Compact Ultra Massive Antenna Array: A Simple Open-Loop Massive Connectivity Scheme

Kai-Kit Wong, *Fellow, IEEE,* Chan-Byoung Chae, *Fellow, IEEE,* and Kin-Fai Tong, *Fellow, IEEE*

***Abstract*—This paper aims to present a simple multiple access scheme for massive connectivity that enables a large number of mobile user equipments (UEs) to occupy the same time-frequency channel without the need of precoding and power control at the base station (BS) and interference cancellation at each UE. The proposed approach does not even require the UEs to know their signal-to-interference ratios (SIRs) and each UE also needs only two radio-frequency (RF) chains to operate. The proposed scheme is inspired by the emerging concept of fluid antenna system (FAS) which enables high-resolution position-changeable antenna to be deployed at each UE. Instead of activating only one port of FAS for reception, each UE activates an ultra massive number of ports to receive the signal. The activated ports are chosen to ensure that the in-phase and quadrature components of the desired signal at the ports are added constructively while the interference signals superimpose randomly. This approach is referred to as compact ultra massive antenna array (CUMA) which can also be realized by deploying a dense, fixed massive antenna array at each UE. We derive the exact probability density function (pdf) of the SIR of a CUMA UE which leads to the data rate analysis. Simulation results demonstrate that even with mutual coupling and under finite scattering, more than 10 UEs can be supported by having a 25×13-port FAS of size $15\text{ cm}\times8\text{ cm}$ at each UE. Considering quadrature phase shift keying (QPSK), CUMA delivers a network data rate of 10.7 bps per channel use serving 10 UEs at 26 GHz, and the rate is risen to 15.1 bps per channel use if 20 UEs are accommodated at 40 GHz with a 40×21-port FAS at every UE. In the case without mutual coupling and under rich scattering, CUMA can even handle hundreds of UEs per channel use.**

***Index Terms*—Compact ultra massive antenna array, FAS, fluid antenna, massive connectivity, multiple access, rate.**

I. Introduction

FORWARD looking beyond the fifth-generation (5G), one ambitious goal is *extreme* massive connectivity, demanding to support 10^7 devices/km^2 [1]–[4]. To realize this ambition, we require a multiple access technology that permits an enormous number of user equipments (UEs) to share the same time-frequency channel. In the current 5G, massive connectivity relies on the use of massive multiple-input multiple-output (MIMO) at the base station (BS) [5]. The super-directivity of massive MIMO in the spatial domain means that many users can occupy the same channel but be separated in space. There has already been suggestion that the sixth-generation (6G) will be heading to an ultra massive MIMO era, benefiting from an even greater number of antennas at the BS [6].

However, is it that simple, we can just increase the number of antennas at the BS to keep up our ever-increasing connectivity demands? A close look reveals a more complex situation. In 5G, massive MIMO with 64 BS antennas is not designed to serve more than 12 UEs on the same channel. This is because the codebook-based Type II New Radio (NR) multiuser MIMO precoding adopted in 5G is not simple [7]. It certainly looks more complex than the matched filtering precoding, originally envisioned in [8]. Upscaling connectivity via massive MIMO causes severe overhead for acquiring the channel states from a large number of UEs and immense complexity for optimizing the precoding matrices. Moreover, MIMO is not an upgrade-friendly solution and any change in the number of BS antennas needs to be carefully addressed in standardization activities.

On the other hand, non-orthogonal multiple access (NOMA) [9], [10] and rate-splitting multiple access (RSMA) [11]–[13] have emerged to offer greater capacity. Due to their aggressive approach to handling multiuser signals, many regard them as a rising massive connectivity solution. However, NOMA and RSMA still require the channel state information (CSI) at the BS for user clustering and power control optimization. Worse, both necessitate UEs to carry out interference cancellation to mitigate the inter-user interference, which is a major obstacle. It seems inconceivable that they will be commanded to deal with more than 3 UEs sharing the same radio channel.

Evidently, we are in need of a much simpler, more scalable multiple access scheme capable of massive connectivity. There is a recent concept, referred to as fluid antenna system (FAS), that can contribute to this [14]–[17]. FAS takes advantage of the recent advances in reconfigurable, flexible antennas such as liquid-based antennas [18] and pixel-based antennas [19]–[21], to envisage that the point of reception over the available space of a UE (i.e., position, also referred to as 'port') can be adaptively and finely changed to improve system performance, e.g., [22]–[25]. In short, FASs include all forms of movable and non-movable flexible-position antenna systems.[1]

Towards massive connectivity, FAS can be used to achieve multiple access in a novel manner. Fast fluid antenna multiple access (FAMA) [26], [27] advocates to switch the port of each UE on a symbol-by-symbol basis in order to receive the signal

The work of Wong and Tong is supported by the Engineering and Physical Sciences Research Council (EPSRC) under grant EP/W026813/1. For the purpose of open access, the authors will apply a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

The work of C. B. Chae is supported by the Institute of Information and Communication Technology Promotion (IITP) grant funded by the Ministry of Science and ICT (MSIT), Korea (2021-0-02208, 2022R1A5A1027646).

K. Wong, and K. Tong are with the Department of Electronic and Electrical Engineering, University College London, London, WC1E 7JE, United Kingdom. K. Wong is also affiliated with Yonsei Frontier Laboratory, Yonsei University, Seoul, 03722, Korea.

C. B. Chae is with the School of Integrated Technology, Yonsei University, Seoul, 03722, Korea.

Corresponding author: Kai-Kit Wong (e-mail: kai-kit.wong@ucl.ac.uk).

[1]FAS is a concept and does not specify how it is made. In fact, techniques for implementing FAS vary. Some used soft materials as a radiating element and shifted it nearly continuously in space either using electric field or small pumps while others investigated the use of a large number of small pixels to form switchable antennas in space with high resolution. Discussion of FAS regarding implementation approaches can be found in [14].

where the instantaneous sum-interference plus noise cancels. The results in [26], [27] demonstrated that tens of UEs could be easily supported but fast FAMA remains highly theoretical because of practical challenges.[2] Recent work in [28] therefore proposed slow FAMA where each UE switches its port once during a coherence time period, to the one in which the signal-to-interference and noise ratio (SINR) can be maximized. Slow FAMA is a lot more practical and can still handle several UEs sharing the same radio channel. Note that the advantage over massive MIMO, NOMA and RSMA is that FAMA does not require CSI at the BS for precoding, user clustering and power control and that the UEs need only single-user decoding. Most recently, [29] investigated the use of opportunistic scheduling for enhancing FAMA. Overall, however, slow FAMA is quite limited in connectivity and the network rate will hit a ceiling if there are too many UEs sharing the same channel.

Motivated by this, in this paper, we propose a novel multiple access scheme which is referred to as compact ultra massive antenna array (CUMA). From the antenna design viewpoint, CUMA can be considered as one form of approximating FAS by packing many antennas compactly in space. By turning on and off some of these antennas, CUMA mimics the effects of switching the antenna positions with high resolution, resemblance to FAS. For multiple access, CUMA can be understood as a variant of slow FAMA. Like slow FAMA, CUMA requires each UE to switch its position only when the channel changes. Instead of activating only one position in the FAS, each UE in CUMA turns on a massive number of selected ports[3] for reception. The signals from the selected ports are aggregated to produce the received signal for decoding. With the hope of realizing massive connectivity, simplicity is our priority.

In the proposed technique, a CUMA UE needs only two radio-frequency (RF) chains,[4] and no complex signal processing is performed to combine the signals from the selected ports. The great performance comes from selecting the correct ports such that the in-phase and quadrature components of the desired signals superimpose constructively whereas the interference signals add randomly. With the increase in the number of activated ports, the interference immunity of a UE improves. Our results will demonstrate that the network rate of CUMA increases with the number of UEs and outperforms slow FAMA greatly. In addition to methods for selecting the ports for activation, we derive the probability density function (pdf) of the signal-to-interference ratio (SIR) for an average UE in the CUMA system. This in turn permits the evaluation of the data rate and ergodic rate of the CUMA network.

[2]Fast FAMA requires the antenna position to be optimized on a symbol-by-symbol basis. While [27] has presented methods to estimate the best position of FAS for maximizing the ratio of the instantaneous energy of the desired signal to that of the sum-interference plus noise signal, the proposed methods require the FAS to observe all the possible ports for the estimation. Knowing that fast FAMA needs a large number of ports (hundreds to even a thousand) to work well, this could be problematic in practice.

[3]In this paper, the terms 'position' and 'port' are used interchangeably.

[4]In fact, CUMA can work with only one RF chain at a UE but it will be shown in Section III that having two RF chains is the most natural choice to deal with both in-phase and quadrature signals.

II. Conventional FAMA Network Model

In this paper, we consider a downlink model in which a BS with N_t fixed antennas transmits to U UEs on the same time-frequency channel. The BS antennas are sufficiently far apart to be spatially independent. For convenience, the UEs are assumed to have identical specifications and each UE is equipped with a two-dimensional (2D) FAS of physical size $\bar{W} = W_1\lambda \times W_2\lambda$ where λ is the wavelength of radiation. The FAS has totally $N = N_1 \times N_2$ ports evenly distributed over the space $\bar{W}$. Each port represents a physical position at which a signal can be received if the port is on. Note that the ports can be closely spaced so their channels are correlated.

Hypothetically, if all the ports are activated, the received signals at UE u in vector form can be written as

$$\begin{aligned}\mathbf{r}_u &= \mathbf{H}_u\mathbf{b}_u s_u + \sum_{\substack{\tilde{u}=1\\ \tilde{u}\neq u}}^{U} \mathbf{H}_u\mathbf{b}_{\tilde{u}} s_{\tilde{u}} + \boldsymbol{\eta}_u,\\ &\equiv \mathbf{g}_{u,u}s_u + \sum_{\substack{\tilde{u}=1\\ \tilde{u}\neq u}}^{U} \mathbf{g}_{\tilde{u},u}s_{\tilde{u}} + \boldsymbol{\eta}_u,\end{aligned} \tag{1}$$

where $\mathbf{H}_u \in \mathbb{C}^{N\times N_t}$ denotes the complex channels from the BS to the ports of UE u, $\mathbf{b}_u \in \mathbb{C}^{N_t}$ denotes the precoding or beamforming vector for transmitting UE u's signals, $\mathbf{g}_{\tilde{u},u} = \mathbf{H}_u\mathbf{b}_{\tilde{u}} \in \mathbb{C}^N$ denotes the effective channel vector from UE $\tilde{u}$ to the ports of UE u, s_u represents the information symbol for UE u, and $\boldsymbol{\eta}_u$ is the complex additive white Gaussian noise vector at the ports of UE u whose elements are independent, identically distributed (i.i.d.) and have zero mean and variance of σ_η^2. Precoding $\{\mathbf{b}_u\}$ at the BS is chosen as any orthonormal basis spanning the range of an $N_t \times N_t$ complex space and needs no optimization with the CSI. As explained in [30], if the channel has limited scatterers, channel randomization via $\{\mathbf{b}_u\}$ becomes essential to inject sufficient channel differentiation between the UEs for FAS to work. On the other hand, note that in (1), vectorization has been used to convert the 2D channel matrix into a one-dimensional (1D) channel vector, with an appropriate mapping between the (n_1, n_2)-th port and the k-th port, i.e., $k = \text{map}(n_1, n_2)$. Specifically, we have

$$n_1 = \begin{cases} N_1 \text{ if } k \bmod N_1 = 0,\\ k \bmod N_1 \text{ otherwise,}\end{cases} \tag{2}$$

and

$$n_2 = \begin{cases} \left\lfloor \dfrac{k}{N_1} \right\rfloor \text{ if } n_1 = N_1,\\ \left\lfloor \dfrac{k}{N_1} \right\rfloor + 1 \text{ otherwise,}\end{cases} \tag{3}$$

where $\lfloor x \rfloor$ returns the largest integer that is smaller than or equal to x. In this paper, for convenience, we set $N_t = U$.

We consider a finite scatterer channel model [31] so that

$$\begin{aligned}[\mathbf{H}_u]_t = &\sqrt{\frac{K}{K+1}} e^{j\delta_{\tilde{u},u}} \mathbf{a}(\theta_0^{(\tilde{u},u)}, \phi_0^{(\tilde{u},u)})\\ &+ \frac{1}{\sqrt{N_p}}\sqrt{\frac{1}{K+1}} \sum_{\ell=1}^{N_p} \kappa_\ell^{(\tilde{u},u)} \mathbf{a}(\theta_\ell^{(\tilde{u},u)}, \phi_\ell^{(\tilde{u},u)}),\end{aligned} \tag{4}$$

where $[\cdot]_t$ returns the t-th column vector of the input matrix,[5] K is the Rice factor, $\delta_{\tilde{u},u}$ is the phase of the line-of-sight (LoS) component, $\kappa_\ell^{(\tilde{u},u)}$ represents the complex channel coefficient of the ℓ-th scattered component, N_p is the number of non-LoS paths, and $\mathbf{a}(\theta,\phi)$ is the steering vector given by

$$\mathbf{a}(\theta,\phi) = \left[1 \; e^{j\left(\frac{2\pi W_1}{N_1-1}\right)\sin\theta\cos\phi} \cdots \; e^{j2\pi W_1 \sin\theta\cos\phi}\right]^T \otimes \left[1 \; e^{j\left(\frac{2\pi W_2}{N_2-1}\right)\sin\theta\cos\phi} \cdots \; e^{j2\pi W_2 \sin\theta\cos\phi}\right]^T, \quad (5)$$

where $\otimes$ denotes the Kronecker tensor product, θ and ϕ are, respectively, the azimuth and elevation angle-of-arrival of the corresponding path, and the superscript T is transposition. In addition, we denote the channel power as $\Omega = \mathrm{E}[|[\mathbf{g}_{\tilde{u},u}]_k|^2]$ where $[\cdot]_k$ returns the k-th entry of the input vector.

A popular special case is the rich scattering scenario where $K = 0$ and $N_p \to \infty$. In this case, precoding is unimportant for FAMA and the precoding matrix $[\mathbf{b}_1 \cdots \mathbf{b}_U]$ can be set as an identity matrix. Also, $[\mathbf{g}_{\tilde{u},u}]_k$ will be complex Gaussian distributed. If 1D FAS is used at the UE under 2D scattering environments, then we have

$$\varrho_{k-m} \triangleq \mathrm{E}\left[[\mathbf{g}_{\tilde{u},u}]_k \left([\mathbf{g}_{\tilde{u},u}]_m\right)^*\right] = \Omega J_0\left(\frac{2\pi(k-m)\bar{W}}{(N-1)\lambda}\right), \quad (6)$$

in which $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind [32]. For the case of 2D FAS under 3-dimensional (3D) scattering, the covariance between any two ports can be derived as [33], [34, Lemma 2]

$$\varrho_{k-m} \triangleq \mathrm{E}\left[[\mathbf{g}_{\tilde{u},u}]_{k=\mathrm{map}(n_1,n_2)} \left([\mathbf{g}_{\tilde{u},u}]_{m=\mathrm{map}(n_3,n_4)}\right)^*\right] = \Omega \mathcal{J}_0\left(2\pi\sqrt{\left(\frac{(n_1-n_3)W_1}{N_1-1}\right)^2 + \left(\frac{(n_2-n_4)W_2}{N_2-1}\right)^2}\right), \quad (7)$$

where $\mathcal{J}_0(x) = \frac{\sin(x)}{x}$ denotes the spherical Bessel function.

Notice that the effects of blockages and path loss are not included in our model. The main reason is that with blockages, it would be difficult to understand the interference mitigation capability of the proposed CUMA scheme, since the random blockages and path loss would affect the SINR considerably. The results based on our model without blockages could also be interpreted as a conservative look on the performance when the interfering signals are always present and strong.

For slow FAMA [28], only the best port that maximizes the received SINR is activated. For UE u, we have

$$k_u^* = \arg\max_k \frac{\sigma_s^2 \left|[\mathbf{g}_{u,u}]_k\right|^2}{\sigma_s^2 \sum_{\substack{\tilde{u}=1 \\ \tilde{u}\neq u}}^{U} \left|[\mathbf{g}_{\tilde{u},u}]_k\right|^2 + \sigma_\eta^2}, \quad (8)$$

where $\sigma_s^2 = \mathrm{E}[|s_u|^2]$ is the symbol power. After the port has been selected, the estimated symbol can be found from

$$\tilde{s}_u = \frac{[\mathbf{r}_u]_{k_u^*}}{[\mathbf{g}_{u,u}]_{k_u^*}}. \quad (9)$$

[5]Note that if the input is a vector, this gives the t-th entry of the vector.

For benchmarking, the average received signal-to-noise ratio (SNR) for each UE is defined as $\Gamma \triangleq \frac{\sigma_s^2\Omega}{\sigma_\eta^2}$.

III. CUMA

A. Signal Model

In this paper, we propose a new FAS receiver architecture for FAMA, referred to as CUMA. Fig. 1 illustrates a downlink system with a MIMO BS communicating to several CUMA UEs. In CUMA, the main difference compared to the conventional FAMA system is that at each user, say UE u, it activates $\bar{N}$ out of the available N ports to obtain the received signal. Let us say the ports that are selected for activation have their indices stored in the set $\mathcal{K}$, hence $\bar{N} = |\mathcal{K}|$. How to determine the set $\mathcal{K}$ for each UE will be addressed in Section III-C. Also, we will first focus on strengthening the in-phase component of the resultant signal here before we include the quadrature component for a full design in Section III-D.

Upon receiving the signals from the activated ports (i.e., the sampled ports) given in the set $\mathcal{K}$, they will be aggregated to produce the resulting received signals

$$\begin{cases} r_u^{\mathrm{I}} = \sum_{k\in\mathcal{K}} \mathrm{real}\left(r_k^{(u)}\right), \\ r_u^{\mathrm{Q}} = \sum_{k\in\mathcal{K}} \mathrm{imag}\left(r_k^{(u)}\right), \end{cases} \quad (10)$$

in which $r_k^{(u)} \triangleq [\mathbf{r}_u]_k$ in (1). They correspond to the in-phase and quadrature components of the aggregated signal. Denoting the complex information symbol as $s_u = s_u^{\mathrm{I}} + j s_u^{\mathrm{Q}}$, we have

$$r_u^{\mathrm{I}} = \left[\sum_{k\in\mathcal{K}} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right)\right] s_u^{\mathrm{I}} + \left[-\sum_{k\in\mathcal{K}} \mathrm{imag}\left([\mathbf{g}_{u,u}]_k\right)\right] s_u^{\mathrm{Q}} + \sum_{k\in\mathcal{K}} \mathrm{real}\left(\left[\sum_{\substack{\tilde{u}=1 \\ \tilde{u}\neq u}}^{U} \mathbf{g}_{\tilde{u},u} s_{\tilde{u}} + \boldsymbol{\eta}_u\right]_k\right) \quad (11)$$

and similarly,

$$r_u^{\mathrm{Q}} = \left[\sum_{k\in\mathcal{K}} \mathrm{imag}\left([\mathbf{g}_{u,u}]_k\right)\right] s_u^{\mathrm{I}} + \left[\sum_{k\in\mathcal{K}} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right)\right] s_u^{\mathrm{Q}} + \sum_{k\in\mathcal{K}} \mathrm{imag}\left(\left[\sum_{\substack{\tilde{u}=1 \\ \tilde{u}\neq u}}^{U} \mathbf{g}_{\tilde{u},u} s_{\tilde{u}} + \boldsymbol{\eta}_u\right]_k\right). \quad (12)$$

Note that to obtain r_u^{I} and r_u^{Q}, UE u simply needs to retrieve the in-phase and quadrature components of the superposition of the received signals from the activated ports. No scaling nor phase shifting is required before aggregation, so only one RF chain suffices to obtain the output signal for detection.

B. Detectors

In this paper, the CSI, i.e., the channel vector $\mathbf{g}_{u,u}$ containing the channel coefficients over all the available ports, is assumed known at UE u. The channels from other interfering users are not needed. Recently, a channel estimation framework for FAS has been presented in [35]. In order to estimate

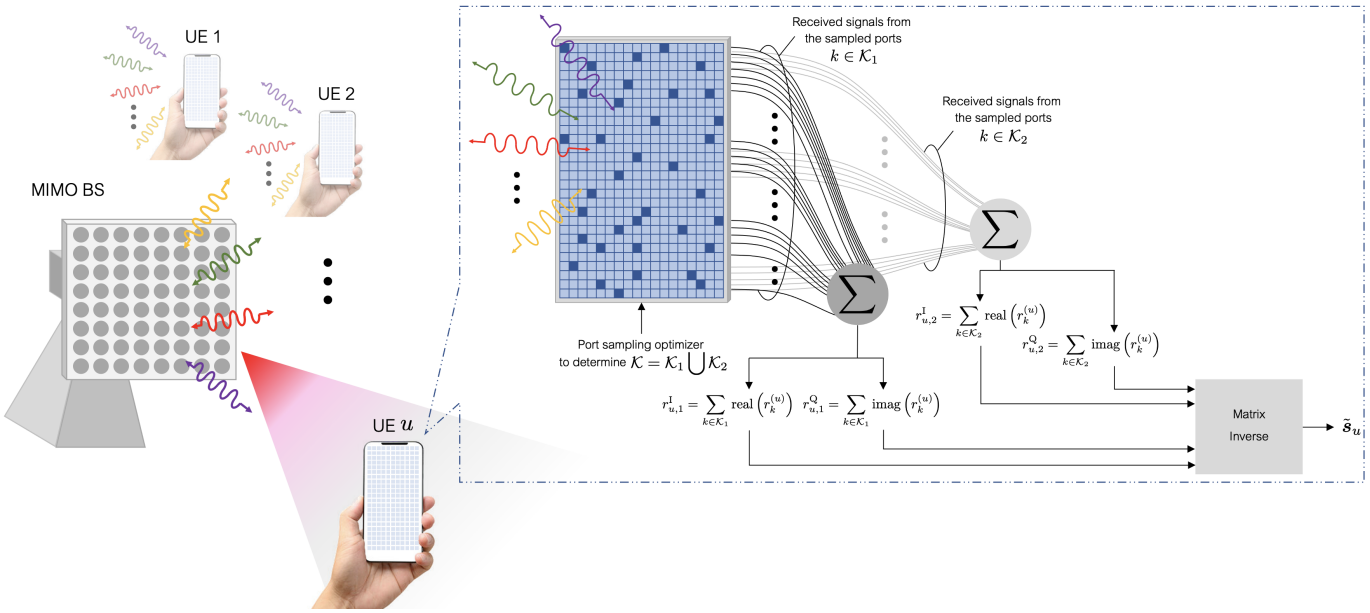


Fig. 1. A downlink system in which a fixed-position MIMO BS communicates to several FAS-enabled UEs. Multiple access is achieved using CUMA at each UE. A massive number of ports sharing the same RF chain at the FAS are activated to receive and aggregate the signals for detection at the UE.

$\mathbf{g}_{u,u}$ regardless of the value of N, it has turned out that only $N_p + 1$ samples in space are needed [36] which aligns with the conclusion for MIMO in the millimeter-wave band in [37]. Note that in the millimeter-wave and terahertz bands, N_p is small and thus, estimating $\mathbf{g}_{u,u}$ for very large N is simple. The channel estimation process can also benefit from easy angle-of-arrival estimation by a programmable metasurface [38].

In the case of rich scattering, i.e., $N_p \to \infty$, it does not require an infinite bandwidth to estimate the channels of FAS. In particular, given the space of FAS, there can only be a finite number of possible independent channels, which depends on the numerical rank of the spatial correlation matrix. For FAS with size $0.5\lambda \times 0.5\lambda$, the complexity of estimating the FAS channel is the same as that with $N_p = 4$. If the size is increased to $2\lambda \times 2\lambda$, then the complexity will be equivalent to the case with $N_p = 25$. For details, see [33, Table II].

With the knowledge of $\mathbf{g}_{u,u}$, we can have coherent detection for the UE symbol $\boldsymbol{s}_u \triangleq [s^{\mathrm{I}}_u \; s^{\mathrm{Q}}_u]^T$. The most straightforward way to do so is Direct Detection which treats interference as noise and attempts to detect s^{I}_u (the in-phase symbol) from r^{I}_u and s^{Q}_u (the quadrature symbol) from r^{Q}_u, respectively, i.e.,

$$\tilde{\boldsymbol{s}}_u = \begin{bmatrix} \tilde{s}^{\mathrm{I}}_u \\ \tilde{s}^{\mathrm{Q}}_u \end{bmatrix} = \frac{1}{\sum_{k\in\mathcal{K}} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right)} \tilde{\boldsymbol{r}}_u, \tag{13}$$

where

$$\tilde{\boldsymbol{r}}_u \triangleq \begin{bmatrix} r^{\mathrm{I}}_u \\ r^{\mathrm{Q}}_u \end{bmatrix}. \tag{14}$$

A more effective detector will be to perform Matrix Inverse to obtain the estimates from r^{I}_u and r^{Q}_u so that

$$\tilde{\boldsymbol{s}}_u = \mathbf{G}^{-1}_u \tilde{\boldsymbol{r}}_u, \tag{15}$$

where

$$\mathbf{G}_u = \begin{bmatrix} \sum_{k\in\mathcal{K}} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right) & -\sum_{k\in\mathcal{K}} \mathrm{imag}\left([\mathbf{g}_{u,u}]_k\right) \\ \sum_{k\in\mathcal{K}} \mathrm{imag}\left([\mathbf{g}_{u,u}]_k\right) & \sum_{k\in\mathcal{K}} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right) \end{bmatrix}. \tag{16}$$

C. Multi-port Activation Schemes

A first glance of the signals, r^{I}_u and r^{Q}_u, might suggest that there is no benefit of activating many ports for reception since the signals will be superimposed randomly. However, the result can be very different if $\mathcal{K}$ is chosen carefully. In particular, we propose to activate those ports, $k_1, \ldots, k_{\bar{N}}$, such that

$$\begin{aligned} \mathrm{sign}\left(\mathrm{real}\left([\mathbf{g}_{u,u}]_{k_1}\right)\right) &= \mathrm{sign}\left(\mathrm{real}\left([\mathbf{g}_{u,u}]_{k_2}\right)\right) = \cdots \\ &= \mathrm{sign}\left(\mathrm{real}\left([\mathbf{g}_{u,u}]_{k_{\bar{N}}}\right)\right) \end{aligned} \tag{17}$$

so that the activated ports can be sure to add constructively to strengthen the desired signal. Evidently, it does not matter if the activated ports lead to all positive in-phase channels, or all negative ones as long as they have the same sign. To decide which set of ports to select, one can perform

$$\left|\sum_{k\in\mathcal{K}^+} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right)\right| \underset{\mathcal{K}^-}{\overset{\mathcal{K}^+}{\gtrless}} \left|\sum_{k\in\mathcal{K}^-} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right)\right|, \tag{18}$$

in which $\mathcal{K}^+$ represents the set of port indices for all positive in-phase channels and $\mathcal{K}^-$ for the negative in-phase channels. The set $\mathcal{K}$ chooses between $\mathcal{K}^+$ and $\mathcal{K}^-$ according to (18).

Presumably, since N is normally very large, the number of ports in $\mathcal{K}$, i.e., $\bar{N}$, can be very large. For practical reasons, one may prefer to reduce the value of $\bar{N}$. Here, we propose two ways to do so. First, we introduce the parameter $0 < \rho \leq 1$ which sets the minimum required level of the in-phase channel to be on. Specifically, the k-th port is selected only if

$$\mathrm{real}\left([\mathbf{g}_{u,u}]_k\right) \geq \rho \max_{m\in\mathcal{K}^+} \mathrm{real}\left([\mathbf{g}_{u,u}]_m\right). \tag{19}$$

The above has considered that $k \in \mathcal{K}^+$ and similar should be done in the case of $k \in \mathcal{K}^-$. The final set $\mathcal{K}$ can be obtained using the criterion (18) but this time only those satisfying (19) is summed. One might feel that a large ρ should be chosen to maximize the network performance. However, it is important

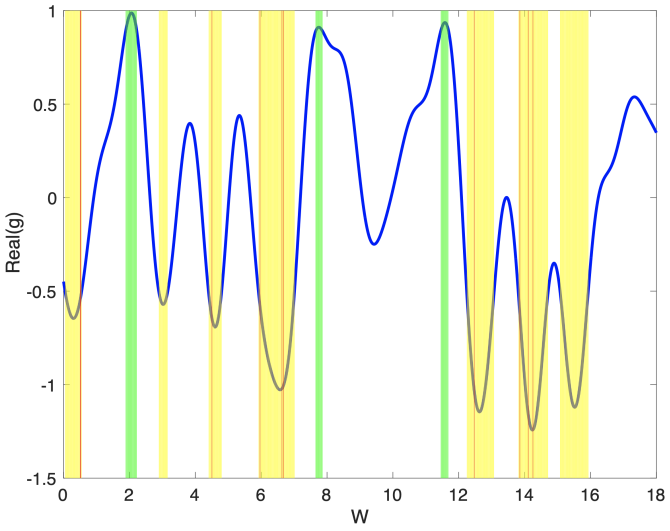


Fig. 2. An example illustrating how smart port sampling works when the FAS at the UE is assumed 1D with length of 18λ, and has 1000 ports. Samples in yellow are for $\rho = 0.4$ while the samples in green consider $\rho = 0.9$. The red samples are for $\rho = 0.4$ with $N_{\max} = 20$.

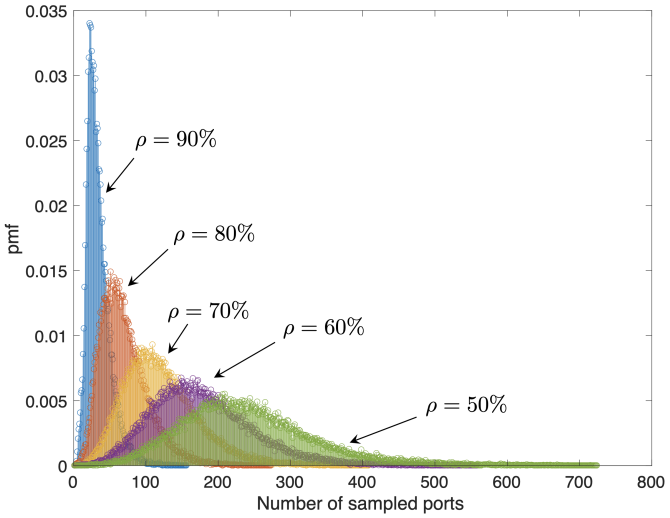


Fig. 3. The pmfs of the number of sampled ports for different values of ρ.

to recognize that strong in-phase channels of $\mathbf{g}_{u,u}$ at the ports do not imply eventual desirable performance as performance depends on the interference as well. It therefore would be too risky to choose the ports entirely based on the strengths of the desired channel and a smaller ρ tends to perform better due to diversity. Another way to reduce $\bar{N}$ is by setting a limit so that $\bar{N} \leq N_{\max}$. Fig. 2 illustrates an example of the in-phase channel for a 1D FAS with $N = 1000$ ports and $\bar{W} = 18\lambda$, highlighting the ports that are selected for different values of ρ. Additionally, Fig. 3 shows the probability mass functions (pmf) for the number of selected ports for different ρ.

D. Doubling Down CUMA

So far, the idea has been to ensure that the resultant signal of a CUMA UE has an aligned in-phase channel component for reception by finding the set $\mathcal{K}_1$. However, same can be done for maximizing the quadrature component of the channel. In particular, we can use the method in Section III-C by focusing on the quadrature component of the channel instead to obtain another set $\mathcal{K}_2$. With the two port sampling sets, we can obtain the received signals given by (11) and (12), $r^{\mathrm{I}}_{u,1}$ and $r^{\mathrm{Q}}_{u,1}$ using $\mathcal{K}_1$, and $r^{\mathrm{I}}_{u,2}$ and $r^{\mathrm{Q}}_{u,2}$ for $\mathcal{K}_2$. By applying the Matrix Inverse method, the estimated symbol can be obtained by

$$\tilde{\boldsymbol{s}}_u = \begin{bmatrix} A & -a \\ a & A \\ b & -B \\ B & b \end{bmatrix}^{-1} \begin{bmatrix} r^{\mathrm{I}}_{u,1} \\ r^{\mathrm{Q}}_{u,1} \\ r^{\mathrm{I}}_{u,2} \\ r^{\mathrm{Q}}_{u,2} \end{bmatrix}, \tag{20}$$

where pseudo-inverse is required, and

$$\begin{cases} A = \sum\limits_{k\in\mathcal{K}_1} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right), \quad a = \sum\limits_{k\in\mathcal{K}_1} \mathrm{imag}\left([\mathbf{g}_{u,u}]_k\right), \\ B = \sum\limits_{k\in\mathcal{K}_2} \mathrm{imag}\left([\mathbf{g}_{u,u}]_k\right), \; b = \sum\limits_{k\in\mathcal{K}_2} \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right). \end{cases} \tag{21}$$

Remarkably, $|A|$ and $|B|$ should be large by the design of $\mathcal{K}_1$ and $\mathcal{K}_2$ for aligning the in-phase and quadrature channels, respectively, while a and b are a lot smaller as the corresponding channels are not aligned and they mix randomly. Also, the inclusion of $\mathcal{K}_2$ provides a diversity signal $(r^{\mathrm{I}}_{u,2}, r^{\mathrm{Q}}_{u,2})$ that will greatly improve the quality of detection.

E. Mutual Coupling

Since there are a large number of radiating elements closely located to each other, mutual coupling does exist regardless of whether the ports are on or not. According to [39], the received signal vector with mutual coupling can be expressed as

$$\tilde{\mathbf{r}}_u = \mathbf{S}_{\underline{\mathcal{K}}} \boldsymbol{\Gamma}_{\mathrm{mc}} \mathbf{S}_{\underline{\mathcal{K}}} \mathbf{r}_u, \tag{22}$$

where $\mathbf{r}_u$ is given in (1), $\mathbf{S}_{\underline{\mathcal{K}}}$ denotes a diagonal matrix with the diagonal element being '1' specifying the activated port and '0' for the unselected port according to the set $\underline{\mathcal{K}} = \mathcal{K}_1 \bigcup \mathcal{K}_2$, and $\boldsymbol{\Gamma}_{\mathrm{mc}}$ is the mutual coupling matrix given by

$$\boldsymbol{\Gamma}_{\mathrm{mc}} = Z_T (\mathbf{Z} + Z_T \mathbf{I})^{-1}, \tag{23}$$

where Z_T is the termination impedance (typically 50 ohms) and $\mathbf{Z}$ is the mutual impedance matrix that is well known for antennas such as dipole and slot antennas [40, p. 417]. When the ports are not too close to each other, $\boldsymbol{\Gamma}_{\mathrm{mc}} \approx \mathbf{I}$.

With mutual coupling, CUMA will work exactly in the same way by considering the effective channel $\tilde{\mathbf{g}}_{u,u} = \boldsymbol{\Gamma}_{\mathrm{mc}} \mathbf{S}_{\underline{\mathcal{K}}} \mathbf{g}_{u,u}$ in place of $\mathbf{g}_{u,u}$ for detection in (20). Note that $\boldsymbol{\Gamma}_{\mathrm{mc}}$ can be pre-computed offline given the antenna technologies used.

It is worth pointing out that the state-of-the-art can provide isolations of more than -20 dB even if the adjacent elements are as close as 0.12λ apart [41]. Latest technologies have further improved the isolations to more than -30 dB [42]. The mutual coupling model in (23) above provides a conservative model assuming no advanced isolation methods.

Note that if fluid-material-based movable antennas are used, then mutual coupling will be much less and achieve better performance for CUMA. Nonetheless, movable antennas will be of limited use because to influence the physical layer, it will require a position change in the order of 10 cm in milliseconds, which equates to having an acceleration of 2×10^5 ms^{-2}. As

a reference, a bullet might exist from a gun at an acceleration of 2000 ms^{-2}. Ignoring whether such unreal acceleration is practically possible, such antenna would be too dangerous to be allowed on a handset or even BS. For this reason, the more realistic option for implementing FAS is using pixel antennas, or packing a lot of sub-wavelength metamaterial-based small antennas in space. As a result, to make sure that our results are relevant, we prefer considering packing a huge number of antennas compactly in space instead. While this causes more mutual coupling which is undesirable, this makes possible the switching of the antennas with no time delay.

IV. Performance Analysis

A. Assumptions and Notations

In this section, our objective is to analyze the performance of CUMA. As the UEs are all i.i.d., it suffices to focus on the performance of any typical UE in the analysis. Furthermore, as the performance depends on the amount of interference each UE suffers, our main effort will be on characterizing the SIR by deriving its pdf before we present the rate analysis.

To make the analysis tractable, we consider a specific type of CUMA where $\rho = 0$ (i.e., ports are selected regardless of their channel strengths)[6] and that $\mathcal{K} = \mathcal{K}^+$; as a result, only those ports with positive in-phase channels are selected in $\mathcal{K}_1$. Same goes to $\mathcal{K}_2$ for the quadrature channels. Direct detector is assumed at each UE. Overall, this particular CUMA can be viewed as a conservative approach in terms of the achievable performance. Evidently, N (hence also $\bar{N}$) is assumed to be extremely large. Furthermore, for simplicity, we assume real-valued modulation such as binary phase shift keying (BPSK), i.e., $s_u = s_u^{\mathrm{I}} \in \{-\sigma_s, \sigma_s\}$ and $s_u^{\mathrm{Q}} = 0$. Hence, we have only $r_{u,1}^{\mathrm{I}}$ and $r_{u,2}^{\mathrm{Q}}$ because $r_{u,1}^{\mathrm{Q}}$ and $r_{u,2}^{\mathrm{I}}$ are too noisy to be useful for reception. As $r_{u,1}^{\mathrm{I}}$ and $r_{u,2}^{\mathrm{Q}}$ are i.i.d., the following analysis will focus upon $r_{u,1}^{\mathrm{I}}$ first and we then replicate the argument for $r_{u,2}^{\mathrm{Q}}$ to complete the analysis by considering diversity of having two signals. Also, rich scattering is assumed, i.e., $K = 0$, $N_p = \infty$ and the effect of mutual coupling is ignored.

As we focus on a typical UE, the user index is unimportant and will be omitted. We also like to simplify our notations by considering the following definitions:

$$X_k \triangleq \mathrm{real}\left([\mathbf{g}_{u,u}]_k\right), \tag{24a}$$

$$X_k^+ \triangleq \max\{0, X_k\}, \tag{24b}$$

$$\alpha_{\mathrm{I}} \triangleq \left(\sum_{k=1}^{N} X_k^+\right)^2, \tag{24c}$$

$$Y_k^{(i)} \triangleq \mathrm{real}\left([\mathbf{g}_{i,u}]_k\right), \tag{24d}$$

$$\tilde{Y}_i \triangleq \sum_{k=1}^{N} t_k Y_k^{(i)}, \tag{24e}$$

$$\beta_{\mathrm{I}} \triangleq \sum_{i=1}^{I} \tilde{Y}_i^2, \tag{24f}$$

where $I \triangleq U - 1$ and $t_k \in \{0, 1\}$ denotes an i.i.d. Bernoulli random variable with equal probability, for modelling the on-off characteristics of a given port k. According to the above, the SINR for a UE based on $r_{u,1}^{\mathrm{I}}$ only is given by

$$\mathrm{SINR}_{\mathrm{I}} = \frac{\sigma_s^2 \alpha_{\mathrm{I}}}{\sigma_s^2 \beta_{\mathrm{I}} + \frac{\bar{N}\sigma_\eta^2}{2}} \approx \frac{\alpha_{\mathrm{I}}}{\beta_{\mathrm{I}}}, \tag{25}$$

in which the approximation is accurate when $\sigma_s \gg \sigma_\eta$ and/or $I \gg 1$, i.e., an interference-limited scenario. Note that we will also have $\mathrm{SINR}_{\mathrm{Q}}$ based on $r_{u,2}^{\mathrm{Q}}$ when the quadrature signal components are considered. The final analysis will combine the two SINRs to complete the rate analysis.

B. Main Results

In this section, we derive the pdf of the SIR, $\frac{\alpha_{\mathrm{I}}}{\beta_{\mathrm{I}}}$, before we get the network rate expression assuming a binary symmetric channel. To this end, we first present the analytical result that characterizes the covariance of the channels $\{X_k^+\}$.

Theorem 1: The covariance of X_k^+ is given by

$$\mathrm{cov}(X_k^+, X_m^+) = \frac{(1-\rho_{k,m}^2)^{\frac{3}{2}}\Omega}{4\pi} - \frac{\Omega}{4\pi} + \frac{\rho_{k,m}}{2\sqrt{\pi\Omega}} \mathcal{W}\left(-\sqrt{\frac{2}{1-\rho_{k,m}^2}}\frac{\rho_{k,m}}{\sqrt{\Omega}}, \frac{1}{\Omega}, \frac{1}{2}\right), \tag{26}$$

where $\rho_{k,m}$ with $k = \mathrm{map}(n_1, n_2)$ and $m = \mathrm{map}(n_3, n_4)$ is the correlation coefficient of X_k given by[7]

$$\rho_{k,m} = \mathcal{J}_0\left(2\pi\sqrt{\left(\frac{(n_1-n_3)W_1}{N_1-1}\right)^2 + \left(\frac{(n_2-n_4)W_2}{N_2-1}\right)^2}\right), \tag{27}$$

and

$$\mathcal{W}(a,b,c) = -\frac{a\Gamma\left(\frac{2c+3}{2}\right)}{\sqrt{2\pi}b^{\frac{2c+3}{2}}} {}_2F_1\left(\frac{1}{2}, \frac{2c+3}{2}; \frac{3}{2}; -\frac{a^2}{2b}\right) + \frac{\Gamma(c+1)}{2b^{c+1}}, \tag{28}$$

where $\Gamma(\cdot)$ is the gamma function and ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ denotes the Gauss hypergeometric function. Note that $\mathcal{W}(a,b,c)$ originates from a special integral of the Q function [43, (2)].

Proof: See Appendix A. ∎

Theorem 2: If N is sufficiently large, $\sqrt{\alpha_{\mathrm{I}}} = \sum_{k=1}^{N} X_k^+$ is approximately Gaussian with mean

$$\mathrm{E}[\sqrt{\alpha_{\mathrm{I}}}] = \frac{N}{2}\sqrt{\frac{\Omega}{\pi}} \tag{29}$$

and variance

$$\mathrm{var}[\sqrt{\alpha_{\mathrm{I}}}] = \frac{N\Omega}{4}\left(1 - \frac{1}{\pi}\right) + 2\sum_{m=2}^{N}\sum_{k=1}^{m-1} \mathrm{cov}(X_k^+, X_m^+). \tag{30}$$

where $\mathrm{cov}(X_k^+, X_m^+)$ is given by (26).

[6]Note that a smaller (or larger) ρ does not imply better or worse performance. Our simulation results in Section V will reveal that there is an optimal value for ρ that results in the best performance.

[7]In the case of having a 1D FAS, $\rho_{m+k,m} = J_0\left(\frac{2\pi k\bar{W}}{(N-1)\lambda}\right)$.

Proof: See Appendix B. ■

Theorem 2 indicates that the number of ports, N, at the FAS of each UE improves the performance by increasing the mean of the desired channel gain, see (29).

Lemma 1: Given a Gaussian random variable X with mean μ and variance σ^2, the pdf of $Z = X^2$ is given by

$$f_Z(z) = \frac{1}{2\sigma^2}\left(\frac{z}{\mu^2}\right)^{-\frac{1}{4}} e^{-\frac{z+\mu^2}{2\sigma^2}} I_{-\frac{1}{2}}\left(\frac{\mu}{\sigma^2}\sqrt{z}\right), \ z \geq 0, \quad (31)$$

where $I_r(u)$ is the modified Bessel function of the first kind and order r [46, (2.13)].

Proof: See Appendix C. ■

Theorem 3: The pdf of $\alpha_{\rm I}$ is given by

$$f_{\alpha_{\rm I}}(\alpha) = \frac{1}{2\sigma_1^2}\left(\frac{\alpha}{\left(\frac{N}{2}\sqrt{\frac{\Omega}{\pi}}\right)^2}\right)^{-\frac{1}{4}} \times e^{-\frac{\alpha+\left(\frac{N}{2}\sqrt{\frac{\Omega}{\pi}}\right)^2}{2\sigma_1^2}} I_{-\frac{1}{2}}\left(\frac{N}{2\sigma_1^2}\sqrt{\frac{\Omega}{\pi}}\sqrt{\alpha}\right), \ \alpha \geq 0, \quad (32)$$

where $\sigma_1^2 \triangleq \text{var}[\sqrt{\alpha_{\rm I}}]$ which is given by (30).

Proof: The result is obtained by applying Theorem 2 and Lemma 1 together, with the fact that N is very large. ■

Theorem 4: For sufficiently large N, $\tilde{Y}_i$ is approximately Gaussian with zero mean and

$$\sigma_2^2 \triangleq \text{var}[\tilde{Y}_i] = \frac{\Omega}{4}\left(N + \sum_{m=2}^{N}\sum_{k=1}^{m-1}\rho_{k,m}\right). \quad (33)$$

Proof: See Appendix D. ■

Corollary 1: The variable $\tilde{\beta}_{\rm I} = \frac{1}{\sigma_2^2}\beta_{\rm I}$ is central chi-square distributed with I degrees of freedom and has the pdf

$$f_{\tilde{\beta}_{\rm I}}(\tilde{\beta}) = \frac{1}{2^{\frac{I}{2}}\Gamma\left(\frac{I}{2}\right)}\tilde{\beta}^{\frac{I}{2}-1}e^{-\frac{\tilde{\beta}}{2}}, \ \tilde{\beta} \geq 0. \quad (34)$$

Proof: According to Theorem 4, we realize that $\frac{1}{\sigma_2}\tilde{Y}_i$ is a standard Gaussian random variable. Therefore, $\tilde{\beta}_{\rm I}$ is a sum of I i.i.d. squared standard Gaussian random variables, which gives a central chi-square random variable with I degrees of freedom having the pdf (34) [46, p. 24]. ■

Theorem 5: The pdf of $Z_{\rm I} = \frac{\alpha_{\rm I}}{\tilde{\beta}_{\rm I}}$ is given by (35) (see top of next page), where $\mathcal{M}_{a,b}(t)$ is the Whittaker M function given by [47, Section 9.220 on p. 1024]

$$\mathcal{M}_{a,b}(t) = t^{b+\frac{1}{2}}e^{-\frac{t}{2}}\phi\left(b-a+\frac{1}{2}, 2b+1; t\right), \quad (36)$$

where $\phi(\cdot,\cdot;\cdot)$ is the confluent hypergeometric function.

Proof: See Appendix E. ■

Lemma 2: The bit error rate (BER) of a typical CUMA UE using BPSK is given by

$$p_e = \frac{1}{2}\int_0^\infty \text{erfc}\left(\frac{\sqrt{z}}{\sigma_2}\right) f_Z(z)dz, \quad (37)$$

where $\text{erfc}(\cdot)$ is the complementary error function and

$$f_Z(z) = \int_0^z f_{Z_{\rm I}}(x) f_{Z_{\rm I}}(z-x)dx, \quad (38)$$

in which $f_{Z_{\rm I}}(z)$ is given in (35) by Theorem 5.

Proof: Given the channels fixed and ignoring the additive noise, the received in-phase and quadrature signals at a typical UE (dropping the UE index) are, respectively, given by

$$\begin{cases} y_{\rm I} = \sqrt{\alpha_{\rm I}}s + \tilde{\eta}_{\rm I}, \\ y_{\rm Q} = \sqrt{\alpha_{\rm Q}}s + \tilde{\eta}_{\rm Q}, \end{cases} \quad (39)$$

where $\tilde{\eta}_{\rm I}$ and $\tilde{\eta}_{\rm Q}$ are, respectively, the resulting additive noises for the in-phase and quadrature components. From Theorem 4, they are independent Gaussian distributed with zero mean and variances of $\sigma_s^2\beta_{\rm I}$ and $\sigma_s^2\beta_{\rm Q}$.

The optimal receiver for detection is to employ maximal-ratio combining (MRC) to mix $y_{\rm I}$ and $y_{\rm Q}$ which results in the equivalent output signal

$$y = \left(\sqrt{\frac{\alpha_{\rm I}}{\beta_{\rm I}} + \frac{\alpha_{\rm Q}}{\beta_{\rm Q}}}\right)s + \sigma_s\eta, \quad (40)$$

where η is a standard Gaussian random variable.

As a result, the BER for a given channel is found as

$$\begin{aligned} \varepsilon &\overset{(a)}{\approx} \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\sigma_s^2\left(\frac{\alpha_{\rm I}}{\beta_{\rm I}} + \frac{\alpha_{\rm Q}}{\beta_{\rm Q}}\right)}{\sigma_s^2}}\right) \\ &\overset{(b)}{=} \frac{1}{2}\text{erfc}\left(\frac{1}{\sigma_2}\sqrt{\frac{\alpha_{\rm I}}{\tilde{\beta}_{\rm I}} + \frac{\alpha_{\rm Q}}{\tilde{\beta}_{\rm Q}}}\right), \\ &\overset{(c)}{=} \frac{1}{2}\text{erfc}\left(\frac{1}{\sigma_2}\sqrt{Z_{\rm I} + Z_{\rm Q}}\right), \\ &\overset{(d)}{=} \frac{1}{2}\text{erfc}\left(\frac{\sqrt{Z}}{\sigma_2}\right), \end{aligned} \quad (41)$$

in which (a) utilizes the standard BER expression for BPSK in additive white Gaussian noise (AWGN) channels [49], and (b) adopts the definition in Corollary 1. Note that $Z_{\rm I}$ and $Z_{\rm Q}$ in (c) are i.i.d. and each has the pdf (35). The random variable in (d), $Z = Z_{\rm I} + Z_{\rm Q}$, hence has the pdf given by (38). Finally, the average BER expression is obtained by averaging (41) over the pdf given in (38), which completes the proof. ■

Corollary 2: With a binary symmetric channel, the data rate for a U-user CUMA network using BPSK is given by

$$R = U\left[1 - p_e\log_2\frac{1}{p_e} - (1-p_e)\log_2\frac{1}{1-p_e}\right], \ (\text{bits/channel-use}), \quad (42)$$

where p_e is found by (37) in Lemma 2.

Proof: Since the UEs are all i.i.d., the network rate is the product between the number of UEs, U, and the data rate of a typical UE. Assuming a binary symmetric channel, the data rate of a typical UE can be obtained using [50, p. 15]. ■

Corollary 3: For high SNR, the ergodic rate of the CUMA network can be found as

$$C_{\rm e} = U\int_0^\infty \log_2\left(1 + \frac{z}{\sigma_2^2}\right) f_Z(z)dz \ (\text{bits/channel-use}). \quad (43)$$

Proof: The output SIR of a CUMA UE is $\frac{z}{\sigma_2^2}$ which has the pdf given by (38). Therefore, the ergodic rate of the whole

$$f_{Z_{\rm I}}(z)=\frac{\Gamma\left(\frac{I+1}{2}\right)}{\Gamma\left(\frac{I}{2}\right)\Gamma\left(\frac{I}{2}\right)}\left(\frac{1}{2^{\frac{I}{2}}}\right)\left(\frac{N}{2}\sqrt{\frac{\Omega}{\pi}}\right)^{-\frac{1}{2}} z^{-\frac{3}{4}}e^{-\frac{1}{4\sigma_1^2}\left(\frac{N}{2}\sqrt{\frac{\Omega}{\pi}}\right)^2\left(\frac{2\sigma_1^2+z}{\sigma_1^2+z}\right)} \times\left(\frac{2}{1+\frac{z}{\sigma_1^2}}\right)^{\frac{2I+1}{4}}\mathcal{M}_{-\frac{2I+1}{4},-\frac{1}{4}}\left(\frac{\left(\frac{N}{2}\sqrt{\frac{\Omega}{\pi}}\right)^2 z}{2\sigma_1^2\left(\sigma_1^2+z\right)}\right),\ z\geq 0, \quad (35)$$

CUMA network with i.i.d. UEs is given by (43). Note that (38) is valid as long as a real-valued modulation scheme is used. Also, achieving (43) would require the BS to adapt the coding rates to the UEs accordingly. ■

Corollary 4: The ε-outage rate of the CUMA network can be obtained by

$$C_{\rm out}=U\log_2\left(1+\gamma_{\rm th}\right)\ \ (\text{bits/channel-use}), \quad (44)$$

where $\gamma_{\rm th}$ satisfies the following condition for a given ε:

$$\int_0^{\sigma_2^2\gamma_{\rm th}} f_Z(z)dz=\varepsilon. \quad (45)$$

Proof: By definition, the outage rate is given by (44) such that the outage probability satisfies

$$\text{Prob}\left(\frac{z}{\sigma_2^2}<\gamma_{\rm th}\right)=\varepsilon. \quad (46)$$

With the pdf of z in (38) and after some simplifications, (46) can be rewritten as (45), which completes the proof. ■

V. Simulation Results

In this section, computer simulation results are reported to understand the effectiveness of CUMA in the downlink setting. The parameters used in the simulations are given in TABLE I considering typical scenarios. For example, the size of the 2D FAS at each UE was assumed to be $15\ \text{cm}\times 8\ \text{cm}$. Each port represents a fixed half-wavelength dipole so that the mutual coupling matrix in (23) can be obtained. Also, the termination impedance was set to be $Z_T=50$ ohms. The 6 GHz case is meant to represent the 5G Mid-band while the 26 GHz and 40 GHz cases are motivated for the upcoming 5G millimeter-wave bands. Notice that the channel parameters (K,N_p) are carefully chosen to reflect the characteristics of the frequency.

The average SNR, $\Gamma\triangleq\frac{\sigma_s^2\Omega}{\sigma_\eta^2}$, was set to be 50 dB, meaning that we focused on the interference-limited scenarios in which noise had a negligible effect. In addition, our attention will be on the large-U massive connectivity cases. Results for slow FAMA without mutual coupling are provided as a benchmark for comparison [28]. It is worth pointing out that both CUMA and slow FAMA do not need any CSI at the BS. Nevertheless, the CSI requirement at each UE for both schemes is slightly different. For slow FAMA, each UE needs to know the SINR at all of its ports for selecting the best one for reception while for a CUMA UE, it only needs to know its local CSI at all the ports to find the set $\mathcal{K}$ for reception. Orthogonal multiple access (OMA) is another benchmark which divides the whole bandwidth into U sub-bands so that all UEs see no inter-user interference. For this reason, OMA has the same data rate as a single-user system. Figs. 4 and 5 assume QPSK before we consider other modulation schemes in Figs. 6, 8 and 9.

A. Rate Performance under Finite Scattering

The results in Fig. 4 are provided to investigate the average data rate performance of CUMA at different frequencies. Note that the conditions of the channels considered reflect the finite scattering phenomenon at those frequencies and the effect of mutual coupling between the antenna ports is fully considered. We like to begin by commenting the results in Fig. 4(a) that at 6 GHz (and lower frequencies), CUMA does not work well and slow FAMA prevails. The relatively poor performance of CUMA is due to the fact that at this frequency, the effective size of FAS is too small, rendering the received signals at the ports too strongly correlated. That said, intriguingly, the data rate of CUMA does not seem to drop when the number of UEs increases. The same cannot be said for slow FAMA whose rate falls if there are too many UEs to deal with. Also, the results show that slow FAMA at this frequency is able to cope with 3 UEs. Another observation is that the data rate of OMA remains constant regardless of the number of UEs since the more the UEs the less the bandwidth each UE gets. Both slow FAMA and CUMA are also shown to outperform OMA though supporting a large number of UEs is still not viable.

Fig. 4(b) and 4(c) illustrate the data rate results at 26 GHz. In Fig. 4(b), the results assume CUMA to have the antenna ports with the minimum separation of half of the wavelength while Fig. 4(c) considers the compact case where the antenna ports are more densely packed. Now, if we focus on the results in Fig. 4(b) first, we see that CUMA performs brilliantly and much better than slow FAMA. While the rate of slow FAMA drops if the number of UEs becomes too large, that of CUMA continues to increase although it seems to plateau when the number of UEs is very large. The results also reveal that $\rho=0.4$ appears to perform best and having the limit $N_{\max}=200$ does not seem to degrade the rate performance. The superior performance of CUMA over slow FAMA can be explained by the fact that it is more effective to use a massive number of ports to average out the interference than find a port in which the aggregate interference suffers from a deep fade.[8] On the other hand, in Fig. 4(c), we place the results of the compact case on top of that of the non-compact case (the latter results are shown in light grey colour), to contrast their performance.

[8]This comment is only valid when CUMA is compared to slow FAMA. If fast FAMA is to be compared, this comment will not be true as it is a lot more likely to find the interference null in fast FAMA.

TABLE I
SIMULATION PARAMETERS FOR CUMA UNDER FINITE SCATTERING SCENARIOS

		f (GHz)			
		6	26		40
Wavelength, λ (cm)		5	1.15		0.75
Size, $\bar{W}^{\ddagger}$		$3\lambda \times 1.6\lambda$	$13\lambda \times 7\lambda$		$20\lambda \times 10\lambda$
Ports, N		150×3	25×13	64×13	40×21
Rice factor, K		0	7		
No. of non-LoS paths, N_p		50	2		
Dipole[§]	Length	0.5λ			
	Width	0.005λ			
	RowSpacing[♯]	0.02λ	0.5λ	0.2λ	0.5λ
	ColumnSpacing	0.5λ			

[‡]Regardless of the frequency, the actual size of the 2D FAS is set to 15 cm $\times$ 8 cm which is roughly the size of a typical mobile phone.
[§]A rectangular array of dipoles is assumed to form the UE's 2D FAS. Other types of antennas such as slot antenna are also possible. The parameters follow from the handle of Antenna Toolbox in Matlab. The commands such as `sparameters` and `rfparam` are used to compute $\mathbf{Z}$, in (23).
[♯]Note that different level of compactness is considered by different distances of RowSpacing.

TABLE II
SIMULATION PARAMETERS FOR CUMA UNDER RICH SCATTERING SCENARIOS

		Compactness or port density, $N_1 \times N_2$		
		I	II	III
f (GHz)	6	7×3	31×3	61×3
	26	27×13	131×13	261×13
	40	41×21	201×21	401×21

The size of the 2D FAS is still 15 cm $\times$ 8 cm regardless of the frequency.
Case I is the non-compact case with minimum spacing of 0.5λ.
Case II is the compact case with minimum spacing of 0.1λ.
Case III is the very compact case with minimum spacing of 0.05λ.

The results illustrate that packing denser antenna ports at the UE does not seem desirable nor harmful as the performance in both cases is very similar. This indicates that mutual coupling does not cause any issue in CUMA. Finally, it is important to realize that the electrical size of the FAS plays a crucial role, which gives rise to the performance difference at 6 GHz and 26 GHz. The results in Fig. 4(d) further confirm this point as we can observe that the rate increases at 40 GHz. Overall, it is astonishing to see that with sufficient electrical size of FAS, CUMA can support tens of UEs on the same channel without CSI at the BS nor multiuser detection at the UEs.

Fig. 5 evaluates how the performance of CUMA changes for different values of $N_{\max}$ while we study the scenarios of different modulation schemes in Fig. 6. The results demonstrate that evidently, the performance improves as $N_{\max}$ increases and $N_{\max} = 100$ appears sufficient to get close to the best performance. On the other hand, we see from the results that CUMA is not particularly effective if higher-level modulation is considered and the rate performance drops quite clearly.

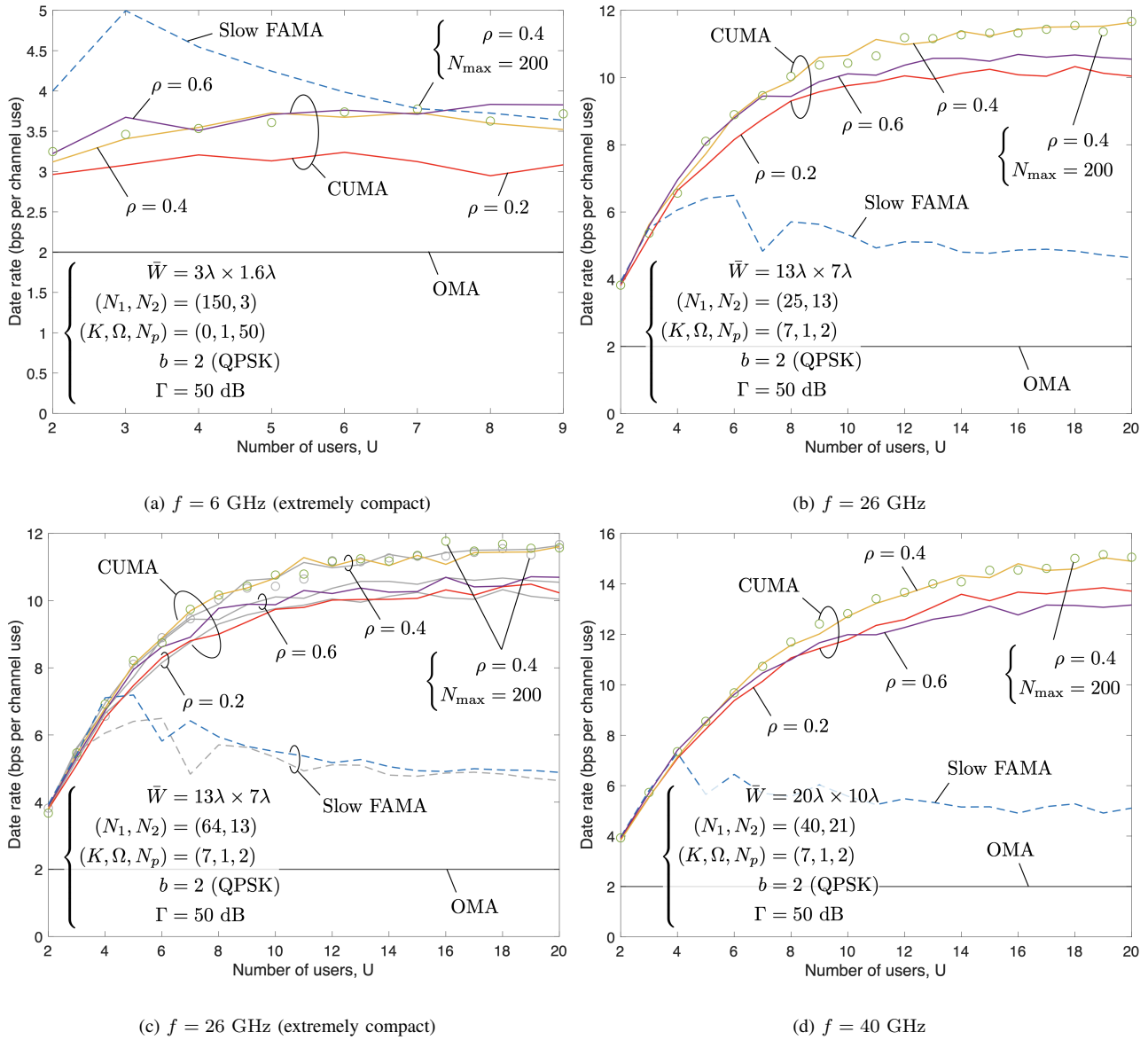
B. Theoretical Performance under Rich Scattering

In this subsection, we consider the rich scattering scenarios (i.e., $K = 0$ and $N_p = \infty$) and apply the analytical results to assess the performance of CUMA. Note that even in the millimeter-wave bands where multipath is few, it is advocated in [17] that in the reconfigurable intelligent surface (RIS) era, there can be a huge number of artificial scatterers, converting the channel into a rich scattering environment. In other words, the rich scattering situation is actually relevant. Our interest will be to study the impact of compactness or port density at each FAS-aided UE on the rate performance. In TABLE II, we detail the configurations considered.

Fig. 7 displays the pdfs obtained from Monte Carlo simulations and the analytical expression in (38) (and hence (35)) for different combinations of the system parameters. Here, the 1D FAS case with 2D scattering was considered. The results confirm that the analytical results are correct and accurate.

The results in Fig. 8 are provided for the average data rate of CUMA when BPSK is considered. As can be observed, when the size of FAS increases (i.e., at higher frequency), the rate performance improves greatly, as expected. Besides, if the number of antenna ports, N, increases, the results reveal that there will be a significant jump in the rate performance which is not seen in the results of Fig. 4 when N_p is very small. In other words, under rich scattering, packing the antenna ports more densely indeed is beneficial. Note that for Case II, the adjacent antenna ports are only 0.1λ apart while Case III is the very compact case, pushing the spacing to as small as 0.05λ apart. The compactness of FAS appears to be a key ingredient for the extraordinary performance of CUMA although there will be a diminishing return by further increasing the level of compactness, as observed in Fig. 8 at all the frequencies. On the other hand, note that under this setting, the upper limit of the average data rate of the CUMA network is U. The results illustrate that CUMA with a good level of compactness (i.e., 0.1λ spacing) and size of FAS can get close to the rate limit. Additionally, in the millimeter-wave band, we see that CUMA can handle hundreds of UEs or more on the same channel, with the rate increasing monotonically with the number of UEs.

Fig. 9 shows the ergodic rate results which further reinforce our understanding in the superiority of CUMA. It should be noted that as far as the ergodic rate is concerned, capacity-achieving codec is assumed but in our case, only real-valued coding is considered due to the analysis of (38). The results in this figure clearly illustrate that the ergodic rate of CUMA improves considerably if the size and/or the resolution (i.e., compactness) of FAS at the UEs increases. Moreover, as seen previously, the rate of CUMA keeps increasing as the number of UEs increases, suggesting that CUMA can be the massive connectivity solution that requires no CSI at the BS nor fancy interference cancellation at each UE. Besides, comparing the results to that in Fig. 8, ergodic rates are much higher and the benefits of compactness appear to be more dramatic.



(a) $f = 6$ GHz (extremely compact)

(b) $f = 26$ GHz

(c) $f = 26$ GHz (extremely compact)

(d) $f = 40$ GHz

Fig. 4. The rate performance of CUMA at different frequencies.

C. Further Discussion

We conclude this section by highlighting several interesting facts about CUMA. First, some may wonder why the proposed approach is named 'CUMA' even if apparently, some results in Fig. 4 considered antenna ports with minimum distance of half wavelength apart. It should be pointed out that a separation of half-λ is still compact in the case with few multipaths. For example, in those results with $N_p = 2$, the spatial correlation is much stronger than what is expected in rich scattering. Thus, a half-λ separation does not ensure independent signals at the antenna ports. This is why such antenna port distribution is still regarded as compact. Besides, under rich scattering, compact antenna ports appear to excel, as already shown in Figs. 8 and 9. Another useful fact for CUMA is that its performance is dependent on but insensitive to the change of the interference signals. The reason is that CUMA relies upon no knowledge of the interference's CSI and the interference is mitigated by being averaged over a massive number of antenna ports at the UE. In addition, there is no repercussion if the antenna ports are overly compact because mutual coupling between the ports does not seem too detrimental to CUMA. Of particular importance is the case of rich scattering in which CUMA has been shown to be able to accommodate hundreds of UEs per channel use if the spacing between adjacent antenna ports is as small as 0.1λ. In this case, the mutual coupling effects can be well addressed by the state-of-the-art [41], [42]. Motivated by the encouraging results under rich scattering, future work should look into ways, e.g., [17], to convert finite scattering environments into rich scattering ones so that we obtain the full benefits of CUMA for massive connectivity.

VI. Conclusion

This paper proposed a new open-loop massive connectivity scheme that could support tens or even hundreds of UEs on the same channel without requiring precoding optimization at the BS nor fancy interference cancellation receivers at the UEs. The proposed approach is based on adopting a large compact antenna array, likening FAS, referred to as CUMA, at each UE.

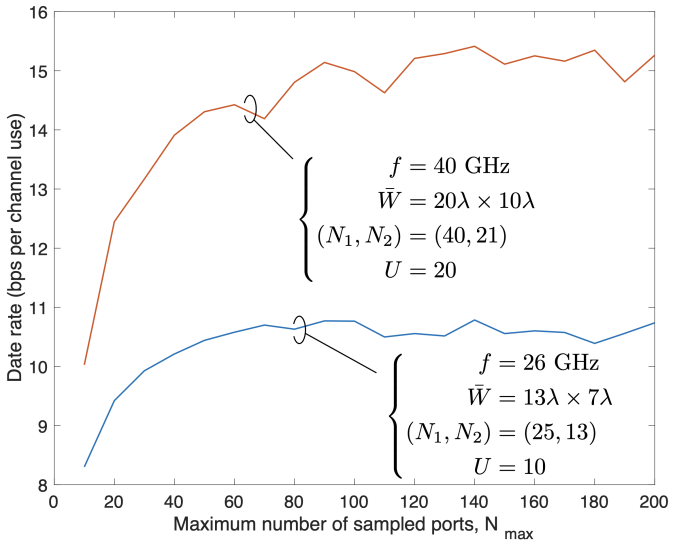


Fig. 5. The rate performance of CUMA with different values of $N_{\max}$ under the channel parameters $(K, N_p) = (7, 2)$, $\rho = 0.4$ and QPSK modulation.

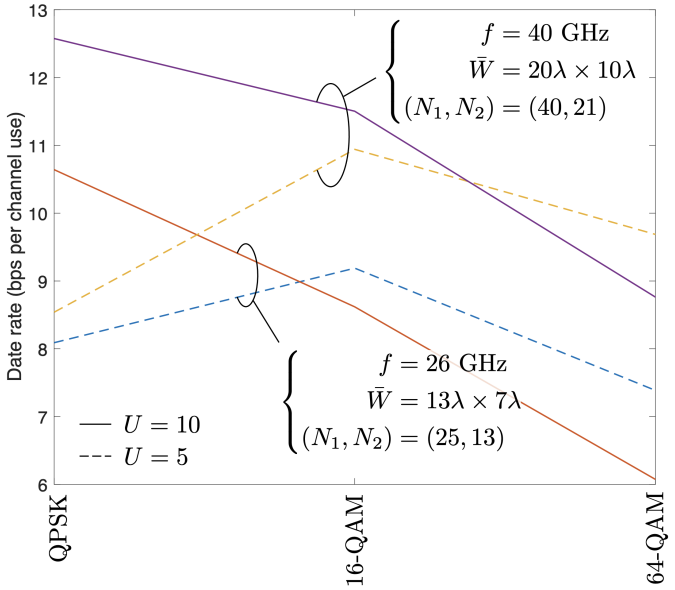


Fig. 6. Rate comparison of CUMA for different modulation schemes (i.e., with different values of b) with $(K, N_p) = (7, 2)$, $\Gamma = 50$ dB and $\rho = 0.4$.

We proposed schemes that cleverly activate certain antenna ports to ensure the desired signal is added constructively while the interference is aggregated randomly. We provided the SIR and rate analysis of CUMA. Our numerical results confirmed the great potential of CUMA as a massive connectivity solution, especially in rich scattering environments.

Appendix A
Proof of Theorem 1

We first present the following useful lemmas.

Lemma 3: $I_1 = \int_0^\infty x e^{-\frac{(x-\mu)^2}{\sigma^2}} dx$ can be evaluated as

$$I_1 = \frac{\sigma^2}{2} e^{-\frac{\mu^2}{\sigma^2}} + \sqrt{\pi}\mu\sigma Q\left(-\frac{\sqrt{2}\mu}{\sigma}\right), \tag{47}$$

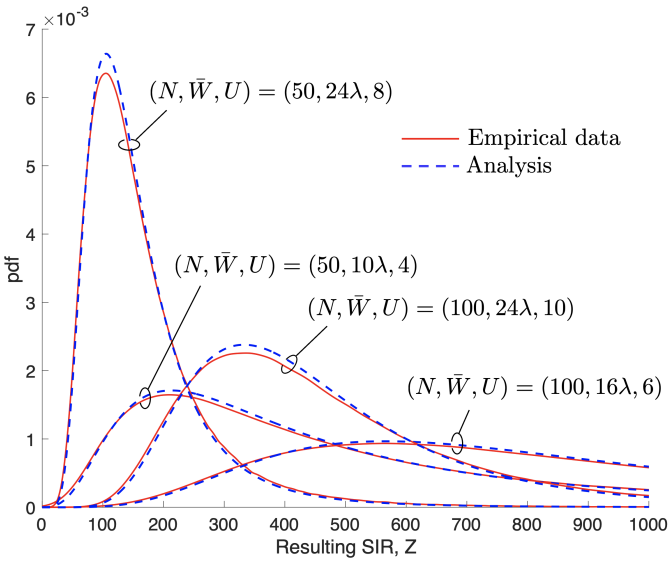


Fig. 7. Empirical and analytical pdfs for CUMA with 1D FAS.

where $Q(\cdot)$ denotes the Q function.

Proof: We start by rewriting I_1 as

$$I_1 = \int_0^\infty (x-\mu) e^{-\frac{(x-\mu)^2}{\sigma^2}} d(x-\mu) + \mu \int_0^\infty x e^{-\frac{(x-\mu)^2}{\sigma^2}} dx. \tag{48}$$

Then if we integrate the first term and also recognize that the second term can be expressed as some form of the Q function, we obtain (47), which completes the proof. ■

Lemma 4: $I_2 = \int_0^\infty y^{2c+1} e^{-by^2} Q(ay) dy$ is given by

$$I_2 = \frac{\mathcal{W}(a, b, c)}{2}, \tag{49}$$

where $\mathcal{W}(a, b, c)$ is given in (28).

Proof: Given the function $\mathcal{W}(a, b, c)$ defined in [43, (2)], we have

$$\mathcal{W}(a, b, c) = \int_0^\infty Q(a\sqrt{x}) e^{-bx} x^c dx \tag{50}$$

which can be rewritten by substitution $\sqrt{x} = y$ as

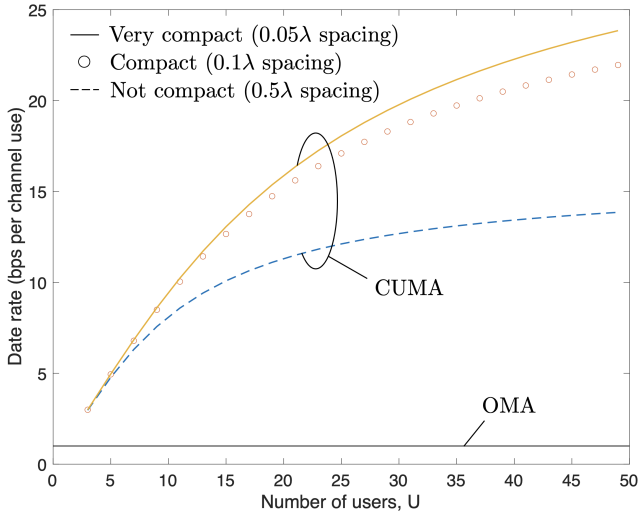
$$\mathcal{W}(a, b, c) = 2\int_0^\infty y^{2c+1} e^{-by^2} Q(ay) dy, \tag{51}$$

which gives (49). Note that (28) comes from [43, (8)]. ■

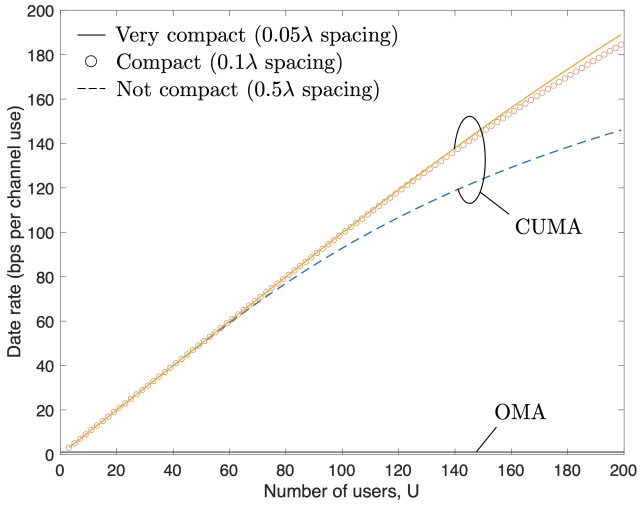
We start off the proof by recognizing that the two random variables, $x = X_k$ and $y = X_m$, are jointly Gaussian but are correlated with the correlation coefficient given by

$$\rho_{k,m} = \frac{\mathrm{E}\left[(X_k - 0)(X_m - 0)\right]}{\sqrt{\frac{\Omega}{2}}\sqrt{\frac{\Omega}{2}}} = \mathcal{J}_0\left(2\pi\sqrt{\left(\frac{(n_1 - n_3)W_1}{N_1 - 1}\right)^2 + \left(\frac{(n_2 - n_4)W_2}{N_2 - 1}\right)^2}\right), \tag{52}$$

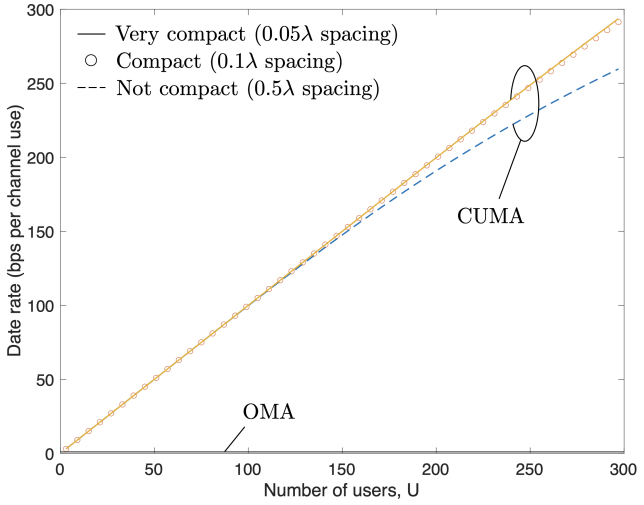
where $k = \mathrm{map}(n_1, n_2)$ and $m = \mathrm{map}(n_3, n_4)$ according to the 2D geometry of the FAS. Hence, the joint pdf of x and y



(a) $f = 6$ GHz

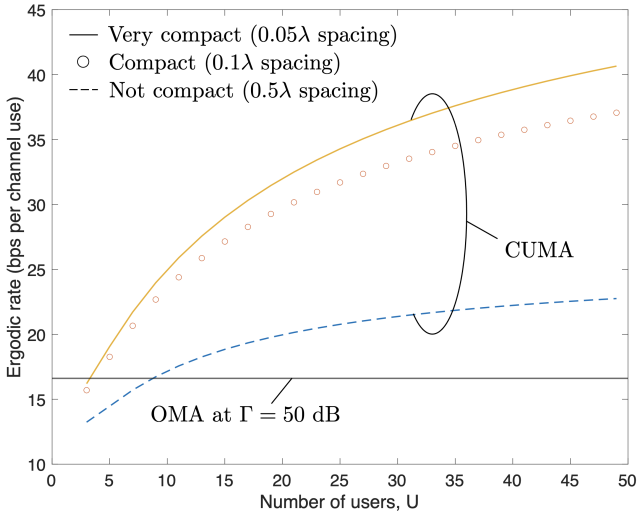


(b) $f = 26$ GHz

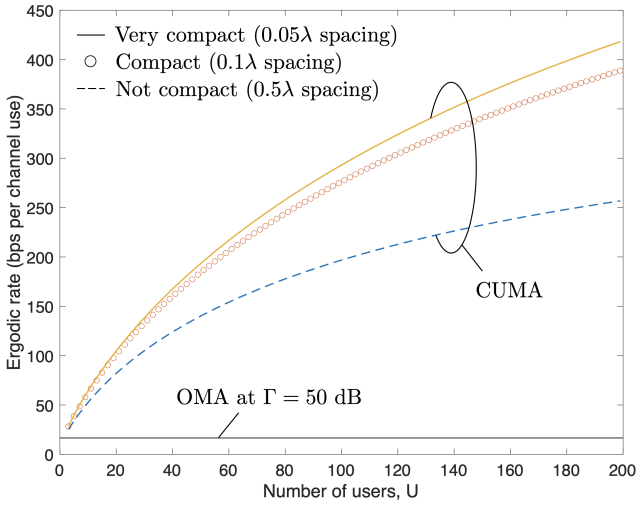


(c) $f = 40$ GHz

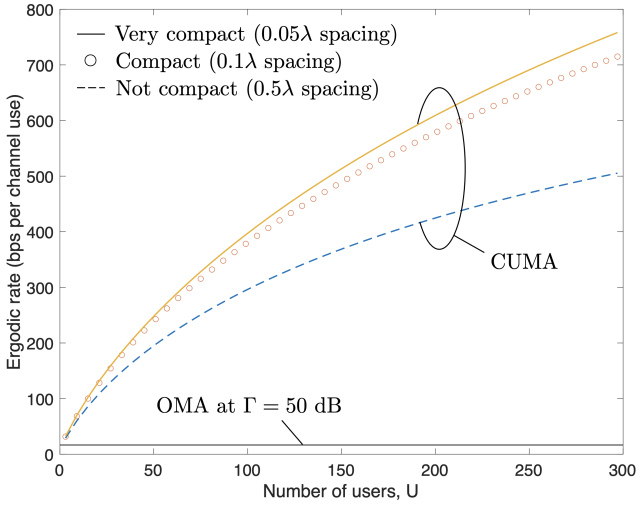
Fig. 8. Average data rate performance of CUMA with BPSK modulation in the absence of mutual coupling and under rich scattering.



(a) $f = 6$ GHz



(b) $f = 26$ GHz



(c) $f = 40$ GHz

Fig. 9. Ergodic rate performance of CUMA assuming real-valued Gaussian signling in the absence of mutual coupling and under rich scattering.

can be found as

$$f_{XY}(x,y) = \frac{1}{\pi\Omega\sqrt{1-\rho_{k,m}^2}} e^{-\frac{1}{1-\rho_{k,m}^2}\left(\frac{x^2-2\rho_{k,m}xy+y^2}{\Omega}\right)}. \tag{53}$$

As a result, we have

$$\mathrm{E}[X_k^+ X_m^+] = \int_0^\infty \int_0^\infty xy f_{XY}(x,y)dxdy, \tag{54}$$

which can be evaluated by

$$\mathrm{E}[X_k^+ X_m^+] = \int_0^\infty \frac{ye^{-\frac{y^2}{\Omega}}}{\pi\Omega\sqrt{1-\rho_{k,m}^2}} \int_0^\infty xe^{-\frac{(x-\rho_{k,m}y)^2}{(1-\rho_{k,m}^2)\Omega}} dxdy. \tag{55}$$

Using Lemma 3 to express the inner integral, we can write

$$\mathrm{E}[X_k^+ X_m^+] = \frac{\sqrt{1-\rho_{k,m}^2}}{2\pi} \int_0^\infty ye^{-\frac{y^2}{(1-\rho_{k,m}^2)\Omega}} dy + \frac{\rho_{k,m}}{\sqrt{\pi\Omega}} \int_0^\infty y^2 e^{-\frac{y^2}{\Omega}} Q\left(-\sqrt{\frac{2}{1-\rho_{k,m}^2}}\frac{\rho_{k,m}}{\sqrt{\Omega}}y\right) dy. \tag{56}$$

The first term of the above can be evaluated using the identity $\int_0^\infty ye^{-\frac{y^2}{\sigma^2}}dy = \frac{\sigma^2}{2}$ while the second term can be found using the result in Lemma 4, which gives

$$\mathrm{E}[X_k^+ X_m^+] = \frac{(1-\rho_{k,m}^2)^{\frac{3}{2}}\Omega}{4\pi} + \frac{\rho_{k,m}}{2\sqrt{\pi\Omega}} \mathcal{W}\left(-\sqrt{\frac{2}{1-\rho_{k,m}^2}}\frac{\rho_{k,m}}{\sqrt{\Omega}}, \frac{1}{\Omega}, \frac{1}{2}\right). \tag{57}$$

Finally, the mean of the rectified Gaussian random variable is $\mathrm{E}[X_k^+] = \frac{1}{2}\sqrt{\frac{\Omega}{\pi}}$ [44, p. 90]. As a result, we can obtain

$$\begin{aligned}
&\mathrm{cov}(X_k^+, X_m^+) \\
&= \mathrm{E}\left[\left(X_k^+ - \frac{1}{2}\sqrt{\frac{\Omega}{\pi}}\right)\left(X_m^+ - \frac{1}{2}\sqrt{\frac{\Omega}{\pi}}\right)\right] \\
&= \mathrm{E}[X_k^+ X_m^+] - 2\mathrm{E}[X_k^+]\left(\frac{1}{2}\sqrt{\frac{\Omega}{\pi}}\right) + \frac{\Omega}{4\pi} \\
&= \mathrm{E}[X_k^+ X_m^+] - \frac{\Omega}{4\pi}.
\end{aligned} \tag{58}$$

Substituting (57) into (58) gives the final result (26).

Appendix B
Proof of Theorem 2

First, (29) is obtained by knowing that $\mathrm{E}[X_k^+] = \frac{1}{2}\sqrt{\frac{\Omega}{\pi}}$. Then the variance of $\sqrt{\alpha_\mathrm{I}}$ can be derived using

$$\mathrm{var}[\sqrt{\alpha_\mathrm{I}}] = N\mathrm{var}[X_m^+] + 2\sum_{m=2}^{N}\sum_{k=1}^{m-1}\mathrm{cov}(X_k^+, X_m^+), \tag{59}$$

which is the classical result for finding the variance of a sum of dependable random variables. Using [44, p. 90], we have

$$\begin{aligned}
\mathrm{var}[X_m^+] &= \left(0+\frac{\Omega}{2}\right)\left(1-\frac{1}{2}\right) + 0 - \left(\frac{1}{2}\sqrt{\frac{\Omega}{\pi}}\right)^2 \\
&= \frac{\Omega}{4}\left(1-\frac{1}{\pi}\right).
\end{aligned} \tag{60}$$

After substituting (60) into (59), we get (30).

The remaining task is to show that under common conditions, when N is very large, $\sqrt{\alpha_\mathrm{I}}$ is approximately Gaussian. Since $\{X_m^+\}$ are dependent random variables, we will use the Lyapunov or Lindeberg version of the Central Limit Theorem (CLT) [45]. To achieve this, we argue that for reasonably large $\bar{W}$, the random variables X_m^+ far apart from one another are nearly independent. According to the definition in [45, p. 364], the sequence $\{X_m^+\}$ is said to be m-dependent (the concept to characterize the level of independency). Then under the condition of a finite twelfth moment, i.e., $\mathrm{E}[(X_m^+)^{12}] < \infty$, [45, Theorem 27.4] adopted the Lyapunov condition to prove that the series $\sqrt{\alpha_\mathrm{I}}$ is asymptotically normal.

Appendix C
Proof of Lemma 1

Let Z be a standard Gaussian random variable, i.e., $\bar{X} \sim \mathcal{N}(0,1)$. Then $Y = (\xi+\bar{X})^2$ is a noncentral chi-square random variable with one degree of freedom, with the pdf [46, p. 26]

$$f_Y(y) = \frac{1}{2}\left(\frac{y}{\xi^2}\right)^{-\frac{1}{4}} e^{-\frac{y+\xi^2}{2}} I_{-\frac{1}{2}}(\xi\sqrt{y}),\ y>0, \tag{61}$$

where $I_r(u)$ is the modified Bessel function of the first kind and order r. From the definition in this lemma, we have

$$Z = X^2 = \sigma^2\left(\frac{\mu}{\sigma}+\bar{X}\right)^2 = \sigma^2 Y. \tag{62}$$

As $f_Z(z) = \frac{1}{\sigma^2}f_Y\left(\frac{z}{\sigma^2}\right)$, (35) is obtained by utilizing (61) with the substitution $\xi = \frac{\mu}{\sigma}$, which completes the proof.

Appendix D
Proof of Theorem 4

To start with, we write

$$\tilde{Y}_i = \sum_{k=1}^{N} t_k Y_k^{(i)} \equiv \sqrt{\frac{\Omega}{2}}\sum_{k=1}^{N} t_k \bar{Y}_k^{(i)} \equiv \sqrt{\frac{\Omega}{2}}\sum_{k=1}^{N}\tilde{t}_k^{(i)}, \tag{63}$$

in which $\tilde{t}_k^{(i)} \triangleq t_k\bar{Y}_k^{(i)}$, t_k is a Bernoulli random variable with probability 0.5, $\bar{Y}_k^{(i)}$ is a standard Gaussian random variable, and t_k and $\bar{Y}_k^{(i)}$ are independent of each other.

The mean of $\tilde{t}_k^{(i)}$ can be easily found as

$$\mathrm{E}[\tilde{t}_k^{(i)}] = \mathrm{E}[t_k]\times\mathrm{E}[\bar{Y}_k^{(i)}] = (0.5)(0) = 0. \tag{64}$$

To find $\mathrm{var}[\tilde{t}_k^{(i)}]$, we know that the pdf of $\tilde{t}$ (with the indices k and i omitted for conciseness) is given by

$$f_{\tilde{T}}(\tilde{t}) = \begin{cases} \frac{1}{2}\delta(\tilde{t}), \text{ if } \tilde{t}=0, \\ \frac{1}{2\sqrt{2\pi}}e^{-\frac{\tilde{t}^2}{2}}, \text{ if } \tilde{t}\neq 0, \end{cases} \tag{65}$$

where $\delta(\cdot)$ is the Kronecker delta function. Then the variance of $\tilde{t}$ can be obtained by

$$\begin{aligned}\mathrm{var}[\tilde{t}] &= \mathrm{E}[\tilde{T}^2] - (\mathrm{E}[\tilde{T}])^2 \\ &= \mathrm{E}[\tilde{T}^2] - 0^2 \\ &= \int_{-\infty}^{\infty} t^2 f_{\tilde{T}}(t) dt \\ &= \int_{-\infty}^{0^-} \frac{t^2 e^{-\frac{t^2}{2}}}{2\sqrt{2\pi}} dt + \int_{0^+}^{\infty} \frac{t^2 e^{-\frac{t^2}{2}}}{2\sqrt{2\pi}} dt \\ &\overset{(a)}{=} \frac{1}{2\sqrt{2\pi}}\sqrt{\frac{\pi}{2}} + \frac{1}{2\sqrt{2\pi}}\sqrt{\frac{\pi}{2}} = \frac{1}{2}, \end{aligned} \tag{66}$$

where (a) has used the result in [47, p. 336].

As a result, we can obtain

$$\mathrm{var}[t_k Y_k^{(i)}] = \frac{\Omega}{2}\mathrm{var}[\tilde{t}_k^{(i)}] = \frac{\Omega}{4}. \tag{67}$$

Also, the covariance of each term in $\tilde{Y}_i$ can be derived by

$$\begin{aligned}&\mathrm{cov}\left(t_k Y_k^{(i)}, t_m Y_m^{(i)}\right) \\ &= \frac{\Omega}{2}\mathrm{E}[\tilde{t}_k^{(i)}\tilde{t}_m^{(i)}] \\ &\overset{(a)}{=} \frac{\Omega}{2}\mathrm{E}[t_k t_m]\mathrm{E}[\bar{Y}_k^{(i)}\bar{Y}_m^{(i)}] \\ &\overset{(b)}{=} \frac{\Omega}{2}\mathrm{E}[t_k]\mathrm{E}[t_m]\rho_{k,m} \\ &\overset{(c)}{=} \frac{\Omega}{2}\left(\frac{1}{2}\right)^2 \rho_{k,m} = \frac{\Omega\rho_{k,m}}{8}, \end{aligned} \tag{68}$$

where (a) uses the fact that t_k and $\bar{Y}_k^{(i)}$ are independent, (b) recognizes that $\{t_k\}$ are i.i.d. and $\mathrm{E}[\bar{Y}_k^{(i)}\bar{Y}_m^{(i)}] = \rho_{k,m}$ comes directly from the correlation structure of the FAS, and (c) uses $\mathrm{E}[t_k] = \frac{1}{2}$. Now, we compute the variance of $\tilde{Y}_i$ by

$$\begin{aligned}\mathrm{var}[\tilde{Y}_i] = N\mathrm{var}[t_k Y_k^{(i)}] \\ + 2\sum_{m=2}^{N}\sum_{k=1}^{m-1}\mathrm{cov}\left(t_k Y_k^{(i)}, t_m Y_m^{(i)}\right). \end{aligned} \tag{69}$$

Substituting (67) and (68) into (69) gives (33). On the other hand, it is easy to see that $\mathrm{E}[\tilde{Y}_i] = 0$ due to (64). Finally, we can use the argument in Appendix B to conclude that the series $\tilde{Y}_i$ is asymptotically normal.

Appendix E
Proof of Theorem 5

Given the pdf of α_{I} in Theorem 3 and that of $\tilde{\beta}_{\mathrm{I}}$ in Corollary 1, we here derive the pdf of $Z_{\mathrm{I}} = \frac{\alpha_{\mathrm{I}}}{\tilde{\beta}_{\mathrm{I}}}$. It is known that

$$\begin{aligned} f_{Z_{\mathrm{I}}}(z) &= \int_{-\infty}^{\infty} |\tilde{\beta}| f_{\alpha,\tilde{\beta}}(z\tilde{\beta},\tilde{\beta}) d\tilde{\beta}, \\ &\overset{(a)}{=} \int_{0}^{\infty} \tilde{\beta} f_{\alpha,\tilde{\beta}}(z\tilde{\beta},\tilde{\beta}) d\tilde{\beta}, \\ &\overset{(b)}{=} \int_{0}^{\infty} \tilde{\beta} f_{\alpha}(z\tilde{\beta}) f_{\tilde{\beta}}(\tilde{\beta}) d\tilde{\beta}, \end{aligned} \tag{70}$$

where (a) comes from the fact that $\tilde{\beta}_{\mathrm{I}} > 0$, and (b) is due to the independence between α_{I} and $\tilde{\beta}_{\mathrm{I}}$. Now, after substituting (32) and (34) into (70) and some simplifications, we get

$$\begin{aligned} f_{Z_{\mathrm{I}}}(z) = &\frac{\left(\frac{\mu^2}{z}\right)^{\frac{1}{4}} e^{-\frac{\mu^2}{2\sigma_1^2}}}{2^{\frac{I}{2}+1}\Gamma\left(\frac{I}{2}\right)\sigma_1^2} \\ &\times \int_0^{\infty} \tilde{\beta}^{\frac{2I-1}{4}} e^{-\left(1+\frac{z}{\sigma_1^2}\right)\frac{\tilde{\beta}}{2}} I_{-\frac{1}{2}}\left(\frac{\mu}{\sigma_1^2}\sqrt{z\tilde{\beta}}\right) d\tilde{\beta}, \end{aligned} \tag{71}$$

where $\mu = \frac{N}{2}\sqrt{\frac{\Omega}{\pi}}$ and $\sigma_1^2 = \mathrm{var}[\sqrt{\alpha_{\mathrm{I}}}]$ is given in (30). To proceed further, we need the following lemma.

Lemma 5: We have the integral identity [47, p. 709]

$$\begin{aligned} &\int_0^{\infty} x^{\psi-\frac{1}{2}} e^{-\zeta x} I_{2\nu}(2\vartheta\sqrt{x}) dx \\ &= \frac{\Gamma\left(\psi+\nu+\frac{1}{2}\right)}{\Gamma(2\nu+1)} \vartheta^{-1} e^{\frac{\vartheta^2}{2\zeta}} \zeta^{-\psi} \mathcal{M}_{-\psi,\nu}\left(\frac{\vartheta^2}{\zeta}\right), \end{aligned} \tag{72}$$

where $\mathcal{M}_{a,b}(t)$ is the Whittaker M function.

Proof: See [48, p. 45]. ∎

To obtain the integral in (71), we apply Lemma 5 by setting $\frac{2I-1}{4} = \psi - \frac{1}{2}$, $\zeta = \frac{1}{2}\left(1+\frac{z}{\sigma_1^2}\right)$, $2\nu = -\frac{1}{2}$ and $2\vartheta = \frac{\mu\sqrt{z}}{\sigma_1^2}$. After simplifications, we get (35) which completes the proof.

References

[1] NTT Docomo, Inc., "White paper: 5G evolution and 6G," Jan. 2020.

[2] Samsung, "6G: The next hyper-connected experience for all," Jul. 2020.

[3] Ericsson, "Ever-present intelligent communication," Nov. 2020.

[4] IMT-2030 Promotion Group, "White paper on 6G vision and candidate technologies," Jun. 2021.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[6] Z. Wang *et al.*, "Extremely large-scale MIMO: Fundamentals, challenges, solutions, and future directions," [Online] arXiv preprint arXiv:2209.12131, 2022.

[7] D. A. Urquiza Villalonga, H. OdetAlla, M. J. Fernández-Getino Garca, and A. Flizikowski, "Spectral efficiency of precoded 5G-NR in single and multi-user scenarios under imperfect channel knowledge: A comprehensive guide for implementation," *Electronics*, vol. 11, no. 24, p. 4237, Dec. 2022.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[9] Y. Saito *et al.*, "Non-orthogonal multiple access (NOMA) for cellular future radio access," *IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, 2-5 Jun. 2013, Dresden, Germany.

[10] Z. Ding *et al.*, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Select. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[11] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 1, pp. 60–70, Jan. 1978.

[12] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[13] B. Clerckx *et al.*, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE J. Select. Areas Commun.*, vol. 41, no. 5, pp. 1265–1308, May 2023.

[14] K. K. Wong, K. F. Tong, Y. Shen, Y. Chen, and Y. Zhang, "Bruce Lee-inspired fluid antenna system: Six research topics and the potentials for 6G," *Frontiers Commun. and Netw., section Wireless Commun.*, vol. 3, no. 853416, Mar. 2022.

[15] K. K. Wong, W. K. New, X. Hao, K. F. Tong, and C. B. Chae, "Fluid antenna system–Part I: Preliminaries," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 1919–1923, Aug. 2023.

[16] K. K. Wong, K. F. Tong, and C. B. Chae, "Fluid antenna system–Part II: Research opportunities," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 1924–1928, Aug. 2023.
[17] K. K. Wong, K. F. Tong, and C. B. Chae, "Fluid antenna system–Part III: A new paradigm of distributed artificial scattering surfaces for massive connectivity," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 1929–1933, Aug. 2023.
[18] Y. Huang, L. Xing, C. Song, S. Wang and F. Elhouni, "Liquid antennas: Past, present and future," *IEEE Open J. Antennas & Propag.*, vol. 2, pp. 473–487, 2021.
[19] D. Rodrigo, B. A. Cetiner and L. Jofre, "Frequency, radiation pattern and polarization reconfigurable antenna using a parasitic pixel layer," *IEEE Trans. Antennas & Propag.*, vol. 62, no. 6, pp. 3422–3427, Jun. 2014.
[20] T. V. Hoang, V. Fusco, T. Fromenteze and O. Yurduseven, "Computational polarimetric imaging using two-dimensional dynamic metasurface apertures," *IEEE Open J. Antennas & Propag.*, vol. 2, pp. 488–497, 2021.
[21] L. Jing, M. Li and R. Murch, "Compact pattern reconfigurable pixel antenna with diagonal pixel connections," *IEEE Trans. Antennas & Propag.*, vol. 70, no. 10, pp. 8951–8961, Oct. 2022.
[22] K. K. Wong, A. Shojaeifard, K. F. Tong, and Y. Zhang, "Fluid antenna systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1950–1962, Mar. 2021.
[23] L. Tlebaldiyeva, S. Arzykulov, K. M. Rabie, X. Li, and G. Nauryzbayev, "Outage performance of fluid antenna system (FAS)-aided Terahertz communication networks," in Proc. *IEEE Int. Conf. Commun. (ICC)*, 28 May-1 Jun. 2023, Rome, Italy.
[24] M. Khammassi, A. Kammoun, and M.-S. Alouini, "A new analytical approximation of the fluid antenna system channel," *IEEE Trans. Wireless Commun.*, early access, DOI:10.1109/TWC.2023.3266411.
[25] W. K. New, K. K. Wong, X. Hao, K. F. Tong, and C.-B. Chae, "Fluid antenna system: New insights on outage probability and diversity gain," *IEEE Trans. Wireless Commun.*, [Online] arXiv preprint arXiv:2301.00073, 2023.
[26] K. K. Wong and K. F. Tong, "Fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4801–4815, Jul. 2022.
[27] K. K. Wong, K. F. Tong, Y. Chen, and Y. Zhang, "Fast fluid antenna multiple access enabling massive connectivity," *IEEE Commun. Letters*, vol. 27, no. 2, pp. 711–715, Feb. 2023.
[28] K. K. Wong, D. Morales-Jimenez, K. F. Tong, and C. B. Chae, "Slow fluid antenna multiple access," *IEEE Trans. Commun.*, vol. 71, no. 5, pp. 2831–2846, May 2023.
[29] K. K. Wong, K. F. Tong, Y. Chen, Y. Zhang, and C. B. Chae, "Opportunistic fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, early access, DOI:10.1109/TWC.2023.3255940.
[30] K. K. Wong, K. F. Tong, Y. Chen, and Y. Zhang, "Extra-large MIMO enabling slow fluid antenna massive access for millimeter-wave bands," *IET Elect. Lett.*, vol. 58, no. 25, pp. 1016–1018, Dec. 2022.
[31] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, 2015.
[32] G. L. Stüber, *Principles of Mobile Communication*, Second Edition, Kluwer Academic Publishers, 2002.
[33] W. K. New, K. K. Wong, X. Hao, K. F. Tong, C. B. Chae, "An information-theoretic characterization of MIMO-FAS: Optimization, diversity-multiplexing tradeoff and q-outage capacity," accepted in *IEEE Trans. Wireless Commun.*, [Online] arXiv preprint arXiv:2303.02269, 2023.
[34] A. Pizzo, T. L. Marzetta and L. Sanguinetti, "Spatially-stationary model for holographic MIMO small-scale fading," *IEEE J. Select. Areas Commun.*, vol. 38, no. 9, pp. 1964–1979, Sept. 2020.
[35] C. Skouroumounis and I. Krikidis, "Fluid antenna with linear MMSE channel estimation for large-scale cellular networks," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1112–1125, Feb. 2023.
[36] R. Wang, Y. Chen, Y. Hou, K. K. Wong, and X. Tao, "Estimation of channel parameters for port selection in millimeter-wave fluid antenna systems," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops)*, 10-12 Aug. 2023, Dalian, China.
[37] Z. Zhou *et al.*, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Select. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.
[38] T. V. Hoang, V. Fusco, M. A. B. Abbasi, and O. Yurduseven, "Single-pixel polarimetric direction of arrival estimation using programmable coding metasurface aperture," *Nature*, Sci. Rep. 11, no. 23830, Dec. 2021.
[39] B. Clerckx, C. Craeye, D. Vanhoenacker-Janvier and C. Oestges, "Impact of antenna coupling on 2 × 2 MIMO communications," *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1009–1018, May 2007.
[40] C. A. Balanis, *Antenna Theory: Analysis and Design*, New York: Wiley, second edition, 1997.
[41] C.-Y. Chiu, C.-H. Cheng, R. D. Murch and C. R. Rowell, "Reduction of mutual coupling between closely-packed antenna elements," *IEEE Trans. Antennas & Propag.*, vol. 55, no. 6, pp. 1732–1738, Jun. 2007.
[42] A. A. Althuwayb *et al.*, "Design technique to mitigate unwanted coupling in densely packed radiating elements of an antenna array for electronic devices and wireless communication systems operating in the millimeter-wave band," *AEU – Int. J. Elect. & Commun.*, vol. 159, no. 154464, Feb. 2023.
[43] M. Bilim, "Some new results for integrals involving Gaussian Q-function and their applications to α-μ and η-μ fading channels," *Wireless Pers. Commun.*, vol. 109, no. 2, pp. 1463–1469, 2019.
[44] M. Beauchamp, "On numerical computation for the distribution of the convolution of N independent rectified Gaussian variables," *J. de la Société Française de Statistique*, vol. 159, no. 1, pp. 88–111, 2018.
[45] P. Billingsley, *Probability and Measure*, 3rd Edition, Probability and Mathematical Statistics, Wiley, 1995.
[46] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Pearson Education, 1998.
[47] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th edition, Academic Press, Elsevier, 2007.
[48] N. W. McLachlan, P. Humbert, and L. Poli, "Supplément au formulaire pour le calcul symbolique," *Mémorial des sciences mathématiques*, no. 113, 1950.
[49] J. Proakis, and M. Salehi, *Digital communications*, McGraw-Hill Education, 2007.
[50] D. J. C. MacKay, *Information theory, inference, and learning algorithms*, Cambridge University Press, 2003.