

## Mm Wave for MIMO systems in fading channel

Shani Mol.K.H<sup>1</sup>, Karthika.A.R<sup>2</sup>, Sumathi.D<sup>3</sup>,D.Ramalingam<sup>4</sup>

<sup>1</sup>PG scholar, Department of Electronics and Communication Engineering,  
James college of Engineering and Technology, Nagercoil, (India)

<sup>2</sup>Assistant professor, Department of Electronics and Communication Engineering,  
James college of Engineering and Technology, Nagercoil, (India)

<sup>3</sup>Assistant professor, Department of Electronics and Communication Engineering,  
James college of Engineering and Technology, Nagercoil, (India)

<sup>4</sup>Principal, James college of Engineering and Technology, Nagercoil, (India)

### ABSTRACT

Millimeter wave (mmWave) multiple-input multiple- output (MIMO) communication with large antenna arrays has been proposed to enable gigabit per second communication for next generation cellular systems and local area networks. A key difference relative to lower frequency solutions is that in mmWave systems, precoding/combining can not be performed entirely at digital baseband, due to the high cost and power consumption of some components of the radio frequency (RF) chain. In this paper we develop a low complexity algorithm for finding hybrid precoders that split the precoding/combining process between the analog and digital domains. Our approach exploits sparsity in the received signal to formulate the design of the precoder/combiners as a compressed sensing optimization problem. We use the properties of the matrix containing the array response vectors to find first an orthonormal analog precoder, since sparse approximation algorithms applied to orthonormal sensing matrices are based on simple computations of correlations. Then, we propose to perform a local search to refine the analog precoder and compute the baseband precoder. We present numerical results demonstrate substantial improvements in complexity while maintaining good spectral efficiency.

### I. INTRODUCTION

Millimeter wave (mmWave) is the new spectral frontier for next generation cellular networks and wireless local area networks [1], [2], [3], [4]. An important requirement in mmWave systems is the use of large arrays at the transmitter and receiver to provide a reasonable link budget. The antennas form a multiple-input multiple-output (MIMO) communication link that can be configured for different objectives. The de facto approach is spatial directivity, which provides beam forming gain needed to achieve a reasonable signal-to-noise ratio (SNR) at the receiver. MmWave channels though also have the ability to support spatial multiplexing of multiple data streams due to scattering and polarization [5], [6], [7], [8]. Unfortunately, power and cost requirements in the mmWave analog front-end make it challenging to implement the typical MIMO precoding transceiver

found in lower frequency systems, which is implemented in entirely in baseband. A solution is the hybrid precoding framework, where the precoding/combining process is divided between analog and digital domains [9], [10], [11].

A popular design of hybrid precoders for mmWave channels based on variable phase shifters was proposed in [9] for a particular mmWave system model incorporating: i) the constraints on the analog precoder/combiner, ii) presence of large antenna arrays, and iii) the limited scattering nature of the mmWave channel. The design of the precoders and combiners is formulated as a sparsity seeking optimization problem with hardware constraints. It resembles the problem of sparse signal recovery via multiple measurement vectors (MMV), also known as the simultaneous sparse recovery problem (S-OMP) [12]. The approach in [9] is elegant yet solving for the precoders still results in high complexity. A limitation of the work in [9], is that perfect channel state information is assumed at the receiver. This has been overcome in work on adaptive channel estimation [10], where the mmWave channel estimation problem is formulated as a compressed sensing problem, so that the channel parameters are estimated using standard CS tools. Training beamforming and combining vectors during the channel estimation phase are designed using a multi-resolution codebook. The main limitation of this work is that it assumes known array geometries for both the transmitter and receiver. Further investigation is also needed to obtain lower complexity solutions to both the channel estimation and the hybrid analog/digital precoding design problems. Hybrid precoding structures based on the use of variable phase shifters have been proposed earlier for general MIMO architectures in [13], but do not take into account the characteristics of millimeter wave propagation or leverage sparsity of the received signal. A related concept called beamspace MIMO communication has been proposed in [14], which uses a high-resolution discrete lens array for analog spatial beamforming. This avoids the need for phase shifters but does not have uniform performance across a broad range of angles.

In this paper we propose a low-complexity solution to the hybrid precoding optimization problem posed in [9]. We take into account the full structure of the optimization problem by exploiting the semi-unitary optimum precoder (optimum in the absence of hardware constraints). This structure reduces significantly the search space in the array manifold and thus leads to a lower complexity procedure versus that found in [9]. The reduction in complexity is due to an orthogonal matching step that fits the optimum precoder with the closest semi-unitary structure in the array manifold that emulates its behavior. The orthogonal matching step eliminates the need for the, slow, greedy matching pursuit steps deployed in the previous approach [9]. This step is then followed by a local search that further improves the solution by using either a fast one-by-one selection procedure or a full matching pursuit search but both only on a reduced section of the array manifold, around the semi-unitary solution previously found. Numerical results show that the computational advantage comes with no significant performance degradation in the proposed method as compared to previous results.

II. PROBLEM FORMULATION

Hybrid precoding in mmWave MIMO systems, i.e., the fully -connected architecture as shown in Fig. 1 (a) and the sub-connected architecture as shown in Fig. 1 (b). In both cases the BS has  $N \times M$  antennas but only  $N$  RF chains. From Fig. 1, we observe that the sub-connected architecture will likely be more energy-efficient, since it only requires  $N \times M$  PSs, while the fully-connected architecture requires  $N^2 \times M$  PSs. To fully achieve the spatial multiplexing gain, the BS usually transmits  $N$  independent data streams to users employing  $K$  receive antennas.

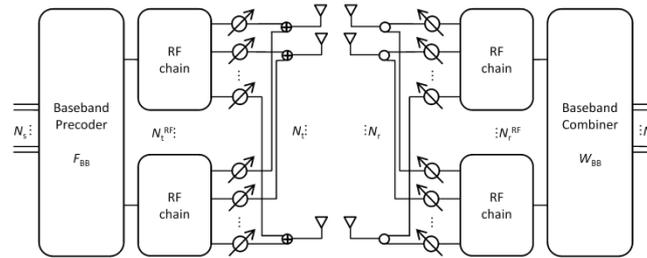


Fig. 1: Block diagram of a mmWave single user system with hybrid precoding: baseband precoding and radio frequency precoding with RF phase shifters.

In the sub-connected architecture as shown in Fig. 1 (b),  $N$  data streams in the baseband are precoded by the digital precoder  $D$ . In cases where complexity is a concern,  $D$  can be further specialized to be a diagonal matrix as

$D = \text{diag}[d_1, d_2, \dots, d_N]$ , where  $d_n \in \mathbb{R}$  for  $n = 1, 2, \dots, N$ . Then the role of  $D$  essentially performs some power allocation. After passing through the corresponding RF chain, the digital-domain signal from each RF chain is delivered to only  $M$  PSs [22] to perform the analog precoding, which can be denoted by the analog weighting vector  $\bar{a}_n \in \mathbb{C}^{M \times 1}$ , whose elements have the same amplitude  $\frac{1}{\sqrt{M}}$  but different phases [22]. After the analog precoding, each data stream is finally transmitted by a sub-antenna array with only  $M$  antennas associated with the corresponding RF chain. Then, the received signal vector  $Y = [y_1, y_2, \dots, y_k]^T$  at the user in a narrowband system

can be presented as

$$Y = \sqrt{\rho}HAD_s + n = \sqrt{\rho}HAD_s + n \tag{1}$$

Where  $\rho$  is the average received power;  $H \in \mathbb{C}^{K \times NM}$  denotes the channel matrix,  $A$  is the  $NM \times N$  analog precoding

matrix comprising  $N$  analog weighting vectors  $\{\bar{a}_m\}_{m=1}^N$  as

$$A = \begin{bmatrix} \bar{a}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \bar{a}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \bar{a}_N \end{bmatrix}_{NM \times N} \tag{2}$$

$S = [s_1, s_2, \dots, s_N]^T$  represents the transmitted signal vector in the baseband. In this paper, we assume the widely used Gaussian signals [10]–[13], [15]–[17] with normalized signal power  $E(SS^H) = \frac{1}{N}I_N$

while the practical system with finite alphabet inputs [23], [24] will be also briefly discussed in Section IV. P = AD presents the hybrid precoding matrix of size  $NM \times N$ , which satisfies  $\text{tr}(P^H P) \leq N$  to meet the total transmit power constraint. Finally  $n = [n_1, n_2, \dots, n_N]^T$  is an additive white Gaussian noise (AWGN) vector, whose entries follow the independent and identical distribution (i.i.d.)  $CN(0, \sigma^2)$ .

It is known that mmWave channel H will not likely follow the rich-scattering model assumed at low frequencies due to the limited number of scatters in the mmWave propagation environment [3]. In this paper, we adopt the geometric SalehValenzuela channel model to embody the low rank and spatial correlation characteristics of mmWave communications as

$$H = \gamma \sum_{l=1}^L \alpha_l \Lambda_r(\phi_l^r, \theta_l^r) \Lambda_t(\phi_l^t, \theta_l^t) \mathbf{f}_r(\phi_l^r, \theta_l^r) \mathbf{f}_t^H(\phi_l^t, \theta_l^t), \tag{3}$$

Where  $\gamma = \sqrt{\frac{NMK}{L}}$  is a normalization factor, L is the number of effective channel paths corresponding to the limited number of scatters, and we usually have  $L \leq N$  for mmWave communication systems.  $\alpha_l \in \mathbb{C}$  is the gain of the L path.  $\phi_l^r, \theta_l^r$  and  $\phi_l^t, \theta_l^t$  are the azimuth (elevation) angles of departure and arrival (AoDs/AoAs), respectively.  $\square_t \phi_l^r, \theta_l^r$  and  $\square_r \phi_l^t, \theta_l^t$  denote the transmit and receive antenna array gain at a specific AoD and AoA, respectively.

### III. LOW COMPLEXITY HYBRID PRECODING SOLUTIONS

#### SIC-BASED HYBRID PRECODING FOR MMWAVE MIMO SYSTEMS:

A low-complexity SIC-based hybrid precoding to achieve the near-optimal performance. The evaluation of computational complexity is also provided to show its advantages over current solutions.

##### A. Structure of SIC-based hybrid precoding

to maximize the total achievable rate R of mmWave MIMO systems, while other criteria such as the max-min fairness criterion [27] are also of interest. Specifically, R can be expressed as [11].

$$R = \log_2 \left( \left| I_K + \frac{\rho}{N\sigma^2} H P P^H H^H \right| \right). \tag{4}$$

According to the system model (1) in Section II, since the hybrid precoding matrix P can be represented as  $P = AD = \text{diag}\{a_1, \dots, a_N\} \cdot \text{diag}\{d_1, \dots, d_N\}$ , there are three constraints for the design of P: Constraint 1:

$P$  should be a block diagonal matrix similar to the form of  $A$  as shown in (2), i.e.,  $P = \text{diag} \{p_1, \dots, p_N\}$ , where  $p_n = d_n a_n$  is the  $M \times 1$  non-zero vector of the  $n$ th column  $p_n$  of  $P$ , i.e.,  $p_n = [0_{1 \times M(n-1)}, a_n, 0_{1 \times M(N-n)}]^T$ ; Constraint 2: The non-zero elements of each column of  $P$  should have the same amplitude, since the digital precoding matrix  $D$  is a diagonal matrix, and the amplitude of non-zero elements of the analog precoding matrix  $A$  is fixed to  $1/\sqrt{M}$ ; Constraint 3: The Frobenius norm of  $P$  should satisfy  $\|P\|_F \leq N$  to meet the total transmit power constraint, where  $N$  is the number of RF chains equal to the number of transmitted data streams. Unfortunately, these non-convex constraints on  $P$  make maximizing the total achievable rate (6) very difficult to be solved. However, based on the special block diagonal structure of the hybrid precoding matrix  $P$ , we observe that the precoding on different sub-antenna arrays are independent. This inspires us to decompose the total achievable rate (6) into a series of sub-rate optimization problems, each of which only considers one sub-antenna array. In particular, we can divide the hybrid precoding matrix  $P$  as  $P = [P_{N-1} \ p_N]$ , where  $p_N$  is the  $N$ th column of  $P$ , and  $P_{N-1}$  is an  $NM \times (N-1)$  matrix containing the first  $(N-1)$  columns of  $P$ . Then, the total achievable rate  $R$  in (6) can be rewritten as where (a) is obtained by defining the auxiliary matrix  $T_{N-1} = IK + \rho N \sigma^2 H P_{N-1} P_{N-1}^H H^H$ , and (b) is true due to the fact that  $|I + XY| = |I + YX|$  by defining  $X = T_{N-1}^{-1} N^{-1} H p_N$  and  $Y = P_n H^H$ . Note that the second term  $\log_2(1 + \rho N \sigma^2 p_N H^H T_{N-1}^{-1} N^{-1} H p_N)$  on the right side of (7) is the achievable sub-rate of the  $N$ th sub-antenna array, while the first term  $\log_2(|T_{N-1}|)$  shares the same form as (6). This observation implies that we can further decompose  $\log_2(|T_{N-1}|)$  using the similar method in (7) as where we have  $T_n = IK + \rho N \sigma^2 H P_n P_n^H H^H$  and  $T_0 = IN$ . From (8), we observe that the total achievable rate optimization problem can be transformed into a series of sub-rate optimization problems of sub-antenna arrays, which can be optimized one by one. After that, inspired by the idea of SIC for multi-user signal detection [21], we can optimize the achievable sub-rate of the first sub-antenna array and update the matrix  $T_1$ . Then, the similar method can be utilized to optimize the achievable sub-rate of the second sub-antenna array. Such procedure will be executed until the last subantenna array is considered. Fig. 2 shows the diagram of the proposed SIC-based hybrid precoding. Next, we will discuss how to optimize the achievable sub-rate of each sub-antenna array.

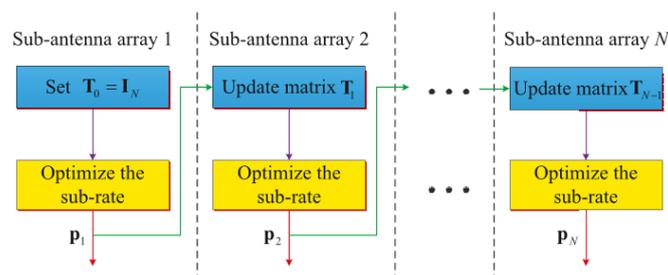


Fig. 2. Diagram of the proposed SIC-based hybrid precoding.

### B. Solution to the sub-rate optimization problem

In this subsection, we focus on the sub-rate optimization problem of the  $n$ th sub-antenna array, which can be directly applied to other sub-antenna arrays. According to (8), the sub-rate optimization problem of the  $n$ th sub-antenna array by

designing the  $n$ th precoding vector  $\mathbf{p}_n$  can be stated as  $\text{popt } n = \arg \max_{\mathbf{p}_n \in \mathcal{F}} \log_2 (1 + \rho N \sigma^2 \mathbf{p}_n^H \mathbf{G}_{n-1} \mathbf{p}_n)$ , where  $\mathbf{G}_{n-1}$  is defined as  $\mathbf{G}_{n-1} = \mathbf{H} \mathbf{T}_{n-1}^{-1} \mathbf{H}^H$ ,  $\mathcal{F}$  is the set of all feasible vectors satisfying the three constraints described in Section III-A. Note that the  $n$ th precoding vector  $\mathbf{p}_n$  only has  $M$  non-zero elements from the  $(M(n-1) + 1)$ th one to the  $(Mn)$ th one. Therefore, the sub-rate optimization problem (9) can be equivalently written as  $\text{popt } n = \arg \max_{\mathbf{p}_n \in \bar{\mathcal{F}}} \log_2 (1 + \rho N \sigma^2 \mathbf{p}_n^H \bar{\mathbf{G}}_{n-1} \mathbf{p}_n)$ , where  $\bar{\mathcal{F}}$  includes all possible  $M \times 1$  vectors satisfying Constraint 2 and Constraint 3,  $\bar{\mathbf{G}}_{n-1}$  of size  $M \times M$  is the corresponding sub-matrix of  $\mathbf{G}_{n-1}$  by only keeping the rows and columns of  $\mathbf{G}_{n-1}$  from the  $(M(n-1) + 1)$ th one to the  $(Mn)$ th one, which can be presented as

$$\bar{\mathbf{G}}_{n-1} = \mathbf{R} \mathbf{G}_{n-1} \mathbf{R}^H = \mathbf{R} \mathbf{H}^H \mathbf{T}_{n-1}^{-1} \mathbf{H} \mathbf{R}^H, \tag{4}$$

Define the singular value decomposition (SVD) of the Hermitian matrix  $\bar{\mathbf{G}}_{n-1}$  as  $\bar{\mathbf{G}}_{n-1} = \mathbf{V} \bar{\mathbf{\Lambda}} \mathbf{V}^H$ , where  $\bar{\mathbf{\Lambda}}$  is an  $M \times M$  diagonal matrix containing the singular values of  $\bar{\mathbf{G}}_{n-1}$  in a decreasing order, and  $\mathbf{V}$  is an  $M \times M$  unitary matrix. It is known that the optimal unconstrained precoding vector of (10) is the first column  $\mathbf{v}_1$  of  $\mathbf{V}$ , i.e., the first right singular vector of  $\bar{\mathbf{G}}_{n-1}$  [11]. However, according to the constraints mentioned in Section III-A, we cannot directly choose  $\text{popt } n$  as  $\mathbf{v}_1$  since the elements of  $\mathbf{v}_1$  do not obey the constraint of same amplitude (i.e., Constraint 2). To find a feasible solution to the sub-rate optimization problem (10), we need to further convert (10) into another form, which is given by the following Proposition 1.

Proposition 1. The optimization problem

$$\bar{\mathbf{p}}_n^{\text{opt}} = \arg \max_{\bar{\mathbf{p}}_n \in \bar{\mathcal{F}}} \log_2 \left( 1 + \frac{\rho}{N \sigma^2} \bar{\mathbf{p}}_n^H \bar{\mathbf{G}}_{n-1} \bar{\mathbf{p}}_n \right) \tag{5}$$

is equivalent to the following problem

$$\bar{\mathbf{p}}_n^{\text{opt}} = \arg \min_{\bar{\mathbf{p}}_n \in \bar{\mathcal{F}}} \|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2, \tag{6}$$

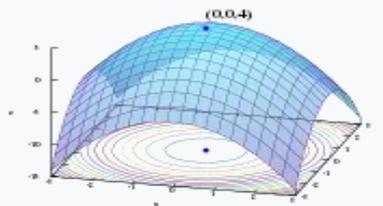
**Proposition 1** indicates that we can find a feasible precoding vector  $\bar{\mathbf{p}}_n$ , which is sufficiently close (in terms of Euclidean distance) to the optimal but unpractical precoding vector  $\mathbf{v}_1$ , to maximize the achievable sub-rate of the  $n$ th subantenna array.

$$\begin{aligned} & \|v_1 - \bar{p}_n\|_2^2 \\ &= (v_1 - d_n \bar{a}_n)^H (v_1 - d_n \bar{a}_n) \\ &= v_1^H v_1 + d_n^2 \bar{a}_n^H \bar{a}_n - 2d_n \operatorname{Re}(v_1^H \bar{a}_n) \\ &\stackrel{(a)}{=} 1 + d_n^2 - 2d_n \operatorname{Re}(v_1^H \bar{a}_n) \\ &= (d_n - \operatorname{Re}(v_1^H \bar{a}_n))^2 + (1 - [\operatorname{Re}(v_1^H \bar{a}_n)]^2), \end{aligned}$$

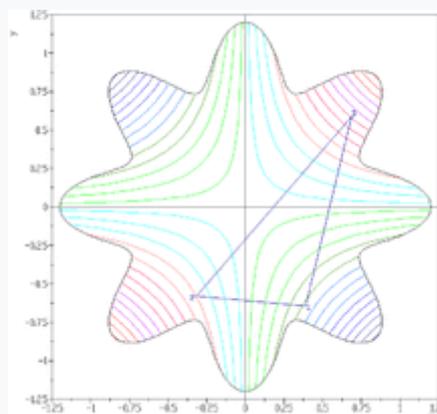
Since  $\bar{p}_n = d_n \bar{a}_n$  according to (1)

------(7)

""Optimization" and "Optimum" redirect here. For other uses, see Optimization(disambiguation) and Optimum (disambiguation).



Graph of a paraboloid given by  $z = f(x, y) = -(x^2 + y^2) + 4$ . The global maximum at  $(x, y, z) = (0, 0, 4)$  is indicated by a blue dot.



Nelder-Mead minimum search of Simionescu's function. Simplex vertices are ordered by their value, with 1 having the lowest (best) value.

In mathematics, computer science and operations research, mathematical optimization or mathematical programming, alternatively spelled optimisation, is the selection of a best element (with regard to some criterion) from some set of available alternatives.<sup>[1]</sup>

In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function. The generalization of optimization theory and techniques to other formulations constitutes a large area of applied mathematics. More generally, optimization includes finding "best available" values of some objective function given a defined domain (or input), including a variety of different types of objective functions and different types of domains.

An optimization problem can be represented in the following way: Given: a function  $f: A \rightarrow \mathbb{R}$  from some set  $A$  to the real numbers Sought: an element  $x_0$  in  $A$  such that  $f(x_0) \leq f(x)$  for all  $x$  in  $A$  ("minimization") or such that  $f(x_0) \geq f(x)$  for all  $x$  in  $A$  ("maximization"). Such a formulation is called an optimization problem or a mathematical programming problem (a term not directly related to computer programming, but still in use for example in linear programming – see History below). Many real-world and theoretical problems may be modeled in this general framework. Problems formulated using this technique in the fields of physics and computer vision may refer to the technique as energy minimization, speaking of the value of the function  $f$  as representing the energy of the system being modeled.

Typically,  $A$  is some subset of the Euclidean space  $\mathbb{R}^n$ , often specified by a set of constraints, equalities or inequalities that the members of  $A$  have to satisfy. The domain  $A$  of  $f$  is called the search space or the choice set, while the elements of  $A$  are called candidate solutions or feasible solutions.

The function  $f$  is called, variously, an objective function, a loss function or cost function (minimization), a utility function or fitness function (maximization), or, in certain fields, an energy function or energy functional. A feasible solution that minimizes (or maximizes, if that is the goal) the objective function is called an optimal solution.

In mathematics, conventional optimization problems are usually stated in terms of minimization. Generally, unless both the objective function and the feasible region are convex in a minimization problem, there may be several local minima. A local minimum  $\mathbf{x}^*$  is defined as a point for which there exists some  $\delta > 0$  such that for all  $\mathbf{x}$  where holds; that is to say, on some region around  $\mathbf{x}^*$  all of the function values are greater than or equal to the value at that point. Local maxima are defined similarly. While a local minimum is at least as good as any nearby points, a global minimum is at least as good as every feasible point. In a convex problem, if there is a local minimum that is interior (not on the edge of the set of feasible points), it is also the global minimum, but a nonconvex problem may have more than one local minimum not all of which need be global minima. A large number of algorithms proposed for solving nonconvex problems—including the majority of commercially available solvers—are not capable of making a distinction between locally optimal solutions and globally optimal solutions, and will treat the former as actual solutions to the original problem. Global optimization is the branch of applied mathematics and numerical analysis that is concerned



with the development of deterministic algorithms that are capable of guaranteeing convergence in finite time to the actual optimal solution of a nonconvex problem.

### C. Low-complexity algorithm to obtain the optimal solution

In computational complexity theory, a language  $B$  (or a complexity class  $B$ ) is said to be **low** for a complexity class  $A$  (with some reasonable relativized version of  $A$ ) if  $A^B = A$ ; that is,  $A$  with an oracle for  $B$  is equal to  $A$ .<sup>[1]</sup> Such a statement implies that an abstract machine which solves problems in  $A$  achieves no additional power if it is given the ability to solve problems in  $B$  at unit cost. In particular, this means that if  $B$  is low for  $A$  then  $B$  is contained in  $A$ . Informally, lowness means that problems in  $B$  are not only solvable by machines which can solve problems in  $A$ , but are "easy to solve." An  $A$  machine can simulate many oracle queries to  $B$  without exceeding its resource bounds. Results and relationships that establish one class as low for another are often called **lowness** results. The set of languages low for a complexity class  $A$  is denoted  $Low(A)$ .

. We start by considering how to avoid the SVD involving high computational complexity as well as a large number of divisions, which are difficult to be implemented in hardware. In computational complexity theory, a language  $B$  (or a complexity class  $B$ ) is said to be low for a complexity class  $A$  (with some reasonable relativized version of  $A$ ) if  $A^B = A$ ; that is,  $A$  with an oracle for  $B$  is equal to  $A$ . Such a statement implies that an abstract machine which solves problems in  $A$  achieves no Computational complexity theory is a branch of the theory of computation in theoretical computer science that focuses on classifying computational problems according to their inherent difficulty, and relating those classes to each other. A computational problem is understood to be a task that is in principle amenable to being solved by a computer, which is equivalent to stating that the problem may be solved by mechanical application of mathematical steps, such as an algorithm.

A problem is regarded as inherently difficult if its solution requires significant resources, whatever the algorithm used. The theory formalizes this intuition, by introducing mathematical models of computation to study these problems and quantifying the amount of resources needed to solve them, such as time and storage. Other complexity measures are also used, such as the amount of communication (used in communication complexity), the number of gates in a circuit (used in circuit complexity) and the number of processors (used in parallel computing). One of the roles of computational complexity theory is to determine the practical limits on what computers can and cannot do. Closely related fields in theoretical computer science are analysis of algorithms and computability theory. A key distinction between analysis of algorithms and computational complexity theory is that the former is devoted to analyzing the amount of resources needed by a particular algorithm to solve a problem, whereas the latter asks a more general question about all possible algorithms that could be used to solve the same problem. More precisely, computational complexity theory tries to classify problems that can or cannot be solved with appropriately restricted resources. In turn, imposing restrictions on the available resources is what distinguishes computational complexity from computability theory: the latter theory asks what kind of problems can, in principle, be solved algorithmically. Several natural complexity

classes are known to be low for themselves. Such a class is sometimes called self-low.<sup>[2]</sup> Scott Aaronson calls such a class a physical complexity class.<sup>[3]</sup> Note that being self-low is a stronger condition than being closed under complement. Informally, a class being low for itself means a problem can use others problems in the class as unit-cost subroutines without exceeding the power of the complexity class.

The following classes are known to be self-low:

- P is self-low (that is,  $P^P = P$ ) because polynomial-time algorithms are closed under composition: a polynomial-time algorithm can make polynomially many queries to other polynomial-time algorithms, while retaining a polynomial running time.
- PSPACE (with restricted oracle access mechanism) is also self-low, and this can be established by exactly the same argument.
- L is self-low because it can simulate log space oracle queries in log space, reusing the same space for each query.
- NC is also self-low for the same reason.
- BPP is also low for itself and the same arguments almost work for BPP, but one has to account for errors, making it slightly harder to show that BPP is low for itself.
- Similarly, the argument for BPP almost goes through for BQP, but we have to additionally show that quantum queries can be performed in coherent superposition.
- Both Parity P and BPP are low for themselves. These were important in showing Toda's theorem.<sup>[5]</sup>
- $NP \cap coNP$  is low for itself.<sup>[1]</sup>

Every class which is low for itself is closed under complement, provided that it is powerful enough to negate the boolean result. This implies that NP isn't low for itself unless  $NP = co-NP$ , which is considered unlikely because it implies that the polynomial hierarchy collapses to the first level, whereas it is widely believed that the hierarchy is infinite. The converse to this statement is not true. If a class is closed under complement, it does not mean that the class is low for itself.

An example of such a class is EXP, which is closed under complement, but is not low for itself. While low-complexity art does not require a priori restrictions of the description size, the basic ideas are related to the size-restricted intro categories of the demoscene, where very short computer programs are used to generate pleasing graphical and musical output. Very small (usually C) programs that create music have been written: the style of this music has come to be called "bytebeat"

#### **D. Summary of the proposed SIC-based hybrid precoding**

Image inpainting refers to filling in the missing parts or modifying the damaged parts of an image in a visually plausible way. Image inpainting, an artistic term used from ancient times, refers to restoration or retouching works of paintings. This technique can be used for restoring the missing parts of an image or for removing the



unwanted objects from an image. In digital images, role of image inpainting techniques grow from mere restoration of images, photographs and lms to powerful image enhancement and image completion. Successive interference cancellation (SIC) is a physical layer capability that allows a receiver to decode packets that arrive simultaneously. While the technique is well known in communications literature, emerging software radios are making practical experimentation feasible. This motivates us to study the extent of throughput gains possible with SIC from a MAC layer perspective. Contrary to our initial expectation, we find that the gains from SIC are not easily available in many realistic situations. Moreover, we observe that the scope for SIC gets squeezed by the advances in bitrate adaptation, casting doubt on the future of SIC based protocols. Let us define collision as the simultaneous arrival of two

or more packet transmissions at a receiver. Traditionally, only the strongest signal can be decoded, treating the other signal as interference. However, SIC facilitates recovery of even the weaker signal. For this, the bits of the stronger

signal are decoded as before. The original (stronger) signal is then reconstructed from these bits, and subtracted (i.e., cancelled) from the combined signal. The bits of the weaker packet are then decoded from this residue. This can be an iterative process to recover multiple packets and hence it is termed successive interference cancellation. It is worth pointing out that the idea of SIC-based hybrid precoding can be also extended to the combining at the user following the similar logic in [11]. When the number of RF chains at the BS is smaller than that at the user, we first compute the optimal hybrid precoding matrix  $P$  according to Algorithm 2, where we assume that the combining matrix  $Q = I$ . Then, given the effective channel matrix  $HP$ , we can similarly obtain the optimal hybrid combining matrix  $Q$  by referring to Algorithm 2, where the input  $_G0$  and the optimal unconstrained solution  $v1$  should be correspondingly replaced. Conversely, when the number of RF chains at the BS is larger than that at the user, we can assume  $P = I$  and obtain the optimal hybrid combining matrix  $Q$ . After that, the optimal precoding matrix  $P$  can be acquired given the effective channel matrix  $QH$ . Additionally, to further improve the performance, we can combine the above method with the “Ping-pong” algorithm [22], which involves an iteration procedure between the BS and the user, to jointly seek the optimal hybrid precoding and combining matrices pair. Further discussion about hybrid combining will be left for future work.

### E. Complexity evaluation

In this subsection, we provide the complexity evaluation of the proposed SIC-based hybrid precoding in terms of the required numbers of complex multiplications and divisions. The notion of communication complexity was introduced by Yao in 1979,<sup>[1]</sup> who investigated the following problem involving two separated parties (Alice and Bob). Alice receives an  $n$ -bit string  $x$  and Bob another  $n$ -bit string  $y$ , and the goal is for one of them (say Bob) to compute a certain function  $f(x,y)$  with the least amount of communication between them. Note that here we are not concerned about the number of computational steps, or the size of the computer memory used. Communication complexity tries to quantify the amount of communication required for such distributed computations. Of course they can always succeed by having Alice send her whole  $n$ -bit string to Bob, who then computes the function, but the idea here is to find clever ways of calculating  $f$  with fewer



than  $n$  bits of communication. This abstract problem, and its general form with more than two parties, is relevant in many contexts: in VLSI circuit design, for example, one wants to minimize energy used by decreasing the amount of electric signals required between the different components during a distributed computation. The problem is also relevant in the study of data structures, and in the optimization of computer networks. For a survey of the field, see the book by Kushilevitz and Nisan. From Algorithm 2, we observe that the complexity of SIC-based hybrid precoding comes from the following four parts:

- 1) The first one originates from the computation of  $\mathbf{G}_0 = \mathbf{R}\mathbf{H}\mathbf{H}^H\mathbf{R}$  according to (11). Note that  $\mathbf{R}$  is a selection matrix and  $\mathbf{H}$  has the size  $K \times NM$ . Therefore, this part involves  $KM^2$  times of multiplications without any division.
- 2) The second one is from executing Algorithm 1. It is observed that in each iteration we need to compute a matrix-to-vector multiplication  $\mathbf{z}(s) = \mathbf{G}_n^{-1}\mathbf{u}(s-1)$  together with the Aitken acceleration method (20). Therefore, we totally require  $S(M^2 + 2) - 4$  and  $(2S - 2)$  times of multiplications and divisions, respectively.
- 3) The third one stems from acquiring the optimal solution  $\mathbf{p}_{opt}$  in step 2 of Algorithm 2. We find that this part is quite simple, which only needs 2 times of multiplications without any division, since  $\mathbf{v}_1$  has been obtained is a fixed constant.
- 4) The last one comes from the update of  $\mathbf{G}_n$ . According to Proposition 2, we know that this part mainly involves a outer product  $\mathbf{v}_1\mathbf{v}_1^H$ . Thus, it requires  $M^2$  times of multiplications with only one division. To sum up, the proposed SIC-based hybrid precoding approximately requires  $O(M^2(NS+K))$  times of multiplications and  $O(2NS)$  times of divisions.

Table 1 provides the complexity comparison between SIC-based hybrid precoding and the recently proposed spatially sparse precoding [11], which requires  $O(N^4M + N^2L^2 + N^2M^2L)$  times of multiplications and  $O(2N^3)$  times of divisions. Here,  $L$  is the number of effective channel paths as defined in (3). Considering the typical mmWave MIMO system with  $N = 8$ ,  $M = 8$ ,  $K = 16$ ,  $L = 3$  [11], we observe that the complexity of SIC-based hybrid precoding is about  $4 \times 10^3$  times of multiplications and 102 times of divisions, where we set  $S = 5$  (note that  $S \geq 5$  is usually sufficient to guarantee the performance, which is verified through intensive simulations). By contrast, the complexity of the spatially sparse precoding is about  $5 \times 10^4$  times of multiplications and 103 times of divisions. Therefore, the proposed SIC-based hybrid precoding enjoys much lower complexity, which is only about 10% as complex as that of the spatially sparse precoding.

#### IV. NUMERICAL RESULTS

the proposed SIC-based hybrid precoding. We compare the performance of SIC-based hybrid precoding with the recently proposed spatially sparse precoding and the optimal unconstrained precoding based on the SVD of the channel matrix, which are both with fully-connected architecture. Additionally, we also include the conventional

analog precoding and the optimal unconstrained precoding (i.e.,  $n = \nu$ ) which are both with sub-connected architecture as benchmarks for comparison.

The simulation parameters are described as follows. We generate the channel matrix according to the channel model described in Section II. The number of effective channel paths is  $L = 3$ . The carrier frequency is set as 28GHz. Both the transmit and receive antenna arrays are ULAs with antenna spacing  $d = \lambda/2$ . Since the BS usually employs the directional antennas to eliminate interference and increase antenna gain, the AoDs are assumed to follow the uniform distribution within  $[-\pi/6, \pi/6]$ . Meanwhile, due to the random position of users, we assume that the AoAs follow the uniform distribution within  $[-\pi, \pi]$ , which means the unidirectional antennas are adopted by users. Furthermore, we set the maximum number of iterations  $S = 5$  to run Algorithm 2. Finally, SNR is defined as  $\frac{P}{\sigma^2}$ . Firstly, we consider the perfect channel state information (CSI) scenario.

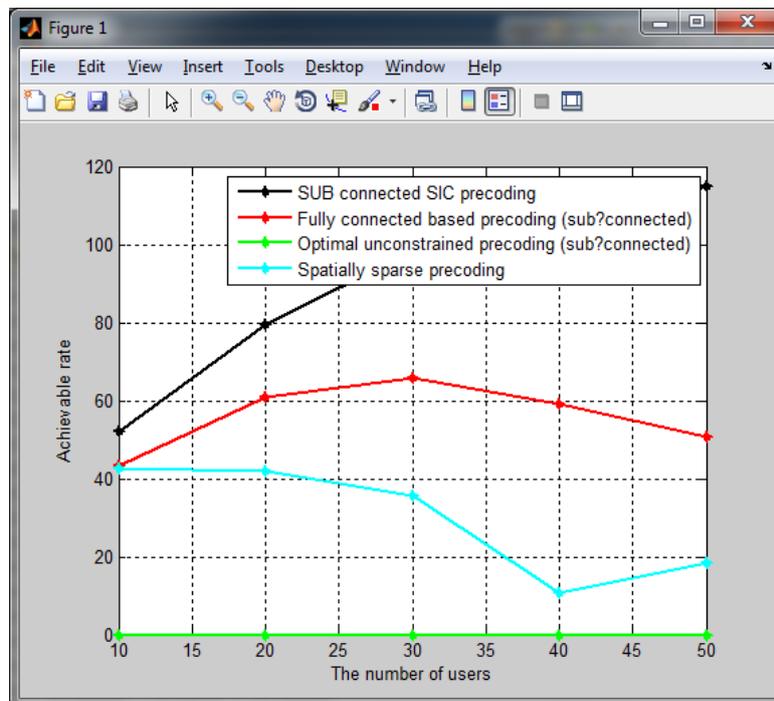


Fig. 3 shows the achievable rate comparison in mmWave MIMO system, where  $NM \times K = 64 \times 16$  and the number of RF chains is  $N = 8$ . We observe from Fig. 3 that the proposed SIC-based hybrid precoding outperforms the conventional analog precoding with sub-connected architecture in whole simulated SNR range. Meanwhile,

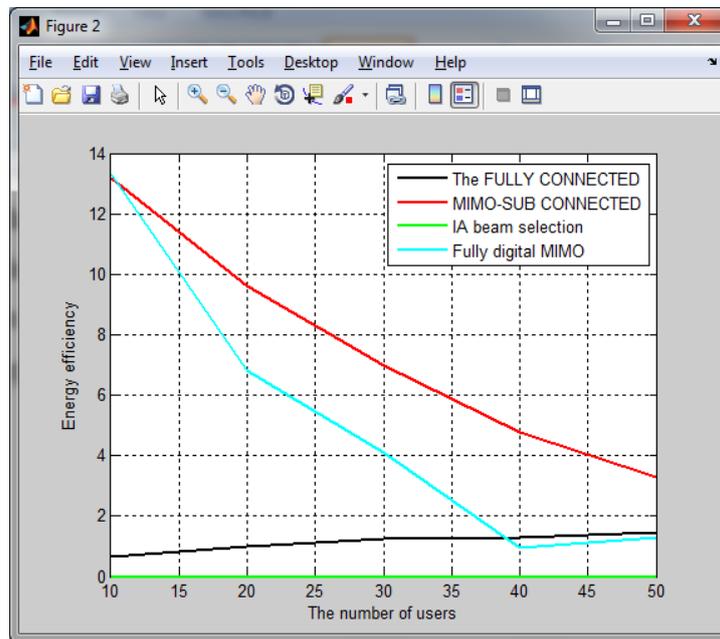


Fig. 3 also verifies the near-optimal performance of SIC-based hybrid precoding, since it can achieve about 99% of the rate achieved by the optimal unconstrained precoding with sub-connected architecture.

## V. CONCLUSIONS

In this paper we developed a new optimization algorithm for the design of hybrid precoders and combiners for mmWave MIMO systems. Our two solutions incorporate constraints that account for the practical hardware limitations at these frequencies: analog beamforming based on quantized variable phase shifters and the use of a limited number of RF chains. The main innovation in our work is to exploit the array geometry in a way that allows us to reduce the search complexity and thus the overall complexity of the algorithm. Simulation results show that the spectral efficiency achieved by using the new algorithms is comparable to the unconstrained solution, yet with substantially lower overall complexity.

## REFERENCES

- [1] T. Bai, A. Alkhateeb, and R. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sep. 2014.
- [2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [3] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

- [5] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 136–143, Dec. 2014.
- [6] M. Samimi, K. Wang, Y. Azar, G. N. Wong, R. Mayzus, H. Zhao, J. K. Schulz, S. Sun, F. Gutierrez, and T. S. Rappaport, "28 GHz angle of arrival and angle of departure analysis for outdoor cellular communications using steerable beam antennas in New York City," in *Proc. IEEE Vehicular Technology Conference (VTC Spring'13)*, May 2013, pp. 1–6.
- [7] A. Alkhateeb, J. Mo, N. Gonz alez-Prelcic, and R. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [8] B. Yin, S. Abu-Surra, G. Xu, T. Henige, E. Pisek, Z. Pi, and J. R. Cavallaro, "High-throughput beamforming receiver for millimeter wave mobile communication," in *Proc. IEEE Global Communications Conference (GLOBECOM'13)*, Dec. 2013, pp. 3697–3702.
- [9] P. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2222, Jun. 2015.
- [10] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
- [11] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [12] Y. Lee, C.-H. Wang, and Y.-H. Huang, "A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 305–317, Jan. 2015.
- [13] C.-E. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. PP, no. 99, Mar. 2015.
- [14] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [15] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [16] T. Kim, J. Park, J.-Y. Seol, S. Jeong, J. Cho, and W. Roh, "Tens of Gbps support with mmwave beamforming systems for next generation communications," in *Proc. IEEE Global Communications Conference (GLOBECOM'13)*, Dec. 2013, pp. 3685–3690.
- [17] T. Kim, C. Kim, and J.-Y. Seol, "A low complexity hybrid beamforming algorithm for multi-beam diversity transmission," in *Proc. IEEE Global Communications Conference (GLOBECOM'14)*, Dec. 2014.
- [18] C. A. Balanis, *Antenna theory: analysis and design*. John Wiley & Sons, 2012.
- [19] D. Schneider, "Could supercomputing turn to signal processors (again)?" *IEEE Spectrum*, vol. 49, no. 10, pp. 13–14, Oct. 2012.

- [20] S. Han, C.-L. I, Z. Xu, and S. Wang, "Reference signals design for hybrid analog and digital beamforming," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1191–1193, Jul. 2014.
- [21] Y.-C. Liang, E. Y. Cheu, L. Bai, and G. Pan, "On the relationship between MMSE-SIC and BI-GDFE receivers for large multiple-input multiple-output channels," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3627–3637, Aug. 2008.
- [22] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [23] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Apr. 2011.
- [24] Y. Wu, C. Xiao, X. Gao, J. D. Matyjas, and Z. Ding, "Linear precoder design for MIMO interference channels with finite-alphabet signaling," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3766–3780, Sep. 2013.
- [25] F. Zhu, F. Gao, M. Yao, and H. Zou, "Joint information- and jamming beamforming for physical layer security with full duplex base station," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6391–6401, Dec. 2014.
- [26] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. IEEE Information Theory and Applications Workshop (ITA'13)*, 2013, pp. 1–5.
- [27] M. Tao and R. Wang, "Linear precoding for multi-pair two-way MIMO relay systems with max-min fairness," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5361–5370, 2012.
- [28] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012.
- [29] A. Björck, "Numerical methods in matrix computations," *Texts in Applied Mathematics*, 2015.
- [30] O. El Ayach, R. W. Heath, S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *Proc. IEEE Global Communications Conference (GLOBECOM'13)*, Dec. 2013, pp. 3476–3480.