Typical Peak Sidelobe Level of Binary Sequences

Typical Peak Sidelobe Level of Binary Sequences

Noga Alon, Simon Litsyn, *Senior Member, IEEE*, and Alexander Shpunt

***Abstract*—For a binary sequence $S_n = \{s_i : i = 1, 2, \ldots, n\} \in \{\pm 1\}^n$, $n > 1$, the peak sidelobe level (PSL) is defined as**

$$M(S_n) = \max_{k=1,2,\ldots,n-1} \left| \sum_{i=1}^{n-k} s_i s_{i+k} \right|.$$

It is shown that the distribution of $M(S_n)$ is strongly concentrated, and asymptotically almost surely

$$\gamma(S_n) = \frac{M(S_n)}{\sqrt{n \ln n}} \in [1 - o(1), \sqrt{2}].$$

Explicit bounds for the number of sequences outside this range are provided. This improves on the best earlier known result due to Moon and Moser that the typical $\gamma(S_n) \in [o(\frac{1}{\sqrt{\ln n}}), 2]$, and settles to the affirmative the conjecture of Dmitriev and Jedwab on the growth rate of the typical peak sidelobe. Finally, it is shown that modulo some natural conjecture, the typical $\gamma(S_n)$ equals $\sqrt{2}$.

***Index Terms*—Aperiodic autocorrelation, concentration, peak sidelobe level (PSL), random binary sequences autocorrelation, second moment method.**

I. Introduction and Definitions

LET $S_n = \{s_i : i = 1, 2, \ldots, n\} \in \mathcal{A}_n, n > 1$, where $\mathcal{A}_n = \boldsymbol{F}^n, \boldsymbol{F} \equiv \{+1, -1\}$. Define

$$M_k(S_n) = \sum_{i=1}^{n-k} s_i s_{i+k}, \quad k = 1, 2, \ldots, n-1.$$

The peak sidelobe level (PSL) $M(S_n)$ of a sequence S_n, is

$$M(S_n) = \max_{k=1,2,\ldots,n-1} |M_k(S_n)|, \quad n > 1.$$

Let μ_n stand for the optimal value of the PSL over the set $\mathcal{A}_n$

$$\mu_n = \min_{S_n \in \mathcal{A}_n} M(S_n).$$

Binary sequences with low PSL are important for synchronization, communications, and radar pulse design, see, e.g., [6], [7], [12], [21], [22]. In theoretical physics, study of the PSL landscape was introduced by Bernasconi via the so-called Bernasconi model [3], which is fascinating for the fact of being completely deterministic, but nevertheless having highly disordered ground states (sequences with the largest merit factor) and thus possessing similarities to the real glasses, with many features of a glass transition exhibited [3], [11].

Study of the problem started in the 1950s. Special attention has been given to the estimation of typical PSL. Since this is our central interest in this paper, let us mention several relevant results. Moon and Moser [19] proved that for almost all sequences

$$\kappa(n) \le M(S_n) \le (2 + \epsilon)\sqrt{n \ln n}$$

for any $\kappa(n) = o(\sqrt{n})$. Mercer [18] showed that

$$\mu_n \le (\sqrt{2} + \epsilon)\sqrt{n \ln n}.$$

Apparently, the suggested approach also allows proving that this bound is indeed true for most of the sequences (see comments in [18, bottom of p. 670]). Dmitriev and Jedwab [4] conjectured and provided experimental evidence that the typical PSL behaves as $\Theta(\sqrt{n \ln n})$. The same was presumed without proof by Ein-Dor, Kanter, and Kinzel [5].

In this paper, we prove that indeed, for almost all binary sequences S_n of length n, $M(S_n) = \Theta(\sqrt{n \ln n})$. Moreover, it is shown that asymptotically almost surely

$$\gamma(S_n) = \frac{M(S_n)}{\sqrt{n \ln n}} \in [1 - o(1), \sqrt{2}]. \tag{1}$$

The results of the paper have application to another problem related to estimation of the "level of randomness" of finite sequences from $\boldsymbol{F}^n$. Mauduit and Sárközy [17] introduced the correlation measure of order r, which is defined for a sequence S_n as

$$C_r(S_n) = \max_{0 \le i_1 < \cdots < i_r \le n-1} \max_{k=1,2,\ldots,n-i_r} \left| \sum_{i=1}^{k} s_{i_1+i} s_{i_2+i} \cdots s_{i_r+i} \right|.$$

In other words, this is just the maximum of the absolute value of the mutual correlation of r continuous runs of vector's entries. In [1], Alon, Kohayakawa, Mauduit, Morreira, and Rödl showed that asymptotically almost surely

$$\sigma_r(S_n) := \frac{C_r(S_n)}{\sqrt{n \ln \binom{n}{r}}} \in \left(\frac{2}{5}, \frac{7}{4}\right). \tag{2}$$

Noticing that $M(S_n) \le C_2(S_n)$, we conclude that any lower bound on the typical $M(S_n)$ is a lower bound for the typical $C_2(S_n)$ as well. Therefore, our results (slightly) improve the lower bound on

$$\liminf_{S_n, n \to \infty} \sigma_2(S_n) = \frac{2}{5} \text{ in (2) to } \frac{1}{\sqrt{2}} \approx 0.7.$$

Manuscript received August 22, 2008; revised February 04, 2009. Current version published December 23, 2009. This work was supported in part by a USA-Israeli BSF grant, by a grant from the Israel Science Foundation, and by an ERC advanced grant. The work of S. Litsyn was supported in part by ISF under Grant 1177/06. The work of A. Shpunt was supported in part by the Di Capua MIT Graduate Fellowship.

N. Alon is with the Schools of Mathematics and Computer Science, Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv, Ramat-Aviv 69978, Israel (e-mail: nogaa@math.tau.ac.il).

S. Litsyn is with the Department of Electrical Engineering–Systems, Tel-Aviv University, Tel-Aviv, Ramat-Aviv 69978, Israel (e-mail: litsyn@eng.tau.ac.i)l.

A. Shpunt is with the Department of Physics, the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: ashpunt@mit.edu).

Communicated by G. Gong, Associate Editor for Sequences.

Digital Object Identifier 10.1109/TIT.2009.2034803

The same improvement can be easily achieved for any r using the method in the paper.

The paper proceeds as follows. In the next short section we sketch a quick proof, based on the approach in [1], that for almost all sequences S_n, $(1-o(1))\sqrt{n\ln n} \le M(S_n) \le (\sqrt{2}+o(1))\sqrt{n\ln n}$. We then proceed to give a detailed analysis that provides a somewhat better control of the error terms in the above estimates. In Section III, we recall a theorem due to Moon and Moser [19] for the number of sequences S_n such that $M_k(S_n)=r$ for any $k=1,2,\dots,n-1$, and $r=-n,\dots,n-1,n$. We then provide estimates for binomial coefficients allowing approximation of the *Moon–Moser* formula by tails of the Gaussian distribution with vanishing error. Section IV is devoted to proving the upper bound in (1). To do so, we relate, via the *Moon–Moser* theorem, the number of sequences S_n with $M(S_n) > \sqrt{2n(\ln n+\delta(n))}$ to certain binomial sums. Accurate estimates using bounds developed in Section III allow to establish the sought inequality. Section V derives a lower bound for the number of sequences S_n with $M(S_n) > \sqrt{n(\ln n+\delta(n))}$, by looking only at autocorrelations with shifts $\ge n/2$. This allows to consider $M_k(S_n), k=n/2+1,\dots,n$ as a collection of linear forms with coefficients $s_1,\dots,s_{n/2}$ and variables $s_{n/2+1},\dots,s_n$. We then apply the *Azuma* inequality to show concentration of $M(S_n)$, and the results of Sections IV and V to accurately locate the mean of $M(S_n)$, and thus establish the lower bound in (1). In Section VI, we argue that modulo a plausible conjecture, and using the *Azuma* inequality, for most S_n

$$\gamma(S_n)=\sqrt{2}(1+o(1)).$$

We attempted to make the paper as self-contained as possible. To achieve this, we have included several sketchy proofs of relevant results from other papers conveying ideas of importance for our presentation.

II. A Quick Sketch

In this short section, we sketch a quick proof that for almost all sequences S_n

$$(1-o(1))\sqrt{n\ln n} \le M(S_n) \le (\sqrt{2}+o(1))\sqrt{n\ln n}.$$

The upper bound is simple (see, e.g., [18]); for each fixed k, the sum $M_k(S_n)$ is easily seen to be a sum of $n-k$ independent random variables, each attaining the values -1 and 1 with equal probability. It thus follows by standard estimates (cf., for example, [2, Corollary A.1.2]) that for a random sequence S_n, the probability that $|M_k(S_n)| > a$ is at most $2e^{-a^2/2(n-k)} < 2e^{-a^2/2n}$. Thus, for $a=(\sqrt{2}+\delta)\sqrt{n\ln n}$, where $\delta>0$ is arbitrarily small, this probability is much smaller than $1/n$ (for all sufficiently large n), and it thus follows that with high probability, all n numbers $M_k(S_n)$ are smaller than $(\sqrt{2}+\delta)\sqrt{n\ln n}$, providing the required upper bound.

To prove the lower bound, we consider only values of k satisfying $k>n/2$. It turns out that for k,ℓ which are both bigger than $n/2$, the $2n-(k+\ell)$ products

$$s_1s_{1+k}, s_2s_{2+k},\dots,s_{n-k}s_n, s_1s_{1+\ell}, s_2s_{2+\ell},\dots,s_{n-\ell}s_n$$

are random and independent members of $\{-1,1\}$. A detailed proof of this simple yet somewhat surprising fact appears in Section V. For each $k, n/2<k\le n/2+n/\ln n$, let X_k denote the indicator random variable whose value is 1 if the event $|M_k(S_n)| \ge (1-\delta)\sqrt{n\ln n}$ (which we denote here by E_k) occurs, and is 0 otherwise. Our objective is to show that asymptotically almost surely, the sum $X=\sum_{n/2<k\le n/2+n/\ln n} X_k$ is positive. By standard estimates, for each fixed k, the probability that E_k occurs is bigger than, say, $\ln^2 n/n$ (for every fixed $\delta>0$ and all sufficiently large n.) This means that the expectation $E(X)$ of X is at least $\ln n$. The crucial point is that since the indicator random variables X_k are pairwise independent, the variance $\mathrm{Var}(X)$ of X is the sum of variances of the variables X_k, and is thus smaller than the expectation of X. Therefore, by Chebyshev's inequality, the probability that X is zero is at most $\mathrm{Var(X)}/(E(X))^2 < 1/E(X) < 1/\ln n$, implying that asymptotically almost surely X is positive, as needed.

The detailed proof, with a more careful treatment of the error terms, is given in the next sections. We present The proof of the lower bound that we present in Section V is slightly different than the one indicated above, as it seems interesting to describe an alternative approach which derives the bound by combining the pairwise independence of the random variables described above with Azuma's inequality.

III. Auxiliary Results

Let $g(n,k,r)$ denote the number of sequences S_n, such that $M_k(S_n)=r$. Throughout, we shall adopt the convention that the binomial coefficient $\binom{m}{x}$ equals 0 if x is not an integer and 1 if $m=x=-1$.

Theorem 3.1 (Moon–Moser [19]): For $r=-n,\dots,n-1,n$, and $k=1,\dots,n-1$

$$g(n,k,r)=2^k\binom{n-k}{(n-k)\cdot\left(\frac{1}{2}+\frac{r}{2(n-k)}\right)}.$$

Proof: See a sketch in the Appendix. □

Note that one of the consequences of Theorem 3.1 is that $g(n,k,-r)=g(n,k,r)$.

In the derivation of our bounds, we will need the following estimates for binomial coefficients.

Lemma 3.2: For $0<\epsilon_1<\sqrt{3/32}$, and all n, such that $n\cdot(\frac{1}{2}-\epsilon_1)$ is an integer

$$2^{-n}\cdot\binom{n}{n\cdot\left(\frac{1}{2}-\epsilon_1\right)} \le \left(1+\varsigma_1^{(\epsilon_1)}\right)\cdot\sqrt{\frac{2}{\pi n}}\cdot e^{-2n\epsilon_1^2} \tag{3}$$

$$\varsigma_1^{(\epsilon_1)}=3\epsilon_1^2. \tag{4}$$

Moreover, for $0<\epsilon_1<(2n)^{-1/4}$ and $n\ge 164$, such that $n\cdot(\frac{1}{2}-\epsilon_1)$ is an integer

$$2^{-n}\cdot\binom{n}{n\cdot\left(\frac{1}{2}-\epsilon_1\right)} \ge \left(1-\varsigma_2^{(\epsilon_1,n)}\right)\cdot\sqrt{\frac{2}{\pi n}}\cdot e^{-2n\epsilon_1^2} \tag{5}$$

$$\varsigma_2^{(\epsilon_1,n)}=\frac{3}{2}n\epsilon_1^4+\frac{1}{2n}. \tag{6}$$

Proof: See the Appendix. □

The next result addresses the question of how well sums of binomial coefficients can be approximated by the Gaussian complementary cumulative distribution function (CCDF). Henceforth, the Gaussian CCDF is defined by

$$P_G(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt = \frac{1}{\sqrt{\pi}} \int_{x/\sqrt{2}}^\infty e^{-t^2} dt.$$

Lemma 3.3: Let $n \geq 164$

$$S(n, d) = \sum_{k=\frac{d}{2}}^{\frac{n}{2}} \binom{n}{n \cdot \left(\frac{1}{2} - \frac{k}{n}\right)}.$$

Then, for $\sqrt{n \ln \ln n} < d < n/2$, the following holdş:

$$2^{-n} \cdot S(n, d) \leq \left(1 + \varsigma_3^{(n,d)}\right) \cdot P_G\left(\frac{d}{\sqrt{n}}\right), \quad \varsigma_3^{(n,d)} = \frac{7d}{n} \quad (7)$$

and for $\sqrt{n \ln \ln n} < d < (2n)^{3/4}$

$$2^{-n} \cdot S(n, d) \geq (1 - \varsigma_4^{(n,d)}) \cdot P_G\left(\frac{d}{\sqrt{n}}\right) - e^{-\sqrt{n/32}} \quad (8)$$

$$\varsigma_4^{(n,d)} = \frac{1}{2n} + \frac{5d^4}{n^3}. \quad (9)$$

Proof: See the Appendix. □

The bounds on $S(n, d)$ in Lemma 3.3 are given in terms of $P_G(\frac{d}{\sqrt{n}})$. In certain cases, we would like to provide more explicit bounds, which can be achieved with the following.

Lemma 3.4: For $d > 0, n > 0$

$$\frac{\sqrt{2n}}{\sqrt{\pi} d} \cdot e^{-\frac{d^2}{2n}} \cdot \left(1 - \frac{n}{d^2}\right) \leq P_G\left(\frac{d}{\sqrt{n}}\right) \leq \frac{\sqrt{2n}}{\sqrt{\pi} d} \cdot e^{-\frac{d^2}{2n}}.$$

Proof: Use, for $x > 0$

$$\frac{e^{-x^2}}{\sqrt{\pi} x}\left(1 - \frac{1}{2x^2}\right) \leq \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \leq \frac{e^{-x^2}}{\sqrt{\pi} x}.$$

□

It will turn out that a specific form of d is of interest. The following explicit bounds will be useful.

Lemma 3.5: For

$$\Delta = \sqrt{2n(\ln n + \delta(n))} < \frac{n}{4}, \quad \text{and } k = 1, 2, \ldots, n-1$$

$$P_G\left(\frac{\Delta}{\sqrt{n-k}}\right) \leq \frac{1}{2n} \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} \cdot e^{-\frac{\ln n + \delta(n)}{n} k}$$

$$P_G\left(\frac{\Delta}{\sqrt{n-k}}\right) \geq \frac{1}{2n} \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} \cdot e^{-\frac{\ln n + \delta(n)}{n-k} k} \times (1 - kn^{-1} - (2(\ln n + \delta(n)))^{-1}).$$

Proof: See Appendix. □

IV. An Upper Bound on $M(S_n)$ for Almost All S_n

For $d > 0$, the number of sequences S_n such that $M_k(S_n) \geq d$, is given by

$$G(n, k, d) \equiv \sum_{r=d}^{n-k} g(n, k, r)$$

$$= 2^k \sum_{r=\frac{d}{2}}^{\frac{n-k}{2}} \binom{n-k}{(n-k) \cdot \left(\frac{1}{2} - \frac{r}{n-k}\right)}$$

$$= 2^k \cdot S(n-k, d).$$

The total number of binary sequences of length n is 2^n, therefore

$$\Pr(M_k(S_n) \geq d) = 2^{-n} \cdot G(n, k, d) = 2^{-(n-k)} \cdot S(n-k, d).$$

Similarly, the number of sequences S_n such that $M_k(S_n) \leq -d$, is given by

$$\hat{G}(n, k, d) \equiv \sum_{r=-d}^{-(n-k)} g(n, k, r) = \sum_{r=d}^{n-k} g(n, k, -r) = \sum_{r=d}^{n-k} g(n, k, r) = G(n, k, d).$$

Therefore

$$\Pr(M_k(S_n) \leq -d) = 2^{-n} \cdot \hat{G}(n, k, d) = 2^{-n} \cdot G(n, k, d) = \Pr(M_k(S_n) \geq d).$$

We have the following lemma.

Lemma 4.1: Let $n - k \geq 164$. For

$$\sqrt{(n-k) \ln \ln(n-k)} < d < (n-k)/2$$

the following holds:

$$\Pr(M_k(S_n) \geq d) \leq P_G\left(\frac{d}{\sqrt{n-k}}\right) \cdot \left(1 + \varsigma_3^{(n-k,d)}\right)$$

$$\varsigma_3^{(n-k,d)} = \frac{7d}{n-k}$$

and for $\sqrt{(n-k) \ln \ln(n-k)} < d < (n-k)^{3/4}$ the following holds:

$$\Pr(M_k(S_n) \geq d) \geq P_G\left(\frac{d}{\sqrt{n-k}}\right) \left(1 - \varsigma_4^{(n-k,d)}\right) - e^{-\sqrt{\frac{n-k}{32}}}$$

$$\varsigma_4^{(n-k,d)} = \frac{5d^4}{(n-k)^3} + \frac{1}{2(n-k)}.$$

Proof: Use Lemma 3.3 with $n - k$ in place of n. □

Combining Lemmas 4.1 and 3.5, and using $e^x > 1 + x$ for $x > 0$, we have

$$2^{-n} \sum_{k=1}^{n-1} G(n, k, d)$$

$$\leq \frac{1}{2n} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} \sum_{k=1}^{n-1} e^{-\frac{\ln n + \delta(n)}{n} k} \cdot \left(1 + \frac{7d}{n-k}\right)$$

$$\leq \frac{1}{2n} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} \times \frac{n}{\ln n + \delta(n)} \cdot (1 + o(1)).$$

Consequently, we have the following corollary.

Corollary 4.2: Under the conditions of Lemma 4.1, for $\Delta = \sqrt{2n(\ln n + \delta(n))}$

$$\Pr(M(S_n) \geq \Delta) \leq \frac{e^{-\delta(n)}}{\sqrt{\pi}(\ln n + \delta(n))^{\frac{3}{2}}} \cdot (1 + o(1)). \quad (10)$$

Proof: Straightforward

$$\Pr\left(\max_{k=1,2,\dots,n-1} |M_k| \geq \Delta\right) \leq \sum_{k=1}^{n-1} \Pr(|M_k| \geq \Delta).$$

$$\sum_{k=1}^{n-1} \Pr(|M_k| \geq \Delta) = 2\sum_{k=1}^{n-1} \Pr(M_k \geq \Delta),$$

$$\sum_{k=1}^{n-1} \Pr(M_k \geq \Delta) = 2^{-n} \sum_{k=1}^{n-1} G(n,k,\Delta)$$

$$\leq \frac{1}{2} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi}(\ln n + \delta(n))^{\frac{3}{2}}} \cdot (1 + o(1)). \qquad \square$$

For example, taking $\delta(n) = -1.5 \ln\ln n + \beta \ln\ln n$, we obtain the following.

Corollary 4.3:

$$\Pr(M(S_n) \geq \sqrt{2n(\ln n - 1.5\ln\ln n + \beta \ln\ln n)}) \leq O\left(\frac{1}{\ln^{\beta} n}\right). \quad (11)$$

For the sake of comparison, let us derive a lower bound for $2^{-n} \cdot \sum_k G(n,k,\Delta)$. For notational convenience, in what follows $f \ll g$ stands for $f = o(g)$.

Lemma 4.4: For

$$\Delta = \sqrt{2n(\ln n + \delta(n))}, \ \sqrt{n} \ll \Delta \ll n^{3/4}$$

$$2^{-n} \sum_{k=1}^{n-1} G(n,k,\Delta) \geq \frac{1}{2e} \frac{e^{-\delta(n)}}{\sqrt{\pi}(\ln n + \delta(n))^{\frac{3}{2}}} (1 - o(1)).$$

Proof: Use Lemma 3.3 and note that for

$$\sqrt{n} \ll \Delta \ll n^{3/4} \text{ and } k \leq 2(n/\Delta)^2$$

$$\varsigma_4^{(n-k)} + \frac{e^{-\sqrt{\frac{n-k}{32}}}}{P_G\left(\frac{\Delta}{\sqrt{n-k}}\right)} = o(1).$$

Hence

$$2^{-n} \sum_{k=1}^{n-1} G(n,k,\Delta)$$

$$\geq \frac{(1-o(1))}{2n} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} \sum_{k=1}^{\frac{n}{\ln n + \delta(n)}} \frac{n-k}{n} \cdot e^{-\frac{\ln n + \delta(n)}{(n-k)/k}}$$

$$\geq \frac{(1-o(1))}{2n} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} e^{-\frac{\ln n + \delta(n)}{\ln n + \delta(n) - 1}} \sum_{k=1}^{\frac{n}{\ln n + \delta(n)}} \frac{n-k}{n}$$

$$= \frac{1}{2e} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi}(\ln n + \delta(n))^{\frac{3}{2}}} \cdot (1 - o(1)). \qquad \square$$

We see that the lower and upper bounds for $2^{-n}\sum_{k=1}^{n-1} G(n,k,\Delta)$ differ only by a multiplicative constant.

V. A Lower Bound on $M(S_n)$ for Almost All S_n

Notice that $M_{\frac{n}{2}}, M_{\frac{n}{2}+1}, \dots, M_{n-1}$ are linear in $s_1, s_2 \dots, s_{\frac{n}{2}}$ and in $s_{\frac{n}{2}+1}, \dots, s_n$, and therefore can be written collectively as a linear system

$$\begin{pmatrix} M_{\frac{n}{2}} \\ M_{\frac{n}{2}+1} \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{pmatrix} = \begin{pmatrix} s_1 & s_2 & s_3 & \cdots & s_{\frac{n}{2}-1} & s_{\frac{n}{2}} \\ 0 & s_1 & s_2 & \cdots & s_{\frac{n}{2}-2} & s_{\frac{n}{2}-1} \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & s_1 & s_2 \\ 0 & 0 & 0 & 0 & \cdots & s_1 \end{pmatrix} \begin{pmatrix} s_{\frac{n}{2}+1} \\ s_{\frac{n}{2}+2} \\ \vdots \\ s_{n-1} \\ s_n \end{pmatrix}.$$

This linearity allows us to prove independence of $M_{\frac{n}{2}+i-1}$ and $M_{\frac{n}{2}+j-1}$ for $1 \leq i < j \leq \frac{n}{2}$, in Lemma 5.1. Using the independence and the inclusion–exclusion principle, we provide a lower bound for the upper tail on the probability of the number of sequences S_n with $M(S_n) \geq \sqrt{n\ln n + \delta(n)}$, Theorem 5.3.

Next, we use Azuma's bound to show that since $M(S_n)$ satisfies a Lipschitz condition, the distribution of $M(S_n)$ is concentrated, though we cannot indicate where its expectation lies. However, noticing that the expectation cannot be too small, since otherwise its upper tail—an upper bound on the probability that $M(S_n) \geq \sqrt{n\ln n + \delta(n)}$—will contradict the earlier derived lower bound on the probability of the same event, we conclude that the expectation cannot be less than approximately $\sqrt{n \ln n}$.

Lemma 5.1: For any $1 \leq i < j \leq \frac{n}{2}$ and $d > 0$

$$\Pr\left(\left|M_{\frac{n}{2}+i-1}\right| > d \bigcap \left|M_{\frac{n}{2}+j-1}\right| > d\right) = \Pr\left(\left|M_{\frac{n}{2}+i-1}\right| > d\right) \cdot \Pr\left(\left|M_{\frac{n}{2}+j-1}\right| > d\right).$$

Remark 5.2: As suggested by a referee, the key idea of the statement is that $2n - (k + \ell)$ products are independent random variables. A consequence is that $M_{n/2+i-1}$ and $M_{n/2+j-1}$ are independent, which in turn implies the above lemma.

Proof: For $1 \leq i < j \leq \frac{n}{2}$, we consider two forms

$$M_{\frac{n}{2}+i-1} = s_1 s_{\frac{n}{2}+i} + s_2 s_{\frac{n}{2}+i+1} + \cdots + s_{\frac{n}{2}-i+1} s_n$$

and

$$M_{\frac{n}{2}+j-1} = s_1 s_{\frac{n}{2}+j} + s_2 s_{\frac{n}{2}+j+1} + \cdots + s_{\frac{n}{2}-j+1} s_n.$$

Notice that the number of product terms in the second form, $\frac{n}{2} - j + 1$, is less than the number of product terms in the first one, $\frac{n}{2} - i + 1$. Let us form a vector of length $n - 2j + 2$, having the first half consisting of the first $\frac{n}{2} - j + 1$ product terms from $M_{\frac{n}{2}+i-1}$ and the second half containing the product terms from $M_{\frac{n}{2}+j-1}$, namely

$$\boldsymbol{x} = (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \left(s_1 s_{\frac{n}{2}+i}, s_2 s_{\frac{n}{2}+i+1}, \dots, s_{\frac{n}{2}-j+1} s_{n-j+i}, s_1 s_{\frac{n}{2}+j}, s_2 s_{\frac{n}{2}+j+1}, \dots, s_{\frac{n}{2}-j+1} s_n\right).$$

Let us show that when

$$\boldsymbol{y} = \left(s_1, s_2, \ldots, s_{\frac{n}{2}-j+1}\right)$$

assumes all possible values from $\boldsymbol{F}_2^{\frac{n}{2}-j+1}$, and

$$\boldsymbol{z} = \left(s_{\frac{n}{2}+i}, s_{\frac{n}{2}+i+1}, \ldots, s_n\right)$$

assumes all possible values from $\boldsymbol{F}_2^{\frac{n}{2}-i+1}$, then $\boldsymbol{x}$ assumes all possible values from $\boldsymbol{F}_2^{n-2j+2}$ equal number of times, 2^{j-i}.

Notice that $\boldsymbol{w} = (\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}) \in (\boldsymbol{F}_2^{n/2-j+1})^2$, assumes all possible values from $\boldsymbol{F}_2^{n-2j+2}$ exactly once if and only if the vector $(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(1)} * \boldsymbol{w}^{(2)})$, where $*$ stands for the coordinate-wise multiplication of vectors, also assumes all possible values of $\boldsymbol{F}_2^{n-2j+2}$ exactly once. Indeed, for a fixed $\boldsymbol{w}^{(1)}$, $\boldsymbol{w}^{(2)}$ assumes all possible values exactly once. The same is clearly true for $\boldsymbol{w}^{(1)} * \boldsymbol{w}^{(2)}$ for a fixed $\boldsymbol{w}^{(1)}$ and $\boldsymbol{w}^{(2)}$ running over all possibilities in $\boldsymbol{F}_2^{\frac{n}{2}-j+1}$. In the opposite direction, the same is correct since the transform is involution.

Using $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ in place of $\boldsymbol{w}^{(1)}$ and $\boldsymbol{w}^{(2)}$ in the preceding expression, and noticing that the variables $s_{\frac{n}{2}-j+2}, \ldots, s_{\frac{n}{2}-i+1}$ do not appear in either $\boldsymbol{x}^{(1)}$ or $\boldsymbol{x}^{(2)}$ we conclude that

$$\left(\boldsymbol{x}^{(1)}, s_{\frac{n}{2}-j+2}s_{n-j+i+1}, \ldots, s_{\frac{n}{2}-i+1}s_n \,\middle|\, \boldsymbol{x}^{(2)}\right)$$

assumes each of its $2^{n-i-j+2}$ possible values exactly 2^{j-i} times. Consequently, $M_{\frac{n}{2}+i-1}$ and $M_{\frac{n}{2}+j-1}$ are independent for all $1 \le i < j \le \frac{n}{2}$. □

Theorem 5.3: Let

$$\Delta = \sqrt{n(\ln n + \delta(n))}, \quad \sqrt{n \ln\ln n} < \Delta \ll n^{3/4}. \tag{12}$$

Then

$$\Pr(M(S_n) \ge \Delta) \ge \frac{2e^{-\delta(n)}}{\ln n \cdot \sqrt{\pi(\ln n + \delta(n))}}(1 - o(1)).$$

Proof: For any subset of sequences S_n from $\mathcal{A}_n$, and any subset $\mathcal{K}$ of indices k belonging to $\{1, 2, \ldots, n-1\}$

$$\Pr\left(\max_{k=1,2,\ldots n-1} |M_k(S_n)| \ge \Delta\right) \ge \Pr\left(\max_{k \in \mathcal{K}} |M_k(S_n)| \ge \Delta\right).$$

For $m = o(n)$ and $\mathcal{K} = \{\frac{n}{2}, \frac{n}{2}+1, \ldots, \frac{n}{2}+m\}$, we have

$$\Pr\left(\max_{k=1,2,\ldots n-1} |M_k(S_n)| \ge \Delta\right) \tag{13}$$

$$\ge \Pr\left(\max_{k=\frac{n}{2},\frac{n}{2}+1,\ldots,\frac{n}{2}+m} |M_k(S_n)| \ge \Delta\right). \tag{14}$$

Moreover, by the inclusion–exclusion we have

$$\Pr\left(\max_{k=\frac{n}{2},\frac{n}{2}+1,\ldots,\frac{n}{2}+m} |M_k(S_n)| \ge \Delta\right) \ge \sum_{k=\frac{n}{2}}^{\frac{n}{2}+m} \Pr(|M_k(S_n)| \ge \Delta) - \sum_{i,j=\frac{n}{2}, i \ne j}^{\frac{n}{2}+m} \Pr\left(|M_i(S_n)| \ge \Delta \bigcap |M_j(S_n)| \ge \Delta\right).$$

The proof of the theorem is to be continued after we demonstrate the following auxiliary result.

Lemma 5.4: For any $i < m, m = o(n)$, and $\delta(n) > -\ln n + \ln\ln n$

$$\Pr\left(\left|M_{\frac{n}{2}+i-1}(S_n)\right| \ge \sqrt{n(\ln n + \delta(n))}\right) \ge \frac{2}{n} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} \cdot (1 - o(1)).$$

Proof: By construction

$$M_{\frac{n}{2}+i-1}(S_n) = s_1 s_{\frac{n}{2}+i} + s_2 s_{\frac{n}{2}+i+1} + \cdots + s_{\frac{n}{2}-i+1} s_n$$

therefore, by Theorem 3.1

$$\Pr\left(M_{\frac{n}{2}+i-1}(S_n) = d\right) = 2^{-(\frac{n}{2}-i+1)} \cdot \binom{\frac{n}{2}-i+1}{\frac{\frac{n}{2}-i+1}{2} - \frac{d}{2}}.$$

Applying Lemmas 3.3, 3.4 with $n \to \frac{n}{2} - i + 1$ and

$$d = \Delta = \sqrt{n(\ln n + \delta(n))}$$

we have

$$\begin{aligned}
\Pr&\left(M_{\frac{n}{2}+i-1}(S_n) \ge \sqrt{n(\ln n + \delta(n))}\right) \\
&= 2^{\frac{n}{2}-i+1} \cdot S\left(\frac{n}{2} - i + 1, \sqrt{n(\ln n + \delta(n))}\right) \\
&\ge \frac{\sqrt{1 - 2(i-1)/n}}{\sqrt{\pi(\ln n + \delta(n))}} \cdot e^{-\frac{\ln n + \delta(n)}{1-2(i-1)/n}} \\
&\quad \times \left(1 - \frac{1 - 2(i-1)/n}{2(\ln n + \delta(n))}\right)(1 - o(1)).
\end{aligned}$$

For $i = 1, 2, \ldots, m, m = o(n)$, we then have

$$\Pr\left(M_{\frac{n}{2}+i-1}(S_n) \ge \sqrt{n(\ln n + \delta(n))}\right) = \frac{1}{n} \cdot \frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n + \delta(n))}} \cdot (1 - o(1)). \tag{15}$$

□

Let us now return to the proof of the theorem. Note that by symmetry

$$\begin{aligned}
\Pr&\left(|M_i(S_n)| \ge \Delta \bigcap |M_j(S_n)| \ge \Delta\right) \\
&= 4 \cdot \Pr\left(M_i(S_n) \ge \Delta \bigcap M_j(S_n) \ge \Delta\right).
\end{aligned}$$

$$\begin{aligned}
\Pr&\left(\max_{k=1,2,\ldots n-1} |M_k(S_n)| \ge \Delta\right) \\
&\ge \frac{2m}{n} \cdot \frac{e^{-\delta(n)} \cdot (1 - o(1))}{\sqrt{\pi(\ln n + \delta(n))}} - \frac{2m^2}{n^2} \cdot \frac{(1 + o(1)) \cdot e^{-2\delta(n)}}{\pi(\ln n + \delta(n))}.
\end{aligned}$$

For $m = n/\ln n$

$$\begin{aligned}
\Pr&\left(\max_{k=1,2,\ldots n-1} |M_k(S_n)| \ge \Delta\right) \\
&\ge \frac{2e^{-\delta(n)} \cdot (1 - o(1))}{\ln n \cdot \sqrt{\pi(\ln n + \delta(n))}} - \frac{2e^{-2\delta(n)}(1 + o(1))}{(\ln n)^2 \cdot \pi(\ln n + \delta(n))} \\
&= \frac{2e^{-\delta(n)}}{\ln n \cdot \sqrt{\pi(\ln n + \delta(n))}} \cdot (1 - o(1)).
\end{aligned}$$

This accomplishes the proof of Theorem 5.3. □

Using the lower bound from Theorem 5.3 and the Azuma inequality, we will provide a lower bound for the mean of $M(S_n)$.

Lemma 5.5 (Azuma, cf. e.g., [15], [16]): Let $z_1, z_2, \ldots, z_n$ be independent random variables, with z_j taking values in a set Λ_j. Assume that a function $f : \Lambda_1 \times \Lambda_2 \times \cdots \Lambda_n \to R$ satisfies, for some constants $b_j, j = 1, 2, 3, \ldots, n$, the following Lipschitz condition: if two vectors $\boldsymbol{z}, \boldsymbol{z}'$ differ only in the jth coordinate, then $|f(\boldsymbol{z}) - f(\boldsymbol{z}')| \le b_j$.

Then, the random variable $X = f(z_1, z_2, \ldots, z_n), \lambda \equiv E(X)$ satisfies, for any $t \ge 0$

$$\Pr(X \ge \lambda + t) \le \exp\left\{-2t^2 \Big/ \sum_1^N b_j^2\right\}$$
$$\Pr(X \le \lambda - t) \le \exp\left\{-2t^2 \Big/ \sum_1^N b_j^2\right\}. \qquad \Diamond$$

For

$$X \equiv \max_{k=1,\ldots,n-1} |M_k|$$

we note that $b_j = 2, j = 1, 2, \ldots, n-1$, and therefore, for $t = \sqrt{n \ln n} - \lambda$

$$\Pr(\max_{k=1,\ldots,n-1} |M_k| \ge \lambda + t) = \Pr(\max_{k=1,\ldots,n-1} |M_k| \ge \sqrt{n \ln n}) \le \exp\left(-\frac{(\sqrt{n \ln n} - \lambda)^2}{2n}\right).$$

On the other hand

$$\Pr(\max_{k=1,\ldots,n-1} |M_k| \ge \sqrt{n \ln n}) \ge \frac{2}{\sqrt{\pi (\ln n)^3}}(1 - o(1)).$$

For consistency, we require

$$\exp\left(-\frac{(\sqrt{n \ln n} - \lambda)^2}{2n}\right) \ge \frac{2}{\sqrt{\pi (\ln n)^3}}(1 - o(1))$$

and therefore

$$(\sqrt{n \ln n} - \lambda)^2 \le 3n \ln \ln n \cdot (1 - o(1))$$

and consequently

$$\lambda \ge \sqrt{n \ln n} - \sqrt{3n \ln \ln n} \cdot (1 - o(1))$$
$$\ge \sqrt{n \ln n}\left(1 - \sqrt{3\frac{\ln \ln n}{\ln n}}\right)(1 - o(1)).$$

Written differently

$$\lambda \ge \sqrt{n(\ln n - 2\sqrt{3 \ln n \ln \ln n} + 3 \ln \ln n)}(1 - o(1)).$$

We have thus shown that

$$E\left\{\max_{k=1,\ldots,n-1} |M_k|\right\} \ge \sqrt{n \ln n} \cdot (1 - o(1)).$$

From here, straightforward application of the Azuma inequality gives us the following.

Corollary 5.6: For almost all $S_n, M(S_n) = \Theta(\sqrt{n \ln n})$

$$\Pr\left(\max_{k=1,\ldots,n-1} |M_k| \le \sqrt{n \ln n} - \sqrt{2\beta n \ln \ln n}\right)$$
$$\le \exp\left(-\frac{2\beta n \ln \ln n}{2n}\right)$$
$$= (\ln n)^{-\beta}. \qquad \Diamond$$

VI. The $M(S_n)$ Concentration, Modulo a Conjecture

For notational convenience, let us introduce the following definitions:

$$A_k(d) \equiv \{|M_k| \ge d\}, \quad k = 1, \ldots, n-1$$
$$A_{k_1 k_2}(d) \equiv A_{k_1}(d) \bigcap A_{k_2}(d), \quad k_1 \ne k_2$$
$$E_{k_1 k_2}(d_1, d_2) \equiv \{|M_{k_1}| = d_1\} \bigcap \{|M_{k_2}| = d_2\}, \quad k_1 \ne k_2.$$

The number of sequences in a set $\mathcal{S}$ is denoted by $\#\mathcal{S}$.

The following idea was suggested to the authors by Alex Koreiko.

Conjecture 6.1 (Koreiko [10]): For $d_1, d_2 = \theta(\sqrt{n \ln n})$

$$\#E_{12}(d_1, d_2) \ge \#E_{k_1 k_2}(d_1, d_2),$$
$$k_1, k_2 = 1, 2, \ldots, n-1, k_1 \ne k_2. \quad \Diamond$$

Though we were not able to prove that the previous is correct, we can show that modulo Conjecture 6.1, the following holds.

Lemma 6.2: Let

$$d = \sqrt{2n(\ln n + \delta(n)))}.$$

For $k_1, k_2 = 1, 2, \ldots, n-1, k_1 \ne k_2$, we have (modulo Conjecture 6.1)

$$\Pr(A_{k_1 k_2}(d)) \le \frac{2(\ln n + \delta(n))}{n^2} \cdot e^{-2\delta(n)} \cdot (1 + o(1)).$$

Proof: As put forward by Moon and Moser in [19]

$$\#E_{12}(d_1, d_2) = 2\left\{\binom{\frac{(n-1)+d_1}{2} - 1}{\frac{1}{2}\frac{(n-2)-d_2}{2}}\binom{\frac{(n-1)-d_1}{2} - 1}{\frac{1}{2}\left(\frac{(n-2)-d_2}{2} - 1\right)} + \binom{\frac{(n-1)+d_1}{2} - 1}{\frac{1}{2}\left(\frac{(n-2)-d_2}{2} - 1\right)}\binom{\frac{(n-1)-d_1}{2} - 1}{\frac{1}{2}\frac{(n-2)-d_2}{2}}\right\}.$$

For

$$d_i = \sqrt{2n(\ln n + \delta_i(n))} = \theta(\sqrt{2n \ln n}), \quad i = 1, 2$$

we have the inequality at the top of the following page. For $\Delta = \sqrt{2n(\ln n + \delta(n))}$

$$\#A_{k_1 k_2}(\Delta) = \sum_{d_1=\Delta}^{n-k_1} \sum_{d_2=\Delta}^{n-k_2} \#E_{k_1 k_2}(d_1, d_2)$$
$$\le \sum_{d_1=\Delta}^{2\Delta} \sum_{d_2=\Delta}^{2\Delta} \#E_{12}(d_1, d_2) + \Delta_G$$
$$\le \Delta^2 \cdot \#E_{12}(\Delta, \Delta) + \Xi_G$$
$$\le 2^n \cdot \frac{2(\ln n + \delta(n))}{n^2} \cdot e^{-2\delta(n)} \cdot (1 + o(1))$$

$$\begin{aligned}\#E_{12}(d_1,d_2) &= 2\binom{\frac{(n-3)+d_1}{2}}{\frac{(n-3)+d_1}{2}\left(\frac{1}{2}-\frac{d_1+d_2-1}{2((n-3)+d_1)}\right)}\binom{\frac{(n-3)-d_1}{2}}{\frac{(n-3)-d_1}{2}\left(\frac{1}{2}+\frac{d_1-d_2-1}{2((n-3)-d_1)}\right)}\\ &\quad+2\binom{\frac{(n-3)+d_1}{2}}{\frac{(n-3)+d_1}{2}\left(\frac{1}{2}-\frac{d_1+d_2+1}{2((n-3)+d_1)}\right)}\binom{\frac{(n-3)-d_1}{2}}{\frac{(n-3)-d_1}{2}\left(\frac{1}{2}+\frac{d_1-d_2+1}{2((n-3)-d_1)}\right)}\\ &\le 2^n\frac{1+o(1)}{\pi\sqrt{(n-3)^2-d_1^2}}\exp\left\{-\frac{0.5(n-3)((d_1-1)^2+d_2^2)-(d_1-1)d_1d_2}{(n-3)^2-d_1^2}\right\}\\ &\quad+2^n\frac{1+o(1)}{\pi\sqrt{(n-3)^2-d_1^2}}\exp\left\{-\frac{0.5(n-3)((d_1+1)^2+d_2^2)-(d_1+1)d_1d_2}{(n-3)^2-d_1^2}\right\}\\ &\le 2^n\cdot\frac{2e^{-(\delta_1(n)+\delta_2(n))}}{n^3}(1+o(1)).\end{aligned}$$

where $\Xi_G \equiv n\cdot(G(n,k_1,2\Delta)+G(n,k_2,2\Delta)) \ll \#A_{k_1k_2}(\Delta)$. □

Lemma 6.2 along with Lemma 4.4 enable us to derive a tight lower bound for

$$\Pr\left(\max_{k=1,2,\ldots,n-1}|M_k|\ge\sqrt{2n(\ln n+\delta(n))}\right).$$

Corollary 6.3: For $3\ln\ln n\le\delta(n)\le\frac{n}{64}-\ln n$

$$\Pr\left(\max_{k=1,2,\ldots,n-1}|M_k|\ge\sqrt{2n(\ln n+\delta(n))}\right)\ge\frac{e^{-\delta(n)}}{e\sqrt{\pi}(\ln n+\delta(n))^{\frac{3}{2}}}(1-o(1)).$$

Proof: Let $d=\sqrt{2n(\ln n+\delta(n))}$

$$\begin{aligned}&\Pr(\max_{k=1,2,\ldots,n-1}|M_k|>d)\\ &=2\cdot\Pr(\max_{k=1,2,\ldots,n-1}M_k>d)\\ &\ge 2\sum_{k=1}^{n-1}\Pr(M_k>d)-4\sum_{k_1=1}^{n-1}\sum_{k_2=1}^{n-1}\Pr(M_{k_1}>d,M_{k_2}>d)\\ &\ge\left(\frac{e^{-\delta(n)}}{e\sqrt{\pi}(\ln n+\delta(n))^{\frac{3}{2}}}-8(\ln n+\delta(n))e^{-2\delta(n)}\right)(1-o(1)).\end{aligned}$$

For $\delta(n)>\frac{5}{2}\ln\ln n(1+o(1))$, the union bound dominates and we obtain the claim. □

Now we may repeat the steps in the end of Section V, but this time using Corollary 6.3, giving

$$\Pr(\max_{k=1,\ldots,n-1}|M_k|\ge\sqrt{2n(\ln n+3\ln\ln n)})\ge\frac{2}{e\sqrt{\pi(\ln n)^9}}(1-o(1)).$$

Together with the Azuma inequality, it provides us with a tighter lower bound for the mean of $M(S_n)$. Indeed, the consistency requirement yields

$$\exp\left\{-\frac{(t-\lambda)^2}{2n}\right\}\ge\frac{1}{e\sqrt{\pi}(\ln n)^{\frac{9}{2}}}(1-o(1)).$$

Therefore

$$(t-\lambda)^2\le\frac{9}{2}n\ln\ln n(1-o(1))$$

and consequently

$$\begin{aligned}\lambda &\ge\sqrt{2n(\ln n+3\ln\ln n)}-3\sqrt{\frac{n\ln\ln n}{2}}(1-o(1))\\ &\ge\sqrt{2n\ln n}\left(1-\frac{3}{2}\sqrt{\frac{\ln\ln n}{\ln n}}\right)(1-o(1)).\end{aligned}$$

Written differently, it is

$$\lambda\ge\sqrt{2n(\ln n-3\sqrt{\ln n\ln\ln n}+2.25\ln\ln n)}(1-o(1)).$$

We have thus shown that

$$E\left\{\max_{k=1,\ldots,n-1}|M_k|\right\}\ge\sqrt{2n\ln n}\cdot(1-o(1)).$$

From here, the straightforward application of the Azuma inequality gives the following result.

Corollary 6.4: For almost all S_n, $M(S_n)\approx\sqrt{2n\ln n}$. ◊

Appendix I

A. Proof of Theorem 3.1

The original proof appears in [19], here it is sketched for completeness. If

$$M_k(S_n)=\sum s_is_{i+k}=r$$

there must be an excess of $r/2$ values of i with $s_is_{i+k}=+1$. Hence, the set $\{1,2,\ldots,n-k\}$ can be partitioned into two subsets A and B, with $(n-k+r)/2$ and $(n-k-r)/2$, respectively, such that

$$s_{i+k}=s_i,\quad \text{if } i\in A \tag{17}$$

and

$$s_{i+k}=-s_i,\quad \text{if } i\in B. \tag{18}$$

There are $\binom{n-k}{(n-k)\cdot(\frac{1}{2}+\frac{r}{2(n-k)})}$ choices for the subsets A and B and there are 2^k choices for the first k elements of

S_n. Once these choices are made, the remaining elements $s_m = s_{(m-k)+k}$ are determined recursively by (17) or (18). $\Diamond$

B. Proof of Lemma 3.2

Let us first show the upper bound. For $0 < \epsilon_1 < 1/2$ and any $n > 0$, we have

$$\frac{n!}{\left(n\cdot\left(\frac{1}{2}-\epsilon_1\right)\right)!\left(n\cdot\left(\frac{1}{2}+\epsilon_1\right)\right)!} \leq \frac{1}{\sqrt{2\pi n\left(\frac{1}{4}-\epsilon_1^2\right)}}\frac{1}{\left(\frac{1}{2}+\epsilon_1\right)^{\left(\frac{1}{2}+\epsilon_1\right)n}\left(\frac{1}{2}-\epsilon_1\right)^{\left(\frac{1}{2}-\epsilon_1\right)n}}$$

where we have used (cf. e.g., [14])

$$\sqrt{2\pi}\cdot n^{n+1/2}\cdot e^{-n+\frac{1}{12n}-\frac{1}{360n^3}} < n! < \sqrt{2\pi}\cdot n^{n+1/2}e^{-n+\frac{1}{12n}}. \tag{19}$$

Therefore, for $0 < \epsilon_1 < \frac{1}{2}$ and any $n > 0$

$$\binom{n}{n\cdot\left(\frac{1}{2}-\epsilon_1\right)} \leq \frac{1}{\sqrt{2\pi n\left(\frac{1}{4}-\epsilon_1^2\right)}}\cdot e^{nH_e\left(\frac{1}{2}-\epsilon_1\right)} \tag{20}$$

where $H_e(x) \equiv -x\ln x - (1-x)\ln(1-x)$ stands for the natural entropy function.

Using also

$$H_e\left(\frac{1}{2}-\epsilon_1\right) \leq \ln 2 - 2\epsilon_1^2, \quad \text{for } 0 < \epsilon_1 < \frac{1}{2} \tag{21}$$

and

$$\frac{1}{\sqrt{1-x}} \leq 1 + \frac{3}{4}x, \qquad \text{for } 0 \leq x \leq \frac{3}{8}$$

we have ($\epsilon_1^2 \leq 3/32$)

$$\binom{n}{n\cdot\left(\frac{1}{2}-\epsilon_1\right)} \leq \frac{1}{\sqrt{2\pi n\left(\frac{1}{4}-\epsilon_1^2\right)}}\cdot e^{n\ln 2-2n\epsilon_1^2} \tag{22}$$

$$\leq 2^n\cdot\left(1+3\epsilon_1^2\right)\cdot\sqrt{\frac{2}{\pi n}}\cdot e^{-2n\epsilon_1^2}. \tag{23}$$

Now to the lower bound. Using (19), we have for $0 < \epsilon_1 < \frac{1}{2}$ and any $n > 0$

$$\frac{n!}{\left(n\cdot\left(\frac{1}{2}-\epsilon_1\right)\right)!\left(n\cdot\left(\frac{1}{2}+\epsilon_1\right)\right)!} \geq \frac{1}{\sqrt{2\pi n\left(\frac{1}{4}-\epsilon_1^2\right)}}\frac{e^{-\frac{1}{12\left(\frac{1}{2}+\epsilon_1\right)n}-\frac{1}{12\left(\frac{1}{2}-\epsilon_1\right)n}}}{\left(\frac{1}{2}+\epsilon_1\right)^{\left(\frac{1}{2}+\epsilon_1\right)n}\left(\frac{1}{2}-\epsilon_1\right)^{\left(\frac{1}{2}-\epsilon_1\right)n}}.$$

Therefore, for $0 < \epsilon_1 < \frac{1}{2}$ and any $n > 0$

$$\binom{n}{n\cdot\left(\frac{1}{2}-\epsilon_1\right)} \geq \frac{1}{\sqrt{2\pi n\left(\frac{1}{4}-\epsilon_1^2\right)}}e^{n\cdot H_e\left(\frac{1}{2}-\epsilon_1\right)-1/\left(n\left(3-12\epsilon_1^2\right)\right)}.$$

For $0 \leq \epsilon_1 \leq (2n)^{-1/4}$, and $n \geq 164$

$$H_e\left(\frac{1}{2}-\epsilon_1\right) \geq \ln 2 - 2\epsilon_1^2 - \frac{3}{2}\epsilon_1^4 \text{ and } \frac{1}{n\left(3-12\epsilon_1^2\right)} \leq \frac{1}{2n}.$$

Since $e^{-x} \geq 1 - x$ for $x > 0$, we have

$$\binom{n}{n\cdot\left(\frac{1}{2}-\epsilon_1\right)} \geq 2^n\cdot\sqrt{\frac{2}{\pi n}}\cdot e^{-2n\epsilon_1^2} \cdot\left(1-\frac{3}{2}n\epsilon_1^4-\frac{1}{2n}\right). \tag{24}$$

$\Diamond$

C. Proof of Lemma 3.3

Let us first prove (7). Let

$$k_0 = \left\lfloor\sqrt{\frac{3}{32}}n\right\rfloor.$$

Then

$$2^{-n}\cdot S(n,d) = \sum_{k=\frac{d}{2}}^{k_0}\binom{n}{n\cdot\left(\frac{1}{2}-\frac{k}{n}\right)} + \sum_{k=k_0+1}^{\frac{n}{2}}\binom{n}{n\cdot\left(\frac{1}{2}-\frac{k}{n}\right)} = S_1(n,d)+S_2(n,d).$$

Taking into account that the terms of $S_2(n,d)$ are monotonically decreasing, let us bound $S_2(n,d)$ from above by the product of the first (biggest) term and the number of terms in the sum, using Lemma 3.2

$$S_2(n,d) < \frac{41}{64}\sqrt{\frac{2}{\pi}}\left(1-\sqrt{\frac{3}{8}}\right)\cdot\sqrt{n}\cdot e^{-\frac{3n}{16}} < \sqrt{\frac{n}{4\pi}}\cdot e^{-3n/16}. \tag{25}$$

As for $S_1(n,d)$, we apply the upper bound of Lemma 3.2, to get

$$S_1(n,d) \leq \sqrt{\frac{2}{\pi n}}\cdot\sum_{k=\frac{d}{2}}^{\infty}\left(1+\frac{3k^2}{n^2}\right)\cdot e^{-2k^2/n}.$$

Bounding the sum with an integral, noting that for $d > \sqrt{n\ln\ln n}$ the integrands are monotonically decreasing functions of k, and recalling

$$P_G\left(\frac{d}{\sqrt{n}}\right) = \frac{1}{\sqrt{\pi}}\int_{\frac{d}{\sqrt{2n}}}^{\infty} e^{-z^2}dz,$$

we have

$$\sqrt{\frac{2}{\pi n}}\cdot\sum_{k=\frac{d}{2}}^{\infty}e^{-2k^2/n} < \sqrt{\frac{2}{\pi n}}\cdot e^{-\frac{d^2}{2n}} + P_G\left(\frac{d}{\sqrt{n}}\right)$$

$$\sqrt{\frac{2}{\pi n}}\cdot\frac{3}{2n}\cdot\sum_{k=\frac{d}{2}}^{\infty}\frac{2k^2}{n}e^{-2k^2/n} < \frac{3d(2d+1)}{\sqrt{32\pi n^5}}\cdot e^{-\frac{d^2}{2n}} + \frac{3}{4n}\cdot P_G\left(\frac{d}{\sqrt{n}}\right).$$

By assumption, we have

$$3d(2d+1)/\sqrt{32} < \sqrt{2}d^2, \quad (d/n)^2+1 < \sqrt{\pi/2}$$

and

$$\sqrt{\frac{2}{\pi n}} + \frac{3d(2d+1)}{\sqrt{32\pi n^5}} < \frac{1}{\sqrt{n}}.$$

Summing up and using (25)

$$2^{-n} \cdot S(n,d) < P_G\left(\frac{d}{\sqrt{n}}\right) \cdot \left(1 + \frac{3}{4n}\right) + \frac{e^{-\frac{d^2}{2n}}}{\sqrt{n}} + \sqrt{\frac{n}{4\pi}} \cdot e^{-3n/16}. \quad (26)$$

Noting that

$$\sqrt{\frac{n}{2\pi d^2}} \cdot e^{-\frac{d^2}{2n}} \cdot \left(1 - \frac{n}{d^2}\right) \le P_G\left(\frac{d}{\sqrt{n}}\right) \le \sqrt{\frac{n}{2\pi d^2}} \cdot e^{-\frac{d^2}{2n}} \quad (27)$$

under the imposed conditions

$$\frac{e^{-\frac{d^2}{2n}}}{\sqrt{n}} \le \frac{\sqrt{2\pi} d}{n} \cdot \left(1 + \frac{n}{d^2 - n}\right) \cdot P_G\left(\frac{d}{\sqrt{n}}\right) < \frac{6.5d}{n} \cdot P_G\left(\frac{d}{\sqrt{n}}\right). \quad (28)$$

We finally have

$$2^{-n} \cdot S(n,d) \le P_G\left(\frac{d}{\sqrt{n}}\right) \cdot \left(1 + \frac{6.55d}{n} + e^{-\frac{3n}{16} + \frac{d^2}{2n}}\right) \quad (29)$$

$$< P_G\left(\frac{d}{\sqrt{n}}\right) \cdot \left(1 + \frac{6.6d}{n}\right) \quad (30)$$

where we have bounded

$$3/4 < d/20, \quad d^2/(2n) < n/8, \quad \text{and} \quad e^{-\frac{n}{16}} < d/(25n).$$

Now, let us prove (8). Starting from the lower bound in Lemma 3.2

$$2^{-n} \cdot S(n,d) \ge \sum_{k=\frac{d}{2}}^{(n^3/2)^{\frac{1}{4}}} \left(1 - \frac{3k^4}{2n^3} - \frac{1}{2n}\right) \cdot \sqrt{\frac{2}{\pi n}} \cdot e^{-2k^2/n}$$

$$\ge \left(1 - \frac{1}{2n}\right)\left[P_G\left(\frac{d}{\sqrt{n}}\right) - P_G\left(\left(\frac{n}{8}\right)^{\frac{1}{4}}\right)\right] - \sqrt{\frac{2}{\pi n}} \cdot \sum_{k=\frac{d}{2}}^{(n^3/2)^{\frac{1}{4}}} \frac{3k^4}{2n^3} \cdot e^{-2k^2/n}.$$

To complete the proof, let us provide an upper bound for

$$S_3(n,d) = \sqrt{\frac{2}{\pi n}} \cdot \sum_{k=\frac{d}{2}}^{(n^3/2)^{\frac{1}{4}}} \frac{3k^4}{2n^3} \cdot e^{-2k^2/n}.$$

Note that the maximum of $k^4 e^{-2k^2/n}$ is reached for $k^2 = n$. If $d > 2\sqrt{n}$, the summands in $S_3(n,d)$ are monotonically decreasing and the sum can be bounded from above by an integral as follows:

$$\sqrt{\frac{2}{\pi n}} \cdot \sum_{k=\frac{d}{2}}^{(n^3/2)^{\frac{1}{4}}} \frac{3k^4}{2n^3} \cdot e^{-2k^2/n} < \frac{3}{8n}\sqrt{\frac{1}{\pi}} \int_{\frac{d-1}{\sqrt{2n}}}^{\infty} x^4 e^{-x^2} dx.$$

On the other hand, if $\sqrt{\ln\ln n} < d/\sqrt{n} \le 2$ (which can only happen when $\ln\ln n < 2$, i.e., for $n < 1619$), the summands increase for $d/2 < k \le \sqrt{n}$ and decrease thereafter. The biggest summand is $\le 4e^{-2} \le 4e^{-(d-1)^2/2}$.

Clearly, $S_3(n,d)$ can be bounded from above in both cases as

$$\sqrt{\frac{2}{\pi n}} \cdot \sum_{k=\frac{d}{2}}^{(n^3/2)^{\frac{1}{4}}} \frac{3k^4}{2n^3} \cdot e^{-2k^2/n}$$

$$\le \frac{3}{8n}\sqrt{\frac{2}{\pi n}} \cdot 4 \cdot e^{-\frac{(d-1)^2}{2n}} + \frac{3}{8n}\sqrt{\frac{1}{\pi}} \int_{\frac{d-1}{\sqrt{2n}}}^{\infty} x^4 e^{-x^2} dx$$

$$= \frac{3}{32n}\left[\frac{d-1}{\sqrt{2\pi n}}\left(\frac{(d-1)^2}{n} + 3 + \frac{32}{d-1}\right) e^{-\frac{(d-1)^2}{2n}} + 3P_G\left(\frac{d-1}{\sqrt{n}}\right)\right].$$

Analogously to (28), we have

$$\frac{e^{-\frac{(d-1)^2}{2n}}}{\sqrt{n}} < \frac{8.26(d-1)}{n} \cdot P_G\left(\frac{d-1}{\sqrt{n}}\right) < \frac{8.26d}{n} \cdot P_G\left(\frac{d-1}{\sqrt{n}}\right). \quad (31)$$

Lumping the contributions

$$\frac{(d-1)^2}{n} + 3 + \frac{32}{d-1} < 4.55\frac{(d-1)^2}{n} < 4.55\frac{d^2}{n} \quad (32)$$

we get

$$\sqrt{\frac{2}{\pi n}} \cdot \sum_{k=\frac{d}{2}}^{(n^3/2)^{\frac{1}{4}}} \frac{3k^4}{2n^3} \cdot e^{-2k^2/n}$$

$$< \frac{3}{8n}\sqrt{\frac{1}{\pi}}\left[\frac{37.6d^4}{4\sqrt{2}n^2} + \frac{3\sqrt{\pi}}{4}\right] P_G\left(\frac{d-1}{\sqrt{n}}\right)$$

$$< \frac{45d^4}{16\sqrt{\pi}n^3} P_G\left(\frac{d-1}{\sqrt{n}}\right) < \frac{5d^4}{n^3} P_G\left(\frac{d}{\sqrt{n}}\right)$$

where in the last inequality we used

$$P_G\left(\frac{d-1}{\sqrt{n}}\right) \le e^{(2d-1)/2n} \cdot \frac{d-1}{d} \cdot \frac{d^2/n}{d^2/n - 1} \cdot P_G\left(\frac{d}{\sqrt{n}}\right) < 3.04 \cdot P_G\left(\frac{d}{\sqrt{n}}\right).$$

Finally, noting that

$$P_G\left(\left(\frac{n}{8}\right)^{\frac{1}{4}}\right) \le \sqrt{\frac{\sqrt{8}}{\pi}} \cdot n^{-1/4} \cdot e^{-\sqrt{n/32}} < e^{-\sqrt{n/32}}$$

we have

$$2^{-n} \cdot S(n,d) \ge \left(1 - \frac{1}{2n} - \frac{5d^4}{n^3}\right) \cdot P_G\left(\frac{d}{\sqrt{n}}\right) - e^{-\sqrt{n/32}}. \quad (33)$$

◊

D. Proof of Lemma 3.5

For $x > 0$, we have

$$\frac{e^{-x^2}}{\sqrt{\pi}x}\left(1-\frac{1}{2x^2}\right) \le \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt \le \frac{e^{-x^2}}{\sqrt{\pi}x}.$$

Noting that

$$P_G\left(\frac{d}{\sqrt{n-k}}\right) = \frac{1}{\sqrt{\pi}}\int_{\frac{d}{\sqrt{2(n-k)}}}^\infty e^{-t^2}dt$$

we have

$$\begin{aligned} x &= \sqrt{\frac{n(\ln n+\delta(n))}{n-k}} \\ &= \sqrt{(\ln n+\delta(n))+\frac{k(\ln n+\delta(n))}{n-k}} \end{aligned}$$

$$\begin{aligned} P_G\left(\frac{\Delta}{\sqrt{n-k}}\right) &\le \frac{1}{2n}\cdot\frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n+\delta(n))}} \\ &\quad\times\sqrt{1-\frac{k}{n}}\left(ne^{\delta(n)}\right)^{-\frac{k}{n}\frac{1}{1-k/n}} P_G\left(\frac{\Delta}{\sqrt{n-k}}\right) \\ &\ge \frac{1}{2n}\cdot\frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n+\delta(n))}}\sqrt{1-\frac{k}{n}} \\ &\quad\times\left(ne^{\delta(n)}\right)^{-\frac{k}{n}\frac{1}{1-k/n}}\left(1-\frac{1}{2(\ln n+\delta(n))}\right). \end{aligned}$$

Further, using

$$1-x \le \sqrt{1-x} \le 1,\ \ 1+x \le \frac{1}{1-x}, \qquad \text{for } 0 \le x < 1$$

$$\begin{aligned} P_G\left(\frac{\Delta}{\sqrt{n-k}}\right) &\le \frac{1}{2n}\cdot\frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n+\delta(n))}}\left(e^{\ln n+\delta(n)}\right)^{-\frac{k}{n}(1+\frac{k}{n})} \\ &\le \frac{1}{2n}\cdot\frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n+\delta(n))}}\cdot e^{-\frac{\ln n+\delta(n)}{n}k} \\ P_G\left(\frac{\Delta}{\sqrt{n-k}}\right) &\ge \frac{1}{2n}\cdot\frac{e^{-\delta(n)}}{\sqrt{\pi(\ln n+\delta(n))}}\left(1-\frac{k}{n}\right) \\ &\quad\times\left(ne^{\delta(n)}\right)^{-\frac{k}{n}\frac{1}{1-k/n}}\left(1-\frac{1}{2(\ln n+\delta(n))}\right). \end{aligned}$$

$\Diamond$

Acknowledgment

The authors wish to thank the referees for insightful comments.

References

[1] N. Alon, Y. Kohayakawa, C. Mauduit, C. Mauduit, C. G. Moreira, and V. Rödl, "Measures of pseudorandomness for finite sequences: Typical values," *Proc. London Math. Soc.*, vol. 95, pp. 778–812, 2007.

[2] N. Alon and J. Spencer*, The Probabilistic Method*. New York: Wiley, 2000.

[3] J. Bernasconi, "Low autocorrelation binary sequences: statistical mechanics and configuration space analysis," *J. Physique*, vol. 48, p. 559, 1987.

[4] D. Dmitriev and J. Jedwab, "Bounds on the growth rate of the peak sidelobe level of binary sequences," *Adv. Math. Commun.*, vol. 1, pp. 461–475, 2007.

[5] L. Ein-Dor, I. Kanter, and W. Kinzel, "Low autocorrelated multiphase sequences," *Phys. Rev. E*, vol. 65, pp. 020102(R)-1–020102(R)-4, 2002.

[6] S. W. Golomb*, Shift Register Sequences*. San Francisco, CA: Holden-Day, 1967.

[7] S. W. Golomb and G. Gong*, Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[8] G. Halász, "On the result of Salem and Zygmund concerning random polynomials," *Studia Scien. Math. Hung.*, pp. 369–377, 1973.

[9] J. Jedwab and K. Yoshida, "The peak sidelobe level of families of binary sequences," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2247–2254, May 2006.

[10] A. Koreiko, personal communication.

[11] W. Krauth and M. Mezard, "Aging without disorder on long time scales," *Z. Phys.*, vol. B 97, pp. 127–131, 1995.

[12] N. Levanon and E. Mozeson*, Radar Signals*. Hoboken, NJ: IEEE Press/ Wiley-Interscience, 2004.

[13] S. Litsyn and A. Shpunt, "On the distribution of Boolean function nonlinearity," *SIAM J. Discr. Math.*, vol. 23, no. 1, pp. 79–95, 2008.

[14] F. J. MacWilliams and N. J. A. Sloane*, The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: Elsevier, 1977.

[15] C. McDiarmid*, On the Methdod of Bounded Differences*, ser. London Math. Soc. Lect. Notes Ser. 141. Cambridge, U.K.: Cambridge Univ. Press, 1989, pp. 148–188.

[16] C. McDiarmid*, Concentration. Probabilistic Moethods for Algorithmic Discrete Mathematics*. Berlin, Germany: Springer-Verlag, 1998, pp. 195–248.

[17] C. Mauduit and A. Sárközy, "On finite pseudorandom binary sequences I. Measure, of pseudorandomness, the Legendre symbol," *Acta Arith.*, vol. 82, no. 4, pp. 365–377, 1997.

[18] I. D. Mercer, "Autocorrelations of random binary sequences," *Combin., Probab. and Comput.*, vol. 15, pp. 663–671, 2006.

[19] J. W. Moon and L. Moser, "On the correlation function of random binary sequences," *SIAM J. Appl. Math.*, vol. 16, no. 2, pp. 340–343, 1968.

[20] J. Spencer, "Six standard deviations suffice," *Trans. AMS*, vol. 289, no. 2, pp. 679–706, 1985.

[21] R. J. Turyn, , H. B. Mann, Ed., "Sequences with small correlations," in *Error Correcting Codes*. New York: Wiley, 1968, pp. 195–228.

[22] D. Wiggert*, Codes for Error Control and Synchronization*. Norwood, MA: Artech House, 1988, pp. 177–181.

Noga Alon received the Ph.D. degree in mathematics from the Hebrew University of Jerusalem, Israel, in 1983.

He is a Baumritter Professor of Mathematics and Computer Science at Tel-Aviv University, Israel. He had visiting positions in various research institutes including MIT, Cambridge, MA; The Institute of Advanced Study, Princeton, NJ; IBM Almaden Center, San Jose, CA; Bell Labs, Murray Hill, NJ; Bellcore, Morristown, NJ; and Microsoft Research, Redmond, WA. He published more than four hundred research papers, mostly in Combinatorics and in Theoretical Computer Science, and one book.

Dr. Alon has been a member of the Israel National Academy of Sciences since 1997 and of the Academia Europaea since 2008. He received the Erdös prize in 1989, the Feher prize in 1991, the Polya Prize in 2000, the Bruno Memorial Award in 2001, the Landau Prize in 2005, the Goedel Prize in 2005, and the Israel Prize in 2008.

Simon Litsyn (M'94-SM'99) was born in Kharkov, U.S.S.R., in 1957. He received the M.Sc. degree from Perm Polytechnical Institute, Perm, U.S.S.R., in 1979, and the Ph.D. degree from Leningrad Electrotechnical Institute, Leningrad, U.S.S.R., in 1982, both in electrical engineering.

Since 1991, he has been with the Department of Electrical Engineering–Systems, Tel-Aviv University, Israel, where he is a Professor. His research interests include coding and information theory, communications, and applications of discrete mathematics. He authored *Covering Codes* (Elsevier, 1997) and *Peak Power Control in Multicarrier Communications* (Cambridge University Press, 2007).

Dr. Litsyn received the Guastallo Fellowship in 1992. During 2000–2003, he served as an Associate Editor for Coding Theory for the IEEE Transactions on Information Theory. He is an editorial board member of *Advances in Mathematics and Communications (AMC)* and *Applicable Algebra in Engineering Communication and Computing (AAECC)*.

Alexander Shpunt received the B.Sc. degree in physics/computer science in 1999, followed by the M.Sc. degree in electrical engineering in 2005, both with high honors, from Tel-Aviv University, Tel-Aviv, Israel. In 2006, he moved to the Department of Physics at the Massachusetts Institute of Technology, Cambridge, where he currently is a fourth year Graduate Fellow.