

# FRI Sampling With Arbitrary Kernels

Jose Antonio Urigüen, *Student Member, IEEE*, Thierry Blu, *Fellow, IEEE*, and Pier Luigi Dragotti, *Senior Member, IEEE*

**Abstract**—This paper addresses the problem of sampling non-bandlimited signals within the Finite Rate of Innovation (FRI) setting. We had previously shown that, by using sampling kernels whose integer span contains specific exponentials (generalized Strang-Fix conditions), it is possible to devise non-iterative, fast reconstruction algorithms from very low-rate samples. Yet, the accuracy and sensitivity to noise of these algorithms is highly dependent on these exponential reproducing kernels — actually, on the exponentials that they reproduce. Hence, our first contribution here is to provide clear guidelines on how to choose the sampling kernels optimally, in such a way that the reconstruction quality is maximized in the presence of noise. The optimality of these kernels is validated by comparing with Cramér-Rao’s lower bounds (CRB). Our second contribution is to relax the exact exponential reproduction requirement. Instead, we demonstrate that arbitrary sampling kernels can reproduce the “best” exponentials within quite a high accuracy in general, and that applying the exact FRI algorithms in this approximate context results in near-optimal reconstruction accuracy for practical noise levels. Essentially, we propose a universal extension of the FRI approach to arbitrary sampling kernels. Numerical results checked against the CRB validate the various contributions of the paper and in particular outline the ability of arbitrary sampling kernels to be used in FRI algorithms.

**Index Terms**—DSP-SAMP, finite rate of innovation, matrix Pencil, MOMS, noise, sampling.

## I. INTRODUCTION

**M**OST signal acquisition systems involve the conversion of signals from analog to digital, and sampling theorems provide the bridge between the continuous and the discrete-time worlds. Usually, the acquisition process is modelled as in Fig. 1, where the smoothing function  $\varphi(t)$  is called the sampling kernel and normally models the distortion due to the acquisition device. The filtered continuous-time signal  $y(t) = x(t) * \varphi(-\frac{t}{T})$  is then uniformly sampled at a rate  $f_s = \frac{1}{T}$ . Following this setup, the measurements are given by

$$y_n = \int_{-\infty}^{\infty} x(t) \varphi\left(\frac{t}{T} - n\right) dt = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right) \right\rangle.$$

Manuscript received December 11, 2012; revised May 03, 2013; revised May 03, 2013 and July 19, 2013; accepted July 24, 2013. Date of publication August 15, 2013; date of current version September 24, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jian-Kang Zhang. This work was supported in part by the European Research Council (ERC) starting investigator Award 277800 (RecoSamp); and in part by a RGC Grant CUHK410110 of the Hong Kong University Grant Council.

J. A. Urigüen and P. L. Dragotti are with the Department of Electrical and Electronic Engineering, Imperial College, London, U.K. (e-mail: jau08@ic.ac.uk; p.dragotti@imperial.ac.uk).

T. Blu is with the Department of Electrical and Electronic Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: thierry.blu@m4x.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2013.2278152

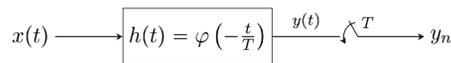


Fig. 1. *Traditional sampling scheme*. The continuous-time input signal  $x(t)$  is filtered with  $h(t)$  and sampled every  $T$  seconds. The samples are then given by  $y_n = (x * h)(t)|_{t=nT}$ .

The fundamental problem of sampling is to recover the original continuous-time waveform  $x(t)$  using the set of samples  $y_n$ . In case the signal is bandlimited, the answer due to Shannon is well known. Recently, it has been shown that it is possible to sample and perfectly reconstruct specific classes of non-bandlimited signals [1]–[3]. Such signals are called signals with finite rate of innovation (FRI) since they are completely described by a finite number of free parameters per unit of time. Perfect reconstruction is achieved by using a variation of Prony’s method, also known as annihilating filter method [4]. Signals that can be sampled within this framework include streams of pulses such as Diracs [1]–[3], [5], piecewise polynomial signals, piecewise sinusoidal signals [6] and classes of 2-D signals [7]–[10]. In the presence of noise, FRI reconstruction techniques become unstable and methods to improve resiliency to noise have been presented in [10]–[15].

Various sampling kernels can be used to perfectly reconstruct FRI signals such as the sinc and Gaussian functions first proposed in the original paper on FRI [1] and compact support kernels such as polynomial and exponential reproducing kernels [2], [3], [16]. While they all allow perfect reconstruction in noiseless settings, their behavior changes in the presence of noise. It is therefore natural to attempt to understand which factors cause a deterioration in performance and also to determine sampling schemes and recovery methods that are resilient to noise.

In this paper, we focus on the family of exponential reproducing kernels [2] for two reasons: First, they can have compact support, which is in itself a nice property when dealing with noisy measurements. Second and more important, any compact support kernel that has so far been used in FRI sampling using the setting of Fig. 1 is a particular instance of the family of exponential reproducing kernels (see Section II-C and Appendix B).

Our contribution is twofold: We first explain how to design the most effective exponential reproducing kernels when sampling and reconstructing FRI signals in noisy environments. Since FRI recovery is equivalent to estimating a set of parameters in noise, we use the Cramér-Rao bound (CRB) of this estimation problem as our optimisation criterion when designing the kernel. For the second contribution we depart from the previous setup and assume we have no control on the acquisition device. In such scenario, we develop a universal FRI reconstruction strategy that works with samples taken by any kernel. In contrast to existing techniques that attempt at finding parameters of the input exactly [1]–[3], [12], we propose an alternative method that finds the parameters approximately. The ad-

vantage of our new method is that, for kernels such as polynomial splines or the Gaussian function which are in practice very unstable, it provides a much more stable and accurate recovery in the presence of noise.

The outline of the paper is as follows. In Section II we review the noiseless scenario in which we sample and perfectly reconstruct the prototypical FRI signal: a train of Diracs. We also discuss exponential reproducing kernels and their generalised Strang-Fix conditions [17], for which we provide a simple proof. In Section II-B we treat the more realistic setup where noise is present in the acquisition process. Here, we describe practical techniques to retrieve the train of Diracs and in Section III we compute the Cramér-Rao bound (CRB) for this problem. We also present a CRB formulation based on the exponential moments of the input that has not been used in the FRI literature to date. In Section IV we design a family of exponential reproducing kernels that is most resilient to noise. Then, in Section V we introduce the approximate FRI framework and develop the basic ideas to sample FRI signals with any kernel. In Section VI we present simulation results to validate the various contributions of the paper. Interestingly, we also show that with the new approximate framework we can improve the accuracy of the reconstruction associated to sampling kernels for which existing exact recovery methods become unstable in the presence of noise. Finally, we conclude the paper in Section VII.

## II. SAMPLING SIGNALS WITH FINITE RATE OF INNOVATION

In this section we provide a brief overview of FRI theory. Specifically, we explain how to reconstruct a stream of Diracs from noisy or noiseless samples taken by an exponential reproducing kernel. We also highlight some of the key properties of this type of kernels which will be useful for the rest of the paper.

### A. Perfect Reconstruction of a Stream of Diracs

Assume that the input  $x(t)$  is a stream of  $K$  Diracs

$$x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k), \quad (1)$$

where  $a_k \in \mathbb{R}$  are the amplitudes and  $t_k \in \mathbb{R}$  are the time locations of the Diracs. We restrict the locations to the interval  $t_k \in [0, \tau)$  for  $k = 0, \dots, K-1$ .

Now, based on the acquisition model of Fig. 1, we filter the input with the kernel  $\varphi(t)$  and obtain the samples

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right) \right\rangle = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right), \quad (2)$$

where  $n = 0, 1, \dots, N-1$  and the sampling period  $T$  satisfies  $\tau = NT$ . Moreover, we assume that  $\varphi(t)$  is an exponential reproducing kernel of compact support. That is,  $\varphi(t)$  is a function satisfying:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t-n) = e^{\alpha_m t}, \quad (3)$$

for proper coefficients  $c_{m,n}$ , with  $m = 0, \dots, P$  and  $\alpha_m \in \mathbb{C}$ . Since  $m = 0, \dots, P$  this is an exponential reproducing kernel of order  $P+1$ . These kernels are discussed in detail in Section II-C.

From now on, we also assume  $\alpha_m = \alpha_0 + m\lambda$  for  $m = 0, \dots, P$ .

Once we have sampled the input, the stream of Diracs can be unambiguously retrieved from the set of measurements  $y_n$  as follows: First we linearly combine the samples  $y_n$  with the coefficients  $c_{m,n}$  of (3), to obtain the new sequence:

$$s_m = \sum_{n=0}^{N-1} c_{m,n} y_n, \quad (4)$$

for  $m = 0, \dots, P$ . Then, given that the signal  $x(t)$  is a stream of Diracs (1) and combining (4) with (2) we have [2]:

$$\begin{aligned} s_m &= \left\langle x(t), \underbrace{\sum_{n=0}^{N-1} c_{m,n} \varphi\left(\frac{t}{T} - n\right)}_{e^{\alpha_m \frac{t}{T}}} \right\rangle \\ &= \sum_{k=0}^{K-1} a_k e^{\alpha_m \frac{t_k}{T}} = \sum_{k=0}^{K-1} x_k u_k^m, \end{aligned} \quad (5)$$

with  $x_k = a_k e^{\alpha_0 \frac{t_k}{T}}$  and  $u_k = e^{\lambda \frac{t_k}{T}}$ . Here it is the choice  $\alpha_m = \alpha_0 + m\lambda$ , where  $m = 0, \dots, P$ , that makes  $s_m$  have a power sum series form. We note that  $s_m$  are precisely the (exponential) moments of the signal  $x(t)$  [6].

The new pairs of unknowns  $\{u_k, x_k\}_{k=0}^{K-1}$  can then be retrieved from the moments  $s_m$  using the annihilating filter method (AFM) [1], [2], [12], also known as Prony's method [4]. Let  $h_m$  with  $m = 0, \dots, K$  be the filter with  $z$ -transform  $\hat{h}(z) = \sum_{m=0}^K h_m z^{-m} = \prod_{k=0}^{K-1} (1 - u_k z^{-1})$ , that is, its roots correspond to the values  $u_k$  to be found. Then, it follows that  $h_m$  annihilates the observed sequence  $s_m$ :

$$h_m * s_m = \sum_{i=0}^K h_i s_{m-i} = \sum_{k=0}^{K-1} x_k u_k^m \underbrace{\sum_{i=0}^K h_i u_k^{-i}}_{\hat{h}(u_k)} = 0. \quad (6)$$

Moreover, the zeros of this filter uniquely define the values  $u_k$  provided the locations  $t_k$  are distinct. The identity (6) can be written in matrix-vector form as:

$$\mathbf{S} \mathbf{h} = 0 \quad (7)$$

which reveals that the Toeplitz matrix  $\mathbf{S}$ , with entries  $s_m$ , is rank deficient. By solving the above system, we find the filter coefficients  $h_m$  and then retrieve  $u_k$  by computing the roots of  $\hat{h}(z)$ . Given  $u_k$  we obtain the locations  $t_k$  since  $u_k = e^{\lambda \frac{t_k}{T}}$ . Finally, we determine the weights  $a_k$  by solving, for instance, the first  $K$  consecutive equations in (5). Notice that the problem can be solved only when there are at least as many equations as unknowns, implying that  $P+1 \geq 2K$ . This indicates that the order  $P+1$  of the exponential reproducing kernel has to be chosen according to the number of degrees of freedom of the input signal  $x(t)$ .

We end the above discussion by noting that all FRI reconstruction setups proposed so far [1]–[3], [12] can be unified as shown in Fig. 2. Here, the samples are represented with the vector  $\mathbf{y} = (y_0, y_1, \dots, y_{N-1})^T$  and the moments are given by  $\mathbf{s} = \mathbf{C} \mathbf{y}$ . The matrix  $\mathbf{C}$ , of size  $(P+1) \times N$  with

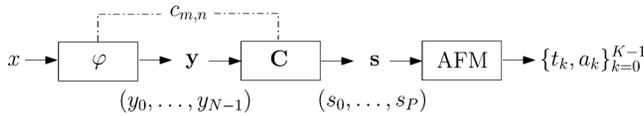


Fig. 2. *Unified FRI sampling and reconstruction*. The continuous-time input signal  $x$  is filtered with  $\varphi$  and uniformly sampled. Then, the vector of samples  $\mathbf{y}$  is linearly combined to obtain the moments  $\mathbf{s} = \mathbf{C}\mathbf{y}$ . Finally, the parameters of the input are retrieved from  $\mathbf{s}$  using the annihilating filter method (AFM).

coefficients  $c_{m,n}$  at position  $(m, n)$ , depends on the sampling kernel and its role becomes pivotal in noisy scenarios as discussed throughout the paper.

### B. Reconstruction of a Stream of Diracs in the Presence of Noise

Any practical acquisition device introduces noise during the acquisition process. We therefore assume that instead of (2) we have access to the noisy samples

$$\tilde{y}_n = y_n + \epsilon_n = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right) + \epsilon_n, \quad (8)$$

with  $n = 0, \dots, N-1$  and where  $\epsilon_n$  are i.i.d. Gaussian random variables, of zero mean and standard deviation  $\sigma$ . When the samples are corrupted by noise, the sequence  $s_m$  of (4) changes and perfect reconstruction is no longer possible. We now have the noisy moments:

$$\begin{aligned} \tilde{s}_m &= \sum_{n=0}^{N-1} c_{m,n} \tilde{y}_n = \underbrace{\sum_{n=0}^{N-1} c_{m,n} y_n}_{s_m} + \underbrace{\sum_{n=0}^{N-1} c_{m,n} \epsilon_n}_{b_m} \\ &= \sum_{k=0}^{K-1} x_k u_k^m + b_m, \end{aligned} \quad (9)$$

for  $m = 0, \dots, P$  and where  $x_k = a_k e^{\alpha_0 \frac{t_k}{T}}$  and  $u_k = e^{\lambda \frac{t_k}{T}}$  with  $k = 0, \dots, K-1$ .

Consequently, in the noisy setting (7) is not satisfied any more because now  $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{B}$  where  $\mathbf{B}$  is Toeplitz with entries  $b_m$  from (9). We may however solve (7) approximately by taking more than the critical number of moments ( $P+1 > 2K$ ) and applying a singular value decomposition (SVD) to  $\tilde{\mathbf{S}}$ . This is the total least-squares (TLS) solution to (7). The procedure may be improved by denoising  $\tilde{\mathbf{S}}$  before applying TLS using the Cadzow iterative algorithm [12], [18]. There exist other methods that attain a similar accuracy and are not iterative. One such approach, based on solving a matrix pencil problem [19], [20], was introduced for FRI in [11]. It has been employed in other FRI publications such as [3], [16] and is used in the simulations of this paper as well.

These methods operate effectively when the perturbation is white, that is when the covariance matrix of the noise  $\mathbf{B}$  satisfies  $\mathbf{R}_B = \mathbf{E}\{\mathbf{B}^H \mathbf{B}\} = \alpha \mathbf{I}$ , where  $\alpha$  is a constant factor and  $\mathbf{I}$  is the identity matrix. However, for many FRI kernels the white Gaussian noise assumption does not hold and in order for SVD to operate correctly it is necessary to “pre-whiten” the noise. In our simulations we use a weighting matrix  $\mathbf{W} = \mathbf{R}_B^{-\dagger/2}$  [21] such that  $\mathbf{R}_A = \mathbf{E}\{\mathbf{A}^H \mathbf{A}\} = \mathbf{I}$  with  $\mathbf{A} = \mathbf{B}\mathbf{W}$ . Here,  $(\cdot)^{\dagger/2}$  is

the square root of the pseudoinverse of  $(\cdot)$ . Therefore, we work with matrix  $\tilde{\mathbf{S}}\mathbf{W}$ , which is now characterised by white noise.

To conclude this part, we summarise the noisy FRI recovery method which we use in our simulations in insert Algorithm 1.

---

**Algorithm 1** Reconstruction of a stream of  $K$  Diracs in the presence of noise.

---

- 1) Calculate the sequence of  $P+1$  moments  $(\mathbf{9})$  from the  $N$  noisy samples  $\tilde{y}_n$  of  $(\mathbf{8})$ . Then, build the Toeplitz matrix  $\tilde{\mathbf{S}}$  with the sequence  $\tilde{s}_m$ . Here  $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{B}$ .
  - 2) Estimate  $\mathbf{R}_B = \mathbf{E}\{\mathbf{B}^H \mathbf{B}\}$  and define the new matrix  $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\mathbf{W}$ , where  $\mathbf{W} = \mathbf{R}_B^{-\dagger/2}$ .
  - 3) Apply the matrix pencil method to  $\tilde{\mathbf{S}}'$ : Obtain the decomposition  $\tilde{\mathbf{S}}' = \mathbf{U}\mathbf{A}\mathbf{V}^H$ , keep the  $K$  columns of  $\mathbf{U}$  corresponding to the  $K$  dominant singular values and estimate  $u_k$  as the eigenvalues of  $\underline{\mathbf{U}}_K^+ \overline{\mathbf{U}}_K$ . Here,  $(\cdot)$  and  $(\overline{\cdot})$  are operations to omit the last and first rows of  $(\cdot)$ .
  - 4) Compute the  $K$  locations of the Diracs as  $t_k = \frac{T}{\lambda} \ln(u_k)$ .
  - 5) Calculate  $a_k$  as the least mean square solution of the  $N$  equations  $\tilde{y}_n - \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right) \approx 0$  for  $n = 0, \dots, N-1$ .
- 

### C. Exponential Reproducing Kernels

An exponential reproducing kernel is any function  $\varphi(t)$  that, together with a linear combination of its shifted versions, can reproduce functions of the form  $e^{\alpha_m t}$ , with complex parameters  $\alpha_m$ . This can be expressed mathematically as follows:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t-n) = e^{\alpha_m t}, \quad (10)$$

for properly chosen coefficients  $c_{m,n} \in \mathbb{C}$  and where  $m = 0, \dots, P$  and  $\alpha_m \in \mathbb{C}$ . Exponential reproducing kernels for which (10) is true satisfy the so-called generalised Strang-Fix conditions [17] (see Appendix A for a simple proof). In particular, Equation (10) holds if and only if

$$\hat{\varphi}(\alpha_m) \neq 0 \text{ and } \hat{\varphi}(\alpha_m + 2j\pi l) = 0, \quad (11)$$

for  $m = 0, \dots, P$  and  $l \in \mathbb{Z} \setminus \{0\}$ , where  $\hat{\varphi}(\alpha_m)$  represents the bilateral Laplace transform of  $\varphi(t)$ , i.e.,  $\hat{\varphi}(s) = \int_{-\infty}^{\infty} \varphi(t) e^{-st} dt$ , at  $s = \alpha_m$ . Moreover, the coefficients  $c_{m,n}$  in (10) are given by

$$\begin{aligned} c_{m,n} &= \langle e^{\alpha_m t}, \tilde{\varphi}(t-n) \rangle = \int_{-\infty}^{\infty} e^{\alpha_m t} \tilde{\varphi}(t-n) dt \\ &= c_{m,0} e^{\alpha_m n}, \end{aligned} \quad (12)$$

where  $\tilde{\varphi}(t)$  forms a biorthonormal set with  $\varphi(t)$  [2], and where  $c_{m,0} = \int_{-\infty}^{\infty} e^{\alpha_m x} \tilde{\varphi}(x) dx$ .

Any exponential reproducing kernel can be written as  $\varphi(t) = \gamma(t) * \beta_{\vec{\alpha}}(t)$  [2], [22], [23], where  $\gamma(t)$  is an arbitrary function, even a distribution, and  $\beta_{\vec{\alpha}}(t)$  is an E-Spline. A function  $\beta_{\alpha}(t)$  with Fourier transform  $\hat{\beta}_{\alpha}(j\omega) = \frac{1 - e^{-\alpha - j\omega}}{j\omega - \alpha}$  is an E-spline of first order. Higher order E-Splines can be obtained through convolution of first order ones. For instance  $\beta_{\vec{\alpha}}(t) = (\beta_{\alpha_0} * \beta_{\alpha_1} * \dots * \beta_{\alpha_P})(t)$ , where  $\vec{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_P)$  is an E-Spline of order  $P+1$ , has compact support  $P+1$  and has  $P-1$  continuous derivatives. This function can reproduce any

exponential in  $\text{span}\{e^{\alpha_0 t}, e^{\alpha_1 t}, \dots, e^{\alpha_P t}\}$  [2], [22]. Moreover, when  $\alpha_m = 0$  for  $m = 0, \dots, P$ , the function  $\beta_{\vec{\alpha}}(t)$  becomes a B-Spline and no longer reproduces exponentials but polynomials up to degree  $P$ .

### Remarks

In this paper we work with real valued sampling kernels, therefore we require that  $\gamma(t)$  and  $\beta_{\vec{\alpha}}(t)$  be real functions. E-Splines  $\beta_{\vec{\alpha}}(t)$  are real if the exponents  $\vec{\alpha} = \{\alpha_m\}_{m=0}^P$  are real or exist in complex conjugate pairs. Since we also restrict the exponents to be of the form  $\alpha_m = \alpha_0 + m\lambda$ ,  $m = 0, \dots, P$ , then either  $\lambda$  and  $\alpha_0$  are real or  $\lambda$  is purely imaginary and  $\alpha_0$  is complex with  $\text{Im}\{\alpha_0\} = -\frac{P\text{Im}\{\lambda\}}{2}$ . For the rest of the paper we use  $\alpha_m = \alpha + j\frac{\pi}{L}(2m - P)$ ,  $m = 0, \dots, P$ , which satisfies all the requirements with  $\alpha_0 = \alpha - j\frac{\pi P}{L}$  and  $\lambda = j\frac{2\pi}{L}$ .

Note that since  $\lambda$  is purely imaginary then there may exist ambiguities when obtaining the locations  $t_k$  from  $u_k$ . This is because  $t_k$  and  $t_k + LT\ell$ , where  $\ell \in \mathbb{Z}$ , produce the same annihilating filter roots  $u_k = e^{\lambda \frac{t_k}{T}} = e^{j\frac{2\pi}{L}t_k} = e^{j\frac{2\pi}{L}(t_k + LT\ell)}$ . It is thus necessary that  $0 \leq t_k < LT$  for  $k = 0, \dots, K - 1$  in order to retrieve the locations unambiguously.

### III. MEASURING THE PERFORMANCE

FRI signals are completely characterised by their innovation parameters. For instance, a stream of  $K$  Diracs can be determined from the locations  $t_k$  and amplitudes  $a_k$ . The goal of FRI reconstruction is to estimate  $\Theta = (t_0, \dots, t_{K-1}, a_0, \dots, a_{K-1})^T$  from the vector of  $N$  noisy samples  $\tilde{\mathbf{y}} = (\tilde{y}_0, \dots, \tilde{y}_{N-1})^T$  given by (8). For simplicity we assume the sampling period is  $T = 1$ . A way to determine the CRB of this estimation problem was given in [12] assuming  $\epsilon_n$  is a zero-mean Gaussian noise with covariance matrix  $\mathbf{R} = \mathbb{E}\{\mathbf{e}\mathbf{e}^H\} = \sigma^2\mathbf{I}$ , where  $\mathbf{e}$  is the vector of length  $N$  with values  $\epsilon_n$ . In this set-up any unbiased estimate of the unknown parameters  $\hat{\Theta}(\tilde{\mathbf{y}}) = (\hat{t}_0, \dots, \hat{t}_{K-1}, \hat{a}_0, \dots, \hat{a}_{K-1})^T$  has a covariance matrix that is lower bounded by

$$\text{cov}(\hat{\Theta}(\tilde{\mathbf{y}})) \geq (\Phi_{\tilde{\mathbf{y}}}^T \mathbf{R}^{-1} \Phi_{\tilde{\mathbf{y}}})^{-1}, \quad (13)$$

where the matrix  $\Phi_{\tilde{\mathbf{y}}}$  is given by (14),<sup>1</sup> shown at the bottom of the page.

While this is one possible way to measure the performance of various FRI recovery techniques [11], [12] given the noisy samples  $\tilde{\mathbf{y}}$ , we also note that in practice FRI reconstruction methods operate on the moments  $\tilde{\mathbf{s}} = \mathbf{C}\tilde{\mathbf{y}}$ . We therefore present the CRB associated to the moments, since this will indicate the best performance that can be achieved when working with  $\tilde{\mathbf{s}}$ . In this case the bound is given by [24]:

$$\text{cov}(\hat{\Theta}(\tilde{\mathbf{s}})) \geq (\Phi_{\tilde{\mathbf{s}}}^H \mathbf{R}_{\tilde{\mathbf{b}}}^{-1} \Phi_{\tilde{\mathbf{s}}})^{-1}, \quad (15)$$

where  $\Phi_{\tilde{\mathbf{s}}}$  takes the form (16) at the bottom of the page and  $\mathbf{R}_{\tilde{\mathbf{b}}} = \mathbb{E}\{\mathbf{b}\mathbf{b}^H\}$ . Here,  $(\cdot)^H$  is Hermitian transpose and  $\mathbf{b}$  is the vector of  $P + 1$  noisy values  $b_m$ .

Furthermore, since we have assumed that the noise  $\epsilon_n$  added to the samples is additive, white and Gaussian (AWGN), with variance  $\sigma^2$ , the covariance matrix of the transformed noise is  $\mathbf{R}_{\tilde{\mathbf{b}}} = \mathbb{E}\{\mathbf{b}\mathbf{b}^H\} = \sigma^2 \mathbf{C}\mathbf{C}^H$ . Then, by taking into account that  $\tilde{\mathbf{s}} = \mathbf{C}(\mathbf{y} + \mathbf{e}) = \mathbf{s} + \mathbf{b}$  and also that  $\Phi_{\tilde{\mathbf{s}}} = \mathbf{C}\Phi_{\tilde{\mathbf{y}}}$ , we may write

$$\text{cov}(\hat{\Theta}(\tilde{\mathbf{y}})) \geq (\Phi_{\tilde{\mathbf{y}}}^H \mathbf{R}^{-1} \Phi_{\tilde{\mathbf{y}}})^{-1} = \sigma^2 (\Phi_{\tilde{\mathbf{y}}}^H \Phi_{\tilde{\mathbf{y}}})^{-1},$$

since the noise  $\epsilon_n$  is AWGN. Consequently, it is true that

$$\begin{aligned} \text{cov}(\hat{\Theta}(\tilde{\mathbf{s}})) &\geq (\Phi_{\tilde{\mathbf{s}}}^H \mathbf{R}_{\tilde{\mathbf{b}}}^{-1} \Phi_{\tilde{\mathbf{s}}})^{-1} \\ &= \sigma^2 (\Phi_{\tilde{\mathbf{y}}}^H \mathbf{C}^H (\mathbf{C}\mathbf{C}^H)^{-1} \mathbf{C} \Phi_{\tilde{\mathbf{y}}})^{-1} \\ &= \sigma^2 (\Phi_{\tilde{\mathbf{y}}}^H \mathbf{C}^\dagger \mathbf{C} \Phi_{\tilde{\mathbf{y}}})^{-1}, \end{aligned}$$

where  $\mathbf{C}^\dagger = \mathbf{C}^H (\mathbf{C}\mathbf{C}^H)^{-1}$  is the left pseudo-inverse of  $\mathbf{C}$ . When the number of moments  $P + 1$  equals the number of samples  $N$ , then (13) and (15) are equivalent. This is easily seen since in such case  $\mathbf{C}^\dagger \mathbf{C} = \mathbf{C}^{-1} \mathbf{C} = \mathbf{I}_N$ . A moment-based algorithm is, of course, likely to underperform a sample-based algorithm because it is the number of these moments that are an indicator of the quality of the retrieval, not the number of samples (which is always larger than the number of moments).

We have seen experimentally that FRI algorithms reach the bound (15) when  $\mathbf{C}$  is sufficiently well conditioned. Therefore

<sup>1</sup>The matrix can be obtained calculating the derivative of  $\tilde{y}_n$  with respect to each parameter in  $\Theta$ . That is, the columns of  $\Phi_{\tilde{\mathbf{y}}}$  to the left of  $|$  are  $\partial \tilde{y}_n / \partial t_k$  and the columns of  $\Phi_{\tilde{\mathbf{y}}}$  to the right of  $|$  are  $\partial \tilde{y}_n / \partial a_k$ .

$$\Phi_{\tilde{\mathbf{y}}} = \left( \begin{array}{ccc|ccc} a_0 \varphi'(t_0) & \dots & a_{K-1} \varphi'(t_{K-1}) & \varphi(t_0) & \dots & \varphi(t_{K-1}) \\ a_0 \varphi'(t_0 - 1) & \dots & a_{K-1} \varphi'(t_{K-1} - 1) & \varphi(t_0 - 1) & \dots & \varphi(t_{K-1} - 1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_0 \varphi'(t_0 - (N - 1)) & \dots & a_{K-1} \varphi'(t_{K-1} - (N - 1)) & \varphi(t_0 - (N - 1)) & \dots & \varphi(t_{K-1} - (N - 1)) \end{array} \right). \quad (14)$$

$$\Phi_{\tilde{\mathbf{s}}} = \left( \begin{array}{ccc|ccc} a_0 \alpha_0 e^{\alpha_0 t_0} & \dots & a_{K-1} \alpha_0 e^{\alpha_0 t_{K-1}} & e^{\alpha_0 t_0} & \dots & e^{\alpha_0 t_{K-1}} \\ a_0 \alpha_1 e^{\alpha_1 t_0} & \dots & a_{K-1} \alpha_1 e^{\alpha_1 t_{K-1}} & e^{\alpha_1 t_0} & \dots & e^{\alpha_1 t_{K-1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_0 \alpha_P e^{\alpha_P t_0} & \dots & a_{K-1} \alpha_P e^{\alpha_P t_{K-1}} & e^{\alpha_P t_0} & \dots & e^{\alpha_P t_{K-1}} \end{array} \right) \quad (16)$$

our goal now is to design kernels that lead to properly conditioned  $\mathbf{C}$  that minimise (15) for any choice of  $P$ .<sup>2</sup>

#### IV. OPTIMAL EXPONENTIAL REPRODUCING KERNELS

As mentioned before, an exponential reproducing kernel can be written as  $\varphi(t) = \gamma(t) * \beta_{\tilde{\alpha}}(t)$ , where  $\gamma(t)$  is arbitrary and  $\beta_{\tilde{\alpha}}(t)$  is an E-Spline. In this section we want to find rules on how to choose the exponential parameters of the E-Spline  $\alpha_m = \alpha + j\frac{\pi}{L}(2m - P)$  for  $m = 0, \dots, P$  and the function  $\gamma(t)$  in order to make FRI recovery techniques with these kernels as stable as possible. Finding the best parameters translates into the optimisation of the matrix of coefficients  $\mathbf{C}$ . Therefore we first determine the properties that  $\mathbf{C}$  has to satisfy and then design the kernels that lead to our choice of  $\mathbf{C}$ .

##### A. How To Choose Matrix $\mathbf{C}$

The first step in the FRI reconstruction stage is to transform the vector of samples  $\mathbf{y}$  into the vector of moments  $\mathbf{s} = \mathbf{C}\mathbf{y}$ , therefore, our first aim is to get a well conditioned  $\mathbf{C}$ . From (12) we note that matrix  $\mathbf{C}$  is composed of elements  $c_{m,n} = c_{m,0}e^{\alpha_m n}$  at position  $(m, n)$ , where  $n = 0, \dots, N - 1$  and  $m = 0, \dots, P$ :

$$\mathbf{C} = \underbrace{\begin{pmatrix} c_{0,0} & 0 & \cdots & 0 \\ 0 & c_{1,0} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{P,0} \end{pmatrix}}_{\mathbf{D}} \underbrace{\begin{pmatrix} 1 & e^{\alpha_0} & \cdots & e^{\alpha_0(N-1)} \\ 1 & e^{\alpha_1} & \cdots & e^{\alpha_1(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{\alpha_P} & \cdots & e^{\alpha_P(N-1)} \end{pmatrix}}_{\mathbf{V}}.$$

Here,  $\mathbf{D}$  is diagonal and  $\mathbf{V}$  Vandermonde. Hence, to have a stable  $\mathbf{C}$  we want the absolute values of the diagonal elements of  $\mathbf{D}$  to be the same, for instance  $|c_{m,0}| = 1$ . Moreover, we want the elements in  $\mathbf{V}$  to lie on the unit circle:

$$e^{\alpha_m n} = e^{j\frac{\pi}{L}(2m-P)n} \quad \text{for } m = 0, \dots, P, \text{ i.e., } \alpha = 0. \quad (17)$$

Clearly, purely imaginary  $\alpha_m$  make the Vandermonde matrix  $\mathbf{V}$  better conditioned [25]. We are therefore only left with the problem of finding the best  $L$  in (17). Since we have experimentally seen that FRI algorithms are able to reach the CRB (15) if  $\mathbf{C}$  is well conditioned, one way to determine  $L$  is to choose the value that minimises (15) for the location of a single Dirac. It turns out the minimum is always achieved when  $L = P + 1$ , as shown in Fig. 3 for various choices of  $P$  and  $L$ , given  $|c_{m,0}| = 1$  for all  $m$ .

To some extent, this is not surprising since this choice ensures that the exponentials span the entire unit circle, which is well known to be the best configuration when recovering the parameters of a power sum series [26]. Finally, when we impose  $P + 1 = N$  with  $L = P + 1$ , besides minimising (15), we also ensure that the moment-based CRB in(15) matches the sample-based bound in(13), leading to the best possible performance. In this situation, the matrix  $\mathbf{C}$  ends up being square and unitary. This is the most stable numerical transformation since its condition number is one.

<sup>2</sup>The condition  $P + 1 = N$  can be imposed only for blockwise sampling, e.g., when sampling periodic signals using  $N$  samples. This condition cannot be imposed on infinite length signals since sequential reconstruction algorithms will operate on blocks with possibly varying number of samples.

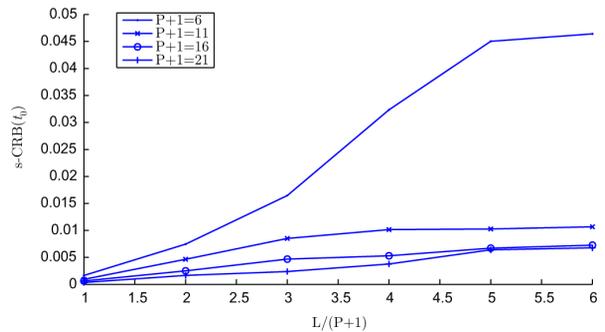


Fig. 3. CRB vs.  $L$ . Here we plot various CRB values (15) ( $\sigma = 1$ ) for coefficients satisfying  $|c_{m,0}| = 1$ ,  $m = 0, \dots, P$  when we vary  $L$  in (17). For any value of  $P$  the CRB is minimised when  $L = P + 1$  (note that all the lines are monotonically increasing).

In summary, the best exponential reproducing kernels should reproduce exponentials with exponents of the form  $\alpha_m = j\frac{\pi}{P+1}(2m - P)$  and have  $|c_{m,0}| = 1$  for  $m = 0, \dots, P$ . Finally, whenever possible, the order of the kernel (which equals the number of moments) should be  $P + 1 = N$ . In the next section we show how to obtain such kernels.

##### B. Exponential MOMS

Equipped with the analysis of the previous section, we now design optimal exponential reproducing kernels of maximum-order and minimum-support (e-MOMS). We require  $|c_{m,0}| = 1$  for  $m = 0, \dots, P$  and exponential parameters of the form:

$$\alpha_m = j\omega_m = j\frac{\pi}{P+1}(2m - P) \quad m = 0, \dots, P. \quad (18)$$

By taking into account that any exponential reproducing kernel  $\varphi(t)$  can be written as  $\varphi(t) = \gamma(t) * \beta_{\tilde{\alpha}}(t)$ , we design  $\gamma(t)$  so that  $|c_{m,0}| = 1$  is satisfied. We note that, by using (3), we have that

$$e^{\alpha_m t} = c_{m,0} \sum_{n \in \mathbb{Z}} e^{\alpha_m n} \varphi(t - n).$$

Consequently

$$\begin{aligned} 1 &= c_{m,0} \sum_{n \in \mathbb{Z}} e^{\alpha_m(n-t)} \varphi(t - n) \\ &\stackrel{(a)}{=} c_{m,0} \sum_{k \in \mathbb{Z}} \hat{\varphi}(\alpha_m + j2\pi k) e^{j2\pi k t} \\ &\stackrel{(b)}{=} c_{m,0} \hat{\varphi}(\alpha_m), \end{aligned}$$

where (a) follows from Poisson summation formula<sup>3</sup> and (b) from the application of the generalised Strang-Fix conditions (11). Therefore, we have that for any exponential reproducing kernel  $c_{m,0} = \hat{\varphi}(\alpha_m)^{-1}$ . We then realise that imposing  $|c_{m,0}| = 1$  is equivalent to requiring  $|\hat{\varphi}(\alpha_m)| = 1$ . Finally, by using the fact that  $\hat{\varphi}(\alpha_m) = \hat{\gamma}(\alpha_m) \hat{\beta}_{\tilde{\alpha}}(\alpha_m)$  and evaluating the Laplace transforms at  $\alpha_m = j\omega_m$ , we arrive at the following condition on  $\hat{\gamma}(j\omega_m)$ :

$$\begin{aligned} |\hat{\varphi}(j\omega_m)| &= |\hat{\gamma}(j\omega_m) \hat{\beta}_{\tilde{\alpha}}(j\omega_m)| = 1 \quad \leftrightarrow \\ |\hat{\gamma}(j\omega_m)| &= |\hat{\beta}_{\tilde{\alpha}}(j\omega_m)|^{-1}, \end{aligned} \quad (19)$$

where we now work with the Fourier transform of each function.

<sup>3</sup>Poisson summation:  $\sum_{n \in \mathbb{Z}} f(t + nT) = \frac{1}{T} \sum_{k \in \mathbb{Z}} \hat{f}(j\frac{2\pi k}{T}) e^{j2\pi k \frac{t}{T}}$ .

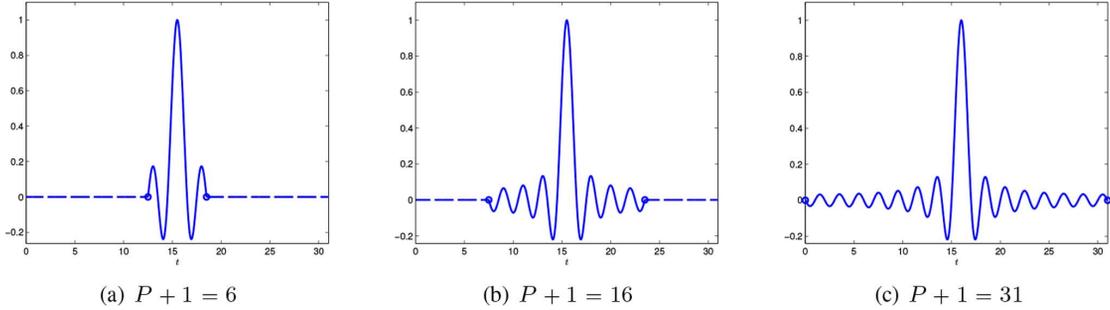


Fig. 4. *Examples of exponential MOMS.* These are 3 of the 30 possible kernels with support  $P + 1 \leq N = 31$  samples. They coincide with one period of the Dirichlet kernel of period  $P + 1$  for  $P$  even or  $2(P + 1)$  for  $P$  odd (see Appendix B). All of them are built selecting the phase of  $c_{m,0}$  such that they are continuous-time functions centred around  $\Delta = \lfloor \frac{P+1}{2} \rfloor$ , where  $\lfloor \cdot \rfloor$  indicates rounded to the nearest integer. They are shown in the middle of the sampling interval for  $T = 1$ . (a)  $P + 1 = 6$ ; (b)  $P + 1 = 16$ ; (c)  $P + 1 = 31$ .

Among all the admissible kernels satisfying (19), we are interested in the one with the shortest support  $P + 1$ . We thus consider the kernels given by a linear combination of various derivatives of the original E-Spline  $\beta_{\tilde{\alpha}}(t)$ , i.e.,:

$$\varphi(t) = \sum_{\ell=0}^P d_{\ell} \beta_{\tilde{\alpha}}^{(\ell)}(t), \quad (20)$$

where  $\beta_{\tilde{\alpha}}^{(\ell)}(t)$  is the  $\ell$ th derivative of  $\beta_{\tilde{\alpha}}(t)$ , with  $\beta_{\tilde{\alpha}}^{(0)}(t) = \beta_{\tilde{\alpha}}(t)$ , and  $d_{\ell}$  is a set of coefficients. This is like saying that  $\gamma(t)$  is a linear combination of the Dirac delta and its derivatives, up to order  $P$  [23]. These kernels are still able to reproduce the exponentials  $e^{\alpha_m t}$  and are a variation of the maximal-order minimal-support (MOMS) kernels introduced in [27]. This is why we call them exponential MOMS (or e-MOMS). They are also a specific case of the broader family of generalised E-Splines presented in [28]. The advantage of this formulation is twofold: first the modified kernel  $\varphi(t)$  is of minimum support  $P + 1$ , the same as that of  $\beta_{\tilde{\alpha}}(t)$ ; second we only need to find the coefficients  $d_{\ell}$  that meet the constraint (19), in order to achieve  $|c_{m,0}| = 1$ . Using the Fourier transform of (20), which is given by:

$$\hat{\varphi}(j\omega) = \hat{\beta}_{\tilde{\alpha}}(j\omega) \sum_{\ell=0}^P d_{\ell} (j\omega)^{\ell},$$

we realise that we can satisfy (19) by choosing the coefficients  $d_{\ell}$  so that the resulting polynomial  $\hat{\gamma}(j\omega) = \sum_{\ell} d_{\ell} (j\omega)^{\ell}$  interpolates the set of points  $(j\omega_m, |\hat{\beta}_{\tilde{\alpha}}(j\omega_m)|^{-1})$  for  $m = 0, 1, \dots, P$ .

Once we have designed the kernels satisfying that  $c_{m,0}$  has modulus one for all  $m$ , we are left with a phase ambiguity. Hypothesising a linear phase behavior, this ambiguity can be reduced to a time shift  $\Delta$  for the E-Spline in (20), introducing an additional degree of freedom. It is possible to show that, in order for the exponential MOMS with  $|c_{m,0}| = 1$  and parameters (18) to be continuous-time functions, then  $c_{m,0} = |c_{m,0}| e^{j\omega_m \Delta}$  for  $m = 0, \dots, P$ , where  $\Delta$  is an integer larger than or equal to 1 and smaller than or equal to  $P$ .

In Fig. 4 we present some of the kernels obtained by implementing the procedure explained above. Interestingly, as shown in Appendix B, these specific functions always equal one period of the Dirichlet kernel. We also point out that when  $P + 1 =$

$N$  the scenario derived using this family of exponential reproducing kernels converges to the original FRI formulation of [1] when we periodise the input or, equivalently, the sampling kernel.

## V. UNIVERSAL SAMPLING OF SIGNALS WITH FRI

In the previous section we have shown how to design optimal exponential reproducing kernels for noisy FRI sampling. In many practical circumstances, however, the freedom to choose the sampling kernel  $\varphi(t)$  is a luxury we may not have.

Essential in the FRI setting is the ability of  $\varphi(t)$  to reproduce exponentials, because this allows us to map the signal reconstruction problem to Prony's method in spectral-line estimation theory. In this section we relax this condition and consider any function  $\varphi(t)$  for which the exponential reproduction property (3) does not necessarily hold. For these functions it is still possible to find coefficients  $c_{m,n}$  such that the reproduction of exponentials is approximate rather than exact. We propose to use this approximate reproduction and the corresponding coefficients  $c_{m,n}$  to retrieve FRI signals from the samples obtained using these kernels.

This new approach has several advantages: First, it is universal in that it can be used with any kernel  $\varphi(t)$ . In fact, as we shall show in the following sections, this new formulation does not even require an exact knowledge of the kernel. Second, while reconstruction of FRI signals with this new method is not going to be exact, we will show that in many cases a proper iterative algorithm can make the reconstruction error arbitrarily small. Finally, it can be used to increase the resiliency to noise of some unstable kernels proposed in the FRI literature. For example, kernels like polynomial splines or the Gaussian function lead to very ill-conditioned reconstruction procedures. We show that by replacing the original  $\mathbf{C}$  with the one formed from properly chosen coefficients  $c_{m,n}$ , based on approximate reproduction of exponentials, we achieve a much more stable reconstruction with the same kernels.

### A. Approximate Reproduction of Exponentials

Assume we want to use the linear combination of a function  $\varphi(t)$  and its integer shifts to approximate the exponential  $e^{\alpha t}$ . Specifically, we want to find the coefficients  $c_n$  such that:

$$\sum_{n \in \mathbb{Z}} c_n \varphi(t - n) \approx e^{\alpha t}. \quad (21)$$

This approximation is exact only when  $\varphi(t)$  satisfies the generalised Strang-Fix conditions (11). For any other function it is of particular interest to find the coefficients  $c_n$  that best fit (21). In order to do so, we directly use<sup>4</sup>  $c_n = c_0 e^{\alpha n}$  and introduce the 1-periodic function

$$g_\alpha(t) = c_0 \sum_{n \in \mathbb{Z}} e^{-\alpha(t-n)} \varphi(t-n). \quad (22)$$

We then find that approximating the exponential  $e^{\alpha t}$  with integer shifts of  $\varphi(t)$  can be transformed into approximating  $g_\alpha(t)$  by the constant value 1. The reason is that we can rewrite (21) in the form of the right-hand side of (22) by substituting  $c_n = c_0 e^{\alpha n}$  and moving  $e^{\alpha t}$  to the left-hand side.

As a consequence of Poisson summation formula, we have that the Fourier series expansion of  $g_\alpha(t)$  is given by

$$g_\alpha(t) = \sum_{l \in \mathbb{Z}} g_l e^{j2\pi l t} = \sum_{l \in \mathbb{Z}} c_0 \hat{\varphi}(\alpha + j2\pi l) e^{j2\pi l t}$$

and that our approximation problem reduces to:

$$g_\alpha(t) = \sum_{l \in \mathbb{Z}} c_0 \hat{\varphi}(\alpha + j2\pi l) e^{j2\pi l t} \approx 1. \quad (23)$$

This shows more deeply the relation between the generalised Strang-Fix conditions (11) and the approximation of exponentials. If  $\varphi(t)$  satisfies the generalised Strang-Fix conditions (11) then  $\hat{\varphi}(\alpha + j2\pi l) = 0$  for  $l \in \mathbb{Z} \setminus \{0\}$  and (23) holds exactly when  $c_0 \hat{\varphi}(\alpha) = 1$ . Otherwise, the terms  $\hat{\varphi}(\alpha + j2\pi l)$  for  $l \in \mathbb{Z} \setminus \{0\}$  do not vanish, and we can only find the coefficient  $c_0$  so that  $g_\alpha(t) \approx 1$ . However, the closer the values  $\hat{\varphi}(\alpha + j2\pi l)$  are to zero, the better the approximation in (21) is.

In general  $\varphi(t)$  can be any function and we can find different sets of coefficients  $c_n$  in order for (21) to hold. Regardless of the coefficients we use, we can determine the accuracy of our approximation by using the Fourier series expansion of  $g_\alpha(t)$ . In fact, the error of approximating  $f(t) = e^{\alpha t}$  by the function  $s(t) = \sum_{n \in \mathbb{Z}} c_n \varphi(t-n)$  with coefficients  $c_n = c_0 e^{\alpha n}$  is equal to:

$$\begin{aligned} \varepsilon(t) &= f(t) - s(t) = e^{\alpha t} [1 - g_\alpha(t)] \\ &= e^{\alpha t} \left[ 1 - c_0 \sum_{l \in \mathbb{Z}} \hat{\varphi}(\alpha + j2\pi l) e^{j2\pi l t} \right]. \end{aligned} \quad (24)$$

Note that, if the Laplace transform of  $\varphi(t)$  decays sufficiently quickly, very few terms of the Fourier series expansion are needed to have an accurate bound for the error.

A natural choice of the coefficients  $c_n = c_0 e^{\alpha n}$  is the one given by the *least-squares approximation*. Despite the fact that  $f(t)$  is not square-integrable, we can still obtain the coefficients by computing the orthogonal projection of  $f(t)$  onto the subspace spanned by  $\varphi(t-n)$  [29]. They are

$$c_n = \frac{\hat{\varphi}(-\alpha)}{\hat{a}_\varphi(e^\alpha)} e^{\alpha n},$$

where  $\hat{a}_\varphi(e^\alpha) = \sum_{l \in \mathbb{Z}} a_\varphi[l] e^{-\alpha l}$  is the  $z$ -transform of  $a_\varphi[l] = \langle \varphi(t-l), \varphi(t) \rangle$ , evaluated at  $z = e^\alpha$ .

<sup>4</sup>The exact exponential reproducing coefficients always satisfy  $c_n = c_0 e^{\alpha n}$ . We now anticipate that different sets of approximation coefficients we derive throughout the section also have the same form.

TABLE I  
COEFFICIENTS FOR THE APPROXIMATE REPRODUCTION (21)

Type	Coefficients
Least-squares	$c_n = \frac{\hat{\varphi}(-\alpha)}{\hat{a}_\varphi(e^\alpha)} e^{\alpha n}$
Constant least-squares	$c_n = \hat{\varphi}(\alpha)^{-1} e^{\alpha n}$
Interpolation	$c_n = \frac{1}{\sum_{l \in \mathbb{Z}} e^{-\alpha l} \varphi(l)} e^{\alpha n}$

The least-squares approximation has the disadvantage that it requires exact knowledge of  $\varphi(t)$ . However, as we stated before, if the Laplace transform of  $\varphi(t)$  decays sufficiently quickly, we can assume the terms  $\hat{\varphi}(\alpha + j2\pi l)$  are close to zero for  $l \in \mathbb{Z} \setminus \{0\}$ . In this case we have that the error in (24) is easily minimised by choosing  $c_0 = \hat{\varphi}(\alpha)^{-1}$ . We denote this second type of approximation *constant least-squares*. Besides its simplicity, a second advantage of choosing  $c_n = \hat{\varphi}(\alpha)^{-1} e^{\alpha n}$  is that it requires only the knowledge of the Laplace transform of  $\varphi(t)$  at  $\alpha$ . If we put ourselves in the FRI setting where we require the approximate reproduction of the exponentials  $e^{\alpha_m t}$  with  $m = 0, \dots, P$ , then this simplified formulation needs only the knowledge of the Laplace transform of  $\varphi(t)$  at  $\alpha_m$ ,  $m = 0, \dots, P$ .

Finally, a third interesting choice of coefficients is the one that ensures that  $s(t)$  *interpolates*  $f(t)$  exactly at integer points in time  $t = \ell \in \mathbb{Z}$  [22], [30]. These coefficients are as follows:

$$c_n = \frac{1}{\sum_{l \in \mathbb{Z}} e^{-\alpha l} \varphi(l)} e^{\alpha n}.$$

Note that in order to use the interpolation coefficients we only need information on  $\varphi(t)$  at integer instants of time. We summarise the previous results in Table I.

According to our experience, in most cases, the constant least-squares approximation is just as good as the least-squares approximation and has the advantage of requiring only the knowledge of the Laplace transform of the kernel at  $s = \alpha$ . Interpolation coefficients are also very easy to compute. However, they always provide a worse approximation quality. Therefore, for the rest of the paper, we use constant least-squares approximation and constant least-squares coefficients.

We show an example of the above analysis in Fig. 5. Here we want to approximate exponentials using linear combinations of integer shifts of a linear spline. First, note that this spline reproduces polynomials of orders 0 and 1 exactly, as shown in Figs. 5(a)–5(b). Then, with the same function, we approximately reproduce 4 complex exponentials  $e^{\alpha_m t} = e^{j \frac{\pi}{16} (2m-7)t}$  for  $m = 0, \dots, 3$ , using the constant least-squares coefficients  $c_{m,n} = \hat{\varphi}(\alpha_m)^{-1} e^{\alpha_m n}$ . We present the approximation of their real part in Figs. 5(c)–5(f). We notice that some exponentials are better approximated than others, in this example the ones with lower frequency. If we used a higher order spline, the approximation quality would improve. However, we have chosen a linear spline for illustration purposes. Also note that the number of exponentials that can be approximated is now independent of the order of the spline.

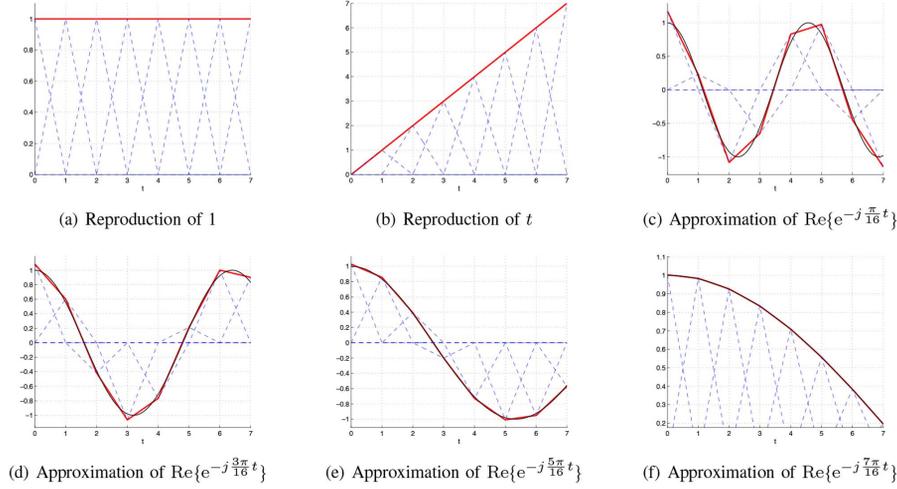


Fig. 5. *B-Spline kernel reproduction and approximation capabilities*. Figures (a-b) show the exact reproduction of polynomials of orders 0 and 1 with a linear spline. Figures (c-f) show the approximation of the real parts of 4 complex exponentials:  $e^{\alpha_m t} = e^{j\frac{\pi}{16}(2m-7)t}$  for  $m = 0, \dots, 3$ , with the constant least-squares coefficients  $c_{m,n} = \hat{\varphi}(\alpha_m)^{-1} e^{\alpha_m n}$ , using a linear spline. We plot the weighted and shifted versions of the splines with dashed blue lines, the reconstructed polynomials and exponentials with red solid lines, and the original functions to be reproduced with solid black lines. (a) Reproduction of 1; (b) Reproduction of  $t$ ; (c) Approximation of  $\text{Re}\{e^{-j\frac{\pi}{16}t}\}$ ; (d) Approximation of  $\text{Re}\{e^{-j\frac{3\pi}{16}t}\}$ ; (e) Approximation of  $\text{Re}\{e^{-j\frac{5\pi}{16}t}\}$ ; (f) Approximation of  $\text{Re}\{e^{-j\frac{7\pi}{16}t}\}$ .

### B. Approximate FRI Recovery

Consider again the stream of Diracs  $x(t) = \sum_{k=0}^{K-1} a_k \delta(t - t_k)$  and the noiseless samples

$$y_n = \left\langle x(t), \varphi\left(\frac{t}{T} - n\right) \right\rangle = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right). \quad (25)$$

We want to retrieve the locations and amplitudes of the Diracs from the samples (25), but now we make no assumption on the sampling kernel. We find proper coefficients for  $\varphi(t)$  to approximate the exponentials  $e^{\alpha_m t}$ , where  $m = 0, \dots, P$ ,  $\alpha_m = \alpha_0 + m\lambda$ , and  $\alpha_0, \lambda \in \mathbb{C}$ . From the previous section we know that a good approximation is achieved if we choose  $c_{m,n} = c_{m,0} e^{\alpha_m n}$  with  $c_{m,0} = \hat{\varphi}(\alpha_m)^{-1}$ . We thus only need to know the Laplace transform of  $\varphi(t)$  at  $\alpha_m$ ,  $m = 0, \dots, P$ . Also, note that  $P$  no longer needs to be related to the support of  $\varphi(t)$ , but we can use any value subject to  $P + 1 \geq 2K$ .

In order to retrieve the innovation parameters  $\{t_k, a_k\}_{k=0}^{K-1}$ , we proceed as in the case of exact reproduction of exponentials, but now using (24) we have that the moments are

$$\begin{aligned} s_m &= \sum_{n=0}^{N-1} c_{m,n} y_n = \left\langle x(t), \underbrace{\sum_{n=0}^{N-1} c_{m,n} \varphi\left(\frac{t}{T} - n\right)}_{e^{\alpha_m \frac{t}{T}} - \varepsilon_m\left(\frac{t}{T}\right)} \right\rangle \\ &= \sum_{k=0}^{K-1} x_k u_k^m - \underbrace{\sum_{k=0}^{K-1} a_k \varepsilon_m\left(\frac{t_k}{T}\right)}_{\zeta_m} \end{aligned} \quad (26)$$

where  $x_k = a_k e^{\alpha_0 \frac{t_k}{T}}$  and  $u_k = e^{\lambda \frac{t_k}{T}}$ . There is a model mismatch due to the approximation error  $\varepsilon_m(t)$  of (24), equal to  $\zeta_m$ . We treat it as noise and retrieve the parameters of the signal using the methods of Section II-B. The model mismatch depends on the quality of the approximation, dictated by the coefficients  $c_{m,n}$ , the parameters  $\alpha_m$  and  $P$ , and the kernel  $\varphi(t)$ .

The estimation of the Diracs can be refined using the iterative algorithm shown in the box Algorithm 2. The basic idea of the algorithm is that, given an estimate of the locations of the Diracs, we can compute an approximation of  $\zeta_m$  and use it to refine the computation of the moments  $s_m$ . In noisy scenarios, if  $\zeta_m$  is negligible when compared to other forms of noise then the procedure is sufficiently good.

#### Algorithm 2 Approximate recovery of a train of $K$ Diracs

- 1) Calculate the approximation coefficients  $c_{m,n}$  by choosing one expression from Table I subject to the available information on the kernel  $\varphi(t)$ .
- 2) Compute the moments  $s_m^0 = \sum_n c_{m,n} y_n$ , from the original data  $y_n$ ,  $n = 0, \dots, N - 1$  and set  $s_m^i = s_m^0$ ,  $m = 0, \dots, P$ . The iteration is  $i = 1$ .
- 3) Obtain  $\{u_k^i, x_k^i\}_{k=0}^{K-1}$  by applying either TLS and Cadzow or matrix pencil to the sequence  $s_m^i$ . Determine the locations  $t_k^i$  and amplitudes  $a_k^i$ , for the  $i$ th iteration.
- 4) Recalculate the moments for the next iteration  $i + 1$  by removing the model mismatch from  $s_m^0$ . This can be done by using (26) as follows:

$$s_m^{i+1} = s_m^0 + \sum_{k=0}^{K-1} a_k^i \varepsilon_m\left(\frac{t_k^i}{T}\right),$$

for  $m = 0, \dots, P$  and where  $\varepsilon_m(t)$  is the error of the approximation (24).

- 5) Repeat steps 3 and 4 until convergence of the values  $\{t_k^i, a_k^i\}_{k=0}^{K-1}$ .

### C. How to Select the Parameters $\alpha_m$

In Section IV we have determined that, if we have full control on the design of the sampling kernel, we should use as many moments as samples:  $P + 1 = N$ , the exponential parameters should be purely imaginary and of the form  $\alpha_m = j\frac{\pi}{P+1}(2m - P)$  and the coefficients  $c_{m,n}$  should be such that  $|c_{m,0}| = 1$  for  $m = 0, \dots, P$ . However, in the approximated FRI scenario, the

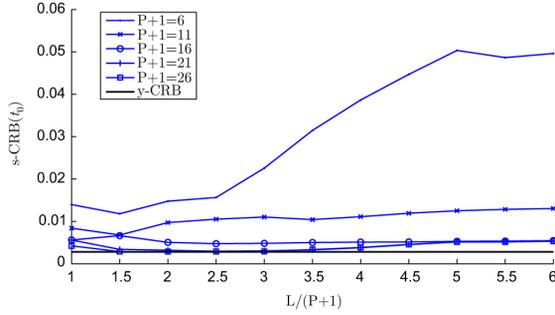


Fig. 6. CRB vs.  $L$ . Here we plot different CRB values (15) ( $\sigma = 1$ ) for exponential parameters (27) when we vary  $L$ . We use  $c_{m,n} = \hat{\varphi}(\omega_m)^{-1} e^{j\omega_m n}$  for  $m = 0, \dots, P$ , where  $\hat{\varphi}(\omega)$  is the Fourier transform of a B-Spline of order 6. Note that the minima are always for  $L$  around  $L = 1.5(P + 1)$ .

sampling kernel is fixed and we can only choose the number of moments  $P + 1$  and the values  $\alpha_m$  but we cannot impose  $|c_{m,0}| = 1$ .

This fact leads to a trade-off in the choice of  $\alpha_m$ . On the one hand, we want them to be purely imaginary:  $\alpha_m = j\omega_m$  and to span the entire unit circle. On the other hand, we want the values  $c_{m,0} = \hat{\varphi}(j\omega_m)^{-1}$  to have modulus as close as possible to 1. Since  $|\hat{\varphi}(j\omega)|$  is normally a low-pass filter, the condition  $|c_{m,0}| \approx 1$  for  $m = 0, \dots, P$  is satisfied when all  $\omega_m$  are very close to zero, which is contrary to spanning the unit circle. We therefore choose the exponential parameters to be of the form:

$$\alpha_m = j\omega_m = j\frac{\pi}{L}(2m - P) \quad m = 0, \dots, P, \quad (27)$$

and then determine  $P$  and  $L$  that optimise the above trade-off.

Again the criterion we follow is to choose the values of  $P$  and  $L$  that minimise the CRB (15) when retrieving the location of a single Dirac. We have found that the best  $P$  is normally greater or equal than the support of the sampling kernel  $\varphi(t)$  and that  $L$  should be in the range  $P + 1 \leq L \leq 4(P + 1)$ . We show an example of the choice of  $L$  in Fig. 6 for the case where  $\varphi(t)$  is a B-Spline of order 6. Here, we first determine a value for  $L$  and compute parameters (27). Then, we calculate the coefficients  $c_{m,n} = \hat{\varphi}(j\omega_m)^{-1} e^{j\omega_m n}$  for  $m = 0, \dots, P$ , where  $\hat{\varphi}(j\omega)$  is the Fourier transform of  $\varphi(t)$ . We finally build matrix  $\mathbf{C}$  and compute (15) for a single Dirac. The minima of the CRB in this example are always around  $L = 1.5(P + 1)$ .

## VI. SIMULATIONS

We now present simulation results to validate the main contributions of the paper. Specifically, we show the performance of the e-MOMS kernels introduced in Section IV-B and of the approximate FRI recovery method introduced in Section V.

### A. The Experimental Setup

We take  $N$  samples following the scheme of Fig. 1 by directly calculating  $y_n = \sum_{k=0}^{K-1} a_k \varphi\left(\frac{t_k}{T} - n\right)$  for  $n = 0, \dots, N - 1$ , since we have a train of  $K$  Diracs as the input. We then either use the noiseless samples or corrupt them with additive white Gaussian noise of variance  $\sigma^2$ . The variance is chosen according to the target signal-to-noise ratio defined as  $\text{SNR}(\text{dB}) = 10 \log \frac{\|y\|^2}{N\sigma^2}$ . We finally compute the  $P + 1$  noisy moments and then retrieve the innovation parameters  $\{a_k, t_k\}_{k=0}^{K-1}$  of the input using the matrix pencil method.

We present results for single realisations of the sampling and reconstruction process or for average performance over multiple trials. For the latter, we are mainly interested in the error in the estimation of the time locations, since these are the most challenging parameters to retrieve. For each Dirac, we show the standard deviation of this error:

$$\Delta t_k = \sqrt{\frac{\sum_{i=0}^{I-1} (\hat{t}_k^{(i)} - t_k)^2}{I}} \quad k = 0, \dots, K - 1, \quad (28)$$

where  $\hat{t}_k^{(i)}$  are the estimated time locations at iteration  $i$  and  $I$  is the total number of iterations. We calculate (28) for a range of fixed signal-to-noise ratios and average the effects using  $I = 1000$  noise realisations for each SNR. We compare the performance (28) with the square root of the variance predicted by the two different Cramér-Rao bounds (CRB) of Section III: the sample-based CRB (13) and the moment-based CRB (15).

### B. Exponential MOMS

In Figs. 7(a)–7(b) we present simulation results when we retrieve  $K = 2$  Diracs from  $N = 31$  samples using a standard E-Spline and the exponential MOMS kernels of Section IV-B. The former are characterised by purely imaginary exponents  $\alpha_m = j\frac{\pi}{2(P+1)}(2m - P)$  for  $m = 0, \dots, P$ . The sampling period is such that  $\tau = NT = 1$ .

We see that for any order  $P + 1$ , e-MOMS outperform E-splines. Moreover, e-MOMS always achieve the moment-based CRB (in red and denoted s-CRB in the legend). This bound gets closer to the sample-based CRB (in black and denoted y-CRB in the legend) as the value of  $P + 1$  increases and as expected matches it when  $P + 1 = N$ .

To further illustrate the stability of e-MOMS, in Fig. 7(c) we show the retrieval of  $K = 20$  Diracs randomly spaced over  $\tau = NT = 1$  and with arbitrary amplitudes. We obtain  $N = 61$  samples, contaminate them with AWGN of signal-to-noise ratio equal to 15 dB and estimate the Diracs from  $P + 1 = N$  moments.

### C. Approximate FRI Recovery

In this section we apply the approximate FRI sampling methods of Section V to two unstable kernels: B-Splines and Gaussian kernel. We show that this new framework leads to much more precise reconstructions than those obtained using the traditional exact FRI recovery methods.

1) *The B-Spline Case:* In this example we compare the performance of our method with the traditional recovery strategy based on the exact reproduction of polynomials [2]. We sample a stream of  $K = 6$  Diracs that has been filtered with a B-Spline kernel of order  $M + 1 = 16$ . We then reconstruct the input from  $N = 31$  noisy samples by obtaining  $P + 1$  moments using the constant least-squares coefficients. These take the form  $c_{m,n} = \hat{\varphi}(\alpha_m)^{-1} e^{\alpha_m n}$ , where  $\hat{\varphi}(s)$  represents the Laplace transform of the B-Spline. We adjust  $L$  as suggested in Section V-C to choose appropriate exponential parameters  $\alpha_m = j\frac{\pi}{L}(2m - P)$ ,  $m = 0, \dots, P$ . The Diracs are located at random over  $\tau = NT = 1$  and have arbitrary amplitudes. The signal-to-noise ratio is  $\text{SNR} = 25$  dB.

In Fig. 8(b) we present the estimation for the method based on reproduction of polynomials with  $M + 1 = 16$  moments.

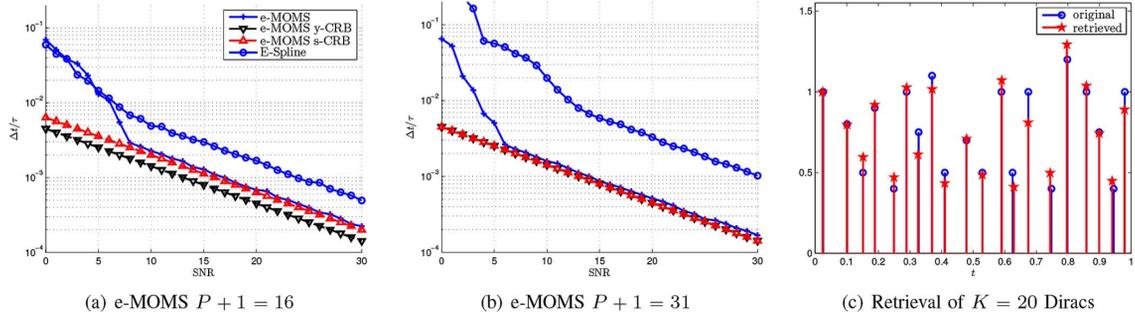


Fig. 7. *Performance of e-MOMS kernels.* (a-b) compare the performance of e-MOMS and E-Splines of different orders  $P + 1$  when noise is added to  $N = 31$  samples. We show the recovery of the first of  $K = 2$  Diracs. Note that e-MOMS always outperform E-splines and achieve the moment-based CRB (s-CRB). This bound gets closer to the sample-based CRB (y-CRB) as the value of  $P + 1$  increases and matches it when  $P + 1 = N$ . Finally, (c) shows the retrieval of  $K = 20$  Diracs randomly spaced over  $\tau = NT = 1$ . The signal-to-noise ratio is 15 dB, and we use  $N = 61$  samples and  $P + 1 = N$  moments. (a) e-MOMS  $P + 1 = 16$ ; (b) e-MOMS  $P + 1 = 31$ ; (c) Retrieval of  $K = 20$  Diracs.

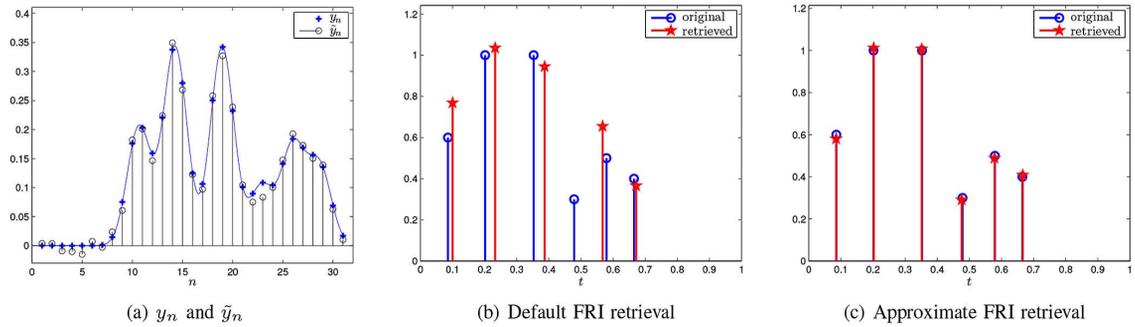


Fig. 8. *Estimation of multiple Diracs with the B-Spline kernel.* We recover  $K = 6$  Diracs from (a)  $N = 31$  noisy samples taken by a kernel of order  $M + 1 = 16$ . (b) Default polynomial recovery of [2], enhanced using pre-whitening. (c) Approximate recovery with  $\alpha_m = j \frac{\pi}{1.5(P+1)}(2m - P)$ ,  $m = 0, \dots, P$  where  $P + 1 = 21$ . The SNR in both cases is 25 dB. (a)  $y_n$  and  $\tilde{y}_n$ ; (b) Default FRI retrieval; (c) Approximate FRI retrieval.

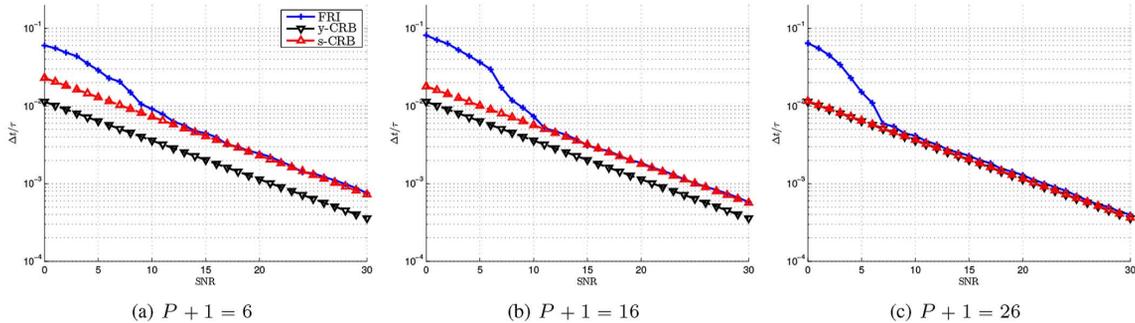


Fig. 9. *Approximate retrieval using a B-Spline.* These figures show the error in the estimation of the first Dirac out of  $K = 2$  retrieved using the approximated FRI recovery. We show how, even when we fix the order of the kernel  $P + 1$ , we can reconstruct any number of moments  $P + 1$  and improve the performance. In fact, with the appropriate choice  $L = 1.5(P + 1)$  the performance improves until the sample-based CRB is reached. (a)  $P + 1 = 6$  (b)  $P + 1 = 16$ ; (c)  $P + 1 = 26$ .

Note that not all the Diracs can be found. In Fig. 8(c) we show the estimation given by the method based on approximation of exponentials, when we select  $L = 1.5(P + 1)$  and generate  $P + 1 = 21$  exponential moments. In this case all the Diracs are retrieved with a root mean squared error on the estimation of the locations of the order of  $10^{-3}$ .

We show further results when we use the approximate method to retrieve  $K = 2$  Diracs from  $N = 31$  noisy samples taken by a B-Spline kernel of order  $M + 1 = 6$ . We use exponential parameters  $\alpha_m = j \frac{\pi}{L}(2m - P)$  with  $m = 0, \dots, P$  and  $L = 1.5(P + 1)$ . In Fig. 9 we show that, even though the order of the kernel is fixed at  $M + 1 = 6$ , we improve the performance by generating more moments, that is, by choosing  $P > M$ . As

the number of moments increases, the performance improves to eventually reach the sample-based CRB as shown in Fig. 9(c).

2) *The Gaussian Kernel Case:* We conclude this set of simulations by showing an example with the Gaussian kernel  $h_\gamma(t) = e^{-t^2/(2\gamma^2)}$  [1]. Since for this kernel the samples are  $y_n = \langle x(t), h_\gamma(nT - t) \rangle$  the approximation problem (21) becomes:

$$e^{\alpha' t} \approx \sum_{n \in \mathbb{Z}} c_n h_\gamma(t - nT).$$

In order to use the coefficients of Table I we need to consider  $\alpha' = \frac{\alpha}{T}$  and  $\varphi(t) = h_\gamma(Tt)$  and manipulate the original expressions. For example, the constant least-squares coefficients

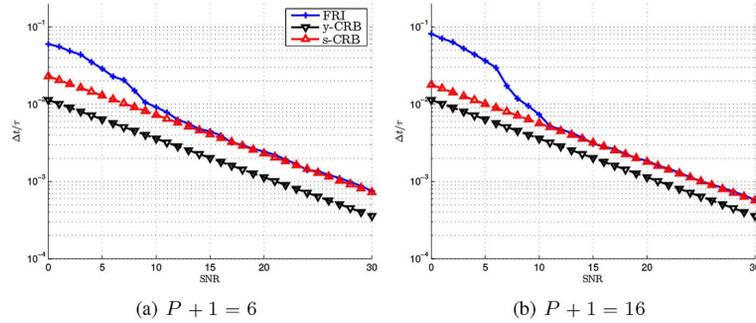


Fig. 10. *Approximate FRI with the Gaussian kernel*. We recover  $K = 5$  Diracs from (a)  $N = 61$  noisy samples taken by a Gaussian kernel of standard deviation  $\gamma = 0.09677$  with period  $T = \frac{1}{31}$ . (b) Recovery using the approximate method with  $\alpha_m = j \frac{\pi}{3.5T(P+1)}(2m - P)$ ,  $m = 0, \dots, P$  where  $P + 1 = 21$ . The SNR is 25 dB. (a)  $y_n$  and  $\hat{y}_n$  (b) Approximate FRI retrieval.

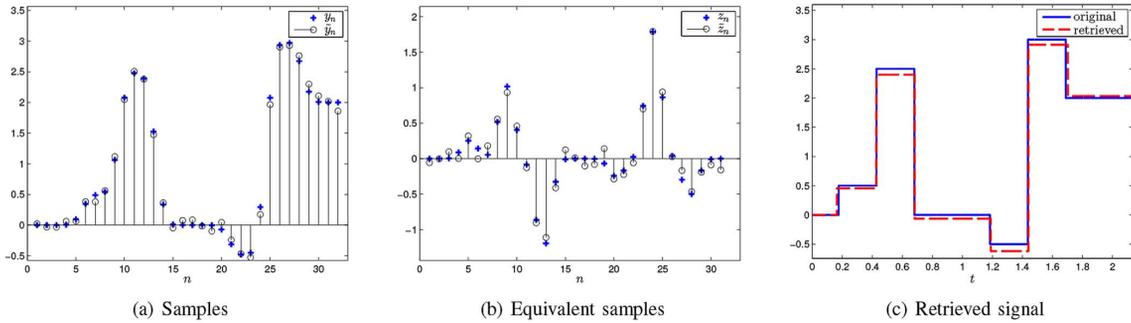


Fig. 11. *Piecewise constant functions and B-Splines*. These figures show the sampling and retrieval process, based on approximation of exponentials, for a piecewise constant function with  $K = 6$  discontinuities in the presence of noise of 25 dB. (a) Samples; (b) Equivalent samples; (c) Retrieved signal.

are now given by  $c_{m,n} = T \hat{h}_\gamma(\alpha'_m)^{-1} e^{\alpha'_m n T}$ . Here, again  $\alpha'_m = j \frac{\pi}{L}(2m - P)$ ,  $m = 0, \dots, P$  and we adjust  $L$  as suggested in Section V-C.

In Fig. 10 we show the reconstruction of  $K = 5$  Diracs from  $N = 61$  noisy samples taken by a Gaussian kernel of standard deviation  $\gamma = 0.09677$  with sampling period  $T = \frac{1}{31}$ . The signal-to-noise ratio is  $\text{SNR} = 25$  dB. We choose  $L = 3.5T(P + 1)$  to compute the constant least-squares coefficients and generate  $P + 1 = 21$  moments to estimate the parameters of the Diracs. The Diracs have random locations in the interval  $[0, 1.2]$ . Fig. 10(b) shows that all the Diracs are correctly retrieved with a root mean squared error on the estimation of the locations of the order of  $3 \cdot 10^{-3}$ . We also note that this kernel is so unstable that the traditional FRI exact recovery method of [11] fails completely in this case.

#### D. Approximation Error and Accuracy of the Reconstruction

In this example we test the hypothesis that better approximation of exponentials leads to more accurate reconstruction of Diracs. Assume we sample a single Dirac with a linear spline and we recover its location using approximation of exponentials. In Fig. 5 we have shown that the linear spline can approximate complex exponentials of lower frequencies better than those with higher frequencies. We now generate four moments  $s_m$  using the constant least-squares coefficients that are associated to the same exponentials  $e^{\alpha_m t} = e^{j \frac{\pi}{16}(2m-7)t}$  for  $m = 0, \dots, 3$  of Fig. 5. Finally, we compare the estimation of the location of the Dirac obtained from the moments associated to the higher frequencies (HF)  $s_0$  and  $s_1$  to the estimation obtained

TABLE II  
ACCURACY OF THE RECONSTRUCTION.

	HF		LF	
	$s_0$	$s_1$	$s_2$	$s_3$
Approximation error	0.061	0.028	0.0093	0.00098
Reconstruction error	0.0018		0.00019	

from the moments associated to the lower frequencies (LF)  $s_2$  and  $s_3$ .

In Table II we show the root mean squared error of the estimation obtained from either pair of moments. The error is averaged over 100 realisations each of which corresponds to placing the Dirac at  $t_0 = 0.15(i - 1)$  for  $i = 1, \dots, 100$ . As expected the approximation with lower frequency achieves a better performance.

#### E. Alternative FRI Signals

We conclude the simulations by showing that it is possible to adapt the approximate FRI framework to sample and reconstruct alternative FRI signals. For example, sampling a piecewise constant function with a kernel  $\varphi(t)$  and calculating the first finite difference of the samples  $z_n = y_n - y_{n+1}$  yields the same measurements as sampling the derivative of the signal with  $\phi(t) = \varphi(t) * \beta_0(t)$ , where  $\beta_0(t)$  is a box function [2]. The derivative of the signal is a train of  $K$  Diracs. Consequently, we may recover the signal by calculating  $c_{m,n}$  for the linear combination of shifted versions of  $\phi(t)$  to approximate exponentials and then applying the annihilating filter method to the moments  $s_m = \sum_n c_{m,n} z_n$ .

We illustrate the process in Fig. 11. Here, we sample a piecewise constant function with  $K = 6$  discontinuities using a B-Spline kernel of order  $M + 1 = 6$ . The sampling period is  $T = \frac{1}{15}$ . In Fig. 11(a) we show the  $N = 32$  samples contaminated with additive white Gaussian noise and in Fig. 11(b) we show their first order difference. Then, we generate  $P + 1 = 21$  moments using the constant least-squares coefficients from exponential parameters  $\alpha_m = j\frac{\pi}{L}(2m - P)$  with  $m = 0, \dots, P$  and  $L = 1.4(P + 1)$ . The signal-to-noise ratio is 25 dB. Note that the order of the spline is not sufficient to apply the retrieval method based on reproduction of polynomials of [2]. On the contrary, we can use the method based on approximation of exponentials as long as  $P + 1 \geq 2K$ . The original and reconstructed signal are shown in Fig. 11(c).

## VII. CONCLUSIONS

In this paper we have considered the FRI reconstruction problem in the presence of noise. We have first revisited existing results in the noiseless setting, and the most effective treatment of noise in the current literature. Then, we have studied robust alternatives to the previous line of work.

More specifically, our contribution is twofold: We have determined how to design optimal exponential reproducing kernels, in the sense that they lead to the most stable signal reconstruction. Moreover, we have departed from the ideal situation in which we have full control of the sampling kernel and considered the case where we are given corrupted samples taken with an arbitrary acquisition device. In this situation, we have developed a universal FRI reconstruction strategy that works with any kernel. In contrast to the original FRI framework, which tries to find the exact parameters of the input signal, we have proposed an approximate recovery of the input based on the approximate reproduction of exponentials. The advantage of our new approach is that it can be applied to any sampling kernel and provides more stable and precise reconstructions than those obtained with specific classes of kernels used in the past.

### APPENDIX A

#### GENERALISED STRANG-FIX CONDITIONS

An exponential reproducing kernel is any function  $\varphi(t)$  that, together with a linear combination of its shifted versions, can generate exponential polynomials of the form  $t^r e^{\alpha_m t}$  [22], [28] for  $m = 0, \dots, P$  and  $r = 0, \dots, R$ . The parameters  $\alpha_m$  are in general complex. In this appendix we prove that exponential reproducing kernels satisfy the generalised Strang-Fix conditions. More specifically, a kernel  $\varphi(t)$  is able to reproduce exponential polynomials, i.e.,:

$$t^r e^{\alpha_m t} = \sum_{n \in \mathbb{Z}} c_{m,n,r} \varphi(t - n),$$

if and only if

$$\hat{\varphi}^{(r)}(\alpha_m) \neq 0 \text{ and } \hat{\varphi}^{(r)}(\alpha_m + j2\pi l) = 0$$

for  $l \neq 0$ ,  $r = 0, \dots, R$  and  $m = 0, \dots, P$ . Here,  $\hat{\varphi}^{(r)}(s)$  represents the  $r$ th order derivative of the double-sided Laplace transform of  $\varphi(t)$ .

The proof is obtained from the Strang-Fix conditions for polynomial reproducing kernels, by considering the function

$\psi(t) = e^{-\alpha_m t} \varphi(t)$  that clearly reproduces polynomials of the form  $t^r$  for  $r = 0, \dots, R$ . The Strang-Fix conditions [2], [31] state that a kernel  $\psi(t)$  is able to reproduce polynomials, i.e.,:

$$t^r = \sum_{n \in \mathbb{Z}} c_{r,n} \psi(t - n),$$

if and only if

$$\hat{\psi}(0) \neq 0 \text{ and } \hat{\psi}^{(r)}(j2\pi l) = 0$$

for  $l \neq 0$  and  $r = 0, \dots, R$ . Here,  $\hat{\psi}(j\omega)$  is the Fourier transform of  $\psi(t)$ , and  $\hat{\psi}^{(r)}(j\omega)$  represents its  $r$ th order derivative. Then, by taking into account that the Fourier transform of  $\psi(t)$  is related to the Laplace transform of  $\varphi(t)$  through  $\hat{\psi}(j\omega) = \hat{\varphi}(\alpha_m + j\omega)$ , the above equation turns into the generalised Strang-Fix conditions for  $\varphi(t)$ :

$$\hat{\varphi}(\alpha_m) \neq 0 \text{ and } \hat{\varphi}^{(r)}(\alpha_m + j2\pi l) = 0$$

for  $l \neq 0$ ,  $r = 0, \dots, R$  and  $m = 0, \dots, P$ . Now  $\hat{\varphi}^{(r)}(s)$  represents the  $r$ th order derivative of the double-sided Laplace transform of  $\varphi(t)$ . This proves that a kernel that reproduces exponential polynomials satisfies the generalised Strang-Fix conditions.

The converse is also true. Consider a kernel  $\varphi(t)$  that satisfies the generalised Strang-Fix conditions. Then, a kernel  $\psi(t)$  with Fourier transform  $\hat{\psi}(j\omega) = \hat{\varphi}(\alpha_m + j\omega)$  is guaranteed to satisfy the Strang-Fix conditions and, consequently, reproduces polynomials  $t^r$  for  $r = 0, \dots, R$ . Finally, due to the relation of the kernels in the Laplace domain it follows that  $\psi(t) = e^{-\alpha_m t} \varphi(t)$  implying that  $\varphi(t)$  reproduces exponential polynomials  $t^r e^{\alpha_m t}$  for  $l \neq 0$ ,  $r = 0, \dots, R$  and  $m = 0, \dots, P$ , which completes the proof.

### APPENDIX B

#### E-MOMS INCLUDE THE DIRICHLET AND SOS KERNELS

Let us consider the exponential reproducing kernel  $\varphi_0(t) = \varphi\left(t + \frac{P+1}{2}\right)$  of support  $P + 1$  and centred in zero, with  $\varphi(t) = \gamma(t) * \beta_{\bar{\alpha}}(t)$ , where  $\beta_{\bar{\alpha}}(t)$  is an E-Spline. We restrict our analysis to  $P$  being even and we use exponential parameters

$$\alpha_m = j\omega_m = j\frac{\pi}{P+1}(2m - P), \quad (29)$$

where  $m = 0, \dots, P$ . We next use the  $P + 1$ -periodic extension of  $\varphi_0(t)$ , that is  $\varphi_{P+1}(t) = \sum_{l \in \mathbb{Z}} \varphi_0(t + l(P + 1))$ , which is equivalent to:

$$\varphi_{P+1}(t) = \frac{1}{P+1} \sum_{k \in \mathbb{Z}} \hat{\varphi}_0\left(j\frac{2\pi k}{P+1}\right) e^{j\frac{2\pi k}{P+1}t}, \quad (30)$$

from the application of Poisson summation formula. The case of  $P$  being odd can be derived likewise, but by periodising over  $2(P + 1)$ . Also note that the Fourier transform of the shifted kernel  $\varphi_0(t)$  is equal to:

$$\hat{\varphi}_0(j\omega) = \gamma(j\omega) \prod_{m=0}^P \left(\frac{\omega - \omega_m}{2}\right). \quad (31)$$

The set of equations

$$\hat{\varphi}_0(j\omega_m) = |\hat{\varphi}(j\omega_m)| = |\hat{\gamma}(j\omega_m)\hat{\beta}_{\bar{\alpha}}(j\omega_m)| = \eta_m, \quad (32)$$

are just like (19) and lead to design exponential reproducing kernels of maximum order and minimum support (e-MOMS), different from those of Section IV-B, but that still correspond to a specific subfamily of the generalised exponential reproducing kernels of [28].

In (30) the Fourier transform  $\hat{\varphi}_0(j\omega)$  is evaluated at  $j\omega_k = j\frac{2\pi k}{P+1}$ . Taking into account (32), we know that  $\hat{\varphi}_0(j\omega_k) = \eta_k$  for  $k = -\frac{P}{2}, \dots, \frac{P}{2}$ . We also have that  $\hat{\varphi}_0(j\omega_k) = 0$  for any other  $k$ , because we can find a term in the product (31) equal to  $\text{sinc}(\ell\pi) = 0$ ,  $\ell \in \mathbb{Z}$ . Therefore, (30) can be reduced to:

$$\varphi_{P+1}(t) = \frac{1}{P+1} \sum_{k=-\frac{P}{2}}^{\frac{P}{2}} \eta_k e^{j\frac{2\pi k}{P+1}t}. \quad (33)$$

Note that when the values  $\eta_k = 1$  for all  $k$ , then (33) reduces to one period of the Dirichlet kernel of period  $P+1$ :

$$\varphi_{P+1}(t) = \frac{1}{P+1} \sum_{k=-\frac{P}{2}}^{\frac{P}{2}} e^{j\frac{2\pi k}{P+1}t} = \frac{1}{P+1} \frac{\sin(\pi t)}{\sin(\frac{\pi t}{P+1})}.$$

And this is precisely the  $P+1$ -periodic extension of the e-MOMS kernels of Section IV-B.

To end, we now consider one period of (33) and denote  $t = \frac{x}{T}$ ,  $N = P+1$  and  $\tau = NT = (P+1)T$ . Then we get the time domain definition of the SoS kernel [3]:

$$\varphi_{P+1}\left(\frac{x}{T}\right) = g(x) = \text{rect}\left(\frac{x}{\tau}\right) \frac{1}{N} \sum_{k \in \mathcal{K}} \eta_k e^{j\frac{2\pi k}{\tau}x}.$$

Here, the number of samples  $N$  needs to be odd, since  $P$  is even, and the set of indices  $\mathcal{K} = \{-\frac{N-1}{2}, \dots, \frac{N-1}{2}\}$ .

## REFERENCES

- [1] 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, Jun. 2002.
- [2] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang-fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.
- [3] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.
- [4] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2000.
- [5] 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, pp. 331–334, May 2011.
- [6] J. Berent, P. L. Dragotti, and T. Blu, "Sampling piecewise sinusoidal signals with finite rate of innovation methods," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 613–625, Feb. 2010.
- [7] I. Maravic and M. Vetterli, "Exact sampling results for some classes of parametric non-bandlimited 2-D signals," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 175–189, Jan. 2004.
- [8] P. Shukla and P. L. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3670–3686, Jul. 2007.
- [9] H. Pan, T. Blu, and P. L. Dragotti, "Sampling curves with finite rate of innovation," in *Proc. 9th Int. Workshop Sampling Theory Appl. (SampTA'11)*, Singapore, May 2011.
- [10] C. Chen, P. Marziliano, and A. C. Kot, "2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2851–2859, Jun. 2012.
- [11] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.
- [12] 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, Apr. 2008.
- [13] V. Y. F. Tan and V. K. Goyal, "Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5135–5146, Oct. 2008.
- [14] A. Erdozain and P. M. Crespo, "A new stochastic algorithm inspired on genetic algorithms to estimate signals with finite rate of innovation from noisy samples," *Signal Process.*, vol. 90, pp. 134–144, Jan. 2010.
- [15] A. Erdozain and P. M. Crespo, "Reconstruction of aperiodic FRI signals and estimation of the rate of innovation based on the state space method," *Signal Process.*, vol. 91, no. 8, pp. 1709–1718, 2011.
- [16] J. A. Urigüen, P. L. Dragotti, and T. Blu, "On the exponential reproducing kernels for sampling signals with finite rate of innovation," in *Proc. 9th Int. Workshop Sampling Theory Appl. (SampTA'11)*, Singapore, May 2011.
- [17] I. Khalidov, T. Blu, and M. Unser, "Generalized L-spline wavelet bases," in *Proc. SPIE Conf. Math. Imaging: Wavelet XI*, San Diego, CA, USA, Aug. 3, 2005, vol. 5914, pp. 59140F–1.
- [18] J. A. Cadzow, "Signal enhancement—a composite property mapping algorithm," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, Jan. 1988.
- [19] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped undamped sinusoids in noise," *IEEE Trans. on Acoust., Speech and Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.
- [20] B. D. Rao, "Model based processing of signals: A state space approach," *Proc. IEEE*, vol. 3980, no. 2, pp. 283–309, Feb. 1992.
- [21] Y. C. Eldar and A. V. Oppenheim, "MMSE whitening and subspace whitening," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1846–1851, Jul. 2003.
- [22] M. Unser and T. Blu, "Cardinal exponential splines: Part I—theory and filtering algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1425–1438, Apr. 2005.
- [23] A. Ron, "Factorization theorems for univariate splines on regular grids," *Israel J. Math.*, vol. 70, no. 1, pp. 48–68, 1990.
- [24] E. Ollila, "On the Cramér-Rao bound for the constrained and unconstrained complex parameters," in *Proc. Sensor Array Multichannel Signal Process. Workshop*, Jul. 2008, pp. 414–418.
- [25] L. Berman and A. Feuer, "On perfect conditioning of Vandermonde matrices on the unit circle," *Electron. J. Linear Algebra*, vol. 16, pp. 157–161, Jul. 2007.
- [26] G. H. Golub, P. Milanfar, and J. Varah, "A stable numerical method for inverting shape from moments," *SIAM J. Sci. Comput. (SISC)*, vol. 21, pp. 1222–1243, Dec. 1999.
- [27] T. Blu, P. Thevenaz, and M. Unser, "MOMS: maximal-order interpolation of minimal support," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1069–1080, Jul. 2001.
- [28] M. Unser, "Cardinal exponential splines: Part II—Think analog, act digital," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1439–1449, Apr. 2005.
- [29] M. Unser, A. Aldroubi, and M. Eden, "Polynomial spline signal approximations: Filter design and asymptotic equivalence with Shannon's sampling theorem," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 95–103, Jan. 1992.
- [30] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
- [31] G. Strang and G. Fix, "Fourier analysis of the finite element variational method," *Constructive Aspect of Functional Anal.*, pp. 796–830, 1971.



**Jose Antonio Urigüen (S'12)** received the "Título de Ingeniero de Telecomunicaciones" from the "Universidad del País Vasco", Spain, in 2000 (third out of 100 students); the M.S. degree in Communications and Signal Processing from Imperial College (with Distinction), London, in 2009; and the Ph.D. degree from Imperial College, London, in 2013.

He was one of the 120 recipients of a scholarship by the "Fundación Caja Madrid", awarded for academic merits, to complete the first two years of his Ph.D. He was then supported by the RecoSamp

project and developed advances in sampling theory for which he has one patent pending.

He is currently a research assistant with the University of Deusto, Spain. His research interests include applied mathematics, interpolation, approximation and sampling theory, sparse sampling and biomedical signal processing.



**Thierry Blu** (M'97–SM'06–F'12) was born in Orléans, France, in 1964. He received the “Diplome d'ingénieur” from École Polytechnique, France, in 1986 and from Télécom Paris (ENST), France, in 1988, and the Ph.D. degree in electrical engineering from ENST in 1996 for a study on iterated rational filterbanks, applied to wideband audio coding.

Between 1998 and 2007, he was with the Biomedical Imaging Group at the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland. He is now a Professor in the Department of Electronic Engineering, The Chinese University of Hong Kong.

Dr. Blu was the recipient of two best paper awards from the IEEE Signal Processing Society (2003 and 2006). He is also coauthor of a paper that received a Young Author best paper award (2009) from the same society. He has been an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2002–2006), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006–2010), and Elsevier Signal Processing (2008–2011). He is currently on the board of Eurasip J. on Image and Video Processing (since 2010) and is a member of the IEEE Signal Processing Theory and Methods Technical Committee (since 2008).

Research interests: (multi)wavelets, multiresolution analysis, multirate filterbanks, interpolation, approximation and sampling theory, sparse sampling, image denoising, psychoacoustics, biomedical imaging, optics, wave propagation.



**Pier Luigi Dragotti** (M'02–SM'11) is currently a Reader (Associate Professor) with the Electrical and Electronic Engineering Department, Imperial College London, London, U.K.

He received the Laurea degree (summa cum laude) in electrical engineering from the University of Naples Federico II, Naples, Italy, in 1997; the master's degree in communications systems from the Swiss Federal Institute of Technology of Lausanne (EPFL), Lausanne, Switzerland, in 1998, and the Ph.D. degree from EPFL in April 2002. He has held

several visiting positions at different universities and research centers. He was a Visiting Student with Stanford University, Stanford, CA, USA, 1996, a Researcher with the Mathematics of Communications Department, Bell Labs, Lucent Technologies, Murray Hill, NJ, USA, 2000, a Visiting Professor with the Politecnico of Milan, Milan, Italy, 2010, and a Visiting Scientist with the Massachusetts Institute of Technology, Cambridge, MA, USA, 2011.

He was a Technical Co-Chair for the European Signal Processing Conference in 2012, an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2006 to 2009 and an Elected Member of the IEEE Image, Video and Multidimensional Signal Processing Technical Committee and the IEEE Signal Processing Theory and Methods Technical Committee.

He is a recipient of the ERC Starting Investigator Award for the project Re-coSamp. His work includes sampling theory, wavelet theory and its applications, image and video compression, image-based rendering, and image super-resolution. He has published more than 100 publications and a patent acquired by Qualcomm, Inc., in 2007.