

Article

On the Downlink Capacity of Cell-Free Massive MIMO with Constrained Fronthaul Capacity

Peng Zhang ^{1,*} and Frans M. J. Willems ²

¹ IMEC the Netherlands, 5656AE Eindhoven, The Netherlands

² Department of Electrical Engineering, Technical University of Eindhoven, 5612AZ Eindhoven, The Netherlands; f.m.j.willems@tue.nl

* Correspondence: peng.zhang@imec.nl; Tel.: +31-404-020-409

Received: 14 February 2020; Accepted: 4 April 2020; Published: 7 April 2020



Abstract: We investigate the downlink of a cell-free massive multiple-in multiple-out system in which all access points (APs) are connected in a linear-topology fronthaul with constrained capacity and send a common message to a single receiver. By modeling the system as an extension of the multiple-access channel with partially cooperating encoders, we derive the channel capacity of the two-AP setting and then extend the results to arbitrary N -AP scenarios. By developing a cooperating mode concept, we investigate the optimal cooperation among the encoders (APs) when we limit the total fronthaul capacity, and the total transmit power is constrained as well. It is demonstrated that achieving capacity requires a water-pouring distribution of the total available fronthaul capacity over the fronthaul links. Our study reveals that a linear growth of total fronthaul capacity results in a logarithmic growth of the beamforming capacity. Moreover, even if the number of APs would be unlimited, only a finite number of them need to be activated. We found an expression for this number.

Keywords: channel capacity; distributed beamforming; cell-free MIMO; constrained fronthaul

1. Introduction

Recently, cell-free massive multiple-input multiple-output (mMIMO) has been considered as a key technology for beyond-5G networks. In such user-centric transmission systems, a large number of distributed access points (APs) are connected to one central processing unit (CPU) via fronthaul links and phase coherently cooperate to cover a wide area for a small number of users in the same time-frequency resource using time-division operation. Compared to cell-based collocated mMIMO solutions, such technology improves energy-spectral efficiency and enhances immunity to shadow fading without extra signal processing burdens. We refer to [1–3] and the references therein for a general overview of current developments of cell-free mMIMO.

Effectively utilizing fronthaul resources is of critical importance for deploying a scalable cell-free mMIMO system. Considering the downlink for instance, simple distributed conjugate beamforming is optimal, as shown in [1]. However, it can already be seen that a large amount of information exchange over fronthaul links is required since all the APs need to know the message that is to be transmitted. A star-topology fronthaul where each APs are individually connected to a CPU was originally modeled and has been widely studied, see, e.g., [4–6] and the references therein. Currently, a serial fronthaul connecting APs in a linear topology is considered for achieving a cost-efficient architecture, both in deployment and maintenance [3]. A novel and promising technique relying on a linear topology is the radio stripe system, where multiple APs are embedded in a cable/strip, see [3,7] in detail. Such radio stripes can be easily and invisibly deployed indoor or outdoor in existing constructions to enable numerous new applications [8].

The focus of prior work in cell-free mMIMO study was on developing wireless signaling techniques. In this paper, we study from an information-theoretic perspective the downlink of a cell-free mMIMO system shown by Figure 1, where single-antenna APs are connected in a linear topology with constrained fronthaul capacities to communicate to one single-antenna terminal receiver (Rx).

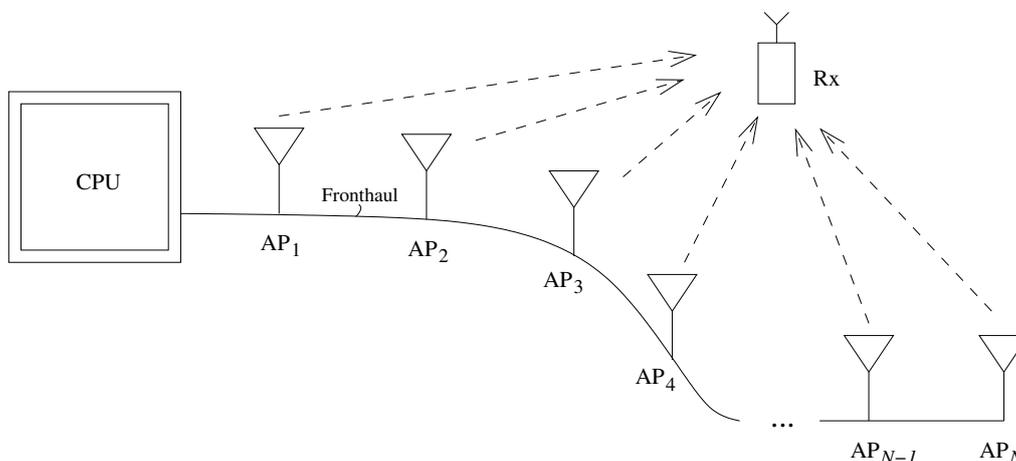


Figure 1. Cell-free transmit beamforming where access points (APs) are connected via serial fronthaul.

The considered multiple-in single-out (MISO) setup forms a distributed massive beamforming system and can be formulated as a multiple-access-channel (MAC) with limited fronthaul capacity, which is defined as the maximal amount of information that can be reliably sent per MAC channel use [9]. By investigating the channel capacity of such a MAC, we reveal essential relations between the three the most fundamental resources of the system, i.e., the total available number of APs (N), the total transmit power (P), and total available fronthaul capacity (C_B). Specifically, in the current cell-free mMIMO literature, the only configuration of APs that is considered is where full cooperation (full beamforming) is realized and where the same information is shared at all involved APs. Therefore, for a real-valued Gaussian MISO channel with N APs and unity channel gains, the maximum downlink rate is given by the channel capacity

$$C_{\text{full}} := \frac{1}{2} \log_2(1 + N \cdot S\mathcal{N}\mathcal{R}) \quad \text{bits/channel use} \tag{1}$$

where $S\mathcal{N}\mathcal{R}$ is the received signal-to-noise ratio (SNR) if only one AP is active with all available transmit power assigned to it. It requires $C_B^{\text{full}} := (N - 1)C_{\text{full}}$ fronthaul capacity among N APs.

In this work, we focus on the case where the available fronthaul capacity is not large enough to support full cooperation of the APs. We were motivated to investigate the achievable downlink rates given that fronthaul resources for communication between the APs is constrained. We call this setting partial beamforming, since $C_B < C_B^{\text{full}}$. We could derive the channel capacity and the optimal cooperation strategies among APs for given total available P and N .

1.1. Related Work

We can model the studied system as a special extension of the multiple-access channel (MAC) with partially cooperating encoders studied by Willems [9]. In particular, we can generalize the system setup in [9] to a network of encoders by considering only one source but employing an arbitrary number of encoders, namely APs, via unidirectional conferences. Since fronthaul links can be treated as separate channels that are orthogonal to the beamforming MAC, our setup might also be viewed as an extension of a special case of the orthogonal-component relay channel due to El Gamal and Zahedi [10], which is generalized to relay networks by Ghabeli and Aref in [11]. In addition, if only two APs are considered, our study is also strongly related to the multiple access diamond channel

as studied in [12,13]. Moreover, the two APs setup looks very similar to the semi-deterministic relay channels [14]. Furthermore, it is also worth to note that in our system, all APs cooperatively send one message to a receiver at a same time. In this sense, our channel setting is “noncausal”, which is related to the relay-with-delay channel studied in [15] in general.

1.2. Contributions and Organization

By investigating the MAC with limited fronthaul capacity in the discrete channel case and in the Gaussian channel cases, the main findings of our research work include

- The channel capacity is found for an arbitrary number of APs for both discrete channel and the Gaussian channel with constrained transit power, where the total fronthaul capacity and the total number of APs are limited.
- When numerous APs are engaged, a linear growth of total fronthaul capacity results in a logarithmic growing of the channel (beamforming) capacity.
- A concept of cooperating modes is developed to demonstrate the optimal cooperation among APs to achieve capacity based on superposition coding.
- When the channel capacity is only limited by the fronthaul capacity, the number of required APs is quasi-linear to the available fronthaul capacity even if the number of APs would be unlimited.
- A new and sharp lower bound of the Lambert-W function is derived for computing the number of required APs given by the total fronthaul constraint.

In the rest of this paper, the system model is first presented in Section 2. In Section 3, we start with investigating a two-APs setting consisting of one fronthaul link. This setting serves as a baseline system where the cooperating mode concept is developed. In Section 4, the study is extended to the case where an arbitrary number of APs is engaged and the behavior and exact solution of the channel capacity is derived. In Section 5, the number of required APs is derived to leverage limited fronthaul resources if the number of available APs is unlimited. Finally, the conclusion and final remarks can be found in Section 6. Detailed proofs and derivations of the presented results are collected in the Appendix A. Partial material in this paper was presented in [16].

2. Problem Setup

2.1. Notation

Although all the paper, capital letters, e.g., X , denote random variables, and their realizations are denoted by small letters, e.g., x . The probability mass or density function according to X is denoted by $p_X(x)$ or simply $p(x)$. The expectation of X is denoted by $E[X]$. The entropy of X is denoted by $H(X)$ and the differential entropy is denoted by $h(X)$. The mutual information between X and Y is denoted by $I(X; Y)$. The consecutive integer range from i to j with $i \leq j$ is denoted by $[i : j]$. In addition, a set of elements x_m with index m in range of i to j is denoted as $\{x_m\}_{m=i}^j$.

2.2. System Model

The investigated system is modeled as Figure 2, where we denote the CPU as the source, the APs as encoders, while for the destination, the receiver is denoted as the decoder. As plotted, one-directional fronthaul links connect N adjacent encoders that simultaneously send a uniformly distributed message $W \in [1 : M]$ to a decoder (receiver). We focus on the study of the fronthaul resource usage among all encoders. The discrete memoryless MAC denoted by $(\mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_N, p(y|x_1, x_2, \dots, x_N), \mathcal{Y}, \{C_{m,m+1}\}_{m=1}^{N-1})$ consists of input alphabets $\{\mathcal{X}_m\}_{m=1}^N$, output alphabet \mathcal{Y} , a transition probability distribution $p(y|x_1, x_2, \dots, x_N)$, and a set of fronthaul capacity constraints $\{C_{m,m+1}\}_{m=1}^{N-1}$ between N encoders.

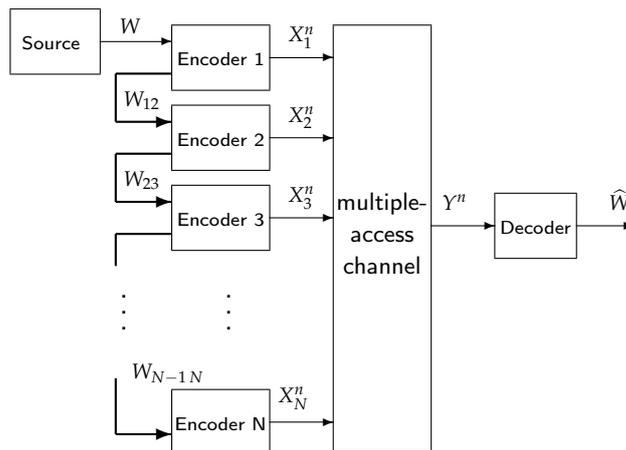


Figure 2. N encoders that cooperate in sending a message W to a decoder with limited fronthaul communication.

Before the beginning of each n channel uses, (partial) information about the generated message W is first shared among N encoders. Let $W_{m,m+1} \in [1 : M_{m,m+1}]$ for $m \in [1 : N - 1]$ be the message sent over the fronthaul link between encoder m to encoder $m + 1$. Then, the encoders map the messages W and $\{W_{m,m+1}\}_{m=1}^{N-1}$ into codewords $\{x_m^n\}_{m=1}^N$ as follows

$$\begin{aligned} e_1(W) &\rightarrow (X_1^n, W_{12}), \\ e_m(W_{m-1,m}) &\rightarrow (X_m^n, W_{m,m+1}), \\ e_N(W_{N-1,N}) &\rightarrow X_N^n, \end{aligned}$$

where $\{e_m(\cdot)\}_{m=1}^N$ are the corresponding encoding functions. Meanwhile, the generated fronthaul messages should satisfy

$$\frac{1}{n} \log_2 M_{m,m+1} \leq C_{m,m+1}. \tag{2}$$

As presented, the corresponding fronthaul link capacity $C_{m,m+1} \geq 0$ is defined as the maximal amount of information that can be reliably sent per channel use of the MAC channel over the link from encoder m to encoder $m + 1$.

At the decoder, a deterministic decoding function $d : \mathcal{Y}^n \rightarrow [1 : M]$ is applied to obtain the message-estimate \hat{W} based on the channel output y^n . We define the average probability of error at the decoder as

$$P_e^{(n)} \triangleq \Pr(\hat{W} \neq W). \tag{3}$$

Now we say that a rate R is achievable with given fronthaul capacities $\{C_{m,m+1}\}_{m=1}^{N-1}$ if there exists N encoders and a corresponding decoder, such that

$$\begin{aligned} \log_2 M &\geq n(R - \delta), \\ \log_2 M_{m,m+1} &\leq nC_{m,m+1}, \\ P_e^{(n)} &\leq \delta, \end{aligned} \tag{4}$$

for all $\delta > 0$ and large enough n . The channel capacity C (of MAC) as a function of the fronthaul capacities is defined as the supremum of all achievable rates given by all the fronthaul constraints. Eventually, we will be interested only in a constraint on the sum of the fronthaul capacities C_B that is defined as

$$C_B \triangleq \sum_{m=1}^{N-1} C_{m,m+1}. \tag{5}$$

To interpret the capacity results for the partial beamforming, we focus on MACs with additive white Gaussian noise. At the output of the Gaussian MAC, the decoder receives

$$Y_i = \sum_{m=1}^N X_{mi} + Z_i, \tag{6}$$

at time i , where X_{mi} is the transmitted symbol by encoder m and Z_i is modeled as independent and identically distributed (i.i.d.) Gaussian noise at the decoder for all $i \in [1 : n]$. For individual encoder m , $m \in [1 : N]$, the transmit power constraint is

$$\frac{1}{n} \sum_{i=1}^n E[X_{mi}^2] \leq P_m \tag{7}$$

for $P_m \geq 0$. Then, the total transmit power is limited as

$$\sum_{m=1}^N P_m \leq P. \tag{8}$$

Without loss of generality, we assume that $Z_i \sim \mathcal{N}(0, 1)$. Therefore, the transmit SNR can be directly represented by the total constrained transmit power P .

3. Two-Encoder Result

We first investigate the simplest system setting where only two encoders are involved. The MAC is now denoted by $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y}, C_{12})$. The fronthaul message $W_{12} \in [1 : M_{12}]$ must satisfy the constraint

$$\frac{1}{n} \log_2 M_{12} \leq C_{12}, \tag{9}$$

which is same to the total fronthaul capacity C_B in this case. The underlying Gaussian MAC is given by

$$Y_i = X_{1i} + X_{2i} + Z_i. \tag{10}$$

Although this two-encoder setting can be considered as a special case of related work, see discussion later, we provide here the capacity proofs for both discrete and Gaussian MACs. The applied approach carries over to the N -encoder setting that is investigated in Section 2.

In the following, the channel capacity as a function of the fronthaul capacity is first obtained for the discrete memoryless MAC. Then, we derive capacity results for the Gaussian case with total transmit power constraint. Within this study, a so-called *cooperating mode* concept is developed that will be very useful to provide cooperation insights among encoders when more of them are engaged.

3.1. Discrete Channel

First, consider the discrete channel setup.

Theorem 1. For the discrete memoryless channel $p(y|x_1, x_2)$, the channel capacity C as a function of the fronthaul capacity C_{12} is given by

$$C(C_{12}) = \max_{p(x_1, x_2)} \min \{I(X_1, X_2; Y), I(X_1; Y|X_2) + C_{12}\}, \tag{11}$$

where distribution $p(x_1, x_2, y) = p(x_1, x_2)p(y|x_1, x_2)$ is determined by the input distribution $p(x_1, x_2)$.

The detailed proof is provided in Section Appendix A.1, where the converse is based on the Markovities of $W \rightarrow (X_1^n, X_2^n) \rightarrow Y^n$ and $(W, W_{12}) \rightarrow (X_1^n, X_2^n) \rightarrow Y^n$, and the achievability is based on applying superposition coding. For the achievability, the source splits the message W into two parts

(W_1, W_{12}) and delivers the index of W_{12} over the fronthaul link to encoder 2 that maps W_{12} into the inner code while encoder 1 of the source encodes W_1 into an outer code-word which is super-imposed on the inner code-word. Although this coding scheme is simple, the *cooperating mode* concept that is important for studying the multi-encoder setup will be developed based on the superposition scheme as discussed later.

Remark 1. By viewing the two-encoder setting as a special setup of ([9], Figure 1), where only one source and one conference link are deployed, we can have Theorem 1 by letting the common message $U = X_2$ and the conference capacity $C_{21} = 0$ in ([9], Thm.). Note that the achievability in [9] which is based on binning becomes superposition coding.

Remark 2. By viewing the two-encoder setting as a special setup of the multiple access diamond channel where one source connects to two encoders (relays) by using two separate noiseless links, see [12,13], Theorem 1 can also be obtained if letting $C_1 = \infty$, the common message $V = X_2$, and the common message rate $R_0 = C_2$ (or $C_2 = \infty$, $V = X_1$, $R_0 = C_1$) in ([12], Thm. 2). Note that the achievability based on superposition and Marton-coding in [12,13] becomes superposition coding only.

3.2. Gaussian Channel

Now we consider the Gaussian MAC of the two-encoder channel setting given by (10) with total power constraint P , i.e., $P_1 + P_2 \leq P$, where

$$\frac{1}{n} \sum_{i=1}^n E[x_{1i}^2] \leq P_1 \quad \text{and} \quad \frac{1}{n} \sum_{i=2}^n E[x_{2i}^2] \leq P_2. \tag{12}$$

This first leads to the following result.

Theorem 2. The channel capacity $C(C_{12}, P)$ of the two-encoder Gaussian MAC is

$$C(C_{12}, P) = \max_{0 \leq \beta \leq 1} \min \left\{ \frac{1}{2} \log_2(1 + (1 + \beta)P), \frac{1}{2} \log_2(1 + (1 - \beta)P) + C_{12} \right\}. \tag{13}$$

The proof is the adaptation of the discrete channel version given in Section Appendix A.1 by considering the transmit power constraints and Gaussian channel noise.

Proof. (i) *Converse.* First note that without loss of generality (and without violating the power constraints) we may assume that all $E[X_{1i}] = E[X_{2i}] = 0$ for all $i \in [1, n]$. If we define (X_1, X_2, Y) being the random triple with density $p_{X_1, X_2, Y}(x_1, x_2, y) = \frac{1}{N} \sum_{i=1}^N p_{X_{1i}, X_{2i}, Y_i}(x_1, x_2, y)$ then converse in Section Appendix A.1 shows that

$$\begin{aligned} I(X_1^n, X_2^n; Y^n) &\leq nI(X_1, X_2; Y), \\ I(X_1^n; Y^n | X_2^n) &\leq nI(X_1; Y | X_2), \end{aligned} \tag{14}$$

where the random variables X_1, X_2 satisfy $E[X_1] = E[X_2] = 0$ and

$$E[X_1^2] \leq \frac{1}{n} \sum_{i=1}^n E[X_{1i}^2] \leq P_1, \tag{15}$$

$$E[X_2^2] \leq \frac{1}{n} \sum_{i=1}^n E[X_{2i}^2] \leq P_2. \tag{16}$$

First consider the random pair (X_1, X_2) . By applying the Cholesky factorization ([17], Thm. 4.2.7) to the covariance matrix of $[X_2, X_1]^T$, the assignment of

$$\begin{cases} X_1 = \alpha_{21}S_2 + \alpha_1S_1, \\ X_2 = \alpha_{22}S_2, \end{cases} \tag{17}$$

can be obtained, where S_1 and S_2 are uncorrelated with zero means and unit variances.

Next, observe that if we take $\alpha'_{21} = \alpha'_{22} = (\alpha_{21} + \alpha_{22})/2 = \alpha_2$ this choice does not affect $I(X_2; Y) = I(S_2; Y)$ and $I(X_1; Y|X_2) = I(S_1; Y|S_2)$, but minimizes the total transmit power for fixed $\alpha_{21} + \alpha_{22}$, since

$$2\alpha_2^2 = 2\left(\frac{\alpha_{21} + \alpha_{22}}{2}\right)^2 \leq \alpha_{21}^2 + \alpha_{22}^2. \tag{18}$$

Therefore we only need to consider assignment

$$\begin{cases} X_1 = \alpha_2S_2 + \alpha_1S_1, \\ X_2 = \alpha_2S_2. \end{cases} \tag{19}$$

Now we take $\alpha_2^2 = P_2$, $\alpha_2^2 + \alpha_1^2 = P_1$, and $2\alpha_2^2 + \alpha_1^2 = P$. By denoting

$$\beta \triangleq \frac{2\alpha_2^2}{P} \tag{20}$$

in $[0, 1]$, we further have

$$\alpha_1^2 = (1 - \beta)P \quad \text{and} \quad \alpha_2^2 = \beta P/2. \tag{21}$$

Taking the signal assignment (19) and the power assignment (21) gives that

$$\begin{aligned} I(X_1, X_2; Y) &= I(S_1, S_2; Y) = h(\alpha_1S_1 + 2\alpha_2S_2 + Z) - h(Z) \\ &\stackrel{(a)}{\leq} \frac{1}{2} \log_2(1 + (1 + \beta)P), \end{aligned} \tag{22}$$

$$\begin{aligned} I(X_1; Y|X_2) &= I(S_1; Y|S_2) = h(\alpha_1S + Z) - h(Z) \\ &\stackrel{(b)}{\leq} \frac{1}{2} \log_2(1 + (1 - \beta)P), \end{aligned} \tag{23}$$

where (a) and (b) follow by the maximum differential entropy theorem, see ([18], Thm. 8.6.5).

(ii) *Achievability.* Taking the assignment (19) by letting $S_1 \sim \mathcal{N}(0, 1)$ and $S_2 \sim \mathcal{N}(0, 1)$. Using the power assignment (21) directly gives

$$I(X_1, X_2; Y) = I(S_1, S_2, Y) = \frac{1}{2} \log_2(1 + (1 + \beta)P), \tag{24}$$

$$I(X_1; Y|X_2) = I(S_1; Y|S_2) = \frac{1}{2} \log_2(1 + (1 - \beta)P). \tag{25}$$

The rest of the proof follows by first establishing a coding theorem for the discrete memoryless channel with input cost (power constraint). The step from discrete to Gaussian channels is justified by the relation between differential entropy and discrete entropy, see, e.g., ([18], Thm. 9.3.1). \square

Now, by optimizing over β in (13), we can further express C as a function only in total transmit power P and total fronthaul capacity C_B , which is C_{12} for this two-encoder setup.

Corollary 1. The channel capacity $C(C_B, P)$ of the total transmit power constrained two-encoder Gaussian MAC can be expressed as

$$C(C_B, P) = \begin{cases} \frac{1}{2} \log_2(1 + 2P), & \text{if } C_B \geq \frac{1}{2} \log_2(1 + 2P) \\ \frac{1}{2} \log_2(1 + P) + \frac{1}{2} \log_2\left(2 - \frac{2}{2^{2C_B} + 1}\right), & \text{otherwise.} \end{cases} \tag{26}$$

Proof. The two logarithms on the RHS of (13) are monotonically increasing and decreasing in β respectively and equal to each other at $\beta = 0$. Hence, we can set

$$\frac{1}{2} \log_2(1 + (1 + \beta)P) = \frac{1}{2} \log_2(1 + (1 - \beta)P) + C_B \tag{27}$$

to obtain the β that maximizes C for $\forall C_B \in [0, \frac{1}{2} \log_2(1 + 2P)]$ as

$$\beta^* = \frac{(1 + P)(2^{2C_B} - 1)}{P(2^{2C_B} + 1)}. \tag{28}$$

This results in the second capacity expression in (26). Then, if $C_B > \frac{1}{2} \log_2(1 + 2P)$, the second term is always larger than the first term for any β in (13). This corresponds to the situation where C_B is large enough and the transmission over the MAC is the bottleneck of the network. In this case, C remains at its global maximum. □

Note that, for $C_B < \frac{1}{2} \log_2(1 + 2P)$, the first term of the capacity result (26) is the channel capacity with no beamforming and the second term directly represents the partial beamforming gain that is independent of transmit power P and only grows as the fronthaul capacity increases. As revealed, the partial beamforming gain increases with a same rate regardless of the transmit power P .

3.3. Cooperating Modes

Based on assignment (19) that possesses a superposition structure, we can naturally denote two *cooperating modes* as what follows to describe the optimal cooperation between the encoders for the capacity achieving.

- *mode 1*: Sending a private message given by $\alpha_1 S_1$ from encoder 1;
- *mode 2*: Coherently sending a common message given by $\alpha_2 S_2$ from encoder 2 and encoder 1.

According to (20), the parameter β represents the fraction of the total transmit power assigned to *mode 2* while $1 - \beta$ represents the remaining fraction assigned to *mode 1*. Note that β^* given by (28) should be taken for achieving the capacity.

Now consider the cooperation scenarios of the two encoders based on the availability of C_B . If $C_B = 0$, the transmission reduces to the point-to-point communication case. This is represented by having only *mode 1* active and encoder 2 is inactive. If $C_B \geq C_{full}$, full cooperation can be achieved by activating *mode 2* only. For $C_{12} \in (0, C_{full})$, two encoders cooperate to achieve partial beamforming capacity by activating both *cooperating modes*. Figure 3 illustrates the *cooperating modes* activating and deactivating at encoders depending on C_B increasing from 0 to C_{full} .

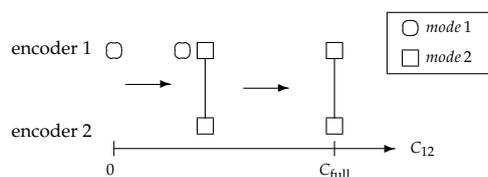


Figure 3. Modes activation for the two-encoder Gaussian setting.

For the two-encoder setting, the modes evolution due to available amount of C_B looks straightforward. Nevertheless, it will be shown that this *cooperating modes* interpretation provides a clear insight of leveraging available encoders for given certain total fronthaul and transmit power constraints, where the optimal cooperation is not trivial as the number of encoders goes largely.

4. N-Encoder Result

Based on the investigations of the two-encoder setting, we extend the study to the system model with arbitrarily N encoders, where $N \geq 2$. The parameter N in principle can be any large integer so that a distributed massive beamforming is obtained. The investigation is focused on the Gaussian MAC under the constraints of the total fronthaul capacity C_B and the total transmit power P , which are defined by (5) and (8), respectively. Before addressing the exact capacity solution for arbitrary N encoders, we first derive capacity bounds of $C(C_B, P)$ to provide a general behavior of channel capacity C in total fronthaul capacity C_B . The obtained result indicates that the growth of C requires an exponential growth of C_B . By using the *compound mode*, the exact capacity solution with the optimal cooperation among encoders are derived. The results show that the distributed beamforming system works most efficiently when it is working in its fronthaul-capacity-limited regime. As a result, we consider the case where encoders are always available to be activated as needed to leverage the entire fronthaul resource.

4.1. Discrete Channel

For simplicity, let the tuple $\underline{X}_l^m \triangleq (X_l, X_{l+1}, \dots, X_m)$ be the collection of ordered transmitted random variables that are generated at encoder l to encoder m with $l \leq m$ for one channel use. In addition, let $\underline{C}_b \triangleq \{C_{j,j+1}\}_{j=1}^{N-1}$ be the collection of the corresponding fronthaul capacities.

Theorem 3. For the discrete memoryless N -encoder setting, channel capacity C of the channel $P(y|x_1, x_2, \dots, x_N)$ as a function of fronthaul capacities \underline{C}_b is

$$C(\underline{C}_b) = \max_{p(x_1, x_2, \dots, x_N)} \min \left\{ I(\underline{X}_1^N; Y), I(\underline{X}_1^m; Y | \underline{X}_{m+1}^N) + C_{m,m+1} \right\}_{m=1}^{N-1}, \quad (29)$$

with $N \geq 2$.

A sketch of the proof is given in Section Appendix A.2. As shown in the achievability, the capacity is achieved by applying an N -layer superposition coding among the encoders, which naturally agrees with the studied linear topology.

4.2. Gaussian Channel under Total fronthaul Constraint

By considering on the total power and separate fronthaul constraints, we first have the following result.

Theorem 4. The the N -encoder Gaussian setting with the total transmit power constraint of P , the channel capacity C as a function of the fronthaul capacities \underline{C}_b is

$$C(\underline{C}_b, P) = \max_{\underline{\beta}} \min \left\{ \frac{1}{2} \log_2 \left(1 + \sum_{l=1}^N l \beta_l P \right), \left\{ \frac{1}{2} \log_2 \left(1 + \sum_{l=1}^m l \beta_l P \right) + C_{m,m+1} \right\}_{m=1}^{N-1} \right\}, \quad (30)$$

where $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_N)^T$ is a probability vector.

Proof. Similar to proof of the two-encoder setting, the generic signal assignment

$$X_m = \sum_{l=m}^N \alpha_{lm} S_l \tag{31}$$

can be used at each encoder for $m \in [1, N]$, where $\{S_l\}_{l=1}^N$ are uncorrelated and have zero mean and unit variance. Again, we can further apply the special signal assignment

$$X_m = \sum_{l=m}^N \alpha_l S_l \tag{32}$$

to minimize the total transmit power without affecting dependency of the different signals $\{S_l\}_{l=1}^N$ at the decoder that determines the beamforming capacity. In this way, the transmit power allocated for signal S_l can be expressed by

$$\beta_l = \frac{l\alpha_l^2}{P} \tag{33}$$

such that $\sum_{l=1}^N \beta_l = 1$.

Thus, for the converse, we can use the assignment (32) to evaluate (29) and the mutual informations on the RHS are bounded as given by (30). Then, for the achievability, by letting $S_l \sim \mathcal{N}(0, 1)$, the result follows. \square

The proof shows that all the transmitted signals at encoders should form a Markov chain $X_N \rightarrow X_{N-1} \rightarrow \dots \rightarrow X_1$. Again, since signal $\alpha_l S_l$ represents the common messages used at first l encoders, we say that *cooperating mode l* is active if the signal S_l is generated and sent and there can be N *cooperating modes* in total for this N -encoder setting.

Now, we can solve the optimization problem

$$\underset{\underline{C}_b}{\text{maximize}} \quad C(\underline{C}_b, P) \tag{34}$$

$$\text{subject to} \quad \sum_{m=1}^{N-1} C_{m,m+1} = C_B, \tag{35}$$

where $C(\underline{C}_b, P)$ is given by (30), to investigate the total power limited capacity C under the constraint of total fronthaul capacity C_B for a given P . To do so, we first prove the following lemma. Note that the full-cooperation capacity is now $C_{\text{full}}(N) = \frac{1}{2} \log_2(1 + NP)$ when N encoders are used. For simplicity, we denote the mutual informations as

$$I_m \triangleq \frac{1}{2} \log_2(1 + \sum_{l=1}^m l\beta_l P) \tag{36}$$

for any $m \in [1 : N]$.

Lemma 1. For the N -encoder setting with any given $C_B \leq (N - 1)C_{\text{full}}(N)$, power distribution $\underline{\beta}$ can only be optimal if equality of all the terms on the RHS of (30) is achieved.

The proof is given in Appendix A.3.

Remark 3. Lemma 1 indicates that asymmetric distribution of C_B over fronthaul link is optimal. This result will be further demonstrated after the capacity result is derived.

Based on the reduced $\underline{\beta}$ set given by Lemma 1, we make the terms on the RHS of (30) equal and have

$$C_{m,m+1} = I_N - I_m \quad \text{with } m \in [1 : N - 1]. \tag{37}$$

Thus, the channel capacity and the required total fronthaul capacity in the power allocation vector $\underline{\beta}$ can now be represented as

$$C(\underline{\beta}, P) = I_N = \frac{1}{2} \log_2 \left(1 + \sum_{l=1}^N l \beta_l P \right), \tag{38}$$

and

$$C_B(\underline{\beta}, P) = (N - 1)I_N - \sum_{m=1}^{N-1} I_m = \frac{N - 1}{2} \log_2 \left(1 + \sum_{l=1}^N l \beta_l P \right) - \sum_{m=1}^{N-1} \frac{1}{2} \log_2 \left(1 + \sum_{l=1}^m l \beta_l P \right) \tag{39}$$

respectively for a fixed P if $C_B \leq (N - 1)C_{\text{full}}(N)$. Based on (38) and (39), we can have the following theorem.

Theorem 5. For the N -encoder Gaussian channel under the total power constraint P and total fronthaul constraint C_B , the channel capacity is given by

$$\begin{aligned} & \underset{\underline{\beta}}{\text{maximize}} && C(\underline{\beta}, P) = I_N \\ & \text{subject to} && (N - 1)I_N - \sum_{m=1}^{N-1} I_m = C_B \end{aligned} \tag{40}$$

$$\sum_{l=1}^N \beta_l = 1 \quad \text{and} \quad 0 \leq \beta_l \leq 1. \tag{41}$$

To evaluate the channel capacity, we only need to maximize the function by introducing a Lagrange multiplier λ as

$$\begin{aligned} g(\underline{\beta}, P, \lambda) &= I_N - \lambda C_B \\ &= (1 - (N - 1)\lambda)I_N + \lambda \sum_{m=1}^{N-1} I_m \\ &= \frac{1 - (N - 1)\lambda}{2} \log_2 \left(1 + \sum_{l=1}^N l \beta_l P \right) + \frac{\lambda}{2} \sum_{m=1}^{N-1} \log_2 \left(1 + \sum_{l=1}^m l \beta_l P \right) \end{aligned} \tag{42}$$

under the constraint that $\underline{\beta}$ is a probability vector to derive the solution of $C(C_B)$ for the general N -encoder case. Note that the parameter λ is the slope of $C(C_B)$. However, before working out the exact solution of this optimization problem, we first derive general bounds of $C(C_B, P)$ to reveal the capacity behavior of the studied distributed beamforming.

4.3. Capacity Behavior Bounds

To obtain a simple but meaningful insight of the relation between C and the constrained C_B and P for an arbitrary N , we propose an upper bound and a lower bound of the channel capacity to draw the following conclusion.

Property 1. For any fixed total transmit power P and number of encoders N , a linear growth of total fronthaul capacity C_B results in a logarithmical growing of the channel capacity as C can be bounded as

$$C \leq \frac{1}{2} \log_2(1 + P) + \frac{1}{2} \log_2(1 + 2 \ln 2 \cdot C_B), \tag{43}$$

and

$$C > \frac{1}{2} \log_2(1 + (2 \ln 2 \cdot PC_B)^{2/3}). \tag{44}$$

for $C_B \leq (N - 1)C_{\text{full}}$.

Proof. (1) *Upper bound.* By considering $L \in [1 : N]$ as a random variable with distribution $\underline{\beta}$, the capacity (38) can be expressed as

$$C = \frac{1}{2} \log_2(1 + \mu_L P), \tag{45}$$

where $\mu_L \triangleq E[L]$. By applying Jensen’s inequality, the corresponding fronthaul capacity C_B in (39) can be lower bounded as

$$\begin{aligned} C_B &\geq (N - 1)C - \frac{N - 1}{2} \log_2\left(\frac{\sum_{m=1}^{N-1}(1 + \sum_{l=1}^m l \beta_l P)}{N - 1}\right) \\ &= (N - 1)C - \frac{N - 1}{2} \log_2\left(1 + \frac{\sum_{l=1}^{N-1}(N - l) l \beta_l P}{N - 1}\right) \\ &\stackrel{(a)}{\geq} (N - 1)C - \frac{N - 1}{2} \log_2\left(1 + \frac{\mu_L(N - \mu_L)P}{N - 1}\right) \\ &= \frac{N - 1}{2} \log_2\left(\frac{1 + \mu_L P}{1 + \frac{\mu_L(N - \mu_L)P}{N - 1}}\right) \\ &\stackrel{(b)}{\geq} \frac{(\mu_L - 1)\mu_L P}{2 \ln 2 \cdot (1 + \mu_L P)} \end{aligned} \tag{46}$$

where (a) follows $\mu_L^2 \leq E[L^2]$ and (b) follows $\ln x \geq 1 - \frac{1}{x}$ for $x > 0$. Since $\mu_L \geq 1$, we can have $C_B \geq \frac{(\mu_L - 1)\mu_L P}{2 \ln 2 \cdot (\mu_L + \mu_L P)} = \frac{(\mu_L - 1)P}{2 \ln 2 \cdot (1 + P)}$ that results in $\mu_L \leq 1 + 2 \ln 2 \cdot (\frac{1+P}{P})C_B$ and thus (43).

2) *Lower bound.* Consider time-sharing of the rates given by only using one cooperating mode. Hence, the channel capacity should be larger than or equal to an achievable rate R as

$$C \geq R \triangleq \frac{1}{2} \log_2(1 + kP), \tag{47}$$

where k is the number of the activated encoders corresponding to the required total fronthaul capacity

$$C_B = (k - 1)R \tag{48}$$

that achieves R . By applying $\ln x \leq \frac{x-1}{\sqrt{x}}$ for $x \geq 1$, see, e.g., ([19], Section 3.6.15), we have

$$R \leq \frac{1}{2 \ln 2} \cdot \frac{kP}{\sqrt{1 + kP}} \tag{49}$$

that gives an upper bound of C_B as

$$\begin{aligned} C_B &\leq \frac{(k - 1)kP}{2 \ln 2 \sqrt{1 + kP}} \\ &< \frac{(k - 1)kP}{2 \ln 2 \sqrt{kP}} \\ &< \frac{k\sqrt{kP}}{2 \ln 2}. \end{aligned} \tag{50}$$

Therefore, we can have $k^3 > (2 \ln 2 \cdot C_B)^2 / P$ that directly gives (44). \square

The upper bound (43) and lower bound (44) thus indicate the logarithmical behavior of C in C_B . Figure 4 gives an illustration of these two bounds for 10-encoder Gaussian setting where total transmit power is set at $P = 21$.

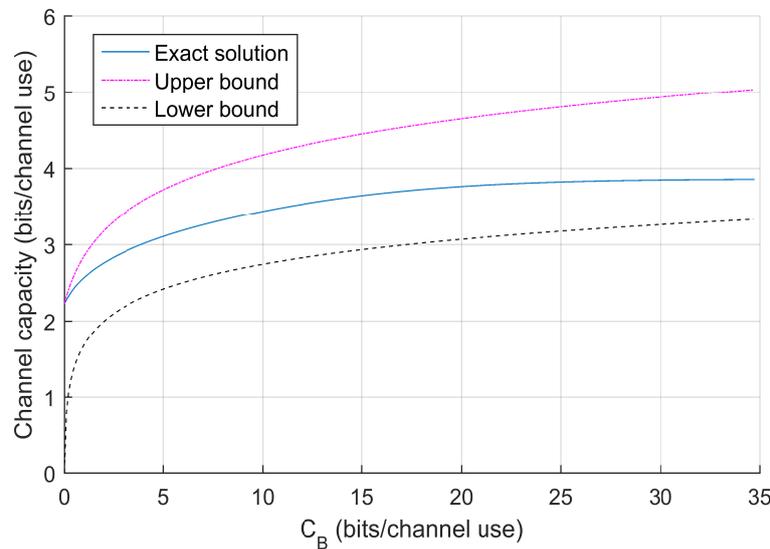


Figure 4. Upper and lower bounds on $C(C_B)$ at $P = 21$ for 10-encoder Gaussian setting.

The exact capacity solution derived shortly is plotted as well as a comparison, showing that the bounds describes the capacity behavior.

4.4. Compound Mode and Exact Solution

In what follows, we perform evaluation of the channel capacity given in Theorem 5 by defining a *compound mode* $\langle j, k \rangle$ as a collection of all consecutive *cooperating modes* between and including *modes* $j, k \in [1 : N]$ with $j \leq k$. A *compound mode* $\langle j, k \rangle$ is referred to as *active* if all $\{\beta_l\}_{l=j}^k$ are nonzero and the other elements in $\underline{\beta}$ are zeros. Note that using a single mode is a special case of compound mode. By denoting

$$b(j) \triangleq \frac{1}{j(2 + (j + 1)P)}, \tag{51}$$

we have the following results.

Corollary 2. For an N -encoder Gaussian setting where $C_B \leq (N - 1)C_{\text{full}}$ with a fixed transmit power P , if there is a compound mode $\langle j, k \rangle$ such that $UB \geq LB$, where

$$UB \triangleq \begin{cases} \frac{1}{2^{(k-1)}} & \text{if } j = 1 \\ \min\{\frac{1}{2^{(k-1)}}, b(j-1)\} & \text{if } j > 1, \end{cases} \tag{52}$$

$$LB \triangleq \begin{cases} \max\{\frac{1}{2^k}, b(j)\} & \text{if } k < N \\ b(j) & \text{if } k = N, \end{cases} \tag{53}$$

the channel capacity corresponding to the slope $\lambda \in [LB, UB]$ is *achieved and only achieved* by using that compound mode which gives

$$C(\lambda) = \frac{1}{2} \log_2 \left(\frac{k(1 - (k - 1)\lambda)(1 + jP)}{j(1 - (j - 1)\lambda)} \right), \tag{54}$$

and

$$C_B(\lambda) = \frac{1}{2} \log_2 \left(\frac{(2\lambda)^{j-k} (1+jP)^{j-1}}{\prod_{l=j}^{k-1} \frac{l(l+1)}{2}} \cdot \frac{(k(1-(k-1)\lambda))^{k-1}}{(j(1-(j-1)\lambda))^{j-1}} \right). \tag{55}$$

The proof is given in Appendix A.4.

Remark 4. The proof in Appendix A.4 shows that if compound mode $\langle j, k \rangle$ achieves the capacity and $k \geq j + 2$, the modes in $[j + 1 : k - 1]$ should be assigned with same power amount as the optimal setting.

Remark 5. By rewriting (55) and comparing it to (54), we can also represent C_B in terms of C as

$$C_B(\lambda) = (j - 1)C(\lambda) + \frac{1}{2} \log_2 \left(\frac{[\lambda^{-1}k(1 - (k - 1)\lambda)]^{k-j}}{\prod_{l=j}^{k-1} l(l + 1)} \right). \tag{56}$$

for a certain slope $\lambda \in [LB, UB]$.

4.5. Modes Selection for Capacity Achieving

The results in Corollary 2 state how the capacity is achieved and expressed over a certain λ range. To further elaborate how to exactly use cooperating modes from no cooperation to full cooperation, a procedure efficiently activating modes is developed based on applying the following result, where an identification of valid compound modes that are the ones resulting in capacity is provided in terms of using a power penalty.

Corollary 3. For any fixed P , a compound mode $\langle j, k \rangle$ achieves the capacity if and only if

$$\frac{2(k - j - 1)}{j(j + 1)} < P \leq \frac{2(k - j + 1)}{j(j - 1) \lceil 1 - \frac{k}{N} \rceil}, \tag{57}$$

where $\lceil \cdot \rceil$ is the ceiling function. If (57) is satisfied, compound modes $\langle j', k' \rangle$ with $j' > j$ and $k' < k$ do not achieve the capacity.

The proof is given in Appendix A.5. The power condition (57) indicates that a compound mode needs certain transmit power to be supported to be optimal. On the other hand, some compound modes can never be optimal if the transmit power is too large.

Now, note that C is monotonically increasing in C_B owing to nonnegative slope λ and monotonically decreasing in λ according to (54) when j and k are fixed. It shows that to achieve the capacity, compound modes should be activated in a way such that the corresponding slope range varies from large to small as C_B increases. Therefore, based on the results in Corollary 3, an algorithm is resulted for computing C and C_B over $C_B \in [0, (N - 1)C_{full}]$ by activating valid compound modes sequentially.

Algorithm 1 represents the cooperating strategy among encoders. It reveals that cooperating modes should be activated one-by-one to form new compound modes with the increase of C_B . At certain point of the growth of C_B , the first mode dies, i.e., deactivated owing to the limited P or N . With the further increasing of C_B , lower modes die in a one-by-one fashion till the full cooperation is obtained.

Figure 5 plots the results for $P = 1$ and $P = 21$ by applying Algorithm 1, $C(C_B)$ over the full range of $C_B \in [0, (N - 1)C_{full}]$. Different number of the available encoders are considered. In the plot, each color segment represents the corresponding activated compound mode. In addition, the pentagram markers label the points where a lower mode has to be deactivated (dead) because of the power penalty (57) or because all N encoders are all used up, namely operations in line 13 and line 9

of the algorithm, respectively. It is shown that for low SNR, i.e., $P = 1$, the modes die fast due to the small power. On the other hand, for large SNR, i.e., $P = 21$, the larger available encoder number the slower the modes die such that higher capacity can be achieved (consider curves of using 2-encoder, 3-encoder, 4-encoder, and 5-encoder).

Algorithm 1 Compute C and C_B from no cooperation to full cooperation

Initialize: $j \leftarrow 1$ and $k \leftarrow 1$
Ensure: $1 \leq j \leq k \leq N$
1: **while** $j < N$ **do**
2: **if** Power condition (57) is satisfied **then**
3: $\lambda \leftarrow [LB, UB]$
4: Compute $C(\lambda)$ and $C_B(\lambda)$ by (54) and (55)
5: **if** $k < N$ **then**
6: $k \leftarrow k + 1$
7: **else**
8: $k \leftarrow N$
9: $j \leftarrow j + 1$
10: **end if**
11: **else**
12: $k \leftarrow k - 1$
13: $j \leftarrow j + 1$
14: **end if**
15: **end while**

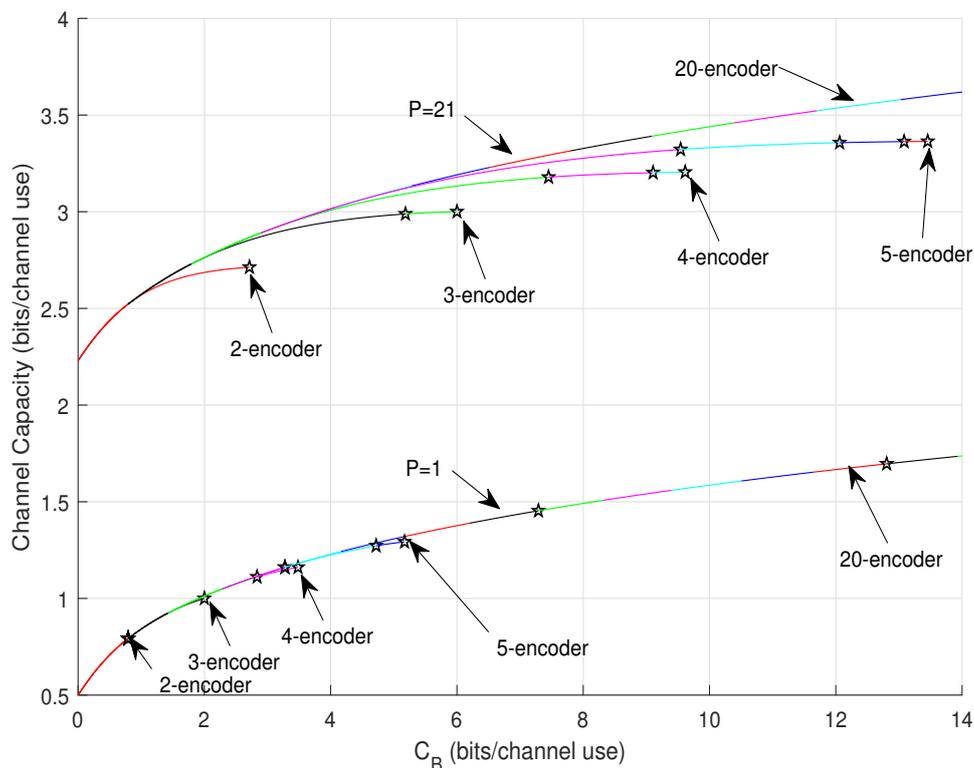


Figure 5. $C(C_B)$ at $P = 1$ and $P = 21$ for different N -encoder settings.

Property 2. As probably the most natural strategy, the way of applying modes in the lower bound proof of Proposition 1, i.e., time-sharing full cooperation of small number of encoders, is not optimal in general. However, it is sub-optimal when SNR is small as the compound modes that achieve capacity reduce to single modes.

To visualize each mode evolution from no beamforming to total beamforming, we can illustrate the power allocation for each cooperating mode as C_B increases. By incorporating calculations of β (given in Appendix A.4) and (37) into Algorithm 1, Figures 6 and 7 show the modes' power evolution of the 10-encoder setting for $P = 5$ and $P = 21$, respectively. It is shown that the first mode dies faster when P is relatively small. They also interestingly show that once C_B is large enough to approach the total beamforming, the last mode dominates as other modes all vanish.

Moreover, we can also elaborate the cooperating of encoders in terms of showing optimal distribution of C_B over fronthaul links. Figure 8 illustrates the distribution of the 10-encoder setting where the bolder curves are for $P = 5$ while the lighter curves are for $P = 21$. In each case, the fronthaul capacity curves for $C_{m m+1}$ for $m = 1$ to $m = 9$ are located from left to right in the plot. This result further demonstrates the asymmetric water-pouring assignment of C_B over fronthaul links, see Remark 3.

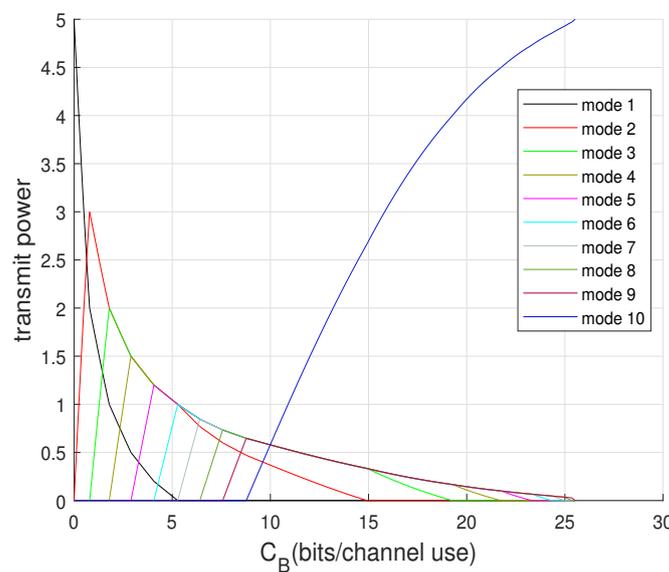


Figure 6. Optimal power allocation for modes distribution for 10-encoder Gaussian setting at $P = 5$.

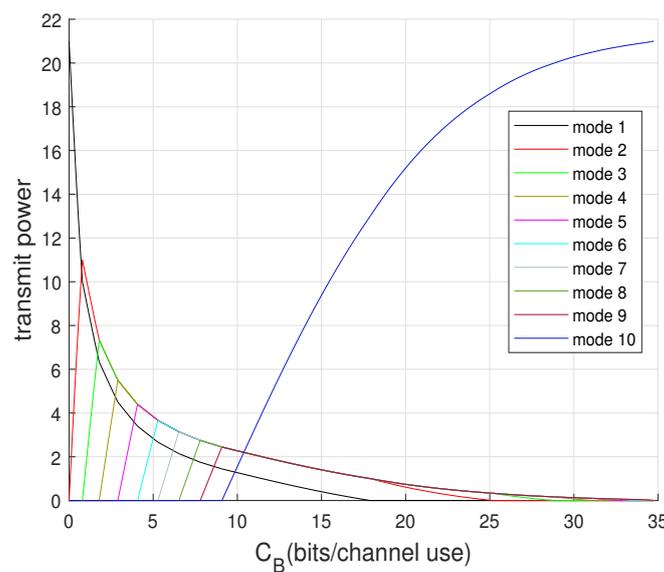


Figure 7. Optimal power allocation for modes distribution for 10-encoder Gaussian setting at $P = 21$.

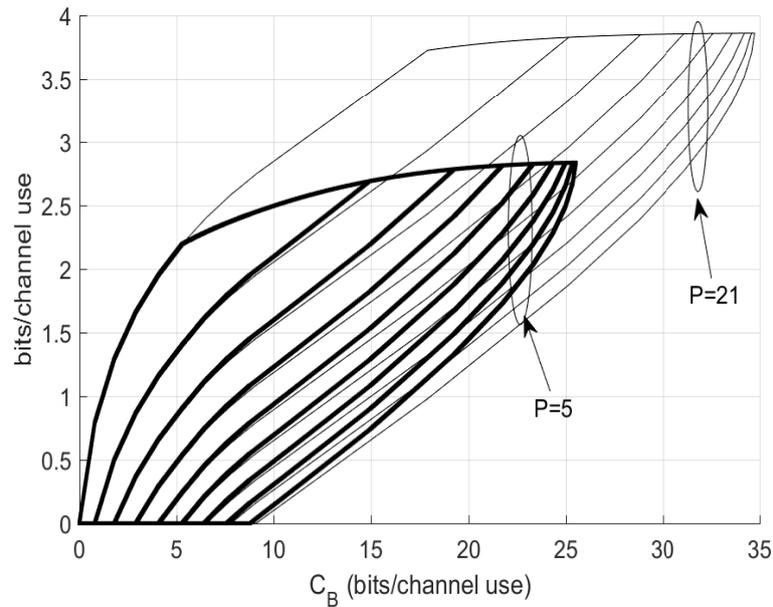


Figure 8. Optimal distribution of C_B at $P = 5$ and $P = 21$ for 10-encoder Gaussian setting.

4.6. $\langle 1, k \rangle$ Mode and Capacity Regimes

Consider the case where $\langle 1, k \rangle$ mode achieves the capacity for $k \leq N$. In this case, the growth rate of $C(C_B)$ is independent of P and N , see the expression of C_B in (56) with $j = 1$. Therefore, we call that the system works in a fronthaul-capacity-limited regime when a $\langle 1, k \rangle$ mode is used. The reason why we are interested in the fronthaul-capacity-limited regime is that $C(C_B)$ achieves the fast growth rate regardless of P and N . As a further increase of C_B , the first mode dies due to either limited P or limited N . We then call the system works in a power-limited regime or encoder-limited regime, respectively. When the system is in either power-limited regime or encoder-limited regime, $C(C_B)$ growth is slowed down compared to when the system works in the fronthaul-capacity-limited regime. This is due to the discontinuities of the slope λ , see the derived optimal upper and lower bounds of λ . The following result shows how to determine which regime the system works in for given C_B, P , and N .

Property 3. For given P and N , if

$$P \leq N - 2, \tag{58}$$

the capacity growth is limited by P and the system works in a fronthaul-capacity-limited regime if

$$C_B \lesssim \frac{1}{2} \log_2 \left(\frac{((1 + \tilde{P})(2 + \tilde{P}))^{\tilde{P}}}{\tilde{P}!(\tilde{P} + 1)!} \right), \tag{59}$$

where $\tilde{P} \triangleq \lceil P \rceil$. Otherwise it works in a power-limited regime.

On the other hand, if $P > N - 2$, the capacity growth is limited by N and the system works in a fronthaul-capacity-limited regime if

$$C_B \leq \frac{1}{2} \log_2 \left(\frac{(N(3 + 2P - N))^{N-1}}{(N - 1)!N!} \right). \tag{60}$$

Otherwise it works in a encoder-number-limited regime.

Proof. Modes dying hampers the growth of C in C_B . Consider that the first mode of the compound mode $\langle 1, k \rangle$ dies because of constrained P not N . In this case, P must satisfy (58) which is given by the lower bound of (57). Consequently, at the moment after the first mode dies, i.e., compound mode

$\langle 2, k \rangle$ is active, we have $j = 2$ and $k \approx \lceil P \rceil + 1$ given by taking the upper bound of (57). This j, k setting results in $\lambda = \frac{1}{2+2P}$ so that (59) is obtained by evaluating (55).

Similarly, considering the compound mode $\langle 1, N \rangle$ can be supported by P , the first mode dies because that no new encoders can be used. At the moment of first mode dying, i.e., compound mode $\langle 1, N \rangle$ is still active, we thus have $j = 1$ and $k = N$, which also result in $\lambda = \frac{1}{2+2P}$. Hence, (60) is resulted. □

Figure 9 plots $C(C_B)$ of 10-encoder setting for $P = 5$ and $P = 21$, respectively, where the regime separations are indicated at the first mode dies for both powers. It is illustrated that in the fronthaul-capacity-limited regime, C has the highest growth rate no matter what its initial value is (point-to-point communication). In the next subsection, we focus on a system working at the fronthaul-capacity-limited regime.

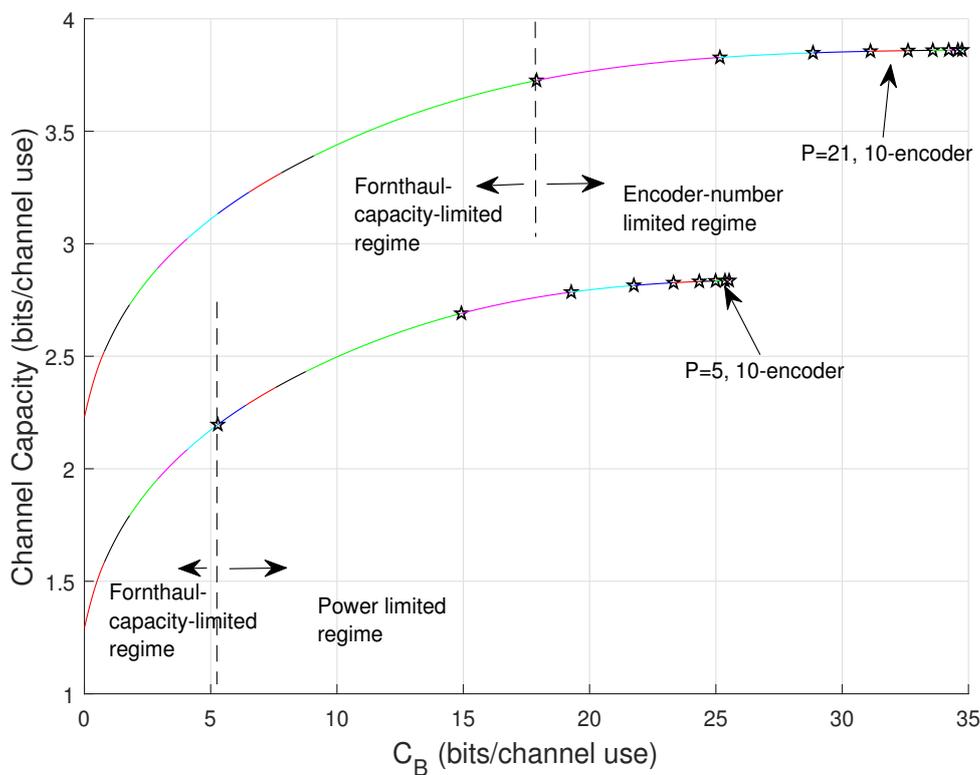


Figure 9. $C(C_B)$ at $P = 5$ and $P = 21$ for 10-encoder Gaussian setting.

5. Infinitely Many Encoders

Consider designing a system in practice when C_B and P are critical resources while available encoders could be many, for instance, the radio stripe system. Based on the previous study, we should always try to let the system work in its fronthaul-capacity-limited regime where the fronthaul capacity is maximally utilized. Hence, we are motivated to determine the number of encoders that are required to be activated for a given C_B by considering infinitely many of them are available when the system is purely fronthaul constrained.

To directly solve k , the highest active mode that is the number of required encoders, from (55) or (56) is not trivial. To achieve an accurate approximate result, we first need the following lemma, of which the proof follows the outline in [20] and is given in Appendix A.6.

Lemma 2. The non-principle branch of Lambert W function $W_{-1}(\cdot)$ defined in the interval $[-e^{-1}, 0)$, see [21], can be bounded as follows

$$W_{-1}(-e^{-(x+1)}) \geq -x - \sqrt[3]{2x} - 1 \tag{61}$$

for $x \geq 0$.

Remark 6. Figure A3 shows that for $\forall x \geq 0.5$, the lower bound (61) is much tighter than $W_{-1}(-e^{-(x+1)}) \geq -x - \sqrt{2x} - 1$ given in [20], which is the tightest bound of $W_{-1}(\cdot)$ reported in the literature so far, to our best knowledge.

Property 4. When the system works in the fronthaul limited regime, the number of required encoders \tilde{k} is quasi-linear to the available fronthaul capacity as

$$\tilde{k} = \lfloor C_B + (2C_B + 1)^{1/3} + 1.5 \rfloor, \tag{62}$$

where C_B in nats per channel use and $\lfloor \cdot \rfloor$ is the floor function.

Proof. As the system works in the fronthaul-limited regime, compound mode $\langle 1, k \rangle$ exists. Thus, according to (56),

$$C_B = \frac{1}{2} \ln \left(\frac{[\lambda^{-1}k(1 - (k-1)\lambda)]^{k-1}}{\prod_{l=1}^{k-1} l(l+1)} \right) \tag{63}$$

in nats. To upper bound k , we lower bound C_B by taking $\lambda = \frac{1}{2(k-1)}$, see the bound (52), which gives

$$\begin{aligned} C_B &\geq \frac{1}{2} \ln \left(\frac{k^{k-2}(k-1)^{k-1}}{((k-1)!)^2} \right) \\ &\stackrel{(a)}{\geq} \frac{1}{2} \ln \left(\frac{e^{2(k-1)}k^{k-2}}{e^2(k-1)^k} \right) \\ &= k - 2 - \ln k + \frac{k}{2} \ln \left(\frac{k}{k-1} \right) \\ &\stackrel{(b)}{\geq} k - \ln k - 1.5 \end{aligned} \tag{64}$$

where (a) follows by applying $(k-1)! \leq e(k-1)^{k-1}e^{-k+1}$ derived based on ([22], 6.1.38), and (b) follows by taking the fact that $(\frac{k}{k-1})^k$ is monotonically decreasing in k and goes to e as $k \rightarrow \infty$. Therefore, we have

$$-\exp(C_B + 1.5) > -k \exp(-k). \tag{65}$$

Now, solving k and applying (61) give

$$\begin{aligned} k &\leq -W_{-1}(-\exp(-(C_B + 1.5))) \\ &\leq C_B + \sqrt[3]{2C_B + 1} + 1.5. \end{aligned} \tag{66}$$

Finally, since the number of encoders is a integer, (62) is resulted. \square

In Figure 10, the bound of k given by (66) and the actual number encoders required to be activated are plotted as a function of C_B . It is revealed that the derived result is accurate enough.

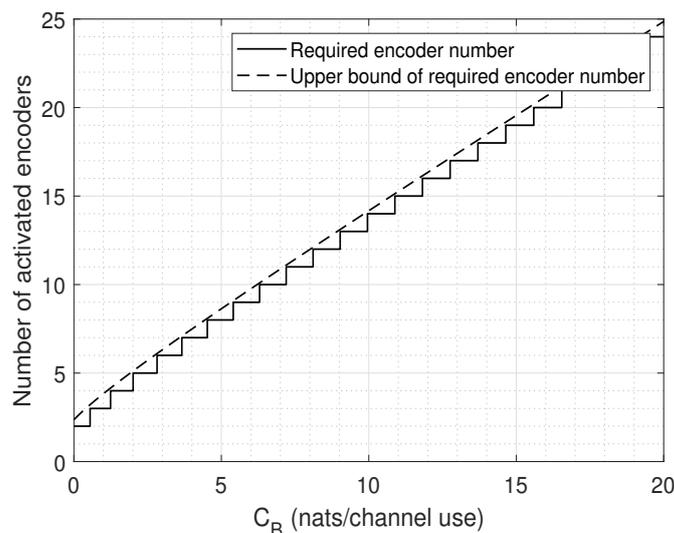


Figure 10. Number of required encoders in relation to C_B for fronthaul resource maximum usage.

6. Concluding Remarks

In this paper, the downlink of a cell-free mMIMO in which multiple APs connected in a linear fronthaul topology serve as a single receiver was studied to reveal relations between the three fundamental network resources, namely the total fronthaul capacity C_B , the total transmit number P , and the number of available APs N . Specifically, we focused on partial distributed beamforming where the total available fronthaul capacity is not enough to support full cooperation between all APs, i.e., beamforming. By formulating the problem as a MAC channel with multiple encoders linked in a feed-and-forward setting, we derived the channel capacity as a function of the total fronthaul capacity for both discrete and Gaussian channels. The derivation was started by considering two encoders and then we extended the analysis multiple encoders. It was demonstrated that capacity is achieved by multi-layer superposition coding from which the concept of cooperating mode was developed for the Gaussian channel. This cooperating mode technique leads to optimal cooperation among encoders. Bounds on the capacity for N -encoder setting demonstrated that this channel capacity grows logarithmically in C_B for a fixed P . The exact capacity solution shows that the capacity is achieved if and only if by certain compound modes are used. An algorithm was derived for computing which compound modes should be activated as a function of C_B , which grows from zero to the value obtaining full beamforming. We demonstrated that C_B should be water-poured over the fronthaul links to obtain optimality. Finally, by considering the case where infinitely many encoders are available, we showed that the number of required encoders is quasi-linear to the available total fronthaul capacity when the system is purely constrained by fronthaul resources.

Future directions include extending the results to channels with links which do not have unit gain as is the case here, and considering multiple receivers. Another interesting direction would be the equivalent uplink case.

Author Contributions: All authors conceived the problem and solution. P.Z. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Appendix A. Proofs and Derivations

Appendix A.1. Proof of Theorem 1

(i) *Converse.* Consider Fano’s inequality $H(W|Y^n) \leq 1 + P_e^{(n)} \log_2 M = F$ for $F \triangleq 1 + P_e^{(n)} \log_2 M$. Since $W \rightarrow (X_1^n, X_2^n) \rightarrow Y^n$ forms a Markov chain, we have that

$$\begin{aligned} \log_2 M = H(W) &= I(W; Y^n) + H(W|Y^n) \leq I(W, X_1^n, X_2^n; Y^n) + F \\ &\leq I(X_1^n, X_2^n; Y^n) + F \leq \sum_{i=1}^n I(X_{1i}, X_{2i}; Y_i) + F \\ &\leq nI(X_1, X_2; Y|Q) + F \leq nI(X_1, X_2; Y) + F. \end{aligned} \tag{A1}$$

Moreover, from the Markovity of $(W, W_{12}) \rightarrow (X_1^n, X_2^n) \rightarrow Y^n$, we obtain that

$$\begin{aligned} \log_2 M = H(W) &= I(W; Y^n) + H(W|Y^n) \leq I(W, W_{12}; Y^n) + F \\ &= I(W_{12}; Y^n) + I(W; Y^n|W_{12}) + F \leq \log_2 M_{12} + I(X_1^n; Y^n|X_2^n) + F \\ &\leq nC_{12} + I(X_1^n; Y^n|X_2^n) + F \leq nC_{12} + \sum_{i=1}^n I(X_{1i}; Y_i|X_{2i}) + F \\ &\leq nC_{12} + nI(X_1; Y|X_2, Q) + F \leq nC_{12} + nI(X_1; Y|X_2) + F. \end{aligned} \tag{A2}$$

In the above derivations, the random variable Q is uniformly distributed on $[1 : n]$ and $\Pr\{X_1 = x_1, X_2 = x_2\} = \frac{1}{N} \sum_{i=1}^N \Pr\{X_{1i} = x_1, X_{2i} = x_2\}$ for $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2$. If now both $\delta \rightarrow 0$ and $n \rightarrow \infty$ we obtain that

$$R \leq \min\{I(X_1, X_2; Y), I(X_1; Y|X_2) + C_{12}\}, \tag{A3}$$

for some distribution $p(x_1, x_2, y) = p(x_1, x_2)p(y|x_1, x_2)$, for all achievable rate R . This concludes the converse for the discrete memoryless two-encoder case.

(ii) *Achievability.* We prove that if the message rate $(1/n) \log_2 M < C_B(C_{12})$ for a given fronthaul capacity C_{12} , the message error probability $P_e^{(n)}$ approaches zero if the codeword length n increases. Our coding method is based on superposition.

Codebook Generation: First fix a joint probability distribution $\{p(x_1, x_2), x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\}$. This distribution determines $p_{X_2}(x_2) = \sum_{x_1 \in \mathcal{X}_1} p(x_1, x_2)$ and $p_{X_1|X_2}(x_1|x_2) = p(x_1, x_2)/p_{X_2}(x_2)$ for x_2 with $p_{X_2}(x_2) > 0$. Now generate at random M_2 i.i.d. sequences $x_2^n \in \mathcal{X}_2^n$ of length n , each drawn according to $\Pr\{X_2^n = x_2^n\} = \prod_{i=1}^n p_{X_2}(x_{2i})$ and index these sequences as $x_2^n(w_2)$ as an inner code, where $w_2 \in [1 : M_2]$. Then, for each such $x_2^n(w_2)$, generate M_1 sequences $x_1^n(w_1, w_2)$ drawn according to $\Pr\{X_1^n = x_1^n|X_2^n = x_2^n(w_2)\} = \prod_{i=1}^n p_{X_1|X_2}(x_{1i}|x_{2i}(w_2))$ in an i.i.d. fashion as an outer code, where $w_1 \in [1 : M_1]$. The resulting codebook is revealed to both encoders and to the decoder.

Encoding: Split the message W that is uniformly distributed on $[1 : M]$ into (W_1, W_2) with $M = M_1 \times M_2$, where the first part W_1 , which is uniformly distributed on $[1 : M_1]$, is transmitted by encoder 1 and the second part W_2 , which is uniformly distributed on $[1 : M_2]$ and is conveyed to encoder 2 by W_{12} , is transmitted by two encoders cooperatively. Hence, when $(W_1, W_2) = (w_1, w_2)$, encoder 2 sends $x_2^n(w_2)$ while encoder 1 inputs $x_1^n(w_1, w_2)$ into the MAC.

Decoding: Let $\epsilon > 0$. Based on the observed channel output sequence y^n , the decoder finds the message pair (w_1, w_2) such that

$$(x_1^n(w_1, w_2), x_2^n(w_2), y^n) \in \mathcal{A}_\epsilon^{(n)}(X_1 X_2 Y), \tag{A4}$$

where set $\mathcal{A}_\epsilon^{(n)}(X_1 X_2 Y)$ is the set of jointly ϵ -typical sequences, see Cover and Thomas [18]. If such a pair cannot be found, or if there are more than one such pairs, an error is declared.

Probability of Error: Due to symmetry, the average probability of error is equivalent to the probability of error for an arbitrary message $w \in \{1, \dots, 2^{nR}\}$. Hence, without loss of generality, we assume $W = w = (w_1, w_2)$. Thus, we have

$$\begin{aligned}
 P_e^{(n)} &= \Pr\{E^c(w_1, w_2) \cup \bigcup_{(\tilde{w}_1, \tilde{w}_2) \neq (w_1, w_2)} E(\tilde{w}_1, \tilde{w}_2)\} \\
 &\leq \Pr\{E^c(w_1, w_2)\} + \sum_{(\tilde{w}_1, \tilde{w}_2) \neq (w_1, w_2)} \Pr\{E(\tilde{w}_1, \tilde{w}_2)\} \\
 &= \Pr\{E^c(w_1, w_2)\} + \sum_{\tilde{w}_1 \neq w_1} \Pr\{E(\tilde{w}_1, w_2)\} + \sum_{(\tilde{w}_1, \tilde{w}_2): \tilde{w}_2 \neq w_2} \Pr\{E(\tilde{w}_1, \tilde{w}_2)\},
 \end{aligned}
 \tag{A5}$$

where

$$E(w_1, w_2) \triangleq \{(x_1^n(w_1, w_2), x_2^n(w_2), Y^n) \in \mathcal{A}_\epsilon^{(n)}(X_1 X_2 Y)\}.
 \tag{A6}$$

Due to the Asymptotic Equipartition Property (AEP), it can be shown that

$$\Pr\{E^c(w_1, w_2)\} \leq \epsilon,
 \tag{A7}$$

for all n large enough. Moreover

$$\begin{aligned}
 \sum_{\tilde{w}_1 \neq w_1} \Pr\{E(\tilde{w}_1, w_2)\} &\leq (M_1 - 1) \sum_{(x_1^n, x_2^n, y^n) \in \mathcal{A}_\epsilon^{(n)}} P(x_1^n | x_2^n) P(x_2^n) P(y^n | x_2^n) \\
 &\leq (M_1 - 1) 2^{-n(I(X_1; Y | X_2) - 4\epsilon)},
 \end{aligned}
 \tag{A8}$$

and

$$\begin{aligned}
 \sum_{(\tilde{w}_1, \tilde{w}_2): \tilde{w}_2 \neq w_2} \Pr\{E(\tilde{w}_1, \tilde{w}_2)\} &= M_1(M_2 - 1) \sum_{(x_1^n, x_2^n, y^n) \in \mathcal{A}_\epsilon^{(n)}} P(x_1^n | x_2^n) P(x_2^n) P(y^n) \\
 &\leq M_1(M_2 - 1) 2^{-n(I(X_1, X_2; Y) - 3\epsilon)}.
 \end{aligned}
 \tag{A9}$$

Now as long as

$$\begin{aligned}
 M_1 &\leq 2^{n(I(X_1; Y | X_2) - 5\epsilon)}, \\
 M_1 M_2 &\leq 2^{n(I(X_1, X_2; Y) - 4\epsilon)},
 \end{aligned}
 \tag{A10}$$

$P_e^{(n)} \leq 2\epsilon$ for all n large enough. Therefore we take

$$\begin{aligned}
 \log_2 M_2 &= \min\{n(I(X_2; Y) + \epsilon), nC_{12}\}, \\
 \log_2 M_1 &= n(I(X_1; Y | X_2) - 5\epsilon),
 \end{aligned}
 \tag{A11}$$

then both (9) and (A10) are satisfied. Note that this implies that

$$\log_2 M_1 M_2 = \min\{n(I(X_1; Y | X_2) - 5\epsilon) + nC_{12}, n(I(X_1, X_2; Y) - 4\epsilon)\}.
 \tag{A12}$$

If we now let $\epsilon \rightarrow 0$, the achievability part of the Theorem 1 is thus established.

Appendix A.2. Proof of Theorem 3

The proof is a generalization of the proof of the two-encoder settings. Consider a simplified block diagram of the N -encoder setting as shown in Figure A1. Now, consider a cut of the fronthaul link between X_m and X_{m+1} for any given $m \in [1 : N - 1]$ such that the nodes in the network are separated in two sets of $\{\underline{X}_1^m\}$ and $\{\underline{X}_{m+1}^N, Y\}$.

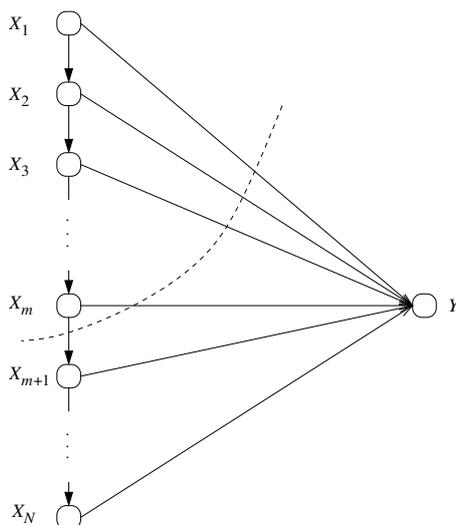


Figure A1. Simplified illustration of the MAC for the N -encoder setting.

(i) *Converse.* Consider the Markovity of $W \rightarrow (X_1^n, X_2^n, \dots, X_N^n) \rightarrow Y^n$. By applying Fano’s inequality, we first have

$$\begin{aligned}
 \log_2 M &\leq I(W, W_{12}, \dots, W_{N-1,N}; Y^n) + H(W|Y^n) \\
 &\leq I(X_1^n, X_2^n, \dots, X_N^n; Y^n) + F \\
 &\leq \sum_{i=1}^n I(X_{1i}, X_{2i}, \dots, X_{Ni}; Y_i) + F \\
 &\leq nI(\underline{X}_1^N; Y) + F.
 \end{aligned}
 \tag{A13}$$

Then, considering the cut between X_m and X_{m+1} , we have that

$$\begin{aligned}
 \log_2 M &= I(W, W_{12}, \dots, W_{N-1,N}; Y^n) + H(W|Y^n) \\
 &\leq I(W_{m,m+1}; Y^n) + I(W, W_{12}, \dots, W_{m-1,m}; Y^n | W_{m,m+1}, \dots, W_{N-1,N}) + F \\
 &\leq \log_2 M_{m,m+1} + I(X_1^n, X_2^n, \dots, X_m^n; Y^n | X_{m+1}^n, X_N^n) + F \\
 &\leq nC_{m,m+1} + \sum_{i=1}^n I(X_{1i}, X_{2i}, \dots, X_{mi}; Y_i | X_{m+1,i}, \dots, X_{N,i}) + F \\
 &\leq nC_{m,m+1} + nI(\underline{X}_1^m; Y | \underline{X}_{m+1}^N) + F.
 \end{aligned}
 \tag{A14}$$

Note that the above result is valid for any m in $[1 : N - 1]$. Thus, by letting $n \rightarrow \infty$ the converse follows.

(ii) *Achievability.* First consider the message W that can be represented by N independent messages as $W = \{W_i\}_{i=1}^N$, where each W_i is uniformly distributed on $[1 : M_i]$ with $\prod_{i=1}^N M_i = M$. Then, given by the linear topology of encoders, we distribute $\{W_i\}_{i=1}^N$ into the network in the manner illustrated by Figure A2, i.e., for the link between any X_m and X_{m+1} , the fronthaul message $W_{m,m+1}$ conveys corresponding messages $\{W_i\}_{i=m}^N$. Therefore, for a fixed distribution $p(x_1, x_2, \dots, x_N)$ and corresponding marginals, we can first generate M_N i.i.d. n -sequences $x_N^n(w_N)$ with $w_N \in [1, M_N]$ according to $\Pr(X_N^n = x_N^n) = \prod_{i=1}^n p_{X_N}(x_{Ni})$ and then for each $x_N^n(w_N)$ generate M_{N-1} i.i.d. n -sequences $x_{N-1}^n(w_{N-1}, w_N)$ with $w_{N-1} \in [1, M_{N-1}]$ according to $\Pr(X_{N-1}^n = x_{N-1}^n | X_N^n = x_N^n(w_N)) = \prod_{i=1}^n p_{X_{N-1}|X_N}(x_{N-1,i} | x_{Ni}(w_N))$ and so on. In this way, an N -layer superposition codebook is generated and revealed at both encoders and decoder.

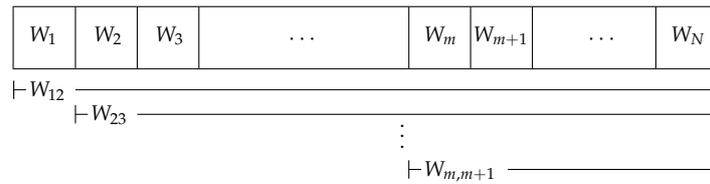


Figure A2. N -layer superposition coding message structure.

Thus, for sending a message w , encoders transmit sequences $\{x_m^n(w_m, w_{m+1}, \dots, w_N)\}_{m=1}^N$ over the MAC channel. At the decoder, a unique message tuple (w_1, w_2, \dots, w_N) is found by using simultaneous typicality decoding as performed for the two-encoder case. By taking the similar probability of error analysis, it gives that, as long as

$$\begin{aligned}
 M_1 &\leq 2^{n(I(X_1; Y | \underline{X}_2^N) - 5\epsilon)}, \\
 M_1 M_2 &\leq 2^{n(I(X_1, X_2; Y | \underline{X}_3^N) - 5\epsilon)}, \\
 &\vdots \\
 \prod_{m=1}^N M_m &\leq 2^{n(I(\underline{X}_1^N; Y) - 4\epsilon)},
 \end{aligned} \tag{A15}$$

we can have $P_e^{(n)} \leq 2\epsilon$ for all sufficiently large n and any $\epsilon > 0$. By further considering $M_N \leq 2^{nC_{N-1,N}}$, $M_{N-1}M_N \leq 2^{nC_{N-2,N-1}}, \dots$, and $\prod_{m=2}^N \leq 2^{nC_{12}}$, we can subsequently take

$$\begin{aligned}
 \log_2 M_N &= \min\{n(I(X_N; Y) + \epsilon), nC_{N-1,N}\} \\
 \log_2 M_{N-1} &= \min\{nI(X_{N-1}; Y | X_N), nC_{N-2,N-1} - \log_2 M_N\} \\
 &\vdots \\
 \log_2 M_1 &= n(I(X_1; Y | \underline{X}_2^N) - 5\epsilon).
 \end{aligned} \tag{A16}$$

Finally, observe that (again subsequently)

$$\begin{aligned}
 \log_2 M_N &= \min\{n(I(X_N; Y) + \epsilon), nC_{N-1,N}\} \\
 \log_2 M_{N-1}M_N &= \min\{I(X_{N-1}, X_N; Y) + \epsilon, \\
 &\quad I(X_{N-1}; Y | X_N) + nC_{N-1,N}, nC_{N-2,N-1}\} \\
 &\vdots \\
 \log_2 M &= \log_2 \prod_{m=1}^N M_m \\
 &= \min\{n(I(\underline{X}_1^N; Y) - 4\epsilon), \\
 &\quad \{I(\underline{X}_1^m; Y | \underline{X}_{m+1}^N) + C_{m,m+1}\}_{m=1}^{N-1}\},
 \end{aligned} \tag{A17}$$

which establishes the achievability for $n \rightarrow \infty$ and $\epsilon \rightarrow 0$.

Appendix A.3. Proof of Lemma 1

For any fixed $C_B \leq (N - 1)C_{\text{full}}(N)$, we first consider a realization of $\underline{\beta}$. Since $I_m \leq I_{m+1}$ for any $m \in [1, N - 1]$, by distributing C_B on top of $\{I_m\}_{m=1}^{N-1}$ in a water-filling fashion, we can always have two possible cases if the equality of all terms on the RHS of (30) can not be achieved, i.e.,

$$\text{Case (a) : } \mathcal{L}(C_B, \underline{\beta}) < I_j, \tag{A18}$$

$$\text{or Case (b): } \mathcal{L}(C_B, \underline{\beta}) > I_N, \tag{A19}$$

where j is the smallest index such that $C_{m,m+1} = 0$ for all $m \in [j - 1, N - 1]$, and $\mathcal{L}(C_B, \underline{\beta}) = I_m + C_{m,m+1}$ for any m such that $C_{m,m+1} \neq 0$, namely the ‘water level’. Once the water-filling is performed, we fix the corresponding distribution of C_B .

For Case (a) where $C = \mathcal{L}$, we now can decrease $\{\beta_m\}_{m=j}^N$ to increase β_1 such that \mathcal{L} is increased. For Case (b) where $C = I_N$, we can look for an $m < N$ with $\beta_m > 0$ such that by decreasing β_m , β_N increases and I_N increases as well. Therefore, the $\underline{\beta}$ satisfying Case (a) and Case (b) are not optimal. The equality of the terms on the RHS of (30) is thus necessary.

Appendix A.4. Proof of Proposition 2

Three steps are taken in the proof. In step (1) we show that an active compound mode $\langle j, k \rangle$ achieves the capacity if $LB \leq UB$ is satisfied. In step (2) we show that using any two separated active modes (all other modes are inactive) does not achieve the capacity. In step (3) we show that exact solutions of (54) and (55) are resulted.

Step (1) Note that function g given in (42) is convex- \cap when $\lambda \leq \frac{1}{k-1}$ if the largest activated mode is k . So, we set the partial derivatives of g with respect to $\{\beta_i\}_{i=1}^N$ according to the Kuhn–Tucker conditions, see ([23], eqn.4.4.10 and eqn.4.4.11), when the active compound mode $\langle j, k \rangle$ achieves the capacity. By considering (42) in nats, the partial derivative of function g with respect to β_i is

$$\begin{aligned} \frac{\partial g}{\partial \beta_i} &= \frac{1 - (k - 1)\lambda}{2} \cdot \frac{iP}{1 + \sum_{l=j}^N l\beta_l P} \\ &+ \frac{\lambda}{2} \sum_{m=i}^{N-1} \frac{iP}{1 + \sum_{l=1}^m l\beta_l P}, \end{aligned} \tag{A20}$$

where $i \in [1 : N]$.

Step (1.1) Firstly, by only considering that compound mode $\langle j, k \rangle$ is active, i.e., all $\{\beta_i\}_{i=j}^k$ are nonzero, while the other $\{\beta_i\}_{i=1}^{j-1}$ and $\{\beta_i\}_{i=k+1}^N$ are zeros, the partial derivative can be expressed as

$$\begin{aligned} \frac{\partial g}{\partial \beta_i} &= \frac{1 - (k - 1)\lambda}{2} \cdot \frac{iP}{1 + \sum_{l=j}^k l\beta_l P} \\ &+ \frac{\lambda}{2} \sum_{m=i}^{k-1} \frac{iP}{1 + \sum_{l=j}^m l\beta_l P}. \end{aligned} \tag{A21}$$

For simplicity, we denote that

$$D(i) \triangleq 1 + \sum_{l=j}^i l\beta_l P. \tag{A22}$$

Now, consider that the partial derivatives corresponding to $i \in [j : k]$ should be all identical to some value μ , i.e.,

$$\frac{\partial g}{\partial \beta_j} = \frac{\partial g}{\partial \beta_{j+1}} = \dots = \frac{\partial g}{\partial \beta_k} := \mu, \tag{A23}$$

to find the capacity solution in terms of optimal distribution of $\underline{\beta}$. Note that

$$\frac{\partial g}{\partial \beta_k} = \frac{1 - (k - 1)\lambda}{2} \cdot \frac{kP}{D(k)} = \mu. \tag{A24}$$

For the case of $k > j$, we can recursively evaluate the equalities in (A23) as $\partial g / \partial \beta_i = \partial g / \partial \beta_{i-1}$ by taking i from k to $j + 1$ in a descending order with the use of (A21). In such a way, it is obtained that

$$\frac{\lambda}{2} \cdot \frac{i(i + 1)P}{D(i)} = \mu \quad \text{for } i \in [j : k - 1]. \tag{A25}$$

Now, based on (A24) and (A25), we can derive expressions of $\{\beta_i\}_{i=j}^k$, by considering two scenarios.

Scenario 1: Consider $k \geq j + 2$, i.e., at least three consecutive modes are active. By taking $i = j$ and $i = j + 1$, (A25) can be used twice to obtain the equality $\frac{j}{D(j)} = \frac{j+2}{D(j+1)}$ that gives the relation

$$\beta_j P = \frac{(1+j)\beta_{j+1}P}{2} - \frac{1}{j}. \tag{A26}$$

For $k > j + 2$, expression (A25) allows us to further obtain

$$\frac{i(i+1)}{D(i)} = \frac{(i+1)(i+2)}{D(i+1)} \tag{A27}$$

by taking i in the order of $j + 1$ to $k - 1$, which results in an interesting and important relation

$$\beta_i = \beta_{j+1} \quad \text{for } i \in [j + 2 : k - 1]. \tag{A28}$$

Therefore, by applying relation (A26), we can express $D(k)$ as

$$D(k) = \frac{k^2 - k}{2} \beta_{j+1} P + k \beta_k P, \tag{A29}$$

which is valid for the case of $k = j + 2$ as well. So, by setting (A25) equal to (A24) with $i = j$ as

$$\frac{1 - (k-1)\lambda}{2} \cdot \frac{k}{D(k)} = \frac{\lambda}{2} \cdot \frac{j(j+1)}{D(j)}, \tag{A30}$$

and substituting $D(k)$ in (A29), we can first derive the power of modes from $j + 1$ to $k - 1$ as

$$\beta_{j+1} P = \frac{2\lambda(1+jP)}{j(1-(j-1)\lambda)}. \tag{A31}$$

According to (A26), we can then obtain the power of the first mode as

$$\beta_j P = \frac{\lambda(1+j)(1+jP)}{j(1-(j-1)\lambda)} - \frac{1}{j}. \tag{A32}$$

Furthermore, owing to $\sum_{i=j}^k \beta_i = 1$, we can finally represent the power of the last mode as

$$\beta_k P = P + \frac{1}{j} - \frac{1}{2}(2k - j - 1)\beta_{j+1} P. \tag{A33}$$

Now, applying the total power constraint, we should have

$$\begin{cases} 0 < \beta_k P < P \\ 0 < \beta_{j+1} P < P \\ 0 < \beta_j P < P. \end{cases} \tag{A34}$$

By substituting (A33), (A31), and (A32), the corresponding slope λ should simultaneously satisfy

$$\left\{ \begin{array}{l} \frac{1}{2(k-1) + jP(2k-j-1)} < \lambda < \frac{1}{2(k-1)} \\ 0 < \lambda < \frac{jP}{2 + jP + j^2P} \\ \frac{1}{j(2 + (j+1)P)} < \lambda < \frac{1}{2j}. \end{array} \right. \tag{A35}$$

Since $k > j + 1$, it is easy to see that the lower bound of λ is $\frac{1}{j(2+(j+1)P)}$. For the upper bound, if $\frac{1}{2(k-1)} > \frac{j^P}{2+j^P+j^2P}$, it leads to $P < \frac{2}{2jk-j^2-3j}$. Due to $j \leq k - 2$, such P results in

$$j(2 + (j + 1)P) < 2j + \frac{2(j + 1)}{2k - j - 3} \leq 2(k - 1), \tag{A36}$$

which contradicts the lower bound. Therefore, to make the *compound mode* exist, the slope should be in the range

$$\frac{1}{j(2 + (j + 1)P)} < \lambda < \frac{1}{2(k - 1)}. \tag{A37}$$

Scenario 2: Consider $k = j + 1$, i.e., the *compound mode* only consists of two *modes*. So, letting (A24) equal to (A25) gives the relation

$$\beta_j P = \frac{\lambda(1 + j)\beta_{j+1}P}{1 - 2j\lambda} - \frac{1}{j}. \tag{A38}$$

By considering $\beta_j + \beta_{j+1} = 1$ now, it is obtained

$$\beta_{j+1}P = \frac{(1 - 2\lambda j)(1 + jP)}{j(1 - (j - 1)\lambda)} \tag{A39}$$

and a same $\beta_j P$ expression as in (A32). If such *compound mode* gives optimal solution, the condition $0 < \beta_j P < P$ should be also satisfied, which gives $\lambda \in (\frac{1}{j(2+(1+j)P)}, \frac{1}{2j})$ that is consistent with the result of (A37).

Step (1.2) Secondly, consider that the derivative $\partial g / \partial \beta_{k+i}$ for $\forall i \in [1 : N - k]$ with $k < N$ should be less than μ as denoted in Step (1.1). Since $\{\beta_{k+i}\}_{i=1}^{N-k} = 0$, we have

$$\frac{\partial g}{\partial \beta_{k+i}} = \frac{1 - (k + i - 1)\lambda}{2} \cdot \frac{(k + i)P}{D(k)} \tag{A40}$$

$$\leq \frac{\partial g}{\partial \beta_k} = \frac{1 - (k - 1)\lambda}{2} \cdot \frac{kP}{D(k)} \tag{A41}$$

which directly gives

$$\lambda \geq \frac{1}{2k}. \tag{A42}$$

Step (1.3) Finally, consider that the derivative $\partial g / \partial \beta_{j-i}$ for $\forall i \in [1 : j - 1]$ with $j > 1$ should be less than μ as well. Since $\{\beta_{j-i}\}_{i=1}^{j-1} = 0$, we have

$$\frac{\partial g}{\partial \beta_{j-i}} = \frac{1 - (k - 1)\lambda}{2} \cdot \frac{(j - i)P}{D(k)} + \frac{\lambda}{2}(j - i)P + \frac{\lambda}{2} \sum_{m=j}^{k-1} \frac{(j - i)P}{D(m)}. \tag{A43}$$

Similarly, by upper bounding this derivative by μ given in (A24), we can have

$$\frac{1 - (k - 1)\lambda}{2} \cdot \frac{iP}{D(k)} + \frac{\lambda}{2} \sum_{m=j}^{k-1} \frac{iP}{D(m)} \geq \frac{\lambda}{2}(j - i)P, \tag{A44}$$

where the summation can be computed by using relation of (A25) as

$$\frac{\lambda}{2} \sum_{m=j}^{k-1} \frac{iP}{D(m)} = i\mu \left(\frac{1}{j} - \frac{1}{k} \right). \tag{A45}$$

By further incorporating μ given by (A24) and $D(k)$ given by (A29) into (A44), the inequality becomes to

$$\frac{1 - (k - 1)\lambda}{j} \geq \frac{(j - 1)\lambda}{i} \left(\frac{k - 1}{2} \beta_{j+1}P + \beta_k P \right). \tag{A46}$$

By substituting (A31) and (A33) for $\beta_{j+1}P$ and $\beta_k P$, it can be easily shown that

$$\lambda \leq \frac{i}{i(j - i) + (j - 1)(1 + jP)} \leq \frac{1}{(j - 1)(2 + jP)}, \tag{A47}$$

where $i \geq 1$ is applied in the bounding for the last step.

Note that, Step (1.2) and (1.3) and resulted bounds of λ also cover the the case of $j = k$, i.e., only one mode is active and achieves the capacity. Now, by considering the ranges given by (A37), (A42), and (A47), the slope bounds in (52) and (53) are resulted.

Step (2) Assume that the capacity can also be achieved by only activating any two separated modes j' and k' , where $j' \in [1 : k' - 1]$ and $k' \in [j' + 1, N]$. Then, the Kuhn–Tucker condition requires $\partial g / \beta_{k'} = \partial g / \beta_{j'}$, which results in the relation of

$$\frac{1 - (k' - 1)\lambda}{2} \frac{1}{j'D(k')} = \frac{\lambda}{2} \frac{1}{1 + j'\beta'_{j'}}. \tag{A48}$$

Note that in $D(k')$ only β'_k and $\beta'_{j'}$ are nonzero. Moreover, for $\forall i \in [1 : k' - j' - 1]$, we have

$$\frac{\partial g}{\partial \beta'_{k'-i}} = \frac{1 - (k' - 1)\lambda}{2} \frac{(k' - i)P}{D(k')} + \frac{\lambda}{2} \frac{i(k' - i)P}{1 + j'\beta'_{j'}} \tag{A49}$$

By substituting the relation (A48) into above derivative, it can be shown that

$$\begin{aligned} \partial g / \beta'_{k'-i} - \partial g / \beta'_k &= P \frac{1 - (k' - 1)\lambda}{2D(k')} \left(k' - i + \frac{i(k' - i)}{j'} - k' \right) \\ &= P \frac{1 - (k' - 1)\lambda}{2D(k')} i \left(\frac{k' - i}{j'} - 1 \right) \\ &> 0 \end{aligned} \tag{A50}$$

where the last step is due to $k' - i > j'$. This result contradicts to the Kuhn–Tucker condition. This demonstrates the only compound modes achieves the capacity.

Step (3) For a slope λ in the range (52) and (53), the obtained optimal $\underline{\beta}$ from Step (1.1) can be used to evaluate (38) and (39) directly. Thus, (54) and (55) are resulted, respectively.

Appendix A.5. Proof of Corollary 3

Consider four possible settings of *UB* and *LB* based on (52) and (53) as

$$\begin{aligned} \text{Case 1 : } & \frac{1}{2(k - 1)} \leq b(j - 1), \quad \frac{1}{2k} \leq b(j), \quad b(j) < \frac{1}{2(k - 1)}, \\ \text{Case 2 : } & \frac{1}{2(k - 1)} \geq b(j - 1), \quad \frac{1}{2k} \leq b(j), \quad b(j) < b(j - 1), \\ \text{Case 3 : } & \frac{1}{2(k - 1)} \leq b(j - 1), \quad \frac{1}{2k} \geq b(j), \quad \frac{1}{2k} < \frac{1}{2(k - 1)}, \\ \text{Case 4 : } & \frac{1}{2(k - 1)} \geq b(j - 1), \quad \frac{1}{2k} \geq b(j), \quad \frac{1}{2k} \leq b(j - 1), \end{aligned}$$

where inequalities given by $\frac{1}{2k}$ are only valid for $k < N$. By evaluating the inequalities, the power condition (57) is resulted. Since the *LB* and *UB* are derived from the Kuhn–Tucker conditions, the

derived power condition is a sufficient and necessary condition. This result directly gives that using compound mode $\langle j', k' \rangle$ with $j' > j$ and $k' < k$ is not a capacity solution.

Appendix A.6. Proof of Lemma 2

By following the method in [20], we define $f(x) = x - \ln(x + 1)$ and prove

$$f(x) + \sqrt[3]{2f(x)} - x \geq 0 \quad (\text{A51})$$

for $x \geq 0$. Once this is done, one can take the procedure applied in ([20], Thm. 1) and the lower bound (61) follows.

Now, we start converting (A51) to an equivalent problem. First, by substituting $f(x)$ in (A51), we need to prove

$$\sqrt[3]{2x - 2\ln(1+x)} \geq \ln(1+x). \quad (\text{A52})$$

By denoting $t := x + 1$ with $t \geq 1$ and considering x^3 and $x^{\frac{1}{3}}$ are monotonically increasing in x for $x \geq 0$, showing (A52) is equivalent to show

$$2t - 2 - \ln t - (\ln t)^3 \geq 0. \quad (\text{A53})$$

let $v(t) := 2t - 2 - \ln t - (\ln t)^3$, we can focus on showing

$$v'(t) = 2 - 2t^{-1} - 3t^{-1}(\ln t)^2 \geq 0, \quad (\text{A54})$$

since $v(1) = 0$. Considering $t \geq 1$, we finally convert proving (A51) to demonstrating

$$\psi(t) := tv'(t) = 2t - 2 - 3(\ln t)^2 \geq 0. \quad (\text{A55})$$

To do so, we take the first derivative of $\psi(t)$ and set it to zero to have

$$3 \ln t = t. \quad (\text{A56})$$

By solving (A56), we can evaluate the local maxima and/or local minima of $\psi(t)$. It can be seen that (A56) has only two real roots for $t \geq 1$, which are

$$t_1 = -3W_0\left(-\frac{1}{3}\right) \quad \text{and} \quad t_2 = -3W_{-1}\left(-\frac{1}{3}\right), \quad (\text{A57})$$

where $W_0(\cdot)$ is the principle branch of the Lambert W function defined over $[-e^{-1}, \infty)$. Based on the property of the Lambert W function, $1 < t_1 < t_2$. Therefore, since $\psi(1) = 0$, we can have two scenarios as

- Scenario 1: if $\psi(t_1) > 0$, we must have $\psi(t_2) < \psi(t_1)$, i.e., t_1 gives a local maxima and t_2 gives a minima;
- Scenario 2: if $\psi(t_1) < 0$, we must have $\psi(t_2) > \psi(t_1)$, i.e., t_1 gives a global minima and t_2 gives a maxima.

Hence, if we can prove that $\psi(t_1) > 0$ and $\psi(t_2) > 0$, both, we can conclude (A55) and thus (A51). We show this in what follows.

By setting $t^* = t_1$ or t_2 , $\psi(t^*)$ can be expressed as

$$\psi(t^*) = \psi(t)|_{t_1, t_2} = 2t^* - 2 - \frac{(t^*)^2}{3}. \quad (\text{A58})$$

If $\psi(t^*) > 0$, it is equivalent to have $(t^*)^2 - 6t^* + 6 < 0$, which results in the range of

$$3 - \sqrt{3} < t^* < 3 + \sqrt{3} \quad (\text{A59})$$

must be satisfied. For t_2 , by applying the bounds given in ([20], Thm. 1) which is $-1 - \sqrt{2x} - x < W_{-1}(-e^{-x-1}) < -1 - \sqrt{2x} - \frac{2}{3}x$, it is easily obtained that $4.53 < t_2 < 4.62$, which is in the range (A59). So, $\psi(t_2) > 0$ follows. For t_1 , we apply an upper bound on $W_0(\cdot)$ given in ([24], Thm. 2.3), which is

$$W_0(x) \leq \ln\left(\frac{x+y}{1+\ln y}\right) \quad (\text{A60})$$

for $x > -e^{-1}$ and $y > e^{-1}$. By taking $y = 0.5$, t_1 can be bounded as $t_1 > 1.8$. By also considering $t_1 < t_2$, t_1 must in the range (A59) as well, which completes showing $\psi(t_1) > 0$.

Numerical evaluation of the bound (61) verifies the proof as illustrated in Figure A3, where the bound in [20] is plotted as a reference as well.

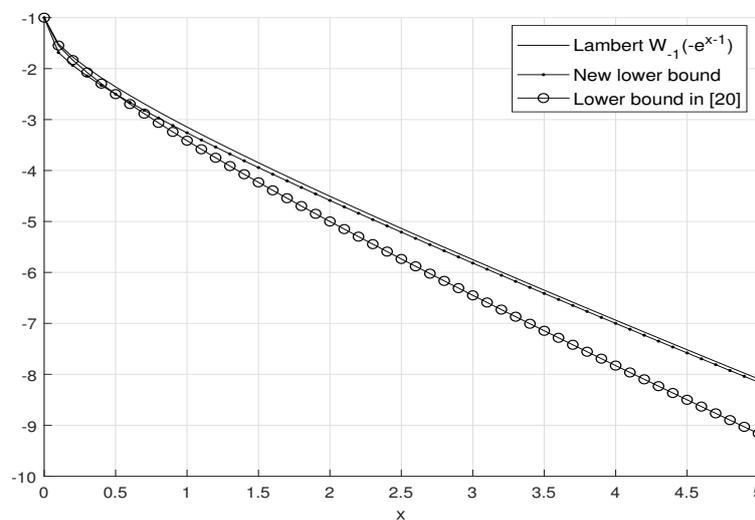


Figure A3. The bounds on the Lambert function $W_{-1}(-e^{-(x+1)})$.

References

1. Ngo, H.Q.; Ashikhmin, A.; Yang, H.; Larsson, E.G.; Marzetta, T.L. Cell-free massive MIMO versus small cells. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 1834–1850. [CrossRef]
2. Zhang, J.; Chen, S.; Lin, Y.; Zheng, J.; Ai, B.; Hanzo, L. Cell-free massive MIMO: A new next-generation paradigm. *IEEE Access* **2019**, *7*, 99878–99888. [CrossRef]
3. Interdonato, G.; Björnson, E.; Ngo, H.G.; Frenger, P.; Larsson, E.G. Ubiquitous cell-free massive MIMO communications. *EURASIP J. Wirel. Commun. Netw.* **2019**, *197*. [CrossRef]
4. Bashar, M.; Cumanan, K.; Burr, A.G.; Ngo, H. Q.; Debbah, M. Cell-Free massive MIMO with limited fronthaul. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas, MO, USA, 20–24 May 2018; pp. 1–7.
5. Bashar, M.; Cumanan, K.; Burr, A.G.; Ngo, H. Q.; Larsson, E.G.; Xiao, P. On the energy efficiency of limited-fronthaul cell-free massive MIMO. In Proceedings of the 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–7.
6. Femenias, G.; Riera-Palou, F. Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity. *IEEE Access* **2019**, *7*, 44596–44612. [CrossRef]
7. Frenger, P.; Hederen, J.; Hessler, M.; Interdonato, G. Improved Antenna Arrangement for Distributed Massive MIMO (2017). Patent Application WO2018103897. Available Online: patentscope.wipo.int/search/en/WO2018103897 (accessed on 21 January 2020).
8. Radio Stripes: Re-Thinking Mobile Networks. Available Online: <https://www.ericsson.com/en/blog/2019/2/radio-stripes> (accessed on 21 January 2020).

9. Willems, F. The discrete memoryless multiple access channel with partially cooperating encoders (corresp). *IEEE Trans. Inf. Theory* **1983**, *29*, 441–445. [[CrossRef](#)]
10. El Gamal, A.; Zahedi S. Capacity of a class of relay channels with orthogonal components. *IEEE Trans. Inf. Theory* **2005**, *51*, 1815–1817. [[CrossRef](#)]
11. Ghabeli, L.; Aref, M.R. A new achievable rate and the capacity of some classes of multilevel relay network. *EURASIP J. Wirel. Commun. Netw.* **2008**, *2008*, 135857. [[CrossRef](#)]
12. Kang, W.; Liu, N.; Chong, W. The Gaussian multiple access diamond channel. *IEEE Trans. Inf. Theory* **2015**, *61*, 6049–6059. [[CrossRef](#)]
13. Saeedi Bidokhti, S.; Kramer, G. Capacity Bounds for Diamond Networks With an Orthogonal Broadcast Channel. *IEEE Trans. Inf. Theory* **2016**, *62*, 7103–7122. [[CrossRef](#)]
14. El Gamal, A.; Aref, M. The capacity of the semideterministic relay channel (corresp.). *IEEE Trans. Inf. Theory* **1982**, *28*, 536–536. [[CrossRef](#)]
15. EL Gamal, A.; Hassanpour, N.; Mammen, J. Relay networks with delays. *IEEE Trans. Inf. Theory* **2007**, *53*, 3413–3431. [[CrossRef](#)]
16. Zhang, P.; Willems, F.; Huang, L. Capacity study of distributed beamforming in relation to constrained backbone communication. In Proceedings of the 2014 International Symposium on Information Theory and its Applications, Melbourne, VIC, Australia, 26–29 October 2014; pp. 473–477.
17. Golub, G.H; Van Loan, C.F. *Matrix Computations*; The Johns Hopkins University Press: Baltimore, MD, USA, 2013.
18. Cover, T; Thomas, J. *Elements of Information Theory*; Wiley: New York, NY, USA, 2006.
19. Mitrinović, D; Vasić, P. *Analytic Inequalities, ser. Grundlehren der Mathematischen Wissenschaften*; Springer: Berlin, Germany, 1970.
20. Chatzigeorgiou, I. Bounds on the Lambert function and their application to the outage analysis of user cooperation. *IEEE Commun. Lett.* **2013**, *17*, 1505–1508. [[CrossRef](#)]
21. Corless, R.M.; Gonnet, G.H.; Hare, D.; Jeffrey, D.J.; Knuth, D.E. On the Lambert W function. *Adv. Comput. Math.* **1996**, *5*, 329. [[CrossRef](#)]
22. Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*; Dover Publications, Incorporated: St. Mineola, NY, USA, 1964.
23. Gallager, G. *Information Theory and Reliable Communication*; John Wiley & Sons, Inc.: New York, NY, USA, 1968.
24. Hoorfar, A.; Hassani, M. Inequalities on the Lambert W function and hyperpower function. *J. Inequal. Pure Appl. Math.* **2008**, *9*, 51.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).