

Estimation of Sparse MIMO Channels with Common Support.

Yann Barbotin, *Student Member, IEEE*, Ali Hormati, *Member, IEEE*, Sundeep Rangan, *Member, IEEE*,
and Martin Vetterli, *Fellow, IEEE*

Abstract—We consider the problem of estimating sparse communication channels in the MIMO context. In small to medium bandwidth communications, as in the current standards for OFDM and CDMA communication systems (with bandwidth up to 20 MHz), such channels are individually sparse and at the same time share a common support set. Since the underlying physical channels are inherently continuous-time, we propose a parametric sparse estimation technique based on finite rate of innovation (FRI) principles. Parametric estimation is especially relevant to MIMO communications as it allows for a robust estimation and concise description of the channels.

The core of the algorithm is a generalization of conventional spectral estimation methods to multiple input signals with common support. We show the application of our technique for channel estimation in OFDM (uniformly/contiguous DFT pilots) and CDMA downlink (Walsh-Hadamard coded schemes). In the presence of additive white Gaussian noise, theoretical lower bounds on the estimation of SCS channel parameters in Rayleigh fading conditions are derived. Finally, an analytical spatial channel model is derived, and simulations on this model in the OFDM setting show the symbol error rate (SER) is reduced by a factor 2 (0 dB of SNR) to 5 (high SNR) compared to standard non-parametric methods — e.g. lowpass interpolation.

Index Terms—Channel estimation, MIMO, OFDM, CDMA, Finite Rate of Innovation.

I. INTRODUCTION

Multiple input multiple output (MIMO) antenna wireless systems enable significant gains in both throughput and reliability [1]–[4] and are now incorporated in several commercial wireless standards [5], [6]. However, critical to realizing the full potential of MIMO systems is the need for accurate channel estimates at the receiver, and, for certain schemes, at the transmitter as well. As the number of transmit antennas is increased, the receiver must estimate proportionally more channels, which in turn increases the pilot overhead and tends to reduce the overall MIMO throughput gains [7].

To reduce this channel estimation overhead, the key insight of this paper is that most MIMO channels have an approximately *sparse common support* (SCS). That is, the

channel in each transmit-receive (TX-RX) antenna pair can be modeled as a discrete multipath channel, with the relative time delays being common across different TX-RX pairs. The commonality across the different antenna pairs reduces the overall number of degree of freedom to estimate, which can in turn be used to reduce the pilot overhead or improve the channel estimate. Also, in communication systems that depend on channel state feedback from the RX to the TX, the SCS model may enable a more compressed representation.

To exploit the SCS property of MIMO channels, we propose a variant along [8], [9] of the finite rate of innovation (FRI) framework, originally developed in [10]. The method, which we call SCS-FRI, uses classical spectral estimation techniques such as Prony’s method, ESPRIT and Cadzow denoising to recover the delay positions in frequency domain. The method is computationally simple, and our simulations demonstrate excellent performance in practical scenarios. The proposed SCS-FRI algorithm applies immediately to channel estimation in multi-output OFDM communication with contiguous or uniformly scattered DFT pilots. Interestingly enough it can be used on other modulation schemes provided a suitable pilot layout. The Walsh-Hadamard transform (WHT), used in CDMA downlink channel among others, qualifies if one controls the pilots layout in the WHT domain.

We also derive a simple scalar formula for the Cramér-Rao bound on the estimation of separable ToAs, and also point to a more general result by Yau and Bresler [11]. Both bounds are extended to Rayleigh fading SCS channels to lower bound the expected estimation error in fading conditions. Our simulations indicate the proposed SCS-FRI method is close to this bound at high SNRs.

A. SCS MIMO models

Due to the physical properties of outdoor electromagnetic propagation, wireless channels are often modeled as having a channel impulse response (CIR) that is sparse in the sense that they contain few significant paths [12]. With multiple antennas, the CIR measured at different antennas share a common support, i.e. the times of arrival (ToA) at different antennas are similar while the paths amplitudes and phases are distinct. This sparse common support (SCS) channels is illustrated in Figure 1. The SCS channel model is usually assumed in the literature, though its physical motivations are not always put forth.

It is important to note that the sparsity and common support assumptions only hold with respect to the channel bandwidth

Y. Barbotin, A. Hormati and M. Vetterli are with Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

S. Rangan is with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, NY.

M. Vetterli is also with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA.

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

This research is supported by *Qualcomm Inc., ERC Advanced Grant Support for Frontier Research - SPARSAM Nr : 247006* and *SNF Grant - New Sampling Methods for Processing and Communication Nr : 200021-121935/1*.

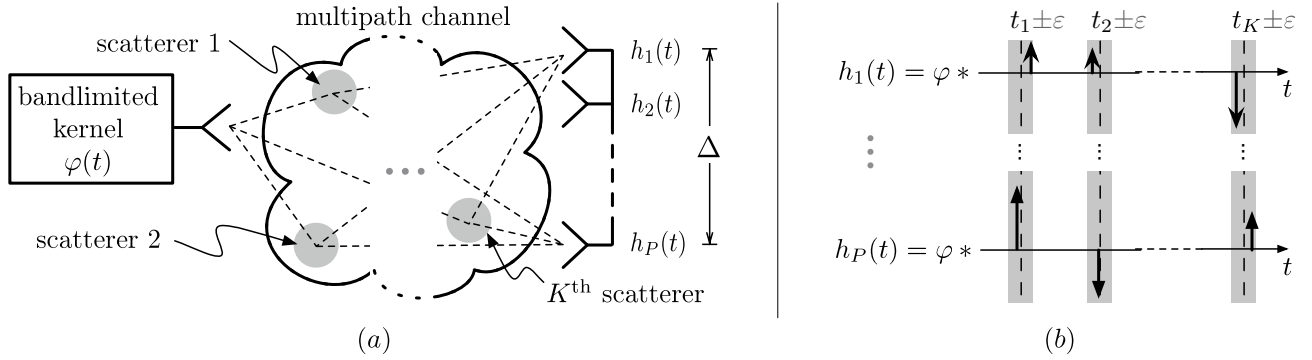


Fig. 1. (a) Transmission over a bandlimited medium with two scatterers and P receiving antennas. (b) The P channels contain two paths arriving at the same time up to $\pm\epsilon$, and are thus no exact SCS channels for $\epsilon > 0$.

TABLE I
CHANNEL BANDWIDTH IN POPULAR WIRELESS SYSTEMS

System	Code	Bandwidth B	Resolvable distance c/B
DVB-T [13]	DFT	5–8 MHz	38–60 m
IS-95 [14]	WHT	1.25 MHz	240 m
3GPP LTE [6]	DFT	1.4–20 MHz	15–215 m
UWB	—	> 500 MHz	< 60 cm

B and the SNR of the channel. Indeed, in the presence of noise, resolution is limited by the inverse bandwidth $1/B$, even if one knows exactly which parametric model the signal obeys. In practice, $1/10^{\text{th}}$ of the inverse bandwidth is a reasonable resolution to shoot for. The limited resolution has the effect of clustering paths from a single scatterer into a single path (promoting sparsity), and the small shift in the ToA due to the distance between antennas becomes negligible (promoting common support). Table I gives the channel bandwidth of several modern standards and c/B which is the distance travelled by an electromagnetic wave in a time lapse equal to the inverse bandwidth.

B. Related work

In OFDM systems, the majority of commercial channel estimators often simply perform some form of linear filtering or interpolation of the pilot symbols [15], [16]. Such non-parametric techniques are computationally very simple, but fundamentally cannot exploit the common sparsity in MIMO channel models. Since the phases and magnitudes are generally independent on the paths on different antenna pairs, the frequency response of sparse common support (SCS) channels are not correlated in any simple manner that can be exploited by basic linear interpolation of pilots.

A different line of work has proposed compressed sensing based methods for sparse channel estimation [17]–[20]. In the compressed sensing context, the SCS property is equivalent to *joint* or *group sparsity* for which there are several methods including group LASSO [21], [22], group OMP [23] and belief propagation [24]. Techniques for mixes of joint and individual

TABLE II
CHANNEL ESTIMATION METHODS ARE NATURALLY CLASSIFIED IN TERMS OF THE CHANNEL PROPERTIES THEY EXPLOIT.

	Algorithm	Exploited channel properties		
		Short delay-spread	Sparsity	Common support
DFT pilots layout	Contiguous	Lowpass	✓	
		FRI		✓
		SCS-FRI	✓	✓
	Scattered	Lowpass	✓	
		FRI	✓	✓
		SCS-FRI	✓	✓

sparsity are considered in [25], [26]. All of these compressed sensing methods, however, require that the delay locations are discretized and exact sparsity is achieved only when the true path locations fall exactly on one of the discrete points. With continuous value path locations, each path components will require a number of terms to approximate well, or demand a larger number of dictionary elements to offer a finer discretization.

Another joint estimation problem with FRI signals is studied in [27].

C. Contributions

The contributions of this work are four-folds:

- Extension of classical FRI sampling and estimation to multiple SCS channels (Section II)
- Derivation of simple scalar formulas for the CRB of SCS channels (Section III)
- Application to OFDM and Walsh-Hadamard coded (e.g. CDMA downlink) communications with contiguous or uniformly scattered DFT pilots (Section IV)
- Characterization of a precise spatial analytical model for SCS channels (Section V)

The proposed SCS-FRI algorithm stands out compared to FRI or lowpass interpolation as it exploits more channel prop-

erties, as indicated in Table II. Lowpass based techniques are a sensible non-parametric approach as they exploit the short delay-spread property. In general, any estimation technique based on uniformly scattered DFT pilots uses this property, as it is a necessary condition to the unicity of the solution.

We conclude our study with numerical simulations showing the efficiency of the SCS-FRI algorithm in a Rayleigh fading scenario, and compare its equalization gain to a standard non-parametric approach, i.e. lowpass interpolation in the DFT domain.

II. SPARSE COMMON SUPPORT FRI: THEORY AND ALGORITHMS

A. Problem formulation

We consider the physical setup described in Figure 1.(a). A periodic signal of limited bandwidth is transmitted over a multipath channel and uniformly sampled by a receiver with P antennas. This leads to P parallel multipath channels as shown in Figure 1.(b). The channels either share a common support exactly, in which case they are called exact SCS channels, or approximately, in which case they are called SCS channels (e.g. Figure 1.(b)).

Consider P exact SCS channels shaped by a kernel φ , with complex baseband equivalent model:

$$h_p(t) = \sum_{k=1}^K c_{k,p} \varphi(t - t_k), \quad c_{k,p} \in \mathbb{C}, \quad t_k \in [0, \tau[\quad (1)$$

where $\varphi(t)$ is the τ -periodic sinc function or Dirichlet kernel:

$$\varphi(t) = \sum_{k \in \mathbb{Z}} \text{sinc}(B(t - k\tau)) = \frac{\sin(\pi Bt)}{B\tau \sin(\frac{\pi t}{\tau})}. \quad (2)$$

The kernel φ is considered periodic as the filtering of a periodically padded signal by a linear shift invariant filter. Therefore, linear convolution of the CIR with the shaping kernel becomes circular.

We assume that the bandwidth parameter B satisfies $B = (2M + 1)/\tau$ for $M \geq K$. The paths coefficients $c_{k,p}$ are treated as complex random variables. N measurements $y_p[n]$ are acquired at a rate $1/T = N/\tau$ (with τ the signal period and $N \geq B\tau = 2M + 1$) and corrupted by AWGN

$$y_p[n] = h_p[n] + q_p[n] \quad n \in \{0, \dots, N - 1\}, \quad (3)$$

where $q_p \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \mathbb{I})$ if the channel is complex-valued or $q_p \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbb{I})$ if real-valued. In the DFT domain, the received signal is:

$$\hat{y}_p[m] = \hat{\varphi}[m] \cdot \sum_{k=1}^K c_{k,p} W^{mt_k} + \hat{q}_p[m]. \quad (4)$$

where $W = e^{-2\pi j/\tau}$ and $\hat{\varphi}[m] = 1/(2M + 1)$ for $|m| \leq M$ and is zero otherwise. The goal is to estimate the support $\{t_k\}_{k=1 \dots K}$ and the paths amplitudes $\{c_{k,p}\}_{k=1 \dots K, p=1 \dots P}$ from the NP samples collected in (3). Once the support is known, estimation of the path amplitudes is simple linear algebra as seen in (4).

B. Support recovery from baseband DFT coefficients

We start from (4). The DFT samples $\hat{y}_p[m]$ in the baseband ($|m| \leq M$) are the DFT coefficients of the channel corrupted by some Gaussian noise.

The noiseless DFT coefficients of a K -multipath channel have the well-known and interesting property to form a linear recurrent sequence of order $K + 1$, i.e., any coefficient $\hat{h}_p[m]$ ($m \geq -M + K$) can be expressed as a unique linear combination of the K previous DFT coefficients common to all indices m :

Lemma 1. *Given $\hat{h}_p[m] = \sum_{k=1}^K c_{k,p} W^{mt_k}$ for $m = -M + K, \dots, M$ and $t_i \neq t_j, \forall i \neq j$, there exists a unique set of coefficients $\{f_k\}_{k=1, \dots, K}$ such that:*

$$\hat{h}_p[m] = f_1 \hat{h}_p[m - 1] + f_2 \hat{h}_p[m - 2] + \dots + f_K \hat{h}_p[m - K]$$

where $x^K - f_1 x^{K-1} - \dots - f_{K-1} x - f_K$ is the polynomial with roots $\{W^{t_k}\}_{k=1, \dots, K}$.

Proof: A linear recursion of degree K can be written as:

$$x_n = f_1 x_{n-1} + \dots + f_K x_{n-K}, \quad f_K \neq 0. \quad (5)$$

Its characteristic equation is:

$$x^K - f_1 x^{K-1} - \dots - f_{K-1} x - f_K = 0. \quad (6)$$

If λ_x is a solution of (6) then multiplying both sides of the equation by λ_x^{n-K} ($\neq 0$ since $f_K \neq 0$) shows that λ_x^n is a solution of (5). Moreover by linearity, any linear combination of solutions of (5) is still a solution, and if (6) has K distinct solutions, $\{f_k\}_{k=1, \dots, K}$ is uniquely defined by a set of K independent linear equations. Hence, for $\sum_{k=1}^K c_{k,p} W^{mt_k}$ “solution” of (5), $t_k \not\equiv t_l \pmod{\tau}$ for all $k \neq l$, there exists a unique set $\{f_k\}_{k=1, \dots, K}$ such that $\{W^{t_k}\}_{k=1, \dots, K}$ are the K distinct roots of

$$x^K - f_1 x^{K-1} - \dots - f_{K-1} x - f_K.$$

The coefficients $\hat{y}_p[m]$ maybe arranged in a tall block-Toeplitz matrix

$$\mathbf{H}^{(L)} = \begin{bmatrix} \mathbf{H}_1^{(L)} \\ \mathbf{H}_2^{(L)} \\ \vdots \\ \mathbf{H}_P^{(L)} \end{bmatrix}$$

such that

$$\mathbf{H}_p^{(L)} = \begin{bmatrix} \hat{y}_{p,L-M-1} & \hat{y}_{p,L-M-2} & \cdots & \hat{y}_{p,-M} \\ \hat{y}_{p,L-M} & \hat{y}_{p,L-M-1} & \cdots & \hat{y}_{p,1-M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{p,M} & \hat{y}_{p,M-1} & \cdots & \hat{y}_{p,M-L+1} \end{bmatrix}, \quad (7)$$

where $\hat{y}_{i,j} = \hat{y}_i[j]$. The data matrix $\mathbf{H}^{(L)}$ is made of P Toeplitz blocks of size $(2M + 2 - L) \times L$, and we assume that both block dimensions are larger or equal to K . It possess interesting algebraic properties which form the core of line spectra estimation techniques. We will use Lemma 1 to

show three well-known spectral estimation tools which extend straightforwardly from Topelitz data matrices to block-Toeplitz ones, i.e. extend from single output to multiple outputs with SCS. We do so, with two propositions:

Proposition 1. [Annihilating filter property]

In the absence of noise ($\hat{y}_p[m] = \hat{h}_p[m]$), a set of exact SCS channels with K distinct paths verifies

$$\mathbf{H}^{(K+1)} \mathbf{f} = \mathbf{0}, \quad (8)$$

where $\mathbf{f} = [1 \ -f_1 \ \dots \ -f_K]^T$ are the annihilating filter coefficients such that the polynomial $p_{\mathbf{f}}(z) = 1 - \sum_{k=1}^K f_k z^k$ has K roots $\{e^{-2\pi j t_k / \tau}\}_{k=1 \dots K}$. The matrix $\mathbf{H}^{(K+1)}$ is built with blocks as in (7) (with $L = K + 1$).

Proof: This is a direct consequence of Lemma 1. ■

Proposition 2. For a set of exact SCS channels with K distinct paths and in the absence of noise, $\mathbf{H}^{(L)}$ satisfies

$$\text{rank } \mathbf{H}^{(L)} = K,$$

for $K \leq L \leq 2M + 2 - K$.

Proof: Let $\tilde{\mathbf{H}}$ be the top-left $K \times K$ minor of $\mathbf{H}^{(L)}$. It can be written as the sum of K rank-1 matrices:

$$\tilde{\mathbf{H}} = \sum_{k=1}^K c_{k,\cdot} W^{(M+1-L)t_k} \boldsymbol{\xi}_k^* \boldsymbol{\xi}_k,$$

such that $\boldsymbol{\xi}_k = [1 \ W^{t_k} \ W^{2t_k} \ \dots \ W^{(K-1)t_k}]$. If $t_k \not\equiv t_l \pmod{\tau}$ for all $k \neq l$, then $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_K\}$ form a set of non colinear vectors. Therefore $\text{rank } \mathbf{H}^{(L)} \geq K$.

Choose K mutually independent rows of $\mathbf{H}^{(L)}$. From Lemma 1, truncating these row vectors to length K preserves the linear independence. Therefore, given a row $\mathbf{h} = [h[0], \dots, h[L-1]]$ of $\mathbf{H}^{(L)}$, there exists a linear combination of these K rows \mathbf{h}' such that

$$h[k] = h'[k], \quad k = 0, \dots, K-1.$$

By construction, since all rows of $\mathbf{H}^{(L)}$ verify the the same linear recursion of degree K , \mathbf{h} and \mathbf{h}' verifies this recursion. Hence $\mathbf{h} = \mathbf{h}'$, implying $\text{rank } \mathbf{H}^{(L)}$ is at most K . ■

a) *Block-Prony algorithm:* Proposition 1 is Prony's method [28], [29] for block-Toeplitz matrices. We call the corresponding algorithm "Block-Prony TLS", listed under Algorithm 1. It solves the annihilating filter equation (8) in the total least-square (TLS) sense. The crucial step is the identification of what shall be the unidimensional null space of $\mathbf{H}^{(K+1)}$ in a noiseless case. Solving this problem in the TLS sense yields the least right singular vector of $\mathbf{H}^{(K+1)}$. Prony's method is notoriously sensitive to noise, which is to be expected as the result relies on identification of the unidimensional complement of the K -dimensional signal space. This sensitivity can be mitigated with prior denoising of the measurements.

b) *Block-ESPRIT algorithm:* Proposition 2 implies each block in the data matrix shares the same signal subspace. Hence the ESPRIT TLS algorithm outlined in [30] applies as-is to the block-Toeplitz data matrix $\mathbf{H}^{(L)}$. The Block-ESPRIT TLS algorithm is outlined in Algorithm 2. The Block-

Algorithm 1 Block-Prony_TLS

Require: An estimate on the number of effective paths K^{est} , $2M + 1$ ($M \geq K$) channel DFT coefficients $\hat{y}_p[m] = \sum_{k=1}^K c_{k,p} W_N^{mt_k} + \hat{q}_p[m]$ for $|m| \leq M$, $p = 1 \dots P$.

- 1: Build $\mathbf{H}^{(K^{\text{est}}+1)}$ according to (7).
- 2: Compute the SVD decomposition of the data matrix: $\mathbf{H}^{(K^{\text{est}}+1)} = \mathbf{U} \mathbf{S} \mathbf{V}^*$.
- 3: $\phi \leftarrow \text{roots}(\mathbf{v})$, such that \mathbf{v} is the right singular vector associated to the least singular value.
- 4: **return** $\{t_k^{\text{est}}\}_{k=1 \dots K^{\text{est}}} \leftarrow -\frac{\tau}{2\pi} \arg \phi$.

Algorithm 2 Block-ESPRIT_TLS

Require: An estimate on the number of effective paths K^{est} , $2M + 1$ ($M \geq K$) channel DFT coefficients $\hat{y}_p[m] = \sum_{k=1}^K c_{k,p} W_N^{mt_k} + \hat{q}_p[m]$ for $|m| \leq M$, $p = 1 \dots P$.

- 1: Build $\mathbf{H}^{(M)}$ according to (7).
- 2: Compute the SVD decomposition of the data matrix: $\mathbf{H}^{(M)} = \mathbf{U} \mathbf{S} \mathbf{V}$.
- 3: Extract the signal subspace basis $\boldsymbol{\Xi}_0 = \mathbf{V}_{1:(M-1), 1:K^{\text{est}}}$.
- 4: Extract the rotated signal space basis $\boldsymbol{\Xi}_1 = \mathbf{V}_{2:M, 1:K^{\text{est}}}$.
- 5: Solve $\boldsymbol{\Xi}_1 = \boldsymbol{\Xi}_0 \boldsymbol{\Psi}$ in the TLS sense.
- 6: **return** $\{t_k^{\text{est}}\}_{k=1 \dots K^{\text{est}}} \leftarrow \{-\frac{\tau}{2\pi} \arg \lambda_k(\boldsymbol{\Psi})\}_{k=1 \dots K^{\text{est}}}$.

Algorithm 3 Block-Cadzw denoising

Require: A block-Toeplitz matrix $\mathbf{H}^{(L)}$ and a target rank K .
Ensure: A block-Toeplitz matrix $\mathbf{H}^{(L)}$ with $\text{rank} \leq K$.

- 1: **repeat**
- 2: Reduce $\mathbf{H}^{(L)}$ to rank K by a truncated SVD.
- 3: Make $\mathbf{H}_p^{(L)}$ $p = 1 \dots P$, Toeplitz by averaging diagonals.
- 4: **until** convergence

Algorithm 4 SCS-FRI channel estimation

Require: An estimate on the number of effective paths K^{est} , $2M + 1$ ($M \geq K$) noisy channel DFT coefficients $\hat{y}_p[m] = \sum_{k=1}^K c_{k,p} W_N^{mt_k} + \hat{q}_p[m]$ for $|m| \leq M$, $p = 1 \dots P$.

Ensure: Support estimate $\{t_k^{\text{est}}\}_{k=1 \dots K^{\text{est}}}$

- 1: Build $\mathbf{H}^{(M)}$ according to (7).
- 2: $\mathbf{H}^{(M)} \leftarrow \text{Block-Cadzw}(\mathbf{H}^{(M)}, K^{\text{est}})$ [optional].
- 3: Update $\hat{y}_p[m]$ with the first row and column of the denoised block $\mathbf{H}_p^{(M)}$.
- 4: Estimate the common support with Block-Prony_TLS or Block-ESPRIT_TLS.
- 5: Estimate $\{c_{k,p}\}$ solving P linear Vandermonde systems (4).

ESPRIT algorithm fulfils the same goal as the Block-Prony algorithm, but its essence is entirely different. Where Prony's method identifies a line with least energy in a K dimensional space, ESPRIT finds the rotation between two K -dimensional subspaces in an M -dimensional space. What makes ESPRIT much more resilient to noise is that the two subspaces are computed from the most energetic part of the signal.

c) *Block-Cadzow denoising*: Proposition 2 used together with the block-Toeplitz structure property yields the “*lift-and-project*” denoising Algorithm 3, which we call *Block-Cadzow* denoising [31]. Using the same argument as in [32], the block-Cadzow algorithm provably converges.

d) *SCS-FRI*: We have all the elements to describe the SCS-FRI algorithm. The Block-Cadzow algorithm may be used to denoise the measurements and is followed by either Block-Prony or Block-ESPRIT estimation of the common ToAs (solved in the TLS sense).

For Cadzow denoising and ESPRIT, it is empirically found that a data matrix with square blocks works well. The last step is to estimate the path amplitudes independently for each channel. This is done by solving a linear Vandermonde system (4) [33]. The processing chain at the receiver is listed in Algorithm 4 and shown in Figure 2. Combination of Cadzow and ESPRIT for estimation of a single OFDM channel is considered in [34]. We assumed the number of paths to be known. For estimation techniques of the number of paths, we refer to [30].

III. ESTIMATION THEORETIC BOUNDS ON SCS-FRI RECOVERY

A. Deterministic multipath channels

In [10], the authors derive the Cramér-Rao lower bound [35], [36] for estimating the positions and weights of the Diracs in FRI signals. Considering a single Dirac with deterministic amplitude in a single-channel real-valued scenario, the minimal relative uncertainties on the location of the Dirac, t_1 , and on its amplitude, c_1 , are given by

$$\mathbb{E} \left[\left(\frac{\Delta t_1}{\tau} \right)^2 \right] \geq \frac{3(2M+1)}{4\pi^2 NM(M+1)} \text{PSNR}^{-1}$$

$$\mathbb{E} \left[\left(\frac{\Delta c_1}{c_1} \right)^2 \right] \geq \frac{2M+1}{N} \text{PSNR}^{-1}$$

where $\text{PSNR} = c_1^2/\sigma^2$ is the input peak signal to noise ratio. When there are more than two Diracs, the Cramér-Rao formula for one Dirac still holds approximately when the locations are sufficiently far apart¹.

B. Jointly Gaussian multipath channels

We derive bounds on the support estimation accuracy with measurements taken according to (3). The paths coefficients $c_{k,p}$ are assumed to be jointly Gaussian, and modeled as the product of $a_{k,p} = \mathbb{E}[|c_{k,p}|]$ by a standard normal random variable $Z_{k,p}$ having the following properties, consistent with the well-known Rayleigh-fading model:

- $Z_{k,p} \sim \mathcal{N}_{\mathbb{C}}(0, \sqrt{1/2}\mathbb{I})$.
- Similar expected path amplitude between antennas: $a_k \stackrel{\text{def}}{=} a_{k,1} = a_{k,2} = \dots = a_{k,P}$.
- Independence between paths: $\mathbb{E}[Z_{k,p}Z_{k',p'}^*] = 0$, $k \neq k'$.

- The random vector $\mathbf{Z}_k = [Z_{k,1} \dots Z_{k,P}]^T$ is defined as $\mathbf{Z}_k = \mathbf{L}_k \mathbf{r}$, where \mathbf{L}_k is the Cholesky factor of the covariance matrix $\mathbf{R}_k = \mathbb{E}[\mathbf{Z}_k \mathbf{Z}_k^*]$ and \mathbf{r} is a vector of iid standard complex Gaussian random variables.

The Rayleigh-fading case can be seen as deterministic if conditioned on the path amplitudes. Thus, the Cramér-Rao bounds for random paths coefficients are random variables for which we can compute statistics. Expectation and standard deviation will respectively give the expected accuracy of the estimator and its volatility. For a single path, and a symmetric or antisymmetric φ (not necessarily a sinc kernel), the Crámer-Rao bound has a concise closed form formula:

Proposition 3. *With complex-valued measurements according to (3), $K = 1$, and \mathbf{Z}_1 be a random Gaussian vector, then*

$$\mathbb{E} \left[\left(\frac{\Delta t_1}{\tau} \right)^2 \right] \geq \frac{\mathbb{E}[(\mathbf{Z}_1^* \mathbf{Z}_1)^{-1}]}{2N \cdot d\text{SNR}}, \quad (9)$$

where $d\text{SNR} = |a_1|^2 \|\varphi'(nT - t_1)\|^2 / (N\sigma^2)$ is the differential SNR and $\lambda_1, \dots, \lambda_P$ are the eigenvalues of the covariance matrix \mathbf{R}_1 and $P > 1$:

- *Uncorrelated paths coefficients*, $\lambda_1 = \dots = \lambda_P = 1$:

$$\mathbb{E}[(\mathbf{Z}_1^* \mathbf{Z}_1)^{-1}] = (P-1)^{-1}, \quad P > 1.$$

- *Correlated path coefficients*, such that $\lambda_1 \neq \dots \neq \lambda_P$:

$$\mathbb{E}[(\mathbf{Z}_1^* \mathbf{Z}_1)^{-1}] = \sum_{p=1}^P (-\lambda_p)^{P-1} \frac{\ln \lambda_p}{\lambda_p} \prod_{p' \neq p} (\lambda_{p'} - \lambda_p)^{-1}$$

Proof: See [37]. The uncorrelated case is found in various statistical handbooks as the first moment of an inverse- χ^2 distributed random variable. For the correlated case, see [38]. ■

This expression is a suitable approximation for multipaths scenario with distant paths (separated by more than twice the inverse bandwidth. It gives an important insight on the evolution of the estimation performance when uncorrelated antennas are added to the system. Namely, the RMSE decays as $1/\sqrt{P-1}$.

In general, multiple paths are interacting with each other and the information matrix cannot be considered diagonal. In this case Yau and Bresler [11] derived the following expression:

Proposition 4. [11] *Let Φ and Φ' be $N \times K$ matrices such that*

$$\Phi_{n,k} = \varphi((n-1)T - t_k), \quad \Phi'_{n,k} = \varphi'((n-1)T - t_k),$$

$n \in \{1, \dots, N\}$, $k \in \{1, \dots, K\}$. Given the stochastic matrix

$$\mathbf{C} = \text{diag}(a_1, \dots, a_K) \left(\sum_{p=1}^P \mathbf{Z}'_p \mathbf{Z}'_p{}^* \right) \text{diag}(a_1^*, \dots, a_K^*),$$

with $\mathbf{Z}'_p = [Z_{1,p} \dots Z_{K,p}]^T$, the Fisher information matrix \mathbf{J} conditioned on the path amplitudes is given by

$$\mathbf{J} = 2\sigma^{-2} \Phi'^* P_{\ker \Phi} \Phi' \odot \mathbf{C}. \quad (10)$$

such that $P_{\ker \Phi} = \mathbb{I} - \Phi \Phi^\dagger$ is the projection into the nullspace of Φ and “ \odot ” denotes the entrywise matrix product.

¹Empirically, the distance should be larger than $2/B$.

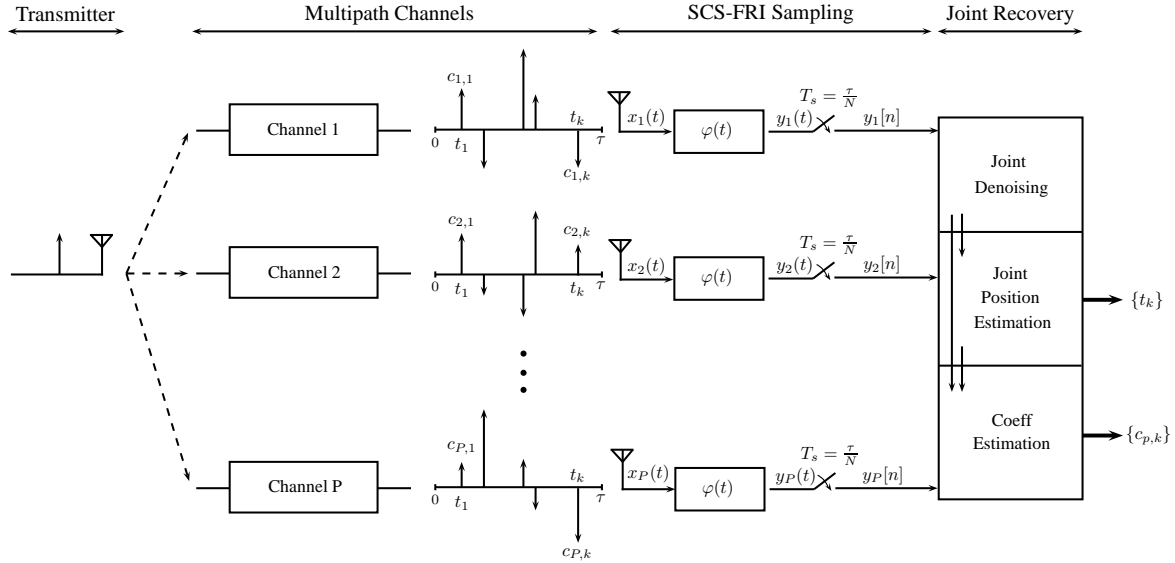


Fig. 2. The SCS-FRI sampling and reconstruction scheme in a multi-antenna channel estimation setting with P receiving antennas.

See [11] for the proof. The Cramér-Rao bounds for the estimation of the normalized times of arrival are on the diagonal of the expectation of \mathbf{J}^{-1} . The matrix \mathbf{J} is a complex Wishart matrix. Computing its inverse moments analytically is not an easy task, nevertheless it can be numerically computed via Monte-Carlo simulations.

IV. APPLICATION TO OFDM AND CDMA DOWNLINK

1) SCS-FRI with uniformly scattered DFT pilots (OFDM):

The theory in Section II is developed for contiguous DFT coefficients. In OFDM communications, pilots are often uniformly laid out in frequency (ETSI DVB-T [13], 3GPP LTE [6],...). The period of pilot insertion D is upper-bounded by Δ^{-1} , the inverse of the delay-spread of the CIR : $D < \tau/\Delta$. If not, the CIR cannot be unambiguously recovered from the pilots because of aliasing. For a fixed number of pilots, D is chosen as large as possible ($D = \lceil \tau/\Delta \rceil$), as interpolation of the CIR spectrum is more robust than extrapolation.

SCS-FRI can take advantage of uniformly scattered pilot layouts [37], [39]. For $\hat{\varphi}$ flat in $\{-MD, \dots, MD\}$, equation (4) becomes:

$$\hat{y}_p[mD] = \sum_{k=1}^K c_{k,p} W_N^{mDt_k} + \hat{q}_p[mD], \quad (11)$$

which corresponds to a dilation by D of the support parameters $\{t_k\}$. By definition $0 \leq t_k < \Delta$, and so the bound on D prohibits aliasing of Dt_k . Therefore, SCS-FRI is applicable without other modification than division of the recovered support parameters by D . The results of Proposition 3 can be extended to scattered pilot with minimal effort.

Corollary 1. *The minimal uncertainties on the estimation of the parameters in the SCS-FRI scenario (11) with P signals*

are given by

$$\begin{aligned} \mathbb{E} \left[\left(\frac{\Delta t_1}{\tau} \right)^2 \right] &\geq \frac{3BT}{4D^2\pi^2 M(M+1)} \mathbb{E} [\text{ESNR}^{-1}] \\ \mathbb{E} \left[\left(\frac{\Delta c_\ell}{c_\ell} \right)^2 \right] &\geq BT \mathbb{E} [\text{PSNR}_\ell^{-1}] \quad \ell = 1, \dots, P. \end{aligned}$$

For real-valued signal and noise $\text{ESNR} = \frac{1}{\sigma^2} \sum_{\ell=1}^P c_\ell^2$ denotes the effective signal to noise ratio and $\text{PSNR}_\ell = c_\ell^2/\sigma^2$. For complex-valued signal and noise $\text{ESNR} = \frac{1}{2\sigma^2} \sum_{\ell=1}^P c_\ell^* c_\ell$ and $\text{PSNR}_\ell = c_\ell^* c_\ell / (2\sigma^2)$.

Proof:

$$\mathbb{E} \left[\left(\frac{\Delta t_1}{\tau} \right)^2 \right] = \mathbb{E} \left[\left(\frac{\Delta(Dt_1)}{\tau} \right)^2 \right] \cdot D^{-2}.$$

Evaluation of $\mathbb{E} \left[\left(\frac{\Delta(Dt_1)}{\tau} \right)^2 \right]$ based on measurements from (11) is answered by Proposition 3. The dSNR is explicitly computed for (11) taking $\varphi = \text{sinc}_B$. ■

2) *Extension to Walsh-Hadamard coded schemes (CDMA):* Numerous applications use the 2^n -WHT to code the channel into 2^n subchannels ($N = 2^n$). Among others, IS-95 uses a 64-WHT to code the downlink channel. The straightforward way to insert pilots is to use one of these subchannels as a pilot itself and use correlation based channel estimation methods as the Rake-receiver for example [40]. The SCS-FRI algorithm works in the DFT domain but can nevertheless be applied in the WHT domain with pilots uniformly scattered by D a power of 2.

Proposition 5. *Let \mathbf{W}_n and \mathbf{S}_n be respectively the 2^n -points DFT and WHT matrices obtained by Sylvester's construction:*

$$\mathbf{S}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \mathbf{S}_{i+1} = \mathbf{S}_1 \otimes \mathbf{S}_i.$$

Then, for any $\ell \in \{1, \dots, n-1\}$ the set of \mathbf{S}_n 's columns with indices in $\{2^\ell + i\}_{1 \dots 2^\ell}$ and the set of \mathbf{W}_n 's columns with indices in $\{(i-1/2) \cdot 2^{n-\ell} + 1\}_{1 \dots 2^\ell}$ span the same subspace.

Proof: We partition the Walsh-Hadamard transform matrix in two “left” and “right” blocks:

$$\mathbf{S}_n = \begin{bmatrix} \mathbf{S}_n^{(l)} & \mathbf{S}_n^{(r)} \end{bmatrix}, \quad \mathbf{S}_n^{(l)} = \begin{bmatrix} \mathbf{S}_{n-1} \\ \mathbf{S}_{n-1} \end{bmatrix}, \quad \mathbf{S}_n^{(r)} = \begin{bmatrix} \mathbf{S}_{n-1} \\ -\mathbf{S}_{n-1} \end{bmatrix}.$$

Given $\mathbf{w}_{2^n}^k = [W_{2^n}^{0k} \dots W_{2^n}^{(2^n-1)k}]$ the k^{th} vector of the DFT basis and $\mathbf{s}^{(r)} \in \text{span } \mathbf{S}_n^{(r)}$:

$$\begin{aligned} \langle \mathbf{w}_{2^n}^k, \mathbf{s}_n^{(r)} \rangle &= \sum_{l=0}^{N-1} W_{2^n}^{kl} s_n^{(r)}[l] \\ &= \sum_{l=0}^{2^{n-1}-1} W_{2^n}^{kl} s_n^{(r)}[l] + W_{2^n}^{k(l+2^{n-1})} s_n^{(r)}[l+2^{n-1}] \\ &= \sum_{l=0}^{2^{n-1}-1} W_{2^n}^{kl} s_n^{(r)}[l] \left(1 - W_{2^n}^{k2^{n-1}}\right) \end{aligned}$$

Hence

$$\langle \mathbf{w}_{2^n}^k, \mathbf{s}_n^{(r)} \rangle = 0, \quad \text{for } k \text{ even.}$$

The spans of $\mathbf{S}_n^{(r)}$ and $\mathbf{S}_n^{(l)}$ partition the original 2^n -dimensional space in two subspaces of dimension 2^{n-1} . Let $\mathbf{W}_n^{(o)} = \{\mathbf{w}_{2^n}^k\}_{0 \leq 2k+1 < N, k \in \mathbb{N}}$ be the DFT basis vectors with odd indices and $\mathbf{W}_n^{(e)} = \{\mathbf{w}_{2^n}^k\}_{0 \leq 2k < N, k \in \mathbb{N}}$ the ones with even indices. The spans of $\mathbf{W}_n^{(o)}$ and $\mathbf{W}_n^{(e)}$ partition the original space into two subspaces of dimension 2^{n-1} . Since $\text{span } \mathbf{S}_n^{(r)} \perp \text{span } \mathbf{W}_n^{(e)}$:

$$\begin{aligned} \text{span } \mathbf{W}_n^{(o)} &= \text{span } \mathbf{S}_n^{(r)}, \\ \text{span } \mathbf{W}_n^{(e)} &= \text{span } \mathbf{S}_n^{(l)}. \end{aligned}$$

This property applies recursively, since for $k \in \{0, \dots, 2^{n-1}-1\}$:

$$\begin{aligned} \langle \mathbf{w}_{2^n}^{2k}, \mathbf{s}_n^{(l)} \rangle &= \left\langle \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{w}_{2^{n-1}}^k \\ \mathbf{w}_{2^{n-1}}^k \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{s}_{n-1} \\ \mathbf{s}_{n-1} \end{bmatrix} \right\rangle, \\ &= \langle \mathbf{w}_{2^{n-1}}^k, \mathbf{s}_{n-1} \rangle, \end{aligned}$$

where $\mathbf{s}_{n-1} \in \text{span } \mathbf{S}_{n-1}$. \blacksquare

Proposition 5 states that one can choose 2^ℓ contiguous Walsh-Hadamard codewords for pilots and get 2^ℓ uniformly spread DFT pilots with layout gap $D = 2^{n-\ell}$. The channel coding is akin CDMA, but the pilot layout matches the one used in OFDM communication. The lesson, is that the Walsh-Hadamard transform alone achieves “scrambling” of data followed by carrier mapping in the DFT domain in a fashion similar to SC-FDMA [41]. In SC-FDMA, the data are first “scrambled” by application of a shorter length DFT.

This result has a nice interpretation in the context of generalized Fourier transforms, the 2^n -WHT being itself the Fourier transform on the finite group $(\mathbb{Z}/2\mathbb{Z})^n$ instead of $\mathbb{Z}/2^n\mathbb{Z}$ for the classical 2^n -points DFT [37], [42]. A similar result holds for DFT on any toric finite group [37].

V. APPLICATION: FADING CHANNEL ESTIMATION IN MULTI-OUTPUT SYSTEMS

A. Channel model

1) *Physical assumptions:* A linear time-invariant channel is characterized by its impulse response h . In mobile communications, channels are transient, but we may assume the channel to be locally invariant around a time τ . This leads to the definition of a time-dependent channel impulse response h_τ . Consider a channel impulse response made of a large number L of echoes:

$$h_\tau(t) = \sum_{l=1}^L \alpha_l(\tau) \delta(t - t_l(\tau)). \quad (12)$$

The number of echoes L , is usually far too large to warrant a finite rate of innovation approach. However individual echoes aggregate in a smaller and manageable number of clusters K [43]. The rationales behind clustering are the same as for the common support assumption: a finite bandwidth combined with background noise allow only for a limited resolution. Table I lists a few examples for which clustering applies in typical operating conditions.

This simplification is at the heart of medium and narrow-band wireless communications [44]. We want to estimate h_τ by sending probes at the input and collecting samples at the output.

Correlation of the channel with respect to time is an important feature to exploit, however we will not consider it, as scheduling in modern communication systems makes its usage uncertain. Hence we settle on a time τ and drop it from the notation.

Communication is carried over a restricted frequency band, which is achieved by pulse-shaping with a template function $\varphi(t)$ and modulation by $e^{j\omega_c t}$. Applying clustering to (12) the channel impulse response becomes:

$$h(t) = \sum_{k=1}^K c_k \varphi(t - t_k) \quad (13)$$

$$\text{s.t. } c_k = a_k Z_k = e^{j\omega_c t_k} \sum_{(\alpha_l, t_l) \in \mathcal{C}_k} \alpha_l e^{j\omega_c (t_l - t_k)}, \quad (14)$$

where Z_k has unit-variance, a_k is the appropriate scaling parameter and \mathcal{C}_k is the k^{th} cluster. Assuming $\{\alpha_l e^{j\omega_c (t_l - t_k)}\}_{(\alpha_l, t_l) \in \mathcal{C}_k}$ contains i.i.d. elements with finite first two moments (echoes of finite energy)

$$\lim_{\# \mathcal{C}_k \rightarrow \infty} Z_k \sim \mathcal{N}_{\mathbb{C}}(0, 1).$$

This is the classical non line of sight fading scenario where the paths amplitudes $|Z_k|$ are independently Rayleigh distributed.

2) *Multipoint communications, one to many:* Communication through fading channels rely on spatial diversity to gain robustness. Spatial diversity is achieved with the deployment of several antennas at the receiver and/or transmitter. We describe a spatial channel model between one transmitter and several receivers, which generalizes to MIMO communications in a straightforward manner. The physical properties of the

channel are the following as shown in Figure 3:

- The distance in between antennas m and n is $d_{m,n}$.
- Each path is characterized by an angle of arrival (AoA) θ_k . To simplify computations it is assumed that the AoA is the same for all antennas (far field assumption). In the near field, a scatterer surrounds the receiver and the distribution becomes almost isotropic. Hence this assumption can be made for both regimes with limited error.
- The direction normal to the segment between antennas m and n points toward azimuth $\theta_{m,n}$.

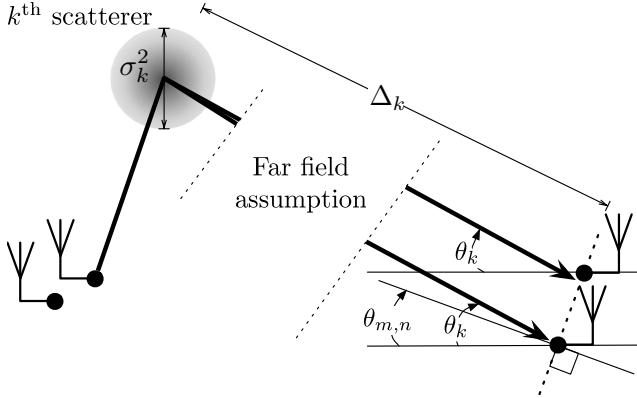


Fig. 3. Channel model with a single scatterer. Each scatterer is characterized by its apparent width (width/distance) σ_k/Δ_k and its azimuth θ_k . This model is considered valid in the near field as well, as the scatterer surrounds the receiver, thus having no intrinsic azimuth.

The channel model in (13) applies to the P subchannels

$$h_m(t) = \sum_{k=1}^K a_{m,k} Z_{m,k} \varphi(t - t_{m,k}), \quad m = 1, \dots, P. \quad (15)$$

We assume the distance in between antennas is smaller than the achievable spatial resolution, hence

$$t_{1,k} = t_{2,k} = \dots = t_{P,k}.$$

To fully characterize the channel, the path correlation across antennas must be known — by assumption $Z_{m,k}$ and $Z_{m',k'}$ are independent for $k \neq k'$. Following Salz and Winters [45] we derive a formula for the autocorrelation matrix of $\mathbf{Z}_k = [Z_{1,k} \dots Z_{P,k}]$. However we put a Gaussian prior on the cluster shape rather than a uniform one with discontinuities at the boundaries.

3) *Spatial correlation of paths*: As in [45], a large number of reflections are assumed to be drawn from a continuous probability distribution for each scatterer.

Proposition 6. *Under the spatial channel model described in Section V-A2, the antenna crosscorrelation is closely approximated by:*

$$\begin{aligned} \mathbb{E}[Z_{k,m} Z_{k,n}^*] = & J_0\left(\frac{\omega_c}{c} d_{m,n}\right) \\ & + 2 \sum_{l=1}^{\infty} j^l \frac{I_l(\kappa_k)}{I_0(\kappa_k)} J_l\left(\frac{\omega_c}{c} d_{m,n}\right) \\ & \cdot \cos\left[l\left(-\theta_{m,n} + \theta_k - \frac{\pi}{2}\right)\right], \end{aligned} \quad (16)$$

TABLE III
SIMULATION PARAMETERS

Parameter	Symbol	Value
Sampling step	T	50ns
Bandwidth	B	20MHz
Center frequency	f_c	2.6GHz
Frame duration (without padding)	τ	25.55μs
Samples per frame	N_{frame}	511
Pilots per frame	N	63
Pilot gap	D	8
Delay spread	Δ	1.6μs

where $\Delta_k^2/\sigma_k^2 \approx (1 - e^{-3\kappa_k/4})\kappa_k$, $J_{(\cdot)}$ is the Bessel function of the first kind and $I_{(\cdot)}$ is the modified Bessel function of the first kind.

Proof: See appendix A. It is only a close approximation since the azimuthal distribution at the receiver is approximated by a Von-Mises distribution. ■

Corollary 2. *For a path width κ_k large enough:*

$$\begin{aligned} \mathbb{E}[Z_{k,m} Z_{k,n}^*] \approx & J_0\left(\frac{\omega_c}{c} d_{m,n}\right) \\ & + 2\sqrt{2\pi\kappa_k} \sum_{l=1}^{\infty} j^l f_{\kappa_k}(l) J_l\left(\frac{\omega_c}{c} d_{m,n}\right) \\ & \cdot \cos\left[l\left(-\theta_{m,n} + \theta_k - \frac{\pi}{2}\right)\right], \end{aligned} \quad (17)$$

such that f_{κ_k} is the centered Gaussian pdf of variance κ_k .

The result is in its form similar to [45], however the derivation stays closer to the original physical model.

VI. NUMERICAL RESULTS

For simulations we use the channel model developed in Section V, and choose its parameters to *loosely* follow the 3GPP-LTE standard. Its characteristics are listed in Table III. We assume 63 pilots which are uniformly spaced in frequency, one every 8. The transmitted frame is circularly padded such as to guarantee circular convolution of the transmitted signal with the CIR. Results are derived from three different experiments:

- The medium has two paths separated by $2T$. The second path's expected amplitude is $1/10^{th}$ of the expected amplitude of the first path. The receiver possesses 1, 2, 4 or 8 uncorrelated antennas. The channels have exact SCS ($\varepsilon = 0$).
- The medium has two paths separated by T or $2T$. Both paths have the same expected amplitude. The receiver has 4 uncorrelated antennas. The channels have either exact SCS ($\varepsilon = 0$) or non-exact SCS ($\varepsilon = T/50 = 1\text{ns}$). The discrepancy in the ToA between antennas is uniformly distributed in $[-\varepsilon, \varepsilon]$. A time lapse of $2T/50$ corresponds to a path length difference of 60 cm.
- This experiment is more realistic from a physical standpoint. The receiver has 5 antennas equispaced on a circle

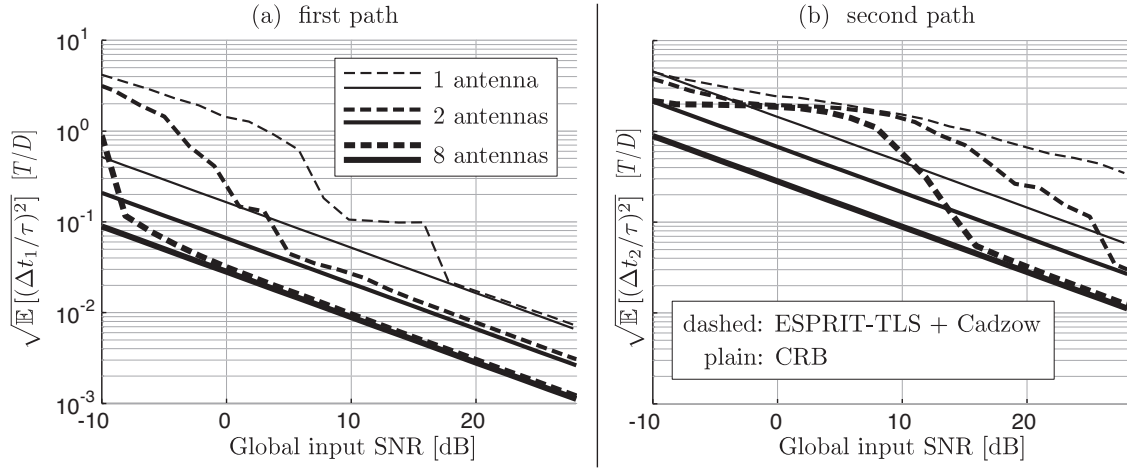


Fig. 4. (Exp. A) For the same global input SNR, a system with more antennas estimates the ToAs more accurately and is more resilient to noise. This is a consequence of the increased receiver diversity. The second path has $1/10^{\text{th}}$ the amplitude of the first path and is thus quickly buried into noise as SNR decreases. The estimation reaches the Cramér-Rao bound as long as it correctly identifies the path.

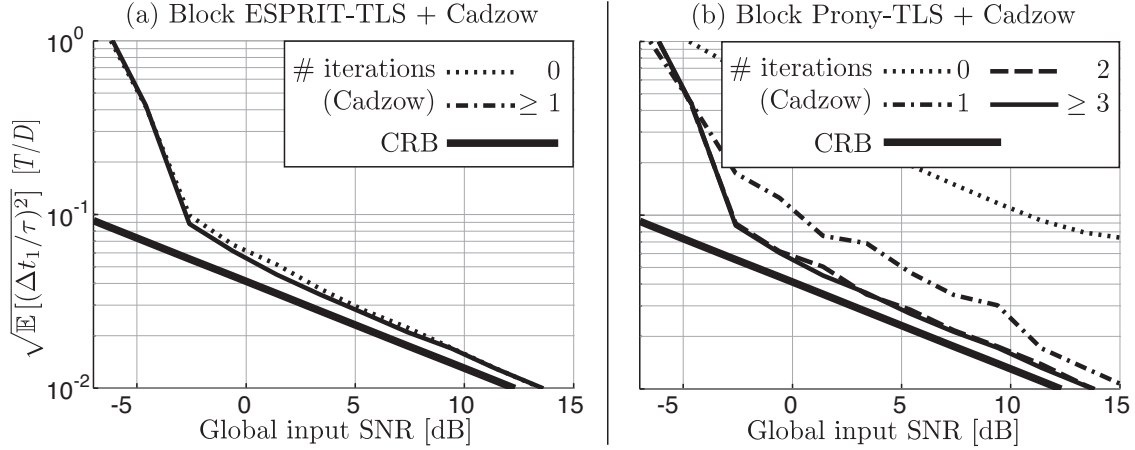


Fig. 5. (Exp. A) Part (a) shows the performances of Block ESPRIT-TLS with or without Block Cadzow denoising. In this setup, the gain obtained with the denoising is relatively small and is achieved after one iteration. Part (b) shows the performances of Block Prony-TLS with or without Block Cadzow denoising. As expected, the performance of Prony's algorithm without denoising is very poor. After 3 denoising iterations, performances of Block Prony-TLS and Block ESPRIT-TLS are indistinguishable.

of radius 10 cm. The propagation medium contains 4 scatterers (Figure ??(a)). The expected CIR modulus is represented in Figure ??(b). We use the spatial correlation model derived in Proposition 6, and provide the antennas cross-correlation in Figure ??(c). Also the channel is not exactly SCS, with a maximum delay $\varepsilon = T/50 = 1\text{ns}$.

Results were obtained on 400 independent noise and fading realisations.

4) *Results on Exp. A* : Figure 4 shows that the SCS-FRI algorithm efficiently estimates the ToA down to a certain SNR where the recovery breaks down. This breaking point is pushed lower as spatial diversity increases, which is to be expected. Figure 5 compares the use and combination of the various subspace identification techniques discussed earlier. The conclusion is that the performances of Block-ESPRIT TLS or Block-Prony TLS are exactly the same on a signal denoised with the Block-Cadzow algorithm. However Block-ESPRIT TLS requires fewer to none Block-Cadzow iterations

than Block-Prony TLS to reach the optimum. It is well-known that Prony TLS is not robust to noise [10], [29].

5) *Results on Exp. B* : Figure 6 shows that the single path CRB given in Proposition 3 is a good approximation of the true bound computed via Proposition 4 for multiple paths separated by more than twice the inverse bandwidth of the channel. This experiment also verifies the usefulness of the SCS assumption when ToAs are slightly perturbed from one antenna to another:

$$t_{k,p} = t_k + E_{k,p}, \quad E_{k,p} \sim \mathcal{U}([- \varepsilon, \varepsilon]), \text{ i.i.d.}$$

The error caused by the random perturbation $E_{k,p}$ is of the order of the perturbation itself, and thus we may say SCS-FRI is robust on non exact SCS channels.

6) *Results on Exp. C* : All estimation algorithms use the fact that the delay spread is much shorter than the frame length. The difference between lowpass interpolation and other techniques is the use of the sparsity property. Using this property alone, the SER is halved at a SNR of 5dB as shown in Figure 7. The addition of the SCS property proves to be

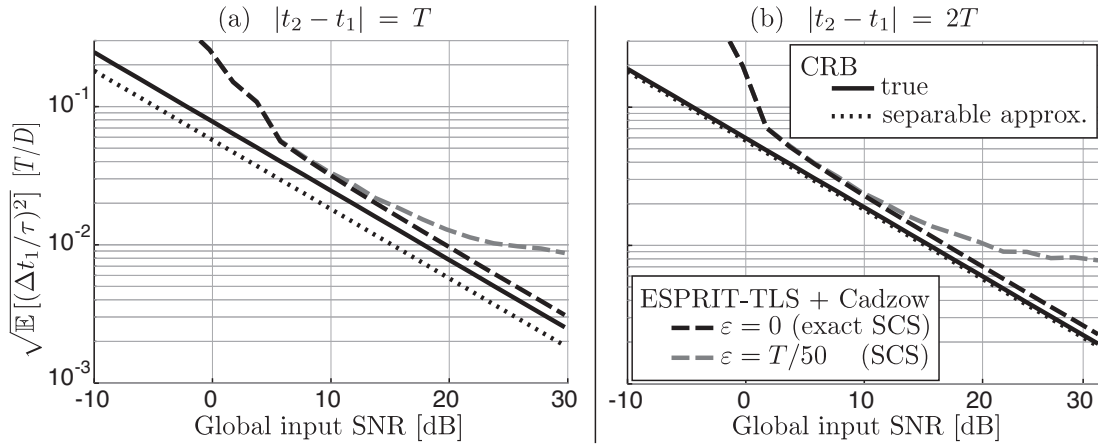


Fig. 6. (Exp. B) This figure shows that the proposed algorithms behave as expected in the presence of ToA mismatches between antennas. Part (b) motivates the separability assumption to compute the CRB of paths located more than $2T$ apart, while Part (a) shows its inadequacy for a smaller delay T . The “true” estimate is obtained via Monte-Carlo simulations.

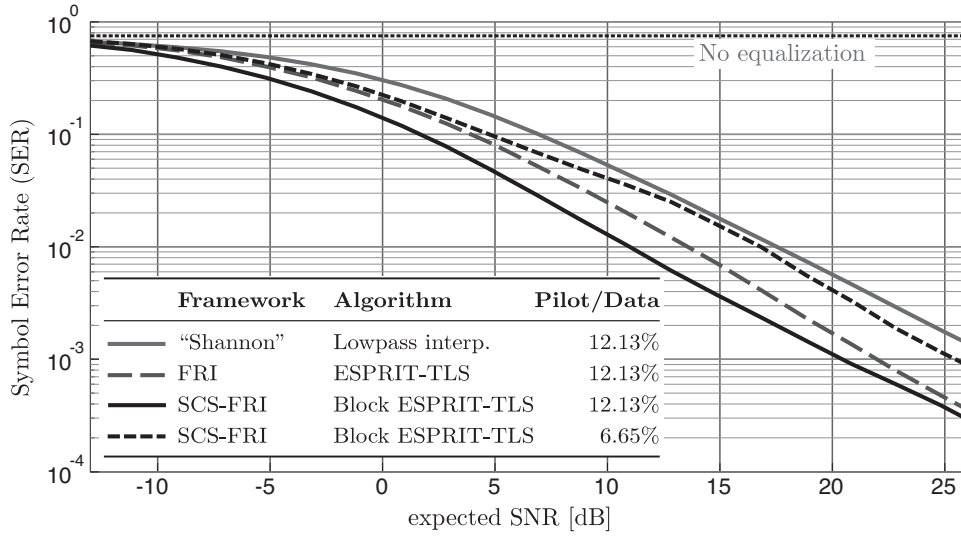


Fig. 7. (Exp. C) Using the SCS property, the SER is decreased by a factor 5 above 10dB of SNR compared to the conventional non-parametric approach. Sparsity alone provides a significant SER improvement, which shall be combined with the common support property below 30dB of SNR. At very high SNR, independent channel estimation across antennas become preferable as the channels only approximately have the common support property. However, below 15dB of SNR the effect of this approximation are undetectable. Another advantage of SCS-FRI is the reduction of pilots, it allows to halve their number while retaining performances superior to lowpass interpolation.

valuable, at 5dB of SNR the SER is decreased by a factor 3. At high SNR, the SCS property provides a factor 5 of improvement over lowpass interpolation. At very high SNR the error due to the approximate SCS nature of the channels diminishes this gain, and eventually the SCS assumption becomes detrimental.

It also shows that the number of pilots can be halved while having SER performances superior to the non-parametric approach (we retained half of the original pilots closest to the carrier frequency). For lowpass interpolation, this cannot be done without introducing aliasing. Reducing the number of pilots below “Nyquist” is relevant at high SNR where little redundancy is required for denoising, leaving some additional spectrum for data transmission. In favorable transmission conditions, it would be possible to reduce the number of pilots down to the rate of innovation of the channel to maximize the data throughput.

VII. CONCLUSION

We outlined the SCS-FRI algorithm, studied its performances on SCS channels estimation and computed theoretical lower-bounds for comparison. A spatial channel model was proposed for simulation purposes. The algorithm takes full advantage of the main properties of outdoor multipath channels and is directly applicable to most OFDM based communication standards. Simulations indicate that SCS-FRI based on the Block-ESPRIT TLS routine seems to be the most suitable since it requires only two partial SVD with size of the model order and provides optimal accuracy.

Future work is needed for estimation of the model order, incorporation of temporal correlation in the model and the algorithm (tracking of the model parameters). Computational complexity is also a crucial point for mobile applications, and improvements could be made with Krylov subspaces

techniques as in [46].

APPENDIX A

SPATIAL CORRELATION FORMULA FOR FADING CHANNELS

A. Azimuthal scatterers density distribution

The reflection density of each scatterer is normally distributed with mean $\boldsymbol{\mu}_k$ (its position) and covariance matrix $\sigma_k^2 \mathbb{I}$ (its “girth”). The number of reflections within a scatterer is assumed to be large enough to warrant their approximation by their continuous probability density function. The azimuthal density is the integral of the scatterer’s pdf over Γ_ϑ the straight path from the receiving antenna at an angle² ϑ :

$$p(\vartheta; \boldsymbol{\mu}_k, \sigma_k^2) = \int_{\mathbb{R}^2} f_{\sigma_k^2}^{(2D)}(\mathbf{x} - \boldsymbol{\mu}_k) \mathcal{I}_{\mathbf{x} \in \Gamma_\vartheta} d\mathbf{x}. \quad (18)$$

Reparametrization in polar coordinates yield:

$$\begin{aligned} p(\vartheta; \boldsymbol{\mu}_k, \sigma_k^2) &= f_{\sigma_k^2}(\|\boldsymbol{\mu}_k\| \sin(\vartheta)) \\ &\quad \cdot \int_{\mathbb{R}_+} f_{\sigma_k^2}(r - \|\boldsymbol{\mu}_k\| \cos(\vartheta)) J_{\mathbf{x}}(r, \vartheta) dr, \\ &= \sigma_k^{-1} f\left(\sqrt{\kappa'_k} \sin(\vartheta)\right) \\ &\quad \cdot \int_{\sqrt{\kappa'_k} \cos(\vartheta)}^{+\infty} \sigma_k^{-1} f(r - \|\boldsymbol{\mu}_k\| \cos(\vartheta)) \\ &\quad \cdot (s + \sqrt{\kappa'_k} \cos(\vartheta)) \sigma_k^2 ds \end{aligned}$$

such that $\kappa'_k = \|\boldsymbol{\mu}_k\|^2 / \sigma_k^2$ and $J_{\mathbf{x}}(r, \vartheta) = r$ is the Jacobian of the cartesian to polar transformation. We performed the change of variable $s = r - \sqrt{\kappa'_k} \cos(\vartheta)$. Hence, the distribution has only one degree of freedom, and after some calculus:

$$\begin{aligned} p_{\kappa'_k}(\vartheta) &= f(\sqrt{\kappa'_k} \sin \vartheta) \\ &\quad \cdot \left[\sqrt{\kappa'_k} \cos \vartheta \cdot F(\sqrt{\kappa'_k} \cos \vartheta) f(\sqrt{\kappa'_k} \cos \vartheta) \right]. \end{aligned} \quad (19)$$

The circular distribution (19) is well approximated by a Von-Mises distribution of scale κ_k :

$$q_{\kappa_k}(\vartheta) = \frac{e^{\kappa_k \cos \vartheta}}{2\pi I_0(\kappa_k)}. \quad (20)$$

where I_0 is the 0th order modified Bessel function of the first kind. Asymptotically, $\kappa'_k \xrightarrow{\kappa'_k \rightarrow \infty} \kappa_k$, and the approximation $\kappa'_k \approx (1 - e^{-3\kappa_k/4})\kappa_k$ was found to be empirically accurate for all κ_k (K-L divergence between $p_{\kappa'_k}$ and q_{κ_k} is less than 0.02 bits).

B. Derivation of the correlation matrix formula

Considering the setup of Figure 3, and from [45]:

$$R_Z^{(k)}[m, n] = \int_{-\pi}^{\pi} q_{\kappa_k}(\vartheta + \theta_{m,n} - \theta_k) e^{j \frac{\omega_c}{c} d_{m,n} \sin \vartheta} d\vartheta.$$

Then, q_{κ_k} is expanded in terms of spherical harmonics via the Jacobi-Anger expansion [47](9.1):

$$\begin{aligned} &q_{\kappa_k}(\vartheta + \theta_{m,n} - \theta_k) \\ &= \frac{1}{2\pi I_0(\kappa_k)} \left\{ J_0(-j\kappa_k) \right. \\ &\quad \left. + \sum_{l=1}^{\infty} j^l J_l(-j\kappa_k) \cos[l(\vartheta + \theta_{m,n} - \theta_k)] \right\}, \\ &= \frac{1}{2\pi} + \frac{1}{\pi I_0(\kappa_k)} \sum_{l=1}^{\infty} I_l(\kappa_k) \cos[l(\vartheta + \theta_{m,n} - \theta_k)], \end{aligned}$$

where the second equality is obtained with $I_l(jx) = j^l I_l(x)$ [47](9.6.3, 9.1.35).

We now have a series for $R_Z^{(k)}[m, n]$ with l^{th} term:

$$\begin{aligned} &\frac{I_l(\kappa_k)}{\pi I_0(\kappa_k)} \int_{-\pi}^{\pi} \cos[l(\vartheta + \theta_{m,n} - \theta_k)] e^{j \frac{\omega_c}{c} d_{m,n} \sin \vartheta} d\vartheta \\ &\stackrel{(a)}{=} \frac{I_l(\kappa_k)}{\pi I_0(\kappa_k)} \left\{ \cos \left[l \left(\theta_k - \theta_{m,n} - \frac{\pi}{2} \right) \right] \right. \\ &\quad \cdot \int_{-\pi}^{\pi} \cos l\vartheta e^{j \frac{\omega_c}{c} d_{m,n} \cos \vartheta} d\vartheta \\ &\quad \left. + \sin \left[l \left(\theta_k - \theta_{m,n} - \frac{\pi}{2} \right) \right] \right. \\ &\quad \left. \cdot \int_{-\pi}^{\pi} \sin l\vartheta e^{j \frac{\omega_c}{c} d_{m,n} \cos \vartheta} d\vartheta \right\} \\ &\stackrel{(b)}{=} \frac{2I_l(\kappa_k)}{I_0(\kappa_k)} I_l \left(j \frac{\omega_c}{c} d_{m,n} \right) \cos \left[l \left(\theta_k - \theta_{m,n} - \frac{\pi}{2} \right) \right] \\ &\stackrel{(c)}{=} \frac{2j^l I_l(\kappa_k)}{I_0(\kappa_k)} J_l \left(\frac{\omega_c}{c} d_{m,n} \right) \cos \left[l \left(\theta_k - \theta_{m,n} - \frac{\pi}{2} \right) \right] \end{aligned}$$

Equality (a) is obtained with some standard trigonometric identities and a shift by $-\frac{\pi}{2}$ of the variable of integration. Equality (b) follows from the standard integral representation of I_l ([47] 9.6.19). The second integrand is antisymmetric which leads the integral over the unit-circle to vanish. Finally (c) is a consequence of $I_l(jx) = j^l I_l(x)$ again. Hence:

$$\begin{aligned} R_Z^{(k)}[m, n] &= J_0 \left(\frac{\omega_c}{c} d_{m,n} \right) \\ &\quad + \frac{2}{I_0(\kappa_k)} \sum_{l=1}^{\infty} j^l I_l(\kappa_k) J_l \left(\frac{\omega_c}{c} d_{m,n} \right) \\ &\quad \cdot \cos \left[l \left(\theta_k - \theta_{m,n} - \frac{\pi}{2} \right) \right]. \end{aligned}$$

REFERENCES

- [1] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European trans. on telecommunications*, Wiley, vol. 10, no. 6, pp. 585–595, 1999.
- [2] G. Foschini and M. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless personal communications*, Springer, vol. 6, no. 3, pp. 311–335, 1998.
- [3] Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451 – 1458, 1998.
- [4] L. Zheng and D. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.
- [5] IEEE Computer Society, *IEEE 802.11n-2009 “IEEE Draft Standard for Information Technology — Telecommunications and Information Exchange between Local and Metropolitan Area Networks — Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC)*

²Without loss of generality the scatterer origin is at azimuth 0, and the antenna is located at position 0

- and Physical Layer (PHY) Specifications” Amendment: Enhancements for Higher Throughput., Computer Society Std., 2009.
- [6] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*. Oxford Ac. Press, 2008.
 - [7] B. Hochwald and B. Hassibi, “How much training is needed in multiple-antenna wireless links?” *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
 - [8] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, “Estimating Sparse MIMO Channels Having Common Support,” in *36th International Conference on Acoustics, Speech and Signal Processing*, 2011.
 - [9] A. Hormati and M. Vetterli, “Compressive sampling of multiple sparse signals having common support using finite rate of innovation principles,” *IEEE Signal Process. Lett.*, vol. 18, no. 5, p. 331, 2011.
 - [10] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, 2008. [Online]. Available: <http://infoscience.epfl.ch/record/130351/files/blu0801.pdf>
 - [11] S. Yau and Y. Bresler, “A compact Cramér-Rao bound expression for parametric estimation of superimposed signals,” *IEEE Trans. Signal Process.*, vol. 40, no. 5, pp. 1226 – 1230, 1992.
 - [12] T. S. Rappaport, *Principles of Wireless Communications*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
 - [13] ETSI, *300 744, DVB; Framing, channel coding and modulation for digital terrestrial television*, ETSI Std. 300 744, Rev. 1.4.1, 01 2001. [Online]. Available: <http://www.etsi.org>
 - [14] V. Garg, *IS-95 CDMA and CDMA2000*. Prentice Hall PTR, 2000.
 - [15] M. Rana, J. Kim, and W.-K. Cho, “Low complexity downlink channel estimation for LTE systems,” in *Int. Conf. on Advanced Communication Technology (ICACT)*, vol. 2, feb. 2010, pp. 1198 –1202.
 - [16] B. Karakaya, H. Arslan, and H. Curpan, “Channel Estimation for LTE Uplink in High Doppler Spread,” in *WCNC*, Las Vegas, NV, Apr. 2008.
 - [17] J. Paredes, G. Arce, and Z. Wang, “Ultra-Wideband Compressed Sensing: Channel Estimation,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 383–395, Oct. 2007.
 - [18] G. Taubock and F. Hlawatsch, “A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots,” in *ICASSP*, 2008.
 - [19] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, “Toeplitz Compressed Sensing Matrices With Applications to Sparse Channel Estimation,” *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010.
 - [20] P. Schniter, “Joint estimation and decoding for sparse channels via relaxed belief propagation,” in *Conf. Rec. Asilomar Conf. on Signals, Syst. & Computer*, Pacific Grove, CA, Nov. 2010.
 - [21] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *J. Royal Statist. Soc.*, vol. 68, pp. 49–67, 2006.
 - [22] P. Zhao, G. Rocha, and B. Yu, “The composite absolute penalties family for grouped and hierarchical variable selection,” *Ann. Stat.*, vol. 37, no. 6, pp. 3468–3497, 2009.
 - [23] A. C. Lozano, G. Świrszcz, and N. Abe, “Group Orthogonal Matching Pursuit for Variable Selection and Prediction,” in *NIPS*, Vancouver, Canada, Dec. 2008.
 - [24] J. Kim, W. Chang, B. Jung, D. Baron, and J. C. Ye, “Belief propagation for joint sparse recovery,” arXiv:1102.3289, Feb. 2011.
 - [25] M. Wakin, M. Duarte, S. Sarvotham, D. Baron, and R. Baraniuk, “Recovery of jointly sparse signals from few random projections,” in *Proc. Neural Information Processing Systems*. NIPS, 2005.
 - [26] M. Duarte, S. Sarvotham, D. Baron, M. Wakin, and R. Baraniuk, “Distributed compressed sensing of jointly sparse signals,” in *Asilomar Conf. on Signals, Systems and Computations*. Citeseer, 2005, pp. 1537–1541.
 - [27] K. Gedalyahu and Y. Eldar, “Time-delay estimation from low-rate samples: A union of subspaces approach,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3017–3031, 2010.
 - [28] R. Prony, “Essai experimental et analytique,” *J. Ecole Polytech. (Paris)*, vol. 2, pp. 24–76, 1795.
 - [29] D. Tufts and R. Kumaresan, “Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood,” *Proc. IEEE*, vol. 70, pp. 975–989, 1982.
 - [30] R. Roy and T. Kailath, “ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.
 - [31] A. Hormati, “Sensing and Recovery under Sparsity Constraints,” Ph.D. dissertation, EPFL, Lausanne, 2010. [Online]. Available: <http://library.epfl.ch/theses/?nr=4837>
 - [32] J. Cadzow, “Signal enhancement-a composite property mapping algorithm,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, Jan 1988.
 - [33] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.
 - [34] V. Prabhu and D. Jaliha, “An improved ESPRIT based time-of-arrival estimation algorithm for vehicular OFDM systems,” in *IEEE 69th Conf. on Vehicular Technology*. IEEE, 2009, pp. 1–4.
 - [35] H. Cramér, *Mathematical methods of statistics*. Princeton NJ: Princeton University Press, 1946.
 - [36] C. R. Rao, “Information and accuracy attainable in the estimation of statistical parameters,” *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.
 - [37] Y. Barbotin, “Joint estimation of multipath channels with FRI: Theory, model and bounds,” EPFL, Tech. Rep., 2010.
 - [38] D. L. Thomas, “Reciprocal moments of linear combinations of exponential variates,” *Journal of the American Statistical Association*, vol. 71, no. 354, pp. 506–512, June 1976.
 - [39] I. Maravic, “Sampling Methods for Parametric Non-Bandlimited Signals,” Ph.D. dissertation, EPFL, Lausanne, 2004. [Online]. Available: <http://library.epfl.ch/theses/?nr=3058>
 - [40] R. Price and P. Green, “A communication technique for multipath channel,” *Proc. of the IRE*, 1958.
 - [41] H. Myung, “Single Carrier Orthogonal Multiple Access Technique for Broadband Wireless Communications,” Ph.D. dissertation, Polytechnic Institute of New York University, 2007.
 - [42] A. Terras, *Fourier Analysis on Finite Groups and Applications*. Cambridge Univ. Press, 1999.
 - [43] G. Turin, “Communication through noisy, random-multipath channels,” Ph.D. dissertation, MIT, 1956.
 - [44] A. Molisch, *Wireless Communications*. Wiley, 2005.
 - [45] J. Salz and J. Winters, “Effect of fading correlation on adaptive arrays in digital mobile radio,” *IEEE Trans. Veh. Technol.*, vol. 43, 1994.
 - [46] Y. Barbotin, “Finite Rate of Innovation sampling techniques for embedded UWB devices,” Master’s thesis, EPFL, 2009.
 - [47] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.