Mercury/Waterfilling: Optimum Power Allocation with Arbitrary Input Constellations

Angel Lozano
Bell Labs (Lucent Technologies)
Holmdel NJ 07733, USA
Email: aloz@lucent.com

Antonia M. Tulino
Universita di Napoli "Federico II"
Napoli 80125, Italy
Email: atulino@princeton.edu

Sergio Verdú
Princeton University
Princeton NJ 08540, USA
Email: verdu@princeton.edu

Abstract— **For parallel independent Gaussian-noise channels with an aggregate power constraint, independent Gaussian inputs whose powers are allocated according to the waterfilling policy maximize the sum mutual information. In practice, however, discrete signalling constellations such as m-PSK or m-QAM are used in lieu of the ideal Gaussian signals. This paper gives the power allocation policy, referred to as *mercury/waterfilling*, that maximizes the sum mutual information over parallel channels with arbitrary input constellations.**

I. INTRODUCTION

A problem often encountered in transmitter design is that of having to allocate a certain amount of power among a bank of parallel noninteracting channels. In the all-important case that the noises impairing each of the parallel channels are Gaussian and independent, the mutual information is maximized if the inputs to those channels are mutually independent and also Gaussian, with the power allocated according to the *waterfilling* policy [1, Section 10.4]. First devised by Shannon in 1949 [2] and rigorously formalized in the context of dispersive channels in [3]–[6], the waterfilling policy is a central result in information theory.

Examples of parallel channels abound in both wireline and wireless communication:

- *Multicarrier transmission.* Signalling takes place over a number of distinct frequency bands, each of which constitutes a parallel channel.
- *Multiantenna communication.* If multiple transmit and receive antennas are employed and the transfer coefficients between them are known, the left singular vectors of the resulting matrix can be used for signaling and the right singular vectors for reception. The outcome is a set of parallel noninteracting channels [7].
- *Power control for fading channels.* When the gain of an individual frequency-flat channel varies over time, it can be seen as a collection of parallel channels where each such channel encompasses a group of symbols over which the fading coefficients are identical [8].
- *Dispersive channels.* In linear dispersive channels, a power-preserving orthonormal linear transformation at transmitter and receiver turns the channel into one with parallel branches having uncorrelated noises.

Although Gaussian inputs are optimum from a mutual information standpoint, they can never be realized in practice. Rather, the inputs are usually drawn from discrete constellations (often with limited peak-to-average ratios), which may significantly depart from the Gaussian idealization. Yet, no solution has been found to date for the power allocation that maximizes the mutual information over parallel channels with non-Gaussian inputs. A direct obstacle in the way of this formulation is the lack of explicit expressions for the corresponding mutual informations.

Also for specific coding schemes, or even in the absence of coding, the allocation of power in order to maximize the throughput with discrete constellations at some target error probability is hampered by the lack of explicit expressions, in this case for the throughput functions [9]. A strategy often adopted to approximate such throughput-maximizing power allocation consists of applying the waterfilling policy, except with the gain of each channel reduced by a *gap* that quantifies the deficit of the corresponding constellation with respect to a Gaussian signal operating at the same rate [10]. This strategy hinges on the assumption that the gap deficit is roughly constant for the range of rates of interest, an assumption that is only valid if the inputs can be drawn from a family of constellations whose size grows with the interval of signal-to-noise ratios of interest.

In this paper, we derive the power allocation policy that maximizes the sum mutual information achieved over parallel noninteracting Gaussian channels with arbitrary (not necessarily identical for each channel) input constellations. Capitalizing on the recently unveiled fundamental relationship between mutual information and MMSE [11], we formulate our power-allocation policy using the readily computable nonlinear MMSE of the input constellations given their noisy outputs. Thus, the need for explicit mutual information expressions is conveniently circumvented and the solution is instilled with operational significance.

In certain dispersive channels (e.g., magnetic recording), where each input to the channel must belong to a given constellation, the joint distribution on the set of transmitted sequences that maximizes mutual information must be found without the benefit of an orthonormal transformation, which would distort the constellations seen by the channel. While an analytical solution to this problem remains open, our solution does apply to any scenario where an orthonormal transformation can be used to create parallel noninteracting

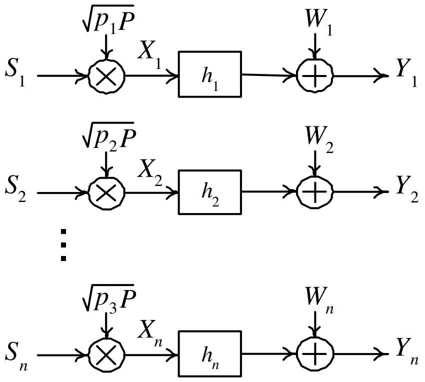

Fig. 1. Bank of n parallel noninteracting channels.

branches (e.g., in multiantenna communication with channel coefficients known at the transmitter).

II. Problem Formulation

Consider n parallel channels as depicted in Fig. 1. On the ith such channel, the input-output relationship is

$$Y_i = h_i X_i + W_i \tag{1}$$

where the complex scalar h_i is a deterministic gain while the noise W_i is a zero-mean unit-variance complex Gaussian random variable independent of the noise on the other channels. Regardless of their marginal distributions, the zero-mean complex inputs $\{X_i\}_{i=1}^n$ should be independent for the mutual information to be maximized. The aggregate power constraint is

$$\frac{1}{n}\sum_{i=1}^{n} E\left[|X_i|^2\right] \leq P. \tag{2}$$

It is convenient to introduce normalized unit-power inputs $\{S_i\}_{i=1}^n$, with distribution dictated by the modulation scheme, such that

$$X_i = \sqrt{p_i P}\, S_i \tag{3}$$

with the (normalized) powers $\{p_i\}_{i=1}^n$ constrained by

$$\frac{1}{n}\sum_{i=1}^{n} p_i \leq 1. \tag{4}$$

We can define, for each parallel channel,

$$\gamma_i = P\,|h_i|^2 \tag{5}$$

which is a measure of the strength of that channel. Specifically, $p_i\gamma_i$ is the signal-to-noise ratio on the ith channel and thus $\{\gamma_i\}_{i=1}^n$ represent the signal-to-noise ratios when the power allocation is uniform ($p_i = 1$ for $i = 1, \ldots, n$).

Define the input-output mutual information on the ith channel, in nats/s/Hz, as

$$\mathcal{I}_i(\rho) = I(S_i; \sqrt{\rho} S_i + W_i). \tag{6}$$

The problem that we pose is the determination of the (normalized) power allocation $\{p_i^\star\}_{i=1}^n$ that maximizes the sum mutual information while satisfying (4), i.e.,

$$[p_1^\star, \ldots, p_n^\star] = \arg \max_{\substack{p_1,\ldots,p_n \\ \frac{1}{n}\sum_i p_i = 1}} \sum_{i=1}^{n} \mathcal{I}_i(p_i\gamma_i) \tag{7}$$

In addition to the power allocation, the receiver is presumed cognizant of the magnitude and phase of the channel gains, $\{h_i\}_{i=1}^n$. The transmitter, on the other hand, need only know the magnitudes $\{|h_i|\}_{i=1}^n$.

III. Optimum Power Allocation Policy

A key ingredient in our power allocation policy is the MMSE incurred in the estimation of the signals $\{S_i\}_{i=1}^n$. The MMSE estimate of S_i is the conditional mean

$$\hat{S}_i(y_i, \rho) = E\left[S_i | \sqrt{\rho} S_i + W_i = y_i\right] \tag{8}$$

which is, in general, a nonlinear function of the observation y_i. (It becomes linear if S_i is Gaussian.) The corresponding mean-square error is then

$$\mathsf{MMSE}_i(\rho) = E\left[\left|S_i - \hat{S}_i(\sqrt{\rho}S_i + W_i, \rho)\right|^2\right] \tag{9}$$

where the expectation is over both S_i and W_i. Since S_i has unit power, $\mathsf{MMSE}_i(\cdot) \in (0, 1]$.

Note that $\mathcal{I}_i(\cdot)$ and $\mathsf{MMSE}_i(\cdot)$ depend on the index i only through the distribution of S_i.

Regardless of the distribution of S_i, the functions in (6) and (9) are related through the following key formula (couched in our notation).

Theorem 1: [11] For any distribution of S_i (not dependent on ρ),

$$\frac{d}{d\rho}\mathcal{I}_i(\rho) = \mathsf{MMSE}_i(\rho). \tag{10}$$

This relationship is the seed for the following optimum power allocation policy, whose proof is given in [12].

Theorem 2: The (normalized) powers $\{p_i^\star\}_{i=1}^n$ that maximize the sum mutual information satisfy

$$p_i^\star = 0 \qquad \gamma_i \leq \eta \tag{11}$$

$$\gamma_i\,\mathsf{MMSE}_i(p_i^\star\gamma_i) = \eta \qquad \gamma_i > \eta \tag{12}$$

with η such that

$$\frac{1}{n}\sum_{i=1}^{n} p_i^\star = 1 \tag{13}$$

Denoting the inverse of $\mathsf{MMSE}_i(\cdot)$ with respect to composition by $\mathsf{MMSE}_i^{-1}(\cdot)$, with domain equal to $[0, 1]$, the (normalized) powers $\{p_i^\star\}_{i=1}^n$ can be recast more explicitly as

$$p_i^\star = \frac{1}{\gamma_i}\,\mathsf{MMSE}_i^{-1}\left(\min\left\{1, \frac{\eta}{\gamma_i}\right\}\right) \qquad i = 1, \ldots, n \tag{14}$$

with η given by the solution to the nonlinear equation

$$\sum_{\substack{i=1 \\ \gamma_i > \eta}}^{n} \frac{1}{n\gamma_i} \mathsf{MMSE}_i^{-1}\left(\frac{\eta}{\gamma_i}\right) = 1. \tag{15}$$

In light of (14) and (15), the allocation of power can be viewed as a two-step process:

i) Solve for η in (15).
ii) Use η to identify $\{p_i^\star\}_{i=1}^n$ via (14).

In the special case that the inputs are Gaussian,

$$\mathsf{MMSE}_i(\rho) = \frac{1}{1+\rho} \tag{16}$$

and, correspondingly,

$$\mathsf{MMSE}_i^{-1}(\xi) = \frac{1}{\xi} - 1 \tag{17}$$

which reduce Theorem 2 to the waterfilling policy

$$p_i^\star = 0 \qquad \gamma_i \le \eta \tag{18}$$

$$p_i^\star = \frac{1}{\eta} - \frac{1}{\gamma_i} \qquad \gamma_i > \eta \tag{19}$$

with $1/\eta$ the water level.

IV. INTERLUDE: NONLINEAR MMSE ESTIMATION

Given the chief role it plays in the formulation of the power allocation policy, a brief review of nonlinear estimation for input constellations of common use is in order. Since a generic channel is dealt with, the index i is dropped.

Consider an m-ary modulation defined by m discrete equiprobable values, $\{s_\ell\}_{\ell=1}^m$. This includes m-PAM, m-PSK, m-QAM, etc. In the presence of Gaussian noise, (8) leads to

$$\hat{S}(y,\rho) = \frac{\sum_{\ell=1}^m s_\ell \, e^{-|y-\sqrt{\rho}\, s_\ell|^2}}{\sum_{\ell=1}^m e^{-|y-\sqrt{\rho}\, s_\ell|^2}} \tag{20}$$

while

$$\mathsf{MMSE}(\rho) = \int \sum_{\ell=1}^{m} \left| s_\ell - \hat{S}(y,\rho) \right|^2 \frac{e^{-|y-\sqrt{\rho}\, s_\ell|^2}}{m\pi} dy \tag{21}$$

$$= 1 - \frac{1}{m\pi} \int \frac{|\sum_{\ell=1}^m s_\ell \, e^{-|y-\sqrt{\rho}\, s_\ell|^2}|^2}{\sum_{\ell=1}^m e^{-|y-\sqrt{\rho}\, s_\ell|^2}} dy \tag{22}$$

with integration over the complex field. The expression in (22) can be further elaborated for specific constellations such as:

- BPSK, for which

$$\mathsf{MMSE}^{\mathrm{BPSK}}(\rho) = 1 - \int_{-\infty}^{\infty} \tanh\left(2\sqrt{\rho}\,\xi\right) \frac{e^{-(\xi-\sqrt{\rho})^2}}{\sqrt{\pi}} d\xi. \tag{23}$$

- QPSK, which amounts to two BPSK constellations in quadrature, each with half the QPSK power, such that

$$\mathsf{MMSE}^{\mathrm{QPSK}}(\rho) = \mathsf{MMSE}^{\mathrm{BPSK}}\left(\frac{\rho}{2}\right). \tag{24}$$

With a modicum of algebra, integrals for other constellations can be similarly obtained from (22). Constellations whose discrete values are not equiprobable can be tackled with a slight generalization thereof.

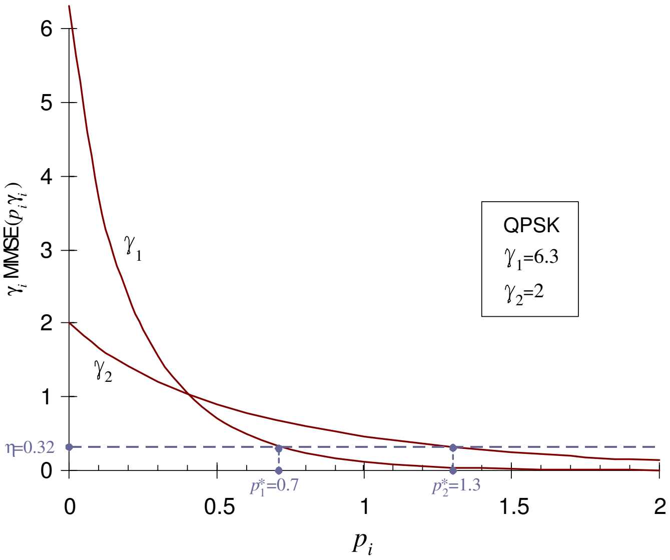

Fig. 2. Two parallel channels with $\gamma_1 = 6.3$ and $\gamma_2 = 2$ and QPSK inputs. Solid lines depict $\gamma_i \mathsf{MMSE}(p_i\gamma_i)$ as function of p_i. Their intersection with the dashed line at $\eta = 0.32$ yields $p_1^\star$ and $p_2^\star$.

V. GRAPHICAL INTERPRETATIONS

The waterfilling policy owes much of its popularity to its very intuitive graphical interpretation. In the same spirit, this section provides graphical interpretations that cast valuable insight on the more general policy spelled out by Theorem 2.

A. *MMSE-Power Charts*

A direct depiction of Theorem 2 can be obtained by charting the functions $\gamma_i \, \mathsf{MMSE}_i(p_i\gamma_i)$, for $i = 1, \ldots, n$, as function of p_i on the interval $[0, n]$. Every power allocation that satisfies (11)–(12) corresponds to drawing a horizontal line whose interception with each of the functions directly gives the (normalized) power on the corresponding channel. This line must be at an ordinate η such that (13) is satisfied. If the line were higher, some power would be left unused; if lower, the power constraint would be exceeded.

Example 1: Let $n = 2$ with $\gamma_1 = 6.3$ and $\gamma_2 = 2$ (respectively 8 dB and 3 dB) and with QPSK inputs on both channels. As shown in Fig. 2, $\eta = 0.32$ yields $p_1^\star = 0.7$ and $p_2^\star = 1.3$.

Example 2: Let $n = 2$ with $\gamma_1 = 1$ and $\gamma_2 = 0.1$ (respectively 0 dB and -10 dB) and with QPSK inputs on both channels. As shown in Fig. 3, $\eta = 0.23$ yields $p_1^\star = 2$ and $p_2^\star = 0$.

Similar charts can be plotted whenever the inputs to the parallel channels exhibit different distributions.

Example 3: Let $n = 3$ with $\gamma_1 = \gamma_2 = \gamma_3 = 2$ (i.e., 3 dB on each channel) but with distinct inputs, respectively BPSK, QPSK and Gaussian. As shown in Fig. 4, $\eta = 0.23$ yields $p_1^\star = 0.47$, $p_2^\star = 0.96$ and $p_3^\star = 1.57$. Despite their equal strengths, channels with richer input distributions seize a larger fraction of the power.

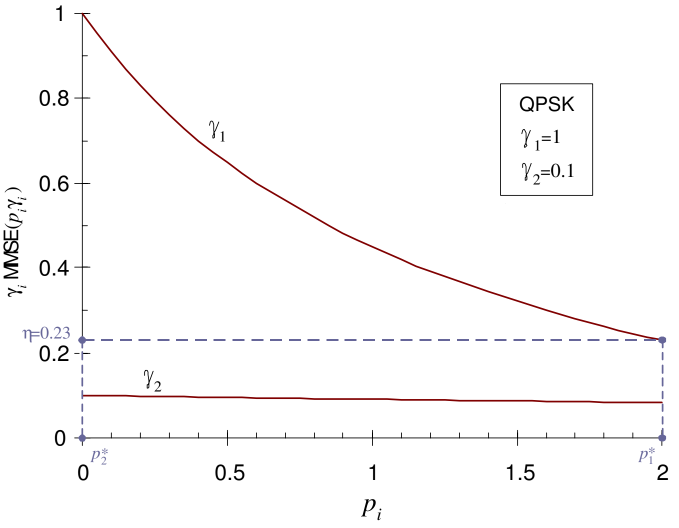

Fig. 3. Two parallel channels with $\gamma_1 = 1$ and $\gamma_2 = 0.1$ and QPSK inputs. Solid lines depict $\gamma_i \mathsf{MMSE}(p_i \gamma_i)$ as function of p_i. Their intersection with the dashed line at $\eta = 0.23$ yields $p_1^\star$ and $p_2^\star$.

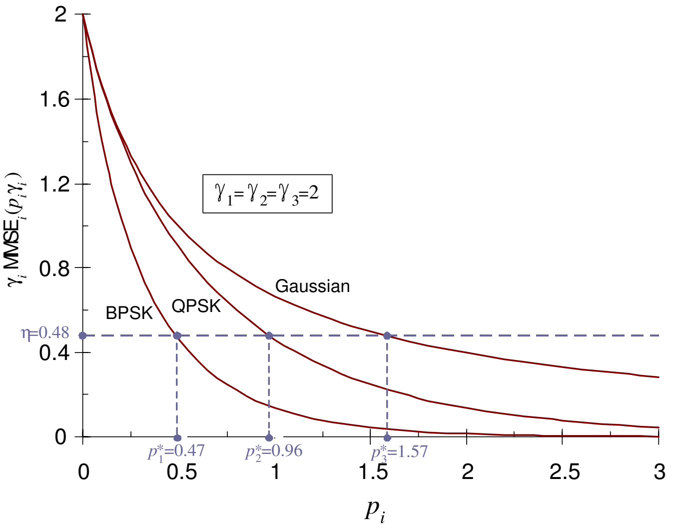

Fig. 4. Three parallel channels with $\gamma_1 = \gamma_2 = \gamma_3 = 2$. The respective inputs are BPSK, QPSK and Gaussian. Solid lines depict $\gamma_i \mathsf{MMSE}(p_i \gamma_i)$ as function of p_i. Their intersection with the dashed line at $\eta = 0.48$ yields $p_1^\star$, $p_2^\star$ and $p_3^\star$.

B. Mercury/Waterfilling

While the MMSE-power charts are very useful, we also seek an alternative interpretation that more directly generalizes the waterfilling process and allows retaining some of its intuition. To that end, it is convenient to relate the functions $\mathsf{MMSE}_i^{-1}(\cdot)$ with their counterpart for a Gaussian input distribution, given in (17). Let us define, for any arbitrary input distribution,

$$G_i(\rho) = \begin{cases} 1/\rho - \mathsf{MMSE}_i^{-1}(\rho) & \rho \in [0,1] \\ 1 & \rho > 1 \end{cases} \tag{25}$$

such that, for a Gaussian input, $G_i(\cdot) = 1$.

The function $G_i(\cdot)$ enables the following interpretation of our power allocation policy, which we refer to as *mercury/waterfilling* (cf. Fig. 5).

(a) For each of the n channels, set up a unit-base vessel solid up to a height $1/\gamma_i$.
(b) Choose η. Pour mercury onto each of the vessels until its height (including the solid) reaches $G_i(\eta/\gamma_i)/\gamma_i$.
(c) Waterfill, keeping identical upper level of water in all vessels, with a volume of water equal to n (or, equivalently, until the water level reaches $1/\eta$).
(d) The water height over the mercury on the ith vessel equals $p_i^\star$.

What distinguishes this interpretation from conventional waterfilling is the mercury pouring stage, which regulates the water admitted by each vessel tailoring the process to arbitrary input distributions. (No mercury is poured onto vessels whose input is Gaussian or for which $\gamma_i < \eta$.) Just as conventional waterfilling, mercury/waterfilling is a parametric procedure leading to optimal mutual information and aggregate power that are continuous monotonically decreasing functions of η.

Note that the mercury plays the role of the *gap* with respect to an ideal Gaussian signal. On the ith channel, specifically, the gap corresponds to $G_i(\eta/\gamma_i)$. The optimum such gap that must be applied to each individual channel depends—in general—not only on the strength and input distribution on that channel but, through η, on the strengths and input distributions of the other channels as well.

In a real-time implementation, $\mathsf{MMSE}_i^{-1}(\cdot)$ can be tabulated for every constellation of interest and hence the problem boils down to solving (15). This can be accomplished through established iterative methods (bisection, secant, Newton, etc). Advantageously, the number of iterations does not depend on the number of channels but only on the desired degree of approximation to the optimum solution. Furthermore, some of these iterative methods (bisection, secant) do not require any derivatives but simply the functions $\mathsf{MMSE}_i^{-1}(\cdot)$ themselves.

VI. SUBOPTIMALITY OF WATERFILLING

The idea that stronger channels should be allocated more power, an ingrained belief that originates in the waterfilling policy, may not hold with constellations of practical interest (cf. Example 1). In fact, in some circumstances the mercury/waterfilling policy mandates exactly the opposite.

Theorem 3: If all the inputs conform to the same m-ary constellation, power equalization is asymptotically optimum in the high-power regime. For $P \to \infty$, the (normalized) powers converge to

$$p_i^\star = \frac{\alpha}{|h_i|^2} \tag{26}$$

with

$$\alpha = \frac{1}{\frac{1}{n}\sum_{\ell=1}^{n} \frac{1}{|h_\ell|^2}}. \tag{27}$$

Since the mutual information of a m-ary constellation cannot exceed $\log_2 m$ bits/s/Hz, there is little incentive to

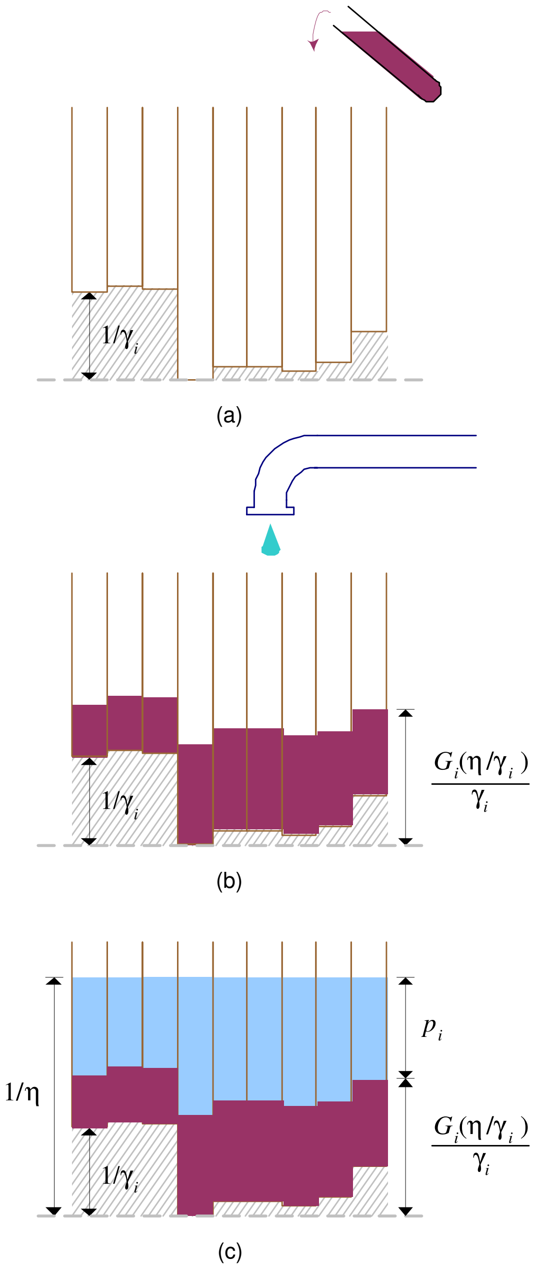

Fig. 5. Mercury/waterfilling. (a) On vessels solid up to heights $\{1/\gamma_i\}$, pour mercury until its height reaches $G_i(\eta/\gamma_i)/\gamma_i$ on each vessel. (b) Waterfill with a volume of water equal to n, after which the water level reaches $1/\eta$. (c) The water height over the mercury on the ith vessel gives $p_i^\star$.

allocate further power to a channel once it nears such point. Rather, that power is better allocated to a weaker channel. The proof of Theorem 3, and its generalization to different m-ary constellations on the various channels, are given in [12].

Given the prevalence of waterfilling in information theory, it is worthwhile to investigate how well it performs with simple constellations. Fig. 6 quantifies the suboptimality of waterfilling for $n = 2$ channels fed with QPSK inputs. Specifically, the figure depicts the aggregate power penalty (in dB) incurred by waterfilling to achieve the same sum mutual information as mercury/waterfilling as function of the channel strengths $\{\gamma_1, \gamma_2\}$ (in dB). For strengths ranging from -4 dB to 8 dB, the excess aggregate power is over 3 dB. In fact, the power penalty incurred by waterfilling grows without bound as the channel strengths become more imbalanced.

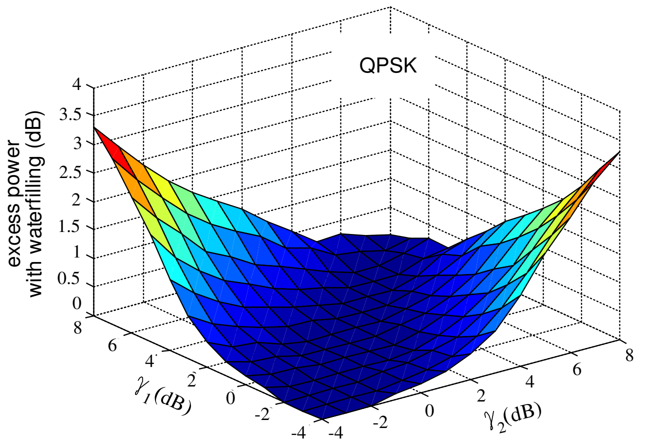

Fig. 6. Excess aggregate power (in dB) required for waterfilling to achieve the same sum mutual information as the optimum policy, for $n = 2$ channels with QPSK inputs, as function of the pair $\{\gamma_1, \gamma_2\}$ (in dB).

Another problem in which our solution can be used as a building block is that of bit allocation, in which a given bit budget needs to be allocated among parallel branches. Mercury/waterfilling gives the best achievable mutual information for any possible allocation of bits.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, Wiley, 1990.

[2] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

[3] J. L. Holsinger, "Digital communications over fixed time-continuous channels with memory, with special application to telephone channel," *M.I.T. Res. Lab. Electron. Rep.*, vol. 430, pp. 460, 1964.

[4] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968.

[5] B. S. Tsybakov, "Capacity of a discrete-time Gaussian channel with a filter," *Problems of Information Transmission*, vol. 6, pp. 253–256, Jul.–Sep. 1970.

[6] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. on Inform. Theory*, vol. 34, pp. 380–388, May 1988.

[7] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585–595, Nov. 1999.

[8] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. on Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[9] E. Baccarelli, A. Fasano, and M. Biagi, "Novel efficient bit-loading algorithms for peak-energy-limited ADSL-type multicarrier systems," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1237–1239, May 2002.

[10] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[11] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. on Inform. Theory*, vol. 51, no. 4, pp. 1261–1283, Apr. 2005.

[12] A. Lozano, A. M. Tulino, and S. Verdú, "Power allocation over parallel Gaussian channels with arbitrary input distributions," *preprint*, 2005.