Article

On the Capacity and the Optimal Sum-Rate of a Class of Dual-Band Interference Channels [†]

Subhajit Majhi * and Patrick Mitran

Department of Electrical and Computer Engineering, University of Waterloo, ON N2L 3G1, Canada;
pmitran@uwaterloo.ca

* Correspondence: smajhi@uwaterloo.ca; Tel.: +1-519-888-4567 (ext. 31779)
† This paper is an extended version of our paper published in the 2016 IEEE International Symposium on Information Theory, Barcelona, Spain, 10–15 July 2016.

Abstract: We study a class of two-transmitter two-receiver dual-band Gaussian interference channels (GIC) which operates over the conventional microwave and the unconventional millimeter-wave (mm-wave) bands. This study is motivated by future 5G networks where additional spectrum in the mm-wave band complements transmission in the incumbent microwave band. The mm-wave band has a key modeling feature: due to severe path loss and relatively small wavelength, a transmitter must employ highly directional antenna arrays to reach its desired receiver. This feature causes the mm-wave channels to become highly directional, and thus can be used by a transmitter to transmit to its designated receiver or the other receiver. We consider two classes of such channels, where the underlying GIC in the microwave band has weak and strong interference, and obtain sufficient channel conditions under which the capacity is characterized. Moreover, we assess the impact of the additional mm-wave band spectrum on the performance, by characterizing the transmit power allocation for the direct and cross channels that maximizes the sum-rate of this dual-band channel. The solution reveals conditions under which different power allocations, such as allocating the power budget only to direct or only to cross channels, or sharing it among them, becomes optimal.

Keywords: Gaussian interference channel; dual-band network; optimal power allocation; millimeter wave band

1. Introduction

Current technology such as 4G (e.g., [1]) is rapidly becoming inadequate to support the exponential growth in wireless traffic [2]. Moreover, the potential for improvement of network throughput in 4G is limited due to the shortage of spectrum in the incumbent microwave band (i.e., carrier frequencies below 6 GHz). Thus, methods to tackle the ever growing amount of mobile traffic has become a key research area (see [3] and references therein). Several new technologies are being considered, among which employing additional spectrum in the 28–300 GHz frequency range, often referred to as the millimeter wave (mm-wave) band, seems to be a promising solution to the problem of spectrum scarcity [3,4]. Specifically, integrating additional spectrum from the mm-wave band to complement transmissions in the microwave band is poised to play a central role in the functioning of 5G networks.

Transmission in the mm-wave band is distinctly different from that in the microwave band. Due to higher operating frequencies, omnidirectional transmission in the mm-wave band is subject to much higher absorption and power loss [5] compared to that in the microwave band, and thus a transmitter needs to employ beamforming with highly directional antenna arrays to counter this loss and reach its receiver [2]. However, beamforming constrains most of the transmission energy to the line-of-sight (LOS) component and very few, if any, significant non-LOS components exist [6,7]. Thus, transmission

in this band is highly directional and point-to-point. Such mm-wave channels can support high data rates due to their vast bandwidth, but nevertheless are prone to blockage and absorption [8] due to their point-to-point nature. In contrast, microwave links are much more reliable due to rich scattering and diffraction, but cannot support as high rates as the mm-wave links. Thus, in a dual-band setting, conventional traffic and control information can be reliably communicated over microwave links, and high data-rate traffic can be sent through mm-wave links [2,7,9–15].

Most studies on microwave and mm-wave dual-band transmission have focused on how to improve network layer performance metrics of cellular access or backhaul networks by using the high-bandwidth, highly directional mm-wave links [9,13]. For example, the authors in [7] posed the problem of optimal resource allocation in a cellular setting in the dual microwave and mm-wave band, and showed that certain network level performance parameters, e.g., the number of simultaneously supported users and the link connection probability, are vastly improved with their proposed solution. Similarly, the authors in [11] study a two-tier cellular network where the 60 GHz band is used to create point-to-point directional links, and the 70 GHz band is used to establish long range connections just as the microwave band is used in our work, and propose a hybrid scheme involving both bands that improves the network throughput. In a more related article [14], the authors characterize the benefits of beamforming over a point-to-point dual band multi-antenna channel, and study the performance of a hybrid adaptive queueing scheme over both bands that maximizes the delay-constrained throughput.

Recent works on joint transmission in the microwave band and mm-wave band [14,16–20] indicate that it is possible to isolate the transmissions in the microwave band from that in the mm-wave band, and communicate over the two bands simultaneously. For example, the authors in [14] designed a queue-based scheme that transmits from a single transmitter to a single receiver simultaneously over the 3 GHz microwave band and the 30 GHz mm-wave band using beamforming, and conducted successful practical experiments in this dual-band setup. In addition, Intel has recently announced the production of a dual-band modem that supports both sub-6 GHz and 28 GHz bands [16]. Moreover, recent works on resource allocation in the microwave and mm-wave dual-band setting in [19,20], and their variant in [21], show that simultaneous transmission in both microwave and mm-wave bands are indeed feasible, and are gaining acceptance as an architecture for cellular access in 5G.

We study here the performance of a two-transmitter two-receiver (2×2) dual-band interference channel from an information theoretic perspective, where a transmitter communicates to its respective receiver over the microwave band and the mm-wave band simultaneously. In the microwave band, each receiver observes the superposition of signals from both transmitters as in a conventional single band Gaussian interference channel (GIC) [22]. However, as the mm-wave channels are considered to be highly directional, a transmitter in this band is well-modeled as being able to transmit towards one intended receiver, while causing negligible to no interference to the other receiver [6]. This raises the question: to which receiver should a transmitter in the mm-wave band transmit? In this 2×2 GIC, the transmitters in the mm-wave band can transmit from: (a) the first transmitter (Tx_1) to the first receiver (Rx_1), and the second transmitter (Tx_2) to the second receiver (Rx_2); (b) Tx_1 to Rx_2, and Tx_2 to Rx_1; (c) Tx_1 and Tx_2 to Rx_2; or (d) Tx_1 and Tx_2 to Rx_1. We focus here on the first two cases where the transmitters in the mm-wave band either transmit (a) from Tx_1 to Rx_1, and from Tx_2 to Rx_2, i.e., in the *direct* channels; or (b) from Tx_1 to Rx_2, and from Tx_2 to Rx_1, i.e., in the *cross* channels; or (c) share the spectrum between the two modes. We denote the resulting channels by Direct-Link IC (DLIC), Cross-Link IC (CLIC) and Direct-and-Cross-Link IC (DCLIC), respectively, and study their capacity.

The capacity of the conventional single band GIC [22] has been characterized when it has strong [22] or very strong interference [23]. We know that, under strong interference, encoding the message at each transmitter with independent Gaussian distributed codewords, and decoding both the desired and the interfering messages at each receiver is the optimal strategy [22]. However, if the GIC has weak interference [24], the capacity and the optimal strategy is still unknown in general.

The 2×2 parallel Gaussian IC (PGIC) consists of several orthogonal parallel channels (sub-channels) such that a 2×2 conventional GIC operates in each sub-channel without interfering with that in other sub-channels [25]. The optimal strategies for the PGIC are not known in general; however, its capacity was characterized in [26] when the GIC in each sub-channel has strong interference. In a related study, the capacity of the ergodic fading GIC was characterized in [27] when each fading state has strong interference. In the dual-band GIC considered here, the number of channel uses in the microwave band and the mm-wave band may differ. We model this with a bandwidth mismatch factor (BMF) in the system model. Note that the dual-band GIC considered here is a special case of the ergodic fading GIC in [27] if one identifies each fading state as a different sub-channel. In the special case that the bandwidth mismatch factor between the microwave and mm-wave bands is 1, it is also a special case of the PGIC in [26].

Moreover, the studies in [26,27] show that if every sub-channel (or fading state) has strong interference, the capacity is achieved by encoding jointly over all sub-channels and decoding messages from both transmitters. Encoding independently over each sub-channel of a PGIC is suboptimal in general [28], except for the GIC in the very weak (noisy) interference regime [29]. In fact, joint encoding over all sub-channels generally achieves better rates as it can potentially offset the weak interference in one or more sub-channels if the other sub-channels have strong interference [28].

In this sequel, we study the capacity of the DCLIC and the CLIC. First, we present a useful result that decomposes the capacity of the DCLIC into that of the underlying CLIC and the set of direct channels. This result shows that the capacity of the DCLIC can be established if the capacity of a corresponding CLIC is known. Hence, we focus on the CLIC next. In particular, we consider two specific classes, the strong CLIC and the weak CLIC, where the underlying GIC in the microwave band has strong and weak interference, respectively, and characterize sufficient conditions on the channel parameters under which their capacity is established.

Resource allocation techniques that maximize the sum-rate or throughput of interference channels have been investigated throughly (e.g., see [25,30,31]) as they indicate how to allocate resources in practice. The DCLIC models a basic multiuser network over the microwave and mm-wave dual-band, whose performance is likely to be dominated by the mm-wave channels due to their large bandwidths [9]. Thus, it is useful to understand how to optimize the performance of the DCLIC over the parameters in the mm-wave band. Therefore, we study the power allocation in the direct and cross channels in the mm-wave band that maximizes the sum-rate of the DCLIC. We derive the optimal powers in closed form, and study how the channel parameters influence the optimal power allocation.

The contributions of this paper are summarized as follows:

- We show that the capacity region of the DCLIC can be decomposed into the capacity region of the underlying CLIC and two non-interfering direct links in the mm-wave band. This illustrates that the cross channels are actively involved in characterizing the capacity of the CLIC, whereas the direct channels improve the rates of individual users.

- We characterize the capacity of the strong CLIC, and observe that the strong interference condition in the microwave band is sufficient to characterize the capacity.

- For the weak CLIC, we characterize sufficient channel conditions under which its capacity is established. This shows that even if the GIC in the microwave band has weak interference, adequately strong cross channels in the mm-wave band are sufficient to characterize the capacity.

- We characterize the optimal power allocation in the direct and cross channels that maximizes the sum-rate of the DCLIC, and study channel conditions under which the optimal power allocation either assigns the entire power budget to a specific subset of channels, or shares the power budget among all channels. We establish a direct relation between the channel parameters and the optimal powers, from which we observe the following:

  - The optimal power allocation distributes the power budget among the direct and cross channels following two properties: a *waterfilling-like* property and a *max-min* property.

- When the power budget is sufficiently small, the optimal allocation assigns power to either both direct channels, or both cross channels and at most one direct channel.
- Due to the max-min property, the optimal allocation imposes a maximum limit on the cross-channel powers. When the power budget exceeds a certain threshold, the limit on the cross-channel powers are reached, and all additional increments to the power budget are then added only to the direct channels that do not have such limits.
- If the underlying GIC in the microwave band has *very strong* interference, the optimal power allocation assigns the power budget entirely to the direct channels.
- If the channel parameters satisfy one of the following criteria, then transmitting only in the direct channels is approximately optimal, in the sense that the difference between the sum-rates resulting from allocating to only direct channels and allocating optimally in all channels, is negligibly small: (a) the transmit powers in the underlying GIC in the microwave band is very small; or (b) the cross channel gains in the mm-wave band are very large.

The rest of the paper is organized as follows. We define the system model in Section 2. We present the decomposition result on the DCLIC and the capacity of the strong CLIC in Section 3. We present the capacity result on the weak CLIC in Section 4. In Section 5, we formulate the optimum sum-rate problem and discuss its solution, and we conclude in Section 6.

Notation: We denote sets in calligraphic (e.g., $\mathcal{A}_k$), except the sets of reals, positive reals, and positive integers, which are denoted by $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{Z}_+$, respectively. Vectors are in bold (e.g., **x**), and $\mathbf{x} \succeq \mathbf{0}$ denotes that each element of **x** is in $\mathbb{R}_+$, where **0** is the zero vector. We denote by $\mathbb{E}(Y)$ the statistical expectation of a random variable (RV) Y, and by $X \sim \mathcal{N}(\mu, \sigma^2)$ an RV following the Gaussian distribution with mean μ and variance σ^2. We also denote an n-length vector $(X_1, X_2, \ldots, X_n)$ by X^n, the empty set by $\varnothing$, $\log_2 x$ by $\log x$, and define $\mathsf{C}(x) := \frac{1}{2} \log(1 + x)$.

2. System Model

First, we define the discrete memoryless (DM) model of the DCLIC, and from that we define the Gaussian models. In the DCLIC, a bandwidth mismatch between the first (microwave) band and the second (mm-wave) band may exist. Thus, we assume that during n accesses of the first band, the second band is accessed $n_1(n)$ times as cross channels, and $n_2(n)$ times as direct channels. We model this by two bandwidth mismatch factors (BMF) $\alpha_k := \lim_{n \to \infty} n_k(n)/n, k = 1, 2$. Thus, the total normalized channel uses in the second band is asymptotically $\alpha := \alpha_1 + \alpha_2$, which is shared between the cross channels (α_1) and direct channels (α_2). For ease of exposition, we denote $n_k(n)$ by $n_k, k = 1, 2$.

The 2×2 DM interference channel is defined by $((\mathcal{X}_k, \mathcal{Y}_k)_{k=1}^2, p(y_1, y_2 | x_1, x_2))$ where $\mathcal{X}_k$ and $\mathcal{Y}_k$ are the discrete input and output alphabets of user k, $k = 1, 2$, and $p(y_1, y_2 | x_1, x_2)$ is the set of channel transition probabilities [32] (Chapter 6.1). We define the 2×2 DM DCLIC similarly by the tuple, $((\mathcal{X}_k, \hat{\mathcal{X}}_k, \bar{\mathcal{X}}_k, \mathcal{Y}_k, \hat{\mathcal{Y}}_k, \bar{\mathcal{Y}}_k)_{k=1}^2, p)$, where $\mathcal{X}_k$ and $\mathcal{Y}_k$ are the input and output alphabets of the interference channel in the first band, $\bar{\mathcal{X}}_k$ and $\bar{\mathcal{Y}}_k$ are the input and output alphabets for the Tx_k to Rx_k direct channels in the second band, $k = 1, 2$, and $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{Y}}_2$ (respectively $\hat{\mathcal{X}}_2$ and $\hat{\mathcal{Y}}_1$) are the input and output alphabets for the Tx_1 to Rx_2 (resp. Tx_2 to Rx_1) cross channel in the second band. The joint probability mass function (pmf) of the DCLIC decomposes as

$$p(x_1^n, \hat{x}_1^{n_1}, \bar{x}_1^{n_2}) p(x_2^n, \hat{x}_2^{n_1}, \bar{x}_2^{n_2}) \prod_{\ell=1}^{n} p(y_{1\ell}, y_{2\ell} | x_{1\ell}, x_{2\ell}) \prod_{\ell=1}^{n_1} p(\hat{y}_{1\ell} | \hat{x}_{2\ell}) p(\hat{y}_{2\ell} | \hat{x}_{1\ell}) \prod_{\ell=1}^{n_2} p(\bar{y}_{1\ell} | \bar{x}_{1\ell}) p(\bar{y}_{2\ell} | \bar{x}_{2\ell}).$$

We define a $(2^{nR_1}, 2^{nR_2}, n, n_1, n_2)$ code for the DCLIC to consist of (a) two uniformly distributed and independent message sets, $\mathcal{M}_1 := \{1, \ldots, 2^{nR_1}\}$ and $\mathcal{M}_2 := \{1, \ldots, 2^{nR_2}\}$, respectively, for Tx_1 and Tx_2; (b) two encoding functions, ϕ_1 and ϕ_2, respectively for Tx_1 and Tx_2 such that

$\phi_k : \mathcal{M}_k \to \mathcal{X}_k^n \times \hat{\mathcal{X}}_k^{n_1} \times \bar{\mathcal{X}}_k^{n_2}$, $k = 1, 2$; and (c) two decoding functions, ψ_1 and ψ_2, respectively, for Rx_1 and Rx_2 such that $\psi_k : \mathcal{Y}_k^n \times \hat{\mathcal{Y}}_k^{n_1} \times \bar{\mathcal{Y}}_k^{n_2} \to \mathcal{M}_k, k = 1, 2$.

We define the decoding probability of error at Rx_k by $P_{e,k}^n := \Pr\big(\psi_k(Y_k^n, \hat{Y}_k^{n_1}, \bar{Y}_k^{n_2}) \neq M_k\big), k = 1, 2$. We say that (R_1, R_2) is an achievable rate pair of the DCLIC, if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n, n_1, n_2)$ codes such that $n_k \leq \alpha_k n$, and $P_{e,k}^n \to 0$ as $n, n_k \to \infty$ for $k = 1, 2$. The capacity region of the DCLIC is defined as the closure of the set of all achievable rate tuples.

Following the above notations, the GIC in the first band of the Gaussian DCLIC is modeled as in [22]. The signals received at Rx_1 and Rx_2 are given by

$$Y_{1\ell} = X_{1\ell} + a_{21} X_{2\ell} + Z_{1\ell}, \tag{1}$$

$$Y_{2\ell} = X_{2\ell} + a_{12} X_{1\ell} + Z_{2\ell}, \quad \ell = 1, \dots, n, \tag{2}$$

where $X_{k\ell}, Y_{k\ell} \in \mathbb{R}$, $a_{km} \in \mathbb{R}$ are coefficients of the channels from Tx_k to Rx_m, $k \neq m \in \{1, 2\}$ (a_{kk}^2 are normalized to 1 as in [22]), and $Z_{k\ell} \sim \mathcal{N}(0, 1)$ are i.i.d. noise, $k = 1, 2$. In addition, the codewords now satisfy the average power constraint, $\frac{1}{n} \sum_{\ell=1}^{n} X_{k\ell}^2 \leq Q_k, k = 1, 2$. The cross links in the second band of the DCLIC are point-to-point, and are modeled as

$$\hat{Y}_{2\ell} = c_{12} \hat{X}_{1\ell} + \hat{Z}_{2\ell}, \tag{3}$$

$$\hat{Y}_{1\ell} = c_{21} \hat{X}_{2\ell} + \hat{Z}_{1\ell}, \quad \ell = 1, \dots, n_1, \tag{4}$$

where $\hat{X}_{k\ell}, \hat{Y}_{k\ell} \in \mathbb{R}$, $c_{km} \in \mathbb{R}$ are coefficients of the channels from Tx_k to Rx_m, $k \neq m \in \{1, 2\}$, and $\hat{Z}_{k\ell} \sim \mathcal{N}(0, 1)$ are i.i.d. noise. The codewords satisfy the average power constraint, $\frac{1}{n_1} \sum_{\ell=1}^{n_1} \hat{X}_{k\ell}^2 \leq \hat{P}_k, k = 1, 2$. The direct links in the second band are similarly modeled as

$$\bar{Y}_{1\ell} = d_1 \bar{X}_{1\ell} + \bar{Z}_{1\ell}, \tag{5}$$

$$\bar{Y}_{2\ell} = d_2 \bar{X}_{2\ell} + \bar{Z}_{2\ell}, \quad \ell = 1, \dots, n_2, \tag{6}$$

where $\bar{X}_{k\ell}, \bar{Y}_{k\ell} \in \mathbb{R}$, $d_k \in \mathbb{R}$ are the direct channel coefficients, $\bar{Z}_{k\ell} \sim \mathcal{N}(0, 1)$ are i.i.d. noise, $k = 1, 2$, and the codewords satisfy the average power constraint, $\frac{1}{n_2} \sum_{\ell=1}^{n_2} \bar{X}_{k\ell}^2 \leq \bar{P}_k, k = 1, 2$, as well. We present the Gaussian DCLIC in Figure 1. We define a $(2^{nR_1}, 2^{nR_2}, n, n_1, n_2)$ code for the Gaussian DCLIC from that of the DM DCLIC, by choosing all input and output variables to be in $\mathbb{R}$, and by imposing the average power constraints on the codewords, X_k^n, $\hat{X}_k^{n_1}$ and $\bar{X}_k^{n_2}, k = 1, 2$, defined above.
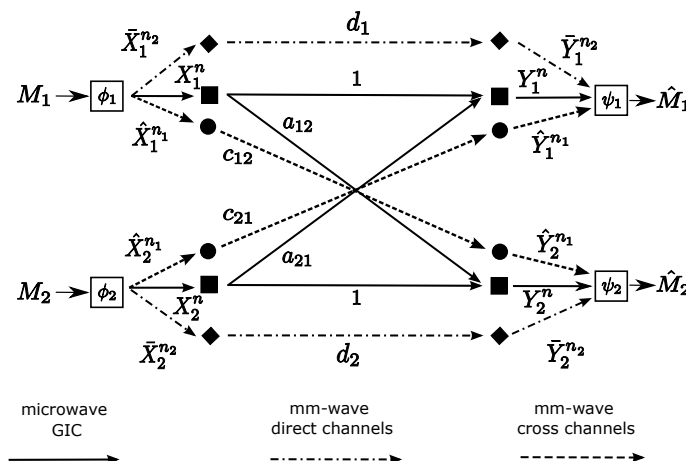

Figure 1. System model of the Gaussian DCLIC, which consists of an underlying GIC in the microwave band and the set of direct channels and cross channels in the mm-wave band.

Next, the Gaussian CLIC is defined from the Gaussian DCLIC by imposing the restrictions, $n_2 = 0$ and $\bar{\mathcal{X}}_k = \bar{\mathcal{Y}}_k = \varnothing, k = 1, 2$, such that the channel outputs in the CLIC are described by (1)–(4), respectively. A $(2^{nR_1}, 2^{nR_2}, n, n_1)$ code for CLIC is defined from a $(2^{nR_1}, 2^{nR_2}, n, n_1, n_2)$ code of the Gaussian DCLIC with the above-mentioned restrictions.

For ease of exposition, we define two classes of the CLICs: we say that a CLIC is a *strong* CLIC or a *weak* CLIC, if the underlying GIC in the first band has strong interference (i.e., $a_{12}^2 \geq 1$, $a_{21}^2 \geq 1$) or weak interference (i.e., $a_{12}^2 < 1$, $a_{21}^2 < 1$), respectively. Moreover, a *symmetric* CLIC is a CLIC where $a^2 := a_{12}^2 = a_{21}^2, c^2 := c_{12}^2 = c_{21}^2, Q_1 = Q_2$, and $\hat{P}_1 = \hat{P}_2$. In addition, a *symmetric* DCLIC is a DCLIC with an underlying symmetric CLIC, and $d^2 = d_1^2 = d_2^2, \bar{P}_1 = \bar{P}_2$.

Finally, the Gaussian model for the DLIC are defined from that of the DCLIC by imposing the restrictions, $n_1 = 0$ and $\hat{\mathcal{X}}_k = \hat{\mathcal{Y}}_k = \varnothing, k = 1, 2$, and thus the channel outputs are described by (1), (2), (5) and (6), respectively. In the sequel, we focus on the Gaussian model of the DCLIC and the CLIC.

3. Decomposition Result on the Capacity of the DCLIC

Recall that, in the DCLIC, there are n channel uses in the first band, n_1 cross channel uses in the second band, and n_2 direct channels uses in the second band. We show below that the capacity of the DCLIC can be decomposed into the capacity of the underlying CLIC, complemented by the direct channels that are used to transmit individual user information to their respective receivers.

Theorem 1. *The capacity region of the DCLIC with BMFs α_1 and α_2 is given by the set of all nonnegative rate tuples (R_1, R_2) that satisfy the decomposition*

$$R_1 = r_1 + \alpha_2 \mathsf{C}(d_1^2 \bar{P}_1), \quad R_2 = r_2 + \alpha_2 \mathsf{C}(d_2^2 \bar{P}_2), \tag{7}$$

where (r_1, r_2) is an achievable rate tuple in the underlying CLIC with BMF α_1.

Proof. The proof is relegated to Appendix A. □

Therefore, an achievable rate pair in the DCLIC with BMFs α_1 and α_2 consists of a rate pair achievable in the CLIC with BMF α_1, and the rate pair achieved in the direct channels, $(\alpha_2 \mathsf{C}(d_1^2 \bar{P}_1), \alpha_2 \mathsf{C}(d_2^2 \bar{P}_2))$. The capacity of the DCLIC can thus be characterized from that of the underlying CLIC. Hence, we focus on the CLIC next. In particular, we consider two specific classes of the CLIC, the strong CLIC and the weak CLIC. First, we present the capacity of the strong CLIC.

Lemma 1. *The capacity region of the strong CLIC with BMF α_1 is given by the set of all nonnegative rate tuples (R_1, R_2) that satisfy*

$$R_1 \leq \mathsf{C}(Q_1), \tag{8}$$
$$R_2 \leq \mathsf{C}(Q_2), \tag{9}$$
$$R_1 + R_2 \leq \mathsf{C}(Q_1 + a_{21}^2 Q_2) + \alpha_1 \mathsf{C}(c_{21}^2 \hat{P}_2), \tag{10}$$
$$R_1 + R_2 \leq \mathsf{C}(Q_2 + a_{12}^2 Q_1) + \alpha_1 \mathsf{C}(c_{12}^2 \hat{P}_1). \tag{11}$$

The proof of Lemma 1 follows from that of [27] in a straightforward manner. Hence, we omit the proof here, and discuss only the key idea. In the strong CLIC, the GIC in the first band has strong interference. Additionally, in the second band, the cross-channel gains are positive and the direct-channel gains are zero, which results in a GIC with strong interference. Hence, the strong CLIC is a parallel GIC with strong interference in both bands. The capacity is thus achieved by encoding jointly over both bands and decoding messages from both transmitters as in [27].

Moreover, the capacity of the DCLIC with strong underlying CLIC, which follows from Theorem 1 and Lemma 1, is characterized below.

Corollary 1. *The capacity region of the DCLIC with BMFs α_1 and α_2 that has a strong underlying CLIC is given by the set of all nonnegative rate tuples (R_1, R_2) that satisfy*

$$R_1 \le \mathsf{C}(Q_1) + \alpha_2\mathsf{C}(d_1^2\bar{P}_1), \tag{12}$$

$$R_2 \le \mathsf{C}(Q_2) + \alpha_2\mathsf{C}(d_2^2\bar{P}_2), \tag{13}$$

$$R_1 + R_2 \le \mathsf{C}(Q_1 + a_{21}^2 Q_2) + \alpha_1\mathsf{C}(c_{21}^2\hat{P}_2) + \alpha_2\mathsf{C}(d_1^2\bar{P}_1) + \alpha_2\mathsf{C}(d_2^2\bar{P}_2), \tag{14}$$

$$R_1 + R_2 \le \mathsf{C}(Q_2 + a_{12}^2 Q_1) + \alpha_1\mathsf{C}(c_{12}^2\hat{P}_1) + \alpha_2\mathsf{C}(d_1^2\bar{P}_1) + \alpha_2\mathsf{C}(d_2^2\bar{P}_2). \tag{15}$$

Furthermore, if the DLIC, where the second band is used as direct channels only, has strong underlying GIC, the capacity region follows from Corollary 1 with $\alpha_1 = 0$ and $\alpha_2 = \alpha$.

4. Capacity of the Weak CLIC

Now, consider the weak CLIC where the underlying GIC has weak interference. Recall that in a conventional GIC with weak interference, decoding both messages is suboptimal in general. However, in the weak CLIC, if messages are encoded jointly over both bands and the cross channels in the second band are sufficiently strong, then decoding both messages is indeed optimal. We characterize the sufficient conditions below.

Lemma 2. *Provided the channel parameters of the CLIC with BMF α_1 satisfy $a_{12}^2 < 1$, $a_{21}^2 < 1$, as well as the conditions*

$$(1 + Q_1)(1 + Q_2) \le (1 + a_{12}^2 Q_1)(1 + c_{12}^2\hat{P}_1)^{\alpha_1}, \tag{16}$$

$$(1 + Q_1)(1 + Q_2) \le (1 + a_{21}^2 Q_2)(1 + c_{21}^2\hat{P}_2)^{\alpha_1}, \tag{17}$$

the following inequalities hold

$$I(X_1^n; Y_1^n | X_2^n) \le I(X_1^n; Y_2^n) + n\alpha_1\mathsf{C}(c_{12}^2\hat{P}_1), \tag{18}$$

$$I(X_2^n; Y_2^n | X_1^n) \le I(X_2^n; Y_1^n) + n\alpha_1\mathsf{C}(c_{21}^2\hat{P}_2), \tag{19}$$

for all X_1^n and X_2^n with product distributions $p(x_1^n)p(x_2^n)$ on $\mathbb{R}^n \times \mathbb{R}^n$, $n \in \mathbb{Z}_+$, which satisfy the average power constraints, $\frac{1}{n}\sum_{\ell=1}^n X_{k\ell}^2 \le Q_k, k = 1, 2$.

Proof. We prove only (18) as (19) follows similarly. From the system model for the first band in (1) and (2), we have

$$I(X_1^n; Y_1^n | X_2^n) - I(X_1^n; Y_2^n) = h(X_1^n + Z_1^n) - h(Z_1^n) - h(X_2^n + a_{12}X_1^n + Z_2^n) + h(X_2^n + Z_2^n)$$

$$\overset{(a)}{\le} h(X_2^n + Z_2^n) - h(Z_2^n) + h(X_1^n + Z_1^n) - h(a_{12}X_1^n + Z_2^n)$$

$$\overset{(b)}{=} I(X_2^n; X_2^n + Z_2^n) - \{h(X_1^n + Z_1^n + Z_3^n) - h(X_1^n + Z_1^n)\} - n\log|a_{12}|$$

$$\overset{(c)}{\le} n\mathsf{C}(Q_2) - I(Z_3^n; Z_3^n + Z_1^n + X_1^n) - n\log|a_{12}|$$

$$\overset{(d)}{=} n\mathsf{C}(Q_2) - \sum_{\ell=1}^n \left(h(Z_{3\ell}|Z_3^{n-\ell}) - h(Z_{3\ell}|Z_3^{n-\ell}, Z_3^n + Z_1^n + X_1^n)\right) - n\log|a_{12}|$$

$$\overset{(e)}{\le} n\mathsf{C}(Q_2) - \sum_{\ell=1}^n \left(h(Z_{3\ell}) - h(Z_{3\ell}|Z_{3\ell} + Z_{1\ell} + X_{1\ell})\right) - n\log|a_{12}|$$

$$\overset{(f)}{=} n\mathsf{C}(Q_2) - \sum_{\ell=1}^n I(Z_3; Z_3 + Z_1 + X_{1\ell}) - n\log|a_{12}|$$

$$\overset{(g)}{\le} n\mathsf{C}(Q_2) - \sum_{\ell=1}^n I(Z_3; Z_3 + Z_1 + X_{1\ell}^*) - n\log|a_{12}|$$

$$\overset{(h)}{=} n\mathsf{C}(Q_2) + \sum_{\ell=1}^{n} \frac{1}{2} \log \frac{Q_{1\ell} + 1}{Q_{1\ell} + 1/a_{12}^2} - n \log|a_{12}|$$

$$\overset{(i)}{\leq} n\mathsf{C}(Q_2) + \frac{n}{2} \log \frac{\frac{1}{n}\sum_{\ell=1}^{n} Q_{1\ell} + 1}{\frac{1}{n}\sum_{\ell=1}^{n} Q_{1\ell} + 1/a_{12}^2} - n \log|a_{12}|$$

$$\overset{(j)}{\leq} \frac{n}{2} \log \frac{(1 + Q_1)(1 + Q_2)}{(1 + a_{12}^2 Q_1)}$$

$$\overset{(k)}{\leq} \frac{\alpha_1 n}{2} \log(1 + c_{12}^2 \hat{P}_1),$$

where (*a*) follows from unconditioning, $h(X_2^n + a_{12}X_1^n + Z_2^n) \geq h(X_2^n + a_{12}X_1^n + Z_2^n|X_2^n) = h(a_{12}X_1^n + Z_2^n)$, and since $h(Z_1^n) = h(Z_2^n)$; (*b*) follows since $1/a_{12}^2 > 1$, and thus $h(a_{12}X_1^n + Z_2^n) = h(a_{12}X_1^n + Z_1^n) = h(X_1^n + Z_1^n/a_{12}) + n \log|a_{12}| = h(X_1^n + Z_3^n + Z_1^n) + n \log|a_{12}|$, where $Z_{3\ell} \sim \mathcal{N}(0, 1/a_{12}^2 - 1)$, are i.i.d.; (*c*) follows since the first term in *b* is maximized by $X_{2\ell} \sim \mathcal{N}(0, Q_2)$, i.i.d.; (*d*) follows from the chain rule; (*e*) follows since $Z_{3\ell}$ are i.i.d., and since unconditioning does not reduce entropy; (*f*) follows by replacing the i.i.d. RVs, $Z_{3\ell}$ and $Z_{1\ell}$, with $Z_3 \sim \mathcal{N}(0, 1/a_{12}^2 - 1)$ and $Z_1 \sim \mathcal{N}(0, 1)$, respectively; (*g*) follows by invoking the worst additive noise result of [33] to every mutual information term inside the summation of (*f*), where $X_{1\ell}^* \sim \mathcal{N}(0, Q_{1\ell})$, with $Q_{1\ell} := \mathbb{E}(X_{1\ell}^2)$, and $\frac{1}{n}\sum_{\ell=1}^{n} Q_{1\ell} \leq Q_1$; (*h*) follows from the differential entropy of Gaussian RVs; (*i*) follows since the log function is concave in $\{Q_{1\ell}\}_{\ell=1}^{n}$; (*j*) follows since $\frac{1}{n}\sum_{\ell=1}^{n} Q_{1\ell} \leq Q_1$; and (*k*) follows from condition (16). □

We characterize the capacity of the weak CLIC under the conditions of Lemma 2 below.

Theorem 2. *The capacity region of the weak CLIC with BMF α_1 that satisfies the conditions in* (16) *and* (17) *is given by the set of all nonnegative rate tuples (R_1, R_2) that satisfy*

$$R_1 \leq \mathsf{C}(Q_1), \tag{20}$$

$$R_2 \leq \mathsf{C}(Q_2), \tag{21}$$

$$R_1 + R_2 \leq \mathsf{C}(Q_1 + a_{21}^2 Q_2) + \alpha \mathsf{C}(c_{21}^2 \hat{P}_2), \tag{22}$$

$$R_1 + R_2 \leq \mathsf{C}(Q_2 + a_{12}^2 Q_1) + \alpha \mathsf{C}(c_{12}^2 \hat{P}_1). \tag{23}$$

Proof. The proof is relegated to Appendix B. □

The result shows that, in the weak CLIC, where sufficient interference forwarding is not possible through the underlying weak GIC, if the cross links in the second band are sufficiently strong, it is possible to forward enough interference by encoding jointly over both bands such that decoding both messages becomes optimal. Hence, these conditions can be classified as being able to push the receivers to the "strong interference regime over both bands".

Next, we illustrate the relationship between the channel parameters in (16) and (17) with an example of a symmetric weak CLIC, where (16) and (17) imply the same condition. We denote by $c_{\min}^2 := (((1 + Q_1)(1 + Q_2)/(1 + a^2 Q_1))^{1/\alpha_1} - 1)/\hat{P}_1$, and $\alpha_{1,\min} := \log((1 + Q_1)(1 + Q_2)/(1 + a^2 Q_1))/\log(1 + c^2 \hat{P}_1)$, respectively the minimum c^2 and the minimum α_1 required to achieve (16), and show the interplay between $a^2, c_{\min}^2$ and $\alpha_{1,\min}$ for $Q_1 = Q_2 = \hat{P}_1 = 1$.

In Figure 2a, we plot $c_{\min}^2$ against $a^2 \in (0, 1)$ for $\alpha_1 \in \{0.5, 1, 2\}$. Note that $c_{\min}^2$ reduces monotonically as a^2 or α_1 increases. This follows since, if a^2 increases, then the first band forwards more interference, and, if α_1 increases, then the pre-log factor of the cross-channel capacity increases. In either case, smaller $c_{\min}^2$ is required to achieve (16). Similarly in Figure 2b, we plot $\alpha_{1,\min}$ against $a^2 \in (0, 1)$ for $c^2 \in \{0.5, 1, 2\}$, and note that $\alpha_{1,\min}$ reduces as a^2 or c^2 increases. Finally, in Figure 2c, we depict the set of cross-channel gains (a^2 and c^2) of a symmetric CLIC with $Q_1 = Q_2 = \hat{P}_1 = 1, \alpha_1 = 2$, and partition it depending on whether the capacity has been characterized in each set. Note that the capacity for the set where $a^2 < 1$ and $c^2 < c_m^2$ has not been characterized yet.
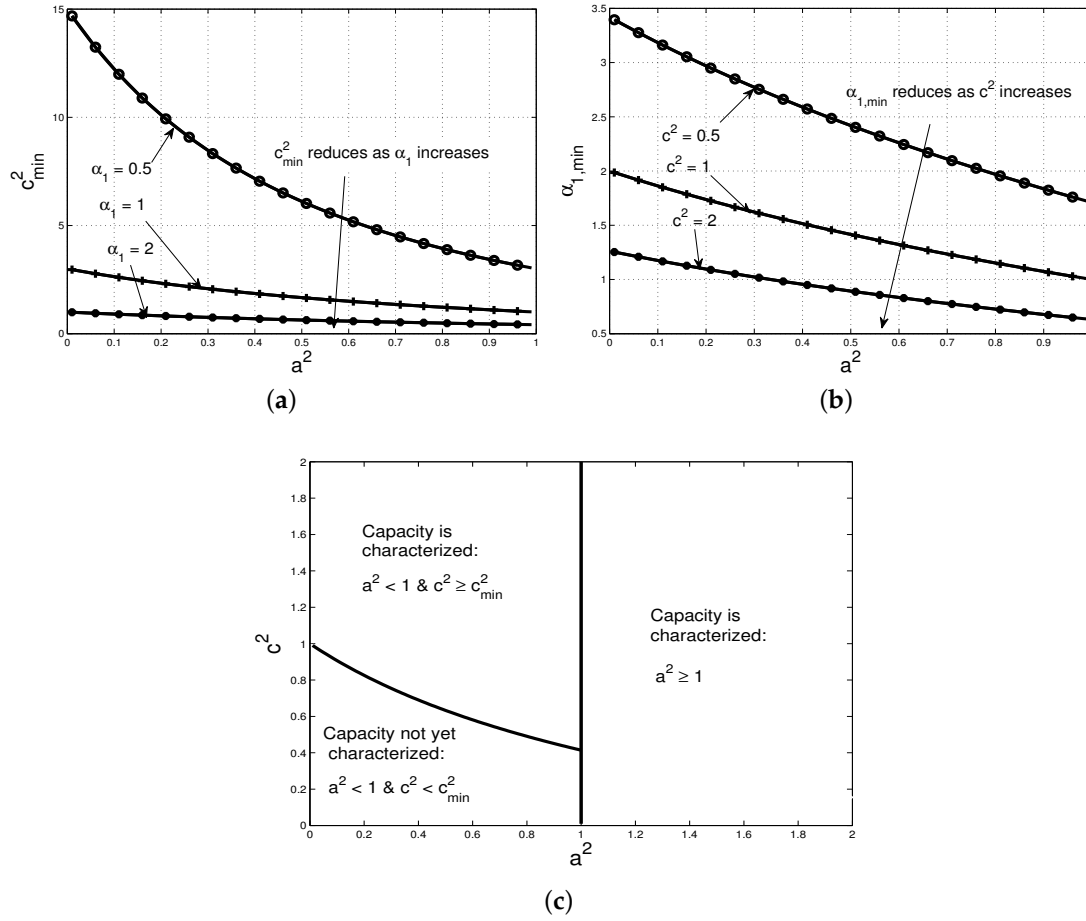
(a)

(b)

(c)

Figure 2. In (**a**,**b**), we plot c_{min}^2 and $\alpha_{1,min}$, respectively. In (**c**), the channel gains of a symmetric CLIC is partitioned based on whether its capacity has been characterized in each set.

5. The Optimal Sum-Rate Problem

In this section, we study the power allocation scheme over the direct and cross channels in the DCLIC that maximizes its sum-rate. Recall from Corollary 1 that, if the underlying GIC of the DCLIC has strong interference (i.e., $a_{12}^2 \geq 1$ and $a_{21}^2 \geq 1$), then the sum-rate of the DCLIC is known for all values of the remaining channel parameters, and, in particular, for all transmit powers in the second band. Therefore, we pose this problem for the class of DCLICs with strong underlying GIC (see [A1] below). Also recall that the normalized bandwidth (α) of the second band is shared between the direct channels (α_2) and the cross channels (α_1). We denote the fraction of α alloted to the direct and cross channels by β and $\bar{\beta}$, respectively, where $\beta := \alpha_2/\alpha$ and $\bar{\beta} := 1 - \beta = \alpha_1/\alpha$, and $\beta, \bar{\beta} \in (0,1)$. Thus, β provides a trade-off between the bandwidths in the direct channels ($\beta\alpha$) and the cross channels ($\bar{\beta}\alpha$).

We formulate the problem under the following assumptions:

1. the underlying GIC of the DCLIC has strong interference, but not very strong interference, $1 \leq a_{12}^2 < 1 + Q_1$ and $1 \leq a_{21}^2 < 1 + Q_2$;
2. the underlying GIC of the DCLIC satisfies: $Q_2 + a_{12}^2 Q_1 = Q_1 + a_{21}^2 Q_2$;
3. β and $\bar{\beta}$ are fixed a priori;
4. the transmission power in the direct channel (p_k) and cross channel (q_k) from transmitter k (Tx_k) satisfy the constraint, $\beta p_k + \bar{\beta} q_k = P, k = 1, 2$, where P is the power budget.

In [A1], we assume that the underlying GIC does not have very strong interference, since the power allocation in that case is trivial (see Section 5.4). For ease of exposition, we assume in [A2] that the underlying GIC receives equal power in both its receivers. Note that the class of GICs that satisfies

[A2] also contains the symmetric GICs. Moreover, the analysis under [A2] reveals enough insight such that it can be extended to the general case (see Remark 1).

In [A3], β is assumed to be fixed a priori and known. This models practical constraints in many wireless networks where dynamically allocating the bandwidth may not be feasible or straightforward [34]. In [A4], we assume that power budget (P) in both transmitters are the same. There is no loss of generality in this assumption as the relative difference between the power budgets of the first and the second user can be absorbed into the channel gains.

5.1. Problem Formulation and Solution

For a fixed power allocation (p_1, q_1, p_2, q_2) in the mm-wave channels, we denote the sum-rate achievable at Rx_1 and Rx_2 in (14) and (15) by Σ_1 and Σ_2, and the interference-free sum-rate given by the sum of individual rates in (12) and (13) by Σ, and present them below

$$\Sigma_1 := A_1 + \frac{\alpha\bar{\beta}}{2}\log(1 + c_{21}^2 q_2) + \frac{\alpha\beta}{2}\big(\log(1 + d_1^2 p_1) + \log(1 + d_2^2 p_2)\big), \tag{24}$$

$$\Sigma_2 := A_2 + \frac{\alpha\bar{\beta}}{2}\log(1 + c_{12}^2 q_1) + \frac{\alpha\beta}{2}\big(\log(1 + d_1^2 p_1) + \log(1 + d_2^2 p_2)\big), \tag{25}$$

$$\Sigma := A + \frac{\alpha\beta}{2}\big(\log(1 + d_1^2 p_1) + \log(1 + d_2^2 p_2)\big), \tag{26}$$

where $A_1 := \frac{1}{2}\log(1 + Q_1 + a_{21}^2 Q_2)$, $A_2 := \frac{1}{2}\log(1 + Q_2 + a_{12}^2 Q_1)$ and $A := \frac{1}{2}\log(1 + Q_1)(1 + Q_2)$. Furthermore, they satisfy $A_1 = A_2 < A$ due to [A2] and [A1]. Therefore, a necessary and sufficient condition for R to be an achievable sum-rate of the DCLIC is $R \le \min\{\Sigma_1, \Sigma_2, \Sigma\}$, for some power allocation (p_1, q_1, p_2, q_2). The optimization problem that maximizes R over the transmit powers (p_1, q_1, p_2, q_2) is then

$$
\begin{aligned}
[\mathbf{P1}] \quad \text{maximize} \quad & R \\
\text{subject to:} \quad & R \le \Sigma_1, && (27) \\
& R \le \Sigma_2, && (28) \\
& R \le \Sigma, && (29) \\
& \beta p_1 + \bar{\beta} q_1 = P, && (30) \\
& \beta p_2 + \bar{\beta} q_2 = P, && (31) \\
& (p_1, q_1, p_2, q_2, R) \succeq \mathbf{0}. && (32)
\end{aligned}
$$

Note that [**P1**] is a convex optimization problem, since its objective function R is linear, the equality constraints (30) and (31) are affine, and the inequality constraints (27)–(29) are convex. Furthermore, [**P1**] satisfies Slater's condition [35] (Chapter 5.2.3), and thus it can be solved using the Karush–Kuhn–Tucker (KKT) conditions [35] (Chapter 5.5.3). We relegate the details to Appendix C.

It is well known that optimal power allocation in parallel Gaussian point-to-point channels follows the Waterfilling (WF) property. Due to this, if the power budget is sufficient small, it is allocated entirely to the "strongest" sub-channel, and, as the power budget is increased, power is allocated to the other "weaker" sub-channels, in addition to the strongest one (see Chapter 3.4.3 in [32]). In the ensuing discussion, it becomes clear that the optimal power allocation in [**P1**] (hereby referred to as "the optimal allocation") has two noticeable properties: a *WF-like* property, due to which it assigns power to the cross and direct channels following a WF-like allocation, and a *max-min* property, due to which it increases the minimum of the sum-rate constraints.

Due to its WF-like property, the optimal allocation assigns the entire power budget (P) to only a subset of all the direct and cross channels, depending on channel conditions that indicate whether the direct channels are stronger than the cross channels or vice versa. Moreover, if P is sufficiently increased, it becomes optimal to allocate power to the remaining set of channels. In addition, since the

objective of the problem is to maximize $\min\{\Sigma_1, \Sigma_2, \Sigma\}$, the optimal allocation assigns powers in such a way that minimizes or eliminates any difference between Σ_1, Σ_2 and Σ. We observe that, due to this property (max-min property), the optimal allocation imposes a maximum limit on the cross-channel powers, which is unlike WF in [32] where no such limit exists.

We study the optimal allocation by partitioning the entire set of channel parameters and P into disjoint sets $(\mathcal{S}_{(.)})$, such that the optimal allocation can be classified according to the power levels in the direct and cross channels in each set. Without loss of generality, we present the optimal allocation under $c_{21}^2 > c_{12}^2$ in Table 1, and study it in detail. In this case, only fours sets, $\mathcal{S}_D$, $\mathcal{S}_C$, $\mathcal{S}_{CD}$ and $\mathcal{S}_{sat}$, are sufficient. The power allocation under $c_{21}^2 < c_{12}^2$ can be readily obtained from Table 1 by swapping the indices 1 and 2, and the case with $c_{21}^2 = c_{12}^2$ is briefly discussed in Section 5.3.

For notational convenience, we express the optimal powers and the conditions of the sets $\mathcal{S}_{(.)}$ in Table 1 in terms of the following functions:

$$g_1(P) := \frac{d_1^2}{c_{12}^2(1 + Pd_1^2/\beta)} + \frac{d_2^2}{c_{21}^2(1 + Pd_2^2/\beta)}, \tag{33}$$

$$g_2(P) := \left(1 + c_{12}^2 P/\bar{\beta}\right)\left(\frac{d_1^2}{c_{12}^2} + \frac{1}{P(c_{21}^2 - c_{12}^2)/\beta + c_{21}^2/d_2^2}\right), \tag{34}$$

$$g_3(P) := \frac{\beta\gamma}{Pc_{21}^2 + \bar{\beta} + \beta c_{21}^2/d_2^2 - \bar{\beta}\gamma} + \frac{\beta\gamma}{Pc_{12}^2 + \bar{\beta} + \beta c_{12}^2/d_1^2 - \bar{\beta}\gamma}, \tag{35}$$

$$F(P) := \left(E_1 + E_2 - \sqrt{(E_1 - E_2)^2 + 4\beta^2 E_1 E_2}\right)/(2\bar{\beta}(1 + \beta)) - 1, \tag{36}$$

$$E_k := Pc_{km}^2 + \bar{\beta} + \beta c_{km}^2/d_k^2, \quad k \neq m \in \{1, 2\}, \tag{37}$$

$$\gamma := \left(\frac{(1 + Q_1)(1 + Q_2)}{1 + Q_1 + a_{21}^2 Q_2}\right)^{1/\alpha\bar{\beta}}, \tag{38}$$

and $P_4^* := \bar{\beta}(\gamma - 1)/c_{12}^2$. We relegate the details of the derivation to Appendix C.

In the following, we use "the optimal allocation" and $\mathbf{OA_1}$ interchangeably to refer to the optimal power allocation for [**P1**] under $c_{21}^2 > c_{12}^2$, and discuss some interesting characteristics of $\mathbf{OA_1}$.

Table 1. Optimal power allocation and conditions of sets in $\mathbf{OA_1}$.

Set	Optimal Powers				Condition
$\mathcal{S}_D$	$p_1 = \dfrac{P}{\beta}$,	$q_1 = 0$,	$p_2 = \dfrac{P}{\beta}$,	$q_2 = 0$,	$g_1(P) > 1$
$\mathcal{S}_C$	$p_1 = 0$,	$q_1 = \dfrac{P}{\bar{\beta}}$,	$p_2 = \dfrac{P(c_{21}^2 - c_{12}^2)}{\beta c_{21}^2}$,	$q_2 = \dfrac{c_{12}^2}{c_{21}^2}\dfrac{P}{\bar{\beta}}$,	$P < P_4^*$, $g_2(P) < 1$
$\mathcal{S}_{sat}$	$p_1 = \dfrac{P}{\beta} - \dfrac{\bar{\beta}(\gamma - 1)}{\beta c_{12}^2}, q_1 = \dfrac{\gamma - 1}{c_{12}^2}$,		$p_2 = \dfrac{P}{\beta} - \dfrac{\bar{\beta}(\gamma - 1)}{\beta c_{21}^2}, q_2 = \dfrac{\gamma - 1}{c_{21}^2}$,		$P \geq P_4^*$, $g_3(P) < 1$
$\mathcal{S}_{CD}$	$p_1 = \dfrac{P - \bar{\beta}q_1}{\beta}$,	$q_1 = \dfrac{F(P)}{c_{12}^2}$,	$p_2 = \dfrac{P - \bar{\beta}q_2}{\beta}$,	$q_2 = \dfrac{c_{12}^2}{c_{21}^2}q_1$,	complement of other conditions

First, the condition $g_1(P) < 1$ in $\mathcal{S}_D$ implies that the direct channels are "stronger" than the cross channels, in the sense that, for sufficiently small P, the sum-rate achieved from allocating P only to both direct channels is larger than that achieved from any other subset of channels. Thus, following its WF-like property, $\mathbf{OA_1}$ allocates P entirely to the direct channels, i.e., $p_1 = p_2 = \frac{P}{\beta}$, and zero power to the cross channels, i.e., $q_1 = q_2 = 0$. This allocation also achieves $\Sigma_2 = \Sigma_1$, which is consistent with the max-min property of $\mathbf{OA_1}$.

Second, the condition $g_2(P) < 1$ in $\mathcal{S}_C$ implies that the cross channels are "stronger" than the direct channels, in the sense that, for sufficiently small P, the sum-rate achieved from allocating P to

both cross channels and a direct channel is larger than that achieved from only direct channels. Note that, unlike in $\mathcal{S}_D$, **OA₁** needs to allocate power to a direct channel in addition to both cross channels in $\mathcal{S}_C$, since allocating P entirely to the cross channels causes an imbalance between Σ_1 and Σ_2 ($\Sigma_2 < \Sigma_1$) due to $c_{21}^2 > c_{12}^2$, which violates the max-min property. Therefore, **OA₁** shares P among the cross and direct channels from Tx_2 to preserve $\Sigma_2 = \Sigma_1$. Moreover, the condition $P < P_4^*$ ensures that power in the cross channels have not yet reached their maximum limits.

Third, as P is increased, **OA₁** shares P among all channels in $\mathcal{S}_{CD}$. As P increases, the additional benefits from transmitting in a particular subset of channels (either direct channels as in $\mathcal{S}_D$, or cross and direct channels as in $\mathcal{S}_C$) begins to diminish. Thus, following its WF-like property, **OA₁** starts sharing P among all channels. In addition, **OA₁** shares P in such a way that preserves $\Sigma_2 = \Sigma_1$.

Finally, if P is increased sufficiently, **OA₁** follows the allocation in $\mathcal{S}_{\text{sat}}$, where the cross-channel powers have reached their maximum limits, $q_{k,\text{sat}} := (\gamma - 1)/c_{km}^2, k \neq m \in \{1,2\}$. Therefore, as P increases further, all subsequent increments of P is allotted to only direct channels. We now say that the cross channels are saturated, in the sense that allocating more power beyond these limits does not improve the sum-rate. Such limits for the cross channels in **OA₁** is unlike the WF allocation in [32].

The cross channels become saturated due to the max-min property of **OA₁**. Recall that, in $\mathcal{S}_{CD}$, **OA₁** allocates powers to all channels, which increase as P increases. The increases in p_1 and p_2 results in the increase of Σ_1, Σ_2 and Σ by an equal margin. However, an increase in q_1 only increases Σ_2, and an increase in q_2 only increases Σ_1. Now, note that, due to its max-min property, **OA₁** preserves $\Sigma_1 = \Sigma_2$ in $\mathcal{S}_D, \mathcal{S}_C$, and $\mathcal{S}_{CD}$. However, there may exist a gap between $\Sigma_1 = \Sigma_2$ and Σ. As q_1 and q_2 are increased, this gap reduces and finally becomes zero in $\mathcal{S}_{\text{sat}}$. At this point, both cross channels become saturated simultaneously. If any more power is allocated to either q_1 or q_2, a suboptimal sum-rate $R = \Sigma < \min\{\Sigma_1, \Sigma_2\}$ will result. Therefore, **OA₁** maintains $q_k = q_{k,\text{sat}}$ and diverts all additional power to the direct channels.

Note that once, in $\mathcal{S}_{\text{sat}}$, the sum-rates achieved by joint decoding (Σ_1 and Σ_2) become equal to the sum of interference-free user rates (Σ), which is somewhat similar to the behavior of the GIC under very strong interference. At this point, all additional increments in P are allocated to the direct channels to increase the individual user rates.

In addition, note that the sets $\mathcal{S}_{(.)}$ form a partition due to their construction using the optimal Lagrange multipliers (defined in Appendix C). Thus, the conditions in Table 1 are mutually exclusive.

Moreover, these conditions can be equivalently described in terms of three *critical* powers, $P_k^* \in \mathbb{R}_+$, defined by $g_k(P_k^*) = 1$, where $g_k(P)$ is defined in (33)–(35) for $k \in \{1,2,3\}$. Specifically, the conditions of $\mathcal{S}_D, \mathcal{S}_C$, and $\mathcal{S}_{\text{sat}}$ are given by $P < P_1^*$, $P < \min\{P_2^*, P_4^*\}$, and $P > \max\{P_3^*, P_4^*\}$, respectively. In addition, we note that the direct channels are "stronger" than the cross channels in the sense specified above, if $P_1^* > 0$. Similarly, if $P_2^* > 0$, the cross channels are "stronger" than the direct channels. Furthermore, if $0 < P_3^* < P_4^* < P_2^*$, the cross channels are said to be "much stronger", in the sense that **OA₁** continues allocating power to the cross channels as in $\mathcal{S}_C$ until they become saturated, and only after that it assigns power to both direct channels.

We note from the mutual exclusiveness of $\mathcal{S}_D$ and $\mathcal{S}_C$ that, if $P_1^* > 0$, then no $P_2^* \in \mathbb{R}_+$ exists that satisfies $g_2(P_2^*) = 1$. This shows that, if the direct channels are stronger, the allocation in $\mathcal{S}_C$ is suboptimal for any $P > 0$. Similarly, if $P_2^* > 0$, then no $P_1^* \in \mathbb{R}_+$ exists such that $g_1(P_1^*) = 1$, and thus the allocation in $\mathcal{S}_D$ is suboptimal for any $P > 0$.

5.2. The Waterfilling-Like Nature of the Optimal Power Allocation

Now, we characterize how **OA₁** adapts the power allocation, as the power budget (P) increases and crosses the critical thresholds $P_k^*, k \in \{1,2,3\}$. We say that **OA₁** follows the sequence $\mathcal{A}_1 \to \mathcal{A}_2 \to \mathcal{A}_3$ where $\mathcal{A}_l \in \{\mathcal{S}_D, \mathcal{S}_C, \mathcal{S}_{CD}, \mathcal{S}_{\text{sat}}\}$, if **OA₁** allocates power as in $\mathcal{A}_1$ for sufficiently small P, and then adapts the powers according to the allocation in $\mathcal{A}_2$ and $\mathcal{A}_3$ as P increases. In this regard, we note that **OA₁** follows one of the three sequences, as explained below and illustrated graphically in Figure 3.
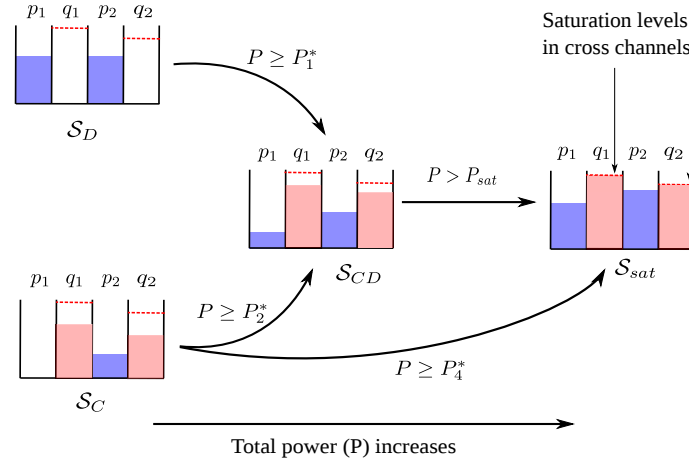
Figure 3. Due to its WF-like property, **OA$_1$** follows one of the three sequences depending on P_k^*. The saturation levels in the cross channels are due to its max-min property.

- If $P_1^* > 0$: **OA$_1$** follows the sequence $\mathcal{S}_D \to \mathcal{S}_{CD} \to \mathcal{S}_{\text{sat}}$ (denoted by [S1]). Since the direct channels are stronger, **OA$_1$** allocates all of P to them as in $\mathcal{S}_D$ when P is sufficiently small (i.e., $P < P_1^*$). However, as P increases, the additional benefit from transmitting only in the direct channels decreases, and thus, when $P \geq P_1^*$, **OA$_1$** begins transmitting in both cross and direct channels as in $\mathcal{S}_{CD}$. This allocation follows from its WF-like property, and remains optimal for all $P_1^* \leq P \leq P_3^*$. Finally, when $P > P_3^*$, the max-min property of **OA$_1$** comes into effect, and thus the cross channels become saturated and **OA$_1$** starts following the allocation in $\mathcal{S}_{\text{sat}}$. Note that, in this case, the saturation threshold for P is $P_{\text{sat}} = P_3^*$.

- If $P_2^* > 0$ but $0 < P_3^* < P_4^* < P_2^*$ is not satisfied: **OA$_1$** follows the sequence $\mathcal{S}_C \to \mathcal{S}_{CD} \to \mathcal{S}_{\text{sat}}$ ([S2]). This case is similar to [S1] above, except for the fact that now the cross channels are stronger. Hence, **OA$_1$** transmits in the cross channels and the direct channel with gain d_2^2 as in $\mathcal{S}_C$ when P is sufficiently small (i.e., $P < \min\{P_2^*, P_4^*\}$). Next, following its WF-like property, **OA$_1$** starts transmitting in all the direct and cross channels as in $\mathcal{S}_{CD}$ when $P \geq \min\{P_2^*, P_4^*\}$. Finally, when $P > \max\{P_3^*, P_4^*\}$, the cross channels become saturated, and **OA$_1$** follows the allocation in $\mathcal{S}_{\text{sat}}$ thereon. The saturation threshold for P in this case is $P_{\text{sat}} = \max\{P_3^*, P_4^*\}$.

  Note that, whenever $P_2^* > 0$, **OA$_1$** follows [S2] irrespective of how P_2^*, P_3^*, and P_4^* compare, except in two cases: (a) $0 < P_3^* < P_4^* < P_2^*$, where **OA$_1$** follows [S3], described next, and (b) $0 < P_4^* < P_3^* < P_2^*$, which is infeasible as they violate the mutual exclusiveness of $\mathcal{S}_C$ and $\mathcal{S}_{\text{sat}}$.

- If $0 < P_3^* < P_4^* < P_2^*$: **OA$_1$** follows the sequence $\mathcal{S}_C \to \mathcal{S}_{\text{sat}}$ ([S3]). In this case, the cross channels are much stronger than the direct channels. Hence, similar to [S2], **OA$_1$** allocates power to both cross channels and a direct channel as in $\mathcal{S}_C$ when P is sufficiently small (i.e., $0 \leq P < P_4^*$). As P increases and $P \geq P_4^*$, the cross channels become saturated, and **OA$_1$** begins assigning powers to all channels as in $\mathcal{S}_{\text{sat}}$. Interestingly, in this case, **OA$_1$** skips $\mathcal{S}_{CD}$. This shows that, since the cross channels are much stronger, it is optimal to allocate power as in $\mathcal{S}_C$ until they become saturated at $P_{\text{sat}} = P_4^*$, beyond which the allocation in $\mathcal{S}_{\text{sat}}$ becomes optimal.

We illustrate the characteristics of **OA$_1$** with two numerical examples where we choose the following parameters, $a_{12}^2 = a_{21}^2 = 1.5, Q_1 = Q_2 = 5, \alpha = 2, \beta = 0.5$. First, we illustrate an example of [S1] by additionally choosing $d_1^2 = 2, d_2^2 = 3, c_{12}^2 = 1, c_{21}^2 = 1.5$, such that the direct channels are stronger than the cross channels in the sense that $P_1^* = 0.619$ and $P_2^* = -0.75 \notin \mathbb{R}_+$. In Figure 4a, we plot the optimal powers against the power budget P, and note the following: (i) when $0 \leq P < P_1^* = 0.619$, **OA$_1$** allocates P entirely to the direct channels as in $\mathcal{S}_D$, and thus p_1 and p_2 increase with P, and $q_1 = q_2 = 0$; (ii) when $P_1^* \leq P \leq P_3^* = 2.73$, **OA$_1$** allocates power to all channels

as in $\mathcal{S}_{CD}$, and thus q_1 and q_2 also increase with P; (iii) finally, when $P > P_3^*$, **OA$_1$** follows $\mathcal{S}_{sat}$ where the cross channels become saturated simultaneously, and all increments of P are added to p_1 and p_2.

We depict the resulting constraints Σ_1, Σ_2, and Σ in Figure 4b. First, note that **OA$_1$** preserves $R = \Sigma_1 = \Sigma_2$ for all P. However, there exists a gap between $\Sigma_1 = \Sigma_2$ and Σ in $\mathcal{S}_D$ and $\mathcal{S}_{CD}$. Specifically, in $\mathcal{S}_D$, the gap remains constant ($A - A_1$); in $\mathcal{S}_{CD}$, it reduces gradually as **OA$_1$** transmits in the cross channels, and, in $\mathcal{S}_{sat}$, it becomes zero as **OA$_1$** achieves $R = \Sigma_1 = \Sigma_2 = \Sigma$, as expected.
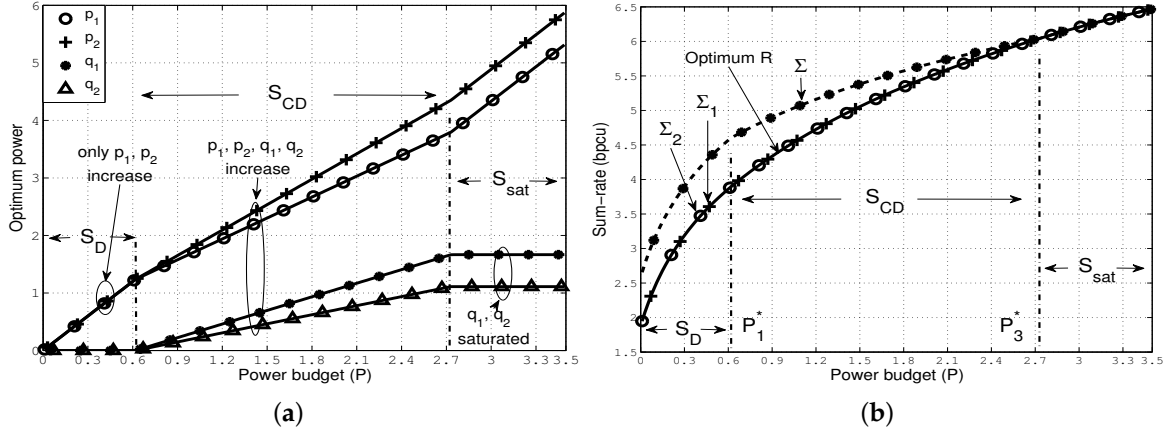

Figure 4. Optimal power allocation that follows the sequence [S1] $\mathcal{S}_D \rightarrow \mathcal{S}_{CD} \rightarrow \mathcal{S}_{sat}$. (**a**) optimal power allocation (p_1, q_1, p_2, q_2); (**b**) the resulting sum-rate constraints.

Next, we illustrate an example of [S2] with the channel gains, $d_1^2 = 0.5, d_2^2 = 1, c_{12}^2 = 1.5, c_{21}^2 = 3$, such that the cross channels are stronger than the direct channels in the sense that $P_2^* = 0.22$ and $P_1^* = -0.21 \notin \mathbb{R}_+$. In Figure 5a, we plot the optimal powers against P, and observe the following: (i) when $P < P_2^* = 0.22$, **OA$_1$** follows the allocation in $\mathcal{S}_C$, and thus p_2, q_1 and q_2 increase with P, whereas $p_1 = 0$; (ii) when $P_2^* \leq P \leq P_3^* = 0.97$, **OA$_1$** allocates power to all channels as in $\mathcal{S}_{CD}$, and thus p_1 now increases with P; (iii) finally, when $P > P_3^*$, **OA$_1$** follows $\mathcal{S}_{sat}$, and thus the cross channels become saturated simultaneously, as expected in [S2]. In Figure 5b, we plot the sum-rate constraints, and note that **OA$_1$** achieves $R = \Sigma_1 = \Sigma_2$ in all the sets. In addition, the gap between $\Sigma_1 = \Sigma_2$ and Σ is gradually offset as **OA$_1$** transmits in the cross channels in $\mathcal{S}_C$ and $\mathcal{S}_{CD}$, and it finally becomes zero in $\mathcal{S}_{sat}$, as expected. We omit an example of [S3] ($\mathcal{S}_C \rightarrow \mathcal{S}_{sat}$), which is somewhat similar to [S2].

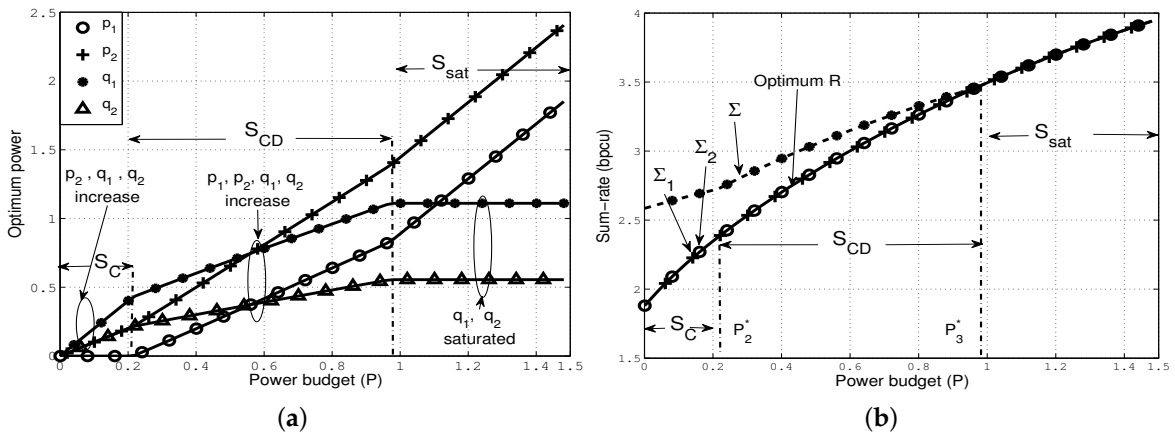
Figure 5. Optimal power allocation that follows the sequence [S2] $\mathcal{S}_C \rightarrow \mathcal{S}_{CD} \rightarrow \mathcal{S}_{sat}$. (**a**) optimal power allocation (p_1, q_1, p_2, q_2); (**b**) the resulting sum-rate constraints.

Remark 1. *Recall that* $\mathbf{OA_1}$ *presents the solution for* [**P1**] *under* $c_{12}^2 < c_{21}^2$ *and* [A2] *(implying* $A_1 = A_2$*). If* $c_{12}^2 < c_{21}^2$ *remains unchanged, and* [**P1**] *is formulated under a more general assumption* [A2G]*; where* $A_2 < A_1$*, then seven disjoint sets are needed to describe the optimal allocation (now denoted by* $\mathbf{OA_G}$*). The conditions and power allocations of these sets are found by solving the corresponding KKT conditions in a similar way to that of* [**P1**] *under* [A2] *in Appendix C. Hence, their explicit derivation is omitted for brevity. Of the seven sets, the first four sets,* $\hat{S}_l, l \in \{D, C, CD, \text{sat}\}$*, are counterparts of the corresponding sets* S_l *in Table 1, in the sense that the power allocation in* $\hat{S}_l$ *is similar to that in* S_l *as discussed below. The remaining three sets* $\hat{S}_k, k = 1, 2, 3$ *are "new" in the sense that the power allocation in these sets do not resemble that in any of the sets* $S_{(.)}$ *in Table 1.*

The power allocation in the first four sets $\hat{S}_l$ *are similar to that of* $S_l, l \in \{D, C, CD, \text{sat}\}$ *in that (a) when the power budget (P) is sufficiently small,* $\mathbf{OA_G}$ *allocates power to only both direct channels (in* $\hat{S}_D$*), or both cross channels and one direct channel (in* $\hat{S}_C$*) depending on whether the direct or the cross channels are "strong" as in* S_D *or* S_C *before; (b) as P is increased* $\mathbf{OA_G}$ *allocates power to all channels (in* $\hat{S}_{CD}$*) as in* S_{CD}*; and (c) when P is sufficiently large, both cross channels become saturated simultaneously, and* $\mathbf{OA_G}$ *allocates all additional increments of P to only the direct channels (in* $\hat{S}_{\text{sat}}$*) as in* S_{sat} *before. Note that the optimal powers and the critical powers of* $\hat{S}_{(.)}$ *differ slightly from that of* $S_{(.)}$ *due to the general assumption* [A2G]*. For conciseness, we omit the conditions, and present the optimal powers in Table 2.*

Table 2. Optimal power allocation in the first four sets in $\mathbf{OA_G}$.

Set		Optimal Powers					
$\hat{S}_D$	$p_1 = \dfrac{P}{\beta},$		$q_1 = 0,$		$p_2 = \dfrac{P}{\beta},$		$q_2 = 0,$
$\hat{S}_C$	$p_1 = 0,$		$q_1 = \dfrac{P}{\bar{\beta}},$		$p_2 = \dfrac{Pc_{21}^2 + \bar{\beta} - (Pc_{12}^2 + \bar{\beta})/\eta}{\beta c_{21}^2},$		$q_2 = \dfrac{Pc_{12}^2/\bar{\beta} + 1 - \eta}{\eta c_{21}^2},$
$\hat{S}_{\text{sat}}$	$p_1 = \dfrac{P}{\beta} - \dfrac{\bar{\beta}(\gamma_1 - 1)}{\beta c_{12}^2},$		$q_1 = \dfrac{\gamma_1 - 1}{c_{12}^2},$		$p_2 = \dfrac{P}{\beta} - \dfrac{\bar{\beta}(\gamma_2 - 1)}{\beta c_{21}^2},$		$q_2 = \dfrac{\gamma_2 - 1}{c_{21}^2},$
$\hat{S}_{CD}$	$p_1 = \dfrac{P - \bar{\beta}q_1}{\beta},$		$q_1 = \dfrac{\hat{F}(P)}{c_{12}^2},$		$p_2 = \dfrac{P - \bar{\beta}q_2}{\beta},$		$q_2 = \dfrac{1 + c_{12}^2 q_1}{\eta c_{21}^2} - \dfrac{1}{c_{21}^2},$

Here $\gamma_k := ((1 + Q_1)(1 + Q_2)/(1 + Q_l + a_{lk}^2 Q_k))^{1/\alpha\bar{\beta}}, l \neq k \in \{1, 2\}$*, and* $\hat{F}(P)$ *and* η *are defined in* (39) *and* (40) *below*

$$\hat{F}(P) := \left(E_1 + \eta E_2 - \sqrt{(E_1 - \eta E_2)^2 + 4\beta^2 \eta E_1 E_2}\right)/(2\bar{\beta}(1 + \beta)) - 1, \tag{39}$$

$$\eta := \left(\frac{1 + Q_1 + a_{21}^2 Q_2}{1 + Q_2 + a_{12}^2 Q_1}\right)^{1/\alpha\bar{\beta}} > 1, \tag{40}$$

with E_1, E_2*, defined in* (37)*. Note that, under* [A2] *(*$\eta = 1$*), the powers in Table 2 simplifies to that in Table 1.*

Next, we present the optimal powers and the critical powers of the three new sets in Table 3, where

$$P_5^* := \bar{\beta}\frac{(\eta - 1)}{c_{12}^2}, \quad P_6^* := \bar{\beta}\left(\frac{1}{d_1^2} - \frac{1}{c_{12}^2}\right), \quad P_7^* := \beta\left(\frac{1}{c_{12}^2} - \frac{1}{d_1^2}\right), \quad P_8^* := \frac{\eta - \bar{\beta}}{c_{12}^2} - \frac{\beta}{d_1^2},$$

$$f_3(P) := \frac{d_2^2}{c_{21}^2(1 + d_2^2 P/\beta)} + \frac{\beta\eta}{Pc_{12}^2 + \bar{\beta} + \beta c_{12}^2/d_1^2 - \bar{\beta}\eta}.$$

The power allocation for the new sets can be understood by following the WF-like and max-min properties of $\mathbf{OA_G}$*. For illustration, we explain the allocations in* $\hat{S}_1$ *and* $\hat{S}_2$*. In* $\hat{S}_1$*, the cross channels are stronger than the direct channels. Thus, due to its WF-like property,* $\mathbf{OA_G}$ *preferably allocates as much of P to the cross channels as possible. However, under* [A2G] *(*$A_2 < A_1$*) and* $c_{12}^2 < c_{21}^2$*, allocating power to both cross channels results in* $\Sigma_2 < \Sigma_1$*, and does not increase* $\min\{\Sigma_2, \Sigma_1\}$ *as much as possible. This results in a suboptimal sum-rate.*

Therefore, **OA$_G$** *allocates* $p_1 = 0, q_1 = P/\bar{\beta}$*, which increases only* Σ_2*, and assigns* $p_2 = P/\beta, q_2 = 0$*, which represents the allocation in* $\hat{\mathcal{S}}_1$*.*

Table 3. Optimal Power Allocation and Conditions of the "New" Sets in **OA$_G$**.

Set	Optimal Powers				Condition		
$\hat{\mathcal{S}}_1$	$p_1 = 0,$	$q_1 = \dfrac{P}{\bar{\beta}},$	$p_2 = \dfrac{P}{\beta},$	$q_2 = 0,$	$P \leq P_5^*,$	$P < P_6^*$	
$\hat{\mathcal{S}}_2$	$p_1 = P + \dfrac{\bar{\beta}}{c_{12}^2} - \dfrac{\bar{\beta}}{d_1^2},$	$q_1 = P + \dfrac{\beta}{d_1^2} - \dfrac{\beta}{c_{21}^2},$	$p_2 = \dfrac{P}{\beta},$	$q_2 = 0,$	$P > P_6^*,$	$P > P_7^*,$	$P < P_8^*$
$\hat{\mathcal{S}}_3$	$p_1 = \dfrac{P}{\beta} - \dfrac{\bar{\beta}(\eta-1)}{\beta c_{12}^2},$	$q_1 = \dfrac{\eta-1}{c_{12}^2},$	$p_2 = \dfrac{P}{\beta},$	$q_2 = 0,$	$P > P_5^*,$	$P \geq P_8^*,$	$f_3(P) > 1$

Furthermore, as P increases and $P > \max\{P_6^*, P_7^*\}$*, additional benefits from transmitting only in the cross channel with gain* c_{12}^2 *reduces, and thus* **OA$_G$** *starts allocating a fraction of P to the direct channel with gain* d_1^2*, due to its WF-like property. Thus,* **OA$_G$** *follows the power allocation in* $\hat{\mathcal{S}}_2$*, which remains optimal as long as* $P < P_8^*$*. The power allocation in* $\hat{\mathcal{S}}_3$ *can be interpreted similarly.*

The important insight is that **OA$_G$** *can be explained with the WF-like and max-min properties as in* **OA$_1$***, and it does not reveal any new fundamental properties.*

5.3. Optimum Power Allocation in the Symmetric DCLIC

We briefly discuss the optimum allocation for the symmetric DCLIC, where $a^2 := a_{12}^2 = a_{21}^2$, $c^2 := c_{12}^2 = c_{21}^2$, $d^2 := d_1^2 = d_2^2$, and $Q_1 = Q_2$. Due to symmetry, considering only symmetric power allocation of the form (p, q, p, q) is sufficient, and does not cause loss of generality. Moreover, any feasible (symmetric) power allocation achieves $\Sigma_1 = \Sigma_2$, rendering the constraint $R \leq \Sigma_2$ in (28) redundant. The sum-rate optimization problem in this case can be formulated and solved as in [**P1**], and is omitted here for brevity. In addition, we denote the optimal power allocation for this case by **OA$_S$**.

In this case, **OA$_S$** can be described by the power allocation in four disjoint sets, $\bar{\mathcal{S}}_l, l \in \{D, C, CD, \text{sat}\}$, which are counterparts of the sets $\mathcal{S}_l$ in Table 1. The conditions and optimal powers in these sets are given in Table 4, and they can be derived following the same procedure as in Appendix C, and thus is omitted here.

Table 4. Optimal power allocation and conditions of sets in **OA$_S$**.

Set	Optimal Powers		Condition		
$\bar{\mathcal{S}}_D$	$p = \dfrac{P}{\beta},$	$q = 0,$	$P < \bar{P}_1^*$		
$\bar{\mathcal{S}}_C$	$p = 0,$	$q = \dfrac{P}{\bar{\beta}},$	$P < \bar{P}_2^*,$	$P < \bar{P}_4^*$	
$\bar{\mathcal{S}}_{CD}$	$p = \dfrac{2}{1+\beta}\left(P + \dfrac{\bar{\beta}}{c^2} - \dfrac{\bar{\beta}}{2d^2}\right),$	$q = \dfrac{1}{1+\beta}\left(P + \dfrac{\beta}{d^2} - \dfrac{2\beta}{c^2}\right),$	$P \geq \bar{P}_1^*,$	$P \geq \bar{P}_2^*,$	$P \leq \bar{P}_3^*$
$\bar{\mathcal{S}}_{\text{sat}}$	$p = \dfrac{P}{\beta} - \dfrac{\bar{\beta}(\bar{\gamma}-1)}{\beta c^2},$	$q = \dfrac{\bar{\gamma}-1}{c^2},$	$P \geq \bar{P}_4^*,$	$P > \bar{P}_3^*$	

Moreover, the critical powers are now given by

$$\bar{P}_1^* := \beta\left(\frac{2}{c^2} - \frac{1}{d^2}\right), \quad \bar{P}_2^* := \frac{\bar{\beta}}{2}\left(\frac{1}{d^2} - \frac{2}{c^2}\right), \quad \bar{P}_3^* := \frac{(1+\beta)\bar{\gamma}}{c^2} - \frac{\beta}{d^2} - \frac{\bar{\beta}}{c^2}, \quad \bar{P}_4^* := \frac{\bar{\beta}(\bar{\gamma}-1)}{c^2},$$

where $\bar{\gamma} := ((1+Q_1)^2/(1+Q_1+a^2Q_1))^{1/\alpha\bar{\beta}} > 1$.

OA$_S$ has the same WF-like and max-min properties as that of **OA$_1$**. In fact, a rudimentary inspection of the conditions reveals that **OA$_S$** follows one of the three possible sequences of sets depending on the channel gains, as before:

- If $c^2 \leq 2d^2$ (direct channels are "stronger"): **OA$_S$** follows the sequence $\bar{\mathcal{S}}_D \rightarrow \bar{\mathcal{S}}_{CD} \rightarrow \bar{\mathcal{S}}_{\text{sat}}$. It transmits only in the direct channels as in $\bar{\mathcal{S}}_D$ when $P < \bar{P}_1^*$, then transmits in all channels as in $\bar{\mathcal{S}}_{CD}$ when $\bar{P}_1^* \leq P \leq P_3^*$, and finally starts following the allocation in $\bar{\mathcal{S}}_{\text{sat}}$ when $P > \bar{P}_3^*$.

- If $2d^2 < c^2 < 2d^2\bar{\gamma}$ (cross channels are "stronger"): **OA$_S$** follows the sequence $\bar{\mathcal{S}}_C \rightarrow \bar{\mathcal{S}}_{CD} \rightarrow \bar{\mathcal{S}}_{\text{sat}}$. It transmits only in the cross channels as in $\bar{\mathcal{S}}_C$ when $P < \bar{P}_2^*$, then transmits in all channels as in $\bar{\mathcal{S}}_{CD}$ when $\bar{P}_2^* \leq P \leq \min\{P_3^*, P_4^*\}$, and finally follows $\bar{\mathcal{S}}_{\text{sat}}$ when $P > \min\{P_3^*, P_4^*\}$.

- If $c^2 \geq 2d^2\bar{\gamma}$ (cross channels are much "stronger"): **OA$_S$** follows the sequence $\bar{\mathcal{S}}_C \rightarrow \bar{\mathcal{S}}_{\text{sat}}$. It follows $\bar{\mathcal{S}}_C$ when $P < \bar{P}_4^*$, and then follows $\bar{\mathcal{S}}_{\text{sat}}$ when $P \geq \bar{P}_4^*$, while skipping $\bar{\mathcal{S}}_{CD}$ altogether.

Thus, **OA$_S$** allocates all of P to either both direct or both cross channels if P is small, and then shares P among all channels as P increases. In addition, when P is increased beyond the saturation threshold, the cross channels become saturated. Note that, due to symmetry, **OA$_S$** needs to transmit in both cross channels only in $\bar{\mathcal{S}}_C$ to preserve $\Sigma_1 = \Sigma_2$, unlike in $\mathcal{S}_C$, where **OA$_1$** additionally transmits in a direct channel.

We note the following, which are revealed due to symmetry. First, if $c^2 < 2d^2$, the direct channels are considered to be stronger than the cross channels. Similarly, if $c^2 > 2d^2$, the cross channels are considered to be stronger. Moreover, in the special case of $c^2 = 2d^2$, neither the direct nor the cross channels are stronger than the other, and thus allocating P entirely to only direct or cross channel would be suboptimal. Therefore, **OA$_S$** transmits in all channels as in $\mathcal{S}_{CD}$ when $P \geq 0$. Finally, the condition $c^2 \geq 2d^2\gamma$ implies that the cross channels are much stronger than the direct channels. Thus, it is optimal to transmit only in the cross channels as in $\bar{\mathcal{S}}_C$ until they become saturated, at which point the allocation in $\bar{\mathcal{S}}_{\text{sat}}$ becomes optimal. Note that these conditions follow trivially from the corresponding conditions of [S1], [S2], and [S3] in Section 5.2 by applying symmetry.

Furthermore, due to symmetry, one can now directly observe that the conditions of the sets in Table 4 are mutually exclusive. For example, consider the conditions of $\bar{\mathcal{S}}_D$, $c^2 < 2d^2$ and $P < \bar{P}_1^*$, and note that they violate the second condition of $\bar{\mathcal{S}}_C$ (since $\bar{P}_2^* < 0$), the first condition in $\bar{\mathcal{S}}_{CD}$ (since $P < \bar{P}_1^*$), and the second condition in $\bar{\mathcal{S}}_{\text{sat}}$ (trivially follows from $\bar{\gamma} > 1$).

In Figure 6, we depict the set of cross and direct channel gains (c^2 and d^2) of a symmetric DCLIC with parameters $a^2 = 1.5, Q_1 = 5, \alpha = 2, \beta = 0.5$, and partition it depending on whether the direct or the cross channels are stronger.

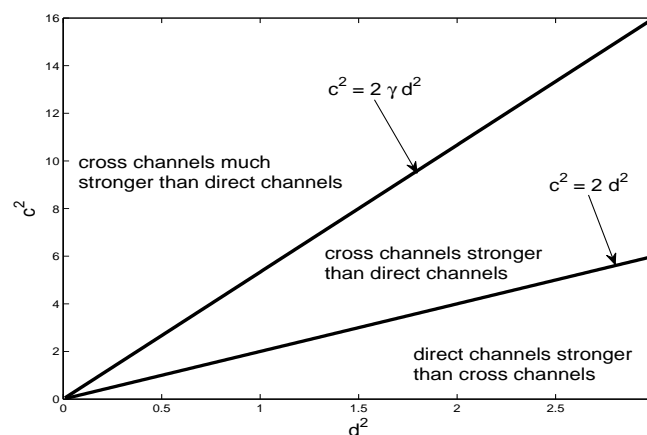**Figure 6.** The set of all c^2 and d^2 is partitioned depending on whether the cross or the direct channels are "stronger".

5.4. Discussion and Insights

Now, we discuss some insights obtained from the power allocation in **OA$_1$** in Table 1.

First, if the underlying GIC in the first band has *very strong interference*, i.e., $a_{12}^2 \geq 1 + Q_1$ and $a_{21}^2 \geq 1 + Q_2$, **OA$_1$** allocates the power budget (P) entirely to direct channels as in $\mathcal{S}_D$ for all values of P. Note that, under very strong interference, we have $\Sigma \leq \min\{\Sigma_1, \Sigma_2\}$ for any feasible power allocation. Therefore, the sum-rate is maximized by maximizing Σ, which is achieved by allocating all of P to the direct channels.

Second, if the transmit powers in the underlying GIC of the strong CLIC (Q_1 and Q_2) are small, i.e., $Q_1 \approx 0, Q_2 \approx 0$, then allocating P entirely to the direct channels as in $\mathcal{S}_D$, is approximately optimal for all values of P, in the sense that, the difference between the sum-rates achieved with the allocation in $\mathcal{S}_D$ and that in **OA$_1$** is very small. Specifically, since $Q_1 \approx 0, Q_2 \approx 0$, the gap between Σ and Σ_1 at $P = 0$ is very small. As P is sufficiently increased, **OA$_1$** starts transmitting in all channels as in $\mathcal{S}_{CD}$. However, since the gap is very small, only a small fraction of P needs to be allocated to the cross channels to offset the gap, and thus almost all of P is redirected to the direct channels. The resulting allocation closely resembles that in $\mathcal{S}_D$.

We can attribute this behavior of **OA$_1$** to the fact that the GIC with strong interference ($a_{12}^2 \geq 1, a_{21}^2 \geq 1$) and small transmit powers ($Q_1 \approx 0, Q_2 \approx 0$) approaches the very strong interference regime ($a_{12}^2 \geq 1 + Q_1, a_{21}^2 \geq 1 + Q_2$), where allocating P entirely to the direct channels as in $\mathcal{S}_D$ is optimal for all P.

Third, if the transmit powers in the underlying GIC of the strong CLIC are large, i.e., $Q_1, Q_2 \gg 1$, **OA$_1$** allocates power in all channels as in $\mathcal{S}_{CD}$ for all but a relatively small range of P. Recall that even if **OA$_1$** transmits in the direct or cross channels as in $\mathcal{S}_D$ or $\mathcal{S}_C$ for small P, eventually when $P \geq P_1^*$ or $P \geq P_2^*$, **OA$_1$** allocates power in all channels as in $\mathcal{S}_{CD}$, where the gap between Σ_1 and Σ starts to reduce. However, if $Q_1, Q_2 \gg 1$, this gap is very large, and thus very large P is needed to offset the gap as in $\mathcal{S}_{sat}$. Therefore, for all moderate P, **OA$_1$** allocates power to all channels following $\mathcal{S}_{CD}$.

Fourth, if the cross channel gains in the second band are very large, allocating P entirely to the direct channels as in $\mathcal{S}_D$ is approximately optimal for all P, in the sense described in the second point above. This follows since with very large c_{12}^2 and c_{21}^2, **OA$_1$** needs to allocate only a very small fraction of P to close the gap between Σ_1 and Σ, and thus redirects the remaining of P to the direct channels. The resulting allocation thus closely resembles that in $\mathcal{S}_D$.

Finally, when $\beta \approx 1$, the DCLIC can be approximated by the DLIC, and thus allocating P entirely to only the direct channels is approximately optimal. On the other hand, when $\beta \approx 0$, the DCLIC can be approximated by the CLIC where allocating P entirely to the cross channels is approximately optimal. However, since the cross channels do not increase Σ, the sum-rate becomes limited to Σ when P is sufficiently large, and does not increase anymore.

6. Conclusions

Motivated by simultaneous transmission in the microwave and mm-wave bands, we considered a special class of dual band 2×2 GIC. Due to their small wavelength and the use of highly directional antenna arrays, the mm-wave band can be used to form highly directional point-to-point links, allowing interference free reception at a designated receiver. For this model, we first showed that the capacity region of a DCLIC can be decomposed into the capacity of a CLIC complemented by direct channels. Moreover, we considered two specific classes of the CLIC with strong and weak underlying GIC, and obtained sufficient conditions on the channel gains under which the capacity region is characterized.

We characterized the power allocation over the mm-wave channels that maximizes the sum-rate of the DCLIC. Channel conditions have been characterized that determine how power is allocated to the channels depending on whether the direct or the cross channels are "strong". In particular, the optimal allocation has a WF-like property due to which, when the power budget is small, it assigns the power budget entirely to either both direct channels, or both cross channels and at most one direct channel. As the power budget is increased, the optimal allocation eventually assigns power to all channels.

It also has a max-min property, and, as a result, imposes a maximum limit on the cross-channel powers. This power allocation evaluates the impact of high bandwidth point-to-point mm-wave channels on the sum-rate performance, and thus can be useful in allocating resources in practice.

Potential future work includes analyzing other multi-user networks that operate with dual microwave and mm-wave bands.

Author Contributions: Both Subhajit Majhi and Patrick Mitran conceived of the idea of a dual-band GIC. Subhajit Majhi derived the results and completed writing the manuscript, with Patrick Mitran providing guidance on improving the quality of the presentation of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GIC	Gaussian interference channel
CLIC	Cross-Link interference channel
DLIC	Direct-Link interference channel
DCLIC	Direct-and-Cross-Link interference channel
PGIC	Parallel Gaussian interference channel
BMF	Bandwidth mismatch factor
DM	Discrete memoryless
RV	Random variable
WF	Waterfilling
KKT	Karush–Kuhn–Tucker

Appendix A. Proof of Theorem 1

Proof. Outer bound: We derive the bounds on R_1 and R_2 as follows. If transmitter Tx_k transmits the message $M_k, k = 1, 2$, we have from Fano's inequality

$$
\begin{aligned}
nR_1 &\leq I(X_1^n, \hat{X}_1^{n_1}, \bar{X}_1^{n_2}; Y_1^n, \hat{Y}_1^{n_1}, \bar{Y}_1^{n_2}) + n\epsilon_n \\
&\overset{(a)}{=} I(X_1^n, \hat{X}_1^{n_1}; Y_1^n, \hat{Y}_1^{n_1}) + I(\bar{X}_1^{n_2}; Y_1^n, \hat{Y}_1^{n_1} | X_1^n, \hat{X}_1^{n_1}) \\
&\quad + I(\bar{X}_1^{n_2}; \bar{Y}_1^{n_2} | Y_1^n, \hat{Y}_1^{n_1}) + I(X_1^n, \hat{X}_1^{n_1}; \bar{Y}_1^{n_2} | \bar{X}_1^{n_2}, Y_1^n, \hat{Y}_1^{n_1}) + n\epsilon_n \\
&\overset{(b)}{=} I(X_1^n, \hat{X}_1^{n_1}; Y_1^n, \hat{Y}_1^{n_1}) + I(\bar{X}_1^{n_2}; \bar{Y}_1^{n_2} | Y_1^n, \hat{Y}_1^{n_1}) + n\epsilon_n \\
&= I(X_1^n, \hat{X}_1^{n_1}; Y_1^n, \hat{Y}_1^{n_1}) + \sum_{\ell=1}^{n_2} h(\bar{Y}_{1\ell} | \bar{Y}_1^{\ell-1}, Y_1^n, \hat{Y}_1^{n_1}) - h(\bar{Y}_{1\ell} | \bar{Y}_1^{\ell-1}, \bar{X}_1^{n_2}, Y_1^n, \hat{Y}_1^{n_1}) + n\epsilon_n \\
&\overset{(c)}{\leq} I(X_1^n, \hat{X}_1^{n_1}; Y_1^n, \hat{Y}_1^{n_1}) + \sum_{\ell=1}^{n_2} h(d_1 \bar{X}_{1\ell} + \bar{Z}_{1\ell}) - h(\bar{Z}_{1\ell}) + n\epsilon_n \\
&\overset{(d)}{\leq} I(X_1^n, \hat{X}_1^{n_1}; Y_1^n, \hat{Y}_1^{n_1}) + n_2 \mathsf{C}(d_1^2 \bar{P}_1) + n\epsilon_n,
\end{aligned}
\tag{A1}
$$

where (*a*) follows from the chain rule; (*b*) follows since the second term of (*a*) is zero due to the Markov chain (MC) $\bar{X}_1^{n_2} \to (X_1^n, \hat{X}_1^{n_1}) \to (Y_1^n, \hat{Y}_1^{n_1})$, and the fourth term of (*a*) is zero due to the MC $(Y_1^n, \hat{Y}_1^{n_1}, X_1^n, \hat{X}_1^{n_1}) \to \bar{X}_1^{n_2} \to \bar{Y}_1^{n_2}$; (*c*) follows since unconditioning does not reduce entropy, and also from the memoryless Gaussian model of the direct channels; and (*d*) follows since $\bar{X}_{1\ell} \sim \mathcal{N}(0, \bar{P}_1)$, i.i.d., maximizes the mutual information. Bounding R_2 similarly, and since $n_2 \leq \alpha_2 n$, we have

$$
R_1 \leq r_1 + \alpha_2 \mathsf{C}(d_1^2 \bar{P}_1), \quad R_2 \leq r_2 + \alpha_2 \mathsf{C}(d_2^2 \bar{P}_2),
$$

where $r_k := \sup_{X_k^n, \hat{X}_k^{n_1}} \frac{1}{n} I(X_k^n, \hat{X}_k^{n_1}; Y_k^n, \hat{Y}_k^{n_1}), k = 1, 2$, for product distribution $\prod_{k=1}^2 p(x_k^n, \hat{x}_k^{n_1})$, with $n_1 \leq \alpha_1 n$.

Achievability: We will code over t blocks of symbols together. We choose (n, n_1, n_2), and define $\mathbf{U}_k := (X_k^n, \hat{X}_k^{n_1}), k = 1, 2$, where $p(\mathbf{U}_1, \mathbf{U}_2) = p(\mathbf{U}_1)p(\mathbf{U}_2)$, and define $\bar{\mathbf{U}}_k := \bar{X}_k^{n_2}, k = 1, 2$ and choose $\bar{X}_{k\ell} \sim \mathcal{N}(0, \bar{P}_k)$, i.i.d., for $\ell = 1, \ldots, n_2$. Suppose Tx_k transmits $M_k \in \mathcal{M}_k, k = 1, 2$. Thus, to encode M_k over t blocks for the underlying CLIC, we generate 2^{tnR_k} i.i.d. sequences $\mathbf{U}_k^t(M_k)$, one for each $M_k \in \mathcal{M}_k$, where $\mathbf{U}_k^t(M_k)$ are distributed according to $p(\mathbf{u}_k^t) = \prod_{\ell=1}^t p(\mathbf{u}_{k,\ell}) = \prod_{\ell=1}^t p(x_{k,(\ell-1)n+1}^{n\ell} \hat{x}_{k,(\ell-1)n_1+1}^{n_1\ell}), k = 1, 2$. Similarly, for the two direct channels, we generate 2^{tnR_k} i.i.d. sequences $\bar{\mathbf{U}}_k^t(M_k)$, distributed as $\prod_{\ell=1}^{n_2 t} p(\bar{x}_{k\ell})$, where $\bar{x}_{k\ell} \sim \mathcal{N}(0, \bar{P}_k)$, for $k = 1, 2$, and $\ell = 1, \ldots, n_2$.

Thus, to transmit $M_k, k = 1, 2$, $\mathbf{U}_1^t(M_1)$ and $\mathbf{U}_2^t(M_2)$ are transmitted through the CLIC, and $\bar{\mathbf{U}}_1^t(M_1)$ and $\bar{\mathbf{U}}_2^t(M_2)$ are transmitted through the two direct channels in the second band. Upon receiving the sequences $(Y_k^{nt}, \hat{Y}_k^{n_1 t})$ and $\bar{Y}_k^{n_2 t}$, Rx_k employs joint typical decoding to estimate the transmitted message as in [32] (Chapter 4.3). The probability of decoding error vanishes with $t \to \infty$ if

$$R_k < \frac{1}{n} I(X_k^n, \hat{X}_k^{n_1}; Y_k^n, \hat{Y}_k^{n_1}) + \frac{n_2}{n} \mathsf{C}(d_k^2 \bar{P}_k), \quad k = 1, 2, \tag{A2}$$

which matches the upper bound on R_1 and R_2, as $n_2 \leq \alpha_2 n$, and $n \to \infty$. Finally, the capacity region of the CLIC with BMF α is defined as the closure of the union of all sets of achievable rate pairs (R_1, R_2) that satisfy (A2) with $n_2 = 0$, where the union is taken over all product distributions, $p(x_1^n, \hat{x}_1^{n_1})p(x_2^n, \hat{x}_2^{n_1})$, and for all $n, n_1 \in \mathbb{Z}_+$, such that $n \leq \alpha_1 n$. $\square$

Appendix B. Proof of Theorem 2

Proof. Outer bound: We derive the bound on R_1 first. Assuming that the transmitter Tx_k transmits the message $M_k, k = 1, 2$, we have from Fano's inequality

$$\begin{aligned}
nR_1 &\leq I(X_1^n, \hat{X}_1^n; Y_1^n, \hat{Y}_1^n) + n\epsilon_n \\
&\leq I(X_1^n, \hat{X}_1^n; Y_1^n, \hat{Y}_1^n, X_2^n, \hat{X}_2^n) + n\epsilon_n \\
&\stackrel{(a)}{=} I(X_1^n, \hat{X}_1^n; Y_1^n, \hat{Y}_1^n | X_2^n, \hat{X}_2^n) + n\epsilon_n \\
&\stackrel{(b)}{=} I(X_1^n, \hat{X}_1^n; Y_1^n | X_2^n, \hat{X}_2^n) + I(X_1^n, \hat{X}_1^n; \hat{Y}_1^n | X_2^n, \hat{X}_2^n, Y_1^n) + n\epsilon_n \\
&\stackrel{(c)}{=} I(X_1^n; Y_1^n | X_2^n) + n\epsilon_n \\
&\stackrel{(d)}{\leq} \sum_{\ell=1}^n I(X_{1\ell}; Y_{1\ell} | X_{2\ell}) + n\epsilon_n \\
&\stackrel{(e)}{\leq} n\mathsf{C}(Q_1) + n\epsilon_n,
\end{aligned} \tag{A3}$$

where (a) follows since $(X_1^n, \hat{X}_1^n) \perp\!\!\!\perp (X_2^n, \hat{X}_2^n)$; (b) follows from the chain rule; (c) follows from the first term in (b) due to the MCs $\hat{X}_2^n \to X_2^n \to Y_1^n$ and $(\hat{X}_1^n, \hat{X}_2^n) \to (X_1^n, X_2^n) \to Y_1^n$, and since the second term in (b) is zero due to $(X_1^n, \hat{X}_1^n, X_2^n, Y_1^n) \to \hat{X}_2^n \to \hat{Y}_1^n$; (d) follows since unconditioning does not reduce entropy, and also from the memoryless Gaussian model; and (e) follows from the Gaussian system model, and by choosing $X_{1\ell} \sim \mathcal{N}(0, Q_1)$ i.i.d., which maximizes the mutual information. We similarly bound R_2 by, $R_2 \leq \mathsf{C}(Q_2)$.

Next, we bound the sum-rate. We have from Fano's inequality

$$
\begin{aligned}
nR_1 + nR_2 &\leq I(X_1^n, \hat{X}_1^n; Y_1^n, \hat{Y}_1^n) + I(X_2^n, \hat{X}_2^n; Y_2^n, \hat{Y}_2^n) + n\epsilon_n \\
&\leq I(X_1^n, \hat{X}_1^n; Y_1^n, \hat{Y}_1^n, X_2^n, \hat{X}_2^n) + I(X_2^n, \hat{X}_2^n; Y_2^n, \hat{Y}_2^n, X_1^n, \hat{X}_1^n) + n\epsilon_n \\
&\overset{(a)}{=} I(X_1^n; Y_1^n | X_2^n) + I(X_2^n; Y_2^n | X_1^n) + n\epsilon_n \\
&\overset{(b)}{\leq} I(X_1^n; Y_2^n) + I(X_2^n; Y_2^n | X_1^n) + n\alpha_1 \mathsf{C}(c_{12}^2 \hat{P}_1) + n\epsilon_n \\
&= I(X_1^n, X_2^n; Y_2^n) + n\alpha_1 \mathsf{C}(c_{12}^2 \hat{P}_1) + n\epsilon_n \\
&\overset{(c)}{\leq} \sum_{\ell=1}^n I(X_{1\ell}, X_{2\ell}; Y_{2\ell}) + n\alpha_1 \mathsf{C}(c_{12}^2 \hat{P}_1) + n\epsilon_n \\
&\overset{(d)}{\leq} n\mathsf{C}(a_{12}^2 Q_1 + Q_2) + n\alpha_1 \mathsf{C}(c_{12}^2 \hat{P}_1) + n\epsilon_n,
\end{aligned}
\tag{A4}
$$

where the first term in (*a*) follows from the steps (*a*)–(*c*) of (A3), and the second term follows similarly; (*b*) follows from (18) in Lemma 2; (*c*) follows since unconditioning does not reduce entropy, and also from the memoryless Gaussian model; and (*d*) follows from choosing $X_{k\ell} \sim \mathcal{N}(0, Q_k), k = 1, 2$, i.i.d., which maximizes the mutual information. Hence as $n \to \infty$, and thus $\epsilon_n \to 0$, we obtain the bound in (23). The other bound in (22) is found similarly. Thus, (20)–(23) provides an outer bound on the capacity region of the CLIC.

　　Achievability: We fix n and n_1, and choose independent input distributions for the codewords of two users in each channel, i.e., $p(x_1, x_2) = p(x_1)p(x_2)$, and $\hat{p}(\hat{x}_1, \hat{x}_2) = \hat{p}(\hat{x}_1)\hat{p}(\hat{x}_2)$. The transmitter Tx_k encodes a message, $M_k \in \mathcal{M}_k$ into two codewords, $x_k^n(M_k)$ and $\hat{x}_k^{n_1}(M_k)$, generated according to the i.i.d. distributions, $\prod_{\ell=1}^n p_k(x_{k\ell})$ and $\prod_{\ell=1}^{n_1} \hat{p}_k(\hat{x}_{k\ell}), k = 1, 2$. The codewords, $x_k^n(M_k)$ and $\hat{x}_k^{n_1}(M_k)$, are then transmitted through the IC in the first band and the cross channels in the second band, respectively. Each receiver then estimates both messages, (M_1, M_2), from the signals observed over both bands as in a multiple access channel. We use standard error analysis technique [32] (Chapter 4.5) to find that the probability of decoding error becomes arbitrarily small as $n \to \infty$, if the rates achieved in the receiver Rx_k satisfies

$$
\begin{aligned}
R_k &\leq I(X_k; Y_k | X_m), \\
R_m &\leq I(X_m; Y_k | X_k) + \alpha_1 I(\hat{X}_m; \hat{Y}_k), \\
R_1 + R_2 &\leq I(X_1, X_2; Y_k) + \alpha_1 I(\hat{X}_m; \hat{Y}_k),
\end{aligned}
\tag{A5}
$$

for $k \neq m \in \{1, 2\}$. Moreover, these achievable rates are maximized by using i.i.d. Gaussian inputs, $X_k^* \sim \mathcal{N}(0, Q_k)$ and $\hat{X}_k^* \sim \mathcal{N}(0, \hat{P}_k), k = 1, 2$. Finally, observe that (18) and (19) in Lemma 2 are valid for the independent Gaussian inputs X_1^* and X_2^* with $n = 1$, which implies that the individual rates must satisfy, $R_k \leq I(X_k^*; Y_k^* | X_m^*) \leq I(X_k^*; Y_m^*) + \alpha_1 \mathsf{C}(c_{km}^2 \hat{P}_k) \leq I(X_k^*; Y_m^* | X_m^*) + \alpha_1 \mathsf{C}(c_{km}^2 \hat{P}_k)$, where Y_k^* is the channel output with the Gaussian inputs $X_k^*, k = 1, 2$. Hence, the individual rates are limited to $R_k \leq I(X_k^*; Y_k^* | X_m^*) = \mathsf{C}(Q_k), k \neq m \in \{1, 2\}$, which matches the bound in (20) and (21). In addition, the sum-rate constraints in (A5) above match the sum-rate bounds in (22) and (23). Thus, the outer bound is achieved with independent Gaussian inputs. □

Appendix C. Derivation of the Optimal Power Allocation

　　The KKT Conditions: We show that [**P1**] is convex, and formulate its KKT conditions. Denote by $\mathbf{x} := (p_1, q_1, p_2, q_2, R) \in \mathbb{R}_+^5$ a feasible point, satisfying the constraints in [**P1**]. Now, the objective of [**P1**] is equivalent to *minimizing* $-R$, which is linear, and the equality constraints (30) and (31) are affine. Next, we note that the constraints (27)–(29) are convex. To illustrate, we define $G_1(\mathbf{x}) := R - A_1 - \frac{\alpha\bar{\beta}\kappa}{2} \ln(1 + c_{21}^2 q_2) - \frac{\alpha\beta\kappa}{2} \ln(1 + d_1^2 p_1) - \frac{\alpha\beta\kappa}{2} \ln(1 + d_2^2 p_2)$ corresponding to (27) with $\kappa := 1/\ln 2$, and derive its Hessian, $\nabla^2 G_1(\mathbf{x}) := \frac{\alpha\kappa}{2} \text{diag}\left[\left(\beta d_1^4/(1 + d_1^2 p_1)^2, \; 0, \; \beta d_2^4/(1 + d_2^2 p_2)^2, \; \bar{\beta} c_{21}^4/(1 + c_{21}^2 q_2)^2, \; 0\right)\right]$, where

$D = \text{diag}\,[a_1, \ldots, a_m]$ is a diagonal matrix with elements $a_1, \ldots, a_m$. Note that $\nabla^2 G_1(\mathbf{x})$ is positive semidefinite, and thus (27) is convex. Likewise, (28) and (29) are found to be convex. In addition, (30) and (31) and $\mathbf{x} \succeq \mathbf{0}$ imply that the feasible set is compact for given $P > 0$ and $\beta \in (0,1)$. Hence, [**P1**] is a convex problem over a compact set. Furthermore, [**P1**] satisfies Slater's condition [35] (Chapter 5.2.3), since the point $\tilde{\mathbf{x}} := (P - \bar{\beta}\epsilon/\beta, \ \epsilon, \ P - \bar{\beta}\epsilon/\beta, \ \epsilon, \ A_1)$ is strictly feasible for sufficiently small $\epsilon > 0$. Therefore, [**P1**] can be solved using the KKT conditions [35] (Chapter 5.5.3).

We define the Lagrange multipliers $\{\lambda_i\}_{i=1}^3$, corresponding to (27)–(29), $\{\mu_i\}_{i=1}^2$, corresponding to (30) and (31), and $\{\rho_i\}_{i=1}^5$, corresponding to the nonnegativity constraints, $(p_1, q_1, p_2, q_2, R) \succeq \mathbf{0}$. The Lagrangian is then

$$\begin{aligned}
\mathcal{L} = -R &+ \lambda_1(R - \Sigma_1) + \lambda_2(R - \Sigma_2) + \lambda_3(R - \Sigma) \\
&+ \mu_1(\beta p_1 + \bar{\beta} q_1 - P) + \mu_2(\beta p_2 + \bar{\beta} q_2 - P) - \rho_1 p_1 - \rho_2 q_1 - \rho_3 p_2 - \rho_4 q_2 - \rho_5 R,
\end{aligned} \tag{A6}$$

where Σ_1, Σ_2 and Σ, are defined in (24)–(26) with $A_1 = A_2$.

With a slight abuse of notation, we denote the optimal primal variable by (p_1, q_1, p_2, q_2, R), and the optimal Lagrange multipliers by $(\lambda_1, \lambda_2, \lambda_3, \mu_1, \mu_2, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5)$. The optimal variables satisfy the KKT conditions below

$$\lambda_1 + \lambda_2 + \lambda_3 = 1, \tag{A7}$$

$$\rho_1 = \beta\left(\mu_1 - \frac{\alpha\kappa}{2}\frac{d_1^2}{1 + d_1^2 p_1}\right), \tag{A8}$$

$$\rho_2 = \bar{\beta}\left(\mu_1 - \frac{\alpha\kappa}{2}\frac{c_{12}^2\lambda_2}{1 + c_{12}^2 q_1}\right), \tag{A9}$$

$$\rho_3 = \beta\left(\mu_2 - \frac{\alpha\kappa}{2}\frac{d_2^2}{1 + d_2^2 p_2}\right), \tag{A10}$$

$$\rho_4 = \bar{\beta}\left(\mu_2 - \frac{\alpha\kappa}{2}\frac{c_{21}^2\lambda_1}{1 + c_{21}^2 q_2}\right), \tag{A11}$$

$$\beta p_1 + \bar{\beta} q_1 = P, \tag{A12}$$

$$\beta p_2 + \bar{\beta} q_2 = P, \tag{A13}$$

$$R - \Sigma_1 \leq 0, \tag{A14}$$

$$R - \Sigma_2 \leq 0, \tag{A15}$$

$$R - \Sigma \leq 0, \tag{A16}$$

$$\lambda_1(R - \Sigma_1) = 0, \tag{A17}$$

$$\lambda_2(R - \Sigma_2) = 0, \tag{A18}$$

$$\lambda_3(R - \Sigma) = 0, \tag{A19}$$

$$\rho_1 p_1 = 0, \tag{A20}$$

$$\rho_2 q_1 = 0, \tag{A21}$$

$$\rho_3 p_2 = 0, \tag{A22}$$

$$\rho_4 q_2 = 0, \tag{A23}$$

$$(p_1, q_1, p_2, q_2, R, \lambda_1, \lambda_2, \lambda_3, \rho_1, \rho_2, \rho_3, \rho_4) \succeq \mathbf{0}, \tag{A24}$$

where we used the fact that $\rho_5 = 0$ since $R \geq A_1 > 0$. Next, in order to characterize the power allocation in **OA$_1$**, where [A2] (implying $A_1 = A_2$) and $c_{12}^2 < c_{21}^2$ apply, we partition the set of optimal Lagrange multipliers.

Partitioning the Set of the Optimal Lagrange Multipliers: First, note that ρ_1 and ρ_2 do not satisfy $\rho_1 > 0$ and $\rho_2 > 0$, as this implies $p_1 = q_1 = 0$, which violates (A12). Similarly, ρ_3 and ρ_4 do

not satisfy $\rho_3 > 0$ and $\rho_4 > 0$, as this violates (A13). Therefore, we partition the set of $(\rho_1, \rho_2) \succeq \mathbf{0}$ into the disjoint subsets, $\mathcal{B}_1 := \{(\rho_1, \rho_2) : \rho_1 > 0, \rho_2 = 0\}$, $\mathcal{B}_2 := \{(\rho_1, \rho_2) : \rho_1 = 0, \rho_2 > 0\}$, and $\mathcal{B}_3 := \{(\rho_1, \rho_2) : \rho_1 = 0, \rho_2 = 0\}$. Similarly, we partition the set $(\rho_3, \rho_4) \succeq \mathbf{0}$ into three disjoint subsets, $\mathcal{C}_k, k = 1, 2, 3$. Therefore, any tuple $(\rho_1, \rho_2, \rho_3, \rho_4)$ must be in one of the nine resulting cases, $\mathcal{B}_k \cap \mathcal{C}_l, k, l = 1, 2, 3$. For conciseness, we denote the tuples $(\rho_1, \rho_2, \rho_3, \rho_4)$ and $(\lambda_1, \lambda_2, \lambda_3)$ by the vectors $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$, respectively.

We similarly partition the set of $\boldsymbol{\lambda} \succeq \mathbf{0}$ into 5 subsets: $\mathcal{D}_1 := \{\boldsymbol{\lambda} : \lambda_1 > 0, \lambda_2 = 0, \lambda_3 \geq 0\}$, $\mathcal{D}_2 := \{\boldsymbol{\lambda} : \lambda_1 = 0, \lambda_2 > 0, \lambda_3 \geq 0\}$, $\mathcal{D}_3 := \{\boldsymbol{\lambda} : \lambda_1 = 0, \lambda_2 = 0, \lambda_3 > 0\}$, $\mathcal{D}_4 := \{\boldsymbol{\lambda} : \lambda_1 > 0, \lambda_2 > 0, \lambda_3 = 0\}$, and $\mathcal{D}_5 := \{\boldsymbol{\lambda} : \lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0\}$. Note that when $\boldsymbol{\lambda} \in \mathcal{D}_1$, it satisfies either (a) $\lambda_1 > 0, \lambda_2 = 0, \lambda_3 > 0$, or (b) $\lambda_1 > 0, \lambda_2 = 0, \lambda_3 = 0$. Specifically, in (a) (A14) and (A16) are tight, and thus $R = \Sigma_1 = \Sigma < \Sigma_2$, and in (b) only (A14) is tight, and thus $R = \Sigma_1 < \min\{\Sigma_2, \Sigma\}$. Thus, in $\mathcal{D}_1$, we have $\Sigma_1 \leq \Sigma, \Sigma_1 < \Sigma_2$. Similarly, we note the following: in $\mathcal{D}_2$, we have $\Sigma_2 \leq \Sigma, \Sigma_2 < \Sigma_1$; in $\mathcal{D}_3$, we have $\Sigma < \min\{\Sigma_1, \Sigma_2\}$; in $\mathcal{D}_4$, we have $\Sigma_2 = \Sigma_1 < \Sigma$, and in $\mathcal{D}_5$, we have $\Sigma_2 = \Sigma_1 = \Sigma$.

Now, we characterize "rules" (R_m) that determines whether any $\boldsymbol{\rho} \in \mathcal{B}_k \cap \mathcal{C}_l$ and $\boldsymbol{\lambda} \in \mathcal{D}_j$ jointly satisfy the KKT conditions (i.e., they are compatible) or not (i.e., incompatible).

- R_1: For any $k \in \{1, 3\}, l \in \{1, 2, 3\}$, if $\boldsymbol{\rho} \in \mathcal{B}_k \cap \mathcal{C}_l$ then $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 1, 3$.

- R_2: For any $k \in \{1, 2, 3\}, l \in \{1, 3\}$, if $\boldsymbol{\rho} \in \mathcal{B}_k \cap \mathcal{C}_l$ then $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 2, 3$.

- R_3: If $\boldsymbol{\rho} \in \mathcal{B}_1 \cap \mathcal{C}_2$ then $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 2, 4, 5$.

- R_4: If $\boldsymbol{\rho} \in \mathcal{B}_2 \cap \mathcal{C}_1$ then $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 1, 4, 5$.

- R_5: If $\boldsymbol{\rho} \in \mathcal{B}_2 \cap \mathcal{C}_3$ then $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 1, 4, 5$.

- R_6: If $\boldsymbol{\rho} \in \mathcal{B}_3 \cap \mathcal{C}_2$ then $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 2, 4, 5$.

- R_7: If $\boldsymbol{\rho} \in \mathcal{B}_2 \cap \mathcal{C}_2$ then $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 1, 2, 3, 5$.

We provide below an outline of how $\{R_m\}_{m=1}^{7}$ follow. For brevity, we indicate the compatibility (incompatibility) between $\mathcal{B}_k \cap \mathcal{C}_l$ and $\mathcal{D}_j$ by $\checkmark (\times)$ in Table A1, along with the conditions [Cm], $m = 1, \ldots, 4$, ($R_m, m = 1, \ldots, 7$) that make them compatible (incompatible). Note that [Cm], $m = 1, \ldots, 4$, are given in (A25)–(A28) in the next two pages.

Table A1. Compatibility of $\mathcal{B}_k \cap \mathcal{C}_l$ and $\mathcal{D}_j$.

$\mathcal{B}_k \cap \mathcal{C}_l$	$\mathcal{D}_1$	Cond.	$\mathcal{D}_2$	Cond.	$\mathcal{D}_3$	Cond.	$\mathcal{D}_4$	Cond.	$\mathcal{D}_5$	Cond.
$\mathcal{B}_1 \cap \mathcal{C}_1$	$\times$	R_1	$\times$	R_2	$\times$	R_1	$\times$	$c_{12}^2 < c_{21}^2$	$\times$	$c_{12}^2 < c_{21}^2$
$\mathcal{B}_1 \cap \mathcal{C}_2$	$\times$	R_1	$\times$	R_3	$\times$	R_1	$\times$	R_3	$\times$	R_3
$\mathcal{B}_1 \cap \mathcal{C}_3$	$\times$	R_1	$\times$	R_2	$\times$	R_1	$\checkmark$	[C1]	$\checkmark$	[C3]
$\mathcal{B}_2 \cap \mathcal{C}_1$	$\times$	R_4	$\times$	R_2	$\times$	R_2	$\times$	R_4	$\times$	R_4
$\mathcal{B}_2 \cap \mathcal{C}_2$	$\times$	R_7	$\times$	R_7	$\times$	R_7	$\checkmark$	[C2]	$\times$	R_7
$\mathcal{B}_2 \cap \mathcal{C}_3$	$\times$	R_5	$\times$	R_2	$\times$	R_2	$\times$	R_5	$\times$	R_5
$\mathcal{B}_3 \cap \mathcal{C}_1$	$\times$	R_1	$\times$	R_2	$\times$	R_2	$\times$	$c_{12}^2 < c_{21}^2$	$\times$	$c_{12}^2 < c_{21}^2$
$\mathcal{B}_3 \cap \mathcal{C}_2$	$\times$	R_1	$\times$	R_6	$\times$	R_1	$\times$	R_6	$\times$	R_6
$\mathcal{B}_3 \cap \mathcal{C}_3$	$\times$	R_1	$\times$	R_2	$\times$	R_1	$\checkmark$	[C4]	$\checkmark$	[C3]

R_1 follows from a specific relation between ρ_2 and λ_2 as follows. Suppose $\boldsymbol{\rho} \in \mathcal{B}_k \cap \mathcal{C}_l$ as in R_1. In $\mathcal{B}_k$, we have $\rho_2 = 0$, and in $\mathcal{D}_j$ we have $\lambda_2 = 0$ for $j, k \in \{1, 3\}$. Now, $\rho_2 = \lambda_2 = 0$ implies $\mu_1 = 0$ from (A9), resulting in $\rho_1 = -\alpha \kappa d_1^2 / (2(1 + d_1^2 p_1)) < 0$ from (A8), which contradicts $\rho_1 \geq 0$ in $\mathcal{B}_k$. Thus, $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 1, 3$. R_2 follows from a similar relation between ρ_4 and λ_1.

Now, consider R_3 where $\boldsymbol{\rho} \in \mathcal{B}_1 \cap \mathcal{C}_2$ satisfy $\rho_1 > 0, \rho_2 = 0, \rho_3 = 0, \rho_4 > 0$. This results in $p_1 = 0, q_1 = P/\bar{\beta}, p_2 = P/\beta, q_2 = 0$, following (A20)–(A23) and (A12) and (A13). This further implies $\Sigma_1 < \Sigma_2$, which contradicts the implications of $\boldsymbol{\lambda} \in \mathcal{D}_j, j = 2, 4, 5$. Therefore, $\boldsymbol{\lambda} \notin \mathcal{D}_j, j = 2, 4, 5$. R_4 follows similarly where the roles of ρ_1, ρ_3, and Σ_1 are exchanged with that of ρ_2, ρ_4, and Σ_2, respectively.

In R_5, $\rho \in \mathcal{B}_2 \cap \mathcal{C}_3$ satisfy $\rho_1 = 0, \rho_2 > 0, \rho_3 = 0, \rho_4 = 0$, which results in $p_1 = P/\beta, q_1 = 0, p_2 > 0$, $q_2 > 0$, from (A20)–(A23) and (A12) and (A13). Since $q_1 = 0$ and $q_2 > 0$, we have $\Sigma_2 < \Sigma_1$ that contradicts the implications of $\lambda \in \mathcal{D}_j, j = 1, 4, 5$ as in R_4 above. The assertion in R_6 follows similarly by exchanging the roles of ρ_1, ρ_2, and Σ_1 with ρ_3, ρ_4, and Σ_2, respectively.

In R_7, $\rho \in \mathcal{B}_2 \cap \mathcal{C}_2$ satisfy $\rho_1 = 0, \rho_2 > 0, \rho_3 = 0, \rho_4 > 0$, which results in $p_1 = P/\beta, q_1 = 0, p_2 = P/\beta, q_2 = 0$, from (A20)–(A23) and (A12) and (A13). This implies $\Sigma_1 = \Sigma_2 < \Sigma$, which is consistent with the implication of $\lambda \in \mathcal{D}_4$ only.

Power Allocation in OA$_1$: We associate the five compatible subsets in Table A1 (denoted by ✓) to the four sets $\mathcal{S}_{(.)}$ in Table A1 according to the resulting power allocations: channel parameters and P are in (a) $\mathcal{S}_C$ if $\rho \in \mathcal{B}_1 \cap \mathcal{C}_3, \lambda \in \mathcal{D}_4$; (b) $\mathcal{S}_D$ if $\rho \in \mathcal{B}_2 \cap \mathcal{C}_2, \lambda \in \mathcal{D}_4$; (c) $\mathcal{S}_{\text{sat}}$ if $\rho \in \mathcal{B}_3 \cap \mathcal{C}_3$ or $\mathcal{B}_1 \cap \mathcal{C}_3$, $\lambda \in \mathcal{D}_5$; and (d) $\mathcal{S}_{CD}$ if $\rho \in \mathcal{B}_3 \cap \mathcal{C}_3, \lambda \in \mathcal{D}_4$. Next, we characterize the conditions, $[Cm], m = 1, \ldots, 4$, in terms of the channel parameters, which give the conditions of sets $\mathcal{S}_{(.)}$.

- $\mathcal{S}_C$: Since $\rho \in \mathcal{B}_1 \cap \mathcal{C}_3$, they satisfy $\rho_1 > 0, \rho_2 = 0, \rho_3 = 0, \rho_4 = 0$, which implies $p_1 = 0$, $q_1 = P/\bar{\beta}$ from (A12), (A20) and (A21), and $p_2 > 0, q_2 > 0$ from (A22) and (A23). In addition, $\lambda \in \mathcal{D}_4$ implies $\Sigma_1 = \Sigma_2$ from (A17) and (A18), which, from the expressions of Σ_1 and Σ_2 in (24) and (25) gives $c_{21}^2 q_2 = c_{12}^2 P/\bar{\beta}$. Thus, we have $q_2 = Pc_{12}^2/(c_{21}^2 \bar{\beta})$, and therefore, $p_2 = P(1 - c_{12}^2/c_{21}^2)/\beta$ from (A13). Note that $P > 0$ and $c_{12}^2 < c_{21}^2$ are sufficient for $(q_1, p_2, q_2) \succ \mathbf{0}$. Moreover, $\lambda_3 = 0$ implies $\Sigma_2 < \Sigma$, resulting in $P < \bar{\beta}(\gamma - 1)/c_{12}^2$, i.e., $P < P_4^*$ where γ, defined in (38), is $\gamma > 1$ due to assumption [A1]. Next, from $\rho_3 = 0, \rho_4 = 0$ and (A10) and (A11), we have $\lambda_1 = (1 + Pc_{12}^2/\bar{\beta})/(P(c_{21}^2 - c_{12}^2)/\beta + c_{12}^2/d_2^2)$. In addition, from $\rho_2 = 0$ and (A8), the condition for $\rho_1 > 0$ is $\lambda_2 > d_1^2(1 + Pc_{12}^2/\bar{\beta})/c_{12}^2$. Since $\lambda_3 = 0$, from (A7) we have $\lambda_2 + \lambda_1 = 1$, which subsequently gives $d_1^2(1 + Pc_{12}^2/\bar{\beta})/c_{12}^2 + (1 + Pc_{12}^2/\bar{\beta})/(P(c_{21}^2 - c_{12}^2)/\beta + c_{12}^2/d_2^2) < 1$, i.e., $g_2(P) < 1$ as in Table 1. Thus, the condition of $\mathcal{S}_C$ is

$$[C1]: \quad P < P_4^*, \quad g_2(P) < 1. \tag{A25}$$

- $\mathcal{S}_D$: Since $\rho \in \mathcal{B}_2 \cap \mathcal{C}_2$, they satisfy $\rho_1 = 0, \rho_2 > 0, \rho_3 = 0, \rho_4 > 0$, and imply $p_1 = P/\beta$, $q_1 = 0, p_2 = P/\beta, q_2 = 0$, following (A20)–(A23) and (A12) and (A13). In addition, $\lambda \in \mathcal{D}_4$ implies $\Sigma_1 = \Sigma_2 < \Sigma$ from (A17)–(A19), for which assumption [A1] is sufficient. Next, using $\rho_1 = \rho_3 = 0$ and (A8)–(A11) the sufficient conditions for $\rho_2 > 0$ and $\rho_4 > 0$ are found to be $\lambda_2 < d_1^2/(c_{12}^2(1 + d_1^2 P/\beta))$ and $\lambda_1 < d_2^2/(c_{21}^2(1 + d_2^2 P/\beta))$ respectively. Since $\lambda_3 = 0$, and thus $\lambda_1 + \lambda_2 = 1$ from (A7), the bounds on λ_1 and λ_2 are combined, which gives $d_1^2/(c_{12}^2(1 + d_1^2 P/\beta)) + d_2^2/(c_{21}^2(1 + d_2^2 P/\beta)) > 1$, i.e., $g_1(P) > 1$ as in Table 1. Finally, $P > 0$ is sufficient for $(p_1, p_2) \succ \mathbf{0}$. Thus, the condition of $\mathcal{S}_D$ is

$$[C2]: \quad g_1(P) > 1. \tag{A26}$$

- $\mathcal{S}_{\text{sat}}$: Since $\rho \in \mathcal{B}_3 \cap \mathcal{C}_3$, they satisfy $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0$, and imply $(p_1, q_1, p_2, q_2) \succ \mathbf{0}$, following (A20)–(A23). In addition, $\lambda \in \mathcal{D}_5$ implies $\Sigma_1 = \Sigma$ and $\Sigma_2 = \Sigma$ from (A17)–(A19), which gives $q_1 = \frac{\gamma - 1}{c_{12}^2}$ and $q_2 = \frac{\gamma - 1}{c_{21}^2}$, respectively, where $\gamma > 1$. Thus, we have $p_1 = \frac{P}{\beta} - \frac{\bar{\beta}(\gamma - 1)}{\beta c_{12}^2}$ and $p_2 = \frac{P}{\beta} - \frac{\bar{\beta}(\gamma - 1)}{\beta c_{21}^2}$ from (A12) and (A13). Note that the condition for $p_1 > 0$ is $P > \bar{\beta}(\gamma - 1)/c_{12}^2$, i.e., $P > P_4^*$, and it is also sufficient for $p_2 > 0$ due to $c_{21}^2 > c_{12}^2$.

  Now, from $\rho = \mathbf{0}$ and (A8)–(A11), and using the expressions of $p_l, q_l, l = 1, 2$, above we find $\lambda_1 = \gamma \beta/(Pc_{21}^2 + \beta c_{21}^2/d_2^2 - \bar{\beta}(\gamma - 1))$, and $\lambda_2 = \gamma \beta/(Pc_{12}^2 + \beta c_{12}^2/d_1^2 - \bar{\beta}(\gamma - 1))$. We also note that $P > P_4^*$ is sufficient for $\lambda_l > 0, l = 1, 2$. Next, to ensure $\lambda_3 > 0$, λ_1 and λ_2 must satisfy $\lambda_1 + \lambda_2 < 1$, which gives the condition $g_3(P) < 1$ as in $\mathcal{S}_{\text{sat}}$ in Table A1. In addition, note that $\gamma > 1$ is sufficient for $(q_1, q_2) \succ \mathbf{0}$.

Finally, the case with $\boldsymbol{\rho} \in \mathcal{B}_1 \cap \mathcal{C}_3$ only differ from that with $\mathcal{B}_3 \cap \mathcal{C}_3$ in that now $\rho_1 > 0$, and thus $p_1 = 0$. We note that $P = P_4^*$ is sufficient for $p_1 = 0$. We also note that the other condition, which follows from the conditions on λ_1 and λ_2 as in $\mathcal{B}_3 \cap \mathcal{C}_3$, is expressed by evaluating $g_3(P) < 1$ above at $P = P_4^*$. Therefore, the conditions of the two cases are combined as

$$[\text{C3}]: \quad P \geq P_4^*, \quad g_3(P) < 1. \tag{A27}$$

- $\mathcal{S}_{CD}$: Since $\boldsymbol{\rho} \in \mathcal{B}_3 \cap \mathcal{C}_3$, they satisfy $\boldsymbol{\rho} = \mathbf{0}$, and imply $(p_1, q_1, p_2, q_2) \succ \mathbf{0}$, following (A20)–(A23). In addition, $\boldsymbol{\lambda} \in \mathcal{D}_4$ imply $\Sigma_1 = \Sigma_2$. Next, from $\boldsymbol{\rho} = \mathbf{0}$, (A8)–(A11), (A12) and (A13), we find $\frac{\alpha \kappa}{2\mu_1}(\beta + \bar{\beta}\lambda_2) = P + \frac{\bar{\beta}}{c_{12}^2} + \frac{\beta}{d_1^2}$, and $\frac{\alpha \kappa}{2\mu_2}(\beta + \bar{\beta}\lambda_1) = P + \frac{\bar{\beta}}{c_{21}^2} + \frac{\beta}{d_2^2}$. Also, since $\Sigma_1 = \Sigma_2$, we have from (24) and (25), $\lambda_2 \mu_2 c_{12}^2 = \lambda_1 \mu_1 c_{21}^2$. In addition, since $\lambda_3 = 0$ we have $\lambda_1 = 1 - \lambda_2$. Combining these conditions, we get a quadratic equation of λ_2, $\hat{A}\lambda_2^2 - \hat{B}\lambda_2 + \hat{C} = 0$, where $\hat{A} := \bar{\beta}(E_1 - E_2)$, $\hat{B} := E_1 - E_2 + 2\beta E_2$, and $\hat{C} := \beta E_2$, with $E_1 := Pc_{12}^2 + \bar{\beta} + \beta \frac{c_{12}^2}{d_1^2}$ and $E_2 := Pc_{21}^2 + \bar{\beta} + \beta \frac{c_{21}^2}{d_2^2}$ as defined in (37). One of its roots, $\lambda_2^{(1)} := (\hat{B} + \sqrt{\hat{B}^2 - 4\hat{A}\hat{C}})/2\hat{A}$, is infeasible as it violates (A7) and the nonnegativity of λ_2, respectively, when $\hat{A} > 0$ and $\hat{A} < 0$. Therefore, the valid solution is $\lambda_2 = \lambda_2^{(2)} := (\hat{B} - \sqrt{\hat{B}^2 - 4\hat{A}\hat{C}})/2\hat{A}$. Next, from (A9) and substituting $\lambda_2^{(2)}$ in $\frac{\alpha \kappa}{2\mu_1}(\beta + \bar{\beta}\lambda_2) = P + \frac{\bar{\beta}}{c_{12}^2} + \frac{\beta}{d_1^2}$, and with some algebraic simplification, we have $q_1 = F(P)/c_{12}^2$ where $F(P) := \left(\frac{E_1 + E_2 - \sqrt{(E_1 - E_2)^2 + 4\beta^2 E_1 E_2}}{2\bar{\beta}(1+\beta)} - 1 \right)$, and E_1 and E_2 defined in (37). Finally, from the mutual exclusiveness of the sets $\mathcal{S}_{(.)}$, the condition of $\mathcal{S}_{CD}$ is given by

$$[\text{C4}]: \text{complement of } ([\text{C1}] \cup [\text{C2}] \cup [\text{C3}]). \tag{A28}$$

References

1. LTE-A 4G Solution, Stockholm, Sweden. Available online: https://www.ramonmillan.com/documentos/bibliografia/4G_Ericsson.pdf (accessed on 13 September 2017).
2. Pi, Z.; Khan, F. An introduction to millimeter-wave mobile broadband systems. *IEEE Commun. Mag.* **2011**, *49*, 101–107.
3. Rappaport, T.S.; Sun, S.; Mayzus, R.; Zhao, H.; Azar, Y.; Wang, K.; Wong, G.N.; Schulz, J.K.; Samimi, M.; Gutierrez, F. Millimeter wave mobile communications for 5G cellular: It Will Work! *IEEE Access* **2013**, *1*, 335–349.
4. Akdeniz, M.R.; Liu, Y.; Samimi, M.K.; Sun, S.; Rangan, S.; Rappaport, T.S.; Erkip, E. Millimeter Wave Channel Modeling and Cellular Capacity Evaluation. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1164–1179.
5. Qingling, Z.; Li, J. Rain Attenuation in Millimeter Wave Ranges. In Proceedings of the IEEE 7th International Symposium on Antennas, Propagation and EM Theory, Guilin, China, 26–29 October 2006; pp. 1–4.
6. Singh, S.; Mudumbai, R.; Madhow, U. Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control. *IEEE/ACM Trans. Netw.* **2011**, *19*, 1513–1527.
7. Qiao, J.; Shen, X.S.; Mark, J.W.; Shen, Q.; He, Y.; Lei, L. Enabling device-to-device communications in millimeter-wave 5G cellular networks. *IEEE Commun. Mag.* **2015**, *53*, 209–215.
8. Collonge, S.; Zaharia, G.; Zein, G.E. Influence of the human activity on wide-band characteristics of the 60 GHz indoor radio channel. *IEEE Trans. Wirel. Commun.* **2004**, *3*, 2396–2406.
9. Zheng, K.; Zhao, L.; Mei, J.; Dohler, M.; Xiang, W.; Peng, Y. 10 Gb/s Hetsnets with Millimeter-wave communications: Access and networking - challenges and protocols. *IEEE Commun. Mag.* **2015**, *53*, 222–231.
10. 5G Radio Access: Requirements, Concept and Technologies. Available online: https://gsacom.com/paper/docomo-5g-radio-access-requirements-concept-and-technologies/ (accessed on 13 September 2017).
11. Mehrpouyan, H.; Matthaiou, M.; Wang, R.; Karagiannidis, G.K.; Hua, Y. Hybrid millimeter-wave systems: A novel paradigm for hetnets. *IEEE Commun. Mag.* **2015**, *53*, 216–221.
12. Majhi, S.; Mitran, P. On the capacity of a class of dual-band interference channels. In Proceedings of the IEEE International Symposium on Information Theory, Barcelona, Spain, 10–15 July 2016; pp. 2759–2763.

13. Feng, W.; Li, Y.; Jin, D.; Su, L.; Chen, S. Millimetre-wave backhaul for 5G networks: Challenges and solutions. *Sensors* **2016**, *16*, 892.

14. Morteza, H.; Koksal, C.E.; Shroff, N.B. Dual Sub-6 GHz—Millimeter Wave Beamforming and Communications to Achieve Low Latency and High Energy Efficiency in 5G Systems. Available online: https://arxiv.org/pdf/1701.06241.pdf (accessed on 13 September 2017).

15. Niu, Y.; Li, Y.; Jin, D.; Su, L.; Vasilakos, A.V. A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges. *Wirel. Netw.* **2015**, *21*, 2657–2676.

16. Intel Announces World's First Global 5G Modem. Available online: https://newsroom.intel.com/newsroom/wp-content/uploads/sites/11/2017/01/5G-modem-fact-sheet.pdf (accessed on 13 September 2017).

17. Ali, A.; González-Prelcic, N.; Heath, R.W. Estimating millimeter wave channels using out-of-band measurements. In Proceedings of the 2016 Information Theory and Applications Workshop, La Jolla, CA, USA, 31 January–5 February 2016; pp. 1–6.

18. Ali, A.; Prelcic, N.; Robert, W.; Heath, J. Millimeter Wave Beam-Selection Using Out-of-Band Spatial Information. Available online: https://arxiv.org/pdf/1702.08574.pdf (accessed on 13 September 2017).

19. Elshaer, H.; Kulkarni, M.N.; Boccardi, F.; Andrews, J.G.; Dohler, M. Downlink and uplink cell association with traditional macrocells and millimeter wave small cells. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 6244–6258.

20. Semiari, O.; Saad, W.; Bennis, M. Joint millimeter wave and microwave resources allocation in cellular networks with dual-mode base stations. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 4802–4816.

21. Rebato, M.; Boccardi, F.; Mezzavilla, M.; Rangan, S.; Zorzi, M. Hybrid spectrum sharing in mmwave cellular networks. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 155–168.

22. Sato, H. The capacity of the Gaussian interference channel under strong interference. *IEEE Trans. Inf. Theory* **1981**, *27*, 786–788.

23. Carleial, A. A case where interference does not reduce capacity. *IEEE Trans. Inf. Theory* **1975**, *21*, 569–570.

24. Kramer, G. Outer bounds on the capacity of Gaussian interference channels. *IEEE Trans. Inf. Theory* **2004**, *50*, 581–586.

25. Chung, S.T.; Cioffi, J. The capacity region of frequency-selective Gaussian interference channels under strong interference. *IEEE Trans. Commun.* **2007**, *55*, 1812–1821.

26. Choi, S.W.; Chung, S.Y. On the separability of parallel Gaussian interference channels. In Proceedings of the IEEE International Symposium on Information Theory, Seoul, South Korea, 28 June–3 July 2009; pp. 2592–2596.

27. Sankar, L.; Shang, X.; Erkip, E.; Poor, H.V. Ergodic fading interference channels: Sum-capacity and separability. *IEEE Trans. Inf. Theory* **2011**, *57*, 2605–2626.

28. Cadambe, V.; Jafar, S. Parallel Gaussian interference channels are not always separable. *IEEE Trans. Inf. Theory* **2009**, *55*, 3983–3990.

29. Shang, X.; Chen, B.; Kramer, G.; Poor, H. Noisy-interference sum-rate capacity of parallel Gaussian interference channels. *IEEE Trans. Inf. Theory* **2011**, *57*, 210–226.

30. Yu, W.; Ginis, G.; Cioffi, J.M. Distributed multiuser power control for digital subscriber lines. *IEEE J. Sel. Areas Commun.* **2002**, *20*, 1105–1115.

31. Chen, H.; Ma, Y.; Lin, Z.; Li, Y.; Vucetic, B. Distributed power control in interference channels with QoS constraints and RF energy harvesting: A game-theoretic approach. *IEEE Trans. Veh. Technol.* **2016**, *65*, 10063–10069.

32. Gamal, A.E.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2012.

33. Diggavi, S.N.; Cover, T.M. The worst additive noise under a covariance constraint. *IEEE Trans. Inf. Theory* **2006**, *47*, 3072–3081.

34. Sankar, L.; Liang, Y.; Mandayam, N.B.; Poor, H.V. Fading multiple access relay channels: Achievable rates and opportunistic scheduling. *IEEE Trans. Inf. Theory* **2011**, *57*, 1911–1931.

35. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.