Optimum Power Allocation for Parallel Gaussian Channels with Arbitrary Input Distributions

Angel Lozano[*], Antonia M. Tulino[†], and Sergio Verdú[‡]

February 17, 2006

Abstract

The mutual information of independent parallel Gaussian-noise channels is maximized, under an average power constraint, by independent Gaussian inputs whose power is allocated according to the waterfilling policy. In practice, discrete signalling constellations with limited peak-to-average ratios (m-PSK, m-QAM, etc) are used in lieu of the ideal Gaussian signals. This paper gives the power allocation policy that maximizes the mutual information over parallel channels with arbitrary input distributions. Such policy admits a graphical interpretation, referred to as *mercury/waterfilling*, which generalizes the waterfilling solution and allows retaining some of its intuition. The relationship between mutual information of Gaussian channels and nonlinear minimum mean-square error proves key to solving the power allocation problem.

Keywords: Gaussian Channels; Power Allocation; Waterfilling; Channel Capacity; Mutual Information; MMSE.

---

[*] Angel Lozano is with Bell Laboratories (Lucent Technologies), Holmdel, NJ07733, USA.

[†] Antonia M. Tulino is with Universita Degli Studi di Napoli "Federico II", 80125 Napoli, Italy

[‡] Sergio Verdú is with Princeton University, Princeton, NJ08544, USA.

I Introduction

A problem often encountered in transmitter design is that of allocating a certain amount of power among a bank of independent parallel channels. Examples abound, both in the wireline and the wireless domains:

- *Multitone transmission.* Signalling takes place over a number of distinct frequency bands, each of which constitutes a parallel channel [1]. These bands may be nonoverlapping or, as in OFDM (orthogonal frequency division multiplexing) and DMT (discrete multitone), overlapping but with spectral shapes designed to ensure orthogonality. Multitone techniques, whose origins can be traced back to early multichannel modulation schemes developed in the 1950s, have matured over several decades and are nowadays thriving with a host of commercial applications. In the wireline world, a prime application is in DSL (digital subscriber lines).
- *Multiantenna communication.* If multiple transmit and receive antennas are employed and the transmitter knows the transfer coefficients between the antennas, the left and right singular vectors (corresponding to nonzero singular values) of the resulting matrix can be used for signaling and reception, respectively. The outcome is a set of independent parallel channels [2]–[4].
- *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 [5].
- *Dispersive channels.* For linear dispersive channels or parallel channels with correlated noises, a power-preserving orthonormal linear transformation at transmitter and receiver turns the channel into one with parallel branches having uncorrelated noises and possibly different signal-to-noise ratios.

Diverse criteria can be invoked in order to decide which portion of the available power is allocated to each of the channels. A particularly enticing criterion, on which we shall focus henceforth, is the maximization of the input-output mutual information. In the all-important case that the noises impairing each of the parallel branches are Gaussian and independent, the mutual information under an average power constraint is maximized if the inputs to the channels are mutually independent and also Gaussian, with their power allocated according to the *waterfilling* policy [6, Sec. 10.4]. First devised by Shannon in 1949 [7] and rigorously formalized in the context of dispersive channels in [8]–[11], the waterfilling[1] policy is a central result in information theory.[2]

Although Gaussian inputs are optimum from a mutual information standpoint, they can never be realized in practice. Rather, the inputs must usually be drawn from discrete constellations (often with very 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 optimization 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 (and bits) 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 [21]. In the absence of a precise solution, a heuristic 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 class of constellations with respect to a Gaussian signal operating at the same rate [22]–[31] (cf. Appendix H). This strategy hinges on the assumption that the gap deficit is roughly constant for all the rates of interest, an assumption that is only valid if the cardinality of the available constellations (within the class for which the gap is defined) grows with the interval of signal-to-noise ratios of interest.

In this paper, we obtain the power allocation policy that maximizes the mutual information achieved over independent parallel Gaussian channels with arbitrary (and not necessarily identical for each channel) input distributions. Capitalizing on the recently unveiled fundamental relationship between mutual information and MMSE (minimum mean-square error) [32], we formulate our power allocation policy using the readily computable nonlinear MMSE of the inputs given their noisy outputs. Thus, the need for explicit mutual information expressions is conveniently circumvented and the solution is instilled with immediate operational significance.

In certain channels subject to intersymbol interference (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 this problem remains open,[3] our solution does apply to any other scenario where an orthonormal transformation can indeed be used to create parallel noninteracting branches. For example, our solution gives the optimum power allocation in multiantenna communication with channel coefficients known at the transmitter, where a pragmatic power-preserving linear transformation coupling the antenna feeds to the input constellations is the orthonormal transformation.

---

[1] The term *waterfilling* (or, equivalently, *waterpouring*) appears to have been coined by Fano [12].

[2] The waterfilling interpretation has also permeated the signal processing literature over the years. Power allocations that can be viewed as multidimensional waterfillings have been derived as solutions to a number of constrained optimization problems involving various uncoded performance criteria [13]–[20].

[3] See [33] for a review of the state-of-the-art on Monte Carlo methods to approximate the maximum mutual information of antipodal-input linear dispersive channels.

The organization of the paper is as follows. In Section II, the quantities to be used throughout are introduced and the power allocation problem is formulated. The solution is given in Section III in the form of an optimum power allocation policy. A key ingredient of this policy is the nonlinear minimum mean-square error, to which Section IV is devoted entirely. Section V presents graphical interpretations of the optimum policy leading to the notion of *mercury/waterfilling*. To gain insight, Sections VI and VII specialize the optimum policy to the low- and high-power regimes, respectively. In turn, Section VIII exemplifies the application of mercury/waterfilling to the problem of multitone communication using a canonical DSL test channel. Another specific application, power control for fading channels, is explored in Section IX. Finally, the paper concludes in Section X with the proofs of several results relegated to the appendices.

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 noise W_i is a zero-mean unit-variance complex Gaussian random variable independent of the noise on the other channels while the complex scalar h_i is a deterministic nonzero gain.[4] Regardless of their marginal distributions, the complex inputs $\{X_i\}_{i=1}^{n}$ should be independent for the mutual information to be maximized [6]. Since any nonzero mean contributes to the power but not to the mutual information, we only consider zero-mean inputs. The 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 $\{S_i\}_{i=1}^{n}$, whose distribution is dictated by the modulation scheme, such that

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

---

[4] Since the power allocated to any channel with zero gain would be strictly zero, such channels can be removed from the problem.

with the power allocation $\{p_i\}_{i=1}^n$ constrained by

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

We can define, for each parallel channel,

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

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 as[5]

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

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

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

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 inputs $\{S_i\}_{i=1}^n$. 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

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

---

[5] Unless otherwise stated, throughout the paper the mutual information is expressed in nats/s/Hz and the logarithms are in natural base.

with expectation 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. Such dependence can be dropped in the case that all the channels are fed with identical distributions.

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 *[32] 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.

Theorem 2 *The power allocation $\{p_i^\star\}_{i=1}^n$ that solves (7) satisfies*

$$\begin{aligned} p_i^\star &= 0 & \gamma_i \leq \eta \quad (11)\\ \gamma_i \, \mathsf{MMSE}_i(p_i^\star \gamma_i) &= \eta & \gamma_i > \eta \quad (12) \end{aligned}$$

with η such that

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

Proof: *See Appendix A.*

Denoting the inverse of $\mathsf{MMSE}_i(\cdot)$ with respect to composition of functions by $\mathsf{MMSE}_i^{-1}(\cdot)$, with domain equal to $[0,1]$ and $\mathsf{MMSE}_i^{-1}(1) = 0$, the power allocation $\{p_i^\star\}_{i=1}^n$ 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 the unique 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,

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

which reduce Theorem 2 to the waterfilling policy

$$p_i^{\mathsf{WF}} = 0 \qquad \gamma_i \leq \eta \tag{18}$$

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

with $1/\eta$ playing the role of the water level. Furthermore, in the Gaussian-input case, Theorem 2 provides an alternative fixed-point form for the waterfilling policy that is worth noting.

Corollary 1 *With Gaussian inputs,*

$$p_i^{\mathsf{WF}} = 0 \qquad \gamma_i \leq \eta \tag{20}$$

$$p_i^{\mathsf{WF}} = \frac{1 - \mathsf{MMSE}_i(p_i^{\mathsf{WF}}\gamma_i)}{\frac{1}{n}\sum_{\ell=1}^{n}(1 - \mathsf{MMSE}_\ell(p_\ell^{\mathsf{WF}}\gamma_\ell))} \qquad \gamma_i > \eta \tag{21}$$

Since the numerator in (21) indicates the signal power at the output of a (linear) MMSE estimator of S_i, the interpretation of Corollary 1 is straightforward: when the inputs are Gaussian, the fraction p_i^{WF}/n of transmit power allocated to each channel should equal the fraction of total received signal power recovered from that channel. Corollary 1 has the additional advantage of generalizing, unlike waterfilling in its traditional form, to the case of parallel interacting channels, i.e., to nondiagonal multiple-input multiple-output channels [34].

A basic attribute of waterfilling is that the larger the gain of a channel, the higher its allocated power. As we shall see once we elaborate on the optimum power allocation policy described by Theorem 2, this property is not upheld in general.

IV Interlude: Nonlinear MMSE Estimation

Given the chief role played by MMSE estimation in the formulation of the power allocation policy, a brief overview of nonlinear estimation for commonly used input distributions is in order. In addition to a review, this section also presents some new results (Theorems 3–4) which, although derived as stepping stones towards information-theoretic derivations given later in the paper, may be of interest on their own in nonlinear estimation theory. Since this section deals with a generic channel, the index i is dropped.

A Discrete Constellations

Consider an m-ary modulation defined by m discrete points, denoted by $\{s_\ell\}_{\ell=1}^m$, taken with probabilities $\{q_\ell\}_{\ell=1}^m$ such that $\sum_{\ell=1}^m q_\ell = 1$. This includes:

- m-PAM (pulse amplitude modulation), where $s_\ell \in \left\{(2\ell - 1 - m)\sqrt{\frac{3}{m^2-1}}\right\}$.
- m-PSK (phase shift keying), where $s_\ell \in \left\{e^{\mathrm{j}\frac{2\pi}{m}\ell+\phi}\right\}$ with ϕ an arbitrary phase.
- m-QAM (quadrature amplitude modulation), which for even m consists of two $m/2$-PAM constellations in quadrature, each with half the power. (Nonsquare m-QAM constellations not admitting such decomposition also exist, but they are less frequently used.)

Each of these classes of constellations is exemplified in Fig. 2, which depicts 4-PAM, 8-PSK and 16-QAM. Although in most practical implementations the signalling is equiprobable, i.e., $q_\ell = 1/m$ for every ℓ, this is (with the obvious exception of m-PSK) in general suboptimal in terms of mutual information [35, 36, 37]. The optimum mass probabilities are usually a function of the input power. In the remainder, the formulation considers arbitrary—but fixed—mass probabilities, $\{q_\ell\}_{\ell=1}^m$. Unless otherwise stated, the examples and plots throughout the paper correspond to equiprobable distributions.

In the presence of Gaussian noise, (8) leads to

$$\hat{S}(y,\rho) = \frac{\sum_{\ell=1}^m q_\ell \, s_\ell \, e^{-|y-\sqrt{\rho}\, s_\ell|^2}}{\sum_{\ell=1}^m q_\ell \, e^{-|y-\sqrt{\rho}\, s_\ell|^2}} \tag{22}$$

while

$$\begin{aligned} \mathsf{MMSE}(\rho) &= \int \sum_{\ell=1}^m q_\ell \left| s_\ell - \hat{S}(y,\rho)\right|^2 \frac{e^{-|y-\sqrt{\rho}\, s_\ell|^2}}{\pi} dy & (23)\\ &= 1 - \frac{1}{\pi}\int \frac{|\sum_{\ell=1}^m q_\ell \, s_\ell \, e^{-|y-\sqrt{\rho}\, s_\ell|^2}|^2}{\sum_{\ell=1}^m q_\ell \, e^{-|y-\sqrt{\rho}\, s_\ell|^2}} dy & (24)\end{aligned}$$

where the integrals are over the complex field. The expression for $\mathsf{MMSE}(\cdot)$ in (24) can be further elaborated for specific constellations:

- BPSK (or equivalently 2-PAM), for which

$$\mathsf{MMSE}^{\mathrm{BPSK}}(\rho) = 1 - \int_{-\infty}^{\infty} \tanh\left(2\sqrt{\rho}\,\xi\right) \frac{e^{-(\xi-\sqrt{\rho})^2}}{\sqrt{\pi}} d\xi. \tag{25}$$

- QPSK (or equivalently 4-QAM), which amounts to two BPSK constellations in quadrature, each with half the QPSK power, resulting in

$$\mathsf{MMSE}^{\mathrm{QPSK}}(\rho) = \mathsf{MMSE}^{\mathrm{BPSK}}\left(\frac{\rho}{2}\right). \tag{26}$$

- 4-PAM, for which

$$\mathsf{MMSE}^{4-\mathrm{PAM}}(\rho) = 1 - \int_{-\infty}^{\infty} \frac{\left(e^{-8\rho/5}\sinh(6\sqrt{\frac{\rho}{5}}\xi) + \sinh(2\sqrt{\frac{\rho}{5}}\xi)\right)^2}{e^{-8\rho/5}\cosh(6\sqrt{\frac{\rho}{5}}\xi) + \cosh(2\sqrt{\frac{\rho}{5}}\xi)} \frac{e^{-\xi^2-\rho/5}}{10\sqrt{\pi}} d\xi. \tag{27}$$

- 16-QAM, for which decomposition in two 4-PAM constellations in quadrature yields

$$\mathsf{MMSE}^{16-\mathrm{QAM}}(\rho) = \mathsf{MMSE}^{4-\mathrm{PAM}}\left(\frac{\rho}{2}\right). \tag{28}$$

With a modicum of algebra, integrals for other constellations can be similarly obtained from (24).

Fig. 3 displays $\mathsf{MMSE}(\cdot)$ for BPSK, QPSK, 8-PSK, 4-PAM and 16-QAM constellations alongside the corresponding $\mathsf{MMSE}(\cdot)$ for a Gaussian input. Note the evident classification that occurs for small arguments: all those constellations that are *proper complex*, i.e., quadrature symmetric (which can be composed as a mixture of rotated and/or scaled QPSK distributions) behave as

$$\mathsf{MMSE}(\rho) = 1 - \rho + \mathcal{O}(\rho^2) \tag{29}$$

which is the same behavior exhibited by a Gaussian input. This reflects the second-order optimality of proper complex constellations in terms of low-power mutual information [38], which by means of Theorem 1 maps onto first-order optimality in terms of minimum mean-square error. More generally,

$$\mathsf{MMSE}(\rho) = 1 + \ddot{\mathcal{I}}(0)\,\rho + \mathcal{O}(\rho^2) \tag{30}$$

where, for constellations that are not proper complex, $\ddot{\mathcal{I}}(0) < -1$. (For BPSK and m-PAM, specifically, $\ddot{\mathcal{I}}(0) = -2$.) The quantity $\ddot{\mathcal{I}}(0)$ is thus the key low-power performance measure of any constellation as already evidenced in [38], where it was shown that the bandwidth required to sustain a given rate with a given (low) power is proportional to $-\ddot{\mathcal{I}}(0)$.

For large ρ, in turn, $\textsf{MMSE}_i(\rho)$ with a Gaussian input expands as

$$\textsf{MMSE}(\rho) = \frac{1}{\rho} + \mathcal{O}(1/\rho^2) \tag{31}$$

which determines the corresponding high-power behavior. For discrete constellations, the key feature in such regime is the minimum distance between any two of its points, i.e.,

$$d = \min_{\substack{k,\ell \\ k\neq\ell}} |s_k - s_\ell| \tag{32}$$

which is listed in Table 1 for m-PAM, m-PSK and m-QAM.

Theorem 3 *For BPSK and QPSK,* $\textsf{MMSE}(\rho)$ *admits the series expansion*

$$\textsf{MMSE}(\rho) = \frac{e^{-\frac{d^2}{4}\rho}}{d\sqrt{\rho}} \left(\sqrt{\pi} + \sum_{\ell=1}^{\infty} \frac{b_\ell}{(d^2\rho)^\ell} \right) \tag{33}$$

with $d = 2$ *for BPSK and* $d = \sqrt{2}$ *for QPSK, and with*

$$b_\ell = (-1)^\ell \, \frac{\mathcal{Z}(2\ell+1, 1/4) - \mathcal{Z}(2\ell+1, 3/4)}{\sqrt{\pi}\, 2^{3\ell}} \prod_{q=1}^{\ell} (2q-1) \tag{34}$$

where $\mathcal{Z}(\cdot,\cdot)$ *is the generalized Riemann Zeta function*

$$\mathcal{Z}(q,\xi) = \sum_{k=0}^{\infty} (k+\xi)^{-q}. \tag{35}$$

Proof: *See Appendix B.*

Theorem 3 is illustrated in Fig. 4, which matches the high-power expansion in (33), truncated at $\ell = 1$, against the exact $\textsf{MMSE}(\cdot)$ for QPSK. Also shown is the corresponding low-power expansion in (29).

From Theorem 3, expansions for the high-power mutual information with BPSK and QPSK constellations can be derived through the fundamental relationship in Theorem 1. For instance, if (33) is truncated as

$$\textsf{MMSE}(\rho) \approx \frac{e^{-\frac{d^2}{4}\rho}}{d\sqrt{\rho/\pi}} \tag{36}$$

Table 1: Minimum distance for unit-variance constellations.

Constellation	d
m-PAM	$2\sqrt{\frac{3}{m^2-1}}$
m-PSK	$2\sin\left(\frac{\pi}{m}\right)$
m-QAM	$\sqrt{\frac{6}{m-1}}$

it follows via integration and application of (81) in Appendix B that, for BPSK and QPSK in the high-power regime,

$$\mathcal{I}(\rho) \approx \log m - \frac{e^{-\frac{d^2}{4}\rho}}{\frac{d}{2}\sqrt{\rho/\pi}}. \tag{37}$$

Beyond BPSK and QPSK, the exponential decay of $\text{MMSE}(\cdot)$ in Theorem 3 can be seen to extend to any m-ary constellation.

Theorem 4 *For large ρ, the function $\text{MMSE}(\rho)$ corresponding to an m-ary constellation with minimum distance d behaves as*

$$\text{MMSE}(\rho) = K(\rho)\, e^{-\frac{d^2}{4}\rho} \tag{38}$$

with

$$K_1(\rho) < K(\rho) < K_2 \tag{39}$$

where $K_1(\rho) = \mathcal{O}(1/\sqrt{\rho})$ and K_2 is a constant.

Proof: *See Appendix C.*

B Continuous Distributions

It is sometimes analytically convenient to approximate discrete constellations using continuous distributions over a suitable region on the complex plane (see, e.g., [39, 40, 41]). For such continuous distributions, asymptotic notions such as the *shaping gain* [39] can then be meaningfully defined. Furthermore, a continuous distribution can be interpreted as the limit of a dense m-ary constellation as $m \to \infty$ (and, correspondingly, $d \to 0$).

For equiprobable m-PSK, m-PAM and square m-QAM, the appropriate unit-variance continuous approximations are:

- ∞-PSK: $S = e^{j\phi}$ with ϕ uniform on $[0, 2\pi]$.
- ∞-PAM: Uniform distribution over the segment $[-\sqrt{3}, \sqrt{3}]$ on the real axis.

- ∞-QAM: Uniform distribution over the square $[-\sqrt{3/2}, \sqrt{3/2}] \times [-\sqrt{3/2}, \sqrt{3/2}]$ on the complex plane.

Note that, for ∞-PAM and ∞-QAM, uniform distributions are not optimal for the respective supports.

For small ρ, $\textsf{mmse}(\rho)$ corresponding to ∞-PSK and ∞-QAM distributions behaves as (29). In the case of ∞-PAM, it behaves as in BPSK.

For large ρ, in turn, the behavior of $\textsf{mmse}(\rho)$ is characterized by the following results.

Theorem 5 *For large ρ, the function $\textsf{mmse}(\rho)$ corresponding to a ∞-PSK distribution behaves as*

$$\textsf{mmse}(\rho) = \frac{1}{2\rho} + \mathcal{O}(1/\rho^2) \tag{40}$$

Proof: *See Appendix D.*

Theorem 6 *For large ρ, the function $\textsf{mmse}(\rho)$ corresponding to a uniform ∞-PAM distribution behaves as*

$$\textsf{mmse}(\rho) = \frac{1}{2\rho} + \mathcal{O}(1/\rho^{3/2}) \tag{41}$$

Proof: *See Appendix E.*

Since a uniform ∞-QAM can be expressed as the cartesian product of two uniform ∞-PAM distributions in quadrature, each with half the power, the function $\textsf{mmse}(\cdot)$ of the former follows trivially from that of the latter.

Corollary 2 *For large ρ, the function $\textsf{mmse}(\rho)$ corresponding to a uniform ∞-QAM distribution behaves as*

$$\textsf{mmse}(\rho) = \frac{1}{\rho} + \mathcal{O}(1/\rho^{3/2}) \tag{42}$$

Note that, for all the above continuous distributions, the high-power $\textsf{mmse}(\rho)$ decays reciprocally in ρ (like in the Gaussian case) rather than exponentially on ρ as in the discrete m-ary case. As will be shown in Section VII, this has direct implications in terms of the optimum power allocation.

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 $[0, n]$ (cf. Figs. 5–7). Every power allocation that satisfies (11) and (12) corresponds to drawing a horizontal line whose interception with each of the functions directly gives the power allocated to the corresponding channel. For the allocation to be optimal, this line must be at an ordinate η such that the average of the allocated powers equals 1 as required by (13). If the intercepting 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. 5,* $\eta = 0.32$ *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. 6,* $\eta = 0.23$ *yields* $p_1^\star = 2$ *and* $p_2^\star = 0$.

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 the main chart of Fig. 7,* $\eta = 0.23$ *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.*

Note that, as long as the input distributions remain unchanged, the charts simply scale if the average input power varies. Specifically, if the set $\{\gamma_i\}_{i=1}^n$ is multiplied by a factor greater than one, the new chart is essentially found by zooming in on the original one with a rescaled horizontal axis that continues to span $[0, n]$. (The scaling of the vertical axis affects the value of η but is inconsequential in terms of $\{p_i^\star\}_{i=1}^n$). If each γ_i is varied by a distinct factor, then constructing the new chart requires scaling the various curves differently.

Example 4 *Consider the same scenario of Example 3 except with* $\gamma_1 = \gamma_2 = \gamma_3 = 1/3$, *a sixth-fold decrease in average power. As shown in the inset of Fig. 7, the chart coincides (except for a vertical scaling) with the leftmost sixth of the chart for Example 3.*

B Mercury/waterfilling

While the charts used in the foregoing examples are quite useful, in this subsection we put forth 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 $\text{MMSE}_i^{-1}(\cdot)$ to their special case, given in (17), for a Gaussian input distribution. Let us define, for any arbitrary input distribution,

$$G_i(\zeta) = \begin{cases} 1/\zeta - \text{MMSE}_i^{-1}(\zeta) & \zeta \in [0,1] \\ 1 & \zeta > 1 \end{cases} \tag{43}$$

such that, for a Gaussian input, $G_i(\zeta) = 1$ for all ζ. It can be verified that $G_i(\cdot)$ is a monotonically decreasing function.

The function $G_i(\cdot)$ enables the following interpretation of our power allocation policy, which we refer to as *mercury/waterfilling* (cf. Fig. 8).

(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,[6] 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 gives $p_i^\star$.

The mercury/waterfilling is a parametric procedure, leading to optimal mutual information and average power that are continuous monotonically decreasing functions of η. For a given value of η, the mutual information is maximized using exactly all of the available power P for which (15) is satisfied and thus the water heights give the corresponding $\{p_i^\star\}_{i=1}^n$.

The mercury pouring stage regulates the water admitted by each vessel thus tailoring the process to arbitrary input distributions. Pouring mercury onto a vessel amounts to raising the noise level in that channel by an amount that depends on the input distribution. No mercury is poured onto vessels whose input is Gaussian or for which $\gamma_i < \eta$. On the ith channel, the increase in noise level due to the mercury equals $G_i(\eta/\gamma_i)|_{\text{d}B}$. Note that this increase 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.

---

[6] Note that the outlined mercury/waterfilling procedure requires that the vessels be non-porous with respect to mercury and porous with respect to water. Fine glass grit is one such material.

In comparison to the existing heuristic methods mentioned in Section I, the mercury can be seen as playing the role of the *gap* with respect to an ideal Gaussian signal, but with the advantage that it is exact for any input power and constellations (as opposed to only approximate for a class of constellations).

In a real-time implementation, $\text{MMSE}_i^{-1}(\cdot)$ can be simply tabulated for every constellation of interest and hence the problem essentially boils down to solving (15). This can be accomplished through established iterative methods (see Appendix F for details).

VI Low-Power Regime

The behavior of the waterfilling policy in the region of low power is easily inferred from (29). If the largest and second-largest entries in the set $\{\gamma_i\}_{i=1}^n$ are dissimilar, it is asymptotically optimal (as $P \to 0$) to allocate power only to the strongest channel. If $\max\{\gamma_i\}$ is plural, then the power should be equally divided among the corresponding channels [38].

It is known from [38, Thm 14] that the first two derivatives of the mutual information at zero power achieved by any proper complex distribution equal those achieved by a Gaussian input with the same power. Thus, the first two derivatives of the mutual information (and, consequently, the minimum energy per bit and the wideband slope) with waterfilling are identical to those with the optimum mercury/waterfilling policy as long as all of the inputs conform to proper complex distributions. This holds for both discrete and continuous input distributions.

For inputs that are not proper complex, mercury/waterfilling and waterfilling are no longer equivalent for $P \to 0$. Specializing $G(\cdot)$ via (30), mercury/waterfilling behaves as follows in the low-power regime:

- Allocating power only to the strongest channel continues to be asymptotically optimal.
- If $\max\{\gamma_i\}$ is plural, then the power should be divided among the corresponding channels in inverse proportion to the second derivative at zero of the mutual information for the respective input distributions, i.e.,

  $$p_i^\star = \frac{\alpha}{\ddot{\mathcal{I}}_i(0)} \tag{44}$$

  with the proportionality constant

  $$\alpha = \frac{1}{\frac{1}{n}\sum_{\ell=1}^{n} 1/\ddot{\mathcal{I}}_\ell(0)} \tag{45}$$

Gross missallocations of power may result if waterfilling is used, without pouring mercury, with inputs that are not proper complex.

Example 5 *Consider* $n = 2$ *channels with* $|h_1|^2 = |h_2|^2 = |h|^2$*, fed respectively with BPSK and QPSK inputs. From (44),* $p_1^\star = 2/3$ *and* $p_2^\star = 4/3$ *resulting in*

$$\frac{\mathcal{I}_1^* + \mathcal{I}_2^*}{2} = |h|^2 P - \frac{10}{9}|h|^4 P^2 + \mathcal{O}(P^3). \tag{46}$$

In contrast, waterfilling yields $p_1^{\mathsf{WF}} = p_2^{\mathsf{WF}} = 1$ *and*

$$\frac{\mathcal{I}_1^{\mathsf{WF}} + \mathcal{I}_2^{\mathsf{WF}}}{2} = |h|^2 P - \frac{3}{4}|h|^4 P^2 + \mathcal{O}(P^3). \tag{47}$$

The bandwidth B^{WF} *required to achieve a certain rate (in bits/s) with a certain average power using waterfilling, relative to the bandwidth* B^* *required with mercury/waterfilling, approaches (as* $P \to 0$*)*

$$\frac{B^{\mathsf{WF}}}{B^*} = \frac{\ddot{\mathcal{I}}^*(0)}{\ddot{\mathcal{I}}^{\mathsf{WF}}(0)} = 1 + \frac{13}{27} \tag{48}$$

indicating that, in this example, waterfilling requires about 48% *excess bandwidth.*

VII High-Power Regime

From (31), the optimum power allocation with Gaussian inputs for $P \to \infty$ behaves as

$$p_i^{\mathsf{WF}} = 1 + \mathcal{O}(1/P). \tag{49}$$

Hence the well-known feature that, to zero-order, waterfilling is tantamount to a uniform allocation in the high-power regime. From Theorems 5–6 and Corollary 2, the same behavior can be seen to extend to continuous ∞-PSK, ∞-PAM and ∞-QAM distributions. For discrete m-ary constellations, however, the limiting allocation at high power is drastically different.

Theorem 7 *If the inputs conform to discrete constellations with minimum distances* $\{d_i\}_{i=1}^n$*, the power allocation behaves (for* $P \to \infty$*) as*

$$p_i^\star = \frac{\alpha}{|h_i|^2 d_i^2} + \mathcal{O}(\frac{\log P}{P}) \tag{50}$$

with

$$\frac{1}{\alpha} = \frac{1}{n}\sum_{\ell=1}^{n} \frac{1}{|h_\ell|^2 d_\ell^2}. \tag{51}$$

Proof: *See Appendix G.*

Theorem 7 prompts two immediate observations concerning discrete constellations in the high-power regime:

- Given equal constellations: the stronger a channel, the less power it is allocated. This is in stark contrast with waterfilling.
- Given equal channel strengths: the richer the constellation, the more the allocated power.

Corollary 3 *If all the inputs conform to the same discrete constellation, power equalization is asymptotically optimal in the high-power regime, i.e.,*

$$p_i^\star = \frac{\alpha}{|h_i|^2} + \mathcal{O}\left(\tfrac{\log P}{P}\right) \tag{52}$$

with

$$\frac{1}{\alpha} = \frac{1}{n} \sum_{\ell=1}^{n} \frac{1}{|h_\ell|^2}. \tag{53}$$

In addition to the limiting power allocation given by Theorem 7, for BPSK and/or QPSK inputs we can also use the high-power expansions of $\textsc{mmse}(\cdot)$ in Theorem 3 in order to approximate the mercury/waterfilling power allocation to any chosen order.

Example 6 *Let $n = 2$ with $|h_2|^2 = 2|h_1|^2$ and with QPSK inputs on both channels. Plugging (36) into Theorem 2, we can approximate $p_1^\star$ and $p_2^\star$ within the high-power regime by solving*

$$|h_i| \frac{e^{-p_i^\star |h_i|^2 P}}{\sqrt{p_i^\star}} = \eta \qquad\qquad i = 1, 2 \tag{54}$$

subject to $p_1^\star + p_2^\star = 2$. Shown in Fig. 9 are the exact $p_1^\star$ and $p_2^\star$ obtained from mercury/waterfilling as well as the solutions to (54). For $P \to \infty$, the power allocation converges to the power-equalizing solution $p_1^\star = 4/3$ and $p_2^\star = 2/3$.

Although, because of the flatness of the mutual information function for any m-ary constellation in the high-power regime, the loss in mutual information with suboptimum power allocations is small, the increase in power required for a certain mutual information can be sizeable.

Example 7 *Consider n channels fed by BPSK and/or QPSK inputs and denote by P^{eq} and P^{unif} the powers required to sustain a mutual information $\mathcal{I}$ with the power allocation in (50) and with a uniform power allocation, respectively. The ratio $P^{\mathsf{eq}}/P^{\mathsf{unif}}$ converges (as $P \to \infty$ and thus $\mathcal{I} \to \frac{1}{n}\sum_{i=1}^{n} \log m_i$) to*

$$\frac{P^{\mathsf{eq}}}{P^{\mathsf{unif}}} = \frac{\min\{|h_i|^2 d_i^2\}}{n} \sum_{\ell=1}^{n} \frac{1}{|h_\ell|^2 d_\ell^2} \tag{55}$$

which can represent as much as an n-fold power penalty.

Clearly, the cardinality of the constellation critically determines the point where the high-power behavior comes into play. The smaller the constellation, the less power required to bring it about.

VIII Application: Multitone for DSL

The qualitative importance of the mercury/waterfilling policy rests on its optimality when the mutual information is the driving performance measure. In order to calibrate its quantitative relevance, it is interesting to assess the shortfall in mutual information when conventional waterfilling is applied in conjunction with constellations of common usage.

The simplest of comparisons, with $n = 2$ channels fed with parallel BPSK and parallel QPSK inputs, are presented in Figs. 10 and 11, respectively. Specifically, the figures depict the excess power required by waterfilling to achieve the same mutual information as mercury/waterfilling as a function of the channel strengths $\{\gamma_1, \gamma_2\}$. For strengths ranging from -4 dB to 8 dB, the excess power is over 5 dB with BPSK inputs and over 3 dB with QPSK inputs. The excess power required by waterfilling grows without bound as the channel strengths become more imbalanced.

More elaborate assessments require postulating specific channel profiles. To that end, in the remainder of this section we shall use a typical wireline channel that complies with one of the canonical loops employed for DSL testing [42]. Its frequency response, sampled into $n = 256$ orthogonal tones and scaled by their respective noise variances to give $\{|h_i|^2\}_{i=1}^{n}$, is depicted in Fig. 12. The total bandwidth spanned is slight over 1 MHz. The channel response accounts for the presence of far-end crosstalkers, modelled as additional Gaussian noise. We let the input signals conform to square m-QAM constellations, possibly distinct on each tone. The results are parameterized by some maximum constellation cardinality that does not exceed 256-QAM. This limit will help illustrate the role of the maximum constellation cardinality on the suboptimality of conventional waterfilling.

A Uniform Loading

For maximum mutual information, the richest available constellation should be loaded on each of the tones. With such uniform loading, a first evaluation of the relative performances of mercury/waterfilling and waterfilling can be conducted.

For a wide range of power constraints $P|_{\text{dB}}$, the excess power required for waterfilling

Table 2: With m-QAM uniform loading, excess power required by waterfilling (respectively on/off) at the mutual informations $\frac{1}{2}\log m$, $\frac{3}{4}\log m$ and $\frac{9}{10}\log m$.

m	$\frac{1}{n}\sum_i \mathcal{I}_i = \frac{1}{2}\log m$	$\frac{1}{n}\sum_i \mathcal{I}_i = \frac{3}{4}\log m$	$\frac{1}{n}\sum_i \mathcal{I}_i = \frac{9}{10}\log m$
16	2.6 dB (resp. 2 dB)	4.4 dB (resp. 3.7 dB)	5 dB (resp. 4.1 dB)
64	1.5 dB (resp. 1.3 dB)	2.8 dB (resp. 2.1 dB)	3.6 dB (resp. 3.5 dB)
256	0.8 dB (resp. 0.6 dB)	1.8 dB (resp. 1.6 dB)	3.2 dB (resp. 3.2 dB)

to meet the mutual information achieved by mercury/waterfilling is depicted in Fig. 13, parameterized by the constellation. We note that:

- The smaller the constellation, the higher the excess power at all values of P. This reflects the fact that smaller m-QAM constellations are a further departure from the Gaussian signal idealization to which conventional waterfilling is tailored.
- Regardless of the constellation, the excess power vanishes for $P \to 0$.
- For large P, as $\frac{1}{n}\sum_i \mathcal{I}_i \to \log m$, excess powers of several dB are observed. The excess powers at the representative operating points where the mutual information equals $1/2$, $3/4$ and $9/10$ its maximum value are summarized in Table 2.

For the range of $P|_{\mathrm{dB}}$ where the power offsets are substantial, the mutual informations achieved by both mercury/waterfilling and waterfilling are displayed in Fig. 14. The horizontal displacement between the curves corresponding to each constellation equals the respective excess power.

In a similar fashion, we can gauge the excess power incurred over mercury/waterfilling by a common variant of waterfilling: an on/off policy activating the same subset of tones but with a uniform power allocation thereon, i.e. [25, 43, 44]

$$p_i^{\mathsf{On/Off}} = 0 \qquad \gamma_i < \eta \tag{56}$$

$$p_i^{\mathsf{On/Off}} = \bar{p} \qquad \gamma_i \geq \eta \tag{57}$$

where η is as in the waterfilling case and $\bar{p}$ is a suitable constant. The excess powers incurred by the on/off allocation, also summarized for a uniform constellation loading in Table 2, equal their waterfilling counterparts for $P \to 0$ and $P \to \infty$ but are somewhat reduced at intermediate points.

B Levin-Campello Loading

Although the mutual information is strictly maximized by loading the richest available constellation onto every tone, there is little loss in using constellations of smaller cardi-

Table 3: With Levin-Campello loading, excess power required by waterfilling (respectively on/off) at the mutual informations $\frac{1}{2}\log m$, $\frac{3}{4}\log m$ and $\frac{9}{10}\log m$ where m is the cardinality of the richest available constellation.

m	$\frac{1}{n}\sum_i \mathcal{I}_i = \frac{1}{2}\log m$	$\frac{1}{n}\sum_i \mathcal{I}_i = \frac{3}{4}\log m$	$\frac{1}{n}\sum_i \mathcal{I}_i = \frac{9}{10}\log m$
16	2.9 dB (resp. 2.6 dB)	4.5 dB (resp. 4 dB)	4.8 dB (resp. 4.6 dB)
64	1.5 dB (resp. 1.3 dB)	3 dB (resp. 2.7 dB)	4 dB (resp. 4 dB)
256	1 dB (resp. 0.9 dB)	2.1 dB (resp. 2.1 dB)	3.5 dB (resp. 3.5 dB)

nality on those tones with reduced γ_i. Thus, by sensibly loading the tones, the complexity (in the sense of the constellation cardinalities being used) can be eased considerably in exchange for only a slight penalty in mutual information. For this purpose, we invoke the Levin-Campello algorithm (cf. Appendix H) that is known to optimally solve the problem of bit loading with discrete constellations at a target error probability. The comparisons with waterfilling and with on/off conducted in the previous section are herein repeated but with the constellations being loaded via Levin-Campello. An example of this loading, for $P|_{\text{dB}} = -15$ dB (corresponding to a mutual information of 4.1 bits/s/Hz when mercury/waterfilling is used), can be found in Fig. 15. Also shown in the same figure are the corresponding mercury/waterfilling and waterfilling power allocations, from which we remark the following:

- In the segment of tones exhibiting high signal-to-noise ratios (i.e, those with low indices), the power allocations evidence their markedly different behaviors. Mercury/waterfilling essentially inverts the channel response in Fig. 12 while waterfilling is close to being uniform.
- Mercury/waterfilling recognizes the specificity of each constellation, as evidenced by the sharp disparities in the powers assigned to tones with similar γ_i but different constellations. In contrast, waterfilling responds solely to γ_i.

For a wide range of power constraints, each loaded via Levin-Campello, the excess power required by waterfilling is depicted in Fig. 16, parameterized by the set of constellations available for loading. By and large, the excess powers are analogous to those encountered with a uniform loading.[7] The excess powers are summed up in Table 3, again for the representative operating points where the mutual information equals $1/2$, $3/4$ and $9/10$ its maximum value. Also listed in Table 3 are the excess powers with on/off, which are slightly reduced.

---

[7] The nonmonotonicities in Fig. 16, as well as in Fig. 13, are illustrative of the behavior with large dynamic range variations in the channel frequency response. They are due to the different power constraints that trigger the use of low-power channels in waterfilling and mercury/waterfilling.

Notice that, although the loading obtained via Levin-Campello is optimum given a specific coding scheme and a target error probability, it does not in general maximize the mutual information when the cardinalities of the constellations are somehow constrained. One could envision, for instance, constraining the sum of the cardinalities on the n tones or the fraction of tones on which a certain constellation is used. Mercury/waterfilling would be a building block in solving the optimum loading under any such constraint.

IX Application: Power Control for Fading Channels

Another application of the mercury/waterfilling optimization is that of power control for a fading channel known by the transmitter, which can be viewed as an infinite sequence of parallel channels each encompassing either a symbol or a block of symbols whose fading coefficients are (approximately) equal.

A Ergodic Channel with Flat Fading

Let us first consider an ergodic channel, where the average power constraint applies over a time interval long enough for the fading distribution to be revealed. For a Gaussian input, the corresponding mutual information is maximized by a temporal waterfilling with a water level that depends only on the average power and the fading distribution [5]. For uncoded transmission with m-ary constellations, waterfilling can be used in combination with power gaps to approximate the power control policy that maximizes the throughput at some target error probability [45]. We next formalize the application of mercury/waterfilling to obtain a power control policy that maximizes the mutual information with any arbitrary input distribution.

Corollary 4 *Consider the fading channel*

$$Y_i = \sqrt{\gamma_i \, p(\gamma_i)}\, S_i + W_i \tag{58}$$

where $\{\gamma_i\}$ is a stationary and ergodic random process, the input has an arbitrary (but fixed) unit-variance distribution, and the dependence of $p(\gamma)$ on γ emphasizes the ability of the transmitter to control the power based on the fading state with the constraint that $E[p(\gamma)] = 1$. The ergodic mutual information is maximized by the mercury/waterfilling power control policy

$$p^\star(\gamma) = \frac{1}{\gamma} \, \mathsf{MMSE}^{-1}\left(\min\left\{ 1, \frac{\eta}{\gamma} \right\} \right) \tag{59}$$

with η satisfying

$$\int_\eta^\infty \frac{\mathsf{MMSE}^{-1}\left(\eta/\xi\right)}{\xi} f_\gamma(\xi)\, d\xi = 1 \tag{60}$$

where $f_\gamma(\cdot)$ denotes the probability density function of the fading.

For any fading distribution of interest, the value of η that satisfies (60) can be precomputed and tabulated for different power constraints, after which the actual power control would simply entail the mapping in (59) via the function $\text{MMSE}^{-1}(\cdot)$ for the constellation in use.

In the case that the input is Gaussian, not only does Corollary 4 default to a temporal waterfilling, but under some fading laws it is further possible to obtain a more explicit fixed-point equation for η. Specifically, if the fading is Rayleigh-distributed with $E[\gamma] = P$, then

$$f_\gamma(\xi) = \frac{e^{-\gamma/P}}{P} \tag{61}$$

and

$$\begin{aligned} p^{\mathsf{WF}}(\gamma) &= 0 & \gamma \leq \eta \qquad (62)\\ p^{\mathsf{WF}}(\gamma) &= \frac{1}{\eta} - \frac{1}{\gamma} & \gamma > \eta \qquad (63) \end{aligned}$$

with η solution to [46]

$$\frac{e^{-\eta/P}}{\eta} - \frac{E_1(\eta/P)}{P} = 1 \tag{64}$$

where $E_1(\zeta) = \int_1^\infty t^{-1} e^{-\zeta t} dt$ is an exponential integral.

Example 8 *Let the channel be Rayleigh-faded. The excess power required, without power control, to achieve the same mutual information attained with temporal mercury/waterfilling is depicted in Fig. 17 for several input distributions (QPSK, 16-QAM and Gaussian).*

For $P \to 0$, the excess power grows without bound for all of the distributions. This is a direct consequence of the well-known fact that the minimum energy per bit, which without power control is -1.59 dB over the noise floor at the receiver, vanishes with temporal waterfilling [47]. Also well understood is that, for $P \to \infty$, the excess power offset wanes when the input is Gaussian. With discrete constellations, in contrast, the excess power rises rapidly and without bound once the mutual information nears its maximum value. For QPSK and 16-QAM specifically, the absence of power control is least suboptimal at $P|_{\text{dB}} \approx 4$ and $P|_{\text{dB}} \approx 10$, respectively.

Example 9 *Let the fading law be log-normal with a standard deviation $\sigma|_{\text{dB}} = 10$, for which*

$$f_\gamma(\xi) = \frac{e|_{\text{dB}}}{\sqrt{2\pi}\,\sigma|_{\text{dB}}\,\xi}\, e^{-\frac{(\xi|_{\text{dB}} - P|_{\text{dB}})^2}{2\,\sigma|^2_{\text{dB}}}}. \tag{65}$$

The excess power required, without power control, to meet the same mutual information achieved with temporal mercury/waterfilling is depicted in Fig. 17 for several input distributions (QPSK, 16-QAM and Gaussian).

Although the limiting behaviors are similar, the excess powers are magnified by the heavy tail of the log-normal fading law. Also, the operating points where not controlling the power is least suboptimal increase significantly.

B Non-Ergodic Block-Fading Channel with Flat Fading

If the time interval over which the power can be averaged is limited, such that the channel variations thereon do not fully reveal the fading distribution, the ergodic framework is no longer valid. A useful set-up in this case is that of a block-fading channel, where the power can be averaged over n blocks each having a fading coefficient that holds constant for the duration of the block and changes independently between blocks [48, 49]. In the limit of a large number of symbols per block, which renders the mutual information of each block operationally significant, the problem of allocating power across the n blocks to minimize the outage probability can be stated as

$$[p_1^\star, \ldots, p_n^\star] = \arg \min_{\substack{p_1,\ldots,p_n \\ \frac{1}{n}\sum_i p_i \leq 1}} \mathcal{P}r \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i(p_i \gamma_i) \leq R \right\} \tag{66}$$

where R is a target spectral efficiency and $\{\gamma_i\}_{i=1}^n$ are known to the transmitter. Note that, in this formulation, not only are the input distributions allowed to be arbitrary but also possibly different on each of the n blocks.

It can be shown that the solution to (66) has the form of separate mercury/waterfilling procedures on each of the n blocks. When the input is Gaussian on each of the blocks, the solution naturally reverts to individual waterfillings on each of the n blocks [49].

X Conclusions

The derivation and formulation of the maximal mutual information power allocation policy for independent Gaussian-noise channels driven by arbitrary (not necessarily identical) input constellations hinge on the nonlinear minimum mean-square error of estimating the point in the constellation given a noisy version. Thus, the solution in this paper provides another application for the nexus between mutual information and MMSE of Gaussian channels found in [32].

We have seen that mercury/waterfilling, a geometric representation of the optimum policy that generalizes the classical solution for ideal Gaussian inputs, leads to valuable insights.

As long as the input constellations are proper complex (i.e., quadrature symmetric), they are almost as good as Gaussian inputs in the low-power region and thus waterfilling is almost optimum for proper complex inputs in that region.

In contrast, in the moderate-to-high power regions, waterfilling and mercury/waterfilling can lead to markedly different policies. The principle that stronger channels should be allocated more power, an ingrained belief stemming from the waterfilling policy, may not hold with constellations of practical interest. In fact, in some circumstances the mercury/waterfilling policy mandates exactly the opposite. Since the mutual information of an m-ary constellation cannot exceed $\log m$, there is little incentive to allocate further power to a channel once it nears such point. Rather, the additional power is better allocated to a weaker channel still far from saturation. For identical constellations in the high-power regime, we have found that stronger channels are always allocated strictly less transmit power so as to ensure that the received powers on the various channels are equalized.

In order to quantify the significance of using mercury/waterfilling rather than suboptimal policies (such as conventional waterfilling or an on/off variant thereof), we have analyzed the setting of multitone communication over a canonical DSL channel. The suboptimality of those power allocation policies when applied with m-QAM constellations was seen to depend critically on the interplay of:

- The dynamic range spanned by $\{\gamma_i\}_{i=1}^{n}$. For given constellations and operating point, a larger dynamic range is likely to heighten the deficit.
- The cardinality of the constellations. On a given channel, richer constellations stretch the range of operating points on which the suboptimal policies are effective. Interestingly, the exact loading appears to have minimal impact as long as it is sensible, i.e., as long as the richest available constellation is loaded onto all those tones that are strong.
- The power constraint. Suboptimal policies are very effective if the power constraint is sufficiently low. Conversely, they tend to incur substantial excess powers at operating points that are sufficiently close to the maximum achievable mutual information. Note that, while DSL systems are designed for a wide range of power levels, the maximum cardinality of the constellations is limited by implementation considerations. Thus, high-power channels (corresponding to subscribers near the central office) may benefit from the mercury/waterfilling power allocation by maintaining the same rates with lower power expenditure (and thus, crosstalk) than that required by the conventional waterfilling.

A similar quantitative evaluation for multitone communication on a frequency-selective wireless channel can be found in [50].

Yet another distinct application, power control for fading channels, has also been treated in detail. The advantage of using mercury/waterfilling is found to depend on the interplay between the fading distribution, the choice of constellation and the power constraint.

Appendices

A Proof of Theorem 2

Define

$$\begin{aligned}
\mathbf{s} &= [S_1 \dots S_n]^T \\
\mathbf{y} &= [Y_1 \dots Y_n]^T \\
\mathbf{G} &= \mathrm{diag}\{\sqrt{\gamma_1}, \dots, \sqrt{\gamma_n}\} \\
\mathbf{P} &= \mathrm{diag}\{\sqrt{p_1}, \dots, \sqrt{p_n}\}
\end{aligned}$$

such that (1) can be written in the equivalent vector form

$$\mathbf{y} = \mathbf{GPs} + \mathbf{w} \tag{67}$$

with $\mathrm{Tr}\{\mathbf{P}\}/n \leq 1$ and $E[\mathbf{ss}^\dagger] = \mathbf{I}$. The vector $\mathbf{w}$ has complex Gaussian entries and covariance $E\left[\mathbf{ww}^\dagger\right] = \mathbf{I}$. The input-output mutual information of the vector channel in (67) is $\frac{1}{n}\mathcal{I}(\mathbf{P})$ with

$$\mathcal{I}(\mathbf{P}) = \sum_{i=1}^{n} \mathcal{I}_i(p_i \gamma_i) \tag{68}$$

Because of its monotonicity with $\mathrm{Tr}\{\mathbf{P}\}/n$, the mutual information is maximized when the power constraint is met with strict equality. In addition, the function $\frac{1}{n}\mathcal{I}(\cdot)$ is strictly concave over the set of feasible power allocations since it represents the average of the mutual information functions on the n scalar parallel channels, each of which is strictly[8] concave. The strict concavity of the mutual information of a scalar channel with respect to its input power follows from the strict monotonicity of the corresponding MMSE$(\cdot)$ in combination with Theorem 1. Consequently, there is a unique $\mathbf{P}^\star = \mathrm{diag}\{p_1^\star, \dots, p_n^\star\}$ such that $\mathcal{I}(\mathbf{P}^\star) > \mathcal{I}(\mathbf{P})$ for every $\mathbf{P} \neq \mathbf{P}^\star$. At $\mathbf{P}^\star$, the derivative of $\mathcal{I}(\cdot)$ in the direction from $\mathbf{P}^\star$ to any other power allocation $\mathbf{P}$ must be negative, i.e.,

$$\frac{d}{d\mu} \mathcal{I}\left(\mu\mathbf{P} + (1-\mu)\mathbf{P}^\star\right)\Big|_{\mu=0} < 0 \tag{69}$$

---

[8] Recall that $|h_i| > 0$ for every i and, therefore, $\gamma_i > 0$ strictly.

which can be rewritten as [51, p. 213]

$$
\begin{aligned}
\frac{d}{d\mu}\,\mathcal{I}\left(\mathbf{P}^\star + \mu(\mathbf{P}-\mathbf{P}^\star)\right)\big|_{\mu=0} &= \sum_{\ell=1}^{n}(p_\ell - p_\ell^\star)\,\frac{\partial \mathcal{I}(\mathbf{P})}{\partial p_\ell}\big|_{p_\ell=p_\ell^\star} && (70)\\
&\leq 0 && (71)
\end{aligned}
$$

Since the right-hand side of (70) is affine on $\mathbf{P}$, it suffices to impose (71) on the n extreme points of the set. The ith such point is defined by $p_i = n$ and $p_\ell = 0,\ \ell \neq i$. The line connecting that point with $\mathbf{P}^\star$ can be extended beyond $\mathbf{P}^\star$ if and only if $p_i^\star > 0$, in which case the corresponding directional derivative at $\mathbf{P}^\star$ vanishes and (71) becomes a strict equality. Conversely, if $p_i^\star = 0$ then (71) remains an inequality. Altogether,

$$
\frac{\partial \mathcal{I}(\mathbf{P})}{\partial p_i}\big|_{p_i=p_i^\star} \leq \frac{1}{n}\sum_{\ell=1}^{n} p_\ell^\star \frac{\partial \mathcal{I}(\mathbf{P})}{\partial p_\ell}\big|_{p_\ell=p_\ell^\star} \tag{72}
$$

with strict equality if $p_i^\star > 0$. Note that, in Theorem 1, identity (10) holds even if the roles of p_i and γ_i are interchanged. Thus,

$$
\frac{\partial \mathcal{I}(\mathbf{P})}{\partial p_i} = \gamma_i \, \mathsf{MMSE}_i(p_i \gamma_i) \tag{73}
$$

where, in (73), we have used (3). Plugging (73) into (72),

$$
\begin{aligned}
\gamma_i &\leq \eta && p_i^\star = 0 && (74)\\
\gamma_i \, \mathsf{MMSE}_i(p_i^\star \gamma_i) &= \eta && p_i^\star > 0 && (75)
\end{aligned}
$$

where we have used $\mathsf{MMSE}_i(0) = 1$ and introduced

$$
\eta = \frac{1}{n}\sum_{\ell=1}^{n} p_\ell^\star \, \gamma_\ell \, \mathsf{MMSE}_\ell(p_\ell^\star \gamma_\ell). \tag{76}
$$

The optimum power allocation is now uniquely identified through (74)–(76), which are equivalent to (11)–(13) or, alternatively, to (14) and (15).

B Proof of Theorem 3

From (25) and (26) we find that, for both BPSK and QPSK,

$$
\mathsf{MMSE}(\rho) = 1 - \int_{-\infty}^{\infty} \tanh\left(d\sqrt{\rho}\,\xi\right) \frac{e^{-(\xi - \frac{d}{2}\sqrt{\rho})^2}}{\sqrt{\pi}} d\xi \tag{77}
$$

with d equal to the corresponding minimum distance, i.e., 2 and $\sqrt{2}$ respectively.

Using the asymptotic ($\xi \to \infty$) expansion

$$\tanh(\xi) = \mathrm{sgn}(\xi) \left(1 + 2 \sum_{\ell=1}^{\infty} (-1)^{\ell} e^{-2\ell|\xi|} \right) \tag{78}$$

we obtain, for large ρ,

$$\text{MMSE}(\rho) = 2\, Q\left(d\sqrt{\rho/2}\right) - 2 \sum_{\ell=1}^{\infty} (-1)^{\ell} \, \frac{e^{-\ell d^2 \rho}\, Q\Big((2\ell-1)d\sqrt{\rho/2}\Big) - e^{\ell d^2 \rho}\, Q\Big((2\ell+1)d\sqrt{\rho/2}\Big)}{e^{-\ell^2 d^2 \rho}} \tag{79}$$

where

$$Q(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\xi}^{\infty} e^{-t^2/2} dt \tag{80}$$

which admits the expansion [52, p. 100]

$$Q(\xi) = \frac{e^{-\xi^2/2}}{\sqrt{2\pi}\, \xi} \left(1 + \sum_{\ell=1}^{\infty} (-1)^{\ell} \, \frac{\prod_{q=1}^{\ell} (2q-1)}{\xi^{2\ell}} \right). \tag{81}$$

Combining (79) and (81) with some algebra, the claimed expression is found.

C Proof of Theorem 4

Applying the law of total probability to (9),

$$\text{MMSE}(\rho) = \sum_{\ell=1}^{m} q_{\ell}\, E\left[\left| s_{\ell} - \hat{S}(Y, \rho) \right|^2 \mid S = s_{\ell} \right]. \tag{82}$$

The first step in the proof is the derivation of a lower bound for $\text{MMSE}(\rho)$ in the large-ρ regime. To that end, we can postulate the existence of a genie that informs the estimator that it should only consider the points s_{ℓ} and s_u equiprobably, where the former is the actual transmitted point and the latter is its closest neighbor situated at a distance d_{ℓ}. (If there is a multiplicity of closest neighbors, the genie selects any of them as s_u.) The genie-aided estimate is thus

$$\hat{S}_{\mathrm{g}}(y, \rho) = \frac{s_{\ell}\, e^{-|y - \sqrt{\rho} s_{\ell}|^2} + s_u\, e^{-|y - \sqrt{\rho} s_u|^2}}{e^{-|y - \sqrt{\rho} s_{\ell}|^2} + e^{-|y - \sqrt{\rho} s_u|^2}} \tag{83}$$

whose mean-square error is lower than that of the actual estimator and hence

$$\begin{aligned} \mathsf{MMSE}(\rho) &\geq \sum_{\ell=1}^{m} q_\ell \, E\left[\left| s_\ell - \hat{S}_{\mathsf{g}}(Y,\rho) \right|^2 \, | \, Y = \sqrt{\rho} s_\ell + W \right] && (84) \\ &= \sum_{\ell=1}^{m} q_\ell \, \frac{d_\ell^2}{4} \, \mathsf{MMSE}^{\mathsf{BPSK}} \left(\frac{d_\ell^2}{4} \rho \right) && (85) \end{aligned}$$

where, by introducing $\mathsf{MMSE}^{\mathsf{BPSK}}(\cdot)$ as defined in (25), we identify the ℓth term in the summation in (84) as the MMSE for a scaled BPSK constellation between two points separated by a distance d_ℓ. Using $d = \min\{d_\ell\}$ and applying Theorem 3 to $\mathsf{MMSE}^{\mathsf{BPSK}}(\cdot)$, we obtain for large ρ the lower bound

$$\mathsf{MMSE}(\rho) \geq K_1(\rho) \, e^{-\frac{d^2}{4}\rho} \tag{86}$$

with

$$K_1(\rho) = \mathcal{O}(1/\sqrt{\rho}). \tag{87}$$

Next, we derive an upper bound, to which end we postulate a suboptimum estimator that outputs the constellation point closest in Euclidean distance to the noisy observation at the receiver, i.e.,

$$\hat{S}_{\mathsf{sub}}(y,\rho) = \arg\min_{s_k} |y - \sqrt{\rho}\, s_k|. \tag{88}$$

We can expand (82) into

$$\begin{aligned} \mathsf{MMSE}(\rho) = \sum_{\ell=1}^{m} q_\ell \Bigg(& E\left[\left| s_\ell - \hat{S}(Y,\rho) \right|^2 \, | \, S = s_\ell \, , \, Y \in \mathcal{V}_\ell \right] \Pr\left\{ Y \in \mathcal{V}_\ell \, | \, S = s_\ell \right\} \\ & + E\left[\left| s_\ell - \hat{S}(Y,\rho) \right|^2 \, | \, S = s_\ell \, , \, Y \notin \mathcal{V}_\ell \right] \Pr\left\{ Y \notin \mathcal{V}_\ell \, | \, S = s_\ell \right\} \Bigg) && (89) \end{aligned}$$

where $\mathcal{V}_\ell$ denotes the Voronoi region of the ℓth constellation point. Applying the suboptimum estimator in (88),

$$\mathsf{MMSE}(\rho) \leq \sum_{\ell=1}^{m} q_\ell \, E\left[\left| s_\ell - \hat{S}_{\mathsf{sub}}(Y,\rho) \right|^2 \, | \, S = s_\ell \, , \, Y \notin \mathcal{V}_\ell \right] \Pr\left\{ Y \notin \mathcal{V}_\ell \, | \, S = s_\ell \right\} \tag{90}$$

where $\hat{S}_{\mathsf{sub}}(Y,\rho) \neq s_\ell$ since Y lies outside $\mathcal{V}_\ell$. Furthermore, the probability that Y lies outside $\mathcal{V}_\ell$ when s_ℓ was transmitted can be bounded by the probability that Y lies outside a circle centered on s_ℓ and completely contained within $\mathcal{V}_\ell$. This circle has a radius that cannot exceed $(d_\ell/2)\sqrt{\rho}$ where d_ℓ is the distance from s_ℓ to its closest neighbor(s) in the constellation. Thus,

$$\begin{aligned} \Pr\left\{ Y \notin \mathcal{V}_\ell \, | \, S = s_\ell \right\} &\leq \frac{1}{\pi} \int_{|\xi| > \frac{d_\ell}{2}\sqrt{\rho}} e^{-|\xi|^2} d\xi && (91) \\ &= e^{-\frac{d_\ell^2}{4}\rho}. && (92) \end{aligned}$$

In turn,

$$E\left[\left|s_\ell - \arg\min_{s_k} |Y - \sqrt{\rho}\, s_k|\right|^2 \mid S = s_\ell\, ,\, Y \notin \mathcal{V}_\ell\right] \leq K_2 \tag{93}$$

with $K_2 = \max\{d_\ell\}$. Combining (90), (92) and (93),

$$\begin{aligned} \textsf{MMSE}(\rho) &\leq K_2 \sum_{\ell=1}^{m} q_\ell\, e^{-\frac{d_\ell^2}{4}\rho} && (94)\\ &\leq K_2\, e^{-\frac{d^2}{4}\rho} \sum_{\ell=1}^{m} q_\ell && (95)\\ &= K_2\, e^{-\frac{d^2}{4}\rho} && (96) \end{aligned}$$

where $d = \min\{d_\ell\}$.

Relating the lower and upper bounds in (86) and (96), the claim is proved.

D Proof of Theorem 5

From (8), the MMSE estimate of an ∞-PSK input S given its noisy output observation y is

$$\hat{S}(y,\rho) = e^{j\arg(y)} \frac{I_1(2\sqrt{\rho}|y|)}{I_0(2\sqrt{\rho}|y|)} \tag{97}$$

where $I_0(\cdot)$ and $I_1(\cdot)$ are modified Bessel functions of the first kind [53]. These functions are related by

$$I_1(\xi) = \frac{d}{d\xi} I_0(\xi). \tag{98}$$

From (97) and (9),

$$\textsf{MMSE}(\rho) = 1 - 2\, e^{-\rho} \int_0^\infty r\, e^{-r^2} \frac{I_1^2(2\sqrt{\rho}\, r)}{|I_0(2\sqrt{\rho}\, r)|} dr. \tag{99}$$

Using the asymptotic ($\xi \to \infty$) expansion

$$I_0(\xi) = \frac{e^\xi}{\sqrt{2\pi\xi}} \left(1 + \frac{1}{8\xi} + \mathcal{O}(1/\xi^2)\right) \tag{100}$$

as well as the relationship in (98), we can expand (99) into

$$\textsf{MMSE}(\rho) = \frac{1}{2\rho} + \mathcal{O}(1/\rho^2) \tag{101}$$

as claimed. This expression is consistent, via the relationship in Theorem 1, with the high-power expansion of the mutual information for a continuous PSK distribution reported in [54]:

$$\mathcal{I}(\rho) = \frac{1}{2}\log\left(\frac{4\pi}{e}\rho\right) + \mathcal{O}(1/\rho). \tag{102}$$

E Proof of Theorem 6

From (8), the MMSE estimate of an ∞-PAM input S given its noisy output observation y is

$$\hat{S}(y,\rho) = \frac{y}{\sqrt{\rho}} - \frac{1}{2\sqrt{\pi\rho}} \frac{e^{-(y-\sqrt{3\rho})^2} - e^{-(y+\sqrt{3\rho})^2}}{Q\left(\sqrt{2}\,y - \sqrt{6\rho}\right) - Q\left(\sqrt{2}\,y + \sqrt{6\rho}\right)} \tag{103}$$

from which

$$\text{MMSE}(\rho) = \frac{1}{2\sqrt{3\pi}} \int_{-\infty}^{\infty}\int_{-\sqrt{3}}^{\sqrt{3}} \left(s - \hat{S}(y,\rho)\right)^2 e^{-(y-\sqrt{\rho}\,s)^2} dy\,ds. \tag{104}$$

For large ρ, the estimate in (103) can be expanded as

$$\hat{S}(y,\rho) = \frac{y}{\sqrt{\rho}} - \frac{e^{-(y-\sqrt{3\rho})^2}}{2\sqrt{\pi\rho}} + \epsilon(\rho) \tag{105}$$

where $\epsilon(\rho)$ decays with ρ exponentially faster than the other terms in (105), i.e.,

$$\lim_{\rho\to\infty} \frac{\log\epsilon(\rho)}{\rho} < -3. \tag{106}$$

Plugging (105) into (104),

$$\text{MMSE}(\rho) = \frac{1}{2\rho} + \mathcal{O}(1/\rho^{3/2}) \tag{107}$$

as claimed.

F Iterative Solution of (15)

The level η for the power constraint in place is obtained by solving (15). This invites the application of iterative methods such as bisection, secant or Newton, all of which have good convergence rates [55]. Remarkably, the number of iterations does not depend on the number of channels but only on the desired tolerance. Furthermore, some of these iterative methods (bisection, secant) do not require derivatives of the function $\text{MMSE}_i^{-1}(\cdot)$. The secant method, in particular, has superlinear convergence, i.e., denoting by $\eta^{(k)}$ the

Table 4: Iterative solution of (15) via the secant method for Examples 1 and 3. The target values are $\eta = 0.32$ and $\eta = 0.48$, respectively.

	Example 1	Example 3
$\eta_{\min}$	0.11	0.14
$\eta_{\max}$	0.46	0.67
$\eta^{(1)}$	0.37	0.62
$\eta^{(2)}$	0.32	0.44
$\eta^{(3)}$	0.32	0.49
$\eta^{(4)}$	0.32	0.48

succession of values taken towards the target η, there exists a sequence $b_k > 0$ with $b_k \to 0$ such that

$$|\eta^{(k+1)} - \eta| \leq b_k |\eta^{(k)} - \eta|. \tag{108}$$

In general, it suffices to search for η within an interval bounded by

$$\eta_{\min} = \min\{\gamma_i \, \text{MMSE}_i(\gamma_i)\} \tag{109}$$

$$\eta_{\max} = \max\{\gamma_i \, \text{MMSE}_i(\gamma_i)\} \tag{110}$$

Graphically, these are the top and bottom intersections of the functions $\gamma_i \, \text{MMSE}_i(p_i \gamma_i)$ with a vertical line at $p_i = 1$. (Smaller search intervals can be foreseen in some cases.)

In order to illustrate the superlinear convergence of the secant method, we next apply it to the examples in Section V. The upper and lower bounds in (109) and (110) are used as starting points. These starting points, and a few of the values obtained thereafter, are listed in Table 4 for Examples 1 and 3. (In Example 2, η is graphically evident by inspection.)

In Example 1, only 2 iterations are required to find η with 2-digit precision. In Example 3, 4 iterations are required for that same precision.

G Proof of Theorem 7

From Theorem 4, $\text{MMSE}(\rho)$ decays exponentially to zero as $\rho \to \infty$ and thus, in Theorem 2, $\eta \to 0$. More precisely,

$$\frac{\log \text{MMSE}(\rho)}{\rho} = \mathcal{O}(1) \tag{111}$$

and it follows that, for sufficiently large input power P, on every channel $p_i^\star > 0$ strictly. From (12) and the strict monotonicity of $\log(\cdot)$,

$$\log |h_i|^2 + \log P + \log \text{MMSE}_i(p_i^\star |h_i|^2 P) = \log \eta \qquad i = 1, \ldots, n \tag{112}$$

with η such that the power constraint in (13) is satisfied. Dividing both sides of (112) by P and applying Theorem 4 we find, for large P,

$$\frac{d_i^2}{4} p_i^\star |h_i|^2 \log(e) = \frac{\log 1/\eta}{P} + \mathcal{O}\left(\frac{\log P}{P}\right) \qquad i = 1, \ldots, n \tag{113}$$

where, we note,

$$\frac{\log 1/\eta}{P} = \mathcal{O}(1). \tag{114}$$

For $P \to \infty$, therefore, the power allocation $\{p_i^\star\}_{i=1}^n$ converges to the solution of

$$d_i^2 p_i^\star |h_i|^2 = \alpha \qquad i = 1, \ldots, n \tag{115}$$

where the constant α is determined by the power constraint in (13). This set of equations leads to the claimed limiting power allocation.

H Levin-Campello Algorithm

The core idea in this algorithm is that a greedy sequential approach can lead to a globally optimum discrete loading [27, 28, 29, 56]. Accordingly, it suffices to sequentially load bits, always assigning them to the tone that requires the least incremental amount of power for their transport, up until the desired number of bits has been loaded or the power constraint has been reached. (In our implementation, the constraint is on the average power.)

Denote by $r_i(\cdot)$ the number of bits/s/Hz that can be transported by tone i at a given error probability as function of the power allocated to such tone. This function must be monotonic and concave for a greedy sequential approach to be globally optimum. Often, $r_i(\cdot)$ is approximated by

$$r_i(p_i) = \log_2\left(1 + \frac{p_i \gamma_i}{\Gamma}\right) \tag{116}$$

where Γ is the power gap between r_i and the mutual information achieved by an ideal Gaussian signal. This gap depends on the constellation, coding format and error probability. (Since our ultimate performance measure is the mutual information itself, in our implementation $\Gamma = 1$.) The class of constellations being used determines the set of discrete values that can be taken by r_i. Typically, $r_i = 0, \beta, 2\beta, \ldots, r_{\max}$, where β is the granularity in bits and $r_{\max}$ is the number of bits in the richest available constellation. (In our case, with square QAM constellations, $\beta = 2$ while $r_{\max}$ is a parameter.)

From (116), the power required to transport r_i bits/s/Hz is

$$p_i(r_i) = \frac{2^{r_i} - 1}{\gamma_i / \Gamma} \tag{117}$$

which allows us to express the incremental power required by tone i to transport β more bits, i.e., to promote to the next-cardinality constellation, as

$$\begin{aligned}\Delta p_i(r_i) &= p_i(r_i+\beta)-p_i(r_i) &(118)\\ &= (2^{\beta}-1)\frac{2^{r_i}}{\gamma_i\Gamma}. &(119)\end{aligned}$$

The Levin-Campello algorithm boils down to sequentially loading groups of β bits, with each group being assigned to the tone that has the smallest $\Delta p_i(r_i)$ among those satisfying $r_i < r_{\max}$, up until the point where assigning another group of bits would bring the average power above its constraint. Reduced-complexity approaches to this procedure can be found, e.g., in [28].

References

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.

[2] B. S. Tsybakov, "The capacity of a memoryless Gaussian vector channel," *Problems of Information Transmission*, vol. 1, pp. 18–29, 1965.

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

[4] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[5] 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.

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

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

[8] 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.

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

[10] 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.

[11] 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.

[12] R. Price, "A conversation with Claude Shannon," *IEEE Communications Magazine*, vol. 22, no. 5, pp. 123–126, May 1984.

[13] T. Berger and D. W. Tufts, "Optimum pulse amplitude modulation. Part I: Transmitter-receiver design and bounds from information theory," *IEEE Trans. Inform. Theory*, vol. 13, no. 2, pp. 196–208, Apr. 1967.

[14] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1283–1290, Dec. 1976.

[15] J. Salz, "Digital transmission over cross-coupled linear channels," *AT&T Tech. J.*, pp. 1147–1159, July–Aug. 1985.

[16] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[17] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 1988–2006, July 1999.

[18] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[19] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beam-forming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[20] D. P. Palomar and J. R. Fonollosa, "Practical algoriths for a family of waterfilling solutions," *IEEE Trans. on Signal Processing*, vol. 53, no. 2, pp. 686–695, Feb. 2005.

[21] 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, no. 5, pp. 1237–1239, May 2002.

[22] 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.

[23] R. Price, "Non-linearly feedback equalized PAM versus capacity for noisy filter channels," *Proc. IEEE Intern. Conf. on Commun. (ICC'72)*, 1972.

[24] J. A. C. Bingham P. S. Chow, J. M. Cioffi, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 773–775, Feb/Mar/Apr 1995.

[25] P. S. Chow, "Bandwidth optimized digital transmission techniques for spectrally shaped channels with impulse noise," *Ph.D. Thesis, Stanford University*, May 1993.

[26] R. F. H. Fisher and J. B. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. of GLOBECOM'96*, Nov. 1996, pp. 724–728.

[27] D. Hughes-Hartogs, "Ensemble modem structure for imperfect transmission media," *U.S. Patent 4 679 227*, July 1987.

[28] J. Campello, "Practical bit loading for DMT," in *Proc. of the Int. Conf. on Communications (ICC'99)*, June 1999, vol. 2, pp. 801–805.

[29] J. Campello, "Optimal discrete bit loading for multicarrier modulation systems," in *Proc. of the Int. Symp. on Inform. Theory (ISIT'98)*, Aug. 1998, p. 193.

[30] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 23–27, Jan. 2000.

[31] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*, Upper Saddle River, NJ; Prentice Hall, 1999.

[32] 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.

[33] S. Yang and A. Kavčić, "Capacity of partial-response channels," *Handbook of Coding and Signal Processing for Recording Systems (Chapter 13), CRC Press*, 2004.

[34] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. on Wireless Communications*, Jan. 2006.

[35] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. on Inform. Theory*, vol. 18, no. 1, pp. 12–20, Jan. 1972.

[36] R. E. Blahut, "Computation of channel capacity and rate distortion functions," *IEEE Trans. on Inform. Theory*, vol. 18, pp. 460–473, July 1972.

[37] N. Varnica, Ma Xiao, and A. Kavcic, "Capacity of power constrained memoryless AWGN channels with fixed input constellations," *Proc. of IEEE GLOBECOM'02*, vol. 2, pp. 1339–1343, Nov. 2002.

[38] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[39] G. D. Forney, R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Quereshi, "Efficient modulation for bandlimited channels," *IEEE J. Select. Areas in Communications*, vol. 2, pp. 632–647, 1984.

[40] G. D. Forney and L.-F. Wei, "Multidimensional constellations—part I: Introduction, figures of merit, and generalized cross constellations," *IEEE Trans. on Inform. Theory*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[41] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signalling on the Gaussian channel," *IEEE Trans. on Inform. Theory*, vol. 36, no. 4, pp. 726–740, July 1990.

[42] K. Sistanizadeh, "Proposed canonical loops for ADSL and their loss characteristics," *ANSI TlEl.4 Technical Subcommittee Working Group Contribution*, pp. 91–116, Aug. 1991.

[43] B. Schein and M. Trott, "Sub-optimal power spectra for colored Gaussian channels," in *IEEE Int'l Symp. on Inform. Theory (ISIT'97)*, June 1997, p. 340.

[44] W. Yu and J. M. Cioffi, "On constant power water-filling," in *IEEE Int'l Conf. on Communications (ICC'01)*, June 2001, vol. 6, pp. 1665–1669.

[45] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[46] M.-S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1165–1181, July 1999.

[47] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. on Inform. Theory*, vol. 47, no. 5, May 2001.

[48] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[49] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, 1999.

[50] A. Lozano, A. M. Tulino, and S. Verdú, "Mercury/waterfilling for fixed wireless OFDM systems," in *IEEE Radio and Wireless Symposium (RWS'06)*, Jan. 2006.

[51] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[52] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.

[53] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, New York: Dover Publications, 1972.

[54] J. M. Geist, "Capacity and cutoff rate for dense M-ary constellation," *Proc. of IEEE Military Comm. Conf. (MILCOM'90)*, pp. 768–770, Sept. 1990.

[55] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM (Classics in Applied Mathematics Series, 16), 1996.

[56] B. Fox, "Discrete optimization via marginal analysis," *Management Science*, vol. 13, pp. 210–216, Nov. 1966.

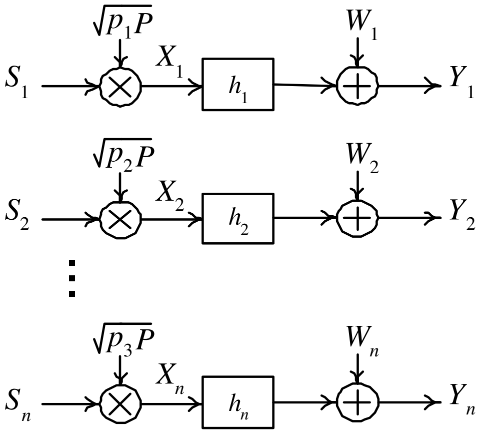
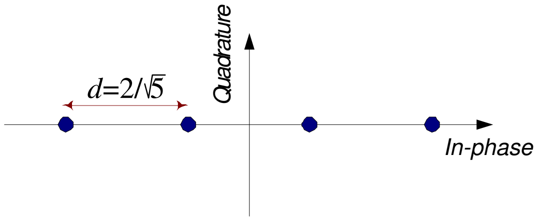
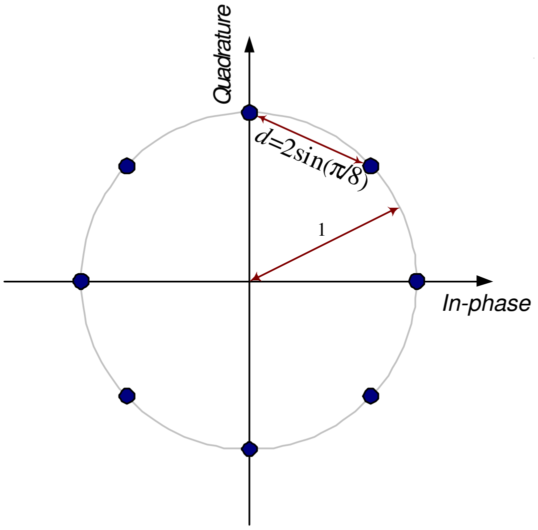

$\sqrt{p_1 P}$
W_1
S_1
X_1
h_1
Y_1
$\sqrt{p_2 P}$
W_2
S_2
X_2
h_2
Y_2
$\sqrt{p_3 P}$
W_n
S_n
X_n
h_n
Y_n

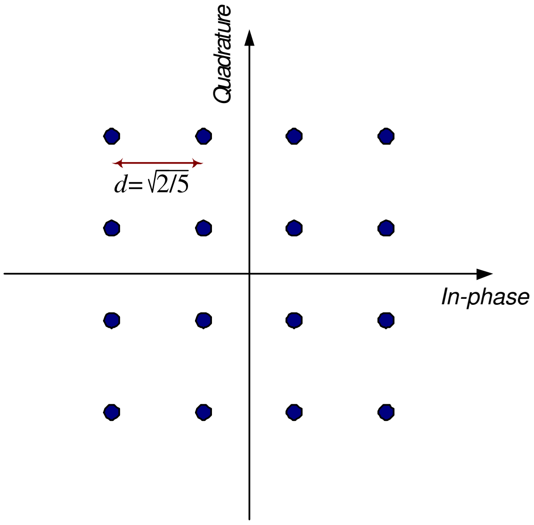
Figure 1: Bank of n independent parallel channels.


Quadrature
d=2/√5
In-phase

(a) 4-PAM


Quadrature
d=2sin(π/8)
1
In-phase

(b) 8-PSK


Quadrature
d=√2/5
In-phase

(c) 16-QAM

Figure 2: Examples of m-ary constellations: (a) 4-PAM, (b) 8-PSK, and (c) 16-QAM.

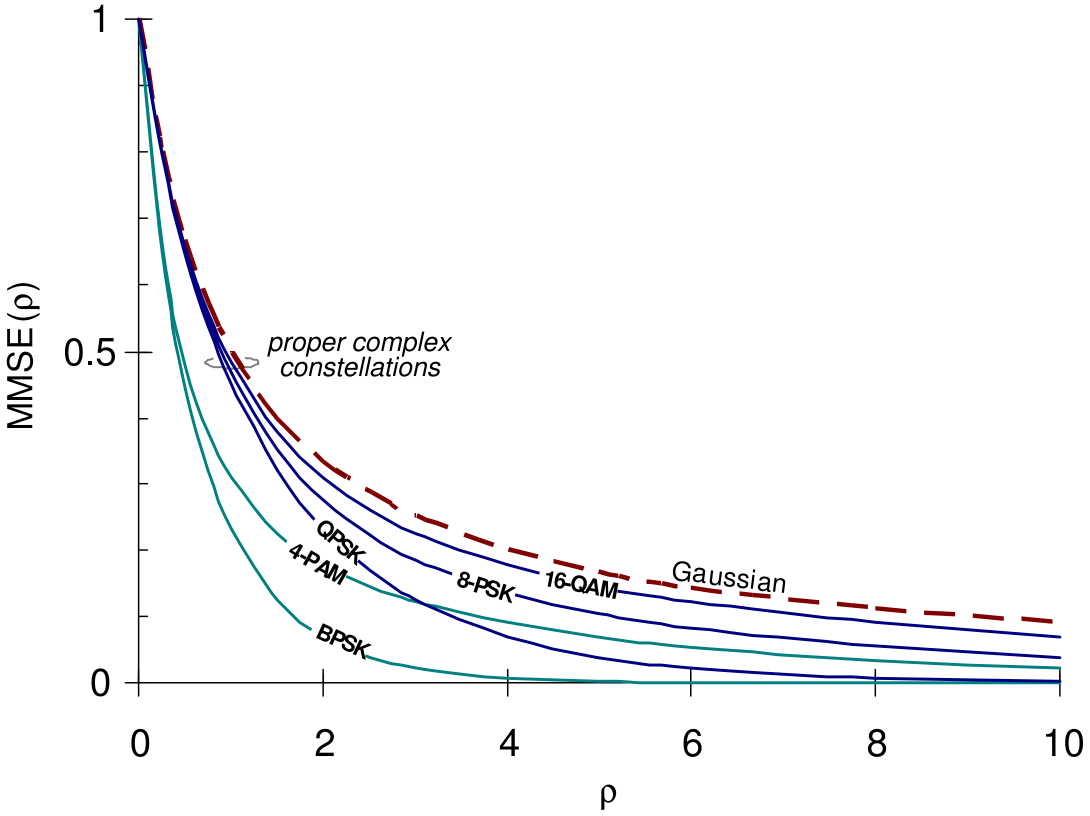


Figure 3: In solid, MMSE$(\cdot)$ for BPSK, QPSK, 8-PSK, 4-PAM and 16-QAM constellations. In dashed, MMSE$(\cdot)$ for a Gaussian input.

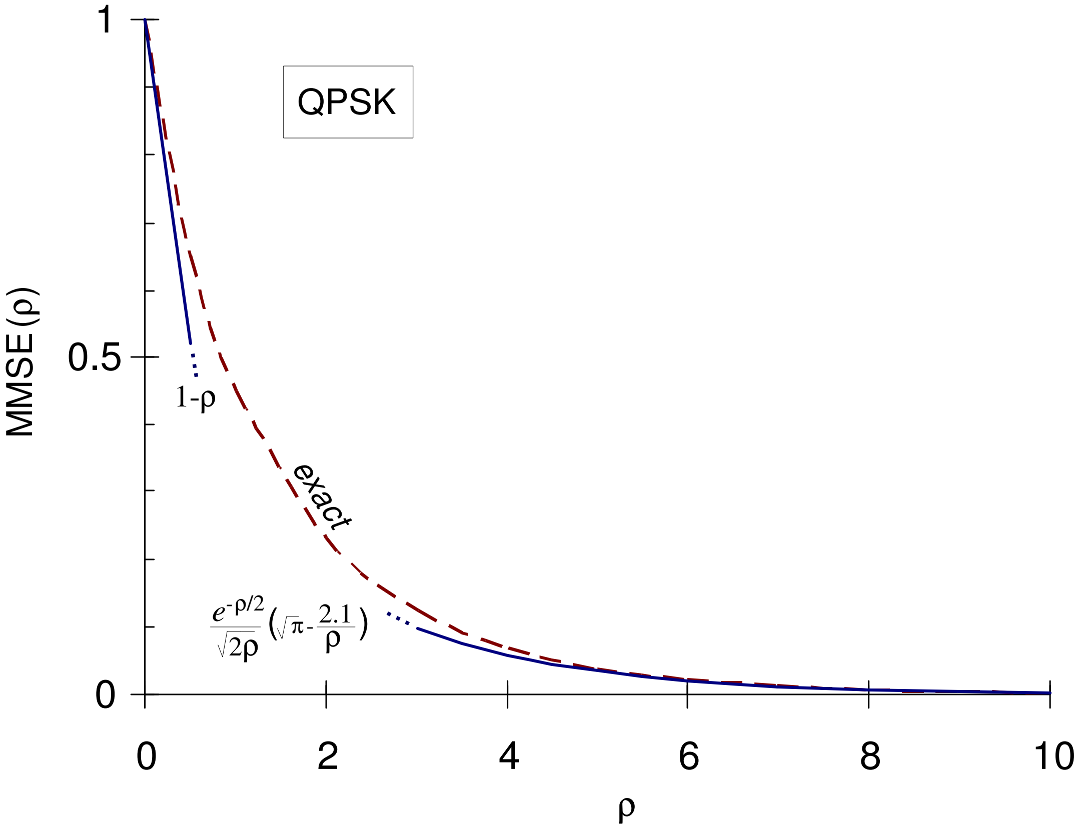


Figure 4: In dashed, exact MMSE$(\cdot)$ for QPSK. In solid, low- and high-power expansions.

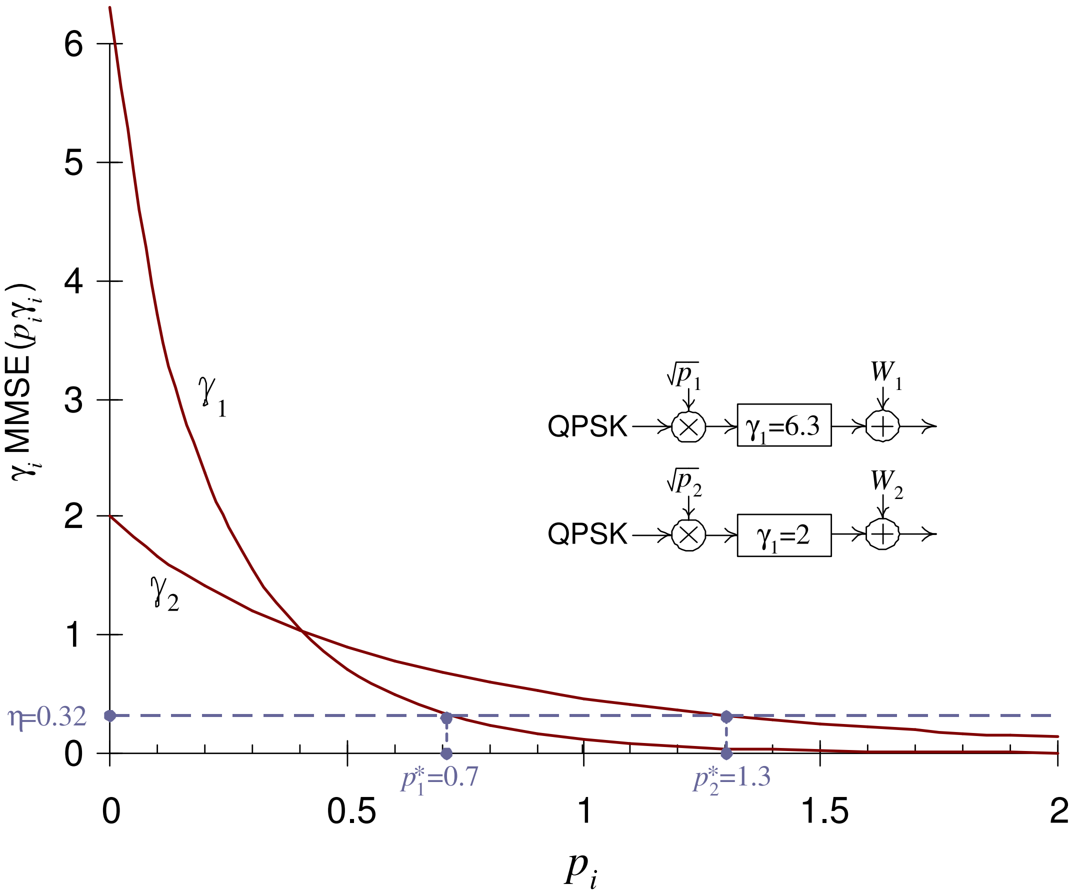


Figure 5: Two parallel channels with $\gamma_1 = 6.3$ and $\gamma_2 = 2$ and QPSK inputs. Solid lines depict $\gamma_i \text{MMSE}(p_i \gamma_i)$ as function of p_i. Their intersection with the dashed line at $\eta = 0.32$ yields $p_1^\star$ and $p_2^\star$.

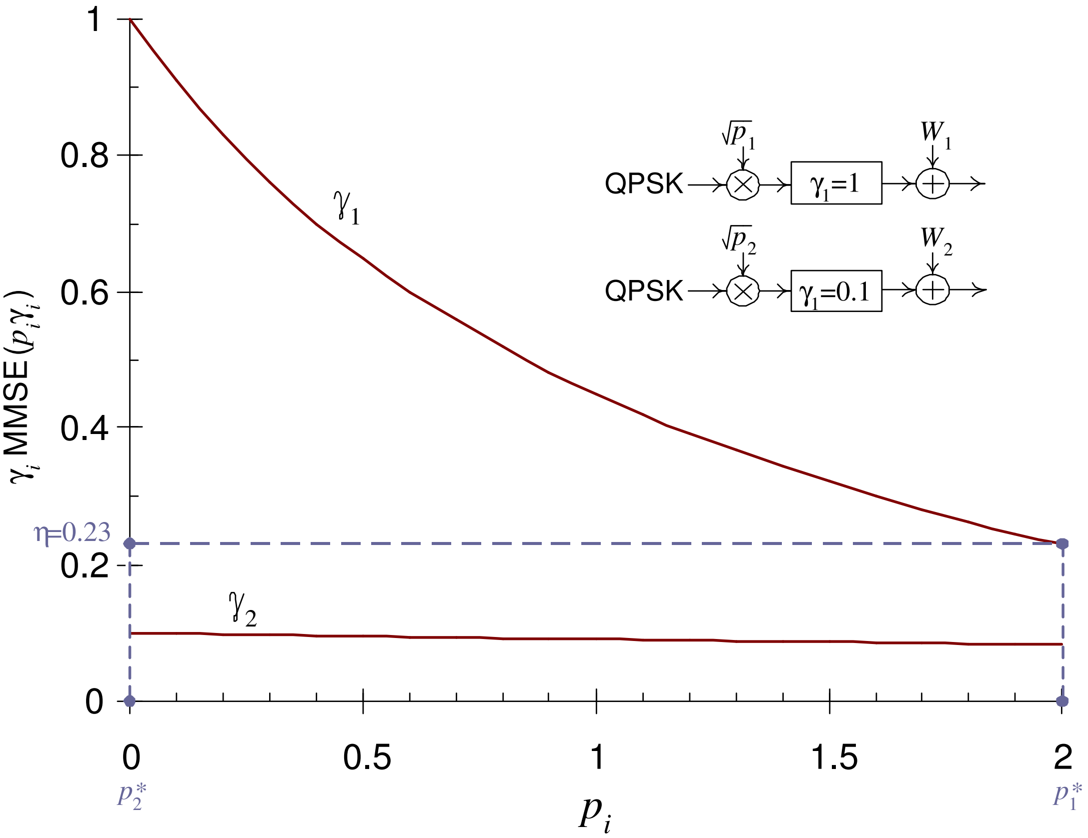


Figure 6: Two parallel channels with $\gamma_1 = 1$ and $\gamma_2 = 0.1$ and QPSK inputs. Solid lines depict $\gamma_i\mathsf{MMSE}(p_i\gamma_i)$ as function of p_i. Their intersection with the dashed line at $\eta = 0.23$ yields $p_1^\star$ and $p_2^\star$.

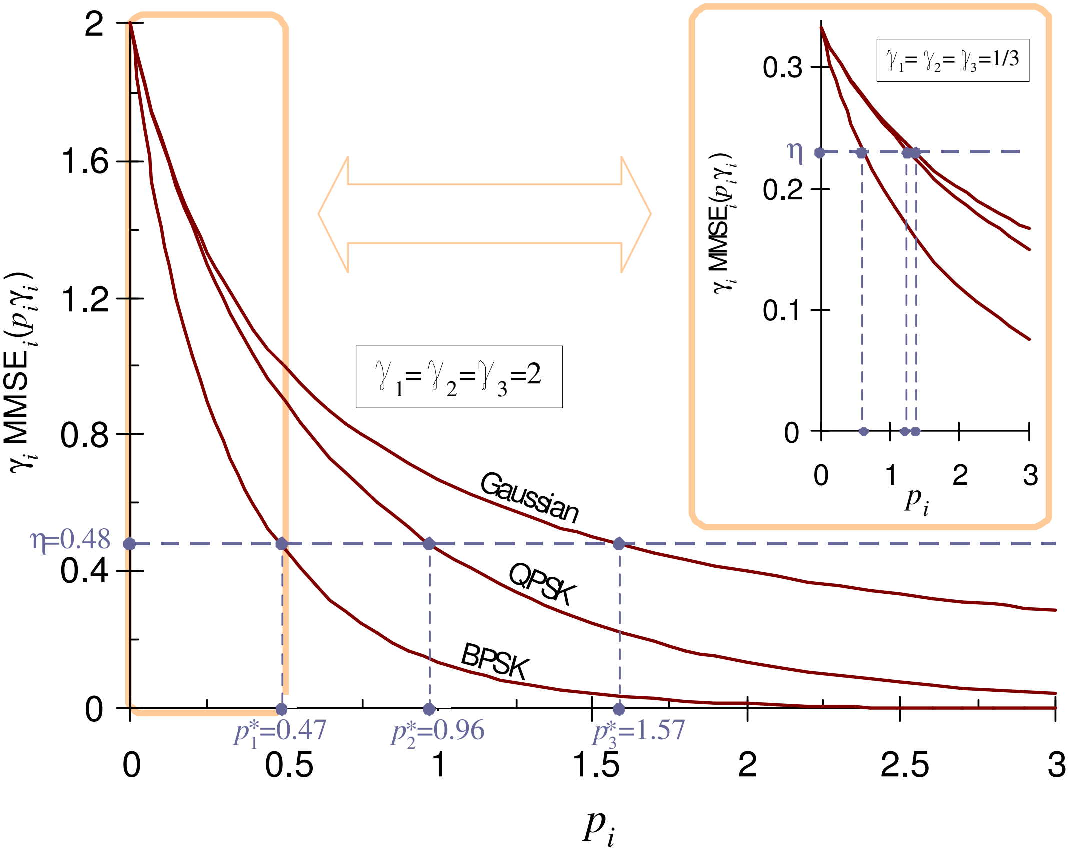


Figure 7: Main chart: Three parallel channels with $\gamma_1 = \gamma_2 = \gamma_3 = 2$. The respective inputs are BPSK, QPSK and Gaussian. Solid lines depict $\gamma_i \mathsf{MMSE}(p_i\gamma_i)$ as function of p_i. Their intersection with the dashed line at $\eta = 0.48$ yields $p_1^\star$, $p_2^\star$ and $p_3^\star$. Inset: identical scenario except for $\gamma_1 = \gamma_2 = \gamma_3 = 1/3$.

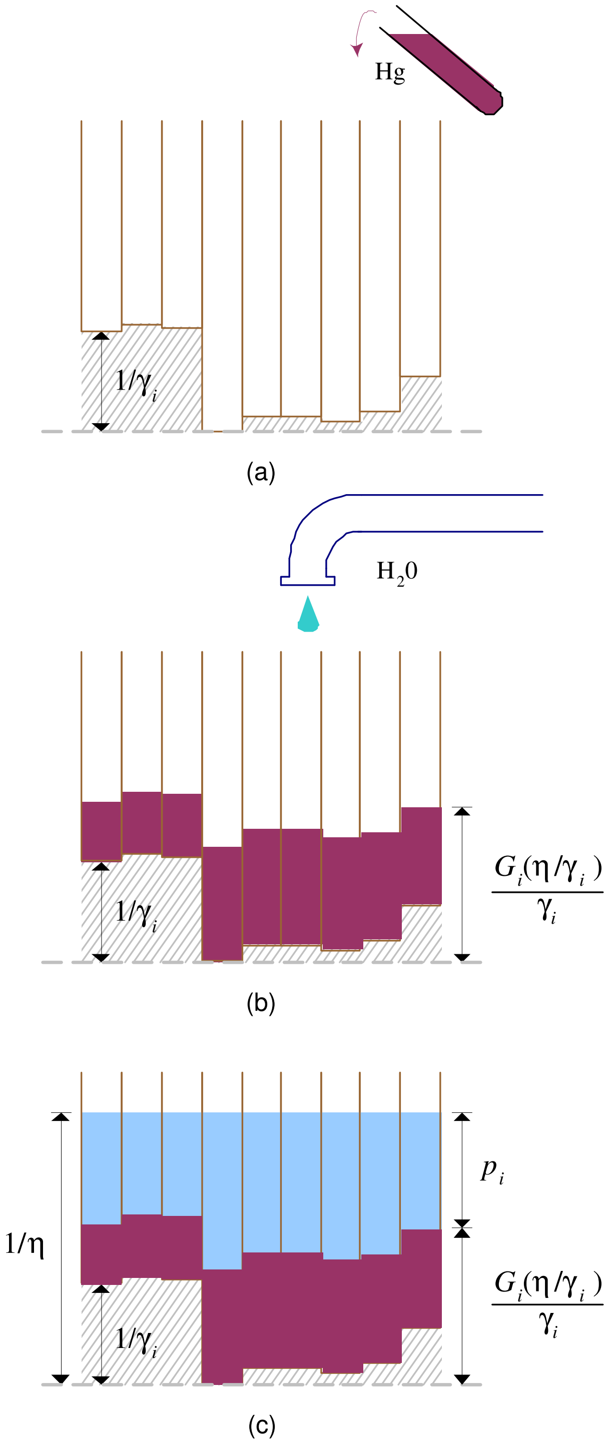

Hg
1/γ_i
(a)
H_20
1/γ_i
G_i(η/γ_i)
γ_i
(b)
1/η
p_i
1/γ_i
G_i(η/γ_i)
γ_i
(c)

Figure 8: 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 $p_i^\star$.

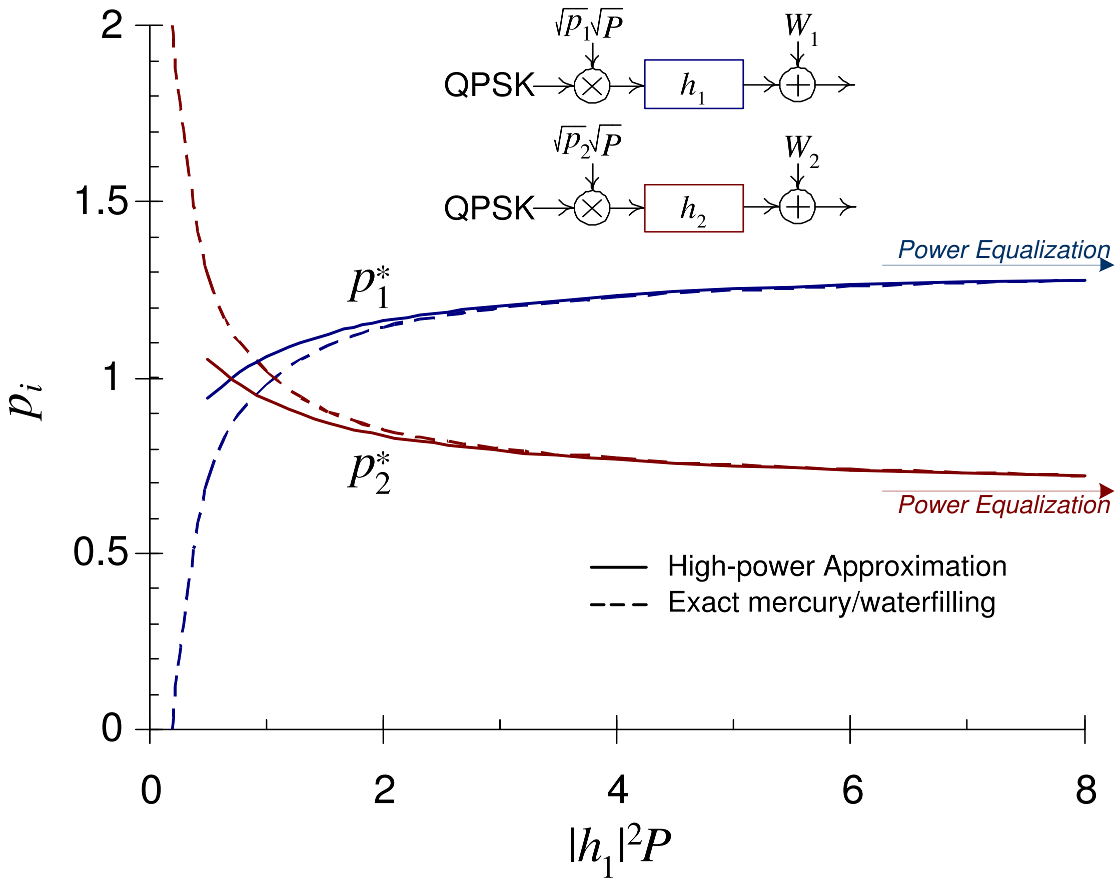


Figure 9: For $n = 2$ with $|h_2|^2 = 2|h_1|^2$, exact values (in dashed) for $p_1^\star$ and $p_2^\star$ as function of $|h_1|^2P$ as well as high-power approximations (in solid) obtained by solving (54). The limiting power-equalizing values are also indicated.

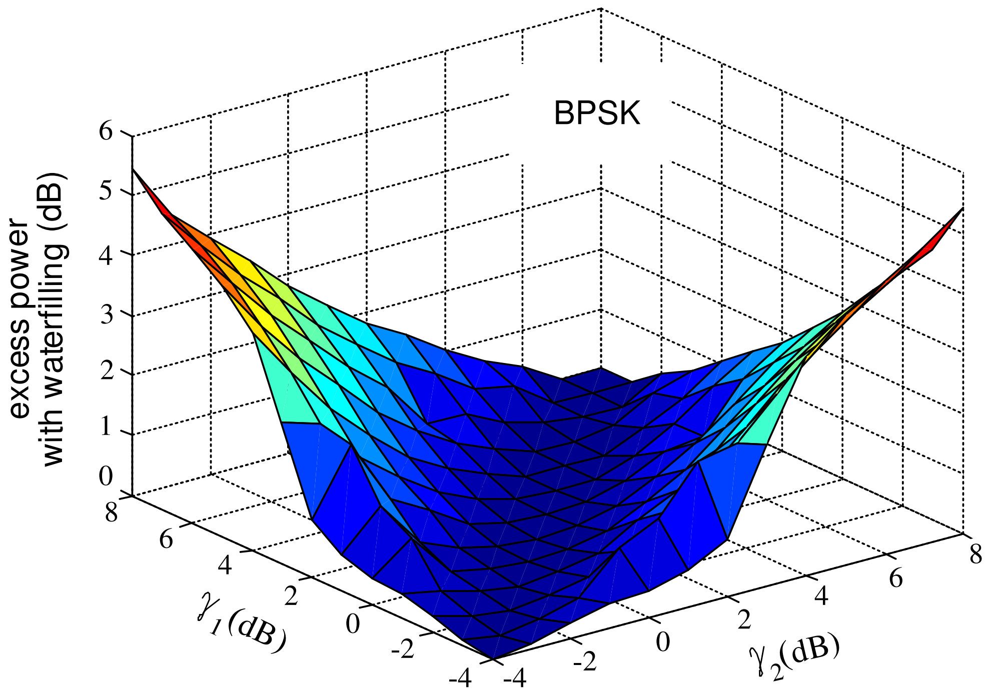


Figure 10: Excess power (dB) required for waterfilling to achieve the same mutual information as the optimum policy, for $n = 2$ channels with BPSK inputs, as function of the pair $\{\gamma_1, \gamma_2\}$ (dB).

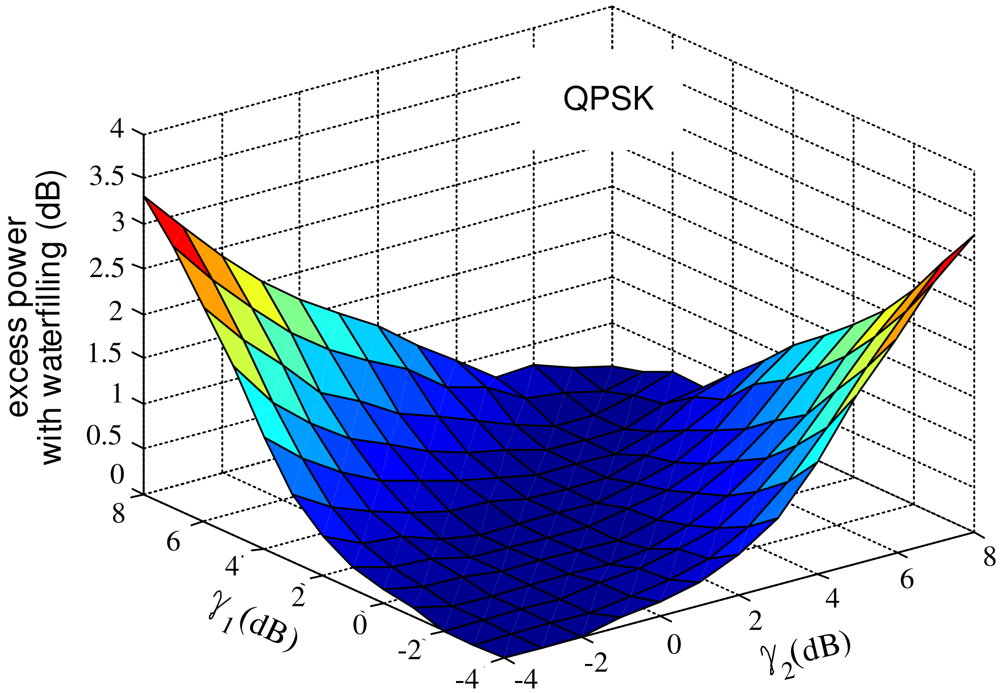


Figure 11: Excess power (dB) required for waterfilling to achieve the same mutual information as the optimum policy, for $n = 2$ channels with QPSK inputs, as function of the pair $\{\gamma_1, \gamma_2\}$ (dB).

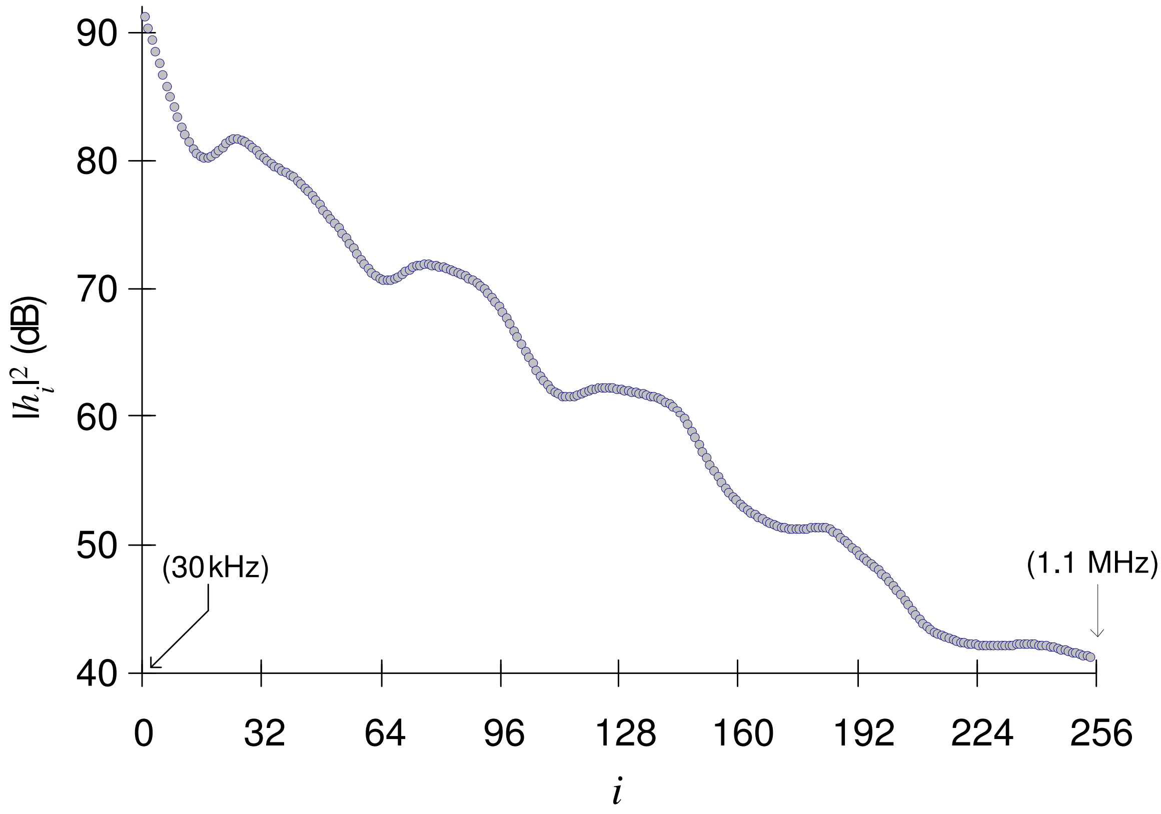


Figure 12: DSL frequency response, sampled into $n = 256$ orthogonal tones and scaled by the respective noise variances.

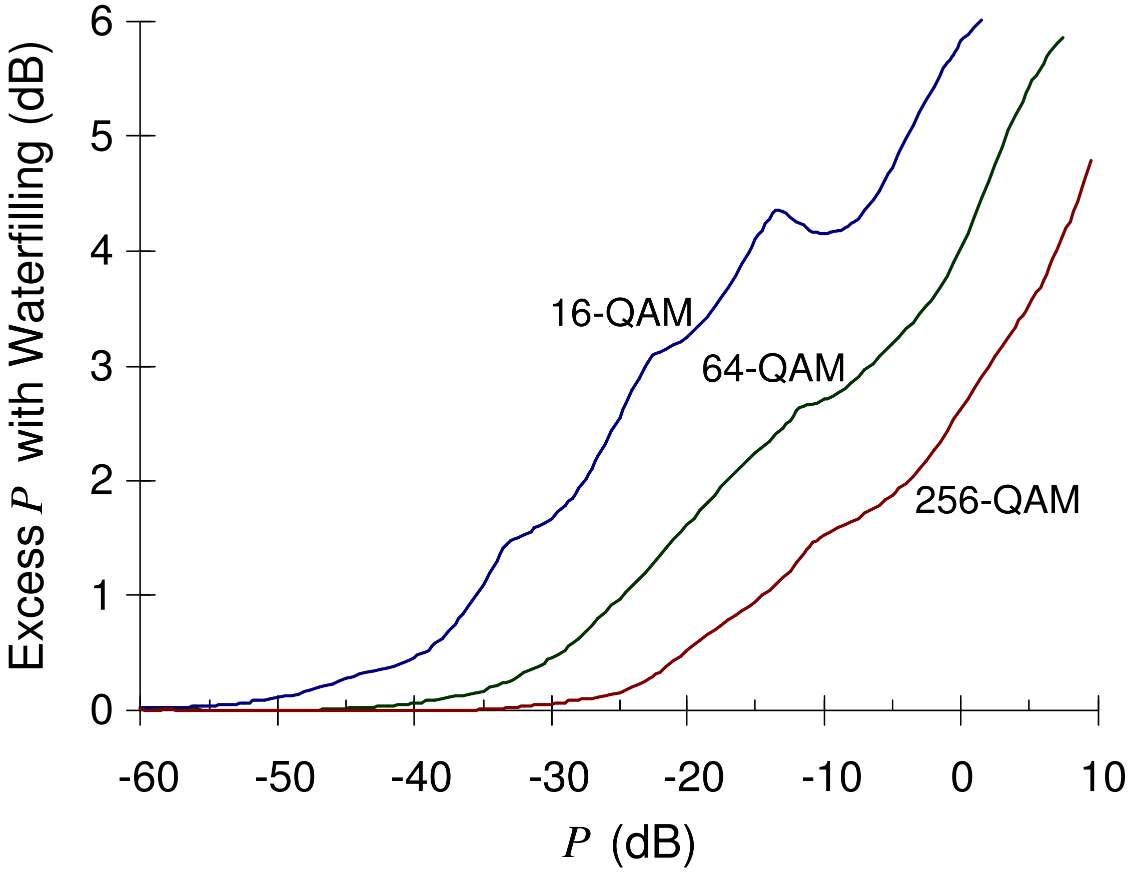


Figure 13: For the channel in Fig. 12 with uniform constellation loading, additional power (in dB) required by waterfilling to achieve the same mutual information achieved by mercury/waterfilling. Results parameterized by the input distribution (16-QAM, 64-QAM and 256-QAM).

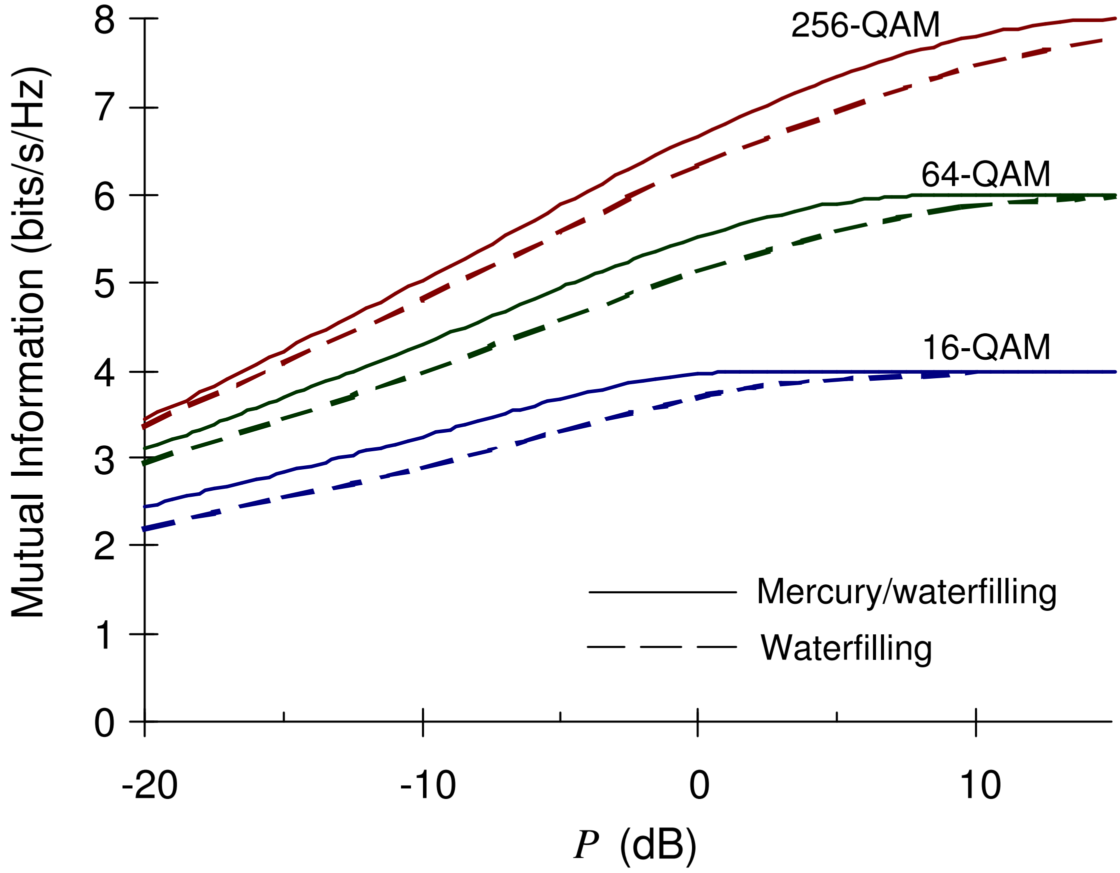


Figure 14: For the channel in Fig. 12 with uniform constellation loading, mutual informations (in bits/s/Hz) achieved with mercury/waterfilling and with waterfilling. Results parameterized by the input distribution (16-QAM, 64-QAM and 256-QAM).

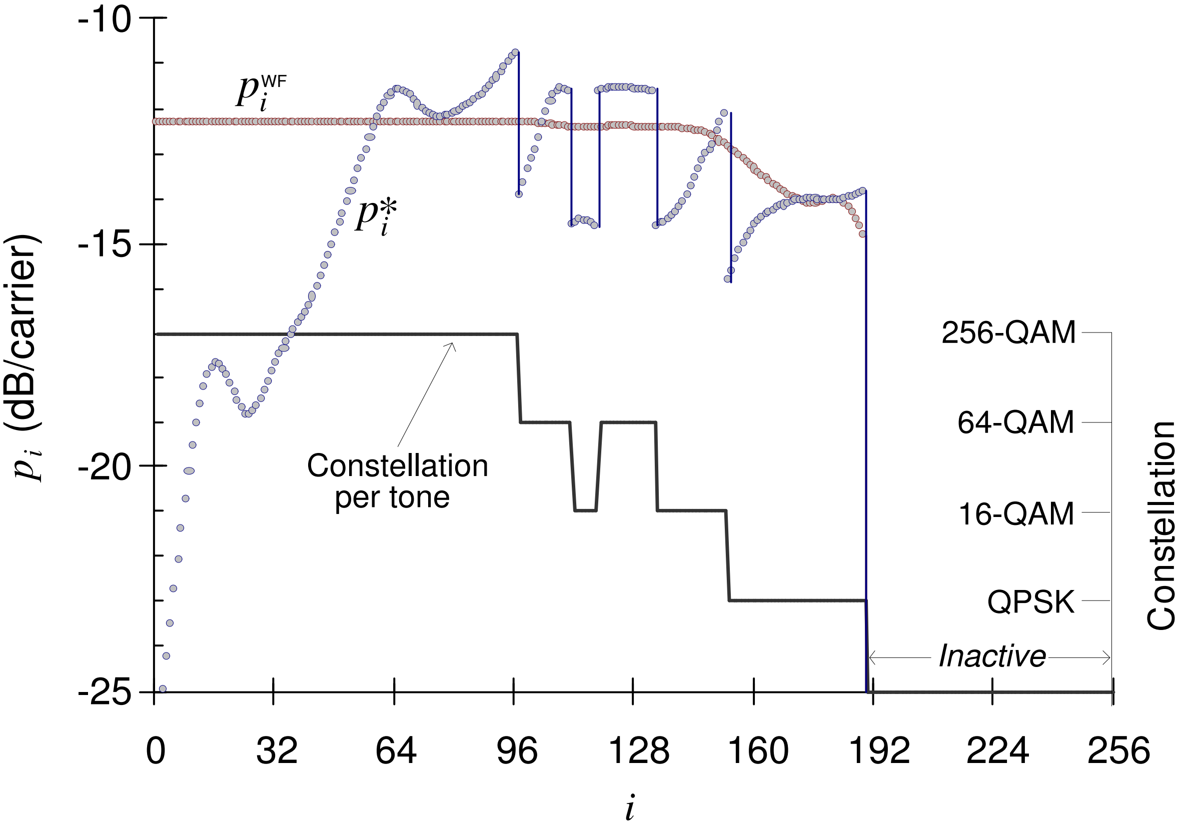


Figure 15: Levin-Campello loading, from the set {QPSK,16-QAM,64-QAM,256-QAM}, at the operating point $P = -15$ dB. Also shown are the mercury/waterfilling and waterfilling power allocations with such loading.

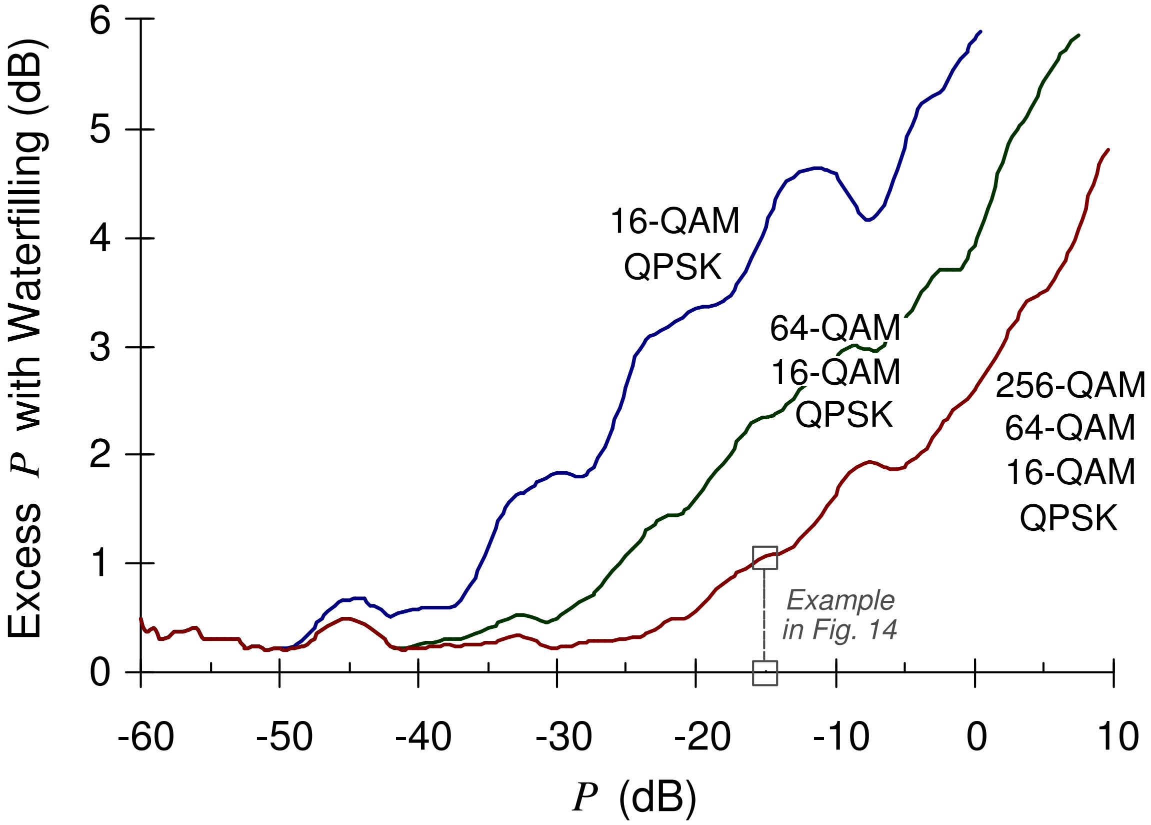


Figure 16: For the channel in Fig. 12 with Levin-Campello loading, additional power (in dB) required by waterfilling to achieve the same mutual information achieved by mercury/waterfilling. Results parameterized by the set of available constellations. Highlighted is the power constraint that corresponds with Fig. 15.

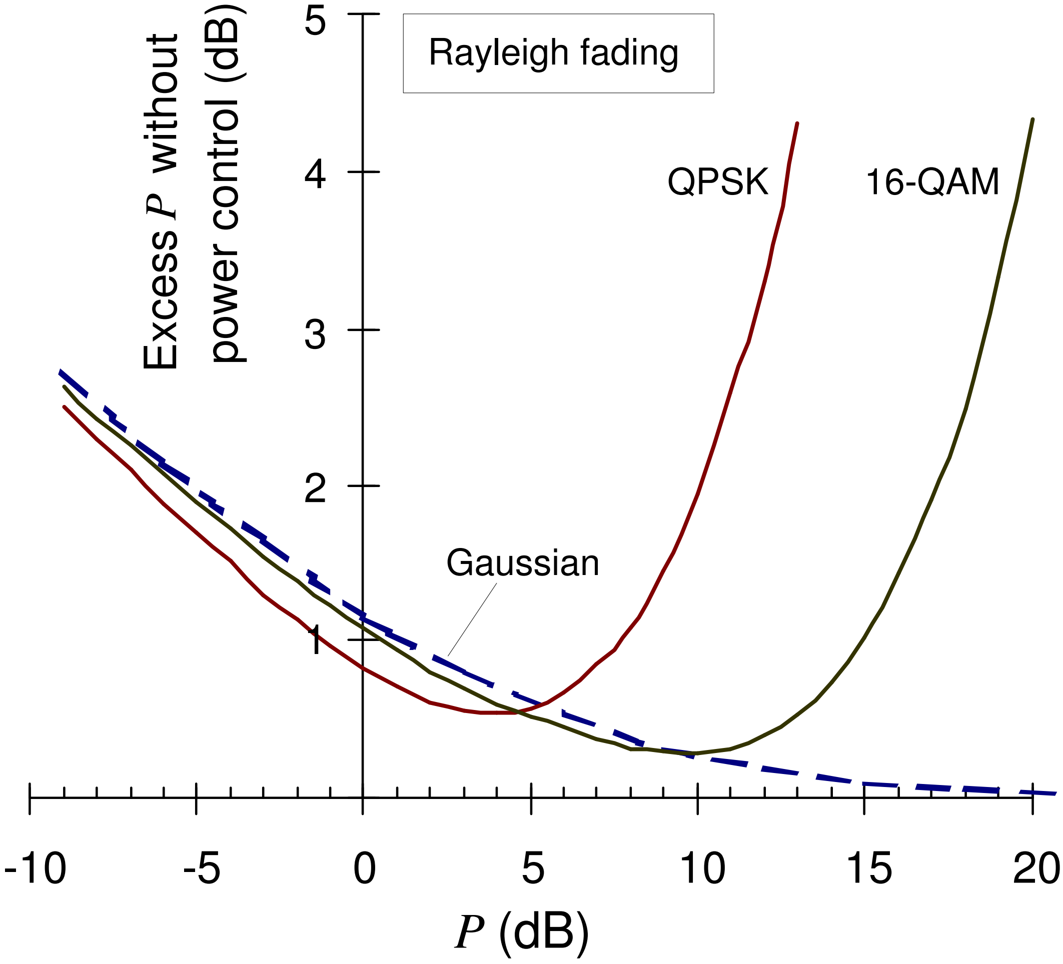


Figure 17: For a Rayleigh-faded channel, additional power (in dB) required to achieve, without power control, the same mutual information achieved with temporal mercury/waterfilling power control at each average power P (in dB). Results parameterized by the input distribution (QPSK, 16-QAM and Gaussian).

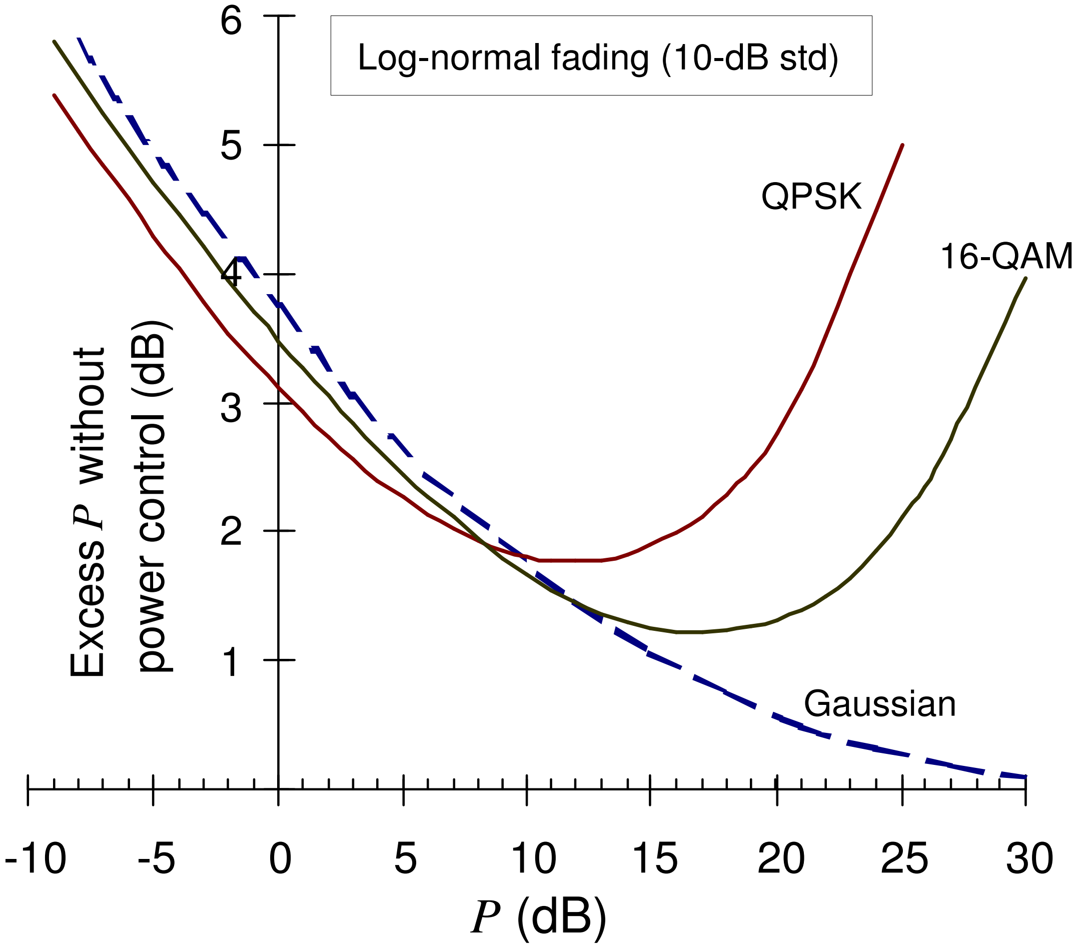


Figure 18: For a log-normal-faded channel with 10-dB standard deviation, additional power (in dB) required to achieve, without power control, the same mutual information achieved with temporal mercury/waterfilling power control at each average power P (in dB). Results parameterized by the input distribution (QPSK, 16-QAM and Gaussian).