

Chemical Reaction Spectrum

Jianchao Lee,* Jianghong Li, Qiannan Duan, Sifan Bi, Ruen Luo, Yachao Lian, Ruixing Tian, Jiayuan Chen, Guodong Ma, Fenli Liu, Yunjin Feng

Lab of Aqueous Chemistry, Department of Environment Science, Shaanxi Normal University, Xi'an 710062, China

*Corresponding author, e-mail: jianchaolee@snnu.edu.cn

Abstract: We proposed a new method of chemical reaction spectrum (CRS) in terms of chemical characterization, and established a method to fulfill it by combining with 3D chemical printing technology and 2D sampling. The CRS can provide a graphical data set for pure or mixed substances, which can comprehensively describe the reaction characteristics of the research object. Compared with common characterization methods (NMR, UV/vis, IR, Raman, GC or LC), it is more capable of revealing chemical behaviors enough, and is much lower in cost. It is expected to be an important data acquisition approach for the application of artificial intelligence in the field of chemistry in the future.

Key word: Chemical Reaction, Machine Learning, Image, Big Data, Experimentation.

There are numerous natural or synthetic chemicals, being as pure substances or mixtures. The scientific aim of exploring them mainly involves the following two aspects. One is the exploration of the intrinsic characteristics, such as composition analysis, spectral characteristics, etc. (hereinafter referred to as the first type of work, I-work). The other is the behaviors on the chemical reactions with many other reactants (the second type of work,

II-work). The chemical science is aimed at exploring the interaction between substances at the levels not lower than molecular or atom, so the knowledge obtained from the latter type of work is closer to the core purpose of chemical research.

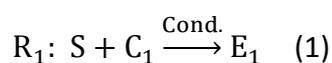
Due to the workload, for most new substances or new mixture systems, scientists actually have no energy to provide a large amount of information in II-work. Consequently, we often have to provide some information related to these substances via some methods derived from physical means, such as NMR, UV/vis, IR, Raman, GC or LC, etc[1]. Due to the lack of a large amount of information of the real chemical reactions, researchers don't have enough knowledge to predict accurately the behaviors in most complex reaction processes.

In recent years, artificial intelligence technology (AI) has developed rapidly, and it has shown excellent performance in many fields[2-5]. It may also lead to a new round of scientific revolution in the natural sciences. The results of research in the field of chemistry in the past several years are also appearing in considerable numbers[6-9]. We have found a feature that most of the research work uses the data of the composition and structural properties to carry out deep prediction and analysis of chemical reaction characteristics, showing extremely high efficiency.

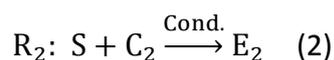
However, in this research model, the existing standardized data is limited in fact. The available data resources are easily exhausted and do not support the explosive growth of this big data research method. Some work began to focus on using the information of II-work. Lately, Science journal reported that some researchers even processed the large number of failed or negative data accumulated in the history of the laboratories to find complex features with the aid of machine learning (ML). The results show that the real reaction data has a startling performance in prediction. In fact, it is clear that such reaction data has been accumulated by conventional experiments for many years, and the number is still too small to be used in further play. This also hints from another aspect that if the experiment data of II-work

can be increased in a large amount, AI will play an important role in chemical research. After all, the complexity of chemical phenomena, an essential barrier on the road of the chemical research in the past or next decades, is very consistent with the superiority of AI. It is foreseeable that after a decade of development, chemical research will enter a new era.

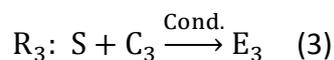
In traditional research, in order to obtain information on a certain type of chemical reaction R_1 of the researched object, S , the researcher to construct a reaction system to realize the interaction between S and other reactants, C , and then monitor the results, E . As follow,



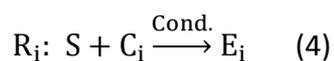
More types of chemical reaction information are obtained by changing the reactant C involved.



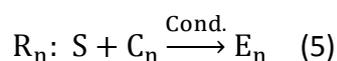
Akin,



...



...



When n is larger, the reaction features of S becomes clearer. This is a repetitive work. Commercial high-throughput facilities and experimental methods allow researchers to quickly conduct hundreds of reactions simultaneously[10-12]. But the exorbitant price and tedious operation destined a rare appearance in most laboratories.

We tried to modify the operating modes of the reactions R_1, R_2, \dots, R_n , and arranged the reactants C_i in a convenient spatial distribution (as shown in Fig. 1). Under optimal conditions, each point has its own specific composition.

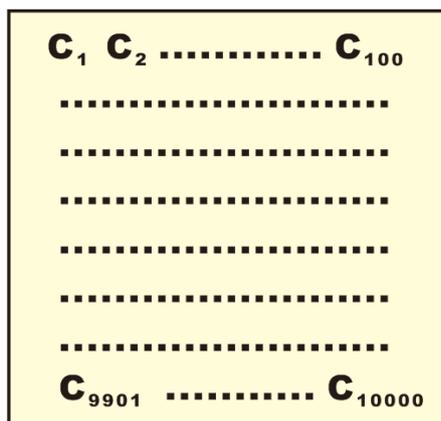


Fig. 1. Schematic of a two-dimensional distribution of the reactants combinations.

When the number of C_i reaches 10^5 or even 10^6 , it will give S a huge information of reaction behaviors. On considering the major categories of C_i reactions and the application of ordinary chemists, it is better to establish several standard groups for the typical C_i on behalf of a wide range of reaction relationships. But how to conveniently realize the structure in Fig. 1, and complete the detection of E_i accordingly is still very creative job.

Below, we present a means to realize the idea, which took full advantage of 3D chemical printing[13, 14], and chemical imaging detection[15].

We use the chemical printing method to complete the construction of the physical structure of C chemicals[15]. First, a print template as shown in Fig. 2A is designed, and each template channel corresponds to a chemical (or a certain mixture) with a 2D distribution. As a preliminary attempt, we selected 8 typical compounds (citric acid, potash, ferric sulfate, copper nitrate, aluminium chloride, bromocresol green, sodium diethyldithiocarbamate, potassium thiocyanate, bromothymol blue) and prepared their solutions (100mg/5ml as working concentration) as printing inks. Using jet printing technology, we build a super C_i array, referring to 3×10^8 ratios of the eight substances. Especially, the last 3 chromogenic compounds play not only roles of the reactants, but also a role in visualizing the reaction results. On the other hand, in the field of analytical chemistry, the synthesis of numerous color (or fluorescence) indicators have accumulated lots of knowledge and products. They are

contributing a treasure trove of creating many novel MCRP detectors. The right part of Fig. 2C is a prepared Ci array on a plate, called in our lab as "million-level combination reactions test paper", or MCRP. The right part of Fig. 2B is a 32-MCRPs batched-plate.

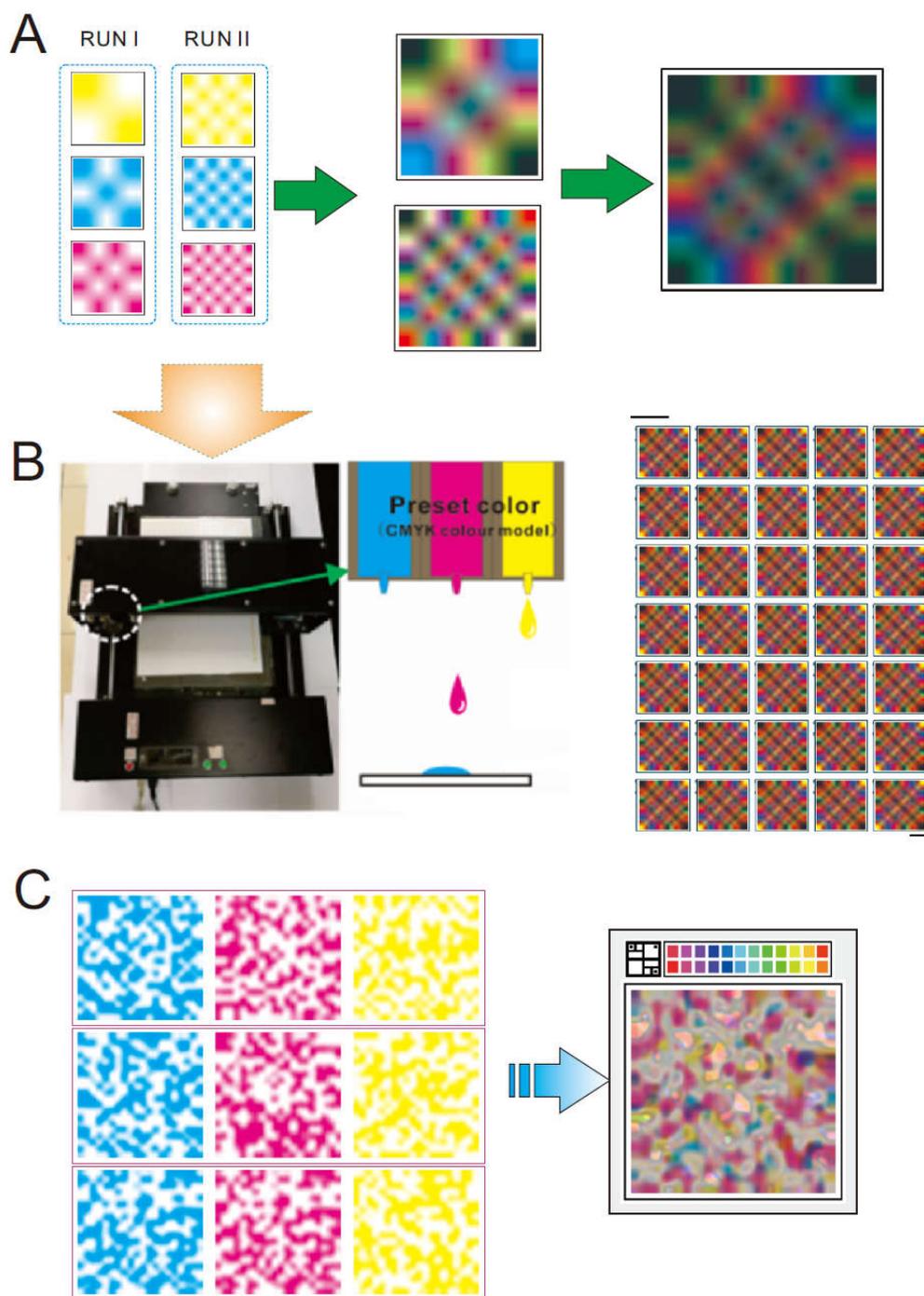


Fig. 2. Preparation technology and method of MCRP. (A) A printing template used to distribute the various reactants. (B) schematic diagram of the distribution of reactants using a modified ink-jet printer. (C) another printing template designed to constructing multidimensional reactant distribution and a sample of the prepared MCRP.

For an analyst to be tested, we completed the reaction of the sample and MCRP in a mode of quantitative spraying[15]. As shown in Fig. 3A, an ultrasonic nebulizer evenly distributed a certain amount of liquid sample on a MCRP, and the microscopic droplets are in state of independence to some extent (schematic in Fig. 3A). The device costs about \$10, and if the camera system is converted into a cellphone camera, it will cost less than \$5. After capturing images and deducting the background, we immediately obtained a complex image derived from the reaction results (Fig. 3B), which we call a chemical reaction spectrum (CRS). This image consisted of the same number of 1 million data points as the MCRP. Each data point has a known ratio of all reactants.

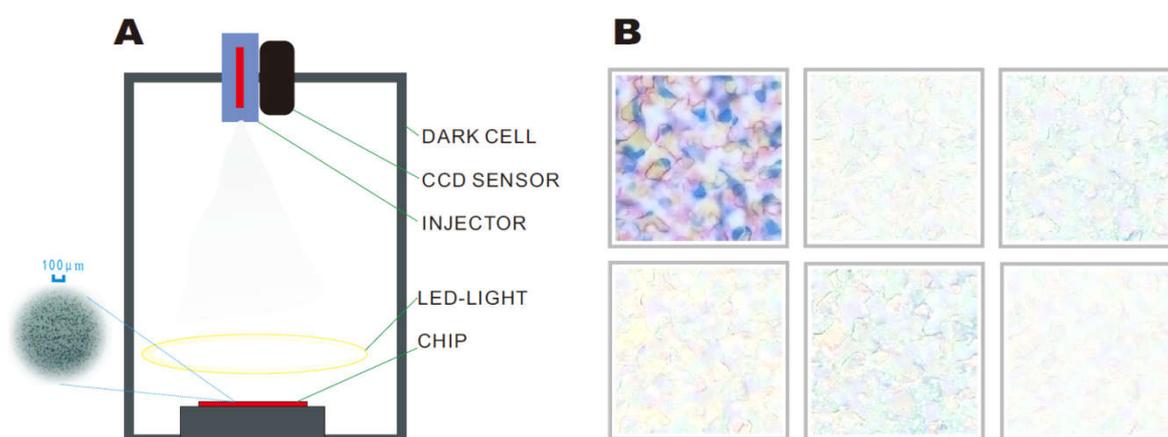


Fig. 3. The way to quantitatively interact the analyte with MCRP. (A) An ultrasonic atomizing device is used to realize quantitative two-dimensional sample addition, and a CCD camera device is used to obtain CRS images. (B) Some captured images after reaction, and the corresponding data of the CRS image obtained through background deduction. In order, blank MCRP, the CRS of nickel chloride, 1:2 mixture of manganese chloride and nickel chloride, disodium hydrogen phosphate, boric acid, and urine.

It can be seen from the selected ink types and their combination that in the above case, numerous chemical reactions and their quantitative fluctuation are included, providing a wealth of reaction information of the sample, S. In addition, the above case was an initial trial for our idea. Later, researchers can screen some more representative inks and increase the ink species to optimize the performance of CRS.

We give some discussion below. For all chemical reactions of a particular substance, researchers should not look at each reaction in isolation. In fact, a specific reaction and other reactions are inextricably linked. From a perspective of functions in algebra, we can even map a specific reaction with a limited number of other reaction processes. And this mapping process or algorithm actually needs to depend on the development of artificial intelligence and machine learning (ML). Nowadays, many ML applications of chemistry have shown that after getting a large amount of reaction information, researchers can give a very good prediction of other reactions to the object of study.

Let's talk about it, why on a MCRP, there are more than 1 million reactions. First, consider on the combination of substances. The printing templates (Fig. 2A and 2C) guarantee the ink mixture at all ratios. It is assumed that the amount range of each ink is divided into ten levels (, in an actual template and printing process, the levels are near 100). Then 8 inks will give out 8^{10} (~ 1 billion) ratios. Considering printing resolution and size, the different ratios of reactants are about 1 million. In the incipient case of the above work, the types of chemical reactions involved at least precipitation/dissolution, complexation, ion exchange, oxidation/reduction, adsorption/desorption and their combination at different levels. At a later stage, the researcher could do a lot of research to establish a more appropriate ink composition to adapt to a wider range of chemical phenomena, even biochemical reactions, organic reactions, etc. The external form of these higher order combinations is extremely complex. Obviously, we can improve the CRS to discover nearly whole chemical behaviors of any concerned substance by only selecting inks and reacting conditions. The fabrication of MCRPs can also be done on a large scale by means of screen printing, etc. This is also the work to be carried out later. We expect a dozen standard MCRPs confirmed after adequate test, and each of them is worthy of any of familiar analysis methods, such as IR, UV/vis, HPLC.

Such a method of obtaining large data of chemical reactions of a substance is called "chemical reaction spectrum", CRS. In our lab, CRS, as image data, is often used for deep neural network learning, so we also call it chemical neural network, ChemNN. After the standardization of the MCRP preparation, the CRS is almost equipment-free, low-cost and convenient for the application to individual users. It can cover almost any small laboratory or even student operations in middle schools. Such a chemical experiment with trace reagents and big data is a typical green method in the future chemistry.

In particular, we would like to mention that this technique can be used not only for pure substances, but also for most mixtures including urine, saliva, fluids, wastewater, etc. These applications have the potential to build bridges between some engineering technologies and AI.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.50309011) and the Scientific Research Foundation for the Returned Overseas Chinese Scholars (08501041585). All the data that support the results of this study are available from the corresponding author.

References

- [1] D.A. Skoog, F.J. Holler, S.R. Crouch, Principles of instrumental analysis, Cengage learning 2017.
- [2] D.S. Kermany, M. Goldbaum, W. Cai, C.C.S. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M.K. Prasadha, J. Pei, M. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V.A.N. Huu, C. Wen, E.D. Zhang, C.L. Zhang, O. Li, X. Wang, M.A. Singer, X. Sun, J. Xu, A. Tafreshi, M.A. Lewis, H. Xia, K. Zhang, Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning, *Cell*, 172 (2018) 1122-1131.e1129.
- [3] M.I. Jordan, T.M. Mitchell, Machine learning: Trends, perspectives, and prospects, *Science*, 349 (2015) 255-260.
- [4] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, 521 (2015) 436-444.
- [5] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis, Mastering the game of Go without human knowledge, *Nature*, 550 (2017) 354-359.

- [6] J.M. Granda, L. Donina, V. Dragone, D.-L. Long, L. Cronin, Controlling an organic synthesis robot with machine learning to search for new reactivity, *Nature* 559 (2018) 377-381.
- [7] P.V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning, *Nature communications*, 9 (2018) 1668.
- [8] A. Keller, R.C. Gerkin, Y. Guan, A. Dhurandhar, G. Turu, B. Szalai, J.D. Mainland, Y. Ihara, C.W. Yu, R. Wolfinger, C. Vens, L. Schietgat, K. De Grave, R. Norel, G. Stolovitzky, G.A. Cecchi, L.B. Vosshall, P. Meyer, Predicting human olfactory perception from chemical features of odor molecules, *Science*, 355 (2017) 820-826.
- [9] Q. Zhou, P. Tang, S. Liu, J. Pan, Q. Yan, S.C. Zhang, Learning atoms for materials discovery, *Proceedings of the National Academy of Sciences of the United States of America*, (2018).
- [10] D. Barata, C. van Blitterswijk, P. Habibovic, High-throughput screening approaches and combinatorial development of biomaterials using microfluidics, *Acta Biomaterialia*, 34 (2016) 1-20.
- [11] D.W. Robbins, J.F. Hartwig, A Simple, Multidimensional Approach to High-Throughput Discovery of Catalytic Reactions, *Science*, 333 (2011) 1423-1427.
- [12] R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm, H. Lam, Combinatorial and high-throughput screening of materials libraries: Review of state of the art, *ACS combinatorial science*, 13 (2011) 579-633.
- [13] X. Yan, Y. Zheng, J. Gao, J. Lee, An Ink-jet Printing Strategy for Extensive Exploration of One Chemical Action with Three Interactive Variations, *Analytical Sciences*, 33 (2017) 1-3.
- [14] L. Wang, J. Lee, M. Zhang, Q. Duan, J. Zhang, H. Qi, Fluorescence imaging technology (FI) for high-throughput screening of selenide-modified nano-TiO₂ catalysts, *Chemical Communications*, 52 (2016) 2944-2947.
- [15] Y. Zheng, J. Lee, Q. Duan, H. Chen, X. Yan, J. Zhang, L. Wang, A Novel Encoded Recording Strategy of Complex Chemical System, *Chemistry Letters*, 46 (2017) 360-363.