On Fidelity Per Unit Cost

Marius Kleiner, Bixio Rimoldi

School of Computer and Communication Sciences

Ecole Polytechnique Fédérale de Lausanne

CH-1015 Lausanne, Switzerland

E-mail: firstname.lastname@epfl.ch

Abstract— We consider source coding with a fidelity criterion, channel coding with a channel input constraint, and the combined problem of reproducing a source across a noisy channel. All three cases face a similar tradeoff between resource and performance, and the operating point with the highest performance per resource is of particular interest. In the case of channel coding, channel input cost is traded for rate, and the optimal tradeoff corresponds to the capacity per unit cost. We define equivalent notions for the other two cases and show how they relate. For each case we give necessary and sufficient conditions for the optimal tradeoff to be achieved.

I. INTRODUCTION

In this paper we consider three standard communication problems, namely source coding with a fidelity criterion, channel coding with a channel input constraint, and the combined problem of reproducing a source across a noisy channel. Our focus is on discrete memoryless sources and channels, but the essence of our results extends to continous alphabets.

We begin by addressing the channel coding problem in Section II. We consider the capacity per unit cost, a concept that is traceable in various forms as far back as Reza [1] (see also references in [2]) and was more recently studied by Gallager ([3], [4, Ch. 14]) and developed by Verdú in his influential paper [2] (see also [5]). The capacity per unit cost for a channel with transition probabilities $P_{Y|X}$ and input cost measure $\rho(x)$ is defined as

$$\hat{C} = \sup_{P_X} \frac{I(X;Y)}{\mathbb{E}[\rho(X)]},$$

where the supremum is taken over all possible input distributions P_X. Usually $P_{Y|X}$ and $\rho(x)$ are given, and to compute the capacity per unit cost one maximizes over the space of input distributions. For the case when a channel input symbol of zero cost exists, however, Verdú showed that the capacity per unit cost can be obtained by a simpler maximization over the channel input alphabet.

We look at the problem from a different angle. We show that for fixed P_X and $P_{Y|X}$ there is a simple expression for the cost measure $\rho(x)$ for which P_X achieves the capacity per unit cost. This criterion is applicable whether or not a zero-cost symbol exists; it follows from refining a similar criterion that Gastpar et al. [6] developed for the purpose of achieving capacity rather than capacity per unit cost.

We illustrate our result with a detailed example of a Gaussian channel with Gaussian input. First we find that the cost measure for which the system operates at capacity per unit cost

is of the form $\rho(x) = x^2 + k$ for some positive constant k (to be specified). We show that a discrete-time channel with this cost measure relates to a continuous-time channel for which power can be traded against bandwidth. Furthermore, the tradeoff that maximizes the rate (in bits/second) of the continuous-time channel corresponds to the operating point at which the discrete-time channel achieves the capacity per unit cost.

For the rate-distortion problem a similar result exists, which we explore in Section III. For reasons that will become apparent, it is more convenient to study this result in terms of a fidelity measure, defined as the negative of the distortion measure. Following this line of thought we define the *fidelity per unit rate* of a source: it is to source coding what the capacity per unit cost is to channel coding. Again refining a result by Gaspar et al., we give a condition for a test channel[1] to achieve the fidelity per unit rate of the source.

In Section IV we consider the end-to-end problem of reproducing a source across a noisy channel. Drawing on the results of Sections II and III, we show that the capacity per unit cost and the fidelity per unit rate can be combined to give the *fidelity per unit cost*, *i.e.*, the maximum ratio of fidelity per cost at which a source-channel coding system can operate. Lastly we provide necessary and sufficient conditions under which a given system achieves the fidelity per unit cost.

II. ACHIEVING CAPACITY PER UNIT COST

Let $P_{Y|X}$ be the transition distribution of a discrete memoryless channel and let P_X and Q_X be two arbitrary but fixed input distributions. Let P_Y and Q_Y be the corresponding output distributions, *i.e.*,

$$P_Y(y) = \sum_x P_X(x) P_{Y|X}(y|x) \quad \text{and}$$

$$Q_Y(y) = \sum_x Q_X(x) P_{Y|X}(y|x).$$

Define $\rho_0(x)$ through P_X as

$$\rho_0(x) = D(P_{Y|X}(\cdot|x)||P_Y(\cdot)). \tag{1}$$

Let $\mathbb{E}_P$ and $\mathbb{E}_Q$ denote expectations with respect to P_X and Q_X, respectively, and let $I_P(X;Y)$ and $I_Q(X;Y)$ denote the mutual informations corresponding to these input distributions.

---

[1] In rate-distortion theory, a test channel is a conditional distribution of the reconstruction given a source symbol.

Proposition 1: We have

$$\frac{I_Q(X;Y)}{\mathbb{E}_Q[\rho_0(X)]} \le \frac{I_P(X;Y)}{\mathbb{E}_P[\rho_0(X)]} = 1, \qquad (2)$$

with equality if and only if $Q_X = P_X$.

Proposition 1 implies that if the channel input cost function is as in (1) (up to scaling), then P_X achieves the capacity per unit cost.

Proof: First note that $\mathbb{E}_P[\rho_0(X)] = I_P(X;Y)$, *i.e.*, $I_P(X;Y)/\mathbb{E}_P[\rho_0(X)] = 1$. On the other hand, if the input distribution is Q_X then

$$\mathbb{E}_Q[\rho_0(X)] - I_Q(X;Y)$$
$$= \sum_{x,y} Q_X(x)P_{Y|X}(y|x)\left[\log\frac{P_{Y|X}(y|x)}{P_Y(y)} - \log\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right]$$
$$= D(Q_Y\|P_Y) \ge 0.$$

It follows that

$$\frac{I_Q(X;Y)}{\mathbb{E}_Q[\rho_0(X)]} \le 1$$

with equality if and only if $P_Y = Q_Y$, or $P_X = Q_X$. ∎

Remark 1: The proposition also holds for channels with continuous alphabets. This follows directly if the sums in the proof are replaced by integrals.

Example 1: Consider a binary symmetric channel (BSC) with crossover probability $\epsilon = 0.1$. Let the input distribution be such that $P_X(0) = 0.7$ and $P_X(1) = 0.3$. Evaluating $\rho_0(x)$ for this situation yields

$$\rho_0(0) \approx 0.23 \quad \text{and} \quad \rho_0(1) \approx 0.99.$$

Example 2: Consider the AWGN channel $Y = X + Z$, where $X \sim \mathcal{N}(0,\sigma_X^2)$ and $Z \sim \mathcal{N}(0,1)$. Evaluating $\rho_0(x)$, we obtain

$$\rho_0(x) = D(P_{Y|X}(\cdot|x)\|P_Y(\cdot)) = ax^2 + b,$$

where

$$a = \frac{1}{2\ln 2(1+\sigma_X^2)}$$

and

$$b = \frac{1}{2}\left(\log_2(1+\sigma_X^2) - \frac{1}{\ln 2} + \frac{1}{\ln 2(1+\sigma_X^2)}\right).$$

The corresponding capacity-cost function is plotted on Figure 1.

When does a cost measure of the form $\rho(x) = x^2 + b$ make sense? The following example gives an answer to this question.

Example 3: Consider the problem of designing a communication system that maximizes the communication rate across a continuous-time Gaussian channel. We are free to trade bandwidth for power, provided that $P + kB \le A$ where B is the (two-sided) bandwidth, P is the power, and k and A are positive constants. Using the sampling theorem at intervals of length T, $BT = 1$, we translate the above problem into its discrete-time equivalent. To do so we first rewrite the constraint as $\frac{1}{T}\int_0^T \mathbb{E}[X^2(t)]dt \le A - kB$ and use the fact that
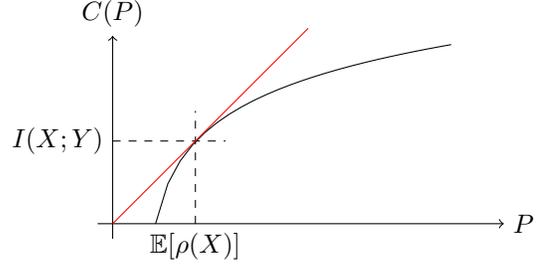

Fig. 1. Capacity-cost function for a Gaussian channel when the cost function is $\rho(x) = cD(P_{Y|X}(\cdot|x)\|P_Y(\cdot))$. Geometrically, the capacity per unit cost is achieved if a tangent through the origin touches the curve $C(P)$ at the operating point.

the discrete-time equivalent of the integral is $\mathbb{E}[X^2]$. Hence the discrete-time constraint becomes $\frac{1}{T}\mathbb{E}[X^2] \le A - \frac{k}{T}$ or, equivalently, $\mathbb{E}[X^2 + k] \le AT$.

With this constraint, the maximum rate at which we can transmit reliably (in bits/second) is

$$\frac{C(AT)}{T} = A\frac{C(AT)}{AT} \le A\hat{C},$$

where $\hat{C}$ is the capacity per unit cost of the discrete-time channel. Notice that we can choose to operate at capacity per unit cost by choosing T such that $\frac{C(AT)}{AT}$ achieves the maximum. This implies the optimal power–bandwidth tradeoff.

In the next proposition we show that under a minor technical condition, the cost function for which an input distribution achieves capacity per unit cost is unique (up to a scaling factor).

Proposition 2: If P_X achieves the capacity per unit cost of the channel $P_{Y|X}$ with cost function $\rho(x)$ and if the derivative of the corresponding capacity-cost function $C(P)$ exists at $P = \mathbb{E}_P[\rho(X)]$, then $\rho(x) = c\rho_0(x)$ for some $c > 0$.

Proof: Assume that P_X achieves the capacity per unit cost of the channel for some cost function $\rho(x)$. This implies that P_X also achieves the "regular" capacity of the channel at average cost $P^* = \mathbb{E}_P[\rho(X)]$, *i.e.*,

$$I_P(X;Y) = C(P^*).$$

As was shown in [6], a necessary condition for this is that $\rho(x) = c\rho_0(x) + \beta$ for some β. We therefore only have to show that if $\beta \ne 0$ then P_X does not achieve the capacity per unit cost.

Assume thus that $\rho(x) = c\rho_0(x) + \beta$, and let $C_0(P)$ and $C_\beta(P)$ be the capacity-cost functions corresponding to $\beta = 0$ and $\beta \ne 0$, respectively. $C_0(P)$ and $C_\beta(P)$ are related by

$$C_0(P) = \max_{P_X:\mathbb{E}[c\rho_0(X)]\le P} I(X;Y)$$
$$= \max_{P_X:\mathbb{E}[c\rho_0(X)+\beta]\le P+\beta} I(X;Y)$$
$$= C_\beta(P+\beta). \qquad (3)$$

If $\beta = 0$ then P_X achieves the capacity per unit cost (cf. Proposition 1) and we have

$$\frac{d}{dP}\left.\frac{C_0(P)}{P}\right|_{P=P^*} = 0,$$

where $P^* = \mathbb{E}_P[\rho(X)]$. This is equivalent to

$$C_0{}'(P^*) = C_0(P^*)/P^*, \tag{4}$$

which says nothing else than the tangent of $C_0(P)$ at P^* goes through the origin, as shown in Figure 2.

If $\beta \neq 0$ then the average cost under P_X is $P^* + \beta$. Then,

$$\begin{aligned}
C_\beta{}'(P^* + \beta) &\stackrel{(a)}{=} C_0{}'(P^*)\\
&\stackrel{(b)}{=} \frac{C_0(P^*)}{P^*}\\
&\stackrel{(c)}{=} \frac{C_\beta(P^* + \beta)}{P^*}\\
&\neq \frac{C_\beta(P^* + \beta)}{P^* + \beta},
\end{aligned}$$

where (a) and (c) follow from (3) and (b) follows from (4). This means that the tangent of $C_\beta(P)$ at $P = P^* + \beta$ does not go through the origin, and so according to (4) we are not at capacity per unit cost. See Figure 2 for a geometric interpretation of this proof. ∎
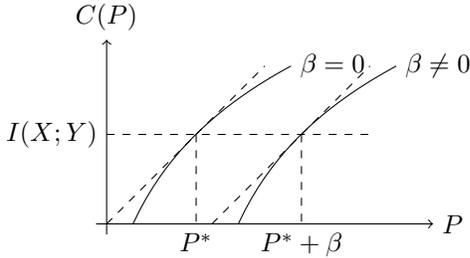

Fig. 2. A geometric interpretation of the proof of Proposition 2. A nonzero β shifts the capacity-cost curve such that the tangent at the operating point no longer goes through the origin.

Propositions 1 and 2 are summarized in the following theorem.

Theorem 3: Under the technical condition of Proposition 2, the channel input distribution P_X achieves the capacity per unit cost of the DMC $P_{Y|X}$ if and only if the cost function satisfies

$$\rho(x) = cD(P_{Y|X}(\cdot|x)\|P_Y(\cdot)), \quad c > 0. \tag{5}$$

For continuous alphabets, the condition is sufficient but not necessary.

Remark 2: This result is very much related to the following result [7, Section 2.3, Problem 2], [6]. The channel input distribution P_X achieves the capacity of the channel $P_{Y|X}$ at expected cost $\mathbb{E}[\rho(X)]$ if and only if

$$\rho(x) = c(D(P_{Y|X}(\cdot|x)\|P_Y(\cdot)) + \beta)$$

for some constants $c > 0$ and $\beta \in \mathbb{R}$. Our result shows that the extra requirement that capacity per unit cost be achieved restricts the possible cost functions to those with $\beta = 0$.

III. A Dual Result for the Source Coding Problem

Consider a discrete memoryless source (DMS) S with distribution P_S. Let $V(\hat{s}|s)$ and $W(\hat{s}|s)$ be two test channels for the given source. Let $P_{\hat{S}}$ and $Q_{\hat{S}}$ be the corresponding unconditional distributions of $\hat{S}$, *i.e.*,

$$P_{\hat{S}}(\hat{s}) = \sum_s P_S(s)V(\hat{s}|s) \quad \text{and}$$

$$Q_{\hat{S}}(\hat{s}) = \sum_s P_S(s)W(\hat{s}|s).$$

Let d_0 be defined through $V(\hat{s}|s)$ as follows:

$$d_0(s, \hat{s}) = -\log_2 \frac{V(\hat{s}|s)}{P_{\hat{S}}(\hat{s})} + \gamma(s), \tag{6}$$

where $\gamma(s)$ is an arbitrary function of s satisfying $\mathbb{E}[\gamma(S)] = 0$. Let $\mathbb{E}_V$ and $\mathbb{E}_W$ denote expectations over S and $\hat{S}$ with respect to V and W, respectively, and let $I_V(S; \hat{S})$ and $I_W(S; \hat{S})$ be the corresponding mutual informations.

Proposition 4:

$$\frac{\mathbb{E}_W[d_0(S, \hat{S})]}{I_W(S; \hat{S})} \geq \frac{\mathbb{E}_V[d_0(S, \hat{S})]}{I_V(S; \hat{S})} = -1, \tag{7}$$

with equality if and only if $V(\hat{s}|s) = W(\hat{s}|s)$ for all s and $\hat{s}$.

Proposition 4 implies that when the distortion measure is as in (6) then $V(\hat{s}|s)$ is the test channel that minimizes the ratio of distortion per rate.

Proof: First note that $\mathbb{E}_V[d_0(S, \hat{S})] = -I_V(S; \hat{S})$, *i.e.*, $\mathbb{E}_V[d_0(S, \hat{S})]/I_V(S; \hat{S}) = -1$. Next,

$$\begin{aligned}
&I_W(S; \hat{S}) + \mathbb{E}_W[d_0(S, \hat{S})]\\
&= \sum_{s,\hat{s}} P_S(s)W(\hat{s}|s)\left[\log_2 \frac{W(\hat{s}|s)}{Q_{\hat{S}}(\hat{s})} - \log_2 \frac{V(\hat{s}|s)}{P_{\hat{S}}(\hat{s})}\right]\\
&= \sum_{s,\hat{s}} P_S(s)W(\hat{s}|s)\left[\log_2 \frac{W(\hat{s}|s)}{V(\hat{s}|s)} - \log_2 \frac{Q_{\hat{S}}(\hat{s})}{P_{\hat{S}}(\hat{s})}\right]\\
&= \sum_s P_S(s)D(W(\cdot|s)\|V(\cdot|s)) - D(Q_{\hat{S}}(\hat{s})\|P_{\hat{S}}(\hat{s}))\\
&\geq 0,
\end{aligned}$$

where the inequality follows from the convexity of the Kullback-Leibler divergence (see *e.g.* [8, Thm 2.7.2]); it becomes an equality iff $V(\hat{s}|s) = W(\hat{s}|s)$. Rearranging the inequality, we obtain

$$\frac{\mathbb{E}_W[d_0(S, \hat{S})]}{I_W(S; \hat{S})} \geq -1.$$

∎

Remark 3: This proposition holds also for continuous alphabets; just replace the sums in the proof by integrals.

The distortion as defined in (6) may take positive and negative values. In fact, Proposition 4 shows that the smallest achievable distortion per rate is -1. The reader may find this awkward; indeed it is hard to imagine an application where one would define a distortion this way. The issue disappears if we define a *fidelity measure* $\phi_0(s, \hat{s}) = -d_0(s, \hat{s})$ and ask for the

largest possible $\mathbb{E}_W[\phi_0(S, \hat{S})]/I_W(S; \hat{S})$, which will of course be 1. Hence, for the rest of the paper we will work with the the notion of fidelity, defined as the negative of the distortion.[2]

For later reference we restate Proposition 4 in terms of fidelity:

Proposition 5: If $\phi_0(s, \hat{s}) = -d_0(s, \hat{s})$, then

$$\frac{\mathbb{E}_W[\phi_0(S, \hat{S})]}{I_W(S; \hat{S})} \leq \frac{\mathbb{E}_V[\phi_0(S, \hat{S})]}{I_V(S; \hat{S})} = 1, \qquad (8)$$

with equality if and only if $V(\hat{s}|s) = W(\hat{s}|s)$ for all s and $\hat{s}$.

Hence, if the fidelity measure is $\phi_0(s, \hat{s})$ then $V(\hat{s}|s)$ is the test channel that maximizes the *fidelity per rate* of the source.

Analog to the rate-distortion function we can define

$$R(\Phi) = \min_{p(\hat{s}|s):\mathbb{E}[\phi(S, \hat{S})] \geq \Phi} I(S; \hat{S})$$

as the *rate-fidelity function* of the source at average fidelity Φ. The rate-fidelity function has the same operational meaning as the rate-distortion function: for each rate larger than $R(\Phi)$ there exists a source code that achieves fidelity Φ, and conversely no such code exists with a rate smaller than $R(\Phi)$.

Example 4 (Gaussian): Consider a memoryless Gaussian source S with zero mean and unit variance, and a test channel $V(\hat{s}|s)$ such that $\hat{S}$ given $S = s$ is distributed as[3]

$$\mathcal{N}\left(\frac{\alpha}{\alpha+1}s, \frac{\alpha}{(\alpha+1)^2}\right)$$

for some $\alpha > 0$. Then $\hat{S}$ is distributed as $\mathcal{N}(0, \alpha/(\alpha+1))$, and $\phi_0(s, \hat{s})$ evaluates to

$$\log_2 \frac{V(\hat{s}|s)}{P_{\hat{S}}(\hat{s})} = c\left(k_1 + k_2 s^2 - (s - \hat{s})^2\right)$$

for some postive constants c, k_1, and k_2.

The rate-fidelity function corresponding to $\phi_0(s, \hat{s})$ is

$$R(\Phi) = \frac{1}{2} \log_2 \frac{\sigma_S^2}{k_1 + k_2 \sigma_S^2 - \Phi/c}.$$

The plot of Φ vs. $R(\Phi)$ on Figure 3 bears a strong resemblance to Figure 1: a tangent of slope 1 touches the curve in the operating point corresponding to P_S and $V(\hat{s}|s)$.

Does the fidelity obtained in the above example make sense? The multiplicative constant c is unavoidable: it accounts for the fact that one is free to choose the units. The term $k_2 s^2 - (s - \hat{s})^2$ indicates that the tolerance for the error $(s - \hat{s})^2$ is relative to s^2. In other words, small errors when s^2 is small are considered at the same level of satisfaction as large errors when s^2 is large, provided that $k_2 s^2 - (s - \hat{s})^2$ is constant. Finally, the constant k_1 accounts for the fact that the degree of satisfaction at the same value of $k_2 s^2 - (s - \hat{s})^2$ may vary from one application to another: a radiologist looking at an x-ray will have a smaller error tolerance than a person listening to music in a noisy environment.

[2]Of course mathematically the two notions are quivalent.
[3]We use $\mathcal{N}(\mu, \sigma^2)$ to denote a Gaussian distribution of mean μ and variance σ^2.
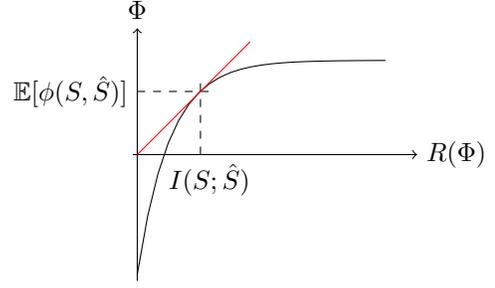

Fig. 3. Average fidelity Φ vs. minimum rate $R(\Phi)$ for Example 4. The similarity to Figure 1 illustrates that the fidelity per cost is to a source what the capacity per cost is to a channel.

The next proposition says that the fidelity measure $\phi(s, \hat{s})$ for which a given test channel maximizes the fidelity per rate is unique up to addition of a function of s with zero mean.

Proposition 6: Let P_S be a discrete memoryless source with fidelity measure $\phi(s, \hat{s})$ and rate-fidelity function $R(\Phi)$. If the test channel $V(\hat{s}|s)$ maximizes the ratio

$$\frac{\mathbb{E}[\phi(S, \hat{S})]}{I(S; \hat{S})}$$

among all test channels and if the derivative of $R(\Phi)$ exists at $\Phi = \mathbb{E}_V[\phi(S, \hat{S})]$, then $\phi(s, \hat{s}) = c\phi_0(s, \hat{s})$ for some $c > 0$.

The proof mimics that for Proposition 2. In particular, after taking expectation over S the function $\gamma(s)$ becomes a constant that plays the same role that β plays in the proof of Proposition 2.

Propositions 5 and 6 are summarized in the following theorem.

Theorem 7: Under the technical condition of Proposition 6, the test channel $V(\hat{s}|s)$ maximizes the fidelity per rate of a discrete memoryless source if and only if the fidelity measure satisfies

$$\phi(s, \hat{s}) = c \log_2 \frac{V(\hat{s}|S)}{P_{\hat{S}}} + \gamma(s), \quad c > 0 \qquad (9)$$

with $\mathbb{E}[\gamma(S)] = 0$. For continous sources, the condition is sufficient but not necessary.

Remark 4: This result is very much related to the following result [7, Section 2.3, Problem 3], [6] (restated here in terms of fidelity). The test channel $V(\hat{s}|s)$ achieves the rate-fidelity function of the source P_S at expected fidelity $\mathbb{E}[\phi(S, \hat{S})]$ if and only if

$$\phi(s, \hat{s}) = c \log_2 \frac{V(\hat{s}|s)}{P_{\hat{S}}} + \gamma(s)$$

for $c > 0$ and an arbitrary function $\gamma(s)$. Our result shows that the extra requirement that the maximum fidelity per rate be achieved restricts the allowed fidelity functions to those with $\mathbb{E}[\gamma(S)] = 0$.

In analogy to the capacity per unit cost we define the *fidelity per unit rate* of a discrete memoryless source as

$$\hat{\Phi} = \sup_{V(\hat{s}|s)} \frac{\mathbb{E}_V[\phi(S, \hat{S})]}{I_V(S; \hat{S})} = \sup_\Phi \frac{\Phi}{R(\Phi)}.$$

In light of the previous observations, this quantity is to the source what the capacity per unit cost is to the channel.

Remark 5: To establish a connection with the perspective offered in Section IV of Verdú's paper [2], let Φ_0 be the maximum fidelity that can be achieved when representing the source by a fixed symbol, *i.e.*, $\Phi_0 = \max_{\hat{s}} \mathbb{E}[\phi(S, \hat{s})]$. The following applies:

- If $\Phi_0 > 0$ then $\hat{\Phi} = \infty$. Indeed, the positive fidelity Φ_0 is achieved at zero rate.
- If $\Phi_0 < 0$ then $0 < \hat{\Phi} < \infty$ (cf. Fig. 3).
- If $\Phi_0 = 0$ then $\hat{\Phi} = \lim_{\Phi \searrow 0} \Phi/R(\Phi)$.

This differs from Verdú's approach in that the latter – if stated in terms of fidelity – defines the counterpart to the capacity per unit cost to be $\lim_{\Phi \searrow \Phi_0} \Phi/R(\Phi)$ regardless of the value of Φ_0.

IV. Fidelity Per Unit Cost

The problems considered in the two previous sections are instances of a general class of problems where resources are traded against performance. In the channel case, we spend a cost (the resource) to achieve a high rate (the performance), and we want to maximize the performance per unit of resource. In the source case, we spend bits (the resource) to achieve a high fidelity (the performance), and once again we want to maximize the performance per unit of resource.

In this section we consider the end-to-end communication problem, where the resource is the channel input cost and the performance is the fidelity of the source reproduction. The goal is to maximize the *fidelity per cost*. The aim of this section is to find a tight bound on this ratio and to characterize the operating point that achieves the bound.

From the separation theorem, the average fidelity and cost of any communication system must satisfy

$$kR(\Phi) \leq mC(P), \qquad (10)$$

where k and m are the number of source symbols and channel uses per time, respectively. Furthermore, the largest achievable Φ for a fixed P is the one for which the above inequality is satisfied with equality. Conversely, any fidelity/cost pair satisfying the inequality can be approached by a suitable communication system using separate source and channel coding.

Using the definitions of $\hat{\Phi}$ and $\hat{C}$,

$$\hat{\Phi} = \sup_{\Phi} \frac{\Phi}{R(\Phi)} \qquad \hat{C} = \sup_{P} \frac{C(P)}{P},$$

together with (10), we obtain the following tight bound on the fidelity per cost:

$$\frac{\Phi}{P} = \frac{\Phi}{R(\Phi)} \cdot \frac{R(\Phi)}{C(P)} \cdot \frac{C(P)}{P} \leq \frac{m}{k} \hat{C}\hat{\Phi}. \qquad (11)$$

We call the right-hand side of (11) the *fidelity per unit cost* of the communication system.

A separation-based communication system achieves the fidelity per unit cost if and only if

1) the source encoder and decoder operate at $\hat{\Phi}$,
2) the channel encoder and decoder operate at $\hat{C}$, and

3) the number of bits produced by the source encoder for k source symbols equals the number of bits transmitted on the channel in n channel uses.

Using the converse to the source-channel coding theorem as done in [9] along with the results of the previous two sections we can translate the conditions for equality in (11) into conditions that apply to any system that achieves the maximum fidelity per cost, regardless of whether the system relies on the separation principle, on joint source-channel block coding, or on uncoded communication. These general conditions are stated in the following theorem.

Theorem 8: A discrete point-to-point communication system of the type shown in Figure 4 achieves the fidelity per unit cost if and only if all of the following conditions are satisfied.

1) The reproduction is related to the source via the test channel, *i.e.*,

$$P_{\hat{S}_1^k | S_1^k}(\hat{s}_1^k | s_1^k) = \prod_{i=1}^{k} V(\hat{s}_i | s_i),$$

where $s_1^k = (s_1, \ldots, s_k)$ and where V relates to the fidelity function according to (9).

2) The channel input is *iid* with distribution P_X that achieves capacity per unit cost, *i.e.*, it relates to the cost function according to (5).

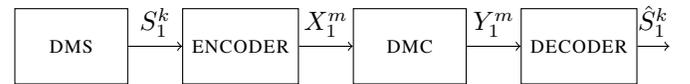3) The encoder and decoder are information lossless in the sense that $kI(S; \hat{S}) = nI(X; Y)$.

Fig. 4. A point-to-point communication system consisting of a discrete memoryless source (DMS) with distribution P_S, a discrete memoryless channel (DMC) with transition probabilities $P_{Y|X}$, an encoder and a decoder.

References

[1] F. M. Reza, *An Introduction to Information Theory*. New York: McGraw-Hill, 1961.
[2] S. Verdú, "On channel capacity per unit cost." *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 1019–1030, 1990.
[3] R. G. Gallager, "Energy limited channels: Coding, multiaccess and spread spectrum," in *Proc. 1988 Conf. Inform. Sci. Syst.*, March 1988, p. 372.
[4] R. Blahut and R. Koetter, Eds., *Codes, Graphs, and Systems*. Springer, 2002.
[5] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, Jun 2002.
[6] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited." *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, 2003.
[7] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
[9] M. Gastpar, "To code or not to code," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2002.