

Compressive tracking with 1000-element arrays: a framework for multi-Gbps mm wave cellular downlinks

Dinesh Ramasamy, Sriram Venkateswaran and Upamanyu Madhow

Abstract—We propose and demonstrate the feasibility of multi-Gbps cellular downlinks using the mm-wave band. The small wavelengths allows deployment of compact base station antenna arrays with a very large number (32×32) of elements, while a compressive approach to channel acquisition and tracking reduces overhead while simplifying hardware design (RF beamforming with four phases per antenna element). The base station array transmits multiple compressive ($\ll 32 \times 32$) training beacons by choosing different sets of phases from $0^\circ, 90^\circ, 180^\circ, 270^\circ$ at random at each of the elements. Each mobile, equipped with a single antenna, reports the observations corresponding to the different beacons (e.g., on an existing LTE link at a lower frequency), allowing the array to estimate the angles of departure. We observe that tracking overhead can be reduced by exploiting the sparsity of the spatial channel to a given mobile (which allows parametric estimation of departure angles for the different paths), and the continuity in the user’s mobility at microsecond timescales (for tracking the evolution of departure angles). We first illustrate the basic feasibility of such a system for realistic values of system parameters, including range of operation, user mobility and hardware constraints. We then propose a compressive channel tracking algorithm that exploits prior channel estimates to drastically reduce the number of beacons and demonstrate the efficacy of the system using simulations.

I. INTRODUCTION

The rapid proliferation of smart phones and tablets in recent years, and the accompanying exponential growth in demand for wireless data, has strained existing mobile networks to their limit. It is clear that orders of magnitude increases in cellular network capacity, especially on the downlink, are required to sustain this growth. In this paper, we propose and investigate an approach that could potentially deliver such increases in downlink capacity via orders of magnitude increases in bandwidth, by exploiting vast amounts of spectrum in the millimeter(mm)-wave band (we focus on the unlicensed 60 GHz band in this paper), and spatial reuse, through a dense deployment of picocellular base stations (e.g., on lampposts), each with range of the order of 100-200m. The small carrier wavelength enables realization of electrically large but physically small antenna arrays (e.g., a 32×32 array easily fits within a square of side 6 cm), which allows us to synthesize highly directive beams from the base station (required to meet the link

budget with sufficient margin for oxygen absorption and rain) while providing flexible coverage. However, a fundamental bottleneck is to adapt such large arrays so as to track channel time variations due to user mobility. We present an architecture which exploits the sparsity of mm-wave channels to track such variations with low overhead.

There have been significant recent developments in indoor 60 GHz systems, which also employ directive transmission and reception in order to overcome the higher path loss at smaller carrier wavelengths. However, the adaptation algorithms developed for indoor 60 GHz communication do not apply to our outdoor system, which requires a significantly higher level of directivity (and hence much larger arrays) due to the significantly higher range, and must contend with a higher degree of user mobility. While this complicates the adaptation strategy at the base station, we employ an asymmetric architecture which potentially allows the same 60 GHz radio in the mobile device to be used both indoors and outdoors. 60 GHz is used only on the downlink, so that the mobile’s 60 GHz radio can be operated in receive-only mode. The mobile uses conventional cellular technologies such as LTE to provide feedback (used by the base station to adapt its array) and for uplink transmission. Another key feature of our architecture that allows scaling to very large base station arrays is the use of RF beamforming with coarse phase control. Conventional architectures that dedicate one RF chain per antenna element do not scale to such large arrays. In this scenario, RF beamforming using a single data stream, which is then upconverted and phase shifted in RF before being supplied to individual antenna elements, becomes extremely attractive. However, hardware scaling to a large number of phase shifters is still difficult, hence we consider highly simplified phase shifters that can apply, for example, one among the four values $\{0, \pm \frac{\pi}{2}, \pi\}$.

The starting point of our design is the observation that mm-wave channels are sparse, owing to reduced diffraction at smaller carrier wavelengths and substantial losses incurred at each reflection. As a result, tracking the channel between the array and the mobile is equivalent to estimating the Direction of Departures (DoDs) of a few paths. Conventional beam-scanning architectures attempt to do this by sending out a sequence of highly directional beams, whose union covers all possible directions by which the array can reach any mobile. When a mobile receives a signal from any beam – either along the Line-of-Sight (LoS) path or a reflected path – it provides feedback to the array, thereby allowing the array to maintain an inventory of paths to each mobile. However, this approach does not work in our setting: forming

The authors are with the Department of ECE, University of California Santa Barbara, CA 93106, USA. Email: {dineshr, sriram, madhow}@ece.ucsb.edu. This work was supported by the National Science Foundation through the grant CNS-0832154 and by the Institute for Collaborative Biotechnologies through the grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

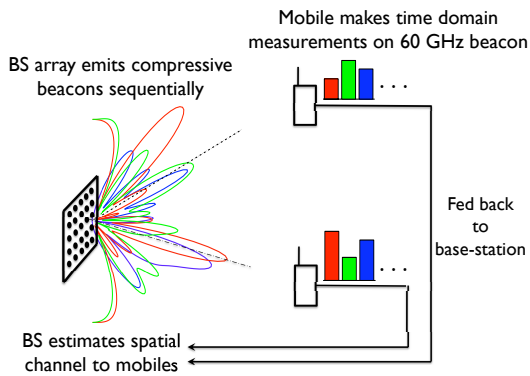


Fig. 1. Schematic of the compressive adaptation architecture. The base station emits a sequence of compressive beacons by choosing phases at random. The mobiles feed back measurements corresponding to the beacons from which the base station infers the channels.

such directional beams requires fine control of the phases, which is unavailable, and the large number of directions to be scanned leads to slow adaptation.

To overcome these problems, we propose a *compressive* adaptation architecture, as shown in Figure 1. The array transmits a sequence of beacons, each of which is produced by choosing a different set of phases randomly from the set $\{\pm 1, \pm j\}$. Since the phases are chosen randomly, the transmission is not focused in any particular direction. As a result, all the mobiles hear each beacon and feed back the complex baseband samples corresponding to the different beacons to the array (perhaps over an existing LTE link). The array uses this feedback to estimate the channel to all the mobiles *in parallel*. We keep the overhead, given by the number of beacons, small using two key ideas. First, we leverage the fact that the channel is sparse and use the recent advances we have made in compressive estimation (based on ideas from compressed sensing, but avoiding the pitfalls of “basis mismatch” that arise with a naïve application of compressed sensing) to lower the number of beacons. We then exploit the fact that the motion of the mobile is continuous at the timescales of communication to track changes in DoDs, instead of re-estimating them, thereby reducing the overhead even further. The compressive estimation algorithm is built on the observation that the array response along each path to the mobile is sinusoidal. At startup, we acquire the frequencies and gains corresponding to the different paths in a sequential manner. In each round of the sequential algorithm, we pick the sinusoid that best matches the observations from a coarse grid and then refine the gains and the frequencies in an iterative manner. As the mobile moves and the frequencies of the different paths change, we track the variations using frequency estimates from prior rounds of beaconing to bootstrap the refinement algorithm. We also provide theoretical insight by estimating the number of beacons required for successful channel acquisition and tracking, thereby quantifying the benefits of using prior channel estimates over re-acquiring the channel repeatedly.

Note that the small physical dimension of the array potentially allows multiple arrays with the same orientation to be deployed (e.g., along each “face” of a cube-like form factor). One array could then be dedicated to compressive beaconing, while others could use the channel estimates obtained thereby to form data-bearing beams towards the users. We do not discuss such details any further, and focus on the basic problem of channel tracking.

Related Work: Indoor channel measurements [1] indicate that mm-wave channels are sparse with only a few reflected paths in addition to the line of sight (LoS) path. Electronic beamsteering for steering around obstacles in the LoS path has been demonstrated, and 32-element arrays are now included in indoor 60 GHz products [2], [3]. Existing outdoor mm-wave links (with ranges up to 2.5 km) [4], on the other hand, use fixed, highly directive antennas which require careful manual pointing. Our work differs from both these scenarios, in that we wish to employ much higher directivities than for existing indoor mm-wave systems, while providing flexible beamsteering unlike existing outdoor point-to-point mm-wave links. While algorithms and protocols for RF beamforming have been developed for indoor settings [5], these involve explicit beam scanning and do not scale to the large arrays or rapid mobility of interest to us.

For regularly spaced arrays, the problem of estimating DoDs maps to one of estimating spatial frequencies (one-dimensional for a linear array, two-dimensional for the two-dimensional array considered here). The sparsity of the mm-wave channel naturally evokes the idea of using compressive sensing to estimate these spatial frequencies. However, a naïve application of standard compressive sensing, discretizing the set of possible frequencies and then using ℓ_1 reconstruction, can lead to large reconstruction errors when the frequencies actually come from a continuum [6]. We proposed a compressive estimation approach that avoids such pitfalls in [7] [8], and demonstrated its efficacy for frequency estimation given a noisy mixture of sinusoids. Our compressive tracking framework builds on this recent work, but goes further in terms of exploiting continuity in time for reducing beaconing overhead. Our work in [7] [8] and in the present paper also draws upon ideas from [9] [10].

While mm-wave cellular links have been proposed before [11], to the best of our knowledge, this is the first attempt to describe and evaluate a detailed system architecture that addresses the fundamental bottleneck of channel tracking in such settings.

II. LINK BUDGET & CHANNEL MODEL

We consider an architecture with interspersed rounds of channel sounding and data communication, where the DoD estimates from the channel sounding stage are used to beamform in the subsequent communication round. We compute a realistic link budget for both the data communication and channel sounding phases, taking into account the mandatory regulations and design constraints. This gives us an estimate of the distances that a mm-wave base station array can serve.

We then specify the channel and measurement models that we will use in the rest of the paper.

A. Link budget

Data communication: Suppose that we wish to establish a link with bandwidth 1 GHz using a 32×32 array whose elements are spaced $\lambda/2$ apart. The Federal Communications Commission (FCC) sets a limit on the power density in the neighborhood of the array, which translates to the following: the Effective Isotropic Radiated Power (EIRP), which is the maximum power transmitted in any direction, can be no larger than 40 dBm [12]. Suppose further that the transmitter has estimated the direction to the receiver and beamforms in this direction to communicate. Since a 32×32 array provides a beamforming gain of roughly 30 dB, the total power supplied to the transmit elements must be smaller than 10 dBm (the power per element is about 30 dB lower). In order to maximize the link range, we set the total transmit power to the largest permissible value of 10 dBm.

The path loss between a transmitter and a receiver that are a distance r apart is given by

$$G(r) = \frac{\lambda^2}{16\pi^2 r^2} e^{-\mu r},$$

where λ is the carrier wavelength and the exponential attenuation factor $\mu = 16$ dB/km accounts for losses in the mm-wave band due to oxygen absorption.

Since the receiver is on a mobile device with limited area, it is essential to keep it simple. Consequently, we assume that the receiver only provides us directivity gains on the order of 10 dBi, which can be realized easily using an array consisting of a few moderately directional elements.

Assuming that a link SNR of 6 dB suffices for communication, we find that we can establish links over 150 m with the parameters described above, with a link margin of 10dB. Thus, a base-station array can serve mobiles over distances on the order of traditional picocells.

Channel Sounding: The channel sounding phase differs from the communication phase in two fundamental ways. First, since the array transmits beacons by choosing the phases at different elements randomly, we no longer have a 30 dB beamforming gain. Secondly, we can no longer assume that receiver beamforms towards the transmitter and provides a 10 dBi directivity gain. The straightforward reason is that, at network startup, when the receiver has no knowledge of the transmitter's direction, it clearly cannot form a beam towards the array. More importantly, though, we would like the system to have the flexibility of estimating new paths that arise as the mobile moves and maintain an *inventory* of all paths to the mobile. Since these new paths need not fall into the mobile's beam towards the transmitter, the receiver is forced to be omnidirectional even if it has some knowledge of the direction to the array (say, from a prior round of training). As a result, the 10 dBi receiver directivity gain that we assumed in the communication phase is not available during channel sounding. Thus, the channel sounding stage loses about 40 dB beamforming gain relative to the data communication phase.

Suppose now that we send M compressive beacons to estimate the channel. For the estimate to have sufficient accuracy for beamforming in the communication stage, we find, from the Cramer-Rao Lower Bound (CRLB) on the estimation error variance, that the match-filtered signal-to-noise ratio (SNR) per path (SNR aggregated across all M measurements) needs to be around 20 dB. Furthermore, we find that we need $M = 30$ beacons for successful tracking (we explain why in Section IV). This implies that, for reliable channel estimation, each compressive beacon measurement must have an SNR of 6 dB. This SNR is similar to the value we obtain in the communication phase *with beamforming gains*. Thus, we need to choose the transmit power and sounding bandwidth in the training phase so that we compensate for the lack of the 40 dB beamforming gain. Channel sounding can be done with a much smaller bandwidth to reduce the noise power considerably. For example, we choose the sounding bandwidth to be $f_s = 1$ MHz which leads to a 30 dB improvement in noise power over the communication stage. This choice implies that the total transmitted power must increase to 20 dBm in the channel sounding stage. This can be done in one of two ways. The first option is to allow the transmit array elements to have a dynamic range of 10 dB, so that the array can be used both for channel sounding and communication. The alternative is to place two separate arrays side-by-side for sounding and communication, with the nominal transmit powers of these arrays set to 20 dBm and 10 dBm respectively.

B. Channel & Measurement Models

Consider a square array with $N_{1D} \times N_{1D}$ elements (we use $N_{1D} = 32$) that is placed in the $x - z$ plane. The mm-wave channel between the array and the mobile is sparse, typically consisting of the LoS path and single reflections off walls and the road. We denote the number of such paths by L . Let (θ_k, ϕ_k) denote the inclination and azimuthal angles of the k th path respectively. The complex gain along the k th path, including reflection losses and phase shifts, is denoted by g_k . The channel between the array and the mobile, denoted by $\mathbf{H} \in \mathbb{C}^{N_{1D} \times N_{1D}}$, is given by summing up the standard array responses along the L paths. Specifically, the (m, n) th entry in \mathbf{H} is

$$h_{mn} = \sum_{k=1}^L g_k e^{j(\omega_{x,k}m + \omega_{z,k}n)}, \quad 0 \leq m, n \leq N_{1D} - 1, \quad (1)$$

where $\omega_{x,k} = 2\pi(d/\lambda) \sin \theta_k \cos \phi_k$ and $\omega_{z,k} = 2\pi(d/\lambda) \sin \theta_k \sin \phi_k$. We see that each path contributes a two-dimensional sinusoid to the channel and, therefore, call $\omega_{x,k}$ and $\omega_{z,k}$ as the spatial frequencies associated with the k th path, which we denote by the vector $\boldsymbol{\omega}_k$.

Suppose that the array transmits a compressive beacon by choosing phases a_{mn} uniformly at random from $\{\pm 1, \pm j\}$. The complex baseband sample received at the mobile y is then

$$y = \sum_{m=0}^{N_{1D}-1} \sum_{n=0}^{N_{1D}-1} a_{mn} h_{mn} + z,$$

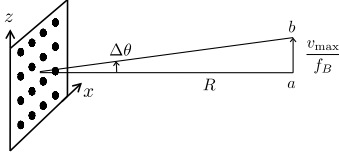


Fig. 2. Geometry corresponding to the maximum change in ω_z . The user moves from point a to b along the z axis at a speed of v_{\max} in the time interval $1/f_B$ between two consecutive channel sounding rounds

where $z \sim CN(0, \sigma^2)$ is the measurement noise, which is assumed to be independent across different beacons. We vectorize the matrices $e^{j(\omega_{x,k}m + \omega_{z,k}n)}$ and a_{mn} appropriately and denote the resulting vectors by $\mathbf{x}(\omega_k)$ and \mathbf{a} respectively. With this notation, we can rewrite the above equation as

$$\mathbf{y} = \mathbf{a}^T \sum_{k=1}^L g_k \mathbf{x}(\omega_k) + z.$$

Thus, if the arrays transmits a sequence of M beacons $\mathbf{a}(1), \mathbf{a}(2), \dots, \mathbf{a}(M)$ and we stack the observations $y(1), y(2), \dots, y(M)$ in a vector \mathbf{y} , we get the measurement model

$$\mathbf{y} = \sum_{k=1}^L g_k \mathbf{A} \mathbf{x}(\omega_k) + \mathbf{z}, \quad (2)$$

where $\mathbf{A} = [\mathbf{a}(1) \ \dots \ \mathbf{a}(M)]^T$ and $\mathbf{z} = [z(1) \ \dots \ z(M)]^T$ is the measurement noise distributed as $CN(0, \sigma^2 \mathbb{I}_M)$. To simplify the explanations, we have assumed here that the gains along the different paths g_k do not change as the M measurements are made. This may not be reasonable if the mobile moves fast and experiences Doppler shifts along the different paths. However, generalizing the algorithm to handle Doppler shifts is straightforward and we postpone the description of this extension to Section V.

III. COMPRESSIVE CHANNEL TRACKING

We now describe the algorithm used at the base station to track the channel variations as the mobile moves around. The base station emits compressive beacons of the form shown in Figure 1 in multiple rounds, with each round consisting of M beacons. Suppose that at the end of round r , the base station has estimates of the gains and the frequencies along the L paths to the mobile. Using the measurements made by the mobile corresponding to the M beacons in round $r+1$, the base station updates the gains and the frequencies along different paths, with the estimates from round r serving to bootstrap the process. We first explain how to choose the time between successive rounds of beacons and then describe the algorithm.

Frequency of beacons: The spatial frequency estimate made at the end of the r th round of beaconing becomes progressively less accurate with time as the mobile moves

around, leading to lower beamforming gains and, in turn, reduced link SNRs. Therefore, we choose the time between two rounds of beaconing to ensure that the beamforming loss is small. We denote the rate of beaconing (inverse of the time between two rounds of beacons) by f_B . Consider a mobile at point a in Figure 2, at a distance of R along the normal to the array. In the time interval between two rounds of beacons, the mobile can move at most v_{\max}/f_B , where v_{\max} is the maximum speed of the mobile. The maximum change in one of the spatial frequencies ω_x or ω_z occurs when the mobile moves parallel to the x or z -axes and is at most $v_{\max}/(f_B R)$ (assuming that the change in the angle subtended at the array $\Delta\theta$ is small enough for the approximation $\tan \Delta\theta \approx \Delta\theta$ to hold). Thus, the spatial frequency changes at most by $(2\pi d/\lambda) \times v_{\max}/(f_B R)$. We choose f_B large enough to ensure that the change in spatial frequency is much smaller than the width of the main-lobe ($2\pi/N_{1D}$), thereby minimizing beamforming losses. In particular, we find that for reasonable choices of the parameters ($v_{\max} = 45$ mi/h = 20 m/s, $N_{1D} = 32$ and $R > 15$ m), a beacon frequency $f_B = 75$ Hz suffices.

A. Tracking Algorithm: Spatial Frequency Refinement

We denote the estimates of the gains and spatial frequencies at the end of the r th round of beaconing by $\{\hat{g}_k, \hat{\omega}_k : k = 1, \dots, L\}$, $\hat{\omega}_k = (\hat{\omega}_{x,k}, \hat{\omega}_{z,k})$ (to keep the notation simple, we do not index the estimates by r). Assume, for simplicity, that no additional path pops up between the r th and the $r+1$ th round of beaconing (we will soon explain how to handle this) and that the true gains and spatial frequencies of the L paths at the start of the $r+1$ th round are given by $\{g_k, \omega_k : k = 1, \dots, L\}$ respectively. Suppose that we make M compressive measurements satisfying (2) in the $r+1$ th round. Since we have ensured that the spatial frequencies do not change much across successive rounds, we can perform a Taylor's series expansion of the array response $\mathbf{x}(\omega_k)$ around the estimate from the previous round $\mathbf{x}(\hat{\omega}_k)$, retain only the linear terms in the expansion and obtain

$$\mathbf{x}(\omega_k) \approx \left(\mathbf{x}(\hat{\omega}_k) + (\omega_{x,k} - \hat{\omega}_{x,k}) \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_x} + (\omega_{z,k} - \hat{\omega}_{z,k}) \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_z} \right). \quad (3)$$

Using (3) in (2), the measurements \mathbf{y} satisfy

$$\mathbf{y} = \sum_{k=1}^L g_k \mathbf{A} \mathbf{x}(\hat{\omega}_k) + \sum_{k=1}^L g_k (\omega_{x,k} - \hat{\omega}_{x,k}) \mathbf{A} \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_x} + \sum_{k=1}^L g_k (\omega_{z,k} - \hat{\omega}_{z,k}) \mathbf{A} \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_z} + \tilde{\mathbf{z}}, \quad (4)$$

where $\tilde{\mathbf{z}}$ includes the measurement noise and the modeling error resulting from the omission of higher order Taylor series terms. Since the wavelength is on the order of a few millimeters, the phase of the gain terms g_k undergoes a complete rotation even for miniscule movements of the

mobile. Thus, the gain g_k is drastically different from the estimate after the previous round \hat{g}_k . Therefore, we cannot linearize the gain g_k in (3) around its previous estimate and we need to compute it afresh.

The compressive tracking algorithm is based on the following observation: the model (4) is not linear in the unknown quantities $\{g_k, \omega_k - \hat{\omega}_k\}$. However, when we fix all the gains $\{g_k\}$, the observations are linear in the frequency refinements $\omega_k - \hat{\omega}_k$ and vice versa. Therefore, we propose an alternating optimization procedure, fixing one set of variables and then the other, to refine the frequencies and update the gains.

For notational simplicity, we denote the frequency refinement $\omega_k - \hat{\omega}_k$ by Δ_k in the following discussion.

Fix frequency refinements, update gains: In this stage, we set the frequency refinements $\Delta_k = 0$ and update the gains. Setting the frequency refinements to zero in (4), we get

$$\mathbf{y} = \sum_{k=1}^L g_k \mathbf{A} \mathbf{x}(\hat{\omega}_k) + \tilde{\mathbf{z}}.$$

Defining $\mathcal{H} = \mathbf{A} [\mathbf{x}(\hat{\omega}_1) \cdots \mathbf{x}(\hat{\omega}_L)]$ and a vector of gains $\mathbf{g} = [g_1 \cdots g_L]^T$, the least squares estimate of \mathbf{g} is given by

$$\hat{\mathbf{g}} = (\mathcal{H}^H \mathcal{H})^{-1} \mathcal{H}^H \mathbf{y}.$$

Fix gains, refine frequencies: Suppose that we are given a vector of gain estimates $\hat{\mathbf{g}} = [\hat{g}_1 \cdots \hat{g}_L]^T$. Using these gains in (4) and noting that $\sum_{k=1}^L \hat{g}_k \mathbf{A} \mathbf{x}(\hat{\omega}_k) = \mathcal{H} \hat{\mathbf{g}}$, we get

$$\mathbf{y} - \mathcal{H} \hat{\mathbf{g}} = \sum_{k=1}^L \hat{g}_k \Delta_{x,k} \mathbf{A} \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_x} + \sum_{k=1}^L \hat{g}_k \Delta_{z,k} \mathbf{A} \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_z} + \tilde{\mathbf{z}}$$

Let $\boldsymbol{\Omega} = [\Delta_{x,1} \cdots \Delta_{x,L} \Delta_{z,1} \cdots \Delta_{z,L}]^T$ be the vector of frequency refinements we wish to compute. Setting $\mathbf{y}_r = \mathbf{y} - \mathcal{H} \hat{\mathbf{g}}$ and

$$\mathcal{D} = \mathbf{A} \left[\hat{g}_1 \frac{\partial \mathbf{x}(\hat{\omega}_1)}{\partial \omega_x} \cdots \hat{g}_L \frac{\partial \mathbf{x}(\hat{\omega}_L)}{\partial \omega_x} \hat{g}_1 \frac{\partial \mathbf{x}(\hat{\omega}_1)}{\partial \omega_z} \cdots \hat{g}_L \frac{\partial \mathbf{x}(\hat{\omega}_L)}{\partial \omega_z} \right],$$

we see that the least squares estimate of the frequency refinements is the minimizer of the cost function $\|\mathbf{y}_r - \mathcal{D} \boldsymbol{\Omega}\|^2$. Since the frequency refinements are constrained to be real, the solution to this optimization problem is given by

$$\hat{\boldsymbol{\Omega}} = (\mathcal{Q}^T \mathcal{Q})^{-1} \mathcal{Q}^T \mathbf{t}_r,$$

where

$$\mathcal{Q} = \begin{pmatrix} \Re \{ \mathcal{D} \} \\ \Im \{ \mathcal{D} \} \end{pmatrix} \quad \mathbf{t}_r = \begin{pmatrix} \Re \{ \mathbf{y}_r \} \\ \Im \{ \mathbf{y}_r \} \end{pmatrix}.$$

We update the frequency estimates by adding the refinements to the estimates from the prior round. We can go through multiple rounds of the alternating optimization procedure to improve the accuracy of refinements.

B. Deleting non-existent paths & adding new paths

When users move around in a cluttered urban setting, paths that are visible in the r th sounding round could be occluded in subsequent rounds and vanish. Similarly, paths that were occluded in earlier rounds may materialize in the r th round. We now explain how to prune non-existent paths and add new paths.

Deleting weak paths: Let $\{\hat{g}_k, \hat{\omega}_k : k = 1, \dots, L\}$ be the gains and spatial frequencies estimated by the tracking algorithm in the r th sounding round. We remove paths that have little evidence in the measurements: if $|\hat{g}_k|^2 < \sigma^2 / (MN)$, the k -th path is deleted. We do this because the energy contributed by such paths to the measurements on the average $\mathbb{E} \|\hat{g}_k \mathbf{A} \mathbf{x}(\hat{\omega}_k)\|^2$ (over the ensemble of random measurement matrices) is smaller than the noise variance σ^2 .

Adding new paths: Let $\mathbf{y}_r = \mathbf{y} - \sum_{k=1}^{k=L} \hat{g}_k \mathbf{A} \mathbf{x}(\hat{\omega}_k)$ be the residual measurement after the refinement stage of the tracking algorithm. If the residual energy $\|\mathbf{y}_r\|^2$ is large, it is likely that a new path has risen between the array and the mobile. We can, therefore, use this as a criterion to add new paths.

The residual \mathbf{y}_r is restricted to an $M - L$ dimensional subspace that is orthogonal to the space spanned by the estimated frequencies $\mathcal{G} = \text{span}\{\mathbf{x}(\hat{\omega}_k) : k = 1, \dots, L\}$. Let us denote this $M - L$ dimensional subspace by \mathcal{G}^\perp . Thus, if there were no new paths, and the only contribution to \mathbf{y}_r was from noise (which is restricted to \mathcal{G}^\perp) the expected energy in the residual $\mathbb{E} \|\mathbf{y}_r\|^2 = (M - L)\sigma^2$. We can, therefore, add a path when $\|\mathbf{y}_r\|^2$ is much larger than this value. In our simulations, we use $\|\mathbf{y}_r\|^2 > 4(M - L)\sigma^2$. In such a scenario, we add a new path in a two stage process: first, we correlate the residual measurement \mathbf{y}_r against the responses $\mathbf{x}(\omega_d)$ corresponding to spatial frequencies from an oversampled DFT grid $\omega_d \in \{2\pi k / (SN_{1D}) : k = 0, \dots, SN_{1D} - 1\}^2$ (with the oversampling factor S set to 2) and pick the frequency that maximizes the correlation. We then refine all the $L + 1$ frequency and gain estimates using the iterative algorithm described above.

IV. BENEFITS OF TRACKING

Since the compressive measurements that the mobile feeds back to the base station count as overhead, we would like to keep them to a minimum. We quantify the overhead by using ideas from compressive parameter estimation to provide an estimate of the number of measurements required by the proposed channel tracking algorithm. We do this in two stages. First, we describe concepts from compressed sensing, such as ϵ -isometry and the Johnson-Lindenstrauss (JL) lemma, that help us quantify the number of measurements needed for compressive estimation to be successful in very general scenarios. We then apply these concepts to our tracking problem to quantify the feedback overhead. These ideas also let us compute the number of measurements required to re-estimate the channel in each round of beaconing without using any prior estimates. By comparing the estimates required for tracking and re-estimation, we find

that exploiting past channel knowledge reduces the feedback overhead considerably.

A. ϵ -isometry and JL Lemma

We begin by describing the intuition behind the conditions necessary for successful compressive parameter estimation and concretize these ideas using the concepts of ϵ -isometry and the JL lemma. Suppose that we wish to estimate a quantity $\mathbf{u} \in \mathcal{P}$ (a subset of \mathbb{C}^N) using M compressive measurements ($M \ll N$) of the form $\mathbf{y} = \mathbf{A}\mathbf{u} + \mathbf{z}$, with $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbb{I}_M)$. The ML estimate of \mathbf{u} is given by

$$\begin{aligned} \hat{\mathbf{u}} &= \arg \min_{\mathbf{v} \in \mathcal{P}} \|\mathbf{y} - \mathbf{A}\mathbf{v}\| \\ &= \arg \min_{\mathbf{v} \in \mathcal{P}} \|\mathbf{A}(\mathbf{u} - \mathbf{v}) + \mathbf{z}\| \end{aligned} \quad (5)$$

If the number of measurements is too small, then it is possible for $\|\mathbf{A}(\mathbf{u} - \mathbf{v})\| \approx 0$ even when $\|\mathbf{u} - \mathbf{v}\|$ is large. In such cases, with a small amount of noise, the optimizing solution $\hat{\mathbf{u}}$ could be drastically different from the true parameter \mathbf{u} . This problem can be avoided if the measurement matrix \mathbf{A} does not distort the geometry of the estimation problem too much. Specifically, if the distances between points in \mathcal{P} are preserved under the action of \mathbf{A} , so that

$$\|\mathbf{A}(\mathbf{u} - \mathbf{v})\| \approx \|\mathbf{u} - \mathbf{v}\| \quad \mathbf{u}, \mathbf{v} \in \mathcal{P},$$

the optimizing solution with compressive measurements will be close to the solution with all the measurements (this can be seen easily from (5) at high SNR by neglecting \mathbf{z} and using the distance preserving property of \mathbf{A}). The ϵ -isometry property captures the idea of distance preservation for a set of points precisely.

ϵ -isometry: The matrix \mathbf{A} is said to satisfy the ϵ -isometry property for the set $\mathcal{W} \subset \mathbb{C}^N$ if for some constant $C > 0$,

$$C(1 - \epsilon) \leq \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \leq C(1 + \epsilon), \quad \forall \mathbf{x} \in \mathcal{W}. \quad (6)$$

Thus, for the parameter estimation problem described above, \mathbf{A} would need to provide ϵ -isometry for all points in the set $\mathcal{W} = \{\mathbf{u} - \mathbf{v} : \forall \mathbf{u}, \mathbf{v} \in \mathcal{P}\}$.

JL Lemma: When the set \mathcal{W} is finite and the elements of \mathbf{A} are chosen i.i.d from suitable distributions, the JL lemma shows that \mathbf{A} provides an ϵ -isometry for points in \mathcal{W} with high probability if the number of measurements $M = O(\epsilon^{-2} \log |\mathcal{W}|)$. In particular, this result holds when the elements of \mathbf{A} are chosen uniformly at random from $\{\pm 1, \pm j\}$ and we use it to quantify the number of measurements needed for tracking the channel. We also note that the number of measurements required depends only on the cardinality of the set \mathcal{W} and not on its geometric structure.

B. Quantifying the measurements needed for tracking

Suppose that we are tracking the channels of Q mobiles and the channel to each mobile consists of L taps (if the number of taps are different, then the following arguments generalize if we set L to be the maximum number of taps across mobiles). Let $\{\hat{\omega}_k\}$ denote the estimates of the spatial

frequencies to one of the Q mobiles after the r th round of beacons. We denote the set of true gains and spatial frequencies of the same mobile at the start of round $r + 1$ by $\{g_k, \omega_k\}$ respectively. From (4), we see that the tracking algorithm tries to pick gains f_k and frequency refinements $\Delta_{x,k}$ and $\Delta_{z,k}$ that minimize the residual $\|\mathbf{A}\mathbf{r}\|$, with

$$\begin{aligned} \mathbf{r} = \sum_{k=1}^L g_k \mathbf{x}(\omega_k) - \left(\sum_{k=1}^L f_k \mathbf{x}(\hat{\omega}_k) + f_k \Delta_{x,k} \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_x} \right. \\ \left. + f_k \Delta_{z,k} \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_z} \right). \end{aligned} \quad (7)$$

Suppose we make enough measurements so that \mathbf{A} provides an ϵ -isometry for the set \mathcal{W} defined as follows:

$$\begin{aligned} \mathcal{W} = \left\{ \sum_{k=1}^L \left(a_k \mathbf{x}(\omega_k) + b_k \mathbf{x}(\hat{\omega}_k) + c_k \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_x} \right. \right. \\ \left. \left. + d_k \frac{\partial \mathbf{x}(\hat{\omega}_k)}{\partial \omega_z} \right) : \forall a_k, b_k, c_k, d_k \in \mathbb{C} \right\}. \end{aligned} \quad (8)$$

Then, for any choice of the true gains g_k , the gain estimates f_k and the frequency refinements $\Delta_{x,k}, \Delta_{z,k}$ in (8), we will have $\|\mathbf{r}\| \approx \|\mathbf{A}\mathbf{r}\|$. Arguing as we did for the general parameter estimation problem, this guarantees that the estimation problem's geometry is preserved and that we have made enough measurements to estimate the gains and frequency refinements accurately. Note though that this condition only suffices to estimate the channel to the specific mobile under consideration and that we need a similar condition for each of the Q mobiles. Thus, \mathbf{A} needs to provide ϵ -isometry for a set \mathcal{W}_T that is a union of Q sets \mathcal{W}_i $1 \leq i \leq Q$, each of which has the general form in (8) i.e. $\mathcal{W}_T = \bigcup_{Q \text{ mobiles}} \mathcal{W}_i$.

We cannot use JL lemma directly on the set \mathcal{W}_T to infer the number of measurements as each of the sets \mathcal{W}_i have an infinite number of elements. We sidestep this problem as follows: first, we discretize the set \mathcal{W}_T finely and use the JL lemma to infer the number of measurements needed for \mathbf{A} to provide an ϵ -isometry for the discretized set. Since the JL lemma depends only the number of points in the discretized version of \mathcal{W}_T , and not its structure, we do not need the true frequencies ω_k in this process. We then use the covering and bootstrapping arguments in [10] to extend this to an isometry for the infinite set \mathcal{W}_T . We omit the details of this argument. From this, we conclude that \mathbf{A} provides an ϵ -isometry for \mathcal{W}_T with $M = O(\epsilon^{-2} L \log(Q\epsilon^{-1}))$ measurements.

C. Quantifying the measurements needed for re-estimation

Suppose that we are required to estimate the gains and the frequencies to each of the Q mobiles in the r th round of beaconing without using any prior estimates. It is hard to quantify the number of measurements needed in this scenario when the frequencies come from a continuum. While the number of measurements can be specified in terms of the condition number and volume of the manifold $\mathcal{M} = \{\sum_k g_k \mathbf{x}(\omega_k) : \omega \in [0, 2\pi]^2\}$ [9], computing these quantities is not straightforward. However, we can get a rough idea of the number of measurements needed for

re-estimation by restricting the frequencies to a discrete set \mathcal{F} that is an S oversampled version of the DFT grid i.e. $\mathcal{F} = \{0, \frac{2\pi}{SN_{1D}}, \dots, \frac{2\pi(SN_{1D}-1)}{SN_{1D}}\}^2$. Suppose that the true gains and frequencies are g_k and ω_k and that we make M measurements satisfying

$$\mathbf{y} = \sum_{k=1}^L g_k \mathbf{A} \mathbf{x}(\omega_k) + \mathbf{z}, \quad \omega_k \in \mathcal{F}.$$

Since we do not have any prior estimates, the ML estimator tries to choose gains $f_k \in \mathbb{C}$ and frequencies $\Omega_k \in \mathcal{F}$ that minimize the residual

$$\begin{aligned} J &= \|\mathbf{y} - \mathbf{A} \mathbf{f} \mathbf{x}(\Omega_k)\| \\ &= \left\| \mathbf{A} \left(\sum_{k=1}^L g_k \mathbf{x}(\omega_k) - \mathbf{f}_k \mathbf{x}(\Omega_k) \right) + \mathbf{z} \right\|. \end{aligned} \quad (9)$$

Arguing as before, we see that, for successful compressive estimation, \mathbf{A} needs to provide an ϵ -isometry for the set

$$\mathcal{W}_A = \left\{ \sum_{k=1}^L g_k \mathbf{x}(\omega_k) - f_k \mathbf{x}(\Omega_k) : \forall g_k, f_k \in \mathbb{C}, \omega_k, \Omega_k \in \mathcal{F} \right\}$$

Note that in (9), we are optimizing over all frequencies $\Omega_k \in \mathcal{F}$, unlike the tracking phase, where we pick frequencies in the vicinity of the prior estimate $\hat{\omega}_k$. As a result the set \mathcal{W}_A is much larger than the corresponding set with the tracking algorithm \mathcal{W}_T . Using the JL lemma and covering arguments from [10], we find that $M = O(\epsilon^{-2} L \log(N\epsilon^{-1}))$ compressive beacons are necessary for \mathbf{A} to provide an ϵ -isometry for \mathcal{W}_A , where $N = N_{1D}^2$ is the total number of antenna elements.

We see that the overhead with the tracking algorithm grows logarithmically with the number of mobiles Q , while the overhead with re-estimation grows logarithmically with the number of antenna elements N . Since the number of antennas $N = 1024$ is much larger than the number of mobiles ($Q \approx 10$ s of users in a picocell), we see that using prior estimates in tracking the channel provides substantial gains.

V. ESTIMATING DOPPLER SHIFTS

We have assumed so far that the channel gains g_k remain unchanged within one round of beacons consisting of M measurements. However, this may not be realistic when the mobile moves reasonably fast. Indeed, the phase of the gains g_k increases linearly with time due to Doppler shifts and the rate of increase can be as large as $(2\pi v_{max})/(f_S \lambda)$ radians per sample (f_S is the bandwidth used for the compressive sounding phase). For reasonable choices of these parameters — $v_{max} = 20$ m/s, $f_S = 1$ MHz — we find that the phase can change by $\pi/4$ radians in the time taken to make $M = 30$ measurements. We generalize our algorithm to handle such Doppler shifts briefly.

Suppose, as before, that the spatial frequencies along the k th path to a mobile are $(\omega_{x,k}, \omega_{z,k})$ and the Doppler shift along this path is $\omega_{t,k}$. Denoting the phases chosen by the

base station for the the p th beacon by $\mathbf{a}(p) \in \{\pm 1, \pm j\}^N$, the corresponding received sample satisfies

$$y(p) = e^{j\omega_{t,k} p} \left(\mathbf{a}(p)^T \sum_{k=1}^L g_k \mathbf{x}(\omega_{x,k}, \omega_{z,k}) \right) + z(p). \quad (10)$$

Suppose that we make M measurements according to (10). In order to make the generalization clear, we use ω_k to denote the *triplet* of frequencies $(\omega_{x,k}, \omega_{z,k}, \omega_{t,k})$. Denoting the contribution of the k th path to the M measurements by $\mathbf{s}(\omega_k)$, we have $\mathbf{s}(\omega_k) = [\mathbf{a}(1)^T \mathbf{x}(\omega_{x,k}, \omega_{z,k}) \quad \mathbf{a}(2)^T \mathbf{x}(\omega_{x,k}, \omega_{z,k}) e^{j\omega_{t,k}} \dots \mathbf{a}(M)^T \mathbf{x}(\omega_{x,k}, \omega_{z,k}) e^{j(M-1)\omega_{t,k}}]^T$. Thus, the M measurements satisfy the model

$$\mathbf{y} = \sum_{k=1}^L g_k \mathbf{s}(\omega_k) + \mathbf{z}. \quad (11)$$

Suppose now that we have an estimate of the spatial and Doppler frequencies $\hat{\omega}_k$ from a prior round of beacons. As before, we perform a Taylor series expansion (11) around $\hat{\omega}_k$ and retain only the linear terms. We obtain an equation similar to (3) with an extra term corresponding to the refinement of the Doppler frequency. From this point, the tracking algorithm proceeds exactly as before, with the only difference being that we compute three frequency refinements per path (two spatial refinements, one Doppler refinement) instead of two.

VI. RESULTS

We simulate the performance of the tracking algorithm in an urban canyon, where the road and the buildings along its side reflect signals transmitted from the base-station array.

Transmitter: The base station array is mounted on a lamp-post of height 6m that is located at a distance of 7m from one of the walls of the 28m wide canyon. We transmit multiple rounds of compressive beacons using a 32×32 , $\lambda/2$ spaced array, whose elements are assumed to be isotropic. Each beacon is formed using randomly chosen array weights from $\{\pm 1, \pm j\}$, has a total transmit power (and EIRP) of 20dBm and a bandwidth of $f_S = 1$ MHz. We use $M = 30$ compressive beacons in each round of sounding and set the frequency of the channel sounding rounds to $f_B = 75$ Hz.

We now describe the rationale behind our choice of the bandwidth f_S and the number of beacons M . The matched filter SNR (SNR aggregated across measurements) grows linearly with the time spent beaconing M/f_S . The proposed tracking algorithm requires a matched filter SNR of 20 dB so that the estimate has sufficient accuracy for use in the data communication stage. This SNR requirement can be translated back to the time spent on beaconing $M/f_S \approx 30 \mu\text{s}$. Now, since the M measurements that the mobile feeds back to the array count as overhead, we would like to keep M as small as possible. However, in order to guarantee ϵ -isometry, M has to be sufficiently large [for tracking, from Section IV, $M = O(\epsilon^{-2} L \log(Q\epsilon^{-1}))$]. Therefore, we choose M as small as possible to guarantee ϵ -isometry and

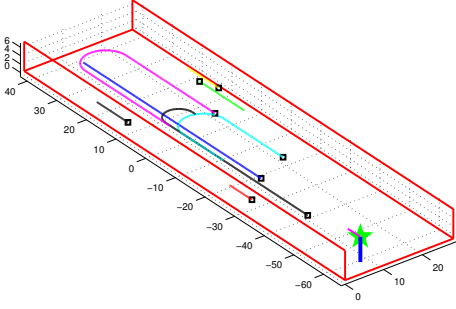


Fig. 3. Scenario of the simulation with the base station array mounted at a height of 6m and a distance of 7m from one of the walls of the 28m wide urban canyon. $Q = 8$ mobile users move along different trajectories in the canyon over the 5s duration of the simulation.

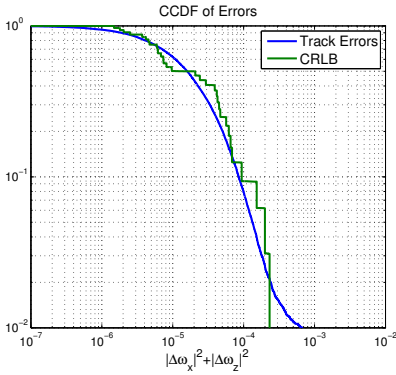


Fig. 4. CCDF of the square of tracking errors of the spatial frequencies along with their corresponding CCDF of the CRLB on error variance of spatial frequencies for the 8 concurrently simulated mobile users. The channel between the array and each mobile user consists of 4 angular taps corresponding to the LoS path and the three first order reflected paths.

then pick the bandwidth f_S to satisfy $M/f_S \approx 30\mu s$, leading to $M = 30$ measurements and $f_S = 1$ MHz.

Channel: The spatial channel between the array and a mobile consists of four paths: the LoS path and three first order reflections, from the road and the two walls. We assume that the reflections are lossless.

Receivers: We consider both pedestrian and vehicular mobile users who are within 100m from the array. The maximum speed of any mobile is 45mi/h. We plot the urban canyon and the trajectories of the users in Fig. 3.

The proposed algorithm correctly identified and tracked all four paths to each of the users. The Complementary Cumulative Distribution Function (CCDF) of the squared tracking errors is plotted in Fig. 4 along with the CCDF of the Cramer-Rao-Lower-Bound (CRLB) on the error variance. Since the curves virtually fall on top of one another, the estimation accuracy of the tracking algorithm is nearly optimal.

Overhead: The time spent in each round of beacons is $M/f_S = 30\mu s$ and the time between two successive rounds is $1/f_B = 1/75s$. Thus, the overhead in terms of the time spent on channel sounding $\frac{M/f_S}{1/f_B} = 0.23\%$ is minuscule.

Since the tracking algorithm requires significantly fewer beacons than a re-estimation based scheme, the feedback

overhead on the link from the mobile to the array is also very small (30 channel measurements per user every $1/f_B = 13ms$).

VII. CONCLUSIONS

We have introduced a novel asymmetric architecture for multi-Gbps mm-wave cellular downlinks which has the potential of relieving the smart phone induced capacity crisis facing mobile networks. The key challenge is channel tracking between a base station array with a very large number of elements and the mobiles it serves. We show that this can be accomplished with low overhead with an architecture that tracks the channels to different mobiles in parallel, exploiting the sparsity of the mm-wave channel, the continuity in a mobile's trajectory at the timescales of communication and recent advances in the theory of compressive estimation.

Important topics for future investigation include the coordination of multiple base stations to share space, time and frequency resources, the details of the protocol-level interaction between the mm-wave downlink and the LTE uplink, and the design of handoff mechanisms for handling blockages. At the physical layer, transceiver design for communication over large bandwidths remains an area of active research; for example, handling the increase in dynamic range due to channel dispersion subject to constraints on analog-to-digital conversion precision at high sampling rates.

REFERENCES

- [1] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Khoryaev, and A. Lomayev, "Experimental investigations of 60 ghz wlan systems in office environment," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 8, pp. 1488–1499, october 2009.
- [2] "Vizio XWH200 with Silicon Image ICs," <http://store.vizio.com/accessories/xwh200.html>.
- [3] X. Tie, K. Ramachandran, and R. Mahindra, "On 60 ghz wireless link performance in indoor environments," in *Proc. of PAM*. Springer, 2012.
- [4] "BridgeWave 60 GHz wireless ethernet point-to-point bridge," http://www.bridgewave.com/products/ar60_ar60x.cfm.
- [5] E. Perahia, C. Cordeiro, M. Park, and L. L. Yang, "IEEE 802.11ad: defining the next generation Multi-Gbps WiFi," January 2010, pp. 1–5.
- [6] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, May 2011.
- [7] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Information Theory and Applications Workshop (ITA), 2012*, feb. 2012, pp. 234–239.
- [8] —, "Compressive estimation in AWGN : General observations and a case study," in *Asilomar Conference on Signals, Systems, and Computers, 2012*, Nov. 2012.
- [9] R. Baraniuk and M. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, no. 1, p. 5177, 2009.
- [10] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, p. 253263, 2008.
- [11] Z. Pi and F. Khan, "System design and network architecture for a millimeter-wave mobile broadband (mmb) system," in *Sarnoff Symposium, 2011 34th IEEE*, may 2011, pp. 1–6.
- [12] S. K. Yong and C.-C. Chong, "An overview of multigigabit wireless through millimeter wave technology: potentials and technical challenges," *EURASIP J. Wirel. Commun. Netw.*, vol. 2007, no. 1, pp. 50–50, Jan. 2007.