

Fast and Robust Parametric Estimation of Jointly Sparse Channels

Yann Barbotin *Student Member*, and Martin Vetterli, *Fellow, IEEE*

Abstract—We consider the joint estimation of multipath channels obtained with a set of receiving antennas and uniformly probed in the frequency domain. This scenario fits most of the modern outdoor communication protocols for mobile access [1] or digital broadcasting [2] among others.

Such channels verify a Sparse Common Support property (SCS) which was used in [3] to propose a Finite Rate of Innovation (FRI) based sampling and estimation algorithm. In this contribution we improve the robustness and computational complexity aspects of this algorithm. The method is based on projection in Krylov subspaces to improve complexity and a new criterion called the Partial Effective Rank (PER) to estimate the level of sparsity to gain robustness.

If P antennas measure a K -multipath channel with N uniformly sampled measurements per channel, the algorithm possesses an $\mathcal{O}(KPN \log N)$ complexity and an $\mathcal{O}(KPN)$ memory footprint instead of $\mathcal{O}(PN^3)$ and $\mathcal{O}(PN^2)$ for the direct implementation, making it suitable for $K \ll N$. The sparsity is estimated *online* based on the PER, and the algorithm therefore has a sense of *introspection* being able to relinquish sparsity if it is lacking.

The estimation performances are tested on field measurements with synthetic AWGN, and the proposed algorithm outperforms non-sparse reconstruction in the medium to low SNR range (≤ 0 dB), increasing the rate of successful symbol decodings by $1/10^{\text{th}}$ in average, and $1/3^{\text{rd}}$ in the best case. The experiments also show that the algorithm does not perform worse than a non-sparse estimation algorithm in non-sparse operating conditions, since it may fall-back to it if the PER criterion does not detect a sufficient level of sparsity.

The algorithm is also tested against a method assuming a “discrete” sparsity model as in Compressed Sensing (CS). The conducted test indicates a trade-off between speed and accuracy.

Index Terms—Channel Estimation, Sparse, Finite Rate of Innovation, Effective Rank, Krylov Subspace.

I. INTRODUCTION

COMMUNICATIONS between two parties are subject to unknowns: noise and filtering by the Channel Impulse Response (CIR). With respect to decoding, the noise is treated as nuisance parameters, and the CIR coefficients as unknowns to be estimated as precisely as possible to maximize the decoding rate. For this purpose, the channel can be used to transmit a signal known at both ends — the *pilots* — to gain knowledge about the CIR. It dictates a trade-off between the portion of the channel reserved to the pilots — thus lost to *data*

Y. Barbotin and M. Vetterli are with the faculty of Informatics and Communications at École Polytechnique Fédérale de Lausanne, Switzerland

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*.

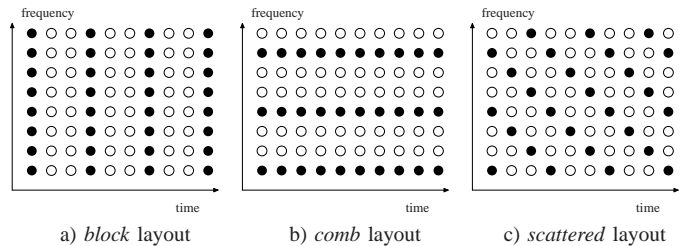


Fig. 1. In pilot assisted OFDM communications, time-frequency slots are reserved (black) for pilots, thus providing a sampling of the CIR in time and frequency.

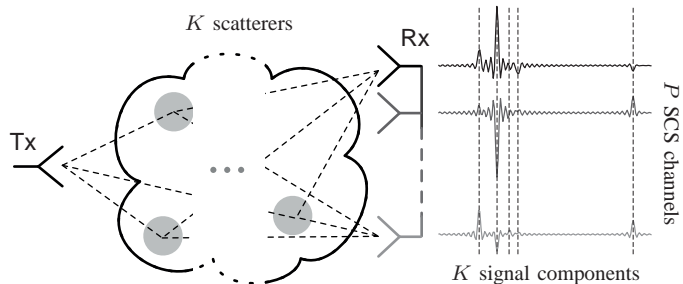


Fig. 2. The ideal SCS channel model is a set of P channels of bandwidth B each having K components aligned in time. Assuming complex valued signal components, the total number of unknowns is $(2P + 1)K$ instead of $3PK$ for a sparse model with independent time of arrivals (ToA), or $2P$ times the Nyquist Rate for a bandlimited model.

— and the decoding error rate due to bad channel estimation, both affecting the communication bitrate.

The two interdependent aspects of channel estimation are therefore the selection of pilots and the design of an estimation algorithm. We focus on the later and use uniform DFT pilot layouts shown in Fig. 1, which are found in modern communication standards using OFDM. The pilots provide information about the CIR, and so does an a priori knowledge about its structure.

In the noiseless case, the CIR can be perfectly recovered with a finite set of samples if it perfectly obeys the a priori known structure, thus providing a *sampling theorem* — e.g. uniform pilots in time at the Nyquist rate characterize uniquely bandlimited signals.

In this paper we study *Sparse Common Support* (SCS) channels, i.e. channels sharing a common structure of very low-dimension. Fig. 2 shows an example of SCS channels.

A. Problem definition

Imposing a structure may not lead to a trivial linear system of equations as it is the case for Shannon/Nyquist (projection

in a linear subspace). The SCS structure is a union of K unidimensional subspaces [4], [3] shared by P channels which leads to an estimation problem exactly and efficiently solvable in the framework of *Finite Rate of Innovation* sampling [5]. In [3] an estimation algorithm SCS-FRI is proposed and studied.

This leads to two questions. First, as seen in Fig. 3.a), the SCS-FRI algorithm selects the K subspaces from an infinite and uncountable set¹ which could be sampled. This is the approach taken by *Compressed sensing* (CS) or the *sparse representation* literature. The limiting case is to reduce the set to form a basis as shown in Fig. 3.b). The CIR can be perfectly and uniquely represented with elements of this set, but it probably won't have a K -sparse representation. A possible trade-off is to enlarge the set to form a *frame* as in Fig. 3.c). Reconstruction becomes more complex, but the shift-invariance provided by the frame [6] gives a better K -sparse representation.

Second, models are an approximation of reality — e.g. Fig. 4 — and it is not immediately clear which amount of modelization error can be tolerated in practice. This is especially important as we rely on a very specific signal structure, and this question can only be answered by trials on field measurements.

B. Contributions

With these considerations in mind, we address the estimation of *Sparse Common Support* (SCS) channels from DFT-domain measurements (pilots). SCS channels estimation with FRI was studied in a previous paper [3], and we hereby focus on computational issues and robustness. A fast algorithm is derived based on Krylov subspace projections. Its main advantage is to have a computation and memory cost proportional to the sparsity level K . It relies on FFT evaluations for the heavy load computations, which is particularly appealing for embedded DSP applications.

The sparsity level is unknown in practice, and we derive a heuristic estimate of it using a new measure called the Partial Effective Rank (PER). The PER tracks the “effective dimension” of the Krylov subspace as its size is increased, and can therefore be estimated online unlike other information criteria such as Akaike, MDL [7] or the EDC [8].

To assess the performances of the proposed algorithm, simulations are performed on measured CIRs to which synthetic AWGN is added. We compare the obtained results with an algorithm exploiting discrete sparsity to see if the input shift sensitivity described in Fig. 3 is a practical issue.

C. Outline

First, we will explicit the SCS channel model for aerial electromagnetic transmissions, and see under which conditions this model is relevant. Our conclusion is that the SCS property may not always be verified, which is confirmed by the data shown in Fig. 4. It establishes the requirements for a robust SCS estimation algorithm aware of the operating conditions.

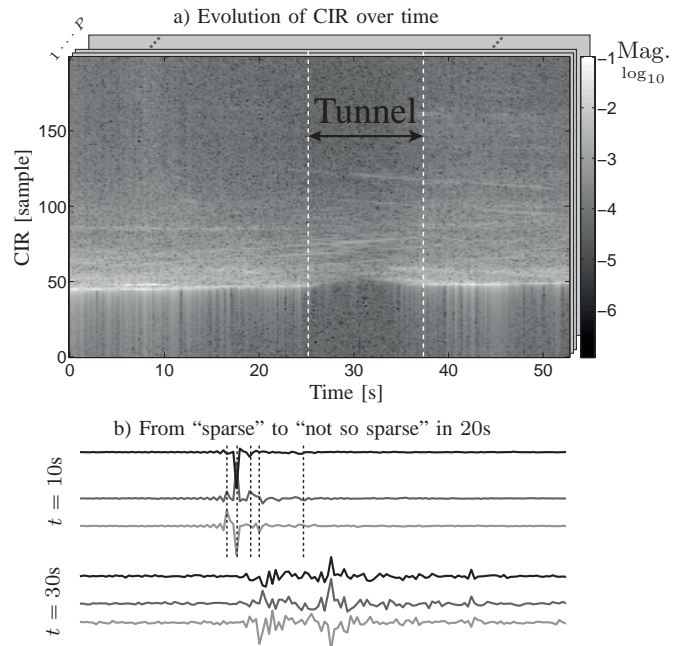


Fig. 4. Field measurements collected in [9]. The receiver is a base-station with $P = 8$ antennas, and the transmitter is mobile. The image a) shows the magnitude of the first antenna's CIR. The channel is qualitatively sparse except when the transmitter goes through a tunnel. The real part of the CIR for three different antennas is shown in b) confirming the common support property and the transient nature of sparsity.

Second, we quickly review the application of FRI and discrete sparse representations to the SCS estimation problem. Details are left out since they are already present in the literature [10], [3], [11].

The third part is devoted to the design of a fast and robust algorithm for the SCS estimation problem, based on Krylov subspace projections with $\mathcal{O}(KPN \log(N))$ operations, requiring $\mathcal{O}(KPN)$ memory, where K is the number of sparse components, P is the number of channels, and N is the number of collected samples per channel. For robustness, we introduce the *Partial Effective Rank* (PER) derived from the work of Roy [12], which only adds a marginal $\mathcal{O}(K^2)$ cost to be evaluated *online*. The PER is used to derive a heuristic estimate of K requiring little overhead. The heuristic may fail if the channel is not sparse enough, giving a beneficial sense of adequacy to the algorithm. In such non-sparse cases, the algorithm can yield to a non-sparse estimation method.

The resulting algorithm — combining FRI sampling, the PER criterion and Krylov subspace projection — named FRI-PERK was implemented in MATLAB and compares favorably to FRI for $N > 200$.

We conclude with the application of FRI-PERK and a CS algorithm (RA-ORMP) to field measurements collected in [9] with synthetic AWGN, and compare them with a simple non-parametric method (spectrum lowpass interpolation). The results show the SCS property can be exploited in the medium to low SNR bracket (below 0dB) to significantly lower the Symbol Error Rate (SER).

Compared with the non-parametric method, FRI-PERK increases the proportion of correctly decoded symbols by 1/10 in average, and 1/3 at best.

¹notwithstanding the limitations of machine precision.

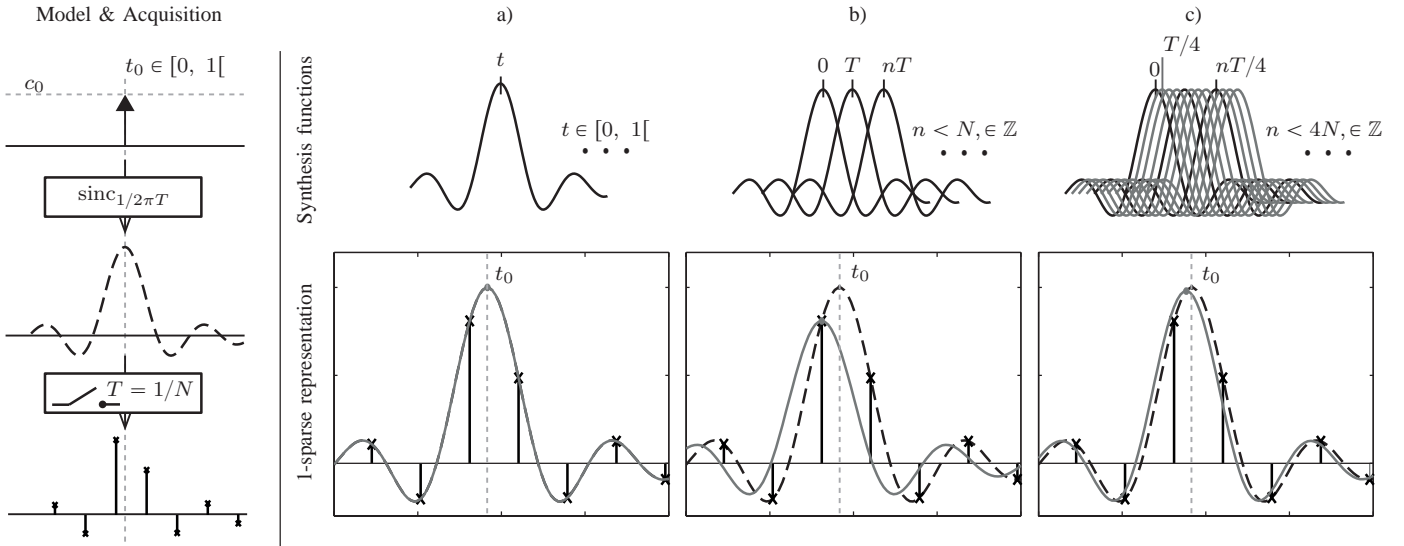


Fig. 3. Consider a single pulse (— —, dashed curve) critically sampled as shown on the left column. From the samples (\times , stems), 1-sparse representations of the original signal are computed. In a), the problem is treated as a parametric estimation of t_0 and c_0 as in the FRI framework. Conceptually, the signal component is chosen from the infinite and uncountable set of the pulse shape and its shifts in $[0, 1[$. The original signal has a perfect 1-sparse representation in this setup. In b), the signal component is chosen in a finite set of functions forming a basis of the signal space. The original signal can be represented by these functions, but it does not have a 1-sparse representation in general. In c), three times more synthesis functions are added to the set, to form a frame. The signal has a much closer 1-sparse representation in this frame thanks to the shift invariance introduced by the redundancy between the synthesis functions, but the estimation becomes combinatorially more complex. The estimation frameworks b) and c) are referred as *discrete sparsity* which is used in *Compressed Sensing* (CS), and the estimation is subject to a trade-off between accuracy and complexity.

II. THE PHYSICS OF SPARSE COMMON SUPPORT CHANNELS

A. Sparsity beyond simulations

Algorithms for the estimation of a sparse signal from noisy measurements are quite mature [13], [14], [3]. The principal hurdle for their application to physical signals is the inadequacy between the simple theoretical model and reality.

For example, indoor electromagnetic channels are in general not sparse, as described in the Saleh and Valenzuela model [15]. In this model, reflections are bundled in clusters having an exponentially decaying energy. This dynamic holds in general [16], [17], and after demodulation, an electromagnetic channel impulse response $h(t)$ can be described as the superposition of clustered reflections,

$$h(t) = \sum_{k=1}^K \sum_{(A_l, \Delta_l) \in \mathcal{C}_k} c_l \varphi(t - t_k - \Delta_l),$$

$$\xleftrightarrow{\mathcal{F}} \sum_{k=1}^K e^{-j\omega t_k} \sum_{(A_l, \Delta_l) \in \mathcal{C}_k} A_l e^{-j\omega \Delta_l} \hat{\varphi}(e^{j\omega}), \quad (1)$$

such that φ is the channel mask in the time domain (after demodulation), and \mathcal{C}_k are clusters of reflections with a delay t_k . Within each cluster we observe reflections shifted by Δ_l from t_k with randomly distributed complex-valued² amplitudes A_l .

The number of clusters K is usually small, but the total number of reflexions is not. However, if the bandwidth Ω_φ of φ and the maximal intra-cluster delays Δ_l are small enough, the 0th order approximation

$$e^{-j\omega \Delta_l} \hat{\varphi}(e^{j\omega}) \approx \hat{\varphi}(e^{j\omega}),$$

²After demodulation, the equivalent baseband channel becomes complex-valued.

holds at all considered frequencies $\omega \in]-\pi, \pi]$. If in addition the amplitudes are distributed identically and independently enough, one of the many avatars of the central-limit theorem may be used to obtain a simplified model known as the “Multipath Rayleigh channel model”:

$$h(t) = \sum_{k=1}^K C_k \varphi(t - t_k), \quad C_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, c_k^2 \mathbb{I}). \quad (2)$$

The channel is called *sparse* if and only if the expected delay-spread τ upper-bounding $t_K - t_1$ verifies

$$K/\tau \ll \Omega_\varphi/2\pi,$$

i.e. if the rate of innovation is substantially lower than the Nyquist rate.

The requirements for sparsity are thus two-folds:

- The “girth” of each cluster must be a fraction of the inverse bandwidth of the channel.
- The density of clusters must be a fraction of the channel bandwidth

The first property hints at channels with a low or medium bandwidth, while the second one requires long-distances propagation where both the transmitter and receivers are in a relatively clear environment, such as outdoor communications.

The high velocity of electromagnetic (EM) waves ensures scatterers of large dimensions, such as trees, generate clusters of modest time spread allowing the use of the simplified model (2) instead of (1) for channels with a bandwidth up to 100MHz approximately. Fig. 5 summarizes the conditions necessary for the channel to be sparse. The channel shown in Fig. 4 has a bandwidth of 120MHz, and an EM wave travels 2.5m in a time-lapse equal to the inverse-bandwidth. Therefore, in typical noisy operating conditions, reflections having

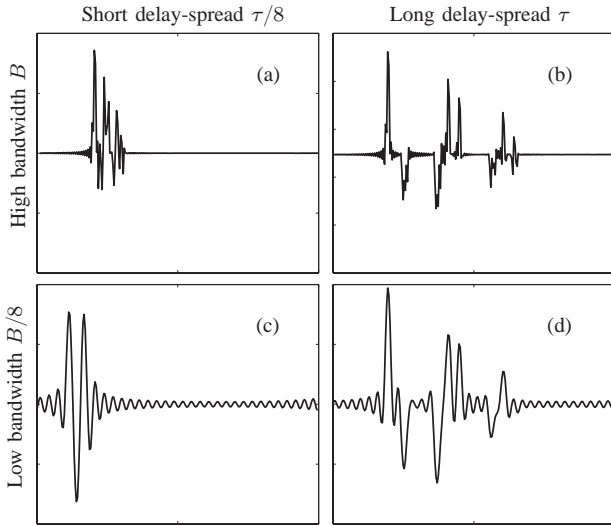


Fig. 5. This figure shows how sparsity relates to the channel bandwidth and its delay-spread. All four panels (a)–(d) have the same number of signal components — 80 of them grouped in 8 clusters with exponentially fast energy decay. Signals (a) and (c) cannot be considered sparse as the rate of innovation is close to/greater than the Nyquist rate on the time-lapse corresponding to the delay-spread, and 0 outside it. Signal (b) is weakly sparse, the rate of innovation is for this reason also high. In this setup the discrete sparsity approach may be suitable. The signal (d) can be considered sparse as only the 8 clusters will be resolvable in the presence of noise. The rate of innovation of this approximation is much lower than the Nyquist rate. Even though (b) and (d) have the same rate of innovation in a strict sense, (d) can be approximated with a signal having $1/10^{\text{th}}$ the rate of innovation of (b) thanks to its low bandwidth. This approximation motivates the use of a model with a low rate of innovation in the low-SNR regime where the model approximation error has less power than the noise.

a path difference of a fraction of 2.5m will be unresolvable.

On the other hand the ratio K/τ will be quite large necessitating long range outdoor transmissions for the channel to be considered sparse.

Fig. 5 shows the magnitude impulse responses of an outdoor channel with a medium bandwidth from the FTW³ MIMO dataset [9], and confirms this analysis.

Note that we do not rule-out the occurrence of sparsity in ultrawide-band communications as the channel dynamics become quite different at very short range [18].

B. Common support

We now consider the case where a receiver (Rx) possess several antennas, as in SIMO and MIMO communications. Therefore, if we consider a 1-to- P communication verifying the multipath model (2), the receiver observes P channels

$$h_p(t) = \sum_{k=1}^K C_{k,p} \varphi(t - t_{k,p}), \quad p = 1, \dots, P,$$

for a total of $3KP$ unknown coefficients parametrizing the channels.

As for sparsity, the high velocity of EM waves is again a crucial factor to establish the common support property, and

the difference in amplitudes

$$t_{k,1} \approx t_{k,2} \approx \dots \approx t_{k,P}, \quad (3)$$

$$C_{k,1} \neq C_{k,2} \neq \dots \neq C_{k,P}. \quad (4)$$

Indeed, if d_{\max} is the maximal distance between the P receiving antennas, the ToA difference between antennas is upper-bounded by $2d_{\max}/c$. The criterion for common support (3) is

$$d_{\max} \ll \pi \frac{c}{\Omega_\varphi}.$$

In Fig. 4, $d_{\max} = 60\text{cm}$ and $\pi \frac{c}{\Omega_\varphi} = 1.25\text{m}$.

To assert (4), we need to quantify the spatial correlation of paths amplitudes between the receiving antennas. Using [3] (Proposition 6), and assuming narrow scatterers, the correlation between antennas separated by a distance $d_{m,n}$ can be approximated as

$$\frac{\mathbb{E} [C_{k,m} C_{k',m}^*]}{\sqrt{\mathbb{E} [|C_{k,m}|^2] \mathbb{E} [|C_{k',n}|^2]}} = \delta_{k-k'} J_0 (d_{m,n} \omega_{\text{carrier}}/c).$$

Using the numbers of Fig. 5, correlation between antennas does not exceed 0.5. In most communication scenario (4) is verified by design as it provides spatial diversity [19].

In conclusion, we see that the common support property is relatively easy to establish as it only depends on the antennas topology. This is not the case for sparsity, in Fig. 4 when the transmitter enters a tunnel, a dense train of reflections is observed. Hence, it is reasonable to say outdoor communications with medium bandwidth have typically —as opposed to “surely”— sparse channel impulse responses.

III. ESTIMATION OF SPARSE COMMON-SUPPORT CHANNELS FROM DFT PILOTS

Assume each frame of period τ_f is uniformly sampled in time

$$x_p[n] = x_p(n\tau_f/N_f), \quad n = 0, \dots, N_f - 1.$$

The pilot layouts in Fig. 1 lead to estimate a set of sparse common support channels from a uniform subset of its DFT coefficients.

To make notation less cumbersome, we assume without loss of generality that $N_f = 2MD + 1$, and use negative indices which shall be understood “modulo the index limit” in general. Also the pilot index is centered on the DC carrier

$$\mathcal{P} = \{-MD, \dots, -2D, -D, 0, D, \dots, MD\}$$

and is supposed to take the value 1. The total number of pilots is $N = 2M + 1$.

Each frame is periodically padded to ensure the convolution with the CIR is circular, thus

$$\begin{aligned} \hat{\mathbf{x}}_p &= \text{diag}(\mathbf{1}_{\mathcal{P}}) \hat{\mathbf{h}}_p, \\ \Leftrightarrow \mathbf{x}_p &= \mathbf{W} \text{diag}(\mathbf{1}_{\mathcal{P}}) \mathbf{W}^H \mathbf{h}_p, \end{aligned}$$

such that \mathbf{W} is the DFT matrix, $\mathbf{1}_{\mathcal{P}}$ is the indicator of \mathcal{P} ,

³Forschungszentrum Telekommunikation Wien, <http://measurements.ftw.at>

and

$$\hat{h}_p[n] = \sum_{k=1}^K C_{k,p} W_{N_f}^{nt_k/\tau_s}. \quad (5)$$

is the sampled SCS channel model in the DFT domain.

The operator $\mathbf{W} \text{diag}(\mathbf{1}_P) \mathbf{W}^H$ is the orthogonal projection in the subspace spanned by the basis vectors of the DFT corresponding to the pilot indices.

With a good synchronization, the ToAs are contained in $]-\tau/2, \tau/2]$, and if

$$D \leq \left\lfloor \frac{\tau_s}{\tau} \right\rfloor, \quad D \in \mathbb{N}, \quad (6)$$

the original spectrum \hat{h}_p can be recovered by ideal lowpass interpolation

$$\mathbf{W}^H \mathbf{h}_p = \mathbf{C}_D \text{diag}(\mathbf{1}_P) \mathbf{W}^H \mathbf{x}_p, \quad (7)$$

such that \mathbf{C}_D is circulant with entries

$$[\mathbf{C}_D]_{m,n} = \frac{\sin(\pi(m-n)/D)}{\sin(\pi(m-n)/N_f)}.$$

This principle is at the core of OFDM communications which interleave data and pilots in the DFT domain.

If the measurements \mathbf{x}_p are corrupted by AWGN, this technique projects orthogonally the measurements in the signal subspace and therefore minimizes the estimation MSE if common-support property or sparsity are ignored.

A. FRI approach

The channels in (5) do not lie in linear subspaces, but in a common union of subspace. Therefore the channels can be estimated in two-steps

- 1) Identify jointly the K subspaces (the common support).
- 2) Compute the orthogonal projection of the measurements in the union of subspaces separately for each channel.

Algorithms and analysis for this problem are found in [3], and are a simple extension of well-known linear array-processing techniques such as ESPRIT [20] or the annihilating filter [21] to common-support channels.

The union of subspaces is identified by studying the column-space of the following data matrices

$$\mathbf{T}_p = \begin{bmatrix} \hat{h}_p[0] & \hat{h}_p[-D] & \hat{h}_p[-2D] & \cdots \\ \hat{h}_p[D] & \hat{h}_p[0] & \hat{h}_p[-D] & \cdots \\ \hat{h}_p[2D] & \hat{h}_p[D] & \hat{h}_p[0] & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

of dimensions $(M+1) \times (M+1)$

These data matrices have a Vandermonde decomposition

$$\mathbf{T}_p = \mathbf{J} \mathcal{V} \text{diag}(C_{p,1}, \dots, C_{p,K}) \mathcal{V}^H,$$

$$\text{such that } \mathcal{V} = \begin{bmatrix} 1 & \cdots & 1 \\ W^{Dt_1/\tau_s} & \cdots & W^{Dt_K/\tau_s} \\ W^{2Dt_1/\tau_s} & \cdots & W^{2Dt_K/\tau_s} \\ \vdots & \vdots & \vdots \end{bmatrix},$$

is an $(M+1) \times K$ Vandermonde matrix and \mathbf{J} is the exchange matrix. This decomposition has the merit to clearly show \mathbf{T}_p has a column-space of dimension K which depends only on the support, and is thus the same for each channel. Moreover any basis for this column-space verifies a *rotation invariance* property. Indeed, let

$$\mathcal{V}^\dagger \stackrel{\text{def}}{=} [\mathcal{V}]_{1:(M-1),:} \quad \text{and} \quad \mathcal{V}^\downarrow \stackrel{\text{def}}{=} [\mathcal{V}]_{2:M,:}.$$

Then

$$\mathcal{V}^\dagger = \mathcal{V}^\downarrow \mathbf{\Psi}, \quad \mathbf{\Psi} = \text{diag}\left(W^{Dt_1/\tau_s}, \dots, W^{Dt_K/\tau_s}\right).$$

Any basis \mathbf{V} having the same span as \mathcal{V} can be written as $\mathbf{V} \stackrel{\text{def}}{=} \mathcal{V} \mathbf{A}$, where \mathbf{A} is a fullrank $K \times K$ matrix, therefore

$$\begin{aligned} \mathbf{V}^\dagger &= \mathcal{V}^\dagger \mathbf{A}, \\ &= \mathcal{V}^\downarrow \mathbf{\Psi} \mathbf{A}, \\ &= \underbrace{\mathcal{V}^\downarrow \mathbf{A}}_{\mathbf{V}^\dagger} \underbrace{\mathbf{A}^{-1} \mathbf{\Psi} \mathbf{A}}_{\mathbf{X}}. \end{aligned}$$

which means the support is recovered from any basis \mathbf{V} of the column-space of \mathbf{T}_p as the phase of the eigenvalues of the solution to

$$\mathbf{V}^\dagger = \mathbf{V}^\downarrow \mathbf{X},$$

which is the ESPRIT algorithm [20]. In the presence of AWGN, the basis \mathbf{V} may be obtained in a robust manner from the SVD of the stacked matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_P \end{bmatrix}, \quad (8)$$

but other subspace identification techniques may be used depending on the measurements model.

To compare with the canonical estimation technique (7), this method not only uses the limited-length of the delay-spread but also sparsity and common-support. The solution of the estimation problem is to be found in a smaller set of candidates.

We expect that with good algorithms the estimation will be more resilient to noise. On the other hand more restrictive models yield higher modelization error.

Note that sparsity is not used to reduce the number of pilots but to make the estimation more robust to noise. See [3] for pilot reduction.

B. CS approach

The measured DFT samples $\hat{\mathbf{x}}_p$ corresponding to the set of pilot subcarriers \mathcal{P} are linked to the channel impulse response coefficients by

$$\mathbf{X} = [\mathbf{W}]_{\mathcal{P},\mathcal{C}} \mathbf{H}, \quad (9)$$

$$\mathbf{X} \stackrel{\text{def}}{=} [[\hat{\mathbf{x}}_1]_{\mathcal{P}}, \dots, [\hat{\mathbf{x}}_P]_{\mathcal{P}}], \quad \mathbf{H} \stackrel{\text{def}}{=} [\mathbf{h}_1, \dots, \mathbf{h}_P],$$

where \mathcal{C} is the index set on which the channel impulse responses are supported a priori. \mathcal{C} contains contiguous indices covering a time-lapse less than or equal to the delay-spread.

The matrix of channel coefficients \mathbf{H} is assumed to be *jointly row-sparse*, i.e. only a few rows of \mathbf{H} are not null.

This is known as the MMV (Multiple Measurement Vectors) problem which is not a trivial extension of the SMV (Single Measurement Vector) problem. It has received a lot of attention in the past few years and good, efficient algorithms exist such as RA-ORMP [13], Lee & Bressler [14],...

First notice that the channel model described in (5) is not jointly row-sparse nor sparse, unless the ToAs coincide with the sampling points.

Second, since (6) holds, the system (9) is invertible, which is typically not the case in compressed sensing. In this study sparsity is used as a regularization technique to gain robustness to noise as in Section III-A. One could select a random subset of pilots to effectively compress the channel measurements leaving additional space for data.

The compressed sensing literature on sparse channel estimation has followed two tracks

- 1) The ToA coincide with the sampling grid.
- 2) The ToA do not coincide with a regular grid resulting in “approximately sparse” channels [11], [10], [22]. It introduces a model mismatch which can be mitigated using frames (increased estimation complexity).

The first assumption is remote from reality and the reported performances shall be taken with a grain of salt, nevertheless it is useful to benchmark algorithms. The second one is more relevant and its limitations due to model mismatch may fade away when applied to real-world data where it is unavoidable anyway.

IV. APPLICATION OF THE FRI APPROACH: PERK

In this section, we consider the number of signal components K and the number of antennas P to be small compared to the number of pilots N .

The FRI based channel estimation technique outlined in Section III-A has two shortcomings if implemented in a straightforward manner

- 1) Its computational complexity and memory footprint are respectively $\mathcal{O}(N^3)$ and $\mathcal{O}(N^2)$. Both are contributed by the SVD decomposition used to estimate the column-space of \mathbf{T} .
- 2) The sparsity level K is unknown.

The complexity 1) is especially important for channel estimation as it is a core signal processing block at the receiver’s physical layer. It is called on several times per second, and must operate in real-time with limited power and hardware resources.

A. An $\mathcal{O}(KPN \log N)$ FRI estimation

Naive computation of an SVD of \mathbf{T} defined in 8 to obtain a K -dimensional subspace of the column-space is wasteful for two reasons. First, only K out of $M+1$ principal singular pairs (σ_m, \mathbf{v}_m) are of interest. Second, \mathbf{T} is well structured — made of Toeplitz blocks — and matrix factorization techniques such as QR will destroy this structure during the factorization process, rendering it unexploitable and requiring an explicit storage of the data matrix.

Since we are interested in the column-space of \mathbf{T} , we will work on the hermitian symmetric *correlation matrix*

$$\mathbf{T}^H \mathbf{T} = \sum_{p=1}^P \mathbf{T}_p^H \mathbf{T}_p,$$

which has eigenpairs $(\sigma_m, \mathbf{v}_m, m = 0, \dots, M)$.

A solution to compute only the leading eigenpairs, is to project the correlation matrix in a *Krylov subspace* [23]. This is an iterative method in which computations are performed on the original matrix, meaning the original structure is preserved. Hence the memory footprint is kept low and the computational complexity is similar for each iteration.

Krylov approximants have received a lot of attention in the numerical analysis [23], [24], [25], [26] or control theory [27], [28] literature, therefore we will only quickly skim through the subject.

Projection into a *Krylov subspace* is done with *Lanczos algorithm*. A comprehensive analysis was done by Xu [29] for the estimation of covariance matrices in linear array processing. He proposed an $\mathcal{O}(N^2)$ algorithm⁴ together with an approximate $\mathcal{O}(N^2)$ estimation of the subspace dimension K . Implementation of the Lanczos algorithm is quite involved in practice and may require costly corrections at each iteration [26], [23], [24], this is why asymptotically more expensive $\mathcal{O}(N^3)$ are generally preferred unless the system is very large.

The additional structure on the original data matrix \mathbf{T} allows us to lower the complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$, making it appealing even for matrices of modest size, and we will derive a novel criterion to estimate the signal subspace dimension K which requires $\mathcal{O}(K^2)$ computations to be run along the subspace estimation process.

A K -dimensional Krylov subspace \mathcal{K} of a M -dimensional hermitian matrix \mathbf{A} has is

$$\mathcal{K}_{K,\mathbf{f}}(\mathbf{A}) = \text{span}_{k=1,\dots,K} \mathbf{A}^k \mathbf{f},$$

where \mathbf{f} is an initial vector which can be randomly chosen. Note that all the properties of Krylov subspaces only hold in probability, an “unlucky” initial draw may compromise them.

The k^{th} basis vector $\mathbf{A}^k \mathbf{f}$ is a monomial of \mathbf{A} of degree k , therefore using a three terms linear recursion on this sequence of monomials one can derive a sequence of orthogonal polynomials [33]. This is equivalent to say that an orthonormal basis \mathbf{Q}_K of $\mathcal{K}_{K,\mathbf{f}}(\mathbf{A})$ is computed by orthogonalization of each of $\mathbf{A}^k \mathbf{f}$ only against the two previous ones and normalization. So, the main cost of the procedure is the computation of the non-orthogonal basis vectors, which is done by recursive matrix-vector multiplications.

The three terms recursion used to orthogonalize the Krylov basis implies that $\mathcal{K}_{K,\mathbf{f}}(\mathbf{A})$ has a tridiagonal decomposition

$$\mathcal{P}_{\mathbf{f},K} \mathbf{A} = \mathbf{Q}_K^H \mathbf{\Gamma}_K \mathbf{Q}_K.$$

where \mathbf{Q}_K is unitary and $\mathbf{\Gamma}_K$ is symmetric and tridiagonal (thanks to the 3-terms recursion). The eigenpairs of $\mathcal{P}_{\mathbf{f},K} \mathbf{A}$ are derived from this factorization at little cost [23], [24], and they are called the *Ritz pairs*.

⁴ K is considered small and constant compared to N and $P = 1$.

TABLE I
APPROXIMATIVE “ \mathcal{O} ” COMPLEXITY FOR SUBSPACE IDENTIFICATION.

Algorithm	Main computation	Storage	Latency	Processing units (pu)
Krylov + PER	$KPN \log N$	$K(N + 1)$	KN [30]	$P \times$ FFT engines ($N + 1$ points)
Full SVD (serial)	PN^3	PN^2	PN^3	1 multipurpose pu.
Full SVD (systolic array) [31], [32]	PN^3	PN^2	$N(\log N + P)$	$N^2 \times$ 2-by-2 SVD pu.

The full SVD is done with Jacobi rotations and can be massively parallelized using the systolic array method of Brent, Luk & Van Loan [31]. Parallelism greatly reduces the latency of the system, but since it does not reduce the number of computations it comes at the cost of using multiple processing units.

A remarkable property is that the Ritz pairs quickly converge to the principal eigenpairs of \mathbf{A} as K grows. This quick convergence does not follow exactly the result of Xu [29] — since in our case the Toeplitz blocks are of approximately square size — but it can be explained with a theorem of Saad [23] (Theorem 12.4.1) which links the rate of convergence of the Ritz pairs to the growth rate of Chebyshev polynomials. The theorem shows that Ritz pairs converge faster to the corresponding eigenpairs if the eigenvalues are farther apart.

Because of the Toeplitz structure of the data matrix, matrix-vector multiplications with $\mathbf{T}^H \mathbf{T}$, which is the central step of a Lanczos iteration has a cost of $\mathcal{O}(PN \log(N))$. Indeed,

$$\mathbf{T}^H \mathbf{T} \mathbf{f} = \sum_{p=1}^P \mathbf{T}_p^H \mathbf{T}_p \mathbf{f}$$

is the sum of P matrix-vector multiplications, each of them realized as two consecutive Toeplitz matrix-vector multiplications. Square Toeplitz matrices of dimension $M+1$ can be embedded in circulant matrices of dimension $2(M+1) = N+1$

$$\mathbf{C}_p \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{T}_p & \bar{\mathbf{T}}_p \\ \bar{\mathbf{T}}_p & \mathbf{T}_p \end{bmatrix},$$

$$\mathbf{T}_p = \text{toeplitz}(t_{p,-M}, \dots, t_{p,0}, \dots, t_{p,M}),$$

$$\bar{\mathbf{T}}_p = \text{toeplitz}(t_{p,1}, \dots, t_{p,M}, 0, t_{p,-M}, \dots, t_{p,-1}).$$

Circulant matrices are diagonalized by the DFT matrix, hence the cost of a circulant matrix-vector multiplication is dominated by the cost of 4 FFT.

Since

$$\begin{bmatrix} \mathbb{I}_M & \mathbb{O}_M \\ \mathbb{O}_M & \mathbb{I}_M \end{bmatrix} \mathbf{C}_p^H \begin{bmatrix} \mathbb{I}_M & \mathbb{O}_M \\ \mathbb{O}_M & \mathbb{I}_M \end{bmatrix} \mathbf{C}_p \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} = \mathbf{T}^H \mathbf{T} \mathbf{f},$$

each Lanczos iteration has a cost dominated by $4P$ FFT⁵ of length $N+1$. Since only the first half of the input and the first half of the output of each FFT are non 0 or not needed, each FFT could be replaced by two FFT of half the size.

The resulting algorithm fulfill all the requisites for a fast embedded implementation since most of the computational load is put on the FFT block, ubiquitous in communication systems. Table I

B. On-line sparsity assessment

In this section we introduce the *Partial Effective Rank* (PER) a criterion to estimate the signal subspace dimension

⁵The DFT of the circulant generators can be precomputed.

working online with the Lanczos algorithm. Its main advantage compared to other methods is to require little information about the noise space. A formal study of the PER is deferred to an upcoming report.

1) *Shortcomings of traditional information criterions:* Information theoretic criteria such as Rissanen’s MDL [7], Akaike criterion or the EDC [34] are powerful mathematical tools which may be used to evaluate the sparsity level K . They all follow a similar pattern, which is to minimize

$$\text{ITC}(\boldsymbol{\sigma}, K) = \mathcal{L}(\boldsymbol{\sigma}, K) + K \cdot (2(M+1) - K) \cdot \text{P}(M),$$

where \mathcal{L} is the log-likelihood function based on $\boldsymbol{\sigma}$ the singular values of \mathbf{T} , and K an estimate of the sparsity level. The term P is a penalty growing at rate between $\mathcal{O}(1)$ and $o(N)$ depending on which one of the criteria is used.

Their performances are well understood [35], but the evaluation of the likelihood function requires to compute the product

$$\prod_{m=K}^M \sigma_m^2 = \det(\mathbf{T}^H \mathbf{T}) \left/ \prod_{k=0}^{K-1} \sigma_k^2 \right.,$$

which has an algorithmic cost superior to the Lanczos algorithm itself.

Also, the argument used in [36] to show consistency of a partial evaluation of $\text{ITC}(\boldsymbol{\sigma}, K)$ cannot be used in our setup because the asymptotic distribution of the noise matrix spectrum⁶ is extremely different. As the number of pilots increases, the probability measure of the noise spectrum concentrates [37], but its support is unbounded. This is a major difference with the setup considered in [36], [29] where the noise power spectrum concentrates around the noise variance. Formally we can say

Proposition 1: Let \mathbf{T} be an $(M+1)P \times (M+1)$ matrix composed of fixed and finite number P of stacked Toeplitz blocks \mathbf{T}_p of square dimensions $(M+1) \times (M+1)$ as in (8). Assuming the generators of each block are sequences of white Gaussian noise

$$\sigma_{\max}(\mathbf{T}^H \mathbf{T} / (M+1)) = \mathcal{O}(\log M),$$

i.e. even with proper normalization, the spectral norm of an all-noise matrix diverges.

Proof: The matrix norm induced by the euclidian vector

⁶“spectrum” shall be understood as the SVD spectrum of the matrix and not the noise Fourier power spectrum when matrices are concerned.

norm on $\mathbb{C}^{(M+1)P}$ is submultiplicative, therefore

$$\|\mathbf{T}_1\| \leq \left\| \begin{bmatrix} \mathbb{I}_{M+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{T} & \mathbf{0} \end{bmatrix} \right\| \leq 1 \cdot \|\mathbf{T}\|. \quad (10)$$

We then use a result of Meckes [38] on the spectral norm of square Toeplitz matrices with independently distributed subgaussian generating coefficients

$$\|\mathbf{T}_1\| \sim \mathcal{O}\left(\sqrt{M \log M}\right). \quad (11)$$

Using the triangle inequality and Meckes' result

$$\|\mathbf{T}^H \mathbf{T}\| \leq \sum_p \|\mathbf{T}_p^H \mathbf{T}_p\| \sim P\mathcal{O}(M \log M),$$

which together with (10) and (11) proves the proposition. \blacksquare

It indicates that for pure AWGN measurements the spectrum of the matrix is unbounded, and one cannot expect the signal and the noise spectrum to separate nicely. Without a proper approximation, the computation of $\prod_{m=K+1}^M \sigma_m^2$ would have a cost of $\mathcal{O}(PN^2 \log(N))$ in general⁷, driving the complexity of the algorithm.

2) *The Partial Effective Rank*: The effective rank, is a matrix functional introduced by Roy [12] which may be seen as a ‘‘convexification’’ of the rank.

Definition 1: Let \mathbf{A} be a matrix with singular values $\boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_M]^T$ in decreasing order, and singular values distribution

$$p_m = \sigma_m / \|\boldsymbol{\sigma}\|_1, \quad m = 1, \dots, M.$$

The *Effective Rank* of \mathbf{A} is

$$\text{erank}(\mathbf{A}) = e^{\mathcal{H}(p_1, \dots, p_M)},$$

where \mathcal{H} is the entropy of the singular values distribution

$$\mathcal{H}(p_1, \dots, p_M) = - \sum_{m=1}^M p_m \log_e p_m.$$

For elementary properties of the effective rank, we defer to [12].

We can now introduce the partial effective rank

Definition 2: For \mathbf{A} and \mathcal{H} as in Definition 1 and

$$p_{K,k} = \sigma_k / \|\boldsymbol{\sigma}_{1:K}\|_1, \quad k = 1, \dots, K \leq M,$$

the *Partial Effective Rank* (PER) is

$$\text{PER}_K(\mathbf{A}) = e^{\mathcal{H}(p_{K,1}, \dots, p_{K,K})}.$$

The PER verifies

Proposition 2:

$$0 \leq \text{PER}_{K+1}(\mathbf{A}) - \text{PER}_K(\mathbf{A}) \leq 1.$$

The lower bound 0 is reached if and only if $\sigma_{K+1} = 0$ and the upper bound 1 if and only if

$$\sigma_1 = \sigma_2 = \dots = \sigma_{K+1}.$$

⁷Using a block-Levinson recursion to estimate the determinant \mathbf{T} , it maybe brought down to $\mathcal{O}((PN)^2)$, but it was not verified.

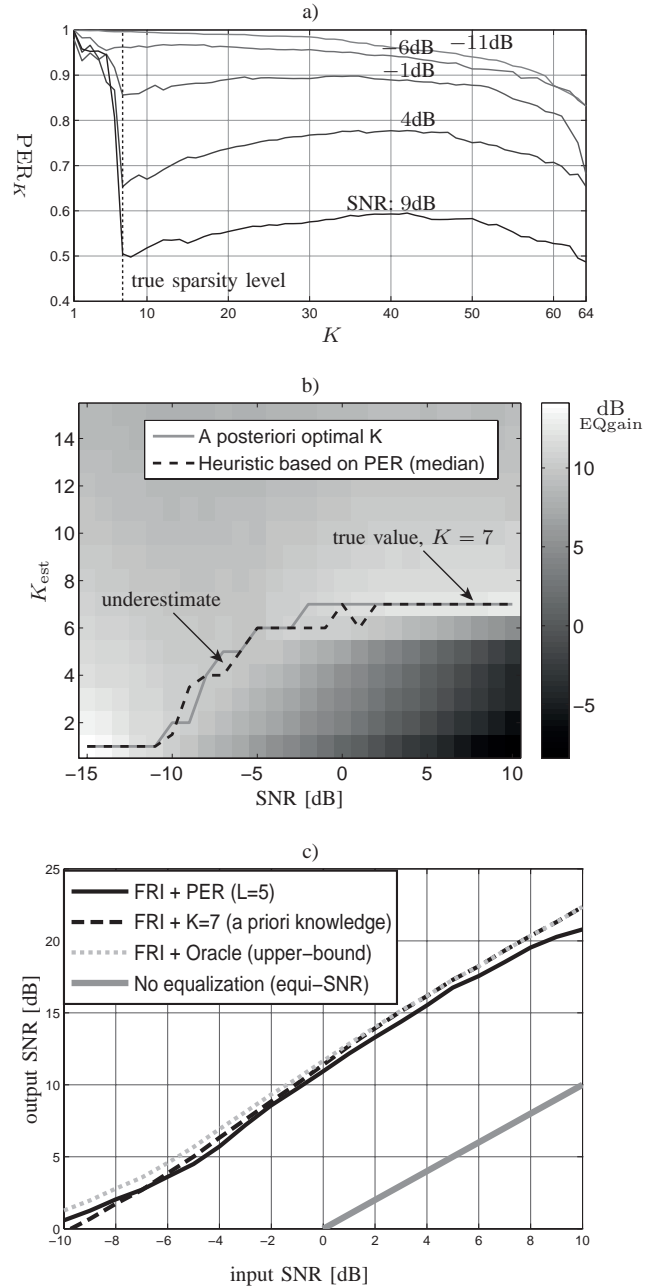


Fig. 6. Simulation on a signal with 7 components. The PER curves in (a) show a clear inflection at $K = 7$. As the SNR diminishes, the inflection occurs at lower values of K and completely disappears at SNRs < -10 dB. The graph (b) empirically shows that the PER indicates the number of components significantly above noise level which can be reliably estimated: the value of K it predicts follows (in median) the one minimizing the channel estimation error (energy of the residual). Finally, in (c) the estimation performances obtained with the PER closely follow the one obtained with an oracle choosing a posteriori the optimal number of components. We see also that below -5 dB the underestimation of K by the PER is beneficial compared to the true value 7. At high SNR, the PER sometimes overestimate the number of paths, mitigating the performances.

The increase of the partial effective rank with K reflects how ‘‘significant’’ is the K^{th} principal component of \mathbf{A} compared to the previous ones.

Fig. 6 provides empirical evidences that the evolution of the PER during the Lanczos process provides a suitable criterion to estimate K .

We settle on a very simple heuristic to estimate K based on the PER. We choose the smallest K such that

$$\text{PER}_K(\mathbf{T}) - \frac{1}{L} (\text{PER}_{K+1}(\mathbf{T}) + \dots + \text{PER}_{K+L}(\mathbf{T})) \leq 0,$$

which is a very simple “positive slope” detector. It introduces a small overhead: $K + L$ dimensions are required to decide if the signal space is of dimension $\leq K$.

Despite its simplicity and its heuristic motivations, this criterion proved to be robust in practice as shown in Fig. 6.

Too many components and/or large modelization error indicate the channels are not sparse.

V. NUMERICAL RESULTS

A. accuracy: FRI vs CS vs lowpass

1) *Setup*: We use the “FTW rural” dataset to compare aforementioned algorithms on the SCS channel estimation problem. The transmitter is a mobile single antenna device⁸. The receiver is a “base-station” with $P = 8$ receiving antennas. A total of 251 frames are transmitted on a 120MHz wide channel with a 2GHz carrier frequency. The transmission has a high SNR, so we consider the samples to be the ground truth (infinite SNR). Various SNR conditions are simulated with the addition of AWGN to the samples.

The DFT pilots are uniformly laid-out every $D = 3$ DFT bin. The estimated channel is used to demodulate 4-PSK coded data symbols occupying the left over DFT bins, and the obtained *Symbol Error Rate* is the quality metric used to benchmark the estimation.

2) *Interpretation of the results*: From Fig. 7 we may conclude that

- The channels do not exactly fit the SCS model, therefore the modelization error becomes larger than the noise at high SNR
- The SCS property helps in lowering the symbol error rate at medium to low SNR (below 0 dB)
- The “sparsity” model assumed by FRI (few reflections) match the field measurements better than the one assumed by CS (few non-0 coefficients) as seen in Fig. 8.
- Any algorithm exploiting sparsity must be “*introspective*”, i.e. it must detect when sparsity does not occur, and fall-back to a more traditional method whenever it happens. It is exemplified by the stroll through the tunnel.

B. Computational benchmark

Both FRI-PERK⁹ and RA-ORMP were implemented in MATLAB¹⁰. Because of the decimation in frequency and the limitation of the delay-spread in time, the projection in RA-ORMP cannot be realized with simple FFTs, results are reported in Fig. 9.(c-d).

⁸Only the first Tx antenna is used.

⁹Uses the ARPACK library [39]. Since the size of the Krylov subspace dimension must be fixed beforehand, it does not have the ability to stop before K_{\max} is reached.

¹⁰v. R2011b for MacOS X, running on a 1.8GHz Intel Core I7 processor and enough 1.3 GHz DDR3 memory

If this is not the case, we implemented a fast version of it and reported results in Fig. 9.(a-b). Note that such a setup does not apply to most standards [1], [2].

Definitive conclusions for hardware implementations cannot be drawn from these experiments, however we can see that the low asymptotic complexity of FRI-PERK is not misleading since it provides improvements for a number of pilots which are encountered in practice [1], [2] (higher bandwidth modes). For small numbers of pilots, the direct implementation may be faster, though we did not use the tracking capacity of Lanczos algorithm used by choosing an initial vector lying in the signal space obtained at the previous step [36].

VI. CONCLUSION

In this study, we addressed computational and robustness issues of FRI techniques applied to channel estimation. The tests conducted on field measurements show the SCS assumption is relevant at medium to low SNR where the noise power exceeds the one of model mismatch, and can be used to substantially lower the SER.

The PER criterion used to estimate the sparsity level gives satisfactory results, but an upcoming thorough analytical study is required.

Comparison with discrete techniques from compressed sensing indicate a trade-off between speed and accuracy compared to FRI-PERK. Using a finer discretization (*frames*) allows one to vary this trade-off, and shall be tested in the future. We believe that with a proper discretization similar speed/accuracy results shall be met.

REFERENCES

- [1] G. T. V9.0, *Requirements for E-UTRA and E-UTRAN*, ETSI Std., 2009.
- [2] 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>
- [3] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, “Estimation of Sparse MIMO Channels with Common Support,” *submitted to IEEE Trans. Commun.*, 2012.
- [4] Y. Lu and M. Do, “Sampling Signals from a Union of Subspaces,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 41–47, march 2008.
- [5] 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, pp. 31–40, 2008.
- [6] A. Aldroubi and K. Gröchenig, “Nonuniform sampling and reconstruction in shift-invariant spaces,” *SIAM review*, pp. 585–620, 2001.
- [7] J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [8] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, “On detection of the number of signals in presence of white noise,” *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 1–25, 1986.
- [9] H. Hofstetter, C. Mecklenbräuer, R. Müller, H. Anegg, H. Kunczler, E. Bonek, I. Viering, and A. Molisch. (2002). [Online]. Available: <http://measurements.ftw.at/MIMO.html>
- [10] C. Berger, S. Zhou, J. Preisig, and P. Willett, “Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, 2010.
- [11] 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*, 2008, pp. 2885–2888.
- [12] O. Roy and M. Vetterli, “The Effective Rank: A Measure of Effective Dimensionality,” in *European Signal Processing Conference (EUSIPCO)*, 2007, pp. 606–610.
- [13] M. Davies and Y. Eldar, “Rank awareness in joint sparse recovery,” *Arxiv preprint arXiv:1004.4529*, 2010.

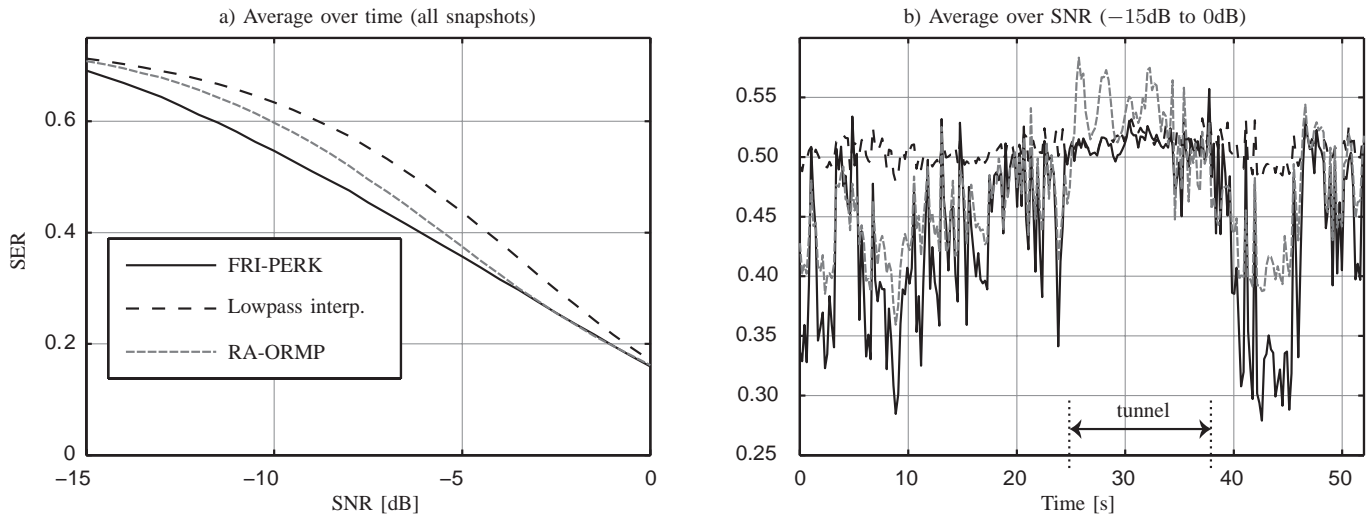


Fig. 7. From a) we conclude that sparse recovery lowers the SER compared to non-sparse recovery below 0dB of SNR. If we look at the SER over time in b), we see that FRI-PERK is robust in the sense that if the input signal is not sparse, its performs approximately as well as a non-sparse recovery.

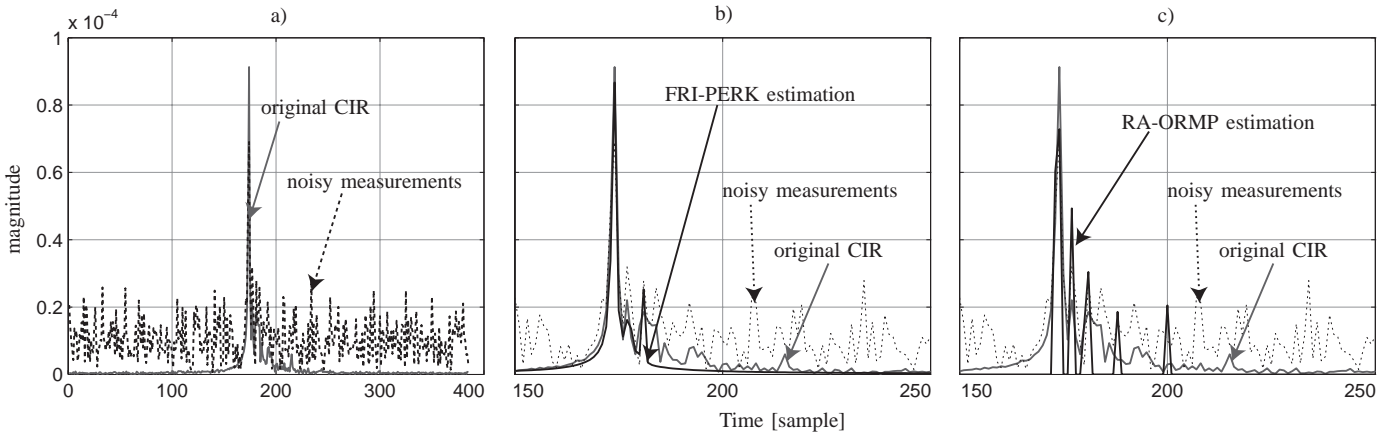


Fig. 8. This figure compares the estimation result of FRI-PERK and RA-ORMP. The input signal is the first frame received at the first antenna corrupted with AWGN to obtain -5 dB of SNR. Panel (a) shows the measurements and the original CIR. Panel (b) shows a portion of interest of the CIR estimated with FRI-PERK. The PER criterion estimates $K = 3$, and by visual inspection the three signal components found match the largest ones of the original signal. Panel (c) shows the result obtained with RA-ORMP. The discrete sparsity model causes the estimation to be more sensitive to uncorrelated noise, as spurious spikes contributed by noise are estimated as signal components.

- [14] K. Lee, Y. Bresler, and M. Junge, "Subspace Methods for Joint Sparse Recovery," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, p. 1, 2012.
- [15] A. Saleh and R. Valenzuela, "A Statistical Model for Indoor Multipath Propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, february 1987.
- [16] D. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)," in *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, vol. 5. IEEE, 2005, pp. 3132–3136.
- [17] J. Salz and J. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, 1994.
- [18] D. Cassioli, M. Win, and A. Molisch, "The ultra-wide bandwidth indoor channel: from statistical model to simulations," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257, aug 2002.
- [19] T. S. Rappaport, *Principles of Wireless Communications*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [20] 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.
- [21] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 975–989, 1982.
- [22] M. Soltanolkotabi, M. Soltanalian, A. Amini, and F. Marvasti, "A practical sparse channel estimation for current OFDM standards," in *ICT '09*, may 2009, pp. 217–222.
- [23] B. Parlett, *The symmetric eigenvalue problem*. SIAM, 1998, vol. 20.
- [24] G. Golub and C. van Van Loan, *Matrix Computations*. The Johns Hopkins Uni. Press, 1996.
- [25] Y. Saad, "On the rates of convergence of the Lanczos and the block-Lanczos methods," *SIAM J. on Num. Analysis*, pp. 687–706, 1980.
- [26] C. Paige, "Computational variants of the Lanczos method for the eigenproblem," *IMA J. of Applied Math.*, vol. 10, no. 3, p. 373, 1972.
- [27] D. Boley and F. Luk, "A general vandermonde factorization of a Hankel matrix," *Int'l Lin. Alg. Soc.(ILAS) Symp. on Fast ...*, 1997.
- [28] Boley, "Fast identification of impulse response modes via Krylov space methods," in *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, 1997, pp. 4384–4388.
- [29] G. Xu, H. Zha, G. Golub, and T. Kailath, "Fast algorithms for updating signal subspaces," *IEEE Trans. Circuits Syst. II*, vol. 41, no. 8, pp. 537–549, 1994.
- [30] W. Yeh and C. Jen, "High-speed and low-power split-radix FFT," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 864–874, 2003.
- [31] R. Brent, F. Luk, and C. Van Loan, "Computation of the singular value decomposition using mesh-connected processors," *Journal of VLSI and computer systems*, vol. 1, no. 3, pp. 242–270, 1985.
- [32] J. Cavallaro and F. Luk, "CORDIC Arithmetic for an SVD Processor," *J. of parallel and dist. computing*, vol. 5, no. 3, pp. 271–290, 1988.

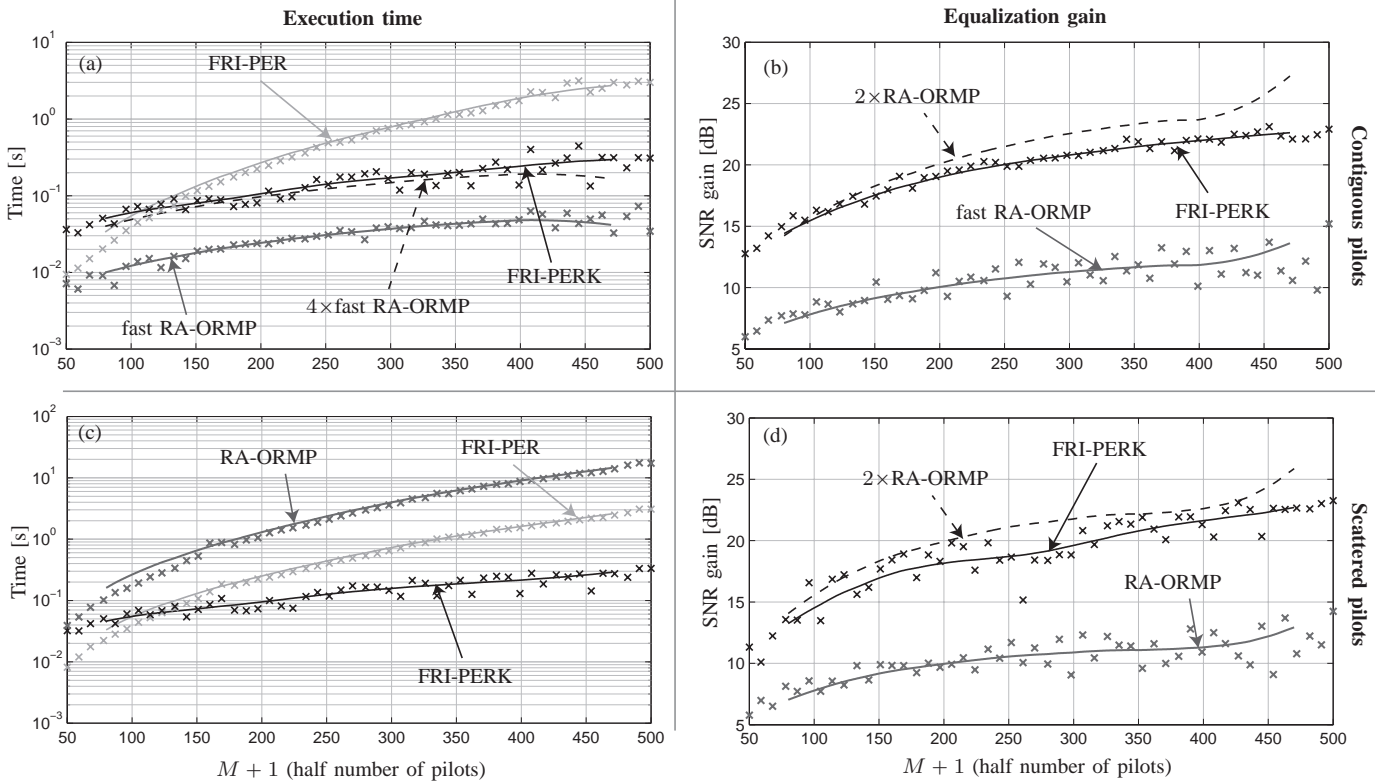


Fig. 9. A benchmark is run for pilot sequences of length $2M + 1 = 101$ to $2M + 1 = 1001$. The top row compares FRI-PERK with FRI-PER — same algorithm using a full SVD instead of Krylov subspace projection — and a fast implementation of RA-ORMP using FFTs, which is possible since the pilots are contiguous in frequency ($D = 1$). The bottom row follows the same procedure, but with scattered pilots ($D = 3$) and a delay-spread smaller than the frame length. There is no straightforward fast implementation of RA-ORMP in this case.

- [33] G. Szegő, *Orthogonal polynomials. Revised ed.* American Mathematical Society (AMS), 1939.
- [34] Y. Yin and P. Krishnaiah, “On some nonparametric methods for detection of the number of signals,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 11, pp. 1533–1538, 1987.
- [35] M. Kaveh, H. Wang, and H. Hung, “On the theoretical performance of a class of estimators of the number of narrow-band sources,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 9, pp. 1350–1352, 1987.
- [36] G. Xu, R. H. Roy, and T. Kailath, “Detection of number of sources via exploitation of centro-symmetry property,” *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 102–112, 1994.
- [37] W. Bryc, A. Dembo, and T. Jiang, “Spectral measure of large random Hankel, Markov and Toeplitz matrices,” *The Annals of Probability*, vol. 34, no. 1, pp. 1–38, 2006.
- [38] M. Meckes, “On the spectral norm of a random Toeplitz matrix,” *Elec. Comm. in Proba.*, vol. 12, pp. 315–325, 2007.
- [39] R. Lehoucq, D. Sorensen, and C. Yang, *ARPACK users’ guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods.* Siam, 1998, vol. 6.



Yann Barbotin was born in Ambilly, France in 1985. He received the B.Sc. and M.Sc. degrees in Communication Systems from the Swiss Federal Institute of Technology, Lausanne (EPFL) in 2006 and 2009, respectively. As an intern at Qualcomm Inc. in 2009 and an ongoing scientific collaboration he is the co-inventor of five patents on sparse sampling. He is now a PhD candidate in the LCAV laboratory at EPFL. His research interests are in fundamental limitations of estimation problems and sampling.



Martin Vetterli Martin Vetterli was born in 1957 and grew up near Neuchâtel. He received the Dipl. El.-Ing. degree from Eidgenössische Technische Hochschule (ETHZ), Zurich, in 1981, the Master of Science degree from Stanford University in 1982, and the Doctorat es Sciences degree from the Ecole Polytechnique Fédérale, Lausanne, in 1986. After his dissertation, he was an Assistant and then Associate Professor in Electrical Engineering at Columbia University in New York, and in 1993, he became an Associate and then Full Professor at the Department

of Electrical Engineering and Computer Sciences at the University of California at Berkeley. In 1995, he joined the EPFL as a Full Professor. He held several positions at EPFL, including Chair of Communication Systems and founding director of the National Competence Center in Research on Mobile Information and Communication systems (NCCR-MICS). From 2004 to 2011 he was Vice President of EPFL and since March 2011, he is the Dean of the School of Computer and Communications Sciences. He works in the areas of electrical engineering, computer sciences and applied mathematics. His work covers wavelet theory and applications, image and video compression, self-organized communications systems and sensor networks, as well as fast algorithms, and has led to about 150 journals papers. He is the co-author of three textbooks, with J. Kovacevic, “Wavelets and Subband Coding” (Prentice-Hall, 1995), with P. Prandoni, “Signal Processing for Communications”, (CRC Press, 2008) and with J. Kovacevic and V. Goyal, of the forthcoming book “Fourier and Wavelet Signal Processing” (2012). His research resulted also in about two dozen patents that led to technology transfers to high-tech companies and the creation of several start-ups. His work won him numerous prizes, like best paper awards from EURASIP in 1984 and of the IEEE Signal Processing Society in 1991, 1996 and 2006, the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, the IEEE Signal Processing Technical Achievement Award in 2001 and the IEEE Signal Processing Society Award in 2010. He is a Fellow of IEEE, of ACM and EURASIP, was a member of the Swiss Council on Science and Technology (2000–2004), and is an ISI highly cited researcher in engineering.