Capacity Analysis of Asymmetric Multi-Antenna Relay Systems using Free Probability Theory

Lucinda Hadley
Computing & Communications
Lancaster University
Lancaster, UK
lucinda.hadley@lancaster.ac.uk

Zhijin Qin
Electronic Engineering & Computer Science
Queen Mary University of London
London, UK
z.qin@qmul.ac.uk

Zhiguo Ding
Electrical & Electronic Engineering
The University of Manchester
Manchester, UK
zhiguo.ding@manchester.ac.uk

Abstract—Random matrix theory (RMT) has been used to derive the asymptotic capacity of multiple-input-multiple-output (MIMO) channels by approximating the asymptotic eigenvalue distributions (AEDs) of the associated channel matrices. A novel methodology is introduced which enables the computation of the asymptotic capacity for a generalised system in which two relays cooperate to facilitate communication between two remote devices. It is computationally demanding to calculate this capacity using RMT when nodes are equipped with large-scale antenna arrays, and impossible in the case where asymmetry exists between channels within the system. This is because deriving the capacity across the combined channels from the relays to the receiver involves polynomials in large and non-commutative random matrix variables. This paper uses free probability theory (FPT) as an efficient alternative tool for analysis in these circumstances. The method described can be applied with no additional complexity for arbitrarily large antenna arrays. The minimum SNR required to achieve a given asymptotic capacity is computed and the simulation results verify the accuracy of the FPT approach.

Index Terms—Operator-valued free probability theory, random matrix theory, massive MIMO, eigenvalue distribution, capacity, co-operative relay.

I. INTRODUCTION

As a consequence of the increasing number of users and devices and the rising demands for adaptable technologies to support the internet of things (IoT), the demand for data has been predicted to increase a thousandfold over the next decade. 5G networks will need to support more connections and increased data traffic, whilst maintaining a high quality of service in terms of delay, security and reliability [1].

Multiple-input-multiple-output (MIMO) technology has been instrumental in the advancement of wireless networks in the past. To address the even greater demands of 5G, the natural extension is to increase diversity further by using many more antennas in 'massive MIMO' arrays. At the Mobile World Congress in February 2017, the capability of massive MIMO to increase cell capacity eightfold by comparison to 4G LTE was demonstrated. Sprint in particular showcased arrays that include 64 antennas at the transmit and receive end of both uplink and downlink wireless channels, and it is speculated that hundreds of antennas could be used in future designs.

Massive MIMO is therefore of critical importance in industry and large-scale arrays are a topic of great interest in current research [1], [2].

The versatility of the IoT necessitates the integration of a diverse range of multi-antenna nodes with channels of varying characteristics in terms of fading, distance and scattering. Extensive work has been carried out in which random matrix theory (RMT) is used to analyse the capacity of MIMO systems [3]. The Shannon capacity formula was first adapted using RMT for small MIMO arrays by Foschini [4]. However, for massive arrays the calculations involved become highly complex. A key advance was the technique introduced in [5] whereby, the eigenvalue distributions of certain random matrices provide an alternative approach to capacity analysis. However, some channels cannot be modelled by a matrix with a known eigenvalue distribution, and in the case where asymmetry exists between channels, computing the capacity for the combined channels is impossible using traditional RMT. Free probability theory (FPT) provides an alternative approach to a number of RMT problems [3], [6]. In [7] FPT is used to find the asymptotic spectral efficiency of massive MIMO channels with transmit and receive correlation, while in [8] it is used to find the asymptotic variance of the mutual information for a channel with insufficient scattering.

Our research analyses the overall capacity limits of a generalised two-relay system that includes channels with asymmetric characteristics. We focus particularly on the case where large-scale antenna arrays are employed. The analysis cannot be carried out using RMT and, to our knowledge, this is an area that has not been previously addressed.

The structure of the remainder of this paper is as follows. In Section II we describe a co-operative system in which two relays are used to facilitate the communication between a source and remote destination. Section III presents an overview of the method including FPT, which is required for the computationally efficient derivation of the capacity limitations of our model. The simulation results are given in Section IV, where we demonstrate the accuracy of the method and calculate the minimum SNR required to achieve a desired capacity across the channel with varying degrees of freedom. In Section V, we summarise our findings.
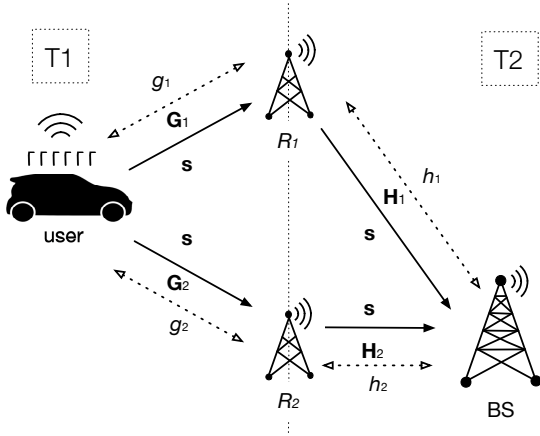
Fig. 1: Relay network

II. SYSTEM MODEL

Consider the situation where a pair of multi-antenna source and destination nodes wish to communicate a wireless signal $\mathbf{s}$, but have no direct link due to a long separation distance. Instead, the signal is forwarded by two multi-antenna relays R_1 and R_2 via channels $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{H}_1$ and $\mathbf{H}_2$ which represent the channel gains between the source node and R_1, the source node and R_2, R_1 and the destination node and R_2 and the destination node and cover distances g_1, g_2, h_1 and h_2 respectively. An example of such a system given in Fig. 1 where the source node is a vehicle, the destination node is a base station (BS) and the relays are road-side units. We suppose that the relays work in half-duplex mode and that the source, relay and destination nodes are equipped with N antennas.

For the duration of this paper the subscript i refers to $i \in \{1, 2\}$. During time slot T1, each R_i receives

$$\mathbf{y}_i = \alpha_i \mathbf{G}_i \mathbf{s} + \mathbf{n}_i, \tag{1}$$

where $\mathbf{n}_i$ represents the noise across sub-channel $\mathbf{G}_i$ and α_i is the attenuation due to path-loss for the channel $\mathbf{G}_i$. This attenuation has been modelled according to the relationship between distance and path-loss $\alpha_i = \frac{1}{g_i^m}$ given in [9], where m is the path-loss exponent. For simplicity, we fix $m = 2$ globally for the channels considered in this paper, since this is the standard path-loss exponent for the type of free-space channels in the far-field depicted in Fig. 1 [10]. However, we note that our method readily extends to include models in which this value varies for different channels. We suppose that each R_i has channel state information (CSI) for $\mathbf{G}_i$ and $\mathbf{H}_i$, and recovers $\mathbf{s}$ using maximum-ratio-combining. In T2, R_i transmits signal $\mathbf{s}$ with optimal matched-filter precoding, by premultiplying the signal by the complex conjugate transpose $\mathbf{H}_i^\dagger$ as described in [11]. This enables the user to decode the received signal:

$$\mathbf{y} = \beta_1 \mathbf{H}_1 \mathbf{H}_1^\dagger \mathbf{s} + \beta_2 \mathbf{H}_2 \mathbf{H}_2^\dagger \mathbf{s} + \mathbf{n} = (\beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2) \mathbf{s} + \mathbf{n}, \tag{2}$$

TABLE I: Rates across relay network for considered cases.

Case	Occurs when	Rate
(i)	$R1$ and $R2$ active	$\min\left(\min\left(\mathfrak{C}_{\mathbf{G}_1}, \mathfrak{C}_{\mathbf{G}_2}\right), \mathfrak{C}_\mathbf{p}\right)$
(ii)	$R1$ only active	$\min\left(\mathfrak{C}_{\mathbf{G}_1}, \mathfrak{C}_{\mathbf{H}_1}\right)$
(iii)	$R2$ only active	$\min\left(\mathfrak{C}_{\mathbf{G}_2}, \mathfrak{C}_{\mathbf{H}_2}\right)$
(iv)	No relays active	$\mathbf{0}$

where $\mathbf{n}$ is the total noise, $\beta_i = \frac{1}{h_i^m}$ is the attenuation due to distance over $\mathbf{H}_i$ and $\mathbf{X}_i = \mathbf{H}_i \mathbf{H}_i^\dagger$.

We show that, for large N, the instantaneous capacities of $\mathbf{G}_i$ and $\mathbf{H}_i$ and also of the combined link in (2) converge to fixed limits. Define the capacities of $\mathbf{G}_i$, $\mathbf{H}_i$ and the combined channel, as $\mathfrak{C}_{\mathbf{G}_i}$, $\mathfrak{C}_{\mathbf{H}_i}$ and $\mathfrak{C}_\mathbf{p}$ respectively. Four possibilities arise according to whether $R1$, $R2$, both or neither are active in our system. Given that the individual capacities converge, the overall capacity in each case is given by the lowest rate across any of the contributing channels as described in Table I.

By finding the combined capacity $\mathfrak{C}_\mathbf{p}$ in T2, for the asymmetric case where $h_1 \neq h_2$, we quantify the benefits of using co-operative relays by comparing the maximum achievable capacity in case (i) with cases (ii), (iii) and (iv).

III. THEORY

Consider the case where the signal-to-noise-ratio is allocated uniformly and given by SNR $= \frac{\rho}{\sigma^2 N_t}$ per transmit antenna, where ρ is the total input power and σ^2 is the variance of the noise vector $\mathbf{n}$. The ergodic capacity of the MIMO channel in (1) is then given by [3]:

$$\mathfrak{C}_{\mathbf{G}_i}^{Erg} = \mathbb{E}\left(\log_2\left|\mathbf{I}_{N_r} + \text{SNR}\, \alpha_i^2\, \mathbf{G_i G_i}^\dagger\right|\right). \tag{3}$$

For large massive MIMO channels, on the other hand, the ergodic capacity is given by the limit of the instantaneous channel capacity as N_r, N_t tend to infinity while the ratio $\frac{N_r}{N_t}$ tends to the fixed limit ζ (in our case $N_r = N_t \Rightarrow \zeta = 1$):

$$\begin{aligned}\mathfrak{C}_{\mathbf{G_i}}^{Asy} &= \log_2\left|\mathbf{I}_{N_r} + \text{SNR}\, \alpha_i^2\, \mathbf{X}\right| \\ &= \sum_{i=1}^{N_r} \log_2\left(1 + \text{SNR}\, \alpha_i^2\, \lambda_\mathbf{X}(k)\right) \\ &= N_r \int_0^\infty \log_2\left(1 + \text{SNR}\, \alpha_i^2\, x\right) f_\mathbf{X}(x)\, dx, \quad (4)\end{aligned}$$

where $\mathbf{X} = \mathbf{G}_i \mathbf{G}_i^\dagger$, $\lambda_\mathbf{X}(k)$ is the kth eigenvalue of $\mathbf{X}$ and $f_\mathbf{X}(x)$ is the limiting distribution of the eigenvalues of $\mathbf{X}$ [7]. When $f_\mathbf{X}(x)$ exists it is called the asymptotic eigenvalue distribution (AED) of $\mathbf{X}$ and we call the capacity found in this manner the asymptotic capacity of $\mathbf{H}$, to emphasise the fact that it is found by taking asymptotic limits, rather than by applying (3). This result demonstrates the importance of the channel matrix, and in particular its eigenvalue distribution, in calculating the capacity for a MIMO channel.

Assuming the communication over the system in Fig. 1 occurs in a rich scattering environment and that each of the paths across $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{H}_1$ and $\mathbf{H}_2$ are independent and subject to Rayleigh fading, the channels can be modelled as zero-mean independently and identically distributed Gaussian

complex random matrices with normalised variance $\frac{1}{N}$. The AED of such matrices are given by the Marčenko-Pastur law [3]:

$$f_{\mathbf{X}}(x) = \left(1 - \frac{1}{\zeta}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+ (b-x)^+}}{2\pi\zeta x}, \quad (5)$$

where $(z)^+ = \max(0,z)$, $a = (1 - \sqrt{\zeta})^2$, $b = (1 + \sqrt{\zeta})^2$ and $\delta(x) = 1$ if $x = 0$ and $\delta = 0$ otherwise.

Inserting this result into (4) we can derive the asyptotic capacities $\mathfrak{C}_{\mathbf{G}_i}$ and $\mathfrak{C}_{\mathbf{H}_i}$. However, it is less straightforward to derive $\mathfrak{C}_{\mathfrak{p}}$. If we attempt to do so using (3) the capacity across the overall channel in T2 requires the computationally demanding calculation of the matrix polynomial $\mathfrak{p}$ where

$$\begin{aligned}\mathfrak{p} &= (\beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2)(\beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2)^\dagger \\ &= \beta_1^2\mathbf{X}_1^2 + \beta_1\beta_2(\mathbf{X}_1\mathbf{X}_2 + \mathbf{X}_2\mathbf{X}_1) + \beta_2^2\mathbf{X}_2^2. \end{aligned} \quad (6)$$

However, taking the asymptotic limit using (4) gives:

$$\mathfrak{C}_{\mathfrak{p}}^{Asy} = N_r \int_0^\infty \log_2\left(1 + \text{SNR } x\right) f_{\mathfrak{p}}(x)\, dx, \quad (7)$$

where the asymptotic capacity is given in terms of $f_{\mathfrak{p}}(x)$, the AED of the polynomial $\mathfrak{p}$, rather than as a function of a matrix polynomial. Now, the AEDs $f_{\mathbf{X}_1}(x)$ and $f_{\mathbf{X}_2}(x)$ are given by (5), and $f_{\mathfrak{p}}(x)$ depends only on $\mathbf{X}_1$ and $\mathbf{X}_2$. Unfortunately, traditional RMT is generally unable to derive the AED for the sum, product or any other polynomial combination of multiple random matrices, given only their AEDs [3]. In [12] and [13], Voiculescu is able to apply FPT to solve this problem for $\mathbf{H}_1 + \mathbf{H}_2$ and $\mathbf{H}_1 \times \mathbf{H}_2$ through the use of the R- and S-transform respectively. However, these results cannot be extended for more general $\mathfrak{p}$ as this would require that the Stieltjes transform of the polynomial AED be given in a closed form, which is not generally the case [3]. We overcome this problem using a method derived by Belinschi, Mai and Speicher in [14], which combines the analytic theory of operator-valued free convolution with Anderson's self-adjoint version of the 'linearization trick' [15] to determine $f_{\mathfrak{p}}(x)$. Armed with this distribution we can use (4) to compute the capacity-SNR relationship and analyse the overall performance of our system, in a novel and compute efficient way.

A. Requirements

When applying FPT we view the random matrix variables $\mathbf{X}_i$ as asymptotically free random linear operators with respect to probability measure $\varphi(\mathbf{X}_i) = \frac{1}{N}Tr(\mathbf{X}_i)$, the normalised trace of $\mathbf{X}_i$. We indicate this change of perspective by using lowercase x_i to represent the operator-valued variables and rewrite (6) as:

$$\mathfrak{p} = \beta_1^2 x_1^2 + \beta_1\beta_2(x_1 x_2 + x_2 x_1) + \beta_2^2 x_2^2. \quad (8)$$

There are two restrictions that must be placed upon x_1 and x_2 in order to apply the method of [14]. First, they must be 'asymptotically free'. The definition and fulfilment of this condition for our case are described in Section 2.4.1

and Example 2.36 of [3] respectively. The second requirement is that x_1 and x_2 must be self-adjoint. This is immediate, since $\mathbf{X}_i^\dagger = (\mathbf{H}_i\mathbf{H}_i^\dagger)^\dagger = \mathbf{H}_i\mathbf{H}_i^\dagger = \mathbf{X}_i$, and the self-adjointness of the matrices $\mathbf{X}_i$ and the operator-valued free random variables x_i are equivalent.

B. Linearization

The first step of the method is to convert our polynomial problem in random matrix variables to one in linear additive convolution by applying Anderson's self-adjoint version of the 'linearization trick' [15].

Definition 1: Let x_1 and x_2 be freely independent variables and let

$$\hat{\mathfrak{p}} = \begin{pmatrix} 0 & \mathbf{u} \\ \mathbf{v} & \mathbf{Q} \end{pmatrix}$$

be an $N \times N$ matrix where $\mathbf{u}$ is an $1 \times (N-1)$, $\mathbf{v}$ an $(N-1) \times 1$ and $\mathbf{Q}$ is an $(N-1) \times (N-1)$ matrix. All entries of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{Q}$ are polynomials in x_1 and x_2 with degree ≤ 1 so that each entry of the matrix $\hat{\mathfrak{p}}$ has the form $\gamma_1 x_1 + \gamma_2 x_2 + \gamma_3$ for $\gamma_1, \gamma_2, \gamma_3 \in \mathbb{C}$. We call $\hat{\mathfrak{p}}$ a linearisation of the polynomial $\mathfrak{p}$ if $\mathfrak{p} = -\mathbf{u}\mathbf{Q}^{-1}\mathbf{v}$.

For $\mathfrak{p}$ as given in (8), we find that taking:

$$\mathbf{u} = \frac{1}{\sqrt{2}}\begin{pmatrix} \beta_1 x_1 + \beta_2 x_2 & \beta_1 x_1 + \beta_2 x_2 \end{pmatrix},$$

$$\mathbf{v} = \frac{1}{\sqrt{2}}\begin{pmatrix} \beta_1 x_1 + \beta_2 x_2 \\ \beta_1 x_1 + \beta_2 x_2 \end{pmatrix} \text{ and } \mathbf{Q} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

produces a linearisation that satisfies Definition 1 and can be written as the operator-valued linear combination $\hat{\mathfrak{p}} = \hat{x}_1 + \hat{x}_2$ where the operator-valued distributions of

$$\hat{x}_1 = \begin{pmatrix} 0 & \frac{\beta_1 x_1}{\sqrt{2}} & \frac{\beta_1 x_1}{\sqrt{2}} \\ \frac{\beta_1 x_1}{\sqrt{2}} & 0 & 0 \\ \frac{\beta_1 x_1}{\sqrt{2}} & 0 & 0 \end{pmatrix}, \quad \hat{x}_2 = \begin{pmatrix} 0 & \frac{\beta_2 x_2}{\sqrt{2}} & \frac{\beta_1 2 x_2}{\sqrt{2}} \\ \frac{\beta_2 x_2}{\sqrt{2}} & 0 & -1 \\ \frac{\beta_2 x_2}{\sqrt{2}} & -1 & 0 \end{pmatrix},$$

can be derived from $f_{x_1}(x)$ and $f_{x_2}(x)$ respectively [14].

C. Subordination Theorem

To find the distribution of the polynomial $\mathfrak{p}$ using this deconstruction requires the use of the Cauchy transform:

$$G_{\mathfrak{p}}(x) = \varphi\left(\frac{1}{x - \mathfrak{p}}\right),$$

for all x for which $x - \mathfrak{p}$ is invertible. Defining $\mathbf{q}$ as the $N \times N$ matrix with x as the top left entry and zeros elsewhere, we notice that

$$\mathbf{q} - \hat{\mathfrak{p}} = \begin{pmatrix} 1 & -\mathbf{u}\mathbf{Q}^{-1} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x - \mathfrak{p} & 0 \\ 0 & -\mathbf{Q} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\mathbf{Q}^{-1}\mathbf{v} & 1 \end{pmatrix},$$

so that the invertibility of $x - \mathfrak{p}$ and $\mathbf{q} - \hat{\mathfrak{p}}$ are equivalent. Standard matrix operations can then be used to show that the $(1,1)$th-entry of $(\mathbf{q} - \hat{\mathfrak{p}})^{-1}$ is $(x - \mathfrak{p})^{-1}$, so that the Cauchy

transform $G_{\mathfrak{p}}(x)$ is just the $(1,1)$th-entry of the operator-valued (3×3 matrix) Cauchy transform:

$$G_{\hat{\mathfrak{p}}}(\mathbf{q}) = \mathbf{I}_3 \otimes \ \varphi\left((\mathbf{q} - \hat{\mathfrak{p}})^{-1}\right)$$

$$= \begin{pmatrix} \varphi\left((x - \mathfrak{p})^{-1}\right) & \varphi(*) \\ \varphi(*) & \varphi(*) \end{pmatrix}, \quad (9)$$

where $\mathbf{I}_3$ is the 3×3 identity matrix [14]. Therefore, we can compute the Cauchy transform of $\mathfrak{p}$ if we can determine $\hat{\mathfrak{p}} = \hat{x}_1 + \hat{x}_2$ from the operator-valued distributions of $\hat{x}_1$ and $\hat{x}_2$. As explained in III-B, we are able to derive the distributions of $\hat{x}_1$ and $\hat{x}_2$ from $f_{x_1}(x)$ and $f_{x_2}(x)$. Moreover, having verified that x_1 and x_2 are freely independent, it follows from the basic properties of freeness that $\hat{x}_1$ and $\hat{x}_2$ are operator-valued freely independent. The main result in [14] tells us that, provided this is the case, it is possible to calculate the operator-valued Cauchy transform of $\hat{\mathfrak{p}} = \hat{x}_1 + \hat{x}_2$ from $G_{\hat{x}_1}(\mathbf{q})$ and $G_{\hat{x}_2}(\mathbf{q})$ using operator-valued free convolution:

[14, Theorem 2.2] *Let $\hat{x}_1$ and $\hat{x}_2$ be self adjoint operator-valued random variables free over an operator-valued non-commutative probability space (M, φ, B) with M a unital Banach algebra, B a unital Banach subalgebra of M and $\varphi : M \mapsto B$ a conditional expectation which preserves the unit. Then there exists a Fréchet analytic map $\omega : \mathbb{H}^+(B) \to \mathbb{H}^+(B)$ such that*

$$G_{\hat{x}_1 + \hat{x}_2}(\mathbf{q}) = G_{\hat{x}_1}(\omega(\mathbf{q})) \text{ for all } \mathbf{q} \in \mathbb{H}^+(B).$$

Moreover, let $\mathbf{q} \in \mathbb{H}^+(B)$ and define

$$\mathbb{H}^+(\mathcal{B}) := \left\{ \mathbf{q} \in \mathcal{B} : \frac{\mathbf{q} - \mathbf{q}*}{2\mathfrak{j}} > 0 \right\} \text{ and } h_{\hat{x}_i}(\mathbf{q}) := \frac{1}{G_{\hat{x}_i}(\mathbf{q})}.$$

Then $\omega(\mathbf{q})$ is the unique fixed point of the map

$$\Omega_{\mathbf{q}} : \mathbb{H}^+(B) \to \mathbb{H}^+(B), \quad \Omega_{\mathbf{q}}(\omega) = h_{\hat{x}_2}(h_{\hat{x}_1}(\omega) + \mathbf{q}) + \mathbf{q}$$

and

$$\omega(\mathbf{q}) = \lim_{n \to \infty} \Omega_{\mathbf{q}}^{\circ n}(\omega) \text{ for any } \omega \in \mathbb{H}^+(B),$$

where $\Omega_{\mathbf{q}}^{\circ n}$ denotes the nth iteration of $\Omega_{\mathbf{q}}$, $\mathbf{q}*$ represents the adjoint of $\mathbf{q}$ in M and $\mathfrak{j}$ is the imaginary number with unit length.

Now the Cauchy transforms $G_{\hat{x}_1}$ and $G_{\hat{x}_2}$ are uniquely determined by $\hat{x}_1$ and $\hat{x}_2$. Therefore, by rearranging the subordination form given in Theorem III-C, we may compute the Cauchy transform, $G_{\hat{\mathfrak{p}}}(\mathbf{q}) = G_{\hat{x}_1}(\omega(\mathbf{q}))$, via the iteration

$$\omega \longmapsto G_{\hat{x}_2}\left(b + G_{\hat{x}_1}(\omega)^{-1} - \omega\right)^{-1} - \left(G_{\hat{x}_1}(\omega)^{-1} - \omega\right).$$

As demonstrated in (9), the Cauchy transform $G_{\mathfrak{p}}(x)$ is just the $(1,1)$th entry of the matrix obtained for $G_{\hat{\mathfrak{p}}}(\mathbf{q})$. The final step is to use the Cauchy inversion formula, as was done in [7], to find the AED $f_{\mathfrak{p}}(x)$ itself from its Cauchy transform $G_{\mathfrak{p}}(x)$, that is:

$$f_{\mathfrak{p}}(x) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \Im\left(G_{\mathfrak{p}}\left(x + \sqrt{-1}\epsilon\right)\right). \quad (10)$$
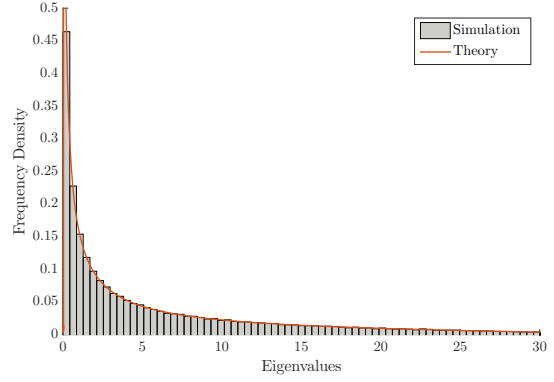

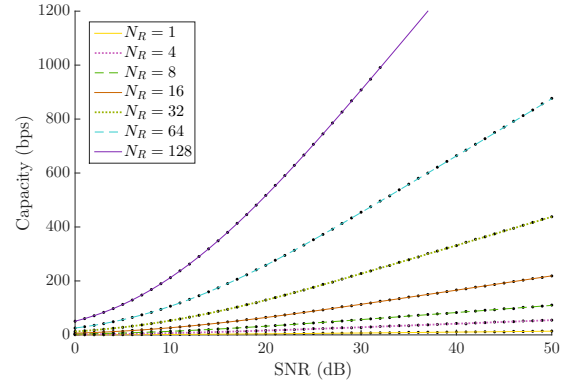Fig. 2: Simulated eigenvalue distribution vs. $f_{\mathfrak{p}}(x)$.

Fig. 3: Optimal capacity for given input SNR derived using simulation vs. theory.

IV. SIMULATION RESULTS

A. Accuracy

In this section we demonstrate the accuracy of the analytic techniques introduced in Section III. The analysis has been carried out for the case where we normalise the distances in the system model so that $g_1 = g_2 = 1$, $h_1 = 0.2^{-\frac{1}{2}}$ and $h_2 = 0.3^{-\frac{1}{2}}$ with $m = 2$ so that $\alpha_1 = \alpha_2 = 1$, $\beta_1 = 0.2$ and $\beta_2 = 0.3$. That is, the distances between the source node and each R_i are equal whereas the distance between R_1 and the destination is greater than that between R_2 and the destination, as illustrated (not to scale) in Fig. 1. Fig. 2 is the histogram of the eigenvalues of $\mathfrak{p}$ approximated by simulating ten thousand instances of $\mathbf{H}_1$ and $\mathbf{H}_2$ in the case where $N = 32$. This is overlaid by the distribution function $f_p(x)$, calculated using (10). It is clear that the results give an accurate approximation of the actual AED in this case.

Fig. 3 gives the capacity per input SNR achievable across the channel for arrays ranging from size $N = 1$ to $N = 128$. The solid lines in the graph represent the asymptotic capacity found using (7) while the points represent the simulation results found by applying (3) with $\mathfrak{p}$ replacing $\mathbf{G}_i \mathbf{G}_i^\dagger$. The asymptotic capacity predicted using FPT matches the simulation results with high accuracy. Surprisingly, even when $N = 1$
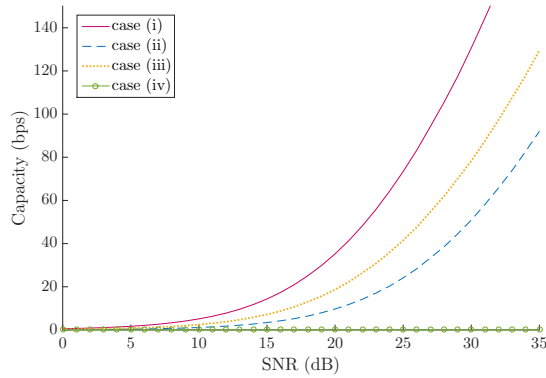
Fig. 4: Capacity for $N = 64$ in cases (i-iv).

we can see that the simulated points are well approximated by the theoretical asymptotic capacity, which means the asymptotic result can also be meaningfully applied to smaller arrays. However, for $N = 128$ we notice that there are no simulation points in Fig. 3 for input SNRs greater than 38 dB. This is due to the limitations a standard desktop-computer's hardware imposes on its ability to calculate the determinants of large matrices. Because these calculations involve both very large and very small numbers, and the computer cannot store these values with an infinite degree of accuracy, increasing the array size drastically magnifies the round-off error and the program outputs an error. As either N or the SNR increase further, this problem worsens, and the maximum number of antennas for which we can derive the capacity for $T2$ across the full range of SNR values, without using FPT, is 128.

We also note that the theoretical asymptotic capacity result took a fixed length of time (approximately three seconds) regardless of the size of the array, whereas the time complexity of the simulation increases as we increase N. In particular, when a thousand channel matrix realisations were computed as part of the simulation approach, it was outperformed by the FPT method for array sizes greater than $N = 2$.

B. Overall System Analysis

Finally, we use the asymptotic capacities for T1 and T2 to analyse the overall system model given in Fig. 1 using the rate equations for cases (i-iv) from Table I. The comparison given in Fig. 4 shows that the best rate is achieved in case (i) when all channels are viable and the FPT result applies. This is what we would expect because the ability for the signal to travel via both relays introduces an extra spatial dimension when compared to cases (ii) and (iii). Moreover case (iii) outperforms case (ii) which can be explained by the fact that $h_2 < h_1$, which means that channel $\mathbf{H}_2$ suffers less from attenuation due to distance. Thus, we have been able to efficiently quantify the benefit of using massive MIMO as part of a co-operative wireless communication system with an unlimited number of antennas, through the use of FPT.

V. CONCLUSIONS

In this paper we demonstrated how to use (7) to find the asymptotic capacity of a massive MIMO, co-operative relay system with multiple channels and hence multiple channel matrices. Our result enables a new, computationally efficient means of quantifying the achievable rate improvements when such systems employ large-scale antenna arrays. The method is used to determine the capacity limits for asymmetric relay arrangements regardless of how large the arrays at each node become. In particular, it is possible to use the FPT method to calculate the capacity of systems using antenna arrays greater than 128×128 in dimension, which is impossible to derive using standard simulation approach due to the nature of the operations involved. The method be readily extended to analyse co-operative networks with more than two relays, different numbers of transmit and receive antennas and even channels with different fading statistics, provided the AED of the channel matrix is known.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
[2] M. Pätzold, "5G developments are in full swing [mobile radio]," *IEEE Veh. Technol. Mag.*, vol. 12, no. 2, pp. 4–12, Jun. 2017.
[3] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers, 2004, vol. 1.
[4] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *IEEE Personal Commun. Mag.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
[5] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2832–2836, Nov. 1998.
[6] R. R. Müller, "Random matrices, free probability and the replica method," in *12th European Signal Process. Conf.*, Vienna, Austria, Sep. 2004.
[7] P. Pan, Y. Zhang, Y. Sun, and L. Yang, "On the asymptotic spectral efficiency of uplink MIMO-CDMA systems over Rayleigh fading channels with arbitrary spatial correlation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 679–691, Feb. 2013.
[8] Z. Zheng, R. S. L. Wei, R. R. Müller, J. Hämäläinen, and J. Corander, "Asymptotic analysis of Rayleigh product channels: A free probability approach," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1731–1745, Mar. 2017.
[9] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[10] Z. Ren, G. Wang, Q. Chen, and H. Li, "Modelling and simulation of Rayleigh fading, path loss, and shadowing fading for wireless mobile networks," *Sim. Modelling Prac. and Theory*, vol. 19, no. 2, pp. 626–637, Feb. 2011.
[11] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.
[12] D. Voiculescu, "Addition of certain non-commuting random variables," *J. Functional Analysis*, vol. 66, no. 3, pp. 323–346, May 1986.
[13] ——, "Multiplication of certain non-commuting random variables," *J. Operator Theory*, vol. 18, no. 2, pp. 223–235, Sep. 1987.
[14] S. T. Belinschi, T. Mai, and R. Speicher, "Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem," *J. Reine Angew. Math. (Crelles J.)*, vol. 2017, no. 732, pp. 21–53, Apr. 2015.
[15] G. W. Anderson, "Convergence of the largest singular value of a polynomial in independent Wigner matrices," *Ann. Probab.*, vol. 41, no. 3B, pp. 2103–2181, May 2013.