How Long to Estimate Sparse MIMO Channels

Yahia Shabara, C. Emre Koksal and Eylem Ekici
Dept. of ECE, The Ohio State University, Columbus, OH 43210
Email: {shabara.1, koksal.2 , ekici.2}@osu.edu

***Abstract*—Large MIMO transceivers are integral components of next-generation wireless networks. However, for such systems to be practical, their channel estimation process needs to be fast and reliable. Although several solutions for fast estimation of sparse channels do exist, there is still a gap in understanding the fundamental limits governing this problem. Specifically, we need to better understand the lower bound on the number of measurements under which accurate channel estimates can be obtained. This work bridges that knowledge gap by deriving a tight asymptotic lower bound on the number of measurements. This not only helps develop a better understanding for the sparse MIMO channel estimation problem, but it also provides a benchmark for evaluating current and future solutions.**

I. Introduction

Through the use of a large number of antennas, wireless transceivers can focus their signal transmission and/or reception through very narrow angular directions [1]. This helps increase the channel capacity in two main ways. First, it improves the spatial multiplexing capability of transceivers, which allows simultaneously serving multiple users while keeping cross interference low. Second, it allows more signal power to be propagated from a transmitter (TX) to a receiver (RX). For the latter reason, large MIMO transceivers have emerged as the prominent solution to solve the severe path loss problem in millimeter-wave (mmWave) systems [2], [3].

The main challenge of large MIMO, however, is that the channel estimation process can be complex [4] since channel matrices have large dimensions. This problem is further exacerbated by the practically-viable transceiver designs used to overcome the cost and power consumption problems attributed with the traditional fully-digital transceiver architectures.

Reducing the number of channel measurements is thus one of the main challenges facing large MIMO implementations. This problem has largely been tackled as an application of **Compressed Sensing** (CS) [5], [6], which relies on channel *sparsity* as a key enabler for reducing the number of measurements[1]. The closest effort to understanding how changing the number of measurements affects the quality of channel estimates, to the best of our knowledge, is [7], where computer simulations were conducted to measure the quality of channel estimates as the number of measurements increases. Nonetheless, there is still a gap in the current literature in understanding the *lower bound* on the number of necessary measurements needed for accurate channel recovery. To the best of our knowledge, the tightest known bound scales as $\Omega\left(k \log \frac{n_t n_r}{k}\right)$ [8], [9], where k is the channel sparsity level and n_t and n_r are the numbers of antennas at TX and RX, respectively. This bound, however, is a naive application of the CS bound for recovery of sparse vectors of length $n = n_t n_r$ and k non-zero values. In fact, the nature of the channel estimation problem poses limitations on how measurements are obtained, as opposed to the standard CS problem. Thus, more attention needs to be paid when deriving measurement lower bounds. In this paper, we show that the aforementioned bound is too loose, and we provide a tighter lower bound which has order of $\Omega\left(k^2 \log\left(\frac{n_t}{k}\right) \log\left(\frac{n_r}{k}\right)\right)$. We argue the tightness of this bound by showing that, under a mild constraint on the channel sparsity level, there exists a solution with a number of measurements upper bounded as $O(k^2 \log\left(\frac{n_t}{k}\right) \log\left(\frac{n_r}{k}\right))$.

Notations: Let x be a scalar quantity, $\boldsymbol{x}$ be a vector and $\boldsymbol{X}$ be a matrix. The Conjugate, Transpose and Hermition of $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^*$, $\boldsymbol{X}^T$ and $\boldsymbol{X}^H$, respectively. The p^{th} norm is denoted by $\|\boldsymbol{x}\|_p$ (if the subscript p is dropped, then assume p=2), and the ℓ_0 norm $\|\boldsymbol{x}\|_0$ is the number of non-zero elements of $\boldsymbol{x}$. Denote by $\text{vec}\left(\boldsymbol{X}\right)$ the vectorization of columns of $\boldsymbol{X}$, and denote by $\otimes$ the Kronecker product. Finally, we use: (i) $\Omega\left(\cdot\right)$ to denote the Big Omega notation, i.e., the asymptotic lower bound[2], (ii) $O\left(\cdot\right)$ to denote the Big O notation, i.e., the asymptotic upper bound[3], and (iii) we say that $f(n){\in}\Theta\left(g(n)\right)$ if both $f(n){\in}\Omega(g(n))$ and $f(n) \in O(g(n))$.

II. System Model

Consider a single-tap, block-fading, ***sparse*** MIMO channel between a TX and RX equipped with n_t and n_r antennas, respectively. Antennas at TX and RX form Uniform Linear Arrays (ULA), with *normalized* antenna spacing of Δ_t and Δ_r, respectively. The normalization is with respect to the carrier wavelength, denoted by λ_c. We consider *analog* transceiver architectures at both TX and RX. That is, only one RF chain exists per transceiver, and all antennas are connected to this RF chain through phase-shifters and variable-gain amplifiers.

Let the maximum number of *resolvable* signal propagation paths in the channel be denoted by k. Recall that we consider sparse channels. By the sparsity assumption [4], [5], [10]–[13], only a few signal propagation paths exist, where $k \ll n_t, n_r$. Note that a wireless transceiver may not be able to resolve multiple channel paths if they are spatially close. However, as the number of antennas increases, the transceiver's ability to resolve more paths also increases due to its ability to form narrower antenna beams. This means that k increases with n. However, the ratio $\frac{k}{n}$ decreases as n increases. We assume that $n_t, n_r \geq k^{1+\epsilon}$, for some $\epsilon{>}0$, which reflects the ability of transceivers to resolve more channel paths as their number of antennas increases. For each propagation path p, let α_p be its path-gain, θ_p be its Angle of Departure (AoD) at TX, ϕ_p be its

[1]Sparsity here means that the number of signal propagation paths is small compared to the number of TX and RX antennas (e.g., mmWave channels).

[2]We say that $f(n) \in \Omega\left(g(n)\right)$ (or loosely, $f(n){=}\Omega\left(g(n)\right)$) if there exists a constant $c > 0$, and $n_0 \in \mathbb{N}$ such that $f(n) \geq cg(n)$, for all $n{\geq}n_0$.

[3]We say that $f(n) \in O\left(g(n)\right)$ (or loosely $f(n){=}O\left(g(n)\right)$) if there exists a constant $c > 0$ and $n_0 \in \mathbb{N}$ such that $f(n) \leq cg(n)$, for all $n{\geq}n_0$.

Angle of Arrival (AoA) at RX, and ρ_p be its path length. The baseband path gain, α_p^b, is given by $\alpha_p^b = \alpha_p\sqrt{n_t n_r}\exp^{-j\frac{2\pi\rho_p}{\lambda_c}}$.

Let $\boldsymbol{Q} \in \mathbb{C}^{n_r\times n_t}$ denote the channel matrix, where $q_{i,j}$, the element at row i and column j in $\boldsymbol{Q}$, is the channel gain between the j^{th} TX antenna and the i^{th} RX antenna. Let us denote the path-loss by μ. Then, we can write $\boldsymbol{Q}$ as

$$\boldsymbol{Q} = \sum_{p=1}^{k} \frac{\alpha_p^b}{\mu} \boldsymbol{e_r}(\omega_{r,p})\boldsymbol{e_t}^H(\omega_{t,p}), \tag{1}$$

where $\boldsymbol{e_t}(\omega)$ and $\boldsymbol{e_r}(\omega)$ are the transmit and receive signal spatial signatures, at angular cosine ω [1, Chapter 7]. The channel $\boldsymbol{Q}$, in this form, is not sparse. However, it can be represented in a sparse form using a simple change of basis:

$$\boldsymbol{Q^a} = \boldsymbol{U_r}^H\boldsymbol{Q}\boldsymbol{U_t}, \tag{2}$$

where $\boldsymbol{Q^a}$ is known as the "*angular channel*" and is sparse. The matrices $\boldsymbol{U_t}$ and $\boldsymbol{U_r}$ are Discrete Fourier Transform matrices whose columns represent an orthonormal basis for the transmit and receive signal spaces, and are defined as:

$$\boldsymbol{U_i} = \left(\boldsymbol{e_i}(0)\ \boldsymbol{e_i}\left(\tfrac{1}{L_i}\right)\ \boldsymbol{e_i}\left(\tfrac{2}{L_i}\right)\ \dots\ \boldsymbol{e_i}\left(\tfrac{n_i-1}{L_i}\right)\right), \quad i \in \{t, r\},$$

where L_t/L_r are the normalized lengths of the TX/RX ULAs.

When transmitting a symbol ζ, the TX uses a precoder vector $\boldsymbol{f} \in \mathbb{C}^{n_t}$ while RX uses a combiner vector $\boldsymbol{w} \in \mathbb{C}^{n_r}$. The received symbol at RX is thus given by:

$$y_{i,j} = \boldsymbol{w}_i^H\boldsymbol{Q}\boldsymbol{f}_j\zeta + \boldsymbol{w}_i^H\boldsymbol{n_{i,j}}, \tag{3}$$

where $y_{i,j}$ denotes the received symbol (i.e., measurement result), $\boldsymbol{w}_i$ denotes the i^{th} receive combiner and $\boldsymbol{f}_j$, the j^{th} transmit precoder. Assume, for simplicity, that $\zeta = 1$. Let the number of rx-combiners be m_r and the number of tx-precoders be m_t. Then, the total number of measurements we can obtain using all combinations of $\boldsymbol{f}_j$ and $\boldsymbol{w}_i$ is $m = m_t \times m_r$. We can also write the measurement equations for all precoders and combiners more compactly as:

$$\boldsymbol{Y} = \boldsymbol{W}^H\boldsymbol{Q}\boldsymbol{F} + \boldsymbol{N}, \tag{4}$$

where $y_{i,j}$ is the element at row i and column j of $\boldsymbol{Y}$. $\boldsymbol{W}$ and $\boldsymbol{F}$ are defined as:

$$\boldsymbol{W} \triangleq \left(\boldsymbol{w}_1\ \boldsymbol{w}_2\ \dots\ \boldsymbol{w}_{m_r}\right), \quad \boldsymbol{F} \triangleq \left(\boldsymbol{f}_1\ \boldsymbol{f}_2\ \cdots\ \boldsymbol{f}_{m_t}\right) \tag{5}$$

The channel estimation problem, i.e., figuring out what the matrix $\boldsymbol{Q}$ is, can be broken down into determining the best set of precoders $\boldsymbol{f}_j$ and combiners $\boldsymbol{w}_i$ using which we can accurately recover $\boldsymbol{Q}$. To speed up the estimation process, the smallest sets of those $\boldsymbol{f}_j$'s and $\boldsymbol{w}_i$'s should be used. In this paper, we do not provide a specific design for such precoders and combiners, but we seek to find a "tight" lower bound on the number of measurements using which $\boldsymbol{Q}$ can be recovered.

Special Cases: Consider the special cases in which either (i) n_t=1 or (ii) n_r=1. In the former case, the channel is Single-Input-Multiple-Output (SIMO), while in the latter it is Multiple-Input-Single-Output (MISO). In both cases, the channel becomes a vector $\boldsymbol{q}$. Tx-precoders in MISO fall back to just a scalar quantity; $f = 1$, while in SIMO rx-combiners fall back to $w = 1$. Thus, we can rewrite the measurement equation (Eq. (4)) for SIMO and MISO, respectively, as:

$$\boldsymbol{y} = \boldsymbol{W}^H\boldsymbol{q} + \boldsymbol{n}, \qquad \boldsymbol{y} = \boldsymbol{F}^H\boldsymbol{q} + \boldsymbol{n}. \tag{6}$$

III. Problem Formulation

In this section, we will provide a brief overview of compressed sensing (CS). Then, we will formulate the problem of channel estimation as a CS problem. To that end, we will reshape the measurement equation given in Eq. (4) to be in the form $\boldsymbol{y_v}$=$\boldsymbol{G_v}\boldsymbol{q_v^a}$+$\boldsymbol{n_v}$, which conforms with the traditional compressed sensing problem, as will be shown in Eq. (7) below. Here, $\boldsymbol{q_v^a}$ is sparse and has dimensions $n_r n_t \times 1$.

A. Compressed Sensing Background

Compressed sensing is a signal processing technique [6] that allows the reconstruction of a signal $\boldsymbol{x} = (x_i)_{i=1}^n$ from a small number of samples *given that* $\boldsymbol{x}$ is either: (i) sparse, or (ii) can be represented in a sparse form, using a linear transformation $\boldsymbol{U}$ such that $\boldsymbol{x} = \boldsymbol{U}\boldsymbol{s}$ where $\boldsymbol{s}$ is sparse. Let the number of measurements be denoted by m where $m < n$ and $m, n \in \mathbb{N}$. Each measurement of $\boldsymbol{x}$ is a linear combination of its components x_i. Such measurements are dictated by the sensing matrix $\boldsymbol{G}$ and are given by $\boldsymbol{y} = \boldsymbol{G}\boldsymbol{x}$, where $\boldsymbol{y}$ denotes the $m\times 1$ measurement vector. The matrix equation $\boldsymbol{y} = \boldsymbol{G}\boldsymbol{x}$ represents an under-determined system of linear equations (since $m < n$). In other words, we have fewer equations than the number of unknowns we want to solve for. While, in general, an infinite number of solutions exist, the sparsity of $\boldsymbol{x}$ allows for perfect signal reconstruction from $\boldsymbol{y}$ *given that* certain conditions are satisfied, among which, is a lower bound on the "*spark*" of the sensing matrix.

Definition III.1. The spark of a given matrix $\boldsymbol{G}$ is the smallest number of its linearly dependent columns.

Theorem 1 (Corollary 1 of [14]). For any vector $\boldsymbol{y} \in \mathbb{R}^m$, there exits at most one vector $\boldsymbol{q^a} \in \mathbb{R}^n$ with $\|\boldsymbol{q^a}\|_0 = k$ such that $\boldsymbol{y} = \boldsymbol{G}\boldsymbol{q^a}$ if and only if $\text{spark}(\boldsymbol{G}) > 2k$.

Theorem 1 provides a mathematical guarantee on the exact recovery of $k-$sparse vectors using m linear measurements. An immediate bound on the number of measurements, m, we get from Theorem 1 is $m \geq 2k$. The lower bound on the spark of $\boldsymbol{G}$ works well under noise-free measurements, but in practice, measurements get corrupted with a vector $\boldsymbol{n}$, i.e.,

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{x} + \boldsymbol{n}. \tag{7}$$

It is necessary to guarantee that the measurement process is not adversely affected by such errors in a significant way. This calls for alternative, stricter requirements on sensing matrices to guarantee "good" sparse recovery. Mathematically, we need to design the sensing matrix such that the energy in the measured signal is preserved. This is quantified using the **Restricted Isometry Property (RIP)**. The RIP property guarantees that the distance between any pair of $k-$sparse vectors is not significantly changed under the measurement process. This RIP property is defined as follows:

Definition III.2. A matrix $\boldsymbol{G}$ satisfies the restricted isometry property (RIP) of order k if there exists a constant $\delta_k \in (0, 1)$ such that for all vectors $\boldsymbol{q^a}$, with $\|\boldsymbol{q^a}\|_0 \leq k$, we have

$$(1-\delta_k)\|\boldsymbol{q^a}\|_2^2 \leq \|\boldsymbol{G}\boldsymbol{q^a}\|_2^2 \leq (1+\delta_k)\|\boldsymbol{q^a}\|_2^2. \tag{8}$$

The smallest δ_k which satisfies Eq. (8) is called the "$k-$restricted isometry constant". Note that in general, a matrix $\tilde{\boldsymbol{G}}$ does not necessarily result in $||\tilde{\boldsymbol{G}}\boldsymbol{q^a}||^2$ that is symmetric about 1. However, a simple scaling of $\tilde{\boldsymbol{G}}$ results in $\boldsymbol{G}$ such that the tightest bounds of $\|\boldsymbol{G}\boldsymbol{q}^{\boldsymbol{a}}\|^2$ in Eq. (8) are symmetric [15]. From now on, we will only consider matrices whose bounds are symmetric as shown in Eq. (8).

The following theorem provides a necessary condition for $m \times n$ matrices that satisfy the RIP property with $\delta_k \in (0,1)$.

Theorem 2 (Theorem 3.5 of [16]). Let $\boldsymbol{G}$ be an $m \times n$ matrix that satisfies RIP of order k with constant $\delta_k \in (0,1)$. Then, m satisfies

$$m \geq c_\delta k \log\left(\frac{n}{k}\right), \tag{9}$$

where $c_\delta = \frac{0.18}{\log\left(\sqrt{\frac{1+\delta}{1-\delta}}+1\right)}$, is a function of δ only.

Theorem 2 demonstrates the popular asymptotic measurement bound: $m = \Omega\left(k \log \frac{n}{k}\right)$. Next, we will formulate the MIMO channel estimation as a compressed sensing problem.

B. The Problem

Recall from Eq. (4) that channel measurements take the form $\boldsymbol{Y} = \boldsymbol{W}^H \boldsymbol{Q} \boldsymbol{F} + \boldsymbol{N}$. This is not the standard form of a noisy CS problem (see Eq. (7)). Thus, it cannot readily be solved using compressed sensing. To put this equation in a CS problem form, let us "*vectorize*" its left and right hand sides as follows: (i) let $\boldsymbol{y_v} \triangleq \mathrm{vec}\,(\boldsymbol{Y})$, (ii) $\boldsymbol{n_v} \triangleq \mathrm{vec}\,(\boldsymbol{N})$, (iii) $\boldsymbol{q_v^a} \triangleq \mathrm{vec}\,(\boldsymbol{Q^a})$, and finally (iv) by properties of *vectorization* [17], we have

$$\begin{aligned} \mathrm{vec}(\boldsymbol{W}^H \boldsymbol{Q} \boldsymbol{F}) &= (\boldsymbol{F}^T \otimes \boldsymbol{W}^H)\,\mathrm{vec}(\boldsymbol{Q}) && (10)\\ &= (\boldsymbol{F}^T \otimes \boldsymbol{W}^H)\,(\boldsymbol{U_t^*} \otimes \boldsymbol{U_r})\,\mathrm{vec}\,(\boldsymbol{Q^a}) && (11)\\ &= \left(\left(\boldsymbol{F}^H \boldsymbol{U_t}\right)^* \otimes \left(\boldsymbol{W}^H \boldsymbol{U_r}\right)\right) \boldsymbol{q_v^a} && (12) \end{aligned}$$

Thus, we can rewrite the measurement equation in (4) as

$$\boldsymbol{y_v} = \boldsymbol{G_v} \boldsymbol{q_v^a} + \boldsymbol{n_v}, \tag{13}$$

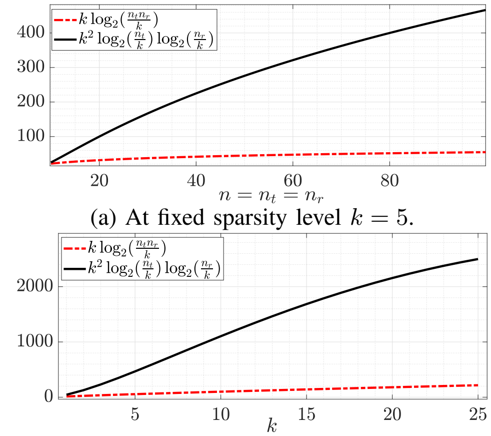
$$\textit{where} \quad \boldsymbol{G_v} = \left(\boldsymbol{F}^H \boldsymbol{U_t}\right)^* \otimes \left(\boldsymbol{W}^H \boldsymbol{U_r}\right) \tag{14}$$

is the sensing matrix, with dimensions $m_t m_r \times n_t n_r$, while $\boldsymbol{y_v}$ has dimensions $m_t m_r \times 1$ and $\boldsymbol{q_v^a}$ has dimensions $n_t n_r \times 1$. This form of the problem allows us to employ CS sparse recovery techniques to estimate $\boldsymbol{q_v^a}$ from $\boldsymbol{y_v}$.

IV. Lower Measurement Bound

We are interested in sensing matrices that preserve the ***distance*** between two different channels $\boldsymbol{q_{v1}^a}$ and $\boldsymbol{q_{v2}^a}$. This distance is the norm of $\boldsymbol{q_{v1}^a} - \boldsymbol{q_{v2}^a}$, which has a sparsity level of $2k$ (recall that the maximum number of channel paths is k). Thus, to be able to accurately estimate $\boldsymbol{q_v^a}$, we need the sensing matrix $\boldsymbol{G_v}$ to satisfy the RIP property of order $2k$ with some RIP constant $\delta_{2k} \in (0,1)$. At sparsity level of $2k$, Theorem 2 shows that the recovery of a sparse vector with dimensions $n = n_t n_r$ requires a number of measurements, m, lower bounded as $m \geq c_\delta (2k) \log\left(\frac{n_t n_r}{(2k)}\right) = 2 c_\delta k \left(\log\left(\frac{n_t}{\sqrt{2k}}\right) + \log\left(\frac{n_r}{\sqrt{2k}}\right)\right)$. This demonstrates the popular $m = \Omega\left(k \log\left(\frac{n_r \times n_t}{k}\right)\right)$ lower bound for sparse channel estimation. Although this bound is valid, it is in fact too loose since it assumes that arbitrary constructions of $\boldsymbol{G}_v$ are possible. This, however, is not the case for sparse MIMO channel estimation since $\boldsymbol{G}_v$ takes a special, Kronecker product form, as derived in Eq. (14).

Next, we will derive a tighter bound on the number of measurements. A bound that considers the special structure of the sensing matrix. This will result in $m = \Omega\left(k^2 \log\left(\frac{n_t}{k}\right) \log\left(\frac{n_r}{k}\right)\right)$. To appreciate how much tighter our



(a) At fixed sparsity level $k = 5$.

(b) At fixed number of antennas $n = n_t = n_r = 100$.

Fig. 1: Unscaled asymptotic measurement lower bounds.

derived bound is, we plot the functions $k \log\left(\frac{n_t \times n_r}{k}\right)$ and $k^2 \log\left(\frac{n_t}{k}\right) \log\left(\frac{n_r}{k}\right)$ without constant scaling in Fig. 1.

A. Main Results: A "Tight" Measurement Bound

In this section, we will derive the relationship between $k-$RIP constants of Kronecker product matrices and those of the blocks that form it. Then, using Theorem 2, we will derive an asymptotic lower bound on the number of rows of $\boldsymbol{G_v}$ and deduce its asymptotic behavior. We will finally show the tightness of our derived asymptotic bound using the solution framework in [18].

Optimum Measurement Length: Among all possible matrices which satisfy the RIP property, we are interested in the ones that have the least number of rows (since the number of rows equals the number of measurements). This leads to the notion of "*Optimum Measurement Length* (OML)". *We define OML as the smallest number of measurements such that the RIP property is satisfied.* OML is dependent on the length of unknown vectors n, the maximum sparsity level k and the $k-$RIP constant δ. Hence, we can define a function μ,

$$\mu : \mathcal{N} \times \mathcal{K} \times (0,1) \rightarrow \mathbb{N}_0^+ \tag{15}$$

which maps the space of all possible values for n, k, and δ, given by[4] $\mathcal{N} \subseteq \mathbb{N}_0^+$, $\mathcal{K} \subseteq \mathbb{N}_0^+$ and $(0,1)$, respectively, to the corresponding OML quantity.

Now, let us focus on the special case of matrices which can be *arbitrarily* constructed. In such case, let μ be denoted by μ_a ('a' stands for 'arbitrary' matrix construction). We define μ_a to be the solution of the following optimization problem:

$$\begin{aligned} P1: \quad & \underset{\boldsymbol{M_a} \in \mathbb{C}^{m_a \times n}}{\text{minimize}} \quad m_a && (16a)\\ & \text{subject to} \quad \boldsymbol{M_a} \in \mathcal{F}_\delta^{(k)} && (16b) \end{aligned}$$

where $\mathcal{F}_\delta^{(k)}$ is the feasible set, and it is defined as

$$\mathcal{F}_\delta^{(k)} \triangleq \{\boldsymbol{M_a} \in \mathbb{C}^{m_a \times n} : (1-\delta)\,\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{M_a}\boldsymbol{x}\|_2^2 \leq (1+\delta)\,\|\boldsymbol{x}\|_2^2, \; \forall \boldsymbol{x} \in \mathbb{C}^n : \|\boldsymbol{x}\|_0 \leq k\}$$

Lemma 3. Let n and k be fixed. Then, $\delta_1 \geq \delta_2$ implies $\mu_a(n,k,\delta_1) \leq \mu_a(n,k,\delta_2)$.

Proof. The proof directly follows by observing that $\delta_1 \geq \delta_2$ implies that $\mathcal{F}_{\delta_2}^{(k)} \subseteq \mathcal{F}_{\delta_1}^{(k)}$. Since the problem is a minimization problem, then $\mu_a(n,k,\delta_1) \leq \mu_a(n,k,\delta_2)$. □

[4]We define $\mathbb{N}_0^+$ to be the set of non-negative integers.

Kronecker Product Matrices: The standard compressed sensing problem assumes that all elements of the sensing matrix are independently chosen. On the contrary, in sparse channel estimation, we are restricted to a specific sensing matrix structure, as shown in Eq. (14). The only free parameters in this sensing matrix are the tx-precoders $\boldsymbol{f_j}$ and the rx-combiners $\boldsymbol{w_i}$. This limitation suggests that more measurements may be needed to achieve the same RIP constant, compared to matrices whose elements are independently selected.

At the heart of our results lies the relationship between the $k-$RIP constant of Kronecker product matrices and the $k-$RIP constants of the matrices that form them. We formally state this relationship in the following lemma.

Lemma 4 (RIP of Kronecker Products). Let δ_a and δ_b be the $k-$RIP constants of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Then, the $k-$RIP constant of $\boldsymbol{A}\otimes\boldsymbol{B}$, denoted by δ, is bounded as

$$\delta \geq \max\{\delta_a, \delta_b\} \tag{17}$$

A similar result to Lemma 4 was derived in [19], but under the stronger assumption of matrices with normalized columns. Our more general result implies that even if the normalized columns assumption is loosened, we still cannot obtain a matrix, through a Kronecker Product, which satisfies the RIP property with a constant smaller than the maximum of the $k-$RIP constants of the matrices that form it. To prove Lemma 4, we define two matrices: $\boldsymbol{C}=\boldsymbol{A}\otimes\boldsymbol{B}$ and $\boldsymbol{C'}=\boldsymbol{B}\otimes\boldsymbol{A}$, whose $k-$RIP constants are δ_c and $\delta_{c'}$, respectively. We will then show that: (1) $\delta_c\geq\delta_b$, (2) $\delta_{c'}\geq\delta_a$, and that (3) $\delta_c=\delta_{c'}$, by which Lemma 4 follows. A proof outline is provided next, while the detailed proof is provided in [20]. To show part (1), let the number of columns of $\boldsymbol{A}$ and $\boldsymbol{B}$ be n_a and n_b, respectively. And define $\mathcal{X}_c$ and $\mathcal{X}_b$ to be the sets of vectors with sparsity levels $\leq k$ and whose lengths are $n_a n_b$ and n_b, respectively. Since $\boldsymbol{C}$ and $\boldsymbol{B}$ satisfy $k-$RIP with δ_c and δ_b. Then, $\|\boldsymbol{C}\boldsymbol{x_c}\|^2$ is bounded between $(1\pm\delta_c)\|\boldsymbol{x_c}\|^2$ for all $\boldsymbol{x_c}\in\mathcal{X}_c$. Similarly, $\|\boldsymbol{B}\boldsymbol{b}\|^2$ is bounded between $(1\pm\delta_b)\|\boldsymbol{b}\|^2$. View $\boldsymbol{x_c}$, whose length is $n_a n_b$, as n_a blocks of length n_b each, and focus on a strict subset of $\mathcal{X}_c$, call it $\mathcal{X}_c^{(b)}$, with vectors defined as $\boldsymbol{x_c^{(b)}}=\left(\boldsymbol{b}^T\ \boldsymbol{0}^T\ \ldots\ \boldsymbol{0}^T\right)^T$ for all $\boldsymbol{b}\in\mathcal{X}_b$. Now examine that $\|\boldsymbol{C}\boldsymbol{x_c^{(b)}}\|^2=\|\boldsymbol{a_1}\|^2\|\boldsymbol{B}\boldsymbol{b}\|^2$, where $\boldsymbol{a_1}$ is the first columns of $\boldsymbol{A}$, which leads to these two bounds:

$$(1-\delta_c)\|\boldsymbol{b}\|^2\leq\|\boldsymbol{a_1}\|^2(1-\delta_b)\|\boldsymbol{b}\|^2\leq\|\boldsymbol{a_1}\|^2\|\boldsymbol{B}\boldsymbol{b}\|^2 \tag{18}$$

$$\|\boldsymbol{a_1}\|^2\|\boldsymbol{B}\boldsymbol{b}\|^2\leq\|\boldsymbol{a_1}\|^2(1+\delta_b)\|\boldsymbol{b}\|^2\leq(1+\delta_c)\|\boldsymbol{b}\|^2 \tag{19}$$

If $\|\boldsymbol{a_1}\|\leq 1$, we can use Eq. (18) to conclude that $\delta_c\geq\delta_b$, otherwise, use Eq. (19) to arrive at the same conclusion. This concludes part (1), and by analogy part (2) follows. Finally, since there exists Permutation matrices $\boldsymbol{P_\rho}$ and $\boldsymbol{P_c}$, such that $\boldsymbol{C'}=\boldsymbol{P_\rho}\boldsymbol{C}\boldsymbol{P_c}$. This leads to $\|\boldsymbol{C'}\boldsymbol{x'_c}\|=\|\boldsymbol{C}\boldsymbol{x_c}\|$, where $\boldsymbol{x'_c}=\boldsymbol{P_c}\boldsymbol{x_c}$ and $\boldsymbol{x_c}\in\mathcal{X}_c$ if and only if $\boldsymbol{x'_c}\in\mathcal{X}_c$. Hence $\delta_c=\delta_{c'}$.

A Generalized Bound: Recall Eq. (14). We will rewrite $\boldsymbol{G_v}$, for brevity, in terms of $\boldsymbol{M_t}$ and $\boldsymbol{M_r}$, where

$$\boldsymbol{M_t}\triangleq\left(\boldsymbol{F}^H\boldsymbol{U_t}\right)^*\in\mathbb{C}^{m_t\times n_t},\quad \boldsymbol{M_r}\triangleq\boldsymbol{W}^H\boldsymbol{U_r}\in\mathbb{C}^{m_r\times n_r}$$

Thus, we have $\boldsymbol{G_v}{=}\boldsymbol{M_t}\otimes\boldsymbol{M_r}$, and $m{=}m_t m_r$ is the number of rows of $\boldsymbol{G_v}$. Now, suppose that $\boldsymbol{G_v}$ satisfies $k-$RIP with constant $\delta\in(0,1)$. Then, both $\boldsymbol{M_t}$ and $\boldsymbol{M_r}$ must satisfy the $k-$RIP with constants $\delta_t\in(0,1)$ and $\delta_r\in(0,1)$, respectively. To show that this is true, assume, without loss of generality (w.l.o.g.), that there does not exist $\delta_t\in(0,1)$ such that $\boldsymbol{M_t}$ satisfies $k-$RIP. Then, there exists a vector $\boldsymbol{v}$ with $\|\boldsymbol{v}\|_0\leq k$ such that $\boldsymbol{M_t}\boldsymbol{v}=\boldsymbol{0}$, which implies the existence of at least k dependent columns of $\boldsymbol{M_t}$, call them $\boldsymbol{a_{t1}},\boldsymbol{a_{t2}},\ldots,\boldsymbol{a_{tk}}$. In turn, there exists at least k dependent columns in $\boldsymbol{G_v}$ (let $\boldsymbol{a_{r1}}$ be a column in $\boldsymbol{M_r}$, then the columns $\boldsymbol{a_{t1}}\otimes\boldsymbol{a_{r1}},\boldsymbol{a_{t2}}\otimes\boldsymbol{a_{r1}},\ldots,\boldsymbol{a_{tk}}\otimes\boldsymbol{a_{r1}}$ are dependent). Hence, $\nexists\delta\in(0,1)$ such that $\boldsymbol{G_v}$ satisfies $k-$RIP with a constant δ. Thus, we arrive at a contradiction. Further, by Lemma 4, we have that $\delta\geq\max\{\delta_t,\delta_r\}$.

Since $\boldsymbol{M_t}$ and $\boldsymbol{M_r}$ can be arbitrarily constructed, then we can lower bound m_t and m_r by their OML values as follows

$$m_t\geq\mu_a(n_t,k,\delta_t)\overset{(i)}{\geq}\mu_a(n_t,k,\delta) \tag{20}$$

$$m_t\geq\mu_a(n_r,k,\delta_r)\overset{(ii)}{\geq}\mu_a(n_r,k,\delta) \tag{21}$$

where inequalities (i) and (ii) follow from Lemma 3. Thus, it follows that the number of rows of $\boldsymbol{G_v}$, m, is bounded as

$$m\geq\mu_a(n_t,k,\delta)\times\mu_a(n_r,k,\delta). \tag{22}$$

Recall that $\mu_a(\cdot)$ is the value that solves problem P1.

Remark. The implication of Inequality (22) is that the number of measurements needed for estimating a sparse MIMO channel, $\boldsymbol{Q}$, is at least equal to (but possibly higher) than the product of the number of measurements needed to solve the following two sub-problems: The first is a SIMO, $1\times n_r$ channel, with $\boldsymbol{M_r}$ as sensing matrix. The second is a MISO, $n_t\times 1$ channel, with $\boldsymbol{M_t}^*$ as sensing matrix, where the sparsity level of both channels is $\leq k$. These two sub-problems are special cases of the original problem, whose measurement equations are shown in Eq. (6). The only difference is the conjugation of $\boldsymbol{M_t}$.

The bound we derive in Eq. (22) highlights the dependence on the channel dimensions n_t and n_r, the maximum sparsity level k and a measure, δ, of how much information the measurements preserve about the channel. This bound, however, is not explicit, but we can use Theorem 2 to derive a more concrete lower bound for $\mu_a(\cdot)$. This leads to our main result:

Theorem 5 (Main Theorem). Fix $\delta\in(0,1)$. If $\boldsymbol{G}_v$ in Eq. (14) satisfies RIP with order $2k$ and constant δ, then the number of measurements m is asymptotically bounded as:

$$\boxed{m=\Omega\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right) \qquad (23)}$$

Proof. Since $\mu_a(n_t,2k,\delta)$ and $\mu_a(n_r,2k,\delta)$ are obtained by solving the problem $P1$ (with their respective n_t, n_r and δ values), then there exists matrices $\boldsymbol{X_t}$ and $\boldsymbol{X_r}$, with dimensions $\mu_a(n_t,2k,\delta)\times n_t$ and $\mu_a(n_r,2k,\delta)\times n_r$ which satisfy $2k-$RIP with constant δ. Thus, it follows by Theorem 2 that:

$$\mu_a(n_t,2k,\delta)\geq c_\delta 2k\log\left(\frac{n_t}{2k}\right),\quad \mu_a(n_r,2k,\delta)\geq c_\delta 2k\log\left(\frac{n_r}{2k}\right)$$

Therefore, by Eq. (22), the following follows

$$m=m_t m_r\geq 4c_\delta^2k^2\log\left(\frac{n_t}{2k}\right)\log\left(\frac{n_r}{2k}\right) \tag{24}$$

Finally, let $c=0.5$ and recall that the ratio $\frac{n_t}{k}$ increases (by assumption). Then, there exists $n_{t0}\in\mathbb{N}$ such that $\log(\frac{n_t}{2k})\geq c\log(\frac{n_t}{k})$ for all $n_t\geq n_{t0}$. Similarly, there exists $n_{r0}\in\mathbb{N}$ such that $\log(\frac{n_r}{2k})\geq c\log(\frac{n_r}{k})$ for all $n_r\geq n_{r0}$. Then, it follows that $m\geq 4c^2c_\delta^2k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)$ where $4c^2c_\delta^2=c_\delta^2$ is a constant, from which Eq. (23) follows. □

B. Tightness of the Measurement Bound

To argue that the measurement lower bound in Theorem 5 is tight, we will show that there exists a solution, based on [18], which yields sensing matrices that satisfy $2k-$RIP with constants $\in(0,1)$ and with $m\in\Theta\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$. We briefly discuss the measurement framework of [18] next.

In [18], a source-coding-based framework for the sparse MIMO channel estimation problem is developed. This solution proposes a method for obtaining a small number of measurements that are sufficient to estimate the channel. Such measurements are designed based on two carefully chosen binary linear source codes, C_t and C_r. These codes dictate the design of tx-precoders (using C_t) and rx-combiners (using C_r) and produce real-valued measurement (sensing) matrices, namely, $\boldsymbol{H_t}$ (of size $m_t\times n_t$) and $\boldsymbol{H_r}$ (of size $m_r\times n_r$), respectively. The matrix $\boldsymbol{H_t}$ can estimate $k-$sparse MISO channel vectors (i.e., produces unique measurements), while $\boldsymbol{H_r}$ can estimate $k-$sparse SIMO channels. Hence, the spark of both matrices is greater than $2k$ (by Theorem 1). Measurements are then obtained using all combinations of m_t tx-precoders and m_r rx-combiners, and can be arranged as $\boldsymbol{y_v}=\boldsymbol{H_v}\boldsymbol{q_v^a}+\boldsymbol{n_v}$ where $\boldsymbol{H_v}=\boldsymbol{H_t}\otimes\boldsymbol{H_r}$. Then, it follows that $\operatorname{spark}(\boldsymbol{H_v})>2k$. This is shown in detail in our technical report [20]. Hence, either $\boldsymbol{H_v}$ or a scaled version of it satisfies $2k-$RIP with a constant $\delta_h\in(0,1)$. This measurement framework is shown to produce a number of measurements, m, that is lower bounded as:

$$m\geq\underline{\mathrm{m}}\triangleq\underbrace{\left\lceil\log_2\left(\sum_{i=0}^{k}\binom{n_r}{i}\right)\right\rceil}_{\leq m_t}\underbrace{\left\lceil\log_2\left(\sum_{i=0}^{k}\binom{n_t}{i}\right)\right\rceil}_{\leq m_r}. \tag{25}$$

This lower bound is achievable with equality for specific examples as shown in [18]. However, it is not immediately clear how this bound compares to our bound in Eq. (23). The following lemma sheds more light on this issue:

Lemma 6. The asymptotic behavior of $\underline{\mathrm{m}}$, defined in Eq. (25) follows: $\underline{\mathrm{m}}=\Theta\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$.

This is the same asymptotic behavior as the lower bound in Theorem 5. To prove this lemma, we use the following bound on $\binom{n}{k}$ [21]: $\left(\frac{n}{k}\right)^k\leq\binom{n}{k}\leq\left(\frac{ne}{k}\right)^k$, and observe that for $k<\frac{n+1}{2}$, we have $\binom{n}{k}\leq\sum_{i=0}^{k}\binom{n_r}{i}\leq(k+1)\binom{n}{k}$. This allows us to establish upper and lower bounds for $\sum_{i=0}^{k}\binom{n}{i}$ using the aforementioned bounds for $\binom{n}{k}$, which leads to showing that $\log\left(\sum_{i=0}^{k}\binom{n}{i}\right)=\Theta\left(k\log\left(\frac{n}{k}\right)\right)$. Note that $\lceil\cdot\rceil$ does not change the asymptotic behavior of its argument. Therefore, we are able to show that $\underline{\mathrm{m}}=\Theta\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$. A rigorous proof of Lemma 6 is provided in our technical report [20]. Next, we will examine a specific solution based on the family of BCH codes, which results in a number of measurements upper bounded as $m=O\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$.

Example 1 (BCH codes). Although BCH codes are natively error-correcting codes, they can be used as syndrome-source-codes, as well[5]. By the properties of BCH codes, we have that for any positive integers $t\geq3$ and $k<2^{t-1}$, there exists a binary BCH code with: i) block length $n=2^t-1$, ii) minimum distance $d_{\min}\geq2k+1$ (hence, it can correct up to k errors), and iii) a number of parity check bits $m\leq tk=k\log_2(n+1)$. Using BCH codes to design C_t and C_r, we obtain a solution whose number of measurements is upper bounded according to the following lemma:

[5]A linear block error-correcting code (LBC) can be utilized as a syndrome source code which can uniquely compress sequences that contain a number of 1's less than or equal to the number of correctable errors of the used code [22]. The parity check matrix of the LBC code is used as the generator matrix for the source code. Hence, the number of parity bits of the LBC code is the length of the compressed sequences for the corresponding source code.

Lemma 7. The number of measurements achievable using BCH codes in the framework of [18] is asymptotically bounded as $m=O\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$.

Proof Sketch. For arbitrary values of $n_t{\geq}7$, there exists t such that $n_t{\leq}2^{t+1}{\triangleq}n_t'$. Then, for all $k{<}\frac{n_t'+1}{2}$, there exists a BCH code with block length n_t' and parity length of $m_t'{=}O(k\log n_t')$. We can then shorten that BCH code by removing $n_t'{-}n_t$ information bits from its codewords while keeping $m_t{=}m_t'$ unchanged. We can then show that $m_t{=}m_t'{=}O(k\log n_t)$, since $\log2^{t+1}{\leq}\frac{4}{3}\log2^t$. Now, recall that $n_t\geq k^{1+\epsilon}$, where $\epsilon>0$ (by assumption). Then, $\frac{1}{\epsilon}\log\frac{n_t}{k}{\geq}\log k$. Therefore, we can show that $\log(n_t){\leq}(1+\frac{1}{\epsilon})\log\left(\frac{n_t}{k}\right)$, by which we have $m_t{=}O(k\log\left(\frac{n_t}{k}\right))$. Similarly, $m_r{=}O(k\log\left(\frac{n_r}{k}\right))$. Thus, $m=O\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$. For a detailed proof, see our technical report [20]. □

Among all solutions in [18], we are interested in the ones whose number of measurements, m, is closest to $\underline{\mathrm{m}}$. These solutions are *"optimum"* in the sense of reducing the number of measurements. Recall that $\underline{\mathrm{m}}$ is the lower bound of all solutions based on [18] (see Eq. (25)). The following theorem shows that these optimum solutions scale similarly to $\underline{\mathrm{m}}$, which in turn shows that the lower bound of Theorem 5 is tight.

Theorem 8. The number of measurements of "Optimum Solutions" of [18] scales as $m=\Theta\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$

Proof. By Lemma 6, we have that all solutions, including the optimal, have $m=\Omega\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$. Moreover, Lemma 7 shows that solutions based on BCH codes result in $m=O\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$. Since optimal solutions have a number of measurements smaller than or equal to those obtained by BCH codes, then they also have the same asymptotic upper bound. Therefore, optimal solutions have $m=\Theta\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$ follows. □

Remark. Even though we have shown that the bound of Theorem 5 is tight, we have demonstrated this tightness in the asymptotic regime of n and k. The dependence on the RIP constant, δ, however, remains an open question.

V. Conclusion

In this paper, we study the fundamental lower bound governing the number of measurements required for estimating sparse, large-MIMO channels. We consider a simple analog transceiver, where each channel measurements is obtained using a specific combination of beamforming vectors at the transmitter and receiver. The currently known lower bound on number of measurements is $\Omega\left(k\log\left(\frac{n_rn_t}{k}\right)\right)$, which we show to be loose. We then derive a tight lower measurement bound, which scales asymptotically as $\Omega\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$. The tightness of our derived bound is demonstrated by showing that there exists a solution with $m=O\left(k^2\log\left(\frac{n_t}{k}\right)\log\left(\frac{n_r}{k}\right)\right)$.

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.

[2] Y. Shabara, C. E. Koksal, and E. Ekici, "Beam Discovery Using Linear Block Codes for Millimeter Wave Communication Networks," *IEEE/ACM Transactions on Networking*, 2019.

[3] D. Fan, F. Gao, Y. Liu, Y. Deng, G. Wang, Z. Zhong, and A. Nallanathan, "Angle Domain Channel Estimation in Hybrid Millimeter Wave Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8165–8179, 2018.

[4] Y. Ding and B. D. Rao, "Dictionary Learning-Based Sparse Channel Representation and Estimation for FDD Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5437–5451, 2018.

[5] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed Sensing for Wireless Communications: Useful Tips and Tricks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1527–1550, 2017.

[6] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] A. Alkhateeb, G. Leusz, and R. W. Heath, "Compressed Sensing Based Multi-User Millimeter Wave Systems: How Many Measurements are Needed?" in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2909–2913.

[8] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed Sensing for Wireless Communications : Useful Tips and Tricks," *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2017.

[9] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[10] Z. Gao, L. Dai, C. Yuen, and Z. Wang, "Asymptotic Orthogonality Analysis of Time-Domain Sparse Massive MIMO Channels," *IEEE Communications Letters*, vol. 19, no. 10, pp. 1826–1829, 2015.

[11] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient Coordinated Recovery of Sparse Channels in Massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 104–118, 2014.

[12] S. Sun and T. S. Rappaport, "Millimeter wave MIMO channel estimation based on adaptive compressed sensing," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 47–53.

[13] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Deep learning for compressed sensing based channel estimation in millimeter wave massive mimo," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2019, pp. 1–6.

[14] D. L. Donoho and M. Elad, "Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via l1 Minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[15] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge university press, 2012.

[16] M. A. Davenport, "Random Observations on Random Observations: Sparse Signal Acquisition and Processing," Ph.D. dissertation, Rice University, 2010.

[17] P. J. Dhrymes, "Mathematics for Econometrics," Springer, Tech. Rep., 1978.

[18] Y. Shabara, E. Ekici, and C. E. Koksal, "Source Coding Based mmWave Channel Estimation with Deep Learning Based Decoding," *arXiv preprint arXiv:1905.00124*, 2019.

[19] S. Jokar and V. Mehrmann, "Sparse Solutions to Underdetermined Kronecker Product Systems," *Linear Algebra and its Applications*, vol. 431, no. 12, pp. 2437–2447, 2009.

[20] Y. Shabara, C. Emre Koksal, and E. Ekici, "How Long to Estimate Sparse MIMO Channels," *arXiv:2101.07287 [cs.IT]*, Jan. 2021.

[21] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT press, 2009.

[22] T. Ancheta, "Syndrome-Source-Coding and its Universal Generalization," *IEEE Transactions on Information Theory*, vol. 22, no. 4, pp. 432–436, Jul 1976.