On Achievable Rates of AWGN Energy-Harvesting Channels with Block Energy Arrival and Non-Vanishing Error Probabilities

Silas L. Fong, Vincent Y. F. Tan, and Ayfer Özgür

Abstract

This paper investigates the achievable rates of an additive white Gaussian noise (AWGN) energy-harvesting (EH) channel with an infinite battery. The EH process is characterized by a sequence of blocks of harvested energy, which is known causally at the source. The harvested energy remains constant within a block while the harvested energy across different blocks is characterized by a sequence of independent and identically distributed (i.i.d.) random variables. The blocks have length $L$, which can be interpreted as the coherence time of the energy arrival process. If $L$ is a constant or grows sublinearly in the blocklength $n$, we fully characterize the first-order term in the asymptotic expansion of the maximum transmission rate subject to a fixed tolerable error probability $\varepsilon$. The first-order term is known as the $\varepsilon$-capacity. In addition, we obtain lower and upper bounds on the second-order term in the asymptotic expansion, which reveal that the second order term scales as $\sqrt{L/n}$ for any $\varepsilon$ less than $1/2$. The lower bound is obtained through analyzing the save-and-transmit strategy. If $L$ grows linearly in $n$, we obtain lower and upper bounds on the $\varepsilon$-capacity, which coincide whenever the cumulative distribution function (cdf) of the EH random variable is continuous and strictly increasing. In order to achieve the lower bound, we have proposed a novel adaptive save-and-transmit strategy, which chooses different save-and-transmit codes across different blocks according to the energy variation across the blocks.

Index Terms

Achievable rates, block energy arrival, energy-harvesting, non-vanishing error probabilities, save-and-transmit

I. INTRODUCTION

The energy-harvesting (EH) channel consists of one source equipped with an energy buffer (also called battery), and one destination. For simplicity, we assume in this paper that the buffer has infinite capacity. At each discrete time $k \in \{1, 2, \ldots\}$, a random amount of energy $E_k \in [0, \infty)$ arrives at the buffer and the source transmits a symbol $X_k \in (-\infty, \infty)$ such that

$$\sum_{i=1}^{k} X_i^2 \leq \sum_{i=1}^{k} E_i \quad \text{almost surely.} \quad (1)$$

S. L. Fong is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: silas_fong@nus.edu.sg).
V. Y. F. Tan is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, and also with the Department of Mathematics, National University of Singapore, Singapore 119076 (e-mail: vtan@nus.edu.sg).
A. Özgür is with the Department of Electrical Engineering, Stanford University, CA 94305, USA (email: aozgur@stanford.edu).
This implies that the total harvested energy \( \sum_{i=1}^{k} E_i \) must be no smaller than the energy of the codeword \( \sum_{i=1}^{k} X_i^2 \) at every discrete time \( k \) for transmission to take place successfully. The knowledge of \( E_k \) is available at the source at time \( k \) before encoding \( X_k \), and the destination has no access to the energy-arrival process.

We assume that \( \{E_i\}_{i=1}^{\infty} \) arrive at the buffer in a block-by-block manner as follows: For each \( \ell \in \mathbb{N} \), let
\[
b_{\ell} \triangleq (\ell - 1)L
\]
such that \( b_{\ell} + 1 \) is the index of the first channel use within the \( \ell \)th block of energy arrival, where \( L \) denotes the length of each block. In other words, the \( \ell \)th block of energy arrival starts at the \( (b_{\ell} + 1) \)th channel use. The EH random variables that mark the starting points of the blocks (i.e., \( \{E_{b_{\ell}+1}\}_{\ell=1}^{\infty} \)) are assumed to be independent and identically distributed (i.i.d.) random variables where\(^1\) \( \mathbb{E}[E_1^3] < +\infty \) and \( \mathbb{E}[E_1] = P \) for some \( P > 0 \). A large class of distributions of practical interests satisfy the third-moment assumption including those with well-defined moment generating function. In addition, we assume
\[
E_{b_{\ell}+1} = E_{b_{\ell}+2} = \ldots = E_{b_{\ell}+L}
\]
for all \( \ell \in \mathbb{N} \). In other words, the harvested energy in each channel use within a block remains constant while the harvested energy across different blocks is characterized by a sequence of i.i.d. random variables with mean equal to \( P \). This block-by-block energy-arrival assumption is useful for modeling practical scenarios when the energy-arrival process evolves at a slower timescale compared to the transmission process [1, Sec. II-C]. This is often the case for most natural energy processes, such as solar energy or wind energy. The block i.i.d. EH process can model, for example, a solar panel which harvests energy from the sun, and the appearance of clouds can change randomly and block certain amounts of sunshine for a certain period of time. Similarly, this is a good model for a device which harvests RF energy from other transmitting devices in its environment. Such transmitting devices typically transmit continuously for certain periods of time and are silent for the remaining periods (as in TDMA, for example), which warrants a block i.i.d. model. Most importantly, the block i.i.d. model provides a simple way to study the impact of correlations in the EH process on the system capacity. Such block i.i.d. models are popularly used in wireless communication as a means to capture correlations in the channel fading process by a simple model. In that context, the block length \( L \) is called the coherence time of the channel, which corresponds to the time duration over which the channel remains approximately constant [2, Sec. 2.3]. Analogously, we refer to \( L \) as the coherence time of the energy arrival process.

The channel noise of the EH channel is modeled as an additive white Gaussian noise (AWGN), which is described as follows. In each time slot \( k \in \mathbb{N} \), after the source has transmitted \( X_k \), the destination receives
\[
Y_k = X_k + Z_k
\]
where \( \{Z_k\}_{k=1}^{\infty} \) are i.i.d. standard normal random variables. The above EH channel is referred to as the AWGN EH

\(^1\)If the constraint \( \mathbb{E}[E_1^3] < +\infty \) is replaced with the less stringent one \( \mathbb{E}[E_1^2] < +\infty \), all the achievability results in this paper continue to hold. In fact, the only place that requires \( \mathbb{E}[E_1^3] < +\infty \) is the use of the Berry-Esséen theorem in Section VI-B in the course of proving the converse of Theorem 1.
channel. It was shown by Ozel and Ulukus [3] that the capacity of the AWGN EH channel for the case $L = 1$ is

$$C(P) \triangleq \frac{1}{2} \log(1 + P) \quad \text{bits per channel use},$$

where $P = \mathbb{E}[E_1]$ is the expectation of the harvested energy for each energy arrival. In this paper, we assume that $L$ can grow with $n$ and would like to investigate how the growth rate of $L$ affects the first- and second-order terms of the asymptotic expansion of the maximum transmission rate. The first-order term is also known as the $\varepsilon$-capacity [4, Sec. 3.4], and the second-order term divided by an appropriate scaling (which is $\frac{1}{\sqrt{n}}$ in many cases including the AWGN channel) is known as the second-order coding rate [5]. The following two cases regarding the growth rate of $L$ will be investigated in this paper:

(i) $L$ is a constant or $L$ grows sublinearly in $n$. The latter statement means

$$\omega(1) = L = o(n).$$

(ii) $L$ grows linearly in $n$.

Note that in practice $L$ and $n$ are two independent parameters. The first one is dictated by the nature of the EH process and the second one is a design parameter typically dictated by the delay and reliability requirements of the application and the complexity constraints at the transmitter and the receiver. Depending on how fast the EH process changes over time, $L$ can be significantly smaller than $n$ or comparable to $n$. In order to reveal the impact of the interplay between these two parameters on the second-order term, we couple these parameters in different ways, say $L = n^\gamma$, and study the limiting case when both $L$ and $n$ approach $\infty$ for different couplings, i.e., different values of $\gamma$ in $[0, 1]$ where $\gamma$ captures how large $L$ is with respect to $n$. This allows us to identify how $L$ and $n$ together impact the second-order term. In particular, we conclude that it is the ratio of the two that determines the second-order term. The impact of correlation in the EH process can be interpreted as effectively decreasing the blocklength by a factor of $L$. Note that keeping $L$ constant while taking $n$ to infinity, i.e., considering only the special case $\gamma = 0$, would lead to a degenerate regime where the correlation in the EH process does not play a significant role. The approach we take here has been extensively used in the wireless information theory literature to obtain asymptotic results in multi-parameter problems where the problem involves multiple independent parameters that can be large or small with respect to each other. (See for example the notion of generalized degrees of freedom in [6] and follow-up work, or Section 3.1 of [7] for a detailed discussion of a similar formulation in the context of scaling laws for wireless networks.)

A. Main Contribution

We use $O(\cdot)$, $o(\cdot)$, $\omega(\cdot)$ and $\Theta(\cdot)$ denote standard asymptotic Bachmann-Landau notations except our convention that they must be non-negative. The contributions of this paper are summarized in the following:

Case (i): When $L$ is a constant or grows sublinearly in $n$

1. First, we prove an achievable finite blocklength bound based on the save-and-transmit strategy of [3]. During the saving phase of the save-and-transmit strategy, energy is saved for a certain number of time slots. During
this period, no information is transmitted. Subsequently, during the transmission phase, the source uses a Gaussian codebook to send information. In order to analyze the performance of this save-and-transmit strategy, we construct a single sequence of random variables that characterizes the probability of the available energy being insufficient to support the Gaussian codeword (i.e., $\sum_{i=1}^{k} E_i < \sum_{i=1}^{k} X_i^2$ for all $k$) and derive a concentration bound related to the random sequence. Our analysis reveals that the backoff from capacity $C(P)$ for the optimal length-$n$ code with error probability less than $\varepsilon$ is no larger than $O\left(\sqrt{\frac{L}{n}}\right)$. More specifically, the maximum alphabet size of the message we can transmit over $n$ channel uses with average probability of error no larger than $\varepsilon \in (0, 1/2)$, denoted by $M^*_{n, \varepsilon}$, satisfies

$$\frac{1}{n} \log M^*_{n, \varepsilon} \geq C(P) + V^- \sqrt{\frac{L}{n}} - o\left(\sqrt{\frac{L}{n}}\right)$$

(7)

where $V^- < 0$ is some constant that does not depend on $n$. We also identify the implied constant $V^-\varepsilon$ in Theorem 1 in Section III. The qualitative interpretation of $V^-\varepsilon$ will be given in Remark 3 in Section III-C. The lower bound (7) is obtained by choosing the lengths of the saving phase and transmission phase to be $\Theta(\sqrt{Ln})$ and $n - \Theta(\sqrt{Ln})$ respectively, which is illustrated in Figure 1 where the accumulated harvested energy is always above the accumulated transmitted energy due to the EH constraints (1).

2. Second, we prove a non-asymptotic upper bound on achievable rates by simplifying the type-II error of a carefully chosen binary hypothesis test. The first-order term of the upper bound is $C(P)$ and the second-order term is proportional to $-\sqrt{\frac{L}{n}}$ for all $\varepsilon \in (0, 1/2)$. More specifically, for all $\varepsilon \in (0, 1/2)$, we have

$$\frac{1}{n} \log M^*_{n, \varepsilon} \leq C(P) + V^+ \sqrt{\frac{L}{n}} + o\left(\sqrt{\frac{L}{n}}\right)$$

(8)

where $V^+ < 0$ is some constant that does not depend on $n$. We also identify the implied constant $V^+\varepsilon$ in Theorem 1 in Section III. The qualitative interpretation of $V^+\varepsilon$ will be given in Remark 4 in Section III-C. Note that (7) and (8) together reveal that the back-off from $C(P)$ for the optimal length-$n$ code with error probability less than $\varepsilon$ is of the order $\sqrt{\frac{L}{n}}$. Therefore, the impact of correlation in the EH process can be interpreted as effectively decreasing the blocklength by a factor of $L$. In other words, to achieve the same reliability, one needs to increase the blocklength by a factor equal to the coherence time of the EH process.

Fig. 1. Save-and-transmit strategy for $L = o(n)$. 
It is readily seen from (7) and (8) that for any fixed $\varepsilon \in (0, 1/2)$, the $\varepsilon$-capacity is $C(P)$ and the second-order term in the asymptotic expansion for the maximum achievable rate is proportional to $-\sqrt{\frac{L}{n}}$. In addition, define

$$V_\varepsilon \triangleq \sup \left\{ S \in \mathbb{R} \mid \liminf_{n \to \infty} \frac{\log M_{n,\varepsilon}^* - nC(P)}{\sqrt{Ln}} \geq S \right\}$$

(9)

to be the second-order coding rate [5]. We can see from (7) and (8) that for any fixed $\varepsilon \in (0, 1/2)$, the second-order coding rate is sandwiched between $V_\varepsilon^-$ and $V_\varepsilon^+$.

Case (ii): When $L$ grows linearly in $n$

1. First, we prove a lower bound on the $\varepsilon$-capacity, as shown in Theorem 2 in Section III, based on a modified version of the save-and-transmit strategy called the *adaptive save-and-transmit strategy*. Under the adaptive save-and-transmit strategy which is described in Section VII-A, different save-and-transmit codes are used across different blocks. In each block $\ell$, the coding rate is adapted to the corresponding EH random variable $E_{b_{\ell+1}}$ so that it is close to $C(E_{b_{\ell+1}})$. In addition, the lengths of the saving phase and transmission phase for block $\ell$ are chosen to be $\Theta(\sqrt{L})$ and $L - \Theta(\sqrt{L})$ respectively as illustrated in Figure 2.

2. Second, we prove an upper bound on the $\varepsilon$-capacity (Theorem 2). We do so by considering a typical set of sequences of EH random variables followed by simplifying the type-II error of a binary hypothesis test conditioned on the aforementioned typical set.

For any EH process whose EH random variable has a continuous and strictly increasing cumulative density function (cdf), the upper and lower bounds in Theorem 2 coincide and hence the $\varepsilon$-capacity is fully characterized. See Remark 6 in Section III-C for a detailed discussion. Case (ii) is useful for modeling the scenario where the energy-harvesting rate changes slowly such that the number of energy-arrival blocks stays constant as $n$ increases. Since the number of energy-arrival blocks stays constant and the length of each energy-arrival block grows with $n$, it is first-order optimal to choose an appropriate save-and-transmit scheme that achieves the maximum coding rate for each block according to the energy level in that block. Therefore, we need an adaptive save-and-transmit scheme rather than the conventional non-adaptive one to achieve the overall maximum coding rate.
B. Related Work

The channel capacity was characterized for the AWGN channel with an i.i.d. EH process in [3] and with a stationary ergodic EH process in [8]. The aforementioned studies showed that with an unlimited battery, the capacity of the AWGN channel with stochastic energy constraints is equal to the capacity of the same channel under an average power constraint as long as the average power equals the average recharge rate of the battery. In this paper, we focus on the AWGN channel with a block EH process, where the energy arrivals remain constant for a block of duration \( L \) and are independent across blocks drawn from an arbitrary distribution. A similar block i.i.d. EH model has been recently considered in [9,10] concurrently with the current paper. However, these papers focus on the power control problem for EH communications with finite battery at the transmitter. In this paper, we rather consider the information-theoretic capacity of the channel and with infinite battery at the transmitter. Characterizing the information theoretic capacity of the channel with a finite battery is known to be a difficult problem even for an i.i.d. model for the energy arrivals and in general remains an open problem. It has been studied in several recent works [11]–[14] and the most recent ones [13,14] characterize the capacity within a constant gap. Due to the lack of a complete characterization of the capacity under a finite battery assumption, in this paper we focus on the AWGN EH channel with infinite battery and develop bounds on the first- and second-order terms in the asymptotic expansion of the maximum transmission rate.

For a fixed tolerable error probability \( \varepsilon \), Fong et al. [15] recently performed a finite blocklength analysis of save-and-transmit schemes proposed in [3] and obtained a non-asymptotic achievable rate for the AWGN channel with an i.i.d. EH process. The first-, second- and third-order terms of the non-asymptotic achievable rate presented in [15, Th. 1] are equal to the capacity, \(-c_1 \sqrt{\frac{\log n}{n}}\) and \(-c_2 \sqrt{\frac{2+\varepsilon}{n\varepsilon}}\) respectively where \( c_1 \) and \( c_2 \) are some positive constants that do not depend on \( n \) and \( \varepsilon \). Subsequently, Shenoy and Sharma [16] refined the analysis in [15] and improved the second-order term to \(-\frac{c_3}{\sqrt{n\varepsilon}}\) where \( c_3 \) is some positive constant that does not depend on \( n \) and \( \varepsilon \). This paper further improves the second-order term to \(-c_4 \sqrt{\frac{\log(1/\varepsilon)}{n}}\) for any \( \varepsilon \in (0, 1/2) \) where \( c_4 \) is some positive constant that does not depend on \( n \) and \( \varepsilon \) (see Remark 2). The aforementioned improvements are due to better analyses of the “energy outage” probability for the same save-and-transmit strategy, where the “energy outage” occurs when the source cannot output the desired codeword due to energy shortage.

C. Paper Outline

This paper is organized as follows. The notation used in this paper is described in the next subsection. Section II states the formulation of the AWGN EH channel with block energy arrival. Section III presents our two main results — the first result fully characterizes the \( \varepsilon \)-capacity and provides lower and upper bounds on the second-order coding rate when \( L \) is a constant or grows sublinearly in \( n \); the second result presents lower and upper bounds on the \( \varepsilon \)-capacity when \( L \) grows linearly in \( n \), where the two bounds coincide for random variables with continuous and strictly increasing cdf. In Section IV, we present the proof of the first main result, which relies on a save-and-transmit achievability lemma and a converse lemma. The proofs of the achievability and converse lemmas are provided respectively in Sections V and VI, which are briefly described as follows. Section V describes the save-and-transmit strategy which is the key to the achievability part of the first result. More specifically, we use
Shannon’s achievability bound [17] to prove a non-asymptotic achievable rate for the save-and-transmit strategy. Section VI proves the converse part of the first result, and the proof technique involves simplifying a non-asymptotic bound derived from the type-II error of a binary hypothesis test. In Section VII, we provide the proof of the second result when \( L \) grows linearly in \( n \). Concluding remarks are provided in Section VIII.

D. Notation

The sets of natural, real and non-negative real numbers are denoted by \( \mathbb{N} \), \( \mathbb{R} \) and \( \mathbb{R}_+ \) respectively. We let \( 1 \{ \mathcal{E} \} \) be the indicator function of the set \( \mathcal{E} \). An arbitrary (discrete or continuous) random variable is denoted by an upper case letter (e.g., \( X \)), and the realization and alphabet of the random variable are denoted by the corresponding lower case letter (e.g., \( x \)) and calligraphic letter (e.g., \( \mathcal{X} \)) respectively. We use \( X^n \) to denote the random tuple \((X_1, X_2, \ldots, X_n)\).

The following notations are used for any arbitrary random variables \( X \) and \( Y \) and any real-valued function \( g \) with domain \( X \). We let \( p_{X,Y} \) and \( p_{Y|X} \) denote the probability distribution of \((X,Y)\) and the conditional probability distribution of \( Y \) given \( X \) respectively. More specifically, \( p_{X,Y} \) is the Radon-Nikodym derivative of a measure with respect to the Lebesgue measure in an appropriate Euclidean space. We let \( p_{X,Y}(x,y) \) and \( p_{Y|X}(y|x) \) be the evaluations of \( p_{X,Y} \) and \( p_{Y|X} \) respectively at \((X,Y) = (x,y)\). To make the dependence on the distribution explicit, we let \( \Pr_{p_X} \{ g(X) \in A \} \) denote \( \int_X p_X(x) 1 \{ g(x) \in A \} \, dx \) for any set \( A \subseteq \mathbb{R} \). The expectation and the variance of \( g(X) \) are denoted as \( \mathbb{E}_{p_X} [g(X)] \) and \( \text{Var}_{p_X} [g(X)] \) respectively. We let \( \mathcal{N} (\cdot; \mu, \sigma^2) : \mathbb{R} \to [0, \infty) \) denote the probability density function of a Gaussian random variable whose mean and variance are \( \mu \) and \( \sigma^2 \) respectively, i.e.,

\[
\mathcal{N}(z; \mu, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{10}
\]

The cdf of the standard normal distribution is denoted by \( \Phi \), i.e.,

\[
\Phi(a) \triangleq \int_{-\infty}^a \mathcal{N}(z; 0, 1) \, dz. \tag{11}
\]

We will take all logarithms to base 2 throughout this paper unless specified otherwise. The logarithm function to base 2 is denoted by \( \log \), and the natural logarithm function is denoted by \( \ln \).

II. ADDITIVE WHITE GAUSSIAN NOISE ENERGY-HARVESTING CHANNEL WITH BLOCK ENERGY ARRIVAL

A. Problem formulation

The AWGN EH channel consists of one source and one destination, denoted by \( s \) and \( d \) respectively. Node \( s \) transmits information to node \( d \) in \( n \) time slots as follows. Node \( s \) chooses message \( W \) and sends \( W \) to node \( d \), where \( W \) is uniformly distributed over \( \{1, 2, \ldots, M\} \) for some \( M \) that denotes the message size. Then for each \( k \in \{1, 2, \ldots, n\} \), node \( s \) transmits \( X_k \in \mathbb{R} \) and node \( d \) receives \( Y_k \in \mathbb{R} \) in time slot \( k \). Let \( \{E_{b+1}\}_{b=1}^{\infty} \) be i.i.d. random variables that satisfy \( \Pr \{ E_1 < 0 \} = 0 \) (\( b_e \) was defined in (2)),

\[
\mathbb{E}[E_1] = P \tag{12}
\]
and $\mathbb{E}[E_1^2] < \infty$ (hence $\mathbb{E}[E_1^2] < \infty$) for some $P > 0$. Each other $E_k$ is equal to the nearest preceding $E_{b\ell+1}$ according to (3). In other words, for each $k \in \{1, 2, \ldots, n\}$ and all $e^k \in \mathbb{R}_+^k$,

$$p_{E_k|E^{k-1}}(e_k|e^{k-1}) = \begin{cases} p_{E_1}(e_k) & \text{if } k = b\ell + 1 \text{ for some } \ell \in \mathbb{N}, \\ 1\{e_k = e_{k-1}\} & \text{otherwise.} \end{cases}$$  \hfill (13)

The knowledge of $E_k$ is available at the source at time $k$ before encoding $X_k$, and the destination has no access to the energy-arrival process. The length of each energy-arrival block $L$ is assumed to remain constant, grow sublinearly in $n$, or grow linearly in $n$. We assume the following for each $k \in \{1, 2, \ldots, n\}$:

(I) $E_k$ and $(W, X^{k-1}, Y^{k-1})$ are independent when conditioned on $E^{k-1}$, i.e.,

$$p_{W,E^k,X^{k-1},Y^{k-1}} = p_{E_k|E^{k-1}}p_{W,E^{k-1},X^{k-1},Y^{k-1}}.$$  \hfill (14)

(II) Every codeword $X^n$ transmitted by $s$ must satisfy the harvested energy constraint

$$\Pr \left\{ \sum_{i=1}^k X_i^2 \leq \sum_{i=1}^k E_i \middle| E^n = e^n, W = w \right\} = 1$$ \hfill (15)

for each $e^n \in \mathbb{R}_+^n$ and each $w \in \mathcal{W}$.

Assumption (I) is a mathematical statement of the following fact due to the block i.i.d. EH process: If $E_k$ is the first energy-arrival random variable in a block, then it is independent of any random variables that are generated before time $k$. Otherwise, $E_k$ equals $E_{k-1}$. In both cases, $E_k$ and $(W, X^{k-1}, Y^{k-1})$ are independent when conditioned on $E^{k-1}$.

After $n$ time slots, node $d$ declares $\hat{W}$ to be the transmitted $W$ based on $Y^n$. The standard definitions are formally stated in the following subsection.

**B. Standard definitions**

**Definition 1:** An $(n, M)$-code consists of the following:

1) A message set $\mathcal{W} \triangleq \{1, 2, \ldots, M\}$ at node $s$. Message $W$ is uniform on $\mathcal{W}$.

2) A sequence of encoding functions $f_k : \mathcal{W} \times \mathbb{R}_+^k \to \mathbb{R}$ for each $k \in \{1, 2, \ldots, n\}$, where $f_k$ is the encoding function for node $s$ at time slot $k$ for encoding $X_k$ such that $X_k = f_k(W, E^k)$ and (15) holds.

3) A decoding function $\varphi : \mathbb{R}^n \to \mathcal{W}$ for decoding $W$ at node $d$ where the message estimate $\hat{W}$ is produced by setting $\hat{W} \triangleq \varphi(Y^n)$.

**Definition 2:** The AWGN EH channel is characterized by $q_{Y|X} \triangleq \mathcal{N}(y - x; 0, 1)$. The distribution induced by any $(n, M)$-code used for the AWGN EH channel follows the channel law below: For each $k \in \{1, 2, \ldots, n\}$,

$$p_{W,E^k,X^k,Y^k} = p_{W,E^k,X^k,Y^{k-1}}p_{Y_k|X_k}$$ \hfill (16)

where

$$p_{Y_k|X_k}(y_k|x_k) = q_{Y|X}(y_k|x_k) = \mathcal{N}(y_k - x_k; 0, 1)$$ \hfill (17)
for all \(x_k\) and \(y_k\). Since \(p_{Y_k|X_k}\) does not depend on \(k\) by (17), the channel is stationary.

For any \((n,M)\)-code defined on the AWGN EH channel, let \(p_{W,E^n,X^n,Y^n}\tilde{W}\) be the joint distribution induced by the code. We can use Definition 1, (14) and (16) to factorize \(p_{W,E^n,X^n,Y^n}\tilde{W}\) as follows:

\[
p_{W,E^n,X^n,Y^n,\tilde{W}} = p_W p_{E^n} \left( \prod_{k=1}^{n} p_{X_k|W,E^k} p_{Y_k|X_k} \right) p_{\tilde{W}|Y^n}.
\] (18)

**Definition 3:** For an \((n,M)\)-code defined on the AWGN EH channel, we can calculate, according to (18), the average probability of decoding error defined as \(\text{Pr}\{\tilde{W} \neq W\}\). We call an \((n,M)\)-code with average probability of decoding error no larger than \(\varepsilon\) an \((n,M,\varepsilon)\)-code.

**Definition 4:** Let \(\varepsilon \in (0,1)\) be a real number. A rate \(R\) is \(\varepsilon\)-achievable for the AWGN EH channel if there exists a sequence of \((n,M,\varepsilon)\)-codes such that
\[
\lim_{n \to \infty} \frac{1}{n} \log M \geq R.
\] (19)

**Definition 5:** Let \(\varepsilon \in (0,1)\) be a real number. The \(\varepsilon\)-capacity of the AWGN EH channel, denoted by \(C_\varepsilon\), is defined to be \(C_\varepsilon \triangleq \sup \{R | R\text{ is }\varepsilon\text{-achievable}\}\).

### III. Main Results

Section III-A contains the first main result in this paper, which concerns the \(\varepsilon\)-capacity and the second-order coding rate when \(L\) is a constant or grows sublinearly in \(n\). Section III-B contains the second main result in this paper, which concerns the \(\varepsilon\)-capacity when \(L\) grows linearly in \(n\).

**A. When \(L\) is a constant or grows sublinearly in \(n\)**

In this section, we assume that \(L\) is a constant or \(\omega(1) = L = o(n)\) so that \(\lim_{n \to \infty} \frac{L}{n} = 0\). Our goal in this section is to formalize the results in (7) and (8). Before presenting the first main result, we define the second-order achievable rate as follows.

**Definition 6:** Let \(\varepsilon \in (0,1)\). A real number \(S\) is said to be a second-order \(\varepsilon\)-achievable rate if there exists a sequence of \((n,M,\varepsilon)\)-codes such that
\[
\lim_{n \to \infty} \frac{1}{n} \log M - n C_\varepsilon \geq S.
\] (20)

The justification of the choice of \(\sqrt{Ln}\) in (20) will be explained after the following definition concerning the second-order coding rate is presented.

**Definition 7:** Let \(\varepsilon \in (0,1)\). The \(\varepsilon\)-second-order coding rate is defined as
\[
V_\varepsilon \triangleq \sup \{S \in \mathbb{R} | S\text{ is a second-order }\varepsilon\text{-achievable rate }\}.
\] (21)

\(^2\)Although \(M\) always depends on \(n\), it is not explicitly indicated to simplify notation.

\(^3\)Although \(L\) can depend on \(n\), it is not explicitly indicated to simplify notation.
The choice of $\sqrt{Ln}$ in (20) can be justified as follows by inspecting (29) in the main theorem. More specifically, if we replace $\sqrt{Ln}$ in (20) with any $f(n)$ such that $\lim_{n \to \infty} \frac{\sqrt{Ln}}{f(n)} \in \{0, \infty\}$ and define $V_{\epsilon}^*$ as in Definition 7, it will then follow from (29) that

$$V_{\epsilon}^* = \begin{cases} 
0 & \text{if } \lim_{n \to \infty} \frac{\sqrt{Ln}}{f(n)} = 0, \\
+\infty & \text{or } -\infty & \text{if } \lim_{n \to \infty} \frac{\sqrt{Ln}}{f(n)} = \infty.
\end{cases}$$

Our choice of $\sqrt{Ln}$ in (20) is analogous to the choice of $n^\beta$ in [18, Sec. II-D] which studies the $\epsilon$-second-order coding rate of channels with states.

We are ready to present the first main result in this paper.

**Theorem 1:** Suppose $L$ is a constant or $\omega(1) = L = o(n)$. Fix any $\epsilon \in (0, 1)$. Recalling the definition of $C(\cdot)$ in (5), we have

$$C_{\epsilon} = C(P).$$

(23)

In addition, define

$$\varrho \triangleq 2 \left( \frac{\mathbb{E}[E^2_1]}{P^2} + 1 \right),$$

(24)

and

$$V_{\epsilon}^- \triangleq \begin{cases} -C(P) \sqrt{\varrho \log \frac{1}{\epsilon}} & \text{if } \omega(1) = L = o(n), \\
\sup_{(\epsilon_1, \epsilon_2) \in (0,1)^2; \epsilon_1 + \epsilon_2 = \epsilon} \left\{-C(P) \sqrt{\varrho \log \frac{1}{\epsilon_1}} + \sqrt{\frac{P(\log e)^2}{L(1+P)}} \Phi^{-1}(\epsilon_2) \right\} & \text{if } L \text{ is a constant},
\end{cases}$$

(25)

and

$$V_{\epsilon}^+ \triangleq \frac{\log e}{2(1 + P)} \sqrt{2P^2 + \mathbb{E}[E^2_1]} \Phi^{-1}(\epsilon).$$

(26)

Then, the $\epsilon$-second-order coding rate $V_{\epsilon}$ satisfies

$$V_{\epsilon}^- \leq V_{\epsilon} \leq V_{\epsilon}^+.$$ 

(27)

In other words, if we let

$$M_{n,\epsilon}^* \triangleq \sup \{ M \in \mathbb{N} \mid \text{There exists an } (n, M, \epsilon)-\text{code} \}$$

(28)

be the maximum alphabet size of the message we can transmit using an $(n, M, \epsilon)$-code, then

$$C(P) + V_{\epsilon}^- \sqrt{\frac{L}{n}} - o\left( \sqrt{\frac{L}{n}} \right) \leq \frac{1}{n} \log M_{n,\epsilon}^* \leq C(P) + V_{\epsilon}^+ \sqrt{\frac{L}{n}} + o\left( \sqrt{\frac{L}{n}} \right).$$

(29)

Theorem 1 presents a complicated lower bound on $V_{\epsilon}$ as stated in (25). The following corollary presents a simpler lower bound, which implies that $V_{\epsilon}$ scales as $-O\left( \sqrt{\log \frac{1}{\epsilon}} \right)$. Since the proof of Corollary 1 is straightforward, it is relegated to Appendix A.
Corollary 1: Fix an \( \varepsilon \in (0, 1/2) \). Following the definitions in Theorem 1, if we define

\[
V_{\varepsilon^-} \triangleq \begin{cases} 
-C(P)\sqrt{\log \frac{1}{\varepsilon}} & \text{if } \omega(1) = L = o(n), \\
-C(P)\sqrt{2\theta} + \sqrt{\frac{4P\log \varepsilon}{1+P}} \sqrt{\log \frac{1}{\varepsilon}} & \text{if } L \text{ is a constant},
\end{cases}
\]

then

\[V_{\varepsilon^-} \leq V_{\varepsilon^-} \leq V_{\varepsilon} \leq V_{\varepsilon^+}.\] (31)

The following corollary presents an explicit bound on \( V_{\varepsilon^+} - V_{\varepsilon^-} \), whose proof relies on Corollary 1 and is relegated to Appendix B.

Corollary 2: Fix an \( \varepsilon \in (0, \Phi(-1)) \) (note that \( \Phi(-1) \approx 0.1586 \)). Following the definitions in Theorem 1, we have

\[
V_{\varepsilon^+} - V_{\varepsilon^-} \leq \begin{cases} 
(C(P)\sqrt{\theta} - \sqrt{\frac{(2P^2+\sum E_j^2)\log e}{2(1+P)^2}}) \sqrt{\log \frac{1}{\varepsilon}} & \text{if } \omega(1) = L = o(n), \\
(C(P)\sqrt{2\theta} + \sqrt{\frac{4P\log \varepsilon}{1+P}} \sqrt{\log \frac{1}{\varepsilon}}) & \text{if } L \text{ is a constant}.
\end{cases}
\]

This work does not intend to optimize the bound in (32), which can be arbitrarily large as \( P \) approaches infinity or \( \varepsilon \) approaches 0.

B. When \( L \) grows linearly in \( n \)

Before presenting the second main result, we make some necessary definitions. Fix an arbitrary \( \lambda \in (0, 1) \) and assume that

\[L = [\lambda n].\] (33)

Define

\[q \triangleq \left\lfloor \frac{1}{\lambda} \right\rfloor\] (34)

and

\[d \triangleq 1 - q\lambda\] (35)

to be the quotient and remainder respectively resulting from dividing 1 by \( \lambda \). The following theorem is our second main result, which provides lower and upper bounds on \( C_\varepsilon \). The proof of the lower and upper bounds will be given in Sections VII-A and VII-B respectively.

Theorem 2: Suppose \( L \) grows linearly in \( n \) according to (33) for some constant \( \lambda \in (0, 1) \). Let \( q \) and \( d \) be as defined in (34) and (35) respectively, and recall that \( p_{E_1} = p_{E_2} = \ldots = p_{E_n} \). Define

\[
R_\varepsilon \triangleq \sup_{\delta > 0} \sup_{r \in \mathbb{R}_+} \Pr_{\Pi_{i=1}^{q+1} p_{E_i}} \left\{ \sum_{\ell=1}^{q} \lambda C(E_\ell) + dC(E_{q+1}) \geq r \right\} \geq 1 - \varepsilon + \delta
\]

and

\[
\overline{R}_\varepsilon \triangleq \inf_{\delta > 0} \inf_{r \in \mathbb{R}_+} \Pr_{\Pi_{i=1}^{q+1} p_{E_i}} \left\{ \sum_{\ell=1}^{q} \lambda C(E_\ell) + dC(E_{q+1}) \geq r \right\} \leq 1 - \varepsilon - \delta.
\] (37)
Then, we have
\[ R_\varepsilon \leq C_\varepsilon \leq R^*_\varepsilon \] (38)
for all \( \varepsilon \in (0,1) \).

The following corollary identifies a sufficient condition under which \( C_\varepsilon \) can be fully characterized by Theorem 2. The proof of Corollary 3 is straightforward and hence relegated to Appendix C.

**Corollary 3:** Under the setting of Theorem 2, if we further assume that \( E_1 \) has a continuous and strictly increasing cdf, then
\[ C_\varepsilon = R^\text{thr}_\varepsilon \] (39)
holds for all \( \varepsilon \in (0,1) \) where \( R^\text{thr}_\varepsilon \) is the unique threshold that satisfies
\[ \Pr_{\prod_{l=1}^{q+1} P_{E_l}} \left\{ \sum_{\ell=1}^{q} \lambda C(E_\ell) + dC(E_{q+1}) < R^\text{thr}_\varepsilon \right\} = \varepsilon. \] (40)

**C. Remarks on Theorem 1 and Corollary 1**

**Remark 1:** It is already known [15, Remark 1] that
\[ C_\varepsilon = C(P) \] (41)
for all \( \varepsilon \in (0,1) \) under an i.i.d. EH process. In other words, the AWGN EH channel admits the **strong converse property** [4, Ch. 3] for \( L = 1 \), meaning that \( C_\varepsilon \) does not depend on \( \varepsilon \in (0,1) \). It follows from (23) in Theorem 1 that (41) remains to hold under a block i.i.d. EH process when \( L = o(n) \). An intuitive explanation about why the strong converse property holds when \( L = o(n) \) is as follows. When \( L = o(n) \), since the number of energy-arrival blocks \( n/L \) grows to infinity, it follows from the strong law of large numbers that the received power \( \frac{1}{n} \sum_{k=1}^{n} E_k \) converges to \( \mathbb{E}[E_1] = P \) with probability 1, which leads to a strong converse proof.

**Remark 2:** Consider the special case where \( L = 1 \) and fix any \( \varepsilon \in (0,1/2) \). Clearly, both \( V^-_\varepsilon \) in (25) and \( V^+_\varepsilon \) in (26) are negative. Therefore, it follows from (29) in Theorem 1 that the second-order term in the asymptotic expansion of \( \frac{1}{n} \log M^* \) scales as \(-\Theta\left(\sqrt{\frac{1}{n}}\right)\). In particular, it follows from Corollary 1 that \( V^-_\varepsilon \leq V_\varepsilon \), meaning that the second-order term scales as \(-O\left(\sqrt{\frac{\log(1/\varepsilon)}{n}}\right)\). This improves the previous findings in [15, Th. 1] and [16] which established that the second-order term scales as \(-O\left(\sqrt{\frac{\log n}{n}}\right)\) and \(-O\left(\frac{1}{\sqrt{\varepsilon}}\right)\) respectively.

**Remark 3:** Suppose \( L = o(n) \) and fix any \( \varepsilon \in (0,1/2) \). Clearly, \( V^-_\varepsilon \) in (25) is negative. In addition, the left hand side (LHS) of (29) which involves \( V^-_\varepsilon \) is the rate achievable by the save-and-transmit strategy (whose details can be found in Section V-B). For a fixed \( P \) and a fixed \( \mathbb{E}[E_1^2] \), since \( V^-_\varepsilon \) is negative, it follows from the LHS of (29) that the rate achievable by the save-and-transmit strategy will increase at a slower rate if \( L \) approaches infinity at a faster rate. This can be explained by the fact that block i.i.d. EH processes with longer \( L \) result in higher probabilities of “energy outage” — the source cannot output the desired codeword due to energy shortage. Similarly for a fixed \( P \), since \( |V^-_\varepsilon| \) increases as the variance \( \text{Var}[E_1] = \mathbb{E}[E_1^2] - P^2 \) increases, it follows that block i.i.d. EH processes with larger variance \( \text{Var}[E_1] \) result in higher probabilities of “energy outage”.

Remark 4: Suppose $L = o(n)$ and fix any $\varepsilon \in (0, 1/2)$. Clearly, $V_\varepsilon^+$ in (26) is negative. For a fixed $P$, since $V_\varepsilon^+$ is negative, it follows that the right hand side (RHS) of (29) increases at a slower rate if the following holds:

(*) $L$ approaches infinity at a faster rate or $\text{Var}[E_1]$ is increased.

In addition, it was shown in the previous remark that the LHS of (29) increases at a slower rate if (*) holds. Consequently, both the LHS and RHS of (29) increase at slower rates if (*) holds, which implies that the maximum rate achievable by an $(n, M^*_{n,\varepsilon,\varepsilon})$-code increases at a slower rate if (*) holds.

Remark 5: Suppose $L = o(n)$. The achievability proof of Theorem 1 is based on analyzing the save-and-transmit strategy, which was illustrated in Figure 1 and will be formally discussed in Section V. Equation (25) in Theorem 1 is indeed a lower bound on the second-order coding rate achieved by the save-and-transmit strategy. By inspecting (25) we see that the two components that dominate the lower bound achieved by save-and-transmit are the saving period (contributed by the two terms with $-C(P)$ in (25)) and the Gaussian noise (contributed by the term with $\varepsilon^2$ in (25)). If $L$ is a constant, both components contribute to the rate loss of the lower bound on the second-order coding rate achieved by save-and-transmit because the length of the saving period is $\Theta(\sqrt{L(n+L)}) = \Theta(\sqrt{n})$ and the minimum rate backoff needed to overcome the Gaussian noise is $\Theta(\sqrt{n})$, which correspond to the quantities $-C(P)\sqrt{\rho \log \frac{1}{\varepsilon}}$ and $\sqrt{\frac{P(\log e)^2}{L(1+P)}}\Phi^{-1}(\varepsilon^2)$ in (25) respectively. If $L = \omega(1)$, the term $\sqrt{\frac{P(\log e)^2}{L(1+P)}}\Phi^{-1}(\varepsilon^2)$ vanishes and the resultant lower bound achieved by save-and-transmit is the quantity $-C(P)\sqrt{\rho \log \frac{1}{\varepsilon}}$ in (25), meaning that the length of the saving period dominates the lower bound.

D. Remarks on Theorem 2 and Corollary 3

Remark 6: Suppose $L$ grows linearly in $n$ and $E_1$ has a continuous and strictly increasing cdf. Using the formula of the $\varepsilon$-capacity provided by Corollary 3, we conclude that $C_\varepsilon$ is strictly increasing on $(0, 1)$, implying that the strong converse property ceases to hold. An intuitive explanation about why the strong converse property does not hold is as follows: Since the number of energy-arrival blocks $n/L$ remains constant and the cdf of $E_1$ is continuous and strictly increasing, the received power $\frac{1}{n} \sum_{k=1}^{n} E_k$ does not converge (with probability 1) to a constant, which leads to the impossibility of a strong converse.

Remark 7: Consider the special case where $L = n$ and the cdf of $E_1$ is continuous and strictly increasing. Let $F_{E_1}(e) = \Pr\{E_1 \leq e\}$ be the cdf of $E_1$. It then follows from Corollary 3 with the identifications $\lambda = 1$, $q = 1$ and $d = 0$ that

$$C_\varepsilon = R_{\varepsilon}^{\text{thr}} = F^{-1}_{E_1}(\varepsilon)$$

for all $\varepsilon \in (0, 1)$, which is analogous to the $\varepsilon$-capacities (outage capacities) of slow fading channels as stated in [19, Sec. 23.3.1], the $\varepsilon$-capacities of channels with mixed states as stated in [18, Example 1], and the $\varepsilon$-capacities of mixed channels as stated in [4, Example 3.4.2].

Remark 8: Suppose $L$ grows linearly in $n$. The achievability proof of Theorem 2 is based on designing an adaptive save-and-transmit code that enables the source to adjust the transmission rate for each energy-arrival block according to the changes of harvested energy across different energy-arrival blocks. The adaptive save-and-transmit code was
illustrated in Figure 2 and will be formally discussed in Section VII-A. Equation (36) in Theorem 2 is the coding rate achievable by the adaptive save-and-transmit strategy. By inspecting (36), we see that the main event that dominates the coding rate achievable by adaptive save-and-transmit is the “slow fading” behavior of the EH process — the energy-harvesting rate changes slowly such that the number of energy-arrival blocks stays constant as \( n \) increases.

**Remark 9:** Suppose \( L \) grows linearly in \( n \). The converse proof of Theorem 2 is proved by considering a typical set of energy-arrival sequences followed by simplifying the conditional type-II errors of some binary hypothesis tests where the type-II errors are conditioned on the sequences in the typical set. In particular, the typical set is defined through (221) in the converse proof in Section VII-B, and the energy-arrival sequence falls into the set with high probability by (226).

### IV. Proof of Theorem 1

The achievability proof of Theorem 1 relies on the following lemma, whose proof will be presented in Section V.

**Lemma 4:** Fix any \( \varepsilon \in (0, 1) \), \( \varepsilon_1 > 0 \) and \( \varepsilon_2 > 0 \) such that

\[
\varepsilon_1 + \varepsilon_2 = \varepsilon. \tag{43}
\]

Recall the definition of \( \varrho \) in (24). Then for all sufficiently large \( n \), there exist a natural number

\[
m \leq \sqrt{\rho L n \log \frac{1}{\varepsilon_1} + O(L)} \tag{44}
\]

and an \((n + m, M, \varepsilon)\)-code such that

\[
\log M \geq nC(P) + \sqrt{\rho P n [\log e]^2} \frac{\Phi^{-1}(\varepsilon)}{1 + P} - \frac{1}{2} \log n - \kappa_1 \tag{45}
\]

for some constant \( \kappa_1 \). More specifically, \( \kappa_1 \) is defined as

\[
\kappa_1 \triangleq \sqrt{\frac{P}{1 + P} (\tau_1 + 1) \log e} \frac{(1 + P)^{3/2}}{\Phi^{-1}(\varepsilon_2)} + 1 \tag{46}
\]

where

\[
\tau_1 \triangleq \left( 15^{1/3} \sqrt{P} + \frac{8 \sqrt{\pi}}{\sqrt{P}} \right)^3 \tag{47}
\]

In addition, equation (45) holds for any sufficiently large \( n \in \mathbb{N} \) that satisfies

\[
n > \frac{2L \log(1/\varepsilon_1)}{E[E_1^2] + P^2} \cdot \max \left\{ 4P^2, \frac{(E_1^2)^2}{4P^2} \right\}, \tag{48}
\]

\[
n \geq \frac{L(E_1^2)^2 \log(1/\varepsilon_1)}{P^4}, \tag{49}
\]

and

\[
\varepsilon_2 - \varepsilon_2^2 - \frac{\tau_1}{\sqrt{n}} = \frac{1}{\sqrt{n}} \geq 0, \tag{50}
\]
and $m$ can be chosen to satisfy
\[
m \leq \frac{\left( E[E_1^2] + \frac{3P^2}{2} \left( 1 - 2\sqrt{L\log(1/\epsilon_1)} \right) \right)^{\frac{1}{2}}}{\frac{P}{2} + \frac{E[E_1^2]}{2} + \sqrt{L\log(1/\epsilon_1)}} \sqrt{(Ln + L^2)\log(1/\epsilon_1)} + 2L + 1.
\]

**Remark 10:** Lemma 4 guarantees the existence of a carefully designed save-and-transmit scheme with the saving phase being no greater than the RHS of (51) and the message size being no less than the RHS of (45). Here, $\epsilon_1$ specifies the probability of energy outage induced by energy shortage and $\epsilon_2$ specifies the probability of decoding error induced by noise for the save-and-transmit scheme. As indicated by (51), the designed saving phase has to be increased as $\epsilon_1$ decreases. In addition, as indicated by (45), the designed message size has to be decreased as $\epsilon_2$ decreases.

The following corollary is a direct consequence of Lemma 4. The proof of Corollary 5 is given in Appendix D for completeness.

**Corollary 5:** Fix any $\epsilon \in (0, 1)$. For any $\epsilon_1 > 0$ and $\epsilon_2 > 0$ that satisfy $\epsilon_1 + \epsilon_2 = \epsilon$, there exists a sequence of $(n^*, M, \epsilon)$-codes such that
\[
\liminf_{n^* \to \infty} \frac{\log M - n^*C(P)}{\sqrt{Ln^2}} \geq \begin{cases} 
-C(P)\sqrt{\log \frac{1}{\epsilon_1}} & \text{if } L = \omega(1), \\
-C(P)\sqrt{\log \frac{1}{\epsilon_1} + \sqrt{\frac{P(\log e)^2}{L} \Phi^{-1}(\epsilon_2)}} & \text{if } L \text{ is a constant}.
\end{cases}
\]

The converse proof of Theorem 1 relies on the following lemma, whose proof will be presented in Section VI.

**Lemma 6:** Fix any $\epsilon \in (0, 1)$. For any sufficiently large $n$ and any $(n, M, \epsilon)$-code, we have
\[
\log M \leq (n + L)C(P) + \frac{\sqrt{n + L\log e}}{2(1 + P)} \sqrt{2P(P + 2) + L(E[E_1^2] - P^2) \Phi^{-1}(\epsilon) + \frac{1}{2} \log(n + L) + \kappa_2}
\]
for some $\kappa_2 = O(L)$. More specifically, $\kappa_2$ is defined as
\[
\kappa_2 \triangleq \frac{\tau_2}{1 + P} \sqrt{2LP(P + 2) + L^2(E[E_1^2] - P^2) \log e} \Phi^{-1}(\min\{\epsilon, \epsilon(1 - \epsilon)\}); 0, 1) - \log(\tau_2\sqrt{L})
\]
where
\[
\tau_2 \triangleq \left( \frac{15^{1/3}P + 2(2\sqrt{2/\pi})^{1/3}}{L}(E[E_1^{3/2}])^{1/3} + (E[E_1^3])^{1/3} \right)^3 \left( \frac{2P(P + 2)}{L} + E[E_1^2] - P^2 \right)^{3/2}.
\]

In addition, equation (53) holds for any sufficiently large $n$ that satisfies
\[
n \geq \frac{4L\tau_2^2}{(1 - \epsilon)^3}.
\]

**Remark 11:** For any $\epsilon \in (0, 1/2)$, since $\Phi^{-1}(\epsilon)$ is negative, it follows from Corollary 5 and Lemma 6 that the second-order term in the asymptotic expansion of $\log M_{n,\epsilon}$ is $-O(\sqrt{L(n + L)}) = -O(\sqrt{Ln})$.

We are now ready to prove Theorem 1.

**Proof of Theorem 1:** For any $\epsilon \in (0, 1)$, the left inequality of (29) follows directly from Corollary 5 and the
V. PROOF OF LEMMA 4 VIA THE SAVE-AND-TRANSMIT STRATEGY

In this section, we investigate the save-and-transmit scheme proposed in [3, Sec. IV] in the finite blocklength regime. We will use this achievability scheme to prove Lemma 4.

A. Prerequisites

The following lemma is useful for obtaining a lower bound on the length of the energy-saving phase. The proof is deferred to Appendix E.

**Lemma 7:** Let $m$ and $n$ be two natural numbers. Suppose $\{X_k\}_{k=1}^n$ and $\{E_k\}_{k=1}^{m+n}$ are two sequences of i.i.d. random variables such that $X^n$ and $E^{m+n}$ are independent,

\[
\Pr\{E_1 < 0\} = 0, \quad \mathbb{E}[E_1] = \mathbb{E}[X_1^2] = P \quad (57)
\]

and

\[
\mathbb{E}[E_2^2] < \infty. \quad (59)
\]

Suppose there exists a sufficiently small $t \in (0, 1)$ such that $\mathbb{E}[X_1^4 e^{tX_1^2}] < \infty$ and

\[
\alpha_t \triangleq P - \frac{t\mathbb{E}[E_1^2]}{2} > 0, \quad (60)
\]

and we define

\[
b_t \triangleq \max \left\{ 0, \frac{\mathbb{E}[E_1^2] + \mathbb{E}[X_1^4 e^{tX_1^2}]}{2} - P^2 + \frac{tP}{2} (\mathbb{E}[E_1^2] - \mathbb{E}[X_1^4 e^{tX_1^2}]) + \frac{t^2 \mathbb{E}[E_1^2] \mathbb{E}[X_1^4 e^{tX_1^2}]}{2} \right\}. \quad (61)
\]

Then,

\[
\Pr_{X^n, E^{m+n}} \left\{ \sum_{k=1}^n \sum_{i=1}^k X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \leq e^{-\alpha_t m + b_t n^2}. \quad (62)
\]

In order to adapt Lemma 7 to the block energy arrival setting, we define the following quantities for each $t > 0$ and each $L \in \mathbb{N}$ (cf. (60) and (61)):

\[
\alpha_t \triangleq L \left( P - \frac{tL \mathbb{E}[E_1^2]}{2} \right), \quad (63)
\]
\[
\beta_0 \triangleq \frac{L^2(\mathbb{E}[E_1^2] + P^2)}{2},
\]
and
\[
\beta_t \triangleq L^2 \max \left\{ 0, \frac{\mathbb{E}[E_1^2]}{2} + \frac{3P^2}{2(1 - 2LP_t)^{5/2}} - P^2 + \frac{tLP}{2} \left( \mathbb{E}[E_1^2] - \frac{3P^2}{(1 - 2LP_t)^{5/2}} \right) + \frac{3t^2L^2P^2\mathbb{E}[E_1^2]}{2(1 - 2LP_t)^{5/2}} \right\}.
\]

The following corollary adapts Lemma 7 to the block energy arrival setting. Since the proof of the corollary is tedious, it is deferred to Appendix F.

**Corollary 8:** Fix a natural number \( L \). Suppose \( \{X_k\}_{k=1}^n \) is a sequence of i.i.d. random variables where

\[
X_1 \sim \mathcal{N}(x_1; 0, P),
\]

and suppose \( \{E_k\}_{k=1}^{m+n} \) is a sequence of random variables that are distributed according to (13) (in an i.i.d.-block manner with block size \( L \)). Fix an \( \varepsilon_1 > 0 \) and define

\[
t_n \triangleq \sqrt{\log(1/\varepsilon_1)} \frac{\beta_0}{\beta_{t_n}}.
\]

If

\[
n > \frac{L^3 \log(1/\varepsilon_1)}{\beta_0} \max \left\{ 4P^2, \frac{\mathbb{E}[E_1^2]^2}{4P^2} \right\},
\]

and

\[
m \geq \sqrt{(n/L + 1) \log(1/\varepsilon_1) L(\beta_{t_n} + \beta_0)} + 2L,
\]

then

\[\Pr_{p_{X^n}p_{E_{m+n}}} \left\{ \bigcup_{k=1}^n \left\{ \sum_{i=1}^k X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right\} \leq \varepsilon_1.\]

The following lemma [17] is standard for proving achievability results in the finite blocklength regime and its proof can be found in [4, Th. 3.8.1].

**Lemma 9 (Implied by Shannon’s bound [17]):** Let \( p_{X^n,Y^n} \) be the probability distribution of a pair of random variables \( (X^n, Y^n) \). Let \( \{X^n(w), Y^n(w)\}_{w=1}^\infty \) be a sequence of i.i.d. random variables where \( (X^n(1), Y^n(1)) \sim p_{X^n,Y^n} \). For each \( \delta > 0 \) and each \( M \in \mathbb{N} \), we have

\[
\Pr \left\{ \bigcup_{w=2}^M \left\{ \log \left( \frac{p_{Y^n|X^n}(Y^n(1)|X^n(w))}{p_{Y^n}(Y^n(1))} \right) > \log M + n\delta \right\} \right\} \leq e^{-n\delta}.
\]

**B. Proof of Lemma 4**

Fix any \( \varepsilon \in (0, 1) \), and fix any \( \varepsilon_1 > 0 \) and \( \varepsilon_2 > 0 \) such that

\[
\varepsilon = \varepsilon_1 + \varepsilon_2.
\]
Define \( \alpha_t, \beta_0, \beta_t \) and \( t_n \) as in (63), (64), (65) and (66) respectively. Fix a sufficiently large \( n \) such that (48), (49) and (50) hold. Since (48) holds, it follows from the definition of \( \beta_0 \) in (64) that (67) also holds. Define

\[
m \triangleq \left[ \frac{\sqrt{(n/L)} + 1)}{\alpha_t} \log(1/\epsilon_1) L (\beta_{t_n} + \beta_0) + 2L \right],
\]

which satisfies (68) and specifies the number of time slots which are used for saving energy. Consider the random code that uses the channel \( m + n \) times as follows:

**Save-and-Transmit Random Codebook Construction**

Let \( 0^m \) denote the length-\( m \) zero tuple. Define the distribution \( p_X \) as

\[
p_X(x) \triangleq \mathcal{N}(x; 0, P).
\]

In addition, define the distribution \( p_{X^n} \) as \( p_{X^n}(x^n) \triangleq \prod_{k=1}^n p_X(x_k) \). Construct \( M \) i.i.d. random tuples denoted by \( X^n(1), X^n(2), \ldots, X^n(M) \) such that \( X^n(1) \) is distributed according to \( p_{X^n} \), where \( M \) will be carefully chosen later when we evaluate the probability of decoding error. Define

\[
\hat{X}^{m+n}(w) \triangleq (0^m, X^n(w))
\]

for each \( w \in \{1, 2, \ldots, M\} \) and construct the random codebook

\[
\{ \hat{X}^{m+n}(w) \mid w \in \{1, 2, \ldots, M\} \}.
\]

The codebook is revealed to both the encoder and the decoder. To facilitate discussion, we let \( X_k(w) \) and \( \hat{X}_k(w) \) denote the \( k \)th symbols in \( X^n(w) \) and \( \hat{X}^{m+n}(w) \) respectively for each \( i \). Since the first \( m \) symbols of each random codeword \( \hat{X}^{m+n}(w) \) are zeros by (74), the source will just transmit 0 with probability 1 until time slot \( m + 1 \) when the amount of energy \( \sum_{k=1}^{m+1} E_k \) is available for encoding \( \hat{X}_{m+1}(W) \).

**Encoding under the EH Constraints**

For each \( w \in \{1, 2, \ldots, M\} \), recalling that \( \hat{X}_k(w) \) is the \( k \)th element of \( \hat{X}^{m+n}(w) \) \( \triangleq (0^m, X^n(w)) \), we construct recursively for \( k = 1, 2, \ldots, m + n \) the random variable

\[
\hat{X}_k(w, E^k) \triangleq \begin{cases} 
\hat{X}_k(w) & \text{if } (\hat{X}_k(w))^2 \leq \sum_{i=1}^k E_i - \sum_{i=1}^{k-1} (\hat{X}_i(w, E^i))^2, \\
0 & \text{otherwise}.
\end{cases}
\]

To send message \( W \) which is uniformly distributed on \( \{1, 2, \ldots, M\} \), the source transmits \( \hat{X}_k(W, E^k) \) in time slot \( k \) for each \( k \in \{1, 2, \ldots, m + n\} \). Note that the source transmits 0 with probability 1 in the first \( m \) times slots by (74) and (76), and the transmitted codeword \( (\hat{X}_1(W, E^1), \hat{X}_2(W, E^2), \ldots, \hat{X}_{m+n}(W, E^{m+n})) \) satisfies the EH constraints (15) by (76).

**Threshold Decoding**

Upon receiving

\[
\hat{Y}^{m+n} = \hat{X}^{m+n}(W, E^{m+n}) + Z^{m+n}
\]


where
\[ \hat{X}^{m+n}(W, E^{m+n}) \triangleq (\hat{X}_1(W, E^1), \hat{X}_2(W, E^2), \ldots, \hat{X}_{m+n}(W, E^{m+n})) \] (78)
denotes the transmitted tuple specified in (76) and \( Z^{m+n} \sim \prod_{k=1}^{m+n} \mathcal{N}(z_k; 0, 1) \) by the channel law (cf. (17)), the destination constructs its subtuple denoted by \( \bar{Y}^n \) by keeping only the last \( n \) symbols of \( \hat{Y}^{m+n} \). Recalling that \( q_{Y|X} \) denotes the channel law and \( p_X(x) \equiv N(x; 0, P) \), we define the joint distribution
\[ p_{X,Y} \triangleq p_X q_{Y|X}, \] (79)
and define the joint distribution \( p_{X^n,Y^n} \) as
\[ p_{X^n,Y^n}(x^n, y^n) \triangleq \prod_{k=1}^{n} p_{X,Y}(x_k, y_k). \] (80)

Then, the decoder declares \( \varphi(\bar{Y}^n) \in \{1, 2, \ldots, M\} \) (with a slight abuse of notation, we write \( \varphi(\bar{Y}^n) \) instead of \( \varphi(\hat{Y}^{m+n}) \)) to be the transmitted message where \( \varphi(\bar{Y}^n) \) is the decoding function defined as follows: If there exists a unique index \( j \) such that
\[ \log \left( \frac{p_{Y^n|X^n}(\bar{Y}^n|X^n(j))}{p_{Y^n}(\bar{Y}^n)} \right) > \log M + \frac{1}{2} \log n, \] (81)
then \( \varphi(\bar{Y}^n) \) is assigned the value \( j \). Otherwise, \( \varphi(\bar{Y}^n) \) is assigned a random value uniformly distributed on \( \{1, 2, \ldots, M\} \).

Calculating the Probability of Violating the EH Constraints
Defining \( \bar{X}^n(W, E^{m+n}) \) to be the tuple containing the last \( n \) symbols of \( \hat{X}^{m+n}(W, E^{m+n}) \), we obtain from (74), (76) and (78) that
\[ \Pr \left\{ \bar{X}^n(W, E^{m+n}) = X^n(W) \ \bigg| \ \bigcap_{k=1}^{n} \left\{ \sum_{i=1}^{k} (X_i(W))^2 \leq \sum_{i=1}^{m+k} E_i \right\} \right\} = 1. \] (82)
Using Corollary 8 and noting that \( E^{m+n} \) and \( (W, X^n(W)) \) are independent by construction, we obtain
\[ \Pr \left\{ \bigcup_{k=1}^{n} \left\{ \sum_{i=1}^{k} (X_i(W))^2 > \sum_{i=1}^{m+k} E_i \right\} \right\} \leq \varepsilon_1. \] (83)
Using (82) and (83), we have
\[ \Pr \left\{ \bar{X}^n(W, E^{m+n}) = X^n(W) \right\} \geq 1 - \varepsilon_1. \] (84)

Calculating the Probability of Decoding Error
Defining \( \bar{Z}^n \) to be the tuple containing the last \( n \) symbols of \( Z^{m+n} \) and recalling \( \bar{X}^n(W, E^{m+n}) \) and \( \bar{Y}^n \) are the tuples containing the last \( n \) symbols of \( \hat{X}^{m+n}(W, E^{m+n}) \) and \( \hat{Y}^{m+n} \) respectively, we obtain from (77) and (84) that
\[ \Pr \left\{ \bar{Y}^n = X^n(W) + \bar{Z}^n \right\} \geq 1 - \varepsilon_1. \] (85)
where $X^n(W)$ and $\bar{Z}^n$ are independent and $\bar{Z}^n \sim \prod_{k=1}^n \mathcal{N}(\bar{z}_k; 0, 1)$. Following (81) and (85), we define the events

$$E_{j|w} \triangleq \left\{ \log \left( \frac{p_{Y^n|X^n}(X^n(w) + \bar{Z}^n)}{p_{Y^n}(X^n(w) + \bar{Z}^n)} \right) \leq \log M + \frac{1}{2} \log n \right\}$$

and consider the following chain of inequalities for each $w \in \{1, 2, \ldots, M\}$:

$$\Pr_{P_W(\prod_{k=1}^n P_{X^n(k)})} \left\{ E_{w|w} \cup \bigcup_{j \in \{1, 2, \ldots, M\} \setminus \{w\}} E_{j|w} \bigg| W = w \right\}$$

$$\leq \Pr_{P_W(\prod_{k=1}^n P_{X^n(k)})} \left\{ E_{1|1} \bigg| W = 1 \right\} + \Pr_{P_W(\prod_{k=1}^n P_{X^n(k)})} \left\{ \bigcup_{j=2}^M E_{j|1} \bigg| W = 1 \right\}$$

(a) follows from symmetry of the random codebook construction and the union bound.

(b) follows from Lemma 9 and (86).

(c) follows from the fact that $X^n(1)$ and $\bar{Z}^n$ are independent copies of $X_1(1)$ and $\bar{Z}_1$ respectively by construction.

**Applying the Berry-Esséen Theorem**

Using (79), (73) and (17) we conclude that $X \sim \mathcal{N}(x; 0, P)$ and $Z \triangleq Y - X$ are independent, $Z \sim \mathcal{N}(z; 0, 1)$, and

$$\log \left( \frac{p_{Y|X}(Y|X)}{p_Y(Y)} \right) = \frac{1}{2} \log(1 + P) + \frac{(-PZ^2 + 2XYZ + X^2) \log e}{2(1 + P)}.$$ 

In order to ensure the first term in (90) can be bounded above by a simple term, we first define the mean $\mu$, the variance $\sigma^2$ and the third absolute moment $T$ of $\log \left( \frac{p_{Y|X}(Y|X)}{p_Y(Y)} \right)$ as follows: $\mu \triangleq \frac{1}{2} \log(1 + P)$, $\sigma \triangleq \sqrt{\frac{P(\log e)^2}{1 + P}}$ and

$$T^{1/3} \triangleq \left( \frac{\log e}{1 + P} \right)^{1/3} \left( \frac{15^{1/3} P + 8\sqrt{P}}{\pi} \right)$$

where the derivation of the last inequality is relegated to Appendix G-A. Clearly,

$$\frac{T}{\sigma^3} \leq \left( \frac{15^{1/3} \sqrt{P} + \frac{8}{\sqrt{\pi}}}{(1 + P)^{3/2}} \right)^3 \leq \tau_1.$$ 

After defining $\mu$, $\sigma$ and $T$, we choose $M$ to be the unique integer that satisfies

$$\log(M + 1) \geq n\mu + \sqrt{n}\sigma^2 \Phi^{-1} \left( \varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}} \right) - \frac{1}{2} \log n > \log M$$

where

$$\varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}} \geq \varepsilon_2 - \frac{\tau_1}{\sqrt{n}} - \frac{1}{\sqrt{n}} > 0.$$
Following (90), we obtain the following inequality where the random variables are distributed according to \(\prod_{k=1}^{n} P_{X_k(1)} P_{Z_k};\)

\[
    \Pr\{\mathcal{E}_{1|1}\} \leq \Pr\left\{ \sum_{k=1}^{n} \log \left( \frac{P_{Y|X}(X_k(1) + Z_k|X_k(1))}{P_{Y}(X_k(1) + Z_k)} \right) \right\} \leq n\mu + \sqrt{n}\sigma^2 \Phi^{-1}\left( \varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}} \right) \quad (97)
\]

\[
    \leq \varepsilon_2 - \frac{1}{\sqrt{n}} \quad (98)
\]

where

(a) follows from (86) and (95).

(b) follows from the Berry-Essén theorem for i.i.d random variables [20], i.e., \(\left| \Pr\left\{ \frac{\sum_{k=1}^{n} V_k - n\mu}{\sqrt{n}\sigma} \leq a \right\} - \Phi(a) \right| \leq \frac{T}{\sigma^3 \sqrt{n}}\) for all \(a \in \mathbb{R}\) where \(\mu, \sigma^2\) and \(T\) denote the mean, the variance and the third absolute moment of \(V_k\) respectively.

We are ready to compute the probability of decoding error as follows, where the random variables are distributed according to \(p_{W,X^n(W)} p_{Z^n|W,X^n(W),Z^n};\)

\[
    \Pr\{\varphi(\tilde{Y}^n) \neq W\} \leq \Pr\{\{\varphi(\tilde{Y}^n) \neq W\} \cap \{\tilde{Y}^n = X^n(W) + \tilde{Z}^n\}\} + \varepsilon_1 \quad (99)
\]

\[
    \leq \Pr\{\varphi(X^n(W) + \tilde{Z}^n) \neq W\} + \varepsilon_1 \quad (100)
\]

\[
    \leq \varepsilon_1 + \varepsilon_2 \quad (101)
\]

\[
    \overset{(71)}{=} \varepsilon_2 \quad (102)
\]

where (a) follows from (86), (85), (90) and (98).

**Obtaining a Lower Bound on the Message Size \(M\)**

Using (95), (102) and the simple fact that \(\log(M + 1) \leq \log M + 1\), we conclude that the constructed code is an \((n + m, M, \varepsilon)\)-code that satisfies

\[
    \log M \geq \frac{n}{2} \log(1 + P) + \sqrt{\frac{nP \log e^2}{1 + P}} \Phi^{-1}\left( \varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}} \right) - \frac{1}{2} \log n - 1. \quad (103)
\]

Using Taylor’s theorem together with the fact by (50) that \(\left[ \varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}}, \varepsilon_2 \right] \subseteq [\varepsilon_2^2, \varepsilon_2]\), we obtain

\[
    \Phi^{-1}\left( \varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}} \right) \geq \Phi^{-1}(\varepsilon_2) - \frac{\left( \frac{T}{\sigma^3 \sqrt{n}} + \frac{1}{\sqrt{n}} \right)}{\sqrt{n} \Phi^{-1}(\varepsilon_2)} \quad (104)
\]

\[
    \overset{(94)}{=} \Phi^{-1}(\varepsilon_2) - \frac{\tau_1 + 1}{\sqrt{n} \Phi^{-1}(\varepsilon_2)} \quad (105)
\]

where the derivation of (104) is relegated to Appendix G-B. Combining (103) and (105) and recalling the definition of \(\kappa_1\) in (46), we have

\[
    \log M \geq \frac{n}{2} \log(1 + P) + \sqrt{\frac{nP \log e^2}{1 + P}} \Phi^{-1}(\varepsilon_2) - \frac{1}{2} \log n - \kappa_1. \quad (106)
\]

**Obtaining an Upper Bound on the Length of Saving Phase \(m\)**

Since the constructed \((n + m, M, \varepsilon)\)-code satisfies (45) by (106), it remains to show that \(m\) satisfies (51). To this
end, recall the definition of \( m \) in (72) and consider the following bounds on \( t_n, \alpha_{t_n} \) and \( \beta_{t_n} \):

\[
\begin{align*}
t_n &\leq \sqrt{\frac{\log(1/\varepsilon_1)}{(n/L)\beta_0}} \\
&\quad \sqrt{\frac{2\log(1/\varepsilon_1)}{nL(E[t_1^2] + P^2)}} \tag{64} \\
&\leq \frac{1}{P} \sqrt{\frac{\log(1/\varepsilon_1)}{L_n}} \tag{12} \\
&\leq \frac{P}{LE[E_1^2]} \tag{49}
\end{align*}
\]

where

\[
\alpha_{t_n} \overset{(63)}{=} L \left( P - \frac{t_n L E[E_1^2]}{2} \right) \tag{111}
\]

\[
\geq \overset{(109)}{\left( P - \frac{E[E_1^2]}{2P} \sqrt{\frac{L \log(1/\varepsilon_1)}{n}} \right)} \tag{112}
\]

and

\[
\beta_{t_n} \overset{(65)}{=} L^2 \max \left\{ 0, \frac{E[E_1^2]}{2} + \frac{3P^2}{2(1-2LPt_n)^{5/2}} - P^2 + \frac{t_nLP}{2} \left( E[E_1^2] - \frac{3P^2}{2} \right) \right\} \tag{113}
\]

\[
\leq L^2 \left( \frac{E[E_1^2]}{2} + \frac{3P^2}{2(1-2LPt_n)^{5/2}} - P^2 + \frac{t_nLP}{2} E[E_1^2] \right) \tag{114}
\]

\[
\leq L^2 \left( \frac{E[E_1^2]}{2} + \frac{3P^2}{2} \left( 1 - \sqrt{\frac{L \log(1/\varepsilon_1)}{n}} \right)^{5/2} - P^2 + \frac{E[E_1^2]}{2} \sqrt{\frac{L \log(1/\varepsilon_1)}{n}} \right). \tag{115}
\]

In order to obtain an upper bound on \( m \), consider the following chain of inequalities:

\[
m \overset{(72)}{\leq} \frac{\sqrt{(n/L + 1) \log(1/\varepsilon_1)L(\beta_{t_n} + \beta_0)}}{\alpha_{t_n} \sqrt{\beta_0}} + 2L + 1 \tag{116}
\]

\[
\overset{(a)}{\leq} \frac{L^3 \left( E[E_1^2] + \frac{3P^2}{2} \left( 1 - \sqrt{\frac{L \log(1/\varepsilon_1)}{n}} \right)^{5/2} - P^2 \right) + \frac{E[E_1^2]}{2} \sqrt{\frac{L \log(1/\varepsilon_1)}{n}}}{\alpha_{t_n} \sqrt{\beta_0}} + 2L + 1 \tag{117}
\]

\[
\overset{(b)}{\leq} \frac{\left( \sqrt{E[E_1^2]} + \frac{3P^2}{2} \left( 1 - \sqrt{\frac{L \log(1/\varepsilon_1)}{n}} \right)^{5/2} - P^2 + \frac{E[E_1^2]}{2} \sqrt{\frac{L \log(1/\varepsilon_1)}{n}} \right) \sqrt{(Ln + L^2) \log(1/\varepsilon_1)}}{\alpha_{t_n} \beta_0} + 2L + 1 \tag{118}
\]

where

(a) follows from (115) and the definition of \( \beta_0 \) in (64).

(b) follows from (112) and the definition of \( \beta_0 \) in (64).

Consequently, the constructed \((n + m, M, \varepsilon)\)-code satisfies (45) and (51) by (106) and (118) respectively. This completes the proof.
VI. Proof of Lemma 6 via Binary Hypothesis Testing

A. Prerequisites

The following definition concerning the non-asymptotic fundamental limits of a simple binary hypothesis test is standard. See for example [21, Section III-E].

Definition 8: Let $p_X$ and $q_X$ be two probability distributions defined on some common alphabet $\mathcal{X}$. Let

$$Q(\{0, 1\}|\mathcal{X}) \triangleq \{ r_{Z|\mathcal{X}} : Z \text{ and } X \text{ assume values in } \{0, 1\} \text{ and } \mathcal{X} \text{ respectively} \}$$

be the set of randomized binary hypothesis tests between $p_X$ and $q_X$ where $\{ Z = 0 \}$ indicates the test chooses $q_X$, and let $\delta \in [0, 1]$ be a real number. The minimum type-II error in a simple binary hypothesis test between $p_X$ and $q_X$ with type-I error no larger than $1 - \delta$ is defined as

$$\beta_\delta(p_X \| q_X) \triangleq \inf_{r_{Z|\mathcal{X}} \in Q(\{0, 1\}|\mathcal{X})} \int_{\mathcal{X}} r_{Z|\mathcal{X}}(1|x) q_X(x) \, dx.$$ (120)

The existence of a minimizing test $r_{Z|\mathcal{X}}$ is guaranteed by the Neyman-Pearson lemma.

We state in the following lemma and proposition some important properties of $\beta_\delta(p_X \| q_X)$, which are crucial for the proof of Theorem 1. The proof of the two statements in the following lemma can be found in [22, Lemma 1] and [23, Sec. 2.3] respectively.

Lemma 10: Let $p_X$ and $q_X$ be two probability distributions defined on some $\mathcal{X}$, and let $g$ be a function whose domain contains $\mathcal{X}$. Then, the following two statements hold:

1. (Data processing inequality (DPI)) $\beta_\delta(p_X \| q_X) \leq \beta_\delta(p_g(X) \| q_g(X))$.
2. For all $\xi > 0$, $\beta_\delta(p_X \| q_X) \geq \frac{\delta}{\int_{\mathcal{X}} p_X(x) \{ p_X(x) \geq \frac{\xi}{q_X(x)} \} \, dx}$.

The proof of the following proposition is similar to Lemma 3 in [22] and therefore omitted.

Proposition 11: Let $p_{U,V}$ and $s_V$ be two probability distributions defined on $\mathcal{W} \times \mathcal{W}$ and $\mathcal{W}$ respectively for some $\mathcal{W}$, and let $p_U$ be the marginal distributions of $p_{U,V}$. Suppose $p_U$ is the uniform distribution, and let $\alpha = \Pr\{ U \neq V \}$ (121)

be a real number in $[0, 1)$. Then,

$$\beta_{1-\alpha}(p_{U,V} \| p_U s_V) \leq \frac{1}{|\mathcal{W}|}.$$ (122)

B. Proof of Lemma 6

Fix an $\varepsilon \in (0, 1)$, an $n \in \mathbb{N}$ which is larger than the RHS of (56) and an $(n, M, \varepsilon)$-code for the AWGN EH channel. Using Definition 1, we have

$$\Pr \left\{ \sum_{k=1}^{n} X_k^2 \leq \sum_{k=1}^{n} E_k \right\} = 1$$ (123)
for the \((\bar{n}, M, \varepsilon)\)-code. Define

\[
\Delta \triangleq \begin{cases} 
L & \text{if } \bar{n} \text{ is a multiple of } L, \\
L \left\lceil \frac{\bar{n}}{L} \right\rceil - \bar{n} & \text{otherwise}
\end{cases}
\]  

(124)

to be the smallest positive integer such that \(\bar{n} + \Delta\) is a multiple of \(L\). Then, we can always construct an \((\bar{n} + \Delta, M, \varepsilon)\)-code by appending carefully chosen \(X_{\bar{n}+1}, X_{\bar{n}+2}, \ldots, X_{\bar{n}+\Delta}\) to each transmitted sequence \(X^{\bar{n}}\) generated by the \((\bar{n}, M, \varepsilon)\)-code such that

\[
\Pr\left\{ \sum_{k=1}^{\bar{n}+\Delta} X_k^2 = \sum_{k=1}^{\bar{n}+\Delta} E_k \right\} = 1.
\]

(125)

The technique of transforming the peak power inequality constraint (123) to a power equality constraint (125) by appending an extra symbol has been employed in [21, Lemma 39] and [24, Theorem 4.4] (and is called the Yaglom-map trick). To simplify notation, we let

\[
n \triangleq \bar{n} + \Delta
\]

(126)

where \(n\) is a multiple of \(L\) and satisfies (56).

**Obtaining a Lower Bound on the Error Probability in Terms of the Type-II Error of a Hypothesis Test**

Let \(p_{W^n, X^n, Y^n, \hat{W}}\) be the probability distribution induced by the \((n, M, \varepsilon)\)-code constructed above, where \(p_{W^n, X^n, Y^n, \hat{W}}\) can be expressed according to (18). In view of (125), we assume without loss of generality that

\[
\int_{\mathcal{A}} p_{W^n, X^n, Y^n, \hat{W}}(w, e^n, x^n, y^n) = \int_{\mathcal{A}} p_{W^n, X^n, Y^n}(w, e^n, x^n, y^n) \mathbf{1}\left\{ \sum_{k=1}^{n} x_k^2 = \sum_{k=1}^{n} e_k \right\}
\]

(127)

for all Borel measurable \(\mathcal{A} \subseteq \mathcal{W} \times \mathbb{R}_+^n \times \mathbb{R}^n \times \mathbb{R}^n\). All the probability and expectation terms in the rest of this proof are evaluated according to \(p_{W^n, X^n, Y^n, \hat{W}}\) unless specified otherwise. Define

\[
s_{Y^n, \hat{W}} \triangleq \left( \prod_{k=1}^{n} s_{Y_k} \right) p_{\hat{W}|Y^n}
\]

(128)

where

\[
s_{Y_k}(y_k) \triangleq \mathcal{N}(y_k; 0, P + 1).
\]

(129)

It follows from Proposition 11 and Definition 1 with the identifications \(U \equiv W, V \equiv \hat{W}, p_{U,V} \equiv p_{W,\hat{W}}, |W| \equiv M\) and \(\alpha \equiv \Pr\{W \neq \hat{W}\} \leq \varepsilon\) that

\[
\beta_{1-\varepsilon}(p_{W,\hat{W}} \| p_W s_{\hat{W}}) \leq \beta_{1-\alpha}(p_{W,\hat{W}} \| p_W s_{\hat{W}}) \leq \frac{1}{M}.
\]

(130)
Using the DPI to Introduce the Channel Inputs and Outputs

Using the DPI of $\beta_1 - \varepsilon$ in Lemma 10, we have

$$\beta_1 - \varepsilon(p_{W,W} \mid p_{W,W}) \geq \beta_1 - \varepsilon \left( p_{W,E^n,X^n,Y^n,W} \left\| p_{W,E^n} \left( \prod_{k=1}^{n} p_{X_k|W,E^n} \right) s_{Y^n,W} \right) \right).$$  \hfill (131)

Fix a $\xi_n > 0$ to be specified later. Since

$$p_{W,E^n,X^n,Y^n,W} = p_{W,E^n} \left( \prod_{k=1}^{n} p_{X_k|W,E^n} p_{Y_k|X_k} \right) p_{W|Y^n},$$  \hfill (132)

it follows from (130), the definition of $s_{Y^n,W}$ in (128), (131) and Lemma 10 that

$$\log M \leq \log \xi_n - \log \left( 1 - \varepsilon - \Pr \left\{ \sum_{k=1}^{n} \log \left( \frac{p_{Y_k|X_k|Y_k}}{s_{Y_k}(Y_k)} \right) \geq \log \xi_n \right\} \right).$$  \hfill (133)

**Simplifying the Non-Asymptotic Bound**

Combining (17) and (129), we have

$$\log \frac{p_{Y_k|X_k}(Y_k|X_k)}{s_{Y_k}(Y_k)} = \frac{1}{2} \log(1 + P) + \frac{\log e}{2(1 + P)} \left( -P(Y_k - X_k)^2 + X_k^2 + 2X_k(Y_k - X_k) \right)$$  \hfill (134)

for each $k \in \{1, 2, \ldots, n\}$. Due to the power equality constraint imposed on the codewords, we have

$$\Pr \left\{ \sum_{k=1}^{n} X_k^2 = \sum_{k=1}^{n} E_k \right\} \overset{(125)}{=} 1.$$  \hfill (135)

Letting

$$U_k \triangleq \log e \frac{-P(Y_k - X_k)^2 + 2X_k(Y_k - X_k) + E_k}{2(1 + P)}$$  \hfill (136)

for each $k \in \{1, 2, \ldots, n\}$, we obtain from (134) and (135) that

$$\Pr \left\{ \sum_{k=1}^{n} \log \frac{p_{Y_k|X_k}(Y_k|X_k)}{s_{Y_k}(Y_k)} = \frac{n}{2} \log(1 + P) + \sum_{k=1}^{n} U_k \right\} = 1.$$  \hfill (137)

Combining (133) and (137), we have

$$\log M \leq \log \xi_n - \log \left( \Pr \left\{ \sum_{k=1}^{n} U_k < \log \xi_n - \frac{n}{2} \log(1 + P) \right\} - \varepsilon \right).$$  \hfill (138)

**Evaluating the Distribution of the Sum of Random Variables $\sum_{k=1}^{n} U_k$**

In order to simplify (138), we now investigate the distribution of the sum of random variables $\sum_{k=1}^{n} U_k$. We will show in the following that the distribution of $\sum_{k=1}^{n} U_k$ can be evaluated in closed form. Define the function $\lambda : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$

$$\lambda(e, x, y) = -P(y - x)^2 + 2x(y - x) + e.$$  \hfill (139)
We begin evaluating the distribution of $\sum_{k=1}^{n} U_k$ by examining the distribution of $\sum_{k=1}^{n} \lambda(E_k, X_k, Y_k)$ (cf. (136)) as follows. Let

$$E \left[ e^{it \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k)} \right]$$

(140)

be the characteristic function of $\sum_{k=1}^{n} \lambda(E_k, X_k, Y_k)$ where $i$ denotes the imaginary unit. In order to evaluate a closed-form expression for (140), we write

$$E \left[ e^{it \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k)} \right]$$

(127) $$E \left[ \exp \left( \frac{it}{2} \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k) - \frac{n^2}{2} \sum_{k=1}^{n} (E_k - X_k^2) \right) \right]$$

(141)

$$= E \left[ \left( E \left[ e^{i \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k)} \right] \right)^n \right]$$

(142)

In order to simplify the RHS of (142), we consider the following chain of equalities for each $r \in \{2, 3, \ldots, n\}$:

$$E \left[ e^{i \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k) - \frac{2i^2}{2 + 2itP} \sum_{k=1}^{n} (E_k - X_k^2)} \right]$$

$$= E \left[ e^{i \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k) - \frac{2i^2}{2 + 2itP} \sum_{k=1}^{n} (E_k - X_k^2)} \right]$$

$$E \left[ E \left[ e^{i \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k) - \frac{2i^2}{2 + 2itP} \sum_{k=1}^{n} (E_k - X_k^2)} \right] E^n \right]$$

$$= E \left[ e^{i \lambda(E_r, X_r, Y_r) - \frac{2i^2}{2 + 2itP} (E_r - X_r^2)} \right] E^n \left. \right|_{E^n, X_r^{r-1}, Y_r^{r-1}}$$

(143)

$$E \left[ e^{i \lambda(E_r, X_r, Y_r) - \frac{2i^2}{2 + 2itP} (E_r - X_r^2)} \right] E^n \left. \right|_{E^n, X_r^{r-1}, Y_r^{r-1}}$$

(144)

Since $Y_r - X_r$ is a standard normal random variable which is independent of $(E^n, X_r^{r-1}, Y_r^{r-1})$ by the channel law in (17), it follows from straightforward calculations based on (139) that the following equality holds with probability 1 for each $i \in \{2, 3, \ldots, n\}$:

$$E \left[ e^{i \lambda(E_r, X_r, Y_r) - \frac{2i^2}{2 + 2itP} (E_r - X_r^2)} \right] E^n \left. \right|_{E^n, X_r^{r-1}, Y_r^{r-1}}$$

$$= \frac{1}{\sqrt{1 + 2itP}} E \left[ e^{it E_r - \frac{2i^2 E_r}{2 + 2itP}} \right] E^n \left. \right|_{E^n, X_r^{r-1}, Y_r^{r-1}}$$

(145)

$$= \frac{1}{\sqrt{1 + 2itP}} e^{it E_r - \frac{2i^2 E_r}{2 + 2itP}}$$

(146)

Using (144) and (146), we have the following equality that holds with probability 1 for each $r \in \{2, 3, \ldots, n\}$:

$$E \left[ e^{i \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k) - \frac{2i^2}{2 + 2itP} \sum_{k=1}^{n} (E_k - X_k^2)} \right] E^n$$

$$= \frac{1}{\sqrt{1 + 2itP}} e^{it E_r - \frac{2i^2 E_r}{2 + 2itP}} \cdot E \left[ e^{i \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k) - \frac{2i^2}{2 + 2itP} \sum_{k=1}^{n} (E_k - X_k^2)} \right] E^n$$

(147)

Combining (142) and (147), we obtain

$$E \left[ e^{i \sum_{k=1}^{n} \lambda(E_k, X_k, Y_k)} \right] = (1 + 2itP)^{-\frac{1}{2}} E \left[ e^{i \sum_{k=1}^{n} E_k - \frac{2i^2 E_k}{2 + 2itP}} \right].$$

(148)
Let \( \{ Z_k \}_{k=1}^n \) be \( n \) independent copies of the standard normal random variable. Straightforward calculations reveal that
\[
\mathbb{E}_{p,E^n} \prod_{k=1}^n p_{Z_k} \left[ e^{it \sum_{k=1}^n (-PZ_k^2 + 2\sqrt{E_k}Z_k + E_k)} \right] = (1 + 2itP)^{-\frac{n}{2}} \mathbb{E} \left[ e^{\sum_{k=1}^n it E_k - \frac{2it^2 E_k}{1 + 2itP}} \right].
\] (149)

Therefore,
\[
\mathbb{E}_{p,W,E^n,X^n,Y^n} \left[ e^{it \sum_{k=1}^n \lambda(E_k, X_k, Y_k)} \right] = \mathbb{E}_{p,E^n} \prod_{k=1}^n p_{Z_k} \left[ e^{it \sum_{k=1}^n (-PZ_k^2 + 2\sqrt{E_k}Z_k + E_k)} \right]
\] (150)

by (148) and (149), i.e., the characteristic functions of \( \sum_{k=1}^n \lambda(E_k, X_k, Y_k) \) and \( \sum_{k=1}^n (-PZ_k^2 + 2\sqrt{E_k}Z_k + E_k) \) are equal. Consequently, the probability distributions of \( \sum_{k=1}^n \lambda(E_k, X_k, Y_k) \) and \( \sum_{k=1}^n (-PZ_k^2 + 2\sqrt{E_k}Z_k + E_k) \) are equal almost everywhere, which implies from (136) and (139) that the probability distributions of \( \sum_{k=1}^n \log \xi_\ell - \frac{n}{2} \log(1 + P) \) and \( \sum_{k=1}^n U_k \) are equal almost everywhere, which then implies from (138) that
\[
\log M \leq \log \xi_n - \log \left( \Pr_{p,E^n} \prod_{k=1}^n p_{Z_k} \left\{ \sum_{k=1}^n \frac{\log e}{2(1 + P)} (-PZ_k^2 + 2\sqrt{E_k}Z_k + E_k) < \log \xi_n - \frac{n}{2} \log(1 + P) \right\} - \varepsilon \right)
\] (151)

We recall from (124) that \( n = \bar{n} + \Delta \) is a multiple of \( L \) and define
\[
\tilde{V}_\ell = \frac{\log e}{2(1 + P)} \sum_{i=1}^L (-PZ_{bi+i}^2 + 2\sqrt{E_{bi+i}}Z_{bi+i} + E_{bi+i})
\] (152)

for each \( \ell \in \{1, 2, \ldots, n/L \} \) (cf. the definition of \( b_\ell \) in (2)). Then, equation (151) can be rewritten as
\[
\log M \leq \log \xi_n - \log \left( \Pr_{p,E^n} \prod_{k=1}^n p_{Z_k} \left\{ \sum_{\ell=1}^{n/L} \tilde{V}_\ell < \log \xi_n - \frac{n}{2} \log(1 + P) \right\} - \varepsilon \right).
\] (153)

**Applying the Berry-Esséen Theorem**

Using the facts that \( \{ Z_k \}_{k=1}^n \) are i.i.d., \( \{ E_{bi+1} \}_{\ell=1}^{n/L} \) are i.i.d. and
\[
E_{bi+1} = E_{bi+2} = \ldots = E_{bi+L}
\] (154)

for each \( \ell \in \{1, 2, \ldots, n/L \} \), we conclude in view of (152) that \( \{ \tilde{V}_\ell \}_{\ell=1}^{n/L} \) are i.i.d. where
\[
\tilde{V}_1 = \frac{\log e}{2(1 + P)} \left( -P \sum_{i=1}^L Z_i^2 + 2\sqrt{E_1} \sum_{i=1}^L Z_i + LE_1 \right).
\] (155)

In order to invoke the Berry-Esséen Theorem to bound the probability term in (153), we define the following quantities related to \( \tilde{V}_1 \):
\[
\mu \triangleq \mathbb{E}_{p,E^1} \prod_{i=1}^{L} p_{Z_i} [\tilde{V}_1]
\] (156)
\[
(\text{156}) \iff 0,
\] (157)
\[ \sigma \triangleq \sqrt{\text{Var}_{p_{E_1}} \prod_{i=1}^{L} p_{Z_i} [\tilde{V}_i]} \]

\[ \left( L \log e \frac{2P(P+2)}{L} + \mathbb{E}[E_1^2] - P^2 \right) \]

(158)

and

\[ T^{1/3} \triangleq \left( \mathbb{E}_{p_{E_1}} \prod_{i=1}^{L} p_{Z_i} [\mathbb{E}_{1}^{3/2}] \right)^{1/3} \]

\[ \leq L \log e \frac{15^{1/3} P + 2(2\sqrt{2/\pi})^{1/3} \cdot \left( \mathbb{E}[E_1^{3/2}] \right)^{1/3} + \left( \mathbb{E}[E_1^{3}] \right)^{1/3} }{2(1 + P)} \]

(159)

(160)

where the derivation of the last inequality is relegated to Appendix H-A. Recalling the definition of \( \tau_2 \) in (55), we use the Berry-Esséen theorem for i.i.d. random variables [20] to obtain

\[ \Pr_{p_{E_1}} \prod_{i=1}^{n} p_{E_i} \left\{ \frac{1}{\sigma \sqrt{L}} \sum_{t=1}^{n} \log e \frac{2}{2(1 + P)} \left( -P Z_{b_t+i} + 2 \sqrt{E_{b_t+i} Z_{b_t+i} + E_{b_t+i}} \right) \leq \Phi^{-1} \left( \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right) \right\} \]

\[ \geq \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} - \frac{T}{\sigma^3 \sqrt{L}} \]

(161)

\[ \geq \varepsilon + \tau_2 \sqrt{\frac{L}{n}} \]

(162)

where the argument of \( \Phi^{-1} \) satisfies

\[ \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \leq \varepsilon + (1 - \varepsilon)^2 < 1 \]

(163)

Following (153) and letting

\[ \xi_n \triangleq \frac{n}{2} \log(1 + P) + \sigma \sqrt{\frac{n}{L}} \Phi^{-1} \left( \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right), \]

(164)

we can express (153) as

\[ \log M \leq \frac{n}{2} \log(1 + P) + \sigma \sqrt{\frac{n}{L}} \Phi^{-1} \left( \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right) \]

\[ - \log \left( \Pr_{p_{E_1}} \prod_{i=1}^{n} p_{E_i} \left\{ \sum_{t=1}^{n/L} \tilde{V}_t < \sigma \sqrt{\frac{n}{L}} \Phi^{-1} \left( \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right) \right\} - \varepsilon \right), \]

(165)

which implies from the definition of \( \sigma \) in (158), the inequality in (162) and the definition of \( \tilde{V}_t \) in (152) that

\[ \log M \leq \frac{n}{2} \log(1 + P) + \sqrt{\frac{n}{L}} \left( 2P(P+2) + L(\mathbb{E}[E_1^2] - P^2) \right) \Phi^{-1} \left( \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right) - \log \left( \tau_2 \sqrt{\frac{L}{n}} \right) \]

(166)

Using Taylor’s theorem together with the fact by (50) that \([\varepsilon, \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}}] \subseteq [\varepsilon, \varepsilon + (1 - \varepsilon)^2]\), we obtain

\[ \Phi^{-1} \left( \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right) \leq \Phi^{-1}(\varepsilon) + \frac{2\tau_2 \sqrt{\frac{L}{n}}}{N(\Phi^{-1}(\min\{\varepsilon, \varepsilon(1-\varepsilon)\}); 0, 1)} \]

(167)
whose derivation is relegated to Appendix H-B. Combining (166) and (167) and recalling the definition of $\kappa_2$ in (54), we have

$$\log M \leq \frac{n}{2} \log(1 + P) + \frac{\sqrt{n} \log e}{2(1 + P)} \sqrt{2P(P + 2) + L(E[E^2_1] - P^2)} \Phi^{-1}(\varepsilon) + \frac{1}{2} \log n + \kappa_2. \quad (168)$$

Using (168) and the fact by (124) and (126) that $n \leq \bar{n} + L$, we have

$$\log M \leq \frac{\bar{n} + L}{2} \log(1 + P) + \frac{\sqrt{\bar{n} + L} \log e}{2(1 + P)} \sqrt{2P(P + 2) + L(E[E^2_1] - P^2)} \Phi^{-1}(\varepsilon) + \frac{1}{2} \log(\bar{n} + L) + \kappa_2. \quad (169)$$

This completes the proof.

VII. WHEN THE LENGTH OF EACH ENERGY ARRIVAL BLOCK GROWS LINEARLY IN BLOCKLENGTH

This section focuses on the scenario $L = \lfloor \lambda n \rfloor$ for some real constant $\lambda \in (0, 1]$. Define

$$\rho \triangleq \left\lfloor \frac{n}{L} \right\rfloor \quad (170)$$

to be the number of length-$L$ energy-arrival blocks. The total number of energy-arrival blocks is $\rho + 1$ where the length of each of the first $\rho$ energy-arrival blocks equals $L$ and the length of the $(\rho + 1)$th energy-arrival block equals

$$n - \rho L. \quad (171)$$

The following proposition gives us a lower bound and an upper bound on the length of the $(\rho + 1)$th energy-arrival block, which are useful for the achievability and the converse proofs of Theorem 2 respectively. The proof of Proposition 12 is straightforward and is deferred to Appendix I.

**Proposition 12:** For all sufficiently large $n \in \mathbb{N}$,

$$\rho = q = \left\lfloor \frac{1}{\lambda} \right\rfloor \quad (172)$$

and

$$nd \leq n - \rho L \leq nd + \left\lfloor \frac{1}{\lambda} \right\rfloor, \quad (173)$$

where $n - \rho L$ is the length of the $(\rho + 1)$th energy-arrival block and $q$ and $d$ were defined in (34) and (35) respectively.

A. Achievability proof of Theorem 2

In this section, we propose an adaptive save-and-transmit code which will be used to prove the achievability part of Theorem 2. The adaptive save-and-transmit code enables the source to transmit information at a rate close to $C(E_{b_{\ell+1}})$ for each block $\ell \in \{1, 2, \ldots, \rho + 1\}$. For each block $\ell$, since the destination does not know the EH random variables $E_{b_{\ell+1}}$, the source first needs to quantize $E_{b_{\ell+1}}$ and convey the quantized version to the destination before adjusting the transmission rate. To facilitate discussion, we let

$$\Delta \triangleq \frac{\sqrt{4P + 2}}{L^{1/6}} \quad (174)$$
and define the set of quantization points
\[ \Gamma \triangleq \{ 2v\Delta \mid v \text{ is a non-negative integer} \}. \] (175)

In addition, define the quantization mapping \( g^\Delta : \mathbb{R}_+ \to \Gamma \) such that \( g^\Delta(a) \) is the unique quantization point that satisfies
\[ g^\Delta(a) \leq a < g^\Delta(a) + 2\Delta. \] (176)

In order to enable communication at a rate close to \( C(g^\Delta(E_{b1+1})) \) and with error probability \( O(\frac{1}{L^{1/3}}) \) in block \( \ell \) for each \( \ell \in \{1, 2, \ldots, \rho + 1\} \), we propose to use an adaptive save-and-transmit code in each block so that node \( s \) can adapt the coding rate to the EH process.

**Definition 9:** An \((L, \Delta, \varepsilon)\)-adaptive code consists of the following:

1) A message alphabet \( U^\infty \triangleq \{0, 1\}^\infty \). The message \( U^\infty \) is a sequence of i.i.d. uniform bits.
2) An adaptive encoding function \( f : \Gamma \times U^\infty \to \mathbb{R}^L \) which depends on \( g^\Delta(E_1) \) such that
\[ X^L = f(g^\Delta(E_1), U^\infty) \] (177)
and
\[ \Pr \left\{ \sum_{i=1}^{k} X_i^2 \leq kE_1 \right\} = 1 \] (178)
for each \( k \in \{1, 2, \ldots, L\} \).
3) A decoding function \( \varphi : \mathbb{R}^L \to U^\infty \) for decoding \( W \) at node \( d \) such that the message estimate \( \hat{U}^\infty \) is produced by setting \( \hat{U}^\infty \triangleq \varphi(Y^L) \). Define the mapping \( \gamma : \mathbb{N} \times \mathbb{R}_+ \to \mathbb{R}_+ \) as
\[ \gamma(\ell, e) \triangleq \left\lfloor \left( \ell - \left\lfloor \ell^{2/3} \right\rfloor \right) C(g^\Delta(e)) - \frac{2\ell^{3/4}}{L} \right\rfloor. \] (179)

Then, the probability of decoding error adapted to \( g^\Delta(E_1) \), which is defined as
\[ \Pr\{ \hat{U}^{\gamma(L,E_1)} \neq U^{\gamma(L,E_1)} \}, \] (180)
is no larger than \( \varepsilon \).

By Definition 9, node \( s \) can use an \((L, \Delta, \varepsilon)\)-adaptive code to transmit \( 2^{\gamma(L,E_1)} \) bits to node \( d \) with small error probability in each length-\( L \) energy-arrival block. We use “adaptive” to describe the code because the number of bits that can be conveyed by the code changes with \( E_1 \). We will prove the achievability part of Theorem 2 by using an adaptive code that has the following two features for every \( \ell \in \{1, 2, \ldots, \rho\} \):

(i) Each of the first \( \left\lfloor \sqrt{L} \right\rfloor \) symbols in the \( \ell^{th} \) block sent by node \( s \) is the constant symbol \( \sqrt{g^\Delta(E_{b1+1})} \) so that with probability larger than \( 1 - \frac{1}{L^{1/3}} \), the destination can estimate \( g^\Delta(E_{b1+1}) \) correctly.

(ii) In the remaining \( L - \left\lfloor \sqrt{L} \right\rfloor \) symbols in the \( \ell^{th} \) block, node \( s \) intends to use a Gaussian codebook with average power \( g^\Delta(E_{b1+1}) \) to transmit i.i.d. uniform bits at a rate close to \( C(g^\Delta(E_{b1+1})) \) and with error probability \( \leq \frac{1}{L^{1/3}} \).
Feature (i) is based on Proposition 13 to be presented later. Feature (ii) will be established through proving the existence of an adaptive code with the desired properties in Lemma 15. The proof of the following proposition is simple and thus deferred to Appendix J.

**Proposition 13:**

$$\Pr_{p_{E_1}, p_{Z_1}} \left\{ \left| \frac{1}{\sqrt{L}} \sum_{i=1}^{\lceil \sqrt{T} \rceil} \left( \sqrt{g^A(E_1)} + Z_i \right)^2 - g^A(E_1) - 1 \right| < \Delta \right\} > 1 - \frac{1}{L^{1/6}}. \quad (181)$$

The following lemma is useful for proving the achievability part of Theorem 2. Since Lemma 14 is a direct consequence of [15, Th. 1], its proof is relegated to Appendix K.

**Lemma 14:** The following statement holds for any sufficiently large $L \in \mathbb{N}$. Fix an arbitrary $\hat{P} > 0$ and suppose $E_1 = E_2 = \ldots = E_L = \hat{P}$ holds with probability 1. Then, there exists an $(L - \lceil \sqrt{T} \rceil, M, \frac{1}{\sqrt{L}})$-code such that

$$\log M \geq \gamma(L, \hat{P}). \quad (182)$$

To facilitate discussion, we call the $(L - \lceil \sqrt{T} \rceil, M, \frac{1}{\sqrt{L}})$-code an $(L - \lceil \sqrt{T} \rceil, M, \hat{P}, \frac{1}{\sqrt{L}})$-code.

The following lemma is based on Proposition 13 and Lemma 14.

**Lemma 15:** For any sufficiently large $L \in \mathbb{N}$, there exists an $(L, \Delta, \frac{1}{L^{1/2}} + \frac{1}{\sqrt{T}})$-adaptive code.

**Proof:** We construct an $(L, \Delta, \frac{1}{L^{1/2}} + \frac{1}{\sqrt{T}})$-adaptive code in two steps as follows.

1) In each of the first $\lceil \sqrt{T} \rceil$ time slots in the length-$L$ block, node $s$ sends the constant symbol $\sqrt{g^A(E_1)}$, which is always possible because $g^A(E_1) \leq E_1$ by (176). Upon receiving $Y[\sqrt{T}]$, node $d$ produces an estimate of $g^A(E_1)$, denoted by $\hat{g}^A(E_1)$, by setting

$$\hat{g}^A(E_1) \triangleq \min \arg \min_{v \in \Gamma} \left\{ \left| \frac{1}{\sqrt{L}} \sum_{i=1}^{\lceil \sqrt{T} \rceil} Y_i^2 - v - 1 \right| \right\}. \quad (183)$$

It follows from the definition of $\Gamma$ in (175), Proposition 13 and (183) that

$$\Pr_{p_{E_1}} \{ \hat{g}^A(E_1) = g^A(E_1) \} > 1 - \frac{1}{L^{1/6}}. \quad (184)$$

2) In the remaining $L - \lceil \sqrt{T} \rceil$ time slots, node $s$ will choose an $(L - \lceil \sqrt{T} \rceil, M, \frac{1}{\sqrt{L}})$-code based on the knowledge of $E_1$ as follows: Node $s$ calculates $g^A(E_1)$ and transmits $\gamma(L, E_1)$ i.i.d. uniform bits $U^\gamma(L, E_1)$ using a predetermined $(L - \lceil \sqrt{T} \rceil, M, g^A(E_1), \frac{1}{\sqrt{T}})$-code whose existence is guaranteed by Lemma 14. The encoding strategy of $s$ is known to node $d$, which will decode the bits using the decoder of the $(L - \lceil \sqrt{T} \rceil, M, \hat{g}^A(E_1), \frac{1}{\sqrt{T}})$-code predetermined by node $s$ and output the bits estimate $\hat{U}^\gamma(L, E_1)$. By the definition of the codes,

$$\Pr \left\{ \hat{U}^\gamma(L, E_1) \neq U^\gamma(L, E_1) \mid \hat{g}^A(E_1) = g^A(E_1) \right\} \leq \frac{1}{\sqrt{L}}. \quad (185)$$
For the adaptive code described above, it follows from (183) and (185) together with the union bound that

\[
Pr \left\{ \hat{U}^\gamma(L,E_1) \neq U^\gamma(L,E_1) \right\} \leq \frac{1}{\sqrt{L}} + \frac{1}{L^{1/6}},
\]

which implies that the adaptive code is an \( (L, \Delta, \frac{1}{L^{1/6}} + \frac{1}{\sqrt{L}}) \)-adaptive code.

We are ready to prove the achievability part of Theorem 2.

**Achievability proof of Theorem 2:** Fix an \( \varepsilon \in (0, 1) \). Our goal is to prove

\[
R_{\varepsilon} \leq C_{\varepsilon}.
\]

It suffices to show that

\[
R_{\varepsilon} - 2\eta \leq C_{\varepsilon}
\]

for all \( \eta > 0 \). Fix an arbitrary \( \eta > 0 \). By the definition of \( R_{\varepsilon} \) in (36), we have

\[
Pr \left\{ \prod_{\ell=1}^{q+1} \left( \sum_{\ell=1}^{q} \lambda C(E_{\ell}) + dC(E_{q+1}) \geq \frac{R_{\varepsilon}}{2} - \eta \right) \right\} \geq 1 - \varepsilon + \delta
\]

for some \( \delta > 0 \). Let \( \chi_{\delta} > 0 \) be a sufficiently large number such that

\[
Pr \left\{ \max_{\ell \in \{1,2,\ldots,q+1\}} C(E_{\ell}) \geq \chi_{\delta} \right\} \leq \frac{\delta}{3}.
\]

We want to show that there exists a sequence of \( (n,M_n,\varepsilon) \)-codes such that

\[
\lim_{n \to \infty} \frac{1}{n} log M_n \geq R_{\varepsilon} - 2\eta,
\]

which will then imply (188). To this end, fix a sufficiently large \( n \in \mathbb{N} \) such that (172) holds, (173) holds,

\[
\sum_{\ell=1}^{q+1 \{ \ell \neq 0 \}} \left( \frac{1}{L_{\ell}^{1/6}} + \frac{1}{\sqrt{L_{\ell}}} \right) \leq \frac{\delta}{3}
\]

and

\[
n\eta \geq (q + 1) \left( L \Delta + \left[ L^{2/3} \right] (\chi_{\delta} + \Delta) + 2L^{3/4} + \chi_{\delta} + 1 \right)
\]

where \( \Delta \) is as defined in (174). The number of i.i.d. uniform bits that can be transmitted by the code is chosen to be

\[
log M_n \triangleq \lfloor nR_{\varepsilon} - 2n\eta \rfloor.
\]

In the rest of the proof, we are devoted to constructing an \( (n,M_n) \)-code followed by showing that the error probability is bounded above by \( \varepsilon \).

**Construction of an \( (n,M_n) \)-code:**

Recall that the length of each of the first \( \rho \) blocks is \( L = \lfloor \lambda n \rfloor \). To facilitate discussion, define

\[
L_{\ell} \triangleq L = \lfloor \lambda n \rfloor
\]
to be the length of the $\ell$th block for each $\ell \in \{1, 2, \ldots, \rho\}$, and define

$$L_{\rho + 1} \triangleq \lfloor dn \rfloor$$

(196)

to be a lower bound on the length of the $(\rho + 1)$th block (due to (173)). Since $\rho L + L_{\rho + 1} \leq n$ by construction, we will construct an $(n, M_n)$-code by concatenating $\rho$ blocks of length-$L$ adaptive codes and one block of length-$L_{\rho + 1}$ adaptive code as described below. The message of the $(n, M_n)$-code is a sequence of $\log M_n$ i.i.d. uniform bits denoted by $U^{\log M_n}$. Then for each block $\ell \in \{1, 2, \ldots, \rho + 1\}$, node $s$ uses an $(L_\ell, \Delta, \frac{1}{L_\ell^{2/3}} + \frac{1}{\sqrt{L_\ell}})$-adaptive code to transmit $\gamma(L, E_{b_{\ell + 1}})$ i.i.d. uniform bits. A decoding error is declared if one of the following cases occurs:

(i) The total number of transmitted bits is less than $\log M_n$, i.e., the following event occurs:

$$\mathcal{F} \triangleq \left\{ \sum_{\ell=1}^{\rho + 1} \gamma(L_\ell, E_{b_{\ell + 1}}) < \log M_n \right\}.$$  

(197)

(ii) Provided that $\mathcal{F}^c$ occurs, the bits estimates output by $d$ denoted by $\hat{U}^{\log M_n}$ are not equal to the transmitted bits, i.e., the following event occurs:

$$\left\{ \hat{U}^{\log M_n} \neq U^{\log M_n} \right\}.$$  

(198)

**Analysis of the Error Probability:**

In the rest of the proof, all the probability terms are evaluated according to the distribution induced by the $(n, M_n)$-code constructed above. Since the $(n, M_n)$-code is a concatenation of $\rho$ blocks of $(L_\ell, \Delta, \frac{1}{L_\ell^{2/3}} + \frac{1}{\sqrt{L_\ell}})$ adaptive codes and one block of $(L_{\rho + 1}, \Delta, \frac{1}{L_{\rho + 1}^{2/3}} + \frac{1}{\sqrt{L_{\rho + 1}}})$ adaptive code, it follows from Definition 9 and the union bound that

$$\Pr \left\{ \hat{U}^{\log M_n} \neq U^{\log M_n} \mid \mathcal{F}^c \right\} \leq \sum_{\ell=1}^{\rho + 1} \left( \frac{1}{L_\ell^{2/3}} + \frac{1}{\sqrt{L_\ell}} \right) \text{ if } d = 0,$$

$$\sum_{\ell=1}^{\rho + 1} \left( \frac{1}{L_\ell^{2/3}} + \frac{1}{\sqrt{L_\ell}} \right) \text{ otherwise},$$

(199)

which together with (172) and (197) implies that the error probability of the $(n, M_n)$-code is bounded above as

$$\Pr \left\{ \hat{U}^{\log M_n} \neq U^{\log M_n} \right\} \leq \sum_{\ell=1}^{q + 1} \gamma(L_\ell, E_{b_{\ell + 1}}) + \Pr \left\{ \sum_{\ell=1}^{q + 1} \gamma(L_\ell, E_{b_{\ell + 1}}) < \log M_n \right\}.$$  

(200)

In order to obtain an upper bound on the last term in (200) in terms of $C(E_{b_{\ell + 1}})$, we consider

$$\Pr \left\{ \sum_{\ell=1}^{q + 1} \gamma(L_\ell, E_{b_{\ell + 1}}) < \log M_n \right\}$$

\[(^{(179)} \leq \Pr \left\{ \sum_{\ell=1}^{q + 1} L_\ell C(g^\Delta(E_{b_{\ell + 1}})) < \log M_n + \sum_{\ell=1}^{q + 1} \left( \left[ L_\ell^{2/3} \right] C(g^\Delta(E_{b_{\ell + 1}})) + 2L_\ell^{3/4} + 1 \right) \right\} \]

\[(^{(a)} \leq \Pr \left\{ \sum_{\ell=1}^{q + 1} L_\ell (C(E_{b_{\ell + 1}}) - \Delta) < \log M_n + \sum_{\ell=1}^{q + 1} \left( \left[ L_\ell^{2/3} \right] (C(E_{b_{\ell + 1}}) + \Delta) + 2L_\ell^{3/4} + 1 \right) \right\} \]

\[= \Pr \left\{ \sum_{\ell=1}^{q + 1} L_\ell C(E_{b_{\ell + 1}}) < \log M_n + \sum_{\ell=1}^{q + 1} \left( L_\ell \Delta + \left[ L_\ell^{2/3} \right] (C(E_{b_{\ell + 1}}) + \Delta) + 2L_\ell^{3/4} + 1 \right) \right\} \]

(201)
(b) \( \leq \Pr \left\{ \sum_{\ell=1}^{q+1} L_\ell C(E_{b_\ell+1}) < \log M_n + \sum_{\ell=1}^{q+1} \left( L_\ell \Delta + \left\lceil L_\ell^{2/3} \right\rceil (\chi_\delta + \Delta) + 2 L_\ell^{3/4} + 1 \right) \right\} + \frac{\delta}{3} \) \hspace{1cm} (204)

\[ \leq \Pr \left\{ \sum_{\ell=1}^{q+1} L_\ell C(E_{b_\ell+1}) < \log M_n + (q + 1) \left( L \Delta + \left\lceil L^{2/3} \right\rceil (\chi_\delta + \Delta) + 2 L^{3/4} + 1 \right) \right\} + \frac{\delta}{3}, \] \hspace{1cm} (205)

where

(a) follows from the definition of \( g^\Delta \) in (176) and the fact that

\[ \log(1 + a + 2b) - 2b \leq \log(1 + a) \leq \log(1 + a + 2b). \] \hspace{1cm} (206)

(b) follows from (190) and the union bound.

On the other hand, combining (189) with the fact that \( \{ E_{b_\ell+1} \}_{\ell=1}^{n/L} \) are i.i.d., we have

\[ \Pr \left\{ \sum_{\ell=1}^{q} \lambda n C(E_{b_\ell+1}) + d n C(E_{b_q+1+1}) < n R_\varepsilon - n \eta \right\} \leq \varepsilon - \delta, \] \hspace{1cm} (207)

which implies that

\[ \Pr \left\{ \sum_{\ell=1}^{q} \left[ \lambda n \right] C(E_{b_\ell+1}) + \left\{ d n \right\} C(E_{b_q+1+1}) < n R_\varepsilon - n \eta - \sum_{\ell=1}^{q+1} C(E_{b_\ell+1}) \right\} \leq \varepsilon - \delta, \] \hspace{1cm} (208)

which then together with (195), (196) and (194) implies that

\[ \Pr \left\{ \sum_{\ell=1}^{q+1} L_\ell C(E_{b_\ell+1}) < \log M_n + n \eta - \sum_{\ell=1}^{q+1} C(E_{b_\ell+1}) \right\} \leq \varepsilon - \delta. \] \hspace{1cm} (209)

Using (209), (190) and the union bound, we obtain

\[ \Pr \left\{ \sum_{\ell=1}^{q+1} L_\ell C(E_{b_\ell+1}) < \log M_n + n \eta - (q + 1) \chi_\delta \right\} \leq \varepsilon - \frac{2\delta}{3}. \] \hspace{1cm} (210)

Combining (205), (210) and (193), we have

\[ \Pr \left\{ \sum_{\ell=1}^{q+1} C(E_{b_\ell+1}) < \log M_n \right\} \leq \varepsilon - \frac{\delta}{3} \] \hspace{1cm} (211)

Using (200), (192) and (211), we have

\[ \Pr \left\{ \sum_{\ell=1}^{q+1} \gamma(L_\ell, E_{b_\ell+1}) < \log M_n \right\} \leq \varepsilon. \] \hspace{1cm} (212)

Therefore, the constructed \((n, M_n)\)-code is an \((n, M_n, \varepsilon)\)-code where \( M_n \) satisfies (194). Consequently, for any \( \eta > 0 \), there exists a sequence of \((n, M_n, \varepsilon)\)-codes where \( M_n \) satisfies (194) such that (191) holds, which then implies (188). Since \( \eta > 0 \) is arbitrary, we have (187).

**B. Converse Proof of Theorem 2**

Fix an \( \varepsilon \in (0, 1) \). Our goal is to prove

\[ C_\varepsilon \leq \overline{R}_\varepsilon. \] \hspace{1cm} (213)
It suffices to show that
\[ C_\varepsilon \leq \overline{R}_\varepsilon + \eta \] (214)
for all \( \eta > 0 \). Fix an arbitrary \( \eta > 0 \) and an \( \varepsilon \)-achievable rate \( R \). By the definition of \( \overline{R}_\varepsilon \) in (37),
\[
\Pr_{\Pi_{i=1}^{q+1} p_{E_i}} \left\{ \sum_{\varepsilon=1}^{q} \lambda C(E_\varepsilon) + dC(E_{q+1}) \geq \overline{R}_\varepsilon + \eta \right\} \leq 1 - \varepsilon - 2\delta
\] (215)
for some \( \delta > 0 \). Let \( \chi_\delta \) be a sufficiently large number such that
\[
\Pr_{\Pi_{i=1}^{q+1} p_{E_i}} \left\{ \max_{\ell \in \{1, 2, \ldots, q+1\}} (E_\ell^2 + 2E_\ell) \geq \chi_\delta \right\} \leq \delta.
\] (216)

In addition, since \( R \) is \( \varepsilon \)-achievable, it follows from Definition 4 that there exists a sequence of \( (n, M, \varepsilon) \)-codes such that
\[
\liminf_{n \to \infty} \frac{1}{n} \log M \geq R.
\] (217)
Fix a sufficiently large \( n \in \mathbb{N} \) such that (172) holds, (173) holds and
\[
\frac{(\log e)^2 \chi_\delta}{\sqrt{n}} \leq \frac{\delta}{2},
\] (218)
and fix the corresponding \( (n, M, \varepsilon) \)-code. Let \( p_{W, E^n, X^n, Y^n, \hat{W}} \) be the probability distribution induced by the \( (n, M, \varepsilon) \)-code. Unless specified otherwise, all the probability, expectation and variance terms are evaluated according to \( p_{W, E^n, X^n, Y^n, \hat{W}} \). Since \( \{E_{b+1}\}_{b=1}^\infty \) are i.i.d. by assumption, it follows from (216) that
\[
\Pr \left\{ \max_{\ell \in \{1, 2, \ldots, q+1\}} (E_{b+1}^2 + 2E_{b+1}) \geq \chi_\delta \right\} \leq \delta.
\] (219)
Define
\[
\varepsilon(e^n) \triangleq \Pr \left\{ \hat{W} \neq W \mid E^n = e^n \right\}
\] (220)
and
\[
\Psi_\delta \triangleq \left\{ e^n \in \mathbb{R}^n \mid \varepsilon(e^n) < 1 - \delta, \max_{\ell \in \{1, 2, \ldots, q\}} (e_{b+1}^2 + 2e_{b+1}) < \chi_\delta \right\}.
\] (221)
Since the average error probability of the code is no larger than \( \varepsilon \), we have
\[
\int_{\mathbb{R}^n_+} p_{E^n}(e^n) \varepsilon(e^n) de^n \leq \varepsilon.
\] (222)
Consider
\[
\Pr \{ E^n \notin \Psi_\delta \} \overset{(221)}{\leq} \Pr \{ \varepsilon(E^n) \geq 1 - \delta \} + \Pr \left\{ \max_{\ell \in \{1, 2, \ldots, q\}} (E_{b+1}^2 + 2E_{b+1}) \geq \chi_\delta \right\}
\] (223)
\[
\overset{(219)}{\leq} \Pr \{ \varepsilon(E^n) \geq 1 - \delta \} + \Pr \{ \max_{\ell \in \{1, 2, \ldots, q\}} (E_{b+1}^2 + 2E_{b+1}) \geq \chi_\delta \}
\] (224)
\[
\overset{(a)}{\leq} \frac{\varepsilon}{1 - \delta} + \delta
\] (225)
where (a) follows from (222) and Markov’s inequality, which implies that

$$\Pr\{E^n \in \Psi_\delta\} \geq 1 - \varepsilon - 2\delta.$$  \hfill (226)

**Obtaining a Lower Bound on the Error Probability in Terms of the Type-II Error of a Hypothesis Test**

Define

$$s_{Y^n,W^n|E^n=e^n} \triangleq \left( \prod_{k=1}^{n} s_{Y_k|E_k=e_k} \right) p_{W^n|Y^n}$$  \hfill (227)

for all $e^n \in \Psi_\delta$ where

$$s_{Y_k|E_k=e_k}(y_k) \triangleq \mathcal{N}(y_k; 0, e_k + 1).$$  \hfill (228)

It follows from Proposition 11 and Definition 1 with the identifications $U \equiv W$, $V \equiv \hat{W}$, $p_{U,V} \equiv p_{W,\hat{W}|E^n=e^n}$, $|W| \equiv M$ and $\alpha \equiv \Pr\{W \neq \hat{W} | E^n = e^n\} \leq \varepsilon(e^n)$ that

$$\beta_{1-\varepsilon(e^n)}(p_{W,\hat{W}|E^n=e^n}||p_{W|E^n=e^n}s_{\hat{W}|E^n=e^n}) \leq \beta_{1-\alpha}(p_{W,\hat{W}|E^n=e^n}||p_{W|E^n=e^n}s_{\hat{W}|E^n=e^n}) \leq \frac{1}{M}$$  \hfill (229)

for all $e^n \in \Psi_\delta$.

**Using the DPI to Introduce the Channel Inputs and Outputs**

Using the DPI of $\beta_{1-\varepsilon(e^n)}$ in Lemma 10, we have

$$\beta_{1-\varepsilon(e^n)}(p_{W,\hat{W}|E^n=e^n}||p_{W|E^n=e^n}s_{\hat{W}|E^n=e^n})$$

$$\geq \beta_{1-\varepsilon(e^n)}(p_{W,E^n,X^n,Y^n,\hat{W}|E^n=e^n}||p_{W|E^n=e^n}\left( \prod_{k=1}^{n} p_{X_k|W,E^k=e_k} \right) s_{Y^n,\hat{W}|E^n=e^n})$$  \hfill (230)

for all $e^n \in \Psi_\delta$. For each $e^n \in \Psi_\delta$, fix a $\xi(e^n) > 0$ to be specified later. Since

$$p_{W,E^n,X^n,Y^n,\hat{W}|E^n=e^n} = p_{W|E^n=e^n} \left( \prod_{k=1}^{n} p_{X_k|W,E^k=e_k} p_{Y_k|X_k} \right) p_{\hat{W}|Y^n},$$  \hfill (231)

it follows from (229), the definition of $s_{Y^n,\hat{W}|E^n=e^n}$ in (227), (230) and Lemma 10 that

$$\log M \leq \log \xi(e^n) - \log \left( 1 - \varepsilon(e^n) - \Pr\left\{ \sum_{k=1}^{n} \log \frac{p_{Y_k|X_k}(Y_k|X_k)}{s_{Y_k|E_k}(Y_k|e_k)} \geq \log \xi(e^n) | E^n = e^n \right\} \right)$$

$$\geq \log \xi(e^n) - \log \left( 1 - \varepsilon(e^n) - \Pr\left\{ \sum_{k=1}^{n} \log \frac{p_{Y_k|X_k}(Y_k|X_k)}{s_{Y_k|E_k}(Y_k|e_k)} \geq \log \xi(e^n) | E^n = e^n \right\} \right)$$  \hfill (232)

for all $e^n \in \Psi_\delta$.

**Simplifying the Non-Asymptotic Bound**

Combining (17) and (228), we have

$$\log \frac{p_{Y_k|X_k}(Y_k|X_k)}{s_{Y_k|E_k}(Y_k|e_k)} = \frac{1}{2} \log(1+e_k) + \frac{\log e}{2(1+e_k)} (-e_k(Y_k - X_k)^2 + X_k^2 + 2X_k(Y_k - X_k))$$  \hfill (234)
for each $e^n \in \Psi_\delta$ and each $k \in \{1, 2, \ldots, n\}$. Let
\[ U_k(e_k) \triangleq \frac{\log e}{2(1 + e_k)} (-e_k(Y_k - X_k)^2 + 2X_k(Y_k - X_k) + e_k) \] (235)
for each $k \in \{1, 2, \ldots, n\}$. Then, it follows from (234) and the energy-harvesting constraints (15) that
\[ \Pr \left\{ \sum_{k=1}^{n} \log \frac{p_Y|x_k}{s_{Y|x_k}^2(Y_k|e_k)} - \sum_{k=1}^{n} \frac{1}{2} \log(1 + e_k) + \sum_{k=1}^{n} U_k(e_k) \left| E^n = e^n \right. \right\} = 1. \] (236)
Combining (233) and (236), we have for each $e^n \in \Psi_\delta$
\[ \log M \leq \log \xi(e^n) - \log \left( 1 - \varepsilon(e^n) \right) - \operatorname{Pr} \left\{ \sum_{k=1}^{n} U_k(e_k) \geq \log \xi(e^n) - \sum_{k=1}^{n} \frac{1}{2} \log(1 + e_k) \left| E^n = e^n \right. \right\}. \] (237)
In order to simplify the RHS of (237), we choose
\[ \log \xi(e^n) \triangleq \sum_{k=1}^{n} \frac{1}{2} \log(1 + e_k) + n^{3/4}, \] (238)
recall the definition of $\Psi_\delta$ in (221) and rewrite (237) as
\[ \log M \leq \sum_{k=1}^{n} \frac{1}{2} \log(1 + e_k) + n^{3/4} - \log \left( \delta - \operatorname{Pr} \left\{ \sum_{k=1}^{n} U_k(e_k) \geq n^{3/4} \left| E^n = e^n \right. \right\} \right). \] (239)
Applying Chebyshev’s inequality
Following (239), we evaluate for each $e^n \in \Psi_\delta$
\[ \mathbb{E} \left[ \sum_{k=1}^{n} U_k(e_k) \left| E^n = e^n \right. \right] \overset{(a)}{=} 0 \] (240)
and
\[ \operatorname{Var} \left[ \sum_{k=1}^{n} U_k(e_k) \left| E^n = e^n \right. \right] \overset{(b)}{=} (\log e)^2 \sum_{k=1}^{n} \frac{e_k^2 + 2\mathbb{E}[X_k^2|E^n = e^n]}{2(1 + e_k)^2} \] (241)
\[ \leq (\log e)^2 \sum_{k=1}^{n} (e_k^2 + 2\mathbb{E}[X_k^2|E^n = e^n]) \] (242)
\[ \overset{(15)}{\leq} (\log e)^2 \sum_{k=1}^{n} (e_k^2 + 2e_k) \] (243)
\[ \overset{(221)}{\leq} (\log e)^2 n\chi_\delta \] (244)
where (a) and (b) are due to the definition of $U_k(e_k)$ in (235) and the following fact: $\{Y_k - X_k\}_{k=1}^{n}$ are i.i.d. standard normal random variables that are independent of $(E^n, X^n)$. Using (240), (244) and Chebyshev’s inequality, we have for each $e^n \in \Psi_\delta$
\[ \Pr \left\{ \sum_{k=1}^{n} U_k(e_k) \geq n^{3/4} \left| E^n = e^n \right. \right\} \leq \frac{(\log e)^2 \chi_\delta}{\sqrt{n}} \overset{(218)}{\leq} \frac{\delta}{2}. \] (245)
Combining (239) and (246), we have
\[
\log M \leq \inf_{e^n \in \Psi} \sum_{k=1}^{n} \frac{1}{2} \log(1 + e_k) + n^{3/4} - \log \delta + 1. \tag{247}
\]

Since the number of energy-arrival blocks of length \( L \) equals \( q \) by (172) and the length of the last block is no larger than \( nd + \left\lceil \frac{1}{\lambda} \right\rceil \) by (173), it follows from (247) that
\[
\log M \leq \inf_{e^n \in \Psi} \left\{ \sum_{\ell=1}^{q} \frac{n \lambda}{2} C(e_{b_{\ell}+1}) + \left( \frac{nd}{2} + \left\lceil \frac{1}{\lambda} \right\rceil \right) C(e_{b_{q+1}+1}) \right\} + n^{3/4} - \log \delta + 1 \tag{248}
\]
\[
\leq n \inf_{e^n \in \Psi} \left\{ \sum_{\ell=1}^{q} \lambda C(e_{b_{\ell}+1}) + dC(e_{b_{q+1}+1}) \right\} + n^{3/4} - \log \delta + 1 + \left\lceil \frac{1}{\lambda} \right\rceil C(\chi_{\delta}). \tag{249}
\]

In order to simplify the first term in (249), we define \( \phi : \mathbb{R}_+^n \to \mathbb{R} \) as
\[
\phi(e^n) \triangleq \sum_{\ell=1}^{q} \lambda C(e_{b_{\ell}+1}) + dC(e_{b_{q+1}+1}) \tag{250}
\]
and consider the following chain of inequalities where the sets \( \Psi \) and \( \Omega \) are assumed to be Borel measurable:
\[
\inf_{e^n \in \Psi} \phi(e^n) \leq \sup_{\Psi \subseteq \mathbb{R}_+^n: \Pr\{E^n \in \Psi\} \geq 1 - \varepsilon - 2\delta} \inf_{e^n \in \Psi} \phi(e^n) \tag{251}
\]
\[
= \sup_{\Psi \subseteq \mathbb{R}_+^n: \Pr\{E^n \in \Psi\} \geq 1 - \varepsilon - 2\delta} \inf \phi(\Psi) \tag{252}
\]
\[
\leq \sup_{\Omega \subseteq \mathbb{R}_+} \left\{ \inf \Omega \Pr\{\phi(E^n) \in \Omega\} \geq 1 - \varepsilon - 2\delta \right\} \tag{253}
\]
\[
= \sup_{r \in \mathbb{R}_+} \left\{ \Pr\left\{ \sum_{\ell=1}^{q} \lambda C(E_{b_{\ell}+1}) + dC(E_{b_{q+1}+1}) \geq r \right\} \geq 1 - \varepsilon - 2\delta \right\} \tag{254}
\]
\[
\leq \overline{R}_c + \eta. \tag{255}
\]

Combining (247) and (255), we obtain
\[
\log M \leq n(\overline{R}_c + \eta) + n^{3/4} - \log \delta + 1 + \left\lceil \frac{1}{\lambda} \right\rceil C(\chi_{\delta}), \tag{256}
\]
which together with (217) implies that
\[
R \leq \overline{R}_c + \eta. \tag{257}
\]

Since \( \eta > 0 \) is arbitrary and \( R \) is an arbitrary \( \varepsilon \)-achievable rate, the inequality (213) follows from (257).

**VIII. Concluding Remarks and Future Work**

This paper studies the \( \varepsilon \)-capacity and the second-order coding rate for the AWGN EH channel with an infinite battery under the assumption that the error probabilities do not vanish as the blocklength increases. The EH process is assumed to be block i.i.d. where the blocks have length \( L \).
For the case where $L$ is a constant or grows sublinearly in the blocklength $n$, we have the following two findings stated in Theorem 1: (i) The $\varepsilon$-capacity is the same for all $\varepsilon \in (0, 1)$, i.e., the strong converse holds; (ii) A lower bound and an upper bound on the second-order coding rate have been obtained, where the lower bound is obtained by analyzing the conventional save-and-transmit strategy [3].

For the case where $L$ grows linearly in $n$, we prove in Theorem 2 a lower bound and an upper bound on the $\varepsilon$-capacity.

Two interesting directions for future research are obtaining the full characterization of the $\varepsilon$-capacity and good approximations on the second-order coding rate for $L = \lambda n$, i.e., a strengthening of Theorem 2. In addition, while this work investigates only optimal codes which have high decoding complexity, future research may compare the performances between an industrial low-complexity yet (first-order) optimal code for the AWGN channel and its (adaptive) save-and-transmit counterpart for the AWGN EH channel (as performed for the binary-input EH channel in [25]). Finally, one may explore analogies among AWGN EH channels, slow fading channels and channels with mixed states for the case in which $L$ grows linearly in $n$.

**APPENDIX A**

**PROOF OF COROLLARY 1**

In view of Theorem 1, it suffices to prove that $V_{e}^{-} \leq V_{e}^{-\varepsilon}$ for the case where $L$ is a constant. To this end, we let $L$ be a fixed constant and consider the following chain of inequalities:

\[
V_{e}^{-} = \sup_{\varepsilon} \left\{ -C(P) \sqrt{\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}} + \sqrt{\frac{P(\log e)^2}{1 + P}} \Phi^{-1}(\varepsilon) \right\}
\]

\[
\geq -C(P) \sqrt{\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}} + \sqrt{\frac{P(\log e)^2}{1 + P}} \Phi^{-1}
\]

\[
\geq -C(P) \sqrt{\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}} - \sqrt{\frac{2P(\log e)^2}{1 + P}} \ln \sqrt{\frac{e}{\varepsilon}}
\]

\[
= \left( C(P) \sqrt{2} - \frac{4P(\log e)}{1 + P} \right) \sqrt{\log \frac{1}{\varepsilon}}
\]

\[
= V_{e}^{-\varepsilon},
\]

where

(a) is due to the easily verified fact that

\[
\Phi \left( -\sqrt{2 \ln \frac{\sqrt{e}}{\varepsilon}} \right) \leq \frac{1}{\sqrt{2\pi}} e^{-\ln \frac{\sqrt{e}}{\varepsilon}} = \frac{\varepsilon}{\sqrt{2\pi e}}.
\]

(b) is due to the fact that $\varepsilon < 1/2 < \min \left\{ 1 - \frac{1}{\sqrt{2\pi e}}, \frac{1}{\sqrt{e}} \right\}$. 

APPENDIX B

PROOF OF COROLLARY 2

Fix any $\varepsilon \in (0, \Phi(-1))$. Using the easily verified fact that $\Phi^{-1}(\varepsilon) \leq -\sqrt{2 \ln \frac{1}{\varepsilon}}$, we obtain that

$$V_{\varepsilon}^+ \leq -\sqrt{\frac{(2P^2 + \mathbb{E}[E_1^2]) \log e}{2(1 + P)^2}} \times \sqrt{\log \frac{1}{\varepsilon}},$$

(265)

which together with (31) and (30) implies that (32) holds. The rest of the proof is dedicated to showing $\Phi^{-1}(\varepsilon) \leq -\sqrt{2 \ln \frac{1}{\varepsilon}}$. Let $a = \Phi^{-1}(\varepsilon)$. Since $a \leq -1$ due to the assumption that $\varepsilon \leq \Phi(-1)$, we have $\varepsilon = \Phi(a) \leq e^{-a^2/2}$, which then implies that $a \leq -\sqrt{2 \ln \frac{1}{\varepsilon}}$.

APPENDIX C

PROOF OF COROLLARY 3

Suppose $E_1$ has a continuous and strictly increasing cdf (i.e., the mapping $a \mapsto \Pr\{E_1 \leq a\}$ is continuous and strictly increasing on $[0, \infty)$). It follows that

$$\Pr_{\prod_{\ell=1}^{p_{E_1}} \sum_{q=1}^{\lambda C(E_\ell)} + dC(E_{q+1}) \geq r}$$

(266)

is continuous in $r$ and strictly increasing, which then implies that

$$R_{\varepsilon} = \sup \left\{ r \in \mathbb{R}_+ \mid \Pr_{\prod_{\ell=1}^{p_{E_1}} \sum_{q=1}^{\lambda C(E_\ell)} + dC(E_{q+1}) < r} = \varepsilon \right\}$$

(267)

$$= R_{\varepsilon}^{thr}$$

(268)

$$= \inf \left\{ r \in \mathbb{R}_+ \mid \Pr_{\prod_{\ell=1}^{p_{E_1}} \sum_{q=1}^{\lambda C(E_\ell)} + dC(E_{q+1}) < r} = \varepsilon \right\}$$

(269)

$$= \overline{R}_{\varepsilon},$$

(270)

which together with Theorem 2 implies that (39) holds for all $\varepsilon \in (0, 1)$.

APPENDIX D

PROOF OF COROLLARY 5

To facilitate discussion, let $m_n$ denote the RHS of (51), and simple calculations reveal that

$$\lim_{n \to \infty} \frac{m_n}{\sqrt{L_n}} = \sqrt{2 \left( \frac{\mathbb{E}[E_1^2]}{P^2} + 1 \right) \log \frac{1}{\varepsilon_1}}$$

(271)

$$= \sqrt{\rho \log \frac{1}{\varepsilon_1}}.$$ 

(272)

For each $n^* \in \mathbb{N}$, let $\hat{n}$ be the unique natural number that satisfies

$$\hat{n} - 1 + |m_{\hat{n} - 1}| \leq n^* < \hat{n} + |m_{\hat{n}}|.$$ 

(273)
It is clear from (272) and (273) that
\[
\lim_{n^* \to \infty} \frac{n^*}{n} = 1. \tag{274}
\]

By Lemma 4, there exists for each sufficiently large \( n \in \mathbb{N} \) an \((n + |m_n|, M, \varepsilon)-code\) such that (45) holds, which implies from the left inequality in (273) that for each sufficiently large \( n^* \in \mathbb{N} \), there exists an \((n^*, M, \varepsilon)-code\) such that
\[
\log M \geq (n - 1)C(P) + \sqrt{(n - 1)P(\log e)^2} \Phi^{-1}(\varepsilon_2) - \frac{1}{2} \log(n - 1) - \kappa_1, \tag{275}
\]
which then implies from the right inequality in (273) that
\[
\log M \geq (n^* - 1 - \lfloor m_{n^*} \rfloor)C(P) + \sqrt{(n^* - 1)P(\log e)^2} \Phi^{-1}(\varepsilon_2) - \frac{1}{2} \log(n^* - 1) - \kappa_1. \tag{276}
\]

Combining the facts that
\[
\frac{\log M - n^*C(P)}{\sqrt{L n^*}} \geq \frac{(-1 - \lfloor m_{n^*} \rfloor)C(P)}{\sqrt{L n^*}} + \sqrt{(n^* - 1)P(\log e)^2} \Phi^{-1}(\varepsilon_2) - \frac{1}{2} \log(n^* - 1) - \kappa_1, \tag{277}
\]
\[
\lim_{n^* \to \infty} \frac{n^*}{n} \tag{274} = 1 \tag{278}
\]
and
\[
\lim_{n^* \to \infty} \frac{\lfloor m_{n^*} \rfloor}{\sqrt{L n^*}} = \lim_{n \to \infty} \frac{|m_n|}{\sqrt{L n}} = \sqrt{\rho \log \frac{1}{\varepsilon_1}}, \tag{279}
\]
we conclude that
\[
\liminf_{n^* \to \infty} \frac{\log M - n^*C(P)}{\sqrt{L n^*}} \geq -C(P) \sqrt{\rho \log \frac{1}{\varepsilon_1}} + \liminf_{n^* \to \infty} \sqrt{\frac{P(\log e)^2}{L(1 + P)} \Phi^{-1}(\varepsilon_2)} \tag{280}
\]
\[
= \begin{cases} 
-C(P) \sqrt{\rho \log \frac{1}{\varepsilon_1}} & \text{if } L = \omega(1), \\
-C(P) \sqrt{\rho \log \frac{1}{\varepsilon_1}} + \sqrt{\frac{P(\log e)^2}{L(1 + P)} \Phi^{-1}(\varepsilon_2)} & \text{if } L \text{ is a constant.} 
\end{cases} \tag{281}
\]

This completes the proof.

**APPENDIX E**

**PROOF OF Lemma 7**

In this proof, all the probability, expectation and variance terms are evaluated according to \( p_{X^n} p_{E_{m+n}} \). In order to obtain an upper bound on \( \Pr \left\{ \bigcup_{k=1}^{n} \left\{ \sum_{i=1}^{k} X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right\} \), we construct the following sequence denoted
by \( \{B_k\}_{k=1}^{m+n} \). For each \( k \in \{1,2,\ldots,m+n\} \), define \( B_k \) recursively\(^4\) as
\[
B_k \overset{\text{(282)}}{=} \begin{cases} 
E_1 & \text{if } k = 1, \\
B_{k-1} + E_k & \text{if } k \in \{2,3,\ldots,m\}, \\
B_{k-1} + E_k - X_{k-m}^2 & \text{if } k \in \{m+1,m+2,\ldots,m+n\} \text{ and } B_{k-1} > 0, \\
B_{k-1} & \text{if } k \in \{m+1,m+2,\ldots,m+n\} \text{ and } B_{k-1} \leq 0.
\end{cases}
\]

By inspecting (282), we have
\[
\{B_{m+n} \leq 0\} = \bigcup_{k=1}^n \left\{ \sum_{i=1}^{m+k} E_i - \sum_{i=1}^k X_i^2 \leq 0 \right\},
\]
which implies that
\[
\Pr\left\{ \bigcup_{k=1}^n \left\{ \sum_{i=1}^k X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right\} = \Pr\{B_{m+n} \leq 0\}.
\]

Define for each \( k \in \{1,2,\ldots,m+n\} \)
\[
U_k \overset{\text{(285)}}{=} \begin{cases} 
B_1 & \text{if } k = 1, \\
B_k - B_{k-1} & \text{otherwise}
\end{cases}
\]
\[
\overset{\text{(282)}}{=} \begin{cases} 
E_k & \text{if } k \in \{1,2,\ldots,m\}, \\
E_k - X_{k-m}^2 & \text{if } k \in \{m+1,m+2,\ldots,m+n\} \text{ and } B_{k-1} > 0, \\
0 & \text{if } k \in \{m+1,m+2,\ldots,m+n\} \text{ and } B_{k-1} \leq 0.
\end{cases}
\]

Following (284), we consider the following chain of inequalities for any \( t > 0 \):
\[
\Pr\{B_{m+n} \leq 0\} \overset{\text{(285)}}{=} \Pr\left\{ \sum_{k=1}^{m+n} U_k \leq 0 \right\} \overset{\text{(287)}}{=} \Pr\left\{ e^{-t\sum_{k=1}^{m+n} U_k} \geq 1 \right\} \overset{\text{(288)}}{\leq} \exp\left\{ -t \sum_{k=1}^{m+n} U_k \right\} \overset{\text{(289)}}{\leq} \exp\left\{ -t \sum_{k=1}^{m+n} U_k \right\}
\]
\[
\overset{\text{where (a) follows from Markov’s inequality. In order to simplify the RHS of (289), we consider the following chain of inequalities for each } i \in \{1,\ldots,n\}}{\overset{\text{(290)}}{=}}
\exp\left\{ -t \sum_{k=1}^{m+i} U_k \right\}
\]
\[
\overset{\text{(291)}}{=} \exp\left\{ -t \sum_{k=1}^{m+i} U_k \right\} \exp\left\{ -t \sum_{k=1}^{m+i} U_k \right\}
\]
\[
\overset{\text{(286)}}{\leq} \exp\left\{ -t \sum_{k=1}^{m+i} U_k \right\} \exp\left\{ -t \sum_{k=1}^{m+i} U_k \right\}
\]
\[
\overset{\text{(293)}}{=} \exp\left\{ -t \sum_{k=1}^{m+i} U_k \right\}
\]
\[\]
\[\]

\(^4\)The construction of \( \{B_k\}_{k=1}^{m+n} \) is inspired by a standard proof of Kolmogorov’s inequality.
where (a) follows from the independence between \((E_{m+i}, X_i)\) and \(U_{m+i-1}\) due to the independence between \((E_{m+i}, X_i)\) and \((E^{m+i-1}, X^{i-1})\). Combining (289) and (294), we have

\[
Pr \{B_{m+n} \leq 0\} \leq \max \left\{ \left( E \left[ e^{-t(E_1 - X_1^2)} \right] \right)^n, 1 \right\} E \left[ e^{-t \sum_{k=1}^m U_k} \right].
\]

(295)

Since

\[
E \left[ e^{-t \sum_{k=1}^m U_k} \right] = E \left[ e^{-t \sum_{k=1}^m E_k} \right]
\]

(296)

by (286), it follows from (295) that

\[
Pr \{B_{m+n} \leq 0\} \leq \max \left\{ \left( E \left[ e^{-t(E_1 - X_1^2)} \right] \right)^n, 1 \right\} \left( E \left[ e^{-tE_1} \right] \right)^m.
\]

(297)

Since \(X_1\) and \(E_1\) are independent, we can rewrite (297) as

\[
Pr \{B_{m+n} \leq 0\} \leq \max \left\{ \left( E \left[ e^{-tE_1} \right] \right)^{m+n} \left( E \left[ e^{tX_1^2} \right] \right)^n, \left( E \left[ e^{-tE_1} \right] \right)^m \right\}.
\]

(298)

In order to simplify the RHS of (298), we use the following two facts, whose proofs can be found in [15, Appendix]:

For any \(y \geq 0\),

\[
1 + y \leq e^y \leq 1 + y + \frac{y^2 e^y}{2}
\]

(299)

and

\[
1 - y \leq e^{-y} \leq 1 - y + \frac{y^2}{2}.
\]

(300)

Fix a sufficiently small \(t > 0\) such that \(E[X_1^4 e^{tX_1^2}] < \infty\) and \(P - \frac{tE[X_1^2]}{2} > 0\). Following (298), we use (299), (300) and (58) to obtain

\[
E \left[ e^{-tE_1} \right] \leq 1 - tP + \frac{t^2 E[E_1^2]}{2}
\]

(301)

and

\[
E \left[ e^{tX_1^2} \right] \leq 1 + tP + \frac{t^2 E[X_1^4 e^{tX_1^2}]}{2},
\]

(302)

which implies that

\[
E \left[ e^{-tE_1} \right] E \left[ e^{tX_1^2} \right] \leq 1 + t^2 \left( \frac{E[E_1^2]}{2} + E[X_1^4 e^{tX_1^2}] \right) - P^2 + \frac{t^3 P}{2} \left( E[E_1^2] - E[X_1^4 e^{tX_1^2}] \right) + \frac{t^4 E[E_1^2] E[X_1^4 e^{tX_1^2}]}{4}.
\]

(303)

Define \(a_t\) and \(b_t\) as in (60) and (61) respectively. It then follows from (301) and (303) that

\[
E \left[ e^{-tE_1} \right] \leq 1 - a_t t
\]

(304)
and
\[
\mathbb{E}\left[ e^{-tE_1} \right] \mathbb{E}\left[ e^{tX_i^2} \right] \leq 1 + b_i t^2.
\] (305)

Combining (284), (298), (304), (305), (299) and (300), we obtain
\[
\Pr \left\{ \bigcup_{k=1}^{n} \left\{ \sum_{i=1}^{k} X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right\} \leq e^{-a_t \max \left\{ e^{b_t^2 n}, 1 \right\}}
\] (306)
\[
= e^{-a_t \max \{ e^{b_t^2 n}, 1 \}}.
\] (307)

**APPENDIX F**

**PROOF OF COROLLARY 8**

Since \( \{E_k\}_{k=1}^{m+n} \) is distributed according to an i.i.d.-block manner with block size \( L \), we cannot apply Lemma 7 directly for \( L > 1 \) to bound \( \Pr_{(X^n_{m+n})} \left( \bigcup_{k=1}^{n} \left\{ \sum_{i=1}^{k} X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right) \). In the following, we will construct two sequences based on \( \{E_k\}_{k=1}^{m+n} \) and \( \{X_k\}_{k=1}^{n} \) so that Lemma 7 can be applied to the resultant sequences. Define
\[
\tilde{n} \triangleq \left\lceil \frac{n}{L} \right\rceil L
\] (308)
and
\[
\tilde{m} \triangleq \left\lfloor \frac{m}{L} \right\rfloor L.
\] (309)

Let \( \{X_k\}_{k=1}^{\tilde{n}} \) be a sequence of i.i.d. random variables where \( X_1 \sim \mathcal{N}(x_1; 0, P) \), and let \( \{E_k\}_{k=1}^{\tilde{m}+\tilde{n}} \) be a sequence of random variables that are distributed according to (13). Since \( \tilde{n} \geq n \) and \( \tilde{m} \leq m \), we have
\[
\Pr_{(X^n_{m+n})} \left( \bigcup_{k=1}^{n} \left\{ \sum_{i=1}^{k} X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right)
\leq \Pr_{(X^n_{m+n})} \left( \bigcup_{k=1}^{\tilde{n}} \left\{ \sum_{i=1}^{k} \tilde{X}_i^2 \geq \sum_{i=1}^{\tilde{m}+k} \tilde{E}_i \right\} \right).
\] (310)

To simplify notation, define
\[
\tilde{E}_\ell \triangleq \sum_{i=1}^{L} E_{b_\ell + i}
\] (311)
for each \( \ell \in \{1, 2, \ldots, (\tilde{m} + \tilde{n})/L \} \), and define
\[
\tilde{X}_\ell \triangleq \sum_{i=1}^{L} X_{b_\ell + i}
\] (312)
for each \( \ell \in \{1, 2, \ldots, \tilde{n}/L \} \). For each \( k \in \{1, 2, \ldots, \tilde{n} \} \), define \( \nu(k) \) to be the unique integer in \( \{1, 2, \ldots, \tilde{n}/L \} \) that satisfies \( k \in \{b_\nu(k) + 1, b_\nu(k) + 2, \ldots, b_\nu(k) + L \} \), i.e., \( \nu(k) \) is the index of the information block that contains the \( k \)th symbol of \( X^n \). Then for each \( k \in \{1, 2, \ldots, \tilde{n} \} \),
\[
\sum_{i=1}^{k} X_i^2 \leq \sum_{i=1}^{\nu(k)} \tilde{X}_i^2
\] (313)
and 
\[ \sum_{i=1}^{m+k} E_i \geq \sum_{i=1}^{\frac{n}{L} + \nu(k) - 1} \tilde{E}_i \]  
(314)

with probability 1. Therefore,
\[ \left\{ \sum_{i=1}^{k} X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \subseteq \left\{ \sum_{i=1}^{\nu(k)} \tilde{X}_i^2 \geq \sum_{i=1}^{\frac{n}{L} + \nu(k) - 1} \tilde{E}_i \right\} \]
for each \( k \in \{1, 2, \ldots, \bar{n}\} \), which implies that
\[ \Pr_{X_{\bar{n}E_{m+n}}} \left\{ \bigcup_{k=1}^{\bar{n}} \left\{ \sum_{i=1}^{k} X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right\} \leq \Pr_{X_{\bar{n}E_{m+n}}} \left\{ \bigcup_{\ell=1}^{\bar{n}} \left\{ \sum_{i=1}^{\ell} \tilde{X}_i^2 \geq \sum_{i=1}^{\frac{n}{L} + \ell - 1} \tilde{E}_i \right\} \right\} \]
(315)

which then implies from (310) that
\[ \Pr_{X_{\bar{n}E_{m+n}}} \left\{ \bigcup_{k=1}^{\bar{n}} \left\{ \sum_{i=1}^{k} X_i^2 \geq \sum_{i=1}^{m+k} E_i \right\} \right\} \leq \Pr_{X_{\bar{n}E_{m+n}}} \left\{ \bigcup_{\ell=1}^{\bar{n}} \left\{ \sum_{i=1}^{\ell} \tilde{X}_i^2 \geq \sum_{i=1}^{\frac{n}{L} + \ell - 1} \tilde{E}_i \right\} \right\} \]
(316)

By construction, \( \{ \tilde{E}_\ell \}_{\ell=1}^{(\bar{n} - L + \bar{n})/L} \) is a sequence of i.i.d. random variables with
\[ \mathbb{E} \left[ \tilde{E}_1 \right] = \mathbb{E} \left[ LE_{b_1+1} \right] = LP \]
(318)

and
\[ \mathbb{E} \left[ \tilde{E}_1^2 \right] = \mathbb{E} \left[ (LE_{b_1+1})^2 \right] = L^2 \mathbb{E} \left[ E_1^2 \right] \]
(319)

and \( \{ \tilde{X}_\ell \}_{\ell=1}^{\bar{n}/L} \) is a sequence of i.i.d. Gaussian random variables with
\[ \mathbb{E} \left[ \tilde{X}_1 \right] = 0 \]
(320)
\[ \mathbb{E} \left[ \tilde{X}_1^2 \right] = LP \]
(321)

and
\[ \mathbb{E} \left[ \tilde{X}_1^2 e^{t \tilde{X}_1^2} \right] \overset{(10)}{=} \frac{3L^2 P^2}{(1 - 2LP t)^{5/2}} \quad \forall t \in \left(0, \frac{1}{2LP} \right) . \]
(322)

Combining (318) and (321) and (322), we have
\[ \mathbb{E} \left[ \tilde{E}_1 \right] = \mathbb{E} \left[ \tilde{X}_1^2 \right] = LP \]
(323)

and
\[ \mathbb{E} \left[ \tilde{X}_1^4 e^{t \tilde{X}_1^2} \right] < \infty \quad \forall t \in \left(0, \frac{1}{2LP} \right) . \]
(324)
Since (323) holds, we can apply Lemma 7 to \( \{ \tilde{X}_\ell \} \) and \( \{ \tilde{E}_\ell \} \) if the following two statements hold for \( t_n \) (which was defined in (66)):

\[
\mathbb{E} \left[ \tilde{X}_1^4 e^{t_n \tilde{X}_1^2} \right] < \infty \tag{325}
\]

and

\[
LP - \frac{t_n \mathbb{E}[\tilde{X}_1^2]}{2} > 0. \tag{326}
\]

To this end, we suppose \( n \) and \( m \) satisfy (67) and (68) respectively. Recalling the definition of \( \beta_0 \) in (64), we obtain from the definition of \( t_n \) in (66) and (67) that

\[
t_n \in \left(0, \min \left\{ \frac{1}{2LP}, \frac{2P}{LE[\tilde{E}_1^2]} \right\} \right), \tag{327}
\]

which implies from (324) that

\[
\mathbb{E} \left[ \tilde{X}_1^4 e^{t_n \tilde{X}_1^2} \right] < \infty. \tag{328}
\]

In addition,

\[
LP - \frac{t_n \mathbb{E}[\tilde{E}_1^2]}{2} \overset{(323)}{=} LP - \frac{t_n L^2 \mathbb{E}[\tilde{E}_1^2]}{2} \overset{(327)}{>} 0. \tag{329}
\]

Consequently, \( \{ \tilde{E}_\ell \} \) \( \{ \tilde{X}_\ell \} \) are two sequences of i.i.d. random variables that satisfy (323), (328) and (330). Therefore, we have the following inequality due to Lemma 7 together with the definitions of \( \alpha_t \) and \( \beta_t \) in (63) and (65) respectively and the equalities (318), (319) and (322):

\[
\Pr_{P,x^n P_{E,m-L+n}} \left( \bigcup_{\ell=1}^n \left\{ \sum_{i=1}^\ell \tilde{X}_i^2 \geq \sum_{i=1}^\ell \tilde{E}_i \right\} \right) \leq e^{-\alpha_t t_n (\bar{m}/L - 1) + \beta_t t_n^2 \tilde{n}/L}. \tag{331}
\]

In order to simplify the RHS of (331), we consider

\[
\bar{m}/L - 1 \overset{(309)}{>} m/L - 2 \tag{332}
\]

\[
\alpha_t \sqrt{\beta_0} \overset{(68)}{>} \frac{\sqrt{(n/L + 1) \log(1/\varepsilon_1)} (\beta_t + \beta_0)}{\alpha_t \sqrt{\beta_0}} \tag{333}
\]

\[
\alpha_t \sqrt{\beta_0} \overset{(308)}{>} \frac{\sqrt{(n/L) \log(1/\varepsilon_1)} (\beta_t + \beta_0)}{\alpha_t \sqrt{\beta_0}} \tag{334}
\]

Following (331), we consider

\[
-\alpha_t t_n (\bar{m}/L - 1) + \beta_t t_n^2 \tilde{n}/L \overset{(334)}{<} t_n \left( -\frac{\sqrt{(n/L) \log(1/\varepsilon_1)} (\beta_t + \beta_0)}{\sqrt{\beta_0}} + \frac{\beta_t t_n \tilde{n}}{L} \right) \tag{335}
\]

\[
\overset{(a)}{=} -t_n \sqrt{\beta_0 (n/L) \log(1/\varepsilon_1)} \tag{336}
\]

\[
\overset{(b)}{=} \log(\varepsilon_1) \tag{337}
\]
where (a) and (b) follow from the fact due to (66) and (308) that $t_n = \sqrt{\frac{\log(1/\varepsilon_1)}{(n/L)\beta_0}}$. Combining (31) and (337), we have

$$\Pr_{P_X \sim P_{E^m-L-n}} \left\{ \bigcup_{\ell=1}^{\frac{n}{L}} \left\{ \sum_{i=1}^{\ell} X_i^2 \geq \sum_{i=1}^{\ell} E_i \right\} \right\} \leq \varepsilon_1,$$

(338)

which implies from (317) that (69) holds.

**Appendix G**

Simple derivations in the proof of Lemma 4

A. Derivation of (93)

Consider the following chain of inequalities:

$$\left( E_{P_X,Y} \left[ \log \left( \frac{P_{Y|X}(Y|X)}{P_Y(Y)} \right) - \mu \right] \right)^{1/3} \leq \frac{\log e}{2(1+P)} \left( P \left( E_{P_Z} [Z^6] \right)^{1/3} + 2 \left( E_{P_X} [X^3] E_{P_Z} [Z^3] \right)^{1/3} + \left( E_{P_X} [X^6] \right)^{1/3} \right)$$

(339)

$$= \frac{\log e}{2(1+P)} \left( 15^{1/3} P + \frac{16\sqrt{P}}{\pi} + 15^{1/3} P \right)$$

(340)

$$= \frac{\log e}{1+P} \left( 15^{1/3} P + \frac{8\sqrt{P}}{\pi} \right).$$

(341)

where (a) follows from (91) and the triangle inequality for the 3-norm.

B. Derivation of (104)

By Taylor's theorem, we have

$$\Phi^{-1} \left( \varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}} \right) = \Phi^{-1}(\varepsilon_2) - \left( \frac{T}{\sigma^3 \sqrt{n}} + \frac{1}{\sqrt{n}} \right) (\Phi^{-1})'(c)$$

(342)

for some real number

$$c \in \left[ \varepsilon_2 - \frac{T}{\sigma^3 \sqrt{n}} - \frac{1}{\sqrt{n}}, \varepsilon_2 \right].$$

(343)

Since

$$(\Phi^{-1})'(c) = \frac{1}{\Phi'(\Phi^{-1}(c))} = \frac{1}{\mathcal{N}(\Phi^{-1}(c); 0, 1)}$$

(344)

and

$$\mathcal{N}(\Phi^{-1}(c); 0, 1) \geq \mathcal{N}(\Phi^{-1}(\min\{\varepsilon_2^2, 1-\varepsilon_2\}); 0, 1)$$

(345)

by (10) and (343) respectively, it follows that

$$(\Phi^{-1})'(c) \leq \frac{1}{\mathcal{N}(\Phi^{-1}(\min\{\varepsilon_2^2, 1-\varepsilon_2\}); 0, 1)}.$$ 

(346)

Consequently, (104) follows from (342) and (346).
APPENDIX H

SIMPLE DERIVATIONS IN THE PROOF OF LEMMA 6

A. Derivation of (160)

Consider the following chain of inequalities:

\[
T^{1/3} \overset{(155)}{=} \left( \mathbb{E}_{p_{E_1}} \prod_{i=1}^{L} p_{Z_i} \left\lfloor -P \sum_{i=1}^{L} Z_i^2 + 2\sqrt{E_1} \sum_{i=1}^{L} Z_i + LE_1 \right\rfloor \right)^{1/3}
\]

\[
\overset{(a)}{\leq} \frac{\log e}{2(1 + P)} \left( LP \left( \mathbb{E}_{p_{Z_1}} [Z_1^6] \right)^{1/3} + 2L \left( \mathbb{E}[E_1^{3/2}] \right)^{1/3} \cdot \left( \mathbb{E}_{p_{Z_1}} [Z_1^3] \right)^{1/3} + L \left( \mathbb{E}[E_1^3] \right)^{1/3} \right)
\]

\[
= \frac{L \log e}{2(1 + P)} \left( 15^{1/3} P + 2(2\sqrt{2}/\pi)^{1/3} \cdot \left( \mathbb{E}[E_1^{3/2}] \right)^{1/3} + \left( \mathbb{E}[E_1^3] \right)^{1/3} \right)
\]

where (a) follows from the triangle inequality for the 3-norm.

B. Derivation of (167)

By Taylor’s theorem, we have

\[
\Phi^{-1} \left( \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right) = \Phi^{-1}(\varepsilon) + 2\tau_2 \sqrt{\frac{L}{n}} (\Phi^{-1})'(c)
\]

for some real number

\[
c \in \left[ \varepsilon, \varepsilon + 2\tau_2 \sqrt{\frac{L}{n}} \right] \subseteq [\varepsilon, \varepsilon + (1 - \varepsilon)^2].
\]

Since

\[
(\Phi^{-1})'(c) = \frac{1}{\Phi'(\Phi^{-1}(c))} = \frac{1}{N(\Phi^{-1}(c); 0, 1)}
\]

by (10) and

\[
N(\Phi^{-1}(c); 0, 1) \geq N(\Phi^{-1}(\min\{\varepsilon, 1 - \varepsilon - (1 - \varepsilon)^2\}); 0, 1)
\]

\[
= N(\Phi^{-1}(\min\{\varepsilon, \varepsilon(1 - \varepsilon)\}); 0, 1)
\]

by (350), it follows that

\[
(\Phi^{-1})'(c) \leq \frac{1}{N(\Phi^{-1}(\min\{\varepsilon, \varepsilon(1 - \varepsilon)\}); 0, 1)}.
\]

APPENDIX I

PROOF OF PROPOSITION 12

Fix any sufficiently large \( n \) such that

\[
\left| \frac{1}{\lambda} \right| = \left| \frac{1}{\lambda - \frac{1}{n}} \right|.
\]
Then,
\[
\rho (\text{170}) = \left\lfloor \frac{n}{\lambda n} \right\rfloor \quad (356)
\]
\[
\geq \left\lfloor \frac{n}{\lambda n} \right\rfloor \quad (357)
\]
\[
\overset{(34)}{=} q. \quad (358)
\]

In addition,
\[
\rho (\text{170}) = \left\lfloor \frac{n}{\lambda n} \right\rfloor \quad (359)
\]
\[
\leq \left\lfloor \frac{n}{\lambda n - 1} \right\rfloor \quad \overset{(355)}{=} \left\lfloor \frac{1}{\lambda} \right\rfloor \quad (360)
\]
\[
\overset{(34)}{=} q. \quad (361)
\]

Combining (358) and (362), we have
\[
\rho = q. \quad (363)
\]

It remains to prove (173). Using (363) and the definition of \( L \) in (33), we have
\[
n - \rho L = n - q\lfloor \lambda n \rfloor, \quad (364)
\]
which together with the definition of \( d \) in (35) implies (173).

APPENDIX J

PROOF OF PROPOSITION 13

Consider the following chain of inequalities where all the probability and expectation terms are evaluated with respect to \( p_{E_1 p_Z L} \):
\[
\Pr_{p_{E_1 p_Z L}} \left\{ \mathbb{1} \mathbb{1}_{\sqrt{L}} \left[ \sum_{i=1}^{\sqrt{L}} \left( \sqrt{g^\Delta(E_1) + Z_i} \right)^2 - g^\Delta(E_1) - 1 \right] \geq \Delta \right\}
\]
\[
\overset{(a)}{\leq} \frac{4 \mathbb{E} \left[ g^\Delta(E_1) \right] + 2}{\Delta^2 \sqrt{L}} \quad (365)
\]
\[
\overset{(176)}{\leq} \frac{4 \mathbb{E} [E_1] + 2}{\Delta^2 \sqrt{L}} \quad (366)
\]
\[
\overset{(174)}{\leq} \frac{1}{L^{1/6}} \quad (367)
\]
where (a) follows from Chebyshev’s inequality.
Since $\text{Var}_{p_{E_1}}[E_1] = 0$ by assumption, it follows that $E_{p_{E_1}}[E_1^2] = \hat{P}^2$. Therefore, for any $\varepsilon > 0$ and any sufficiently large $L$ that satisfies
\[ L \geq \left( \log \left( \frac{2 + \varepsilon}{\varepsilon^2} \right) \right)^4 \]  
and
\[ \frac{L}{\log L} \geq \max \left\{ 12\sqrt{2}, \frac{e^{0.4(2 + \varepsilon)}}{\varepsilon} \right\}, \]  
we can use [15, Th. 1] to conclude that there exists an $(L - \left\lceil L^{2/3} \right\rceil + m, M, \varepsilon)$-code such that
\[ \log M \geq \left( L - \left\lceil L^{2/3} \right\rceil \right) C(\tilde{P}) - \sqrt{3L} - \left( L - \left\lceil L^{2/3} \right\rceil \right)^{\frac{1}{2}} - 1 \]  
where
\[ m \triangleq \left\lceil 12\sqrt{3\sqrt{2}} (L - \left\lceil L^{2/3} \right\rceil) \log (L - \left\lceil L^{2/3} \right\rceil) \right\rceil \]  
(372)
denotes the length of the initial saving period before any transmission occurs and $L - \left\lceil L^{2/3} \right\rceil$ denotes the length of the actual transmission period. Let $\varepsilon \triangleq \frac{1}{2\sqrt{L}}$ and fix a sufficiently large $L$ that satisfies
\[ L \geq \left( \log \left( 2L + \sqrt{L} \right) \right)^4, \]  
(373)
\[ \frac{L}{\log L} \geq \max \left\{ 12\sqrt{2}, e^{0.4} \left( 2\sqrt{L} + 1 \right) \right\} \]  
(374) and
\[ (2 - \sqrt{3}) L^{3/4} \geq \left( L - \left\lceil L^{2/3} \right\rceil \right)^{\frac{1}{2}} + 1. \]  
(375)
Since (368) and (369) hold by (373) and (374), it follows from (371) and (372) that there exists an $(L - \left\lceil \sqrt{L} \right\rceil, M, \frac{1}{\sqrt{L}})$-code such that
\[ \log M \geq \left( L - \left\lceil L^{2/3} \right\rceil \right) C(\tilde{P}) - \sqrt{3L} - \left( L - \left\lceil L^{2/3} \right\rceil \right)^{\frac{1}{2}} - 1 \]  
(376)
\[ \geq \left( L - \left\lceil L^{2/3} \right\rceil \right) C(\tilde{P}) - 2L^{3/4} \]  
(377)
\[ \geq \gamma(L, \tilde{P}). \]  
(378)

ACKNOWLEDGEMENTS

The authors would like to thank Associate Editor Michele Wigger and the three anonymous reviewers for the useful comments that greatly improve the presentation of this work.
REFERENCES

[1] F. Zhang and V. K. N. Lau, “Closed-form delay-optimal power control for energy harvesting wireless system with finite energy storage,” IEEE Trans. Signal Process., vol. 62, no. 21, pp. 5706–5715, 2014.
[2] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge, U.K.: Cambridge University Press, 2005.
[3] O. Ozel and S. Ulukus, “Achieving AWGN capacity under stochastic energy harvesting,” IEEE Trans. Inf. Theory, vol. 58, no. 10, pp. 6471–6483, 2012.
[4] T. S. Han, Information-Spectrum Methods in Information Theory. Springer Berlin Heidelberg, 2003.
[5] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” IEEE Trans. Inf. Theory, vol. 55, no. 11, pp. 4947–4966, 2009.
[6] R. H. Etkin, D. N. C. Tse, and H. Wang, “Gaussian interference channel capacity to within one bit,” IEEE Trans. Inf. Theory, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.
[7] A. Özgür, O. Lévêque, and D. Tse, “Operating regimes of large wireless networks,” Foundations and Trends in Networking, vol. 5, no. 1, pp. 1–107, 2011. [Online]. Available: http://dx.doi.org/10.1561/1300000016
[8] M. Hayashi, “Capacity of Gaussian channels with energy harvesting and processing cost,” IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2563–2575, 2014.
[9] D. Shaviv and A. Özgür, “Online power control for block i.i.d. Bernoulli energy harvesting channels,” in Proc. IEEE Wireless Commun. and Networking Conference, San Francisco, CA, USA, Mar. 2017.
[10] ——, “Online power control for block i.i.d. energy harvesting channels,” in to be presented in IEEE GLOBECOM, Singapore, Dec. 2017.
[11] W. Mao and B. Hassibi, “On the capacity of a communication system with energy harvesting and a limited battery,” in Proc. IEEE Intl. Symp. Inf. Theory, Istanbul, Turkey, Jul. 2013, pp. 1789–1793.
[12] V. Jog and V. Anantharam, “An energy harvesting AWGN channel with a finite battery,” in Proc. IEEE Intl. Symp. Inf. Theory, Honolulu, HI, USA, Jun. 2014, pp. 806–810.
[13] D. Shaviv, P.-M. Nguyen, and A. Özgür, “Capacity of the energy harvesting channel with a finite battery,” IEEE Trans. Inf. Theory, vol. 62, no. 11, pp. 6436 – 6458, 2016.
[14] D. Shaviv and A. Özgür, “A communication channel with random battery recharges,” IEEE Trans. Inf. Theory, vol. 63, no. 5, 2017.
[15] S. L. Fong, V. Y. F. Tan, and J. Yang, “Non-asymptotic achievable rates for energy-harvesting channels using save-and-transmit,” IEEE J. Sel. Areas Commun., vol. 34, no. 12, pp. 3499 – 3511, 2016.
[16] K. G. Shenoy and V. Sharma, “Finite blocklength achievable rates for energy harvesting AWGN channels with infinite buffer,” in Proc. IEEE Intl. Symp. Inf. Theory, Barcelona, Spain, Jul. 2016, pp. 465 – 469.
[17] C. E. Shannon, “Certain results in coding theory for noisy channels,” Information and Control, vol. 1, pp. 6–25, 1957.
[18] M. Tomamichel and V. Y. F. Tan, “Second-order coding rates for channels with state,” IEEE Trans. Inf. Theory, vol. 60, no. 8, pp. 4427–4448, 2014.
[19] A. El Gamal and Y.-H. Kim, Network Information Theory. Cambridge, U.K.: Cambridge University Press, 2012.
[20] V. Y. Korolev and I. G. Shtesleva, “On the upper bound for the absolute constant in the Berry-Esseen inequality,” Theory of Probability and Its Applications, vol. 54, no. 4, pp. 638–658, 2010.
[21] 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, 2010.
[22] L. Wang, R. Colbeck, and R. Renner, “Simple channel coding bounds,” in Proc. IEEE Intl. Symp. Inf. Theory, Seoul, Korea, Jul. 2009, pp. 1804 – 1808.
[23] Y. Polyanskiy, “Channel coding: Non-asymptotic fundamental limits,” Ph.D. dissertation, Princeton University, 2010.
[24] V. Y. F. Tan, “Asymptotic estimates in information theory with non-vanishing error probabilities,” Foundations and Trends in Communications and Information Theory, vol. 11, no. 1-2, pp. 1–183, 2014.
[25] S. L. Fong and V. Y. F. Tan, “On the scaling exponent of polar codes for binary-input energy-harvesting channels,” IEEE J. Sel. Areas Commun., vol. 34, no. 12, pp. 3540 – 3551, 2016.