Article
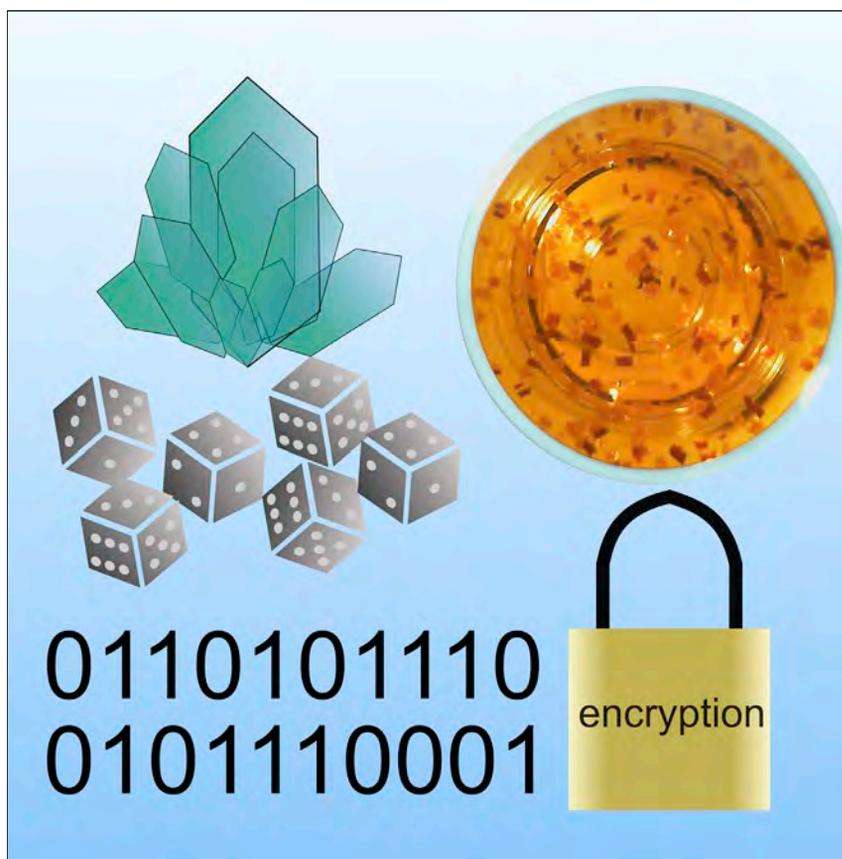
A Crystallization Robot for Generating True Random Numbers Based on Stochastic Chemical Processes

Crystallization of compounds can be used as an entropy pool in the generation of random numbers, which have applications in data encryption, and to investigate stochastic chemical processes. Automation of chemical reactions and crystal detection has enabled the collection and processing of the required large amounts of data on crystal growth and formation to allow production of such numbers and investigations.

Edward C. Lee, Juan M. Parrilla-Gutierrez, Alon Henson, Euan K. Brechin, Leroy Cronin

lee.cronin@glasgow.ac.uk

HIGHLIGHTS

Automated system for obtaining statistical data on crystal formation

First example of generation of true random numbers using stochasticity of chemistry

Encryption capability better than Mersenne Twister pseudorandom number generator

Analysis of entropy in different crystallizing compounds

2 | **Benchmark**
First qualification/assessment of material properties and/or performance

Article

A Crystallization Robot for Generating True Random Numbers Based on Stochastic Chemical Processes

Edward C. Lee,[1] Juan M. Parrilla-Gutierrez,[1] Alon Henson,[1] Euan K. Brechin,[2] and Leroy Cronin[1,3,*]

SUMMARY

Chemistry inherently involves a wide range of stochastic processes, yet chemists do not typically explore stochastic processes at the macroscale due to the difficulty in gathering data. We wondered whether it was possible to explore such processes, in this case crystallization, in a systematic way using an autonomous robotic platform. By performing inorganic reactions in an automated system, and observing the resultant occupied macrostate (crystallization images), we developed a powerful entropy source for generation of true random numbers. Randomness was confirmed using tests described by the National Institute for Standards and Technology ($p_{uniformity} \gg 0.0001$). Deficit from maximum approximate entropy was found be different between compounds ($p_{ANOVA} \ll 0.01$), and encryption security of these numbers was found to be greater than that of a frequently used pseudorandom number generator. This means that we can now use random number generation as a probe of the stochastic process, as well as explore potential real-world applications.

INTRODUCTION

Recently the reproducibility and bias in chemical reaction data have been discussed.[1] This is because machine learning in chemical and materials synthesis will require large amounts of reliable data, but also so that algorithms can be validated.[2] The recent development of automated platforms for chemistry are not only transformative for chemical synthesis[3,4] and discovery,[5,6] but also for exploration of data reliability. Another important aspect of exploring the extent to which experimental data are reproducible is the fact that many processes are intrinsically stochastic. Such stochasticity can also be useful, such as in the generation of random numbers.

Random numbers are used extensively in many applications, such as cryptography[7] and scientific modeling,[8] where their non-deterministic properties and unpredictability are essential. The source of randomness (i.e., the entropy source) is a crucial factor for whether correlations between generated numbers are present, and this has a great impact on the randomness and utility of the output.[9] As such, random number generation not only can help us understand the reproducibility of a process, but also are more desirable than computational methods of number generation (e.g., pseudorandom number generators). This is because they extract their randomness from a physical system with a large available pool of entropy.[10,11] Importantly, the generation of random outcomes is of profound importance as a source of noise, and potentially allows the generation of unanticipated data.

Progress and Potential

As chemistry and materials synthesis is starting to embrace an era of automation and the use of machine learning, it is becoming vital that the quality and reliability of that data is assessed. By automating and parallelizing batch chemical reactions, enough samples may be run that statistical data can be obtained on the reaction system. We monitored the crystallization for hundreds of parallel reactions using a webcam and found that crystal features in the images obtained could be used to generate true random numbers. We also found that the approximate entropy of these numbers was different for different types of chemical reaction, and that the encryption capability of these numbers was greater than a commonly used pseudorandom number generator. This is the first time that stochasticity of chemistry has been investigated in large datasets from experimental data.

Matter
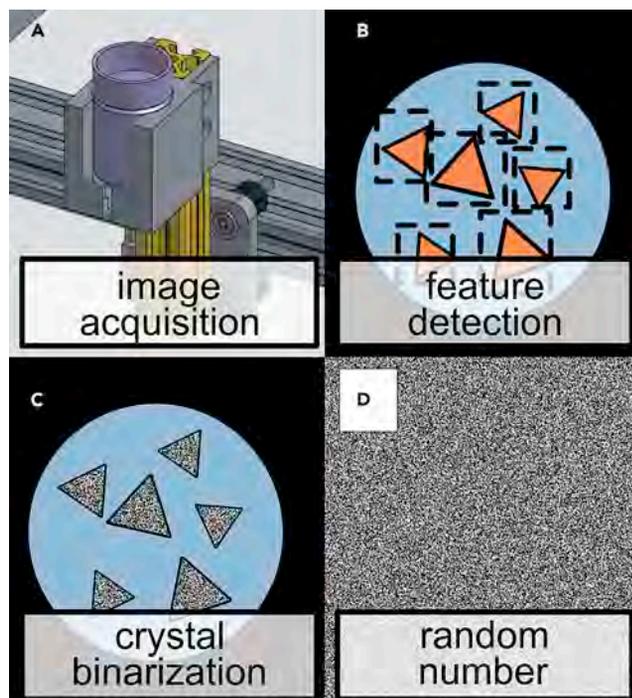
Figure 1. Schematic of Procedure for Generating Random Numbers Using Crystallization

(A) Images are acquired using a mobile tracking camera attached to the underside of the platform.

(B) Feature-detection and image-segmentation algorithms locate the pixels corresponding to crystals and the crystallization vial.

(C) Application of a binarization algorithm converts the feature-detection data in (B) to a sequence of 0s and 1s.

(D) Sequences from subsequent binarization applications are joined to form a longer sequence, suitable for randomness testing.

In a chemical system, each time a reaction is performed there is an almost infinite number of energetically equivalent ways for particular reagents to combine, resulting in both high uncertainty and entropy, and the exact pathway undertaken will never be repeated. The overall outcome of such a reaction is therefore an example of one specific state out of an almost infinite number of possible macrostates.[12] In chemistry, the energies of different configurations of molecules under thermodynamic control can be considered as part of a canonical ensemble, in which an almost infinite number of energetically degenerate states can be accessed, and of which only one macrostate will be occupied during a particular observation.[13] As such, the entropy of such a chemical system is extraordinarily high,[14] and may therefore serve as a very good entropy pool for application of random number generation.

In this regard, we hypothesized that one such system with a large entropy pool is that of compound formation and subsequent crystallization, where the detectable ensemble macrostates considered are the locations and morphologies of each crystal that has grown in a period of time as a result of these processes. To explore this idea, we set out to develop a fully automated system to not only do the chemical reactions but also grow crystals of the products using a camera as a detector. The platform was designed so that it converts these data into binary sequences, as shown schematically in Figure 1, which are assessed for randomness using the methods specified in National Institute for Standards and Technology (NIST) special publication 800-22a.[15] We find that the numbers generated in this way are random,

[1]University of Glasgow, Joseph Black Building, University Avenue, Glasgow G12 8QQ, UK

[2]University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh EH9 3FJ, UK

[3]Lead Contact

*Correspondence: lee.cronin@glasgow.ac.uk

https://doi.org/10.1016/j.matt.2020.01.024

Matter
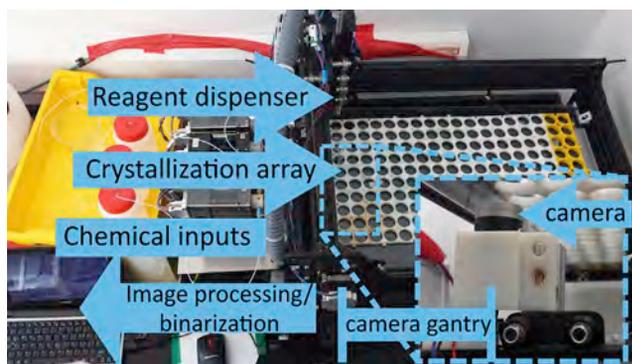
Figure 2. Setup of the Robotic System
Photograph showing the crystallization array inside the CNC framework and its relative position to the input stock solutions, pumps, camera, and controlling computer.

demonstrating the possibility to investigate and use crystallization as an entropy pool for random numbers, and we show that this is possible by encrypting a word and validating the difficulty in breaking the code.

RESULTS

A robotic platform (Figure 2) was designed to generate images of fresh crystallizations for random number generation from chemistry, based on a Computer Numerical Control (CNC) machine. Using rapid prototyping techniques described previously,[16] an additional set of motorized linear axes were attached to the underside of the device to support a camera on a mobile gantry. The mobility in the main CNC framework and auxiliary framework were controlled using technology originally designed for the open source "RepRap" 3D printer.[17] Reagent stock solutions were located adjacent to the platform, and could be transferred to vials in the crystallization array using a combination of tubing and pumps.

Additional 3D-printed components were incorporated to direct the reagent outlets, fix the positions of the vials in a 10 × 10 array, and support the camera (Figures S1–S8). Experiments consisted of pre-set volumes of stock solutions being pumped into each 14-mL vial in a vial array, sequentially. The subsequent growth of crystals in each of the vials was recorded by a mobile camera at regular 10-min intervals at a resolution of 1,280 × 800 pixels. Image analysis using an object-detection and image-segmentation algorithm (Mask R-CNN)[18,19] was employed to locate the crystals in the vial. The full methodology for platform construction and operation is described in Supplemental Experimental Procedures, with supporting software found in the repository linked below.

Chemical inputs were chosen primarily such that they would produce macroscopically observable crystals in a time scale of minutes to hours without the formation of precipitate.[20] The location, size, shape, orientation, and color of crystal formation within the vial were detected and taken as the entropy source for this system (see Supplemental Information for details). We performed reactions that involved different chemical processes, namely (1) crystallization alone, (2) cluster formation, and (3) ligand attachment to cluster, and hypothesized that by increasing the number of chemical processes prior to observation of crystallization a larger number of microstates would be accessible, increasing the system's entropy, and therefore increase the randomness of the number generated (Figure 3A).[14] The three investigated reactions were: recrystallization of the inorganic salt $CuSO_4 \cdot 5H_2O$; the
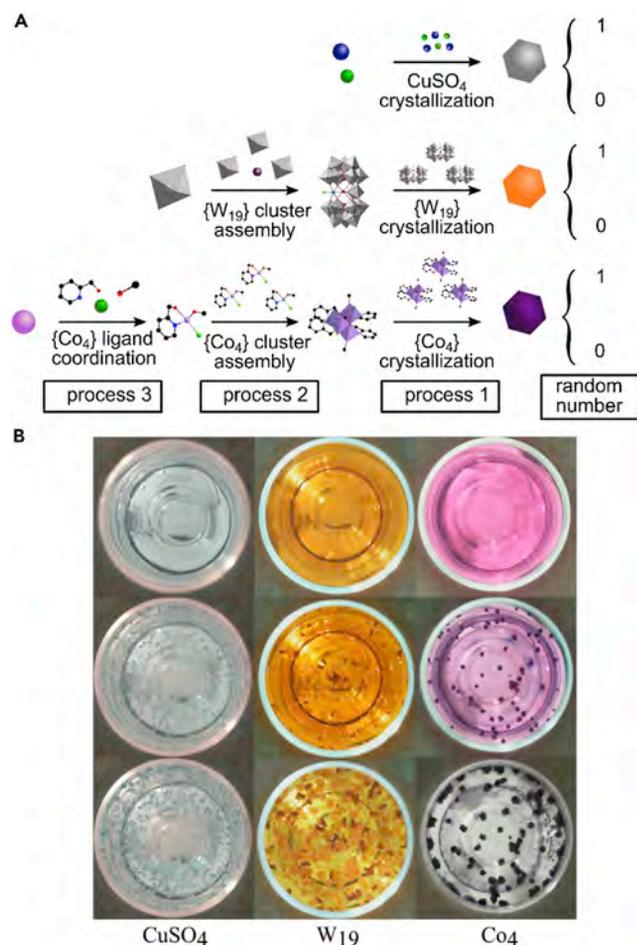
Figure 3. Chemical Schemes for Process Investigation

(A) Chemical schemes for process investigation. $CuSO_4$ requires the stochastic process of crystallization alone, whereas $\{W_{19}\}$ and $\{Co_4\}$ require cluster formation in addition, and $\{Co_4\}$ requires the further step of ligand attachment.

(B) Reactions to form crystals of $CuSO_4$, W_{19}, and Co_4. Top: initial reaction solutions (time = 0 min). middle: partially complete crystallization (time = 40 min). Bottom: crystallizations at the end of the experiment (time = 150 min).

synthesis and crystallization of the polyoxometalate salt $(C_2H_8N)_8Na_3[W_{19}Mn_2O_{61}Cl(SeO_3)_2(H_2O)_2]Cl_2 \cdot 6H_2O$,[21] hereafter referred to as $\{W_{19}\}$; and the synthesis and crystallization of the coordination cluster $[Co_4(2\text{-pyridinemethanol})_4(MeOH)_4Cl_4]$,[22] hereafter referred to as $\{Co_4\}$. Images of these crystallizations at different times are shown in Figure 3B and their chemical structures are shown in Figures S9–S11. These reactions involve the chemical processes of (1), (1 and 2), and (1, 2, and 3), respectively. The synthetic procedure files are included in the online repository and were used to perform the experiments in a fully automated manner. Compound confirmation was obtained by performing single-crystal X-ray diffraction.

Images of each vial were obtained at 10-min intervals for each reaction. The crystals in each image were isolated using an object-detection and image-segmentation algorithm (Mask R-CNN) and their locations within the vial were determined using computer vision techniques (see Figures S12 and S13 for details). A raw binary sequence was then generated by assigning 5 bytes per crystal pixel based on size, orientation, and color, with crystals ordered from top left to bottom right in the
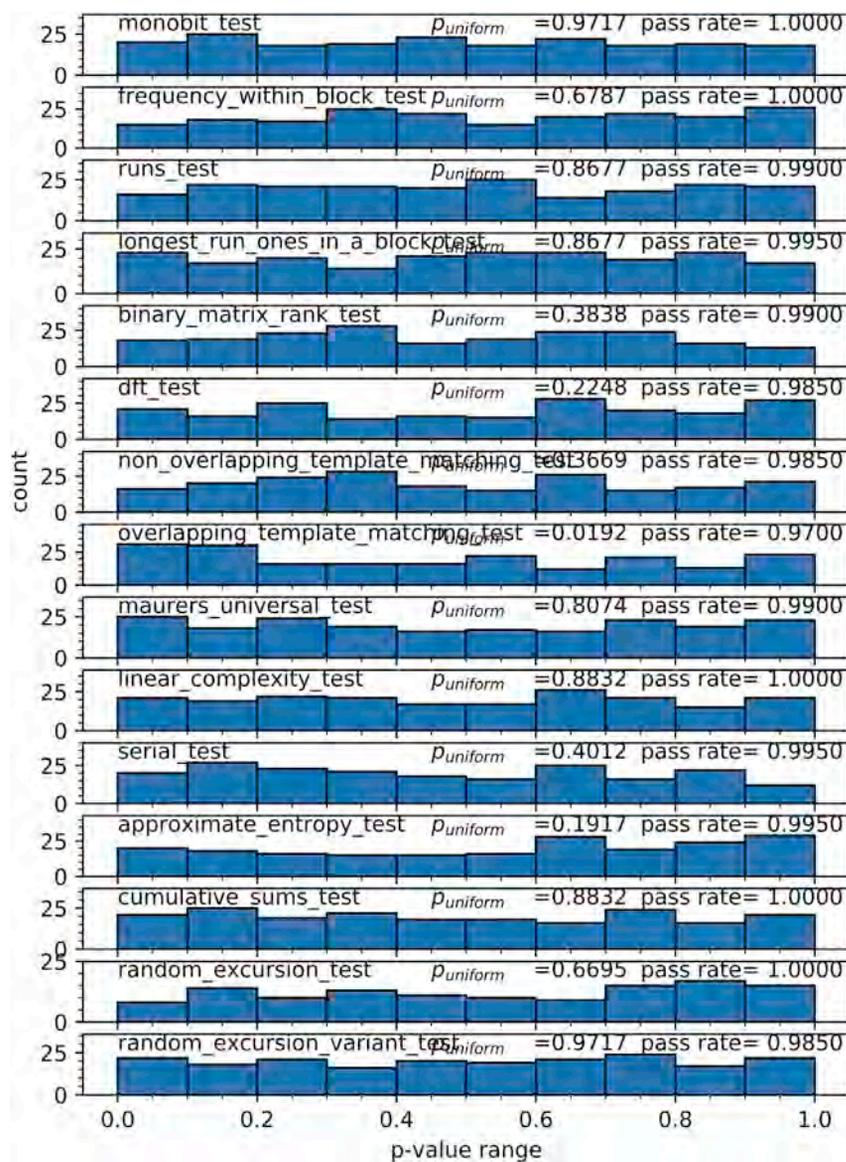
Figure 4. Results of NIST Testing for a Sequence Generated by {CuSO₄} at a Time of 2 h

The histogram consists of p values obtained by running the first-level testing on a single sequence of binary integers divided into 200 blocks of length 1,120,378.

vial region (see Supplemental Information for details). These bytes were concatenated with those of adjacent crystal pixels and of subsequent crystals to generate a long binary sequence. This sequence was then split into sections of 64 bytes, and the sha512 algorithm was applied to each of these sections. The resulting binary sequences were concatenated to form a larger-output binary sequence.

The output binary sequences were evaluated for randomness using the tests for randomness recommended by NIST and published in NIST special publication 800-22a.[15] We assessed the randomness of each experiment in different reaction vials at the same time. The p value histogram of each test for {CuSO₄} is shown in Figure 4 and confirm that each in each case the numbers generated were random, due to the pass rates and uniformity of p values. Similar results for {W₁₉} and {Co₄}, along
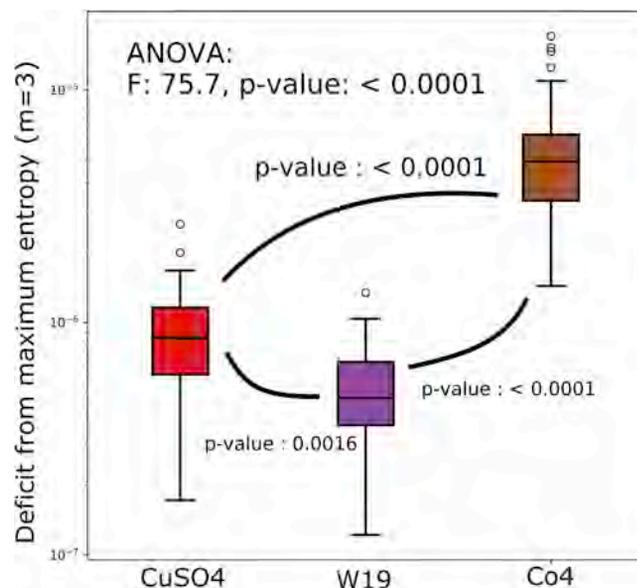
Figure 5. Deficit from Total Maximum Approximate Entropy for Reactions with $CuSO_4$, W_{19}, and Co_4 across 30 Different Reactions in an Experiment

Results of the one-way ANOVA show that the samples are statistically different, while Dunn's test shows that all compounds are different from each other with order of increasing DEF_3 being $\{W_{19}\}, > CuSO_4, \{Co_4\}$.

with figures showing bar charts of p value uniformity and pass rates, are presented in Figures S14–S16. It is worth noting that assessment of only one feature produced strings that were too short to be reliably assessed using the entire NIST package, and as such randomness of features individually were not assessed.

We also wanted to assess the entropy content of each random sequence. To do so, we calculated the deficit from maximum entropy for each sequence as described in Pincus and Singer.[23] Results were obtained by calculating deficit from maximum approximate entropy (DEF_m) for independent reactions in an experiment for different values of m for one compound, and comparing these against DEF_m values from other compound reactions at an equal time index. The results from a one-way ANOVA test show that there are statistically significant differences between the three samples for $m = 1,2,3$ (F values = 20.2, 31.8, 75.7; p values $<< 0.001$), and results from application of Dunn's test indicate that each of the compounds are different from each other. Box plots showing this information for DEF_3 for each reaction are shown in Figure 5, while other values of m are shown in Figure S17.

Finally, since random numbers are commonly used to encrypt data, we considered the encryption capability of this random number generator versus that of a frequently used pseudorandom number generator, the Mersenne Twister (MT).[24] Since the MT output is determined based on its internal state, knowledge of this state allows accurate prediction of future output. However, this is not possible in the case of the true random number generator, whose internal state is either (seemingly) non-existent[25] or unknowable due to the amount of uncertainty/entropy of the physical process.

To do so, we encrypted the ASCII string "crystal!" using numbers generated from both the chemical random number generator and the MT, and compared the average time taken to obtain the correct key to decrypt the message. In both
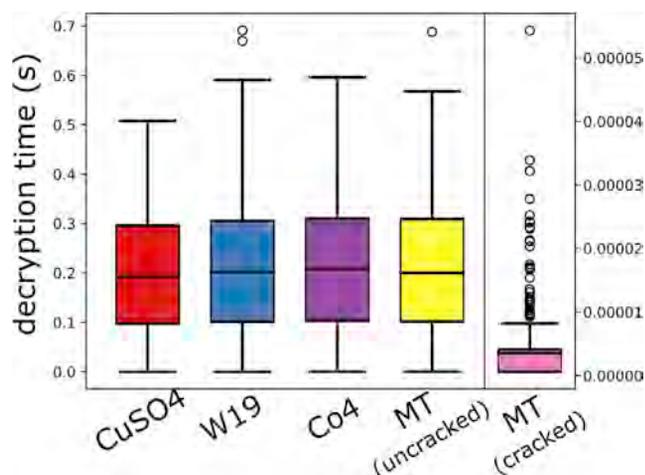
Figure 6. Box Plots of Times for the Message "Crystal!" to be Decrypted after Encryption Using 2,496 16-Bit Keys Generated by Different Methods

Keys generated using the compound random number generator on average take as long to crack as those produced using the uncracked Mersenne Twister (MT); however, once the state of the MT is determined, the average decryption time for this method is negligible when compared with the crystallization method.

instances, brute-force decryption was used; however in the MT case, knowledge of previous successful keys allowed determination of the MT state, and hence allowed prediction of all future output. Box plots for the decryption of 5,000 8-bit keys are presented in Figure 6, and show that while the average times to decrypt keys generated by each compound and the uncracked MT are similar (~5 ms), average decryption time for the cracked MT is reduced to almost zero seconds.

DISCUSSION

We have shown that the large entropy pool present during compound formation and crystallization is able to be accessed and utilized in the creation of a random number generator. We also found that the deficit from maximum entropy of the strings generated by each chemical reaction is not identical, suggesting that this may be a useful means of comparing entropy created during different chemical reactions. However, the order predicted, which supposed that the entropy would correlate with the number of pre-crystallization processes, was not observed, indicating that other factors may be having a greater influence. Finally, we were able to show that numbers generated in this manner could produce sequences of encryption keys with better cryptographic properties than those of the MT.

Although the bit-generation rate is significantly lower than that in other methods (up to ~25 kB/s), this method represents the first time that sampling from the compound formation and crystallization entropy pools has been approached, and future optimization of this process will likely permit much greater bit-production rates. On the other hand, using chemistry to generate random numbers may enhance the output's cryptographic security, as factors such as concentration, temperature, and chemical composition may affect the output number. An attacker would need to know which type of chemistry had been used in order to create the number, and this becomes increasingly infeasible as more types of chemistry and conditions are used to generate random numbers. Additionally, measurement of the entropy from these reactions may be a useful means to assess the chemical complexity of these systems. For instance, it may be possible to distinguish between multivariate metal-organic

Matter

CellPress

frameworks with different internal ordering[26] using this method, since a change in entropy of the crystal structure may influence the entropy of the random number output.

In practice, this feature may be miniaturized and incorporated into computer hardware as a monolithic sealed device in which crystallization is temperature controlled, allowing for repeated cycles of random number generation. Such a sealed device could be embedded in conventional electronic computers, allowing access to a powerful and convenient random number generator powered by chemical processes.

EXPERIMENTAL PROCEDURES

Full experimental procedures are provided in Supplemental Information.

DATA AND CODE AVAILABILITY

All experimental data are available from the authors upon request. Code for the binarization of images, sequence randomness testing, and figure preparation is located at https://github.com/croningp/xtl-rng.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.matt.2020.01.024.

AUTHOR CONTRIBUTIONS

The idea was conceived by L.C. and developed by E.C.L. with help from J.M.P.-G., A.H., and E.K.B. E.C.L. built the robot, wrote the code, did the reactions, and analyzed the data with help from J,M.P.-G. and A.H. All the authors helped to write the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Jia, X., Lynch, A., Huang, Y., Danielson, M., Lang'at, I., Milder, A., Ruby, A.E., Wang, H., Friedler, S.A., Norquist, A.J., and Schrier, J. (2019). Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis. Nature 573, 251–255.

2. Dodge, J., Gururangan, S., Card, D., Schwartz, R., and Smith, N.A. (2019). Show your work: improved reporting of experimental results. arXiv. https://arxiv.org/abs/1909.03004.

3. Steiner, S., Wolf, J., Glatzel, S., Andreou, A., Granda, J.M., Keenan, G., Hinkley, T., Aragon-Camarasa, G., Kitson, P.J., Angelone, D., and Cronin, L. (2019). Organic synthesis in a modular robotic system driven by a chemical programming language. Science 363, eaav2211.

4. Coley, C.W., Thomas, D.A., Lummiss, J.A.M., Jaworski, J.N., Breen, C.P., Schultz, V., Hart, T., Fishman, J.S., Rogers, L., Gao, H., et al. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. Science 365, eaax1566.

5. Granda, J.M., Donina, L., Dragone, V., Long, D.-L., and Cronin, L. (2018). Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature 559, 377–381.

6. Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A., Veselov, M.S., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Polykovskiy, D.A.,

Matter

CellPress

Kuznetsov, M.D., Asadulaev, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat. Biotechnol. *37*, 1038–1040.

7. Baur, S., Deppe, C., and Boche, H. (2019). Secure storage for identification; random resources and privacy leakage. IEEE Trans. Inf. Forensics Security *14*, 2013–2027.

8. He, S., Tang, S., and Wang, W. (2019). A stochastic SIS model driven by random diffusion of air pollutants. Physica A *532*, 121759.

9. Stipčević, M., and Koç, Ç.K. (2014). True random number generators. In Open Problems in Mathematics and Computational Science, Ç.K. Koç, ed. (Springer International Publishing), pp. 275–315.

10. Li, P., Wang, Y.-C., and Zhang, J.-Z. (2010). All-optical fast random number generator. Opt. Express *18*, 20360–20369.

11. Noll, L.C.; Mende, R.G.; Sisodiya, S.. (1996). Method for seeding a pseudo-random number generator with a cryptographic hash of a digitization of a chaotic system.US Patent US5732138A, filed January 29, 1996 and granted March 24, 1998.

12. Feistel, R. (2019). Distinguishing between Clausius, Boltzmann and Pauling entropies of frozen non-equilibrium states. Entropy *21*, 799.

13. Greiner, W., Neise, L., and Stocker, H. (2000). Thermodynamics and Statistical Mechanics (Classical Theoretical Physics), Second Edition (Springer).

14. Krivovichev, S.V. (2014). Which inorganic structures are the most complex? Angew. Chem. Int. Ed. *53*, 654–661.

15. Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., et al. (2010). A statistical test suite for the validation of random number generators and pseudo random number generators for cryptographic applications, NIST Special Publication 800-22, revison 1a. https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-22r1a.pdf.

16. Gutierrez, J.M.P., Hinkley, T., Taylor, J.W., Yanev, K., and Cronin, L. (2014). Evolution of oil droplets in a chemorobotic platform. Nat. Commun. *5*, 5571.

17. Jones, R., Haufe, P., Sells, E., Iravani, P., Olliver, V., Palmer, C., and Bowyer, A. (2011). RepRap—the replicating rapid prototyper. Robotica *29*, 177–191.

18. Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. GitHub repository. https://github.com/matterport/Mask_RCNN.

19. He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask R-CNN. arXiv, arXiv:1703.06870. https://arxiv.org/abs/1703.06870.

20. Duros, V., Grizou, J., Xuan, W.M., Hosni, Z., Long, D.-L., Miras, H.N., and Cronin, L. (2017). Human versus robot in the discovery and crystallization of gigantic polyoxometalates. Angew. Chem. Int. Ed. *56*, 10815–10820.

21. Symes, M.D., Kitson, P.J., Yan, J., Richmond, C.J., Cooper, G.J.T., Bowman, R.W., Vilbrandt, T., and Cronin, L. (2012). Integrated 3D-printed reactionware for chemical synthesis and analysis. Nat. Chem. *4*, 349–354.

22. Yang, E.-C., Hendrickson, D.N., Wernsdorfer, W., Nakano, M., Zakharov, L.N., Sommer, R.D., Rheingold, A.L., Ledezma-Gairaud, M., and Christou, G. (2002). Cobalt single-molecule magnet. J. Appl. Phys. *91*, 7382–7384.

23. Pincus, S., and Singer, B.H. (1996). Randomness and degrees of irregularity. Proc. Natl. Acad. Sci. U S A *93*, 2083–2088.

24. Matsumoto, M., and Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. *8*, 3–30.

25. Hensen, B., Bernien, H., Dréau, A.E., Reiserer, A., Kalb, N., Blok, M.S., Ruitenberg, J., Vermeulen, R.F.L., Schouten, R.N., Abellán, C., et al. (2015). Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. Nature *526*, 682–686.

26. Kong, X., Deng, H., Yan, F., Kim, J., Swisher, J.A., Smit, B., Yaghi, O.M., and Reimer, J.A. (2013). Mapping of functional groups in metal-organic frameworks. Science *341*, 882–885.