Power-Efficient Downlink Communication Using Large Antenna Arrays: The Doughnut Channel

Saif Khan Mohammed and Erik G Larsson

Saif Khan Mohammed and Erik G Larsson, Power-Efficient Downlink Communication Using Large Antenna Arrays: The Doughnut Channel, 2012, IEEE International Conference on Communications (ICC).

Power-Efficient Downlink Communication Using Large Antenna Arrays: The Doughnut Channel

Saif Khan Mohammed and Erik G. Larsson
Communication Systems Division, Dept. Electrical Engineering (ISY),
Linköping University, Sweden. E-mail: saif@isy.liu.se and erik.larsson@isy.liu.se.

***Abstract*—Large antenna arrays at the base station can facilitate *power efficient* single user downlink communication due to the inherent *array power gain*, i.e., under an *average only* total transmit power constraint, for a fixed desired information rate, the required total transmit power can be reduced by increasing the number of base station antennas (e.g. with i.i.d. fading, the required total transmit power can be reduced by roughly 3 dB with every doubling in the number of base station antennas, i.e., an $O(N)$ array power gain can be achieved with N antennas). However, in practice, building power efficient large antenna arrays would require power efficient amplifiers/analog RF components. With current technology, highly linear power amplifiers generally have low power efficiency, and therefore linearity constraints on power amplifiers must be *relaxed*. Under such relaxed linearity constraints, the transmit signal that suffers the least distortion is a signal with *constant envelope* (CE). In this paper, we consider a single user Gaussian multiple-input single-output (MISO) downlink channel where the signal transmitted from each antenna is constrained to have a constant envelope (i.e., for *every channel-use* the amplitude of the signal transmitted from each antenna is *constant*, *irrespective* of the channel realization). We show that under such a per-antenna CE constraint, the complex noise-free received signal lies in the interior of a "doughnut" shaped region in the complex plane. The per-antenna CE constrained MISO channel is therefore equivalent to a *doughnut channel*, i.e., a single-input single-output (SISO) AWGN channel where the channel input is constrained to lie inside a "doughnut" shaped region. Using this equivalence, we analytically compute a *closed-form* expression for an achievable information rate under the per-antenna CE constraint. We then show that, for a broad class of fading channels (i.i.d. and direct-line-of-sight (DLOS)), even under the more stringent per-antenna CE constraint (compared to the average only total power constraint), an $O(N)$ *array power gain* can still be achieved with N base station antennas. We also show that with $N \gg 1$, compared to the average only total transmit power constrained channel, the extra total transmit power required under the per-antenna CE constraint, to achieve a desired information rate is *small and bounded* for a broad class of fading channels (i.i.d. and DLOS). We also propose novel CE precoding algorithms. The analysis and algorithms presented are general and therefore applicable to conventional systems with a small number of antennas. Analytical results are supported with numerical results for the i.i.d. Rayleigh fading channel.**

I. Introduction

The high electrical power consumption in cellular base stations has been recognized as a major problem worldwide

This work was supported by the Swedish Foundation for Strategic Research (SSF) and ELLIIT. E. G. Larsson is a Royal Swedish Academy of Sciences (KVA) Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

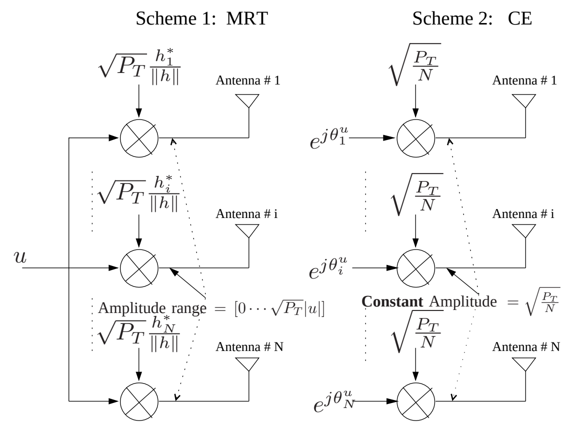

Fig. 1. Maximum Ratio Transmission (MRT) versus per-antenna Constant Envelope (CE) constrained transmission, for a given average total transmit power constraint of P_T. $\mathbf{h} = (h_1, \cdots, h_N)^T$ is the vector of complex channel gains.

[1]. One way of reducing the power consumed is to reduce the total radiated radio-frequency (RF) power. In theory, the total radiated power from a base station can be reduced (without affecting downlink throughput), simply by increasing the number of antennas. This has been traditionally referred to as the "array power gain" [2]. In addition to improving power-efficiency, there has been a great deal of recent interest in multi-user Multiple-Input Multiple-Output (MIMO) systems with *large antenna arrays*, due to their ability to substantially reduce intra-cell interference with very simple signal processing (see [3] for a recent work on communication with an *unlimited* number of antennas).

To illustrate the improvement in power efficiency with large antenna arrays, let us consider a downlink channel with the base station having $N > 1$ antennas and a single-antenna user. With knowledge of the channel vector ($\mathbf{h} = (h_1, h_2, \cdots, h_N)^T$) at the base station and an *average only* total transmit power constraint of P_T, information u (with mean energy $\mathbb{E}[|u|^2] = 1$) can be beamformed in such a way (i-th antenna transmits $\sqrt{P_T} h_i^* u / \|\mathbf{h}\|_2$) that the signals from different base station antennas add up *coherently* at the user (user receives $\sqrt{P_T}\|\mathbf{h}\|_2 u$), thereby resulting in an effective channel with a received signal power that is $\|\mathbf{h}\|_2^2/|h_1|^2$ times higher compared to a scenario where the base station has only one antenna. For a broad class of fading channels (i.i.d. fading, DLOS) $\|\mathbf{h}\|_2^2 = |h_1|^2 O(N)$, and *therefore, for a fixed desired received signal power, the total transmit power can be reduced by roughly half with every doubling in the number of base* *station antennas.* This type of beamforming is referred to as "Maximum Ratio Transmission" (MRT) (see Fig. 1).

In theory, to achieve an order of magnitude reduction in the total radiated power (without affecting throughput) would need base stations with "large" number of antennas (by large, we mean tens). However, building very large arrays in practice requires that each individual antenna, and its associated RF electronics, be *cheaply manufactured* and implemented in *power efficient* RF technology. It is known that conventional base stations are *highly power inefficient* (the ratio of radiated power to the total power consumed is less than 5 percent), the main reason being the usage of highly *linear* and power inefficient analog electronic components like the power amplifier [4].[2] Generally, high linearity implies low power efficiency and vice-versa. Therefore, non-linear *but highly power efficient* amplifiers must be used. With non-linear power amplifiers, the signal transmitted from each antenna must have a *low* peak-to-average-power-ratio, so as to avoid significant signal distortion. The type of signal that facilitates the use of most power-efficient and cheap power amplifiers/analog components is therefore a *constant envelope* (CE) signal.

With this motivation, in this paper, we consider the downlink of a single-user Gaussian MISO fading channel with the signal transmitted from each base station antenna constrained to have constant envelope. Fig. 1 illustrates the proposed signal transmission under a per-antenna CE constraint. Essentially, for a given information symbol u to be communicated to the end-user, the signal transmitted from the i-th antenna is $\sqrt{P_T/N}e^{j\theta_i^u}$. The transmitted phase angles $(\theta_1^u, \cdots, \theta_N^u)$ are determined in such a way that the noise-free signal received at the end-user *matches closely* with u. As shown in Fig. 1, under a per-antenna CE constraint, the amplitude of the signal transmitted from each antenna is *constant* (i.e., $\sqrt{P_T/N}$) for every channel-use, irrespective of the channel realization. In contrast, with MRT, the amplitude of the transmitted signal *depends upon the channel realization* as well as u, and can vary from 0 to $\sqrt{P_T}|u|$.[3] Since, the CE constraint is much more *restrictive* than the average only total power constraint, a natural question which arises now, is *whether, and how much array power gain can be achieved with the stringent per-antenna CE constraint ?*

So far, in reported literature, this question has not been addressed. For the special case of $N = 1$ (SISO AWGN), channel capacity under an input CE constraint has been reported in [5]. However, for $N > 1$, the only known reported works on per-antenna power constrained communication consider an average-power only constraint (see [6] and references therein). In contrast, in this paper, we consider a more *stringent* constraint i.e., the instantaneous per-antenna *per-channel-use* power is *constant* i.e., P_T/N (where P_T is the constant total power radiated per-channel-use, and is independent of the channel realization).

Specific contributions made in this paper are, i) we show that, under a per-antenna CE constraint the MISO downlink channel reduces to a SISO AWGN channel with the noise-free received signal being constrained to lie in a "*doughnut*" shaped region in the complex plane, ii) using the equivalent doughnut channel model, we compute a *closed-form* analytical expression for an achievable information rate, iii) we also propose *novel* algorithms for downlink precoding under the per-antenna CE constraint. Our results show that for a large class of fading channels (i.i.d. fading, DLOS), i) under the per-antenna CE constraint, an array power gain of $O(N)$ is *indeed achievable* with N antennas, ii) by choosing a sufficiently large antenna array, at high total transmit power P_T, the ratio of the information rate achieved under the CE constraint to the capacity of the average only total power constrained channel can be guaranteed to be *close to* 1, with high probability. This is in *contrast* to Wyner's result in [5] for $N = 1$, where this ratio is *only* $1/2$ at high P_T. Analytical results are supported with numerical results for the i.i.d. Rayleigh fading channel. We believe that the results and algorithms presented in this paper are novel and are expected to have a *profound* impact in *significantly* improving the *power efficiency* of cellular base stations by deploying *large antenna arrays at low cost*.

Notations: $\mathbb{C}$ and $\mathbb{R}$ denote the set of complex and real numbers. $|x|$ and $\arg(x)$ denote the absolute value and argument of $x \in \mathbb{C}$. For any $p \geq 1$ and $\mathbf{h} = (h_1, \cdots, h_N) \in \mathbb{C}^N$, $\|\mathbf{h}\|_p \stackrel{\Delta}{=} (\sum_i |h_i|^p)^{1/p}$. $\mathbb{E}[.]$ is the expectation operator. Abbreviations: r.v. (random variable), bpcu (bits-per-channel-use).

II. System model

We consider the downlink of a single user MISO system. The complex channel gain between the i-th transmit antenna and the user's receive antenna is denoted by h_i, and the channel vector by $\mathbf{h} = (h_1, h_2, \cdots, h_N)^T$. The base station is assumed to have perfect knowledge of $\mathbf{h}$, whereas the user is required to have only partial knowledge (we shall discuss this later in more detail). Let the complex symbol transmitted from the i-th antenna be denoted by x_i. The complex symbol received by the user is given by

$$y = \sum_{i=1}^{N} h_i x_i + w \tag{1}$$

where w denotes the circularly symmetric distributed AWGN noise having mean zero and variance σ^2. Due to the same CE constraint on each antenna and a total transmit power constraint of P_T, we must have $|x_i|^2 = P_T/N$, $i = 1, \ldots, N$. Therefore x_i must be of the form

$$x_i = \sqrt{\frac{P_T}{N}} e^{j\theta_i} \ , \ i = 1, 2, \ldots, N \tag{2}$$

where $j \stackrel{\Delta}{=} \sqrt{-1}$, and $\theta_i \in [-\pi \, , \, \pi)$ is the phase of x_i. We refer to the type of signal transmission in (2) as "CE

[2] In conventional base stations, about $40-50$ percent of the total operational power is consumed by the power amplifier and associated RF electronics which have a low power efficiency of about $5-10$ percent [4].

[3] The proposed transmission scheme is also different from Equal Gain Transmission (EGT). In a MISO channel with EGT, the signal transmitted from the i-th antenna is $\sqrt{P_T/(N\mathbb{E}[|u|^2])}\ u\ e^{j\theta_i}$, where the angles θ_i are chosen independently of u and depend only on the channel gains (for e.g. $e^{j\theta_i} = h_i^*/|h_i|$). Therefore in EGT, the envelope of the signal transmitted from each antenna depends upon u and therefore varies over time.

transmission". Note that under an average only total transmit power constraint, the transmitted signals are *only* required to satisfy $\mathbb{E}[\sum_i |x_i|^2] = P_T$, which is much *less restrictive* than (2). For the sake of notation, let $\Theta \stackrel{\Delta}{=} (\theta_1, \theta_2, \cdots, \theta_N)^T$ denote the vector of transmitted phase angles. With CE transmission, the signal received by the user is given by (using (1) and (2))

$$y = \sqrt{\frac{P_T}{N}} \sum_{i=1}^{N} h_i e^{j\theta_i} + w. \tag{3}$$

Let $u \in \mathcal{U} \subset \mathbb{C}$, denote the information symbol to be communicated to the user ($\mathcal{U}$ is the information symbol alphabet). For a given u, the precoder in the base station uses a map $\Phi(\cdot) : \mathcal{U} \rightarrow [-\pi, \pi)^N$ to generate the transmit phase angle vector, i.e.

$$\Theta = \Phi(u). \tag{4}$$

The range of the AWGN noise-free received signal scaled down by $\sqrt{P_T}$, i.e., $\sqrt{\frac{1}{N}} \sum_{i=1}^{N} h_i e^{j\theta_i}$, is given by

$$\mathcal{M}(\mathbf{h}) \stackrel{\Delta}{=} \left\{ \frac{\sum_{i=1}^{N} h_i e^{j\theta_i}}{\sqrt{N}} \ , \ \theta_i \in [-\pi, \pi) \ i = 1, \ldots, N \right\} \tag{5}$$

By choosing $\mathcal{U} \subseteq \mathcal{M}(\mathbf{h})$, for any $u \in \mathcal{U}$, it is implied that $u \in \mathcal{M}(\mathbf{h})$, and therefore from (5) it follows that, there exists a phase angle vector $\Theta^u = (\theta_1^u, \cdots, \theta_N^u)$ such that[4]

$$u = \sqrt{\frac{1}{N}} \sum_{i=1}^{N} h_i e^{j\theta_i^u}. \tag{6}$$

With the precoder map

$$\Phi(u) \stackrel{\Delta}{=} \Theta^u \tag{7}$$

where Θ^u satisfies (6), the received signal is given by

$$y = \sqrt{P_T}\, u + w \tag{8}$$

i.e., the AWGN noise-free received signal is the same as the intended information symbol u scaled up by $\sqrt{P_T}$.

If we choose $\mathcal{U} \not\subseteq \mathcal{M}(\mathbf{h})$, then it is clear that there exists some information symbol $u' \notin \mathcal{M}(\mathbf{h})$, for which any transmitted phase angle vector Θ would result in a received signal

$$y = \sqrt{P_T} u' + \sqrt{P_T} \Big(\frac{\sum_{i=1}^{N} h_i e^{j\theta_i}}{\sqrt{N}} - u' \Big) + w \tag{9}$$

where the energy of the *interference* term $\sqrt{P_T} \Big(\frac{\sum_{i=1}^{N} h_i e^{j\theta_i}}{\sqrt{N}} - u' \Big)$ is *strictly positive* for any Θ, since $u' \notin \mathcal{M}(\mathbf{h})$. This interference could then result in a loss in information rate.[5]

Motivated by the above arguments, subsequently in this paper, we propose to choose

$$\mathcal{U} \subseteq \mathcal{M}(\mathbf{h}) \tag{10}$$

and also that the precoder map is as defined in (7) and (6). With $\mathcal{U} \subseteq \mathcal{M}(\mathbf{h})$ it is clear that the information rate would depend upon $\mathcal{M}(\mathbf{h})$, and therefore we characterize it in the next Section.

III. Characterization of $\mathcal{M}(\mathbf{h})$

We characterize $\mathcal{M}(\mathbf{h})$ through a series of intermediate results. Due to lack of space we are unable to present proof for the intermediate results. Firstly, we define the maximum and minimum absolute value of any complex number in $\mathcal{M}(\mathbf{h})$.

$$\begin{aligned} M(\mathbf{h}) &\stackrel{\Delta}{=} \max_{\Theta=(\theta_1,\cdots,\theta_N)\,,\,\theta_i \in [-\pi,\pi)} \left| \frac{\sum_{i=1}^{N} h_i e^{j\theta_i}}{\sqrt{N}} \right| \\ m(\mathbf{h}) &\stackrel{\Delta}{=} \min_{\Theta=(\theta_1,\cdots,\theta_N)\,,\,\theta_i \in [-\pi,\pi)} \left| \frac{\sum_{i=1}^{N} h_i e^{j\theta_i}}{\sqrt{N}} \right| \end{aligned} \tag{11}$$

Lemma 1: If $z \in \mathcal{M}(\mathbf{h})$ then so does $z e^{j\phi}$ for all $\phi \in [-\pi, \pi)$.

The following two lemmas characterize $M(\mathbf{h})$ and $m(\mathbf{h})$.

Lemma 2: $M(\mathbf{h})$ is given by

$$M(\mathbf{h}) = \frac{\sum_{i=1}^{N} |h_i|}{\sqrt{N}} = \frac{\|\mathbf{h}\|_1}{\sqrt{N}}. \tag{12}$$

Lemma 3:

$$m(\mathbf{h}) \leq \frac{\|\mathbf{h}\|_\infty}{\sqrt{N}} = \frac{\max_{i=1,\ldots,N} |h_i|}{\sqrt{N}}. \tag{13}$$

The next theorem characterizes the set $\mathcal{M}(\mathbf{h})$.

Theorem 1:

$$\mathcal{M}(\mathbf{h}) = \Big\{ z \mid z \in \mathbb{C}\,,\, m(\mathbf{h}) \leq |z| \leq M(\mathbf{h}) \Big\}. \tag{14}$$

Proof – Let

$$(\theta_1^\star, \theta_2^\star, \cdots, \theta_N^\star) \stackrel{\Delta}{=} \arg \min_{\theta_i \in [-\pi,\pi)\,,\, i=1,2,\ldots,N} \left| \frac{\sum_{i=1}^{N} h_i e^{j\theta_i}}{\sqrt{N}} \right| \tag{15}$$

Consider the single variable function

$$f(t) \stackrel{\Delta}{=} \left| \frac{\sum_{i=1}^{N} h_i e^{j\theta_i(t)}}{\sqrt{N}} \right|^2 \ , \ t \in [0, 1] \tag{16}$$

where the functions $\theta_i(t)\,,\, i = 1, 2, \ldots, N$ are defined as

$$\theta_i(t) \stackrel{\Delta}{=} (1-t)\theta_i^\star - t \arg(h_i) \ , \ t \in [0, 1]. \tag{17}$$

Note that $f(t)$ is a differentiable function of t, and is therefore continuous for all $t \in [0, 1]$. Also from (15), Lemma 2 and (11) it follows that

$$f(0) = m(\mathbf{h})^2 \ , \ f(1) = M(\mathbf{h})^2 \tag{18}$$

[4] $\mathcal{U} \subseteq \mathcal{M}(\mathbf{h})$ implies that the information symbol alphabet is chosen adaptively with $\mathbf{h}$, and therefore the user must be informed about the newly chosen $\mathcal{U}$, every time it changes. By appropriately choosing $\mathcal{U}$ (whenever $\mathbf{h}$ changes), the base station need not send control information to the user about each element of the chosen $\mathcal{U}$. To be precise, we shall see in Section III that the set $\mathcal{M}(\mathbf{h})$ is the interior of a "doughnut" in the 2-dimensional complex-plane and can therefore be fully characterized with only 2 non-negative real numbers (the inner and the outer radius). Therefore, as an example, if we choose $\mathcal{U}$ to be square-QAM with its four maximal energy elements lying on the outer boundary of $\mathcal{M}(\mathbf{h})$, then the only information required to be sent to the user is the QAM alphabet size and the outer radius of $\mathcal{M}(\mathbf{h})$. A similar observation holds true for PSK sets also.

[5] For $\mathcal{U} \not\subseteq \mathcal{M}(\mathbf{h})$, it may be possible to consider a precoder map which for any $u \notin \mathcal{M}(\mathbf{h})$, finds the phase angle vector which minimizes the energy of the interference term. However, even with this interference-minimizing precoder, through simulations, it has been observed that for conventional alphabets like QAM,PSK, having $\mathcal{U} \not\subseteq \mathcal{M}(\mathbf{h})$, does not increase the achievable information rate compared to when $\mathcal{U} \subseteq \mathcal{M}(\mathbf{h})$.

Since $f(t)$ is continuous, it follows that for any non-negative real number c with $m(\mathbf{h})^2 \leq c^2 \leq M(\mathbf{h})^2$, there exists a value of $t = t' \in [0, 1]$ such that $f(t') = c^2$. Let

$$z' \triangleq \frac{\sum_{i=1}^{N} h_i e^{j\theta_i(t')}}{\sqrt{N}}. \tag{19}$$

From the definition of the set $\mathcal{M}(\mathbf{h})$ in (5), and (19) it is clear that $z' \in \mathcal{M}(\mathbf{h})$. From (19) and (16) it follows that

$$|z'| = \sqrt{f(t')} = c. \tag{20}$$

Therefore, we have shown that for any non-negative real number $c \in [m(\mathbf{h})\,,\, M(\mathbf{h})]$, there exists a complex number having modulus c and belonging to $\mathcal{M}(\mathbf{h})$.

Further, from Lemma 1, we already know that the set $\mathcal{M}(\mathbf{h})$ is circularly symmetric, and therefore all complex numbers with modulus c belong to $\mathcal{M}(\mathbf{h})$. Since the choice of $c \in [m(\mathbf{h})\,,\, M(\mathbf{h})]$ was arbitrary, any complex number with modulus in the interval $[m(\mathbf{h})\,,\, M(\mathbf{h})]$ belongs to $\mathcal{M}(\mathbf{h})$.

A. The proposed precoder map $\Phi(u) = \Theta^u$

The proof of theorem 1 is constructive and for a given $u \in \mathcal{U} \subseteq \mathcal{M}(\mathbf{h})$, it gives us a method to find the corresponding phase angle vector $\Theta^u = (\theta_1^u, \cdots, \theta_N^u)$ which satisfies (6). For a given $u \in \mathcal{U} \subseteq \mathcal{M}(\mathbf{h})$, we define the function

$$f_u(t) \triangleq f(t) - |u|^2 \ ,\ t \in [0, 1] \tag{21}$$

where $f(t)$ is given by (16). Using Newton-type methods or simple brute-force enumeration, we can find a $t = t_u$ satisfying $f_u(t_u) = 0$ (the existence of such a t_u is guaranteed by the constructive proof of theorem 1). The phase angles which satisfy (6) are then given by

$$\theta_i^u = \theta_i(t_u) + \phi \tag{22}$$

where $\theta_i(t)$ is given by (17), and ϕ is given by

$$e^{j\phi} = \frac{u\sqrt{N}}{\sum_{i=1}^{N} h_i e^{j\theta_i(t_u)}} \tag{23}$$

For large N, it has been observed that, most local minima of the error norm function $e^u(\Theta) \triangleq |u - \sum_{i=1}^{N} h_i e^{j\theta_i}/\sqrt{N}|^2$ have small error norms, and therefore *low-complexity* methods like the *gradient descent* method can be used to find Θ^u by minimizing $e^u(\Theta)$. However, with small N, for a significant fraction of local minima, the value of the error norm function *may not be small*, which leads to poor performance of the gradient descent method. Therefore, for small N, we propose the following two-step algorithm[6].

In the first step, we find a value of $\Theta = \tilde{\Theta}^u$ such that $|u - \sum_{i=1}^{N} h_i e^{j\tilde{\theta}_i^u}/\sqrt{N}|^2$ is *sufficiently* small. This step ensures that with high probability, $\tilde{\Theta}^u$ is *inside the region of attraction* of the global minimum of the error norm function. In the second step, with this $\Theta = \tilde{\Theta}^u = (\tilde{\theta}_1^u, \cdots, \tilde{\theta}_N^u)$ as the initial vector, a simple gradient descent algorithm would then converge to the global minimum.

The first step of the proposed algorithm is based on the Depth-First-Search (DFS) technique. Basically, for a given u, we start with enumerating the possible values taken by $\tilde{\theta}_N^u$ such that (6) is satisfied with $\Theta^u = \tilde{\Theta}^u$. To satisfy (6), it is clear that $\tilde{\theta}_N^u$ must equivalently satisfy

$$u - \frac{h_N e^{j\tilde{\theta}_N^u}}{\sqrt{N}} = \sqrt{\frac{N-1}{N}} \frac{\sum_{i=1}^{N-1} h_i e^{j\tilde{\theta}_i^u}}{\sqrt{N-1}}. \tag{24}$$

Using theorem 1 this then equivalently implies that, $(\sqrt{N}/\sqrt{N-1})(u - \frac{h_N e^{j\tilde{\theta}_N^u}}{\sqrt{N}}) \in \mathcal{M}((h_1, \cdots, h_{N-1})^T)$ i.e.

$$m(\mathbf{h}^{(N-1)}) \leq \sqrt{\frac{N}{N-1}} \Big| u - \frac{h_N e^{j\tilde{\theta}_N^u}}{\sqrt{N}} \Big| \leq M(\mathbf{h}^{(N-1)}) \tag{25}$$

where $\mathbf{h}^{(N-1)} \triangleq (h_1, \ldots, h_{N-1})^T$ and $m(\cdot), M(\cdot)$ are defined in (11). For example $M(\mathbf{h}^{(N-1)}) = \|\mathbf{h}^{(N-1)}\|_1/\sqrt{N-1}$. Equation (25) gives us an *admissible* set $I_N^u \subset [-\pi, \pi)$ to which $\tilde{\theta}_N^u$ must belong for (24) to be satisfied. We call this as the $k = 0$-th "depth" level of the proposed DFS technique.

Next, for a given value of $\tilde{\theta}_N^u \in I_N^u$, we go to the next "depth" level (i.e., $k = 1$) and find the set of admissible values for $\tilde{\theta}_{N-1}^u$. Essentially, at the k-th depth level, for a given choice of values of $(\tilde{\theta}_N^u, \tilde{\theta}_{N-1}^u, \ldots, \tilde{\theta}_{N-k+1}^u)$, with $\tilde{\theta}_{N-i+1}^u \in I_{N-i+1}^u, i = 1, \cdots, k$, we solve for the set of admissible values for $\tilde{\theta}_{N-k}^u$ such that (6) is satisfied with $\Theta^u = \tilde{\Theta}^u$. From theorem 1, this set (i.e., I_{N-k}^u) is given by the values of $\tilde{\theta}_{N-k}^u$ satisfying

$$\Big| u^{(k)} - \frac{h_{N-k} e^{j\tilde{\theta}_{N-k}^u}}{\sqrt{N}} \Big| \geq \sqrt{\frac{N-k-1}{N}}\ m(\mathbf{h}^{(N-k-1)})$$
$$\Big| u^{(k)} - \frac{h_{N-k} e^{j\tilde{\theta}_{N-k}^u}}{\sqrt{N}} \Big| \leq \sqrt{\frac{N-k-1}{N}}\ M(\mathbf{h}^{(N-k-1)}) \tag{26}$$

where $u^{(k)} \triangleq (u - \sum_{i=1}^{k} \frac{h_{N-i+1}}{\sqrt{N}} e^{j\tilde{\theta}_{N-i+1}^u})$ and $\mathbf{h}^{(N-k-1)} \triangleq (h_1, \ldots, h_{N-k-1})^T$. If there exists no solution to (26) (i.e., I_{N-k}^u is empty), then the algorithm backtracks to the previous depth level i.e., $k-1$, and picks the next possible unexplored admissible value for $\tilde{\theta}_{N-k+1}^u$ from the set I_{N-k+1}^u. If there exists a solution to (26), then the algorithm simply moves to the next depth level, i.e., $k+1$. The algorithm terminates once we reach a depth level of $k = N-1$ with a non-empty admissible set I_1. Since $u \in \mathcal{M}(\mathbf{h})$, the algorithm is guaranteed to terminate (by theorem 1). It can be shown that for depth levels less than $k = N-2$, the admissible set is generally an infinite set (usually a union of intervals in $\mathbb{R}$). Therefore, due to complexity reasons, at each depth level it is usually suggested to consider only a finite subset of values from the admissible set (e.g. values on a very fine grid), and terminate once the algorithm reaches a sufficiently high pre-defined depth level K with the current error norm i.e., $|u^{(K)}|$ below a pre-defined threshold.

In the second step, a gradient descent algorithm starting with the initial vector $\Theta = (\tilde{\theta}_N^u, \ldots, \tilde{\theta}_{N-K+1}^u, 0, \ldots, 0)$, converges to the global minimum of the error norm function $e^u(\Theta)$.

[6] It is to be noted here, that for $N = 2, 3$ there exist closed-form expressions for Θ^u and therefore the following algorithm is only required when N is greater than 3 and generally less than 10 (Since with large enough N, the low-complexity gradient descent method suffices).

IV. The Doughnut Channel and an Achievable Information Rate

From theorem 1 it is clear that, geometrically the set $\mathcal{M}(\mathbf{h})$ resembles a "*doughnut*" in the complex plane. Since we propose to use an information symbol set $\mathcal{U} \subseteq \mathcal{M}(\mathbf{h})$, and the precoder map as defined in (7), we effectively have a "*doughnut channel*" (see (8))

$$y = \sqrt{P_T}\, u + w \tag{27}$$

where the information symbol u is constrained to belong to the "doughnut" set $\mathcal{M}(\mathbf{h})$. For $N = 1$, the doughnut set contracts to a circle in the two-dimensional complex plane, and for which capacity is achieved when the input u is uniformly distributed on this circle, i.e., u has uniform phase [5].

For $N > 1$, we propose $\mathcal{U} = \mathcal{M}(\mathbf{h})$, with u "*uniformly*" distributed inside the doughnut, i.e., its probability density function (p.d.f.) is given by

$$P_u^{\text{unif}}(z) = \frac{1}{\pi(M(\mathbf{h})^2 - m(\mathbf{h})^2)}\ ,\ z \in \mathcal{M}(\mathbf{h}). \tag{28}$$

The information rate achieved with uniformly distributed u is given by

$$\begin{aligned} I(y;u)^{\text{unif}} &= h\Big(u + \frac{w}{\sqrt{P_T}}\Big) - h\Big(\frac{w}{\sqrt{P_T}}\Big) \\ &\geq \log_2(2^{h(u)} + 2^{h(w/\sqrt{P_T})}) - h(w/\sqrt{P_T}) \\ &= \log_2(1 + 2^{h(u) - h(w/\sqrt{P_T})}) \end{aligned} \tag{29}$$

where $h(s) \stackrel{\Delta}{=} -\int P_s(z)\log_2(P_s(z))dz$ denotes the differential entropy of the r.v. s ($P_s(\cdot)$ denotes the p.d.f. of s). The third step in (29) follows from the Entropy Power Inequality (EPI) [7][7] which states that for any complex r.v. (essentially a 2-real dimensional r.v.) $y = u + v$, which is the sum of two independent complex r.v's u and v, the differential entropy of y (in bits) satisfies the inequality $2^{h(y)} \geq 2^{h(u)} + 2^{h(v)}$. Since u is uniformly distributed inside $\mathcal{M}(\mathbf{h})$, we have $h(u) = \log_2(\pi(M(\mathbf{h})^2 - m(\mathbf{h})^2))$. Using this in (29), we have

$$I(y;u)^{\text{unif}} \geq \log_2\Big(1 + \frac{P_T}{\sigma^2}\frac{M(\mathbf{h})^2 - m(\mathbf{h})^2}{e}\Big) \tag{30a}$$

$$I(y;u)^{\text{unif}} \geq \log_2\Big(1 + \frac{P_T}{\sigma^2}\frac{\|\mathbf{h}\|_1^2 - \|\mathbf{h}\|_\infty^2}{Ne}\Big).\ \text{using lemma 2,3} \tag{30b}$$

We therefore have an achievable information rate given by the R.H.S. in the equations above. Note that, to achieve the information rate in the R.H.S of (30a), the receiver needs to have partial CSI only, i.e., it needs to only know $m(\mathbf{h})$ and $M(\mathbf{h})$, since these real non-negative numbers totally characterize the set $\mathcal{M}(\mathbf{h})$.

[7] With $N > 1$, a condition that is required for the usage of EPI to be valid is that $M(\mathbf{h}) > m(\mathbf{h})$, since otherwise the set $\mathcal{M}(\mathbf{h})$ has a zero Lebesgue measure leading to undefined $h(u)$. From Lemma 2 and 3 it follows that the condition $\|\mathbf{h}\|_1 > \|\mathbf{h}\|_\infty$ implies $M(\mathbf{h}) > m(\mathbf{h})$. Since, $\|\mathbf{h}\|_1 > \|\mathbf{h}\|_\infty$ holds for any $\mathbf{h}$ having more than one non-zero component, the required condition is met for most channel fading scenarios of practical interest.

V. Information rate comparison : CE vs. MRT

With an average only total transmit power constraint, MRT with Gaussian information alphabet achieves the capacity of the single user Gaussian MISO channel, which is given by

$$C = \log_2\Big(1 + \|\mathbf{h}\|_2^2\frac{P_T}{\sigma^2}\Big). \tag{31}$$

For a desired information rate, let the ratio of the total transmit power required under the per-antenna CE constraint to the power required under the average only total power constraint (APC) be referred to as the "power gap". From (30b) and (31) it now follows that the power gap can be upper bounded by $1/\kappa$, where

$$\kappa \stackrel{\Delta}{=} \frac{\|\mathbf{h}\|_1^2 - \|\mathbf{h}\|_\infty^2}{Ne\|\mathbf{h}\|_2^2} = \frac{\Big(\frac{\sum_i |h_i|}{N}\Big)^2 - \max_i \frac{|h_i|^2}{N^2}}{e\frac{\sum_i |h_i|^2}{N}} \tag{32}$$

Clearly $0 \leq \kappa < 1/e$ for any $\mathbf{h}$. From (30b) and (31) it also follows that

$$1 > \frac{I(y;u)^{\text{unif}}}{C} \geq 1 - \frac{\log_2\left(\frac{1}{\kappa}\right)}{C}. \tag{33}$$

For practical fading scenarios of interest like i.i.d. fading and DLOS, with sufficiently large N, κ can be shown to be greater than some strictly positive constant μ, with high probability. For example, for a single-path only direct-line-of-sight (DLOS) channel, we have $|h_1| = \ldots = |h_N|$. Using this fact, it can be shown that $1/\kappa \to e$ as $N \to \infty$, for any $\mathbf{h}$. With i.i.d. fading, as $N \to \infty$, using the law of large numbers and Slutsky's theorem [8] it can be shown that

$$\kappa \to_p \frac{(\mathbb{E}[|h_i|])^2}{e\mathbb{E}[|h_i|^2]} \tag{34}$$

where $\to_p$ means convergence in probability (as $N \to \infty$) w.r.t. the distribution of $\mathbf{h}$.[8] This then implies that, for any arbitrary $\epsilon > 0$, there exists an integer $N(\epsilon)$ such that with $N > N(\epsilon)$, the probability that a channel realization will have a value of $\kappa \geq \frac{(\mathbb{E}[|h_i|])^2}{e\mathbb{E}[|h_i|^2]} - \epsilon$ is greater than $1 - \epsilon$. From (34) it also follows that, the asymptotic ($N \to \infty$) power gap limit is $e\mathbb{E}[|h_i|^2]/(\mathbb{E}[|h_i|])^2$. For example, with i.i.d. Rayleigh fading this asymptotic power gap limit is $4e/\pi$, i.e., 5.4 dB.

For $N = 1$, it is known that, at large P_T/σ^2 (i.e., large C), for a given P_T/σ^2 the maximum information rate achieved with CE transmission is roughly *half* of the channel capacity under APC [5]. In contrast, with $N \gg 1$, from (33) it follows that CE transmission can achieve an information rate *close* to the capacity C under APC, since $1 - \frac{\log_2(1/\kappa)}{C}$ is *close to* 1 (as C is large, and κ is greater than a positive constant with high probability (as discussed in the paragraph above)). This fact is illustrated through Fig. 2, where we plot the ergodic information rate achieved with CE transmission (i.e., information rate averaged over the channel fading statistics which is assumed to be i.i.d. $\mathcal{CN}(0,1)$ Rayleigh fading). In Fig. 2, the exact $I(y;u)^{\text{unif}}$ has been computed numerically whereas the EPI lower bound is given by the R.H.S. of (30a).

[8] Here we have also used the fact that $\max_i |h_i|/N$ converges to zero in probability as $N \to \infty$. Results from order statistics, show that for large N, $\|\mathbf{h}\|_\infty = \max_i |h_i| = \mathbb{E}[|h_i|]O(\log(N))$, and therefore $\max_i |h_i|/N = \mathbb{E}[|h_i|]O(\frac{\log(N)}{N})$. (see [9] and references therein for more details)

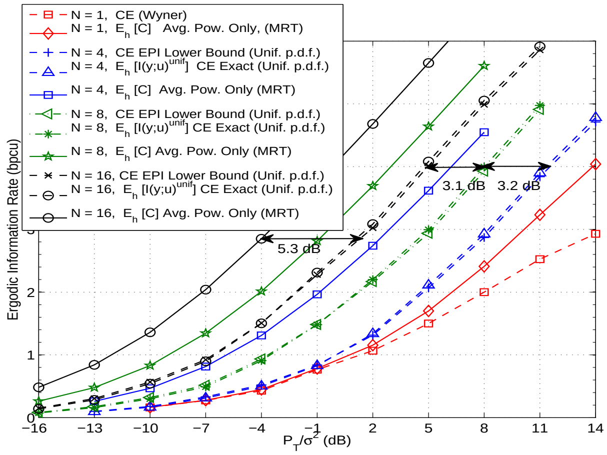

Fig. 2. Comparison between the ergodic information rate achieved with an average only total power constraint (MRT) to that achieved with constant envelope (CE) transmission. i.i.d. $\mathcal{CN}(0,1)$ Rayleigh fading assumed.

We observe that for $N > 1$, the information rate curve with CE transmission *runs parallel* to the capacity curve for an average only total power constrained (APC) channel.[9] This observation supports our analytical claim that, with high probability (w.r.t. the distribution of $\mathbf{h}$) the ratio $I(y;u)^{\text{unif}}/C$ is close to 1 for large C. However for $N = 1$ (as also reported in [5] for non-fading SISO AWGN channel), we observe that the CE information rate curve has a much *smaller* slope w.r.t. $\log(P_T/\sigma^2)$, when compared to the slope of the capacity curve for the APC channel. In Fig. 2, we also observe that with $N = 16$ and over a wide range of values of P_T/σ^2, the power gap is about $5.3 - 5.5$ dB (which matches closely with the asymptotic power gap limit of 5.4 dB as discussed earlier).

VI. Achievable Array Power Gain

For a desired rate R and a given precoding scheme; with N antennas, the *array power gain* achieved by this scheme is defined to be the factor of reduction in the total transmit power required to achieve a rate of R bpcu, when the number of base station antennas is increased from 1 to N. With an *average only total power constraint*, with N antennas the MRT precoder achieves an array power gain of (using (31))

$$G_N^{\text{MRT}}(R) = \frac{\sum_{i=1}^N |h_i|^2}{|h_1|^2} \tag{35}$$

which is $O(N)$ for i.i.d. fading and DLOS. With CE transmission, using the R.H.S of (30b) as the information rate, the array power gain achieved with N antennas is given by

$$G_N^{\text{CE}}(R) = N \, \frac{G_2^{\text{CE}}(R)}{2} \, \frac{\left\{ \left\{ \sum_{i=1}^N |h_i|/N \right\}^2 - \max_i |h_i|^2/N^2 \right\}}{\left\{ \left\{ \sum_{i=1}^2 |h_i|/2 \right\}^2 - \max_{i=1,2} |h_i|^2/4 \right\}}$$

where $G_2^{\text{CE}}(R)$ is the array power gain achieved with only 2 antennas and depends only on h_1 and h_2. From the equation above, it is clear that $G_N^{\text{CE}}(R)$ is $O(N)$ for i.i.d. fading and DLOS (for i.i.d. fading $\sum_i |h_i|/N \to_p \mathbb{E}[|h_i|]$ and $\max_i |h_i|/N \to_p 0$ as $N \to \infty$). The important result is therefore that, for practical fading scenarios like i.i.d. fading and DLOS, *an $O(N)$ array power gain can indeed be achieved even with CE transmission.*[10] This conclusion is validated in Fig. 2, where we observe that in increasing the number of antennas from $N = 4$ to $N = 8$ to $N = 16$, the required total transmit power to achieve a fixed desired information rate of 4 bpcu, *reduces* by a factor of roughly 3.0 dB for every *doubling* in the number of antennas. Similar conclusions can be drawn from Table I, where the required P_T/σ^2 to achieve an ergodic rate of 3 bpcu is listed as a function of N (i.i.d. $\mathcal{CN}(0,1)$ Rayleigh fading assumed). Also, CE transmission with even *small* N can *save* power, e.g., in Table I, the required total power with CE transmission and $N = 3$ is *less* than that required with $N = 1$ and an average only power constraint.

TABLE I
P_T/σ^2 (dB) required to achieve an ergodic rate of 3 bpcu

	N=1	N=2	N=3	N=4	N=8	N=16
MRT	10.1	6.4	4.3	2.9	-0.4	-3.5
CE	14.3	10.4	9.0	8.2	5.0	1.8

VII. Conclusions and Future Work

In this paper, we derived an achievable rate for a single-user Gaussian MISO downlink channel under the constraint that the signal transmitted from each antenna has a constant envelope. We showed that for i.i.d. fading and DLOS, even with the stringent per-antenna CE constraint, an $O(N)$ array power gain can still be achieved with N antennas. Also, compared to the average only total transmit power constrained channel, the extra total transmit power required under the CE constraint to achieve a desired rate (i.e., power gap), is shown to be bounded and small. We believe that these results hold true for a much broader class of fading channels, and are not limited to i.i.d. fading and DLOS. Future work involves deriving the capacity of the equivalent "doughnut" channel in order to exactly characterize the power gap. We would also extend results in this paper to the multi-user setting.

References

[1] A. Fehske, G. Fettweis, J. Malmodin and G. Biczok, "The Global Footprint of Mobile Communications: The Ecological and Economic Perspective," *IEEE Communications Magazine*, pp. 55-62, August 2011.

[2] D. N. C. Tse, "Fundamentals of Wireless Communications," *Cambridge University Press*, 2005.

[3] T. L. Marzetta, "Non-cooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE. Trans. on Wireless Communications*, pp. 3590–3600, vol. 9, no. 11, Nov. 2010.

[4] V. Mancuso and S. Alouf, "Reducing Costs and Pollution in Cellular Networks," *IEEE Communications Magazine*, pp. 63-71, August 2011.

[5] A. D. Wyner, "Bounds on Communication with Polyphase Coding," *Bell Sys. Tech. Journal*, pp. 523-559, vol. 45, Apr. 1966.

[6] M. Vu, "MISO Capacity with Per-Antenna Power Constraint," *IEEE Trans. on Communications*, pp. 1268-1274, vol. 59, no. 5, May 2011.

[7] S. Verdu and D. Guo, "A Simple Proof of the Entropy-Power Inequality," *IEEE Transactions on Information Theory*, pp. 2165-2166, vol. 52, no. 5, May 2006.

[8] T. S. Ferguson, "A Course in Large Sample Theory," Chapman and Hall, 1996.

[9] M. Sharif and B. Hassibi, "On the Capacity of MIMO Broadcast Channels with Partial Side Information," *IEEE Transactions on Information Theory*, pp. 506-522, vol. 51, no. 2, Feb. 2005.

[9] This is also true for small $N = 2, 3$, which we are unable to plot here due to space constraints.

[10] Even if $\mathcal{U}$ were chosen to be a *conventional finite information alphabet (e.g., QAM,PSK)*, using the fact that the outer radius of $\mathcal{M}(\mathbf{h})$ increases as $O(\sqrt{N})$ and the inner radius shrinks to 0, it can be analytically shown that, with sufficiently large N, the array power gain achieved would still be $O(N)$.