Energy Efficiency Maximization Under Delay-Outage Probability Constraints Using Fluid Antenna Systems

Yicong Xu, Yu Chen, Yanzhao Hou, Kai-Kit Wong, and Xiaofeng Tao

Abstract—**Fluid antenna system (FAS) is a new wireless technology that enables reconfigurable antenna position to enhance communication performance. In wireless networks, spectral efficiency, delay and energy efficiency are some of the most important performance indicators. To jointly optimize these indicators for FAS-assisted point-to-point communication systems, we adopt the energy efficiency (EE) metric which is defined as the delivered data rate divided by the total power consumption. The optimal power allocation strategy is obtained to maximize the EE subject to a delay-outage probability constraint. We then study the effects of delay bounds and the number of FAS's ports on the maximum EE. Simulation results are presented to show the effectiveness of the proposed delay-aware FAS-assisted system.**

Index Terms—**Cross-layer, Delay, Energy efficiency, Fluid antenna, Outage, Two-mode transceiver circuitry.**

I. Introduction

FLUID antenna system (FAS) is a new antenna technology; it utilizes a flexible location-reconfigurable antenna to aid communication [1]–[3]. By integrating FAS into the physical layer of wireless communication systems, significant diversity and multiplexing gains are shown to be possible using a single radio-frequency (RF) chain FAS receiver without the need of complex precoding optimization and/or decoding architecture. Increasing interest in FAS is evident by the recent work, e.g., [4]–[10] which investigated various benefits.

As FAS is still in its infancy, researchers should gear up to understand how FAS can be utilized to impact future wireless communication networks. In 5G and beyond systems, spectral efficiency (SE), energy efficiency (EE) and delay are three key performance indicators (KPIs) [11]. Understanding the trade-off between the KPIs in FAS would be desirable. To do so, it would be useful to review the evolution of the performance metric that has been used for wireless system designs.

Back in 1997, Zorzi and Rao [12] proposed a new EE metric which was defined as the amount of data delivered through a link divided by the consumed energy. It was usually assumed that the link had a buffer and the data rate was constant at the buffer. Moreover, the delay-outage probability (DOP) was the only delay requirement for applications. Based on the effective capacity (EC) model [13], the constant value of data rate and a required DOP together determine the EC of a time-varying channel. The constant data rate assumption usually implies the equivalence between EE and the effective EE (EEE), which is defined as the EC divided by the total power consumption, first proposed by Musavian and Le-Ngocin in [14], [15].

EEE is a unified metric that can be applied to any buffer-aided wireless network with general arrivals and capacities. It has played a critical role in the EE optimization under delay constraints for various wireless systems such as simultaneous wireless information and power transfer (SWIPT) [16], Internet-of-Things (IoT) [17], non-orthogonal multiple access (NOMA) [18] and heterogeneous cloud radio access networks (C-RAN) [19], to name a few. Motivated by this, in this paper, we aim to consider the same for FAS-based systems.

Instead of maximizing the EEE, our objective is to maximize Zorzi and Rao's EE[1] subject to the DOP constraint in a point-to-point FAS-based communication system. Particularly, we derive the EC in the FAS-based communication system. Based on the EC model, the DOP is characterized by two EC functions, namely, the quality-of-service (QoS) exponent and the nonempty buffer probability (NBP). Although no expression for the NBP has been widely accepted today, there exists a simple yet accurate approximation formula for the NBP, which was derived by Chen and Darwazeh for exponential arrival processes in 2015 [20]. This formula was adopted to optimize the DOP-constrained EE in different systems [21]–[23] and will be used in our work as well. The above EE optimization problem under the DOP constraint is a concave maximization problem, and the optimal power allocation can be determined by using a Lagrangian optimization approach.

II. System Model and Problem Formulation

Fig. 1(a) shows a simplified FAS-based point-to-point communication system model from an upper-layer data source to a data sink; the FAS is installed at the receiver. Moreover, the communication system inside the bracket of Fig. 1(a) is termed an FAS-based communication system while the details of such a system at slot n are shown in Fig. 1(b).

A. FAS

We assume that a FAS is installed at the receiver end. In particular, FAS means that a flexible fluid antenna is employed

The work is supported in part 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.

Y. Xu, Y. Chen, Y. Hou, and X. Tao are with the National Engineering Laboratory for Mobile Network Technologies, Beijing University of Posts and Telecommunications, China.

K. K. Wong is with the Department of Electronic and Electrical Engineering, University College London, Torrington Place, United Kingdom.

[1]In this work, data sizes are assumed to be exponentially distributed, i.e., non-constant. Therefore, the EE of [12] and the EEE are not equivalent.

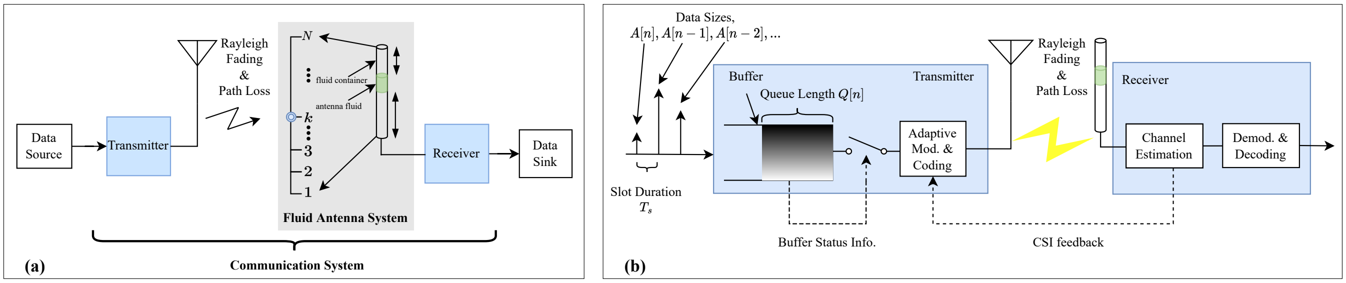

Fig. 1. A delay-constrained communication system using FAS at the receiver end: (a) System model and (b) the scheduling architecture.

for communication. The fluid antenna permits the receiver to choose one of N evenly distributed positions/ports over a given space of $W\lambda$ where W denotes the normalized size of the fluid antenna and λ is the wavelength. Based on the correlation model in [24], the received signal at the k^{th} port (for $k = 1, 2, \ldots, N$) and slot n can be written as

$$r_k[n] = h_k s[n] + \eta_k[n], \quad n = \{1, 2, \ldots\}, \tag{1}$$

where $s[n]$ and $r_k[n]$ denote the transmitted and the received signals, respectively, and $\eta_k[n]$ is a complex Gaussian random variable with zero mean and variance of σ_η^2. For the rest of this paper, the time index n is omitted only if no ambiguity is raised. In (1), h_k is the channel gain, modelled as

$$h_k = \sigma \Big[(\sqrt{1-\mu^2} x_k + \mu x_0) + j(\sqrt{1-\mu^2} y_k + \mu y_0) \Big], \; k = \{1, 2, \ldots, N\}, \tag{2}$$

in which x_0, y_0, x_k, y_k in (2) are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance of 0.5, μ denotes the spatial correlation parameter which can be chosen by [24]

$$\mu = \sqrt{2} \sqrt{ {}_1F_2\left(\frac{1}{2}; 1, \frac{3}{2}; -\pi^2 W^2\right) - \frac{J_1(2\pi W)}{2\pi W} }, \tag{3}$$

where ${}_aF_b(\cdot;\cdot;\cdot)$ is the generalized hypergeometric function and $J_1(\cdot)$ is the first-order Bessel function of the first kind.

The average received signal-to-noise ratio (SNR) at each port is given by

$$\Gamma = \sigma^2 \frac{\mathbb{E}\left[|s|^2\right]}{\sigma_\eta^2}, \tag{4}$$

where $\mathbb{E}[\cdot]$ is an expectation operator. In order to obtain the best performance at a given slot, the FAS selects the port with the highest channel power gain, i.e.,

$$g^* = \max\{|h_1|^2, |h_2|^2, \ldots, |h_N|^2\}. \tag{5}$$

B. Upper Layer Queueing Model

The communication system can be described as a discrete-time queueing model as shown in Fig. 1(b). A fluid model is assumed, i.e., the length of a packet is infinitesimal. Let us denote the system bandwidth by B, the transmit power by P_t, and the noise spectral density by N_0. This system is discrete-time and has the slot duration T_s. Provided the path loss is L_p and the noise variance σ_η^2 is $N_0 B$ [25], the average SNR in (4) is given by

$$\Gamma = \sigma^2 \frac{\mathbb{E}\left[|s|^2\right]}{\sigma_\eta^2} = \sigma^2 \frac{P_\mathrm{t}}{N_0 B} \xrightarrow[\sigma^2]{L_p^{-1}} \frac{1}{L_p} \frac{P_\mathrm{t}}{N_0 B} = \frac{P_\mathrm{t}}{L_p \sigma_\eta}. \tag{6}$$

Upper-layer data from the data source are first pushed into a first-in-first-out (FIFO) buffer at the transmitter; and then transmitted to the receiver over a block-fading channel (i.e., the channel gains $\{h_k\}_{\forall k}$ remain constant during a time slot). Under this model, we write

$A[n]$ for the amount of data in bits from the data source at slot n. The random variables $A[1], A[2], \ldots$ are assumed to be i.i.d. exponential distributed, each with the probability density function (PDF)

$$f_A(a) = \begin{cases} \lambda \exp(-\lambda a) & \text{if } a \geq 0, \\ 0 & \text{if } a < 0. \end{cases} \tag{7}$$

The average data rate is $R = (T_s \lambda)^{-1}$ (bits per second);

$S[n]$ for the amount of data in bits that the transmitter is capable of transmitting at slot n. The random variables $S[1], S[2], \ldots$ are i.i.d. from some fixed distribution;

$Q[n]$ for the backlog size in bits at slot n.

We also assume that the instantaneous channel gain g^* is perfectly known at the transmitter side. The FAS-based system capacity C at slot n could be approximated by

$$C \approx B \log_2\left(1 + \frac{g^* P_\mathrm{t}}{N_0 B}\right) = B \log_2\left(1 + \frac{g^* P_\mathrm{t}}{\sigma_\eta^2}\right) \tag{8}$$

if the capacity-achieving coding is adopted. Thus, the capacity C and the service S have a simple relation $S = C T_s$.

C. EE and Problem Formulation

Consider that the total transmission power can be expressed as $P_\mathrm{tot} = P_c + P_\mathrm{t}$, where P_c is the constant circuit power and P_t is the transmit power consumed by power amplifier. Based on the definition in [12], the EE of an N-port FAS-based system is given by

$$\eta_N(P_\mathrm{t}) = \frac{R}{P_c + P_\mathrm{t}}. \tag{9}$$

As shown in Fig. 1(a), upper-layer data experience delays in a queueing buffer. We first define a maximum delay bound D_max and a tolerance ϵ, and specify a target DOP constraint

$\{D_{\max}, \epsilon\}$ in order to support a given QoS for a hypothetical application. The system has to satisfy

$$\text{Prob}\left(D(\infty) > D_{\max}\right) \leq \epsilon, \tag{10}$$

where $D(\infty)$ is the steady-state delay. Given that the transmitter decides the allocation of the transmit power $P_{\rm t}$ on the basis of the constraint $\{D_{\max}, \epsilon\}$ and the remaining parameters are fixed, EE can be maximized via

$$\mathbb{P}_1 : \begin{cases} \max\limits_{P_{\rm t} \geq 0} \eta_N(P_{\rm t}) \\ \text{s.t. } \text{Prob}(D(\infty) > D_{\max}) \leq \epsilon. \end{cases} \tag{11}$$

III. Proposed Solution

Based on a specified DOP constraint $\{D_{\max}, \epsilon\}$, the DOP $\text{Prob}(D(\infty) > D_{\max})$ depends on the transmit power $P_{\rm t}$ and can be reexpressed as

$$S_N(P_{\rm t}; D_{\max}) = \text{Prob}(D(\infty) > D_{\max}). \tag{12}$$

Therefore, $\mathbb{P}_1$ is equivalent to the following problem:

$$\mathbb{P}_2 : \begin{cases} \max\limits_{P_{\rm t} \geq 0} \eta_N(P_{\rm t}) \\ \text{s.t. } S_N(P_{\rm t}; D_{\max}) \leq \epsilon. \end{cases} \tag{13}$$

An analytical expression of $S_N(P_{\rm t})$ can be obtained by the use of the effective bandwidth and the EC models. Because A has the PDF (7), the effective bandwidth of the arrival process, denoted as $\alpha^{(b)}(u)$, can be found as [26]

$$\alpha^{(b)}(u) = \frac{1}{T_s u} \log\left(\frac{\lambda}{\lambda - u}\right), \tag{14}$$

where u is the QoS exponent (the first EC function). To have the EC of the service process, the following lemma is required.

Lemma 1: The PDF of the channel power gain g^* in an N-port FAS-based system is given by

$$f_{g^*}(x) = \begin{cases} \dfrac{1}{\sigma^2} \exp\left(-\dfrac{x}{\sigma^2}\right) \text{ if } N = 1, \\ \dfrac{N}{\sigma^2} \displaystyle\int_0^\infty F_X(x|r)^{n-1} f_X(x|r) e^{-\frac{r}{\sigma^2}} dr \text{ if } N > 1, \end{cases} \tag{15}$$

where

$$F_X(x|r) = 1 - Q_1\left(\sqrt{\frac{\mu^2 r}{\sigma^2(1-\mu^2)}}), \sqrt{x}\right), \tag{16}$$

and

$$f_X(x|r) = \frac{1}{2} e^{-\frac{x + \frac{\mu^2 r}{\sigma^2(1-\mu^2)}}{2}} J_0\left(\sqrt{\frac{\mu^2 g_0 x}{\sigma^2(1-\mu^2)}}\right) \tag{17}$$

in which $Q_1(a, b)$ is the first-order Marcum Q-function, $J_0(\cdot)$ is the zero-order Bessel function of the first kind, and $g_0 = x_0^2 + y_0^2$ which comes from the channels in (2).

Proof: See Appendix A. ■

Based on the PDF of $f_{g^*}(x)$, the EC, $\alpha^{(c)}(u; P_{\rm t})$, is given by [26]

$$\begin{aligned} &\alpha^{(c)}(u; P_{\rm t}) \\ &= \frac{-1}{T_s u} \log \int_0^\infty e^{-u T_s B \log_2\left(1 + \frac{x P_t}{\sigma_\eta^2}\right)} f_{g^*}(x) dx \\ &= \frac{-1}{T_s u} \log \int_0^\infty \left(1 + \frac{x P_t}{\sigma_\eta^2}\right)^{-\frac{u T_s B}{\log(2)}} f_{g^*}(x) dx. \end{aligned} \tag{18}$$

If the assumptions of the Gartner-Ellis theorem hold and if there is a unique QoS exponent $u^* > 0$ that satisfies

$$\alpha^{(b)}(u^*) = \alpha^{(c)}(u^*; P_{\rm t}), \tag{19}$$

then the DOP can be accurately approximated by [26]

$$S_N(P_{\rm t}; D_{\max}) \approx \left(1 - \lambda^{-1} u^*\right)^{D_{\max}+1}, D_{\max} \in \mathbb{N}_0, \tag{20}$$

where $\mathbb{N}_0$ is the set of all natural numbers including zero. Based on (20), we have the following lemma.

Lemma 2: Provided that the DOP constraint $\{D_{\max}, \epsilon\}$, the average packet length $1/\lambda$ at a slot are known, the target QoS exponent $u^\dagger$ is expressed as

$$u^\dagger = \lambda\left(1 - \epsilon^{\frac{1}{D_{\max}+1}}\right). \tag{21}$$

Proof: The above expression (21) is an inverse of the function (20) with respect to $D_{\max}$ and ϵ. ■

Now we have the optimal power control strategy.

Proposition 1: The optimal power $P_{\rm t}^*$ is unique and satisfies the following equation:

$$\lambda\left(\int_0^\infty \left(1 + \frac{x P_{\rm t}}{\sigma_\eta^2}\right)^{-\frac{u^\dagger T_s B}{\log(2)}} f_{g^*}(x; P_{\rm t}) dx - 1\right) = u^\dagger. \tag{22}$$

Proof: Problem $\mathbb{P}_2$ with respect to $P_{\rm t}$ is a strictly convex optimization problem. This is because 1) $\eta_N(P_{\rm t})$ of (9) is a strictly convex function; 2) the feasible set of $P_{\rm t}$ defined by (13) is a convex set as $S_N(P_{\rm t})$ is a strictly decreasing function of $P_{\rm t}$. Therefore, it has a unique optimal solution. Moreover, for the Karush-Kuhn-Tucker (KKT) condition, allocating

$$S_N(P_{\rm t}; D_{\max}) = \epsilon \tag{23}$$

is an optimal solution to $\mathbb{P}_2$. By substituting u^* in (19) with $u^\dagger$ into (21), we then have

$$\begin{aligned} &\alpha^{(b)}(u^\dagger) = \alpha^{(c)}(u^\dagger; P_{\rm t}^*) \Leftrightarrow \\ &\frac{1}{T_s u^\dagger} \log\left(\frac{\lambda}{\lambda - u^\dagger}\right) \\ &= \frac{-1}{T_s u^\dagger} \log \int_0^\infty \left(1 + \frac{x P_{\rm t}}{\sigma_\eta^2}\right)^{-\frac{u^\dagger T_s B}{\log(2)}} f_{g^*}(x) dx \\ &\Leftrightarrow \lambda\left(\int_0^\infty \left(1 + \frac{x P_{\rm t}}{\sigma_\eta^2}\right)^{-\frac{u^\dagger T_s B}{\log(2)}} f_{g^*}(x) dx - 1\right) = u^\dagger, \end{aligned} \tag{24}$$

where "$\Leftrightarrow$" is an equivalence sign. ■

Finally, $P_{\rm t}^*$ is a root of $f(P_{\rm t}) = \alpha^{(b)}(u^\dagger) - \alpha^{(c)}(u^\dagger; P_{\rm t})$, which can be solved by any root-finding numerical methods, e.g., binary search method.

IV. Results and Discussion

In this section, we report the simulation results of the N-port FAS-based system. In the simulations, the path loss L_p is based on the 3GPP model for a carrier frequency between 1400MHz and 2600MHz [27]:

$$L_p = 128.1 + 37.6\log_{10}(d) + 21\log_{10}\left(\frac{f_c}{2}\right), \tag{25}$$

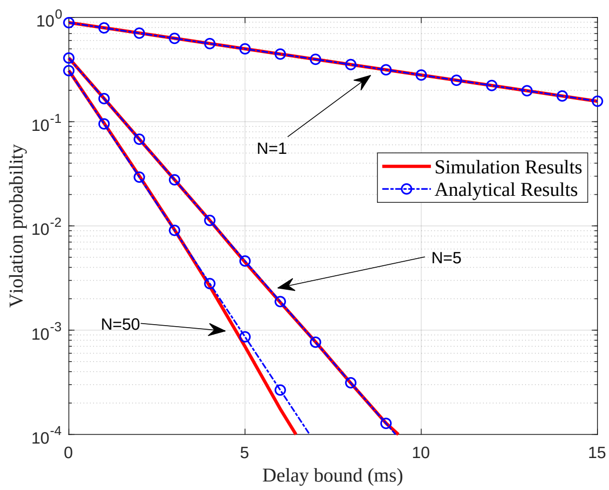

Fig. 2. Simulation and approximation results of $\mathrm{Prob}(D > D_{\max})$.

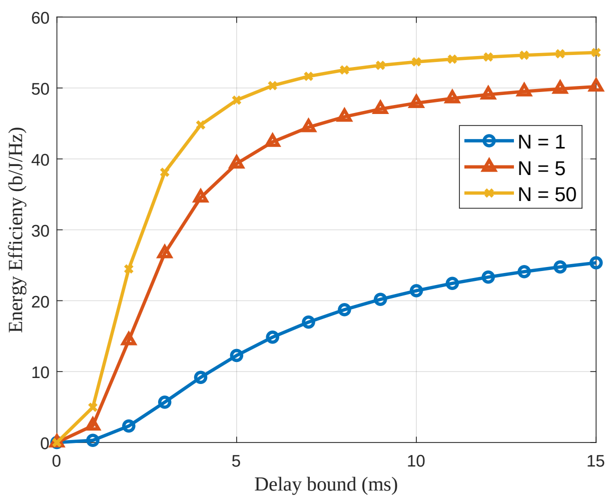

Fig. 3. Optimal EE under different delay-outage probability and number of ports of FAS.

where d represents the distance between the transmitter and the receiver in Km and f_c is the carrier frequency in GHz. In particular, we have fixed the distance $d = 0.2\mathrm{Km}$, $f_c = 2\mathrm{GHz}$, and the bandwidth is set to $10\mathrm{MHz}$ and the constant circuit power is $P_c = 0.1\mathrm{W}$. Also, the noise spectral density is $N_0 = -174\mathrm{dBm/Hz}$ and the slot duration is $1\mathrm{ms}$.

First, consider a FAS-based system with $W = 1$ which gives $W\lambda = 15\mathrm{cm}$. Given that the transmit power is $20\mathrm{dBm}$ and the average data rate R is $60\mathrm{Mbps}$, we show the simulation and analytical results of the DOP with respect to different values of the maximum delay bound $D_{\max}$ in Fig. 2 when $N = 1, 5$ and 50. The x-axes are delay bounds (the unit is millisecond) and the y-axes are violation probabilities in log scale. Under any number of ports, the simulation result is obtained based on one-million samples of backlog size (1000 seconds ≈ 0.277 hours). The unique QoS exponent u^* is numerically obtained based on the algorithm in [26] and the analytical result is calculated by (20). As shown in Fig. 2, the simulation and analytical results almost overlap with each other in all three scenarios. This shows that (20) can accurately approximate the DOP of the FAS system with exponential arrival processes. On the other hand, the performance of DOP can be significantly improved by increasing the number of ports.

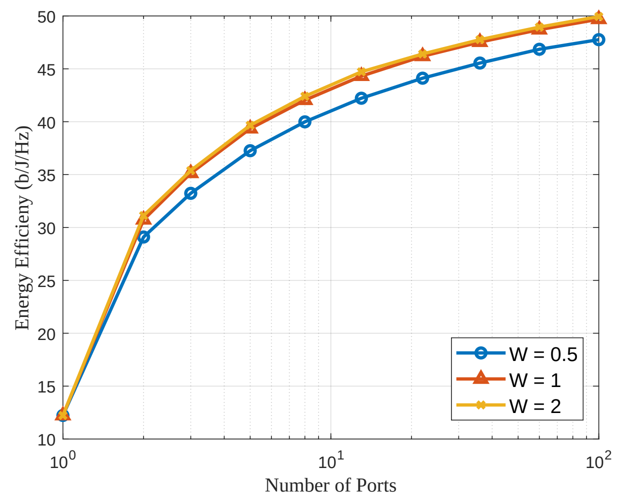

Fig. 4. Optimal EE using different number of ports of FAS and W under a DOP constraint $\{D_{\max} = 5\mathrm{ms}, \epsilon = 0.02\}$.

Fig. 3 illustrates the numerical results on the maximum EE obtained from Proposition 1 for different DOP constraints, i.e., $D_{\max} = \{0, 1, 2, \dots, 15\}\mathrm{ms}$ where $\epsilon = 0.02$. We observe that the EE increases when extending the $D_{\max}$ and relaxing the delay outage probability. When $D_{\max}$ goes to infinity, the optimal transmit power converges to a value that satisfies $R = \mathbb{E}[C]$, which is the condition for a stable system. Again, the EE in FAS-based communication systems is significantly higher compared to conventional systems with a single fixed antenna. In the case of $D_{\max} = 1$ or $10\mathrm{ms}$ (required in the 6G and 5G networks [28], respectively), the EE in the 50-port FAS-based system is $16.5\times$ or $2.5\times$ higher than the fixed antenna system.

Finally, the results in Fig. 4 are provided for the optimal EE under a target delay-outage constraint $\{D_{\max} = 5\mathrm{ms}, \epsilon = 0.02\}$ for different number of ports, N, and size of the fluid antenna, W. As we can observe, EE increases as N increases, but the growth slows down when N is large. Finally, it can be observed that increasing the length of fluid antenna does improve the EE but the improvements are moderate.

V. Conclusion

In this paper, we considered the maximization of EE under a DOP constraint for FAS-aided communication system where trade-offs between rate, power consumption and delay could be formulated. This was done through an accurate approximation of the DOP for a FAS-based system, and then solving the EE maximization using a convex optimization approach. Simulation results confirmed the accuracy of the DOP approximation and illustrated significant gains in the EE using FAS.

Appendix A
Proof of Lemma 1

Let a normalized random variable X be

$$X = \frac{g_k}{\sigma(1-\mu^2)} = \frac{|h_k|^2}{\sigma(1-\mu^2)} = \left(x_k + \frac{\mu}{\sigma\sqrt{1-\mu^2}}x_0\right)^2 + \left(y_k + \frac{\mu}{\sigma\sqrt{1-\mu^2}}y_0\right)^2. \tag{26}$$

Given that $g_0 = x_0^2 + y_0^2$ is known, X_k is a noncentral chi-square random variable with its conditional PDF being

$$f_X(x|g_0) = \frac{1}{2}e^{-\frac{x+\frac{\mu^2 g_0}{\sigma^2(1-\mu^2)}}{2}} J_0\left(\sqrt{\frac{\mu^2 g_0 x}{\sigma^2(1-\mu^2)}}\right) \tag{27}$$

where $J_0(\cdot)$ denotes the zero-order modified Bessel function of the first kind, and the conditional CDF being

$$F_X(x|g_0) = 1 - Q_1\left(\sqrt{\frac{\mu^2 g_0}{1-\mu^2}}, \sqrt{x}\right), \tag{28}$$

where $Q_1(a,b)$ is the first-order Marcum Q-function. Also, the normalized random variables at the ports given g_0 are i.i.d. random variables. Let $X^* = \max\{X_1, X_2, \dots, X_N\}$. The conditional PDF of X^* given g_0 is then given by

$$f_{X^*}(x|g_0) = N F_X(x|g_0)^{n-1} f_X(x|g_0). \tag{29}$$

Obviously, the PDF of g_0 is given by

$$f_{g_0}(x) = \frac{1}{\sigma^2}e^{-\frac{x}{\sigma^2}}. \tag{30}$$

Therefore, the unconditional PDF of X^* is found as

$$\begin{aligned} f_{X^*}(x) &= \int_0^\infty f_{X^*}(x|r) f_{g_0}(r)\,dr \\ &= \int_0^\infty N F_X(x|r)^{n-1} f_X(x|r) \frac{1}{\sigma^2} e^{-\frac{r}{\sigma^2}}\,dr \\ &= \frac{N}{\sigma^2}\int_0^\infty F_X(x|r)^{n-1} f_X(x|r) e^{-\frac{r}{\sigma^2}}\,dr. \end{aligned} \tag{31}$$

References

[1] 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, 2021.

[2] A. Shojaeifard *et al.*, "MIMO evolution beyond 5G through reconfigurable intelligent surfaces and fluid antenna systems," *Proc. IEEE*, vol. 110, no. 9, pp. 1244–1265, 2022.

[3] 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 in Commun. and Netw., section Wireless Commun.*, 3:853416, Mar. 2022.

[4] M. Khammassi, A. Kammoun, and M.-S. Alouini, "A new analytical approximation of the fluid antenna system channel," [Online] arXiv preprint arXiv:2203.09318, 2022.

[5] C. Psomas, G. M. Kraidy, K. K. Wong, and I. Krikidis, "On the diversity and coded modulation design of fluid antenna systems," [Online] arXiv preprint arXiv:2205.01962, 2022.

[6] P. Mukherjee, C. Psomas, and I. Krikidis, "On the level crossing rate of fluid antenna systems," [Online] arXiv preprint arXiv:2205.01711, 2022.

[7] L. Tlebaldiyeva, G. Nauryzbayev, S. Arzykulov, A. Eltawil, and T. Tsiftsis, "Enhancing QoS through fluid antenna systems over correlated Nakagami-m fading channels," in Proc. *IEEE Wireless Commun. & Netw. Conf. (WCNC)*, pp. 78–83, 10-13 Apr. 2022, Austin, TX, USA.

[8] C. Skouroumounis and I. Krikidis, "Fluid antenna with linear MMSE channel estimation for large-scale cellular networks," to appear in *IEEE Trans. Commun.*, 2023.

[9] L. Zhu, W. Ma, and R. Zhang, "Modeling and performance analysis for movable antenna enabled wireless communications," [Online] arXiv preprint arXiv:2210.05325, 2022.

[10] W. Ma, L. Zhu, and R. Zhang, "MIMO capacity characterization for movable antenna systems," [Online] arXiv preprint arXiv:2210.05396, 2022.

[11] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[12] M. Zorzi and R. R. Rao, "Energy-constrained error control for wireless channels," *IEEE Pers. Commun.*, vol. 4, no. 6, pp. 27–33, 1997.

[13] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, 2003.

[14] L. Musavian and T. Le-Ngoc, "Energy-efficient power allocation for delay-constrained systems," in Proc. *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 3554–3559, 3-7 Dec. 2012, Anaheim, CA, USA.

[15] L. Musavian and Q. Ni, "Effective capacity maximization with statistical delay and effective energy efficiency requirements," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3824–3835, 2015.

[16] X. Zhang, J. Wang, and H. V. Poor, "Statistical delay and error-rate bounded QoS provisioning for SWIPT over CF M-MIMO 6G mobile wireless networks using FBC," *IEEE J. Select. Topics Sig. Proc.*, vol. 15, no. 5, pp. 1272–1287, 2021.

[17] F. Qasmi, M. Shehab, H. Alves, and M. Latva-Aho, "Effective energy efficiency and statistical QoS provisioning under Markovian arrivals and finite blocklength regime," *IEEE Int. Things J.*, vol. 9, no. 18, pp. 17 741–17 755, 2022.

[18] J. Zeng, C. Xiao, T. Wu, W. Ni, R. P. Liu, and Y. J. Guo, "Uplink non-orthogonal multiple access with statistical delay requirement: Effective capacity, power allocation, and α fairness," to appear in *IEEE Trans. Wireless Commun.*, 2022.

[19] N. Moosavi, M. Sinaie, P. Azmi, P. H. Lin, and E. Jorswieck, "Cross layer resource allocation in H-CRAN with spectrum and energy cooperation," *IEEE Trans. Mobile Comput.*, vol. 22, no. 1, pp. 145–158, 2023.

[20] Y. Chen and I. Darwazeh, "A close approximation of the nonempty buffer probability over Nakagami-m fading channels," *IEEE Commun. Letters*, vol. 19, no. 7, pp. 1121–1124, Jul. 2015.

[21] J. Xu, Y. Chen, Q. Cui, and X. Tao, "Use of two-mode transceiver circuitry and its cross-layer energy efficiency analysis," *IEEE Commun. Letters*, vol. 21, no. 9, pp. 2065–2068, Sept. 2017.

[22] J. Xu, Y. Chen, H. Chen, Q. Cui, and X. Tao, "Use of two-mode circuitry and optimal energy-efficient power control under target delay-outage constraints," in Proc. *IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, 8-13 Oct. 2017, Montreal, QC, Canada.

[23] B. Cai, Y. Chen, Q. Cui, X. Zhu, Y. Yang, and A. N. S. Mode, "Accurate energy-efficient power control for uplink NOMA systems under delay constraint," in Proc. *IEEE/CIC Int. Conf. Commun. China (ICCC)*, pp. 682–687, 16-18 Aug. 2018, Beijing, China.

[24] K. Wong, K. Tong, Y. Chen, and Y. Zhang, "Closed-form expressions for spatial correlation parameters for performance analysis of fluid antenna systems," *Elect. Letters*, vol. 58, no. 11, pp. 454–457, 2022.

[25] G. Ozcan and M. C. Gursoy, "Optimal power control for fading channels with arbitrary input distributions and delay-sensitive traffic," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4333–4344, 2018.

[26] Y. Chen and I. Darwazeh, "An accurate approximation of delay with Nakagami-m channels and exponential arrivals," in Proc. *IEEE Global Commun. Conf. (GLOBECOM)*, 6-10 Dec. 2015, San Diego, CA, USA.

[27] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios," *3GPP TR 36.942 version 17.0.0 Release 17*, vol. 3, pp. 107, 2022.

[28] M. Series, "IMT vision–framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation ITU*, vol. 2083, no. 0, 2015.