

# Transporting Information and Energy Simultaneously

Lav R. Varshney

Laboratory for Information and Decision Systems and Research Laboratory of Electronics  
Massachusetts Institute of Technology

**Abstract**—The fundamental tradeoff between the rates at which energy and reliable information can be transmitted over a single noisy line is studied. Engineering inspiration for this problem is provided by powerline communication, RFID systems, and covert packet timing systems as well as communication systems that scavenge received energy. A capacity-energy function is defined and a coding theorem is given. The capacity-energy function is a non-increasing concave  $\cap$  function. Capacity-energy functions for several channels are computed.

## I. INTRODUCTION

The problem of communication is usually cast as one of transmitting a message generated at one point to another point. During the pre-history of information theory, a primary accomplishment was the abstraction of the message to be communicated from the communication medium. As noted, “electricity in the wires became merely a carrier of messages, not a source of power, and hence opened the door to new ways of thinking about communications” [1]. This understanding of signals independently from their physical embodiments led to modern communication theory, but it also blocked other possible directions. As Norbert Wiener said, “Information is information, not matter or energy. No materialism which does not admit this can survive at the present day” [2, p. 132]. This separation of messages and media arguably led to the division of electrical engineering into two distinct subfields, electric power engineering and communication engineering.

Some have argued that the greatest inventions of civilization either transform, store, and transmit energy or they transform, store, and transmit information [3]. Although quite reasonable, many engineering systems actually deal with both energy and information. Representation of signals requires the modulation of energy, matter, or some such thing. The separation of messages and media is not always warranted.

Are there scenarios where one would want to transmit energy and information simultaneously over a single line? If there is a power-limited receiver that can harvest received energy, then one should want both things. The earliest telegraphs, telephones, and crystal radios had no external power sources [1], providing historical examples of such systems. Modern communication systems that operate under severe energy constraints may also benefit from harvesting received energy [4]. A powerful base station or other special node [5], [6], may effectively be used to recharge mobile devices. In RFID systems, the energy provided through the forward

channel is used to transmit over the backward channel [7]. There are also extant mudpulse telemetry systems in the oil industry where energy and information are provisioned to remote instruments over a single line [J. Kusuma, personal communication]. For a truly space-age application, one might consider furnishing photons to spacecraft with space sails [8] and optical receivers for both information and propulsion.

Back on earth, power line communication has received significant attention [9], [10], but the literature has focused on the informational aspect under the constraint that modulation schemes not severely degrade power delivery. This need not be the case in future engineering systems.

Except for papers on reversible computing [11], the fact that matter/energy must go along with information does not seem to have been considered in information theory. Similarly, the information carried in power transmission seems not to have been considered in power engineering. Though not implemented in current systems, a receiver constructed from reversible gates would allow received energy to perform additional work and would need not be dissipated as heat [11].

Electricity, of course, is not the only commodity in which signals can be modulated. Information can be physically manifested in almost any substance. Examples include water, railroad cars, and packets in communication networks (whose timing is modulated [12]); the results presented apply equally to these scenarios.

This work deals with the fundamental tradeoff between transmitting energy and transmitting information over a single noisy line. Although this tradeoff must be known to other researchers, it does not seem to appear in the literature. A characterization of communication systems that simultaneously meet two goals:

- 1) large received energy per unit time, and
- 2) large information per unit time

is found. Notice that unlike traditional transmitter power constraints, where small transmitted power is desired, here large received power is desired. One previous study has looked at maximum received power constraints [13].

## II. CAPACITY-ENERGY FUNCTION

In order to achieve the first goal, one would want the most energetic symbol received all the time, whereas to achieve the second goal, one would want to use the unconstrained capacity-achieving input distribution. This intuition is formalized for discrete memoryless channels, following [14].

Work supported in part by an NSF Graduate Research Fellowship.

A discrete memoryless channel (DMC) is characterized by the input alphabet  $\mathcal{X}$ , the output alphabet  $\mathcal{Y}$ , and the transition probability assignment  $Q_{Y|X}(y|x)$ . Furthermore, each output letter  $y \in \mathcal{Y}$  has an energy  $b(y)$ , a nonnegative real number. Channel inputs are described by random variables  $X_1^n = (X_1, X_2, \dots, X_n)$  with distribution  $p_{X_1^n}(x_1^n)$ ; the corresponding outputs are random variables  $Y_1^n = (Y_1, Y_2, \dots, Y_n)$  with distribution  $p_{Y_1^n}(y_1^n)$ . The average received energy is

$$E[b(Y_1^n)] = \sum_{y_1^n \in \mathcal{Y}^n} b(y_1^n) p(y_1^n).$$

An optimization problem that precisely captures the tradeoff between the two goals is as follows. Maximize information rate under a minimum received power constraint. For each  $n$ , the  $n$ th capacity-energy function  $C_n(B)$  of the channel is defined as

$$C_n(B) = \max_{X_1^n: E[b(Y_1^n)] \geq nB} I(X_1^n; Y_1^n).$$

An input vector  $X_1^n$  is a test source; one that satisfies  $E[b(Y_1^n)] \geq nB$  is  $B$ -admissible. The maximization is over all  $n$ -dimensional  $B$ -admissible test sources. The set of  $B$ -admissible  $p(x_1^n)$  is a closed subset of  $\mathbb{R}^{|\mathcal{X}|^n}$  and is bounded since  $\sum p(x_1^n) = 1$ . Since the set is closed and bounded, it is compact. Mutual information is a continuous function of the input distribution and since continuous, real-valued functions defined on compact subsets of metric spaces achieve their supremums (see Theorem 5), defining the optimization as a maximum is not problematic. The  $n$ th capacity-energy functions are only defined for  $0 \leq B \leq B_{\max}$ , where  $B_{\max}$  is the maximum element of  $b^T Q$ ;  $b$  is a column vector of the  $b(y)$  and  $Q$  is  $Q_{Y|X}$ .

The capacity-energy function of the channel is defined as

$$C(B) = \sup_n \frac{1}{n} C_n(B). \quad (1)$$

A coding theorem can be proven that endows this informational definition with operational significance.

A code is a pair of mappings  $(f, g)$  where  $f$  maps a message alphabet  $\mathcal{M}$  to  $\mathcal{X}$  and  $g$  maps  $\mathcal{Y}$  to  $\mathcal{M}$ . The rate of an  $n$ -length block code is  $\frac{1}{n} \log |\mathcal{M}|$ . An  $n$ -length block code with maximum probability of error bounded by  $\epsilon$  is an  $(n, \epsilon)$ -code.

*Definition 1:* Given  $0 \leq \epsilon < 1$ , a non-negative number  $R$  is an  $\epsilon$ -achievable rate for the channel  $Q_{Y|X}$  with constraint  $(b, B)$  if for every  $\delta > 0$  and every sufficiently large  $n$  there exist  $(n, \epsilon)$ -codes of rate exceeding  $R - \delta$  for which  $b(y_1^n) < B$  implies  $g(y_1^n) \notin \mathcal{M}$ .  $R$  is an achievable rate if it is  $\epsilon$ -achievable for all  $0 < \epsilon < 1$ . The supremum of achievable rates is called the *capacity* of the channel under constraint  $(b, B)$  and is denoted  $C_O(B)$ .

*Theorem 1:*  $C_O(B) = C(B)$ .

*Proof:* Follows by reversing the output constraint inequality in the solution to [15, P20 on p. 117]. See also [13]. ■

### III. PROPERTIES OF THE CAPACITY-ENERGY FUNCTION

The coding theorem provides operational significance to the capacity-energy function. Some properties of this function may also be developed.

It is immediate that  $C_n(B)$  is non-increasing, since the feasible set in the optimization becomes smaller as  $B$  increases. The function is also concave  $\cap$ .

*Theorem 2:*  $C_n(B)$  is a concave  $\cap$  function of  $B$  for  $0 \leq B \leq B_{\max}$ .

*Proof:* Let  $\alpha_1, \alpha_2 \geq 0$  with  $\alpha_1 + \alpha_2 = 1$ . The inequality to be proven is that that for  $B_1, B_2 \leq B_{\max}$ ,

$$C_n(\alpha_1 B_1 + \alpha_2 B_2) \geq \alpha_1 C_n(B_1) + \alpha_2 C_n(B_2).$$

Let  $X_1$  and  $X_2$  be  $n$ -dimensional test sources distributed according to  $p_1(x_1^n)$  and  $p_2(x_1^n)$  that achieve  $C_n(B_1)$  and  $C_n(B_2)$  respectively. Denote the corresponding channel outputs as  $Y_1$  and  $Y_2$ . It follows that  $E[b(Y_i)] \geq nB_i$  and  $I(X_i; Y_i) = C_n(B_i)$  for  $i = 1, 2$ . Define another source  $X$  distributed according to  $p(x_1^n) = \alpha_1 p_1(x_1^n) + \alpha_2 p_2(x_1^n)$  with corresponding output  $Y$ . Then

$$\begin{aligned} E[b(Y)] &= b^T Q p = b^T Q [\alpha_1 p_1 + \alpha_2 p_2] \\ &= \alpha_1 b^T Q p_1 + \alpha_2 b^T Q p_2 \\ &= \alpha_1 E[b(Y_1)] + \alpha_2 E[b(Y_2)] \\ &\geq n(\alpha_1 B_1 + \alpha_2 B_2), \end{aligned} \quad (2)$$

where  $b$  and  $Q$  have been suitably extended. Thus,  $X$  is  $(\alpha_1 B_1 + \alpha_2 B_2)$ -admissible. Now, by definition of  $C_n(\cdot)$ ,  $I(X; Y) \leq C_n(\alpha_1 B_1 + \alpha_2 B_2)$ . However, since  $I(X; Y)$  is a concave  $\cap$  function of the input probability,

$$\begin{aligned} I(X; Y) &\geq \alpha_1 I(X_1; Y_1) + \alpha_2 I(X_2; Y_2) \\ &= \alpha_1 C_n(B_1) + \alpha_2 C_n(B_2). \end{aligned}$$

Linking the two inequalities yields the desired result:

$$C_n(\alpha_1 B_1 + \alpha_2 B_2) \geq I(X; Y) \geq \alpha_1 C_n(B_1) + \alpha_2 C_n(B_2). \quad \blacksquare$$

It can also be shown that  $C_1(B) = C(B)$ .

*Theorem 3:* For any DMC,  $C_n(B) = nC_1(B)$  for all  $n = 1, 2, \dots$  and  $0 \leq nB \leq nB_{\max}$ .

*Proof:* Let  $X = (X_1, \dots, X_n)$  be a  $B$ -admissible test source with corresponding output  $Y$  that achieves  $C_n(B)$ , so  $E[b(Y)] \geq nB$  and  $I(X; Y) = C_n(B)$ . Since the channel is memoryless,  $I(X; Y) \leq \sum_{i=1}^n I(X_i; Y_i)$ . Let  $B_i = E[b(Y_i)]$ , then  $\sum_{i=1}^n B_i = \sum_{i=1}^n E[b(Y_i)] = E[b(Y)] \geq nB$ . By the definition of  $C_1(B_i)$ ,  $I(X_i; Y_i) \leq C_1(B_i)$ . Now since  $C_1(B)$  is a concave  $\cap$  function of  $B$ , by Jensen's inequality,

$$\frac{1}{n} \sum_{i=1}^n C_1(B_i) \leq C_1\left(\frac{1}{n} \sum_{i=1}^n B_i\right) = C_1\left(\frac{1}{n} E[b(Y)]\right).$$

But since  $\frac{1}{n} E[b(Y)] \geq B$  and  $C_1(B)$  is a non-increasing function of  $B$ ,

$$\frac{1}{n} \sum_{i=1}^n C_1(B_i) \leq C_1\left(\frac{1}{n} E[b(Y)]\right) \leq C_1(B),$$

that is,

$$\sum_{i=1}^n C_1(B_i) \leq nC_1(B).$$

Combining yields  $C_n(B) \leq nC_1(B)$ .

For the reverse, let  $X$  be a random variable with corresponding output  $Y$  that achieves  $C_1(B)$ . That is,  $E[b(Y)] \geq B$  and  $I(X; Y) = C_1(B)$ . Now let  $X_1, X_2, \dots, X_n$  be i.i.d. according to  $p(X)$  with outputs  $Y_1, \dots, Y_n$ . Then

$$E[b(Y_1^n)] = \sum_{i=1}^n E[b(Y_i)] \geq nB.$$

Moreover by memorylessness,

$$I(X_1^n; Y_1^n) = \sum_{i=1}^n I(X_i; Y_i) = nC_1(B).$$

Thus,  $C_n(B) \geq nC_1(B)$ . Since  $C_n(B) \geq nC_1(B)$  and  $C_n(B) \leq nC_1(B)$ ,  $C_n(B) = nC_1(B)$ . ■

The theorem implies that single-letterization is valid:  $C(B) = C_1(B)$ .

#### IV. THREE BINARY CHANNELS

Closed form expressions of the capacity-energy function for some particular channels may provide insight. Here, three binary channels with output alphabet energy function  $b(0) = 0$  and  $b(1) = 1$  are considered. Such an energy function corresponds to discrete particles and packets, among other commodities.

Consider a noiseless binary channel. The optimization problem is solved by the maximum entropy method, hence the capacity-achieving input distribution is in Gibbsian form. It is easy to show that the capacity-energy function is

$$C(B) = \begin{cases} \log(2), & 0 \leq B \leq \frac{1}{2} \\ h_2(B), & \frac{1}{2} \leq B \leq 1, \end{cases}$$

where  $h_2(\cdot)$  is the binary entropy function. The capacity-energy functions for other discrete noiseless channels are similarly easy to work out using maximum entropy methods.

Consider a binary symmetric channel with crossover probability  $\omega$ . It can be shown that the capacity-energy function is

$$C(B) = \begin{cases} \log(2) - h_2(\omega), & 0 \leq B \leq \frac{1}{2} \\ h_2(B) - h_2(\omega), & \frac{1}{2} \leq B \leq 1 - \omega. \end{cases}$$

Recall that for the unconstrained problem, equiprobable inputs are capacity-achieving, which yield output power  $\frac{1}{2}$ . For  $B > \frac{1}{2}$ , the distribution must be perturbed so that the symbol 1 is transmitted more frequently. The maximum power receivable through this channel is  $1 - \omega$ , when 1 is always transmitted.

A third worked example is the Z-channel; the unconstrained capacity expression and associated capacity-achieving input distribution [16] are used. Consider a Z-channel with 1 to 0 crossover probability  $\omega$ . The capacity-energy function is

$$C(B) = \begin{cases} \log\left(1 - \omega \frac{1}{1-\omega} + \omega \frac{\omega}{1-\omega}\right), & 0 \leq B \leq (1 - \omega)\pi^* \\ h_2(B) - \frac{B}{1-\omega} h_2(\omega), & (1 - \omega)\pi^* \leq B \leq 1 - \omega, \end{cases}$$

where

$$\pi^* = \frac{\omega \frac{\omega}{1-\omega}}{1 + (1 - \omega)\omega \frac{\omega}{1-\omega}}.$$

A Z-channel models quantal synaptic failure [17] and other “stochastic leaky pipes” where the commodity may be lost en route.

#### V. A GAUSSIAN CHANNEL

Attention now turns to discrete-time, continuous-alphabet, memoryless channels. The coding theorem (Theorem 1) can be extended in the usual way [18]. Continuous additive noise systems have the interesting property that for the goal of received power, noise power is actually helpful, whereas for the goal of information, noise power is hurtful. In discrete channels, such an interpretation is not obvious. When working with real-valued alphabets, some sort of transmitter constraint must be imposed so as to disallow arbitrarily powerful signals. Hard amplitude constraints that model rail limitations in power circuits are suitable. Assume that the channel transition pdf  $Q(y|x)$  exists.

Rather than working with the output energy constraint directly, it is convenient to think of the output energy function  $b(y)$  as inducing costs on the input alphabet  $\mathcal{X}$ :

$$\rho(x) = \int Q(y|x)b(y)dy.$$

By construction, this cost function preserves the constraint:

$$\begin{aligned} E[\rho(X)] &= \int \rho(x)dF(x) = \int dF(x) \int Q(y|x)b(y)dy \\ &= \int \int Q(y|x)b(y)dF(x)dy = E[b(Y)], \end{aligned}$$

where  $F(x)$  is the input distribution function. Basically,  $\rho(x)$  is the expected output energy provided by input letter  $x$ . Consider the AWGN channel  $\mathcal{N}(0, \sigma_N^2)$  and  $b(y) = y^2$ . Then

$$\rho(x) = \int_{-\infty}^{\infty} \frac{y^2}{\sigma_N \sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2\sigma_N^2}\right\} dy = x^2 + \sigma_N^2,$$

that is, the output power is just the sum of the input power and the noise power.

Since Theorem 3 extends directly to continuous alphabet channels,

$$C(B) = \sup_{X: E[\rho(X)] \geq B} I(X; Y). \quad (3)$$

Consider the AWGN channel,  $\mathcal{N}(0, \sigma_N^2)$ , with input alphabet  $\mathcal{X} = [-A, A] \subset \mathbb{R}$ , and energy function  $b(y) = y^2$ . Denote the capacity-energy function as  $C(B; A)$ . Following lockstep with Smith [19], [20], it is shown that the capacity-energy achieving input distribution consists of a finite number of mass points. Before proceeding, two optimization theorems are quoted [20]:

*Theorem 4:* Let  $\Omega$  be a convex metric space, and  $f$  and  $g$  concave  $\cap$  functionals on  $\Omega$  to  $\mathbb{R}$ ; assume there exists an  $x_1 \in \Omega$  such that  $g(x_1) < 0$  and let

$$D' \triangleq \sup_{x \in \Omega: g(x) \leq 0} f(x).$$

If  $D'$  is finite, then there exists a constant  $\lambda \geq 0$  such that

$$D' = \sup_{x \in \Omega} [f(x) - \lambda g(x)].$$

Moreover if the supremum in the first equation is achieved by  $x_0$  and  $g(x_0) \leq 0$ , then the supremum is achieved by  $x_0$  in the second equation and  $\lambda g(x_0) = 0$ .

*Theorem 5:* Let  $f$  be a continuous, weakly-differentiable, strictly concave  $\cap$  map from a compact, convex, topological space  $\Omega$  to  $\mathbb{R}$ . Define

$$D \triangleq \sup_{x \in \Omega} f(x).$$

Then the following two properties hold:

- 1)  $D = \max f(x) = f(x_0)$  for some unique  $x_0 \in \Omega$ , and
- 2) A necessary and sufficient condition for  $f(x_0) = D$  is for  $f'_{x_0}(x) \leq 0$  for all  $x \in \Omega$ , where  $f'_{x_0}$  is the weak derivative.

Let  $\mathcal{F}_A$  be the space of input probability distribution functions having all points of increase on the finite interval  $[-A, A]$ .

*Lemma 1 ([19]):*  $\mathcal{F}_A$  is convex and compact in the Levy metric.

Since the channel is fixed, mutual information can be written as a function of the input distribution,  $I(F)$ .

*Lemma 2 ([19]):* Mutual information  $I : \mathcal{F}_A \rightarrow \mathbb{R}$  is a concave  $\cap$ , continuous, weakly differentiable functional.

Let us denote the input squared value under input distribution  $F$  as

$$\sigma_F^2 \triangleq \int_{-A}^A x^2 dF(x).$$

Recall the energy constraint

$$B \leq E[\rho(X)] = E[x^2 + \sigma_N^2] = \sigma_N^2 + \sigma_F^2,$$

which is equivalent to  $B - \sigma_N^2 - \sigma_F^2 \leq 0$ . Now define the functional  $J : \mathcal{F}_A \rightarrow \mathbb{R}$  as

$$J(F) \triangleq B - \sigma_N^2 - \int_{-A}^A x^2 dF(x).$$

*Lemma 3:*  $J$  is a concave  $\cap$ , continuous, weakly differentiable functional.

*Proof:* Clearly  $J$  is linear in  $F$  (see (2) for basic argument). Moreover,  $J$  is bounded as  $B - \sigma_N^2 - A^2 \leq J \leq B - \sigma_N^2$ . Since  $J$  is linear and bounded, it is concave  $\cap$ , continuous, and weakly differentiable. ■

Returning to the optimization problem to be solved,

*Theorem 6:* There exists a constant  $\lambda \geq 0$  such that

$$C(B; A) = \sup_{F \in \mathcal{F}_A} [I(F) - \lambda J(F)].$$

*Proof:* The result follows from Theorem 4 since  $I$  is a concave  $\cap$  functional (Lemma 2),  $J$  is a concave  $\cap$  functional (Lemma 3), since capacity is finite whenever  $A < \infty$  and  $\sigma_N^2 > 0$ , and since there is obviously an  $F_1 \in \mathcal{F}_A$  such that  $J(F_1) < 0$ . ■

*Theorem 7:* There exists a unique capacity-energy achieving input  $X_0$  with distribution function  $F_0$  such that

$$C(B; A) = \max_{F \in \mathcal{F}_A} [I(F) - \lambda J(F)] = I(F_0) - \lambda J(F_0).$$

Moreover, a necessary and sufficient condition for  $F_0$  to achieve capacity-energy is

$$I'_{F_0}(F) - \lambda J'_{F_0}(F) \leq 0 \text{ for all } F \in \mathcal{F}_A. \quad (4)$$

*Proof:* Since  $I$  and  $J$  are both concave  $\cap$ , continuous, and weakly differentiable (Lemmas 2, 3), so is  $I - \lambda J$ . Since  $\mathcal{F}_A$  is a convex, compact space (Lemma 1), Theorem 5 applies and yields the result. ■

For our function  $I - \lambda J$ , the optimality condition (4) is,

$$\int_{-A}^A [i(x; F_0) + \lambda x^2] dF(x) \leq I(F_0) + \lambda \int_{-A}^A x^2 dF_0(x),$$

for all  $F \in \mathcal{F}_A$ , where  $i$  is

$$i(x; F) = \int Q(y|x) \log \frac{Q(y|x)}{p(y; F)} dy$$

and is variously known as the marginal information density [20], the Bayesian surprise [21], or without name [22, Eq. 1]. This follows since the mutual information weak derivative is

$$I'_{F_1}(F_2) = \int_{-A}^A i(x; F_1) dF_2(x) - I(F_1)$$

and the energy weak derivative is  $J'_{F_1}(F_2) = J(F_2) - J(F_1)$ . If  $\int x^2 dF_0(x) > B - \sigma_N^2$ , then the moment constraint is trivial and the constant  $\lambda$  is zero, thus the optimality condition can be written as

$$\int_{-A}^A [i(x; F_0) + \lambda x^2] dF(x) \leq I(F_0) + \lambda [B - \sigma_N^2]. \quad (5)$$

The optimality conditions (5) may be rejiggered to a condition on the input alphabet.

*Theorem 8:* Let  $F_0$  be an arbitrary distribution function in  $\mathcal{F}_A$  satisfying the energy constraint. Let  $E_0$  denote the points of increase of  $F_0$  on  $[-A, A]$ . Then  $F_0$  is optimal if and only if, for some  $\lambda \geq 0$ ,

$$i(x; F_0) \leq I(F_0) + \lambda [B - \sigma_N^2 - x^2] \text{ for all } x \in [-A, A],$$

$$i(x; F_0) = I(F_0) + \lambda [B - \sigma_N^2 - x^2] \text{ for all } x \in E_0.$$

*Proof:* If both conditions hold for some  $\lambda \geq 0$ ,  $F_0$  must be optimal and  $\lambda$  is the one from Theorem 7. This is because integrating both sides of the conditions by an arbitrary  $F$  yield satisfaction of condition (4).

For the converse, assume that  $F_0$  is optimal but that the inequality condition is not satisfied. Then there is some  $x_1 \in [-A, A]$  and some  $\lambda \geq 0$  such that  $i(x_1; F_0) > I(F_0) + \lambda [B - \sigma_N^2 - x_1^2]$ . Let  $F_1(x)$  be the unit step  $\mathbf{1}(x - x_1) \in \mathcal{F}_A$ ; but then

$$\begin{aligned} \int_{-A}^A [i(x; F_0) + \lambda x^2] dF_1(x) &= i(x_1; F_0) + \lambda x_1^2 \\ &> I(F_0) + \lambda [B - \sigma_N^2]. \end{aligned}$$

This violates (5), thus the inequality condition must be valid with  $\lambda$  from Theorem 7.

Now assume that  $F_0$  is optimal but that the equality condition is not satisfied, i.e. there is set  $E' \subset E_0$  such that the following is true.

$$\int_{E'} dF_0(x) = \delta > 0 \text{ and } \int_{E_0 - E'} dF_0(x) = 1 - \delta,$$

$$i(x; F_0) + \lambda x^2 < I(F_0) + \lambda[B - \sigma_N^2] \text{ for all } x \in E',$$

$$i(x; F_0) + \lambda x^2 = I(F_0) + \lambda[B - \sigma_N^2] \text{ for all } x \in E_0 - E'.$$

Then,

$$\begin{aligned} 0 &= \int [i(x; F_0) + \lambda x^2] dF_0(x) - I(F_0) - \lambda[B - \sigma_N^2] \\ &= \int_{E_0} [i(x; F_0) + \lambda x^2] dF_0(x) - I(F_0) - \lambda[B - \sigma_N^2] \\ &< \delta[I(F_0) + \lambda(B - \sigma_N^2)] + (1 - \delta)[I(F_0) + \lambda(B - \sigma_N^2)] \\ &\quad - I(F_0) - \lambda[B - \sigma_N^2] = 0, \end{aligned}$$

a contradiction. Thus the equality condition must be valid. ■

At a point like this, one might try to develop measure-matching conditions like Gastpar et al. [22] for undetermined  $b(\cdot)$ , but this path is not pursued here. To show that the input distribution is supported on a finite number of mass points requires Smith's *reductio ab absurdum* argument (see [23] for a slight correction).

*Theorem 9:*  $E_0$  is a finite set of points.

The proof uses optimality conditions from Theorem 8 to derive a contradiction using the analytic extension property of the marginal entropy density  $h(x; F)$ ,

$$h(x; F) = - \int Q(y|x) \log p(y; F) dy.$$

Since the capacity-energy achieving input distribution is a pmf, a finite numerical optimization algorithm may be used [24]. Consider the AWGN channel  $\mathcal{N}(0, 1)$  and find the point  $C(B = 0; A = 1.5)$ . The capacity achieving input density is  $p(x) = \frac{1}{2}\delta(x + 1.5) + \frac{1}{2}\delta(x - 1.5)$ . The achieved rate is

$$C(0; 1.5) = \int_{-1}^{+1} \frac{2/3}{\sqrt{2\pi(1-y^2)}} e^{-\frac{(1-(4/9)\tanh^{-1}(y))^2}{8/9}} \log(1+y) dy.$$

The achieved output power is  $E[Y^2] = 3.25$ . In fact, this is the maximum output power possible over this channel, since  $E[Y^2] = E[X^2] + \sigma_N^2$ , and  $E[X^2]$  cannot be improved over operating at the edges  $\{-A, A\}$ . Thus,

$$C(B; 1.5) = C(0; 1.5), \quad 0 \leq B \leq B_{\max} = 3.25.$$

For this particular channel, there actually is no tradeoff between information and power: antipodal signaling should be used all the time. This is not a general phenomenon, however. This is not true for the same noise, but for say  $A \geq 1.7$  rather than  $A = 1.5$  [19].

## VI. CONCLUSION

Information is patterned matter-energy. In this work, the fundamental tradeoff between the rate of transporting a commodity between one point and another and the rate of simultaneously transmitting information by modulating in that commodity has been studied. As extensions, one might consider a "wideband regime" formulation [25], a multiterminal problem where different users have different energy and information requirements, or a deeper look into reversible decoding [11].

## ACKNOWLEDGEMENT

Comments from V. K. Goyal & S. K. Mitter are appreciated.

## REFERENCES

- [1] D. A. Mindell, "Opening Black's box: Rethinking feedback's myth of origin," *Technol. Cult.*, vol. 41, pp. 405–434, July 2000.
- [2] N. Wiener, *Cybernetics*. MIT Press, 1961.
- [3] J. Aczél and Z. Daróczy, *On Measures of Information and Their Characterization*. Academic Press, 1975.
- [4] J. A. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," *IEEE Pervasive Comput.*, vol. 4, pp. 18–27, 2005.
- [5] L. R. Varshney and S. D. Servetto, "A distributed transmitter for the sensor reachback problem based on radar signals," in *Advances in Pervasive Computing and Networking*, B. K. Szymanski and B. Yener, Eds. Kluwer, 2005, pp. 225–245.
- [6] B. Ananthasubramaniam and U. Madhoo, "On localization performance in imaging sensor nets," *IEEE Trans. Signal Process.*, vol. 55, pp. 5044–5057, Oct. 2007.
- [7] R. Want, "Enabling ubiquitous sensing with RFID," *IEEE Computer*, vol. 37, pp. 84–86, Apr. 2004.
- [8] C. H. M. Jenkins, *Recent Advances in Gossamer Spacecraft*. American Institute of Aeronautics and Astronautics, Inc., 2006.
- [9] K. Dostert, *Powerline Communications*. Prentice Hall, 2001.
- [10] *Special Issue on Power Line Communications*, *IEEE J. Sel. Areas Commun.*, vol. 24, July 2006.
- [11] R. Landauer, "Computation, measurement, communication and energy dissipation," in *Selected Topics in Signal Processing*, S. Haykin, Ed. Prentice Hall, 1989, pp. 18–47.
- [12] J. Giles and B. Hajek, "An information-theoretic and game-theoretic study of timing channels," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2455–2477, Sept. 2002.
- [13] M. Gastpar, "On capacity under receive and spatial spectrum-sharing constraints," *IEEE Trans. Inf. Theory*, vol. 53, pp. 471–487, Feb. 2007.
- [14] R. J. McEliece, *The Theory of Information and Coding*. Addison-Wesley, 1977.
- [15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 3rd ed. Akadémiai Kiadó, 1997.
- [16] S. W. Golomb, "The limiting behavior of the Z-channel," *IEEE Trans. Inf. Theory*, vol. IT-26, p. 372, May 1980.
- [17] W. B. Levy and R. A. Baxter, "Using energy efficiency to make sense out of neural information processing," in *Proc. ISIT*, 2002, p. 18.
- [18] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.
- [19] J. G. Smith, "On the information capacity of peak and average power constrained Gaussian channels," Ph.D. dissertation, Univ. California, Berkeley, 1969.
- [20] —, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, pp. 203–219, Apr. 1971.
- [21] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," in *Proc. NIPS*, 2005, pp. 547–554.
- [22] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1147–1158, May 2003.
- [23] A. Tchamkerten, "On the discreteness of capacity-achieving distributions," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2773–2778, Nov. 2004.
- [24] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inf. Theory*, vol. 5, pp. 2336–2351, July 2005.
- [25] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1319–1343, June 2002.