Information and Energy Transmission with Experimentally-Sampled Harvesting Functions

Daewon Seo, Graduate Student Member, IEEE and Lav R. Varshney, Senior Member, IEEE

Abstract—This paper considers the problem of simultaneous information and energy transmission (SIET), where the energy harvesting function is only known experimentally at sample points, e.g., due to nonlinearities and parameter uncertainties in harvesting circuits. We investigate the performance loss due to this partial knowledge of the harvesting function in terms of transmitted energy and information. In particular, we assume harvesting functions are a subclass of Sobolev space and consider two cases, where experimental samples are either taken noiselessly or in the presence of noise. Using constructive function approximation and regression methods for noiseless and noisy samples respectively, we show that the worst loss in energy transmission vanishes asymptotically as the number of samples increases. Similarly, the worst loss in information rate vanishes in the interior of the energy domain, however, does not always vanish at maximal energy. We further show the same principle applies in multicast settings such as medium access in the Wi-Fi protocol. We also consider the end-to-end source-channel communication problem under source distortion constraint and channel energy requirement, where distortion and harvesting functions both are known only at samples.

Index Terms—Energy harvesting, information theory, multicast, joint source-channel coding, Sobolev spaces

I. INTRODUCTION

There is growing interest in simultaneous information and energy transmission (SIET) where a single patterned energy signal carries both over a noisy channel. Information-theoretic investigation in this direction started in [1], and has now spawned hundreds of results in the wireline [2] and especially the wireless setting (referred to as SWIPT (simultaneous wireless information and power transmission) in literature), see e.g. [3] for a recent survey. These classes of problems are important for sensor networks, Internet of Things (IoT), and similar settings where terminals may require energy.

Past theoretical works typically assume simple energy harvesting functions such as quadratic [4], so the amount of energy obtained from received signal \( y(t) \) is \( \int_0^T y^2(t) dt \), where \( T \) is the symbol duration. However, practical energy harvesting circuits have nonlinearities and nonidealities that complicate the relationship between channel output symbol values and their harvested energy [5]–[9]. Indeed, this energy harvesting function may only be available through samples from experiments [10]–[14] or perhaps from analog electronic circuit simulations. See Fig. 1 for examples of harvesting circuits and their nonlinear energy harvesting functions, known only at samples [15]. Since our knowledge of harvesting functions will only be partial, it leads to a general problem of energy-requiring channel coding (and joint source-channel coding) with partial knowledge of the energy harvesting function.

Unlike the received symbol, which is uncontrollable due to channel noise—e.g., in the low signal to noise ratio (SNR) regime, thus, it results in uncontrolled harvested energy as well—the transmitted symbol is always under control. Motivated by this limitation, unlike [1], we think of the harvesting function as a function (or a stochastic function, e.g. in the case of low SNR) of the transmitted symbol, which is a sufficiently general model for many modern communication systems.

The goal of this work is to investigate how much worst-case loss in SIET energy and information performance is incurred due to the partial knowledge of the harvesting function from samples. In particular, we study fundamental limits of point-to-point SIET systems when the signalling scheme is optimally designed based not on the full harvesting function but based on the given samples under the assumption the harvesting function is from some class of smooth functions. We consider two settings separately: when samples are noiseless or when samples are noisy. We draw on results from approximation theory including the spline method in function approximation [16] for noiseless samples, and the local polynomial estimator in non-parametric regression [17] for noisy samples. We prove that the worst-case amount of energy transmission is asymptotically the same as if we knew the full harvesting function.

The authors are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (e-mail: {dseoo9, varshney}@illinois.edu).

1As far as we can tell, joint source-channel coding has not been considered in the SIET literature even in the full knowledge setting.
Consider the now-standard formulation of SIET systems from [1], where the goal is to use a patterned energy signal to simultaneously transmit reliable information and energy over a noisy channel. Recall that in a standard SIET system, first at the transmitter, messages are encoded into a codeword $x^n \in X^n$ to protect against channel noise, where $n$ is codeword length. Then, the codeword is modulated into a sequence of $n$ baseband signals using a given modulation scheme, and then up-converted into a sequence of physical radio frequency (RF) waves. Attenuation and noise corrupt the RF waves so that the receiver observes a noisy version of RF waves, which is denoted by $Y^n \in Y^n$. The receiver repeats the process in reverse, that is, down-converts into a baseband signal, demodulates, and decodes.

The received RF signal is also passed through an energy-harvesting circuit as in Fig. 1—either directly or through a signal splitting architecture [1], [22]—to capture energy. We suppose information decoder and energy harvester both process the same signal. Our mathematical formulation subsumes a signal splitting scheme with a certain ratio $\rho$, called static power splitting [4], with proper scaling of harvesting function. Since the receiver obtains energy from the received RF signal, in addition to maximizing information transmission between the transmitter and the receiver, a guarantee on the amount of energy delivery, say $B$, via the RF signal is also required.

As shown in [1], the fundamental limits of this problem are governed by the capacity-energy function:

$$C_B(B) = \max_{P_X \sqsupseteq [b(Y)] \geq B} I(X; Y),$$

where $X \in X, Y \in Y$ are transmitted and received symbols, respectively, and $b(Y)$ is the energy harvesting function for the received symbol $Y$. Note that the minimum energy requirement of [1] can be also written in terms of $x$ using conditional expectation, i.e., letting $\beta(x) := E_Y | x [b(Y)]$,

$$E_Y [b(Y)] = E_X [E_Y | X [b(Y)] ] = E_X [\beta(X)].$$

Hence we can think of the harvesting function as a (perhaps random) function of the transmission alphabet symbols, with the following equivalent capacity-energy expression: for a harvesting function $f$ and a set of harvesting functions $F$,

$$C_f(B) = \max_{P_X \sqsupseteq [f(X)] \geq B} I(X; Y),$$

$$C_F(B) = \sup_{P_X \sqsupseteq [f(X)] \geq B \forall f \in F} I(X; Y),$$

which are used throughout the sequel. $C_F(B)$ indicates the maximal information rate at which we can send energy no smaller than $B$ for any harvesting function in $F$. Note that $C_F(B) \leq C_f(B)$ since underlying probability space of $\xi$ belongs to that of $\xi$. As illustrated in Fig. 2, the tradeoff is non-increasing and concave.
We also define energy-capacity function $B_f(R), B_F(R)$ as

$$B_f(R) = \max_{P_X: E[f(X)] \geq R} E[f(X)],$$  \hspace{1cm} (4)

$$B_F(R) = \max_{P_X: E[f(X)] \geq R} \inf_{f \in F} E[f(X)].$$  \hspace{1cm} (5)

It is easy to see that $B_f(R), B_F(R)$ are the inverse functions of $C_f(B), C_F(B)$, provided that the inverse functions exist.

Probability distribution on $X$ that achieves $C_f(B)$ is called capacity-achieving distribution, i.e.,

$$P_X^* \in \arg \max_{P_X: E[f(X)] \geq B} I(X; Y),$$

where ‘$*$’ indicates that such capacity achieving distribution is not necessarily unique. The maximizers with respect to $C_F(B), B_f(R), B_F(R)$ are similarly defined and also called capacity-achieving distributions with abuse of notation. In this case, the objective function will be clear from the context. Also note that when a certain $P_X$ is given, it can be thought of as the Shannon’s random codebook with rate $I(X; Y)$, generated from $P_X$.\textsuperscript{23}

**A. Channel Alphabets**

In this work, we take $X = [0, 1]$ and $\mathcal{Y}$ as the set of all possible received signals, as determined by the physics of the system. Taking the input alphabet as the unit interval rather than the real line imposes a peak power constraint\textsuperscript{1, 2, 24} and is motivated by practical discrete-time analog or dense constellation digital communication systems, as follows.

- **AWGN channel:** The standard AWGN channel has $X = \mathcal{Y} = \mathbb{R}$ and codewords $x^n \in \mathbb{R}^n$. However, due to limitations on RF front end, we may assume $X = [-a, a]$ so it is possible to assume $X = [0, 1]$ without loss of generality.

- **AM in discrete-time:** In amplitude modulation (AM), at each time slot analog information $x \in [0, 1] = X$ is modulated and up-converted to $x \cos(2\pi f_c t)$, where $f_c$ is the carrier frequency.

- **Dense constellation QAM:** Although the constellation set is discrete in 2-dimensional space, it can be thought of as a 2-dimensional continuous interval when sufficiently dense, say $[0, 1]^2 = X^2$. As an example, in dense quadratic amplitude modulation (QAM), a constellation point $x = [x_1, x_2] \in [0, 1]^2$ generates the RF wave $x_1 \sin(2\pi f_c t) - x_2 \cos(2\pi f_c t)$.

- **Dense constellation OFDM:** Consider a binary sequence of length $2N$, $x = [x_1, x_2, \ldots, x_{2N}] \in \{0, 1\}^{2N}$. Using a $2N$-bit binary representation of real values in $[0, 1]$, it can be thought of as $\{0, 1\}^{2N} \approx [0, 1] = X$ when $N$ is large enough. Once $x = [x_1, \ldots, x_{2N}] \in X$ is chosen, the generated baseband signal is $\sum_{k=1}^N x_{2k-1} \sin(2\pi k t/T) - x_{2k} \cos(2\pi k t/T)$.

- **Dense constellation DSSS:** Similar to OFDM, we can assume $x = [x_1, x_2, \ldots, x_{2N}] \in X \approx [0, 1]$. Each bit of $x$ is XORed with an assigned pseudo-noise (PN) sequence.

**B. Continuity**

We make two continuity assumptions. The first is to assume that the channel is continuous in the sense that when $x_1, x_2 \in X$ are close, the distributions $Y_1$ and $Y_2$ are also close. More precisely, when a sequence $x_n \to x$, the resulting received signals $Y_n \to Y$ in distribution.\textsuperscript{23} The second is to assume the energy harvesting function $\beta(\cdot)$ is smooth on $X$, due to physical continuity of electromagnetic signals and circuits. To define the smoothness rigorously, let us first introduce the $L_q$ norm and the Sobolev space $W^\lambda_q$.

**Definition 1:** For a Lebesgue-measurable function $f$ on $X$, let the $L_q$ norm for $q \in [1, \infty]$ be

$$||f||_q = \left\{ \frac{1}{q} \right\} \text{ess sup}_{x \in X} |f(x)|$$

Let $L_q = L_q(X)$ be the set of all $L_q$-integrable functions on $X$, i.e., $||f||_q < \infty$ if $f \in L_q$.

**Definition 2:** For $\lambda \in \mathbb{N}, q \in [1, \infty]$, the Sobolev space $W^\lambda_q(X)$ is defined as the set of functions in $L_q$ such that derivatives of order equal or less than $\lambda$ exist and are in $L_q$, i.e.,

$$W^\lambda_q(X) = \{ f \in L_q(X) : f^{(k)} \in L_q \ \forall k \leq \lambda \},$$

where $f^{(k)}$ is the $k$th derivative of $f$.

We define our class of energy harvesting functions, $\Gamma^K$, as a subset of $W^\lambda_\infty(X)$ satisfying:

$$\Gamma^K = \{ \beta \in W^\lambda_\infty(X) : ||\beta^{(k)}||_\infty \leq K \ \forall k \leq \lambda \}.$$

When the argument of $|| \cdot ||_q$ for $q \in [1, \infty]$ is a real-valued vector $x \in \mathbb{R}^d$, $||x||_q$ denotes the $\ell_q$ norm with slight abuse of notation.

$$||x||_q = \left\{ \frac{1}{q} \right\} \sum_{i=1}^d |x_i|^q$$

$^2$This makes particular sense when noise is signal-independent, such as in OFDM or DSSS, where a set of length-$2N$ binary sequences in examples above can be rearranged in a Gray code manner so two successive elements differ only in one bit out of $2N$ bits. Then the one-bit difference results in RF signals that also differ only by one subcarrier element in OFDM and one PN sequence duration in DSSS, respectively. Due to the independence of noise, received signals are also similarly distributed so that the channel is continuous in the above sense.

$^3$Note that this notion of continuity has nothing to do with capacity-achieving input distributions and their discreteness\textsuperscript{24}. Such discreteness does appear in the conditions for Thm. 7.
C. Sampling and Losses

We consider regular fixed design of samples, that is, \( m \) samples are evenly-spaced on \( X = [0, 1] \) so that \( x_i = \frac{i}{m-1} \) where \( i = 0, 1, \ldots, m-1 \). Energy samples are experimentally taken either in the absence of noise or in the presence of noise, depending on which we will consider different strategies. However, those strategies do not make a substantial difference as we will see.

For noiseless samples \( \{ (\frac{i}{m-1}, \beta(\frac{i}{m-1})) \}_{i=0}^{m-1} \), let \( \Gamma(\beta, m) \subseteq \Gamma^K \) be the set of harvesting functions that agree on the sample points. Upon observing samples, one takes a conservative strategy to transmit energy no smaller than \( B \) for any harvesting function in \( \Gamma(\beta, m) \). In other words, one seeks the codebook that achieves \( C_{\Gamma(\beta, m)}(B) \).

So for a given \( \beta \), the energy and information losses incurred by partial knowledge are defined as

\[
\Delta_{\text{E}}(R; \beta, \Gamma(\beta, m)) = B_{\beta}(R) - B_{\Gamma(\beta, m)}(R),
\]
\[
\Delta_{\text{I}}(B; \beta, \Gamma(\beta, m)) = C_{\beta}(B) - C_{\Gamma(\beta, m)}(B),
\]

and since the true \( \beta \) is unknown, the losses we investigate are the worst losses:

\[
\Delta_{\text{E}}(R) = \sup_{\beta \in \Gamma^K} \Delta_{\text{E}}(R; \beta, \Gamma(\beta, m)),
\]
\[
\Delta_{\text{I}}(B) = \sup_{\beta \in \Gamma^K} \Delta_{\text{I}}(B; \beta, \Gamma(\beta, m)).
\]

For noisy samples, we assume i.i.d. additive measurement noise \( Z_i \) with mean zero and variance \( \delta^2 \) so that samples are \( \{ (\frac{i}{m-1}, \beta(\frac{i}{m-1}) + Z_i) \}_{i=0}^{m-1} \). Since samples are noisy, unlike noiseless samples, one cannot certify the set of true harvesting functions and design codebook for all functions in the set. Hence, one reconsucts \( \hat{\beta}_m \) as accurately as possible and designs the codebook as if \( \hat{\beta}_m \) is the true harvesting function. Noting that \( \hat{\beta}_m \) depends on observational noise as well as \( \beta \), we know that \( \hat{\beta}_m \) is a stochastic mapping from \( \beta \). Those facts lead us to the expected losses and minimax definitions of them as follows.

\[
\Delta_{\text{E}}(R; \beta, \hat{\beta}_m) = \mathbb{E} \left[ |B_{\beta}(R) - B_{\hat{\beta}_m}(R)| \right],
\]
\[
\Delta_{\text{I}}(B; \beta, \hat{\beta}_m) = \mathbb{E} \left[ |C_{\beta}(B) - C_{\hat{\beta}_m}(B)| \right],
\]
\[
\Delta_{\text{E}}(R) = \inf_{\beta_m \in \Gamma^K} \sup_{\beta \in \Gamma^K} \Delta_{\text{E}}(R; \beta, \hat{\beta}_m),
\]
\[
\Delta_{\text{I}}(B) = \inf_{\beta_m \in \Gamma^K} \sup_{\beta \in \Gamma^K} \Delta_{\text{I}}(B; \beta, \hat{\beta}_m).
\]

Notice from the definition, it is immediate that \( \Delta_{\text{I}}(B) \) is upper-bounded by the unconstrained capacity \( C_{\text{max}} \), i.e., for any \( B \),

\[
\Delta_{\text{I}}(B) \leq C_{\text{max}} := \max_{P_X} I(X; Y),
\]

which will be shown to be tight at maximum energy.

III. SAMPLING LOSS IN ENERGY AND INFORMATION

This section addresses point-to-point SIET performance losses due to \( m \)-sample knowledge of the harvesting function. As will be seen later, the best transmitted energy based on \( \hat{\beta}_m \) is arbitrary close to that based on \( \beta \), and one can still design near-optimal codewords in terms of transmitted energy. Also the speed of convergence is optimal for noiseless samples under some conditions. The loss in information due to sampled knowledge vanishes at interior points of energy transmission, however, it could be arbitrary at the maximum energy transmission, say \( B_{\text{max}} \) for noiseless samples. Thus, a system designer needs to be careful when targeting \( B_{\text{max}} \) or should design with a small margin away from \( B_{\text{max}} \). We constructively propose kernel-based reconstruction for noiseless and noisy samples, yielding near-optimal performance guarantees on transmitted energy.

A. Noiseless Samples

Consider noiseless samples. Reconstructing a continuous signal from samples has been a popular topic in signal processing [25], [26], approximation theory [27], and many other engineering fields. Among numerous reconstruction methods, consider the spline method (our converse argument in Thm. 7 will show this to be a good choice), which has piecewise polynomials as interpolant kernels to achieve efficient implementation. Since it is a local technique, rather than a global polynomial approximation method such as Lagrange interpolation, the value of the reconstructed function \( \hat{f}_m(x) \) only depends on a few neighboring samples of \( x \) and numerical instability called Runge’s phenomenon does not appear [26]. See surveys [25], [28] for introductory material and [29] for details.

Before giving our main theorems and proofs, first recall the following result on spline reconstruction in Sobolev spaces.

Lemma 3 (Prop. 3.1 in [16]): For \( f \in W_{\lambda}^\infty \), let \( \hat{f}_m \in \Gamma(f, m) \) be the spline reconstructed function. Then, for some constant \( c \),

\[
\|f - \hat{f}_m\|_{\infty} \leq cm^{-\lambda}\|f(\lambda)\|_{\infty} \quad \forall f \in W_{\lambda}^\infty.
\]

Now we give a main result, which shows one can attain near-optimal transmitted energy despite the sampled harvesting function.

Theorem 4: \( \Delta_{\text{E}}(R) = O(m^{-\lambda}) \quad \forall R \geq 0 \).

Proof: Note that the best codebooks for \( B_{\beta}(R) \) and \( B_{\Gamma(\beta, m)}(R) \) are not necessarily identical. However, as will be seen, any codebook performs almost the same under \( \beta \) and \( \hat{\beta}_m \) in \( \Gamma(\beta, m) \).

First consider an arbitrary distribution \( P_X \) and Shannon’s random codebook generated from it. Then,

\[
\left| \mathbb{E}_{P_X} [\beta(X)] - \mathbb{E}_{P_X} [\hat{\beta}_m(X)] \right| \leq \mathbb{E}_{P_X} \left| \beta(X) - \hat{\beta}_m(X) \right| = \int_X P_X(x) |\beta(x) - \hat{\beta}_m(x)| dx
\]

\[
\leq \int_X P_X(x) |\beta - \hat{\beta}_m|_{\infty} dx = ||\beta - \hat{\beta}_m||_{\infty},
\]

where the last inequality follows from the sup-norm definition,

\[
||\beta - \hat{\beta}_m||_{\infty} = \text{ess sup}_{x \in X} |\beta(x) - \hat{\beta}_m(x)|.
\]

Furthermore, using triangular inequality, we have

\[
||\beta - \hat{\beta}_m||_{\infty} \leq ||\beta - \beta_m||_{\infty} + ||\beta_m - \hat{\beta}_m||_{\infty}.
\]

The first term is bounded by \( cm^{-\lambda}\|\beta(\lambda)\|_{\infty} \) by Lem. 3. Furthermore, note that \( \beta_m \) can be seen as a spline reconstruction for another \( \beta' \in \Gamma(\beta, m) \) since \( \beta, \beta' \) both agree on
sample points. This means the second term is also bounded by $cm^{-\lambda}||\beta(\lambda)||_\infty$. Therefore, from the definition of $\Gamma^K$,

$$|E_{P_X}[\beta(X)] - E_{P_X}[\hat{\beta}_m(X)]| \leq 2ckm^{-\lambda}. \quad (12)$$

It should be noted that (12) is independent of $P_X, \beta, \hat{\beta}_m$.

Next, fix $R \geq 0$ and consider $A := \{P_X : I(X:Y) \geq R\}$. Also define two capacity achieving distributions $P_X^*, Q_X^* \in A$ for $B_\beta(R), B_{\hat{\beta}_m}(R)$, respectively. Then, we have a chain of inequalities

$$B_\beta(R) \geq B_{\Gamma(\beta,m)}(R) = \min_{\hat{\beta}_m \in \Gamma(\beta,m)} E_{Q_X}[\hat{\beta}_m(X)]$$

$$(b)$$

$$\geq \min_{\hat{\beta}_m \in \Gamma(\beta,m)} E_{P_X}[\hat{\beta}_m(X)]$$

$$(c)$$

$$\geq E_{P_X}[\beta(X)] - 2ckm^{-\lambda}$$

$$= B_\beta(R) - 2ckm^{-\lambda},$$

where (a) follows from the definitions (4) and (5), (b) follows since $P_X^*$ is suboptimal for $B_{\Gamma(\beta,m)}(R)$, and (c) follows since (12) holds for all $\beta \in \Gamma^K$ and $\hat{\beta}_m \in \Gamma(\beta,m)$. Hence, we conclude that $\Delta_\beta(R;\beta,\Gamma(\beta,m)) = O(m^{-\lambda})$ for all $\beta \in \Gamma^K$. Since $R$ is arbitrary and the bound does not depend on $\beta$, $\Delta_\beta(R) = O(m^{-\lambda})$ for all $R$.

From the result, we know that the conservative transmission scheme performs near-optimally in terms of energy. However, the scheme needs optimization with respect to uncountably many $\hat{\beta}_m \in \Gamma(\beta,m)$, which does not reveal a clear codebook design. The following corollary suggests that $\hat{\beta}_m^{SP}$ is a good proxy for unknown $\beta$ enabling us to design near-optimal codewords as if $\hat{\beta}_m^{SP}$ is the true harvesting function.

Corollary 5: Codewords designed based on $\hat{\beta}_m^{SP}$ achieves $O(m^{-\lambda})$ loss of transmitted energy with respect to $B_\beta(R)$.

Proof: Fix an arbitrary $R \geq 0$ and consider $B_{\hat{\beta}_m^{SP}}(R), B_\beta(R)$. Two optimal codebooks are generated from the capacity achieving distributions for $B_{\hat{\beta}_m^{SP}}(R), B_\beta(R)$, say $P_X^{SP}, Q_X^{SP}$.

Then, under $\beta$ the optimal codebook for $\hat{\beta}_m^{SP}$ (i.e., $P_X^{SP}$) performs as:

$$|B_{\hat{\beta}_m^{SP}}(R) - E_{P_X^{SP}}[\beta(X)]|$$

$$\leq \left|E_{P_X}[\hat{\beta}_m^{SP}(X)] - E_{P_X}[\beta(X)]\right|$$

$$(a)$$

$$\leq ||\beta - \hat{\beta}_m^{SP}||_\infty \leq cKm^{-\lambda},$$

where (a) follows from (12). As $P_X^*$ is suboptimal for $\beta$, we know that

$$B_\beta(R) \geq B_{\hat{\beta}_m^{SP}}(R) - cKm^{-\lambda}.$$

Similarly, exchanging roles of $\beta, \hat{\beta}_m^{SP}$ and considering the optimal codebook for $\beta$ (i.e., $Q_X^*$) gives

$$|B_\beta(R) - E_{Q_X}[\hat{\beta}_m^{SP}(X)]| \leq ||\beta - \hat{\beta}_m^{SP}||_\infty \leq cKm^{-\lambda}.$$

As $Q_X^*$ is suboptimal for $\hat{\beta}_m^{SP}$, we know that

$$B_{\hat{\beta}_m^{SP}}(R) \geq B_\beta(R) - cKm^{-\lambda}.$$

Combining the two, we have

$$B_{\hat{\beta}_m^{SP}}(R) + cKm^{-\lambda} \leq B_\beta(R) \leq B_{\hat{\beta}_m^{SP}}(R) - cKm^{-\lambda}.$$

Hence, we can conclude that the codebook designed based on $\hat{\beta}_m^{SP}$ is nearly optimal within $O(m^{-\lambda})$.

It should be noted that Thm. 4 is not tight in general, e.g., consider a peak-power constrained AWGN channel [24] and suppose the capacity achieving distribution, which is discrete, is supported on (a part of) sample points. As $\beta, \hat{\beta}_m$ always agree on sample points, $\Delta_\beta(R) = 0$. However, there are cases such that the bound in Thm. 4 is tight. Before proceeding to demonstration, we define function-wise loss.

$$\Delta_\beta(R;\beta,\hat{\beta}_m) = |B_\beta(R) - B_{\hat{\beta}_m}(R)|,$$

$$\Delta_\beta(R) = \sup_{\beta \in \Gamma^K, \hat{\beta}_m \in \Gamma(\beta,m)} \Delta_\beta(R;\beta,\hat{\beta}_m).$$

Lemma 6: $\Delta_\beta(R) \leq \Delta_\beta(R)$. Proof: Consider the left side

$$\Delta_\beta(R) = \sup_{\beta,\hat{\beta}_m} \left|B_\beta(R) - B_{\hat{\beta}_m}(R)\right|$$

and note that $\hat{\beta}_m$ is a candidate for $\beta$, but, $\beta$ is also a candidate for $\hat{\beta}_m$ since they both agree on the sample points. Hence, we can exchange $\beta, \hat{\beta}_m$ and without loss of generality, it is sufficient to consider pairs $(\beta, \hat{\beta}_m)$ such that $B_\beta(R) \geq B_{\hat{\beta}_m}(R)$. For any such a $(\beta, \hat{\beta}_m)$,

$$B_\beta(R) - B_{\hat{\beta}_m}(R) \leq B_\beta(R) - B_{\Gamma(\beta,m)}(R)$$

by definition of $B_{\Gamma(\beta,m)}(R)$. Taking supremum over all such $(\beta, \hat{\beta}_m)$ does not change the inequality, which completes the proof.

Therefore, to show the lower bound of $\Delta_\beta(R)$, it is sufficient to show the lower bound of $\Delta_\beta(R)$. The following theorem states conditions for which $\Delta_\beta(R) = \Omega(m^{-\lambda})$, i.e., the bound is tight.

Theorem 7: Fix some $B \in (0, B_{max})$. Suppose the capacity-achieving distribution $P_X^*$ yielding the Shannon’s random codebook of rate $R = C_{\hat{\beta}_m}(B)$ satisfies one of the following conditions:

1) $P_X^*$ is continuous and non-vanishing on $\mathcal{X}$, i.e., $P_X^*(x) \geq c$ for some $c$.

2) $P_X^*$ is supported on a finite set of mass points disjoint from the sample points, as specified in the proof.

Then, $\Delta_\beta(R) = \Omega(m^{-\lambda})$ at $R$.

Proof: We consider $\Delta_\beta(R;\beta,\hat{\beta}_m)$ and the lower bound can be shown by a bumpy function. Thm. 4.3 in [33] states that there exists a non-negative function $f$ such that $f(x_i) = 0$ at every $x_i$ and $\|f\|_1 \geq c'm^{-\lambda}$. First consider the case 1). Take $\beta, \hat{\beta}_m$ as

$$\beta(x) = M \quad \forall x \in \mathcal{X},$$

$$\hat{\beta}_m(x) = M(1 - f(x)) \quad \forall x \in \mathcal{X},$$

4 The discrete distribution is particularly important because the optimal input distribution is discrete in many cases especially when $\mathcal{X}$ is compact and convex and channel noise is additive, see [21], [24], [29], [31]. Also refer to [22] for general channels.
where $M$ is a constant. Then, $B_\beta(R) = M$ for any codebook. Also,

$$B_{\beta_m}(R) = \mathbb{E}\left[\hat{\beta}_m(X)\right] = \int_X P_X(x)\hat{\beta}_m(x)dx$$

$$= \int_X P_X^*(x)M(1 - f(x))dx = M - M \int_X P_X^* f(x)dx$$

$$\leq M - M \int_X cf(x)dx = M - cM\|f\|_1$$

$$\leq M - c'cMm^{-\lambda}.$$  

Thus, $\Delta_E(R; \beta, \hat{\beta}_m) = |B_\beta(R) - B_{\beta_m}(R)| \geq c'cMm^{-\lambda}$. We have the desired lower bound of $\Delta_E(R)$ as $\Omega(m^{-\lambda}).$

For the case 2), we repeat the above argument of $\beta(x) = M, \beta_m = M(1 - f(x))$. Since $P_X$ is supported on a discrete set, say $\{x_k\}$,

$$B_\beta(R) - B_{\beta_m}(R) = M \int_X P_X^*(dx)f(x)$$

$$= M \sum_k P_X^*(x_k)f(x_k).$$

Note that by the norm monotonicity with respect to a bounded measure, $\|f\|_\infty \geq \|f\|_1 \geq c'c'cMm^{-\lambda}$, there is a disjoint point from samples such that $f(x) \geq c'cMm^{-\lambda}$. So when $\{x_k\}$ satisfy $f(x_k) \geq c'cMm^{-\lambda}$,

$$B_\beta(R) - B_{\beta_m}(R) \geq M \sum_k P_X^*(x_k)c'cMm^{-\lambda} = M \cdot c'cMm^{-\lambda},$$

which proves $\Delta_E(R) = \Omega(m^{-\lambda}).$

The next theorem and corollary deal with the information loss incurred by sampling. As will be seen below, the loss is negligible on most of the targeted energy range, however, the trivial unconstrained capacity upper bound on $\Delta_{\beta}(I)$ given as [10] could be indeed tight at $B_{\max}$.

**Theorem 8:** For $B \in [0, B_{\max})$,

$$\Delta_{\beta}(I) \to 0 \quad \text{as} \quad m \to \infty.$$  

Furthermore, there is a channel for which $\Delta_{\beta}(I_{\max}) = C_{\max}.$

**Proof:** Let us prove the first claim. At $B = 0$, note that it is the same as the unconstrained capacity, i.e., $C_{\beta}(0) = C_{\hat{\beta}}(0) = C_{\max}$. So $\Delta_{\beta}(I_{\max}) = 0$. For $B \in (0, B_{\max})$, recall that since $C_{\beta}(B)$ is concave, it is continuous over the interior of its domain, i.e., continuous on $(0, B_{\max})$. Thm. 4 guarantees that for every $B$, there exists a $B'$ that attains $C_{\Gamma(\beta, m)}(B) = C_{\beta}(B')$ for some close $B, B'$ with $|B - B'| = O(m^{-\lambda})$, so that at $B \in (0, B_{\max}),$

$$\Delta_{\beta}(B; \beta, \Gamma(\beta, m)) = C_{\beta}(B) - C_{\Gamma(\beta, m)}(B)$$

$$= C_{\beta}(B) - C_{\beta}(B')$$

$$= C_{\beta}(B) - C_{\beta}(B + O(m^{-\lambda})).$$

Due to the continuity of $C_{\beta}$, $\Delta_{\beta}(B; \beta, \Gamma(\beta, m)) \to 0$ as $B + O(m^{-\lambda}) \to B$. It means $\Delta_{\beta}(I) \to 0$ as well.

To show the second claim, fix a large $m$. We will prove by a counterexample. Take a constant $\beta$, that is, $\beta(x) = M$ over all $x$. Then, as any $P_X$ is admissible for $B \leq M$ and none is for $B > M,$

$$C_{\beta}(B) = \begin{cases} 0 & \text{if } B < M \vphantom{\int} \\ C_{\max} & \text{if } B > M \end{cases}.$$  

However, $\Gamma(\beta, m)$ definitely has an element such that $\hat{\beta}_m(x) < \beta(x) = M$ except for given sample points. In other words, $\hat{\beta}_m < \beta$ almost everywhere, so that $\mathbb{E}[\hat{\beta}_m(X)] < M$ unless $P_X$ only has point masses on the sample points. Therefore, discrete $P_X$ s are the only admissible probability distributions for the energy requirement $M(= B_{\max})$.

For such a discrete $P_X$, consider an adversarial channel

$$Y = (X + Z) \mod 1,$$

and $Z$ is an input-dependent additive noise on $X = [0, 1]$. The dependency is as follows: $Z$ is uniform over $[0, 1]$ when $X \in \left\{\frac{i}{m-1}\right\}_{i=0}^{m-1}$, and the probability density of $Z$ is more concentrated around 0 as $X$ is more distant from $\left\{\frac{i}{m-1}\right\}_{i=0}^{m-1}$. Since the discrete $P_X$ only sees uniform noise, $I(X; Y)$ is zero, i.e., $C_{\Gamma(\beta, m)}(M) = 0$, however, we can send information using a non-discrete $P_X$ because noise is biased toward 0 except for sample points. Hence, $\Delta_{\beta}(I_{\max}) = C_{\max}$ for this channel.

Although Thm. 8 describes the convergence of $\Delta_{\beta}$, it does not characterize $\Delta_{\beta}$ in terms of the number of samples. As the next corollary shows, the Lipschitz continuity enables us to characterize $\Delta_{\beta}(I)$ in terms of $m$ for all $B \in (0, B_{\max})$.

**Corollary 9:** Suppose the channel yields Lipschitz continuous $C_{\beta}(B)$ with Lipschitz coefficient $M$ for all $\beta \in \Gamma^K$ except for its end points, i.e., for any $\beta \in \Gamma^K$ and $B_1, B_2 \in (0, B_{\max})$,

$$|C_{\beta}(B_1) - C_{\beta}(B_2)| \leq M|B_1 - B_2|.$$  

Then, $\Delta_{\beta}(I) = O(m^{-\lambda})$ for any $B \in (0, B_{\max})$.

**Proof:** When $B = 0$, it is unconstrained capacity, so $\Delta_{\beta}(I) = 0$. For $B \in (0, B_{\max})$ and a given $\beta \in \Gamma^K$,

$$\Delta_{\beta}(I; \beta, \Gamma(\beta, m)) = C_{\beta}(B) - C_{\Gamma(\beta, m)}(B)$$

$$\leq C_{\beta}(B) - C_{\beta}(B + O(m^{-\lambda}))$$

$$\leq MC_{\beta}(B) - O(m^{-\lambda}).$$

where the last inequality follows from [13]. The bound is independent of $\beta$ so taking supremum does not increase the bound. Therefore, $\Delta_{\beta}(I) = O(m^{-\lambda})$ for any $B \in (0, B_{\max})$.

Thm. 4 and Cor. 5 ensure Shannon’s random codebook designed as $\beta_{\max}$ is nearly close to the optimal codebook for $\beta$ in terms of transmitted energy. Further, Thm. 8 shows that its performance is in fact asymptotically tight under some conditions on $P_X^*.$

From the same argument, Thm. 8 and Cor. 9 both basically ensure that the codebook designed as $\beta_{\max}$ is the true harvesting function also delivers nearly maximal information. However, please be careful when interpreting the second statement of Thm. 8. The statement does not imply the codebook fails to be decoded correctly at $B_{\max}$; rather it means that partial knowledge of the harvesting function may lower (or set higher) the targeted information rate by a non-vanishing amount in the codebook design stage. However such a mismatched codebook is always decodable since the channel remains the same regardless of sampling. This pitfall leads a system designer to stepping back from $B_{\max}$, i.e., setting a safety energy margin from $B_{\max}$. 


### B. Noisy Samples

Consider noisy samples. In particular, received signal varies even for the same transmission signal. Or the noise could be due to errors in measuring battery status. In particular, we consider i.i.d. additive noise $Z_i$, with mean zero and variance $\sigma^2$ so that samples are $\{(x_i, Y_i)\}_{i=0}^{m-1}$, where $x_i = \frac{i}{m-1}, Y_i = \beta(\frac{i}{m-1}) + Z_i$.

As a constructive reconstruction method, we consider local polynomial estimation of order $\lambda$ [17], denoted by $\hat{\beta}^{LP}_m$, since $\Gamma^K$ is differentiable up to order $\lambda$. Consider a symmetric kernel $\phi(x)$ on $[-1, 1]$ such that $|\phi(x)| \leq \phi_{max} < \infty$ and let $h$ be bandwidth. Then, $\hat{\beta}^{LP}_m(x)$ for a particular $x$ is obtained from $\{w_t\}_{t=0}^{m-1}$ that solves

$$\min_{w_t} \sum_{i=0}^{m-1} \phi\left(\frac{x_i - x}{h}\right) \left(Y_i - \sum_{t=0}^{\lambda} w_t(x_i - x)^t\right)^2.$$  

(14)

To express $\hat{\beta}^{LP}_m(x)$ in closed form, it is convenient to introduce vector and matrix representations:

$$X_x = \begin{bmatrix}
1 & (x_0 - x) & \cdots & (x_0 - x)^\lambda \\
1 & (x_1 - x) & \cdots & (x_1 - x)^\lambda \\
\vdots & \vdots & \ddots & \vdots \\
1 & (x_{m-1} - x) & \cdots & (x_{m-1} - x)^\lambda
\end{bmatrix}^T,$$

$$Y = [Y_0, Y_1, \ldots, Y_{m-1}]^T,$$

$$w = [w_0, w_1, \ldots, w_{\lambda}]^T,$$

$$\Phi_x = \begin{bmatrix}
\phi\left(\frac{x_0 - x}{h}\right) & \cdots & 0 \\
\phi\left(\frac{x_1 - x}{h}\right) & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \phi\left(\frac{x_{m-1} - x}{h}\right)
\end{bmatrix}.$$  

Then, (14) is rewritten as a least square problem

$$\min_w (Y - X_x w)^T \Phi_x (Y - X_x w),$$

and the solution to this is

$$w^* = [w_0^*, w_1^*, \ldots, w_{\lambda}^*]^T = (X_x^T \Phi_x X_x)^{-1}(X_x^T \Phi_x Y).$$

Then, $\hat{\beta}^{LP}_m(x) = w_0$, in other words,

$$\hat{\beta}^{LP}_m(x) = e_1^T (X_x^T \Phi_x X_x)^{-1}(X_x^T \Phi_x Y),$$  

(15)

where length $(\lambda + 1)$ vector $e_1$ has a 1 in the first coordinate and 0s otherwise. In particular when the order is zero, it is called the Nadaraya-Watson estimator [17].

**Lemma 10 (Thm. 1.6 in [17]):** If $h = h_m = \alpha m^{-\frac{\lambda+1}{2\lambda+1}}$ for some $\alpha > 0$, the following estimation error bound holds for $\beta \in \Gamma^K$:

$$\sup_{x \in \mathcal{X}} \mathbb{E} \left[(\beta(x) - \hat{\beta}^{LP}_m(x))^2\right] = O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right).$$  

(16)

For further results in nonparametric regression, see [17], [34].

Like for noiseless samples, the following theorem shows that the average loss $\bar{\Delta}_E(R)$ incurred due to sampled knowledge about $\beta$ is asymptotically negligible.

**Theorem 11:** For $R \geq 0$,

$$\bar{\Delta}_E(R) = O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right).$$

**Proof:**

First note that due to the Jensen’s inequality,

$$\sup_{x \in \mathcal{X}} \mathbb{E} \left[(\beta(x) - \hat{\beta}^{LP}_m(x))^2\right] \leq \mathbb{E} \left[(\beta(x) - \hat{\beta}^{LP}_m(x))^2\right] = O \left(m^{-\frac{2(\lambda+1)}{2\lambda+1}}\right),$$

which implies

$$\mathbb{E} \left[(\beta(x) - \hat{\beta}^{LP}_m(x))^2\right] = O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right) \quad \forall x \in \mathcal{X}. \quad (17)$$

Now fix $P_X$ so that rate $R = I(X; Y)$ is also fixed. The expectation in (16) is over the sampling noise distribution,

$$\mathbb{E}_Z \left[\mathbb{E}_X [\beta(X)] - \mathbb{E}_X [\hat{\beta}^{LP}_m(X)]\right]$$

$$\leq \mathbb{E}_Z \left[\mathbb{E}_X [\beta(X) - \hat{\beta}^{LP}_m(X)]\right]$$

$$= \mathbb{E}_X \left[\mathbb{E}_Z [\beta(X) - \hat{\beta}^{LP}_m(X)]\right]$$

$$\leq \mathbb{E}_X \left[O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right)\right] \equiv O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right),$$

where (a) follows from (17) and (b) follows since (17) holds for every $x$. By the same argument as in the proof of Thm. 4 we know that

$$\bar{\Delta}_E(R; \beta, \hat{\beta}^{LP}_m) = \mathbb{E}_Z \left[|B_\beta(R) - B_{\hat{\beta}^{LP}_m}(R)|\right] = O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right),$$

which does not depend on $\beta$.

As $\beta \in \Gamma^K$, $R \geq 0$ are arbitrary, and the local polynomial estimator is a particular choice of estimator, taking the infimum over all estimators implies $\bar{\Delta}_E(R) \leq \bar{\Delta}_E(R; \beta, \hat{\beta}^{LP}_m) = O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right).$

Paralleling arguments for noiseless samples, the information loss can be also specified.

**Corollary 12:** The following are true:

1) For $B \in [0, B_{max})$, $\bar{\Delta}_1(B) \to 0$ as $m \to \infty$.

2) Suppose the channel yields $M$-Lipschitz continuous $C_\beta(B)$ for all $\beta \in \Gamma^K$. Then, $\bar{\Delta}_1(B) = O \left(m^{-\frac{\lambda+1}{2\lambda+1}}\right)$ for all $B \in [0, B_{max})$.

Proofs are basically the same as the proofs of Thm. 8 and Cor. 9 so omitted.

### IV. Sampling Loss in SiET Multicast

Now we investigate the multicast setting in Fig. 2. Consider a single transmitter (i.e., access point) and $L$ receiver nodes. The transmitter sends a signal $X^u$ which conveys not only a common message $W$, but also energy to operate each node. These nodes observe $Y(t)^u$ through individual channels and have their own harvesting functions $\beta(t) \in \Gamma^K, t = 1, \ldots, L$ and energy requirements $b(t)$, which are not necessarily identical since physical devices may be different. As before, we are limited in knowing the harvesting functions only at sample points either in the absence or presence of noise.

The next proposition states the capacity-energy tradeoff for the SiET multicast problem with full knowledge of harvesting functions [35]. Here, superscript (MC) explicitly denotes that
it is a multicast quantity. For notational simplicity, we use vector notations

$$B = [B^{(1)}, \ldots, B^{(L)}],$$
$$\beta = [\beta^{(1)}, \ldots, \beta^{(L)}],$$
$$\hat{\beta}_m = [\hat{\beta}_m^{(1)}, \ldots, \hat{\beta}_m^{(L)}],$$
$$\Gamma(\beta, m) = [\Gamma(\beta^{(1)}, m), \ldots, \Gamma(\beta^{(L)}, m)].$$

Proposition 13 (Thm. 1 in [33]): For L-user SIET multicast, the capacity-energy function is given by

$$C_{\beta}^{(MC)}(B) = \max_{P_X} \min_{1 \leq \ell \leq L} I(X; Y^{(\ell)}).$$

Also like [3], it is easy to extend to the set of possible harvesting functions.

$$C_{\Gamma(\beta, m)}(B) = \max_{P_X} \min_{1 \leq \ell \leq L} I(X; Y^{(\ell)}).$$

Let $B^{(\ell)}(R), B_{\Gamma(\beta, m)}^{(\ell)}(R)$ be the amounts of energy delivered to $\ell$th node using the rate $R$ codebook designed for $\beta$ and $\Gamma(\beta, m)$, respectively, that is,

$$B^{(\ell)}(R) = \max_{P_X \in \mathcal{H}} \mathbb{E}[\beta^{(\ell)}(X)] I(X; Y^{(\ell)}),$$
$$B_{\Gamma(\beta, m)}^{(\ell)}(R) = \max_{P_X \in \mathcal{H}} \min_{\beta_m \in \Gamma(\beta^{(\ell)}, m)} \mathbb{E}[\hat{\beta}_m(X)] I(X; Y^{(\ell)}).$$

Hence, sampling losses (6–9) defined for the point-to-point case extend to multicast as follows. Note that $\Delta_{E}^{(MC)}(R), \Delta_{E}^{(MC)}(B)$ are for noiseless samples and $\Delta_{E}^{(MC)}(R), \Delta_{E}^{(MC)}(B)$ are for noisy samples.

$$\Delta_{E}^{(MC)}(R) = \sup_{\beta \in \mathbb{K}} \max_{1 \leq \ell \leq L} B^{(\ell)}(R) - B_{\Gamma(\beta, m)}^{(\ell)}(R),$$
$$\Delta_{E}^{(MC)}(B) = \sup_{\beta \in \mathbb{K}} C_{\beta}^{(MC)}(B) - C_{\Gamma(\beta, m)}^{(MC)}(B),$$
$$\Delta_{E}^{(MC)}(R) = \inf_{\beta \in \mathbb{K}} \sup_{1 \leq \ell \leq L} \mathbb{E} \left[ |B^{(\ell)}(R) - B_{\Gamma(\beta, m)}^{(\ell)}(R)| \right],$$
$$\Delta_{E}^{(MC)}(B) = \inf_{\beta \in \mathbb{K}} \sup_{1 \leq \ell \leq L} \mathbb{E} \left[ |C_{\beta}^{(MC)}(B) - C_{\Gamma(\beta, m)}^{(MC)}(B)| \right].$$

Note that $\Delta_{E}^{(MC)}(B), \Delta_{E}^{(MC)}(B)$ do not have maximum over $\ell$ because all nodes receive the same information in multicast.

Theorem 4 (Noisless samples): The asymptotic bounds in Thms. 4 and Cor. 3 hold for multicast when samples are noiseless, that is:

1. $\Delta_{E}^{(MC)}(R) = O(m^{-\lambda}).$
2. $\Delta_{E}^{(MC)}(B) \to 0$ as $m \to \infty$ if $B^{(\ell)} \in [0, B_{\max})$ for all $\ell$.
3. Letting $C_{\beta}^{(MC)} := \max_{P_X} \min_{1 \leq \ell \leq L} I(X; Y^{(\ell)})$, there exists a channel such that $\Delta_{E}^{(MC)}(B) = C_{\beta}^{(MC)}$ if some $B^{(\ell)} = B_{\max}^{(\ell)}$.
4. Suppose $C_{\beta}^{(MC)}(B)$ is $M$-Lipschitz with $\ell_q$ norm, where $1 \leq q \leq \infty$, that is,

$$|C_{\beta}(B_1) - C_{\beta}(B_2)| \leq M \|B_1 - B_2\|_q.$$
(\hat{a}_m, \hat{\beta}_m). Informally, \((\hat{a}_m, \hat{\beta}_m)\) is close to the true pair when the number of samples is large. Analogous to our main result in Sec. [III] for the SIET channel coding problem, the source coding problem with sampled distortion measure was studied by Niesen, et al. [21] who showed that the distortion loss vanishes as the number of samples increases. See the Appendix for detailed problem setting and results with its extension to noisy samples. Further, we have shown that designing codebooks as if \((\hat{\beta}_m, \hat{d}_m)\) are the true functions is nearly optimal for noiseless and noisy cases. Hence, the question that naturally follows is whether \(\pi(\hat{\beta}_m, \hat{d}_m)\) is also close to \(\pi(\beta, d)\).

For two optimal tradeoff curves \(\pi(\beta, d), \pi(\hat{\beta}_m, \hat{d}_m)\), let us define loss incurred by sampling. Let \(\Pi_{\beta}(B, D)\) be the projection of \((B, D)\) onto curve \(\pi\) under \(\ell_1\) distance; when there are several projection points, pick any one arbitrarily. Then we define two component losses for noiseless and noisy samples, respectively, as:

\[
\Delta(\beta, d, \hat{\beta}_m, \hat{d}_m) = \sup_{(B', D') \in \pi(\hat{\beta}_m, \hat{d}_m)} \|\Pi_{\beta}(B', D') - \Pi_{\pi(\beta, d)}(B', D')\|_1,
\]

\[
\tilde{\Delta}(\beta, d, \hat{\beta}_m, \hat{d}_m) = \sup_{(B', D') \in \pi(\hat{\beta}_m, \hat{d}_m)} \mathbb{E}[\|\Pi_{\beta}(B', D') - \Pi_{\pi(\beta, d)}(B', D')\|_1].
\]

By definition, \(\Delta\), \(\tilde{\Delta}\) are the maximal possible losses from the true optimal curve when we design optimal end-to-end transmission as if \((\hat{\beta}_m, \hat{d}_m)\) is the true harvesting and distortion function pair.\(^5\) By Shannon’s separation theorem \([37]\), any operating point \((B, D)\) in \(\pi\) can be attained by a separately designed pair of good source and channel codes. Moreover, distortion loss in source coding and harvesting loss in channel coding due to sampling vanish by results in \([21]\) (restated in Appendix) and Sec. [III] Thus one might conjecture that a system design based on \((\hat{\beta}_m, \hat{d}_m)\) is nearly optimal, i.e., \(\Delta, \tilde{\Delta} \to 0\) as \(m \to \infty\). This is partially true with additional restriction on harvesting and distortion functions. The following theorem formally shows it.

**Theorem 17:** Define two sets, \(\mathcal{B} := \{\beta \in \Gamma^K : C_{\beta}(B)\text{ is Lipschitz over all } B \geq 0\}\), \(\mathcal{D} := \{d(\cdot, \hat{s}) \in \Gamma^K \forall \hat{s} : R_d(D)\text{ is Lipschitz over all } D \geq 0\}\), and two minmax losses:

\[
\Delta := \inf_{\beta_m, \hat{d}_m} \sup_{\beta \in \mathcal{B}, d \in \mathcal{D}} \Delta(\beta, d, \hat{\beta}_m, \hat{d}_m),
\]

\[
\tilde{\Delta} := \inf_{\beta_m, \hat{d}_m} \sup_{\beta \in \mathcal{B}, d \in \mathcal{D}} \tilde{\Delta}(\beta, d, \hat{\beta}_m, \hat{d}_m).
\]

Then, \(\Delta = O(m^{-\lambda})\) and \(\tilde{\Delta} = O\left(m^{-\frac{\lambda + 1}{2\lambda + 2}}\right)\).

**Proof:** Consider \((\hat{\beta}_m, \hat{d}_m)\) are estimated by the spline method for noiseless samples and by the local polynomial regression for noisy samples, i.e., \((\hat{\beta}_m, \hat{d}_m) = (\hat{\beta}_m^{SP}, \hat{d}_m^{SP})\) for noiseless and \((\hat{\beta}_m, \hat{d}_m) = (\hat{\beta}_m^{LP}, \hat{d}_m^{LP})\) for noisy samples. Let us only focus on noiseless samples. Proof will be shown by contradiction: suppose that there exists \((B', D') \in \pi(\hat{\beta}_m^{SP}, \hat{d}_m^{SP})\) such that the \(\ell_1\)-balls centered at \((B', D')\) with radius \(O(m^{-\lambda})\) has no intersection with \(\pi(\beta, d)\).

First consider the optimal codebook pair at \((B', D')\). Although the channel codebook is designed for \(\beta_m^{SP}\), actual harvested energy \(B\) is also close to \(B'\), i.e., \(B = B' + O(m^{-\lambda})\). Similarly, the source codebook also achieves the actual distortion \(D = D' + O(m^{-\lambda})\). Since these codebooks are suboptimal for the true \((\beta, d)\), there will be a point \((B^{*}, D^{*})\) on \(\pi(\beta, d)\) such that \(B^{*} > B \geq B^{*} - cm^{-\lambda}\) and \(D^{*} \leq D < D^{*} + cm^{-\lambda}\).

Pick a point \((B^{*}, D^{*}) \in \pi(\beta, d)\) such that \(D^{*} = D'\). We know that this point exists from the Lipschitz continuity. From the assumption, we know that \(B^{*}\) is outside of the \(\ell_1\)-ball, i.e., \(B^{*} > B^{*} + cm^{-\lambda}\). Consider the optimal codebook pair at \((B^{*}, D^{*})\). From the first argument of the proof of Thm. [4] we know that the channel codebook delivers energy \(B^{*} + O(m^{-\lambda})\) under harvesting function \(\beta_m^{SP}\). However, this codebook is definitely suboptimal for \(\beta_m^{SP}\), which means that \(\pi(\beta_m^{SP}, d_m^{SP})\) has a point \((D^{*'}, B')\) such that \(B' > B^{*} - cm^{-\lambda}\).

This implies \(|B' - B^{*}| \leq cm^{-\lambda}\), a contradiction. Therefore, \(\Delta(\beta, d, \beta_m^{SP}, d_m^{SP}) = O(m^{-\lambda})\). Since the bound is independent of \((\beta, d)\) and \((\beta_m^{SP}, d_m^{SP})\) are specific reconstructions, we can further reduce the loss. Therefore, \(\Delta = O(m^{-\lambda})\) holds. The argument is illustrated in Fig. [5].

For noisy sample, the arguments still hold with \((\beta_m^{LP}, d_m^{LP})\) so \(\tilde{\Delta} = O\left(m^{-\frac{\lambda + 1}{2\lambda + 2}}\right)\).

Fig. 5. The proof of Thm. [17] Solid curves and dotted curves denote \(\pi(\beta, d)\) and \(\pi(\hat{\beta}_m, \hat{d}_m)\), respectively. The left illustrates that there is no point in the \(\ell_1\)-ball centered at \((B', D')\), drawn in red. The right illustrates that the channel codebook at \(B^{*}\) performs \(B'\), marked as triangle, under \(\hat{\beta}_m\).

Fig. 6. An example in the proof of Thm. [15] Solid curves denote true quantities \(C_{\beta}(B), R_d(D), \pi(\beta, d)\) and dotted curves denote quantities for estimated functions. Note that \(\hat{d}_m\) and \(\pi(\beta, d)\) is degenerate. \(\kappa = \frac{k_1}{k_2}\) indicates the ratio that \(k_1\) source symbols are mapped to \(k_2\) channel symbols.

\(^5\)Note that \(\Delta(\beta, d, \beta_m, d_m), \tilde{\Delta}(\beta, d, \beta_m, d_m)\) are well-defined even when \(\pi(\beta, d)\) or \(\pi(\beta_m, d_m)\) is degenerate.

\(^6\)Also we can consider the other direction of projection, which is projection from \(\pi(\beta, d)\) onto \(\pi(\beta_m, d_m)\). But, since what we want to know is how close our estimation is to the true one, this makes less sense in practice.
\( \Delta(\beta, d, \hat{\beta}_m, \hat{d}_m) \) and \( \Delta(\beta, d, \tilde{\beta}_m, \tilde{d}_m) \) could be arbitrary large unless \( \beta \in B, d \in D \), even when \( \hat{\beta}_m, \hat{d}_m \) are sufficiently accurate. It suggests the possibility that accurate reconstruction may not be enough to provide performance guarantee for end-to-end communication.

**Theorem 18:** There exists a case where \( \Delta(\beta, d, \hat{\beta}_m, \hat{d}_m), \Delta(\beta, d, \tilde{\beta}_m, \tilde{d}_m) \) are bounded away from 0 even when \( m \to \infty \).

**Proof:** Consider an example with noiseless samples illustrated in Fig. 6. For the source coding part, suppose the \( R_d \) curve is strictly convex and assume that our estimate is perfect, i.e., \( d = \hat{d}_m \) so that \( R_d(D) = R_{d_m}(D) \).

For the channel part, suppose \( \beta(x) = \tilde{M} \) for all \( x \in \mathcal{X} \) for some constant \( \tilde{M} \). Then, every \( P_X \) is admissible with respect to energy requirement \( M \) since every \( P_X \) achieves \( \mathbb{E}[\beta(X)] = M \). Let \( C_{\text{max}} = \max_{P_X} I(X;Y) \) and \( P_X^* \) be the unique capacity-achieving distribution which is non-vanishing everywhere on \( \mathcal{X} \). By the separation theorem, this combination yields a degenerate JSCC curve \( \pi(\beta, d) = (M, D_{\text{min}}) \).

\( C_{\beta}(B), R_d(D), \pi(\beta, d) \) are illustrated with solid line.

On the other hand, suppose our estimate is \( \hat{\beta}_m(x) = \tilde{M} \) where \( f(x) \) is a small non-negative bumpy function such that \( f(x_i) = 0 \) only at every \( x_i \). There are two end points in \( C_{\hat{\beta}_m} \): One point is induced by \( P_X^* \), which still best in information delivery, however, \( \mathbb{E}_{P_X^*}[\hat{\beta}_m] = M - \epsilon \) for some \( \epsilon > 0 \). The other is by some discrete probability, that is, engineers design a codebook that only utilizes a finite number of points in \( \mathcal{X} \), which is strictly suboptimal in information transmission. Since \( \beta_m(x) = \tilde{M} \) only at \( x_i \), the transmitted energy is maximized when \( P_X \) has only point mass on \( x_i \), but such restriction on distribution incurs non-vanishing mutual information loss. Therefore resulting \( \pi(\hat{\beta}_m, \hat{d}_m) \) is a convex curve connecting \((M - \epsilon, D_{\text{min}}) \) and \((M, D') \). Therefore,

\[
\| ( \mathcal{M}, D' ) - \Pi_{\epsilon}(\beta, d)( \mathcal{M}, D' ) \| = D' - D_{\text{min}},
\]

which is non-vanishing, so \( \Delta(\beta, d, \hat{\beta}_m, \hat{d}_m) \) is also non-vanishing.

The argument for \( \Delta(\beta, d, \beta_m, \tilde{d}_m) \) is immediate since \( \Delta(\beta, d, \hat{\beta}_m, \hat{d}_m) \geq \Delta(\beta, d, \beta_m, \tilde{d}_m) \),

**VI. CONCLUSION**

We have studied performance loss in SIET due to experimentally-sampled harvesting functions. To our knowledge, this is the first study of how sampled knowledge of perhaps inexact and nonideal harvesting circuits affects SIET (or SWIPT). Energy loss and information loss are separately considered for noiseless and noisy samples, and extended to multicasting settings. We show theoretical asymptotics for these losses that energy loss asymptotically vanishes as \( O(m^{-\lambda}) \) for noiseless samples and it is indeed asymptotically optimal under some technical conditions. For noisy samples, the speed of convergence in energy loss is lowered to \( O(m^{-\lambda - \lambda_1} + \epsilon) \) due to noise in characterizing the harvesting circuit.

We also suggest spline and local polynomial reconstruction as practical reconstruction methods that attain the above asymptotics. B-spline (basis-spline) method requires \( O(m^2) \) complexity [38] and the local polynomial estimator at each

\( x \) requires complexity at most polynomial in \( m \) since [15] resulted from matrix algebra.

With regard to information loss, large number of samples does not always guarantee vanishing information loss. To get a vanishing information loss, a certain energy margin from \( B_{\text{max}} \) needs to be guaranteed. Hence, it is necessary for system designers to set a sufficient energy transmission margin from \( B_{\text{max}} \).

Another important problem is end-to-end information transmission. Motivated by [21], which shows the optimal source code for a sampled distortion function is also near-optimal for the true distortion function, one might guess that Shannon’s separation theorem would yield a combination of near-optimal source code and channel code that combine to be near-optimal in the energy-distortion tradeoff. It is true when further restriction is given on harvesting and distortion functions.

**APPENDIX**

Let us restate the main result of [21], which considers the lossy source coding problem with noiseless samples of the distortion function. The following assumptions are made on the source component. Suppose \( \mathcal{S} = [0, 1], \mathcal{S} \) is some discrete set, and \( d(\cdot, \hat{s}) \in \mathcal{F}^K \) for all \( \hat{s} \in \mathcal{S} \). For instance, \( \mathcal{S} \) is a set of images, \( \mathcal{S} \) is a set of quantized images or labels of images, and \( d(s, \hat{s}) \) is human perception loss which is unknown. Let a harvesting function, only a finite number of even-spaced sample points of \( d \) are known. In particular, for each \( \hat{s} \in \mathcal{S}, \{ (s_i, d(s_i, \hat{s})) \}_{i=0}^{m-1} \) are given by experiment, where \( s_i = \frac{i}{m-1} \). So \( m \times |\mathcal{S}| \) samples are given. In the case of noisy samples, \( \{ (s_i, d(s_i, \hat{s}) + Z_i) \}_{i=0}^{m-1} \) are given for each \( \hat{s} \in \mathcal{S} \), where \( Z_i \) is i.i.d. additive noise with mean zero and variance \( \sigma_z^2 \).

For a distortion function \( f \) and a set of distortion functions \( F, \text{rate-distortion} \) functions are defined as

\[
R_f(D) = \inf_{P_{\beta,S}:I(f(S,\hat{s})) \leq D} \mathbb{E}(f(S,\hat{s})),
\]

\[
R_F(D) = \min_{P_{\beta,S}:I(f(S,\hat{s})) \leq D} \forall f \in F \mathbb{E}(f(S,\hat{s})).
\]

**Distortion-rate functions** are defined as

\[
D_f(R) = \min_{P_{\beta,S}:I(f(S,\hat{s})) \leq R} \mathbb{E}[f(S,\hat{s})],
\]

\[
D_F(R) = \min_{P_{\beta,S}:I(f(S,\hat{s})) \leq R} \mathbb{E}[f(S,\hat{s})].
\]

Then, the sampling loss in distortion for noiseless samples is defined as

\[
\Delta_D(R) = \sup_{d \in \mathcal{F}^K} D_{f(d,\hat{d})(R) - D_d(R)}.
\]

For noisy samples, we can generalize the distortion loss to noisy samples, similarly to [8].

\[
\Delta_D(R) = \inf_{\hat{d}_m} \sup_{d \in \mathcal{F}^K} \mathbb{E}[|D_d(R) - D_{d_m}(R)|],
\]

where \( \hat{d}_m \) is the estimate of the distortion function. Then, we have the following distortion bound for noiseless samples.
Lemma 19 (Thm. 1 in [27]): If \( P_S(s) < c \) \( \forall s \in \mathcal{S} \) with some constant \( c \),
\[
\Delta_D(R) = O(m^{-\lambda}).
\]

We generalize to the noisy samples case as follows.

Lemma 20: If \( P_S(s) < c \) \( \forall s \in \mathcal{S} \) with some constant \( c \),
\[
\Delta_D(R) = O \left( m^{-\frac{\lambda+1}{2\lambda+3}} \right).
\]

Proof: Pick an arbitrary compression kernel \( P_{\tilde{S}|\mathcal{S}} \). Then, rate \( R = I(S; \tilde{S}) \) is also fixed. For given \( (d, \hat{d}_m^p) \), noting that the expectation is over the noise distribution,
\[
\begin{align*}
\mathbb{E}_{Z} \left[ \mathbb{E}_{S,\tilde{S}}[|d(s, \tilde{s}) - \hat{d}_m^p(s, \tilde{s})|] \right] \\
\leq \mathbb{E}_{Z} \left[ \mathbb{E}_{S,\tilde{S}}[|d(s, \tilde{s}) - \hat{d}_m^p(s, \tilde{s})|] \right] \\
= \sum_{\hat{s} \in \mathcal{S}} \int_{\mathcal{S}} P_S(s) P_{\tilde{S}|\mathcal{S}}(s|\hat{s}) \mathbb{E}_{Z}[|d(s, \tilde{s}) - \hat{d}_m^p(s, \tilde{s})|] ds.
\end{align*}
\]
(18) can be further bounded.
\[
\begin{align*}
\leq c \sum_{\hat{s} \in \mathcal{S}} \int_{\mathcal{S}} \mathbb{E}_{Z}[|d(s, \tilde{s}) - \hat{d}_m^p(s, \tilde{s})|] ds \\
\leq c' \sum_{\hat{s} \in \mathcal{S}} \int_{\mathcal{S}} m^{-\frac{\lambda+1}{2\lambda+3}} ds \\
= c'|\mathcal{S}| m^{-\frac{\lambda+1}{2\lambda+3}} = O \left( m^{-\frac{\lambda+1}{2\lambda+3}} \right),
\end{align*}
\]
where the last inequality follows from the local polynomial estimator in Lem. 10. By the same argument as in the proof of Thm. 4 we have
\[
\begin{align*}
\mathbb{E}_{Z} \left[ |D_d(R) - D_{\hat{d}_m}(R)| \right] \\
= O \left( m^{-\frac{\lambda+1}{2\lambda+3}} \right).
\end{align*}
\]
Since the bound does not depend on the choice of \( d(\cdot, \hat{s}) \in \Gamma^K \), infimum over estimators only further improves the loss of the local polynomial estimator,
\[
\Delta_D(R) = \inf_{d_m \in \Gamma^K} \sup_{d \in \Gamma^K} \mathbb{E} \left[ |D_d(R) - D_{\hat{d}_m}(R)| \right] = O \left( m^{-\frac{\lambda+1}{2\lambda+3}} \right).
\]

References

[1] L. R. Varshney, “Transporting information and energy simultaneously,” in Proc. 2008 IEEE Int. Symp. Inf. Theory, Jul. 2008, pp. 1612–1616.
[2] ———, “On energy/information cross-layer architectures,” in Proc. 2012 IEEE Int. Symp. Inf. Theory, Jul. 2012, pp. 1361–1365.
[3] R. Zhang, R. G. Maudner, and L. Hanzo, “Wireless information and power transfer: From scientific hypothesis to engineering practice,” IEEE Commun. Mag., vol. 53, no. 8, pp. 99–105, Aug. 2015.
[4] X. Zhou, R. Zhang, and C. K. Ho, “Wireless information and power transfer: Architecture design and rate-energy tradeoff,” IEEE Trans. Commun., vol. 61, no. 11, pp. 4754–4767, Nov. 2013.
[5] T. Soyata, L. Copeland, and W. Heinzelman, “RF energy harvesting for embedded systems: A view of tradeoffs and methodology,” IEEE Circuits Syst. Mag., vol. 16, no. 1, pp. 22–37, 2016.
[6] C. R. Valenta and G. D. Durgin, “Harvesting wireless power: Survey of energy-harvester conversion efficiency in far-field, wireless power transfer systems,” IEEE Microw. Mag., vol. 15, no. 4, pp. 108–120, Jun. 2014.
[7] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, “Practical non-linear energy harvesting model and resource allocation for SWIPT systems,” IEEE Commun. Lett., vol. 19, no. 12, pp. 2082–2085, Dec. 2015.
[8] J.-M. Kang, I.-M. Kim, and D. I. Kim, “Wireless information and power transfer: Rate-energy tradeoff for nonlinear energy harvesting,” IEEE Commun. Lett., vol. 17, no. 3, pp. 1966–1981, Mar. 2013.
[9] M. Varasteh, B. Rassouli, and B. Clerckx, “Wireless information and power transfer over an AWGN channel: Nonlinearity and asymmetric Gaussian signaling,” in Proc. IEEE Inf. Theory Workshop (ITW’17), Nov. 2017, pp. 181–185.
[10] T. Le, K. Mayaram, and T. Fiez, “Efficient far-field radio frequency energy harvesting for passively powered sensor networks,” IEEE J. Solid-State Circuits, vol. 43, no. 5, pp. 1287–1302, May 2008.
[11] M. Stoopman, S. Keyrouz, H. J. Visser, K. Philips, and W. A. Serdjin, “A self-calibrating RF energy harvester generating 1V at –20.3 dBm,” in 2013 Symp. VLSI Circuits Dig. Tech. Pap., Jun. 2013, pp. C226–C227.
[12] ———, “Co-design of a CMOS rectifier and small loop antenna for highly sensitive RF energy harvesters,” IEEE J. Solid-State Circuits, vol. 49, no. 3, pp. 622–634, Mar. 2014.
[13] A. P. Sample, A. N. Parks, S. Southwood, and J. R. Smith, “Wireless ambient radio power,” in Wirelessly Powered Sensor Networks and Computational RFID, J. R. Smith, Ed. New York: Springer, 2013, pp. 223–234.
[14] U. Baroudi, A. Qureshi, and S. Mekid, “Characterization and modeling of received signal strength and charging time for wireless energy transfer,” Adv. Electr. Eng., vol. 2015, 2015.
[15] P. Ntantavongsu, U. Muncuk, D. R. Lewis, and K. Roy Chowdhury, “Design optimization and implementation for RF energy harvesting circuits,” IEEE J. Emerg. Sel. Topics Circuits Syst., vol. 2, no. 1, pp. 24–33, Mar. 2012.
[16] M. Unser and I. Daubechies, “On the approximation power of convolution-based least squares versus interpolation,” IEEE Trans. Signal Process., vol. 45, no. 7, pp. 1697–1711, Jul. 1997.
[17] A. B. Tsybakov, Introduction to Nonparametric Estimation. New York: Springer-Verlag, 2009.
[18] A. M. Fouladgar and O. Simeone, “On the transfer of information and energy in multi-user systems,” IEEE Commun. Lett., vol. 16, no. 11, pp. 1733–1736, Nov. 2012.
[19] S. B. Amor, S. M. Perlaza, I. Krikidis, and H. V. Poor, “Feedback enhances simultaneous wireless information and energy transmission in multiple access channels,” IEEE Trans. Inf. Theory, vol. 65, no. 8, pp. 5244–5265, Aug. 2017.
[20] V. Talla, B. Kellogg, B. Ransford, S. Naderiparizi, J. R. Smith, and S. Gollakota, “Powering the next billion devices with Wi-Fi,” Commun. ACM, vol. 60, no. 3, pp. 83–91, Mar. 2017.
[21] U. Niesen, D. Shah, and G. Wornell, “Sampling distortion measures,” in Proc. 44th Annu. Allerton Conf. Commun. Control Comput., Sep. 2006.
[22] L. R. Varshney, “Unreliable and resource-constrained decoding,” Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, Jun. 2010.
[23] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: John Wiley & Sons, 1991.
[24] J. G. Smith, “The information capacity of amplitude- and variance-constrained scalar Gaussian channels,” Inf. Control, vol. 18, no. 3, pp. 203–219, Apr. 1971.
[25] M. Unser, “Sampling—50 years after Shannon,” Proc. IEEE, vol. 88, no. 4, pp. 569–587, Apr. 2000.
[26] M. Vetterli, J. Kovačević, and V. K. Goyal, Foundations of Signal Processing. Cambridge: Cambridge University Press, 2014.
[27] R. A. DeVore and G. G. Lorentz, Constructive Approximation. Berlin: Springer-Verlag, 1993.
[28] M. Unser, “Splines: A perfect fit for signal and image processing,” IEEE Signal Process. Mag., vol. 16, no. 6, pp. 22–38, 1999.
[29] C. de Boor, A Practical Guide to Splines. New York: Springer-Verlag, 1978.
[30] A. Tchamkerten, “On the discreteness of capacity-achieving distributions,” IEEE Trans. Inf. Theory, vol. 50, no. 11, pp. 2773–2778, Nov. 2004.
[31] A. ElMoslimy and T. M. Duman, “On the discreteness of capacity-achieving distributions for fading and signal-dependent noise channels with amplitude-limited inputs,” IEEE Trans. Inf. Theory, vol. 64, no. 2, pp. 1163–1177, Feb. 2018.
[33] S. N. Kudryavtsev, “Recovering a function with its derivatives from function values at a given number of points,” Russian Academy of Sciences Izvestiya Mathematics, vol. 45, no. 3, pp. 505–528, 1995.
[34] L. Györffy, M. Kohler, A. Krzyżak, and H. Walk, A Distribution-Free Theory of Nonparametric Regression. New York: Springer-Verlag, 2002.
[35] T.-Y. Wu, A. Tandon, L. R. Varshney, and M. Motani, “Multicasting energy and information simultaneously,” arXiv:1806.11271v1 [cs.IT], Jun. 2018.
[36] M. Gastpar, “To code or not to code,” Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Switzerland, Jan. 2003.
[37] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” in IRE Nat. Conv. Rec., Part 4, Mar. 1959, pp. 142–163.
[38] K. Toraichi, K. Katagishi, I. Sekita, and R. Mori, “Computational complexity of spline interpolation,” Int. J. Systems Sci., vol. 18, no. 5, pp. 945–954, 1987.