Rate-Optimal Power Adaptation in Average and Peak Power Constrained Fading Channels

Sébastien de la Kethulle de Ryhove and Geir E. Øien
Norwegian University of Science and Technology, Dept. of Electronics and Telecommunications, N-7491 Trondheim, Norway. Tel: +47-73594315. Emails: {delaketh, oien}@iet.ntnu.no

Abstract—The power adaptation strategy which maximises the average information rate that can be reliably transmitted over average and peak power constrained block-fading single-user discrete-time channels, in which perfect transmitter and receiver channel state information (CSI) are available and in which any transmitted codeword spans a single fading block, is characterised by means of a theorem and subsequently computed numerically in different scenarios. The results reveal striking differences between the well known water-pouring power adaptation strategy which is optimal from a capacity point of view in the case of stationary and ergodic fading channels where the input is subject to an average power constraint only, and the rate-optimal power adaptation strategy of the average and peak power constrained block-fading channels studied in this document.

I. INTRODUCTION

The capacity of single-user, discrete-time communication channels subject to different constraints on the input signal has been extensively studied since the dawn of information theory [1]–[5], the most commonly used input constraint being the average power constraint. The capacity of the Gaussian channel – in which the input signal is subject to such a constraint – was originally derived by Shannon [1]. Smith [2] later studied the scalar Gaussian channel subject simultaneously to average and peak power constraints, and proved that the capacity-achieving distribution in this case is discrete with a finite number of mass points. These results were subsequently extended to the quadrature (two-dimensional) additive Gaussian channel by Shamai and Bar-David [3]. They showed that the capacity-achieving distribution of such channels subject to average and peak power constraints has a uniformly distributed phase, and a magnitude which is discrete with a finite number of mass points; that is, that it consists of a finite number of concentric circles which are centred about the origin in the complex plane.

The capacity of average power constrained stationary and ergodic fading channels with perfect channel state information (CSI) at the transmitter and the receiver was studied by Goldsmith and Varaiya [4]. They showed that the optimal power adaptation strategy is water-pouring in time, and proposed a variable-rate, variable-power (multiple codebook) coding scheme to achieve capacity. These results were subsequently generalised to the case of imperfect transmitter CSI by Caire and Shamai [5]. They additionally proved that a conventional constant-rate, constant-power codebook is sufficient to achieve capacity provided that the codebook symbols are dynamically scaled by an appropriate power allocation function (which depends on the transmitter CSI only) before transmission.

In this paper, we study block-fading channels in which the input symbols are subject simultaneously to average and peak power constraints, and in which perfect CSI is available both at the transmitter and at the receiver. After characterising, by means of a theorem, the power allocation strategy which maximises the average information rate that can be transmitted over such channels with an arbitrarily low probability of error – assuming any transmitted codeword spans a single fading block – we numerically compute it in different scenarios. For simplicity, we hereafter refer to this power allocation strategy as the rate-optimal power allocation strategy.

Our results reveal striking differences between the rate-optimal power allocation strategy for block-fading channels where the input is subject to an average power constraint only (which also is the capacity-achieving power allocation strategy for the stationary and ergodic fading channel studied in [4] and thereafter in [5], as we will show), and the rate-optimal power allocation strategy for the average and peak power constrained block-fading channels studied in this document.

The remainder of this document is organised as follows: the system model is introduced in Sec. II, and a theorem which characterises the corresponding rate-optimal power allocation strategy is subsequently proved in Sec. III. The methodology we use to numerically compute this rate-optimal power allocation strategy is then discussed in Sec. IV, where numerical results are also presented. Conclusions are finally drawn in Sec. V.

II. SYSTEM MODEL

We consider a frequency-flat fading, discrete-time channel with time-varying channel gain denoted by \( h_k \in \mathbb{C} \) at time \( k \in \{0,1,2,\ldots\} \). Let the transmitted signal have complex baseband representation \( X_k \in \mathbb{C} \). The received baseband signal \( Y_k \in \mathbb{C} \) can then be written

\[
Y_k = h_k X_k + W_k,
\]

where the elements of the sequence \( \{W_k\} \) are assumed to be zero mean, unit variance i.i.d. circularly symmetric complex Gaussian random variables, viz. \( W_k \sim \mathcal{CN}(0,1) \).

The channel gain \( h_k \) is assumed to be nonzero and perfectly estimated at the receiver (high-quality channel gain estimates can for example be achieved by the transmission of a known sequence of pilot symbols), and subsequently fed back to the transmitter via a delay and error free feedback channel.
The CSI $h_k$ will hence be perfectly known to the transmitter and the receiver at all times $k \in \{0, 1, 2, \ldots \}$, provided that it remains constant between successive estimations. To ensure that this is true in our case, we will assume a block-fading scenario throughout this document – that is, that $h_k$ remains constant for blocks of $N$ consecutive symbols – and therefore that $h_k$ is perfectly known to both the transmitter and the receiver at all times. We furthermore assume that the maximum length $N$ is sufficiently long to allow for capacity-achieving coding within each block (we exclusively consider the case where any transmitted codeword spans a single fading block).

In addition to the generic time index $k$, it is convenient for future use to introduce a block index $q \in \{0, 1, 2, \ldots \}$, and an index $j \in \{0, 1, \ldots, N-1\}$ to designate the elements within each block. We thus have $k = Nq + j$.

III. Maximum Average Information Rate and Rate-Optimal Power Adaptation Strategy

Let $C(P_{av}, P_{peak})$ denote the operational capacity [6] of a non-fading complex additive white Gaussian noise channel with unit noise power, and input samples which are subject to an average power constraint $P_{av}$ and a peak power constraint $P_{peak}$. The information capacity [6] $C_{info}(P_{av}, P_{peak})$ of such a channel was studied in [3], where it was shown that the capacity-achieving distribution has a uniformly distributed phase, and that its magnitude is discrete with a finite number of mass points. It can be shown using a time-sharing argument detailed in Appendix I (see also [7]) that $C_{info}(P_{av}, P_{peak})$ is a concave function of $P_{av}$ for fixed $P_{peak}$. Bearing this in mind, it becomes a simple matter to adapt the proofs for the coding theorem and the converse to the coding theorem – given respectively in [6, Sec. 10.1] and [6, Sec. 10.2] for the case of the scalar discrete-time additive white Gaussian noise channel in which the input is subject to an average power constraint only – to establish equality between $C_{info}(P_{av}, P_{peak})$ and $C(P_{av}, P_{peak})$.

For given average and peak powers $P_{av}$ and $P_{peak}$, let us introduce the pre-adaptation SNR

$$
\gamma_q \triangleq P_{av} |h_q|^2
$$

for block $q$, where $f_q(\gamma) = f_q(\gamma = \gamma_q)$ denote the corresponding pre-adaptation SNR probability density function (which will be assumed to be nonzero for all $\gamma \in (0, \infty)$), and let in addition

$$
P_{peak} \triangleq \frac{P_{peak}}{P_{av}}.
$$

Let us now suppose we allow the transmit power $P(\gamma)$ to change as a function of $\gamma = \gamma_q$ (the block index $q$ will be omitted in the remainder of this document for simplicity). The post-adaptation SNR is then given by

$$
\frac{\gamma \cdot P(\gamma)}{P_{av}} = \gamma \cdot p(\gamma),
$$

where we have defined $p(\gamma) \triangleq P(\gamma)/P_{av}$.

For a given power allocation function $p(\gamma)$, the maximum information rate that can be transmitted through channel (1) with arbitrarily small probability of error during a fading block with pre-adaptation SNR $\gamma$ is thus given by $C(\gamma \cdot p(\gamma), \gamma \cdot p_{peak})$. Moreover, for a given power allocation function $p(\gamma)$, the maximum average information rate that can be transmitted through channel (1) with arbitrarily small probability of error is simply the average of the maximum information rates than can be transmitted via channel (1) during each fading block, i.e.

$$
\int_0^\infty C(\gamma \cdot p(\gamma), \gamma \cdot p_{peak}) f_\gamma(\gamma) \, d\gamma.
$$

We have the following theorem:

**Theorem 1:** The maximum average information rate that can be transmitted through channel (1) with arbitrarily low probability of error under the assumptions of Sec. II, with the input samples subject to the average and peak power constraints

$$
E \{ |X_k|^2 \} \leq P_{av}
$$

and

$$
|X_k|^2 \leq P_{peak} \quad \forall k \in \{0, 1, 2, \ldots \},
$$

where\(^1\) $P_{av} \leq P_{peak}$, is given by

$$
\sup_{p(\gamma)} \int_0^\infty C(\gamma \cdot p(\gamma), \gamma \cdot p_{peak}) f_\gamma(\gamma) \, d\gamma,
$$

with $p_{peak}$ having been defined in (3) and the supremum being taken over all power allocation functions $p(\gamma)$ such that

$$
\int_0^\infty p(\gamma) f_\gamma(\gamma) \, d\gamma \leq 1
$$

and

$$
0 \leq p(\gamma) \leq p_{peak}.
$$

**Proof:** It can easily be seen that the maximum average information rate that can be reliably transmitted through channel (1) under the assumptions of Sec. II is given by (8) provided that the average and peak power constraints on the input samples (6) and (7) are satisfied if and only if the power allocation function $p(\gamma)$ satisfies constraints (9) and (10). We now proceed to prove this. For a given fading block with pre-adaptation SNR $\gamma$, and for fixed $p(\gamma)$ and $p_{peak}$, let

$$
g_{R, \theta}(r, \theta) = \frac{1}{2\pi} \sum_{i=1}^s a_i \delta(r - r_i),
$$

where $\delta(\cdot)$ is the Dirac delta function and $\{a_1, \ldots, a_s\}$ and $\{r_1, \ldots, r_s\}$ respectively denote the weights and radii of $s$ concentric circles (all of which depend on $\gamma$, $p(\gamma)$, and $p_{peak}$), be the input distribution (in polar coordinates) achieving the capacity $C(\gamma \cdot p(\gamma), \gamma \cdot p_{peak})$ of a non-fading quadrature additive white Gaussian noise channel with unit power gain, unit noise power, average power constraint $\gamma \cdot p(\gamma) = |h|^2 P(\gamma)$, and peak power constraint $\gamma \cdot p_{peak} = |h|^2 P_{peak}$. Transmission

\(^1\)We do not consider the case $P_{av} > P_{peak}$ since in this case the supremum in (8) is trivially achieved by setting $p(\gamma) = p_{peak}$ for all $\gamma \in (0, \infty)$.
on a non-fading quadrature additive white Gaussian noise channel with power gain $|h|^2$, unit noise power, average power constraint $\gamma \cdot p(\gamma)/|h|^2 = P(\gamma)$ and peak power constraint $\gamma \cdot p_{\text{peak}}/|h|^2 = P_{\text{peak}}$ — corresponding to channel (1) during the given fading block with pre-adaptation SNR $\gamma$ — should then be performed using a codebook with elements which are i.i.d. [6] according to the distribution
\begin{equation}
g_{R,\Theta}(r, \theta) = \frac{1}{2\pi} \sum_{i=1}^{8} a_i \delta \left( r - \frac{r_i}{|h|} \right) \tag{12}
\end{equation}
in order to achieve the information rate $C(\gamma \cdot p(\gamma), \gamma \cdot p_{\text{peak}})$ during this fading block. When using such a codebook, we have that
\begin{equation}
E\{|X_j|^2\} \leq P(\gamma) \tag{13}
\end{equation}
and
\begin{equation}
|X_j|^2 \leq P_{\text{peak}} \quad \forall j \in \{0, 1, \ldots, N - 1\}, \tag{14}
\end{equation}
where the expectation in (13) is taken over all the elements of the codebook used during the fading block under consideration. By averaging (13) over all fading blocks, it is easy to see that constraint (6) on the input samples will be satisfied if and only if the power allocation function $p(\gamma)$ satisfying (9). Moreover, since (14) must be satisfied during all fading blocks, we also see that condition (7) is equivalent to condition (10).

**Remark:** One could consider as in [4], [5] a scenario without the restriction to block-fading introduced in Sec. II, but still with perfect transmitter and receiver CSI (with e.g. $h_k$ i.i.d.). We conjecture that the capacity of channel (1) under such fading assumptions and average and peak power constraints (6) and (7) is given by (8) with $p(\gamma)$ having to satisfy constraints (9) and (10). In order to establish this, one would have to extend the coding theorem and corresponding converse from [4] or [5] to the case where the input signal is also subject to a peak power constraint. Although we believe this can be done without any particular difficulties, we have chosen not to do so because in our opinion such a scenario is less realistic than the one considered in this paper for the reasons given in Sec. II.

We outline in Appendix II the arguments needed to prove that if $p^*(\gamma)$ is to be a supremum of (8) subject to constraints (9) and (10), it must be continuous and such that
\begin{equation}
p^*(\gamma) = \begin{cases} 
0 & \text{if } \gamma \leq \lambda^* \\
\gamma \cdot C'(\gamma \cdot p^*(\gamma), \gamma \cdot p_{\text{peak}}) & \text{if } \gamma > \lambda^* \land \lambda^* \cdot C'_{p_{\text{peak}}} < \frac{\lambda^*}{\gamma} \\
p_{\text{peak}} & \text{if } \gamma > \lambda^* \land \lambda^* \cdot C'_{p_{\text{peak}}} \geq \frac{\lambda^*}{\gamma}
\end{cases} \tag{15}
\end{equation}
for all $\gamma \in (0, \infty)$, where $C'(\rho, \gamma \cdot p_{\text{peak}}) \triangleq \frac{\partial}{\partial \rho} C(\rho, \gamma \cdot p_{\text{peak}})$, $C'_{p_{\text{peak}}} \triangleq C'(\gamma \cdot p_{\text{peak}}, \gamma \cdot p_{\text{peak}})$, and $\lambda^*$ is such that $\int_{\lambda^*}^{\infty} p^*(\gamma) f_{\gamma}(\gamma) d\gamma = 1$.

It is also important to observe that
\begin{equation}
C'(0, \gamma \cdot p_{\text{peak}}) = 1, \tag{16}
\end{equation}
since when $\rho \ll \gamma \cdot p_{\text{peak}}$, $C(\rho, \gamma \cdot p_{\text{peak}})$ is approximately equal to the capacity $\log(1 + \rho)$ [nats/channel use] of a non-fading quadrature additive Gaussian noise channel with unit average noise power and average input power constraint $\rho$.

Note additionally that the concavity of $C(\rho, \gamma \cdot p_{\text{peak}})$ implies that $C'(\rho, \gamma \cdot p_{\text{peak}})$ is decreasing in $\rho$.

Finally, note that in the special case where the peak power constraint is absent (i.e. when $p_{\text{peak}} = \infty$ and $C'(\gamma \cdot p(\gamma), \infty) = \frac{1}{1 + \gamma p(\gamma)}$), the solution $p^*(\gamma)$ satisfying (15) becomes simply
\begin{equation}
p^*(\gamma) = \max \left( \frac{1}{\lambda^*} - \frac{1}{\gamma}, 0 \right) \tag{17}
\end{equation}
for all $\gamma \in (0, \infty)$, where $\lambda^*$ is such that $\int_{\lambda^*}^{\infty} \left( \frac{1}{\lambda^*} - \frac{1}{\gamma} \right) f_{\gamma}(\gamma) d\gamma = 1$. This is exactly the water-pouring solution established in [4], which is capacity-optimal in the case of stationary and ergodic fading channels with perfect transmitter and receiver CSI (the constant $\lambda^*$ corresponds to the cutoff SNR value $\gamma_0$ from [4]).

**IV. Numerical Simulations**

By using (15), we numerically calculated an approximation for the rate-optimal normalised allocated power $p^*(\gamma)$ which achieves the supremum in (8) subject to constraints (9) and (10) in different cases. A Nakagami-$m$ block-fading channel [8] was assumed, resulting in a pre-adaptation SNR $\gamma$ per block distributed according to the law
\begin{equation}
f_{\gamma}(\gamma) = \left( \frac{m}{\gamma_{av}} \right)^{m-1} \frac{\Gamma(m)}{\Gamma(m)} \exp \left( - \frac{m}{\gamma_{av}} \frac{\gamma}{\gamma_{av}} \right), \tag{18}
\end{equation}
where $m$ is the fading severity parameter (the special case $m = 1$ corresponds to a Rayleigh-fading channel), $\gamma_{av}$ is the average pre-adaptation SNR, and $\Gamma(\cdot)$ denotes the Gamma function [9].

We now briefly describe the details of the numerical optimisation procedure. The main difficulty lies in the computation of $C'(P_{av}, P_{\text{peak}})$ for different values of $P_{av}$ and $P_{\text{peak}}$. An approximation for this quantity can be obtained via the centred finite difference
\begin{equation}
C'(P_{av}, P_{\text{peak}}) \approx \frac{C(P_{av} + \Delta, P_{\text{peak}}) - C(P_{av} - \Delta, P_{\text{peak}})}{2\Delta}, \tag{19}
\end{equation}
where we chose to set $\Delta = 10^{-5}$ after some experimentation. Note that in order for (19) to yield a good approximation for $C'(P_{av}, P_{\text{peak}})$, it is imperative that $C(P_{av} + \Delta, P_{\text{peak}})$ and $C(P_{av} - \Delta, P_{\text{peak}})$ be computed with good accuracy.

In order to evaluate $C(P_{av}, P_{\text{peak}})$ for given $P_{av}$ and $P_{\text{peak}}$, one needs to find the positions and probabilities of the mass points of the capacity-achieving magnitude distribution [3]. This can be done by using steepest descent-like algorithms (routines from the CFSQP optimisation software [10] were used to obtain the results presented in this section) together with the necessary and sufficient optimality condition provided in [3, p. 1064]. Numerical experimentation showed that this can only be done with sufficient accuracy for our purposes when the capacity-achieving magnitude distribution possesses up to six or seven mass points. The reason for this is that as
the number of mass points in the capacity-achieving magnitude distribution increases, the optimality condition from [3] becomes more difficult to verify due to the growing effect of numerical inaccuracies.

The number of mass points in the capacity-achieving magnitude distribution increases as $P_{\text{av}} \to 0$ for fixed $P_{\text{peak}}$, and as $P_{\text{av}}$ and $P_{\text{peak}}$ approach infinity while the ratio $P_{\text{av}}/P_{\text{peak}}$ remains constant. The approximations

$$C(P_{\text{av}}, P_{\text{peak}}) \approx \log(1 + P_{\text{av}}),$$

$$(20)$$

$$C'(P_{\text{av}}, P_{\text{peak}}) \approx 1/(1 + P_{\text{av}}),$$

$$(21)$$
corresponding to a quadrature additive Gaussian channel which is subject to a quadrature additive Gaussian channel

In order to find an approximation for $p^*(\gamma)$ for given $f_2(\gamma)$ and $p_{\text{peak}}$, we proceed by first tabulating approximations for

$$\gamma \cdot C'(\gamma \cdot p(\gamma), \gamma \cdot p_{\text{peak}})$$

for a number of different $\gamma \in [0, 30]$ belonging to a suitably chosen set $2 G$, and for $p(\gamma) \in \{0, 2P_{\text{peak}}/100, 5P_{\text{peak}}/100, \ldots, P_{\text{peak}}\}$. This can be done by using (19)–(21) and following the procedure described above. Fig. 1 shows plots of this quantity for $p_{\text{peak}} = 2$ and $\gamma \in \{1, 2, 6\}$. For comparison, plots of $\gamma \cdot C'(\gamma \cdot p(\gamma), \infty)$ (corresponding to a quadrature additive Gaussian channel which is subject to an average power constraint only) have been included as well. Fig. 1 confirms that $\gamma \cdot C'(\gamma \cdot p(\gamma), \gamma \cdot p_{\text{peak}})$ is a monotonously decreasing function of $p(\gamma)$ (this is to be expected since $C(P_{\text{av}}, P_{\text{peak}})$ is concave in $P_{\text{av}}$ for fixed $P_{\text{peak}}$ as explained in Sec. III), and also shows that $\gamma \cdot C'(\gamma \cdot p(\gamma), \infty)$ is a good approximation for $\gamma \cdot C'(\gamma \cdot p(\gamma), \gamma \cdot p_{\text{peak}})$ when $P_{\text{peak}}$ is small.

Once this has been done, one can find an approximation for the solution $p^*(\gamma)$ of (15) $\forall \gamma \in G$ and any given $\lambda^*$, and subsequently an approximation for $\int_0^\infty p^*(\gamma)f_2(\gamma)d\gamma$. The latter approximation can be computed by assuming that $p^*(\gamma)$ varies linearly between the values $\gamma \in G$ for which it is approximately known, and by neglecting $\int_{\gamma_{\text{max}} \in G} p^*(\gamma)f_2(\gamma)d\gamma$. (The value of the latter integral is always negligible in the case of the results presented in this section.) Numerical root-finding techniques can therefore be used to find an approximation for the value of $\lambda^*$ such that $\int_0^\infty p^*(\gamma)f_2(\gamma)d\gamma = 1$, and an approximation for the corresponding rate-optimal normalised power allocation function $p^*(\gamma)$ follows.

The resulting $p^*(\gamma)$ has been plotted in Fig. 2 for $P_{\text{peak}} = 2$, values of the fading severity parameter $m \in \{1, 2\}$, and values of the average pre-adaptation SNR $\gamma_{\text{av}} \in \{\frac{1}{2}, 1, 2\} = \{-3.01 \ldots \text{dB}, 0 \text{dB}, 3.01 \ldots \text{dB}\}$; and in Fig. 3

$3$ We set $G = G_0 \triangleq \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, 16.0, 17.0, 18.0, 19.0, 20.0, 25.0, 30.0\}$ when $P_{\text{peak}} = 2$, and $G = G_0 \triangleq \{20.0, 25.0, 30.0\}$ when $P_{\text{peak}} = 3$. The points in set $G$ are chosen in a way which is approximately such that they be closer to each other in intervals where $p^*(\gamma)$ changes significantly than in intervals where $p^*(\gamma)$ does not change as much.
These figures also show that the behaviour of the rate-optimal power allocation function when the input is subject to an average and peak power constraints is the fact that for fixed pre-adaptation SNR probability density functions \(f(x)\), the sharpness of the peak in the rate-optimal normalised power allocation function \(p^*(\gamma)\) becomes less pronounced as \(p_{peak}\) increases.

A striking and important difference between the rate-optimal power adaptation strategy when the input is subject to an average power constraint only (see Fig. 4) – also corresponding to the water-pouring solution from [4] – and the rate-optimal power adaptation strategy when the input is simultaneously subject to average and peak power constraints is the fact that in the latter case the rate-optimal power allocation function is not necessarily a monotonously increasing function of \(\gamma\) (see Figs. 2 and 3). This can also be seen by observing in Fig. 1 that the solid lines corresponding to \(\gamma \cdot C(\gamma \cdot p(\gamma), \gamma \cdot P)\) for \(\gamma \in \{1, 2, 6\}\) cross, whereas the dashed lines corresponding to \(\gamma \cdot C(\gamma \cdot p(\gamma), \infty)\) for \(\gamma \in \{1, 2, 6\}\) do not.

V. CONCLUSIONS

In this paper, we characterised by means of a theorem the power adaptation strategy which maximises the maximum average information rate that can be reliably transmitted over average and peak power constrained block-fading Gaussian channels in which perfect transmitter and receiver channel state information (CSI) are available, and in which any transmitted codeword spans a single fading block. We subsequently numerically computed it under different block-fading scenarios and different values of the normalised peak power \(P_{peak}\). Our results reveal that, as opposed to the case of Gaussian stationary and ergodic fading channels with perfect transmitter and receiver CSI and an input which is only subject to an average power constraint (when the well known water-pouring solution is capacity-achieving), the rate-optimal power allocation function is not necessarily a monotonously increasing function of the pre-adaptation SNR.

APPENDIX I

CONVEXITY OF \(C_{info}(P_{av}, P_{peak})\) FOR FIXED \(P_{peak}\)

Consider the discrete-time quadrature additive Gaussian noise channel

\[ Y_k = X_k + W_k, \tag{22} \]

where \(Y_k, X_k,\) and \(W_k\) respectively denote the channel output, input, and additive noise at time \(k\). The sequence \(\{W_k\}\) is assumed to consist of zero mean, unit variance i.i.d. circularly symmetric complex Gaussian random variables. In this appendix, we show that the information capacity

\[ C_{info}(P_{av}, P_{peak}) = \lim_{n \to \infty} \sup \frac{1}{n} I(X^n; Y^n) \tag{23} \]

of channel (22), with \(X^n \triangleq \{X_1, \ldots, X_n\}, Y^n \triangleq \{Y_1, \ldots, Y_n\}\), and where the supremum is taken over all probability density functions \(\phi_{X^n}(x^n)\) such that

\[ \frac{1}{n} \sum_{i=1}^{n} \int_C |x_i|^2 \phi_{X_i}(x_i) \, dx_i \leq P_{av} \quad \text{(24)} \]

and

\[ \phi_{X_i}(x_i) = 0 \text{ if } |x_i|^2 > P_{peak} \quad \forall i \in \{1, \ldots, n\}, \quad \text{(25)} \]

is a concave function of \(P_{av}\) for fixed \(P_{peak}\). (Note that for memoryless channels, definition (23) reduces to

\[ C_{info}(P_{av}, P_{peak}) = \sup_{\phi_X(x)} I(X; Y), \quad \text{(26)} \]

where the supremum is now taken over all probability density functions \(\phi_X(x)\) such that \(\int_C |x|^2 \phi_X(x) \, dx \leq P_{av}\) and \(\phi_X(x) = 0 \text{ if } |x|^2 > P_{peak}\). We need to show that for any \(P_{av1}, P_{av2} \geq 0\) and any \(0 \leq \lambda \leq 1,\) we have

\[ C_{info}(P_{av}, P_{peak}) \geq \lambda C_{info}(P_{av1}, P_{peak}) + (1 - \lambda) C_{info}(P_{av2}, P_{peak}), \quad \text{(27)} \]

where \(P_{av} = \lambda P_{av1} + (1 - \lambda) P_{av2}\). Let \(\phi_{1,X^n}(x^n)\) and \(\phi_{2,X^n}(x^n)\) be the distributions achieving the supremum in (23) respectively when \(P_{av} = P_{av1}\) and \(P_{av} = P_{av2}\). Since
channel (22) is memoryless, we have that $\phi_{k,X_1}(x_1) = \phi_{k,X_2}(x_2) = \cdots = \phi_{k,X_n}(x_n)$ for $k \in \{1, 2\}$, i.e., $X_1,\ldots,X_n$ are i.i.d according to $\phi_{k,X_i}(x_1)$ for $k \in \{1, 2\}$. Let us now define
\begin{equation}
\phi_{\lambda,X^n}(x^n) = \prod_{i=1}^{[n\lambda]} \phi_{1,X_i}(x_i) \prod_{i=[n\lambda]+1}^{n} \phi_{2,X_i}(x_i)
\end{equation}
(28)

Note that $[n\lambda] = n\lambda$ for $n$ sufficiently large. We then have
\begin{align*}
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ |x_i|^2 \phi_{\lambda,X^n}(x^n) \right] &\leq \lim_{n \to \infty} \frac{1}{n} \left[ n\lambda |P_{av_1}| + (n - |n\lambda|)P_{av_2} \right] \\
&= \lambda P_{av_1} + (1 - \lambda)P_{av_2} = P_{\lambda},
\end{align*}
(29)
which means that the average power constraint is satisfied. Note that the peak power constraint is also trivially satisfied when $X^n$ is distributed according to $\phi_{\lambda,X^n}(x^n)$.

Moreover, when $X^n \sim \phi_{\lambda,X^n}(x^n)$, we have that
\begin{align*}
\lim_{n \to \infty} \frac{1}{n} \int I(X^n, Y^n) &\leq \lim_{n \to \infty} \frac{1}{n} \left[ \sum_{i=1}^{n} I(X_i, Y_i) + \sum_{i=[n\lambda]+1}^{n} I(X_i, Y_i) \right] \\
&= \lim_{n \to \infty} \frac{1}{n} \left[ n\lambda C_{info}(P_{av_1}, P_{peak}) \\
&\quad + (n - |n\lambda|) C_{info}(P_{av_2}, P_{peak}) \right] \\
&= \lambda C_{info}(P_{av_1}, P_{peak}) + (1 - \lambda)C_{info}(P_{av_2}, P_{peak}),
\end{align*}
(30)
and (27) follows immediately.

**Appendix II**

**Characterisation of $p^*(\gamma)$**

In this appendix, we discuss how to find a $p(\gamma)$ for which the supremum in (8), subject to constraints (9) and (10), is attained. This is equivalent to finding a $p(\gamma)$ minimising
\begin{equation}
-\int_{0}^{\infty} C(\gamma \cdot p(\gamma), \gamma \cdot p_{peak}) f_\gamma(\gamma) \ d\gamma
\end{equation}
(31)
subject to
\begin{align*}
\int_{0}^{\infty} p(\gamma) f_\gamma(\gamma) \ d\gamma &= 1, \\
-p(\gamma) &\leq 0,
\end{align*}
(32)
(33)
and
\begin{equation}
p(\gamma) - p_{peak} \leq 0,
\end{equation}
(34)
where, as assumed in Sec. III, $f_\gamma(\gamma) \neq 0$ for all $\gamma \in (0, \infty)$. (It is elementary to show that (9) must be satisfied with equality whenever $P_{av} \leq P_{peak}$.) Note that $J_1(p) \triangleq p(\gamma) f_\gamma(\gamma)$, $J_2(p) \triangleq -p(\gamma)$, and $J_3(p) \triangleq p(\gamma) - p_{peak}$ are convex in $p$ [11, p. 63]. In addition, since $C(\rho, \gamma \cdot p_{peak})$ is a concave function of $\rho$ for fixed $\gamma \cdot p_{peak}$ as explained in Sec. III, it follows [11, p. 63] that $-C(\gamma \cdot p, \gamma \cdot p_{peak}) f_\gamma(\gamma)$ is convex in $p$. Consequently [11], we can solve this problem by minimising
\begin{align*}
J(p) &= \int_{0}^{\infty} \left[ -C(\gamma \cdot p(\gamma), \gamma \cdot p_{peak}) f_\gamma(\gamma) + \lambda p(\gamma) f_\gamma(\gamma) \\
&\quad -\mu_1(\gamma)p(\gamma) + \mu_2(\gamma) \left[ p(\gamma) - p_{peak} \right] \right] \ d\gamma,
\end{align*}
(35)
where $\lambda \in \mathbb{R}$ is such that condition (32) is satisfied, and $\mu_1(\gamma)$ and $\mu_2(\gamma)$ are nonnegative and such that, for all $\gamma \in (0, \infty)$,
\begin{align*}
\mu_1(\gamma)p(\gamma) &= 0, \\
\mu_2(\gamma) \left[ p(\gamma) - p_{peak} \right] &= 0.
\end{align*}
(36)
(37)
In order to minimise (35), it can be shown that $p(\gamma)$ should satisfy the Euler-Lagrange equation [11]
\begin{equation}
\gamma \cdot C'(\gamma \cdot p(\gamma), \gamma \cdot p_{peak}) - \lambda \gamma f_\gamma(\gamma) = -\mu_1(\gamma) + \mu_2(\gamma)
\end{equation}
(38)
in all intervals excluding a corner point [11], where we have introduced the notation $C'(\rho, \gamma \cdot p_{peak}) \triangleq \frac{\partial}{\partial \rho} C(\rho, \gamma \cdot p_{peak})$. Furthermore, it can also be shown that $p(\gamma)$ should be continuous at each corner point [11, p. 207] if it is to minimise (35).

By analysing (32)–(34) together with (36)–(38), we find after a minute of thought that any solution $p^*(\gamma)$ minimising (31) subject to the constraints (32)–(34) should be such that
\begin{equation}
p^*(\gamma) = 0 \quad \text{if} \quad \gamma \leq \lambda^* \\
p^*(\gamma) = \lambda^* \quad \text{if} \quad \gamma > \lambda^* \wedge C'(\gamma \cdot \gamma \cdot p_{peak} < \lambda^*) \\
p^*(\gamma) = \gamma \cdot p_{peak} \quad \text{if} \quad \gamma > \lambda^* \wedge C'(\gamma \cdot \gamma \cdot p_{peak} \geq \lambda^*)
\end{equation}
(39)
for all $\gamma \in (0, \infty)$, where $C'(\gamma \cdot \gamma \cdot p_{peak} \triangleq C'(\gamma \cdot \gamma \cdot p_{peak})$, and $\lambda^*$ is such that $\int_{0}^{\infty} p^*(\gamma) f_\gamma(\gamma) \ d\gamma = 1$.

This can be seen by observing that $C'(\gamma \cdot \gamma \cdot p_{peak}) \leq 1$ by virtue of the fact that $C'(0, \gamma \cdot p_{peak}) = 1$ and that $C'(\rho, \gamma \cdot p_{peak})$ is decreasing in $\rho$ as explained in Sec. III, and combining this with the fact that $\mu_1(\gamma) = 0$ ($\mu_2(\gamma) = 0$) whenever $p(\gamma) > 0$ ($p(\gamma) < p_{peak}$).

**References**