown from Co catalysts with nominal thicknesses of 0.025, 0.2, and 1 nm. (e)–(g) SEM images of the SWCNTs corresponding to (b)–(d).

diameter was observed. In addition, the percentage of s-SWCNTs was found to be related to the thickness of Co film, revealing a fluctuating pattern. The observed dependence of RBM modes on the growth parameters shows the potential of applying our method in controlled growth of SWCNTs.

2.3 Modeling of SWCNT growth using machine learning

The features of the Raman spectra, i.e. average I_G/I_D, were then extracted automatically to relate to the growth conditions to construct a database. In total, 1,280 different growth conditions and corresponding Raman spectra were collected from 20 growth batches (Fig. S6 in the ESM). Statistical analysis of the growth parameters and measured I_G/I_D values are presented in Figs. S7 and S8 in the ESM. Growth parameters are set as follows: The nominal thicknesses of Co film are distributed uniformly from 0 to 1.575 nm, the growth temperature ranges from 800 to 950 °C, the reduction time from 3 to 10 min, and the flow rate through ethanol from 20 to 100 sccm (Fig. S7 in the ESM). Only ~ 8.3% and ~ 0.3% of the I_G/I_D values are higher than 60 and 100, respectively (Fig. S8 in the ESM). Such distribution shows the existence and sensibility of the optimum growth conditions for the high quality SWCNTs.

For modeling of SWCNT growth using machine learning, I_G/I_D was used as the output target and four growth conditions were set as descriptors, including the thickness of Co film, growth temperature, reduction time and flow rate through ethanol. Supervised linear regression (LR) and nonlinear regression models, such as support vector regression (SVR), random forest regression (RFR) and ANN were implemented comparatively to find the most suitable algorithm to predict the growth conditions for high-quality SWCNTs. 20% of the dataset was randomly withheld for testing and the other 80% was used for training. During the SVR, RFR and ANN models training, a grid search method was carried out to optimize the hyperparameters (Table S1 in the ESM). In addition, the robustness of the trained models was evaluated by 10-fold cross validation (CV) [23] on the training data.

The coefficient of determination (R^2), root-mean-square error (RMSE) and mean absolute error (MAE) were used as metrics to evaluate the performances. For an accurate model, R^2 should be close to 1, and RMSE and MAE should be small. Table 1 summarizes the performance of the trained machine learning models. Nonlinear regression models, SVR, RFR and ANN, show much higher accuracy in predicting I_G/I_D than the

Table 1 Performance of the machine learning models

Model	RMSE	MAE	R^2
LR	16.22	12.23	0.20
SVR	8.45	5.84	0.78
RFR	6.25	4.57	0.88
ANN	7.72	5.62	0.82

linear regression model indicating a non-linear relationship between the growth parameters and the I_G/I_D values of the SWCNTs. I_G/I_D can be best predicted using RFR with the highest R^2 value of 0.88 and the lowest RMSE and MAE values. And the RFR model achieved the highest CV score of 0.83 (Table S2 in the ESM), indicating that RFR is a suitable algorithm to model the relation of growth parameters and quality of SWCNTs.

Figure 3(a) shows the I_G/I_D values from experiments and from predictions using the RFR model. Most of the predicted data points are located close to the $y = x$ line, indicating that the general trend and relation of the growth parameters to the quality of the SWCNTs is captured by the model. The error for the higher I_G/I_D values (≥ 60) is larger, with the predicted values lower than the experimental ones. One possible reason is that there are fewer data points in the high I_G/I_D region (Fig. S8 in the ESM). One advantage of RFR is that the relative importance of input descriptors can be obtained by replacing each descriptor with random values and measuring the change of performance [24]. Figure 3(b) shows the calculated importance of the growth parameters for controlling the quality of the SWCNTs. The thickness of Co film appears to be the most important parameter followed by the growth temperature, flow rate through ethanol and reduction time. It seems that the growth time was not as influential as other parameters, as shown by the ranking of importance from the modeling including growth time as a variable (Figs. S9 and S10 in the ESM). Therefore, the growth conditions were optimized at a fixed growth time of 5 min.

Generally, the growth of CNTs involves (1) catalytic decomposition of carbon precursors, (2) dissolution and diffusion of carbon atoms, and (3) nucleation and growth of CNTs. All three steps are thermally activated; therefore, the temperature is important to provide sufficient thermal energy. The thermodynamical properties of carbon precursors determine the decomposition and supply of carbon atoms. Only when the thermodynamical and kinetical conditions are optimized, and the mass transport through the three stages is balanced, CNT growth with high yield and high quality occurs. Among the growth parameters, the role of the catalyst NPs is essential. In the 1st step, the catalytic activities affect reaction rate. In the 2nd step, the solubility and diffusion rate determine the carbon transport. In the 3rd step, the catalyst NPs are the templates for CNTs nucleation and growth. The central role of the catalyst is consistent with previous success of controlled synthesis of CNTs by catalyst design [25–27]. Such a sequence of importance is consistent with the prior knowledge about SWCNT growth, suggesting that the RFR model can indeed "learn" and capture the relations between the growth parameters and the I_G/I_D values rather than just memorizing and interpolating the training data.

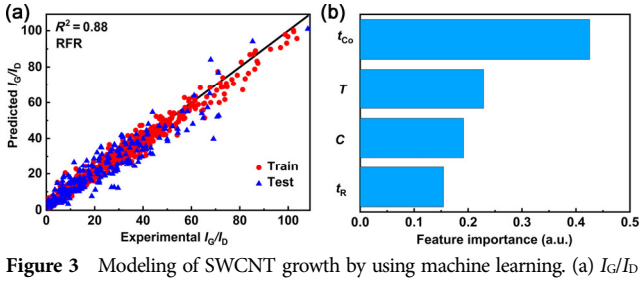

Figure 3 Modeling of SWCNT growth by using machine learning. (a) I_G/I_D values from experimental measurements and predictions from the RFR model. Red circles and blue triangles are predicted I_G/I_D values using the training and testing sets, respectively. (b) Ranking of the calculated importance of growth parameters for controlling the quality of SWCNTs. t_{Co}: thickness of Co film. T: growth temperature. C: flow rate through ethanol. t_R: reduction time.

The trained RFR model was used to further explore the parameter space for the optimum growth conditions for high-quality SWCNTs. I_G/I_D values were predicted using the RFR model for 38,016 combinations of growth parameter, as shown in Table S3 in the ESM. Figure 4 shows the predicted I_G/I_D values of SWCNTs with catalyst thickness as the primary variable and growth temperature, reduction time and flow rate through ethanol as secondary variables. For each combination of growth parameters, there is an optimal range ($I_G/I_D \geq 60$): 0.2–0.4 nm of Co film thickness, 880–920 °C of growth temperature, 5–8 min of reduction time, and 30–40 sccm of flow rate through ethanol.

2.4 Experimental validation

The optimum conditions for the growth of high quality SWCNTs were obtained from the predictions of the RFR model, as shown in Table S4 in the ESM. Such predictions were verified by further CVD growth. Figure S11 in the ESM shows the Raman spectra of the SWCNTs grown with the predicted growth parameters. All the average experimental I_G/I_D values are larger than 80, confirming the validity of the predictions for high quality SWCNTs. The average I_G/I_D values are in the range from ~ 83 to ~ 138, which is larger than the range of the predicted ones (96–101), indicating the sensitivity of the SWCNT quality to the specific experimental conditions. The measured and predicted I_G/I_D values are in agreement for 5 out of the 6 validation experiments, with the RMSE and MAE calculated to be ~ 10 and ~ 8, respectively. For the other predicted condition, the experimentally measured I_G/I_D value (~ 138) is substantially higher than the prediction (100). Thus there may be some overfitting of the RFR model to the training data, but it is general enough for the prediction of optimal conditions for high quality SWCNT growth.

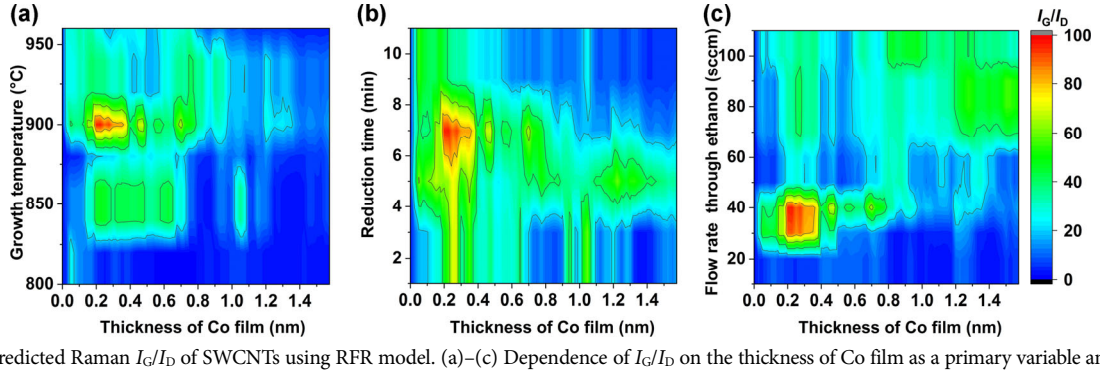

Figure 4 Predicted Raman I_G/I_D of SWCNTs using RFR model. (a)–(c) Dependence of I_G/I_D on the thickness of Co film as a primary variable and (a) growth temperature, (b) reduction time, and (c) flow rate through ethanol as secondary variables.

2.5 Enhanced optimization efficiency

By combining the high throughput method with machine learning, the efficiency of optimizing the growth conditions of high-quality SWCNTs has been dramatically improved (Table S5 in the ESM). We can see that the total time for a high-throughput experiment is around 4 h, where half of the time is spent on Raman mapping. Up to 4 experiments could be done within 1 day to get 256 growth conditions, and more than 1,000 growth conditions could be tested within a week. After collecting the data, the RFR model can be trained in a few minutes to obtain the optimum growth conditions.

3 Conclusions

We have developed a high-throughput method coupled with machine learning for the efficient optimization of the growth parameters of high-quality SWCNTs. The high throughput method includes the preparation of discrete catalyst patterns, CVD growth, Raman characterization and automatic I_G/I_D extraction. Machine learning was implemented to model the relationship of the growth parameters to the quality of the SWCNTs. By comparing typical regression models, the RFR model shows the highest performance with a coefficient of determination (R^2) value of 0.88. The predicted optimum growth parameters from machine learning were further validated to achieve a high I_G/I_D value of 138 in conventional CVD growth process. The efficient modeling, predicting, and learning ability of the combined high-throughput and machine learning method in this work shows the potential to speed up the controlled synthesis of SWCNTs with specific structures and properties.

4 Methods

4.1 Fabrication of numerically marked catalyst patterns

Three quaternary masks (Fig. S1(a) in the ESM) were used sequentially for depositing catalyst patterns of 64 (4^3) different thicknesses on a numerically marked Si/SiO_2 wafer (Fig. S1(b) in the ESM). An ion-beam sputtering (Gatan model 681) was used to deposit the catalyst films at room temperature with a deposition rate of ~ 0.3 nm/min. The averaged deposition rate was calculated by measuring the thickness of Co film from 10 min deposition. Figure S1(c) in the ESM schematically illustrates the masks (A, B and C) used to generate the Co catalyst patterns. The steps for depositing the Co catalyst patterns are as follows A_1: 0 nm; A_2: 0.4 nm; A_3: 0.8 nm; A_4: 1.2 nm; B_1: 0 nm; B_2: 0.1 nm; B_3: 0.2 nm; B_4: 0.3 nm; C_1: 0 nm; C_2: 0.025 nm; C_3: 0.05 nm; and C_4: 0.075 nm. As a result, an array of 64 numerically indexed Co catalyst patterns, each with a size of 0.8 mm × 0.8 mm, spaced 0.4 mm apart, was prepared on a marked silicon wafer (Fig. 1(a)).

4.2 Growth of SWCNTs

A quartz tube reactor with an inner diameter of 2.5 cm was used to grow the SWCNTs. To activate the catalyst, the Co catalyst patterns were annealed at 500 °C in air for 10 min. After the furnace was heated to the growth temperature of CNTs under argon (Ar) flow, the Ar was turned off and then the catalyst film was reduced under 200 sccm H_2 to form NPs. Afterwards, a flow of Ar carrier gas through an ethanol bubbler (in a 30 °C water bath) as the carbon source and 100 sccm H_2 were introduced to grow SWCNTs for 5 min. The furnace was then cooled to room temperature under Ar flow.

4.3 Characterization

The size distribution of the catalyst NPs after reduction was characterized by an AFM (Bruker MultiMode 8-HR) operated in the tapping mode. The morphology of the SWCNTs was characterized by SEM (Nova Nano SEM 430 and Verios G4 UC) with an acceleration voltage of 1 kV. The structure of the SWCNTs was characterized by a TEM (FEI Tecnai F20). The quality of SWCNTs was characterized by a Raman spectrometer (Jobin Yvon HR800), excited by a 532 nm He-Ne laser with spot size of ~1 μm^2, and a mapping step of 0.1 mm. The Raman spectra in the RBM region were collected under both the 532 and 633 nm lasers. The Raman data were cleaned by normalization, removal of universal noise and subtraction of the baseline, using LabSpec 5 software from Horiba. An Excel template was designed to automatically extract the averaged intensities of the D (I_D) and G bands (I_G) from each patterned catalyst area, and to construct a database mapping the growth parameters to the I_G/I_D values.

4.4 Machine learning model

Machine learning was used to model and predict the crystallinity of the grown SWCNTs under a combination of growth parameters. The learning process is summarized as follows:

1. Dataset: collect data from 1,280 growth conditions and Raman spectra.

2. Modelling: map growth conditions to I_G/I_D, by training and testing.

3. Prediction: calculate I_G/I_D of ~ 10^4 new conditions using the trained model.

4. Validation: CVD growth of SWCNTs under predicted optimum conditions and measurement of the I_G/I_D by Raman spectroscopy.

Python's scikit-learn package [28] (version: 0.22.1) and Keras API [29] (version: 2.3.1) were used to build the machine learning models. A min-max scaler was used to normalize the input parameter values to the (0,1) range during the construction of the LR, SVR and ANN models.

ANN is a parallel interconnected network which can map the connection of the features to output target in the network and calculate the prediction under new conditions [30]. The hidden layers of the ANN model were optimized by a trial-and-error method and set to 3 hidden layers with 512, 256, 128 neurons, respectively. Relu (Rectified linear unit) [31] and Adam [32] were used as activation function and optimizer, respectively. After building the ANN architecture, the batch size and number of epochs were adjusted to improve the predictions. By using a grid search method, the optimum batch size and number of epochs (Table S1 in the ESM) were 500 and 3,000, respectively.

Support vector machine (SVM) uses a kernel function to project input data onto a higher dimensional space and then calculate a hyperplane to classify training data with largest margin between the classes for classification problems or fit training data to the hyperplane for regression problems [33]. In this work, Gaussian radial basis function (RBF) was used as the kernel function. Two user dependent parameters, C and γ, were tuned to improve the predictions. By using a grid search method, the optimal C and γ (Table S1 in the ESM) were 10 and 100, respectively.

Random forest (RF) is an ensemble of classification or regression trees through bootstrap sampling and random feature selection [24, 34]. The prediction is determined by majority vote for classification or averaging for regression, of the predictions from all trees in the forest. To simplify the RFR

model, only the number of trees was adjusted while other hyper parameters were used as follows: criterion = "mse", max_depth = None, max_features = "auto", min_samples_leaf = 1, min_samples_split = 2. By using a grid search method, the optimized RFR model was constructed by an ensemble of 800 regression trees (Table S1 in the ESM).

Acknowledgements

The authors thank Zexin Tian and Jianqi Huang for help with training the artificial neural network model, and Shemon Baptiste and Hao-Wei "Ric" Tu for repeating and confirming the model predictions. The authors also thank Hui Li for help with AFM characterization. This project is supported by the National Key Research and Development Program of China (No. 2016YFA0200101), the National Natural Science Foundation of China (Nos. 51522210, 51972311, 51625203, 51532008, 51761135122 and 52001322), JSPS KAKENHI Grant Number JP20K05281 and JP25820336, and MOST 108-2634-F-006-009 and MOST 109-2224-E-006-003.

References

[1] Rao, R.; Pint, C. L.; Islam, A. E.; Weatherup, R. S.; Hofmann, S.; Meshot, E. R.; Wu, F. Q.; Zhou, C. W.; Dee, N.; Amama, P. B. et al. Carbon nanotubes and related nanomaterials: Critical advances and challenges for synthesis toward mainstream commercial applications. *ACS Nano* **2018**, *12*, 11756–11784.

[2] Ding, F.; Rosén, A.; Bolton, K. Dependence of SWNT growth mechanism on temperature and catalyst particle size: Bulk versus surface diffusion. *Carbon* **2005**, *43*, 2215–2217.

[3] Ding, F.; Bolton, K.; Rosén, A. Molecular dynamics study of SWNT growth on catalyst particles without temperature gradients. *Comput. Mater. Sci.* **2006**, *35*, 243–246.

[4] Xu, Z. W.; Yan, T. Y.; Ding, F. Atomistic simulation of the growth of defect-free carbon nanotubes. *Chem. Sci.* **2015**, *6*, 4704–4711.

[5] Agrawal, A.; Choudhary, A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Mater.* **2016**, *4*, 053208.

[6] Kind, H.; Bonard, J. M.; Emmenegger, C.; Nilsson, L. O.; Hernadi, K.; Maillard-Schaller, E.; Schlapbach, L.; Forró, L.; Kern, K. Patterned films of nanotubes using microcontact printing of catalysts. *Adv. Mater.* **1999**, *11*, 1285–1289.

[7] Cassell, A. M.; Verma, S.; Delzeit, L.; Meyyappan, M.; Han, J. Combinatorial optimization of heterogeneous catalysts used in the growth of carbon nanotubes. *Langmuir* **2001**, *17*, 260–264.

[8] Cassell, A. M.; Ye, Q.; Cruden, B. A.; Li, J.; Sarrazin, P. C.; Ng, H. T.; Han, J.; Meyyappan, M. Combinatorial chips for optimizing the growth and integration of carbon nanofibre based devices. *Nanotechnology* **2003**, *15*, 9.

[9] Noda, S.; Tsuji, Y.; Murakami, Y.; Maruyama, S. Combinatorial method to prepare metal nanoparticles that catalyze the growth of single-walled carbon nanotubes. *Appl. Phys. Lett.* **2005**, *86*, 173106.

[10] Oliver, C. R.; Westrick, W.; Koehler, J.; Brieland-Shoultz, A.; Anagnostopoulos-Politis, I.; Cruz-Gonzalez, T.; Hart, A. J. Robofurnace: A semi-automated laboratory chemical vapor deposition system for high-throughput nanomaterial synthesis and process discovery. *Rev. Sci. Instrum.* **2013**, *84*, 115105.

[11] Nikolaev, P.; Hooper, D.; Perea-Lopez, N.; Terrones, M.; Maruyama, B. Discovery of wall-selective carbon nanotube growth conditions via automated experimentation. *ACS Nano* **2014**, *8*, 10214–10222.

[12] Sugime, H.; Sato, T.; Nakagawa, R.; Cepek, C.; Noda, S. Gd-enhanced growth of multi-millimeter-tall forests of single-wall carbon nanotubes. *ACS Nano* **2019**, *13*, 13208–13216.

[13] Hasegawa, K.; Noda, S. Millimeter-tall single-walled carbon nanotubes rapidly grown with and without water. *ACS Nano* **2011**, *5*, 975–984.

[14] Chen, Z. M.; Kim, D. Y.; Hasegawa, K.; Noda, S. Methane-assisted chemical vapor deposition yielding millimeter-tall single-wall carbon nanotubes of smaller diameter. *ACS Nano* **2013**, *7*, 6719–6728.

[15] Kluender, E. J.; Hedrick, J. L.; Brown, K. A.; Rao, R.; Meckes, B.; Du, J. S.; Moreau, L. M.; Maruyama, B.; Mirkin, C. A. Catalyst discovery through megalibraries of nanomaterials. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 40–45.

[16] Nikolaev, P.; Hooper, D.; Webber, F.; Rao, R.; Decker, K.; Krein, M.; Poleski, J.; Barto, R.; Maruyama, B. Autonomy in materials research: A case study in carbon nanotube growth. *npj Comput. Mater.* **2016**, *2*, 16031.

[17] Abad, S. N. K.; Ganjeh, E.; Zolriasatein, A.; Shabani-Nia, F.; Siadati, M. H. Predicting carbon nanotube diameter using artificial neural network along with characterization and field emission measurement. *Iran. J. Sci. Technol. Trans. A: Sci.* **2017**, *41*, 151–163.

[18] Iakovlev, V. Y.; Krasnikov, D. V.; Khabushev, E. M.; Kolodiazhnaia, J. V.; Nasibulin, A. G. Artificial neural network for predictive synthesis of single-walled carbon nanotubes by aerosol CVD method. *Carbon* **2019**, *153*, 100–103.

[19] Chang, J.; Nikolaev, P.; Carpena-Núñez, J.; Rao, R.; Decker, K.; Islam, A. E.; Kim, J.; Pitt, M. A.; Myung, J. I.; Maruyama, B. Efficient closed-loop maximization of carbon nanotube growth rate using bayesian optimization. *Sci. Rep.* **2020**, *10*, 9040.

[20] Wang, H.; Yuan, Y.; Wei, L.; Goh, K.; Yu, D. S.; Chen, Y. Catalysts for chirality selective synthesis of single-walled carbon nanotubes. *Carbon* **2015**, *81*, 1–19.

[21] Wang, J. S.; Yoo, Y.; Gao, C.; Takeuchi, I.; Sun, X. D.; Chang, H.; Xiang, X. D.; Schultz, P. G. Identification of a blue photoluminescent composite material from a combinatorial library. *Science* **1998**, *279*, 1712–1714.

[22] Dresselhaus, M. S.; Jorio, A.; Filho, A. G. S.; Saito, R. Defect characterization in graphene and carbon nanotubes using Raman spectroscopy. *Philos. Trans. Royal Soc. A: Math. Phys. Eng. Sci.* **2010**, *368*, 5355–5377.

[23] Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 111–133.

[24] Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

[25] Yang, F.; Wang, X.; Zhang, D. Q.; Yang, J.; Luo, D.; Xu, Z. W.; Wei, J. K.; Wang, J. Q.; Xu, Z.; Peng, F. et al. Chirality-specific growth of single-walled carbon nanotubes on solid alloy catalysts. *Nature* **2014**, *510*, 522–524.

[26] Zhang, F.; Hou, P. X.; Liu, C.; Wang, B. W.; Jiang, H.; Chen, M. L.; Sun, D. M.; Li, J. C.; Cong, H. T.; Kauppinen, E. I. et al. Growth of semiconducting single-wall carbon nanotubes with a narrow band-gap distribution. *Nat. Commun.* **2016**, *7*, 11160.

[27] Zhang, S. C.; Kang, L. X.; Wang, X.; Tong, L. M.; Yang, L. W.; Wang, Z. Q.; Qi, K.; Deng, S. B.; Li, Q. W.; Bai, X. D. et al. Arrays of horizontal carbon nanotubes of controlled chirality grown using designed catalysts. *Nature* **2017**, *543*, 234–238.

[28] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R. Dubourg, V. et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

[29] Chollet, F. *Keras: The Python Deep Learning library*; Astrophysics Source Code Library, 2018. https://keras.io/ (accessed Nov 9, 2020).

[30] Kohonen, T. An introduction to neural computing. *Neural Netw.* **1988**, *1*, 3–16.

[31] Nair, V.; Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, 2010, pp 807–814.

[32] Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. 2014, arXiv:1412.6980. arXiv.org e-Print archive. https://arxiv.org/abs/1412.6980v1 (accessed Dec 22, 2014).

[33] Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297.

[34] Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.; Sheridan, R. P.; Feuston, B. P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947–1958.