Peak-to-average power ratio of good codes for Gaussian channel

Yury Polyanskiy and Yihong Wu

Abstract

Consider a problem of forward error-correction for the additive white Gaussian noise (AWGN) channel. For finite blocklength codes the backoff from the channel capacity is inversely proportional to the square root of the blocklength. In this paper it is shown that codes achieving this tradeoff must necessarily have peak-to-average power ratio (PAPR) proportional to logarithm of the blocklength. This is extended to codes approaching capacity slower, and to PAPR measured at the output of an OFDM modulator. As a by-product the convergence of (Smith's) amplitude-constrained AWGN capacity to Shannon's classical formula is characterized in the regime of large amplitudes. This converse-type result builds upon recent contributions in the study of empirical output distributions of good channel codes.

Index Terms

Shannon theory, channel coding, Gaussian channels, peak-to-average power ratio, converse

I. Introduction

In the additive white Gaussian noise (AWGN) communication channel a (Nyquist-sampled) waveform $x^n = (x_1, \dots, x_n) \in \mathbb{R}^n$ experiences an additive degradation:

$$Y_j = x_j + Z_j\,, Z_j \sim \mathcal{N}(0,1) \tag{1}$$

where $Y^n = (Y_1, \dots, Y_n)$ represent a (Nyquist-sampled) received signal. An (n, M, ϵ, P) error-correcting code is a pair of maps $f : \{1, \dots, M\} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \{1, \dots, M\}$ such that

$$\mathbb{P}[W \neq \hat{W}] \leq \epsilon\,,$$

where $W \in \{1, \dots, M\}$ is a uniformly distributed message, and

$$X^n = f(W) \tag{2}$$

$$\hat{W} = g(Y^n) = g(f(W) + Z^n)\,, \tag{3}$$

Y. Polyanskiy is with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 02139 USA, e-mail: yp@mit.edu. Y. Wu is with Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA, e-mail: yihongwu@illinois.edu.

The work was supported in part by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant CCF-0939370 and NSF CAREER award Grant no. CCF-12-53205.

are the (encoded) channel input and the decoder's output, respectively. The channel input is required to satisfy the power constraint

$$\|X^n\|_2 \triangleq \left(\sum_{j=1}^{n} |X_j|^2\right)^{\frac{1}{2}} \le \sqrt{nP}\,. \tag{4}$$

The non-asymptotic fundamental limit of information transmission over the AWGN channel is given by

$$M^*(n,\epsilon,P) \triangleq \max\{M : \exists(n,M,\epsilon,P)\text{-code}\}\,.$$

It is known that [1][1]

$$\log M^*(n,\epsilon,P) = nC(P) - \sqrt{nV(P)}Q^{-1}(\epsilon) + O(\log n)\,, \tag{6}$$

where the capacity $C(P)$ and the dispersion $V(P)$ are given by

$$C(P) = \frac{1}{2}\log(1+P)\,, \tag{7}$$

$$V(P) = \frac{\log e}{2}\frac{P(P+2)}{(P+1)^2}\,. \tag{8}$$

The peak-to-average power ratio (PAPR) of a codeword x^n is defined as

$$\text{PAPR}(x^n) \triangleq \frac{\|x^n\|_\infty^2}{\frac{1}{n}\|x^n\|_2^2}\,,$$

where $\|x^n\|_\infty = \max_{j=1\ldots n}|x_j|$. This definition of PAPR corresponds to the case when the actual continuous time waveform is produced from x^n via pulse-shaping and heterodyning:

$$s(t) = \sum_{j=1}^{n} x_j g(t-j)\cdot\cos(f_c t)\,,$$

where $g(t)$ is a bounded pulse supported on $[-1/2, 1/2]$ and f_c is a carrier frequency. Alternatively, one could employ an (ideal) DAC followed by a low-pass filter. Such implementation is subject to peak regrowth due to filtering: the maximal amplitude of the signal may be attained in between Nyquist samples, and thus the PAPR observed by the high-power amplifier may be even larger.

In this paper we address the following question: What are the PAPR requirements of codes that attain or come reasonably close to attaining the performance of the best possible codes (6)? In other words, we need to assess the penalty on $\log M^*$ introduced by imposing, in addition to (4), an amplitude constraint:

$$\|X^n\|_\infty \le A_n\,, \tag{9}$$

where A_n is a certain sequence. If A_n is fixed, then even the capacity term in (6) changes according to a well-known result of Smith [2]. Here, thus, we focus on the case of growing A_n.

[1]As usual, all logarithms log and exponents exp are taken to an arbitrary fixed base, which also specifies the information units. Q^{-1} is the inverse of the standard Q-function:

$$Q(x) = \int_x^\infty \frac{e^{-y^2}}{\sqrt{2\pi}}dy\,. \tag{5}$$

Previously, we have shown, [3, Theorem 6] and [4], that very good codes for AWGN automatically satisfy $A_n = O(\sqrt{\log n})$. Namely, for any constant $\gamma > 0$ there exists $\gamma' > 0$ such that *any* code with

$$\log M \geq nC - \sqrt{nV(P)}Q^{-1}(\epsilon) - \gamma \log n \tag{10}$$

has at least $\frac{M}{2}$ codewords with

$$\|x^n\|_\infty \leq \gamma' \sqrt{\log n}$$

In other words, very good codes cannot have PAPR worse than $O(\log n)$. On the other hand, for capacity-achieving input $X_*^n \sim \mathcal{N}(0, P)$, classical results from extremal value theory shows that the peak amplitude behaves with high probability according to $\|X_*^n\|_\infty = \sqrt{2P \log n} + o_{\mathbb{P}}(1)$ [5]. Therefore it is reasonable to expect that good codes must also have peak amplitude scaling as $\sqrt{2 \log n}$. Indeed, in this paper we show that, even under much weaker assumptions on coding performance than (10), the PAPR of at least half of the codewords must be $\Omega(\log n)$.

Interestingly, the $\log n$ behavior of PAPR has been recently observed for various communication systems implementing orthogonal frequency division multiplexing (OFDM) modulation. To describe these results we need to introduce several notions. Given $x^n \in \mathbb{C}^n$ the baseband OFDM (with n subcarriers) signal $s_b(t)$ is given by

$$s_b(t) = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} x_k e^{2\pi i \frac{kt}{n}} ,$$

whereas the transmitted signal is

$$s(t) = \mathrm{Re}\left(e^{2\pi i f_c t} s_b(t)\right) , \qquad 0 \leq t < n \tag{11}$$

where f_c is the carrier frequency. For large f_c, we have that PAPR of $s(t)$ may be approximated as [6, Chapter 5]

$$\text{OFDM-PAPR}(x^n) \triangleq \frac{\max_{t\in[0,n]} |s(t)|^2}{\frac{1}{n}\int_0^n |s(t)|^2 dt} \approx \frac{\max_{t\in[0,n]} |s_b(t)|^2}{\frac{1}{n}\sum_{k=0}^{n-1} |x_k|^2} \triangleq \text{PMEPR}(x^n) , \tag{12}$$

where the quantity on the right is known as the peak-to-mean envelope power (PMEPR).

Note that values of $s_b(\cdot)$ at integer times simply represent the discrete Fourier transform (DFT) of x^n. Thus PMEPR is always lower bounded by

$$\text{PMEPR}(x^n) \geq \frac{\|Fx^n\|_\infty^2}{\frac{1}{n}\|x^n\|_2^2} , \tag{13}$$

where F is the $n \times n$ unitary DFT matrix

$$F_{k,\ell} = \frac{1}{\sqrt{n}} e^{2\pi i \frac{k\ell}{n}} .$$

In view of (13), it is natural to also consider the case where the amplitude constraint (9) is replaced with

$$\|Ux^n\|_\infty \leq A_n , \tag{14}$$

where U is some fixed orthogonal (or unitary) matrix. Note that for large n there exist some ("atypical") $x \in \mathbb{C}^n$ such that the lower bound (13) is very non-tight [6, Chapter 4.1]. Thus, the constraint (14) with $U = F$ is weaker than constraining inputs to those with small OFDM-PAPR(x^n). Nevertheless, it will be shown even with this relaxation A_n is required to be of order $\log n$.

The question of constellations in $\mathbb{C}^n$ with good minimum distance properties and small OFDM-PAPR has been addressed in [7]. In particular, it was shown in [7, Theorems 7 – 8] that the (Euclidean) Gilbert-Varshamov bound is achievable with codes whose OFDM-PAPR is $O(\log n)$ – however, see Remark 2 below. Furthermore, a converse result is established in [7, Theorem 5] which gives a lower bound on the PAPR of an arbitrary code in terms of its rate, blocklength and the minimum distance. When $x^n \sim \mathcal{N}_c(0,P)^n$, the resulting distribution of OFDM-PAPR was analyzed in [8]. For so distributed x^n as well as x^n chosen uniformly on the sphere, OFDM-PAPR tightly concentrates around $\log n$, cf. [6, Chapter 6]. Similarly, if the components of x^n are independently and equiprobably sampled from the M-QAM or M-PSK constellations OFDM-PAPR again sharply peaks around $\log n$, cf. [9]. If x^n is an element of a BPSK modulated BCH code, then again OFDM-PAPR is around $\log n$ for most codewords [6], [9].

Thus, it seems that most good constellations have a large OFDM-PAPR of order $\log n$. Practically, this is a significant detriment for the applications of OFDM. A lot of research effort has been focused on designing practical schemes for *PAPR reduction*. Key methods include amplitude clipping and filtering [10], partial transmit sequence [11], selected mapping [12], tone reservation and injection [13], active constellation extension [14], and others – see comprehensive surveys [15], [16]. In summary, all these techniques take a base code and transform it so as to decrease the PAPR at the output of the OFDM modulator. In all cases, transformation degrades performance of the code (either probability of error, or rate). Therefore, a natural question is whether there exist (yet to be discovered) techniques that reduce PAPR without sacrificing much of the performance.

This paper answers the question in the negative: the $\Theta(\log n)$ PAPR is unavoidable unless a severe penalty in rate is taken.

II. Main results

We start from a simple observation that achieving capacity (without stronger requirements like (10)) is possible with arbitrarily slowly growing PAPR:

Proposition 1: Let $A_n \to \infty$. Then for any $\epsilon \in (0,1)$ there exists a sequence of (n, M_n, ϵ, P) codes satisfying (9) such that

$$\frac{1}{n}\log M_n \to C(P), \quad n \to \infty\,.$$

Proof: Indeed, as is well known, e.g. [17, Chapter 10], selecting $M_n = \exp\{nC(P) + o(n)\}$ codewords with i.i.d. Gaussian entries $X_j \sim \mathcal{N}(0,P)$ results (with high probability) in a codebook that has vanishing probability of error under maximum likelihood decoding. Let us now additionally remove all codewords violating (9). This results in a codebook with a random number $M_n' \le M_n$ of codewords. However, we have

$$\begin{aligned}
\mathbb{E}[M_n'] &= M_n \mathbb{P}[\|X^n\|_\infty \le A_n] && (15)\\
&= M_n \left(1 - 2Q\left(\frac{A_n}{\sqrt{P}}\right)\right)^n && (16)\\
&= M_n \cdot \exp\{o(n)\} = \exp\{nC(P) + o(n)\}\,. && (17)
\end{aligned}$$

The usual random coding argument then shows that there must exist a realization of the codebook that simultaneously has small probability of error and number of codewords no smaller than $\frac{1}{3}\mathbb{E}[M_n']$. ■

Remark 1: Clearly, by applying U^{-1} first and using the invariance of the distribution of noise Z^n to rotations we can also prove that there exist capacity-achieving codes satisfying "post-rotation" amplitude constraint (14). A more delicate question is whether there exist good codes with small PMEPR (which approximates OFDM-PAPR). In that regard, [8] and [6, Chapter 5.3] show that if $X^n \sim \mathcal{CN}(0, PI_n)$ we have

$$\mathbb{P}[\text{PMEPR}(X^n) \le A_n^2] \approx e^{-\sqrt{\frac{\pi}{3}} n A_n e^{-A_n^2}}. \tag{18}$$

Thus, repeating the expurgation argument in (17) we can show that there exists codes with arbitrarily slowly growing OFDM-PAPR and achieving capacity. Furthermore, there exist codes achieving expansion in (6) to within $O(\sqrt{n})$ terms with OFDM-PAPR of order $\log n$.

Remark 2: [2] Not only capacity, but also the Gilbert-Varshamov (GV) bound on the sphere in $\mathbb{R}^n$ can be achieved with arbitrarily slow growing PMEPR, that is, $A_n = \omega(1)$. Note that previously [7, Theorems 7 – 8] only showed the attainability of the GV bound with $A_n = \Theta(\sqrt{\log n})$. Indeed, since the GV bound follows from a greedy procedure, it is sufficient to show that for arbitrary $A_n \to \infty$ we have

$$\mathbb{P}[\text{PMEPR}(X^n) \le A_n^2] = e^{o(n)}, \tag{19}$$

where X^n is uniformly distributed on a unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$. Furthermore, we may take $X^n = Z^n/\|Z^n\|_2$ with $Z^n \sim \mathcal{N}(0, I_n)$. Since $\|Z^n\|_2$ exponentially concentrates around $(1 \pm \epsilon)\sqrt{n}$, statement (19) is equivalent to

$$\mathbb{P}[\text{PMEPR}(Z^n) \le \text{const} \cdot nA_n^2] = e^{o(n)}, \tag{20}$$

Notice that for Z^n being uniform on the hypercube $\{-1, +1\}^n$ the estimate (20) was shown by Spencer [18, Section 5], and it implies achievability of the binary GV bound with $\omega(1)$ PMEPR – see [6, Section 5.4]. From [18, (5.4)] there exist vectors $L_j \in \mathbb{R}^n, j = 1, \ldots, 4n$ with norms $\|L_j\|_2 = \sqrt{n}$ and such that (20) is equivalent to

$$\mathbb{P}\Big[\max_j |(L_j, Z^n)| \le \text{const} \cdot \sqrt{n} A_n\Big] = e^{o(n)}. \tag{21}$$

Note that $\mathbb{P}[(L_j, Z^n) \le \text{const} \cdot \sqrt{n} A_n] = 1 - Q(A_n^{-1}) = e^{o(1)}$. Finally, (21) follows from Šidák's lemma (see, e.g., [19, (2.8)]):

$$\mathbb{P}\Big[\max_j |(L_j, Z^n)| \le \text{const} \cdot \sqrt{n} A_n\Big] \ge (1 - Q(A_n^{-1}))^{4n} = e^{o(n)}.$$

From Proposition 1 it is evident that the question of minimal allowable PAPR is only meaningful for good codes, i.e. ones that attain $\log M^*(n, \epsilon, P)$ to within, say, terms of order $O(n^\alpha)$. The following lower bound is the main result of this note:

Theorem 2: Consider an (n, M, ϵ, P)-code for the AWGN channel with $\epsilon < 1/2$

$$\log M \ge nC(P) - \gamma n^\alpha \tag{22}$$

[2] This result was obtained in collaboration with Dr. Yuval Peres <peres@microsoft.com>.

for some $\alpha \in [1/2, 1)$ and $\gamma > 0$. Define

$$\delta_{\alpha,P} = (1-\alpha)(\sqrt{1+P}-1)^2. \tag{23}$$

Then for any $\delta < \delta_{\alpha,P}$, there exists an $N_0 = N_0(\alpha, P, \delta, \gamma, \epsilon)$, such that if $n \geq N_0$, then for any $n \times n$ orthogonal matrix U at least $\frac{M}{2}$ codewords satisfy

$$\|Ux^n\|_\infty \geq \sqrt{2\delta \log n}\,. \tag{24}$$

Remark 3: The function $\alpha \mapsto \delta_{\alpha,P}$ suggests there exists a tradeoff between the convergence speed and the peak amplitude for a fixed average power budget P. Choosing U to be the identity matrix, Theorem 2 implies that any sequence of codes with rate $C(P) - O(n^{-(1-\alpha)})$ needs to have PAPR at least

$$\frac{2\delta_{\alpha,P}}{P}\log n = \frac{2(1-\alpha)(\sqrt{1+P}-1)^2}{P}\log n\,.$$

In particular for $\alpha = \frac{1}{2}$, note that $\frac{\delta_{\frac{1}{2},P}}{P} \leq \frac{1}{2}$ for $P > 0$. On the other hand, X^n independently drawn from the optimal input distribution $\mathcal{N}(0, P)$ has PAPR $2\log n(1 + o(1))$ with high probability regardless of P. It is unclear what the optimal α-δ tradeoff is or whether it depends on the average power P.

Proof: We start with a few simple reductions of the problem. First, any code $\{\mathbf{c}_1, \ldots, \mathbf{c}_M\} \subset \mathbb{R}^n$ can be rotated to $\{U^{-1}\mathbf{c}_1, \ldots, U^{-1}\mathbf{c}_M\}$ without affecting the probability of error. Hence, it is enough to show (24) with $U = I_n$, the $n \times n$ identity matrix. Second, by taking some $\epsilon' > \epsilon$ and reducing the number of codewords from M to $M' = c_\epsilon M$ we may further assume that the resulting (n, M', ϵ') subcode has small *maximal* probability of error, i.e.

$$\mathbb{P}[\hat{W} \neq i | W = i] \leq \epsilon'\,, \qquad i \in \{1, \ldots, M\}\,.$$

Note that by Markov's inequality, $c_\epsilon \geq 1 - \frac{\epsilon}{\epsilon'}$. Since $\epsilon < 1/2$ we may have $c_\epsilon > 1/2$ by choosing $\epsilon' \in (2\epsilon, 1)$. Third, if a resulting code contains less than $\frac{M}{2}$ codewords satisfying (24), then by removing those codewords we obtain an (n, M'', ϵ', P) code such that

$$\log M' \geq nC(P) - \gamma n^\alpha - \log\left(c_\epsilon - \frac{1}{2}\right) \triangleq nC(P) - \gamma' n^\alpha\,.$$

Thus, overall by replacing γ with γ', M with M'' and ϵ with ϵ' it is sufficient to prove: Any (n, M, ϵ, P) code with maximal probability of error ϵ satisfying (22) must have at least one codeword such that

$$\|x^n\|_\infty \geq \sqrt{2\delta \log n}\,, \tag{25}$$

provided $n \geq N_0$ for some $N_0 \in \mathbb{N}$ depending only on $(\alpha, \epsilon, P, \gamma, \delta)$. We proceed to showing the latter statement.

In [20, Theorem 7] (see also [21]) it was shown that for any (n, M, ϵ, P) code with maximal probability of error ϵ we have

$$D(P_{Y^n} || P^*_{Y^n}) \leq nC(P) - \log M + a\sqrt{n}\,,$$

where $a > 0$ is some constant depending only on (ϵ, P), $P_{Y^n}^* = \mathcal{N}(0, 1+P)^n$ and P_{Y^n} is the distribution induced at the output of the channel (1) by the uniform message $W \in \{1, \ldots, M\}$. In the conditions of the theorem we have then

$$D(P_{Y^n}||P_{Y^n}^*) \le \gamma n^\alpha + a\sqrt{n} \le \gamma' n^\alpha, \tag{26}$$

where γ' can be chosen to be $\gamma + a$.

Next we lower bound $D(P_{Y^n}||P_{Y^n}^*)$ by solving the following I-projection problem:

$$u_n(A) = \inf_{P_{Y^n}} D(P_{Y^n}||\mathcal{N}(0, 1+P)^n)\,, \tag{27}$$

where P_{Y^n} ranges over the following convex set of distributions:

$$P_{Y^n} = P_{X^n} * \mathcal{N}(0,1)^n\,, \quad P_{X^n}[\|X^n\|_\infty \le A] = 1.$$

Since the reference measure in (27) is of product type and $D(P_{U^n}||\prod_{i=1}^n Q_{U_i}) \ge \sum_{i=1}^n D(P_{U_i}||Q_{U_i})$, we have

$$u_n(A) = n u_1(A)\,. \tag{28}$$

To lower bound $u_1(A)$, we use the Pinsker inequality [22, p. 58]

$$D(P||Q) \ge 2\log e\, \mathrm{TV}^2(P, Q), \tag{29}$$

where the total variation distance is defined by $\mathrm{TV}(P,Q) = \sup_E |P(E) - Q(E)|$ with E ranging over all Borel sets. Next we lower bound $\mathrm{TV}(P_{Y_1}, \mathcal{N}(0, 1+P))$ in a similar manner as in [23, Section VI-B]. To this end, let $Y_1^* \sim \mathcal{N}(0, 1+P)$. Fix $r > \frac{1}{\sqrt{1+P}-1}$. Since $\mathbb{P}[|X_1| \le A] = 1$, applying union bound yields

$$\begin{aligned}\mathbb{P}\left[|Y_1| > r\sqrt{1+P}A\right] &\le \mathbb{P}\left[|Z_1| > A(r\sqrt{1+P} - 1)\right] \\ &= 2Q(A(r\sqrt{1+P} - 1)).\end{aligned} \tag{30}$$

On the other hand,

$$\mathbb{P}\left[|Y_1^*| > r\sqrt{1+P}A\right] = 2Q(rA). \tag{31}$$

Assembling (30) and (31) gives

$$\mathrm{TV}(P_{Y_1}, \mathcal{N}(0, 1+P)) \ge Q(rA) - 2Q(A(r\sqrt{1+P} - 1)). \tag{32}$$

Combining (29) and (32), we have

$$u_1(A) \quad \ge \quad \left(Q(rA) - Q((r\sqrt{1+P} - 1)A)\right)^2 8\log e\,. \tag{33}$$

Suppose that $A_n \triangleq \|X^n\|_\infty \le \sqrt{2\delta \log n}$. Let $r = \frac{1}{\sqrt{1+P}-1} - \tau$ with $\tau > 0$. Note that for all $x > 0$,

$$\frac{x\varphi(x)}{1+x^2} \le Q(x) \le \frac{\varphi(x)}{x} \tag{34}$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the standard normal density. Assembling (26), (27), (28) and (33), we have

$$\begin{aligned}\gamma' n^{\alpha-1} &\geq \left(Q(r\sqrt{2\delta \log n}) - Q((r\sqrt{1+P}-1)\sqrt{2\delta \log n})\right)^2 8 \log e \\ &\geq c_1 \frac{n^{-\delta r^2}}{\sqrt{\log n}}. \end{aligned} \tag{35}$$

for all $n \geq N_0$, where c_1 and N_0 only depend on P and τ. Hence

$$\delta \geq \frac{1 - \alpha - \frac{c_2 \log\log n}{\log n}}{r^2}.$$

for some constant c_2 only depends on P and τ. By the arbitrariness of τ, we complete the proof of (25). ■

Theorem 3: Any (n, M, ϵ, P) code with maximal probability of error ϵ must contain a codeword x^n such that

$$\|x^n\|_\infty \geq A \tag{36}$$

where A is determined as the solution to

$$\left(Q(r^* A) - Q((r^*\sqrt{1+P}-1)A)\right)^2 8 \log e = C - \frac{1}{n}\log M + \sqrt{\frac{6(3+4P)}{n}} \log e + \frac{1}{n} \log \frac{2}{1-\epsilon},$$

where

$$r^* = \frac{\sqrt{A^2 + P\log(P+1)} + A\sqrt{P+1}}{AP}. \tag{37}$$

Remark 4 (Numerical evaluation): Consider SNR=20 dB ($P = 100$), $\epsilon = 10^{-3}$ and blocklength $n = 10^4$. Then, any code achieving 95%, 99% and 99.9% of the capacity is required to have PAPR -1.2 dB (trivial bound), 1.99 dB and 3.85 dB, respectively.

Proof: The proof in [20] actually shows

$$D(P_{Y^n} || P^*_{Y^n}) \leq nC - \log M + \sqrt{6n(3+4P)} \log e + \log \frac{2}{1-\epsilon}.$$

Let $A_n = \|x^n\|_\infty$. Using $D(P_{Y^n} || P^*_{Y^n}) \geq n u_1(A_n)$ and the lower bound on $u_1(A)$ in (33), we obtain the result after noticing that the right-hand side of (33) is maximized by choosing r as in (37). ■

III. Amplitude-constrained AWGN capacity

As an aside of the result in the previous section, we investigate the following question: How fast does the amplitude-constrained AWGN capacity converges to the classical AWGN capacity when the amplitude constraint grows? To this end, let us define

$$C(A, P) = \sup_{\substack{\mathbb{E}[X^2] \leq P \\ |X| \leq A \text{ a.s.}}} I(X; X+Z) \tag{38}$$

This quantity was first studied by Smith [2], who proved the following: For all $A, P > 0$, $C(A, P) < C(\infty, P) = \frac{1}{2}\log(1+P)$. Moreover, the maximizer of (38), being clearly non-Gaussian, is in fact finitely supported. Little is known about the cardinality or the peak amplitude of the optimal input. Algorithmic progress has been made in [24] where an iterative procedure for computing the capacity-achieving input distribution for (38) based on cutting-plane methods is proposed. On the other hand, the lower semi-continuity of mutual information immediately implies that

$C(A,P) \to \frac{1}{2}\log(1+P)$ as $A \to \infty$. A natural ensuing question is the speed of convergence. The next result shows that the backoff to Gaussian capacity due to amplitude constraint vanishes at the same speed as the Gaussian tail.

Theorem 4: For any $P > 0$ and $A \to \infty$ we have

$$e^{-\frac{A^2}{(\sqrt{1+P}-1)^2}+O(\ln A)} \le \frac{1}{2}\log(1+P) - C(A,P) \le e^{-\frac{A^2}{2P}+O(\ln A)}. \tag{39}$$

Remark 5: Non-asymptotically, for any $A, P > 0$, the lower (converse) bound in (39) is

$$\left(\frac{(\sqrt{1+P}-1)\log(1+P)}{A+A_1}\right)^2 \varphi^2\left(\sqrt{1+P}\frac{A_1}{P}+\frac{A}{P}\right) 8\log e\,, \tag{40}$$

and the upper (achievability) bound is

$$\frac{1}{1-2Q(\theta)}\left\{Q(\theta)\log\left(1+\frac{A\sqrt{P}}{1+P}\cdot\frac{\varphi(\theta)}{Q(\theta)}\right)+h(2Q(\theta))\right\} \tag{41}$$

where $\theta \triangleq \frac{A}{\sqrt{P}}$, $A_1 \triangleq \sqrt{A^2+P\log(1+P)}$, and $h(\cdot)$ denotes the binary entropy function.

Remark 6: Theorem 4 focuses on the fixed-P-large-A regime where the achievability is done by choosing a truncated Gaussian distribution as the input. It is interesting to compare our results to the case where A and $\sqrt{P}$ grow proportionally. To this end, fix $\alpha > 1$ and let $A = \sqrt{\alpha P}$. It is proved in [25, Theorem 1] that as $P \to \infty$, $\frac{1}{2}\log(1+P) - C(\sqrt{\alpha P}, \sqrt{P}) \to L(\alpha)$, where $L(\alpha)$ can be determined explicitly [25, Eq. (21)]. Moreover, let us denote the capacity-achieving input for (38) by $X^*_{A,P}$. Then as $P \to \infty$, $\frac{1}{\sqrt{P}}X^*_{\sqrt{\alpha P},P}$ converges in distribution to the uniform distribution (resp. a truncated Gaussian distribution) on $[-\sqrt{\alpha}, \sqrt{\alpha}]$ if $\alpha \le 3$ (resp. $\alpha > 3$). In particular, $L(3) = \frac{1}{2}\log\frac{\pi \mathrm{e}}{6}$ corresponds to the classical result of 1.53dB shaping loss [26]. The non-asymptotic bounds in Remark 5 yields a suboptimal estimate to $L(\alpha)$ in the proportional-growth regime.

Proof: The lower bound follows from the proof of Theorem 2 by noting that for any X such that $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] \le P$ and $|X| \le A$,

$$\begin{aligned}
\frac{1}{2}\log(1+P) - I(X;X+Z) &\ge \frac{1}{2}\log(1+\mathbb{E}[X^2]) - I(X;X+Z) \\
&= D(P_{X+Z}\,||\,\mathcal{N}(0,1+\mathbb{E}[X^2])) \\
&\ge D(P_{X+Z}\,||\,\mathcal{N}(0,1+P)) && (42)\\
&\ge u_1(A) && (43)\\
&\ge \left(Q(r^*A) - Q((r^*\sqrt{1+P}-1)A)\right)^2 8\log e, && (44)
\end{aligned}$$

where (42) follows from the fact that $\inf_{s>0} D(P_Y\,||\,\mathcal{N}(0,s)) = D(P_Y\,||\,\mathcal{N}(0,\mathbb{E}[Y^2]))$ for all zero-mean Y, while (43) and (44) follow from (27) and (33) with $r = r^*$ as in (37), respectively. We can then further lower bound (44) by $8\log e\varphi^2(b)(b-a)^2$, where

$$b \triangleq \sqrt{1+P}\frac{A_1}{P}+\frac{A}{P} > a \triangleq \sqrt{1+P}\frac{A}{P}+\frac{A_1}{P}$$

The proof of (40) is completed upon noticing that

$$b - a = \frac{(\sqrt{1+P}-1)\log(1+P)}{A+A_1}\,.$$

To prove the upper bound, we use the following input distribution: Let $X_* \sim \mathcal{N}(0,P)$. Let X_A and $\bar{X}_A$ be distributed according to X_* conditioned on the event $|X_*| \le A$ and $|X_*| > A$, i.e., $\mathbb{P}\left[X_A \in \cdot\right] = \frac{\mathbb{P}[X_* \in \cdot \cap [-A,A]]}{\mathbb{P}[X_* \in [-A,A]]}$. Then in view of (34) we have

$$\mathbb{E}\left[X_A^2\right] = P - \frac{2\theta P \varphi(\theta)}{1-2Q(\theta)} < P \tag{45}$$

$$\mathbb{E}[\bar{X}_A^2] = P + \frac{\theta P \varphi(\theta)}{Q(\theta)}. \tag{46}$$

Then

$$\begin{aligned}
\frac{1}{2}\log(1+P) &= I(X_*; X_* + Z) \\
&= I(X_*, \mathbf{1}_{\{|X_*|>A\}}; X_* + Z) \\
&\le I(X_A; X_A + Z)\mathbb{P}\left[|X_*| \le A\right] + I(\bar{X}_A; \bar{X}_A + Z)\mathbb{P}\left[|X_*| > A\right] + H(\mathbf{1}_{\{|X_*|>A\}}).
\end{aligned}$$

In view of (46), we have

$$(1-2Q(\theta))I(X_A; X_A+Z) \ge \frac{1}{2}\log(1+P) - Q(\theta)\log\left(1 + P + A\sqrt{P}\frac{\varphi(\theta)}{Q(\theta)}\right) - h(2Q(\theta)),$$

completing the proof of (41). ■

ACKNOWLEDGMENT

The authors thank Dr. T. Koch for stimulating discussions and for pointing out the reference [25].

REFERENCES

[1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[2] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, pp. 203 – 219, 1971.

[3] Y. Polyanskiy, "ℓ_p-norms of codewords from capacity-and dispersion-achieveing Gaussian codes," in *Proc. 2012 50th Allerton Conference*, Allerton Retreat Center, Monticello, IL, USA, Oct. 2012.

[4] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability (extended version)," *arXiv*, arXiv:1309.0141. [Online]. Available: http://arxiv.org/abs/1309.0141

[5] H. David and H. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, New Jersey, USA: Wiley-Interscience, 2003.

[6] S. Litsyn, *Peak power control in multicarrier communications*. Cambridge University Press, 2007.

[7] K. Paterson and V. Tarokh, "On the existence and construction of good codes with low peak-to-average power ratios," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1974–1987, 2000.

[8] H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 282–289, 2001.

[9] S. Litsyn and G. Wunder, "Generalized bounds on the crest-factor distribution of OFDM signals with applications to code design," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 992–1006, 2006.

[10] X. Li and L. J. Cimini Jr., "Effects of clipping and filtering on the performance of OFDM," *IEEE Commun. Lett.*, vol. 2, no. 5, pp. 131–133, 1998.

[11] S. H. Muller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electronics letters*, vol. 33, no. 5, pp. 368–369, 1997.

[12] R. W. Bauml, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electronics Lett.*, vol. 32, no. 22, pp. 2056–2057, 1996.

[13] J. Tellado, *Multicarrier modulation with low PAR: Applications to DSL and wireless*. New York, NY, USA: Kluwer Academic Press, 2000.

[14] B. S. Krongold and D. L. Jones, "PAR reduction in OFDM via active constellation extension," *IEEE Trans. Broadcast.*, vol. 49, no. 3, pp. 258–268, 2003.

[15] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun. Mag.*, vol. 12, no. 2, pp. 56–65, 2005.

[16] T. Jiang and Y. Wu, "An overview: peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Trans. Broadcast.*, vol. 54, no. 2, pp. 257–268, 2008.

[17] T. M. Cover and J. A. Thomas, *Elements of information theory, 2nd Ed.* New York, NY, USA: Wiley-Interscience, 2006.

[18] J. Spencer, "Six standard deviations suffice," *Transactions of the American Mathematical Society*, vol. 289, no. 2, pp. 679–706, 1985.

[19] A. A. Giannopoulos, "On some vector balancing problems," *Studia Mathematica*, vol. 122, no. 3, pp. 225–234, 1997.

[20] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," *IEEE Trans. Inf. Theory*, submitted. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/optcodes_journal.pdf

[21] Y. Polyanskiy and S. Verdú, "Relative entropy at the channel output of a capacity-achieving code," in *Proc. 2011 49th Allerton Conference*, Allerton Retreat Center, Monticello, IL, USA, Oct. 2011.

[22] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., 1982.

[23] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 29 – October 1 2010, pp. 620 – 628.

[24] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2336 – 2351, 2005.

[25] T. Koch, A. Martinez, and A. G. i Fábregas, "The capacity loss of dense constellations," in *2012 IEEE International Symposium on Information Theory Proceedings (ISIT)*, Boston, MA, USA, 2012, pp. 135–139.

[26] G. D. Forney Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, 1998.