A Communication Channel With Random Battery Recharges

Dor Shaviv, *Student, IEEE*, Ayfer Özgür, *Member, IEEE*, and Haim H. Permuter, *Senior Member, IEEE*

***Abstract*— Motivated by the recent emergence of energy harvesting and wirelessly powered transceivers, we study communication over a memoryless channel with a transmitter, whose battery is recharged at random or deterministic times known to the receiver. We characterize the capacity of this channel as the limit of an n-letter maximum mutual information rate under various assumptions: causal and noncausal transmitter knowledge of the battery recharges, with or without feedback from the receiver to the transmitter. While the resultant n-letter capacity expressions are not computable in the general case, we demonstrate their usefulness by focusing on two important special cases, namely, the binary erasure channel (BEC) and the additive white Gaussian noise (AWGN) channel, where they lead to some interesting, and somewhat surprising, insights. By focusing on the BEC, we show that output feedback can strictly increase the capacity of this channel, even though the channel is memoryless and the battery recharging process is independent over time. Interestingly, this provides a counter example to an old claim by Shannon stated without proof in his 1956 paper. On the other hand, by focusing on the AWGN channel, we are able to show that the capacity with noncausal knowledge of the battery recharging times at the transmitter is strictly larger than that with causal knowledge, even though the battery recharging process is independent over time and known to the receiver. The n-letter expressions can also be used to derive explicit upper and lower bounds on capacity. In particular, we derive simple upper and lower bounds on the capacity of the AWGN channel with random battery recharges, which are within 1.05 b/s/Hz of each other for all parameter values.**

***Index Terms*— Energy harvesting, feedback capacity, channel with state, causal and noncausal side information, capacity bounds.**

I. Introduction

THERE has been significant recent progress in building wireless radios that possess no conventional batteries but are powered by either energy harvesting or wireless energy transfer, with latest developments reporting smaller device sizes, better harvesting efficiencies and increased communication ranges and data rates [3], [4]. For example, the ant-sized radios of [3] use the energy provided through the downlink channel in order to transmit over the uplink channel. We model communication with such externally powered transmitters by using a simple model. See Fig. 1. Here a transmitter equipped with a battery (storage unit) of size $\bar{B}$ is communicating to a receiver over a discrete memoryless channel (DMC). The transmitter's battery is recharged either periodically or at random times; in the random case we assume that the recharging process E_t is i.i.d. Bernoulli with recharging probability p, i.e.,

$$E_t = \begin{cases} \bar{B} & \text{w.p. } p \\ 0 & \text{w.p. } 1-p. \end{cases} \tag{1}$$

At each time, the energy $\phi(X_t)$ of the symbol X_t transmitted by the transmitter is limited by the amount of energy B_t available in the battery at time t, which in turn depends on the energy utilized in the previous time-slots as well as the recharging process. This leads to the following constraint on the input:

$$\begin{aligned} \phi(X_t) &\le B_t, \\ B_t &= \min\{B_{t-1} - \phi(X_{t-1}) + E_t, \bar{B}\}. \end{aligned} \tag{2}$$

This dynamic energy constraint is most relevant in the regime where the energy capacity of the battery is comparable to the energy required for transmission of a codeword. This is well motivated by the trend to shrink down the size of wireless devices, which limits the amount of energy they can harvest at any given time as well as the capacity of the battery that can be accommodated on the device. We assume that the

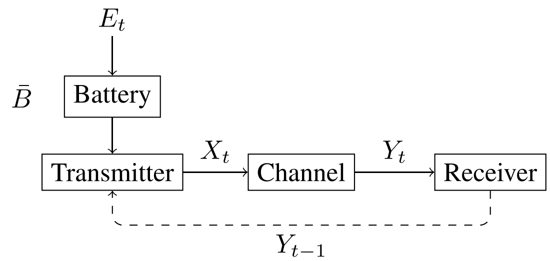

Fig. 1. Model of an energy harvesting communication system. The transmitter is equipped with a battery of size $\bar{B}$, and is communicating to a receiver over a discrete memoryless channel. The transmitter's battery is recharged by an external process E_t. We consider the cases when the transmitter may or may not observe feedback of the channel output, indicated by the dashed line.

Manuscript received August 16, 2016; revised March 13, 2017; accepted May 24, 2017. Date of publication May 31, 2017; date of current version December 20, 2017. D. Shaviv and A. Özgür were supported in part by a Robert Bosch Stanford Graduate Fellowship, in part by the National Science Foundation under Grant CCF-1618278, and in part by the Center the Center for Science of Information, an NSF Science and Technology Center, under Grant CCF-0939370. H. H. Permuter was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC under Grant 337752 and in part by the Israeli Science Foundation. This paper was presented in part at the 2015 IEEE International Symposium on Information Theory [1] and the 2015 Information Theory Workshop [2].

D. Shaviv and A. Özgür are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: shaviv@stanford.edu; aozgur@stanford.edu).

H. H. Permuter is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel (e-mail: haimp@bgu.ac.il).

Communicated by V. M Prabhakaran, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2017.2710121

recharging times (or equivalently the process E_t) are known either causally or noncausally both at the transmitter and the receiver.

This setup is a special case of the energy harvesting communication channel, where more generally one can assume E_t to be any arbitrary process. Such channels have been of significant interest in the recent literature [5]–[11]. The difficulty in characterizing their capacity lies in the fact that although the channel between the transmitter and the receiver is memoryless, the energy constraints (2) on the transmitter lead to a random state B_t for the system which has memory and is input-dependent. In general, even deriving an n-letter expression for the capacity of such channels with state is known to be a difficult problem.

In this paper, we make progress on this problem by focusing on the special case where E_t represents periodic or Bernoulli battery recharges as given in (1); we call this the *random battery recharges* (RBR) channel. This special case is motivated by applications such as [3] and [4], where the transmitter is powered by a targeted energy transfer process. In this case, it is natural to assume that the battery is fully charged whenever there is an energy transfer. In a network setting, a single charger is often responsible for powering multiple transmitters and therefore energy transfers to a given transmitter are typically intermittent since the charger has to time-share beamforming to different transmitters according to a deterministic or random pattern. In the most common scenario, when the charger is the actual receiver, it is natural to assume that the receiver is aware of the battery recharging times at the transmitter. This setting can also be used to model a transmitter harvesting energy from the natural resources in its environment, for example from ambient RF signals, in the small battery regime. When the battery is small, each non-zero energy harvest can be approximately modeled as fully charging the battery. The receiver can be aware of the battery recharging times if it is itself harvesting energy from a correlated physical process.

We provide a characterization of the capacity of this channel as an n-letter mutual information rate under various assumptions: periodic or random battery recharges; causal or noncausal knowledge of the battery recharging times; with or without output feedback from the receiver to the transmitter. While in most of the cases we investigate, the resultant n-letter capacity expressions are not computable, we demonstrate their usefulness by focusing on two important special cases, namely the BEC and the AWGN channel, where they lead to some interesting and surprising insights.

First by focusing on a BEC with periodic recharges, we show that output feedback can strictly increase capacity. The fact that feedback increases the capacity of this channel, and more generally the capacity of energy harvesting channels, is indeed surprising. In a classical wireless channel, it is clear that feedback can increase the capacity by allowing the transmitter to learn the state of the channel which is typically available at the receiver. However, in an energy harvesting channel the state of the system (captured by the available energy in the battery) is readily known at the transmitter (but only partially known at the receiver) and communication occurs over a memoryless channel. It is tempting to believe that, since all information regarding the state of the channel is already available at the transmitter, feedback from the receiver will not provide the transmitter with any additional information and therefore will not increase the capacity of this channel. Indeed, this is also what Shannon claims in his 1956 paper [12]. Theorem 6 therein proves that feedback does not increase the capacity of a discrete memoryless point-to-point channel. The proof is accompanied by the following interesting comment:

"*It is interesting that the first sentence of Theorem 6 can be generalized readily to channels with memory provided they are of such a nature that the internal state of the channel can be calculated at the transmitting point from the initial state and the sequence of letters that have been transmitted.*"[1]
We show that the BEC with periodic battery recharges indeed corresponds to a time-invariant finite state channel where the state is computable at the transmitter from the initial state and the transmitted symbol sequence. As such, it provides a counter-example to Shannon's claim.[2] The gain from feedback, at least in this example, comes from the fact that even though the transmitter knows the state of the system, the receiver does not, and feedback allows the transmitter to learn the receiver's belief regarding the state of the system, which it uses to communicate at higher rates.

For the AWGN channel, we use the n-letter capacity expressions to show that noncausal transmitter knowledge of the i.i.d. Bernoulli battery recharging times strictly increases capacity over causal knowledge, even though the receiver also knows the battery recharging times. This result can be surprising given that for channels with i.i.d. states known both at the transmitter and the receiver, noncausal and causal knowledge of the states lead to the same capacity. The difference here comes from the fact that even though the receiver knows the battery recharging times, it cannot necessarily track the state of the battery. We show this result by explicitly identifying the maximizing input distribution for the n-letter capacity expression when the battery recharging times are known noncausally at the transmitter, and build on properties of this distribution. We then proceed to derive upper and lower bounds on the capacity of this channel, which leads to an approximation of the capacity within 1.05 bits/channel use for all parameter values. An interesting intermediate step for this approximation result is to connect the information-theoretic capacity of this channel to its long-term average throughput and explicitly characterize the optimal online power control strategy maximizing this throughput, a problem that has been of significant interest in the recent literature [14]–[25].

[1] The first sentence of Theorem 6 reads "*In a memoryless discrete channel with feedback, the forward capacity is equal to the ordinary capacity C (without feed-back).*"

[2] In an independent work [13], it was shown that feedback may increase the capacity of a BEC with no consecutive ones (RLL$(1,\infty)$ input constraint), in the asymptotic regime where the erasure probability goes to zero. Note that both for this channel and for the energy harvesting channel, the state represents memory embedded in the input constraint, while the physical channel itself is memoryless.

A. Relation to Prior Work

The setup we consider in this paper corresponds to a special case of the energy-harvesting communication channel, the capacity of which has been of significant recent interest [5], [7]–[10], [26], [27]. In particular, [7] considers a noiseless binary channel with a unit-sized battery where the battery recharges are known causally only at the transmitter. Our model resembles theirs in the fact that the energy arrival process is i.i.d. Bernoulli and each energy arrival fully recharges the battery. However, our model is more general in the fact that we consider noisy channels and allow the size of the battery and the input alphabet to be arbitrary and not necessarily binary. With a binary channel and unit battery, information can be only encoded in the timing of the unit-energy pulse which makes the setup of [7] equivalent to a timing channel. In the general case, for example when the battery size and the inputs to the channel are real numbers, information can be encoded through real valued codewords. The most closely related references to our work are [9] and [10] which derive an approximation for the capacity of an AWGN channel with i.i.d. Bernoulli energy harvests without explicitly providing an n-letter characterization for the capacity. The capacity approximation we develop for the AWGN channel in this paper improves upon the approximation provided in [9] and [10] (roughly improving the gap from 1.58 bits/channel use to 1.05 bits/channel use). The n-letter characterizations for the capacity obtained in the current paper for the AWGN channel with i.i.d. Bernoulli battery recharges have been extended to general i.i.d. energy harvesting processes in [11], though with a different proof technique. The corresponding approximation results also extend to general i.i.d. energy harvests though with a larger approximation gap (3.85 bits/channel use in [11] in comparison to 1.05 bits/channel use for the Bernoulli case).

The rest of the paper is organized as follows: Section II defines the energy harvesting channel model and specifies the various cases under consideration. Section III states the main results of the paper, namely the capacity expressions. In Section IV we study the BEC, and show that feedback can increase the capacity of the energy harvesting channel. Finally, Section V considers the Gaussian channel and its capacity approximations.

II. System Model and Preliminaries

A. Notation

We begin by introducing the notation used throughout the paper. Let uppercase, lowercase, and calligraphic letters denote random variables (RVs), specific realizations of random variables, and alphabets, respectively. For two jointly distributed RVs (X, Y), let $p(x)$, $p(x, y)$, and $p(y|x)$, respectively denote the marginal of X, the joint distribution of (X, Y), and the conditional distribution of Y given X. Let $\mathbb{E}[\cdot]$ denote expectation. For $m \le n$, $X_m^n = (X_m, X_{m+1}, \dots, X_{n-1}, X_n)$, and $X^n = X_1^n$. When the length is clear from the context, we sometimes denote vectors by boldface letters, e.g. $\mathbf{x} \in \mathcal{X}^n$. All logarithms are to base 2 (ln will denote log to base e).

B. System Model

We consider a transmitter powered by RF energy transfer which communicates to a receiver over a DMC. The channel consists of a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, and a transition probability matrix $p(y|x)$. We assume that the transmitter has a battery with finite capacity $\bar{B}$ which is recharged with probability p at each channel use, i.e. the energy arrivals E_t are i.i.d. Bernoulli RVs:

$$E_t = \begin{cases} \bar{B} & \text{w.p. } p \\ 0 & \text{w.p. } 1-p. \end{cases}$$

The effort to shrink down the size of wireless sensors and actuators limits the amount of energy that can be harvested at any given time, as well as the capacity of the storage unit that can be accommodated by the device. This necessitates the recharging process to operate at a scale comparable to the symbol duration [3]. The randomness in the energy transfer process can be due to fluctuations in the alignment of antennas and the position of nodes, as well as randomness in the energy transfer times. We assume that the recharging times are known causally to both the transmitter and the receiver. The knowledge of the energy arrivals at the receiver is motivated by the fact that often it is the receiver that powers the transmitter, and the transmitter can acknowledge its battery exceeding a certain threshold by sending a short pulse to simplify operation. In such cases, it is also natural to consider feedback from the receiver to the transmitter, and assess its advantage (if any). We also consider the case of noncausal energy arrival information at the transmitter, with and without feedback, mainly for comparison with the causal case. Additionally, we will also treat the case of deterministic energy arrivals, which are known ahead of time at both the transmitter and the receiver. This models a scenario in which the transmitter is being recharged by an exogenous charging device, which operates according to a predetermined sequence of charging times.

The channel has an associated cost function $\phi : \mathcal{X} \to \mathbb{R}_+$, denoting the amount of energy used for transmission by each symbol. We assume that there is at least one symbol $x \in \mathcal{X}$ such that $\phi(x) = 0$. We call this symbol the zero symbol and denote it by $x = 0$ (if there is more than one such symbol, it is immaterial which one of them is so labeled). This assumption is necessary in order to ensure the model is well-defined; in our model, the transmitter is forced to transmit a zero symbol when the battery is empty. The energy of the channel input symbol at each time slot is limited by the available energy in the battery. Let B_t represent the available energy in the battery at time t. The system energy constraints can be described as

$$\phi(X_t) \le B_t, \tag{3}$$
$$B_t = \min\{B_{t-1} - \phi(X_{t-1}) + E_t, \bar{B}\}. \tag{4}$$

This implies that at time t, either $B_t = \bar{B}$ w.p. p, or $B_t = B_{t-1} - \phi(X_{t-1})$ w.p. $1-p$. We assume without loss of generality that $B_0 = \bar{B}$, which implies that we can also assume $E_1 = \bar{B}$ w.p. 1.

Our baseline case will be the DMC with causal energy arrival observations and without feedback. An (M, n) code for this channel is a set of encoding functions f_t and a decoding function g:

$$f_t : \mathcal{M} \times \mathcal{E}^t \to \mathcal{X}, \quad t = 1, \ldots, n, \tag{5}$$

$$g : \mathcal{Y}^n \times \mathcal{E}^n \to \mathcal{M}, \tag{6}$$

where $\mathcal{E} = \{0, \bar{B}\}$ and $\mathcal{M} = \{1, \ldots, M\}$. To transmit message $w \in \mathcal{M}$, at time $t = 1, \ldots, n$ the transmitter sets $X_t = f_t(w, E^t)$. The battery state B_t is a deterministic function of (X^{t-1}, E^t), therefore also of (w, E^t). The functions f_t must satisfy the energy constraint (3): $\phi(f_t(w, E^t)) \le B_t(w, E^t)$. The receiver sets $\hat{W} = g(Y^n, E^n)$. The probability of error is

$$P_e = \frac{1}{M} \sum_{w=1}^{M} \Pr(\hat{W} \ne w \mid w \text{ was transmitted}).$$

The rate of an (M, n) code is $R = \frac{\log M}{n}$. A rate R is achievable if for every $\varepsilon > 0$ there exists a sequence of (M, n) codes that satisfy $\frac{\log M}{n} \ge R - \varepsilon$ and $P_e \to 0$ as $n \to \infty$. The capacity C is the supremum of all achievable rates.

When noncausal energy arrival information is available at the transmitter, the symbol transmitted at time t can depend on the entire realization of the energy arrival process E^n. In this case, equation (5) becomes

$$f_t : \mathcal{M} \times \mathcal{E}^n \to \mathcal{X}, \quad t = 1, \ldots, n, \tag{7}$$

with remaining definitions unchanged. The capacity in this case is denoted by C^{nc}.

When the energy arrivals are deterministic, the encoding and decoding functions become

$$f : \mathcal{M} \to \mathcal{X}^n, \tag{8}$$

$$g : \mathcal{Y}^n \to \mathcal{M}, \tag{9}$$

just like the classical DMC, where the dependence on the energy arrival sequence e^n is implicit. To keep things simple, we will focus in this work on *periodic* energy arrival sequences, i.e.

$$e_t = \begin{cases} \bar{B}, & t = 1 \mod k \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

for a given $k \in \mathbb{N}$. We will denote the capacity in this case by $C^{\text{per}}(k)$. This special case is of interest due to its analytical simplicity, and, as will be seen later, due to its connection to C^{nc}.

When there is feedback from the receiver to the transmitter, the encoding functions in all three cases are modified accordingly. Specifically, when the energy arrivals sequence is observed causally, (5) is changed to

$$f_t : \mathcal{M} \times \mathcal{E}^t \times \mathcal{Y}^{t-1} \to \mathcal{X}, \tag{11}$$

when the energy arrival sequence is observed noncausally, (7) is changed to

$$f_t : \mathcal{M} \times \mathcal{E}^n \times \mathcal{Y}^{t-1} \to \mathcal{X}, \tag{12}$$

and when the energy arrival sequence is deterministic, (8) is changed to

$$f_t : \mathcal{M} \times \mathcal{Y}^{t-1} \to \mathcal{X}. \tag{13}$$

The feedback capacities in the causal, noncausal, and periodic(k) cases are denoted by C_{fb}, $C_{\text{fb}}^{\text{nc}}$, and $C_{\text{fb}}^{\text{per}}(k)$, respectively.

C. Directed Information and Causal Conditioning

When characterizing capacity of channels with feedback, mutual information is no longer the right quantity. *Directed information* was introduced my Massey [28] in 1990, and has since been employed extensively to characterize the capacity of channels with feedback [29]–[34]. The directed information from X^N to Y^N is defined as

$$I(X^N \to Y^N) \triangleq \sum_{t=1}^{N} I(X^t; Y_t | Y^{t-1}). \tag{14}$$

It can be shown that in the absence of feedback, directed information is equivalent to mutual information $I(X^N; Y^N)$.

Generally, in channels with feedback the input at time t can depend on the past inputs X^{t-1} as well as past output Y^{t-1}. This is captured by the *causal conditioning* notation, introduced by Kramer [29]:

$$p(x^N \| y^{N-1}) \triangleq \prod_{t=1}^{N} p(x_t | x^{t-1}, y^{t-1}). \tag{15}$$

III. Main Results

We state the main results of the paper, which are general capacity expressions for each of the cases mentioned in the previous section. The capacity expressions are of n-letter form, therefore not computable in general. However, by focusing on specific channels – the BEC and AWGN channels in Sections IV and V respectively – we show that they can be used to deduce interesting insights regarding the capacity of such systems, as well as simple upper and lower bounds that are within a constant gap.

Theorem 1: The capacity of the RBR channel with causal energy arrival observations and without feedback is given by

$$C = \lim_{N \to \infty} \max_{\substack{p(x^N): \\ \sum_{i=1}^{N} \phi(X_i) \le \bar{B}}} \sum_{k=1}^{N} p^2 (1-p)^{k-1} I(X^k; Y^k), \tag{16}$$

and the capacity with feedback is given by

$$C_{\text{fb}} = \lim_{N \to \infty} \max_{\substack{p(x^N \| y^{N-1}): \\ \sum_{i=1}^{N} \phi(X_i) \le \bar{B}}} \sum_{k=1}^{N} p^2 (1-p)^{k-1} I(X^k \to Y^k). \tag{17}$$

The capacity expressions have an intuitive interpretation: The battery recharge times divide transmission into *epochs* – the periods between two consecutive energy arrivals. Each battery recharge effectively erases the memory in the channel, hence these epochs are essentially independent. The length of an epoch is a Geometric(p) RV, so the average rate should be given by $\sum_{k=1}^{\infty} p(1-p)^k R_k$, where R_k is the rate achieved for epoch length k. This is further normalized by the average epoch length $1/p$ (as is done in renewal theory). When there is no feedback, the rate R_k is given by $I(X^k; Y^k)$ as usual. Since the transmitter has causal observations of the energy arrivals, it does not know the epoch lengths in advance. Therefore it must have one sufficiently long code to accommodate all epoch lengths. Hence we should optimize over an input probability distribution $p(x^N)$, where N is large enough, such that the total energy does not exceed the battery capacity, which is the available energy for one epoch. A similar discussion applies when feedback is present, along with the observation that each energy arrival erases the memory in the channel, hence there is no benefit from using feedback across epochs.

The proof of Theorem 1 is provided in Appendix A. The conference version of this work [1] derived capacity expressions using an equivalence to the *clipping channel*, which is a conceptually simpler memoryless channel. The clipping channel admits real vectors as inputs, and outputs a clipped version of the input vector corrupted by noise, where the clipping length is random and corresponds to the length of each epoch, i.e. the time period between consecutive battery recharges. This equivalence enabled a simple derivation of n-letter capacity expressions for the RBR channel. In this extended version, however, we can no longer depend on this equivalence, since in the presence of feedback the transmitter can choose the input based on its past observations of the output. Therefore, the capacity expressions in this paper are derived directly, in a way which is more involved than [1] but accommodates feedback as well.

With *noncausal* observations of the energy arrivals, the capacity expressions take the following form.

Theorem 2: The capacity of the RBR channel with noncausal energy arrival observations and without feedback is given by

$$C^{\text{nc}} = \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \max_{\substack{p(x^k):\\ \sum_{i=1}^{k}\phi(X_i)\le \bar{B}}} I(X^k; Y^k), \tag{18}$$

and the capacity with feedback is given by

$$C_{\text{fb}}^{\text{nc}} = \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \max_{\substack{p(x^k\|y^{k-1}):\\ \sum_{i=1}^{k}\phi(X_i)\le \bar{B}}} I(X^k \to Y^k). \tag{19}$$

See Appendix B for the proof. Note that the difference between these expressions and the expressions in Theorem 1 is that the max is moved inside the summation (which is in fact expectation over the epoch length), and a different input distribution is assigned to each epoch length. Here, since the transmitter knows the epoch lengths ahead of time, it can choose an appropriate code designed specifically for the current epoch.

Finally, we bring the capacity expressions for the case of deterministic periodic energy arrivals.

Theorem 3: The capacity of the RBR channel with periodic energy arrivals, arriving every k time slots, without feedback, is given by

$$C^{\text{per}}(k) = \frac{1}{k} \max_{\substack{p(x^k):\\ \sum_{i=1}^{k}\phi(X_i)\le \bar{B}}} I(X^k; Y^k), \tag{20}$$

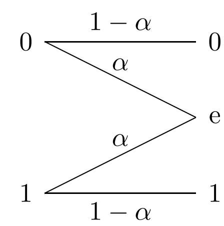

Fig. 2. Binary erasure channel.

and the capacity with feedback is given by

$$C_{\text{fb}}^{\text{per}}(k) = \frac{1}{k} \max_{\substack{p(x^k\|y^{k-1}):\\ \sum_{i=1}^{k}\phi(X_i)\le \bar{B}}} I(X^k \to Y^k). \tag{21}$$

Without feedback, by considering blocks of size k, the channel and the input constraint are both memoryless and (20) follows trivially. With feedback, the same observation applies with some modifications, which are very similar to the techniques utilized in the proofs of Theorem 1 and 2. Therefore, it will not be repeated here, and the reader is referred to [34] for more details.

The following corollary is an immediate result of Theorems 2 and 3:

Corollary 1: The noncausal capacity is given in terms of the periodic(k) capacity:

$$C^{\text{nc}} = \sum_{k=1}^{\infty} p^2(1-p)^{k-1} k C^{\text{per}}(k), \tag{22}$$

$$C_{\text{fb}}^{\text{nc}} = \sum_{k=1}^{\infty} p^2(1-p)^{k-1} k C_{\text{fb}}^{\text{per}}(k). \tag{23}$$

This corollary brings an interesting decomposition of the noncausal capacity into periodic capacities. We will use this result later in Section IV to show that properties that apply to the periodic capacity carry over to the noncausal capacity – specifically the fact that feedback can increase capacity.

IV. Binary Erasure Channel

The binary erasure channel with random battery recharges (BEC-RBR) has input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y} = \{0, 1, \text{e}\}$, and erasure probability α. The channel transition probabilities are given in Fig. 2. The battery has capacity $\bar{B} = 1$, and the cost function is $\phi(x) = x$, i.e. transmitting $x = 1$ costs 1 unit of energy, and transmitting $x = 0$ does not use up energy. Hence, the available energy in the battery is either $B_t = 1$, in which case the transmitter can send either $x = 0$ or $x = 1$, or $B_t = 0$, in which case the transmitter is constrained to send $x = 0$. Tutuncuoglu et al. [7] considered similar binary channels with a unit sized battery, with a noiseless channel and a binary symmetric channel instead of the BEC, however they did not assume knowledge of battery recharge times at the receiver.

In what follows, we will study the capacity of this channel under periodic energy arrivals with period $k = 2$, with and without feedback. As will be seen shortly, the feedback capacity is strictly larger than the no-feedback capacity, which is in direct contradiction with the claim made by Shannon [12], which was mentioned in Section I. Next, we use this result along with Corollary 1 to show that the same applies even when the energy arrivals are random and known noncausally at the transmitter.

A. Periodic Recharges

We focus on a simple case of deterministic periodic energy arrivals, with period $k = 2$. That is,

$$E_t = \begin{cases} 1, & t \text{ odd} \\ 0, & t \text{ even} \end{cases}$$

Therefore, B_t can be written as

$$B_t = \begin{cases} 1, & t \text{ odd} \\ 1 - X_{t-1}, & t \text{ even} \end{cases} \tag{24}$$

It is interesting to note that this channel, with or without feedback, is equivalent to a finite state time-invariant channel where the transmitter can compute the state from the initial state of the channel and the transmitted symbol sequence, satisfying the conditions of Shannon's claim discussed in Section I. Let $S_t = (P_t, B_t)$ where $B_t, P_t \in \{0, 1\}$ and

$$B_{t+1} = \begin{cases} 1, & P_t = 0 \\ 1 - X_t, & P_t = 1 \end{cases} \tag{25}$$

$$P_{t+1} = 1 - P_t. \tag{26}$$

Here the state variable B_t corresponds to the battery level in the energy harvesting channel and $P_t \in \{0, 1\}$ is a state variable indicating whether the time t is odd or even (P stands for "parity"). The state diagram is shown in Fig. 4. Consider a binary channel with no input constraints, but instead assume that when $B_t = 1$, the channel behaves as a standard BEC and when $B_t = 0$, the channel behaves as a BEC with $X = 0$ at its input, regardless of the actual input X_t. This channel is illustrated in Fig. 3. Assume the initial state is $s_1 = (p_1, b_1) = (1, 1)$, and it is known beforehand both at the transmitter and the receiver. Note that at odd times, the state always reverts to $s = (1, 1)$. At even times, the state is a deterministic function of the past input, therefore it is computable at the transmitter, but unknown at the receiver. It is easy to see that this time-invariant finite state channel with no input constraints is equivalent to our original BEC with periodic recharges, as codes designed for one channel can be easily translated to the other with the same probability of error.

1) Capacity Without Feedback: We denote capacity without feedback by $C^{\text{BEC}}(\alpha)$. This is given by Theorem 3:

$$\begin{aligned} C^{\text{BEC}}(\alpha) = C^{\text{per}}(2) &= \frac{1}{2} \max_{X_1+X_2 \le 1} I(X^2; Y^2) \\ &= \frac{1}{2} \max_{X^2 \ne (1,1)} I(X^2; Y^2) \\ &= \frac{1}{2} \max_{X^2 \ne (1,1)} \{H(Y^2) - H(Y^2|X^2)\} \\ &= \frac{1}{2} \max_{X^2 \ne (1,1)} H(Y^2) - h_2(\alpha), \end{aligned} \tag{27}$$

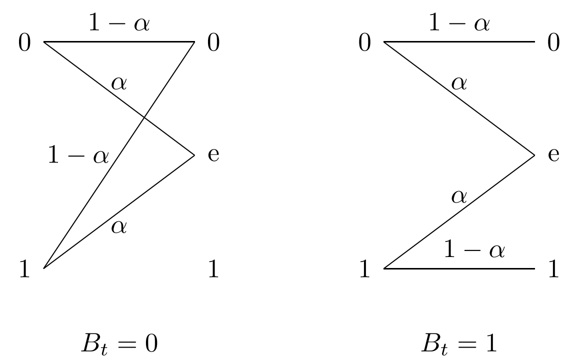

Fig. 3. Finite state binary erasure channel with periodic recharges.

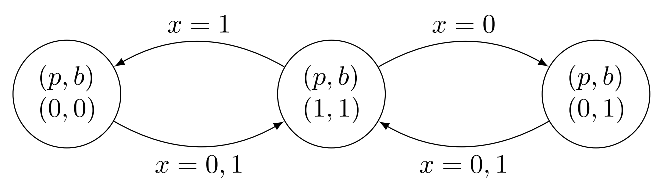

Fig. 4. State diagram of the binary erasure channel with periodic recharges.

where $h_2(\cdot)$ is the binary entropy function, i.e. $h_2(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$.

To find the optimal input distribution, we first observe that since the channel is memoryless, then by the symmetry and the concavity of the mutual information, the inputs (0, 1) and (1, 0) must have the same probability, denoted $\pi < 0.5$. Then $p(x^2 = (0, 0)) = 1 - 2\pi$. The entropy of the output can be readily computed, yielding

$$I(X^2; Y^2) = (1-\alpha)^2[h_2(2\pi) + 2\pi] + 2\alpha(1-\alpha)h_2(\pi). \tag{28}$$

This is a concave function of π. To find the maximum, we take derivative w.r.t. π and equate to 0:

$$(1-\alpha)^2 \left[2\log\frac{1-2\pi}{2\pi} + 2\right] + 2\alpha(1-\alpha)\log\frac{1-\pi}{\pi} = 0$$

$$\left(\frac{2\pi}{1-2\pi}\right) \cdot \left(\frac{\pi}{1-\pi}\right)^{\frac{\alpha}{1-\alpha}} = 2.$$

Denoting $\zeta = \frac{\pi}{1-\pi}$, we get

$$\zeta^{1/(1-\alpha)} + \zeta - 1 = 0. \tag{29}$$

This can be solved numerically for any value of $0 < \alpha < 1$. Specifically, for $\alpha = 0.5$ we can solve analytically to obtain $\pi = (3 - \sqrt{5})/2 \approx 0.382$. Substituting in the expression for capacity, we have

$$\begin{aligned} C^{\text{BEC}}(0.5) &= \frac{1}{8}[h_2(3-\sqrt{5}) + 2h_2((3-\sqrt{5})/2) + 3 - \sqrt{5}] \\ &= 0.4339. \end{aligned} \tag{30}$$

2) Capacity With Feedback: Capacity of the BEC-RBR with feedback is denoted by $C^{\text{BEC}}_{\text{fb}}(\alpha)$. By Theorem 3:

$$C^{\text{BEC}}_{\text{fb}}(\alpha) = C^{\text{per}}_{\text{fb}}(2) = \frac{1}{2} \max_{\substack{p(x^2 \| y_1): \\ X_1+X_2 \le 1}} I(X^2 \to Y^2). \tag{31}$$

The input constraint $X_1 + X_2 \leq 1$ can be equivalently written as

$$p(x_2 = 1|x_1 = 1, y_1) = 0 \qquad \forall y_1 \in \mathcal{Y}.$$

Let

$$\begin{aligned} p(x_1 = 1) &= p_1, \\ p(x_2 = 1|x_1 = 0, y_1 = 0) &= p_{20}, \\ p(x_2 = 1|x_1 = 0, y_1 = \mathrm{e}) &= p_{2e}, \end{aligned}$$

where $0 \leq p_1, p_{20}, p_{2e} \leq 1$. Then the directed information in (31) can be written as

$$\begin{aligned} I(X^2 \to Y^2) &= I(X_1; Y_1) + I(X^2; Y_2|Y_1) \\ &= H(Y_1) + H(Y_2|Y_1) - 2h_2(\alpha), \end{aligned} \tag{32}$$

where

$$\begin{aligned} H(Y_1) &= h_2(\alpha) + (1-\alpha)h_2(p_1), \\ H(Y_2|Y_1) &= (1-p_1)(1-\alpha)H(Y_2|Y_1 = 0) \\ &\quad + p_1(1-\alpha)H(Y_2|Y_1 = 1) + \alpha H(Y_2|Y_1 = \mathrm{e}). \end{aligned}$$

Clearly, $Y_1 = 1$ implies $X_1 = 1$, which in turn implies $X_2 = 0$. Therefore $H(Y_2|Y_1 = 1) = h_2(\alpha)$. When $Y_1 = 0$, the input is necessarily $X_1 = 0$, then the input X_2 is Bernoulli(p_{20}), which yields

$$H(Y_2|Y_1 = 0) = h_2(\alpha) + (1-\alpha)h_2(p_{20}). \tag{33}$$

Finally, when $Y_1 = \mathrm{e}$, we have $X_2 = 1$ w.p. p_{2e} only if $X_1 = 0$, and $X_2 = 0$ otherwise. Therefore $X_2 \sim \text{Bernoulli}\big(p_{2e}(1-p_1)\big)$, giving

$$H(Y_2|Y_1 = \mathrm{e}) = h_2(\alpha) + (1-\alpha)h_2\big(p_{2e}(1-p_1)\big). \tag{34}$$

Summing up all terms, we get

$$\begin{aligned} I(X^2 \to Y^2) = (1-\alpha)\big[&h_2(p_1) + (1-\alpha)(1-p_1)h_2(p_{20}) \\ &+ \alpha h_2\big(p_{2e}(1-p_1)\big)\big]. \end{aligned} \tag{35}$$

This can be maximized by choosing $p_{20} = 0.5$ and $p_{2e} = \min\{\frac{1}{2(1-p_1)}, 1\}$. We are then left with maximizing the expression

$$I(X^2 \to Y^2) = \begin{cases} (1-\alpha)[h_2(p_1) - (1-\alpha)p_1 + 1], \\ \qquad\qquad 0 \leq p_1 \leq 0.5 \\ (1-\alpha^2)h_2(p_1) + (1-\alpha)^2(1-p_1), \\ \qquad\qquad 5 < p_1 \leq 1 \end{cases}$$

over $p_1 \in [0, 1]$. Taking derivative over the region $p_1 \in [0, 0.5]$ gives

$$\log \frac{1-p_1}{p_1} - (1-\alpha) = 0.$$

Observe that this value of p_1 is in $[0, 0.5]$ for any value of $0 \leq \alpha \leq 1$, and since the function is concave in p_1, this is the absolute maximum. We finally get

$$C_{\mathrm{fb}}^{\mathrm{BEC}}(\alpha) = \frac{1-\alpha}{2}\big[\log(1 + 2^{1-\alpha}) + \alpha\big]. \tag{36}$$

For $\alpha = 0.5$, we get

$$C_{\mathrm{fb}}^{\mathrm{BEC}}(0.5) = 0.4429 > 0.4339 = C^{\mathrm{BEC}}(0.5). \tag{37}$$

For all other values of $0 \leq \alpha \leq 1$, the capacities with and without feedback are plotted in Fig. 5.

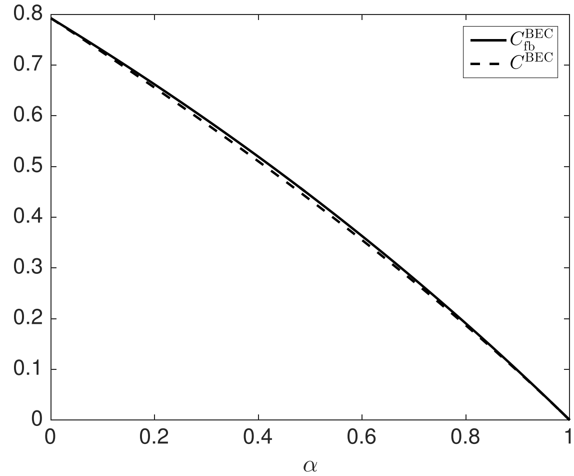

Fig. 5. Capacity of the BEC-RBR with periodical recharges, with and without feedback.

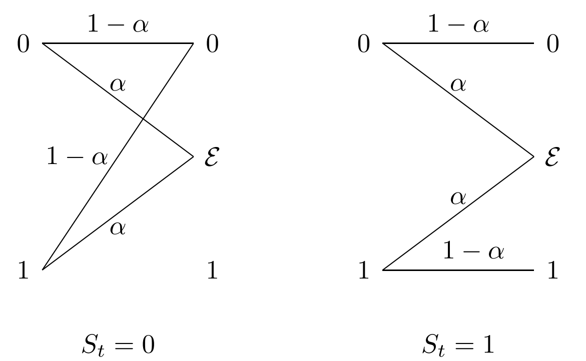

Fig. 6. Equivalent channel with i.i.d. states.

3) Equivalent State-Dependent Model: In this section, we will try to illustrate the intuition behind the usefulness of feedback in this scenario. Recall that the state of the battery is

$$B_t = \begin{cases} 1, & t \text{ odd} \\ 1 - X_{t-1}, & t \text{ even} \end{cases}$$

We focus on even times: the transmitter knows the state, and the receiver observes a noisy version of it (Y_{t-1} is the output of a BEC with $1 - B_t$ at its input).

To understand how feedback increases capacity in this case, consider the following channel with two i.i.d. state processes: the channel transition probabilities depend on S_t as in Fig. 6 (note that this is exactly the same as Fig. 3, with B_t replaced by S_t). The state $S_t = 1$ w.p. p. The second state process is $\tilde{S}_t$, given as the output of a BEC(α) with S_t at its input. This channel is of course different than the BEC-RBR and will have different capacity, however it is instructive to study this channel in order to better understand how feedback helps in our original BEC-RBR with periodic recharges.

The transmitter observes S_t causally, and the receiver observes $\tilde{S}_t$. This is the counterpart of the BEC-RBR without feedback. The capacity is obtained by using

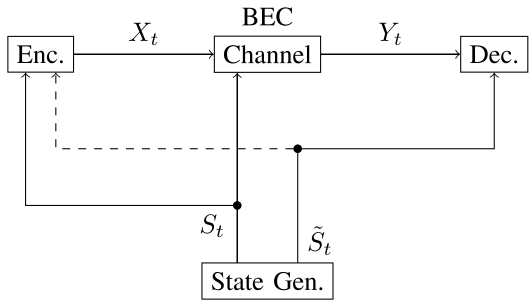

Fig. 7. Equivalent channel models. The dashed line corresponds to the channel equivalent to the case with feedback.

Shannon strategies [35], [36, Sec. 7.5]:

$$C = \max_{p(u)} I(U; Y, \tilde{S}) = \max_{p(u)} I(U; Y|\tilde{S}), \tag{38}$$

where $U : \mathcal{S} \to \mathcal{X}$. When there is feedback in the BEC-RBR, this corresponds to the transmitter observing $\tilde{S}_t$ as well. The capacity is

$$C_{\text{fb}} = \max_{p(u|\tilde{s})} I(U; Y|\tilde{S}). \tag{39}$$

The two types of channels can be seen Fig 7.

The increase in capacity follows from allowing U to depend on $\tilde{S}$. Explicit expressions are given by:

$$C = (1-\alpha) \max_{0\le r\le 1} \left[p(1-\alpha)h_2(r) + \alpha(h_2(pr) - rh_2(p))\right], \tag{40}$$

$$C_{\text{fb}} = (1-\alpha) \max_{0\le r\le 1} \left[p(1-\alpha) + \alpha(h_2(pr) - rh_2(p))\right]. \tag{41}$$

We can see that $C_{\text{fb}} \ge C$ with equality iff $r = 1/2$, which is true only when $p = 0$ or $p = 1$, or when $\alpha = 0$ or $\alpha = 1$.

This shows, at least intuitively, how feedback increases the capacity of the BEC-RBR with periodic recharges: the transmitter can increase the rate by observing the noisy version of the channel state which is observed by the receiver, simply by matching his own codeword to the noisy state.

B. Random Noncausal Energy Arrivals

We showed that feedback can increase capacity when the energy arrivals are deterministic with period 2. However, the model usually studied in the literature involves i.i.d. energy arrivals (see e.g. [7]). We will show that feedback can help in this case as well, at least when noncausal observations of the energy arrivals are available at the transmitter and the receiver.

Proposition 1: Feedback strictly increases the capacity of the BEC-RBR with noncausal energy arrival observations. That is, $C_{\text{fb}}^{\text{nc}} > C^{\text{nc}}$.

Proof: By Corollary 1, and by memorylessness and symmetry of the BEC, we have:

$$C^{\text{nc}}(\alpha) = \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \max_{\substack{p(x^k):\\ \sum_{i=1}^k X_i \le 1}} H(Y^k) - h_2(\alpha), \tag{42}$$

$$C_{\text{fb}}^{\text{nc}}(\alpha) = \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \max_{\substack{p(x^k\|y^{k-1}):\\ \sum_{i=1}^k X_i \le 1}} H(Y^k) - h_2(\alpha). \tag{43}$$

Observe that for every k we have

$$\max_{\substack{p(x^k\|y^{k-1}):\\ \sum_{i=1}^k X_i \le 1}} H(Y^k) \ge \max_{\substack{p(x^k):\\ \sum_{i=1}^k X_i \le 1}} H(Y^k), \tag{44}$$

and the results of Sections IV-A.1 and IV-A.2 imply that the inequality is strict for $k = 2$ and $\alpha = 0.5$, that is

$$\max_{\substack{p(x^2\|y_1):\\ X_1+X_2 \le 1}} H(Y^2) > \max_{\substack{p(x^2):\\ X_1+X_2 \le 1}} H(Y^2). \tag{45}$$

Therefore, we conclude that feedback can strictly increase capacity for i.i.d. energy arrivals. □

The question of whether feedback can increase capacity with causal observations of the energy arrivals remains open. Moreover, even though in this work we assumed the receiver observes the energy arrivals, it is yet unclear if feedback can help when the receiver does not (which is perhaps the more natural assumption for most energy harvesting channels).

V. GAUSSIAN CHANNEL

Although the results in Section III were derived for the DMC, they can be extended to channels with arbitrary input and output alphabets, and specifically the additive white Gaussian noise channel with random battery recharges (AWGN-RBR). This can be done as in [37, Ch. 7] by restricting the input X to a finite set of input symbols, quantizing the output Y according to some finite partition of the real line, and then applying the achievable scheme in Appendix A for the resulting DMC. By taking arbitrarily large input sets and successively refining the partition of the output, the achievable rates will converge to the expressions in Section III. See Appendix C for a detailed derivation.

Alternatively, in this section we will be interested in capacity without feedback, which can also be derived from first principles as done in [1] and [38] (using the equivalent "clipping channel"; see discussion following Theorem 1).

For the AWGN-RBR, $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and $Y_t = X_t + Z_t$, where $Z_t \sim \mathcal{N}(0, 1)$ independent of the input. The energy cost function is quadratic, i.e. $\phi(x) = x^2$. The capacity without feedback is given by analogs of Theorems 1 and 2:

$$C = \lim_{N\to\infty} \max_{\substack{p(x^N):\\ \|X^N\|^2 \le \bar{B}}} \sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k; Y^k), \tag{46}$$

$$C^{\text{nc}} = \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \max_{\substack{p(x^k):\\ \|X^k\|^2 \le \bar{B}}} I(X^k; Y^k), \tag{47}$$

where $\|X^k\|^2 \triangleq \sum_{i=1}^k X_i^2$.

We will first use these expressions in Section V-A to show that noncausal observations of the energy arrivals strictly increase capacity, as compared to causal observations. Next, in Section V-B, we will find computable upper and lower bounds to both capacity expressions which are separated by a constant gap, independent of problem parameters.

A. Noncausal Side Information Strictly Increases Capacity

It is possible to explicitly identify the maximizing input distribution in (47) by using the results of [39]–[41], which characterize the capacity of amplitude-constrained channels. In particular, [41] shows that the maximizing X^k in (47) is distributed over a finite set of k-dimensional spheres with uniform phase, where the number of spheres is determined by the value of $\bar{B}$ (ex. when $\bar{B}$ is very small, X^k is uniformly distributed over a single sphere of radius $\sqrt{\bar{B}}$). Using this result, we suggest the following proposition.

Proposition 2: Noncausal observations of the energy arrival process strictly increase capacity of the AWGN-RBR. That is, $C^{\rm nc} > C$.

This result may be surprising given that for a memoryless channel with i.i.d. state S, the capacity with side information at both the transmitter and the receiver is given by $I(X;Y|S)$, whether the side information is available causally or noncausally. The difference here is that even though the battery recharges E_t are i.i.d. and known to both the transmitter and the receiver, the state of the system is captured by B_t rather than E_t, which has memory and which is unknown to the receiver due to its input-dependence. The fact that noncausal knowledge of the energy arrivals strictly increases capacity can also be observed by using the upper and lower bounds on the causal and noncausal capacities developed in the next section.

Proof of Proposition 2: We would like to show that $C < C^{\rm nc}$, where C and $C^{\rm nc}$ are given by (46) and (47) respectively.

Let $g_n(x^n)$ and $f_k(x^k)$ be the maximizing distributions in (46) and (47) respectively, that is:

$$g_n(x^n) = \underset{p(x^n):\ \|X^n\|^2\le\bar{B}}{\operatorname{argmax}} \left\{\sum_{k=1}^{n}(1-p)^{k-1}I(X^k;Y^k)\right\}, \quad (48)$$

$$f_k(x^k) = \underset{p(x^k):\ \|X^k\|^2\le\bar{B}}{\operatorname{argmax}} I(X^k;Y^k). \quad (49)$$

The maximizing distributions $f_k(x^k)$ are unique and are found explicitly in [39]–[41]. For any $1 \le l \le k$, let $f_k(x^l)$ denote the marginal distribution of $f_k(x^k)$, that is

$$f_k(x^l) = \int f_k(x^k)dx_{l+1}^k,$$

and similarly for $g_n(x^l)$. Denote $I(p(X^k))$ as the mutual information $I(X^k;Y^k)$ computed with input distribution $p(x^k)$.

Using this notation, we rewrite (46) and (47) as follows:

$$C = \lim_{n\to\infty}\sum_{k=1}^{n}p^2(1-p)^{k-1}I(g_n(X^k)), \quad (50)$$

$$C^{\rm nc} = \lim_{n\to\infty}\sum_{k=1}^{n}p^2(1-p)^{k-1}I(f_k(X^k)). \quad (51)$$

Obviously $C \le C^{\rm nc}$. Suppose that also $C = C^{\rm nc}$. This will imply

$$\lim_{n\to\infty}\sum_{k=1}^{n}(1-p)^{k-1}[I(f_k(X^k)) - I(g_n(X^k))] = 0. \quad (52)$$

Since each term in the sum is non-negative, we get in particular that the sum of the first two elements must vanish, or

$$\lim_{n\to\infty}\{I(g_n(X_1)) + (1-p)I(g_n(X^2))\} = I(f_1(X_1)) + (1-p)I(f_2(X^2)).$$

Next, consider $g_2(x^2)$ as defined in (48). Since $\|X^n\|^2 \le \bar{B}$ implies $\|X^2\|^2 \le \bar{B}$, we get for every $n \ge 2$:

$$\begin{aligned} I(g_n(X_1)) &+ (1-p)I(g_n(X^2)) \\ &\le \max_{p(x^2):\ \|X^2\|^2\le\bar{B}} \left\{I(X_1;Y_1) + (1-p)I(X^2;Y^2)\right\} \\ &= I(g_2(X_1)) + (1-p)I(g_2(X^2)), \end{aligned}$$

which implies

$$I(f_1(X_1)) + (1-p)I(f_2(X^2)) \le I(g_2(X_1)) + (1-p)I(g_2(X^2)). \quad (53)$$

Next, since $\|X^2\|^2 \le \bar{B}$ implies $|X_1|^2 \le \bar{B}$, we get that $I(g_2(X_1)) \le I(f_1(X_1))$. Substituting in (53), we get $I(f_2(X^2)) \le I(g_2(X^2))$. Since $f_2(x^2)$ is the unique maximizer of $I(X^2;Y^2)$, this implies $f_2(x^2) = g_2(x^2)$. Substituting this back in (53), we get $I(f_1(X_1)) \le I(g_2(X_1))$. Again, from uniqueness, this implies $f_1(x) = g_2(x)$. Together, we see that

$$f_1(x_1) = g_2(x_1) = \int g_2(x_1,x_2)dx_2 = \int f_2(x_1,x_2)dx_2, \quad (54)$$

which is a contradiction, since $f_1(x_1)$ is discrete [39], whereas $f_2(x_1,x_2)$ has discrete amplitude and uniform phase [40]. Therefore we must have $C < C^{\rm nc}$. □

B. Capacity Bounds

Despite being relatively simpler than previous results,[3] (46) and (47) are difficult to compute explicitly. In particular, (46) is a multi-letter expression that involves optimization over an infinite dimensional space. Therefore, we wish to find suitable approximations. More specifically, we provide an upper and a lower bound, separated by a constant gap of approximately 1.05 bits:

Proposition 3 (Capacity Bounds): The capacity of the AWGN-RBR channel with causal energy arrival observations is bounded by:

$$\bar{C} - \frac{1}{2}\log\left(\frac{\pi e}{2}\right) \le C \le \bar{C}, \quad (55)$$

where

$$\bar{C} \triangleq \lim_{N\to\infty} \max_{\substack{\{\mathcal{E}_i\}_{i=1}^N:\\ \mathcal{E}_i\ge 0\ ,i=1,\ldots,N\\ \sum_{i=1}^N \mathcal{E}_i\le\bar{B}}} \sum_{i=1}^{N} p(1-p)^{i-1}\frac{1}{2}\log(1+\mathcal{E}_i). \quad (56)$$

The proof is provided at the end of this section.

[3]For example, [11] characterizes the capacity of the energy harvesting AWGN channel with general i.i.d. energy arrival distribution, in the form

$$C = \lim_{n\to\infty}\frac{1}{n}\sup I(X^n;Y^n|E^n),$$

where the domain of the optimization problem is suitably defined. Note that the capacity expressions in (46) and (47) are much more explicit, and in particular, it is this explicit form that allows us to identify the maximizing input distribution in (47).

It can be shown that the upper bound $\bar{C}$ in (56) corresponds to the online power control problem, extensively studied in the literature in the general framework of energy-harvesting channels [14]–[16], [42]. Here, one assumes that there is an underlying transmission scheme operating at a finer time-scale, such that allocating power P to this scheme yields an information rate $r(P) = \frac{1}{2}\log(1+P)$, and focuses on the optimal power allocation policy satisfying the energy constraints on the transmitter. For the specific channel of interest here, this online power control problem can be explicitly solved [42], [43]. In particular, we can apply the KKT conditions to the optimization problem in (56), to obtain the optimal values of $\mathcal{E}_i$ (see Appendix D):

$$\mathcal{E}_i = \begin{cases} (\tilde{N}+\bar{B})\dfrac{p(1-p)^{i-1}}{1-(1-p)^{\tilde{N}}} - 1, & i = 1,\dots,\tilde{N} \\ 0, & i > \tilde{N} \end{cases} \tag{57}$$

where $\tilde{N}$ is the smallest positive integer satisfying

$$1 > (1-p)^{\tilde{N}}[1 + p(\bar{B}+\tilde{N})].$$

This gives the following expression for $\bar{C}$:

$$\bar{C} = \frac{1-(1-p)^{\tilde{N}}}{2}\log\left(\frac{p(\bar{B}+\tilde{N})}{1-(1-p)^{\tilde{N}}}\right) \tag{58}$$

$$+\frac{1-p-(1-p)^{\tilde{N}}(1-p+\tilde{N}p)}{2p}\log(1-p). \tag{59}$$

Combined with (55), this is the capacity of the AWGN-RBR channel within 1.05 bits/channel use.

It was shown in [5] that the capacity of an AWGN energy harvesting channel with infinite battery size is $\frac{1}{2}\log(1+\mathbb{E}[E_t])$. Clearly, this is an upper bound to the capacity of our channel, and this can be readily obtained from the result of Proposition 3. Using concavity of the log function in (56):

$$\bar{C} \le \lim_{N\to\infty} \max_{\substack{\{\mathcal{E}_i\}_{i=1}^N: \\ \mathcal{E}_i \ge 0 \;, i=1,\dots,N \\ \sum_{i=1}^N \mathcal{E}_i \le \bar{B}}} \frac{1}{2}\log\left(1 + \sum_{i=1}^N p(1-p)^{i-1}\mathcal{E}_i\right)$$

$$= \frac{1}{2}\log(1+p\bar{B}), \tag{60}$$

where the last step follows because the optimal values for the first line are $\mathcal{E}_1 = \bar{B}$ and $\mathcal{E}_i = 0$ for $i \ge 2$. Dong and Özgür [9] used this upper bound corresponding to infinite battery size to bound the capacity of the energy harvesting channel with Bernoulli energy arrivals. Fig. 8 illustrates that the upper bound we provide here is strictly smaller than the infinite battery upper bound. Similarly, our lower bound here is based on the optimal power allocation strategy we characterize in (57), while the lower bound in [9] is based on a suboptimal power allocation policy.

Similar bounds can be obtained for (47), which we state in the following proposition.

Proposition 4: The capacity of the AWGN-RBR channel with noncausal energy arrival observations is bounded by:

$$\bar{C}^{\text{nc}} - \frac{1}{2}\log\left(\frac{\pi e}{2}\right) \le C^{\text{nc}} \le \bar{C}^{\text{nc}}, \tag{61}$$

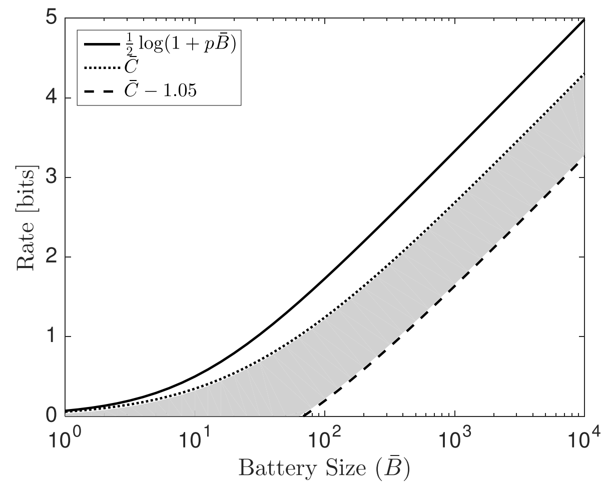

Fig. 8. Upper and lower bounds for $p = 0.1$. The shaded region indicates where the capacity of the energy harvesting channel can lie.

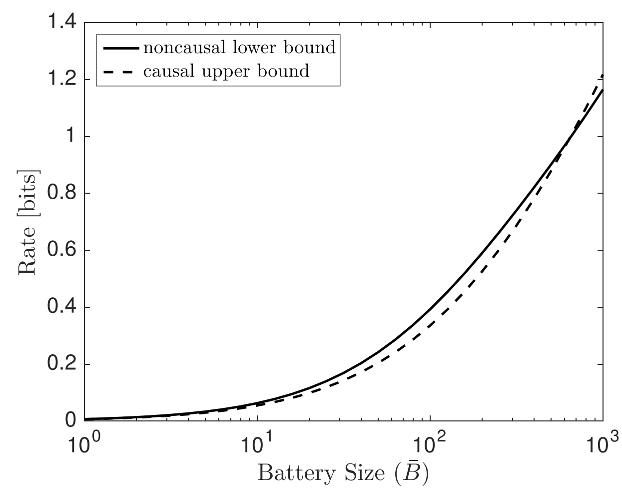

Fig. 9. Noncausal lower bound and causal upper bound for $p = 0.01$. Noncausal capacity is strictly greater than causal capacity for some values of $\bar{B}$.

where

$$\bar{C}^{\text{nc}} = \sum_{k=1}^{\infty} p^2(1-p)^{k-1}\frac{k}{2}\log(1+\bar{B}/k). \tag{62}$$

The proof can be found in Appendix E, where we also consider another lower bound for the capacity with noncausal energy arrival information at the transmitter, which is tighter than the one in (61), and plot it in Fig. 9 together with the upper bound in (55) on the capacity with causal energy arrival information. It is clear from the graph that for some values of $\bar{B}$, the noncausal capacity is strictly greater than the causal capacity, further illustrating the observation we state in Proposition 2.

Proof of Proposition 3:

1) Upper Bound: We can relax the energy constraint in (46) to be only in expectation, thus giving an upper bound:

$$C \le \lim_{N\to\infty} \max_{\substack{p(x^N): \\ \mathbb{E}\|X^N\|^2 \le \bar{B}}} \sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k; Y^k)$$

$$\stackrel{\text{(i)}}{\le} \lim_{N\to\infty} \max_{\substack{p(x^N):\\ \mathbb{E}\|X^N\|^2\le\bar{B}}} \sum_{k=1}^{N} p^2(1-p)^{k-1}\sum_{i=1}^{k} I(X_i;Y_i)$$
$$\stackrel{\text{(ii)}}{=} \lim_{N\to\infty} \max_{\substack{p(x^N):\\ \mathbb{E}\|X^N\|^2\le\bar{B}}} \sum_{i=1}^{N}\sum_{k=i}^{N} p^2(1-p)^{k-1} I(X_i;Y_i)$$
$$= \lim_{N\to\infty} \max_{\substack{p(x^N):\\ \mathbb{E}\|X^N\|^2\le\bar{B}}} \sum_{i=1}^{N} p(1-p)^{i-1}[1-(1-p)^{N-i+1}] I(X_i;Y_i)$$
$$\le \lim_{N\to\infty} \max_{\substack{p(x^N):\\ \mathbb{E}\|X^N\|^2\le\bar{B}}} \sum_{i=1}^{N} p(1-p)^{i-1} I(X_i;Y_i)$$
$$\stackrel{\text{(iii)}}{=} \lim_{N\to\infty} \max_{\substack{\{\mathcal{E}_i\}_{i=1}^N:\\ \mathcal{E}_i\ge 0,\ i=1,\dots,N\\ \sum_{i=1}^N \mathcal{E}_i\le\bar{B}}} \sum_{i=1}^{N} p(1-p)^{i-1}\frac{1}{2}\log(1+\mathcal{E}_i),$$

where (i) is because the channel is memoryless; (ii) is by changing order of summation; and (iii) is obtained by choosing $X_i \sim \mathcal{N}(0,\mathcal{E}_i)$ independent of each other. Note that the limit exists in each step because the sequences are non-decreasing and bounded above. This gives the RHS in (55).

2) Lower Bound: To lower bound (46), we can choose a suboptimal distribution for which the X_i's are independent, i.e. $p(x^N) = \prod_{i=1}^N p(x_i)$, and each of them satisfies $|X_i|^2 \le \mathcal{E}_i$ a.s. for some $\mathcal{E}_i \ge 0$. To satisfy the total energy constraint we must have $\sum_{i=1}^N \mathcal{E}_i \le \bar{B}$. Under this input distribution, we have for every i: $I(X_i;Y_i) \le \frac{1}{2}\log(1+\mathcal{E}_i) \le \frac{1}{2}\log(1+\bar{B})$, and thus for every N:

$$\sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k;Y^k)$$
$$= \sum_{k=1}^{N} p^2(1-p)^{k-1}\sum_{i=1}^{k} I(X_i;Y_i)$$
$$= \sum_{i=1}^{N}\sum_{k=i}^{N} p^2(1-p)^{k-1} I(X_i;Y_i)$$
$$= \sum_{i=1}^{N} p(1-p)^{i-1}[1-(1-p)^{N-i+1}] I(X_i;Y_i)$$
$$\ge \sum_{i=1}^{N} p(1-p)^{i-1} I(X_i;Y_i) - Np(1-p)^N\frac{1}{2}\log(1+\bar{B}).$$

Taking $N\to\infty$, the second term vanishes, and we are left with the following lower bound:

$$C \ge \lim_{N\to\infty} \max_{\substack{\{\mathcal{E}_i\}_{i=1}^N:\\ \mathcal{E}_i\ge 0,\ i=1,\dots,N\\ \sum_{i=1}^N \mathcal{E}_i\le\bar{B}}} \sum_{i=1}^{N} p(1-p)^{i-1} I(X_i;Y_i).$$

Again, the sequence is non-decreasing and bounded above, so the limit exists. Since $p(x_i)$ was arbitrary, we can choose it to maximize $I(X_i;Y_i)$. We obtain

$$C \ge \lim_{N\to\infty} \max_{\substack{\{\mathcal{E}_i\}_{i=1}^N:\\ \mathcal{E}_i\ge 0,\ i=1,\dots,N\\ \sum_{i=1}^N \mathcal{E}_i\le\bar{B}}} \sum_{i=1}^{N} p(1-p)^{i-1} C_{\text{Smith}}(\mathcal{E}_i), \quad (63)$$

where

$$C_{\text{Smith}}(\mathcal{E}) \triangleq \max_{p(x):\ X^2\le\mathcal{E}} I(X;Y) \quad (64)$$

is the capacity of the amplitude constrained scalar AWGN channel studied in [39], where the optimal value for this mutual information maximization problem is found. Unfortunately, it does not have a closed-form expression, however it can be lower bounded using the following lemma.

Lemma 1: The capacity of the amplitude constrained scalar AWGN channel with noise variance 1 can be lower bounded as follows:

$$C_{\text{Smith}}(\mathcal{E}) \ge \frac{1}{2}\log(1+\mathcal{E}) - \frac{1}{2}\log\left(\frac{\pi e}{2}\right). \quad (65)$$

Proof: Let X be uniform on the interval $[-\sqrt{S},\sqrt{S}]$. Then, using the entropy power inequality:

$$\begin{aligned} I(X;X+N) &= h(X+N) - h(N)\\ &\ge \frac{1}{2}\log\left(2^{2h(X)}+2^{2h(N)}\right) - h(N)\\ &= \frac{1}{2}\log(4S+2\pi e) - \frac{1}{2}\log(2\pi e)\\ &= \frac{1}{2}\log\left(1+\frac{2S}{\pi e}\right)\\ &\ge \frac{1}{2}\log(1+S) - \frac{1}{2}\log\left(\frac{\pi e}{2}\right). \end{aligned}$$

□

Plugging (65) into (63) gives the LHS of (55). □

VI. Conclusion

In this paper, motivated by recent developments in energy harvesting and remotely powered wireless radios, we studied the capacity of a communication channel with periodic and random i.i.d. battery recharges. We derived n-letter expressions for the capacity and used them to show that feedback can strictly increase the capacity of an energy harvesting channel, providing a counter example to an old claim by Shannon, and that noncausal knowledge of the battery recharges can strictly increase capacity over causal knowledge. An interesting extension of our setting is provided in [44], where the charger, and its corresponding charging strategy, and the transmitter are jointly optimized for maximum capacity. When the charger is the receiver itself, this more general framework allows to embed feedback information in the charging strategy.

Appendix A
Capacity With Causal Energy Arrival Observations: Proof of Theorem 1

First, we will show that the limits in (16) and (17) exist. To this end, observe that (16) is a limit of a non-decreasing sequence: Let $p^\star(x^N)$ be the maximizing distribution in (16) for a fixed N, and let $X_{N+1} = 0$ w.p. 1 (recall we assumed the existence of the symbol $x = 0$ in Section II-B). The resulting distribution $p(x^{N+1})$ satisfies $\sum_{k=1}^{N+1} \phi(X_k) \leq \bar{B}$, and we have

$$\begin{aligned}\sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k; Y^k) &\leq \sum_{k=1}^{N+1} p^2(1-p)^{k-1} I(X^k; Y^k) \\ &\leq \max_{\substack{p(x^{N+1}): \\ \sum_{k=1}^{N+1} \phi(X_k) \leq \bar{B}}} \sum_{k=1}^{N+1} p^2(1-p)^{k-1} I(X^k; Y^k),\end{aligned}$$

which is the desired result. Similarly, the sequence in (17) is non-decreasing. Next, observe that both sequences in (16) and (17) are upper bounded by $\max_{p(x): \phi(X) \leq \bar{B}} I(X; Y)$ for every N. Therefore, the limits exist and are given by:

$$C = \sup_{N\geq 1} \max_{\substack{p(x^N): \\ \sum_{k=1}^{N} \phi(X_k) \leq \bar{B}}} \sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k; Y^k), \quad (66)$$

$$C_{\text{fb}} = \sup_{N\geq 1} \max_{\substack{p(x^N \| y^{N-1}): \\ \sum_{k=1}^{N} \phi(X_k) \leq \bar{B}}} \sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k \to Y^k). \quad (67)$$

We prove Theorem 1 for the feedback case first.

A. Achievability

Transmission takes place during n time slots. Communication is divided into *epochs*, where an epoch refers to the time between two consecutive energy arrivals. Each energy arrival fully recharges the battery. Therefore, the memory, which is embedded in the state of the battery, is essentially erased. Hence, we think of each epoch as a super-symbol, and the epoch length can be thought of as a random state of the channel, which determines the size of the super-symbol. This state is available only at the receiver, since the transmitter does not know in advance what the epoch length will be.

Next, for each epoch length, or state, we use the technique in [36, Sec. 17.6.3] and [32]. The epochs are divided into *interleaved* blocks, so that the first block consists of all the symbols that appear first in each epoch, the second block consists of all the symbols that appear second in those epochs which are of length at least 2, and so forth. Note that each block has a different length, and this length is a random quantity. In each block, the channel input and output sequences from previous blocks are treated as causal side information available at both the transmitter and the receiver. This technique can be implemented *online*, i.e. the transmitter does not need to know the length of the current epoch in order to code correctly.

Rate Splitting: Fix N. Divide the message w into N independent messages $(w_1, \ldots, w_N)$, one for each of the N states of the super-channel. In practice the epoch length can be larger than N; in that case, we will treat it as N. Let $S_i = (X^{i-1}, Y^{i-1})$. Divide each message w_i into the messages $(w_i(s_i) : s_i \in \mathcal{S}_i = \mathcal{X}^{i-1} \times \mathcal{Y}^{i-1})$. Thus, $R = \sum_{i=1}^{N} \sum_{s_i \in \mathcal{S}_i} R_i(s_i)$.

Codebook Generation: Fix a causally conditioned pmf $p(x^N \| y^{N-1}) = \prod_{i=1}^{N} p(x_i | x^{i-1}, y^{i-1})$ satisfying $\sum_{i=1}^{N} \phi(X_i) \leq \bar{B}$ w.p. 1. For every $i \in \{1, \ldots, N\}$ and $s_i \in \mathcal{S}_i$, randomly and independently generate $2^{nR_i(s_i)}$ codewords $x^n(w_i(s_i), s_i)$, each i.i.d. according to $p(x_i | x^{i-1}, y^{i-1})$.

Encoding: At time t, the transmitter observes the energy arrivals e^t, the past input symbols x^{t-1}, and the past output symbols y^{t-1}. Define $j_t = j_t(e^t)$ as the time since the last energy arrival (including the current time slot):

$$j_t = \{\min\ 1 \leq \tau \leq t :\ e_{t-\tau+1} = \bar{B}\},$$

If $j_t > N$, the zero symbol is transmitted: $x_t = 0$. Otherwise, the transmitter treats the sequences

$$(x_{t-j_t+1}^{t-1}, y_{t-j_t+1}^{t-1}) \in \mathcal{S}_{j_t}$$

as causal state information, available at both the transmitter and the receiver. As in [36, Sec. 7.4.1], the codewords are stored in a FIFO buffer, and at time t, the first untransmitted symbol corresponding to the current state is transmitted. Note that this guarantees the energy constraint (3) is satisfied.

Decoding Step 1 (Demultiplexing): The decoder observes (y^n, e^n). The sequence e^n can be mapped in a sequence of epoch lengths $\ell^m = (\ell_1, \ldots, \ell_m)$, where $m = m(e^n)$ is the total number of energy arrivals, $m(e^n) = \sum_{t=1}^{n} 1\{e_t = \bar{B}\}$. Denote by $\tau_1 < \tau_2 < \ldots < \tau_m$ the energy arrival times, that is

$$E_t = \begin{cases} \bar{B}, & t \in \{\tau_1, \ldots, \tau_m\} \\ 0, & \text{otherwise} \end{cases}$$

Recall that by assumption, $e_1 = \bar{B}$, hence $\tau_1 = 1$. The epoch length $\ell_i = \ell_i(e^n)$ is the time between the i-th and $(i+1)$-th energy arrivals: $\ell_i = \tau_{i+1} - \tau_i$ for $i = 1, \ldots, m-1$, and we let $\ell_m = n + 1 - \tau_m$. The decoder demultiplexes the sequence y^n into subsequences of various super-symbols in the following way: For $1 \leq k < N$, the sequence $\mathbf{y}^k = ((y^k)_1, \ldots, (y^k)_{m_k})$ consists of all the tuples y^k that form epochs of length k. For $k = N$, the sequence $\mathbf{y}^N$ consists of all the tuples y^N that form the first N symbols of epochs of length N or greater. We denote the length of the sequence $\mathbf{y}^k$ by m_k, which is exactly the number of epochs of length k for $k < N$, and m_N is the number of epochs of length $\geq N$. Hence $m = \sum_{k=1}^{N} m_k$. The empirical distribution of the epoch lengths is given by

$$\pi(k|\ell^m) = \frac{m_k}{m} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}\{\ell_i \wedge N = k\}, \qquad k = 1, \ldots, N,$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and $a \wedge b \triangleq \min(a, b)$. Let $q(k)$ be the pmf of $L \wedge N$, where $L \sim \text{Geometric}(p)$:

$$q(k) = \begin{cases} p(1-p)^{k-1}, & k = 1, \ldots, N-1 \\ (1-p)^{N-1}, & k = N \end{cases} \quad (68)$$

Since there is a one-to-one mapping between the sequence e^n and the sequence ℓ^m, we define the following typical set in $\mathcal{E}^n$ with a slight abuse of notation:

$$\mathcal{T}_\epsilon^{(n)}(L) = \{e^n :\ |m/n - p| \leq \epsilon p \text{ and } |\pi(k|\ell^m) - q(k)| \leq \epsilon q(k), \forall 1 \leq k \leq N\},$$

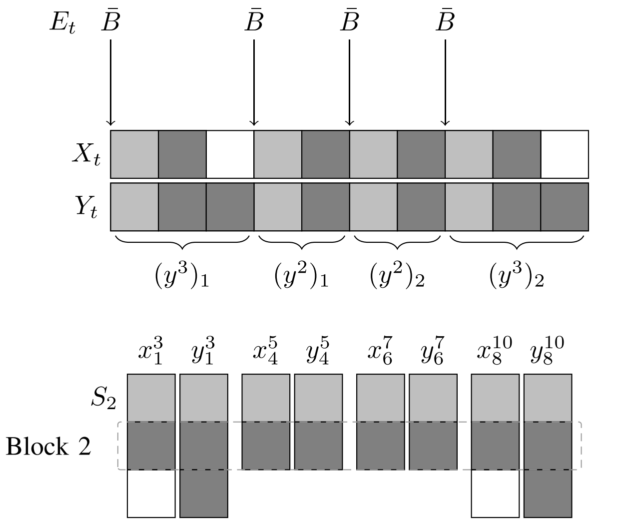

Fig. 10. Illustration of the demultiplexing and decoding performed at the receiver. The top part shows the signals X_t and Y_t, both of length 10, along with the energy arrivals at times 1, 4, 6, and 8. The receiver demultiplexes the signal Y^n to obtain the bottom part of the figure. The bottom part illustrates decoding of block 2: In this specific example, the receiver decides on a codeword $x^n(\hat{w}_2)$ using the state information S_2. For the purpose of this illustration, suppose the state S_2 is the same for all symbols of block 2, i.e. $x_1 = x_4 = x_6 = x_8$ and $y_1 = y_4 = y_6 = y_8$. Then, the receiver will look for a codeword $\hat{x}^4$ for which the pair $(\hat{x}_1, \hat{x}_4)$ is jointly typical with (y_2^3, y_9^{10}), while simultaneously the pair $(\hat{x}_2, \hat{x}_3)$ is jointly typical with (y_5, y_7).

which is the set of all typical e^n sequences for which $\ell^m = \ell^m(e^n)$ is also typical. The receiver declares an error if $e^n \notin \mathcal{T}_\epsilon^{(n)}(L)$. Otherwise, it continues to the next decoding step.

Decoding Step 2 (Successive Decoding): The decoder recovers $w_1, \ldots, w_N$ successively. We start with block 1, which consists of the first symbol in each epoch. Note that there are $m = \sum_{k=1}^N m_k$ such symbols. Assuming $e^n \in \mathcal{T}_\epsilon^{(n)}(L)$, we have

$$m_k \geq m(1-\epsilon)q(k) \geq n(1-\epsilon)^2 pq(k), \qquad \forall k \in \{1, \ldots, N\}.$$

The decoder finds $\hat{w}_1$ such that the codeword subsequences $x^{n(1-\epsilon)^2 pq(k)}$, consisting of the first symbol of each epoch of length k, are jointly typical with $(y^k)^{n(1-\epsilon)^2 pq(k)}$, simultaneously for all $k = 1, \ldots, N$. Note that for successful decoding to occur, there has to be one such $\hat{w}_1$ for which all the appropriate subsequences of $x^n(\hat{w}_1)$ are jointly typical with the output subsequences as described above. Next, for block j, $1 < j \leq N$, which consists of the j-th symbol in each epoch of length $\geq j$, the decoder forms the state sequence estimate $\hat{\mathbf{s}}_j = (\hat{\mathbf{x}}^{j-1}, \mathbf{y}^{j-1})$ from the codeword estimates and output sequences from the previous blocks. It then further demultiplexes each of the output subsequences into $|\mathcal{S}_j|$ sub-subsequences accordingly, and finds a unique index $\hat{w}_j(s_j)$ for each s_j, according to joint typicality across all subsequences for $k = j, \ldots, N$ as was done for the first block. Note that subsequences $1, \ldots, j-1$ are not included since block j does not contain symbols from epochs of length less than j. The demultiplexing and decoding steps are illustrated in Fig. 10.

Analysis of the Probability of Error: By the strong law of large numbers $m(e^n)/n \to p$ a.s., hence $m \to \infty$, which implies $\pi(k|\ell^m) \to q(k)$ a.s. Therefore $E^n \in \mathcal{T}_\epsilon^{(n)}$ with high probability. Next, by the i.i.d. nature of the encoding and the memorylessness of the channel, the probability of error for decoding w_1 is

$$2^{nR_1} \cdot 2^{-n(1-\epsilon)^2 pq(1) I(X_1;Y_1)} \times 2^{-n(1-\epsilon)^2 pq(2) I(X_1;Y^2)} \cdots 2^{-n(1-\epsilon)^2 pq(N) I(X_1;Y^N)},$$

which goes to zero if $R_1 < \sum_{k=1}^N pq(k) I(X_1; Y^k)$. Similarly, for block j, $1 < j \leq N$, we get vanishing probability of error if

$$\begin{aligned} R_j &< \sum_{k=j}^N pq(k) I(X_j; Y_j^k | S_j) \\ &= \sum_{k=j}^N pq(k) I(X_j; Y_j^k | X^{j-1}, Y^{j-1}) \\ &= \sum_{k=j}^N pq(k) \sum_{i=j}^k I(X_j; Y_i | X^{j-1}, Y^{i-1}). \end{aligned}$$

The total rate must satisfy

$$\begin{aligned} R &< \sum_{j=1}^N \sum_{k=j}^N pq(k) \sum_{i=j}^k I(X_j; Y_i | X^{j-1}, Y^{i-1}) \\ &= \sum_{k=1}^N pq(k) \sum_{j=1}^k \sum_{i=j}^k I(X_j; Y_i | X^{j-1}, Y^{i-1}) \\ &= \sum_{k=1}^N pq(k) \sum_{i=1}^k \sum_{j=1}^i I(X_j; Y_i | X^{j-1}, Y^{i-1}) \\ &= \sum_{k=1}^N pq(k) \sum_{i=1}^k I(X^i; Y_i | Y^{i-1}) \\ &= \sum_{k=1}^N pq(k) I(X^k \to Y^k). \end{aligned}$$

By (68), $q(k) \geq p(1-p)^{k-1}$ for all $1 \leq k \leq N$, hence

$$C_{\text{fb}} \geq \max_{\substack{p(x^N \| y^{N-1}): \\ \sum_{k=1}^N \phi(X_k) \leq \bar{B}}} \sum_{k=1}^N p^2 (1-p)^{k-1} I(X^k \to Y^k) \tag{69}$$

for every $N \geq 1$, which implies the achievability part of (67).

B. Converse

By Fano's inequality:

$$\begin{aligned} nR &\leq I(W; Y^n, E^n) + n\epsilon_n \\ &= I(W; Y^n | E^n) + n\epsilon_n \\ &= \sum_{t=1}^n I(W; Y_t | Y^{t-1}, E^n) + n\epsilon_n \\ &\overset{\text{(i)}}{=} \sum_{t=1}^n I(X_t; Y_t | Y^{t-1}, E^n) + n\epsilon_n \\ &\overset{\text{(ii)}}{\leq} \sum_{t=1}^n I(X_t; Y_t | Y^{t-1}, E^t) + n\epsilon_n \end{aligned} \tag{70}$$

where (i) and (ii) are because $X_t = f_t(W, E^t, Y^{t-1})$ and because the channel is memoryless. The final term in (70) can be written as:

$$\begin{aligned} I(X_n; Y_n|Y^{n-1}, E^n) &= pI(X_n; Y_n|Y^{n-1}, E^{n-1}, e_n = \bar{B}) \\ &\quad + (1-p)I(X_n; Y_n|Y^{n-1}, E^{n-1}, e_n = 0). \end{aligned} \tag{71}$$

We upper bound the first term as follows:

$$\begin{aligned} pI(X_n; Y_n|Y^{n-1}, E^{n-1}, e_n = \bar{B}) &\overset{\text{(i)}}{\leq} p[H(Y_n) - H(Y_n|X_n)] \\ &\overset{\text{(ii)}}{\leq} \max_{p(x):\ \phi(X)\leq\bar{B}} pI(X;Y) \\ &\triangleq a_1, \end{aligned} \tag{72}$$

where (i) is because the channel is memoryless and (ii) is because $e_n = \bar{B}$ implies $B_n = \bar{B}$. Next, we can write the second term in (71) as follows:

$$\begin{aligned} &(1-p)I(X_n; Y_n|Y^{n-1}, E^{n-1}, e_n = 0) \\ &\quad = p(1-p)I(X_n; Y_n|Y^{n-1}, E^{n-2}, e_{n-1}^n = [\bar{B}\ 0]) \\ &\quad\quad + (1-p)^2 I(X_n; Y_n|Y^{n-1}, E^{n-2}, e_{n-1}^n = [0\ 0]). \end{aligned} \tag{73}$$

Similarly, the $t = n-1$ term in (70) can be written as:

$$\begin{aligned} &I(X_{n-1}; Y_{n-1}|Y^{n-2}, E^{n-1}) \\ &\quad = pI(X_{n-1}; Y_{n-1}|Y^{n-2}, E^{n-2}, e_{n-1} = \bar{B}) \\ &\quad\quad + (1-p)I(X_{n-1}; Y_{n-1}|Y^{n-2}, E^{n-2}, e_{n-1} = 0). \end{aligned} \tag{74}$$

The first term in (73) and the first term in (74) can be upper bounded in a similar manner to (72), which is done in equations (75)–(77) at the bottom of the page. The input pmf $p(x^2\|y_1) = p(x_1)p(x_2|x_1, y_1)$ in (76) is due to the encoder function (11).

Combining (71)–(74) and (77), we obtain the following upper bound on the last two terms of (70):

$$\begin{aligned} &\sum_{t=n-1}^{n} I(X_t; Y_t|Y^{t-1}, E^t) \\ &\quad \leq a_1 + a_2 \\ &\quad\quad + (1-p)I(X_{n-1}; Y_{n-1}|Y^{n-2}, E^{n-2}, e_{n-1} = 0) \\ &\quad\quad + (1-p)^2\ I(X_n; Y_n|Y^{n-1}, E^{n-2}, e_{n-1}^n = [0\ 0]). \end{aligned}$$

Continuing in this fashion, we obtain:

$$\begin{aligned} &\sum_{t=2}^{n} I(X_t; Y_t|Y^{t-1}, E^t) \\ &\quad \leq \sum_{k=1}^{n-1} a_k + \sum_{t=2}^{n} (1-p)^{t-1} I(X_t; Y_t|Y^{t-1}, E_1, e_2^t = 0^{t-1}), \end{aligned} \tag{78}$$

where 0^{t-1} is a vector of $t-1$ zeros, and

$$a_k \triangleq \max_{\substack{p(x^k\|y^{k-1}):\\ \sum_{t=1}^k \phi(X_t)\leq\bar{B}}} \left\{ \sum_{t=1}^{k} p(1-p)^{t-1} I(X_t; Y_t|Y^{t-1}) \right\}$$

for $k = 1, \ldots, n$. By assumption, $E_1 = \bar{B}$ w.p. 1, hence the second term in (78) along with $I(X_1; Y_1|E_1)$ is upper bounded by a_n/p, giving:

$$\begin{aligned} nR &\leq \sum_{t=1}^{n} I(X_t; Y_t|Y^{t-1}, E^t) + n\epsilon_n \\ &\leq \sum_{k=1}^{n-1} a_k + a_n/p + n\epsilon_n. \end{aligned} \tag{79}$$

Since (79) is true for any $n \geq 1$ and any R, we can take the limit $n \to \infty$ and sup over all rates to obtain:

$$\begin{aligned} C_{\text{fb}} &\leq \liminf_{n\to\infty} \frac{1}{n} \left[\sum_{k=1}^{n} a_k + \tfrac{1-p}{p} a_n \right] \\ &\overset{\text{(i)}}{=} \lim_{n\to\infty} a_n \\ &= \lim_{n\to\infty} \max_{\substack{p(x^n\|y^{n-1}):\\ \sum_{t=1}^n \phi(X_t)\leq\bar{B}}} \sum_{t=1}^{n} p(1-p)^{t-1} I(X_t; Y_t|Y^{t-1}), \end{aligned} \tag{80}$$

where (i) is because the sequence $\{a_n\}_{n=1}^{\infty}$ converges. This follows by the same reasoning as in the beginning of this proof (see discussion leading to (66) and (67)), i.e. the sequence a_k is non-decreasing and bounded above, hence it has a limit. Continuing to upper bound (80), we have for every n and every $p(x^n\|y^{n-1})$ for which $\sum_{t=1}^n \phi(X_t) \leq \bar{B}$ w.p. 1:

$$\begin{aligned} &\sum_{t=1}^{n} p(1-p)^{t-1} I(X_t; Y_t|Y^{t-1}) \\ &\quad \overset{\text{(i)}}{=} \sum_{t=1}^{n} p(1-p)^{t-1} I(X^t; Y_t|Y^{t-1}) \\ &\quad \overset{\text{(ii)}}{\leq} \sum_{t=1}^{n} p[(1-p)^{t-1} - (1-p)^n] I(X^t; Y_t|Y^{t-1}) \\ &\quad\quad\quad + np(1-p)^n \max_{p(x):\phi(X)\leq\bar{B}} I(X;Y) \\ &\quad \overset{\text{(iii)}}{=} \sum_{t=1}^{n}\sum_{k=t}^{n} p^2(1-p)^{k-1} I(X^t; Y_t|Y^{t-1}) + \varepsilon_n \\ &\quad = \sum_{k=1}^{n}\sum_{t=1}^{k} p^2(1-p)^{k-1} I(X^t; Y_t|Y^{t-1}) + \varepsilon_n \end{aligned} \tag{81}$$

$$\begin{aligned} &p(1-p)I(X_n; Y_n|Y^{n-1}, E^{n-2}, e_{n-1}^n = [\bar{B}\ 0]) + pI(X_{n-1}; Y_{n-1}|Y^{n-2}, E^{n-2}, e_{n-1} = \bar{B}) \\ &\quad \leq pI(X_{n-1}; Y_{n-1}|e_{n-1} = \bar{B}) + p(1-p)I(X_n; Y_n|Y_{n-1}, e_{n-1}^n = [\bar{B}\ 0]) && (75) \\ &\quad \leq \max_{\substack{p(x^2\|y_1):\\ \phi(X_1)+\phi(X_2)\leq\bar{B}}} \{pI(X_1; Y_1) + p(1-p)I(X_2; Y_2|Y_1)\} && (76) \\ &\quad \triangleq a_2 && (77) \end{aligned}$$

$$= \sum_{k=1}^{n} p^2(1-p)^{k-1} I(X^k \to Y^k) + \varepsilon_n, \qquad (82)$$

where (i) is because the channel is memoryless, implying the Markov chain $X^{t-1} - X_t - Y_t$; (ii) is because $I(X^t; Y_t|Y^{t-1}) \le \max_{p(x):\phi(X)\le \bar{B}} I(X;Y)$; and (iii) follows from the geometric series $\sum_{k=t}^{n} p(1-p)^{k-1} = (1-p)^{t-1} - (1-p)^n$ and denoting $\varepsilon_n = np(1-p)^n \max_{p(x):\phi(X)\le \bar{B}} I(X;Y)$. Note that $\lim_{n\to\infty} \varepsilon_n = 0$. Substituting (82) in (80) yields the converse part of (17).

C. No-Feedback Case

The proof for the no-feedback case is very similar, and follows with some minor modifications. In the achievability proof, the state variable is $S_i = X^{i-1}$, since the transmitter does not observe Y^{i-1}. Accordingly, the input pmf is $p(x^N) = \prod_{i=1}^{N} p(x_i|x^{i-1})$. Hence rate R_j is upper bounded by

$$R_j < \sum_{k=j}^{N} pq(k) I(X_j; Y^k|X^{j-1})$$

and the total rate is

$$\begin{aligned} R &< \sum_{j=1}^{N}\sum_{k=j}^{N} pq(k) I(X_j; Y^k|X^{j-1}) \\ &= \sum_{k=1}^{N}\sum_{j=1}^{k} pq(k) I(X_j; Y^k|X^{j-1}) \\ &= \sum_{k=1}^{N} pq(k) I(X^k; Y^k). \end{aligned}$$

Similarly to (69), the achievability part of (66) follows.

For the converse part, we repeat the steps leading to (82) where the input pmf is now of the form $p(x^N) = \prod_{i=1}^{N} p(x_i|x^{i-1})$. In this case, we have the Markov chain $Y^{i-1} - X^{i-1} - X_i$, and thus for any $t \le k$:

$$\begin{aligned} I(X^k; Y_t|Y^{t-1}) &= I(X^t; Y_t|Y^{t-1}) + I(X_{t+1}^k; Y_t|X^t, Y^{t-1}) \\ &= I(X^t; Y_t|Y^{t-1}) + \sum_{i=t+1}^{k} I(X_i; Y_t|X^{i-1}, Y^{t-1}) \\ &= I(X^t; Y_t|Y^{t-1}). \end{aligned}$$

Substituting in (81) yields

$$\begin{aligned} &\sum_{k=1}^{n}\sum_{t=1}^{k} p^2(1-p)^{k-1} I(X^t; Y_t|Y^{t-1}) \\ &\quad = \sum_{k=1}^{n}\sum_{t=1}^{k} p^2(1-p)^{k-1} I(X^k; Y_t|Y^{t-1}) \\ &\quad = \sum_{k=1}^{n} p^2(1-p)^{k-1} I(X^k; Y^k), \end{aligned}$$

which is the converse part of (16).

Appendix B
Capacity With Noncausal Energy Arrival Observations: Proof of Theorem 2

As before, it is not hard to see that (18) and (19) are limits of non-decreasing sequences bounded above, so convergence is guaranteed, and the limits are given by the supremum over $N \ge 1$. Again, we start with the feedback case; the proof without feedback will follow exactly the same lines.

A. Achievability

Fix N and maximizing distributions $\{p(x^k \| y^{k-1})\}_{k=1}^{N}$ in (19). Divide the message w into N messages $w_1, \ldots, w_N$, such that $R = \sum_{k=1}^{N} R_k$. Upon observing e^n, the transmitter and receiver divide the transmission into epochs. Each epoch can be considered as a super-symbol and the epoch length can be thought of as the random state of the channel determining the size of the inputted super-symbol. We communicate a codeword of rate R_k over each state k (Since the epoch length can take any value between 1 and n, we treat all epochs longer than or equal to N as the state $k = N$). For each state k, we generate a codeword where each super-symbol x^k is generated according to the pmf $p(x^k \| y^{k-1})$. Note that this guarantees the energy constraint (3) is satisfied. For decoding the codeword corresponding to state k, we use the technique in [32] and [36, Sec. 17.6.3] as before. The subcodeword formed by the j-th symbol inside the super-symbol x^k is decoded separately for $1 \le j \le k$ by treating the earlier decoded subcodewords and the corresponding channel outputs as side information. In this case, however, we use only symbols belonging to the same state, or epoch length, since symbols in epochs of different lengths will be distributed according to different pmf's.

Recall m_k and $q(k)$, $k = 1, \ldots, N$, from Appendix A-A on page 49. Assuming e^n is typical as defined in Appendix A-A, there are $m_k \ge n(1-\epsilon)^2 pq(k)$ epochs of length k. The probability of error for decoding w_k is

$$2^{nR_k} \cdot 2^{-m_k I(X_1; Y^k)} \times 2^{-m_k I(X_2; Y_2^k|X_1, Y_1)} \cdots 2^{-m_k I(X_k; Y_k|X^{k-1}, Y^{k-1})}.$$

The achievable rate for state k is then given by

$$\begin{aligned} R_k &= pq(k) \sum_{j=1}^{k} I(X_j; Y_j^k|X^{j-1}, Y^{j-1}) \\ &= pq(k) \sum_{j=1}^{k}\sum_{i=j}^{k} I(X_j; Y_i|X^{j-1}, Y^{i-1}) \\ &= pq(k) \sum_{i=1}^{k} I(X^i; Y_i|Y^{i-1}) \\ &= pq(k) I(X^k \to Y^k) \end{aligned}$$

and the total rate is

$$\begin{aligned} R &= \sum_{k=1}^{N} pq(k) I(X^k \to Y^k) \\ &\ge \sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k \to Y^k). \end{aligned}$$

This is a lower bound to capacity for every $N \geq 1$, therefore we can take the supremum to obtain

$$C_{\text{fb}}^{\text{nc}} \geq \sup_{N\geq 1} \sum_{k=1}^{N} p^2(1-p)^{k-1} I(X^k \to Y^k),$$

which is the achievability part of (19).

B. Converse

By Fano's inequality:

$$\begin{aligned} nR &\leq I(W; Y^n|E^n) + n\epsilon_n \\ &= \sum_{t=1}^{n} I(W; Y_t|Y^{t-1}, E^n) + n\epsilon_n \\ &= \sum_{t=1}^{n} I(X_t; Y_t|Y^{t-1}, E^n) + n\epsilon_n \\ &= \sum_{e^n} p(e^n) \sum_{t=1}^{n} I(X_t; Y_t|Y^{t-1}, E^n = e^n) + n\epsilon_n. \end{aligned} \quad (83)$$

Recall the definition of $m = m(e^n)$, ℓ^m, and τ^m in Appendix A-A, and let $\tau_{m+1} = n+1$. For a fixed e^n, we can further bound the expression in (83) as

$$\begin{aligned} &\sum_{t=1}^{n} I(X_t; Y_t|Y^{t-1}, E^n = e^n) \\ &\quad \overset{\text{(i)}}{\leq} \sum_{i=1}^{m} \sum_{t=\tau_i}^{\tau_{i+1}-1} I(X_t; Y_t|Y_{\tau_i}^{t-1}, E^n = e^n) \\ &\quad \overset{\text{(ii)}}{=} \sum_{i=1}^{m} \sum_{t=\tau_i}^{\tau_{i+1}-1} I(X_{\tau_i}^t; Y_t|Y_{\tau_i}^{t-1}, E^n = e^n) \\ &\quad = \sum_{i=1}^{m} I(X_{\tau_i}^{\tau_{i+1}-1} \to Y_{\tau_i}^{\tau_{i+1}-1}|E^n = e^n) \\ &\quad \overset{\text{(iii)}}{\leq} \sum_{i=1}^{m} \ell_i C_{\text{fb}}^{\text{per}}(\ell_i), \end{aligned}$$

where (i) and (ii) are because the channel is memoryless, and (iii) is due to (21). Taking $n \to \infty$, we get

$$C_{\text{fb}}^{\text{nc}} \leq \liminf_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^{m(E^n)} \ell_i(E^n) \cdot C_{\text{fb}}^{\text{per}}\big(\ell_i(E^n)\big)\right], \quad (84)$$

where the expectation is over the RV E^n. By the strong law of large numbers for regenerative processes [45, Ch. VI, Thm. 31]:

$$\frac{1}{n} \sum_{i=1}^{m(E^n)} \ell_i(E^n) \cdot C_{\text{fb}}^{\text{per}}\big(\ell_i(E^n)\big) \to p \cdot \mathbb{E}[L \cdot C_{\text{fb}}^{\text{per}}(L)] \quad \text{a.s.},$$

where L is a geometric RV with parameter p. Moreover, since $C_{\text{fb}}^{\text{per}}(k) \leq \max_{p(x):\phi(X)\leq \bar{B}} I(X;Y)$ and $n = \sum_{i=1}^{m(e^n)} \ell_i(e^n)$ for any e^n:

$$\frac{1}{n} \sum_{i=1}^{m(E^n)} \ell_i(E^n) \cdot C_{\text{fb}}^{\text{per}}\big(\ell_i(E^n)\big) \leq \max_{\substack{p(x):\\ \phi(X)\leq \bar{B}}} I(X;Y) \quad \text{w.p. } 1,$$

Therefore, by dominated convergence [46, Thm. 1.5.6], the lim inf in (84) is a lim and it is given by

$$\begin{aligned} C_{\text{fb}}^{\text{nc}} &\leq p \cdot \mathbb{E}[L \cdot C_{\text{fb}}^{\text{per}}(L)] \\ &= p \sum_{k=1}^{\infty} p(1-p)^{k-1} k C_{\text{fb}}^{\text{per}}(k) \\ &= \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \max_{\substack{p(x^k\|y^{k-1}):\\ \sum_{i=1}^{k}\phi(X_i)\leq \bar{B}}} I(X^k \to Y^k), \end{aligned}$$

which completes the proof of the converse part for (19).

The proof for the channel without feedback follows by repeating the above steps and applying the modifications appearing in Appendix A-C.

Appendix C
Capacity Formulas for Arbitrary Alphabets

In this section, we will show the capacity expressions (16)–(21) are valid also when the input and output alphabets are arbitrary sets, such as the AWGN channel. First, observe that the converse parts follow without change, since the finite-alphabet assumption was not used in the proofs. It therefore remains to show the achievability parts; we will focus on the capacity expression for periodic energy arrivals (21) for simplicity, however the derivation applies to all capacity results in Section III.

We follow the technique in [37, Ch. 7]. Fix a finite set of input symbols $\bar{\mathcal{X}} \subseteq \mathcal{X}$, $|\bar{\mathcal{X}}| < \infty$, and denote by $\bar{X}$ an RV taking values in this set. Next, fix a partition $\mathcal{P}$ of the output space $\mathcal{Y}$ into a finite number of disjoint sets, and denote by $[Y]_{\mathcal{P}}$ the RV indicating the partition to which the output Y belongs. This defines a DMC with transition probability $\Pr([Y]_{\mathcal{P}} = B \mid \bar{X} = a) = \int_B p(dy|a)$, where $p(y|x)$ is the transition probability of the original channel.

We can apply our achievability scheme for this DMC, and obtain the following achievable rate for the original channel, for any finite set $\bar{\mathcal{X}}$ and any partition $\mathcal{P}$ of the output space $\mathcal{Y}$:

$$R = \max_{\substack{p(\bar{x}^k\|[y]_{\mathcal{P}}^{k-1}):\\ \sum_{i=1}^{k}\phi(\bar{X}_i)\leq \bar{B}}} \frac{1}{k} I(\bar{X}^k \to [Y]_{\mathcal{P}}^k).$$

Taking the supremum over all finite sets $\bar{\mathcal{X}}$, and taking the limit of finer and finer partitions of $\mathcal{Y}$, we have the following lower bound on capacity:

$$C_{\text{fb}}^{\text{per}}(k) \geq \sup_{\bar{\mathcal{X}}} \lim_{\substack{\text{finer}\\ \mathcal{P}}} \max_{\substack{p(\bar{x}^k\|[y]_{\mathcal{P}}^{k-1}):\\ \sum_{i=1}^{k}\phi(\bar{X}_i)\leq \bar{B}}} \frac{1}{k} I(\bar{X}^k \to [Y]_{\mathcal{P}}^k). \quad (85)$$

The above limit of finer and finer partitions should be understood as a sequence of partitions, where each partition "approximates" the output space $\mathcal{Y}$ better than its predecessor. For example, one can take the limit as $n \to \infty$ of $\mathcal{P}_n = \{-n\Delta, -(n-1)\Delta, \ldots, -\Delta, 0, \Delta, \ldots, n\Delta\}$ with $\Delta = 1/\sqrt{n}$ as in [36, Sec. 3.4.1], where the quantization $[Y]_{\mathcal{P}_n}$ is the closest point to Y.

To show (85) reduces to (21), we start with the following lemma.

Lemma 2: For a fixed set of input symbols $\bar{\mathcal{X}} \subseteq \mathcal{X}$, $|\bar{\mathcal{X}}| < \infty$, the following holds:

$$\lim_{\substack{\text{finer} \\ \mathcal{P}}} \max_{\substack{p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to [Y]_{\mathcal{P}}^k) = \max_{\substack{p(\bar{x}^k \| y^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to Y^k). \tag{86}$$

Applying Lemma 2 to (85) yields

$$C_{\text{fb}}^{\text{per}}(k) \ge \sup_{\bar{\mathcal{X}}} \max_{\substack{p(\bar{x}^k \| y^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} \frac{1}{k} I(\bar{X}^k \to Y^k) = \max_{\substack{p(x^k \| y^{k-1}): \\ \sum_{i=1}^k \phi(X_i) \le \bar{B}}} \frac{1}{k} I(X^k \to Y^k),$$

where the optimal input distribution $p(x^k \| y^{k-1})$ can be approached arbitrarily closely by an input distribution on a large enough set of discrete input symbols. Along with the converse part, this concludes the proof of (21) for general input and output alphabets. It remains to prove Lemma 2.

Proof of Lemma 2: We show inequalities in both directions. Fix a partition $\mathcal{P}$, and fix an input distribution $p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})$ for which $\sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}$ almost surely. One can construct an input distribution $p(\bar{x}^k \| y^{k-1})$ causally conditioned on the original (unpartitioned) output space $\mathcal{Y}$, by setting $p(\bar{x}^k \| y^{k-1}) = p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})$ for all $y^{k-1} \in [y]_{\mathcal{P}}^{k-1}$. Similarly, for any partition $\mathcal{P}'$ which is a subpartition of $\mathcal{P}$ (that is, every set in $\mathcal{P}$ is a union of sets in $\mathcal{P}'$), we construct an input distribution $p(\bar{x}^k \| [y]_{\mathcal{P}'}^{k-1})$ in the same way. Denote the directed information evaluated under this input distribution by $I(\bar{X}^k \to [Y]_{\mathcal{P}'}^k)\big|_{p(\bar{x}^k \| [y]_{\mathcal{P}}^k)}$. We have, for any partition $\mathcal{P}'$ which is a subpartition of $\mathcal{P}$:

$$\max_{\substack{p(\bar{x}^k \| [y]_{\mathcal{P}'}^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to [Y]_{\mathcal{P}'}^k) \ge I(\bar{X}^k \to [Y]_{\mathcal{P}'}^k)\Big|_{p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})}.$$

Note that for any partition $\mathcal{P}'$ which is finer than $\mathcal{P}$, if it is not a subpartition of $\mathcal{P}$, one can find an even finer partition which is a subpartition of $\mathcal{P}$. Hence, taking the limit of finer and finer partitions:

$$\begin{aligned}\liminf_{\substack{\text{finer} \\ \mathcal{P}'}} \max_{\substack{p(\bar{x}^k \| [y]_{\mathcal{P}'}^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to [Y]_{\mathcal{P}'}^k) &\ge \liminf_{\substack{\text{finer} \\ \mathcal{P}'}} I(\bar{X}^k \to [Y]_{\mathcal{P}'}^k)\Big|_{p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})} \\ &= I(\bar{X}^k \to Y^k)\Big|_{p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})}.\end{aligned}$$

Since the LHS does not depend on $\mathcal{P}$ and the distribution $p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})$, we can take $\mathcal{P}$ to be arbitrarily fine and maximize over all input distributions to obtain:

$$\liminf_{\substack{\text{finer} \\ \mathcal{P}'}} \max_{\substack{p(\bar{x}^k \| [y]_{\mathcal{P}'}^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to [Y]_{\mathcal{P}'}^k) \ge \max_{\substack{p(\bar{x}^k \| y^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to Y^k).$$

For the other direction, again fix a partition $\mathcal{P}$ and an input distribution $p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})$. As before, construct a distribution $p(\bar{x}^k \| y^{k-1})$ such that $p(\bar{x}^k \| y^{k-1}) = p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})$ for all $y^{k-1} \in [y]_{\mathcal{P}}^{k-1}$. Observe that this induces the Markov chains

$$\bar{X}_t - (\bar{X}^{t-1}, [Y]_{\mathcal{P}}^{t-1}) - Y^{t-1}, \quad t = 1, \ldots, k. \tag{87}$$

The directed information under this distribution can be lower bounded as follows:

$$\begin{aligned} I(\bar{X}^k \to Y^k) &= \sum_{t=1}^k I(\bar{X}^t; Y_t | Y^{t-1}) \\ &\overset{\text{(i)}}{=} \sum_{i=1}^k I(\bar{X}_i; Y_i^k | \bar{X}^{i-1}, Y^{i-1}) \\ &\overset{\text{(ii)}}{=} \sum_{i=1}^k \Big[H(\bar{X}_i | \bar{X}^{i-1}, Y^{i-1}, [Y]_{\mathcal{P}}^{i-1}) \\ &\qquad - H(\bar{X}_i | \bar{X}^{i-1}, Y^k, [Y]_{\mathcal{P}}^k) \Big] \\ &\overset{\text{(iii)}}{\ge} \sum_{i=1}^k \Big[H(\bar{X}_i | \bar{X}^{i-1}, [Y]_{\mathcal{P}}^{i-1}) - H(\bar{X}_i | \bar{X}^{i-1}, [Y]_{\mathcal{P}}^k) \Big] \\ &= I(\bar{X}^k \to [Y]_{\mathcal{P}}^k), \end{aligned}$$

where (i) is by using the chain rule and changing the order of summation; (ii) is because $[Y]_{\mathcal{P}}$ is a function of Y; and (iii) is due to the Markov chain (87) and because conditioning reduces entropy.

We get

$$\begin{aligned} I(\bar{X}^k \to [Y]_{\mathcal{P}}^k)\Big|_{p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})} &\le I(\bar{X}^k \to Y^k)\Big|_{p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})} \\ &\le \max_{\substack{p(\bar{x}^k \| y^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to Y^k). \end{aligned}$$

Since the RHS does not depend on the distribution $p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1})$, we maximize to obtain:

$$\max_{\substack{p(\bar{x}^k \| [y]_{\mathcal{P}}^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to [Y]_{\mathcal{P}}^k) \le \max_{\substack{p(\bar{x}^k \| y^{k-1}): \\ \sum_{i=1}^k \phi(\bar{X}_i) \le \bar{B}}} I(\bar{X}^k \to Y^k).$$

Taking the limit of finer and finer partitions, we obtain the other direction of (86). □

Appendix D
Optimal Online Power Control

We solve the maximization problem in (56). Writing the problem in standard form and using KKT conditions, we have for $i = 1, \ldots, N$:

$$-p(1-p)^{i-1} \frac{1}{2} \frac{1}{1 + \mathcal{E}_i} - \lambda_i + \tilde{\lambda} = 0,$$

with $\lambda_i, \tilde{\lambda} \ge 0$ and the complementary slackness conditions: $\lambda_i \mathcal{E}_i = 0$ and $\tilde{\lambda}(\sum_{i=1}^N \mathcal{E}_i - B) = 0$.

To obtain the non-zero values of $\mathcal{E}_i$, we set $\lambda_i = 0$:

$$\mathcal{E}_i = \frac{p(1-p)^{i-1}}{2\tilde{\lambda}} - 1. \tag{88}$$

Since $\mathcal{E}_i \geq 0$, this implies $\tilde{\lambda} \leq \frac{p(1-p)^{i-1}}{2}$ for all i for which $\mathcal{E}_i > 0$. This is a decreasing function of i, therefore there exists an integer $\tilde{N}$ such that $\mathcal{E}_i > 0$ for $i = 1, \ldots, \tilde{N}$ and $\mathcal{E}_i = 0$ for $i > \tilde{N}$. Therefore $\tilde{N}$ is the largest integer satisfying $\tilde{N} \leq N$ and

$$\tilde{\lambda} \leq \frac{1}{2} p(1-p)^{\tilde{N}-1}. \tag{89}$$

Next, consider the total energy constraint $\sum_{i=1}^{N} \mathcal{E}_i \leq \bar{B}$. Since increasing $\mathcal{E}_i$ for any i will only increase the objective, this constraint must hold with equality:

$$\begin{aligned}\bar{B} &= \sum_{i=1}^{N} \mathcal{E}_i \\ &= \sum_{i=1}^{\tilde{N}} \left(\frac{p(1-p)^{i-1}}{2\tilde{\lambda}} - 1 \right) \\ &= \frac{1-(1-p)^{\tilde{N}}}{2\tilde{\lambda}} - \tilde{N}.\end{aligned}$$

This yields:

$$\tilde{\lambda} = \frac{1-(1-p)^{\tilde{N}}}{2(\bar{B}+\tilde{N})}. \tag{90}$$

Along with (89), we deduce that $\tilde{N}$ is the largest integer satisfying $\tilde{N} \leq N$ and

$$\frac{1-(1-p)^{\tilde{N}}}{2(\bar{B}+\tilde{N})} \leq \frac{p(1-p)^{\tilde{N}}}{2},$$

or equivalently

$$1 \leq (1-p)^{\tilde{N}}[1 + p(\bar{B}+\tilde{N})].$$

Observe that for N large enough, the optimal $\tilde{N}$ does not depend on N, and will simply be the smallest positive integer satisfying $1 > (1-p)^{\tilde{N}}[1+p(\bar{B}+\tilde{N})]$. The optimal sequence $\{\mathcal{E}_i\}_{i=1}^{\infty}$ in (57) is obtained by substituting (90) in (88) for $i = 1, \ldots, \tilde{N}$, and $\mathcal{E}_i = 0$ for $i > \tilde{N}$. Substituting the optimal $\mathcal{E}_i$ in the optimization objective yields (59). □

Appendix E
Noncausal Capacity Bounds: Proof of Proposition 4

We prove Proposition 4 following the same steps as in the previous sections. For the upper bound, we can similarly relax the energy constraint in (47) to be only in expectation, thus giving an upper bound:

$$\begin{aligned}C^{\text{nc}} &\leq \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \max_{\substack{p(x^k):\\ \mathbb{E}\|X^k\|^2 \leq \bar{B}}} I(X^k; Y^k) \\ &= \sum_{k=1}^{\infty} p^2(1-p)^{k-1} \frac{k}{2} \log(1 + \bar{B}/k),\end{aligned}$$

where the last line is a known result for vector Gaussian channels.

We derive a lower bound for (47) by considering a suboptimal input distribution. Let $p(x^k) = \prod_{i=1}^{k} p_k(x_i)$, where $p_k(x)$ is some distribution for which $X^2 \leq \bar{B}/k$ a.s. We then have the following lower bound:

$$C^{\text{nc}} \geq \sum_{k=1}^{\infty} p^2(1-p)^{k-1} k C_{\text{Smith}}(\bar{B}/k), \tag{91}$$

where $C_{\text{Smith}}(\mathcal{E})$ is defined in (64). The expression has been evaluated numerically using the algorithm suggested in [39] and plotted in Figure 9.

Next, using Lemma 1, we can further lower bound (91) to obtain (61).

References

[1] D. Shaviv and A. Özgür, "Capacity of the AWGN channel with random battery recharges," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 136–140.

[2] D. Shaviv, A. Özgür, and H. Permuter, "Can feedback increase the capacity of the energy harvesting channel?" in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jerusalem, Israel, Apr. 2015, pp. 1–5.

[3] M. Tabesh, N. Dolatsha, A. Arbabian, and A. M. Niknejad, "A power-harvesting pad-less millimeter-sized radio," *IEEE J. Solid-State Circuits*, vol. 50, no. 4, pp. 962–977, Apr. 2015.

[4] S. Pellerano, J. Alvarado, and Y. Palaskas, "A mm-wave power-harvesting RFID tag in 90 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 45, no. 8, pp. 1627–1637, Aug. 2010.

[5] 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.

[6] O. Ozel and S. Ulukus, "AWGN channel under time-varying amplitude constraints with causal information at the transmitter," in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2011, pp. 373–377.

[7] K. Tutuncuoglu, O. Ozel, A. Yener, and S. Ulukus, "Binary energy harvesting channel with finite energy storage," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1591–1595.

[8] W. Mao and B. Hassibi, "On the capacity of a communication system with energy harvesting and a limited battery," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1789–1793.

[9] Y. Dong and A. Özgür, "Approximate capacity of energy harvesting communication with finite battery," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 801–805.

[10] Y. Dong, F. Farnia, and A. Özgür, "Near optimal energy control and approximate capacity of energy harvesting communication," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 540–557, Mar. 2015.

[11] D. Shaviv, P.-M. Nguyen, and A. Özgür, "Capacity of the energy-harvesting channel with a finite battery," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6436–6458, Mar. 2016.

[12] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[13] Y. Li and G. Han. (2016). "Asymptotics of input-constrained erasure channel capacity." [Online]. Available: https://arxiv.org/abs/1605.02175

[14] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.

[15] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.

[16] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1180–1189, Mar. 2012.

[17] C. M. Vigorito, D. Ganesan, and A. G. Barto, "Adaptive control of duty cycling in energy-harvesting wireless sensor networks," in *Proc. 4th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw. (SECON)*, San Diego, CA, Jun. 2007, pp. 21–30.

[18] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1326–1336, Apr. 2010.

[19] R. Rajesh, V. Sharma, and P. Viswanath, "Capacity of fading Gaussian channel with an energy harvesting sensor node," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–6.

[20] R. Srivastava and C. E. Koksal, "Basic performance limits and tradeoffs in energy-harvesting sensor nodes with finite data and energy storage," *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1049–1062, Aug. 2013.

[21] Q. Wang and M. Liu, "When simplicity meets optimality: Efficient transmission power control with stochastic energy harvesting," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 580–584.

[22] M. Zafer and E. Modiano, "Optimal rate control for delay-constrained data transmission over a wireless channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4020–4039, Sep. 2008.

[23] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications powered by energy harvesters," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 2368–2372.

[24] A. Sinha, "Optimal power allocation for a renewable energy source," in *Proc. Nat. Conf. Commun. (NCC)*, Feb. 2012, pp. 1–5.

[25] P. Blasco, D. Gunduz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1872–1882, Apr. 2013.

[26] V. Jog and V. Anantharam, "An energy harvesting AWGN channel with a finite battery," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 806–810.

[27] O. Ozel, K. Tutuncuoglu, S. Ulukus, and A. Yener, "Capacity of the discrete memoryless energy harvesting channel with side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 796–800.

[28] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, Nov. 1990, pp. 303–305.

[29] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[30] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 780–798, Mar. 2005.

[31] S. Tatikonda, "A Markov decision approach to feedback channel capacity," in *Proc. 44th IEEE Conf. Decision Control Eur. Control Conf. (CDC-ECC)*, Dec. 2005, pp. 3213–3218.

[32] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory.*, vol. 54, no. 4, pp. 1488–1499, Apr. 2008.

[33] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

[34] G. Kramer, "Information networks with in-block memory," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2105–2120, Apr. 2014.

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

[36] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[37] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[38] D. Shaviv and A. Özgür. (2015). *Capacity of the AWGN Channel With Random Battery Recharges*. [Online]. Available: https://arxiv.org/abs/1506.02792

[39] J. G. Smith, "The information capacity of amplitude- and variance-constrained sclar Gaussian channels," *Inf. Control*, vol. 18, no. 3, pp. 203–219, 1971.

[40] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.

[41] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.

[42] D. Shaviv and A. Özgür, "Universally near optimal online power control for energy harvesting nodes," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3620–3631, Dec. 2016.

[43] A. Kazerouni and A. Özgür, "Optimal online strategies for an energy harvesting system with Bernoulli energy recharges," in *Proc. IEEE Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, Mumbai, India, May 2015, pp. 235–242.

[44] D. Shaviv, A. Özgür, and H. H. Permuter, "Capacity of remotely powered communication," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1364–1391, Mar. 2017.

[45] S. Asmussen, *Applied Probability and Queues*. New York, NY, USA: Springer, 2008, vol. 51.

[46] R. Durrett, *Probability: Theory and Examples*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

Dor Shaviv (S'13) received his B.Sc. (summa cum laude) degrees in electrical engineering and physics from the Technion-Israel Institute of Technology, Haifa, Israel, in 2007. He received his M.Sc. degree in electrical engineering from the same institution in 2012. During 2007-2013 he worked as an R&D engineer in the Israel Defense Forces. He is currently a Ph.D. candidate in the Electrical Engineering Department at Stanford University, and is a recipient of a Robert Bosch Stanford Graduate Fellowship.

Ayfer Özgür (M'06) received her B.Sc. degrees in electrical engineering and physics from Middle East Technical University, Turkey, in 2001 and the M.Sc. degree in communications from the same university in 2004. From 2001 to 2004, she worked as hardware engineer for the Defense Industries Development Institute in Turkey. She received her Ph.D. degree in 2009 from the Information Processing Group at EPFL, Switzerland. In 2010 and 2011, she was a post-doctoral scholar with the Algorithmic Research in Network Information Group at EPFL. She is currently an Assistant Professor in the Electrical Engineering Department at Stanford University. Her research interests include network communications, wireless systems, and information and coding theory. Dr. Özgür received the EPFL Best Ph.D. Thesis Award in 2010 and a NSF CAREER award in 2013.

Haim H. Permuter (M'08–SM'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY.