Generation of correlated binary sequences from white noise

F. M. Izrailev,[1] A. A. Krokhin,[1,2] N. M. Makarov,[3,*] and O. V. Usatenko[3]

[1]*Instituto de Física, Universidad Autónoma de Puebla, Apartado Postal J-48, Puebla, Puebla, 72570, Mexico*

[2]*Department of Physics, University of North Texas, P. O. Box 311427, Denton, Texas 76203, USA*

[3]*A. Ya. Usikov Institute for Radiophysics and Electronics, Ukrainian Academy of Science, 12 Proskura Street, 61085 Kharkov, Ukraine*

We suggest a method for generation of random binary sequences of elements 0 and 1, with prescribed correlation properties. It is based on a modification of the widely used convolution method of constructing continuous random processes. Using this method, a binary sequence with a power-law decaying pair correlator can be easily generated.

PACS number(s): 05.10.−a, 05.40.−a, 02.50.Ga, 87.10.+e

Generators of random numbers or white-noise signals are customary elements in modern digital electronics. Different algorithms are used for this purpose. The quality of a generated white noise is determined by the length of the sequence, elements of which can be considered as uncorrelated. In many areas of physics, such as engineering and signal processing, it is required to generate a colored noise, i.e., a correlated random process. Since the pair correlations usually give the principal contribution to the observable quantities, the problem of generation of a random sequence with a prescribed *pair correlator* is of particular interest. It has been known for a long time that continuous colored noise with exponential correlations is generated by a linear Ornstein-Uhlenbeck process, which is based on integration of a linear Langevin equation driven by a white-noise fluctuating force. [1] A more general method, valid for the generation of continuous random sequences with an arbitrary correlator, is based on the convolution of white noise with the *modulation function* defined by the pair correlator. Originally, the convolution method was proposed by Rice. [2] Applications of modern versions of this method for the generation of random sequences with specific correlations, including the long-range nonexponential correlations, can be found in Refs. [3–8].

In the theory of spatially disordered systems the role of the pair correlator (for random potentials) is emphasized by the fact that it determines the scattering cross section in the Born approximation (for weak potentials). As a result, many linear transport characteristics (conductance, transmission and reflection coefficients, localization length, etc.) are expressed through the pair correlator [7,9,10].

There are some examples of the systems (or processes) with correlated disorder, for which the fluctuating parameter takes discrete values. An example of such a system is the DNA macromolecule. Here, the genetic information is written using four symbols that are the basic nucleotides. In the digital devices the information is transmitted in a form of a telegraphic signal, i.e., a binary code. The binary sequence is the limiting case of sequences with the least number of values accessible for the random variable. For practical applications it is desirable to develop a method of generation of a binary sequence with prescribed pair correlator—a kind of colored noise containing two elements, e.g., 0 and 1. Although there have been some attempts to obtain a robust algorithm for generation of a correlated binary sequence with the purpose of increasing the performance of a pulse radar [11], the problem is still lacking a general solution. It is worth mentioning that there are methods of generation of a correlated binary sequence that are not based on the properties of the correlation function; see, e.g., Ref. [12]. It is not known yet what are the constraints (if there are any) for the pair correlator, imposed by the fact that the sequence is *dichotomous* (binary). For example, an attempt to generate a dichotomous sequence with the correlations decaying according to the inverse power law was unsuccessful [13].

Recently, we addressed the mathematical problem of generation of a dichotomous sequence with prescribed correlation properties [14]. We concentrated our attention on statistical properties of binary additive Markov chains. It was shown that the dichotomous sequences with short-range correlations and some particular sequences with long-range correlations can be reconstructed with the use of the so-called memory function. The latter is related to the pair correlator through a quite complicated linear integral equation. Analytical solution of this equation can be obtained in some special cases only, which narrows the area of practical validity of the method [14].

In this paper, we present an approach based on the convolution method that was modified for the generation of correlated binary sequences. The relation between the *filtering function*—the kernel of the convolution operator—and the pair correlator turns out to be relatively simple. The advantage of this numerically simple method is that it is valid for generation of sequences with power-law decaying correlations—a problem attracting a lot of attention in physics and technology.

The convolution method of generation of a continuous colored noise $\beta(n)$ starting from a white noise $\alpha(n)$ is based on a linear transformation with the use of the modulation function $G(n)$ [7,15]

$$\beta(n) = \bar{\beta} + \sqrt{\frac{C_\beta(0)}{C_\alpha(0)}} \sum_{n'=-\infty}^{\infty} G(n-n')[\alpha(n') - \bar{\alpha}]. \quad (1)$$

Here $\bar{\alpha}$ and $\bar{\beta}$ are the mean values, and $C_\alpha(0) = \overline{\alpha^2(n)} - (\bar{\alpha})^2$ and $C_\beta(0) = \overline{\beta^2(n)} - (\bar{\beta})^2$ are the variances of the white and

colored noise, respectively. For a homogeneous sequence $\alpha(n)$, the generated sequence $\beta(n)$ is also homogeneous. In what follows, we introduce the normalized pair correlator $K_\beta(r)=K_\beta(-r)=C_\beta(r)/C_\beta(0)$. Substituting the linear transformation (1) into the correlation function $C_\beta(r)$ $=\overline{\beta(n+r)\beta(n)}-(\overline{\beta})^2$ and taking into account that the sequence $\alpha(n)$ is δ correlated, the following relation between the pair correlator and the modulation function is readily obtained:

$$K_\beta(r) = \sum_{n=-\infty}^{\infty} G(n)G(n+r). \qquad (2)$$

From the condition $K_\beta(0)=1$, one gets

$$\sum_{n=-\infty}^{\infty} G^2(n) = 1. \qquad (3)$$

Since $K_\beta(n)$ and $G(n)$ are even functions, we apply the cosine Fourier transform to both sides of the convolution (2). This results in the following relation:

$$\mathcal{K}_\beta(k) = \mathcal{G}^2(k), \qquad (4)$$

where

$$\mathcal{K}_\beta(k) = 1 + 2\sum_{r=1}^{\infty} K_\beta(r)\cos(kr), \qquad (5)$$

$$K_\beta(r) = \frac{1}{\pi}\int_0^\pi \mathcal{K}_\beta(k)\cos(kr)dk.$$

Similar relations can be written for $\mathcal{G}(k)$ and $G(n)$.

The expression (4) determines the modulation function in terms of the Fourier transform of the pair correlator,

$$G(n) = \frac{1}{\pi}\int_0^\pi \mathcal{K}_\beta^{1/2}(k)\cos(kn)dk. \qquad (6)$$

Evidently, the solution (6) satisfies the normalization condition (3).

For different white-noise sequences $\alpha(n)$, the convolution method Eqs. (1), (5), and (6) defines an ensemble of colored-noise sequences $\beta(n)$, where $-\infty < \beta(n) < \infty$. All the sequences in this ensemble have the same pair correlator $K_\beta(n)$. The number of terms contributing to the series (1) depends on the sharpness of the correlator $K_\beta(n)$. For short-range correlations, when $K_\beta(n)$ decays very fast, the modulation function $G(n)$ is also sharp; therefore, the principal contribution is mainly given by a single term with $n'=n$. The sequence $\beta(n)$ in this case is practically δ correlated for both continuous and binary sequences $\alpha(n)$. In the opposite case of long-range correlations, when the correlation length R_c is large ($R_c \gg 1$), many terms contribute to Eq. (1). In this case, even for a binary sequence $\alpha(n)$, employing the method of characteristic functions, one can obtain that the probability density $\rho_B(\beta)$ for the stochastic variable $\beta(n)$ has the Gaussian form

$$\rho_B(\beta) = [1/\sqrt{2\pi C_\beta(0)}]\exp[-(\beta-\overline{\beta})^2/2C_\beta(0)], \qquad (7)$$

provided the condition

$$(\beta-\overline{\beta})^2/2C_\beta(0) \ll R_c \qquad (8)$$

is satisfied. The deviations from the Gaussian shape may appear only at the far tails, where $(\beta-\overline{\beta})^2/2C_\beta(0) \gg R_c$. Note that, for a continuous Gaussian distribution of $\alpha(n)$, Eq. (7) is exact independently of R_c.

From the above consideration, it is clear that the correlated sequence $\beta(n)$ may be generated using very different uncorrelated sequences $\alpha(n)$, including binary white noise. However, if we start with the binary white noise $\alpha(n)$, the sequence $\beta(n)$ is obviously a nonbinary one, since the value of $\beta(n)$ in Eq. (1) results from a linear superposition of binary entries. Thus, the direct application of the convolution method does not generate a binary correlated sequence.

Let us now consider the problem of generating a binary sequence $\varepsilon(n)$ with prescribed correlations, assuming that the sequence $\alpha(n)$ is also binary. We suppose that both the sequence $\varepsilon(n)$ and $\alpha(n)$ contain 0's and 1's. Let the nth site for $\varepsilon(n)$ be associated with the number P_n $(0 \leq P_n \leq 1)$, which is the probability of 1 obtaining at this site. In order to calculate the *filtering probabilities* P_n from white noise $\alpha(n)$, in analogy with Eq. (1) we propose the linear transformation

$$P_n = \overline{\varepsilon} + \sqrt{C_\varepsilon(0)/C_\alpha(0)} \sum_{n'=-\infty}^{\infty} F(n-n')[\alpha(n')-\overline{\alpha}]. \qquad (9)$$

Here $C_\varepsilon(0)=\overline{\varepsilon}(1-\overline{\varepsilon})$ and $C_\alpha(0)=\overline{\alpha}(1-\overline{\alpha})$ are the variances of $\varepsilon(n)$ and $\alpha(n)$, respectively, and $F(n)$ is an unknown modulation function to be determined. Having the value of P_n, the nth symbol is generated by drawing randomly a number from the interval $[0,1]$. If this number is less than P_n, then $\varepsilon(n)=1$, otherwise, $\varepsilon(n)=0$. Thus, a binary sequence $\varepsilon(n)$ can be generated, once the set of numbers $\{P_n\}_{n=-\infty}^{\infty}$ is known due to Eq. (9). The values of P_n are correlated in the following way:

$$\overline{[P_{n+r}-\overline{\varepsilon}][P_n-\overline{\varepsilon}]} = C_\varepsilon(0)\sum_{n=-\infty}^{\infty} F(n)F(n+r) \quad \text{for } r \neq 0.$$

$$(10)$$

According to the method of generation of the sequence $\varepsilon(n)$ from the filtering probabilities, the probability of the symbol $\varepsilon(n+r)$ obtaining at the $(n+r)$th site does not depend on the emergence of the symbol $\varepsilon(n)$ at the nth site (for $r \neq 0$). Therefore, the product $P_{n+r}P_n$ gives the joint probability of 1 obtaining at the nth and $(n+r)$th sites. If 0 appears at either of these sites, the corresponding pair does not contribute to the product $\varepsilon(n+r)\varepsilon(n)$. Hence, the correlation function for sequence $\varepsilon(n)$ can be expressed through the correlation function of the filtering probabilities

$$\overline{[\varepsilon(n+r)-\bar{\varepsilon}][\varepsilon(n)-\bar{\varepsilon}]} = \lim_{N\to\infty} \frac{1}{2N+1}\sum_{n=-N}^{N} P_{n+r}P_n - \bar{\varepsilon}^2$$

$$= \overline{[P_{n+r}-\bar{\varepsilon}][P_n-\bar{\varepsilon}]}. \quad (11)$$

As one can see, the correlations in the sequence $\varepsilon(n)$ occur because of the correlations between the filtering probabilities. The latter are enforced by the modulation function [see Eq. (9)]. Thus, using Eqs. (10) and (11) the relation between the correlator of the binary sequence and the modulation function can be written as

$$K_\varepsilon(r) = \sum_{n=-\infty}^{\infty} F(n)F(n+r) \quad \text{for } r \neq 0, \quad (12)$$

$$K_\varepsilon(0) = 1. \quad (13)$$

The normalization condition (13) is the property of the correlator $K_\varepsilon(r)$. It should be stressed that, unlike Eq. (2), which is valid for all values of r including $r=0$, in the case of binary sequences the derived Eq. (12) is not valid for $r=0$. Therefore, the sum $\sum_n F^2(n)$ remains undefined and has to be considered as a free constant,

$$A = \sum_{n=-\infty}^{\infty} F^2(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{F}^2(k)dk. \quad (14)$$

This constant appears now in the Fourier transform of Eq. (12) as follows:

$$\mathcal{K}_\varepsilon(k) = 1 - A + \mathcal{F}^2(k). \quad (15)$$

Using Eq. (15), the relation between the modulation function and the pair correlator of the binary sequence can be written in the form

$$F(n) = \frac{1}{\pi}\int_0^{\pi} [\mathcal{K}_\varepsilon(k) - 1 + A]^{1/2} \cos(kn)dk. \quad (16)$$

Thus, Eqs. (9) and (16) define the algorithm of the generation of a binary sequence with the prescribed pair correlator.

The values of P_n in Eq. (9) can be associated with probability if $0 \leq P_n \leq 1$. This condition leads to the following constraints for the unknown constant A,

$$\sum_{n=-\infty}^{\infty} |F(n)| \leq \frac{\min(\bar{\varepsilon}, 1-\bar{\varepsilon})}{\sqrt{\bar{\varepsilon}(1-\bar{\varepsilon})}} \frac{\sqrt{\bar{\alpha}(1-\bar{\alpha})}}{\max(\bar{\alpha}, 1-\bar{\alpha})} \leq 1. \quad (17)$$

Taking into account that the argument of the square root in Eq. (16) must be positive, this inequality can be rewritten (for $\bar{\alpha}=\bar{\varepsilon}=1/2$) in terms of the constant A,

$$1 - \mathcal{K}_\varepsilon(k) \leq A = \sum_{n=-\infty}^{\infty} F^2(n) \leq \sum_{n=-\infty}^{\infty} |F(n)| \leq 1. \quad (18)$$

The constraints imposed by Eqs. (17) and (18) limit the class of generated binary sequences that can have a given pair correlator. In particular, a binary sequence with the slow-decaying correlator $K_\varepsilon(r)=\sin(ar)/ar$ cannot be generated with the use of the proposed algorithm, since the sum $\sum_n |F(n)|$ diverges in this case. As is known, the power-law

decaying correlators provide the emergence of a kind of mobility edge in systems with random one-dimensional potentials [7,16]. In this case, a sharp transition from localized to delocalized eigenstates occurs when crossing some value in the energy spectrum, specified by the mobility edge. Although such mobility edges have been experimentally observed for a continuous distribution of fluctuations in artificially fabricated site potentials with long-range correlations [15], it is not clear yet if they do exist for a correlated binary sequence. That is why the existence of a mobility edge in a sequence of nucleotides in a DNA molecule is still questionable [13,17].

As an example of short-range correlations, let us consider the exponential binary correlator with the corresponding Fourier representation,

$$K_\varepsilon(r) = \exp(-\gamma|r|), \quad \mathcal{K}_\varepsilon(k) = \sinh\gamma/(\cosh\gamma - \cos k). \quad (19)$$

Since $\mathcal{K}_\varepsilon(k)$ reaches its minimum at $k=\pi$, the condition $1 - \mathcal{K}_\varepsilon(k) \leq A$ is satisfied for $A = 1 - \mathcal{K}_\varepsilon(\pi) = 1 - \tanh(\gamma/2)$. Therefore, we have

$$\mathcal{F}^2(k) = [(1 + \cos k)/(\cosh\gamma - \cos k)]\tanh\gamma/2, \quad (20)$$

and the function $F(n)$ reads

$$F(n) = \frac{1}{\pi}\sqrt{\tanh\frac{\gamma}{2}}\int_0^{\pi} \cos(kn)\sqrt{\frac{1 + \cos k}{\cosh\gamma - \cos k}}dk. \quad (21)$$

For $n \gg 1$ the modulation function decays as follows:

$$F(n) \approx [(-1)^{n+1}/2n^2\cosh(\gamma/2)]\sqrt{\tanh\gamma/2} \propto 1/n^2; \quad (22)$$

therefore, the sum $\sum_n |F(n)|$ converges. However, it exceeds 1 for $\gamma < \gamma_{cr} \approx 1.60$, thus violating the last inequality in Eq. (18). The numerical simulation shows that for $\gamma > \gamma_{cr}$ the method works quite well, giving the possibility of designing binary chains with exponentially decaying correlations.

Since a method of generation of a binary sequence with exponentially decaying correlations is already known [1], it is important to demonstrate the validity of the proposed algorithm for generation of binary sequences with long-range correlations. Let us consider a power-law decaying correlation function

$$K_\varepsilon(n) = (\pm 1)^n a/n^2, \quad n \neq 0. \quad (23)$$

For these correlators, the modulation function can be explicitly calculated,

$$F(n) = \pm\sqrt{a/2}(1/\pi n^2)[1 - (-1)^n], \quad F(0) = (\pi/2)\sqrt{a/2}. \quad (24)$$

The constant a is obtained from Eq. (18), which leads to the following inequality:

$$\sum_{n=-\infty}^{\infty} |F(n)| = \frac{3\pi}{4}\sqrt{\frac{a}{2}} \leq 1. \quad (25)$$

It is satisfied for $0 < a \leq 32/(9\pi^2) \approx 0.36$. For numerical calculations, we chose $a=0.2$. Two filtering functions Eq. (24) were used to generate two binary sequences of length 10^5 each. The correlators of these sequences are plotted in Fig. 1. A good agreement between the theoretical Eq. (23) and nu-
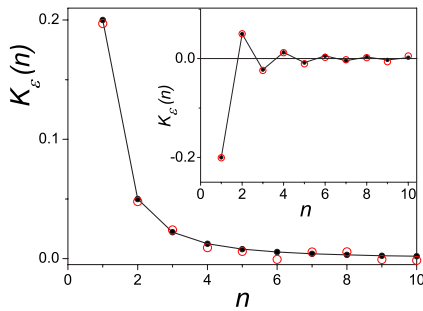
FIG. 1. (Color online) Correlation function of the binary sequence with power-law decaying correlations. The inset shows the correlation function with alternating sign. Solid circles are for the theoretical formulas Eq. (23), and open circles are for the numerical data for generated sequences of length 10^5. The value $K(0)=1$ is not shown.

merical results is obtained for the sequence with monotonically decaying correlations and for the sequence with alternating-sign correlations. Note that the sequences with monotonically power-law decaying correlations are used for modeling of superdiffusion [6], and anticorrelations [i.e., a long negative tail for $K_\beta(r)$] are necessary in the fluctuations of the force, driving a subdiffusion process [8].

Thus, the proposed method can be used for generation of binary sequences with long-range correlations. To the best of our knowledge, this is the first successful attempt to generate a random binary sequence with power-law decaying correlations from white noise. It is worth mentioning that a *renewal* binary sequence with prescribed correlations can be generated using the relation $d^2K_\varepsilon(n)/dn^2=\psi(n)$ between the correlation function $K_\varepsilon(n)$ and the probability density $\psi(n)$ of

occurrence of a laminar layer of length n [18]. The binary sequences that can be designed from this relation are those with $K_\varepsilon''(n)>0$. This constraint does not allow generation of sequences with alternating-sign correlations. Numerous applications of the aforementioned relation for stochastic dynamical processes with intermittency can be found in Ref. [19].

As compared to a sequence of continuously distributed random elements, a binary sequence leaves much less freedom to implant correlations. Therefore, the proposed modification of the convolution method allows generation of random binary sequences with a much narrower class of correlation functions than the traditional convolution method. In the latter case, practically any function, decaying at $r\rightarrow\infty$ and possessing a nonnegative Fourier transformation [see Eq. (4)], can serve as $K_\beta(r)$—a correlator of the random sequence (1). More examples of sequences with short- and long-range correlations generated by the convolution method can be found in Refs. [6–8,15,16].

In conclusion, we suggest a method of filtering probabilities to construct a binary correlated sequence from a white noise. The proposed algorithm consists of the following steps. First, starting from the prescribed mean value $\bar{\varepsilon}$, variance $C_\varepsilon(0)$, and power spectrum $\mathcal{K}_\varepsilon(k)$, one calculates the filtering function $F(n)$ by making use of Eq. (16). The next step is optimization of the value of the constant A according to Eq. (18). Then, the filtering probabilities P_n are calculated from Eq. (9). Finally, for any site n, by comparing the value of P_n with a number drawn randomly from the interval $[0,1]$, one gets the number 0 or 1 that creates the binary sequence $\varepsilon(n)$.

[1] See, e.g., N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).

[2] S. O. Rice, Bell Syst. Tech. J. **23**, 282 (1944); in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954), p. 180.

[3] D. Saupe, in *The Science of Fractal Images*, edited by H.-O. Peitgen and D. Saupe (Springer, New York, 1988); J. Feder, *Fractals* (Plenum Press, New York, 1988).

[4] C. S. West and K. A. O'Donnell, J. Opt. Soc. Am. A **12**, 390 (1995).

[5] A. Czirok *et al.*, Phys. Rev. E **52**, 446 (1995); H. A. Makse *et al.*, *ibid.* **53**, 5445 (1996).

[6] A. Romero and J. Sancho, J. Comput. Phys. **156**, 1 (1999); J. García-Ojalvo and J. Sancho, *Noise in Spatially Extended Systems* (Springer-Verlag, New York, 1999).

[7] F. M. Izrailev and A. A. Krokhin, Phys. Rev. Lett. **82**, 4062 (1999); F. M. Izrailev and N. M. Makarov, J. Phys. A **38**, 10613 (2005).

[8] R. Cakir *et al.*, Phys. Rev. E **74**, 021108 (2006).

[9] See, for example, I. M. Lifshitz *et al.*, *Introduction to the Theory of Disordered Systems* (Wiley, New York, 1988); V. I. Klyatskin, *Dynamics of Stochastic Systems* (Elsevier, Amsterdam, 2005); P. A. Mello and N. Kumar, *Quantum Transport in Mesoscopic Systems: Complexity and Statistical Fluctuations*

(Oxford University Press, Oxford, U.K., 2004).

[10] N. M. Makarov and I. V. Yurkevich, Zh. Eksp. Teor. Fiz. **96**, 1106 (1989) [Sov. Phys. JETP **69**, 628 (1989)] V. D. Freylikher *et al.*, Phys. Rev. B **41**, 8033 (1990).

[11] R. J. Polge and H. Stern, in SOUTHEASTCON '81, Proceedings of the Region 3 Conference and Exhibit, Huntsville, AL, 1981 (IEEE, Piscataway, NJ, 1981), p. 164; R. J. Polge, AGARD Multifunction Radar for Airborne Applications **9**, 11 (1986).

[12] Sh. Hod and U. Keshet, Phys. Rev. E **70**, 015104(R) (2004); I. Avgin, Phys. Rev. B **73**, 052201 (2006).

[13] P. Carpena *et al.*, Nature (London) **418**, 955 (2002); **421**, 764 (2003).

[14] S. S. Melnyk *et al.*, Phys. Rev. E **72**, 026140 (2005); Physica A **361**, 405 (2005); F. M. Izrailev *et al.*, *ibid.* **372**, 279 (2006).

[15] U. Kuhl *et al.*, Appl. Phys. Lett. **77**, 633 (2000).

[16] F. A. B. F. de Moura and M. L. Lyra, Phys. Rev. Lett. **81**, 3735 (1998).

[17] R. A. Caetano and P. A. Schulz, Phys. Rev. Lett. **95**, 126601 (2005); A. Sedrakyan and F. Domínguez-Adame, *ibid.* **96**, 059703 (2006); E. Diaz *et al.*, Phys. Rev. B **75**, 014201 (2007).

[18] T. Geisel *et al.*, Phys. Rev. Lett. **54**, 616 (1985).

[19] P. Grigolini, Adv. Chem. Phys. **133**, 357 (2006).