RESEARCH ARTICLE | OCTOBER 09 2024

Colloidal particles as noise source for random number generation

Alexander Scholz ; Pooja Arya ; Jasmin Aghassi-Hagmann

AIP Advances 14, 105114 (2024)
https://doi.org/10.1063/5.0223329

Colloidal particles as noise source for random number generation

Cite as: AIP Advances 14, 105114 (2024); doi: 10.1063/5.0223329
Submitted: 17 June 2024 • Accepted: 23 September 2024 •
Published Online: 9 October 2024

Alexander Scholz,[a] Pooja Arya, and Jasmin Aghassi-Hagmann[b]

AFFILIATIONS

Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), 76128 Karlsruhe, Germany

[a]Electronic mail: alexander.scholz2@kit.edu
[b]Author to whom correspondence should be addressed: jasmin.aghassi@kit.edu

ABSTRACT

In this work, we investigate colloidal particle patterns as a possible noise source for random number generation. We systematically analyze the minimum entropy of the noise source over different particle concentrations of $\{1, 3, 5, 7, 10, 12, 15\}$ mg/ml according to the recommendations of the National Institute of Standards and Technology Special Publication 800-90B. The estimated minimum entropy of the non-independent and identically distributed particle pattern noise source is $H_{\min} = 0.5896/1$ bit at a particle amount of 5 mg/ml. For further entropy extraction on the noise source data, the secure hash algorithm is used to construct an entropy source. The randomness of the derived entropy source is verified according to the National Institute of Standards and Technology Special Publication 800-22 Rev. 1a and the accompanying statistical test suite. The entropy source passes all randomness tests of the statistical test suite and shows an estimated minimum entropy of $H_{\min} = 0.9992/1$ bit.

I. INTRODUCTION

Utilization of random data plays a pivotal role in various fields, including weather forecast, simulation, and cryptographic protocols.[1] The latter is especially important for securing the exchange of sensitive and private data and establishes trust in our everyday digital life. To achieve the generation of unpredictable and random data, different approaches have shown feasibility. One established method is the so-called deterministic random bit generator (DRBG), or pseudo-random number generator.[2] Here, a deterministic function generates the subsequent random number sequence from a seed using algorithms, which allows for high throughput random number generation in a deterministic fashion. As a seed for such a DRBG, a high quality, unpredictable entropy source is required, as the DRBG can be compromised when the seed is found. Another way for random number generation are the so-called true random number generators (TRNGs), which provide random numbers directly from a high quality entropy source and not from a deterministic function. The output of TRNGs can be used as a seed for DRBGs and also in real time, usually at a lower data rate when compared to DRBGs. However, providing high quality entropy sources requires resources, often not available on classical computers and devices deployed in the Internet of things. For computers, hard-to-guess events from user inputs, such as keyboard stroke timings, and network events are utilized to seed internal DRBGs.[3] These events unfortunately do not carry high entropy and provide a weak spot in the overall system security architecture.

Here, possible solutions such as entropy as a service provide possibilities to seed DRBGs on client devices from external high quality entropy sources.[3] A high quality entropy source exploits intrinsic stochastic phenomena from a physical noise source. This includes noise in electronic devices and meta-stability in logic circuits,[4–8] photon emission,[9,10] radioactive decay,[11] magnetoresistive and resistive random access memory cells,[12–14] and random speckle patterns.[1,15–17] Nonetheless, physical stochastic phenomena need to be recorded and digitized for further processing. In this step, next to possible systematic biases in the physical noise source itself, further bias can be introduced within the seemingly random data by quantization. Often, post-processing methods, in the form of randomness or entropy extractors, are required to increase the randomness in the bit strings extracted from the noise source and hence provide unpredictable and random data.[10] For this purpose,

cryptographic hash functions, such as the secure hash algorithm (SHA-2), can be deployed. The SHA-2 is able to generate unpredictable and randomized bit strings, which can be deployed for entropy source construction.[2] However, as the SHA-2 is a deterministic function, the quality and minimum entropy of the noise source, which is input to the SHA-2, should be verified.

In this work, we experimentally investigate colloidal particle patterns in an aqueous dispersion as a possible physical noise source for entropy source construction. Based on the kinetic theory of gases, molecules in fluids exhibit Brownian motion.[18] This results in collisions of molecules with each other and with other particles suspended in that fluid. During collision with particles, molecules transfer a small amount of their kinetic energy to the particle. As those collisions occur from every direction on the particles at different angles and with different impact energies, the direction of motion of the particle is unpredictable and thus random at every step.[19,20] To investigate the quality of the noise and entropy source, we follow the recommendations of the National Institute of Standards and Technology (NIST) Special Publication (SP) 800-90B.[2] For standardized minimum entropy ($H_{\min}$) estimation, we use the NIST entropy assessment tool. The obtained minimum entropy of the non-independent and identically distributed (non-iid) noise source is $H_{\min} = 0.5896/1$ bit at a particle amount of 5 mg/ml. To generate random and unpredictable bit strings, the obtained raw data of the noise source are post-processed by the SHA-512 to derive the entropy source. The minimum entropy and randomness of the entropy source are investigated on the basis of the NIST SP 800-90B, the NIST SP 800-22 Rev. 1a, and the accompanying statistical test suite (sts-2.1.2).[21] The entropy source passes all randomness tests successfully and shows a minimum entropy of $H_{\min} = 0.9992/1$ bit.

II. EXPERIMENTAL SETUP

Aqueous dispersions with 1 μm-sized spherical silica particles (Sicastar-micromod, Partikeltechnologie) are prepared in milli-Q pore water (≥18.2 MΩ) using different particle concentrations of $\{1, 3, 5, 7, 10, 12, 15\}$ mg/ml. The obtained dispersions are ultrasonicated for 15 min to avoid particle aggregation. Each dispersion is filled in an enclosed chamber (Gene Frames, Thermo Fisher Scientific) with a size of 16 × 15 × 0.25 mm^3 and a volume of 65 µl. The prepared samples in the frames are kept as they are for over two hours for particle sedimentation and thus reaching equilibrium.

An inverted microscope (Eclipse-Ti2, Nikon) equipped with a low-noise monochrome microscope camera (DS-Qi2, Nikon) with a resolution of 14 bits per pixel and a high power LED light source is used for the measurements. The optical images are acquired over a region of interest in the center of the particle pattern sample of 1024 × 1024 pixels at 10× optical magnification and a sampling time of 200 ms. Due to the random motion, the particle pattern changes constantly. The raw particle pattern acquisition process from the random particle sample is shown in Fig. 1. The analyzed and investigated grayscale images of the noise source are stored with a bit depth of 8 bits per pixel in the lossless tagged image file format. At the given sampling rate, ≈41.9 Mbit/s of raw noise source data is generated. The irradiation intensity of the light source is kept constant and stable over the whole imaging area throughout the experiments. The experiments were conducted in a laboratory environment at a temperature of $T = 23\,^{\circ}$C.

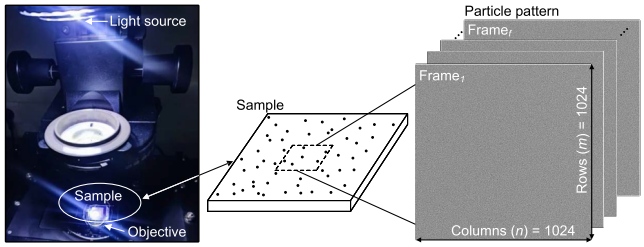

FIG. 1. Schematic of the raw particle pattern acquisition process from the physical noise source sample (spherical silica particles in water). The region of interest is centered on the sample with a size of 1024 × 1024 pixels. Each 200 ms, a new greyscale frame is captured, which changes over time and differs from the preceding image.

III. RESULTS

A. Investigation of the noise source

To investigate the impact of external noise on the experimental setup, which does not stem from the particle dispersion, we acquire 300 consecutive frames of a milli-Q pore water-filled sample container without particles and calculate the Euclidean distances over all observations for each of the 300 frames and compare it with the Euclidean distances over each of 300 frames of a sample, which contains particles with a particle amount of 5 mg/ml. The Euclidean distance for two distinct rows i, j in an $m \times n$ matrix A is calculated according to the following equation:

$$D_{i,j} = \sum_{k=1}^{n} \sqrt{(A_{ik} - A_{jk})^2}. \tag{1}$$

The obtained results are visualized as two histograms in Fig. 2(a). The Euclidean distances of the sample container without particles (black histogram) show a mean value of 68.24 with a maximum value of 92.98, whereas the sample with a particle amount of 5 mg/ml (blue histogram) shows a mean value of 1649.50 and a minimum value of 938.23, due to outliers. This gives evidence that the origin of the acquired data from the noise source stems primarily from the particles and their motion, while system noise shows only a minor influence on the overall acquired grayscale values.

In the following, we investigate the impact of different particle concentrations of the prepared samples of $\{1, 3, 5, 7, 10, 12, 15\}$ mg/ml on the obtainable minimum entropy. The seven samples are systematically evaluated using the NIST SP 800-90B entropy assessment test battery. For each sample, 300 consecutive frames were recorded and analyzed. For minimum entropy estimation using the entropy assessment test battery, a consecutive string of at least 1 Mbit of sample data is required. As the amount of acquired data over 300 frames is too large to be efficiently tested using the SP 800-90B test battery, four sections of six frames at the frame positions {1–6, 95–100, 195–200, 295–300} are investigated for each sample concentration. This assessment helps observe possible changes in $H_{\min}$ over time. To provide a consecutive string, each of the 1024 rows consisting of 1024 column grayscale values in the range of 0–255 is concatenated over all six frames, which results in a bit string of ≈ 50.3 Mbit per frame section. In our initial assessment of the noise source, the χ^2 test is not passed. Therefore, the output of the noise source is not

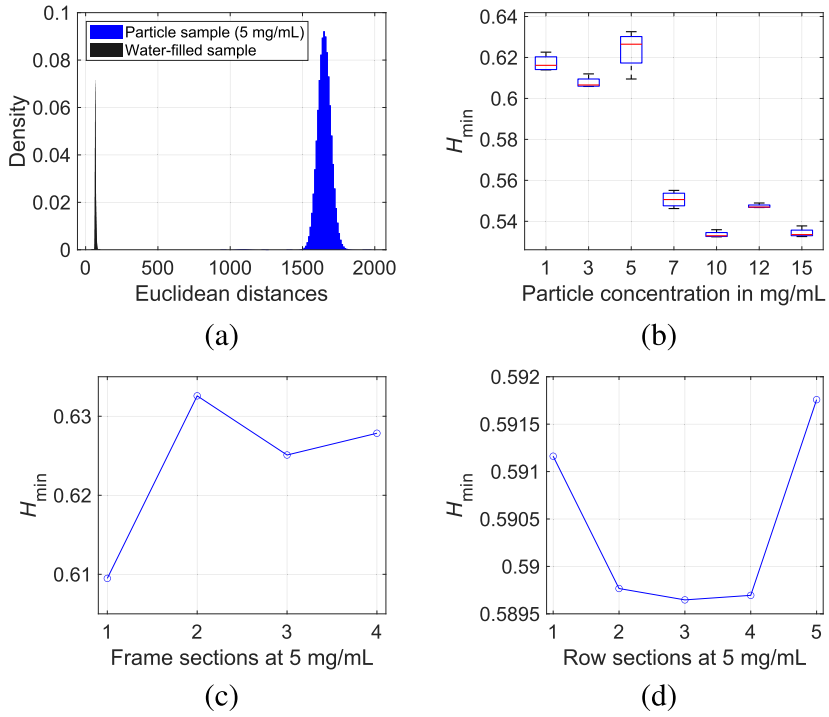

FIG. 2. (a) Euclidean distances visualized as histograms over 300 frames. The black histogram is obtained from a water-filled sample container, which visualizes the impact of the system noise on the particle noise source. The blue histogram is obtained from a sample with a particle concentration of 5 mg/ml. (b) Estimated minimum entropy over different particle concentrations over 300 frames. Particle concentrations between 1 and 5 mg/ml show on average higher H_{min} as compared to higher particle concentrations ranging from 7 to 15 mg/ml. (c) H_{min} over slices of six consecutive frames at a particle concentration of 5 mg/ml and 300 frames. (d) H_{min} over five row sections containing two neighboring rows each and 3000 consecutive frames.

independent and identically distributed (iid). While single particles suspended in a fluid exhibit random motion due to continuous collisions of fluid molecules on the particle, in our case, the particle concentration is very high. This leads to collision of water molecules and to collisions from neighboring particles, which can affect the frequency and force of the collisions and alters the particle trajectory. Furthermore, when a particle moves through a fluid, it creates a disturbance in the fluid that affects the motion of nearby particles. This disturbance can persist for some time and influence the future motion of the original particle and others in its vicinity. The motion of a single particle in a crowded environment can exhibit a memory effect resulting in a long-range correlation and can be understood as a phenomenon, where the past interactions and collisions of a particle can influence its current trajectory, resulting in a sub-diffusive behavior.[22,23] Therefore, the non-iid tests of the NIST SP 800-90B entropy assessment test battery are used to estimate H_{min}. For further in-depth information of the applied non-iid entropy estimators, we refer to the NIST SP 800-90B.[2]

The results of the estimated minimum entropy are visualized as box plots in Fig. 2(b). Particle concentrations in the range of 1–5 mg/ml show on average higher H_{min} values, when compared to increased particle amounts. At increasing particle amounts, ranging from 7 to 15 mg/ml, H_{min} is slightly reduced and a saturation effect with a stable H_{min} can be observed. In Fig. 2(c), H_{min} at a particle concentration of 5 mg/ml over the four frame sections is shown. A slight dip is observed at frames 1–6 when compared to the other frame sections. This could be explained due to possible areal agglomeration of particles at a specific position, which dissolved again after a certain time. However, it is also visible that H_{min} does not degrade over the investigated time. Due to optical magnification of the microscope, there is a chance that a particle, which has a fixed diameter of 1 μm in our test scenario, can be observed on a neighboring pixel. This can be the case for both directions on the x- or y-axis, respectively. To investigate possible correlation of two neighboring rows, we construct five row sections over two neighboring rows {1–2, 256–257, 512–513, 768–769, 1023–1024} over 3000 consecutive frames at a particle concentration of 5 mg/ml and apply the non-iid tests of the NIST SP 800-90B test battery to each section. We can observe in Fig. 2(d) that H_{min} is slightly reduced, due to increased susceptibility of the noise source with respect to compression estimation, with a minimum value of $H_{min} = 0.5896/1$ bit in the third section (rows 512–513). The estimator results for the lowest obtained results of the frame and row section tests are listed in Table I.

For both datasets, restart tests according to the NIST SP 800-90B recommendations are applied. Therefore, a restart of the physical noise source over 1000× and collection of 1000 data sample points per each session should be performed. This is in our case impractical as we cannot turn the microscope on and off 1000×. However, we applied the restart tests to a consecutive string

TABLE I. Results of the H_{min} estimation for the noise source, based on the non-iid track of the NIST SP 800-90B entropy assessment test battery. The results shown highlight the lowest obtained results of the bit strings in the frame and row section tests at a particle concentration of 5 mg/ml.

Non-iid test estimator	Frame section	Row section
Most common value	0.995 392	0.960 168
Collision	1.000 000	1.000 000
Markov	0.994 940	0.893 143
Compression	0.771 517	0.589 647
T-tuple	0.609 485	0.784 990
Longest repeated substring	0.924 404	0.859 415
Multi-MCW	0.700 999	0.962 313
Lag	0.863 063	0.888 489
Multi-MMC	0.884 087	0.763 084
LZ78Y	0.994 431	0.862 079

of 1 Mbit from a singular frame for each frame section and to 1 Mbit for each row section. All restart tests were successfully passed.

B. Minimum entropy and randomness investigation of the entropy source

For entropy extraction and construction of the entropy source, we apply the SHA-512 to our obtained bit strings from the physical noise source. Therefore, we use a full row of a frame, which contains 1024×8 bit values. This results in a 8192 bit wide input string with $H_{min} = 0.5896/1$ bit, and thus ≈4830 bits of raw entropy at a particle concentration of 5 mg/ml, which is input into the SHA-512 compression function. This strategy is applied to each row of a frame. For each frame, 1024×512 random bits are generated. This results in a theoretically available ≈2.6 Mbit/s of random numbers from the entropy source. We estimate the entropy of the entropy source in a similar fashion as we did on the noise source. Therefore, we concatenate the 512 bit wide output strings from the SHA-512 to obtain a 6 Mbit wide bit string of the entropy source. Unlike the noise source, the hashed output bit strings are highly random and pass all the iid tests of the NIST entropy assessment test battery. The obtained entropy of the entropy source is $H_{min} = 0.9992/1$ bit.

For the investigation of the entropy source randomness, in total, 3000 consecutive frames at a particle concentration of 5 mg/ml were acquired and post-processed using the SHA-512 as described earlier. From these data, the first 1 Gbit is used to construct 1000 test sequences with a length of 1 Mbit each and further analyzed using the NIST SP 800-22 Rev. 1a statistical test suite (sts-2.1.2). The results of the sts-2.1.2 are shown Table II, which contains the applied statistical test, block length—where applicable—and the calculated P and proportion values. For tests, which report multiple results, the lowest scoring value is included in the table. To pass a test, the calculated P-value has to be larger than the significance level of $\alpha = 0.01$. The tests where $P > \alpha$ require a proportion value of ≥0.980 with the exception of the random excursion and random excursion variant test, where a proportion value of ≥0.976 is needed to obtain a successful result. For in-depth information of the performed randomness tests and parameter selection of sts-2.1.2, we refer to the NIST SP 800-22 Rev. 1a.[21] The table shows that all tests are successfully passed and no anomalies in the randomness of the entropy source could be detected. It should be noted that the results of sts-2.1.2 can only assess the randomness and cannot substitute for further cryptographic analysis on the entropy source. Nonethe-

TABLE II. Results and settings used in the NIST sts-2.1.2. In total, 1 Gbit of data of the entropy source is analyzed over 1000 sequences and a sequence length of 1 Mbit. The significance level for hypothesis testing is $\alpha = 0.01$.

Statistical test	Block length	P-value	Proportion	Result
Frequency	···	0.603 841	0.990	Success
Block frequency	$M = 20\,000$	0.039 073	0.989	Success
Cumulative sums (forward)	···	0.435 430	0.990	Success
Cumulative sums (backward)	···	0.221 317	0.991	Success
Runs	···	0.488 534	0.987	Success
Longest run	···	0.299 736	0.990	Success
Rank	···	0.339 271	0.987	Success
FFT	···	0.160 805	0.987	Success
Non-overlapping template	$m = 9$	0.991 147	0.984	Success
Overlapping template	$m = 9$	0.668 321	0.991	Success
Universal	···	0.036 352	0.982	Success
Approximate entropy	$m = 10$	0.899 171	0.986	Success
Random excursions	···	0.910 293	0.985	Success
Random excursions variant	···	0.910 293	0.983	Success
Serial (first)	$m = 16$	0.512 137	0.991	Success
Serial (second)	$m = 16$	0.900 569	0.991	Success
Linear complexity	$m = 1000$	0.164 425	0.988	Success

less, as the SHA-2 is built on the Merkle–Damgård construction, its output is considered collision resistant and secure.[24]

IV. CONCLUSION

We systematically investigated particle patterns that were derived from spherical silica particles in water over different particle concentrations of $\{1, 3, 5, 7, 10, 12, 15\}$ mg/ml as a possible noise source for random number generation. The estimated minimum entropy of the non-independent and identically distributed particle pattern noise source is $H_{\min} = 0.5896/1$ bit at a particle amount of 5 mg/ml following the recommendations of the NIST SP 800-90B using the entropy assessment tool for standardized minimum entropy assessment. For further entropy extraction on the noise source data, the SHA-512 is deployed to construct an entropy source. The randomness of the derived entropy source is verified according to the NIST SP 800-22 Rev. 1a and NIST sts-2.1.2. All randomness tests are successfully passed. Furthermore, the entropy of the entropy source is verified similar to the noise source. The acquired minimum entropy of the iid entropy source is $H_{\min} = 0.9992/1$ bit. The investigated spherical silica particles in water are promising candidates as physical noise source for random number generators.

ACKNOWLEDGMENTS

Alexander Scholz acknowledges financial support from the BMBF-funded project SensIC. Pooja Arya acknowledges financial support through the Helmholtz Association, as part of the programme "MSE-materials science and engineering/VirtMat" No. 43.31.01.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Alexander Scholz: Conceptualization (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Pooja Arya**: Conceptualization (equal); Investigation (equal); Validation (equal); Writing – review & editing (equal). **Jasmin Aghassi-Hagmann**: Conceptualization (equal); Investigation (equal); Resources (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] L. M. S. Dias, T. F. S. Silvério, R. A. Sá Ferreira, and P. S. de Brito André, "Random bit sequence generation from speckle patterns produced with multimode waveguides," IET Optoelectr. **16**(4), 174–178 (2022).
[2] M. S. Turan, E. Barker, J. Kelsey, K. A. McKay, M. L. Baish, M. Boyle *et al.*, "Recommendation for the entropy sources used for random bit generation," (NIST Special Publication 800-90B, 2018).
[3] A. Vassilev and R. Staples, "Entropy as a service: Unlocking cryptography's full potential," Computer **49**, 98–102 (2016).
[4] S. K. Tawfeeq, "A random number generator based on single-photon avalanche photodiode dark counts," J. Lightwave Technol. **27**, 5665–5667 (2009).
[5] B. Lampert, R. S. Wahby, S. Leonard, and P. Levis, "Robust, low-cost, auditable random number generation for embedded system security," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM* (Association for Computing Machinery, 2016), pp. 16–27.
[6] X. Zhang, C. Jiang, G. Dai, L. Zhong, W. Fang, K. Gu, G. Xiao, S. Ren, X. Liu, and S. Zou, "Improved performance of SRAM-based true random number generator by leveraging irradiation exposure," Sensors **20**, 6132 (2020).
[7] L. Baldanzi, L. Crocetti, F. Falaschi, M. Bertolucci, J. Belli, L. Fanucci, and S. Saponara, "Cryptographically secure pseudo-random number generator IP-core based on SHA2 algorithm," Sensors **20**, 1869 (2020).
[8] N. Torii, H. Kokubo, D. Yamamoto, K. Itoh, M. Takenaka, and T. Matsumoto, "ASIC implementation of random number generators using SR latches and its evaluation," EURASIP J. Inf. Secur. **2016**, 10.
[9] B. Sanguinetti, A. Martin, H. Zbinden, and N. Gisin, "Quantum random number generation on a mobile phone," Phys. Rev. X **4**, 031056 (2014).
[10] X. Ma, F. Xu, H. Xu, X. Tan, B. Qi, and H.-K. Lo, "Postprocessing for quantum random-number generators: Entropy evaluation and randomness extraction," Phys. Rev. A **87**, 062327 (2013).
[11] K. Park, S. Park, B. G. Choi, T. Kang, J. Kim, Y.-H. Kim, and H.-Z. Jin, "A lightweight true random number generator using beta radiation for iot applications," ETRI J. **42**, 951–964 (2020).
[12] F. Ferdaus, B. B. Talukder, M. Sadi, and M. T. Rahman, "True random number generation using latency variations of commercial MRAM chips," in *2021 22nd International Symposium on Quality Electronic Design (ISQED)* (IEEE, 2021), pp. 510–515.
[13] B. Cambou, D. Telesca, S. Assiri, M. Garrett, S. Jain, and M. Partridge, "TRNGs from pre-formed ReRAM arrays," Cryptography **5**, 8 (2021).
[14] R. Gu, Y. Sun, Y. Wang, W. Wang, and Q. Li, "A rate-adjustable true random number generator based on the stochastic delay of a TiN/NbO_x/Pt memristor," AIP Adv. **11**, 125301 (2021).
[15] M. Akriotou, C. Mesaritakis, E. Grivas, C. Chaintoutis, A. Fragkos, and D. Syvridis, "Random number generation from a secure photonic physical unclonable hardware module," in *Security in Computer and Information Sciences: First International ISCIS Security Workshop 2018, Euro-CYBERSEC 2018, London, UK, February 26-27, 2018, Revised Selected Papers 1* (Springer International Publishing, 2018), pp. 28–37.
[16] K. Chen, F. Huang, P. Wang, Y. Wan, D. Li, and Y. Yao, "Fast random number generator based on optical physical unclonable functions," Opt. Lett. **46**, 4875–4878 (2021).
[17] K. Chen, P. Wang, F. Huang, X. Leng, and Y. Yao, "Analysis of entropy source for random number generation based on optical PUFs," J. Appl. Phys. **133**, 174502 (2023).
[18] J. Renn, "Einstein's invention of Brownian motion," Ann. Phys. **14**, 23–37 (2005).
[19] S. G. Brush, "A history of random processes: I. Brownian movement from Brown to Perrin," Arch. Hist. Exact Sci. **5**, 1–36 (1968).
[20] F. Mainardi and P. Pironi, "The fractional Langevin equation: Brownian motion revisited," arXiv:0806.1010 (2008).
[21] L. E. Bassham III, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks *et al.*, "Sp 800-22 rev. 1a a statistical test suite for random and pseudorandom number generators for cryptographic applications," National Institute of Standards & Technology, 2010.
[22] B. Cichocki and B. Felderhof, "Memory effects in the self-diffusion of interacting Brownian particles," J. Chem. Phys. **96**, 9055–9059 (1992).
[23] J.-H. Jeon, E. Barkai, and R. Metzler, "Noisy continuous time random walks," J. Chem. Phys. **139**, 121916 (2013).
[24] W. Easttom, *Modern Cryptography: Applied Mathematics for Encryption and Information Security* (Springer, 2022).