Achievable Rate for Energy Harvesting Channel with Finite Blocklength

Jing Yang
Department of Electrical Engineering,
University of Arkansas, Fayetteville, AR, 72701, U.S.A.

***Abstract*—This paper characterizes an achievable channel coding rate for a noiseless binary communication channel with an energy harvesting (EH) transmitter at a given blocklength n and error probability ϵ. As energy arrives randomly at the transmitter, codewords must obey the cumulative stochastic energy constraints. The coupling of the energy constraints on the symbols in a codeword makes the analysis fundamentally different from that of discrete memoryless channels. We first adopt a random coding scheme to construct the codebook with statistical information of the EH process. We then analyze the statistics of the corresponding output sequence. Specifically, we prove that the average number of mismatches between the input codeword and the output sequence scales as $O(\sqrt{n})$. Based on such characterization, we then propose a decoding scheme, and analyze the corresponding probability of decoding error. Finally, we explicitly characterize the maximum size of the length-n codebook generated by the random coding scheme in order to achieve the average probability of error ϵ. This leads to a lower bound on the maximum achievable channel coding rate for the EH communication channel. We show that the gap between the lower bound and the corresponding channel capacity under an equivalent average power constraint scales in $O(l \log n/\sqrt{n})$, where l is a constant depending on the error probability ϵ, and the statistics of the energy harvesting process.**

I. Introduction

We consider an energy harvesting communication system, where energy needed for communication is harvested by the transmitter during the course of communication. As energy arrives randomly at the transmitter, codewords must obey the cumulative stochastic energy constraints. The impact of the stochastic energy supply on the channel capacity has been characterized in the asymptotic regime under different assumptions on the battery size. With an unlimited battery, [1] shows that the capacity of the additive white Gaussian noise (AWGN) channel is equal to the capacity of the same channel under an average power constraint equal to the average recharge rate of the battery. The fluctuations in the energy arrivals are averaged out in the long run, which essentially reduces the sample path-wise energy constraints to a single average power constraint. When the battery-size is zero, [2] shows that an AWGN channel with EH status causally available at the transmitter can be modeled as a state-dependent channel, and its capacity can be achieved by using a Shannon's coding scheme [3]. With finite battery size, [4] investigates the capacity of a discrete memoryless channel (DMC). It is shown that the capacity can be described using the Verdu-Han general framework [5]. If the transmitted symbol only depends on the currently available energy, the system reduces to a finite state channel. A special case of the problem, i.e., the capacity of noiseless binary channel with binary energy arrivals and unit-capacity battery is discussed in [6]. It shows that the channel is equivalent to an additive geometric-noise timing channel with causal noise information available at the transmitter.

Different from previous work which focus on the identification of the capacity of energy harvesting channels in the asymptotic regime, in this paper, we aim to characterize the impact of stochastic energy constraints on the maximum achievable channel coding rate in the finite blocklength regime. The same problem without the stochastic energy constraints imposed by the energy harvesting process has drawn great attentions recently due to its practical importance [7]. Generally speaking, in the nonasymptotic regime, there is no exact formula for such characterization. [7] proves that in the finite blocklength regime, the backoff of the channel coding rate from channel capacity is parameterized by the *channel dispersion*, which measures the stochastic variability of the channel relative to a deterministic channel with the same capacity. However, the problem under stochastic energy constraints at the transmitter has not been considered before.

Our objective is to evaluate the impact of stochastic variability of the energy harvesting process on the channel capacity in the finite blocklength regime. Intuitively, in the EH scenario, the backoff from channel capacity is a function of the channel dispersion, as well as the characteristics of the EH process. In order to decouple the impact of energy variability from the channel variability on the coding rate, as a first step, we consider a deterministic channel without any variability. Further, we assume that the statistics of the EH process are available at the transmitter beforehand, and governs the codebook design; See Fig. 1. Such assumptions make our problem "dual" to that studied in [7]. Although the generation of the codebook is independent with the realization of the EH process, the transmission of symbols in a codeword is subject to the energy available at the transmitter, which implies that a portion of the codeword may be erased due to the stochastic energy supply. In the nonasymptotic regime, such erasures are not negligible, and inevitably lead to certain rate loss. Compared with the same channel with equal average power constraint, our purpose is to analytically characterize the rate loss caused by the fluctuations of the EH process.

For the analytical tractability, we assume the physical layer is a noiseless binary channel, and the energy harvesting process is i.i.d. from slot to slot. We assume the transmission

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

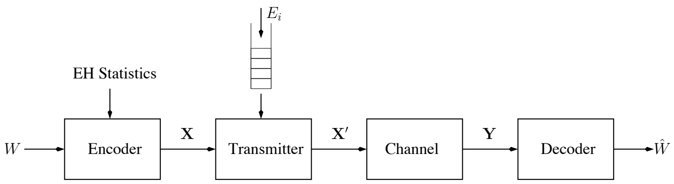

Fig. 1: EH communication system

of symbol 1 causes one unit of energy while the energy cost for the transmission of symbol 0 is zero. Assume the average EH rate is p, $0 < p < 1/2$. In order to match the capacity of the channel without energy variability in the asymptotically regime, we adopt the random coding strategy with symbols generated according to an i.i.d Bernoulli process with parameter p. We then analyze the statistics of the transmitted symbols for a random codeword. Due to the coupling of the energy constraints, erasures of the codeword during the transmission do not happen in an i.i.d fashion, and this makes the analysis complicated. Exploiting tools from stochastic processes and probability theory, we are able to explicitly characterize the distribution of the number of erasures during the transmission. We then propose a simple decoding strategy, based on which we explicitly calculate an lower bound on the maximize size of the codebook in order to achieve an average error probability ϵ with a finite blocklength n, which leads to a lower bound on the maximum achievable rate for the channel.

II. System Model and Problem Formulation

Consider a *noiseless* binary channel with an energy harvesting transmitter. When a channel input symbol, namely 0 or 1, is transmitted through the channel, the receiver gets the same symbol. We assume the energy cost for the channel input symbol 1 is one unit, while the energy cost for the channel input symbol 0 is zero.

Energy arrives at the battery at each channel use and is consumed for the transmission of channel input symbols. Let E_i be the amount of energy harvested in the i-th channel use. We assume E_is are identically and independently distributed (i.i.d) random variables with $E_i \in \mathbb{Z}_+$ and $\mathbb{E}[E_i] = p$. To make the problem nontrivial, we assume p is less than $1/2$. We assume the statistics of the EH process are available at the encoder beforehand, which utilizes such information to construct a length-n codebook.

Since the encoding step does not depend on the instantaneous battery level at each channel use, transmitting symbols in a codeword may not always be feasible. Let X_i be the intended input symbol at the i-th channel use, X'_i be the actual input symbol and Y_i be the output symbol in the same channel use. Then, when $X_i = 0$, $X'_i = Y_i = 0$; when $X_i = 1$ and there exists at least a unit of energy in the battery, $X'_i = Y_i = 1$; otherwise, $X'_i = Y_i = 0$, i.e., the corresponding symbol of the codeword is erased at the receiver.

Let B_i be the battery level at the beginning of the i-th channel use. We assume the size of the battery is sufficiently large such that energy overflow never happens. At each channel use, the transmitter first harvests energy and then transmits a symbol. Then, the battery level evolves according to

$$B_{i+1} = (B_i + E_i - X_i)^+, \quad i = 1, 2, \dots \tag{1}$$

where $(x)^+ = \max\{x, 0\}$. We assume $B_1 = 0$, i.e., the system starts with an empty state. This is the worse case scenario for the problem studied in this paper. Therefore, the results obtained here are valid for any finite B_1.

The encoder encodes M different messages into M length-n binary sequences, denoted as $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$. Let $\mathbf{y}$ be the output sequence from the channel corresponding to a codeword input. Due to the energy constraint imposed by the EH process, $\mathbf{y}$ is not equal to the input sequence in general. The decoder then maps $\mathbf{y}$ to one of the M messages. A decoding error happens if the decoder does not map $\mathbf{y}$ to the message correctly. Given a fixed blocklength n, and $\epsilon \in (0, 1)$, our objective is to characterize the maximum size of the codebook, denoted as $M^*(n,\epsilon)$, such that the average decoding error probability is upper bounded by ϵ.

III. Random Coding and Output Sequences

The achievability of our lower bound relies on a random coding scheme and a distance based decoding scheme. In this section, we first describe our random coding scheme and analyze the statistics of the corresponding output sequences of the channel under the stochastic energy constraint.

Specifically, let us generate M codewords with length n independently at random according to an i.i.d. Bernoulli process with parameter p. To send message $m \in \{1, 2, \dots, M\}$, the transmitter transmits $\mathbf{x}_m$. We are interested in characterizing the mismatch between the received sequence $\mathbf{y}$ and $\mathbf{x}_m$. We have the following theorem.

Theorem 1 *Let $\{X_i\}_{i=1}^{\infty}$ be an i.i.d Bernoulli process with parameter p, and $\{E_i\}_{i=1}^{\infty}$ be an i.i.d EH process with $\mathbb{E}[E_i] = p$, and $\mathbb{E}[\exp(\alpha E_i)]$ continuously differentiable in a small neighborhood of $\alpha = 0$. Let $\{Y_i\}_{i=1}^{\infty}$ be the corresponding output sequence, and $K_n = \sum_{i=1}^{n} |X_i - Y_i|$. Then, for almost all realizations of $\{X_i\}$ and $\{E_i\}$, we have*

$$\lim_{n \to +\infty} \frac{K_n}{n} = 0 \quad a.s.$$

The proof of Theorem 1 is provided in Appendix A. Due to the coupling of energy constraints on the input symbols, the output sequence $\{Y_i\}$ is no longer an i.i.d process. We first define a renewable process embedded in the output sequence, and then prove that the tails of its renewal frequency decay exponentially as n increases, which leads to the convergence in Theorem 1.

Theorem 1 indicates that for a very general class of EH processes, randomly generating a codeword according to its average EH rate leads to $o(n)$ mismatches between $\{X_i\}$ and $\{Y_i\}$. Intuitively, when n is sufficiently large, the receiver can still reliably decode the message regardless of those $o(n)$ mismatches. This is because a joint typicality based decoding strategy is based on the laws of large numbers, which are unaffected by $o(n)$ alterations, as n goes to infinity.

Besides the characterization of K_n in the asymptotic regime, we derive the characteristics of K_n for any finite n, which is summarized in the following theorem.

Theorem 2 *Let $V = Var(E_i) + Var(X_i)$. Then,*

$$\mathbb{P}[K_n \geq \gamma] \leq \exp\left(-\frac{c_1(2-c_1)}{2c_2 V n}\gamma^2\right)$$

where $c_1 \leq 1$ and $c_2 \geq 1$ are positive constants depending on the characteristic of $\{E_i\}$.

Corollary 1

$$\frac{\mathbb{E}[K_n]}{n} \leq \sqrt{\frac{\pi c_2}{2c_1(2-c_1)}\frac{V}{n}} \tag{2}$$

The proofs of Theorem 2 and Corollary 1 are given in Appendix B.

Theorem 2 indicates that for any finite blocklength n, the tails of the distribution of K_n decay super exponentially. The exponent depends on the EH process variability, as well as the randomness of symbols in the codeword. This is because both the energy arrivals and the input symbols are random, and the occurrence of erasures depends on both of them. Corollary 1 characterizes an upper bound on the expected portion of infeasible symbols as a function of the blocklength, which decays in $O(1/\sqrt{n})$. Since the average number of erased symbols in a length-n codeword is on the order of $\sqrt{n}$, in order to reliably distinguish two codewords at the receiver, intuitively, the Hamming distance between them should be in $O(\sqrt{n})$ as well. Such results will guide the codebook design and the associated decoding rules, as discussed in the next section.

IV. A Lower Bound on the Achievable Rate

Before we proceed to characterize the achievable rate based on random coding, we describe the decoding procedure at the receiver, and analyze the conditions required for a reliable decoding. The maximum size of the codebook is then determined to meet the average error probability requirements.

A. Decoding Rule

As we discussed in Section III, some symbols (1's) in the codebook may be erased during the transmission due to the energy constraints, which introduces uncertainty on the decoding. Define $\mathcal{S}_\mathbf{x} := \{i : \mathbf{x}[i] = 1\}$, i.e., the set of indices of symbol 1 in the sequence $\mathbf{x}$. Assume the m-th message is sent through the channel. Then, we have $\mathcal{S}_\mathbf{y} \subseteq \mathcal{S}_{\mathbf{x}_m}$.

However, due to the random erasures happening during the transmission, two different codewords may end up with the same output sequence, i.e., there may exist other codewords $\mathbf{x}_{m'} \neq \mathbf{x}_m$ satisfying $\mathcal{S}_\mathbf{y} \subseteq \mathcal{S}_{\mathbf{x}_{m'}}$. To mitigate decoding error caused by such cases, during the decoding, we restrict to the codewords that are sufficiently close to $\mathbf{y}$. Besides, in order to analytically track the probability of such decoding error, we focus on a subset of codewords, defined as follows.

Define

$$\mathcal{T}_\delta^n = \{\mathbf{x} : \mathbf{x} \in \{0,1\}^n, |\mathcal{S}_\mathbf{x} - np| \leq n\delta\}$$

i.e., $\mathcal{T}_\delta^n$ is the set of typical sequences under the δ-convention generated according to $Bernoulli(p)$ [8]. The value of δ will be decided later.

Then, the decoder declares that message $\hat{m}$ was sent if and only if there exists such a unique $\hat{m}$ satisfying

$$\mathbf{x}_{\hat{m}} \in \mathcal{T}_\delta^n$$

$$\mathcal{S}_\mathbf{y} \subseteq \mathcal{S}_{\mathbf{x}_{\hat{m}}} \tag{3}$$

$$|\mathcal{S}_{\mathbf{x}_{\hat{m}}} \backslash \mathcal{S}_\mathbf{y}| \leq \gamma \tag{4}$$

(3) is a necessary condition for $\mathbf{x}_{\hat{m}}$ to be the codeword just sent. The left hand side of (4) is the number of erasures if $\mathbf{x}_{\hat{m}}$ was sent. We bound it by a constant γ to ensure that the received codeword does not deviate much from the declared codeword. Based on the result in Corollary 1, intuitively, γ should be on the order of $\sqrt{n}$. The selection of γ, as well as the value of δ, determines the probability of error, and affects the codebook design.

B. Error Analysis

Assume $m = 1$ is sent. There are three error events:

$$\mathcal{E}_0 := \{\mathbf{X}_1 \notin \mathcal{T}_\delta^n\}$$

$$\mathcal{E}_1 := \{|\mathcal{S}_{\mathbf{X}_1} \backslash \mathcal{S}_\mathbf{Y}| > \gamma\}$$

$$\mathcal{E}_2 := \{\exists \tilde{m} \neq 1 : \mathcal{S}_\mathbf{Y} \subseteq \mathcal{S}_{\mathbf{X}_{\tilde{m}}}, |\mathcal{S}_{\mathbf{X}_{\tilde{m}}} \backslash \mathcal{S}_\mathbf{Y}| \leq \gamma\}$$

A decoding error happens if any of those events occurs, i.e., $\mathcal{E} := \mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_2$. We bound the probability of those error events as follows. First, we note that

$$\mathbb{P}[\mathcal{E}_0] = \mathbb{P}\left[\left|\sum_{i=1}^n X_i - np\right| \geq n\delta\right] \leq 2\exp\left(-2n\delta^2\right) \tag{5}$$

where (5) follows from the Hoeffding's inequality [9]. Next, $\mathbb{P}[\mathcal{E}_1]$ can be bounded directly according to Theorem 2. The analysis of $\mathbb{P}[\mathcal{E}_2]$ is a bit complicated, and we summarize our result as follows.

Lemma 1 *Let $\bar{p} = 1 - p$. Then,*

$$\mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_0^c \cap \mathcal{E}_1^c] \leq (M-1)\exp\Bigg(-nH(p) - n\delta\log(p\bar{p}) - \gamma\log\frac{\gamma^2}{n^2(p+\delta)(\bar{p}+\delta)} + 2\gamma + O\left(\frac{\gamma^2}{n(p-\delta)} + \frac{\gamma^2}{n(\bar{p}-\delta)}\right)\Bigg)$$

The proof of Lemma 1 is provided in Appendix C. Since $\mathbb{P}[\mathcal{E}] = \mathbb{P}[\mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_2] \leq \mathbb{P}[\mathcal{E}_0] + \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_0^c \cap \mathcal{E}_1^c]$, with each individual error probability given, we are able to derive the conditions under which the average probability of error is upper bounded by ϵ.

C. The Lower Bound

In order to make sure that average probability of error is less than ϵ, we need to select the parameters γ, δ, as well as the size of the codebook M carefully. In order to bound

$\mathbb{P}[\mathcal{E}_0]$, it requires that $\delta = \Omega(1/\sqrt{n})$. Similarly, to bound $\mathbb{P}[\mathcal{E}_1]$, we must have $\gamma = \Omega\ (\sqrt{n})$. We choose $\delta = \sqrt{\frac{\log n}{n}}$, $\gamma = l\sqrt{n}$, where $l = \sqrt{-\frac{2c_2 V \log(\epsilon/3)}{c_1(2-c_1)}}$. Then, we have the following lower bound.

Theorem 3 *Let $M^*(n,\epsilon)$ denote the maximum size of a length-n code over the energy harvesting noiseless binary channel with average probability of error $\epsilon \in (0,1)$. Then,*

$$\frac{1}{n}\log M^*(n,\epsilon) \geq H(p) - l\frac{\log n}{\sqrt{n}} + O\left(\sqrt{\frac{\log n}{n}}\right)$$

The proof of Theorem 3 is provided in Appendix D.

We note that as n goes to infinity, $\frac{1}{n}\log M^*(n,\epsilon)$ approaches $H(p)$, which is exactly the channel capacity of the noiseless binary channel under an average power constraint p. This coincides with the intuition that as the fluctuations of the energy arrivals are averaged out when the blocklength goes to infinity, and the energy constraints on individual symbols essentially become an average power constraint.

In the finite blocklength regime, the backoff of the maximum channel coding rate from the channel capacity $H(p)$ scales in $O(l\frac{\log n}{\sqrt{n}})$, where l is a function of ϵ and the statistics of the EH process. We point out that as the variance of E_i increases, the backoff increases as well, which implies that the EH variability impacts the maximum achievable rate negatively.

APPENDIX

A. Proof of Theorem 1

Based on B_i, we define a renewal process, which renews whenever $B_i + E_i - X_i = -1$. Define $\tau_1, \tau_2, \ldots$ as the lengths of the corresponding renewal intervals. Within each renewal interval, $B_i + E_i - X_i$ follows a random walk starting with state 0 and finishes when it hits state -1 for the first time. Define $D := E_i - X_i$, and

$$\Lambda(\alpha) = \log \mathbb{E}[\exp(-\alpha D)] \tag{6}$$

Then, $\Lambda(\alpha)$ is continuously differentiable in a small neighborhood of $\alpha = 0$. We note that $\Lambda(0) = 0$ and the Taylor expansion of $\Lambda(\alpha)$ around 0 equals

$$\Lambda(\alpha) = \frac{V}{2}\alpha^2 + o(\alpha^2) \tag{7}$$

where $V := \mathbb{E}[D^2] = \mathrm{Var}(E_i) + \mathrm{Var}(X_i)$.

Consider a "random walk" $\{\Omega_k\}_{k=0}^{\infty}$, which starts with 0 and increments with D. Denote the first -1-hitting time as κ. Then, $\Omega_0 = 0, \Omega_\kappa = -1$.

Define a Martingale process as $\{\exp(-\alpha\Omega_k - \Lambda(\alpha)k)\}_{k=0}^{\infty}$, where $\alpha > 0$ and $\Lambda(\alpha)$ is defined in (6). Based on the property of Martingale processes, we have

$$\begin{aligned}&\mathbb{E}\left[\exp(-\alpha\Omega_\kappa - \Lambda(\alpha)\kappa)\right] \qquad (8)\\ &= \mathbb{E}[\mathbb{E}\ldots[\mathbb{E}[\exp(-\alpha\Omega_\kappa - \Lambda(\alpha)\kappa)|\Omega_{\kappa-1}]\ldots]|\Omega_0]] = 1\end{aligned}$$

On the other hand, we have

$$\begin{aligned}&\mathbb{E}\left[\exp(-\alpha\Omega_\kappa - \Lambda(\alpha)\kappa)\right]\\ &= \mathbb{E}\left[\mathbf{1}_{\kappa<\infty} \cdot \exp(-\alpha\Omega_\kappa - \Lambda(\alpha)\kappa)\right]\end{aligned}$$

where $\mathbf{1}_x$ is an indicator function, and it equals one when x is true. Let $\alpha \to 0^+$, then based on (7) and $\Omega_\kappa = -1$, we have

$$1 = \mathbb{E}\left[I_{\kappa<\infty}\right] = \mathbb{P}\left[\kappa < \infty\right]$$

i.e., the probability of hitting -1 in finite time is 1.

(8) implies that

$$\mathbb{E}\left[\exp(-\Lambda(\alpha)\kappa)\right] = \exp(-\alpha).$$

Let K_n be the total number of renewal intervals up to time slot n. For such K_n i.i.d random walks with -1-hitting times τ_i,

$$\mathbb{E}\left[\exp\left(-\Lambda(\alpha)\left(\sum_{i=1}^{K_n}\tau_i\right)\right)\right] = \exp(-K_n\alpha),$$

Therefore, for any $\epsilon > 0$, we have

$$\begin{aligned}&\mathbb{P}\left[\frac{K_n}{\sum_{i=1}^{K_n}\tau_i} > \epsilon\right] = \mathbb{P}\left[\sum_{i=1}^{K_n}\tau_i < \frac{K_n}{\epsilon}\right]\\ &\leq \frac{\exp(-K_n\alpha)}{\exp(-\Lambda(\alpha)\frac{K_n}{\epsilon})} = \exp\left(-K_n\left(\alpha - \frac{\Lambda(\alpha)}{\epsilon}\right)\right) \qquad (9)\end{aligned}$$

Since $\Lambda(\alpha) = O(\alpha^2)$, $\Lambda(0) = 0$, then, for any $\epsilon > 0$, we can always find a sufficiently small α_ϵ to have $\alpha_\epsilon - \frac{\Lambda(\alpha_\epsilon)}{\epsilon} > 0$, which ensures that the probability decays exponentially in K_n. According to Borel-Cantelli lemma [10], we have

$$\lim_{n\to\infty} \frac{K_n}{\sum_{i=1}^{K_n}\tau_i} = 0, \quad a.s.$$

Since

$$0 \leq \frac{K_n}{n} \leq \frac{K_n}{\sum_{i=1}^{K_n}\tau_i},$$

we have $\lim_{n\to\infty}\frac{K_n}{n} = 0$ almost surely.

B. Proofs of Theorem 2 and Corollary 1

Under the assumption that $\mathbb{E}[\exp(\alpha E_i)]$ is continuously differentiable in a small neighborhood of $\alpha = 0$, there must exist a positive constant $c_1 \leq 1$ such that $\Lambda(\alpha)$ is continuously differentiable for $\alpha \in [0, \frac{c_1}{V}]$, and another constant $c_2 \geq 1$ such that $\Lambda(\alpha) \leq \frac{c_2 V}{2}\alpha^2$ for $\alpha \in [0, \frac{c_1}{V}]$. Let $\alpha_\gamma = \frac{\gamma}{n}\frac{c_1}{c_2 V}$. Then,

$$\begin{aligned}\mathbb{P}[K_n \geq \gamma] &= \mathbb{P}[\tau_1 + \tau_2 + \ldots + \tau_\gamma \leq n]\\ &\leq \exp(-\gamma\alpha_\gamma + \Lambda\left(\alpha_\gamma\right)n) \qquad (10)\\ &\leq \exp\left(-\gamma\alpha_\gamma + \frac{c_2 V}{2}\alpha_\gamma^2 n\right) \qquad (11)\\ &= \exp\left(-\frac{c_1(2-c_1)}{2c_2 V n}\gamma^2\right)\end{aligned}$$

where (10) follows from (9) by replacing K_n, ϵ with γ and n/γ, respectively, and (11) follows from the assumption that $\Lambda(\alpha) \leq \frac{c_2 V}{2}\alpha^2$. The proof of Theorem 2 is complete.

To prove Corollary 1, we note

$$\mathbb{E}[K_n] = \sum_{i=1}^{n} \mathbb{P}[K_n \geq i] \leq \sum_{i=1}^{n} \exp\left(-\frac{c_1(2-c_1)}{2c_2Vn} i^2\right)$$
$$\leq \int_0^\infty \exp\left(-\frac{c_1(2-c_1)}{2c_2Vn} x^2\right) dx \quad (12)$$
$$= \sqrt{\frac{\pi c_2 V}{2c_1(2-c_1)} n} \quad (13)$$

where (12) follows from the fact that $\exp\left(-i^2\right) \geq \exp\left(-x^2\right)$ for $x \in [i-1, i]$. The result then follows by dividing both sides of (13) by n.

C. Proof of Lemma 1

First, based on the union bound, we have

$$\mathbb{P}[\mathcal{E}_2] \leq (M-1)\mathbb{P}[\mathcal{S}_{\mathbf{Y}} \subseteq \mathcal{S}_{\mathbf{X}_2}, |\mathcal{S}_{\mathbf{X}_2} \backslash \mathcal{S}_{\mathbf{Y}}| \leq \gamma] \quad (14)$$

Thus,

$$\begin{aligned}
&\mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_1^c] \\
&\leq (M-1)\mathbb{P}[K_n \leq \gamma, \mathcal{S}_{\mathbf{Y}} \subseteq \mathcal{S}_{\mathbf{X}_2}, |\mathcal{S}_{\mathbf{X}_2} \backslash \mathcal{S}_{\mathbf{Y}}| \leq \gamma] \\
&\leq (M-1)\mathbb{P}[|\mathcal{S}_{\mathbf{X}_1} \backslash \mathcal{S}_{\mathbf{X}_2}| \leq \gamma, |\mathcal{S}_{\mathbf{X}_2} \backslash \mathcal{S}_{\mathbf{X}_1}| \leq \gamma] \\
&= (M-1) \sum_{\mathbf{x} \in \{0,1\}^n} \mathbb{P}[|\mathcal{S}_{\mathbf{X}_1} \backslash \mathcal{S}_{\mathbf{X}_2}| \leq \gamma, \\
&\qquad |\mathcal{S}_{\mathbf{X}_2} \backslash \mathcal{S}_{\mathbf{X}_1}| \leq \gamma | \mathbf{X}_1 = \mathbf{x}] \cdot \mathbb{P}[\mathbf{X}_1 = \mathbf{x}] \\
&= (M-1) \sum_{\mathbf{x} \in \{0,1\}^n} \Big(\mathbb{P}[|\mathcal{S}_{\mathbf{X}_1} \backslash \mathcal{S}_{\mathbf{X}_2}| \leq \gamma | \mathbf{X}_1 = \mathbf{x}] \\
&\qquad \cdot \mathbb{P}[|\mathcal{S}_{\mathbf{X}_2} \backslash \mathcal{S}_{\mathbf{X}_1}| \leq \gamma | \mathbf{X}_1 = \mathbf{x}]\Big) \cdot \mathbb{P}[\mathbf{X}_1 = \mathbf{x}] \quad (15)
\end{aligned}$$

where (15) follows from the fact that when $\mathbf{X}_1$ is fixed, the distribution of $\mathbf{X}_2$ over $\mathcal{S}_{\mathbf{X}_1}$ and $\mathcal{S}_{\mathbf{X}_1}^c$ are independent.

Next, we have

$$\mathbb{P}[|\mathcal{S}_{\mathbf{X}_1} \backslash \mathcal{S}_{\mathbf{X}_2}| \leq \gamma | \mathbf{X}_1 = \mathbf{x}] = \mathbb{P}\left[\sum_{i \in \mathcal{S}_{\mathbf{x}}} \mathbf{X}_2[i] \geq |\mathcal{S}_{\mathbf{x}}| - \gamma\right]$$
$$\leq \exp\left\{-|\mathcal{S}_{\mathbf{x}}| D\left(1 - \frac{\gamma}{|\mathcal{S}_{\mathbf{x}}|} \middle\| p\right)\right\} \quad (16)$$

where $D(x\|y) = x \log\left(\frac{x}{y}\right) + (1-x)\log\left(\frac{1-x}{1-y}\right)$, and (16) follows from the Chernoff-Hoeffding theorem [9]. Similarly,

$$\mathbb{P}[|\mathcal{S}_{\mathbf{X}_2} \backslash \mathcal{S}_{\mathbf{X}_1}| \leq \gamma | \mathbf{X}_1 = \mathbf{x}] \leq \exp\left\{-|\mathcal{S}_{\mathbf{x}}^c| D\left(1 - \frac{\gamma}{|\mathcal{S}_{\mathbf{x}}^c|} \middle\| \bar{p}\right)\right\}$$

By Taylor expansion, we have

$$D(1-x\|y) = -\log y + x \log x + x\left[\log\left(\frac{y}{1-y}\right) - 1\right] + O(x^2)$$

Therefore,

$$- |\mathcal{S}_{\mathbf{x}}| D\left(1 - \frac{\gamma}{|\mathcal{S}_{\mathbf{x}}|} \middle\| p\right)$$
$$= |\mathcal{S}_{\mathbf{x}}| \log p - \gamma \log \frac{\gamma}{|\mathcal{S}_{\mathbf{x}}|} - \gamma\left(\log \frac{p}{\bar{p}} - 1\right) + O\left(\frac{\gamma^2}{|\mathcal{S}_{\mathbf{x}}|}\right)$$

$$- |\mathcal{S}_{\mathbf{x}}^c| D\left(1 - \frac{\gamma}{|\mathcal{S}_{\mathbf{x}}^c|} \middle\| \bar{p}\right)$$
$$= |\mathcal{S}_{\mathbf{x}}^c| \log \bar{p} - \gamma \log \frac{\gamma}{|\mathcal{S}_{\mathbf{x}}^c|} - \gamma\left(\log \frac{\bar{p}}{p} - 1\right) + O\left(\frac{\gamma^2}{|\mathcal{S}_{\mathbf{x}}^c|}\right)$$

In addition, based on the definition, for $\mathbf{x} \in \mathcal{T}_\delta^n$, we have

$$n(p-\delta) \leq |\mathcal{S}_{\mathbf{x}}| \leq n(p+\delta)$$
$$n(\bar{p}-\delta) \leq |\mathcal{S}_{\mathbf{x}}^c| \leq n(\bar{p}+\delta)$$

The proof can thus be finished by putting those pieces together.

D. Proof of Theorem 3

Let $\gamma = l\sqrt{n}$, $\delta = \sqrt{\frac{\log n}{n}}$. Then, we have $\mathbb{P}[\mathcal{E}_0] = 2/n^2$, $\mathbb{P}[\mathcal{E}_1] = \exp\left(-\frac{c_1(2-c_1)l^2}{2c_2V}\right)$, and

$$\mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_0^c \cap \mathcal{E}_1^c] \leq (M-1)\exp\big(-nH(p) + l\sqrt{n}\log n + c_3\sqrt{n \log n}\big)$$

where c_3 is some positive constant. Let

$$M = \exp\left(nH(p) - l\sqrt{n}\log n - (c_3+1)\sqrt{n\log n}\right) \quad (17)$$

Then, $\mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_0^c \cap \mathcal{E}_1^c] \leq \exp(-\sqrt{n \log n})$. When n is sufficiently large, we have $\mathbb{P}[\mathcal{E}_0] < \epsilon \not\beta$, $\mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_0^c \cap \mathcal{E}_1^c] \leq \epsilon/3$. Let $l = \sqrt{-\frac{2c_2V\log(\epsilon/3)}{c_1(2-c_1)}}$, then $\mathbb{P}[\mathcal{E}_1] < \epsilon \not\beta$. In summary, we have $\mathbb{P}[\mathcal{E}] = \mathbb{P}[\mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_2] \leq \epsilon$.

Therefore, for all sufficiently large blocklength n and $\epsilon \in (0,1)$, we can ways achieve a channel coding rate described in Theorem 3 through the proposed random coding and decoding scheme, which provides a lower bound for the maximum achievable rate of the EH channel.

References

[1] O. Ozel and S. Ulukus, "Achieving AWGN capacity under stochastic energy harvesting," *IEEE T ransactions on Information Theory*, vol. 58, no. 10, pp. 6471–6483, 2012.

[2] ——, "AWGN channel under time-varying amplitude constraints with causal information at the transmitter," *45th Asilomar Conference on Signals, Systems and Computers*, November 2011.

[3] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Dev.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.

[4] W. Mao and B. Hassibi, "On the capacity of a communication system with energy harvesting and a limited battery," in *IEEE International Symposium on Information Theory*, 2013, pp. 1789–1793.

[5] S. Verdu and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theor.*, vol. 40, no. 4, pp. 1147–1157, Sep. 2006.

[6] K. Tutuncuoglu, O. Ozel, A. Yener, and S. Ulukus, "Binary energy harvesting channel with finite energy storage," in *IEEE International Symposium on Information Theory*, 2013, pp. 1591–1595.

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

[8] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic Press, Inc., 1982.

[9] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, March 1963.

[10] A. Klenke, *Probability Theory: A Comprehensive Course (Universitext)*. Springer, 2007.