

The Binary Energy Harvesting Channel With a Unit-Sized Battery

Kaya Tutuncuoglu, Omur Ozel, *Member, IEEE*, Aylin Yener, *Fellow, IEEE*, and Sennur Ulukus, *Fellow, IEEE*

Abstract—We consider a binary energy harvesting communication channel with a finite-sized battery at the transmitter. In this model, the channel input is constrained by the available energy at each channel use, which is driven by an external energy harvesting process, the size of the battery, and the previous channel inputs. We consider an abstraction where energy is harvested in binary units and stored in a battery with the capacity of a single unit, and the channel inputs are binary. Viewing the available energy in the battery as a state, this is a state-dependent channel with input-dependent states, memory in the states, and causal state information available at the transmitter only. We find an equivalent representation for this channel based on the timings of the symbols, and determine the capacity of the resulting equivalent timing channel via an auxiliary random variable. We present achievable rates based on certain selections of this auxiliary random variable, which resemble lattice coding for the timing channel. We develop upper bounds for the capacity by using a genie-aided method, and also by quantifying the leakage of the state information to the receiver. We show that the proposed achievable rates are asymptotically capacity achieving for small energy harvesting rates. We extend the results to the case of ternary channel inputs. We numerically observe that our achievable rates are notably close to the upper bounds, and outperform basic Shannon strategies that only consider instantaneous battery states, for all parameter values.

Index Terms—Shannon capacity, energy harvesting, timing channel, state dependent channels.

I. INTRODUCTION

WE CONSIDER an energy harvesting communication channel, where the transmitter harvests energy from an exogenous source to sustain power needed for its data transmission. The transmitter stores harvested energy in a

Manuscript received August 27, 2014; revised November 16, 2015 and October 14, 2016; accepted March 11, 2017. Date of publication April 5, 2017; date of current version June 14, 2017. This work was supported by NSF under Grant CNS 09-64364/CNS 09-64632 and Grant CCF 14-22347/CCF 14-22111. This paper was presented in part at the 2013 IEEE International Symposium on Information Theory and the 2014 IEEE International Symposium on Information Theory. (*Corresponding author: Sennur Ulukus.*)

K. Tutuncuoglu was with the Department of Electrical Engineering, Pennsylvania State University, University Park, PA 16802 USA. He is now with Facebook, Menlo Park, CA 94025 USA (e-mail: kayatutun@gmail.com).

O. Ozel was with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA. He is now with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: omurozel@ieee.org).

A. Yener is with the Department of Electrical Engineering, Pennsylvania State University, University Park, PA 16802 USA (e-mail: yener@ee.psu.edu).

S. Ulukus is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: ulukus@umd.edu).

Communicated by D. Tuninetti, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2017.2690997

finite-sized battery, and each channel input is constrained by the remaining energy in the battery. Consequently, stored energy can be viewed as the state of this channel, which is naturally known causally at the encoder, but unknown at the decoder. This state is correlated over time, and is driven by the exogenous energy harvesting process, energy storage capacity of the battery, and the past channel inputs. As such, this channel model introduces unprecedented constraints on the channel input, departing from traditional channels with average or peak power constraints, and requires new approaches to determine its capacity.

References [1]–[5] study the capacity of channels with energy harvesting transmitters with an infinite-sized battery [1], with no battery [2], and with a finite-sized battery [3]–[6]. Reference [1] shows that the capacity of an AWGN with an infinite-sized battery at the energy harvesting encoder is equal to the capacity with an average power constraint equal to the average recharge rate. This reference proposes save-and-transmit and best-effort-transmit schemes, both of which are capacity achieving when the battery size is unbounded. At the other extreme, [2] studies the case with no battery, which is equivalent to a time-varying stochastic amplitude-constrained channel. Reference [2] views harvested energy as a causally known state, and combines the results of Shannon on channels with causal state at the transmitter [7] and Smith on amplitude constrained channels [8]. More recent work [3]–[6] consider models with finite-sized battery. Reference [3] provides a multi-letter capacity expression that is hard to evaluate, since it requires optimizing multi-letter Shannon strategies [7] for each channel use. The authors conjecture that instantaneous Shannon strategies are optimal for this case, i.e., strategies that only observe the current battery state to determine the channel input are sufficient to achieve the capacity. Reference [4] finds approximations to the capacity of the energy harvesting channel within a constant gap of 2.58 bits/channel use, with more general results presented in [6]. For a deterministic energy harvesting profile, [5] provides a lower bound on the capacity by exploiting the volume of energy-feasible input vectors.

We consider a single-user communication scenario with an energy harvesting encoder that has a finite-sized battery, as shown in Fig. 1. In each channel use, the encoder harvests energy that is a multiple of a fixed unit, and stores it in a battery which has a capacity that is also a multiple of this unit. Each channel input then consumes an integer number of units of energy. In this paper, we consider the binary version of this setting, which we refer to as the binary energy harvesting

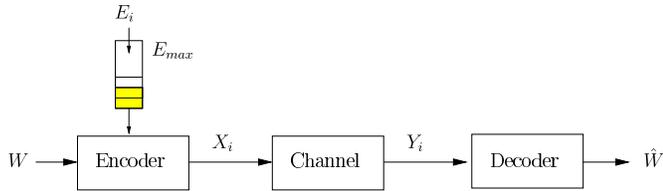


Fig. 1. The binary energy harvesting channel (BEHC) with an energy harvesting encoder and a finite-sized battery.

channel (BEHC). In a BEHC, energy is harvested in binary amounts (0 or 1 unit), the battery has unit size, and the channel inputs are binary. Sending a 1 through the channel requires one unit of energy per channel use, while sending a zero is free in terms of energy. Hence, the encoder can send a 1 only when it has the required energy in the battery, and a 0 anytime. A similar abstraction of communicating with energy packets over an interactive link can be found in [9].

In an energy harvesting channel, the channel input in each channel use is constrained by the battery state of the transmitter. Since the battery is at the transmitter, this state is naturally causally available at the encoder, but is not available at the decoder. This results in a channel with causally known state information at the encoder. In such channels, if the state is independent and identically distributed (i.i.d.) over time, and is independent of the channel inputs, then the capacity is achieved using Shannon strategies [7]. However, in the BEHC, the battery state has memory since the battery stores the energy through channel uses. Further, the evolution of the battery state depends on the past channel inputs since different symbols consume different amounts of energy. Therefore, Shannon strategies of [7] are not necessarily optimal for this channel. This channel model resembles the model of [10] with action dependent states, where the encoder controls the state of the channel through its own actions. However, different from [10], in the case of BEHC, actions and channel inputs are equal, i.e., the two cannot be chosen independently. This yields a conflict between choosing inputs with the purpose of communicating, and with the purpose of controlling the state.

In this paper, we consider a special case of the BEHC with no channel noise. Even in this special case, finding the capacity is challenging due to the memory in the state, the lack of battery state information at the receiver, and the inter-dependence of the battery state and the channel inputs. In essence, the uncertainty that causes the communication errors in this model is not due to the channel, but due to the random energy harvests and the battery state that impose intricate constraints on the channel inputs. For this BEHC, we first propose achievable rates using Shannon strategies in [7]. Next, we develop an equivalent representation for the channel in terms of the time differences between consecutive 1s sent through the channel. This is analogous to the timing channel in [11], or its discrete-time versions [12], [13], where the message is encoded in the arrival times of packets to a queue. Observing that the states are i.i.d. in this equivalent representation, we find a single-letter expression for the capacity of the BEHC by combining approaches from [7] and [11]. This expression is difficult to evaluate due to an involved

auxiliary random variable. We give achievable rates based on certain selections of this auxiliary random variable which resemble lattice coding for the timing channel. We develop upper bounds for the capacity by using a genie-aided method, and also by quantifying the leakage of the state information to the receiver. We find that our bounds are tight asymptotically as energy harvesting rate goes to zero. We extend our results to the case of ternary channel inputs. We numerically evaluate the achievable rates and the upper bounds and observe that our achievable schemes perform notably close to the upper bounds. We observe that the proposed timing channel based achievable schemes outperform Shannon strategies that consider only instantaneous battery state [3], for all parameter values, for this noiseless binary case.

II. CHANNEL MODEL

We consider the binary channel with an energy harvesting transmitter shown in Fig. 1. The battery at the transmitter is of size E_{max} . The harvested energy is first stored in the battery before being used for transmission. The encoder transmits a symbol $X_i \in \{0, 1\}$ in channel use i . At each channel use, the channel input X_i is constrained by the energy available in the battery at that channel use. Hence, for the transmitter to send an $X_i = 1$, it must have a unit of energy in the battery; the transmitter can send an $X_i = 0$ anytime. Next, the encoder harvests an energy unit with probability q , i.e., E_i is Bernoulli(q), and stores it in its battery of size E_{max} units. The harvests are i.i.d. over time. If the battery is full, harvested energy is lost, i.e., E_i cannot be used immediately in the same time slot without storing. We refer to this particular sequence of events within a channel use as the *transmit first* model, since the encoder first sends X_i and then harvests energy E_i .

The battery state S_i denotes the number of energy units available in the battery at the beginning of channel use i , and evolves as

$$S_{i+1} = \min\{S_i - X_i + E_i, E_{max}\} \quad (1)$$

where $X_i = 0$ if $S_i = 0$ due to the energy constraint. We assume that the battery is initially empty, i.e., $S_0 = 0$ with probability one. As shown in [6, Proposition 1], the initial battery level does not alter the capacity. Therefore this assumption does not lead to any loss of optimality. The encoder knows the battery state S_i causally,¹ i.e., at the beginning of time slot i , but does not know what E_i or S_{i+1} will be until after sending X_i . The decoder is unaware of the energy harvests at the encoder, and therefore the battery state. As seen from (1), the battery state S_i has memory, is affected by the channel inputs X_j for $j \leq i$, and imposes a constraint on the channel input X_i . In this work, we focus on the case of a unit-sized battery, i.e., $E_{max} = 1$, and a noiseless channel, i.e., $Y_i = X_i$.

It is worthwhile to reiterate that the capacity of a finite battery energy harvesting channel is a challenging problem. The capacity analysis requires the general capacity formula

¹Note that using S_i , S_{i-1} , and X_{i-1} , the encoder can calculate E_{i-1} as long as the battery has room for the harvested energy. As such, our model is identical to knowing E_i strictly causally at the transmitter whenever it can be stored. The alternative case of causal E_i knowledge at the transmitter is an interesting extension.

in [14] since the channel is not necessarily information stable. This is difficult to evaluate unless certain assumptions and conjectures about the channel are established [3]. Instead, in this work, we consider a simple channel model for which we can make more progress and find a capacity expression.

III. ACHIEVABLE RATES WITH SHANNON STRATEGIES

For a channel with i.i.d. and causally known states at the transmitter, Shannon shows in [7] that the capacity is achieved by code functions, which are since termed Shannon strategies. In particular, the codebook consists of i.i.d. strategies $U_i \in \mathcal{U}$, which are functions from channel state S_i to channel input X_i . In channel use i , the encoder observes S_i and puts $X_i = U_i(S_i)$ into the channel. The capacity of this channel is given by

$$C_{CSIT} = \max_{p_U} I(U; Y) \quad (2)$$

where p_U is the distribution of U over all functions from S_i to X_i .

In the BEHC, the state of the channel, i.e., the battery state of the encoder, is not i.i.d. over time. Therefore, (2) does not yield the capacity for this system. To overcome the memory in the state, [3] uses strategies that are functions of all past battery states to express the capacity in a multi-letter form. However, since the dimension of such strategies grow exponentially with the number of channel uses, this approach is intractable. Alternatively, it is possible to use the method in [7] to develop encoding schemes based on Shannon strategies to obtain achievable rates. One tractable such scheme is obtained when strategies are functions of the current battery state only, which is proposed as an achievable rate in [3] and [15]; and is conjectured to be capacity achieving in [3]. In this section, we consider such encoding schemes, mainly as a basis for comparison for the proposed timing channel based achievable scheme which will be presented in Section IV.

Specifically, for the $E_{max} = 1$ case, we have two states, $S_i \in \{0, 1\}$. We denote a strategy U as $U = (X, X')$, where $U(0) = X$ and $U(1) = X'$, i.e., X is the channel input when $S = 0$ and X' is the channel input when $S = 1$. Due to the inherent energy constraint of the BEHC, $X = 1$ requires $S = 1$, and thus, we have two feasible strategies, namely $(0, 0)$ and $(0, 1)$. We denote $U = (0, 0)$ as 0 and $U = (0, 1)$ as 1.

Given this setup, we consider two strategies that we term *optimal i.i.d. Shannon strategies (NIID)* and *naïve i.i.d. Shannon strategies (OIID)*, which yield the achievable rates

$$R_{NIID} = \max_{p \in [0, 1]} H_2\left(\frac{pq}{p+q-pq}\right) - pH_2\left(\frac{q}{p+q-pq}\right) \quad (3)$$

$$R_{OIID} = \max_{p \in [0, 1]} \lim_{n \rightarrow \infty} \frac{1}{n} I(U^n; Y^n) \quad (4)$$

respectively, where $H_2(p) = -p \log(p) - (1-p) \log(1-p)$ is the binary entropy function. For details of these schemes, see Appendix A.

While the NIID scheme permits an easy analysis, it fails to make use of the memory in the channel. Instead, the decoder can exploit the memory in the OIID scheme. On the other hand, calculating the limit of the n -letter mutual information $I(U^n; Y^n)$ in the OIID scheme is challenging. For this

calculation, we use the message passing algorithm proposed in [16]. This algorithm requires that the joint probability $p(y_i, u_i, s_{i+1}|s_i)$ is time invariant, i.e., independent of the channel index i . In our case, we have i.i.d. u_i , which yields

$$p(y_i, u_i, s_{i+1}|s_i) = p(y_i, s_{i+1}|u_i, s_i)p(u_i) \quad (5)$$

where $p(y_i, s_{i+1}|u_i, s_i)$ is independent of i by the definition of the channel. Thus, we can use the algorithm in [16] to exhaustively search p and solve (4).

We note that it is possible to further improve achievable rates by constructing more involved codebooks. For example, [3] considers generating codewords with Markov processes, which introduces additional memory to the system through the codewords. This approach improves the achievable rate as shown in [3] at the cost of increased computational complexity in the Markov order of the codebook. We evaluate and compare these achievable rates in Section IX.

IV. TIMING REPRESENTATION OF THE BEHC

In this section, we propose an alternative representation of the BEHC, which yields a simpler analysis via a single-letter expression for the capacity. In particular, we equivalently represent channel outputs Y_i with the number of channel uses between instances of $Y_i = 1$. We show that this transformation eliminates the memory in the state of the system, and allows constructing tractable achievable rates and upper bounds for the BEHC.

The input X_i and the output Y_i of the noiseless BEHC are both binary. Let $T_1 \in \{1, 2, \dots\}$ be defined as the number of channel uses before the first instance of output $Y = 1$, and $T_k \in \{1, 2, \dots\}$ for $k \geq 2$ be defined as the number of channel uses between the $(k-1)$ th instance of output $Y = 1$ and the k th instance of output $Y = 1$. In other words, the sequence T^m represents the differences between the channel uses where 1s are observed at the output of the channel. Clearly, T^m and Y^n are equivalent since there is a unique sequence T^m corresponding to each Y^n and vice versa.

When a 1 is transmitted in the i th channel use, the entire energy stored in the unit-sized battery of the encoder is consumed. Hence, the encoder cannot transmit another 1 until another energy unit is harvested. We define the idle time $Z_k \in \{0, 1, \dots\}$ of the encoder as the number of channel uses the encoder waits for energy after the $(k-1)$ st 1 is transmitted. For $k = 1$, since initial battery state is zero, this corresponds to the number of channel uses the encoder waits for the first energy harvest, and therefore has the same distribution as $k = 2, 3, \dots$. Since the probability of harvesting an energy unit is distributed i.i.d. with Bernoulli(q), Z_k is also i.i.d. and distributed geometric(q) on $\{0, 1, \dots\}$. Note that during the idle period, the encoder cannot send any 1s. Once the energy is harvested, the encoder observes Z_k and chooses to wait $V_k \in \{1, 2, \dots\}$ channel uses before sending the next 1. Hence, we have a *timing channel* with causally known state Z_k , channel input V_k , and channel output T_k , satisfying

$$T_k = V_k + Z_k \quad (6)$$

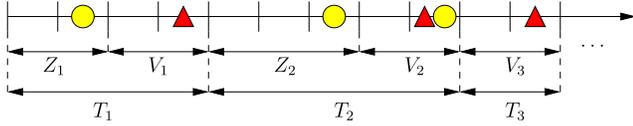


Fig. 2. Graphical representation of T_k , V_k and Z_k . Note that since energy is harvested immediately after sending a 1, we have $Z_3 = 0$.

We illustrate the variables T_k , V_k and Z_k in Fig. 2. In slots representing one use of the BEHC, an energy arrival, i.e., $E_i = 1$, is marked with a circle and sending a 1, i.e., $X_i = 1$, is marked with a triangle. Note that one use of the timing channel spans T uses of the BEHC.

We remark that the timing channel constructed from the time difference between consecutive 1s resembles the noiseless channel in [17] with the addition of varying symbol durations. The symbol durations in [17] are fixed, while the symbol durations in our model depend on the energy harvesting process, and therefore may change each time a symbol is sent. Hence, while [17] studies the problem of packing the most information within a given block length, our problem is also concerned with the randomness introduced by energy harvesting. In this sense, the timing channel defined here is analogous to the telephone signaling channel in [11] and its discrete time counterparts in [12] and [13], with the exception of causal knowledge of Z_k at the encoder in our model.

A. Equivalence of the BEHC and the Timing Channel

In the timing channel, the decoder observes T^m , which can be used to calculate the BEHC output sequence Y^n . The encoder observes Z^m causally, which can be combined with past timing channel inputs V^{m-1} to find the state sequence S^n causally. Hence, any encoding/decoding scheme for the BEHC can be implemented in the timing channel, and vice versa, implying that the two channels are equivalent. However, note that in the timing channel, the k th channel use consists of T_k uses of the BEHC. To take the time cost of each timing channel use into consideration, we define the timing channel capacity C_T as the maximum achievable message rate per use of the BEHC channel. In particular, given a timing channel codebook consisting of M codewords of length m , we define $\sum_{i=1}^m \mathbb{E}[T_i] \leq n$ as the number of channel uses of the BEHC averaged over the codebook and the energy arrival sequence, and the corresponding rate as

$$R = \frac{\log M}{n} \quad (7)$$

The capacity of this timing channel is the maximum rate R such that there exists a sequence of codebooks with vanishing probability of error. See Definition 3 in Appendix B for a formal definition of codes in the timing channel.

When the rates of the BEHC and the timing channel are both defined per use of the binary channel, these two channels can be shown to have the same capacity. This is due to the inputs and outputs of these channels being different but equivalent representations of the same underlying energy

harvesting channel model. We state this fact in the next lemma, the proof of which is provided in Appendix B.

Lemma 1: The timing channel capacity with additive causally known state at the encoder, C_T , and the BEHC capacity, C_{BEHC} , are equal, i.e., $C_{BEHC} = C_T$.

B. Capacity of the Timing Channel

The timing channel defined in (6) is memoryless since Z_k are independent. For such channels, the capacity is given by (2), or more explicitly by the following expression [18, Th. 7.2]²:

$$C_{CSIT} = \sup_{p(u), v(u, z)} I(U; T) \quad (8)$$

where U is an auxiliary random variable with countably infinite support, and $v(U, Z)$ is a mapping from auxiliary U and state Z to the channel input V . As stated in [18, Th. 7.2], a deterministic $v(u, z)$ can be assumed without losing optimality. Hence, solving (8) requires finding the optimal distribution for U , $p(u)$, and the optimal deterministic mapping $v(u, z)$.

In view of Lemma 1, channels defined in the timing domain and the binary domain yield the same capacities. In the following theorem, we combine the approach in [18, Th. 7.2] for channels with causal state information at the transmitter with that in [11] for timing channels, and determine the capacity in the timing domain.

Theorem 1: The capacity of the timing channel with additive causally known state, C_T , is

$$C_T = \sup_{p(u), v(u, z)} \frac{I(U; T)}{\mathbb{E}[T]} \quad (9)$$

Proof: Let W denote the message which is uniform on $\{1, \dots, M\}$. Let n be the maximum number of binary channel uses, averaged over the energy arrivals E_i , to send a message $W = w$. We note that by definition, we have

$$\sum_{k=1}^m \mathbb{E}[T_k] \leq n \quad (10)$$

where the expectation is over the energy arrival sequence E_i and the message W .

1) *Converse:* Define $U_k = (W, T^{k-1})$ and assume $\mathbb{E}[T_k] < \infty$ for all k . We justify this assumption right after we conclude the first part of the converse proof under this assumption. Since E_i is an i.i.d. random process, Z_k is independent of W and T^{k-1} , and therefore U_k . We write

$$\log(M) - H(W|T^m) = H(W) - H(W|T^m) \quad (11)$$

$$= I(W; T^m) \quad (12)$$

$$= \sum_{k=1}^m I(W; T_k|T^{k-1}) \quad (13)$$

²Although [18, Th. 7.2] is stated for finite alphabet case, its achievability and converse extend for the countable infinite alphabet case with minor modifications. In particular, the converse proof extends verbatim, cardinality bound on the auxiliary variable is removed and max is replaced with sup.

$$\leq \sum_{k=1}^m I(W, T^{k-1}; T_k) \quad (14)$$

$$= \sum_{k=1}^m I(U_k; T_k) \quad (15)$$

$$\leq \frac{n}{\sum_{k=1}^m \mathbb{E}[T_k]} \sum_{k=1}^m I(U_k; T_k) \quad (16)$$

$$\leq n \sup_U \frac{I(U; T)}{\mathbb{E}[T]} = nC_T \quad (17)$$

where (16) follows from (10), and (17) follows from U_i being independent of Z_i and the inequality $\frac{\sum_i a_i}{\sum_i b_i} \leq \max_i \frac{a_i}{b_i}$, for $a_i, b_i > 0$ along with the fact³ that $H(T_k) < \infty$ whenever $\mathbb{E}[T_k] < \infty$. When $m \rightarrow \infty$, if the probability of error goes to zero, then Fano's inequality implies $\frac{H(W|T^m)}{\log(M)} \rightarrow 0$. Combining this with (7) and (17), we get $\frac{\log(M)}{n} = R \leq C_T$.

To complete the converse proof, we next justify the assumption that $\mathbb{E}[T_k] < \infty$ for all k . Let m be sufficiently large so that $\frac{H(W|T^m)}{\log(M)} < \epsilon < 1$. We have

$$R(1 - \epsilon) \leq \frac{1}{n} H(T^m) \quad (18)$$

$$\leq \frac{\sum_{k=1}^m H(T_k)}{\sum_{k=1}^m \mathbb{E}[T_k]} \quad (19)$$

$$\leq \frac{\sum_{k=1}^m \mathbb{E}[T_k] H_2\left(\frac{1}{\mathbb{E}[T_k]}\right)}{\sum_{k=1}^m \mathbb{E}[T_k]} \quad (20)$$

where (18) follows from (12) and (19) follows from independence bound on entropy and $\sum_{k=1}^m \mathbb{E}[T_k] \leq n$. The inequality in (20) is due to $H(T_k) \leq \mathbb{E}[T_k] H_2\left(\frac{1}{\mathbb{E}[T_k]}\right)$. Now, we remark that $H_2\left(\frac{1}{\mathbb{E}[T_k]}\right) \rightarrow 0$ as $\mathbb{E}[T_k] \rightarrow \infty$. Hence, n tends to infinity and the right hand side of (19) tends to zero whenever $\mathbb{E}[T_k] \rightarrow \infty$ for some subset of k indices. Therefore, any communication scheme with vanishing probability of error that has $\mathbb{E}[T_k] \rightarrow \infty$ for some subset of k indices cannot achieve a positive rate and is ruled out from optimization without loss of optimality. This justifies the assumption that $\mathbb{E}[T_k] < \infty$ for all k .

2) *Achievability*: We use the typicality based achievability scheme in [18, Th. 7.2] in the timing channel in (6) with state information at the transmitter. Let us fix an auxiliary random variable U of countable cardinality with probability mass function p_U as well as a function $f(U, Z)$. Denote the average value of T according to U and $f(\cdot, \cdot)$ as $\mathbb{E}[T]$ and assume $\mathbb{E}[T] < \infty$. We show that message rates arbitrarily close to $R = \frac{I(U; T)}{\mathbb{E}[T]}$ are achievable. We generate m i.i.d. realizations $U_i(w)$ for each codeword $w \in [1 : M]$ and independently across codewords where each codeword is equally likely with probability $\frac{1}{M}$. To send the codeword w , the transmitter puts

$V_i = f(U_i(w), Z_i)$ upon observing the noise Z_i at timing channel use i . With this encoding scheme, we have:

$$n = \sum_{i=1}^m \mathbb{E}[T_i] = m\mathbb{E}[T] \quad (21)$$

where expectation is with respect to the messages and the energy arrivals. Upon receiving T^m sequence, the receiver performs typicality decoding. This scheme has vanishing probability of error as $m \rightarrow \infty$ for $\frac{\log(M)}{m} < I(U; T)$. More precisely, one can truncate the alphabets of the input U and the output T to obtain a finite alphabet discrete memoryless approximation of the original channel and use standard typical set decoding for such channels. Define \bar{U} and \bar{T} as the truncated versions of U and T , respectively. It follows directly by countable infiniteness of U and T that $I(\bar{U}; \bar{T}) \rightarrow I(U; T)$ and $\mathbb{E}[\bar{T}] \rightarrow \mathbb{E}[T]$ as the truncations in U and T tend to infinity. This proves that rates arbitrarily close to $R = \frac{I(U; T)}{\mathbb{E}[T]}$ are achievable. ■

We remark that our approach of obtaining the channel capacity is in two steps: First, we obtain an equivalent timing channel in Lemma 1 and then we evaluate the capacity in single-letter form in Theorem 1. We note that in recent related work [20], this problem is addressed by defining operational channel capacity first and then showing that it is identical to the single-letter channel capacity formula.

We note in Section III that the optimal distribution over Shannon strategies can be found numerically for the BEHC. This is due to the fact that for a binary input X_i and binary state S_i , there are only two feasible Shannon strategies. However, for the timing channel, both the input $V_k \in \{1, 2, \dots\}$ and the state $Z_k \in \{0, 1, \dots\}$ have infinite cardinalities. Therefore, although (9) is a single-letter expression, it is difficult to evaluate explicitly. In the following sections, we first develop upper bounds for the capacity using a genie-aided method and using a method that quantifies the leakage of the state information to the receiver; and then develop lower bounds, i.e., explicit achievable schemes, by certain specific selections for $p(u)$ and $v(u, z)$; and compare these achievable rates and the upper bounds.

V. UPPER BOUNDS ON THE CAPACITY OF THE BEHC

A. Genie Upper Bound

We first provide the timing channel state Z_k to the decoder as genie information. This yields an upper bound since the decoder can choose to ignore Z_k in decoding. However, with the knowledge of Z_k , the decoder can calculate $V_k = T_k - Z_k$, and thus we obtain the upper bound

$$C_{UB}^{genie} = \max_{p(v)} \frac{H(V)}{\mathbb{E}[V] + \mathbb{E}[Z]} \quad (22)$$

$$= \max_{\mu \geq 1} \frac{1}{\mu + \mathbb{E}[Z]} \max_{\mathbb{E}[V] \leq \mu} H(V) \quad (23)$$

Note that in (23), we partition the maximization into choosing the optimal $\mathbb{E}[V] = \mu$ and choosing the optimal distribution of V with $\mathbb{E}[V] \leq \mu$. The equality in (23) holds since the term $(\mu + \mathbb{E}[Z])^{-1}$ is decreasing in μ , and therefore the optimal μ equals the expectation of the optimal V . The second

³This fact could be reasoned as follows: As a direct application of [19, Th. 12.1.1], under a mean constraint, entropy maximizing distribution is geometric for a positive integer valued random variable. Entropy of a positive valued geometric random variable T with mean $\mathbb{E}[T] \geq 1$ is $\mathbb{E}[T] H_2\left(\frac{1}{\mathbb{E}[T]}\right)$. Therefore, $H(T_k) \leq \mathbb{E}[T_k] H_2\left(\frac{1}{\mathbb{E}[T_k]}\right)$ and $H(T_k) < \infty$ whenever $\mathbb{E}[T_k] < \infty$.

maximization in (23) involves finding the entropy maximizing probability distribution over the discrete support set $\mathbb{Z}^+ = \{1, 2, \dots\}$ with the constraint $\mathbb{E}[V] \leq \mu$. The solution to this problem is a geometric distributed V with parameter $p = \frac{1}{\mu}$. Its entropy is given by $H(V) = \frac{H_2(p)}{p}$, where $H_2(p)$ is the binary entropy function. Noting that Z is also geometrically distributed with parameter q , the genie upper bound reduces to

$$C_{UB}^{\text{genie}} = \max_{p \in [0,1]} \frac{H_2(p)/p}{\frac{1}{p} + \frac{1-q}{q}} = \max_{p \in [0,1]} \frac{qH_2(p)}{q + p(1-q)} \quad (24)$$

The genie upper bound in (24) overcomes the state dependence of the timing channel by effectively removing the state Z_k from the channel. Although this neglects the main challenges of our model, we will show in Section VI-B that this is a useful upper bound which in fact is asymptotically optimal as $q \rightarrow 0$.

B. State Leakage Upper Bound

Another approach to obtain an upper bound is to quantify the minimum amount of information T^m carries about Z^m . Since Z^m is independent of the message, information *leaked* about it via T^m reduces the potential information that can be carried in T^m about the message. Following this intuition, in this subsection, we find an upper bound on $H(Z|T = t, U = u)$, which yields the state leakage upper bound for the timing channel capacity.

An example that relates to this idea can be found in [21]. This reference considers communicating through a queue with a single packet buffer, where the encoding is performed over arrival times to the buffer. The decoder recovers the message by observing the buffer departure times of packets, which have suffered random delays through the buffer. What this example suggests is that it is possible to achieve a positive message rate through a buffer that causes random delays. In a similar manner, we can consider timing channel input V as random delay, and achieve a positive rate between the harvesting process and the decoder in addition to the message rate of the timing channel. Since the total message rate is limited to $H(Y)$ or $H(T)/\mathbb{E}[T]$ by the cutset bound, quantifying this nonzero rate between the harvesting process and the decoder is useful in finding an upper bound.

We first present the following lemma, where we provide an upper bound for $H(Z|T = t, U = u)$. This conditional entropy represents the amount of uncertainty remaining in Z after the decoder receives T and successfully decodes U .

Lemma 2: For the timing channel $T = V + Z$, where Z is geometric with parameter q , and $V = v(U, Z)$ with the auxiliary random variable U independent of Z , we have

$$H(Z|T = t, U = u) \leq H(Z_t) \quad (25)$$

where Z_t is a truncated geometric random variable on $\{0, 1, \dots, t-1\}$ with the probability mass function

$$p_{Z_t}(z) = \begin{cases} \frac{q(1-q)^z}{1-(1-q)^t}, & \text{if } z < t \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

	$T = 1$	$T = 2$	\dots	$T = t$	\dots		
$Z = 0$	0	0	\dots	0	0	p_0	0
$Z = 1$		0	\dots	p_1	0	0	0
\vdots			\ddots	\vdots	\vdots	\vdots	\vdots
$Z = t-1$				p_t	0	0	0
\vdots					0	0	0

Fig. 3. The joint probability matrix $p(z, t|u)$ for a fixed strategy u . There is one non-zero term in each row, which equals $p(z)$. When calculating $H(Z|T = t, U = u)$, only the values in the bold rectangle are required.

Proof: We first examine the joint distribution $p(z, t|u)$ resulting from a deterministic $v(U, Z)$, which is depicted as a two-dimensional matrix in Fig. 3. Given $Z = z$ and $U = u$, the output of the channel is $T = v(u, z) + z$. Therefore, each row of $p(z, t|u)$ in the figure contains one non-zero term. We also have

$$p(z, t|u) = 0, \quad z \geq t \quad (27)$$

since $v(u, z)$ is positive by definition. This is denoted by the shaded area in the figure. Moreover, we write

$$p(z, v(u, z) + z|u) = \sum_{t=1}^{\infty} p(z, t|u) \quad (28)$$

$$= p(z|u) = p(z) \quad (29)$$

implying that the non-zero term in row z is equal to $P[Z = z]$. Here, the second equality in (29) follows from the independence of U and Z .

To find $H(Z|T = t, U = u)$, we focus on column t of the probability matrix $p(z, t|u)$, which is marked with a bold rectangle in the figure. Let $\mathcal{A} \subset \{0, 1, \dots, t-1\}$ denote the set of indices $z \in \{0, 1, \dots, t-1\}$ for which $p(z, t|u) = p(z)$. As such, we can write $p(z|t, u)$ as

$$p_{\mathcal{A}}(z) = p(z|t, u) = \frac{p(z, t|u)}{\sum_{z=1}^{\infty} p(z, t|u)} \quad (30)$$

$$= \begin{cases} \frac{q(1-q)^z}{\sum_{a \in \mathcal{A}} q(1-q)^a}, & \text{if } z \in \mathcal{A} \\ 0, & \text{otherwise} \end{cases} \quad (31)$$

We next prove that $H(Z|T = t, U = u)$ is maximized when $\mathcal{A}^* = \{0, 1, \dots, t-1\}$, i.e., when all terms in the bold rectangle in Fig. 3 are non-zero. To this end, we show that the distribution $p_{\mathcal{A}^*}(z)$ is majorized by $p_{\mathcal{A}}(z)$ for all index sets $\mathcal{A} = \{a_0, a_1, \dots, a_{k-1}\} \subset \{0, 1, \dots, t-1\}$, $k \leq t$. Without loss of generality, we assume that $a_0 < a_1 < \dots < a_{k-1}$, which implies the ordering

$$p_{\mathcal{A}}(a_0) > p_{\mathcal{A}}(a_1) > \dots > p_{\mathcal{A}}(a_{k-1}) \quad (32)$$

for any \mathcal{A} . For $0 \leq n \leq k-1$, we write

$$\sum_{i=0}^n p_{\mathcal{A}}(a_i) = \frac{\sum_{i=0}^n q(1-q)^{a_i}}{\sum_{i=0}^{k-1} q(1-q)^{a_i}} \quad (33)$$

$$\geq \frac{\sum_{i=0}^n (1-q)^{a_n+i-n}}{\sum_{i=0}^n (1-q)^{a_n+i-n} + \sum_{i=n+1}^{k-1} (1-q)^{a_i}} \quad (34)$$

$$\geq \frac{\sum_{i=0}^n (1-q)^{a_n+i-n}}{\sum_{i=0}^{k-1} (1-q)^{a_n+i-n}} \quad (35)$$

$$\geq \frac{\sum_{i=0}^n (1-q)^i}{\sum_{i=0}^{t-1} (1-q)^i} = \sum_{i=0}^n p_{\mathcal{A}^*}(i) \quad (36)$$

where we obtain (34) by subtracting

$$\delta_1 = \sum_{i=0}^n (1-q)^{a_i} - \sum_{i=0}^n (1-q)^{a_n+i-n} \quad (37)$$

from both the numerator and the denominator, and we obtain (35) by adding

$$\delta_2 = \sum_{i=n+1}^{k-1} (1-q)^{a_n+i-n} - \sum_{i=n+1}^{k-1} (1-q)^{a_i} \quad (38)$$

to the denominator. Note that both δ_1 and δ_2 are non-negative since $a_n - a_i \geq n - i$, for $n \geq i$. Finally, (36) follows from $k \leq t$.

Due to the concavity of $f(x) = -x \log(x)$, and since the set \mathcal{A} is finite, the majorization shown in (33)-(36) implies that $H(Z|T=t, U=u)$ is maximized for $\mathcal{A}^* = \{0, 1, \dots, t-1\}$. In this case, the conditional distribution of Z given t and u is truncated geometric. Hence, for any $v(U, Z)$, $H(Z|T=t, U=u)$ is upper bounded by the entropy of a truncated geometric random variable, $H(Z_t)$. ■

Using the bound obtained in Lemma 2, we next present the leakage upper bound on the timing channel capacity C_T .

Theorem 2: The capacity of the timing channel and therefore the BEHC is upper bounded by

$$C_{UB}^{leakage} = \sup_{p_T(t) \in \mathcal{P}} \frac{H(T) - \sum_{t=1}^{\infty} \frac{H_2((1-q)^t)}{1-(1-q)^t} p_T(t)}{\mathbb{E}[T]} \quad (39)$$

where $H_2(\cdot)$ is the binary entropy function, and

$$\mathcal{P} = \left\{ p_T(t) \left| \sum_{t=1}^s p_T(t) \leq 1 - (1-q)^s, \quad s = 1, 2, \dots \right. \right\} \quad (40)$$

Proof: Using the chain rule of mutual information, we write the numerator of (9) as⁴

$$I(U; T) = I(U, Z; T) - I(Z; T|U) \quad (41)$$

$$= H(T) - H(T|U, Z) - I(Z; T|U) \quad (42)$$

$$= H(T) - I(Z; T|U) \quad (43)$$

where the last equality follows since $T = v(U, Z) + Z$ is a deterministic function of U and Z . Note that the $I(Z; T|U)$ term in (43) quantifies the information leaked to the decoder

⁴Following the justification in the proof of Theorem 1, we assume in the following algebraic steps that $\mathbb{E}[T], H(T) < \infty$.

about the energy harvesting process Z . We lower bound this term as

$$I(Z; T|U) = H(Z|U) - H(Z|T, U) \quad (44)$$

$$= H(Z) - H(Z|T, U) \quad (45)$$

$$= \sum_{t=1}^{\infty} \sum_u p(t, u) [H(Z) - H(Z|T=t, U=u)] \quad (46)$$

$$\geq \sum_{t=1}^{\infty} [H(Z) - H(Z_t)] \sum_u p(t, u) \quad (47)$$

$$= \sum_{t=1}^{\infty} [H(Z) - H(Z_t)] p_T(t) \quad (48)$$

where (45) is due to the independence of Z and U , and (47) is due to Lemma 2. Substituting (43) and (48) in (9), we get

$$C_T \leq \sup_{p(u), v(u, z)} \frac{H(T) - \sum_{t=1}^{\infty} [H(Z) - H(Z_t)] p_T(t)}{\mathbb{E}[T]} \quad (49)$$

Note that the objective is a function of $p_T(t)$ only. Therefore, without loss of generality, we can perform the maximization over distributions $p_T(t)$ that are achievable by some auxiliary $p_U(u)$ and function $v(U, Z)$. Since $T > Z$ by definition, such a distribution must satisfy

$$\sum_{t=1}^s p_T(t) \leq \sum_{z=0}^{s-1} p(z) = 1 - (1-q)^s, \quad s = 1, 2, \dots \quad (50)$$

As a result, the distribution $p_T(t)$ induced by any $p_U(u)$ and $v(U, Z)$ lies in the set of distributions \mathcal{P} defined in (40). We finally note that for geometrically distributed Z and truncated geometric distributed Z_t , we have

$$H(Z) - H(Z_t) = \frac{H_2((1-q)^t)}{1-(1-q)^t} \quad (51)$$

Substituting (50) and (51) in (49), we arrive at the upper bound in (39)-(40). ■

C. Computing the Approximate Value of the State Leakage Upper Bound

Solving (39) requires finding the optimal $p_T(t)$ distribution in \mathcal{P} on an infinite dimensional space. While this is analytically challenging, it is possible to numerically approximate the upper bound in (39) by restricting T to a sufficiently large but finite support of $[1, 2, \dots, |\mathcal{T}|]$. In the following, we use \mathcal{P} to denote this probability simplex with finite support of $[1, 2, \dots, |\mathcal{T}|]$. In this case, properties of the optimal distribution $p_T^*(t)$ can be utilized in the calculation for larger values of $|\mathcal{T}|$. To find such properties, we begin by rewriting the maximization problem in (39) as

$$C_{UB}^{leakage} = \sup_{\beta} \frac{1}{\beta} \sup_{p_T(t) \in \mathcal{P}, \mathbb{E}[T] \leq \beta} H(T) - \sum_{t=1}^{|\mathcal{T}|} \Delta_t p_T(t) \quad (52)$$

where we have defined $\Delta_t = \frac{H_2((1-q)^t)}{1-(1-q)^t}$. The inner maximization in (52) is a convex program since it has a concave

objective and linear constraints. For this problem, we write the KKT optimality conditions [22] for $t = 1, 2, \dots, |\mathcal{T}|$ as

$$p_T(t) = \exp\left(-\mu t - \Delta_t + \lambda_t - \sum_{s=t}^{|\mathcal{T}|} \gamma_s - \eta - 1\right) \quad (53)$$

$$\lambda_t p_T(t) = 0, \quad \lambda_t \geq 0 \quad (54)$$

$$\gamma_t \left(\sum_{s=1}^t p_T(s) - 1 + (1-q)^t \right) = 0, \quad \gamma_t \geq 0 \quad (55)$$

$$\mu (\mathbb{E}[T] - \beta) = 0, \quad \mu \geq 0 \quad (56)$$

$$\eta \left(\sum_{s=1}^{|\mathcal{T}|} p_T(s) - 1 \right) = 0 \quad (57)$$

where λ_t , γ_t , μ and η are the Lagrange multipliers for the constraints $p_T(t) \geq 0$, $\sum_{s=1}^t p_T(s) \leq 1 - (1-q)^t$, $\mathbb{E}[T] \leq \beta$, and $\sum_{s=1}^{|\mathcal{T}|} p_T(s) = 1$, respectively.

In order to have $p_T(t) = 0$ for some t , we need the exponent term in (53) to go to $-\infty$. This makes λ_t in the expression of $p_T(t)$ redundant due to (54). Hence, we assign $\lambda_t = 0$ for all t , and obtain

$$p_T^*(t) = A \exp\left(-\mu t - \Delta_t - \sum_{n=t}^{|\mathcal{T}|} \gamma_n\right) \quad (58)$$

where we have defined $A = e^{-\eta-1}$. We find A from (57) for all $\mu \geq 0$ and γ_i as

$$A = \left(\sum_{t=1}^{|\mathcal{T}|} e^{-\mu t - \Delta_t - \sum_{n=t}^{|\mathcal{T}|} \gamma_n} \right)^{-1} \quad (59)$$

which, together with (58), gives us a class of distributions with parameters γ_t and μ . In addition, from (55), we know that γ_t is positive only when the constraint in (50) is satisfied with equality. We can approximate the optimal distribution $p_T^*(t)$ numerically using finely quantized β values by searching the class of distributions in (58) for the optimal γ_t and μ satisfying the above conditions.

VI. ACHIEVABLE RATES FOR THE BEHC

In this section, we propose two choices for the auxiliary random variable U and the mapping $v(u, z)$ in (9) and find lower bounds on the timing channel capacity and hence the BEHC capacity.

A. Modulo Encoding With Finite Cardinality Auxiliary Random Variables

Let U be distributed over the finite support set $\{0, 1, \dots, N-1\}$, where N is a parameter to be optimized. We choose the mapping

$$v(U, Z) = (U - Z \bmod N) + 1 \quad (60)$$

which gives a channel input $V = v(U, Z)$ in $\{1, 2, \dots, N\}$. The output of the timing channel becomes $T = V + Z = (U - Z \bmod N) + 1 + Z$. The decoder calculates

$$T' = (T - 1 \bmod N) = ((U - Z \bmod N) + Z \bmod N) \quad (61)$$

$$= U \bmod N = U \quad (62)$$

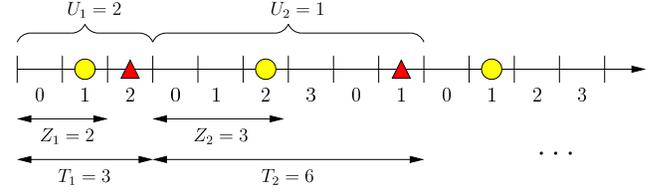


Fig. 4. Modulo encoding: each message symbol U_i is conveyed by transmitting a 1 at the earliest channel use possible with index equal to U_i . Here, $N = 4$.

and therefore perfectly recovers U in each channel use. Hence, the achievable rate for this N is

$$R_A^{(N)} = \max_{p(u), U \in \{0, \dots, N-1\}} \frac{H(U)}{\mathbb{E}[V + Z]} \quad (63)$$

We then find the best rate achievable with this scheme by optimizing over N as

$$R_A^{mod} = \max_N R_A^{(N)} \quad (64)$$

This encoding scheme has the following interpretation for the BEHC: Consider that after each instance of $X_i = 1$, future channel uses are indexed cyclically with the numbers $\{0, 1, \dots, N-1\}$, as illustrated in Fig. 4 for $N = 4$. These indices are available to both the encoder and the decoder since the channel is noiseless. The encoder can then convey any symbol $U \in \{0, 1, \dots, N-1\}$ to the decoder by sending a 1 in a channel use indexed with U . This is performed at the earliest possible such channel use in which the required energy is available. For example, $U_1 = 2$ in the figure is conveyed in the first channel use indexed with a 2 (in the first frame of N channel uses) as the energy becomes available for that transmission. However, $U_2 = 1$ in the figure is conveyed in the second channel use indexed with a 1 (in the second frame of N channel uses), since energy is not yet harvested in the first channel use indexed by a 1 (in the first frame of N channel uses). As such, in this coding scheme, the encoder partitions future channel uses into frames of length N , and uses the earliest feasible frame to convey its symbol U_k .

This encoding scheme resembles the idea of *concentration* proposed by Willems in [23] and [24] for Gaussian channels with causal state information. In particular, part of the channel input in [23] and [24] is used to concentrate the channel state onto a set of values so that it can be decoded and eliminated at the decoder. Here, by waiting for the next frame of length N when necessary, the effective state Z_k is concentrated onto the lattice of the integer multiples of N . The concentrated state is then removed by the decoder with the modulo operation when calculating T' .

B. Asymptotic Optimality of Modulo Encoding

We next show that the modulo encoding scheme proposed in Section VI-A is asymptotically optimal as the harvest rate $q \rightarrow 0$. We establish this by comparing the achievable rate of the modulo encoding scheme in (63)-(64) with the genie-aided upper bound in (24).

Theorem 3: The modulo encoding scheme for the timing channel with auxiliary $U \in \{0, 1, \dots, N-1\}$ and the channel input given in (60) is asymptotically optimal as energy harvest rate $q \rightarrow 0$.

Proof: We show that the upper bound C_{UB}^{genie} and the achievable rate R_A^{mod} scale with the same rate as q goes to zero, i.e.,

$$\lim_{q \rightarrow 0} \frac{C_{UB}}{R_A} = 1 \quad (65)$$

For fixed q , the solution of (24) satisfies

$$\frac{q(\log(1-p^*) - q \log(p^*))}{(p^* + q - p^*q)^2} = 0 \quad (66)$$

which yields

$$q = \frac{\log(1-p^*)}{\log(p^*)} \quad (67)$$

for $q > 0$. This follows from the objective of (24) being continuous, having a positive derivative for $p < p^*$, and having a negative derivative for $p > p^*$. Consequently, there exists an optimal $0 < p^* \leq 0.5$ for all harvest rates $0 < q \leq 1$, which approaches zero with q , i.e.,

$$\lim_{q \rightarrow 0} p^* = 0 \quad (68)$$

We choose the parameters of the encoding scheme as $N = \lceil \frac{1}{p^*} \rceil$, and $p(u) = 1/N$ for $0 \leq u \leq N-1$, i.e., U is uniformly distributed. Note that $p^* \leq 0.5$ implies $N \geq 2$. Since U is uniform and independent of Z , from (60), we observe that V is distributed uniformly on $\{1, 2, \dots, N\}$. This gives $\mathbb{E}[V] = (N+1)/2$, and the achievable rate for this scheme becomes

$$R_A^N \geq \frac{H(U)}{\mathbb{E}[V] + \mathbb{E}[Z]} = \frac{\log(N)}{\frac{N+1}{2} + \frac{1-q}{q}} \quad (69)$$

$$\geq \frac{q \log(N)}{Nq + 1 - q} \quad (70)$$

where $\mathbb{E}[Z] = (1-q)/q$. We next show that (70) is increasing in N within the interval $\left[\frac{1}{p^*}, \lceil \frac{1}{p^*} \rceil\right]$. The derivative of (70) with respect to N is

$$\frac{\partial}{\partial N} \frac{q \log(N)}{Nq + 1 - q} = \frac{Nq - Nq \log(N) - q + 1}{N(Nq + 1 - q)/q} \quad (71)$$

which is positive when its numerator is positive. Substituting $N = \frac{1}{p^*}$ and $N = \lceil \frac{1}{p^*} \rceil$ both yield a positive numerator in (71) as $p^* \rightarrow 0^+$. To verify that the numerator of (71) is positive everywhere in $\left[\frac{1}{p^*}, \lceil \frac{1}{p^*} \rceil\right]$, we show that it is monotonic by checking that its derivative

$$\frac{\partial}{\partial N} (Nq - Nq \log(N) - q + 1) = -q \log(N) \quad (72)$$

is negative for all $N > e$. As a result, we can further lower bound R_A^{mod} as

$$R_A^{mod} \geq \frac{q \log(N)}{Nq + 1 - q} \geq \frac{-qp^* \log(p^*)}{q + p^*(1 - q)} = \bar{R}_A \quad (73)$$

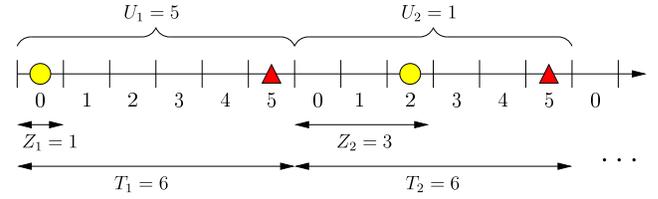


Fig. 5. Extended modulo encoding for $N = 4$.

and upper bound the left hand side of (65) as

$$\lim_{q \rightarrow 0} \frac{C_{UB}^{genie}}{R_A^{mod}} \leq \lim_{q \rightarrow 0} \frac{C_{UB}^{genie}}{\bar{R}_A} \quad (74)$$

$$= \lim_{q \rightarrow 0} \frac{qH(p^*)}{q + p^*(1 - q)} \cdot \frac{q + p^*(1 - q)}{-qp^* \log(p^*)} \quad (75)$$

$$= 1 + \lim_{p^* \rightarrow 0} \frac{(1 - p^*) \log(1 - p^*)}{p^* \log(p^*)} = 1 \quad (76)$$

Since $C_{UB}^{genie} \geq R_A^{mod}$ by definition, this proves (65) and thus the theorem. ■

Theorem 3 states that as $q \rightarrow 0$, the capacity achieving encoding scheme approaches a uniformly distributed U over $\{0, \dots, N-1\}$, where $N \rightarrow \infty$. This gives us a simple and asymptotically optimal encoding scheme for scenarios with very low energy harvesting rates.

C. Extended Modulo Encoding

To improve the rates achievable with modulo encoding of Section VI-A, we propose an extended version of the scheme with $U \in \{0, 1, \dots\}$ and

$$v(U, Z) = \begin{cases} U - Z + 1, & U \geq Z \\ (U - Z \bmod N) + 1, & U < Z \end{cases} \quad (77)$$

The interpretation of this encoding scheme for the BEHC is given in Fig. 5 for $N = 4$. Unlike modulo encoding, we index channel uses with $\{0, 1, \dots\}$ in this case. If the required energy is harvested by the channel use indexed with U_k , then the encoder sends a 1 in that channel use, as is the case for U_1 in the figure. However, if the intended channel use is missed due to lack of energy, the encoder sends a 1 within N channel uses after harvesting energy, such that the channel index and U_k are equal in modulo N . An example is U_2 in the figure, where the channel index and U_2 are equal in modulo N , i.e.,

$$(T - 1) \bmod N = U \bmod N, \quad (78)$$

The achievable rate for this scheme is calculated by solving

$$R_A^{ext} = \max_N \max_{p(u), U \in \{0, 1, \dots\}} \frac{I(U; Y)}{\mathbb{E}[V + Z]} \quad (79)$$

numerically by searching distributions of U . Although this problem is more difficult than that in (64), it is more tractable than (9) since the function $v(U, Z)$ is fixed.

We note that this scheme is an extended version of the modulo encoding scheme in Section VI-A, where U is not restricted to be within $[0, N-1]$. Therefore, the extended

modulo scheme also includes the modulo scheme as a special case when $p(u) = 0$ for $u \geq N$. In fact, this scheme can be interpreted as a combination of modulo encoding and a *best effort* encoding scheme where the closest feasible symbol is transmitted. As an example, consider two random variables $W_1 \in \{0, 1, \dots, N-1\}$ and $W_2 \in \{0, 1, \dots\}$, and let $U = W_1 + W_2N$. Then, the W_1 component is always perfectly recovered at the decoder using $(T-1) \bmod N$, as in modulo encoding. On the other hand, the W_2 component is estimated as $\lfloor (T-1)/N \rfloor$, which is as close to W_2 as can be given Z_k .

As a final remark, we note that the Shannon strategies that consider only the current state, i.e., those presented in Section III, are directly related to the strategies in the timing channel. For example, if the binary Shannon strategies are chosen i.i.d. with $P[U = (0, 1)] = p$, then a geometric distributed timing input V with parameter p yields the same channel input distribution and thus the same rate. In particular, in the timing channel, choosing $v(U, Z) = U$ independent of Z and having $U \sim \text{Geometric}(p)$ generates a codebook which yields an i.i.d. binary codebook with $P[U = (0, 1)] = p$ when translated into the BEHC. Similarly, if binary Shannon strategies are chosen by a first order Markov process, an i.i.d. timing input strategy U that yields the same input distribution can be constructed. In particular, for Shannon strategies in the BEHC given by

$$P[U_i = (0, 1)] = \begin{cases} p_0, & \text{if } U_{i-1} = (0, 0), \\ p_1, & \text{if } U_{i-1} = (0, 1), \end{cases} \quad (80)$$

we construct the timing input with distribution

$$p_{V|Z}(v|z) = \bar{p}_z(1-p_0)^{v-1}p_1 + (1-\bar{p}_z)(1-p_1)(1-p_0)^{v-2}p_1, \quad (81)$$

where \bar{p}_z is the probability of having $U_z = (0, 0)$ after starting from $U_0 = (0, 1)$ and following (80). This is equivalent to using the timing auxiliary U_k as a seed to construct BEHC strategies U_i according to (80). Hence, encoding schemes for the timing channel include the Shannon strategy schemes of Section III. However, for codebooks generated with higher order Markov processes, it is necessary to have timing auxiliary sequences U^m with memory, and a function $v_k(U_k, Z^k)$ that utilizes the history of the states. As a result, it is unclear whether encoding schemes for an i.i.d. timing channel cover higher order Markov inputs in the BEHC or not.

VII. CAPACITY WITH NO BATTERY AND INFINITE-SIZED BATTERY

For the purposes of comparison, in this section, we present two extreme cases, the case of no energy storage, and the case of infinite-sized energy storage.

A. Capacity With Zero Energy Storage

We first consider an encoder without energy storage capability. That is, we allow a non-zero channel input $X_i = 1$ only if energy is harvested within that channel use, i.e., $E_i = 1$. We note that this is slightly different than the *transmit first* model described in Section II, where the channel input is sent

before energy harvesting in each channel use. In contrast, here we consider a *harvest first* model. For this model, E_i can be considered as an i.i.d. channel state known at the encoder [2], for which the capacity is given in (8). Using the Shannon strategies $U_1 = (0, 0)$ and $U_2 = (0, 1)$, with $P[U_2] = p$, the capacity in this case becomes

$$C_{ZS} = \max_p H_2(pq) - pH_2(q) \quad (82)$$

where $H_2(p)$ is the binary entropy function.

B. Capacity With Infinite Energy Storage

Next, we consider the case with an infinite-sized battery at the encoder. Reference [1] studies the Gaussian counterpart of this channel, showing that the save-and-transmit scheme is optimal. A similar argument applies for the binary case, implying that a rate of $H(X)$ can be achieved, where X is constrained as $\mathbb{E}[X] \leq q$. Hence, the capacity of the channel with an infinite-sized storage is

$$C_{IS} = \begin{cases} H_2(q), & q \leq \frac{1}{2} \\ 1, & q > \frac{1}{2} \end{cases} \quad (83)$$

VIII. EXTENSION TO THE TERNARY CHANNEL

The equivalence of the energy harvesting channel and the timing channel extends beyond binary channels. As an example, in this section, we present results for the ternary energy harvesting channel (TEHC). The TEHC has three input and output symbols, $X, Y \in \{-1, 0, 1\}$, and both $X = -1$ and $X = 1$ require one unit of energy to be transmitted. This extension can further be generalized to M -ary channels, with each symbol consuming either 0 or 1 unit of energy.

A. Achievable Rates With Shannon Strategies

In this section, we consider achievable rates with Shannon strategies in the actual channel use index of TEHC. As in the BEHC case, we only have two states, $S_i \in \{0, 1\}$. A strategy U is in the form $U = (X, X')$, where $U(0) = X$ and $U(1) = X'$. Note that $X = 1$ or $X = -1$ is possible only when $S = 1$, and thus we only have three feasible strategies, namely $(0, 0)$, $(0, -1)$ and $(0, 1)$.

We first consider codebooks generated by choosing U_i i.i.d. for each codeword and channel use. Let the probability of choosing $U_i = (0, -1)$ and $U_i = (0, 1)$ be p_2 and p_3 , respectively, for all i and all codewords. First, note that this construction yields an ergodic battery state process, with the transition probabilities

$$P[S_{i+1} = 1 | S_i = 0] = q, \quad P[S_{i+1} = 0 | S_i = 1] = (p_2 + p_3)(1 - q) \quad (84)$$

yielding the stationary probability

$$P[S = 1] = \frac{q}{p_2 + p_3 + q - (p_2 + p_3)q} \quad (85)$$

Note that the stationary probability is a function of $p_2 + p_3$, rather than p_2 and p_3 individually. Denoting $U = (0, 0)$ as 0,

$U = (0, -1)$ as -1 and $U = (0, 1)$ as 1 , the channel in the case of naïve Shannon strategies is expressed as

$$p(y|u) = P[S = 1]\delta(y - u) + P[S = 0]\delta(y) \quad (86)$$

The best achievable rate with this scheme is given by

$$R_{NIID} = \max_{p_2, p_3 \geq 0, p_2 + p_3 \leq 1} H(Y) - (p_2 + p_3)H_2\left(\frac{q}{p_2 + p_3 + q - (p_2 + p_3)q}\right) \quad (87)$$

where $H_2(p)$ is the binary entropy function. We observe that whenever $p_2 + p_3$ is kept constant, the channel in (86) and the term $(p_2 + p_3)H_2\left(\frac{q}{p_2 + p_3 + q - (p_2 + p_3)q}\right)$ in (87) remain unchanged. On the other hand, $H(Y)$ is a concave function of the distribution of Y . Hence, by Jensen's inequality, when we fix $p_2 + p_3 = 2p$, selecting $p_2 = p_3 = p$ yields the highest rate in (87). Therefore, the optimum selection is $p_2 = p_3 = p$, and we obtain the following simpler rate expression:

$$R_{NIID} = \max_{p \in [0, 0.5]} H(Y) - 2pH_2\left(\frac{q}{2p + q - 2pq}\right) \quad (88)$$

Similar to the BEHC case, the decoder can exploit the memory by using the n -letter joint probability $p(u^n, y^n)$ for the channel and obtain *optimal i.i.d. Shannon strategy* (OIID), which achieves the following rate:

$$R_{OIID} = \max_{p \in [0, 0.5]} \lim_{n \rightarrow \infty} \frac{1}{n} I(U^n; Y^n) \quad (89)$$

where again $p_2 = p_3 = p$, whose optimality follows from a similar argument to that we presented for R_{NIID} . In particular, $p(S^n = s^n)$ is fixed from (84) whenever $p_2 + p_3$ is fixed. $p(Y^n|U^n = u^n) = \sum_{s^n} p(Y^n|S^n = s^n, U^n = u^n)p(S^n = s^n)$ has identical entropies if we exchange any $u_i = 1$ with $u_i = -1$ and vice versa. Therefore, $H(Y^n|U^n)$ is a function of only $p_2 + p_3$ and remains unchanged whenever $p_2 + p_3$ is fixed. Similarly, $H(Y^n)$ is a concave function of p_2, p_3 for fixed $p_2 + p_3$. This is due to the fact that $p(S^n = s^n)$ is fixed and $p(Y^n|S^n = s^n)$ is a linear function of p_2, p_3 . Hence, we conclude by Jensen's inequality that $p_2 = p_3 = p$ without loss of optimality in the maximization of $I(U^n; Y^n)$. Calculating the limit of the n -letter mutual information rate $\frac{1}{n} I(U^n; Y^n)$ is possible by using the algorithm in [16]. Moreover, we can further improve such achievable rates by constructing codebooks with Markovian Shannon strategies. We evaluate and compare these achievable rates in Section IX.

B. Timing Equivalence and Related Bounds

In order to find a timing equivalent for the TEHC, we represent the channel output $Y^n \in \{-1, 0, 1\}$ with two sequences, $T^m \in \{1, 2, \dots\}^m$ and $L^m \in \{-1, 1\}^m$. Here, T_k is the duration between the $(k-1)$ st and the k th non-zero outputs in Y^n , and L_k is the sign of the k th non-zero output. As in the binary case, (T^m, L^m) and Y^n are different and complete representations of the same channel output, and therefore are equivalent.

The timing equivalent of the TEHC consists of two parallel channels, namely a timing channel and a sign channel,

expressed as

$$T_k = V_k + Z_k, \quad L_k = Q_k \quad (90)$$

where Q_k is the sign of the k th non-zero input. Extending Lemma 1 to include the sign channel, we observe that the sum capacity of the two independent channels in (90) is equal to the capacity of the TEHC. The capacity of the noiseless sign channel is $\log_2 |L| = 1$ bit per channel use. One use of the sign channel also requires $\mathbb{E}[T]$ uses of the TEHC on average. Considering this, the capacity of the TEHC is given in the following theorem.

Theorem 4: The capacity of the ternary energy harvesting channel is

$$C_{TEHC} = \max_{p(u), v(u, z)} \frac{I(U; T) + 1}{\mathbb{E}[T]} \quad (91)$$

Proof: Consider the timing channel given in (90) with channel input (V_k, Q_k) and channel output (T_k, L_k) , where each timing channel use takes T binary channel uses. Note that the energy constraints for this compound channel are the same as the binary input channel, since L_k does not affect consumed energy. Hence, we can extend Theorem 1 to use the compound channel input, and obtain its capacity as:

$$C_{TEHC} = \max_{p(u), v(u, z), q(u, z)} \frac{I(U; T, L)}{\mathbb{E}[T]}, \quad (92)$$

where Shannon strategies U_k give channel inputs through the functions $v_k(U_k, Z_k)$ and $q_k(U_k, Z_k)$. We can write

$$I(U; T, L) = I(U; T) + I(U; L|T) \quad (93)$$

$$\leq I(U; T) + H(L) \quad (94)$$

$$= I(U; T) + H(Q) = I(U; T) + 1, \quad (95)$$

where the inequality can be achieved with equality by choosing L such that $H(L) = 1$ and $H(L|U) = 0$. Hence, provided that $H(Q) = 1$, $q(u, s) = q(u)$ is sufficient for the maximization in (92), yielding (91). ■

This result is parallel to those in [11] on queues with information-bearing packets. In the timing equivalent of the TEHC, each non-zero channel input can be interpreted as a packet bearing one bit of information. Hence, as in [11], coding for the two channels in (90) is performed independently, yielding the capacity in (91).

The upper and lower bounds for the BEHC immediately extend to the TEHC, since the capacity for the sign channel is simple. The two upper bounds on C_{TEHC} become

$$C_{UB}^{genie} = \max_{p \in [0, 1]} \frac{H_2(p)/p + 1}{\frac{1}{p} + \frac{1-q}{q}} = \max_{p \in [0, 1]} \frac{qH_2(p) + pq}{q + p(1-q)} \quad (96)$$

$$C_{UB}^{leakage} = \max_{p_T(t) \in \mathcal{P}} \frac{H(T) - \sum_{t=1}^{\infty} \frac{H_2((1-q)^t)}{1-(1-q)^t} p_T(t) + 1}{\mathbb{E}[T]} \quad (97)$$

where \mathcal{P} is given in (40), and the two achievable rates become

$$R_A^{mod} = \max_N \max_{p(u), U \in \{0, 1, \dots, N-1\}} \frac{H(U) + 1}{\mathbb{E}[V + Z]} \quad (98)$$

$$R_A^{ext} = \max_N \max_{p(u), U \in \{0, 1, \dots\}} \frac{I(U; Y) + 1}{\mathbb{E}[V + Z]} \quad (99)$$

TABLE I
UPPER BOUNDS AND ACHIEVABLE RATES FOR THE BEHC

Arrival prob. (q)	C_{UB}^{genie}	$C_{UB}^{leakage}$	R_A^{ext}	R_A^{mod}	R_{M2}	R_{M1}	R_{OIID}
0	0	0	0	0	0	0	0
0.1	0.2600	0.2651	0.2317	0.2313	0.2199	0.2188	0.2178
0.2	0.4056	0.3871	0.3546	0.3529	0.3415	0.3384	0.3351
0.3	0.5184	0.4740	0.4487	0.4451	0.4364	0.4320	0.4301
0.4	0.6125	0.5485	0.5297	0.5230	0.5178	0.5130	0.5115
0.5	0.6942	0.6164	0.6033	0.5914	0.5890	0.5880	0.5861
0.6	0.7669	0.6807	0.6729	0.6562	0.6617	0.6591	0.6555
0.7	0.8326	0.7442	0.7403	0.7205	0.7301	0.7301	0.7270
0.8	0.8927	0.8101	0.8088	0.7881	0.8005	0.7997	0.7987
0.9	0.9483	0.8846	0.8845	0.8678	0.8808	0.8807	0.8797
1	1	1	1	1	1	1	1

with $v(U, Z)$ is given in (60) for the modulo encoding scheme, and in (77) for the extended modulo encoding scheme.

C. Capacities With Zero and Infinite Storage

Consider first the capacity with zero energy storage. That is, allow a non-zero channel input $X_i = 1$ or $X_i = -1$ only when energy is harvested in that channel use, i.e., $E_i = 1$. Using the Shannon strategies $U_1 = (0, 0)$, $U_2 = (0, -1)$ and $U_3 = (0, 1)$, with $P[U_2] = p_2$ and $P[U_3] = p_3$, the capacity becomes

$$C_{ZS} = \max_{p_2, p_3} H(Y) - (p_2 + p_3)H_2(q) \quad (100)$$

where Y has the ternary distribution $(p_2 q, 1 - (p_2 + p_3)q, p_3 q)$ and $H_2(p)$ is the binary entropy function. Since $H(Y)$ is a concave function of the distribution of Y , when $p_2 + p_3$ is fixed, by Jensen's inequality $p = p_2 = p_3$ is the optimal selection. Therefore, we get

$$C_{ZS} = \max_p H(Y) - 2pH_2(q) \quad (101)$$

where Y has the distribution $(pq, 1 - 2pq, pq)$.

Next, we consider the capacity with an infinite-sized battery. Similar to the binary case, a rate of $H(X)$ can be achieved, where X is a ternary variable that is constrained as $\mathbb{E}[X^2] \leq q$. Hence, the capacity of the channel with infinite-sized storage is

$$C_{IS} = \begin{cases} H(q/2, 1 - q, q/2), & q \leq \frac{2}{3} \\ \log_2(3), & q > \frac{2}{3} \end{cases} \quad (102)$$

where $H(q/2, 1 - q, q/2)$ denotes the entropy of the ternary distribution $(q/2, 1 - q, q/2)$.

IX. NUMERICAL RESULTS

In this section, we compare the timing channel upper bounds and achievable rates in Sections V and VI, Shannon strategy based achievable rates in Section III and capacity results for extreme cases in Section VII for the BEHC, followed by the results in Section VIII for the TEHC. The upper bounds and

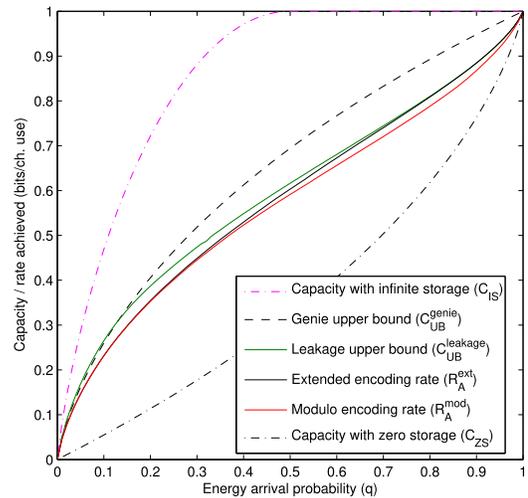


Fig. 6. Upper bounds and achievable rates for the BEHC.

achievable rates for the BEHC evaluated at $q \in \{0, 0.1, \dots, 1\}$ are given in Table I.

Fig. 6 shows the genie upper bound C_{UB}^{genie} in (24), the leakage upper bound $C_{UB}^{leakage}$ in (39) evaluated using finite-dimensional approximations, the modulo encoding achievable rate R_A^{mod} in (64), and the extended encoding achievable rate R_A^{ext} in (79) in comparison with the zero storage capacity C_{ZS} in (82) and the infinite-sized storage capacity C_{IS} in (83). All of these quantities are zero at $q = 0$, because in this case, no energy is harvested, and thus no communication is possible. Moreover, they are all equal to 1 at $q = 1$, because in this case, the battery is always full, and the channel is equivalent to a binary noiseless discrete memoryless channel without any energy constraints.

From Fig. 6, we first observe that the leakage upper bound $C_{UB}^{leakage}$ and the achievable rate with the extended encoding scheme, R_A^{ext} , provide a small gap for the channel capacity. For small q , both upper bounds and both achievable rates get very close, as expected from the asymptotic optimality of R_A^{mod} as $q \rightarrow 0$. On the other hand, for large q , we observe that the genie upper bound C_{UB}^{genie} is looser compared to the

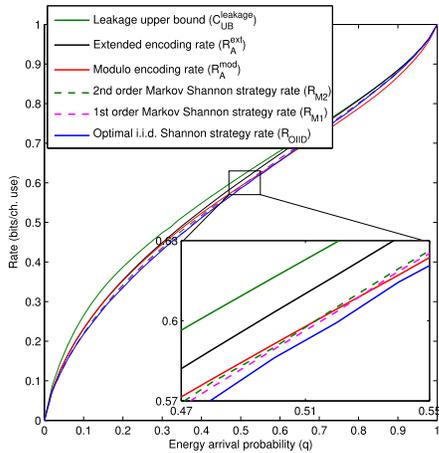


Fig. 7. Achievable rates with timing encoding compared with instantaneous Shannon strategies for the BEHC.

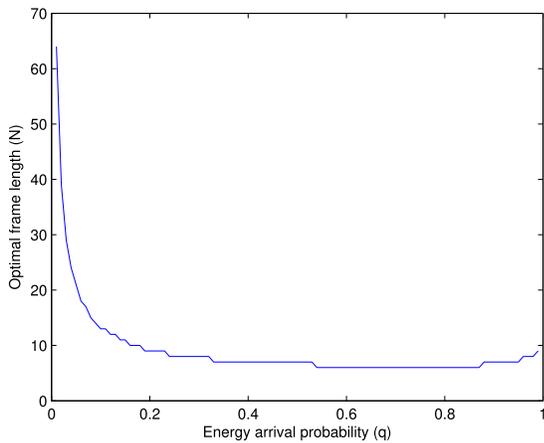


Fig. 8. Optimal choice of frame length N for the modulo encoding scheme.

leakage upper bound $C_{UB}^{leakage}$. This implies that the correlation between the harvesting process and the channel outputs is high in this regime. Finally, we note that although the gap between the infinite storage capacity C_{IS} and the zero storage capacity C_{ZS} is large, a unit-sized energy storage device recovers a significant amount of this difference. This demonstrates that even the smallest sized energy storage device can be very beneficial in energy harvesting communication systems.

We next compare the modulo and extended achievable rates, R_A^{mod} and R_A^{ext} , with the Shannon strategy based achievable rates described in Section III. We remind that the schemes in Section III, which are also studied in [3], only observe the instantaneous battery state in each channel use. Thus, we have simple Shannon strategies, but we allow a Markovian dependence over time in the codewords. Fig. 7 shows R_A^{mod} and R_A^{ext} along with the optimal i.i.d. Shannon strategy rate R_{OIID} in (4) and the optimal 1st and 2nd order Markov Shannon strategy rates R_{M1} and R_{M2} . We observe that although R_A^{mod} outperforms R_{OIID} for all q , the 1st and 2nd order Markov Shannon strategies outperform R_A^{mod} for large q , as seen in the inset in Fig. 7. However, the extended encoding rate R_A^{ext} outperforms both R_{M1} and R_{M2} , for all harvesting rates q .

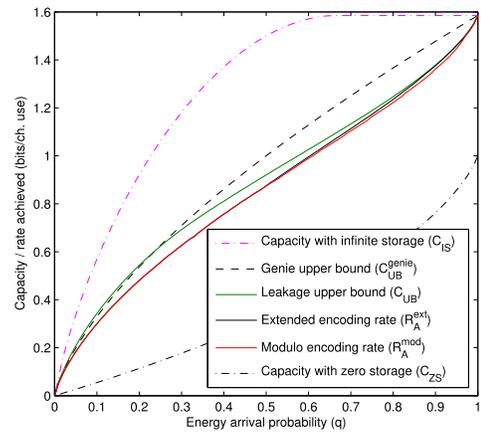


Fig. 9. Upper bounds and achievable rates for the TEHC.

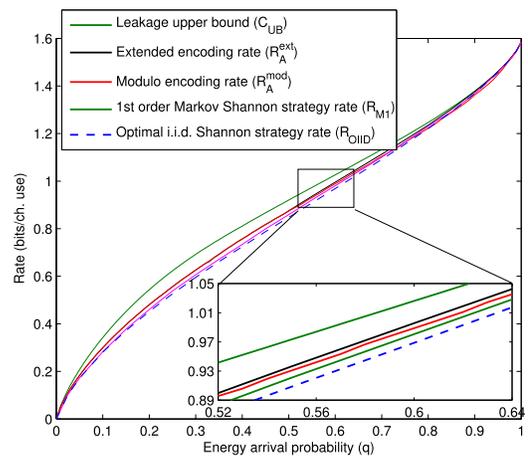


Fig. 10. Achievable rates with timing encoding compared with instantaneous Shannon strategies for the TEHC.

These can also be observed partially (for harvesting rates $q \in \{0, 0.1, \dots, 1\}$) from Table I. We note that the increase in the achievable rate with the Markov order of the input seems to be small. However, due to the exponential increase in the computational complexity with the Markov order, it was not tractable to simulate and compare inputs of higher Markov orders, i.e., 3rd and higher Markov orders.

A parameter of interest is the optimal frame length N for the modulo encoding scheme in Section VI-A, which we present in Fig. 8. The larger N is, the larger the support of U is, and more information can be packed into a single use of the timing channel. However, as N increases, so does $\mathbb{E}[T]$, and thus each symbol takes more time, and more harvested energy is potentially wasted. Thus, for small harvest rates, e.g., $q \leq 0.7$, optimal N decreases with increasing q so that less harvested energy is wasted. On the other hand, for $q > 0.7$, the node is receiving excessive energy, and thus the optimal N increases to pack more information in each timing channel use.

Finally, we present the upper bounds and the achievable rates for the ternary channel, given in (96)-(99), together with the zero and infinite-sized battery capacities C_{ZS} and C_{IS} given in (101)-(102), in Fig. 9. We also compare the achievable rates in Section VI with the optimal i.i.d. and the 1st order Markov Shannon strategies for the ternary channel in Fig. 10.

Note that in the ternary channel, the $q = 1$ case corresponds to a ternary noiseless discrete memoryless channel, and thus has a capacity of $\log_2(3) = 1.58$ bits per channel use. We observe that similar to the binary case, the approximate leakage upper bound $C_{UB}^{leakage}$ and the extended encoding rate R_A^{ext} provide a small gap for channel capacity, and the extended encoding rate outperforms the i.i.d. and the 1st order Markov Shannon strategies, for all harvesting rates q .

X. CONCLUSION

Finding the capacity of the binary energy harvesting channel is challenging due to the memory and the input dependence of the battery state. In this paper, we have addressed a simple instantiation of this channel we call the binary energy harvesting channel, with unit-sized energy storage and without channel noise. For this case, we have shown that the binary channel can be represented as a timing channel, where the states do not have memory and are not input dependent. Using this equivalence, we have derived two upper bounds: the genie upper bound by providing battery state to the decoder, and the leakage upper bound by quantifying the information leaked to the decoder about energy harvests. We have also proposed two encoding schemes based on a modulo encoding strategy, showing that they are asymptotically optimal for small energy harvesting rates. We have extended these results to the ternary energy harvesting channel. We have also observed that the timing channel based achievable rates outperform i.i.d. and the 1st and 2nd order Markov Shannon strategies that only consider instantaneous battery states.

APPENDIX A SHANNON STRATEGIES FOR THE BEHC

We select an encoding strategy as follows: We consider U_i as the input and Y_i as the output of a fictitious channel. In this case, the channel between U_i and Y_i is a finite state channel with memory:

$$P[S_{i+1} = 1 | S_i = 0, U_i = 0] = q, \quad (103)$$

$$P[S_{i+1} = 0 | S_i = 1, U_i = 0] = 0 \quad (104)$$

$$P[S_{i+1} = 1 | S_i = 0, U_i = 1] = q, \quad (105)$$

$$P[S_{i+1} = 0 | S_i = 1, U_i = 1] = 1 - q \quad (106)$$

In particular, the state of this channel is the battery level S_i and there is no state information available. Thus, we can use the development in [25, Sec. 4.6] for finite state channels to analyze achievable rates for the fictitious channel between U_i and Y_i . Note that any achievable rate in this fictitious channel is directly achievable in the BEHC channel; however, these schemes are not necessarily capacity achieving.

We note that the channel between U_i and Y_i is indecomposable as the condition in [25, Th. 4.6.3] holds for the state evolution equations in (103)-(106). Consequently, the initial state of the battery does not affect the capacity of the fictitious channel between U_i and Y_i due to [25, Th. 4.6.4] and it is expressed as:

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{p_{U^n}} I(U^n; Y^n) \quad (107)$$

We impose that the transmitter chooses an i.i.d. U_i sequence with $P[U_i = 1] = p$ for all i and the best achievable rate in this case is given by (4). Since this is the highest achievable rate with an i.i.d. codebook, we refer to this scheme as the *optimal i.i.d. Shannon strategy* (OIID).

Note that since the battery state is initially set to $S_0 = 0$ deterministically, the i.i.d. U_i selection yields an asymptotically mean stationary battery state and output processes, with the transition probabilities

$$P[S_{i+1} = 1 | S_i = 0] = q, \quad P[S_{i+1} = 0 | S_i = 1] = p(1 - q) \quad (108)$$

yielding the stationary probability

$$P[S = 1] = \frac{q}{p + q - pq} \quad (109)$$

Since (U^n, Y^n) is jointly asymptotically mean stationary, we have from [26, Th. 3.1.1]

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} I(U^n; Y^n) &= \lim_{n \rightarrow \infty} \frac{1}{n} H(U^n) + \lim_{n \rightarrow \infty} \frac{1}{n} H(Y^n) \\ &\quad - \lim_{n \rightarrow \infty} \frac{1}{n} H(U^n, Y^n) \end{aligned} \quad (110)$$

$$\begin{aligned} &= H(U) + \lim_{n \rightarrow \infty} H(\tilde{Y}_n | \tilde{Y}_n^{n-1}) \\ &\quad - \lim_{n \rightarrow \infty} H(U_n, \tilde{Y}_n | U^{n-1}, \tilde{Y}_n^{n-1}) \end{aligned} \quad (111)$$

$$\geq H(U) - \lim_{n \rightarrow \infty} H(U_n | U^{n-1}, \tilde{Y}_n) \quad (112)$$

$$\geq H(U) - \lim_{n \rightarrow \infty} H(U_n | \tilde{Y}_n) \quad (113)$$

where \tilde{Y}_n is the unique stationary and ergodic component of the output process Y_n (see also [3]). Note that the right hand side of (113) is the evaluation of the mutual information $I(U; Y)$ for the following channel:

$$p(y|u) = P[S = 1]\delta(y - u) + P[S = 0]\delta(y) \quad (114)$$

where $\delta(u)$ is 1 at $u = 0$, and zero elsewhere. This shows that performing joint typicality decoding assuming the discrete memoryless channel (114) yields an achievable rate. This scheme ignores the memory in the model and considers a channel with i.i.d. states with the state probability given in (109). Since the decoder treats the channel as if it is memoryless and performs decoding over its average statistics rather than its n -letter evolution, it performs worse than the OIID scheme. We refer to this scheme as the *naïve i.i.d. Shannon strategy* (NIID). The best achievable rate for the NIID scheme is given by (3).

APPENDIX B PROOF OF LEMMA 1

The lemma is proved by showing that an encoding scheme designed in the binary domain can be employed in the timing domain with the same probability of error, and vice versa.⁵ For this purpose, we begin with two explicit definitions of codes for the BEHC, which we show to be equivalent to each other. Recall that a transmit-first model is considered in the BEHC

⁵We note that a similar approach is followed in [27] for sticky channels, which are shown to have the same capacity per unit cost as the sticky block channel.

model in Section II, implying that only E^{i-1} is available to the encoder when choosing X_i .

Definition 1: An (n, M, ϵ) -code for the BEHC consists of a codebook of M codewords, each of which is a vector of n functions mapping past realizations of energy harvests into channel inputs, e.g., the i th element of the vector is a mapping from E^{i-1} to X_i ; such that $X_i = 1$ only if energy is present in the battery, i.e., $S_i = 1$, and a decoder observing channel output Y^n can select the correct codeword with probability $1 - \epsilon$. The channel input X_i can equivalently be expressed as the output of the function $X_i(W, E^{i-1})$. The rate of an (n, M, ϵ) -code is defined as $\log(M)/n$.

Definition 2: An (n, M, ϵ) -timing-code for the BEHC consists of a codebook of M codewords, each of which is a vector of n functions mapping past realizations of idle time into timing domain inputs, e.g., the k th element of the vector is a mapping from Z^k to V_k ; such that the decoder observing channel output T^m for the maximum m satisfying $\sum_{k=1}^m T_k \leq n$ can select the correct codeword with probability $1 - \epsilon$. The channel input V_k can equivalently be expressed as the output of the function $V_k(W, Z^k)$. The rate of an (n, M, ϵ) -timing-code is defined as $\log(M)/n$.

We remark that although the channel input and coding in Definition 2 is based on timing, both Definition 1 and Definition 2 are constructed on the binary channel, i.e., the number of channel uses n corresponds to uses of the BEHC.

Lemma 3: For each (n, M, ϵ) -code for the BEHC, there exists an (n, M, ϵ) -timing-code. Conversely, for each (n, M, ϵ) -timing-code for the BEHC, there exists an (n, M, ϵ) -code.

Proof: We first show the existence of an (n, M, ϵ) -timing-code given the existence of an (n, M, ϵ) -code for the BEHC by constructing the corresponding encoding function $V_k(W, Z^k)$ from $X_i(W, E^{i-1})$ and ensuring the same error probability. Let

$$V_k(W, Z^k) = \begin{cases} \min v \text{ s.t. } X_{j_k+v}(W, \bar{E}^{j_k+v-1}) = 1, \\ 1 \leq v \leq n - j_k, \text{ if such } v \text{ exists,} \\ n - j_k, \text{ otherwise,} \end{cases} \quad (115)$$

where $j_k = \sum_{\ell=1}^{k-1} T_\ell$ and $j_1 = 0$. Here, $\{\bar{E}_i\}$ is a dummy harvesting sequence that the timing-code constructs to mimic the harvested energy in the BEHC, and is defined as

$$\bar{E}_i = \begin{cases} 1, & \text{if } i = j_k + Z_k \text{ for some } k, \\ 0, & \text{if } j_k \leq i < j_k + Z_k \text{ for some } k, \\ A_i, & \text{otherwise,} \end{cases} \quad (116)$$

and A_i is an i.i.d. Bernoulli sequence with parameter q . Note that since Z_k are i.i.d. geometric, \bar{E}_i is an i.i.d. Bernoulli(q) sequence regardless of the realizations of j_k . Let the BEHC decoder use the mapping $\hat{W}(Y^n)$ to find the transmitted message with probability $1 - \epsilon$, i.e.,

$$\sum_w p_W(w) \sum_{y^n \in \mathcal{Y}_w} p(y^n | W = w) = 1 - \epsilon. \quad (117)$$

where $\mathcal{Y}_w = \{y^n | \hat{W}(y^n) = w\}$. The timing-code decoder can construct an output sequence \bar{Y}^n as

$$\bar{Y}_i = \begin{cases} 1, & \text{if } i = \sum_{j=1}^k T_j \text{ for some } k, \\ 0, & \text{otherwise,} \end{cases} \quad (118)$$

and use the decoding function $\hat{W}(\bar{Y}^n)$. By the construction of $\{V_k(W, Z^k)\}$ in (115) and the equal distributions of E^n and \bar{E}^n , the two sequences Y^n and \bar{Y}^n have the same distribution for each message w , i.e., $p(Y^n | W = w) = p(\bar{Y}^n | W = w)$. Hence, the error probability for the timing-code is also ϵ , and a (n, M, ϵ) -timing-code is constructed.

We next show the reverse, i.e., we show the existence of an (n, M, ϵ) -code for the BEHC given the existence of an (n, M, ϵ) -timing-code. In the BEHC, let the encoder construct a codebook with

$$X_i(W, E^{i-1}) = \begin{cases} 1, & \text{if } i = j_k + \bar{Z}_k + V_k(W, \bar{Z}^k) \text{ for some } k, \\ 0, & \text{otherwise,} \end{cases} \quad (119)$$

where j_k is the position of the $k - 1$ th nonzero input in the channel,

$$j_k = \sum_{\ell=1}^{k-1} (\bar{Z}_\ell + V_\ell(W, \bar{Z}^\ell)), \quad (120)$$

and $\{\bar{Z}_k\}$ is a dummy idle time sequence constructed by the BEHC, defined as

$$\bar{Z}_k = \begin{cases} \min z \text{ s.t. } E_{j_k+z} = 1, z \geq 0, & j_k + z < n, \\ \infty, & \text{otherwise.} \end{cases} \quad (121)$$

In other words, the code constructs a sequence \bar{Z}^m for the first n uses of the binary channel based on its harvests E^n and past transmissions. By construction, this ensures that $X_i(W, E^i) = 1$ occurs only when there is energy in the battery, i.e., $S_i = 1$. Given the definition in (121) and an i.i.d. Bernoulli harvesting process E^n , the constructed idle time sequence \bar{Z}^k for $j_k + \bar{Z}_k < n$ has the same i.i.d. Geometric distribution as the original state sequence Z^k , and the remaining states do not affect the output observed by the decoder.

Let the timing-code decoder use the mapping $\hat{W}(T^m)$ to find the transmitted message with probability $1 - \epsilon$. The BEHC decoder can then construct an output sequence \bar{T}^m as

$$\bar{T}_k = j_{k+1} - j_k, \quad (122)$$

and use the decoding function $\hat{W}(\bar{T}^m)$. The two sequences T^m and \bar{T}^m have the same distribution for a given message W due to the construction of $\{X_i(W, E^i)\}$ in (119). Hence, the error probabilities are equal and a (n, M, ϵ) -code for the BEHC is constructed. ■

Although Lemma 3 establishes the equivalence of the codes in Definitions 1 and 2, both codes in Definition 2 are defined on the BEHC, i.e., on n binary channel uses rather than timing channel uses. To link the timing-code in Definition 2 to the timing channel, we present the following definition for codes

for the timing channel, which is in line with the discrete timing codes in [13].

Definition 3: An (m, M, n, ϵ) -code for the timing channel consists of a codebook of M codewords, each of which is a vector of m functions mapping past realizations of idle time into timing channel inputs, e.g., the k th element of the vector is a mapping from Z^k to V_k ; such that the sum of outputs T_k for $k = 1, \dots, m$ is on average less than n , and the decoder observing channel output T^m can select the correct codeword with probability $1 - \epsilon$. The rate of an (m, M, n, ϵ) -code is defined as $\log(M)/n$.

The codes in Definitions 2 and 3 differ in the length of the code. The former is terminated at exactly n uses of the binary channel, while the latter is terminated at exactly m uses of the timing channel with the additional constraint that the corresponding number of binary channel uses is on average less than n . We next show that this difference does not affect the maximum achievable rate of these codes for arbitrarily small ϵ as $n \rightarrow \infty$, and thus the two channels have the same capacity.

Let the timing channel have capacity C , then for every $\epsilon, \delta \geq 0$, for sufficiently large n there exists an (m, M, n, ϵ) -code for the timing channel with $\log(M) \geq n(C - \delta)$. We use ℓ independent copies of these codes in ℓ blocks, and denote the sum of outputs T_k in the j th block as τ_j . By definition, we have $\mathbb{E}[\tau_j] \leq n$, which implies that the sum of all outputs in ℓ blocks has an average less than or equal to ℓn . In addition, by weak law of large numbers [28], for all $\alpha > 0$ there exists some finite ℓ such that

$$\mathbb{P} \left[\sum_{j=1}^{\ell} \tau_j - \bar{\ell} n \geq \bar{\ell} \beta \right] \leq \alpha, \quad \text{for all } \bar{\ell} > \ell \quad (123)$$

for all $\beta > 0$. We interpret this block code as a $(\ell n + \ell \beta, M^\ell, \ell \epsilon + \alpha)$ -timing-code for the BEHC. The rate of this code is $\log(M^\ell)/(\ell n + \ell \beta) = \log(M)/(n + \beta)$, which can be made arbitrarily close to $\log(M)/n$ by choosing β sufficiently small. The probability of error is equal to the probability of failure in at least one block, which is upper-bounded by $\ell \epsilon$ due to union bound, plus the probability of ℓ blocks taking more than $\ell n + \ell \beta$ binary channel uses, which is upper-bounded by α due to (123). Hence, the error probability can be made arbitrarily small by choosing ϵ and α sufficiently small.

Conversely, let the BEHC have capacity C , then for every $\epsilon, \delta \geq 0$, for sufficiently large n there exists an (n, M, ϵ) -timing-code for the BEHC with $\log(M) \geq n(C - \delta)$. We use ℓ independent copies of these codes in ℓ blocks, and denote the number of 1s in the j th block as ϕ_j . Let $\mathbb{E}[\phi_j] = m$, which implies that the total number of 1s in ℓ blocks has a mean of ℓm . In addition, by weak law of large numbers, for all $\alpha > 0$ there exists some finite ℓ such that

$$\mathbb{P} \left[\sum_{j=1}^{\ell} \phi_j - \bar{\ell} m \geq \bar{\ell} \beta \right] \leq \alpha, \quad \text{for all } \bar{\ell} > \ell \quad (124)$$

for all $\beta > 0$. We interpret this block code as a $(\ell m + \ell \beta, M^\ell, \ell n, \ell \epsilon + \alpha)$ -code for the timing channel, for which the rate is given by $\log(M^\ell)/\ell n = \log(M)/n \geq C - \delta$. The

probability of error is equal to the probability of failure in at least one block, which is upper-bounded by $\ell \epsilon$ due to union bound, plus the probability of ℓ blocks requiring more than $\ell m + \ell \beta$ timing channel uses, which is upper-bounded by α due to (124). Hence, the error probability can be made arbitrarily small by choosing ϵ and α sufficiently small.

REFERENCES

- [1] O. Ozel and S. Ulukus, "Achieving AWGN capacity under stochastic energy harvesting," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6471–6483, Oct. 2012.
- [2] O. Ozel and S. Ulukus, "AWGN channel under time-varying amplitude constraints with causal information at the transmitter," in *Proc. ASILOMAR*, Nov. 2011, pp. 373–377.
- [3] W. Mao and B. Hassibi, "On the capacity of a communication system with energy harvesting and a limited battery," in *Proc. IEEE ISIT*, Jul. 2013, pp. 1789–1793.
- [4] Y. Dong and A. Ozgur, "Approximate capacity of energy harvesting communication with finite battery," in *Proc. IEEE ISIT*, Jun./Jul. 2014, pp. 801–805.
- [5] V. Jog and V. Anantharam, "An energy harvesting AWGN channel with a finite battery," in *Proc. IEEE ISIT*, Jun./Jul. 2014, pp. 806–810.
- [6] D. Shaviv, P.-M. Nguyen, and A. Ozgur. (2015). "Capacity of the energy harvesting channel with finite battery." [Online]. Available: <https://arxiv.org/abs/1506.02024>
- [7] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.
- [8] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, pp. 203–219, Apr. 1971.
- [9] P. Popovski, A. M. Fouladgar, and O. Simeone, "Interactive joint transfer of energy and information," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 2086–2097, May 2013.
- [10] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.
- [11] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.
- [12] B. Prabhakar and R. Gallager, "Entropy and the timing capacity of discrete queues," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 357–370, Feb. 2003.
- [13] A. S. Bedekar and M. Azizoglu, "The information-theoretic capacity of discrete-time queues," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 446–461, Mar. 1998.
- [14] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
- [15] K. Tutuncuoglu, O. Ozel, A. Yener, and S. Ulukus, "Binary energy harvesting channel with finite energy storage," in *Proc. IEEE ISIT*, Jul. 2013, pp. 1591–1595.
- [16] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.
- [17] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.
- [18] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.
- [20] D. Shaviv and A. Ozgur. (2015). "Capacity of the AWGN channel with random battery recharges." [Online]. Available: <https://arxiv.org/abs/1506.02792>
- [21] M. Tavan, R. D. Yates, and W. U. Bajwa, "Bits through bufferless queues," in *Proc. ALLERTON*, Oct. 2013, pp. 755–762.
- [22] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
- [23] F. M. J. Willems, "On Gaussian channels with side information at the transmitter," in *Proc. 9th Symp. Inf. Theory Benelux*, 1988, pp. 129–135.
- [24] F. M. J. Willems, "Signaling for the Gaussian channel with side information at the transmitter," in *Proc. IEEE ISIT*, Jun. 2000, p. 348.
- [25] R. G. Gallager, *Information Theory and Reliable Communication*, vol. 2. New York, NY, USA: Wiley, 1968.
- [26] R. M. Gray, *Entropy and Information Theory*. New York, NY, USA: Springer-Verlag, 1990.

- [27] M. Mitzenmacher, "Capacity bounds for sticky channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 72–77, Jan. 2008.
- [28] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. Oxford, U.K.: Oxford Univ. Press, 2001.

Kaya Tutuncuoglu received the two B.S. degrees in electrical and electronics engineering and in physics from the Middle East Technical University, Ankara, Turkey, in 2008 and 2009, respectively. He received his PhD degree in Electrical Engineering from Pennsylvania State University, University Park, PA, USA in August 2015 and subsequently joined Facebook Inc. as a research scientist. His research interests include green communications, energy harvesting, and resource allocation for wireless networks. He received the AT&T Graduate Fellowship at Pennsylvania State University in 2012, the IEEE Marconi Prize Paper Award in Wireless Communications in 2014, and the Leonard A. Doggett Award for Outstanding Writing in Electrical Engineering at Penn State in 2014.

Omur Ozel (S'08–M'15) received B.Sc. and M.S. degrees with honors in Electrical and Electronics Engineering from the Middle East Technical University (METU), Ankara, Turkey in 2007 and 2009, respectively. He received his Ph.D. degree in Electrical and Computer Engineering (ECE) from the University of Maryland (UMD), College Park in 2014. He was a postdoctoral researcher in the Electrical Engineering and Computer Sciences Department at the University of California Berkeley in 2015–2016. Currently, he is a postdoctoral research associate in the Electrical and Computer Engineering Department at the Carnegie Mellon University. His research interests are in wireless communications, information theory, signal processing and controls.

Aylin Yener (S'91–M'01–SM'14–F'15) received the B.Sc. degree in electrical and electronics engineering and the B.Sc. degree in physics from Bogazici University, Istanbul, Turkey, and the M.S. and Ph.D. degrees in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, NJ, USA. She is a Professor of Electrical Engineering at The Pennsylvania State University, University Park, PA, USA, since 2010, where she joined the faculty as an Assistant Professor in 2002. Since 2017, she is a Dean's Fellow in the College of Engineering at The Pennsylvania State University. She is currently also a Visiting Professor at the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. From 2008 to 2009, she was a Visiting Associate Professor with the same department. Her research interests include information theory, communication theory, and network science, with recent emphasis on green communications and information security. She received the NSF CAREER award in 2003, the Best Paper Award in Communication Theory in the IEEE International Conference on Communications in 2010, the Penn State Engineering Alumni Society (PSEAS) Outstanding Research Award in 2010, the IEEE Marconi Prize Paper Award in 2014, the PSEAS Premier Research Award in 2014, and the Leonard A. Doggett Award for Outstanding Writing in Electrical Engineering at Penn State in 2014.

Dr. Yener is currently a member of the Board of Governors of the IEEE Information Theory Society, where she was previously the treasurer (2012–2014). She served as the student committee chair for the IEEE Information Theory Society 2007–2011, and was the co-founder of the Annual School of Information Theory in North America co-organizing the school in 2008, 2009, and 2010. She was a technical (co)-chair for various symposia/tracks at the IEEE ICC, PIMRC, VTC, WCNC, and Asilomar (2005–2014). She served as an editor for the IEEE TRANSACTIONS ON COMMUNICATIONS (2009–2012), an editor and an editorial advisory board member for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2001–2012), and a guest editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY (2011) and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (2015). Currently, she serves on the editorial board of the IEEE TRANSACTIONS ON MOBILE COMPUTING and as a senior editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

Sennur Ulukus (S'90–M'98–SM'15–F'16) is a Professor of Electrical and Computer Engineering at the University of Maryland at College Park, where she also holds a joint appointment with the Institute for Systems Research (ISR). Prior to joining UMD, she was a Senior Technical Staff Member at AT&T Labs-Research. She received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University, and the Ph.D. degree in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University. Her research interests are in wireless communications, information theory, signal processing, networking, information theoretic physical layer security, and energy harvesting communications.

Dr. Ulukus is a fellow of the IEEE, and a Distinguished Scholar-Teacher of the University of Maryland. She received the 2003 IEEE Marconi Prize Paper Award in Wireless Communications, the 2005 NSF CAREER Award, the 2010–2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Education Award. She is currently an Area Editor for the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING (2016 -). She was an Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Series on Green Communications and Networking (2015–2016), the IEEE TRANSACTIONS ON INFORMATION THEORY (2007–2010), and the IEEE TRANSACTIONS ON COMMUNICATIONS (2003–2007). She was a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (2015 and 2008), *Journal of Communications and Networks* (2012), and the IEEE TRANSACTIONS ON INFORMATION THEORY (2011). She is a general TPC Co-Chair of the 2017 IEEE ISIT, the 2016 IEEE Globecom, the 2014 IEEE PIMRC, and the 2011 IEEE CTW.