

# Energy-Efficient, Large-Scale Distributed-Antenna System (L-DAS) for Multiple Users

Jingon Joung, *Member, IEEE*, Yeow Kiang Chia, *Member, IEEE*, and Sumei Sun, *Senior Member, IEEE*

**Abstract**—Large-scale distributed-antenna system (L-DAS) with very large number of distributed antennas, possibly up to a few hundred antennas, is considered. A few major issues of the L-DAS, such as high latency, energy consumption, computational complexity, and large feedback (signaling) overhead, are identified. The potential capability of the L-DAS is illuminated in terms of an energy efficiency (EE) throughout the paper. We firstly and generally model the power consumption of an L-DAS, and formulate an EE maximization problem. To tackle two crucial issues, namely the huge computational complexity and large amount of feedback (signaling) information, we propose a channel-gain-based antenna selection (AS) method and an interference-based user clustering (UC) method. The original problem is then split into multiple subproblems by a cluster, and each cluster's precoding and power control are managed in parallel for high EE. Simulation results reveal that i) using all antennas for zero-forcing multiuser multiple-input multiple-output (MU-MIMO) is energy inefficient if there is nonnegligible overhead power consumption on MU-MIMO processing, and ii) increasing the number of antennas does not necessarily result in a high EE. Furthermore, the results validate and underpin the EE merit of the proposed L-DAS complied with the AS, UC, precoding, and power control by comparing with non-clustering L-DAS and colocated antenna systems.

**Index Terms**—Energy efficiency, distributed antenna system, large-scale networks, large MIMO, clustering.

## I. INTRODUCTION

**L**ARGE-SCALE (or massive) multiple-input multiple-output (L-MIMO) techniques have been rigorously studied to tremendously improve spectral efficiency (SE, b/s/Hz) of wireless communications. The L-MIMO employs very large number of colocated antennas, which can effectively mitigate *small-scale* (local) distortion such as noise at the receiver and fast fading (see [1]–[4] and the references therein). On the other hand, a distributed-antenna system (DAS) is also one of the promising technologies to effectively improve SE of wireless communications. The DAS is implemented with multiple distributed antennas (DAs) through base stations (BSs) located in different cells, i.e., coordinated multipoint (CoMP) transmission (see [5] and references therein), or through distributed radio remote heads typically located in the same cell.

Manuscript received September 03, 2013; revised January 11, 2014; accepted February 18, 2014. Date of publication March 05, 2014; date of current version September 11, 2014. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Geoffrey Li.

The authors are with the Institute for Infocomm Research (I<sup>2</sup>R), A\*STAR, Singapore 138632 (e-mail: jgjoung@i2r.a-star.edu.sg; chiayk@i2r.a-star.edu.sg; sunsm@i2r.a-star.edu.sg).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2014.2309942

Thus, the DAS can mitigate *large-scale* fading (path loss) using the many antennas distributed geographically. The SE of DAS has been mainly studied (see, e.g., [6]–[10]). In [6], the authors show proportional relationship between SE and the number of DAs. A suboptimal power control method and a simple antenna selection (AS) method are proposed to improve the SE [7]. It is shown that a single DA usage is preferable to full DA usage in multicell scenario [8], while the opposite results are observed in a single, isolated cell [9] and also in a multiuser (MU) scenario [10].

DAS's channels are typically modeled as the composite channels including *uncorrelated* large- and small-scale fading channels, which are a crucial part of motivation of a DAS technique and differentiate the DAS from L-MIMO techniques. Note that all colocated antennas suffer almost *identical* large-scale fading and *highly correlated* small scale fading. With the common goal, i.e., high SE, it is natural step to consider very *large-scale* DAS (L-DAS) to obtain the synergy for further SE improvement}. Nevertheless, the L-DAS has rarely been studied due to the ambiguity of cost of *very large-size* networks. To observe the tradeoff between the cost and benefit, and to quantify the efficacy of L-DAS, we consider an energy efficiency (EE, b/J) that is the total number of reliably decoded bits normalized by the consumed energy. The EE is a widely used metric in wireless communications recently to find a Pareto optimality between throughput and energy consumption (see e.g., [11], [12]). The SE-EE tradeoff has been analyzed for a single user (SU) in DAS systems [13], and EE optimal power control has been proposed for DAS to support an SU [14]. In [15], the authors show convergence of total transmit power and per-user sum rate when the number of DAs and users approach to *infinity* like an L-DAS, and provide an asymptotical EE that is a ratio between per-user sum rate and total transmit power. However, total transmit power does not imply total power/energy consumption in practice due to the overhead power consumption at the transmitter. Though the overhead to process additional DAs has been recently addressed in [16], the EE behavior of L-DAS is still unclear and difficult to be conjectured from the existing studies, especially for an MU scenario. Note that the power consumption of L-DAS prohibitively increases as the network size increases.

In this paper, the EE of L-DAS is studied in MU scenario. We model an L-DAS transmitter and its power consumption including the overhead. An EE maximization problem is formulated under constraints on per-antenna transmit power and per-user rate to select DAs, design MU precoding, and control transmit power. To tackle the original, computationally intractable optimization problem, we propose a channel-gain-based AS method and an interference-based user clustering (UC) method, which enable us to split the original

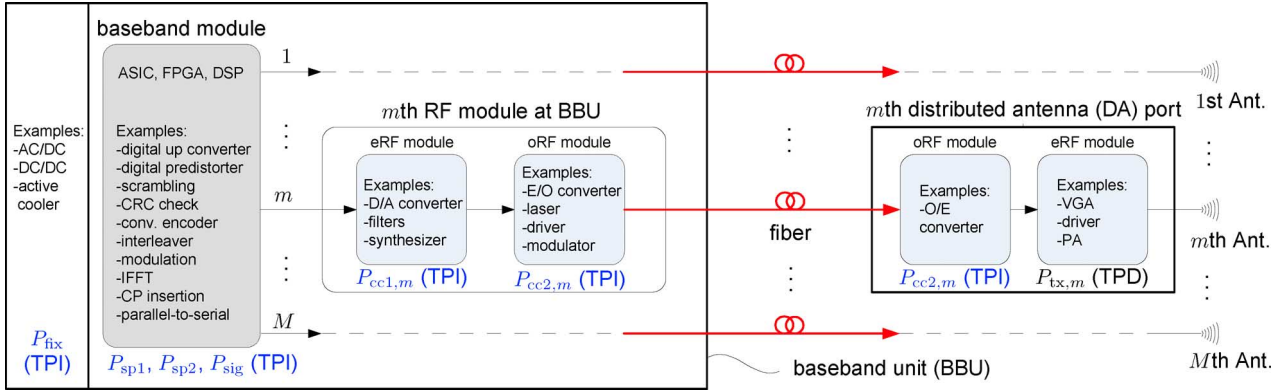


Fig. 1. Illustration of an L-DAS transmitter with a BBU and  $M$  DAs. The BBU consists of one baseband,  $M$  electric-RF (eRF), and  $M$  optical-RF (oRF) modules. Each DA consists of one-pair of remote oRF and eRF modules. The TPD term  $P_{tx,m}$  is the  $m$ th summand in (5).

problem into multiple cluster-based subproblems. The multiple subproblems can be solved in parallel, resulting in computational complexity and feedback (signaling) reduction. Each subproblem is further divided into precoding design and power control problems. EE-aware precoding is derived for each cluster, and the per-cluster optimal and heuristic power control algorithms are then proposed. To further improve EE, additional DA assignment and clustering threshold adaptation are considered. Simulation with practical parameters is performed to observe an average EE over clustering threshold, number of users, and network size. The results of the paper will be a useful reference for further study of energy efficient L-DAS. The summary of main contributions and results of our work:

- We propose an L-DAS, which is a new, natural extension of L-MIMO to DAS systems.
- We provide a practical power consumption model for the L-DAS, which can be readily modified and applied to any types of distributed systems (Section III).
- We formulate an EE maximization problem for a general L-DAS setup (Section III), and solve it through a sub-optimal, decomposition strategy.
- We propose simple AS and UC methods to split the original problem (Section IV), which enables a cluster-based design, resulting in reduction of computational complexity and signaling overhead.
- We generalize the results in [17] for the precoding (Section V) and per-cluster optimal and heuristic power control (Section VI) of L-DAS.
- We show informative simulation results, under the practical power consumption model (Section VII).
  - Using all DAs for MU-MIMO could be an energy-inefficient strategy if there is nonnegligible overhead power consumption for MU-MIMO precoding.
  - Increasing the number of DAs does not necessarily result in a high EE. In other words, there exists EE optimal network size.
  - The L-DAS complied with the AS, UC, precoding, and power control improves EE compared to non-clustering L-DAS and colocated antenna systems.

## II. L-DAS SYSTEM AND ITS ISSUES

We consider an L-DAS with one central unit, called a baseband unit (BBU) or a signal processing center,  $M$  DAs, and  $U$

user equipments (UEs) (refer to Fig. 2(a) in Subsection IV-C). L-DAS has a very large number of DAs compared to the number of UEs, i.e.,  $M \gg U$ . For simple demonstration, a grid antenna layout is depicted in Fig. 2(a), yet any type of antenna layouts, such as circular and random layouts, can be applied to our system model<sup>1</sup>. All DA ports are connected to the BBU through a noise-free wired fronthaul for coordinated and cooperative communications. Since a passive optical network (PON) can support data rate up to 2.4 Gbps with low power consumption of around 1 W per subscriber [22], PON can be one possible implementation of the optical fronthaul in L-DAS.

Particularly, the BBU consists of a baseband module and  $M$  RF modules, each of which consists of an electric RF (eRF) module and an optical RF (oRF) module as shown in Fig. 1. The baseband module performs various digital signal processing, such as digital up conversion, channel encoding, and modulation (see more examples in Fig. 1), and distributes the digital signals to  $M$  eRF modules. Each eRF module converts the digital electric-signals to analogue electric-signals, which are subsequently converted to optic-signals through an electric-to-optic (E/O) converter at the oRF module. The oRF module is connected to the remote oRF module at DA port through optical fiber. Each DA port consists of the remote oRF and eRF modules. The remote oRF module converts optic-signals to electric-signals via an optic-to-electric (O/E) converter, and the eRF module simply emits the electric-signals through a variable gain amplifier (VGA), a driver, and a power amplifier (PA) followed by RF antenna. The power consumption of each module is modeled precisely in Section III.

Since an L-DAS is characterized with the very large number of DAs, the following issues will need to be addressed.

- *Processing delay (computational complexity)*: It occurs from the centralized, huge computation at the BBU. A cloud-processing at a cloudlet can support a BBU of L-DAS to relieve the high processing delay [23]. Through the proposed AS and UC methods in Section IV, parallel

<sup>1</sup>A generalized DAS employing multiple distributed transmitters with multiple colocated antennas, {e.g., CoMP [18], [19], or employing multiple DAs [17], [20], [21]} can be exploited as the L-DAS. Note that multiple transmitters can be implemented by multiple DAs with identical power amplifiers, which is a reasonable approach based on the results in [21], which reports that *equal* power output capability (POC) of transmitters provides further EE merit compared to the unequal POC.

processing per cluster is possible, resulting in complexity and processing delay reduction.

- *Fronthaul delay*: Distributing signals from BBU to DA ports experiences a fronthaul delay. The fronthaul delay is different from a backhaul delay, which is caused by the limited backhaul capacity and signaling among different cells [18], [19], as it occurs within the same cell. Significant fronthaul delay can be avoided from delay-tolerant optical fiber (around 5  $\mu$ s/km) and a residual optical delay tuner (up to 15 ns) [24].
- *Propagation delay*: There is a delay in propagation of transmit signal from DA port to UE. Due to the large number of DAs distributed over the coverage area, we can almost always find the proper DA which is sufficiently close to the UE such that the propagation delay is negligible. With 25 to 900 DAs, *intra-antenna distance* (IAD) that is a minimum distance of neighboring DAs varies from 30 m to 200 m (refer to Section VII), which is the coverage of small cells, such as pico and femto cells, where the propagation delay is not critical issues.
- *Feedback (signaling) overhead*: For the MU-MIMO, the required amount of channel state information (CSI) is enormous due to the large number of DAs (refer to Subsection IV-C). The burden of severe signaling overhead can be effectively reduced from the proposed AS and UC followed by the cluster-based parallel processing.
- *Energy consumption*: To activate many DAs, high energy or power consumption is expected. However, as we pointed out in Section I, the energy consumption behavior of L-DAS is unclear. Throughout the paper, we focus on EE characterization of L-DAS and propose EE-improving baseband algorithms.
- *Remaining issues*: A synchronization issue is tightly related to the delay issues. To completely resolve the recognized delay issues and synchronization issues is out of scope of our work. Please refer to Section VIII for other notable, remaining issues.

Parts of issues of L-DAS, such as high computational complexity, resource consumption for feedback overhead, and energy consumption, have been rigorously considered in the paper. For convenience, some notations and symbols used throughout the paper are listed in Table I.

### III. EE MAXIMIZATION PROBLEM FORMULATION

For simplicity, we assume that i) each UE has a single receive antenna and ii) any channel matrix constructed from the selected DAs is full rank. Denoting a received signal at UE  $u$  by  $y_u$ , its vector form  $\mathbf{y} = [y_1 \cdots y_U]^T$  is written as

$$\mathbf{y} = \mathbf{H}(\mathbf{S} \circ \mathbf{W})\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{n}, \quad (1)$$

where  $\mathbf{H}$  is a  $U$ -by- $M$  MU-MIMO channel matrix<sup>2</sup>;  $\mathbf{S}$  is an  $M$ -by- $U$  binary, transmit AS matrix whose  $(m, u)$ th element  $s_{mu} = 1$  if the  $m$ th DA is selected for UE  $u$ , and  $s_{mu} = 0$

<sup>2</sup>The channels are assumed to be sufficiently static for MU-MIMO precoding, i.e., large coherence time of channels. This assumption is supported by realizing a user association or scheduling, which groups  $U$  users who move with a low mobility less than 1 km/h for example.

otherwise;  $\mathbf{W}$  is an  $M$ -by- $U$  precoding matrix;  $\mathbf{P}$  is a  $U$ -dimensional diagonal matrix whose  $u$ th diagonal element  $p_{uu}$  determines a power portion assigned to UE  $u$ ;  $\mathbf{x} = [x_1 \cdots x_U]^T$  is a transmit signal vector where  $x_u$  is a transmit symbol to UE  $u$  with  $E|x_u|^2 = 1$ ; and  $\mathbf{n} = [n_1 \cdots n_U]^T$  is an additive white Gaussian noise (AWGN) vector whose  $u$ th element  $n_u$  is an AWGN at UE  $u$  and obeys the complex normal distribution with a zero mean and a  $\sigma^2$  variance, i.e.,  $\mathcal{CN}(0, \sigma^2)$ . The  $(u, m)$ th element of  $\mathbf{H}$  represents a channel gain  $\sqrt{A_{um}}h_{um}$  consisting of the path loss  $\sqrt{A_{um}}$  and the small scale fading  $h_{um}$  between DA  $m$  and UE  $u$ . The channels  $\{h_{um}\}$  are assumed to be independent and identically distributed (i.i.d.).

The received signal-to-interference-plus-noise ratio (SINR) of UE  $u$  is derived from  $y_u$  in (1) as

$$\begin{aligned} \text{SINR}_u(\mathbf{S}, \mathbf{W}, \mathbf{P}) &= \frac{E|\mathbf{h}_u^r(\mathbf{s}_u^c \circ \mathbf{w}_u^c)\sqrt{p_{uu}}x_u|^2}{\sum_{u'=1, u' \neq u}^U E|\mathbf{h}_u^r(\mathbf{s}_{u'}^c \circ \mathbf{w}_{u'}^c)\sqrt{p_{u'u'}}x_{u'}|^2 + \sigma^2} \\ &= \frac{|\mathbf{h}_u^r(\mathbf{s}_u^c \circ \mathbf{w}_u^c)|^2 p_{uu}}{\sum_{u'=1, u' \neq u}^U |\mathbf{h}_u^r(\mathbf{s}_{u'}^c \circ \mathbf{w}_{u'}^c)|^2 p_{u'u'} + \sigma^2}. \end{aligned} \quad (2)$$

Under sufficient input backoff assumption, a PA input signal is linearly amplified and the PA output signal has a Gaussian distribution [12]. Hence, we can further assume that UE  $u$  can achieve throughput over bandwidth  $\Omega$  Hz as<sup>3</sup>

$$R_u(\mathbf{S}, \mathbf{W}, \mathbf{P}) = \Omega \log_2(1 + \text{SINR}_u(\mathbf{S}, \mathbf{W}, \mathbf{P})), \forall u \in \mathcal{U}.$$

The system throughput per unit time (bits/sec) is then written as

$$R(\mathbf{S}, \mathbf{W}, \mathbf{P}) = \sum_{u \in \mathcal{U}} R_u(\mathbf{S}, \mathbf{W}, \mathbf{P}). \quad (3)$$

Now, we propose a power consumption model which coincides with signal model (1) and captures the effect of the core design factors of L-DAS, such as  $U$ ,  $M$ , and maximum output power of DA. The power consumption function  $C(\cdot)$  of L-DAS transmitter is basically modeled as two parts as

$$C(\mathbf{S}, \mathbf{W}, \mathbf{P}) = f(\mathbf{S}, \mathbf{W}, \mathbf{P}) + g(\mathbf{S}, \mathbf{W}), \quad (4)$$

where the first part  $f(\cdot)$  is a transmit power dependent (TPD) term and the second part  $g(\cdot)$  is a transmit power independent (TPI) term (refer to Fig. 1).

The TPD power consumption is the sum of all TPD terms of DAs as

$$f(\mathbf{S}, \mathbf{W}, \mathbf{P}) = c \sum_{m \in \mathcal{M}} \frac{1}{\eta_m} [(\mathbf{S} \circ \mathbf{W})\mathbf{P}(\mathbf{S} \circ \mathbf{W})^H]_{mm}, \quad (5)$$

where  $c$  is a system dependent power loss coefficient ( $c > 1$ ), which can be empirically measured,  $\eta_m$  is the efficiency of PA at the  $m$ th DA ( $0 < \eta_m < 1$ ), and the  $(m, m)$ th element of

<sup>3</sup>We consider the system performance after equalization. The equalizer can be located at the receiver or the transmitter in the form of orthogonal frequency-division multiplexing (or similar techniques), and is assumed to be able to remove the effects of inter-symbol interference (ISI) over any frequency band. The ISI therefore is ignored from the performance evaluation of our system with arbitrary  $\Omega$ .

the matrix inside a bracket is the transmit power of DA  $m$ . The average transmit power of DA  $m$  is derived from (1) as

$$\begin{aligned} & \mathbb{E} \left| (\mathbf{s}_m^r \circ \mathbf{w}_m^r) \sqrt{\mathbf{P}} \mathbf{x} \right|^2 \\ &= \mathbb{E} \left( (\mathbf{s}_m^r \circ \mathbf{w}_m^r) \sqrt{\mathbf{P}} \mathbf{x} \mathbf{x}^H \sqrt{\mathbf{P}} (\mathbf{s}_m^r \circ \mathbf{w}_m^r)^H \right) \\ &= (\mathbf{s}_m^r \circ \mathbf{w}_m^r) \mathbf{P} (\mathbf{s}_m^r \circ \mathbf{w}_m^r)^H \\ &= [(\mathbf{S} \circ \mathbf{W}) \mathbf{P} (\mathbf{S} \circ \mathbf{W})^H]_{mm}. \end{aligned}$$

The TPI power consumption is modeled as

$$g(\mathbf{S}, \mathbf{W}) = g_{\text{rf}}(\mathbf{S}) + g_{\text{bb}}(\mathbf{W}) + g_{\text{net}}(M) + P_{\text{fix}}, \quad (6)$$

where  $g_{\text{rf}}(\mathbf{S})$  is the power consumption of an RF circuit, which is proportional to the number of RF chains and depends on the type of eRF and oRF modules;  $g_{\text{bb}}(\mathbf{W})$  is the power consumption of signal-processing at BBU, which depends on baseband processing including precoding  $\mathbf{W}$ ;  $g_{\text{net}}(M)$  is the overhead power consumption for signaling to obtain CSI, which depends on network size  $M$ ; and  $P_{\text{fix}}$  is the fixed power consumption including a part of power consumption at, for example, a power supply, an alternating current to direct current (AC/DC) converter, a DC/DC converter, and an active cooling system at BBU and/or DAs. We further precisely model  $g_{\text{rf}}(\mathbf{S})$ ,  $g_{\text{bb}}(\mathbf{W})$ , and  $g_{\text{net}}(M)$  in (6) as follows (refer to Fig. 1):

$$g_{\text{rf}}(\mathbf{S}) = \sum_{m \in \mathcal{M}} (P_{\text{cc1},m} + P_{\text{cc2},m} \sum_{u \in \mathcal{U}} R_u) \max_u s_{mu}, \quad (7a)$$

$$g_{\text{bb}}(\mathbf{W}) = \Omega P_{\text{sp1}} [\dim(\mathbf{W})]^{\beta+1} + \Omega P_{\text{sp2}}, \quad (7b)$$

$$g_{\text{net}}(M) = M \Omega P_{\text{sig}}. \quad (7c)$$

In (7a),  $P_{\text{cc1},m}$  is the power consumption of eRF module at BBU for DA  $m$ , which includes digital-to-analogue (D/A) converter, filters, synthesizer, and mixer;  $P_{\text{cc2},m}$  is the power consumption per unit-bit-and-second of oRF modules connected the  $m$ th fiber line, which includes possibly modulator driver, laser, optical amplifier, and E/O and O/E converters; and  $R_u$  is a target rate of UE  $u$ .

In (7b), the first term is proportional to the number of active RF chains with order of  $\beta \geq 0$ . The active RF-chain number is the same as the dimension (number of columns) of precoding matrix  $\mathbf{W}$ . The exponent  $\beta$  implies the overhead power consumption of MU processing compared to SU processing. For example, if  $\beta = 0$ , there is no overhead for MU-MIMO signal processing computation<sup>4</sup>. If  $\beta > 0$ , MU-MIMO signal processing computation consumes relatively higher power than SU signal processing. The maximum of exponent  $\beta$  is assumed to be no greater than two as the computational complexity for  $m$ -dimensional MU-MIMO precoding, e.g., zero-forcing (ZF) MU-MIMO precoding, is roughly  $\mathcal{O}(m^3)$ , while that for SU is

<sup>4</sup>This ( $\beta = 0$ ) is the same power consumption model as [17], [21], in which the high power consumption for MU-MIMO precoding was not addressed. For example, if  $\beta = 0$  and  $\dim(\mathbf{W}) = 4$ , there is no difference of signal processing power consumption between four-individual SU processing and single 4-by-4 MU processing.

$\mathcal{O}(m)$ ; therefore,  $0 \leq \beta \leq 2$  is a reasonable assumption. In the second term of (7b),  $P_{\text{sp2}}$  is the signal processing related power consumption per unit frequency at the baseband module, which is independent of the number of active RF chains.

In (7c),  $P_{\text{sig}}$  is the signaling power consumption for channel estimation per frequency at the baseband module, which depends on a network size. Here, we simply model the network size as a linear function  $M$ , which actually depends on network topology. The reasonable value of  $P_{\text{sig}}$  is assumed to be between 0.5% and 50% of  $P_{\text{sp1}}$  (see Table II in Section VII).

From (3) and (4), we express a *system* EE (not the sum of *per-user* EE) as a function of  $\mathbf{S}$ ,  $\mathbf{W}$ , and  $\mathbf{P}$  as

$$\text{EE}(\mathbf{S}, \mathbf{W}, \mathbf{P}) \triangleq \frac{\text{R}(\mathbf{S}, \mathbf{W}, \mathbf{P})}{\text{C}(\mathbf{S}, \mathbf{W}, \mathbf{P})},$$

and formulate the EE maximization problem as follows:

$$\max_{\{\mathbf{S}, \mathbf{W}, \mathbf{P}\}} \text{EE}(\mathbf{S}, \mathbf{W}, \mathbf{P}) \quad (8a)$$

$$\text{s.t.} \quad [(\mathbf{S} \circ \mathbf{W}) \mathbf{P} (\mathbf{S} \circ \mathbf{W})^H]_{mm} \leq P_m, \quad \forall m \in \mathcal{M}, \quad (8b)$$

$$R_u(\mathbf{S}, \mathbf{W}, \mathbf{P}) \geq R_u, \quad \forall u \in \mathcal{U}, \quad (8c)$$

$$p_{u_1 u_2} = 0, \quad \forall u_1 \neq u_2 \in \mathcal{U}, \quad (8d)$$

$$s_{mu} \in \{0, 1\}, \quad \forall m \in \mathcal{M}, \quad \forall u \in \mathcal{U}, \quad (8e)$$

where (8a) is the objective function; (8b) follows a *per-antenna average power* constraint<sup>5</sup>, which is induced by different maximum power capability of the PA, denoted by  $P_m$ , and radio regulations; the inequalities in (8c) are *per-user rate* constraints, i.e., quality-of-service (QoS) constraints; (8d) follows from the diagonal structure of  $\mathbf{P}$ ; and (8e) is for AS.

The problem (8) may be infeasible as (8b) and (8c) give the upper and lower bounds of transmit power, respectively, which may not be satisfied simultaneously for any selected DAs and precoding. In the infeasible case, two possible options may be to i) discard the outage UEs who do not achieve their own target rate, and solve again the new optimization problem with the remaining feasible UE set, and ii) reduce the target rates of the outage UEs. Note that it is difficult to immediately check the feasibility of the problem. Assigning all DAs to all UEs does not guarantee the feasibility. Rather than that, the use of all DAs may increase infeasibility because it increases the dimension of channel matrix, resulting in high probability of ill-conditioned channel matrix. The ill-conditioned channel matrix increases transmit power while the transmitter performs ZF-MU-MIMO precoding. Similar effect on L-MIMO with ZF-MU-MIMO was reported in [3]. If a particular DA exceeds its maximum output power, transmit power of all DAs coupled through the ZF property should be scaled down, resulting in significant throughput degradation. This gives us a strong motivation to consider AS for the L-DAS with ZF-MU-MIMO later.

<sup>5</sup>Per-antenna *instantaneous* transmit power constraint is considered to avoid PA clipping effect in [21], while per-antenna *average* power constraint is considered in [17]. The average transmit power is also an important metric to characterize the transmitter's efficiency and used typically in transmitter design (see e.g., [25]).

TABLE I  
SOME NOTATIONS AND SYMBOLS USED IN THIS PAPER

$a, \mathbf{a}, \mathbf{A}$	scalars, vectors, matrices
$ a , \ \mathbf{a}\ , \ \mathbf{A}\ _F$	absolute value, 2-norm, Frobenius norm
$\dim(\mathbf{A})$	dimension of matrix $\mathbf{A}$
$\mathcal{A},  \mathcal{A} $	sets, cardinality of set
$(\cdot)^T, (\cdot)^H, (\cdot)^\dagger$	transposition, Hermitian transposition, pseudo-inverse
$\text{tr}(\cdot), E(\cdot)$	trace of matrix, expectation of a random variable
$\mathbf{a}_1 \circ \mathbf{a}_2, \mathbf{A}_1 \circ \mathbf{A}_2$	an element-wise product of vectors or matrices
$[\mathbf{a}]_m, [\mathbf{A}]_{mu}$	the $m$ th and $(m, u)$ th elements
$\mathbf{a}_m^c, \mathbf{a}_m^r$	the $m$ th column and row vectors of a matrix $\mathbf{A}$
$(\cdot)^c$ and $(\cdot)^r$	a given constant value, an optimized variable
$U, M, L$	numbers of UEs, DAs, and clusters in network
$u, m, \ell$	indices of UE, DA, and cluster
$U_\ell, M_\ell$	numbers of UEs and DAs allocated to cluster $\ell$
$M_u$	number of DAs assigned to UE $u$
$\mathcal{U}$	set of all UEs, defined as $\{1, \dots, U\} \ni u$
$\mathcal{M}$	set of all DAs, defined as $\{1, \dots, M\} \ni m$
$\mathcal{L}$	set of clusters, defined as $\{1, \dots, L\} \ni \ell$
$\mathcal{U}_\ell, \mathcal{M}_\ell$	sets of UEs and DAs in cluster $\ell$
$\mathcal{M}_u$	set of DAs assigned to UE $u$
$\mathbf{H} \in \mathbb{C}^{U \times M}$	channel matrix
$\mathbf{H}_\ell \in \mathbb{C}^{U_\ell \times M}$	channel matrix for cluster $\ell$
$\mathbf{S} \in \mathbb{R}^{M \times U}$	antenna selection matrix (binary)
$\mathbf{S}_\ell \in \mathbb{R}^{M \times U_\ell}$	antenna selection matrix for cluster $\ell$ (binary)
$\mathbf{W} \in \mathbb{C}^{M \times U}$	precoding matrix
$\mathbf{W}_\ell \in \mathbb{C}^{M \times U_\ell}$	precoding matrix for cluster $\ell$
$\mathbf{P} \in \mathbb{R}^{U \times U}$	power control matrix (diagonal)
$\mathbf{P}_\ell \in \mathbb{R}^{U_\ell \times U_\ell}$	power control matrix for cluster $\ell$ (diagonal)
$\mathbf{y} \in \mathbb{C}^{U \times 1}$	received symbol vector
$\mathbf{x} \in \mathbb{C}^{U \times 1}$	transmit symbol vector
$\mathbf{n} \in \mathbb{C}^{U \times 1}$	AWGN vector
$P_m$	maximum output power of DA $m$
$R_u$	target rate of UE $u$

Obtaining  $\{\mathbf{S}, \mathbf{W}, \mathbf{P}\}$  jointly by directly solving (8) is difficult due to i) the non-convex objective function and constraints<sup>6</sup>, ii) the integer optimization variables  $\{s_{mu}\}$ , and iii) the enormous computational complexity of handling the very large size of matrices  $\{\mathbf{S}, \mathbf{W}, \mathbf{P}\}$ . Moreover, obtaining the full CSI, i.e.,  $U$ -by- $M$  complex-valued matrix  $\mathbf{H}$ , at the BBU is a burden for the network and resource management. Therefore, instead of solving (8) directly, we propose i) a suboptimal, cluster-based decomposition approach by determining  $\mathbf{S}$  in Section IV, and ii) further decomposition of the per-cluster subproblem into two optimization problems to find  $\mathbf{W}$  and  $\mathbf{P}$  in Sections V and VI, respectively.

#### IV. ANTENNA SELECTION AND USER CLUSTERING

To resolve the complexity issues of solving (8), we proposed the antenna selection (AS) and user clustering (UC) methods which enable us to decompose (8) into multiple subproblems by a cluster. We cluster the UEs (or equivalently the selected DAs) based on an SINR threshold such that inter-cluster interferences (ICIs) are small enough to split the original optimization problem (8) into the cluster-based subproblems. Consequently, cluster-based parallel computation and feedback can reduce computational complexity and feedback (or signaling) information, and make an L-DAS practicable.

<sup>6</sup>Contrary to the *power minimization* problem in [5], which can be relaxed to second-order cone program and solved efficiently with the branch-and-cut (BnC) method, (8) is still non-convex even after the continuous relaxation. Furthermore, since computable upper bound for problem (8) is unavailable, BnC type of methods is not applicable to (8).

#### A. Antenna Selection (AS) Algorithms

As mentioned in the previous section, AS could be a crucial strategy to improve EE of the L-DAS with ZF-MU-MIMO. Moreover, our previous studies on a fundamental SE-EE tradeoff [12] motivate us to consider an AS strategy, which can control SE-EE tradeoff and achieve the Pareto optimal tradeoff [11], [17], [21], [26], [27].

There are a few heuristic algorithms for the AS, such as channel norm based (CNB) greedy, precoding norm based (PNB) greedy, and power consumption based (PCB) greedy algorithms [17]. However, the CNB-, PNB-, and PCB-greedy algorithms are irrelevant for the large-size network. The three greedy methods initially assign all DAs and discard one DA in each iteration sequentially; thus, they prohibitively require high computational complexity as  $M$  or  $U$  increases. Precisely, the time complexities of CNB-, PNB-, and PCB-greedy algorithms are  $\mathcal{O}(MU^3)$ ,  $\mathcal{O}(MU^3)$ , and  $\mathcal{O}(M^2U^3)$ , respectively [17]. To circumvent the high complexity, we propose two simple-yet-effective AS algorithms that determine the set of DAs assigned to UE  $u$ , denoted by  $\mathcal{M}_u$ , and the corresponding AS matrix  $\mathbf{S}$ .

1) *Channel-Gain-Based (CGB)-Greedy AS Algorithm*: The CGB-greedy algorithm assigns each UE to a single DA based on the channel gain, e.g., received signal strength indicator (RSSI) used for 3GPP-LTE [28]. In other words, a UE and a DA are paired with each other whose channel gain is as large as possible. Since the BBU is able to differentiate RSSIs received from different DAs, no additional resource is required.

Let  $M_u$  be a predetermined number of DAs that are supposed to be assigned to UE  $u$ . We assign DA  $m$  to UE  $u$  whose channel gain is the strongest, and discard the DA  $m$  from the subsequent allocation procedure. If  $M_u$  DAs are assigned to UE  $u$ , discard the UE  $u$  from the subsequent allocation procedure. This allocation procedure is repeated until all UEs are discarded.

Note that full CSI is not required for the AS procedure, and that the original AS is an  $\mathcal{O}(2^{MU})$  combinatorial problem, yet the greedy AS algorithm requires only  $\mathcal{O}(U)$  time complexity with  $\mathcal{O}(MU \log(MU))$  for sorting  $MU$  channel gains. Thus, the computational complexity is also reduced dramatically compared to the existing greedy algorithms. The CGB-greedy AS algorithm is summarized in Algorithm 1.

#### Algorithm 1: CGB/MDB-greedy AS algorithm

1. Initial setup:  $\mathcal{U} = \{1, \dots, U\}$ ,  $\mathcal{M} = \{1, \dots, M\}$ ,  $s_{mn} = 0, \forall m \in \mathcal{M}, \forall u \in \mathcal{U}, \mathcal{M}_u = \emptyset, \forall u \in \mathcal{U}$ , and given  $M_u$ 's.
2. **while**  $\mathcal{U} \neq \emptyset$  **do**
3.   find  $\{m^*, u^*\} = \arg \text{metric}_{m \in \mathcal{M}, u \in \mathcal{U}}$ .  
   `metric' = max  $|h_{um}|$  for CGB, min  $d_{um}$  for MDB
4.   set  $s_{m^*u^*} = 1$ ,  $\mathcal{M} = \mathcal{M} \setminus m^*$  and  $\mathcal{M}_{u^*} = \mathcal{M} \cup m^*$
5.   **if**  $|\mathcal{M}_{u^*}| = M_{u^*}$  **then**
6.      $\mathcal{U} = \mathcal{U} \setminus u^*$
7.   **end if**
8. **end while**

2) *Minimum-Distance-Based (MDB)-Greedy AS Algorithm*: Instead of channel gains, distance information between DA  $m$  and UE  $u$ , i.e.,  $d_{um}$ , can be considered for the AS. This strategy is typically considered for a simple system without preprocessing at the transmitter, to reduce signaling information and

backhaul overhead [7]. If the BBU has the location information of UEs with free, the MDB-greedy method can simplify the AS procedure as the BBU can omit the RSSI detection procedure. This motivates us to consider a simple MDB-greedy algorithm. However, if BBU needs to perform the localization, the signaling power consumption model in (7c), which is designed for CGB-greedy algorithm, should be modified. To observe the effect of signaling power consumption on EE, we consider various values of  $P_{\text{sig}}$  for EE evaluation in Section VII.

The basic structure of MDB-greedy algorithm is the same as the CGB-greedy algorithm except the metric at line 3 in Algorithm 1.

### B. User Clustering (UC) Algorithm

After the AS, DAs are dedicated to the specific UEs according to  $\mathcal{M}_u$ , and we can then define a minimum SINR between two UEs  $u$  and  $u'$  as (9) at the bottom of this page. The SINR in (9) is defined under the assumptions of maximum transmit power and maximum ratio combining, and it can be interpreted as a distance metric between UE  $u$  and  $u'$  for UC. Based on the distance (not physical distance) between UEs, we cluster UEs who are close to one another. In other words, if two UEs are located too close to each other and the corresponding SINR is too low due to the strong inter-user-interferences (IUIs), we cluster and support them by using MU-MIMO precoding.

Denote the set of UEs in cluster  $\ell$  by  $\mathcal{U}_\ell$  such that  $\bigcup_\ell \mathcal{U}_\ell = \mathcal{U}$  and  $\mathcal{U}_\ell \cap \mathcal{U}_{\ell'} = \emptyset$  where  $\ell \neq \ell'$  and  $|\mathcal{U}_\ell| = U_\ell$ . With a given minimum distance  $\gamma$ , UE  $u'$  will be included to cluster  $\ell$  if their distance is shorter than  $\gamma$  as follows:

$$\mathcal{U}_\ell = \mathcal{U}_\ell \cup \{u'\}, \text{ if } D(\mathcal{U}_\ell, u') \leq \gamma, \quad (10)$$

where the distance metric between two clusters is defined as

$$D(\mathcal{U}_\ell, \mathcal{U}_{\ell'}) \triangleq \min_{u \in \mathcal{U}_\ell, u' \in \mathcal{U}_{\ell'}} d(u, u'). \quad (11)$$

The complexity of the UC is linear, e.g., between  $\mathcal{O}(U^2)$  and  $\mathcal{O}(U^3)$  for a hierarchical clustering algorithm [29]. Denoting a cluster set by  $\mathcal{L} = \{1, \dots, L\}$ , the proposed UC algorithm is summarized in Algorithm 2.

---

#### Algorithm 2: User clustering (UC) algorithm

---

1. Initial setup: distance  $d_o = 0$ , clusters  $\mathcal{U}_\ell = \{\ell\}$  where  $\ell \in \mathcal{L} = \{1, \dots, U\}$ , and given  $\gamma$ .
  2. **while**  $d_o < \gamma$  **do**
  3. find the distance of most closest pair of clusters  $\mathcal{U}_\ell$  and  $\mathcal{U}_{\ell'}$ , i.e.,  $d_o = \min_{\ell, \ell' \in \mathcal{L}, \ell \neq \ell'} D(\mathcal{U}_\ell, \mathcal{U}_{\ell'})$  in (11).
  4. **if**  $d_o < \gamma$  **then**
  5. merge clusters as  $\mathcal{U}_\ell = \mathcal{U}_\ell \cup \mathcal{U}_{\ell'}$ .
  6. update  $\mathcal{L}$ .
  7. **end if**
  8. **end while**
- 

### C. Cluster-Based Subproblems

Once we complete the UC with  $\gamma$ , the ICIs are suppressed such that the minimum SINR in (9) is greater than  $\gamma$ . Assuming sufficiently high  $\gamma$  and the correspondingly negligible ICIs<sup>7</sup>, the SINR (2) of UE  $u$  in a cell is rewritten with only IUIs within the cluster  $\ell$  as

$$\text{SINR}_u^\ell(\mathbf{S}_\ell, \mathbf{W}_\ell, \mathbf{P}_\ell) = \frac{|\mathbf{h}_u^r(\mathbf{s}_u^c \circ \mathbf{w}_u^c)|^2 p_{uu}}{\sum_{u' \in \mathcal{U}_\ell, u' \neq u} |\mathbf{h}_u^r(\mathbf{s}_{u'}^c \circ \mathbf{w}_{u'}^c)|^2 p_{u'u'} + \sigma^2}, \quad (12)$$

where  $\mathbf{S}_\ell \in \mathbb{R}^{M \times U_\ell}$ ,  $\mathbf{W}_\ell \in \mathbb{C}^{M \times U_\ell}$ , and  $\mathbf{P}_\ell \in \mathbb{R}^{U_\ell \times U_\ell}$  are per-cluster AS, precoding, and power control matrices, respectively.  $\mathbf{S}_\ell$  and  $\mathbf{W}_\ell$  consist of column vectors  $\mathbf{s}_u^c$  and  $\mathbf{w}_u^c$ , respectively, where  $u \in \mathcal{U}_\ell$ .  $\mathbf{P}_\ell$  is a diagonal matrix whose diagonal elements consist of  $p_{uu}$ , where  $u \in \mathcal{U}_\ell$ . Hence, the per-cluster EE is defined accordingly as

$$\text{EE}(\mathbf{S}_\ell, \mathbf{W}_\ell, \mathbf{P}_\ell) \triangleq R(\mathbf{S}_\ell, \mathbf{W}_\ell, \mathbf{P}_\ell) / C(\mathbf{S}_\ell, \mathbf{W}_\ell, \mathbf{P}_\ell), \quad (13)$$

where, from (3)–(7),

$$\begin{aligned} R(\mathbf{S}_\ell, \mathbf{W}_\ell, \mathbf{P}_\ell) &= \sum_{u \in \mathcal{U}_\ell} \Omega \log_2 \left( 1 + \text{SINR}_u^\ell(\mathbf{S}_\ell, \mathbf{W}_\ell, \mathbf{P}_\ell) \right) \\ C(\mathbf{S}_\ell, \mathbf{W}_\ell, \mathbf{P}_\ell) &= c \sum_{m \in \mathcal{M}_\ell} \frac{1}{\eta_m} [(\mathbf{S}_\ell \circ \mathbf{W}_\ell) \mathbf{P}_\ell (\mathbf{S}_\ell \circ \mathbf{W}_\ell)^H]_{mm} \\ &\quad + \sum_{m \in \mathcal{M}_\ell} (P_{cc1,m} + P_{cc2,m} \sum_{u \in \mathcal{U}_\ell} R_u) \max_{u \in \mathcal{U}_\ell} s_{mu} \\ &\quad + \Omega P_{\text{sp1}} [\dim(\mathbf{W}_\ell)]^{\beta+1} + M_\ell \Omega P_{\text{sig}} \\ &\quad + \Omega P_{\text{sp2}} / L + P_{\text{fix}} / L. \end{aligned}$$

Here,  $M_\ell$  is the number of DAs assigned in cluster  $\ell$  and  $\mathcal{M}_\ell$  is the set of  $M_\ell$  DAs assigned to cluster  $\ell$ .

Using  $\mathbf{S}_\ell^*$  obtained from AS to (13), we split the original problem (8) into multiple subproblems by the cluster  $\ell \in \mathcal{L}$  as

$$\max_{\{\mathbf{P}_\ell, \mathbf{W}_\ell\}} \text{EE}(\mathbf{S}_\ell^*, \mathbf{W}_\ell, \mathbf{P}_\ell) \quad (14a)$$

$$\text{s.t. } [(\mathbf{S}_\ell^* \circ \mathbf{W}_\ell) \mathbf{P}_\ell (\mathbf{S}_\ell^* \circ \mathbf{W}_\ell)^H]_{mm} \leq P_m, \forall m \in \mathcal{M}_\ell, \quad (14b)$$

$$R_u(\mathbf{S}_\ell^*, \mathbf{W}_\ell, \mathbf{P}_\ell) \geq R_u, \forall u \in \mathcal{U}_\ell, \quad (14c)$$

$$p_{u_1 u_2} = 0, \forall u_1 \neq u_2 \in \mathcal{U}_\ell. \quad (14d)$$

<sup>7</sup>For example, there are no ICIs if  $\gamma = \infty$ , as there is only one cluster. In the case, supporting all UEs with MU-MIMO through the DAs in the cluster yields different (long) propagation delays and synchronization issues. However, with a typical  $\gamma$  between 20 dB and 30 dB, both assumptions of negligible ICI and short propagation hold, because i) neighboring UEs typically consist a cluster, ii) the clusters are generally separated far away from one another for the high EE as shown in our numerical results in Section VII, and iii) there is no joint beamforming (i.e., MU-MIMO precoding) among the clusters.

---


$$d(u, u') \triangleq \min \left\{ \frac{\sum_{m \in \mathcal{M}_u} |h_{um}|^2 P_m}{\sigma^2 + \sum_{m' \in \mathcal{M}_{u'}} |h_{um'}|^2 P_{m'}}, \frac{\sum_{m' \in \mathcal{M}_{u'}} |h_{u'm'}|^2 P_{m'}}{\sigma^2 + \sum_{m \in \mathcal{M}_u} |h_{u'm}|^2 P_m} \right\}, \quad (9)$$

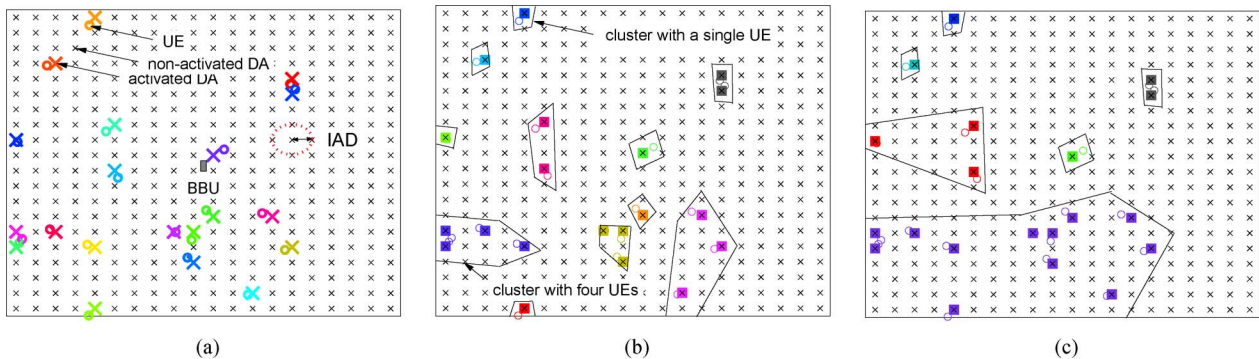


Fig. 2. Illustration of AS and UC for 20 UEs ( $U = 20$ ) and 400 distributed antennas (DAs). (a) AS through a CGB/MDB-greedy algorithm with  $M_u = 1, \forall u \in \mathcal{U}$ . (b) Clustering ( $L = 11$ ) through Algorithm 2 with  $\gamma = 25$  dB. (c) Clustering ( $L = 6$ ) through Algorithm 2 with  $\gamma = 32$  dB.

Since the subproblem (14) can be solve in parallel over  $\ell$ , the computational complexity of the original problem can be significantly reduced. Furthermore, since  $\mathbf{h}_u^r(\mathbf{s}_u^c \circ \mathbf{w}_u^c) = \sum_{m \in \mathcal{M}_\ell} [\mathbf{h}_u^r]_m [\mathbf{w}_u^c]_m$  in (12), the required CSI for cluster  $\ell$  is the cluster-based channel matrix instead of the full channel matrix  $\mathbf{H}$ . Let  $\mathbf{H}_\ell \in \mathbb{C}^{U_\ell \times M}$  be a channel matrix of cluster  $\ell$  that consists of row vectors  $\mathbf{h}_u^r, u \in \mathcal{U}_\ell$ . Then, precisely, the required CSI information to solve (14) is only the  $m$ th columns of  $\mathbf{H}_\ell$ , where  $m \in \mathcal{M}_\ell$ . In other words, we need only  $M_\ell U_\ell$  complex values for the cluster  $\ell$ . The remaining values in other columns of  $\mathbf{H}_\ell$  can be set to be zeros as they will be discarded by AS matrix  $\mathbf{S}_\ell$  during the optimization. Hence, the feedback (or signaling) information can be significantly reduced from  $MU$  complex values for  $\mathbf{H}$  to  $\sum_\ell M_\ell U_\ell$  complex values for  $\{\mathbf{H}_1, \dots, \mathbf{H}_L\}$ . In the example of Figs. 2(b) and (c), 48 and 160 complex values are required for the CSIs, respectively, which is huge reduction compared with the 8 000 complex values for a naive system without AS and UC.

From Jensen's inequality, i.e.,  $\text{EE}(\mathbf{S}^*, \mathbf{W}, \mathbf{P}) \leq \sum_\ell \text{EE}(\mathbf{S}_\ell^*, \mathbf{W}_\ell, \mathbf{P}_\ell)$ , we see that the per-cluster EE maximization in (14) is equivalent to maximize the EE upper bound of the original problem (8). Thus, optimality loss arises from the EE upper bound maximization in (14). However, the decomposition strategy provides reasonable performance if the bound is sufficiently tight, and the tightness depends on  $\gamma$  as follows. As the clustering threshold  $\gamma$  increases, cluster size increases and the number of clusters decreases to one (see Fig. 2). In the case, (14a) is identical to the original objective function (8a) without any optimality loss. On the other hand, as  $\gamma$  decreases, the number of clusters increases resulting in optimality loss. To mitigate the optimality loss, we consider adaptive algorithms for the number of assigned DAs to each UE  $M_u$  and the clustering threshold  $\gamma$ .

#### D. Adaptive Algorithms for $M_u$ and $\gamma$

To mitigate the optimality loss and avoid the outages, it is worth exploring additional DA assignment to the outage UE if there are unallocated DAs in a network, i.e.,  $\sum_{u \in \mathcal{U}} M_u < M$ . For the additional DA assignment, we increase  $M_u$  at line 1 of Algorithm 1. With a limit of iterations, denoted by  $Q_{AS}$ , the AS adaptation algorithm is summarized in Algorithm 3. In

Algorithm 3, precoding and power control in lines 7 and 8 will be introduced in the subsequent sections.

---

#### Algorithm 3: $M_u$ adaptation algorithm for AS

---

1. Initial setup:  $M_u = 1, \forall u \in \mathcal{U}, q = 0, \gamma$ , and  $Q_{AS} \geq 0$ .
  2. **while** feasibility = 0 &  $q < Q_{AS}$  **do**
  3. AS: Algorithm 1.
  4. UC: Algorithm 2 with a threshold  $\gamma$ .
  5. feasibility = 1
  6. **for** cluster  $\ell = 1, \dots, L$  **do**
  7. precoding: (20).
  8. power control: Algorithm 5 or (28).
  9. **if** power control is infeasible &  $\sum_{u \in \mathcal{U}} M_u < M$
  10. **then** add one additional DA to UE  $u \in \mathcal{U}_\ell$  who has the weakest channel gain, i.e.,  $M_u = M_u + 1$  where  $u \in \mathcal{U}_\ell$  s.t.,  $u = \arg \min_{u \in \mathcal{U}_\ell} |h_{um}|$ .
  11. feasibility = 0 **end if**
  12. **end for**
  13.  $q = q + 1$
  14. **end while**
- 

The clustering threshold  $\gamma$  in (10) can be also adjusted to avoid outage or to further improve EE. If we increase  $\gamma$ , the cluster size will increase, while the number of clusters will decrease. Accordingly, throughput increases due to the reduced actual ICIs after precoding and power allocation in lines 7 and 8, respectively, while the processing complexity will increase due to the enlarged cluster size, i.e., MU-MIMO matrix size (see example in Fig. 2). On the other hand, if we decrease  $\gamma$ , the cluster size will decrease, while the number of clusters will increase. Accordingly, the processing complexity can be decreased due to the parallel processing with small MU or SU matrices, while throughput may decrease due to the increased actual ICIs. Therefore, there exists an optimal threshold  $\gamma$  for achieving EE maximum. If the network has sufficiently high capability to adapt  $\gamma$ , the optimal  $\gamma$  can be found numerically by using for example an one-dimensional line search in Algorithm 4 and a bisection search. From the numerical results in Section VII, we observe the existence of optimal  $\gamma$  and the adaptation of  $\gamma$  can manage the actual ICIs to

improve EE. Depending on the network requirement of the computational complexity and latency, maximum number of adaptations  $Q_C$  will be limited.

---

**Algorithm 4:  $\gamma$  adaptation algorithm for UC**


---

1. Initial setup:  $\gamma, \delta > 0$ , stop = 0,  $q = 0$ , and  $Q_{UC} \geq 0$ .
  2. compute  $EE_p$  with  $\gamma$ : Algorithm 3.
  3. compute  $EE_c$  with  $\gamma = \gamma + \delta$ : Algorithm 3.
  4. **if**  $EE_c > EE_p$  **then**
  5.    $EE_p = EE_c$  and  $\xi = 1$ .
  6. **else**
  7.   compute  $EE_c$  with  $\gamma = \gamma - 2\delta$ : Algorithm 3.
  8.   **if**  $EE_c > EE_p$  **then**  $EE_p = EE_c$  and  $\xi = -1$ .
  9.   **else** stop = 1 **end if**
  10. **end if**
  11. **while** stop = 0 &  $q < Q_C$  **do**
  12.   compute  $EE_c$  with  $\gamma = \gamma + \xi\delta$ : Algorithm 3.
  13.   **if**  $EE_c > EE_p$  **then**  $EE_p = EE_c$ .
  14.   **else** stop = 1 **end if**
  15.    $q = q + 1$
  16. **end while**
- 

## V. PRECODING DESIGN

Multiple UEs in a cluster are supported by MU-MIMO precoding to overcome strong IUIs. Note that a single UE in a cluster is a special case of the MU scenario. We assume that  $\mathbf{P}'_\ell$  will be designed to satisfy (14b)–(14d) for given  $\mathbf{S}_\ell^*$  from AS. Then, the  $\ell$ th subproblem (14) is simply rewritten as

$$\mathbf{W}_\ell^* = \max_{\mathbf{W}_\ell} EE(\mathbf{S}_\ell^*, \mathbf{W}_\ell, \mathbf{P}'_\ell). \quad (15)$$

Since a ZF-based precoding is near optimal with respect to the SE if the signal-to-noise ratio (SNR) is high enough [30], [31], it is employed for the MU-MIMO precoding of DAS [32]. Assuming the ZF-MU-MIMO precoding  $\mathbf{W}_\ell$ , all UEs in  $\mathcal{U}_\ell$  share the selected DAs with one another; therefore, the AS matrix is reconstructed as  $\bar{\mathbf{S}}_\ell = [\mathbf{s}_\ell^{c,*} \dots \mathbf{s}_\ell^{c,*}] \in \mathbb{R}^{M \times U_\ell}$  where  $\mathbf{s}_\ell^{c,*} = \sum_{u \in \mathcal{U}_\ell} \mathbf{s}_u^{c,*}$ . Equivalently, we can write the effective channel matrix as  $\mathbf{H}_\ell (\bar{\mathbf{S}}_\ell \circ \mathbf{W}_\ell) = \mathbf{H}_\ell \mathbf{S}_\ell^d \mathbf{W}_\ell$ , where  $\mathbf{S}_\ell^d = \text{diag}(\mathbf{s}_\ell^{c,*}) \in \mathbb{R}^{M \times M}$  is a diagonal matrix whose diagonal elements are the elements of vector  $\mathbf{s}_\ell^{c,*}$ . The ZF-MU-MIMO precoding cancels perfectly IUIs in the cluster.

Following a general ZF property, the effective channel matrix  $\mathbf{H}_\ell \mathbf{S}_\ell^d \mathbf{W}_\ell$  is a diagonal matrix. Noting that the different values of diagonal elements can be implemented by power control  $p_{uu}$ , without loss of generality (w.l.o.g.), the ZF property is degenerated to  $\mathbf{H}_\ell \mathbf{S}_\ell^d \mathbf{W}_\ell = \mathbf{I}_{U_\ell}$ , where  $\mathbf{I}_a$  is an  $a$ -dimensional identity matrix. Thus, we express the structure of ZF-MU-MIMO precoding matrix as

$$\mathbf{W}_\ell = (\mathbf{H}_\ell \mathbf{S}_\ell^d)^\dagger + \text{null}(\mathbf{H}_\ell \mathbf{S}_\ell^d) \mathbf{A}_\ell, \quad (16)$$

where  $\mathbf{A}_\ell \in \mathbb{C}^{U_\ell \times U_\ell}$  is a  $U_\ell$ -dimensional arbitrary matrix.

Using the structure in (16), we can simplify the per-cluster SINR of UE  $u$  in (12) to an SNR as

$$\text{SNR}_u(\mathbf{P}'_\ell) = \frac{p'_{uu}}{\sigma^2}, u \in \mathcal{U}_\ell, \quad (17)$$

and obtain an optimization problem equivalent to (15):

$$\begin{aligned} \mathbf{W}_\ell^* &= \max_{\mathbf{W}_\ell} EE(\mathbf{S}_\ell^*, \mathbf{W}_\ell, \mathbf{P}'_\ell) \\ &\stackrel{(a)}{\equiv} \min_{\mathbf{W}_\ell} C(\mathbf{S}_\ell^*, \mathbf{W}_\ell, \mathbf{P}'_\ell) \\ &\stackrel{(b)}{\equiv} \min_{\mathbf{W}_\ell} c \sum_{m \in \mathcal{M}_\ell} \frac{1}{\eta_m} [(\mathbf{S}_\ell^* \circ \mathbf{W}_\ell) \mathbf{P}'_\ell (\mathbf{S}_\ell^* \circ \mathbf{W}_\ell)^H]_{mm} \\ &\stackrel{(c)}{\equiv} \min_{\mathbf{W}_\ell} \sum_{m \in \mathcal{M}_\ell} [(\mathbf{S}_\ell^* \circ \mathbf{W}_\ell) \mathbf{P}'_\ell (\mathbf{S}_\ell^* \circ \mathbf{W}_\ell)^H]_{mm} \\ &= \min_{\mathbf{W}_\ell} \left\| \mathbf{S}_\ell^d \mathbf{W}_\ell \sqrt{\mathbf{P}'_\ell} \right\|_F^2, \end{aligned} \quad (18)$$

where (a) follows the fact that rate does not depend on  $\mathbf{W}_\ell$  as the SNR (17) is not a function of  $\mathbf{W}_\ell$ ; (b) follows that  $\mathbf{W}_\ell$  affects only on TPD term for given  $\mathbf{S}_\ell^*$  and  $\mathbf{P}'_\ell$ ; and (c) follows that equal POC is preferred for high EE as reported in [21], and  $c/\eta_m$  is then a constant.

Again, using (16) to (18), we have an optimization problem with respect to  $\mathbf{A}_\ell$  as

$$\mathbf{A}_\ell^* = \min_{\mathbf{A}_\ell} \left\| \mathbf{S}^d (\mathbf{H}_\ell \mathbf{S}_\ell^d)^\dagger \sqrt{\mathbf{P}'_\ell} + \mathbf{S}_\ell^d \text{null}(\mathbf{H}_\ell \mathbf{S}_\ell^d) \mathbf{A}_\ell \sqrt{\mathbf{P}'_\ell} \right\|_F^2, \quad (19)$$

where we use the property of a Frobenius norm that  $\text{tr}(\mathbf{A}_\ell \mathbf{A}_\ell^H) = \|\mathbf{A}_\ell\|_F^2$ . Since the lower bound of the objective function in (19) is obtained when  $\mathbf{A}_\ell$  is a zero matrix (refer to the Appendix in [17]), the EE-aware precoding matrix becomes a conventional ZF-MU-MIMO precoding matrix as

$$\mathbf{W}_\ell^* = (\mathbf{H}_\ell \text{diag}(\mathbf{s}_\ell^{c,*}))^\dagger. \quad (20)$$

For an SU cluster, refer to Remark 1.

*Remark 1:* There is no loss of SE optimality of cluster  $\ell$  when cluster  $\ell$  includes a single UE, i.e.,  $U_\ell = 1$ , because (20) is an optimal beamforming for the SU cluster.

## VI. POWER CONTROL

We now propose a cluster-based power control method. Per-cluster optimal power control methods are proposed for SU and MU clusters. A simple heuristic power control method is also proposed for the MU cluster.

### A. Optimal Power Control for MU Cluster

Consider an MU cluster  $\ell$ , which supports  $U_\ell$  multiple UEs. For given  $\mathbf{S}_\ell^*$  and  $\mathbf{W}_\ell^*$ , which are obtained in Sections IV and V, respectively, (14) is rewritten as

$$\mathbf{P}_\ell^* = \max_{\mathbf{P}_\ell} EE(\mathbf{S}_\ell^*, \mathbf{W}_\ell^*, \mathbf{P}_\ell) \quad (21a)$$

$$\text{s.t. } [(\mathbf{S}_\ell^d \mathbf{W}_\ell^*) \mathbf{P}_\ell (\mathbf{S}_\ell^d \mathbf{W}_\ell^*)^H]_{mm} \leq P_m, \forall m \in \mathcal{M}_\ell, \quad (21b)$$

$$R_u(\mathbf{P}_\ell) \geq R_u, \forall u \in \mathcal{U}_\ell, \quad (21c)$$

$$p_{u_1 u_2} = 0, \forall u_1 \neq u_2 \in \mathcal{U}_\ell. \quad (21d)$$



By introducing an additional variable  $\xi$ , we rewrite (21) as

$$\mathbf{P}_\ell^* = \max_{\mathbf{P}_\ell} \xi \text{ s.t. (21b), (21c), (21c), and } \quad (22a)$$

$$\sum_{u \in \mathcal{U}_\ell} R_u(\mathbf{P}_\ell) \geq \xi C(\mathbf{S}_\ell^*, \mathbf{W}_\ell^*, \mathbf{P}_\ell). \quad (22b)$$

This rewriting of the optimization problem introduces an additional constraint (22b) to the problem. However, for fixed  $\xi$ , the (21b) and (21c) are convex constraints and (21d) is linear constraint; therefore, the feasibility of this optimization problem can be checked through solving a *convex feasibility problem* [33]. This optimization problem is quasi-convex, and the optimal  $\xi$  can be found through sequentially solving the convex feasibility problem at each step of the bisection of  $\xi$ . We present the bisection search in Algorithm 5.

---

**Algorithm 5: Per-cluster optimal power control for MU**


---

1. setup:  $\xi_{LB} = 0$ ,  $\xi_{UB} \simeq \infty$ , and a tolerance value,  $\delta > 0$
  2. **while**  $\xi_{UB} - \xi_{LB} > \delta$  **do**
  3.    $\xi \leftarrow (\xi_{UB} + \xi_{LB})/2$
  4.   Solve convex feasibility problem with constraints (21b), (21c), (21d) and (22b), and find (update)  $\mathbf{P}_\ell^*$ .
  5.   **if** infeasible **then**  $\xi_{UB} \leftarrow \xi$
  6.   **else**  $\xi_{LB} \leftarrow \xi$  **end if**
  7. **end while**
  8.  $\mathbf{P}_{optimal,\ell}^* = \mathbf{P}_\ell^*$
- 

The complexity of Algorithm 5 for cluster  $\ell$  is approximately  $\mathcal{O}(M_\ell^{3.5} \log 1/\epsilon)$ , where  $\epsilon$  is the degree of accuracy we desire in finding the optimal energy efficiency. This complexity analysis follows from the fact that we need to solve, essentially, a semi-definite program at every iteration, which costs about  $\mathcal{O}(M_\ell^{3.5})$ , and the number of required iterations to get the optimal energy efficiency within  $\epsilon$  is  $\log 1/\epsilon$  [33]. To circumvent the high complexity of per-cluster optimal power control, we consider a non-iterative power control method in the next subsection.

### B. Heuristic(Optimal) Power Control for MU(SU) Cluster

For simple closed-form solution, we modify (21). To this end, we decompose the power control matrix as

$$\mathbf{P}_\ell = \alpha_\ell \bar{\mathbf{P}}_\ell, \quad (23)$$

where  $\alpha_\ell$  is a common power scaling factor for power limit of PAs and target rate of UEs in cluster  $\ell$ ;  $\bar{\mathbf{P}}_\ell$  is a diagonal matrix with the diagonal element  $\bar{p}_{u,\ell}$ ; and  $\bar{p}_{u,\ell}$  is the relative power portion of UE  $u \in \mathcal{U}_\ell$ , such that  $p_{u,\ell} = \alpha_\ell \bar{p}_{u,\ell}$  and  $\sum_u \bar{p}_{u,\ell} = 1$ . The relative power portion factors are determined, *heuristically*, based on the minimum required power for the target rate as follows [21]:

$$\bar{p}_{u,\ell} = \frac{\tilde{p}_{u,\ell}}{\sum_{k \in \mathcal{U}_\ell} \tilde{p}_{k,\ell}}, \quad \forall u \in \mathcal{U}_\ell, \quad (24)$$

where  $\tilde{p}_{u,\ell}$  is the minimum required power to satisfy (8c) when ZF-MU-MIMO is employed, and it is derived as

$$\tilde{p}_{u,\ell} = \sigma^2 \left( 2^{\frac{R_u}{\Omega}} - 1 \right).$$

Using (23) and (24) to the power constraint (21b), we have the upper bound of  $\alpha_\ell$  as

$$\alpha_\ell \leq \min_{m \in \mathcal{M}_\ell} \left( P_m / \left[ \mathbf{S}_\ell^d \mathbf{W}_\ell^* \bar{\mathbf{P}}_\ell (\mathbf{S}_\ell^d \mathbf{W}_\ell^*)^H \right]_{mm} \right) \triangleq \alpha_{UB,\ell}. \quad (25)$$

On the other hand, using (23) and (24) to the QoS constraint (21c), we can get the lower bound of  $\alpha_\ell$  as follows:

$$\alpha_\ell \geq \frac{\sigma^2 \left( 2^{\frac{R_u}{\Omega}} - 1 \right)}{\bar{p}_{u,\ell}} = \frac{\tilde{p}_{u,\ell}}{\bar{p}_{u,\ell}} = \sum_{u \in \mathcal{U}_\ell} \tilde{p}_{u,\ell} \triangleq \alpha_{LB,\ell}. \quad (26)$$

Thus, if  $\alpha_\ell$  satisfies (25) and (26), i.e.,  $\alpha_{LB,\ell} \leq \alpha_\ell \leq \alpha_{UB,\ell}$ , any  $\alpha_\ell \bar{\mathbf{P}}_\ell$  satisfies (21b) and (21c).

Now, for simple closed-form solution of (21), we maximize EE lower bound instead of EE. Using (23) and an inequality that  $\sum_{u \in \mathcal{U}_\ell} \log_2(f_u) \geq U_\ell \log_2(\min_u(f_u))$ , problem (21) can be modified to maximize the EE lower bound, which is tight as observed in [21], as follows:

$$\alpha_\ell^* = \arg \max_{\alpha_\ell} \frac{\Omega U_\ell \log_2(1 + c_{1,\ell} \alpha_\ell)}{c_{2,\ell} \alpha_\ell + c_{3,\ell}} \quad (27a)$$

$$\text{s.t. } \alpha_{LB,\ell} \leq \alpha_\ell \leq \alpha_{UB,\ell}, \quad (27b)$$

where  $c_{1,\ell} \triangleq \min_u \{\bar{p}_{u,\ell}\} \sigma^{-2}$ ;  $c_{2,\ell} \triangleq \sum_{m \in \mathcal{M}_\ell} \frac{c}{n_m} [\mathbf{S}_\ell^d \mathbf{W}_\ell^* \bar{\mathbf{P}}_\ell (\mathbf{S}_\ell^d \mathbf{W}_\ell^*)^H]_{mm}$ ; and  $c_{3,\ell} = \sum_{m \in \mathcal{M}_\ell} (P_{cc1,m} + P_{cc2,m} \sum_{u \in \mathcal{U}_\ell} R_u) \max_{u \in \mathcal{U}_\ell} s_{mu} + \Omega P_{sp1} [\dim(\mathbf{W}_\ell)]^{\beta+1} + M_\ell \Omega P_{sig} + \Omega P_{sp2}/L + P_{fix}/L$ . Note that all  $c_{1,\ell}$ ,  $c_{2,\ell}$ ,  $c_{3,\ell}$ ,  $\alpha_{LB,\ell}$ , and  $\alpha_{UB,\ell}$  in (27) are constant values for given  $\mathbf{W}_\ell^*$ ,  $\mathbf{S}_\ell^*$ , and  $\bar{\mathbf{P}}_\ell$ ; and the objective function (27a) is a quasi-concave function over  $\alpha_\ell$ . Therefore, we can readily find the maximizer  $\alpha_{o,\ell}$ , which makes the first derivative of the objective function in (27a) to zero, as

$$\alpha_{o,\ell} = \frac{1}{c_{1,\ell}} \left( \exp \left( 1 + \mathbb{W} \left( \frac{-1}{\exp(1)} + \frac{c_{1,\ell} c_{3,\ell}}{c_{2,\ell} \exp(1)} \right) \right) - 1 \right).$$

Considering the feasible region (27b), we get the optimal feasible solution of (27) as

$$\alpha_\ell^* = [\alpha_{o,\ell}]_{\alpha_{LB,\ell}}^{\alpha_{UB,\ell}},$$

and obtain the heuristic power control matrix as

$$\mathbf{P}_{heuristic,\ell}^* = \alpha_\ell^* \bar{\mathbf{P}}_\ell. \quad (28)$$

Since the solution in (28) is obtained from heuristic approach, namely EE lower bound maximization and fixed  $\bar{\mathbf{P}}_\ell$  in (24), it yields performance degradation compared to  $\mathbf{P}_{optimal,\ell}^*$  in Subsection VI-A. However, it is noticeable that the solution in (28) has a tractable, closed-form expression, and the performance gap is marginal as shown in the next section. Moreover, there is no optimality loss for SU cluster as stated in Remark 2.

*Remark 2:* Since the EE lower bound is identical to the EE of the SU cluster and  $\bar{p}_{u,\ell} = 1$  in (24), there is no optimality loss from power control (28) for the SU cluster.

The heuristic power control method is an  $\mathcal{O}(M_\ell^3)$  algorithm. This can be seen from the fact that the bottleneck procedure includes the multiplication of two  $M_\ell$ -dimensional matrices to find  $c_{2,\ell}$  in (27). Hence, the highest complexity order is  $\mathcal{O}(M_\ell^3)$ .

TABLE II  
 SIMULATION PARAMETERS FOR L-DAS/L-CAS [12], [28], [34], [35]

Cell model	square grid (1 km <sup>2</sup> )
Number of DAs/CAs	25 ≤ M ≤ 900
Intra-antenna distance (IAD)	from 33 m to 200 m
Number of UEs	2 ≤ U ≤ 20
UE distribution	Uniform (10 <sup>4</sup> -realization)
Path loss model (f <sub>c</sub> = 2 GHz)	A <sub>um</sub> = g - 128 + 10 log <sub>10</sub> (d <sub>u,m</sub> <sup>-μ</sup> )
feeder loss and antenna gain	g = 5 dB
Path loss exponent	μ = 3.76
Small scale fading	h <sub>um</sub> ~ C.N.(0, 1)
Bandwidth	Ω = 10 MHz
Target rate	R <sub>u</sub> = 10 Mbps
Maximum Tx power	P <sub>m</sub> = 17 dBm
AWGN standard deviation	σ <sup>2</sup> = -174 dBm/Hz
Power loss coefficient	c = 2.63
eRF circuit pow.cons.	P <sub>cc1,m</sub> = 5.7 W
oRF circuit pow.cons.	P <sub>cc2,m</sub> = 0.5/0 pW/bit/s
Fixed pow.cons.	P <sub>fix</sub> = 34 W
Signal processing pow.cons.	P <sub>sp1</sub> = 0.94 × 1/1.1 μW/Hz P <sub>sp2</sub> = 0.54 × 1/1.1 μW/Hz
Signaling pow.cons./antenna	5 ≤ P <sub>sig</sub> ≤ 500 nW/Hz
preprocessing pow.cons. ratio	0 ≤ β ≤ 2
PA efficiency	η <sub>m</sub> = 0.08/0.6
Clustering threshold	-∞ ≤ γ ≤ ∞ dB

## VII. PERFORMANCE EVALUATION AND DISCUSSION

Computer simulations are conducted to examine the average EE performance of the proposed L-DAS. Since an instantaneous EE is set to be zero when an outage happens, the outage performance is already involved in the average EE performance. The EE performance depends highly on the power consumption models, i.e., the power consumption of TPD and TPI terms in (5) and (6), respectively. The simulation is performed under a typical L-DAS scenario, in which the TPI power consumption is dominant compared to the TPD power consumption due to the low power transmission of DAs. With  $c = 2.63$ , the TPD power consumption is observed to be the portion of the TPI power consumption lower than 3% in our simulation. The PA efficiency<sup>8</sup> of all DAs is set by 8%, i.e.,  $\eta_m = 0.08$ , at the maximum transmit power 17 dBm. The BBU performs the centralized, complex processing as a macro BS, yet DA port covers small areas like the small BSs. Hence, we follow macro BS's power consumption model for BBU, while follow a small cell BS, such as pico and femto BSs, for DA's power consumption model. Refer to Table II for other detailed parameters, which are obtained from recent studies [12], [28], [34], [35]. Note that providing the actual, accurate measurement of the parameters is out of scope of our work.

### A. Average EE Over Clustering Threshold

Fig. 3 shows average EEs over clustering threshold  $\gamma$  with various MU-MIMO-processing power consumption order  $\beta$  when  $M = 400$  and  $U = 20$ . One extreme case with  $\gamma = -\infty$  yields  $U$  SU clusters, each of which includes an SU; thus, it is called a *full SU* scheme. Other extreme case with  $\gamma = \infty$  yields a single MU cluster including all UEs who are supported

<sup>8</sup>High input backoff is desired to avoid nonlinearity at the PA, around 12 dB, because a sophisticated, complex linearization method, e.g., predistortion, is not available the simple DA port. Hence, the PA efficiency is very low.

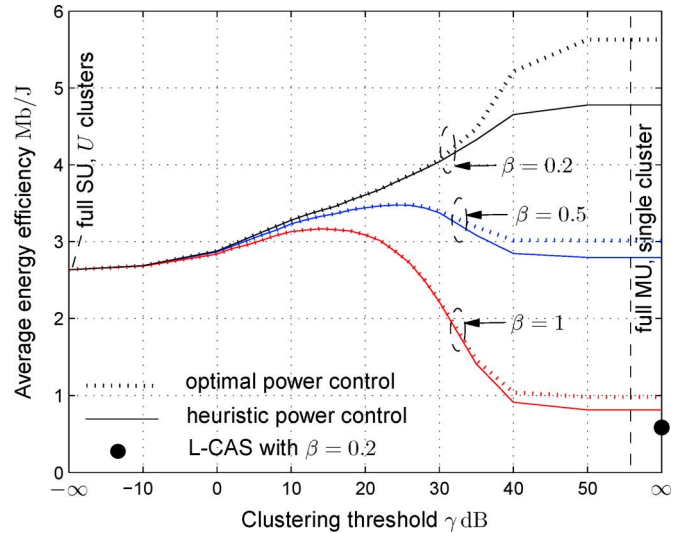


Fig. 3. Average EE over clustering threshold  $\gamma$  with various power consumption models when  $M = 400$ ,  $U = 20$ , and  $P_{\text{sig}} = 50$  nW/Hz.

by ZF-MU-MIMO simultaneously, which is called a *full MU* scheme. If there is less penalty for MU precoding computation compared to SU precoding computation, i.e.,  $\beta = 0.2$ , EE increases as  $\gamma$  increases up to the saturation of EE when  $\gamma = \infty$ . In other words, the full MU scheme achieves the highest EE. This is because MU-MIMO achieves the higher throughput than interference-limited SU scheme and there is small additional power consumption for the MU signal processing. Note that ICIs are inversely proportional to the threshold.

On the other hand, if the power consumption penalty for MU precoding increases, i.e.,  $\beta = 1$ , the overhead MU-MIMO-processing power consumption decreases significantly the EE. We observe that there exists the optimal  $\gamma^*$  and the EE turns to decrease if  $\gamma > \gamma^*$ . In other words, the EE can be severely reduced if there are too many MU clusters in the network. As  $\beta$  increases, MU signal processing has more penalty on MU-MIMO-processing power consumption, and  $\gamma^*$  decreases, resulting in more SU clusters for high EE. Note that the  $\gamma^*$  depends on various parameters, such as  $U$ ,  $M$ , and the power consumption model, and thus, it is difficult to be found analytically. An EE gap between per-cluster optimal and heuristic power control is negligible in the points contiguous to the maximum EE when the penalty of power consumption for MU precoding is nonnegligible.

For the sake of comparison, we add the average EE of a large-size collocated antenna system (L-CAS). The L-CAS can be interpreted as one naive implementation of L-MIMO system that employs a full MU with a simple AS method that selects  $U$  antennas which give the largest average channel gains. The L-CAS may have sufficiently powerful processor to compensate nonlinear effects at the PAs; therefore, the L-CAS BS can employ a PA with much higher efficiency than the DAs in L-DAS. In simulation, we set the PA efficiency of L-CAS by 60%, i.e.,  $\eta_m = 0.6$ , with the corresponding signal processing power consumption increased by 10%. Furthermore, we set  $P_{\text{cc2},m} = 0$  pW/bit/s as there is no oRFs. Numerical result shows that EE performance of L-CAS is very poor because of the high power consumption to overcome the large path losses.

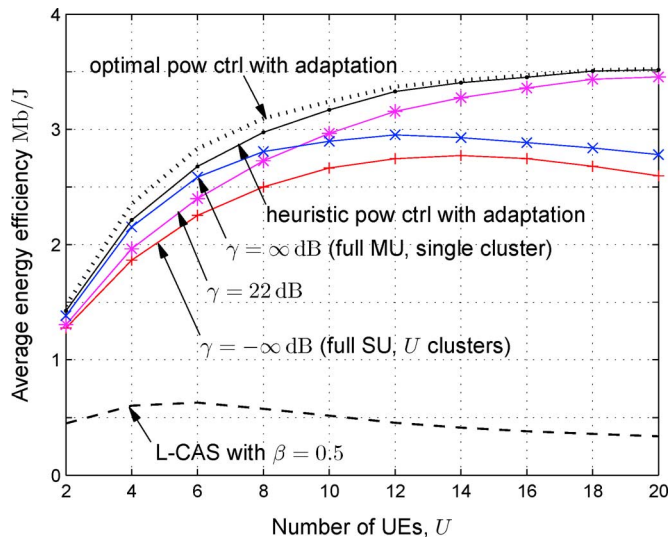


Fig. 4. Average EE over number of UEs,  $U$ , with various UC thresholds  $\gamma$  when  $M = 400$ ,  $\beta = 0.5$ , and  $P_{\text{sig}} = 50$  nW/Hz.

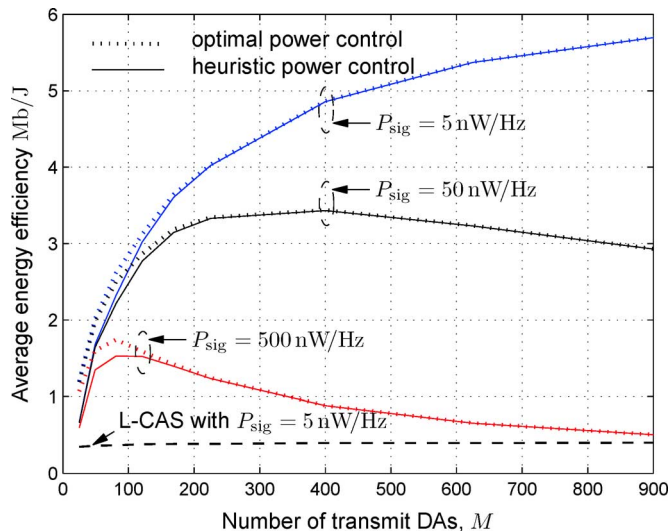


Fig. 5. Average EE over network size,  $M$ , with various  $P_{\text{sig}}$ 's when  $U = 20$ ,  $\beta = 0.5$ , and  $\gamma = 22$  dB.

### B. Average EE Over Number of Users

Fig. 4 shows average EEs over  $U$  with various UC thresholds  $\gamma$  when  $M = 400$  and  $\beta = 0.5$ . Two extreme cases of UC schemes with  $\gamma = \infty$  and  $\gamma = -\infty$  are compared with a UC with fixed  $\gamma$  by 22 dB, which is a proper threshold based on the results in Fig. 3 and allows both MU and SU schemes. The full MU scheme achieves higher EE than the full SU scheme with heuristic power control, because the MU-MIMO-processing power consumption is not dominant when  $\beta = 0.5$  and the full MU-MIMO achieves higher throughput than the full SU case. As  $U$  increases, UC with  $\gamma = 22$  dB achieves higher EE than full MU scheme because the MU-MIMO-processing power consumption increases severely and becomes dominant, which is noticeable as the large number of UEs is our interest of this work. However, when there are small number of UEs, less than nine, full MU scheme outperforms the system with  $\gamma = 22$  dB.

Using Algorithms 3 and 4 in Subsection IV.D,  $M_u$  and  $\gamma$  can be adapted and EE can be improved over any number of UEs. For the adaptation parameters, we use  $Q_{\text{AS}} = 5$ ,  $Q_{\text{UC}} = 10$ ,  $\delta = 5$  dB, and the initial threshold  $\gamma = -10$  dB. Per-cluster optimal power control further improves the EE.

### C. Average EE Over Network Size

Fig. 5 shows average EEs over network size  $M$ , i.e., the number of DAs (not active DAs), when  $U = 20$ ,  $\beta = 0.5$ , and  $\gamma = 22$  dB. We evaluate EEs with three different signaling power consumption models with  $P_{\text{sig}} = \{5, 50, 500\}$  nW/Hz. The average EE increases as  $M$  increases because severe path loss can be circumvented with the increased degree of freedom of AS. On the other hand, the network power consumption will also increase as  $M$  increases, due to the nonzero  $P_{\text{sig}}$  even with a proper AS. Therefore, an EE increases and turns to decrease as  $M$  increases, and the optimal network size is observed, e.g., around  $M = 400$  with  $P_{\text{sig}} = 50$  nW/Hz. As expected, the EE optimal network size decreases as  $P_{\text{sig}}$  increases.

## VIII. CONCLUSION

In this paper, we have considered EE maximization problem for an L-DAS. The power consumption of L-DAS transmitter has been modeled. A simple channel-gain-based antenna selection and SINR-threshold-based user clustering methods have been proposed to reduce the computational complexity of precoding and power control and, at the same time, to reduce the signaling overhead. Iterative algorithms to adapt the number of assigned antennas and the clustering threshold have been considered. Numerical results have validated the potential of the L-DAS.

Remaining issues for further work regarding deployment, implementation, and operation of L-DAS include cell planning, regular/irregular deployment of antennas, synchronization for large cluster, robustness against CSI error, infrastructure cost for wired optical fronthaul, and a comparative, quantitative study of L-DAS and L-CAS considering capital expenditure and operational expenditure.

## REFERENCES

- [1] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sep. 2012.
- [2] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," <http://arxiv.org> [Online]. Available: <http://arxiv.org/abs/1304.0553>
- [3] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?," <http://arxiv.org> [Online]. Available: <http://arxiv.org/abs/1310.3843>
- [4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Trans. Signal Process.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [5] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3972–3987, Aug. 2013.
- [6] L. Dai, S. Zhou, and Y. Yao, "Capacity analysis in CDMA distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2613–2620, Nov. 2005.

- [7] H. Kim, S.-R. Lee, K.-J. Lee, and I. Lee, "Transmit schemes based on sum rate analysis in distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1201–1209, Mar. 2012.
- [8] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 69–73, Jan. 2007.
- [9] J. Zhang and J. G. Andrews, "Distributed antenna systems with randomness," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3636–3646, Sep. 2008.
- [10] R. W. Heath, S. Peters, and J. Z. Y. Wang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 161–167, Apr. 2013.
- [11] J. Joung, C. K. Ho, and S. Sun, "Power amplifier switching (PAS) for energy efficient systems," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 14–17, Feb. 2013.
- [12] J. Joung, C. K. Ho, and S. Sun, "Spectral efficiency and energy efficiency of OFDM systems: Impact of power amplifiers and countermeasures," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 208–220, Feb. 2014.
- [13] O. Onireti, F. Hélot, and M. A. Imran, "On the energy efficiency-spectral efficiency trade-off of distributed MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3741–3753, Sep. 2013.
- [14] H. Kim, S.-R. Lee, C. Song, and I. Lee, "Optimal power allocation for energy efficiency maximization in distributed antenna systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 4362–4366.
- [15] J. Gan, W. Li, S. Zhou, and J. Wang, "On sum rate of multi-user distributed antenna system with circular antenna layout," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Baltimore, MD, USA, Sep. 2007, pp. 596–600.
- [16] D. Castanheira and A. Gameiro, "Distributed antenna system capacity scaling," *IEEE Wireless Commun. Mag.*, vol. 17, no. 3, pp. 68–75, Jun. 2010.
- [17] J. Joung, Y. K. Chia, and S. Sun, "Energy efficient multiuser MIMO systems with distributed transmitters," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013.
- [18] A. Tölli, H. Pannanen, and P. Komulainen, "Decentralized Minimum Power Multi-Cell Beamforming With Limited Backhaul Signaling," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 570–580, Feb. 2011.
- [19] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.
- [20] X. Wang, P. Zhu, and M. Chen, "Antenna location design for generalized distributed antenna systems," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 315–317, May 2009.
- [21] J. Joung and S. Sun, "Energy efficient power control for distributed transmitters with ZF-based multiuser MIMO precoding," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1766–1769, Sep. 2013.
- [22] W. Vereecken, W. V. Heddeghem, M. Deruyck, B. Puype, B. Lannoo, W. Joseph, D. Colle, L. Martens, and P. Demeester, "Power consumption in telecommunication networks: Overview and reduction strategies," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 62–69, Jun. 2011.
- [23] F. Liu, P. Shu, H. Jin, L. Ding, J. Yu, D. Niu, and B. Li, "Gearing resource-poor mobile devices with powerful clouds: Architectures, challenges, and applications," *IEEE Wireless Commun. Mag.*, vol. 20, no. 3, pp. 14–22, Jun. 2013.
- [24] Y. Okawachi, M. S. Bigelow, J. E. Sharping, Z. Zhu, A. Schweinsberg, D. J. Gauthier, R. W. Boyd, and A. L. Gaeta, "Tunable all-optical delays via Brillouin slow light in an optical fiber," *Phys. Rev. Lett.*, vol. 94, pp. 15 3902-1–15 3902-4, Apr. 2005.
- [25] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.
- [26] C. Jiang and L. J. Cimini, "Antenna selection for energy-efficient MIMO transmission," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 577–580, Dec. 2012.
- [27] J. Xu and L. Qiu, "Energy efficiency optimization for MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690–701, Feb. 2013.
- [28] "LTE; E-UTRA; RF Requirements for LTE Pico Node B," ETSI, Tech. Rep. 136 931 V9.0.0, 2011 [Online]. Available: <http://www.etsi.org/deliver/>
- [29] F. Leighton, *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes*, 2nd ed. Burlington, MA: Morgan Kaufmann, 1992.
- [30] J. Joung, E. Y. Kim, S. H. Lim, Y.-U. Jang, W.-Y. Shin, S.-Y. Chung, J. Chun, and Y. H. Lee, "Capacity evaluation of various multiuser MIMO schemes in downlink cellular environments," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Helsinki, Finland, Sep. 2006.
- [31] J. Lee and N. Jindal, "High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4787–4792, Dec. 2007.
- [32] X.-H. You, D.-M. Wang, B. Sheng, X.-Q. Gao, X.-S. Zhao, and M. Chen, "Cooperative distributed antenna systems for mobile communications," *IEEE Wireless Commun. Mag.*, vol. 17, no. 3, pp. 35–42, Jun. 2010.
- [33] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2004.
- [34] "Energy aware radio and network technology (EARTH) project," [Online]. Available: <https://www.ict-earth.eu>
- [35] R. S. Tucker, "Energy consumption and energy density in optical and electronic signal processing," *IEEE Photon. J.*, vol. 3, no. 5, pp. 820–833, Oct. 2011.



**Jington Joung** (S'04–M'07) received B.S. degree in EE from Yonsei University, Korea, and the M.S. and Ph.D. degrees in EECS from the Korea Advance Science and Technology (KAIST), Korea. From March 2007 to August 2008, he was a Postdoctoral Research Scientist at EE Department, KAIST and, at the same time, he worked as a part time Researcher at Lumicomm, Inc., Korea. From October 2008 to September 2009, he was a Postdoctoral Fellow at the EE Department, University of California, Los Angeles, USA.

Since October 2009, he has been working as a Scientist at Institute for Infocomm Research, Singapore. His current research has focused on the study of energy efficient wireless communications.

Dr. Joung has served on the editorial board of the *APSIPA Transactions on Signal and Information Processing* (2014–present). He is the recipient of the first prize at the Intel-ITRC Student Paper Contest in 2006, and recognized as the Exemplary Reviewers in 2012 from IEEE COMMUNICATIONS LETTERS, and in 2012 and 2013 from IEEE WIRELESS COMMUNICATIONS LETTERS.



**Yeow Kiang Chia** is currently a research scientist at the Institute for Infocomm Research, Singapore. He received his M.Eng. (with first class honors) in electrical engineering from Imperial College, London. He received his M.S. and Ph.D., both in electrical engineering, from Stanford University in 2012. He was a recipient of the Stanford Graduate Fellowship (SGF) and the Agency for Science, Technology and Research (A\*STAR), Singapore National Science Scholarship (NSS) for his graduate studies at Stanford University. His research interests include

information theory, optimization, machine learning and signal processing.



**Sumei Sun** (SM'12) obtained the B.Sc.(Honors) Degree from Peking University, China, the M.Eng. Degree from Nanyang Technological University, and Ph.D. Degree from National University of Singapore. She has been with Institute for Infocomm Research (formerly Centre for Wireless Communications) since 1995 and she is currently Head of Modulation & Coding Dept, developing physical layer-related solutions for next-generation communication systems. Her recent research interests are in energy efficient multiuser cooperative MIMO systems, joint source-channel processing for wireless multimedia communications, and wireless transceiver design.

Dr. Sun has served as the TPC Chair of 12th IEEE International Conference on Communications in 2010 (ICCS 2010), General Co-Chair of 7th and 8th IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS), and Track Co-Chair of Transmission Technologies, IEEE VTC 2012 Spring. She has also an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and an Editor of IEEE WIRELESS COMMUNICATION LETTERS. She is co-recipient of IEEE PIMRC'2005 Best Paper Award.