A Simple Proof of the Water Filling Theorem for Multichannel Modulation

Mark Flanagan[†] **and Anthony D. Fagan**[*]

Department of Electronic and Electrical Engineering
University College Dublin, Belfield, Dublin 4
IRELAND

E-mail: [†]mark.flanagan@ee.ucd.ie [*]tony.fagan@ucd.ie

Abstract — **The water filling theorem for power loading of DMT subchannels in order to optimise the overall bit rate is well known. Previous proofs of this theorem relied on the use of the method of Lagrange multipliers with the application of the Kuhn-Tucker conditions. In this paper we present a simple proof of this theorem which does not require calculus. We believe the proof given here provides insight into the optimality of 'water filling' for power allocation.**

Keywords — **Discrete Multitone Modulation, Channel Capacity, Water Filling Algorithm.**

I Expression for the Information Rate

We shall set up the problem in a manner similar to textbooks [1, 2]; however, we shall state and prove the theorem in a novel manner.

According to Shannon's information capacity theorem [3] the capacity of a narrowband AWGN channel of bandwidth B is given by

$$C = B\log_2(1+SNR) \qquad \text{Bits/sec}$$

In practice, a physically realisable system (e.g. DMT) transmits data at a rate $R < C$, incorporating an modulation gap $\Gamma > 1$:

$$R = B\log_2\left(1+\frac{SNR}{\Gamma}\right) \qquad \text{Bits/sec}$$

Consider a channel bandlimited to B Hertz with frequency response function $H(f)$, transmit power spectral density $S_X(f)$ and additive noise power spectral density $S_N(f)$, $|f| \leq B$. We divide the bandwidth $[0,B]$ into N subchannels $\left[f_k - \frac{\Delta f}{2}, f_k + \frac{\Delta f}{2}\right]$ of bandwidth $\Delta f = \frac{B}{N}$, where the center frequencies are $f_k = \frac{\Delta f}{2} + k\Delta f$, $k = 0,1,\ldots N-1$. We then make the approximation that each subchannel is an AWGN narrowband channel. This approximation is good for large N. The information rate for subchannel k is then

$$R_k = \Delta f \log_2\left(1+\frac{|H(f_k)|^2 P_k}{\sigma_k^2\Gamma}\right) \qquad \text{Bits/sec}$$

where $P_k = 2S_X(f_k)\Delta f$ is the transmit signal power allocated to subchannel k, and $\sum_{k=0}^{N-1} P_k = P$. The noise variance for subchannel k is $\sigma_k^2 = 2S_N(f_k)\Delta f$, and we assume the modulation gap $\Gamma \geq 1$ is the same for all subchannels. Define the gain to noise ratio (GNR) for subchannel k as

$$\alpha_k = \frac{|H(f_k)|^2}{\sigma_k^2\Gamma}$$

Then the information rate for subchannel k is

$$R_k = \Delta f \log_2\left(1+\frac{P_k}{\alpha_k}\right) \qquad \text{Bits/sec}$$

II Lemma

The maximum of

$$\prod_{k=0}^{N-1}(P_k+\alpha_k) = \prod_{k=0}^{N-1} x_k$$

where $\alpha_k, k = 0,1,\ldots N-1$ are fixed positive real numbers, subject to the constraints

$$P_k \geq 0 \quad \text{for} \quad k = 0,1,\ldots N-1$$

and

$$\sum_{k=0}^{N-1} P_k = P$$

occurs when the following two conditions are satisfied:

$$x_k \leq x_m \qquad \forall k,m \in \{0,1,\ldots N-1\}\; with\; P_k > 0$$

and

$$\sum_{k=0}^{N-1} P_k = P.$$

In particular we note that if $P_k > 0$ and $P_m > 0$, then we can conclude that $x_k \leq x_m$ and $x_m \leq x_k$, and thus $x_k = x_m$.

Proof: Suppose the optimal configuration $\{x_i\}$ is such that $\sum_{i=0}^{N-1} x_i = S$ and $\prod_{i=0}^{N-1} x_i = Q = Q_1 x_k x_m$, and $x_k > x_m$ for some k with $P_k > 0$. Then, for any α with $0 < \alpha < \text{Max}\left\{P_k, \frac{x_k - x_m}{2}\right\}$, we can perform the mapping

$$\begin{aligned} x_k &\rightarrow x'_k = x_k - \alpha \\ x_m &\rightarrow x'_m = x_m + \alpha \\ x_i &\rightarrow x'_i = x_i \qquad\qquad i \neq k, m \end{aligned}$$

The sum $S' = \sum_{i=0}^{N-1} x'_i = S$ remains the same while the product

$$\begin{aligned} Q' &= \prod_{i=0}^{N-1} x'_i \\ &= Q_1 (x_k - \alpha)(x_m + \alpha) \\ &= Q_1 \left(x_k x_m + \alpha (x_k - x_m) - \alpha^2\right) \\ &> Q \end{aligned}$$

i.e. the product has increased. This contradicts the assumption that the configuration is optimal. The statement of the lemma follows directly.

Q.E.D.

III Optimal power loading of subchannels for DMT

The water pouring theorem states:

The distribution of power among subchannels which maximises information rate is such that the following two conditions are satisfied:

$$P_k + \alpha_k \leq P_m + \alpha_m$$

$$\forall k, m \in \{0, 1, \ldots N-1\} \; with \; P_k > 0, \text{ and}$$

$$\sum_{k=0}^{N-1} P_k = P.$$

Proof:

The overall information rate is

$$\begin{aligned} R &= \sum_{k=0}^{N-1} R_k \\ &= \Delta f \sum_{k=0}^{N-1} \log_2 \left(1 + \frac{P_k}{\alpha_k}\right) \\ &= \Delta f \log_2 \prod_{k=0}^{N-1} \left(1 + \frac{P_k}{\alpha_k}\right) \\ &= \Delta f \log_2 \left\{ \prod_{k=0}^{N-1} \left(\frac{1}{\alpha_k}\right) \prod_{n=1}^{N} (P_k + \alpha_k) \right\} \end{aligned}$$

The maximization of this function is easily seen to reduce to the problem described in the Lemma above, and thus the water pouring theorem follows directly.

Q.E.D.

References

[1] S. Haykin, *Communication Systems*. New York: Wiley, 1994.

[2] J. G. Proakis, *Digital Communications*. USA: McGraw-Hill, third edition ed., 1995.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.