Minimum energy to send k bits with and without feedback

Yury Polyanskiy, H. Vincent Poor, and Sergio Verdú

***Abstract*—The question of minimum achievable energy per bit over memoryless channels has been previously addressed in the limit of number of information bits going to infinity, in which case it is known that availability of noiseless feedback does not lower the minimum energy per bit. This paper analyzes the behavior of the minimum energy per bit for memoryless Gaussian channels as a function of the number of information bits. It is demonstrated that in this non-asymptotic regime, noiseless feedback leads to significantly better energy efficiency. A feedback coding scheme with zero probability of block error and finite energy per bit is constructed. For both achievability and converse, the feedback coding problem is reduced to a sequential hypothesis testing problem for Brownian motion.**

***Index Terms*—Shannon theory, minimum energy per bit, feedback, non-asymptotic analysis, AWGN channel, Brownian motion.**

I. Introduction and problem statement

A problem of broad practical interest is to transmit a message with minimum energy. For the additive white Gaussian noise (AWGN) channel, the key parameters of the code are: the number of degrees of freedom n, the number of information bits k, the probability of block error ϵ and the total energy budget E. Determining the region of feasible (n, k, ϵ, E) has received considerable attention in information theory, primarily in various asymptotic regimes.

The first asymptotic result, due to Shannon [1], demonstrates that in the limit of $\epsilon \to 0$, $k \to \infty$, $n \to \infty$ and $\frac{k}{n} \to 0$ the smallest achievable energy per bit $E_{\mathsf{b}} \triangleq \frac{E}{k}$ is

$$\left(\tfrac{E_{\mathsf{b}}}{N_0}\right)_{min} = \log_e 2 = -1.59 \text{ dB}\,, \tag{1}$$

where $\frac{N_0}{2}$ is the noise power per degree of freedom. The limit does not change if ϵ is fixed, if noiseless causal feedback is available at the encoder, if the channel is subject to fading, or even if the modulation is suitably restricted.

Alternatively, if one fixes $\epsilon > 0$ and rate $\frac{k}{n} = R$ then as $k \to \infty$ and $n \to \infty$ we have (e.g., [2])

$$\tfrac{E_{\mathsf{b}}}{N_0} \to \tfrac{1}{2R}(4^R - 1). \tag{2}$$

Thus in this case the minimum energy per bit becomes a function of R, but not ϵ. In contrast to (1), (2) is sensitive to modulation and fading scenarios; see [3].

Non-asymptotically, in the regime of fixed rate R and ϵ, bounds on the minimum E_{b} for finite k have been proposed [4], [5], studied numerically [6]–[8] and tightly approximated [5], [9].

In this paper we investigate the minimal energy E required to transmit k bits allowing error probability $\epsilon \ge 0$ and $n \to \infty$. Equivalently, we determine the maximal number of bits of information that can be transmitted with a fixed (non-asymptotic) energy budget and an error probability constraint, but without any limitation on the number of degrees of freedom used. This is different from [1] in that we do not take $k \to \infty$, and from [4]–[9] in that we do not fix a non-zero rate $\frac{k}{n}$. By doing so, we obtain the *bona fide* energy-information tradeoff for the AWGN channel. Even though (2) results in (1) by letting $R \to 0$, the minimum energy for finite k cannot be obtained from the asymptotic limit in (2).

The AWGN channel acts between input space $\mathsf{A} = \mathbb{R}^\infty$ and output space $\mathsf{B} = \mathbb{R}^\infty$ by addition:

$$\mathbf{y} = \mathbf{x} + \mathbf{z}\,, \tag{3}$$

where $\mathbb{R}^\infty$ is the vector space of real valued sequences[1] $(x_1, x_2, \ldots, x_n, \ldots)$, $\mathbf{x} \in \mathsf{A}$, $\mathbf{y} \in \mathsf{B}$ and $\mathbf{z}$ is a random vector with independent and identically distributed (i.i.d.) Gaussian components $Z_k \sim \mathcal{N}(0, N_0/2)$ independent of $\mathbf{x}$.

Definition 1: An (E, M, ϵ) code is a list of codewords $(\mathbf{c}_1, \ldots, \mathbf{c}_M) \in \mathsf{A}^M$, satisfying

$$||\mathbf{c}_j||^2 \le E\,, j = 1, \ldots, M\,, \tag{4}$$

and a decoder $g : \mathsf{B} \to \{1, \ldots, M\}$ satisfying

$$\mathbb{P}[g(\mathbf{y}) \ne W] \le \epsilon\,, \tag{5}$$

where $\mathbf{y}$ is the response to $\mathbf{x} = \mathbf{c}_W$, and W is the message which is equiprobable on $\{1, \ldots, M\}$. The fundamental energy-information tradeoff is given by

$$M^*(E, \epsilon) = \max\{M : \exists (E, M, \epsilon)\text{-code}\}\,. \tag{6}$$

Equivalently, we define the minimum energy per bit:

$$E_{\mathsf{b}}^*(k, \epsilon) = \frac{1}{k} \inf\{E : \exists (E, 2^k, \epsilon)\text{-code}\}\,. \tag{7}$$

Although, we are interested in (7), $M^*(E, \epsilon)$ is more suitable for expressing our results.

Definition 2: An (E, M, ϵ) code with feedback is a sequence of encoder functions $\{f_k\}_{k=1}^\infty$ determining the channel input as a function of the message W and the past channel outputs,

$$X_k = f_k(W, Y_1^{k-1})\,, \tag{8}$$

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544 USA. e-mail: {ypolyans,poor,verdu}@princeton.edu.

The research was supported by the National Science Foundation under Grants CCF-06-35154 and CNS-09-05398.

[1]In this paper, boldface letters $\mathbf{x}$, $\mathbf{y}$ etc. denote the infinite dimensional vectors with coordinates X_k, Y_k etc., correspondingly.

satisfying

$$\mathbb{E}\left[||\mathbf{x}||^2|W=j\right] \le E\,, j=1,\ldots,M\,, \qquad (9)$$

and a decoder $g : \mathsf{B} \to \{1,\ldots,M\}$ satisfying (5). The fundamental energy-information tradeoff with feedback is

$$M^*_{\mathsf{f}}(E,\epsilon) = \max\{M : \exists (E,M,\epsilon)\text{-feedback code}\}. \qquad (10)$$

And $E^*_{\mathsf{f}}(k,\epsilon)$ is defined similar to (7).

In the context of finite-blocklength codes without feedback, we showed in [5] that the maximum rate compatible with a given error probability ϵ for finite blocklength n admits a tight analytical approximation which can be obtained by proving an asymptotic expansion under fixed ϵ and $n\to\infty$. We follow the same approach in this paper obtaining upper and lower bounds on $\log M^*(E,\epsilon)$ and $\log M^*_{\mathsf{f}}(E,\epsilon)$ and corresponding asymptotics for fixed ϵ and $E\to\infty$.

II. Main results

A. No feedback

Theorem 1: For every $M > 0$ there exists an (E,M,ϵ) code for the channel (3) with[2]

$$\epsilon \;=\; \mathbb{E}\left[\min\left\{MQ\left(\sqrt{\tfrac{2E}{N_0}}+Z\right),1\right\}\right]\,, \qquad (11)$$

and $Z\sim\mathcal{N}(0,1)$. Conversely, any (E,M,ϵ) code without feedback satisfies

$$\frac{1}{M} \ge Q\left(\sqrt{\tfrac{2E}{N_0}}+Q^{-1}(1-\epsilon)\right)\,. \qquad (12)$$

Proof: To prove (11), notice that a codebook with M orthogonal codewords under a maximum likelihood decoder has probability of error equal to

$$P_e = 1-\frac{1}{\sqrt{\pi N_0}}\int_{-\infty}^{\infty}\left[1-Q\left(\sqrt{\frac{2}{N_0}}z\right)\right]^{M-1} e^{-\frac{(z-\sqrt{E})^2}{N_0}}dz\,. \qquad (13)$$

A change of variables $x=\sqrt{\frac{2}{N_0}}z$ and application of the bound $1-(1-y)^{M-1}\le\min\{My,1\}$ weakens (13) to (11).

To prove (12) fix an arbitrary codebook $(\mathbf{c}_1,\ldots,\mathbf{c}_M)$ and a decoder $g:\mathsf{B}\to\{1,\ldots,M\}$. We denote the measure $P^j = P_{\mathbf{y}|\mathbf{x}=\mathbf{c}_j}$ on $\mathsf{B}=\mathbb{R}^\infty$ as the infinite dimensional Gaussian distribution with mean $\mathbf{c}_j$ and independent components with individual variances equal to $\frac{N_0}{2}$; i.e.,

$$P^j = \prod_{k=1}^{\infty}\mathcal{N}(c_{j,k},N_0/2)\,,\quad n=1,2,\ldots \qquad (14)$$

where $c_{j,k}$ is the k-th coordinate of the vector $\mathbf{c}_j$. We also define an auxiliary measure

$$\Phi = \prod_{k=1}^{\infty}\mathcal{N}(0,N_0/2)\,,\quad n=1,2,\ldots \qquad (15)$$

Assume for now that for each j and event $F\in\mathcal{B}^\infty$ we have

$$P^j(F)\ge\alpha \Longrightarrow \Phi(F)\ge\beta_\alpha(E)\,, \qquad (16)$$

[2]As usual, $Q(x)=\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt\,.$

where the right-hand side of (12) is denoted by

$$\beta_\alpha(E) = Q\left(\sqrt{\tfrac{2E}{N_0}}+Q^{-1}(\alpha)\right)\,. \qquad (17)$$

From (16) we complete the proof of (12):

$$\begin{aligned}\frac{1}{M} &= \frac{1}{M}\sum_{j=1}^{M}\Phi(g^{-1}(j)) &(18)\\ &\ge \frac{1}{M}\sum_{j=1}^{M}\beta_{P^j(g^{-1}(j))}(E) &(19)\\ &\ge \beta_{1-\epsilon}(E)\,, &(20)\end{aligned}$$

where (18) follows because $g^{-1}(j)$ partitions the space B, (19) follows from (16), and (20) follows since the function $\alpha\to\beta_\alpha(E)$ is non-decreasing convex for any E and

$$\frac{1}{M}\sum_{j=1}^{M}P^j(g^{-1}(j))\ge 1-\epsilon \qquad (21)$$

is equivalent to (5), which holds for every (E,M,ϵ) code.

To prove (16) we compute the Radon-Nikodym derivative

$$\log_e\frac{dP^j}{d\Phi}(\mathbf{y}) = \sum_{k=1}^{\infty}\left(-\tfrac{1}{2}c_{j,k}^2+c_{j,k}Y_k\right)\,, \qquad (22)$$

which is Gaussian under both P^j and Φ. A simple analysis [15] then shows (16). This method closely parallels the meta-converse in [5, Theorem 26]. ■

Asymptotic analysis of (11) and (12) shows:

Theorem 2: In the absence of feedback, the number of bits that can be transmitted with energy E and error probability $0<\epsilon<1$ satisfies as $E\to\infty$:

$$\log M^*(E,\epsilon) = \frac{E}{N_0}\log e - \sqrt{\frac{2E}{N_0}}Q^{-1}(\epsilon)\log e + \frac{1}{2}\log\frac{E}{N_0} + O(1). \qquad (23)$$

B. Communication with feedback

We start by stating non-asymptotic converse and achievability bounds whose proofs are given in the appendix.

Theorem 3: Let $0\le\epsilon<1$. Any (E,M,ϵ) code with feedback for the channel (3) must satisfy

$$d\left(1-\epsilon||\tfrac{1}{M}\right)\le\frac{E}{N_0}\log e\,, \qquad (24)$$

where $d(x||y)=x\log\frac{x}{y}+(1-x)\log\frac{1-x}{1-y}$ is the binary relative entropy.

In the special case $\epsilon=0$ (24) reduces to $\log M\le\frac{E}{N_0}\log e$.

Theorem 4: For any $E>0$ and positive integer M there exists an (E,M,ϵ) code with feedback for the channel (3) satisfying

$$\epsilon\le\inf\left\{1-\alpha+(M-1)\beta\right\}\,, \qquad (25)$$

where the infimum is over all $0<\beta<\alpha\le 1$ satisfying

$$d(\alpha||\beta)=\frac{E}{N_0}\log e\,. \qquad (26)$$

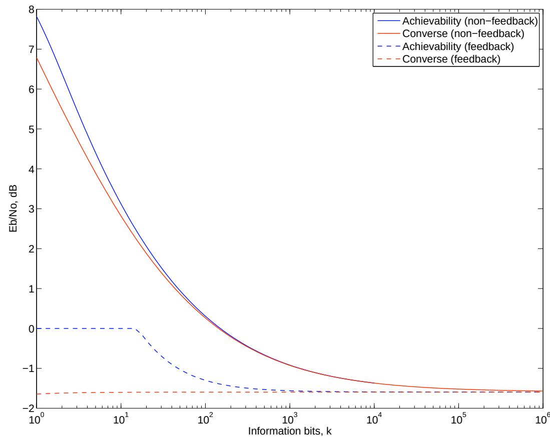

Fig. 1. Bounds on the minimum energy per bit as a function of the number of information bits with and without feedback; block error rate $\epsilon = 10^{-3}$.

Moreover, there exists an (E, M, ϵ) *decision feedback* code, which uses the feedback link only once to send a "ready-to-decode" signal; its probability of error is bounded by (25) with $\alpha = 1$, namely,

$$\epsilon \le (M-1)e^{-\frac{E}{N_0}} . \tag{27}$$

Asymptotic analysis of (24) and (27) shows:

Theorem 5: In the presence of feedback, the number of bits that can be transmitted with energy E and error probability $0 < \epsilon < 1$ satisfies as $E \to \infty$

$$\log M_{\mathrm{f}}^*(E, \epsilon) = \frac{E}{N_0} \frac{\log e}{1-\epsilon} + O\left(\log \frac{E}{N_0}\right) . \tag{28}$$

Note that as $\epsilon \to 0$, the leading term in (28) coincides with the leading term in (23). As we know, in the regime of arbitrarily reliable communication (and therefore $k \to \infty$) feedback does not help.

C. Zero-error communication

At first sight it may be plausible that infinite bandwidth may allow finite energy per bit when zero-error is required. However, a simple consequence of [10] is that without feedback

$$\log M^*(E, 0) = 0 \tag{29}$$

for all $E > 0$. With noiseless feedback the situation changes.

Theorem 6: For any positive integer k and $E > kN_0$ there exists an $(E, 2^k, 0)$-code with feedback. Equivalently, for all positive integers k we have

$$E_{\mathrm{f}}^*(k, 0) \le N_0 . \tag{30}$$

Proof: An $(E_1, M_1, 0)$ code and an $(E_2, M_2, 0)$ code can be combined into an $(E_1 + E_2, M_1 M_2, 0)$ code by using the first code on odd channel inputs and the second code on even. Therefore, to prove the theorem, it is sufficient to prove that for any $E > N_0$ there exists an $(E, 2, 0)$ code with feedback. To this end, we construct the following binary communication scheme. Fix an arbitrary $d > 0$, assume $W = \pm 1$ and consider the following code with feedback:

$$f_n(W, Y_1^{n-1}) = \begin{cases} Wd, & i(W; Y^{n-1}) \le i(-W; Y^{n-1}) , \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

where we have defined information densities

$$i(w; y_1^k) = \sum_{j=1}^{k} \log \frac{P_{Y_j|X_j}(y_j | f_j(w; y_1^{j-1}))}{P_{Y_j|Y_1^{j-1}}(y_j | y_1^{j-1})} . \tag{32}$$

Since the alternative in (31) depends on the difference of the information densities, it is convenient to define

$$\begin{aligned} S_n &= \log \frac{\mathbb{P}[W = +1 | Y^n]}{\mathbb{P}[W = -1 | Y^n]} & (33) \\ &= i(+1; Y_1^n) - i(-1; Y_1^n) . & (34) \end{aligned}$$

The main observation is that assuming $W = +1$ and regardless of the alternative in (31) we have for each $n > 1$

$$S_n = S_{n-1} + \tfrac{1}{2} d^2 + d\, Z_n . \tag{35}$$

From (35) we see that under $W = +1$, S_n is a submartingale drifting towards $+\infty$. Since the transmitter outputs $X_n = +d$ only when $S_n < 0$ and otherwise outputs $X_n = 0$, the positive drift of S_n implies that only finitely many X_n's will be different from zero with probability one. Another conclusion is that $P_{Y^\infty | W = +1}$ and $P_{Y^\infty | W = -1}$ are mutually singular and therefore W can be recovered from Y^∞ with zero error.

To finish the proof, we need to compute the average energy spent by our scheme. Conditioning on $W = +1$ we see that

$$||\mathbf{x}||^2 = \sum_{j=1}^{\infty} ||X_j||^2 = \sum_{j=1}^{\infty} d^2 1\{S_j \le 0\} . \tag{36}$$

To simplify the computation of $\mathbb{E}\left[||\mathbf{x}||^2\right]$, we replace $d\, Z_n$ in (35) with $W_{nd^2} - W_{(n-1)d^2}$, where W_t is a standard Wiener process. In this way, we can write

$$S_n = \left(\frac{s}{2} + \sqrt{\frac{N_0}{2}} W_s \right) |_{s = nd^2} , \tag{37}$$

i.e. S_n is just a sampling of W_t on a d^2-spaced grid. According to (36), $||\mathbf{x}||^2$ is a total number of negative samples multiplied by a grid step. Since every realization of W_t is continuous, as $d \to 0$ the $||\mathbf{x}||^2$ tends to the total time T the Brownian motion $\frac{t}{2} + \sqrt{\frac{N_0}{2}} W_t$ spends below zero:

$$\lim_{d \to 0} ||\mathbf{x}||^2 = T \triangleq \int_0^\infty 1_{\left\{\frac{t}{2} + \sqrt{\frac{N_0}{2}} W_t \le 0\right\}} dt . \tag{38}$$

Then, taking expectations we get that the average energy spent to transmit 1 bit is

$$\mathbb{E}\left[T\right] = \int_0^\infty \mathbb{P}\left[\frac{t}{2} + \sqrt{\frac{N_0}{2}} W_t \le 0\right] dt = N_0 . \tag{39}$$

Hence, $M_{\mathrm{f}}^*(E, 0) \ge 2$ for any $E > N_0$, as required. ∎

The weaker result that $M_{\mathrm{f}}^*(E, 0) \ge 2$ for sufficiently large E follows from [11, Lemma 4.2], which analyzes a modification of an original method of Zigangirov [12]. In contrast, our method is motivated by the Brownian motion analysis and antipodal signaling arising in the achievability proof of Section II-B. At the expense of a significantly more involved analysis, the bound in Theorem 6 can be further improved by using multidimensional constellations. It remains to be seen whether such a method could close the gap with (24).

III. Conclusion

As the number of information bits k goes to infinity, the minimum energy per bit required for arbitrarily reliable communication is equal to -1.59 dB with or without feedback. However, in the non-asymptotic regime, feedback substantially reduces the minimum energy per bit. Comparing Theorems 2 and 5, we observe a double benefit: feedback reduces the leading term in the minimum energy by a factor of $1-\epsilon$, and the penalty due to the second-order term in (23) disappears. According to Theorem 6 feedback enables zero-error transmission of any number of bits with finite energy per bit (with infinite degrees of freedom).

A quantitative analysis of the dependence of the required energy on the number of information bits is given in Fig. 1. The non-feedback upper (11) and lower (12) bounds are tight enough to conclude that for messages of size $k \sim 100$ bits the minimum $\frac{E_b}{N_0}$ is 0.20 dB, whereas the Shannon limit of -1.59 dB is only approachable at $k \sim 10^6$ bits. In contrast, with feedback the upper bound, which is the best of (25) and (30), and the lower bound (24) demonstrate that with feedback, -1.5 dB is achievable already at $k \sim 200$.

Surprisingly, virtually all of the discussed benefits of feedback can be achieved via decision feedback codes only; see [15]. In this way, the results of Section II-B extend to noisy and/or finite capacity feedback links.

Note that (3) also models an infinite-bandwidth continuous-time Gaussian channel (without feedback) with noise spectral density $\frac{N_0}{2}$ observed over an interval $[0,T]$. If we denote by $M_c^*(T,\epsilon)$ the maximum number of messages that is possible to communicate over such a channel with probability of error ϵ and power constraint P, then by taking $E = PT$ in Theorem 2 we have

$$\log M_c^*(T,\epsilon) = \frac{PT}{N_0}\log e - \sqrt{\frac{2PT}{N_0}}Q^{-1}(\epsilon)\log e + \frac{1}{2}\log\frac{PT}{N_0} + O(1) \quad (40)$$

sharpening the capacity result of Shannon [1].

References

[1] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10-21, Jan. 1949.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., New York, 1968.

[3] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319-1343, Jun. 2002.

[4] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Tech. Journal*, vol. 38, pp. 611-656, 1959.

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

[6] D. Slepian, "Bounds on communication," *Bell System Technical Journal*, vol. 42, pp. 681-707, 1963

[7] S. Dolinar, D. Divsalar, and F. Pollara, "Code performance as a function of block size," *JPL TDA Progress Report*, 42-133, Jet Propulsion Laboratory, Pasadena, CA, 1998.

[8] D. Buckingham and M. C. Valenti, "The information-outage probability of finite-length codes over AWGN channels," *Proc. 2008 Conf. on Info. Sci. and Sys. (CISS)*, Princeton, NJ, Mar. 2008

[9] Y. Polyanskiy, H. V. Poor and S. Verdú, "Dispersion of Gaussian channels," *Proc. 2009 IEEE Int. Symp. Inform. Theory (ISIT)*, Seoul, Korea, 2009.

[10] L. A. Shepp, "Distinguishing a sequence of random variables from a translate of itself," *Ann. Math. Stat.*, vol. 36, no. 4, pp. 1107-1112, 1965.

[11] R. G. Gallager and B. Nakiboglu, "Variations on a theme by Schalkwijk and Kailath, " *arXiv:0812.2709v4*, Nov. 2009.

[12] K. Sh. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Prob. Peredachi Inform.*, vol. 6, no. 2, pp. 87-92, 1970.

[13] A. N. Shiryaev, *Optimal Stopping Rules*, Springer: New York, 1978.

[14] A. D. Wyner, "On the Schalkwijk-Kailath coding scheme with a peak energy constraint," *IEEE Trans. Inform. Theory*, vol. 14, no. 1, pp. 129-134, Jan. 1968.

[15] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits with and without feedback," *IEEE Trans. Inf. Theory*, Mar. 2010, submitted for publication.

Appendix

Proof of Theorem 3: Consider an arbitrary (E, M, ϵ) code with feedback, namely a sequence of encoder functions $\{f_n\}_{n=1}^\infty$ and a decoder map $g : \mathsf{B} \to \{1,\dots,M\}$. The "meta-converse" part of the proof proceeds step by step as in the non-feedback case (14)-(21), with the exception that measures $P^j = P_{\mathbf{Y}|W=j}$ on B are defined as

$$P^j = \prod_{k=1}^\infty \mathcal{N}(f_k(j, Y_1^{k-1}), \tfrac{1}{2}N_0) \quad (41)$$

and β_α is replaced by $\tilde\beta_\alpha$, a unique solution $\tilde\beta < \alpha$ of

$$\tilde\beta_\alpha : \quad d(\alpha||\tilde\beta) = \frac{E}{N_0}\log e\,. \quad (42)$$

We need only to show that (16) holds with these modifications, i.e. for any $\alpha \in [0,1]$

$$\inf_{F\subset \mathsf{B}: P^j(F)\ge\alpha} \Phi(F) \ge \tilde\beta_\alpha\,. \quad (43)$$

Once $W = j$ is fixed, channel inputs X_k become functions on the space B defined as $X_k = f_k(j, Y_1^{k-1})$. To find the critical set F achieving the infimum in the hypothesis testing problem (43) we compute the Radon-Nikodym derivative:

$$R \stackrel{\triangle}{=} \log_e \frac{dP^j}{d\Phi} = \sum_{k=1}^\infty X_k Y_k - \tfrac{1}{2}X_k^2\,. \quad (44)$$

To prove a lower bound (43) we need to optimize over the choice of both X_k and the critical region F. The key simplification comes from identifying the noise random variables Z_k with increments of the Wiener process.

Formally, define a standard Wiener process W_t with the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ and two Brownian motions:

$$B_t = \frac{t}{2} + \sqrt{\tfrac{N_0}{2}}W_t\,, \quad (45)$$

$$\bar B_t = -\frac{t}{2} + \sqrt{\tfrac{N_0}{2}}W_t\,. \quad (46)$$

Then we can see that under P^j we have $Y_k = X_k + Z_k$ and hence we can assume

$$X_kY_k - \tfrac{1}{2}X_k^2 = B_{\tau_k} - B_{\tau_{k-1}}\,, \quad (47)$$

where we have denoted the instants $\tau_k = \sum_{m=1}^k X_m^2$. Then under P^j the distribution of R coincides with that of B_τ, where the random variable τ is defined as

$$\tau = \sum_{k=1}^\infty X_k^2\,. \quad (48)$$

Similarly, under Φ, we have $R \sim \bar{B}_\tau$.

Note that without loss of generality $X_k \neq 0$ since having $X_k = 0$ does not help in discriminating P^j vs. Φ. Then each Y_k can be recovered from $X_k Y_k - \frac{1}{2}X_k^2$ since X_k is known. Consequently, each X_k is a function of only the past observations $(B_0, B_{\tau_1}, \ldots, B_{\tau_{k-1}})$. This implies that each τ_k, and thus τ, is a stopping time of the filtration $\mathcal{F}_t$ satisfying

$$\mathbb{E}_{P^j}[\tau] \leq E \tag{49}$$

by the energy constraint (under P^j). Therefore, the encoder maps $\{f_n\}_{n=1}^{\infty}$ and the minimizing set F in (43) define a sequential hypothesis test, namely a stopping time τ and a decision region $F \in \mathcal{F}_\tau$, for discriminating between a Brownian motion with a positive drift B_t (under P) and a Brownian motion with a negative drift $\bar{B}_t$ (under Φ). According to Shiryaev [13, Section 4.2], among all (τ, F) satisfying (49) and having $P(F) \geq \alpha$ there exists an optimal one achieving[3]

$$\Phi(F) = \tilde{\beta}_\alpha \,, \tag{50}$$

where $\tilde{\beta}_\alpha$ is defined in (42). Any other test (τ, F) has a larger value of $\Phi(F)$, which proves (43). ∎

Proof of Theorem 4: Fix a list of elements $(\mathbf{c}_1, \ldots, \mathbf{c}_M) \in \mathsf{A}^M$ to be chosen later; $||\mathbf{c}_j||^2$ need not be finite. Upon receiving channel outputs $Y_1, \ldots, Y_n$ the decoder computes the likelihood $S_{j,n}$ for each codeword $j = 1, \ldots, M$, cf. (22) and (44):

$$S_{j,n} = \sum_{k=1}^{n} C_{j,k} Y_k - \tfrac{1}{2} C_{j,k}^2 \,, \quad j = 1, \ldots, M \,. \tag{51}$$

Fix two scalars $\gamma_0 < 0 < \gamma_1$ and define M stopping times

$$\tau_j = \inf\{n > 0 : S_{j,n} \notin (\gamma_0, \gamma_1)\} \,. \tag{52}$$

The decoder output $\hat{W}$ is the index j of the process $S_{j,n}$ that is the first to upcross γ_1 without having downcrossed γ_0 previously. The encoder conserves energy by transmitting only up until time τ_j (when the true message $W = j$):

$$X_n \triangleq f_n(j, Y_1^{n-1}) = C_{j,n} 1\{\tau_W \geq n\} \,. \tag{53}$$

To complete the construction of the encoder-decoder pair we need to choose $(\mathbf{c}_1, \ldots, \mathbf{c}_M)$. This is done by a random-coding argument. Fix $d > 0$ and generate each $\mathbf{c}_j$ independently with equiprobable antipodal coordinates:

$$\mathbb{P}[C_{j,k} = +d] = \mathbb{P}[C_{j,k} = -d] = \frac{1}{2} \,, \; j = 1, \ldots, M. \tag{54}$$

By symmetry the probability of error averaged over the choice of the codebook equals $\mathbb{P}[\hat{W} \neq 1 | W = 1]$ and thus henceforth probabilities conditioned on $W = 1$. We have

$$\mathbb{P}[\hat{W} \neq 1] \leq \mathbb{P}[S_{1,\tau_1} \leq \gamma_0] + \sum_{j=2}^{M} \mathbb{P}[S_{j,\tau_j} \geq \gamma_1, \tau_j \leq \tau_1] \tag{55}$$

because there are only two error mechanisms: S_1 downcrosses γ_0 before upcrossing γ_1, or some other S_j upcrosses γ_1 before S_1. Notice that in computing probabilities $\mathbb{P}[S_{1,\tau_1} \leq \gamma_0]$ and $\mathbb{P}[S_{2,\tau_2} \geq \gamma_1, \tau_2 \leq \tau_1]$ on the right-hand side of (55) we are interested only in time instants $0 \leq n \leq \tau_1$ and thus we can assume that $X_n = C_{j,n}$. Then under $W = 1$ the process S_1 can be rewritten as

$$S_{1,n} = B_{nd^2} \,, \tag{56}$$

where we define B_t and $\bar{B}_t$ as in (45) and (46). The stopping time τ_1 then becomes

$$d^2 \tau_1 = \inf\{t > 0 : B_t \notin (\gamma_0, \gamma_1) \,, t = nd^2, n \in \mathbb{Z}\} \,. \tag{57}$$

If we now define

$$\tau \;=\; \inf\{t > 0 : B_t \notin (\gamma_0, \gamma_1)\} \,, \tag{58}$$

$$\bar{\tau} \;=\; \inf\{t > 0 : \bar{B}_t \notin (\gamma_0, \gamma_1)\} \,, \tag{59}$$

then the path-continuity of B_t implies that $d^2 \tau_1 \searrow \tau$ as $d \to 0$. Similarly, still under the condition $W = 1$ we rewrite

$$S_{2,n} = d^2 \sum_{k=1}^{n} L_k + \bar{B}_{nd^2} \,, \tag{60}$$

where $L_k = \pm 1$ are i.i.d., independent of $\bar{B}_t$ and $\mathbb{P}[L_k = +1] = \frac{1}{2}$. It can be shown [15] that as $d \to 0$ we also have

$$\mathbb{P}[S_{1,\tau_1} \leq \gamma_0] \;\to\; 1 - \alpha(\gamma_0, \gamma_1) \,, \tag{61}$$

$$\mathbb{P}[S_{2,\tau_2} \geq \gamma_1, \tau_2 < \infty] \;\to\; \beta(\gamma_0, \gamma_1) \,, \tag{62}$$

where $\alpha(\gamma_0, \gamma_1)$ and $\beta(\gamma_0, \gamma_1)$ are

$$\alpha(\gamma_0, \gamma_1) \;=\; \mathbb{P}[B_\tau = \gamma_1] \,, \tag{63}$$

$$\beta(\gamma_0, \gamma_1) \;=\; \mathbb{P}[\bar{B}_{\bar{\tau}} = \gamma_1, \bar{\tau} < \infty] \,. \tag{64}$$

Thus, the interval (γ_0, γ_1) determines the boundaries of the sequential probability ratio test. As shown by Shiryaev [13, Section 4.2], α and β satisfy

$$d(\alpha(\gamma_0, \gamma_1) || \beta(\gamma_0, \gamma_1)) = \frac{\log e}{N_0} \mathbb{E}[\tau] \,. \tag{65}$$

By (55) and from (61) and (62), as $d \to 0$ we must have

$$\mathbb{P}[\hat{W} \neq 1] \leq 1 - \alpha(\gamma_0, \gamma_1) + (M-1)\beta(\gamma_0, \gamma_1) \,, \tag{66}$$

whereas the average energy spent by our scheme is

$$\lim_{d \to 0} \mathbb{E}[||\mathbf{x}||^2] = \lim_{d \to 0} \mathbb{E}[d^2 \tau_1] = \mathbb{E}[\tau] \,, \tag{67}$$

because $d^2 \tau_1 \searrow \tau$.

Finally, comparing (26) and (65) it follows that optimizing (66) over all $\gamma_0 < 0 < \gamma_1$ satisfying $\mathbb{E}[\tau] = E$ we obtain (25). To prove (27) simply notice that when $\alpha = 1$ we have $\gamma_0 = -\infty$, and hence the decision is taken by the decoder the first time any S_j upcrosses γ_1. Therefore, in the encoder rule (53) the time τ_j, whose computation requires the full knowledge of Y_k, can be replaced with the time of decoder decision. Obviously, this modification will not change the probability of error and will conserve energy even more (since $\gamma_0 = -\infty$ prohibits τ_j to occur before the decision time). ∎

[3] If instead of (9) we impose the maximum energy constraint: $||\mathbf{x}||^2 \leq E$ (a.s.), then $\tau \leq E$ and hence instead of $F \in \mathcal{F}_\tau$ we would have $F \in \mathcal{F}_E$, thus obtaining a usual, fixed-sample-size, binary hypothesis test. Then $\tilde{\beta}_\alpha$ should be replaced with β_α from (17) and thus, such an energy constraint renders feedback useless. This parallels the result of Wyner [14].