PHYSICAL REVIEW E, VOLUME 65, 020102(R)

Low autocorrelated multiphase sequences

Liat Ein-Dor,[1] Ido Kanter,[1] and Wolfgang Kinzel[2]
[1]*Department of Physics, Bar-Ilan University, Ramat-Gan, 52900, Israel*
[2]*Institut für Theoretische Physik, Universität Würzburg Am Hubland, D-97074 Würzburg, Germany*
(Received 26 February 2001; published 15 January 2002)

The interplay between the ground-state energy of the generalized Bernasconi model to multiphase, and the minimal value of the maximal autocorrelation function, $C_{\max}=\max_K|C_K|$, $K=1,\ldots,N-1$, is examined analytically in the thermodynamic limit where the main results are (a) For the binary case, the minimal value of $C_{\max}$ over all sequences of length N, $\min C_{\max}$, is $0.435\sqrt{N}$, significantly smaller than the typical value for random sequences $O(\sqrt{\log N}\sqrt{N})$. (b) A new method to approximate $F_{\max}$ is obtained using the observation of data collapse. (c) $\min C_{\max}$ is obtained in an energy which is about 30% above the ground-state energy of the generalized Bernasconi model, independent of the number of phases m. (d) For a given m, $\min C_{\max}\propto\sqrt{N/m}$ indicating that for $m=N$, $\min C_{\max}=1$, i.e., a generalized Barker code exists. The analytical results are confirmed by simulations.

DOI: 10.1103/PhysRevE.65.020102 PACS number(s): 05.20.−y, 87.10.+e

In many applications of communication science [1,2], as well as a variety of other fields, it is necessary to find sequences with low autocorrelation. Some of these applications utilize the pulse compression feature of low autocorrelated sequences to obtain high resolution in radars and sonars. In other applications, the shifts of such periodic sequences can be used to identify users in multiuser systems. Due to their importance, low autocorrelated sequences have evoked a wide spread interest accompanied by the development of various methods for constructing such sequences [1].

In order to construct a binary sequence $S=(s_1,\ldots,s_N)$ with low off-peak autocorrelations, one has to define which quantity has to be minimized. The different applications divide the low autocorrelated sequences into two types in correspondence with the quantity which they minimize. The first kind of low autocorrelated sequences minimizes the Hamiltonian of the Bernasconi model [3–8] which is given by

$$H=\sum_{K=1}^{N-1} C_K^2, \tag{1}$$

where for the case of nonperiodic boundary conditions, which is at the center of our study,

$$C_K=\sum_{j=1}^{N-K} s_j s_{j+K}, \tag{2}$$

and for the periodic case $C_K=\sum_{j=1}^{N} s_j s_{[(j+K-1)\bmod N]+1}$. This is a deterministic model without explicit disorder. However, since its ground states are highly disordered, this model exhibits features of a glass transition like a jump in the specific heat [3] and slow dynamics and aging [4]. Note that for random sequences the average value of H in the nonperiodic case is $N^2/2$, whereas for low autocorrelated sequences this energy is reduced by a merit factor $F>1$ to $N^2/2F$.

However, there are applications for which the maximal off peak autocorrelation, $C_{\max}=\max_K|C_K|$, has to be minimized. The second kind of low autocorrelated sequences are the solutions of this minimization problem. Note that for random sequences $C_{\max}$ is typically $O(\sqrt{\log N}\sqrt{N})$. For a sequence of length N, the maximal possible ratio between the peak, $|C_0|=N$, and the maximal off-peak autocorrelations, $|C_K|$ with $K=1,\ldots,N-1$ is $N/1$. The only known binary sequences with this ratio are the Barker codes of length 2, 3, 4, 5, 7, 11, and 13. Obviously, the Barker sequences [9], when they exist, furnish a minimum for the two minimization problems. However, Turyn [10] has shown that no other binary codes such as this exist for any length less than 144 or for odd length greater than 13. An exact solution for these two minimization problems is known only for systems (up to $N=59$) which are small enough to permit an effective exhaustive search [7]. Extrapolation of the ground-state energies which were found for small systems using an exhaustive search indicates that the maximal value of F, $F_{\max}=\lim_{N\to\infty}N^2/2H_{\min}\sim 8.5$ [7]. Moreover, $F_{\max}$ was conjectured by Golay [11,12] to be bounded from above by 12.324.

The following questions regarding low autocorrelated sequences are still open and are at the center of our study: Do the sequences that minimize $C_{\max}$ minimize the energy of the Bernasconi model as well? In case where the two minimization problems are not equivalent, how far are the energy values of the sequences which minimize $C_{\max}$ from the values of the ground-state energy? Additional interesting questions arise when multiphase sequences, whose terms are complex mth roots of 1 for $m>2$, are considered [13–15]. For such sequences $s_l=\exp(i(2\pi/m)l)$ where $l=1,\ldots,m$ and the correlations are defined as

$$C_K=\sum_{j=1}^{N-K} s_j s_{j+K}^*, \tag{3}$$

where s_{j+K}^* is the conjugate of s_{j+K}. What is the influence of the number of phases both on $F_{\max}$ and on the minimal value of $C_{\max}$, $\min C_{\max}$? Does a generalized Barker code exist for lengths larger than $N=45$ [16] which is the maxi-

mal length for which a multiphase Barker code has been found? How does the minimal number of phases required for a Barker code grow with N?

In this paper we suggest an analytical technique to calculate $F_{\max}$ and $\min C_{\max}$ for both binary and multiphase sequences. Since these two quantities are very important performance criteria in many engineering applications, it is tempting to investigate the relations between F and the minimal upper bound, $\min C_{\max}(F)$, over all sequences in the microcanonical ensemble with a merit factor of F. For our analytical calculation we consider the nonperiodic Bernasconi model. Following Golay's approximation [11] we consider the autocorrelations to be independent variables, where the probability distribution, $P(C_K)$, follows a Gaussian with zero mean and $N-K$ variance. Thus, the probability $P_F(C_{\max})$ that the autocorrelations of a sequence have an upper bound $C_{\max}$ (such that $|C_K| \leqslant C_{\max}$) as well as a merit factor $F=N^2/2H$ is given by

$$P_F(C_{\max}) = \int_{-C_{\max}}^{C_{\max}} \prod_K dC_K P(C_K)\, \delta\left(\frac{1}{N}\sum_{K=1}^{N-1} C_K^2 - \frac{N}{2F}\right). \tag{4}$$

Since there are 2^N distinct binary sequences of length N, it is necessary that $P_F \geqslant 2^{-N}$ in order for a sequence with the corresponding features, namely, a merit factor F and an autocorrelation upper bound $C_{\max}$, to exist. Hence, in order to find $\min C_{\max}(F)$, one has to minimize $P_F(C_{\max})$ under the constraint that $P_F \geqslant 2^{-N}$. To fulfill this constraint, $\min C_{\max}(F)$ must be of $O(\sqrt{N})$. Assigning $\min C_{\max}(F) = B(F)\sqrt{N}$ and inserting the integral representation of δ function in Eq. (4) , the saddle-point method can be used to obtain the following set of equations:

$$\frac{\lambda}{2F} - \frac{1}{\mu^2\lambda}[(1-\mu\lambda)\ln(1-\mu\lambda)+\mu\lambda] - \frac{2}{\mu}\int_0^1 \ln \operatorname{erf}[B(F)g(y,\lambda)]\, dy = \chi, \tag{5}$$

$$\frac{1}{2F} + \frac{1}{\mu^2\lambda^2}[\ln(1-\mu\lambda)+\mu\lambda] + \frac{2}{\mu}\frac{B(F)}{\sqrt{\pi}}\int_0^1 \frac{\exp[-B(F)^2 g(y,\lambda)^2]}{\operatorname{erf}[B(F)g(y,\lambda)]\, g(y,\lambda)}\, dy = 0, \tag{6}$$

where $g(y,\lambda)=\sqrt{1-\mu\lambda(1-y)/\mu(1-y)}$, $\mu=2$, and $\chi \leqslant \ln(2)$ is the maximal value for which a solution exists. Solving numerically this set of equations, it turns out that a solution exists only for a bounded region of F, where $0.213 \leqslant F \leqslant 12.324$. These two limit values of F are alternatively obtained by taking the limits of the integration in Eq. (4) to infinity. Note that the highest autocorrelated sequences, for example (1,1,1, ...), has $F_{\min} \sim 1/N$, which tends to zero as $N \to \infty$ in contradiction to the analytical result $F_{\min} \sim 0.213$. Our analytical method fails to predict the right value of $F_{\min}$ for two main reasons: (a) The assumption that the correlations are independent variables is strongly violated for high autocorrelated sequences. For such sequences the correlations are typically of O(N). Hence given C_K of O(N), $C_{K'}$ must be of O(N) as well for all K' of O(1). (b) The correlations are typically of O(N) and, therefore, are at the tail of the Gaussian distribution well beyond the standard deviation.

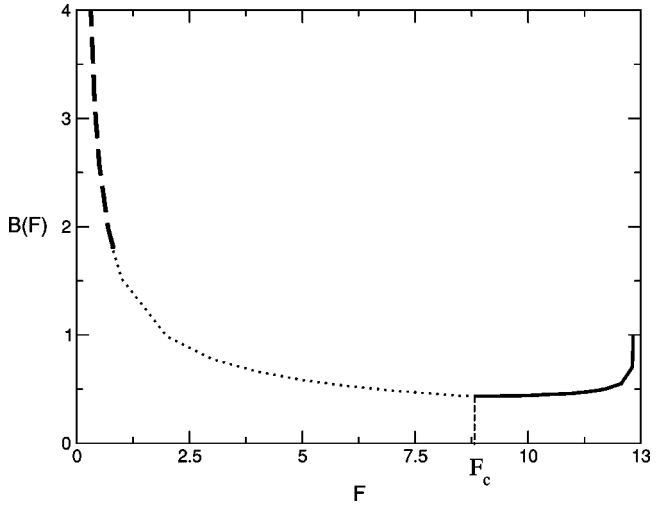

FIG. 1. The behavior of $B(F)$ vs F is composed of three regions. In the dashed region $0.213 \leqslant F \leqslant 0.8$, the fraction of sequences $P_F(C_{\max})$ is getting larger as F increases, while in the dotted-line region, $0.8 < F < F_c$, this fraction decreases to 2^{-N} as F approaches F_c. For $F > F_c$ (solid line) the fraction of sequences with the minimal value $B(F)\sqrt{N}$ is constant and equals 2^{-N}.

The analysis of $B(F)$ reveals that the minimal value of $B(F_c)=0.435$ is obtained for $F_c=8.839$ whereas the maximal merit factor $F=12.324$ corresponds to a higher value of $B(F)=1$. Thus, the merit factor of sequences, which minimizes $C_{\max}$, is around 71% lower than $F_{\max}$. Exhaustive search results in the region $N \in [28,40]$ indicate that the average ratio between the two merit factors, $\langle\langle\langle F_c\rangle\rangle\rangle / F_{\max}\rangle_N$, is ~ 0.653 where $\langle\langle\rangle\rangle$ denotes the average over all sequences which minimize $C_{\max}$ whereas $\langle\rangle_N$ denotes the average over the different N. The deviation from the analytical result ~ 0.717 is attributed to finite-size effects. Figure. 1 shows the behavior of $B(F)$ derived from the numerical solution of Eqs. (5) and (6). The behavior of χ [Eq. (5)] as a function of F divides this graph into three regimes. In the first region, where $F > F_c$, there are only $O(1)$ sequences whose $C_{\max}$ equals $B(F)\sqrt{N}$. In the second region, $0.8 < F < F_c$ the number of sequences increases exponentially as F decreases. This increment terminates at $F \sim 0.8$ and for smaller values of F the number of sequences reduces to $O(1)$ as F approaches 0.213. Aiming to compare between analytical results ($N \to \infty$) for $B(F)$ and simulation results for different values of N, one has to overcome the mismatch between the different domains of F. The rescaling of F by $F_{\max}$ yields a new parameter, $F^* = F/F_{\max}$, whose range of values, (0,1], is independent of N. The rescaling *surprisingly results in data collapse* of the function $B(F^*)$ (Fig. 2). The meaning of the data collapse is that for a given B_0 the corresponding value of F^*, F_0^*, is almost independent of N. This observation

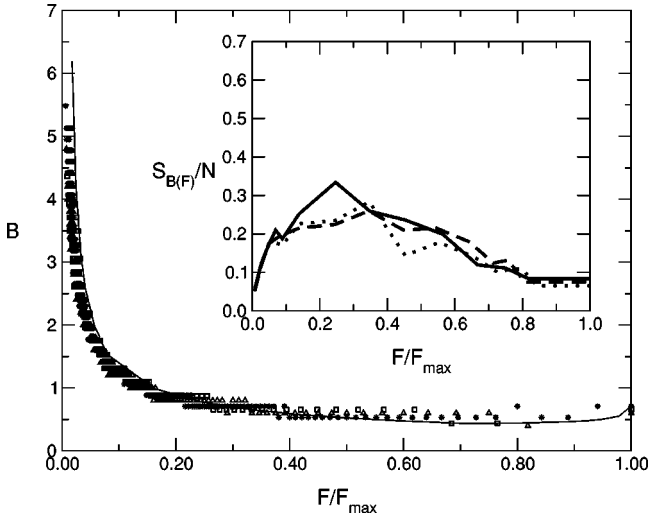

FIG. 2. B vs $F/F_{\max}$. Analytical results (solid line) and exhaustive search results for $N=32$ (triangles), $N=25$ (stars), and $N=21$ (squares). Note that the analytical $F_{\max}$ is 12.324, whereas for $N=32$, 25, and 21, $F_{\max}\sim 8$, 8.68, and 8.48. Inset: The entropy per bit, $S_{B(F)}/N$, of the set of all sequences with a merit factor F and $C_{\max}=B(F)\sqrt{N}$ as a function of $F/F_{\max}$ for $N=32$ (solid), $N=28$ (dashed), and $N=25$ (dotted). The entropies are averaged over windows of size 0.02.

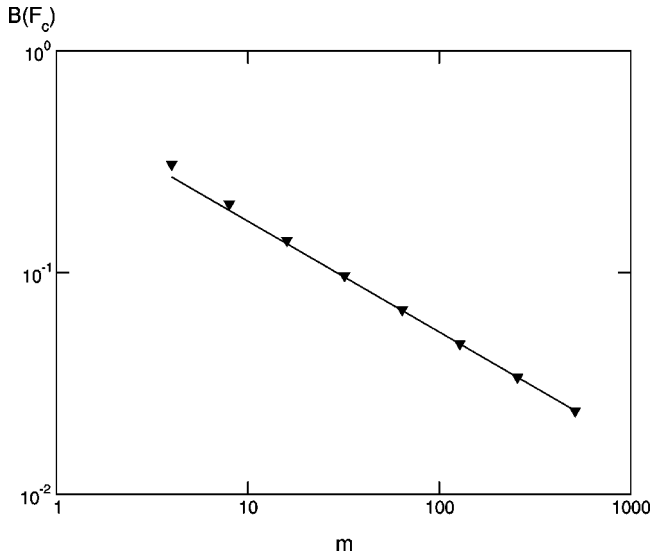

FIG. 3. $B(F_c)$ vs m as was obtained from the numerical solution of Eqs. (5) and (6) with $\mu=1$ and $\chi\leqslant\ln(m)$. The solid line is the least-square fit $0.54/\sqrt{m}$.

provides an alternative way to approximate $F_{\max}(N)$ for N's which are still too large to allow an effective exhaustive search. For such sequence lengths, the traditional way to approximate $F_{\max}(N)$ is to use the simulated annealing method to find sequences which maximize F. However, using our data collapse, the problem of finding $F_{\max}(N)$ can be solved indirectly by choosing a certain $F_0^*\in(0,1]$ and then finding the minimal F under the constraint that $C_{\max}=B_0\sqrt{N}$, where $B_0=B(F_0^*)$ (for $F_0^*=0.05$ and $N=100$, for instance, $B_0=2$ independent of N, and the constraint is $C_{\max}=2\times 10$). This minimal F, which we denote by $F(B_0,N)$, when divided by F_0^* yields an approximated upper bound for $F_{\max}(N)$. The problem of finding $F(B_0,N)$ is easier than finding $F_{\max}(N)$ directly because as we show both analytically and numerically, there are regions of F in which there is an exponential number of sequences with an upper bound $B(F/F_{\max})\sqrt{N}$. We have applied the simulated annealing method to find $F(B_0,N)$ for $N=100,144$, and $B_0=2$. The resultant upper bounds $F_{\max}(100)\leqslant 7.4120$ and $F_{\max}(144)\leqslant 8.92$ are much closer to the values obtained from extrapolation of the exhaustive search results [7] than the values obtained using the traditional method.

Moreover, the existence of the three aforementioned regions of $P_F(C_{\max})$ is confirmed by counting the number of sequences with the minimal autocorrelation upper bound $B(F)\sqrt{N}$ for each of the microcanonical ensembles with a merit factor F. The entropy per bit of such sequences, $S_{B(F)}/N$, is demonstrated in the inset of Fig. 2, for $N=25$, 28, and 32, where the merit factor is rescaled to $F/F_{\max}$. Note that for finite sequences, the entropies of sequences with a given $B(F)$ have large fluctuations. To overcome these fluctuations, we have averaged the entropies over windows of size 0.02. In the small F regime, the average number of sequences increases with F and then decreases up to the *plateau* region in which the number of sequences is $O(1)$. Now we turn to study the influence of the number of phases both on the maximal value of F and on the minimal value of $C_{\max}$. We assign D_K and E_K to be the real and the imaginary parts of C_K, respectively. In case that $m=2^p$ where p is an integer number greater than 1, the probability distributions, $P(D_K)$ and $P(E_K)$, follow the same Gaussian with zero mean and $N-K$ variance. Under the assumption that D_K and E_K are independent variables together with the same assumptions which are used in the binary case, Eq. (4) becomes

$$P_F^m(C_{\max})=\int_{-C_{\max}}^{C_{\max}}\prod_K dD_K dE_K P(D_K)P(E_K)\times\delta\left[\frac{1}{N}\sum_{K=1}^{N-1}(D_K^2+E_K^2)-\frac{N}{2F}\right]. \tag{7}$$

A similar procedure to that which is used to calculate $B(F)$ for the binary case results in Eqs. (5) and (6) with $\mu=1$ and $\chi\leqslant\ln(m)$. Solving numerically these two equations, the behaviors of $B(F_c)=\min C_{\max}(F)/\sqrt{N}$ and $F_{\max}$ as a function of m are obtained. Figure 3 shows that $B(F_c)\propto 1/\sqrt{m}$ with slight deviations for small m. This relation between $B(F_c)$ and m implies that such a relation holds for $B(F_{\max})$ as well. Since the variance of $P(C_K)$ equals $N-K$, the typical values of C_K drop linearly with K. Hence, $H=N^2/2F$ is approximated by a sum of an algebraic series of $N-1$ terms, C_K^2, with an autocorrelation upper bound, $C_{\max}^2$, of $O(N/m)$. A homogeneous distribution of C_K^2 between 0 to $C_{\max}^2$ yields a linear increment of $F_{\max}$ as a function of m in agreement with the numerical solution of $F_{\max}$, which is depicted in the inset of Fig. 4. The results of $F_{\max}$ and $B(F_c)$ as a function of m

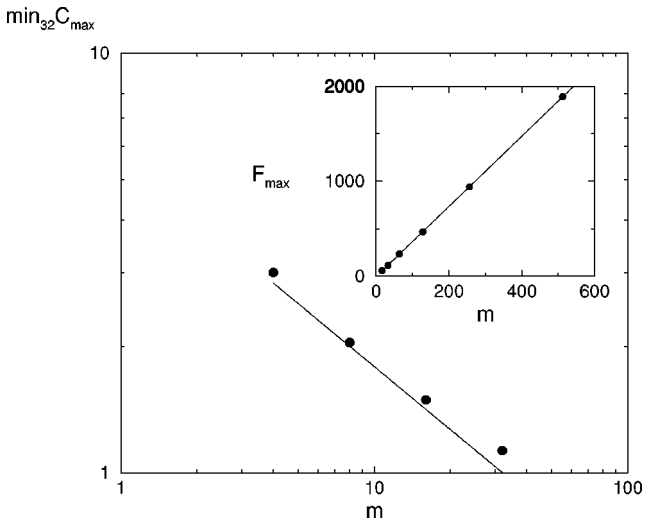

FIG. 4. Results for $\min C_{\max}$ as a function of m for $N=32$. The filled circles stand for the simulated annealing results. The deviation from the analytical results (solid line) is attributed to finite-size effects and to the suboptimal solution obtained in our limited running times of the simulations. Inset: Analytical results for $F_{\max}$ as a function of m obtained from the numerical solution [Eqs. (5) and (6)]. The solid line is the least-square fit $3.7m$.

show that the solutions of the two minimization problems are improved by increasing the number of phases m.

These results raise the question whether it is possible to increase m such that $\min_N C_{\max}$ becomes 1. An asymptotic expansion of Eqs. (5) and (6) with $\mu=1$ and $\chi \leqslant \ln(m)$ reveals that for $m=N$, $\min_N C_{\max}=1$. Note that $|C_{N-1}|$ is always 1 and, therefore, $\min_N C_{\max}=1$ holds for the entire regime $m>N$.

We used the simulated annealing method for sequences of length $N=32$ with different number of phases m, to find $\min C_{\max}(32)$. The relatively small sequence length was chosen to enable an appropriate scan of the configuration space in reasonable computational time. The results are displayed in Fig. 4, and support the anticipated behavior $\min C_{\max} \propto \sqrt{N/m}$. For $m=32$, there is a deviation of $\min C_{\max}$ from the analytical prediction $\min C_{\max}=1$, probably since the simulated annealing method yields only suboptimal solutions.

Finally, we would like to examine the simulated annealing method in light of the results of this study. Simulations show that for the same running times, the simulated annealing method yields $C_{\max}$, which is closer to its minimal value than F. This can be explained by the larger degeneracy of $C_{\max}$ compared with that of F. Moreover, it turns out that in order to minimize $C_{\max}$, it is preferable to start the searching process with the minimization of the energy function, $\Sigma_{K=1}^{N-1} C_K^2$, and then replace it with $C_{\max}$ [17]. In this way the system avoids the plateaus which characterize the landscape of $C_{\max}$ in the configuration space. However, future research is necessary to find out how the results of this study can be applied to further improve the searching processes of low autocorrelated sequences.

Critical comments on the manuscript by R. Metzler are acknowledged. The research is supported in part by the Israel Academy of Science.

[1] M. R. Schroeder, *Number Theory in Science and Communication* (Springer Verlay, Berlin, 1984).
[2] I.I. Shapiro, G.H. Pettengill, M.E. Ash, M.L. Stone, W.B. Smith, R.P. Ingalls, and R.A. Brockelman, Phys. Rev. Lett. **20**, 1265 (1968).
[3] J. Bernasconi, J. Phys. **48**, 559 (1987).
[4] W. Krauth and M. Mezard, Z. Phys. B: Condens. Matter **97**, 127 (1995).
[5] J.P. Bouchaud and M. Mezard, J. Phys. I **4**, 1109 (1994).
[6] E. Marinari, G. Parisi, and F. Ritort, J. Phys. A **27**, 7615 (1994).
[7] S. Mertens, J. Phys. A **29**, 473 (1996).
[8] S. Mertens and C. Bessenrodt, J. Phys. A **31**, 3731 (1998).
[9] R. H. Barker, *Communication Theory* (Butterworths, London, England 1953).
[10] R. Turyn, Sylvania Electronic Systems, Final Rept. Contract, AF19(604)-5473, January, 1960.
[11] M.J.E. Golay, IEEE Trans. Inf. Theory **IT-28**, 543 (1982).
[12] M.J.E. Golay, IEEE Trans. Inf. Theory **IT-29**, 934 (1983).
[13] N. Zhang and S.W. Golomb, IEEE Trans. Inf. Theory **39** (3), 1085 (1993).
[14] M. Friese, IEEE Trans. Inf. Theory **42**, 1248 (1996).
[15] W.H. Mow and S.-Y.R. Li, IEEE Trans. Inf. Theory **43** (3), 1000 (1997).
[16] A.R. Brenner, Electron. Lett. **32** (15), 1364 (1996).
[17] L. Ein-Dor (unpublished).