Waterfilling Gains O(1/SNR) at High SNR

Emin Martinian

***Abstract*— We show that the gain for using a waterfilling power allocation instead of a flat allocation over non-singular channel components is negligible at high signal-to-noise ratios.**

Consider the standard additive noise communication channel with a quadratic power constraint. Although waterfilling provides the optimal input distribution for Gaussian noise channels, sub-optimal distributions are often not too bad. For example, [1] shows that transmitting at only 2 levels (either on or off) in each sub-channel loses very little. Similarly, [2], [3] show that the mutual information for an optimal waterfilling power allocation on any additive noise channel is at most 0.5 bits per real channel use higher than the mutual information for an independent, identically distributed, Gaussian input.

Here we use elementary arguments to show that the gain in mutual information for the optimal waterfilling power allocation versus a flat power allocation over non-singular channel modes is at most $O(\texttt{SNR}^{-1})$ where $\texttt{SNR}$ denotes the signal-to-noise ratio parameter. When power is allocated across all n modes instead of only the k non-singular modes, the mutual information penalty (in bits per complex channel mode) is $(k/n)\log_2(n/k) + O(\texttt{SNR}^{-1})$ (which is at most $e^{-1}\log_2 e + O(\texttt{SNR}^{-1}) \approx .53 + O(\texttt{SNR}^{-1})$). In [3] this bound is tightened to show that the mutual information loss for a flat Gaussian input is at most $e^{-1}\log_2 e \approx .53$ bits per complex channel mode for *any* $\texttt{SNR}$.

A. Problem Model and Notation

We denote vectors and sequences in bold (*e.g.*, $\mathbf{x}$) with the ith element denoted as x_i. Matrices are capitalized bold letters. Random variables are denoted using the sans serif font (*e.g.*, x) while random vectors and sequences are denoted with bold sans serif (*e.g.*, $\boldsymbol{\mathsf{x}}$).

A variety of scenarios involve transmitting over a channel with different gains or noise powers which can be ergodic, non-ergodic, time-selective or frequency-selective. Since the role of waterfilling is essentially the same in all these cases, we focus on the following model and briefly mention how to translate the main results to other scenarios.

The transmitter selects an n-vector, $\boldsymbol{\mathsf{x}}$, as input to the channel which produces the output $\boldsymbol{\mathsf{y}}$ according to

$$\boldsymbol{\mathsf{y}} = \mathbf{H} \cdot \boldsymbol{\mathsf{x}} + \boldsymbol{\mathsf{w}} \tag{1}$$

where $\boldsymbol{\mathsf{w}}$ are independent, complex, zero mean, Gaussian random variables with an identity covariance matrix. We model a power constraint by requiring $E[||\boldsymbol{\mathsf{x}}||^2] < \texttt{SNR}$. Furthermore, we consider the case where the receiver always has perfect knowledge of the deterministic channel matrix $\mathbf{H}$.

For example, (1) may represent a multi-antenna channel. Alternatively, we can model a time or frequency selective fading channel by considering a diagonal $\mathbf{H}$ where the coefficients correspond to separate time-slots or frequency bands. In any case, maximizing the average mutual information between the channel input sequence and channel output sequence is often desirable for maximizing capacity or minimizing outage probability. In the following we will consider the difference between the mutual information for the optimal Gaussian waterfilling input distribution and a Gaussian input distribution with a flat power allocation either over the non-singular channel modes or all channel modes.

B. Non-singular Channels

Theorem 1. *Consider the setting in* (1) *with* $\mathbf{H}$ *non-singular. If we let* $I^*(\texttt{SNR})$ *represent the mutual information for the optimal waterfilling input distribution and let* $I_{\text{flat}}(\texttt{SNR})$ *represent the mutual information for an independent, complex, circularly symmetric, zero mean, Gaussian input distribution with flat power allocation, then*

$$I^*(\texttt{SNR}) - I_{\text{flat}}(\texttt{SNR}) < O(\texttt{SNR}^{-1}). \tag{2}$$

Proof. Without loss of generality we assume that $\mathbf{H}$ is diagonal. If this is not the case, we can always take the Singular Value Decomposition (SVD).[1] Thus we consider the equivalent channel

$$\mathsf{y}_i = s_i \cdot \mathsf{x}_i + \mathsf{w}_i \tag{3}$$

where the w_i are zero mean, unit variance, complex, circularly symmetric, Gaussian random variables and the s_i are the singular values of $\mathbf{H}$.

The resulting mutual information for a flat power allocation is

$$I_{\text{flat}}(\texttt{SNR}) = \sum_{i=1}^{n} [h(\mathsf{y}_i) - h(\mathsf{y}_i|\mathsf{x}_i)] \tag{4}$$

$$= \sum_{i=1}^{n} [h(s_i \cdot \mathsf{x}_i + \mathsf{w}_i) - h(s_i \cdot \mathsf{x}_i + \mathsf{w}_i|\mathsf{x}_i)] \tag{5}$$

$$= \sum_{i=1}^{n} \log\left(|s_i|^2 \cdot \texttt{SNR}/n + 1\right) \tag{6}$$

$$= \sum_{i=1}^{n} \log\left[\frac{\texttt{SNR}}{n}\left(|s_i|^2 + \frac{n}{\texttt{SNR}}\right)\right] \tag{7}$$

$$= n\log\frac{\texttt{SNR}}{n} + \sum_{i=1}^{n} \log(|s_i|^2 + n/\texttt{SNR}) \tag{8}$$

[1] Using the SVD we can write $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger$ where $\mathbf{U}$, $\mathbf{V}$ are unitary and $\mathbf{D}$ is diagonal. When both transmitter and receiver know $\mathbf{H}$, they can pre/post-multiply by $\mathbf{U}$, $\mathbf{V}$ to make the channel diagonal. When only the receiver knows $\mathbf{H}$, it can deal with $\mathbf{U}$, but the transmitter can no longer strip off $\mathbf{V}$. For a complex, circularly symmetric, zero mean, Gaussian random vector with flat power allocation, however, $\mathbf{V}^\dagger \cdot \boldsymbol{\mathsf{x}}$ has the same distribution as $\boldsymbol{\mathsf{x}}$ and we can ignore $\mathbf{V}$ in computing mutual information.

When the transmitter knows s_i, the optimal input distribution is to make each x_i an independent, zero mean, complex, circularly symmetric, Gaussian random variable with variance $P_i = \min[0, \nu - |s_i|^{-2}]$ for some ν such that the power constraint is met. If we define $P_{\max} = \max_i P_i$ then we can bound the waterfilling mutual information via

$$I^*(\texttt{SNR}) = \sum_{i=1}^{n} [h(y_i) - h(y_i|x_i)] \tag{9}$$
$$= \sum_{i=1}^{n} \log(|s_i|^2 \cdot P_i + 1) \tag{10}$$
$$\leq \sum_{i=1}^{n} \log(|s_i|^2 \cdot P_{\max} + 1) \tag{11}$$
$$= \sum_{i=1}^{n} \log\left[P_{\max} \cdot (|s_i|^2 + 1/P_{\max})\right] \tag{12}$$
$$= n \log P_{\max} + \sum_{i=1}^{n} \log(|s_i|^2 + 1/P_{\max}) \tag{13}$$
$$\leq n \log P_{\max} + \sum_{i=1}^{n} \log(|s_i|^2 + n/\texttt{SNR}) \tag{14}$$

Note that for SNR large enough all the modes are active so we can bound the difference between $P_{\max}$ and P_i via

$$P_{\max} - P_i = (\nu - |s_{\max}|^{-2}) - (\nu - |s_i|^{-2}) \tag{15}$$
$$= |s_i|^{-2} - |s_{\max}|^{-2} \tag{16}$$
$$\leq |s_{\min}|^{-2} - |s_{\max}|^{-2} \tag{17}$$

where $s_{\min} = \min_i |s_i|$ and $s_{\max} = \max_i |s_i|$. In particular, this implies that the maximum power is not much larger than the average in the sense that

$$P_{\max} \leq \frac{\texttt{SNR}}{n} + |s_{\min}|^{-2} - |s_{\max}|^{-2}. \tag{18}$$

Combining (8), (14), and (18) yields

$$I^*(\texttt{SNR}) - I_{\text{flat}}(\texttt{SNR}) \leq n \log \frac{P_{\max}}{\texttt{SNR}/n} \tag{19}$$
$$\leq n \log \frac{\texttt{SNR}/n + |s_{\min}|^{-2} - |s_{\max}|^{-2}}{\texttt{SNR}/n} \tag{20}$$
$$= n \log \left(1 + \frac{|s_{\min}|^{-2} - |s_{\max}|^{-2}}{\texttt{SNR}/n}\right) \tag{21}$$
$$\leq n \cdot \left(\frac{|s_{\min}|^{-2} - |s_{\max}|^{-2}}{\texttt{SNR}/n}\right) \tag{22}$$
$$= n^2 \cdot (|s_{\min}|^{-2} - |s_{\max}|^{-2}) \cdot \texttt{SNR}^{-1} \tag{23}$$

□

1) A Time-Selective Fading Example: To gain some idea of why n^2 appears in (23), consider a time-selective fading example. Specifically, let $\mathbf{H}$ be a diagonal matrix where each entry is selected from some distribution $p_s(\cdot)$. In this case, it makes sense to consider the power constraint on a per sample basis and replace $\texttt{SNR}/n$ in (23) with SNR. Similarly, it makes sense to measure mutual information on a per sample basis and divide (23) by n. With these modifications the gain of a waterfilling distribution in this scenario is simply $(|s_{\min}|^{-2} - |s_{\max}|^{-2}) \cdot \texttt{SNR}^{-1}$. Thus we see that the factors of n^2 in (23) are simply normalization factors for the vector scenario and disappear if we are interested in scalar settings.

C. Singular Channels

For channels where $\mathbf{H}$ is singular, we obtain the same result as the previous section provided the transmitter knows which modes of the channel are usable and only spends power on these.

Corollary 1. *Consider the setting in* (1)*. If we let* $I^*(\texttt{SNR})$ *represent the mutual information for the optimal waterfilling input distribution and let* $I_{\text{flat}}(\texttt{SNR})$ *represent the mutual information for an independent, complex, circularly symmetric, zero mean, Gaussian input distribution with flat power allocation over non-singular modes of the channel, then*

$$I^*(\texttt{SNR}) - I_{\text{flat}}(\texttt{SNR}) < O(\texttt{SNR}^{-1}). \tag{24}$$

Proof. In both cases, the transmitter only spends power on the non-singular modes of $\mathbf{H}$ and thus the problem is reduced to the one in Theorem 1. □

If the transmitter does not know which singular values of $\mathbf{H}$ are zero, it essentially wastes power on these modes and suffers a loss.

Corollary 2. *Consider the setting in* (1) *where only* k *out of* n *singular values of* $\mathbf{H}$ *are positive. If we let* $I^*(\texttt{SNR})$ *represent the mutual information for the optimal waterfilling input distribution and let* $I_{\text{flat}}(\texttt{SNR})$ *represent the mutual information for an independent, complex, circularly symmetric, zero mean, Gaussian input distribution with flat power allocation, then*

$$I^*(\texttt{SNR}) - I_{\text{flat}}(\texttt{SNR}) < k \log \frac{n}{k} + O(\texttt{SNR}^{-1}). \tag{25}$$

Proof. Without loss of generality assume that the first k singular values are non-zero (if this is not the case simply permute $\mathbf{H}$ or its SVD representation). Using the same arguments as (9)–(12) we obtain

$$I^*(\texttt{SNR}) \leq \sum_{i=1}^{k} \log\left[P_{\max} \cdot (|s_i|^2 + 1/P_{\max})\right] \tag{26}$$
$$= k \log P_{\max} + \sum_{i=1}^{k} \log\left(|s_i|^2 + 1/P_{\max}\right) \tag{27}$$
$$\leq k \log P_{\max} + \sum_{i=1}^{k} \log\left(|s_i|^2 + n/\texttt{SNR}\right). \tag{28}$$

When a flat power allocation is used, arguments analogous to (4)–(8) yield

$$I_{\text{flat}}(\texttt{SNR}) = k \log \frac{\texttt{SNR}}{n} + \sum_{i=1}^{k} \log(|s_i|^2 + n/\texttt{SNR}). \tag{29}$$

As before we can derive a bound like (18) for $P_{\max}$ (with n replaced by k) which when combined with (28) and (29) yields

$$I^*(\texttt{SNR}) - I_{\text{flat}}(\texttt{SNR}) \leq k \log P_{\max} - k \log \frac{\texttt{SNR}}{n} \tag{30}$$

$$= k \log \frac{n \cdot P_{\max}}{\mathtt{SNR}} \tag{31}$$

$$\leq k \log \left[\frac{n}{k} + \frac{n}{\mathtt{SNR}} \cdot (|s_{\min}|^{-2} - |s_{\max}|^{-2})\right] \tag{32}$$

$$= k \log \frac{n}{k}$$

$$+ k \log \left[1 + \frac{k}{\mathtt{SNR}} \cdot (|s_{\min}|^{-2} - |s_{\max}|^{-2})\right] \tag{33}$$

$$\leq k \log \frac{n}{k} + \frac{k^2}{\mathtt{SNR}} \cdot (|s_{\min}|^{-2} - |s_{\max}|^{-2}) \tag{34}$$

□

Note that if we consider a time-selective or frequency selective channel as in Section B.1, then when (34) is properly normalized, the penalty in bits per channel use becomes $(k/n)\log(n/k) + O(\mathtt{SNR}^{-1})$. If we let $x = k/n$, then the asymptotic penalty is $-x \log x$ which through some elementary calculus can be easily verified to be at most $e^{-1} \log e$ or approximately 0.53 bits per complex channel mode.

D. Non-Gaussian Additive Noise

The previous results hold even when the additive noise, w, is non-Gaussian provided each component is independent and has a finite differential entropy. We show this by demonstrating that, at high $\mathtt{SNR}$, the maximum mutual information of each component channel can be approximated by the capacity of a Gaussian channel.

Theorem 2. *Consider the additive noise channel*

$$\mathsf{y} = \mathsf{x} + \mathsf{w} \tag{35}$$

with complex input x *and complex, zero-mean, unit-variance, additive noise* w *having finite differential entropy* $h(\mathsf{w})$. *With the input constraint* $E[|\mathsf{x}|^2] \leq \mathtt{SNR}$, *the difference between the mutual information for the optimal input distribution and the mutual information with a Gaussian input and* w *replaced by a zero-mean, unit-variance, complex, circularly symmetric Gaussian is at most* $\mathtt{SNR}^{-1}$.

Proof. The capacity of a Gaussian noise channel is $\log(1 + \mathtt{SNR})$. We can lower and upper bound the mutual information of the actual channel by using a Gaussian input distribution for x as follows:

$$I^* = \max_{p_{\mathsf{x}}(x): E[\mathsf{x}^2] \leq \mathtt{SNR}} h(\mathsf{x} + \mathsf{w}) - h(\mathsf{w}) \tag{36}$$

$$\geq \max_{p_{\mathsf{x}}(x): E[\mathsf{x}^2] \leq \mathtt{SNR}} h(\mathsf{x}) - h(\mathsf{w}) \tag{37}$$

$$\geq \log \mathtt{SNR} + \log 2\pi e - h(\mathsf{w}) \tag{38}$$

$$= \log \mathtt{SNR} \tag{39}$$

$$= \log(1 + \mathtt{SNR}) - \log\left(\frac{1 + \mathtt{SNR}}{\mathtt{SNR}}\right) \tag{40}$$

$$I^* = \max_{p_{\mathsf{x}}(x): E[\mathsf{x}^2] \leq \mathtt{SNR}} h(\mathsf{x} + \mathsf{w}) - h(\mathsf{w}) \tag{41}$$

$$\leq \log(\mathtt{SNR} + 1) + \log 2\pi e - h(\mathsf{w}) \tag{42}$$

$$= \log(1 + \mathtt{SNR}) \tag{43}$$

$$|I^* - \log(1 + \mathtt{SNR})| \leq \log\left(\frac{1 + \mathtt{SNR}}{\mathtt{SNR}}\right) \tag{44}$$

$$= \log\left(1 + \mathtt{SNR}^{-1}\right) \tag{45}$$

$$\leq \mathtt{SNR}^{-1}. \tag{46}$$

□

Using this result, we can show that Theorem 1 holds for arbitrary additive noise with finite differential entropy.

Corollary 3. *Theorem 1 holds even if the noise is not Gaussian provided it has a finite differential entropy.*

Proof. We can use the same arguments as in Theorem 1 provided we add an additional term of $O(\mathtt{SNR}^{-1})$ to account for the non-Gaussianity of the noise. □

E. Concluding Remarks

While these results show that waterfilling is essentially a low $\mathtt{SNR}$ strategy, they do not imply that transmitter knowledge of the channel is useless. First, as illustrated by Corollary 2, not knowing which channel modes are singular can result in wasted power and a non-zero penalty in mutual information. Thus, it is at least important to know which channel modes are non-singular. Second, for channels where the range between singular values is large, "high $\mathtt{SNR}$" may be too high to be reached in practice. Third, knowing the channel is often essential for choosing the proper transmission rate. Thus while for non-singular channels the transmitter can achieve the same mutual information regardless of its channel knowledge, it can not necessarily achieve the same capacity unless it knows how many bits the channel can support. Finally, practical considerations such as error rates at finite block lengths or complexity may be affected by channel knowledge.

Acknowledgments

Comments from U. Erez, G. D. Forney, Jr., M. D. Trott, and G. W. Wornell were helpful in clarifying these ideas.

References

[1] B. Schein and M. Trott, "Sub-optimal power spectra for colored gaussian channels," in *International Symposium on Information Theory*, p. 340, 1997.

[2] R. Zamir, "The half a bit loss of robust source/channel codebooks," in *Information Theory Workshop*, pp. 123–126, 2002.

[3] U. Erez and R. Zamir, "A Gaussian input is not too bad," *submitted to IEEE Trans. Inform. Theory*, June 2002.