Capacity Analysis and Optimal Spacing Design for Compact Array Massive MIMO Systems With Finite Aperture

Lihua Pang, *Member, IEEE*, Minglei Wang, Yang Zhang, *Member, IEEE*, Yunhui Guo, Yijian Chen, and Anyi Wang

***Abstract*— To achieve large-scale antenna deployment in limited space, compact arrays should be widely used. In this letter, we discuss the mathematical relationship between system capacity and the aperture of compact uniform linear arrays (ULAs). In particular, the theoretical analysis benefits from the Marcenko-Pastur law and Toeplitz feature of the mutual coupling and correlation matrices. The element spacing optimization problem is then formulated to maximize the system capacity under a finite aperture. After illustrating its convexity, the optimal spacing can be obtained by applying one-dimensional search methods. Further, the discussions can be readily extended to compact uniform planar arrays (UPAs) with finite apertures. Numerical results show the superiority of optimal spacing in system capacity for both ULAs and UPAs to half-wavelength spacing.**

***Index Terms*— Massive MIMO, finite array aperture, mutual coupling, capacity, element spacing.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) has been a key enabling technology for 5G wireless systems, and is believed to play a more important role in the future beyond 5G (B5G) and 6G communications [1]. For the sake of achieving better performance, it is necessary to deploy more antenna elements, which usually leads to a larger array aperture. However, commercial factors promote the development of base stations (BSs) in the direction of miniaturization. In this context, array deployment will be restricted by the actual available physical space. The element spacing is required to be reduced to deploy more antennas, so that the arrangement of the array will be more compact.

For compact arrays, the elements are close to each other, causing the interaction of electric field between them [2], i.e., mutual coupling occurs between antennas. This mutual coupling effect will change the characteristics of antennas, affecting the radiation pattern of the array and ultimately producing an impact on the system performance (e.g., capacity, error rate, energy efficiency). Reference [3] depicts the influence of mutual coupling on capacity for arrays under fixed spacings, which shows that element coupling will enhance the correlation between sub-channels and thus reduce capacity. With invariant antenna number and decreasing of the element spacing, the research in [4] illustrates the decrease of capacity with the increase of mutual coupling between elements. However, these classical conclusions are obtained under the changing array aperture along with the antenna number or element spacing, which is impractical for real deployment. In the practical limited physical space, the reduction in spacing makes it possible to deploy more antennas, which will bring array gain. Therefore, there exists a trade-off between the array gain caused by the increase of the antenna numbers and the mutual coupling as well as the correlation effect created by the decrease in spacing. The question is, what is the optimal array layout under a fixed aperture to reach the trade-off point?

Under fixed physical space, [5] investigates the influence of mutual coupling to provide lower bounds of sum rate for different precoding algorithms. [6] declares that the optimal spacing depends on the array aperture and is irrelevant to the signal-to-noise ratio (SNR). By deriving an expression for the capacity at low SNR, it transforms the spacing optimization problem into a simpler problem, but without providing a definite solution. In [7], the optimal spacings of uniform linear arrays (ULAs) with finite aperture are discussed under different conditions through numerical simulation methods rather than theoretical analysis, so that they cannot reveal the fundamental principle of capacity change. As far as we know, research on this issue is still quite limited and cannot meet the technical requirements of B5G/6G system design.

In order to fill the theory gap and satisfy the practical requirements, we address the optimal layout of uniform compact arrays under a fixed aperture in this letter. The specific contributions lie in three aspects: (1) The mathematical relationship between system capacity and the aperture of ULA is derived with mutual coupling and correlation effects taking into account. (2) On account of the theoretical derivation, a topology design scheme is presented to determine the optimal spacing of ULA with finite aperture, which can achieve superior capacity performance in contrast to the traditional half-wavelength spacing array. (3) The proposed scheme is scalable. It can be readily extended to compact uniform planar arrays (UPAs) with little modification of the details. Therefore, this study provides beneficial insights into the uniform compact array design in future large-scale MIMO systems.

Manuscript received 31 May 2022; accepted 6 July 2022. Date of publication 14 July 2022; date of current version 10 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grants 61871300 and U19B2015, in part by the Key Research and Development Program of Shaanxi under Grant 2021GY-050, in part by the Excellent Youth Science Foundation of Xi'an University of Science and Technology under Grant 2019YQ3-13, and in part by the Fundamental Research Funds for the Central Universities under Grant JB210112. The associate editor coordinating the review of this letter and approving it for publication was C. D'andrea. *(Corresponding author: Yang Zhang.)*

Lihua Pang is with the Xi'an Key Laboratory of Heterogeneous Network Convergence Communications, Xi'an University of Science and Technology, Xi'an 710054, China, and also with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China.

Minglei Wang, Yang Zhang, and Yunhui Guo are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: yangzhang1984@gmail.com).

Yijian Chen is with the Algorithm Department, ZTE Corporation, Shenzhen 518057, China.

Anyi Wang is with the Xi'an Key Laboratory of Heterogeneous Network Convergence Communications, Xi'an University of Science and Technology, Xi'an 710054, China.

Digital Object Identifier 10.1109/LCOMM.2022.3190492

II. System Models

A. System Description

Consider a typical single-cell multiuser downlink massive MIMO system containing a BS and K single-antenna users. A ULA is deployed at BS with restricted space. Denote d as the element spacing, the number of the ULA elements N is related to d. For this (N, K) massive MIMO system, the received signal can be expressed as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{K\times N}$ represents the channel matrix, $\mathbf{x} \in \mathbb{C}^{N\times 1}$ denotes the transmitted signal, and $\mathbf{n} \in \mathbb{C}^{K\times 1}$ is the zero-mean Gaussian noise vector with power spectral density $\sigma^2/2$ per element.

We adopt the widely used narrowband channel model where the multipaths are inseparable and can be superimposed together. It can be regarded as the summation of a large number of statistically independent random variables. According to the Lindeberg-Levy central limit theorem, the distribution of the sum converges to a normal distribution. Therefore, the entries of $\mathbf{H}$ can be assumed to be independently identical distributed (i.i.d.) with zero-mean and variance $1/N$.

B. Coupling Model

For antennas with fixed lengths and polarization, the mutual coupling mainly depends on spacing d. Generally speaking, the smaller the spacing, the stronger the coupling. The relationship between them is usually modeled by induced electromotive force (EMF) method [8]. In this method, however, d is implicitly included in a composite function and its integral limits, which is not mathematically amenable to our subsequent derivations. This motivates us to look for alternative ways to revisit it. Through numerical analysis of the coupling modeled by EMF, we found that the negative exponential model can effectively fit the relationship between coupling and spacing. Hence, the mutual coupling between element i and j can be remodeled as

$$\alpha(i,j) = \exp(-A * \hat{d}_{ij}) \tag{2}$$

where $i, j = 1, 2, \cdots, N$, $\hat{d}_{ij} = d_{ij}/\lambda$ is the normalized spacing with d_{ij} as the physical distance between element i and j, λ denotes the wavelength corresponding to the center frequency. The parameter $A > 0$ is a coefficient, which depicts the degree of the mutual coupling. Specifically, the smaller the value of A, the stronger the mutual coupling. Referring to the numerical value of the mutual coupling obtained by the EMF method, set $A = 13.4$.

For a compact array, a matrix $\mathbf{C}_t$ can be used to characterize the mutual coupling relationship between elements. It is noteworthy that when $d \geq \lambda/2$, the mutual coupling effect is weak and can be generally ignored. Besides, in consideration of hardware cost and to avoid strong mutual coupling, the spacing in practical engineering will not be less than $\lambda/4$. Therefore, when $\lambda/4 \leq d \leq \lambda/2$, only the mutual coupling between adjacent elements needs to be considered. In this sense, for the N-element ULA, $\mathbf{C}_t$ is modeled as a tridiagonal Toeplitz matrix of order N. The elements on the main diagonal are 1 and the elements on the two sub-diagonals are α, which can be obtained from (2).

Further, in order to ensure that the incorporation of $\mathbf{C}_t$ will not physically affect the total power of the transmitter, each column of $\mathbf{C}_t$ should be normalized as

$$c'_{i,j} = \frac{c_{i,j}}{\sum_{t=1}^{N} c_{t,j}} \tag{3}$$

where $c_{i,j}$ represents the element in row i and column j of $\mathbf{C}_t$, which should be replaced by $c'_{i,j}$ after normalization.

C. Correlation Model

When antennas are densely deployed in a limited space, the lack of physical distance will lead to spatial fading correlation of MIMO channels. Using the Kronecker model, the relevant channel matrix $\mathbf{H}_\Sigma$ can be described as

$$\mathbf{H}_\Sigma = \mathbf{H}{\mathbf{\Sigma}_T}^{1/2} \tag{4}$$

In this system, only the correlation matrix at the transmitter $\mathbf{\Sigma}_T$ should be considered, which can refer to the model proposed in [9]. In the case of Rayleigh-fading environment caused by multipaths, the generic element of $\mathbf{\Sigma}_T$ is represented by

$$\mathbf{\Sigma}_T(i,j) = I_0\left(\sqrt{\eta^2 - 4\pi^2\hat{d}_{ij}^2 + 4\pi\eta\sin(\mu)\,\hat{d}_{ij}}\right) \Big/ I_0(\eta) \tag{5}$$

where $i, j = 1, 2, \cdots, N$. The parameter $\eta \in (0, \infty)$ accounts for the width of the angle-of-departure (AOD), with $\eta = 0$ means isotropic scattering. $\mu \in [-\pi, \pi)$ is the mean direction of AOD, and $I_0(\cdot)$ is the zero-order modified Bessel function.

Similar to the mutual coupling, only the correlation between two adjacent elements is considered, which is denoted by r. The $\mathbf{\Sigma}_T$ modeled in this letter is also a tridiagonal Toeplitz matrix, in which the elements on the main diagonal are 1 and that on the two sub-diagonals are r.

III. Capacity Analysis

Before spacing optimization for the compact ULA deployed in a limited area, it is essential to analyze the mathematical relationship between the capacity and the array aperture. The first step is to construct an equivalent channel $\tilde{\mathbf{H}}$ through multiplying $\mathbf{H}$ by $\mathbf{C}_t$ and ${\mathbf{\Sigma}_T}^{1/2}$, and then we can arrive at the average capacity per unit bandwidth per user as

$$C_{coupling} = \frac{1}{K}\log_2\det\left(\mathbf{I} + \frac{P}{N\sigma^2}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right) = \frac{1}{K}\log_2\det \times \left(\mathbf{I} + \frac{P}{N\sigma^2}\mathbf{H}{\mathbf{\Sigma}_T}^{1/2}\mathbf{C}_t{\mathbf{C}_t}^H\left({\mathbf{\Sigma}_T}^{1/2}\right)^H\mathbf{H}^H\right) \tag{6}$$

where P is the total transmitting power. The problem is that parameter d is implicitly contained in $\mathbf{H}$, $\mathbf{\Sigma}_T$ and $\mathbf{C}_t$, which makes it hard to analyze the influence of d mathematically.

Fortunately, asymptotic analysis is an effective way to simplify the problem and will provide guidance for system design. Consequently, we conduct a further analysis in the high SNR region. (6) can be approximated as

$$C_{coupling} \approx \frac{1}{K}\log_2\det\left(\frac{P}{N\sigma^2}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right) = \frac{1}{K}\log_2\det \times \left(\frac{P}{N\sigma^2}\mathbf{H}{\mathbf{\Sigma}_T}^{1/2}\mathbf{C}_t{\mathbf{C}_t}^H\left({\mathbf{\Sigma}_T}^{1/2}\right)^H\mathbf{H}^H\right) \tag{7}$$

Because of the property that the determinant of the matrix product is equal to the product of the matrix determinant, (7) can be further rewritten as

$$C_{coupling} = \frac{1}{K}\log_2\left[\det\left(\frac{P}{N\sigma^2}\mathbf{H}\mathbf{H}^H\right)\right] + \frac{2}{K}\times\log_2\left[\det\left(\mathbf{C}_t\right)\right] + \frac{1}{K}\times\log_2\left[\det\left(\mathbf{\Sigma}_T\right)\right] \tag{8}$$

As can be seen, the capacity is decomposed into three parts. The first part is related to $\mathbf{H}$ and is irrelevant to $\mathbf{C}_t$ and $\mathbf{\Sigma}_T$, while the second part is only related to $\mathbf{C}_t$ and the third part is only related to $\mathbf{\Sigma}_T$.

For the first part of (8), according to the theory of asymptotic spectrum [10], with an i.i.d. Gaussian input signal, the normalized capacity can be derived as

$$\begin{aligned} C_N &= \frac{1}{\min(N,K)}\log_2\det\left(\mathbf{I}+\frac{P}{N\sigma^2}\mathbf{H}\mathbf{H}^H\right) \\ &= \frac{1}{\min(N,K)}\sum_{i=1}^{\min(N,K)}\log_2\left(1+\frac{P}{N\sigma^2}\lambda_i\right) \\ &\to \beta\int_a^b \log_2\left(1+\frac{P}{N\sigma^2}x\right)f_\beta(x)dx \end{aligned} \tag{9}$$

where λ_i represents the i-th eigenvalue of $\mathbf{H}\mathbf{H}^H$, $\to$ denotes the limit, $\beta{=}N/K$ indicates the ratio of the number of elements at the transmitter and the receivers, $f_\beta(x)$ is the empirical probability density function of the eigenvalues, and $a=\left(1-\sqrt{\beta}\right)^2$ and $b=\left(1+\sqrt{\beta}\right)^2$ represent the lower and the upper bounds of $f_\beta(x)$, respectively. From the perspective of high-dimensional random matrix, when $N\to\infty$, $K\to\infty$, and $\beta\in(0,+\infty)$, the eigenvalues of $\mathbf{H}\mathbf{H}^H$ are non-randomly distributed [10]. By Marcenko-Pastur law, their empirical probability density function can be expressed as

$$f_\beta(x)=\left(1-\frac{1}{\beta}\right)^+\delta(x)\ +\frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\beta x} \tag{10}$$

where $(z)^+=\max(0,z)$ and $\delta(x)$ denotes the impulse function. Inserting (10) into (9), we can get

$$\begin{aligned} C_N = {}& \beta\log_2\left(1+\frac{P}{N\sigma^2}-\frac{1}{4}F\left(\frac{P}{N\sigma^2},\beta\right)\right) \\ &+\log_2\left(1+\frac{P}{N\sigma^2}\beta-\frac{1}{4}F\left(\frac{P}{N\sigma^2},\beta\right)\right) \\ &-\frac{\log_2 e}{4P/(N\sigma^2)}F\left(\frac{P}{N\sigma^2},\beta\right) \end{aligned} \tag{11}$$

where $F(x,y)=\left(\sqrt{x\left(1+\sqrt{y}\right)^2+1}-\sqrt{x\left(1-\sqrt{y}\right)^2+1}\right)^2$.

For the second and third parts of (8), the analysis operations are very similar, because $\mathbf{C}_t$ and $\mathbf{\Sigma}_T$ are both tridiagonal Toeplitz matrices. We take the second part as an example.

With regard to matrix $\mathbf{C}_t$, it can be noticed that there is a recursive relationship between the determinants of $\mathbf{C}_t$ of different orders, which can be depicted as

$$D_n = D_{n-1}+\alpha^2 D_{n-2} \tag{12}$$

where $D_i\,(1\le i\le N)$ denotes the ith-order determinant of $\mathbf{C}_t$ and α is the element on the sub-diagonals of $\mathbf{C}_t$ as aforementioned. Owing to the existence of this recursive relationship, D_n can be obtained by solving this characteristic equation $x^2{-}x{+}\alpha^2=0$ [11]. Specifically, while $1{-}4\alpha^2\neq 0$, the two roots are $x_1=\frac{1+\sqrt{1-4\alpha^2}}{2}$ and $x_2=\frac{1-\sqrt{1-4\alpha^2}}{2}$. Then, the determinant of $\mathbf{C}_t$ can be represented by

$$D_n=\theta\cdot {x_1}^n+\varphi\cdot {x_2}^n \tag{13}$$

Combined with $\det(\mathbf{C}_1)=1$ and $\det(\mathbf{C}_2)=\alpha^2-1$, it can be readily arrived at $\theta=\left(1+\sqrt{1-4\alpha^2}\right)/2\sqrt{1-4\alpha^2}$ and $\varphi=-\left(1-\sqrt{1-4\alpha^2}\right)/2\sqrt{1-4\alpha^2}$. More specifically, the determinant of $\mathbf{C}_t$ is

$$\det(\mathbf{C_t})=\frac{\left(1+\sqrt{1-4\alpha^2}\right)^{N+1}-\left(1-\sqrt{1-4\alpha^2}\right)^{N+1}}{2^{N+1}\sqrt{1-4\alpha^2}} \tag{14}$$

According to the above mathematical analysis, system capacity with the deployment of a compact ULA can be obtained in (15), as shown at the bottom of the page. When P/σ^2 and K are determined, the normalized capacity now is a function of N and $\hat{d}$. In other words, the relationship between the capacity and the array aperture has been established theoretically.

IV. Capacity Optimization

Based on the above results, the optimal element spacing of the compact ULA will be further studied in this section. The optimization problem can be formulated as

$$\mathbf{P1}:\max_{d,N}\ \ C_{coupling} \quad s.t.\ \begin{cases} \mathrm{c1}: L\le L_0,\ with\ L=(N-1)\times d \\ \mathrm{c2}: \lambda/4\le d\le\lambda/2 \end{cases} \tag{16}$$

where (c1) represents the finite aperture restriction of the array and (c2) is the coupling distance constraint, respectively. It is worth noting that the actual array aperture is also determined by N and d. So the non-convex constraint (c1) in (16) can be transformed to an affine constraint:

$$\mathrm{c3}: L=L_0, with\ L=(N-1)\times d \tag{17}$$

The equation holds when $d=L_0/(N-1)$. In this context, $\mathbf{P1}$ can be simplified as

$$\mathbf{P2}:\max_{N}\ \ C_{coupling} \quad s.t.\ \lambda/4\le L_0/(N-1)\le\lambda/2 \tag{18}$$

$$C_{coupling}=\beta\log_2\left(1+\frac{P}{N\sigma^2}-\frac{1}{4}F\left(\frac{P}{N\sigma^2},\beta\right)\right)+\log_2\left(1+\frac{P}{N\sigma^2}\beta-\frac{1}{4}F\left(\frac{P}{N\sigma^2},\beta\right)\right)-\frac{N\sigma^2}{4P}\times\log_2 e\times F\left(\frac{P}{N\sigma^2},\beta\right) + \frac{2}{K}\log_2\left[\frac{\left(1+\sqrt{1-4\alpha^2}\right)^{N+1}-\left(1-\sqrt{1-4\alpha^2}\right)^{N+1}}{2^{N+1}\sqrt{1-4\alpha^2}}\right]+\frac{1}{K}\log_2\left[\frac{\left(1+\sqrt{1-4r^2}\right)^{N+1}-\left(1-\sqrt{1-4r^2}\right)^{N+1}}{2^{N+1}\sqrt{1-4r^2}}\right] \tag{15}$$

TABLE I

OPTIMAL RESULTS FOR N AND $\hat{d}$

One-Dimensional Search Algorithm	N	$\hat{d}$
Newton Method	279.7	0.36
Dichotomy	279.3	0.35
Golden Section Method	278.7	0.35
Success-Failure Algorithm	279.3	0.35
Three-Point Quadratic Interpolation	279.3	0.35

This is motivated by the fact that a larger array aperture generally provides more degrees to maximize the system capacity. Therefore, $\mathbf{P1}$ and $\mathbf{P2}$ are equivalent and have the same optimal solution.

By introducing the objective function $f(N) = C_{coupling}$ into the definition of concave function, it can be declared that the capacity is a concave function of N under fixed array aperture. Detailed derivations are omitted here due to space limitation. Therefore, a certain optimal N can be found to maximize the capacity.

For the convex optimization problem $\mathbf{P2}$, one-dimensional search algorithms can be employed to find the optimal solution. The popular methods include: (1) Newton method; (2) Dichotomy; (3) Golden section method; (4) Success-failure algorithm; (5) Three-point quadratic interpolation. In case of SNR $= 20$ dB and $K = 30$, for compact ULA with $L_0 = 100\lambda$, the optimal N and $\hat{d}$ are shown in Table I. For the sake of achieving greater spatial degrees of freedom, we tend to choose the method that obtains bigger spacing, i.e., Newton Method.

V. EXTENSION TO COMPACT UPA

It is worthwhile to mention that the preceding discussions can be readily extended to the scenario with compact UPA deployed in a limited space. In this case, mutual coupling and correlation between adjacent elements in horizontal, vertical, and diagonal directions must be considered simultaneously. $\mathbf{C}_t$ and $\mathbf{\Sigma}_T$ will be more complex, but still Toeplitz matrices. Similarly to the theoretical operations in Sections III and IV, the capacity as a function of the UPA aperture can be derived, and the optimal spacing can be obtained by establishing and solving the subsequent optimization problem. Detailed derivations are omitted here.

VI. SIMULATION RESULTS

In this section, numerical simulations are conducted to verify the correctness of the theoretical analysis and evaluate the optimality of our proposal. The array apertures are fixed and determined by half-wavelength spacing ULA and UPA with 200 and 64 antennas, i.e., $L_0 = 100\lambda$ and $S_0 = 7\lambda/2 \times 7\lambda/2$, respectively. The other simulation parameters are set as follows: (1) SNR $= 20$ dB; (2) $K = 30$; (3) The omnidirectional scattering model is used; (4) N changes with d, and the relationship between them is constrained by L_0 and S_0. To ensure the fairness of comparison, the total transmitting power is always fixed to be 1. All the following results are obtained by averaging more than 1,000 random implementations.

Fig. 1 shows the comparisons between the capacity results obtained by the theoretical analysis in this letter and the Monte-Carlo simulations. As can be seen, no matter how $\hat{d}$ changes, the analytical results are quite close to the simulation results for both ULA and UPA. This verifies the accuracy of our theoretical derivations. With the increase of $\hat{d}$, the capacity shows a trend of increasing first and then decreasing due to the fact that the increase of $\hat{d}$ reduces the harmful effect imposed by the correlation and mutual coupling between antennas. However, in a limited space, the larger the $\hat{d}$, the fewer antenna elements can be deployed in the array, and consequently the smaller the multiplexing gain. Clearly, there is a trade-off point between the array gain and the mutual coupling loss. At this point, the maximum capacity is achieved. In particular, the optimal $\hat{d}$ for ULA setting is identical to that in Table I. Moreover, it can be seen that UPA utilizes fewer antennas to achieve a similar capacity as ULA. This is enabled by UPA's planar topology advantage in a three-dimensional propagation environment.

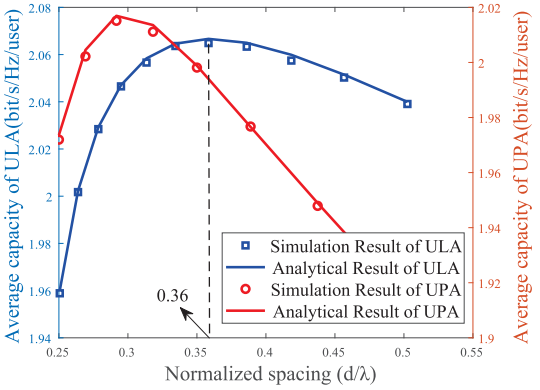


Fig. 1. Simulation and analytical results comparison for ULA and UPA.

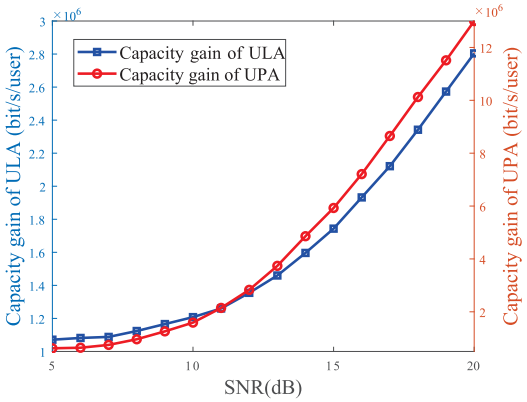


Fig. 2. Capacity gain of ULA and UPA with the optimal spacing.

Fig. 2 illustrates the capacity gain of the optimal spacing array presented in this letter in contrast to the traditional half-wavelength spacing array under the same array aperture. Simulations are executed under the 5G NR standard with a channel bandwidth of 100MHz. The optimal spacing for both ULA and UPA can increase system capacity compared to the half-wavelength spacing in the moderate-to-high SNR region. In the case of SNR $= 20$ dB, the optimal ULA can increase capacity by 2.8×10^6 bps for each user, while the optimal UPA can increase capacity by 1.3×10^7 bps. It shows that UPA can provide higher gains than ULA at the same SNR. This is because under the above simulation configurations,

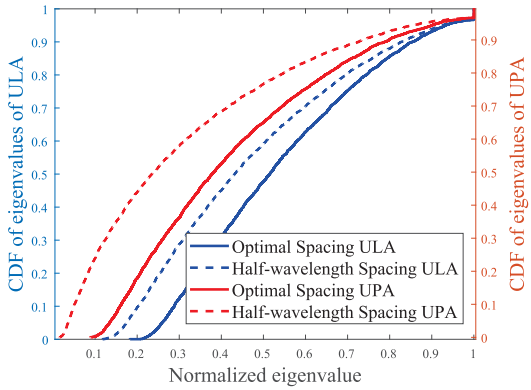
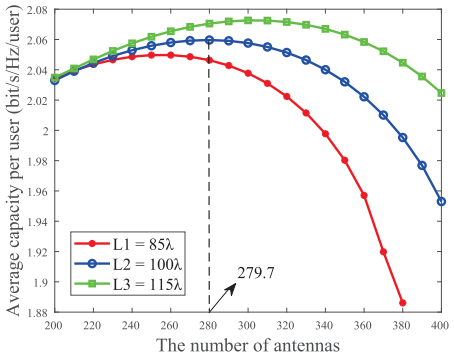
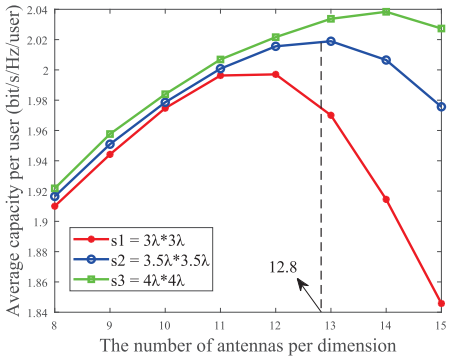


Fig. 3. CDFs of channel eigenvalues corresponding to different spacings.



(a) ULA



(b) UPA

Fig. 4. Capacity comparison under different fixed array apertures.

the physical space constraint for the UPA is stronger from both horizontal and vertical directions, and thus a greater performance gain can be attained after optimization. Furthermore, although the theoretical analysis is under the assumption of non-line of sight (NLOS) scenario, additional simulation results under Rician channel not shown here indicate the effectiveness of our proposal when the LOS path exists.

To explore the essential reason for capacity improvement, Fig. 3 depicts the cumulative distribution functions (CDFs) of the eigenvalues of $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H$. A close look at this figure reveals that: (1) The eigenvalues with the optimal spacing array are larger than those with the half-wavelength spacing array; (2) The channels corresponding to the optimal spacing array are less ill-conditioned, i.e., the degree of orthogonality among users in this case is relatively higher. It is for these two reasons that our proposal can achieve superior capacity.

To show the performance under different array apertures, the capacity metrics corresponding to $L_1 = 85\lambda$, $L_2 = 100\lambda$, $L_3 = 115\lambda$, $S_1 = 3\lambda \times 3\lambda$, $S_2 = 7/2\lambda \times 7/2\lambda$, and $S_3 = 4\lambda \times 4\lambda$ are illustrated in Fig. 4 versus the antenna numbers. It is obvious that for the same N, a larger aperture corresponds to a greater capacity for both ULA and UPA. This can be attributed to the fact that for a fixed N, the array with larger aperture has higher degrees of freedom, resulting in better performance. Moreover, it can be seen that the optimal spacings for both ULA and UPA with larger apertures are slightly increased, as the optimization problem is less constrained in this case.

VII. Conclusion

In this letter, we study the optimal topology design for compact uniform linear and planar massive MIMO arrays with finite apertures. We first shed some light on the mathematical relationship between system capacity and the array aperture analytically. Then, the optimal element spacings of both ULA and UPA are derived to maximize the capacity. Simulation results show the superiority of the proposed layout to the traditional half-wavelength spacing under the same aperture. In the near future, we will further verify the performance of the proposed topology using a 5G prototype verification system.

References

[1] J. Jeon *et al.*, "MIMO evolution toward 6G: Modular massive MIMO in low-frequency bands," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 52–58, Nov. 2021.

[2] Y. Xiao and Y. Wang, "Deep learning-based mutual coupling modeling and baseband decoupling algorithm for MIMO systems," *IEEE Commun. Lett.*, vol. 24, no. 9, pp. 1986–1990, Sep. 2020.

[3] X. Chen, S. Zhang, and Q. Li, "A review of mutual coupling in MIMO systems," *IEEE Access*, vol. 6, pp. 24706–24719, 2018.

[4] F. Wang, M. E. Bialkowski, and X. Liu, "Performance of block diagonalization broadcasting scheme for multiuser MIMO system operating in presence of spatial correlation and mutual coupling," *Int. J. Commun., Netw. Syst. Sci.*, vol. 3, no. 3, pp. 266–272, 2010.

[5] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-scale MIMO transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2794–2804, Jul. 2013.

[6] S. Shen, M. R. McKay, and R. D. Murch, "MIMO systems with mutual coupling: How many antennas to pack into fixed-length arrays?" in *Proc. Int. Symp. Inf. Theory Appl.*, Taichung, Taiwan, Oct. 2010, pp. 531–536.

[7] C. Masouros, J. Chen, and K. Tong, "Towards massive-MIMO transmitters: On the effects of deploying increasing antennas in fixed physical space," in *Proc. Future Netw. Mobile Summit*, Lisboa, Portugal, Jul. 2013, pp. 1–10.

[8] C. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Hoboken, NJ, USA: Wiley, 2005, pp. 478–486.

[9] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr. 2002.

[10] A. Tulino and S. Verd, *Random Matrix Theory and Wireless Communications*. Boston, MA, USA: Now Publishers Inc., 2004, pp. 3–11.

[11] Z. Xu, K. Y. Zhang, and Q. Lu, *Fast Algorithm Based on Toeplitz Matrix*. Xi An, China: Northwestern Polytech. Univ., 1999, pp. 101–103.