Finite Rate of Innovation Channel Models and DoF of MIMO Multi-User Systems with Delayed CSIT Feedback

Yohan Lejosne, Dirk Slock, Yi Yuan-Wu

Finite Rate of Innovation Channel Models and DoF of MIMO Multi-User Systems with Delayed CSIT Feedback

Yohan Lejosne and Dirk Slock[†]
EURECOM, Mobile Communications Dept.
Campus SophiaTech, 450 route des Chappes
06410 Biot Sophia Antipolis, FRANCE
Email: {yohan.lejosne,dirk.slock}@eurecom.fr

Yi Yuan-Wu
Orange Labs, RESA/WASA
38-40 Rue du General Leclerc
92794 Issy Moulineaux Cedex 9, FRANCE
Email: {yohan.lejosne, yi.yuan}@orange.com

***Abstract*—Channel State Information at the Transmitter (CSIT) is of utmost importance in multi-user wireless networks, in which transmission rates at high SNR are characterized by Degrees of Freedom (DoF, the rate prelog). In recent years, a number of ingenious techniques have been proposed to deal with delayed and imperfect CSIT. However, we show that the precise impact of these techniques in these scenarios depends heavily on the channel model (CM). We introduce the use of linear Finite Rate of Information (FRoI) signals (which could also be called Basis Expansion Model (BEM)) to model time-selective channel coefficients. The FRoI dimension turns out to be well matched to Degree of Freedom (DoF) analysis since the FRoI CM allows compressed feedback (FB) and captures the DoF of the channel coefficient time series. Both the block fading model and the stationary bandlimited channel model are special cases of the FRoI CM. However, the fact that FRoI CMs model stationary channel evolutions allows to exploit one more dimension: arbitrary time shifts. In this way, the FroI CM allows to maintain the DoF unaffected in the presence of CSIT feedback (FB) delay, by increasing the FB rate. We call this Foresighted Channel FB (FCFB). FRoI CM relates also to (predictive) filterbanks and we work out the optimization details in the biorthogonal case (different analysis and synthesis filters). The FRoIC model with multiple basis functions accommodates FB delay beyond the coherence time and handling of users with unequal coherence times.**

I. Introduction

In this paper, Tx and Rx denote transmit/transmitter/ transmitting/transmission and receive/receiver/receiving/reception. Interference is undoubtedly the main limiting factor in multi-user wireless communication systems. Tx side or Rx side zero-forcing (ZF) beamforming (BF) or joint Tx/Rx ZF BF (signal space interference alignment (IA)) allow to obtain significant Degrees of Freedom (DoFs) (= multiplexing factor, or rate prelog). These technique require very good Channel State Information at Tx and Rx (CSIT/CSIR). Especially CSIT is problematic since it requires feedback (FB) which involves delay, which may be substantial if FB Tx is slot based. We shall remark here up front that these observations advocate the design of wireless systems in which the FB delay is made as short as possible. In a TDD system this may be difficult but in a FDD system the FB delay can be made as short as the roundtrip delay! These considerations are independent of the fact that we can find ways to get around FB delay, as we elaborate below, because a reduction in FB delay always leads to improvements (be it in terms of DoF, or NetDoF or at finite SNR).

II. Delayed CSIT State of the Art

It therefore came as a surprise that with totally outdated Delayed CSIT (DCSIT), the MAT scheme [1] is still able to produce significant DoF gains for multi-antenna transmission compared to TDMA. In the DCSIT setting, (perfect) CSIT is available only after a FB delay T_{fb} (T_{fb} taken as the unit of time in number of the following schemes). The channel correlation over T_{fb} can be arbitrary, possibly zero. Perfect overall CSIR is assumed (which leads to significant NetDoF reduction due to CSIR distribution overhead [2], [3]). The MISO BC (Broadcast Channel) and IC (Interference Channel) cases of [1] have been extended to some MIMO cases in [4].

Using a sophisticated variation of the MAT scheme, [5] was able to propose an improved scheme for the case where the FB delay T_{fb} is less than the channel coherence time T_c (define as the inverse of the Doppler bandwidth (BW)). Let's focus on the temporal correlation of one channel coefficient h. The channel FB leads to an estimate and estimation error: $h = \widehat{\widehat{h}} + \widetilde{\widetilde{h}}$ with FB SNR $\frac{\sigma^2_{\widehat{\widehat{h}}}}{\sigma^2_{\widetilde{\widetilde{h}}}} = \mathcal{O}(\rho)$ where ρ is the system SNR. At the Tx, on the basis of $\widehat{\widehat{h}}$, channel prediction over a horizon T_{fb} leads to a prediction with error: $h = \widehat{h} + \widetilde{h}$ with prediction SNR $\frac{\sigma^2_{\widehat{h}}}{\sigma^2_{\widetilde{h}}} = \mathcal{O}(\rho^{1-\frac{T_{fb}}{T_c}})$. The scheme of [5] attains for MISO BC or IC with $K = 2$ users a sumDoF = $2(1 - \frac{T_{fb}}{3T_c}) = 2(\frac{2}{3}\frac{T_{fb}}{T_c} + 1 - \frac{T_{fb}}{T_c})$ using a sophisticated combination of analog and digital FB. The scheme is limited to mostly MISO and to $K = 2$. They also consider: imperfect CSIT (apart from delayed) and the DoF region.

It was generally believed that any delay in the feedback necessarily causes a DoF loss. However, Lee and Heath in [6] proposed a scheme that achieves N_t (sum) DoF in the block fading underdetermined MISO BC with N_t transmit antennas and $K = N_t + 1$ users if the feedback delay is small enough ($\leq \frac{T_c}{K}$). We introduce FRoI channel models and exploit their approximately stationary character to propose a simple ZF scheme based on Foresighted Channel FB (FCFB). The DoF of FCFB ZF are also insensitive to FB delay.

III. Some Channel Model State of the Art

One category of popular channel models is the (first-order) autoregressive (Gauss-Markov) channel model, see e.g. [7]. However, these models (at finite and especially low order) do not allow perfect prediction and hence do not lead to interesting DoF results. These models are called regular in [8]. The two classical (nonregular) channel models that allow permanent perfect CSIT for Doppler rate perfect channel feedback are block fading and bandlimited (BL) stationary channels. The block fading model dates back to the time of GSM where it was quite an appropriate model for the case of frequency hopping. However, though this model is very convenient for very tractable analysis (e.g. for sinngle-user MIMO [9]), it is inappropriate for DoF analysis which works at infinite SNR and requires exact channel models. Now, whereas exact channel models do not exist, channel models for DoF analysis should at least be good approximations. Indeed, mobile speeds and Doppler shifts are finite. This leads to a strictly BL Jakes Doppler spectrum. However, in the Jakes model, the mobile terminal has a certain speed without

[†] EURECOM's research is partially supported by its industrial members: BMW Group R&T, iABG, Monaco Telecom, Orange, SAP, SFR, ST Microelectronics, Swisscom, Symantec, and also by the EU FP7 projects WHERE2 and NEWCOM#.

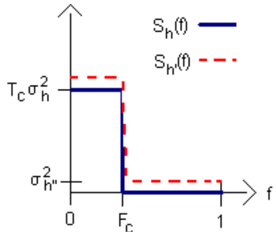

Fig. 1. A bandlimited (BL) Doppler spectrum and its noisy version.

ever moving (attenuation, directions of arrival, path delays, speed vector etc. are all constant forever). In reality, the channel evolution constantly evolves from one temporarily BL Doppler spectrum to another, leading to a possibly overall stationary process but that is not BL.

Another aspect is that there is a difference between channel modeling for CSIR only and for CSIT. In the CSIR case, causality is not much of an issue and channel estimation can be done in a non-causal fashion. Hence block processing and associated channel models as in [7] and references therein are acceptable. In the CSIT case however, the CSI needs to be fed back for adaptation of the Tx. Due to the feedback delay, the channel estimation in the CSIT scenario is necessarily causal (case of prediction). Hence different channel models are required.

IV. The Bandlimited (BL) Doppler Spectrum Case

In an optimal approach, all channel coefficients (in the channel impulse response) need to be treated jointly. However, if no deterministic relations exist between the channel coefficients, then for the purpose of DoF analysis, we may consider the case of i.i.d. channel coefficients. In what follows we consider one such generic channel coefficient h. Its temporal evolution is a stationary discrete-time process, at the sampling rate (channel uses) of the communications channel. We assume this sampling rate to be normalized to 1. We assume the Doppler spectrum $S_h(f)$, the spectrum of the process h, to be bandlimited to F_c, which is the total Doppler Bandwidth (as the channel coefficients are complex, the position of the Doppler spectrum w.r.t. the carrier frequency is less crucial, so we can assume the Doppler support to be $[0, F_c]$ as in Fig. 1; also, $S_h(f)$ is periodic in frequency f with period 1). We denote the coherence time as $T_c = 1/F_c$. Due to the (deterministic) estimation of the channel in the downlink, and its imperfect feedback to the Tx, the Tx has a noisy version $\hat{\hat{h}}$ with additive estimation noise $\tilde{\tilde{h}}$ (h', h'' in Fig. 1) (note that the use of a prior channel distribution in a Bayesian approach can be postponed until the prediction operation to follow). The noisy spectrum is $S_{\hat{\hat{h}}}(f) = S_h(f) + S_{\tilde{\tilde{h}}}(f) = S_h(f) + \sigma^2_{\tilde{\tilde{h}}}$ assuming independent white noise $\tilde{\tilde{h}}$. Let T_{fb} be the delay with which the channel estimate $\hat{\hat{h}}$ arrives at the Tx for (instantaneous) adaptation of the transmitter. That means that the Tx has to perform channel prediction over a horizon of T_{fb}. Assuming a Gaussian channel and estimation noise, linear minimum mean squared error (LMMSE) prediction is optimal (if MMSE is the optimality criterion). Prediction over a horizon of T_{fb} samples will become prediction by one sample if we downsample the channel estimate signal by a factor T_{fb}. Downsampling in the time domain leads to a expansion of the spectrum support by a factor T_{fb} (of course, prediction from a subsampled version is of a degraded quality in the noisy case). Considering Fig. 1, as long as $F_cT_{fb} < 1$ (or $T_{fb} < T_c$), the downsampled channel signal remains bandlimited. Let $S(f)$ denote the downsampled version of $S_{\hat{\hat{h}}}(f)$. Then we get for the (infinite order) one sample ahead prediction MSE

$$\tilde{\sigma}^2 = e^{\int_0^1 \ln S(f)\, df} \sim \sigma_{\tilde{\tilde{h}}}^{2(1-F_cT_{fb})}\,. \tag{1}$$

A similar behavior is obtained for the T_{fb} ahead prediction error from the original unsubsampled process. The prediction error $\tilde{h}'$ considered in (1) is actually the error in estimating $\hat{\hat{h}}$ from its past. However, what we are really interested in is estimating h from the past of $\hat{\hat{h}}$, with prediction error $\tilde{h}$. Now, since $\tilde{\tilde{h}}$ is white noise, we get in fact $\sigma^2_{\tilde{h}} = \sigma^2_{\tilde{h}'} - \sigma^2_{\tilde{\tilde{h}}}$. When $T_{fb} > 0$, the dominating term at high SNR is still $\sigma^2_{\tilde{h}'}$, though.

Let $P(f) = 1 - \sum_{n=1}^{\infty} p_n e^{-j2\pi fn}$ be the (one sample ahead) prediction error filter for $\hat{\hat{h}}$ (monic: $p_0 = 1$). The $-p_k$, $k > 0$ are the coefficients for predicting both $\hat{\hat{h}}$ or h. As infinite order prediction succeeds in whitening the prediction error, we have that

$$S_{\hat{\hat{h}}}(f) = \frac{\tilde{\sigma}^2}{|P(f)|^2} \tag{2}$$

which is the Kolmogorov representation, an infinite order autoregressive (AR(∞)) model. Since $|P(f)|$ is a scaled version of $1/\sqrt{S_{\hat{\hat{h}}}(f)}$, it can easily be seen that $P(f)$ is a high-pass filter, and converges to an ideal high-pass filter as the SNR increases [10]. This has led a number of researchers (see [10] and references therein) to construct predictors for bandlimited signals simply by approximating ideal high-pass filters. These FIR filters are typically chosen to be linear phase and are made monic ($p_0 = 1$) by dividing the filter by its first coefficient. However, the prediction error filter $P(f)$ is not only monic but also minimum-phase.

A. The Noiseless BL Case: two-time scale model

Now consider the noiseless case, $\sigma^2_{\tilde{\tilde{h}}} = 0$. Then clearly the prediction errors become zero, $\sigma^2_{\tilde{h}} = \sigma^2_{\tilde{h}'} = 0$. Hence the signal can be perfectly predicted from its past. For simplicity let T_c be an integer. Let h_k denote the channel coefficient at discrete time k and consider one sample ahead prediction, then $h_k = \sum_{n=1}^{\infty} p_n\, h_{k-n}$. Note that the prediction error filter $P(f)$, which is an ideal high-pass filter, can be chosen to be independent of the actual Doppler spectrum $S_h(f)$ within its support, and can be chosen to be only a function of the Doppler spread $F_c = \frac{1}{T_c}$. Let us denote this spectrum independent prediction error filter as $P_{T_c}(f)$. As we have perfect prediction, we can repeat the one sample ahead prediction recursively to perfectly predict multiple samples ahead. Can this be repeated indefinitely? Yes if we have all samples available to predict from, but no if T-ahead prediction is based on a T times downsampled version. In that case, when we hit prediction horizon T_c, T_c-ahead prediction being here (in terms of zero prediction error) equivalent to 1-ahead prediction on a T_c times downsampled signal, downsampling (and hence stretching its support) $S_h(f)$ by a factor T_c makes it non-singular at all frequencies (non bandlimited). Note also that due to the perfect predictibility over the horizon $\{1, \ldots, T_c - 1\}$, linear estimation in terms of the complete past is equivalent to linear estimation in terms of a T_c times downsampled version of the past, since the samples in between can be filled up causally from a downsampled version. At prediction horizon T_c now, from a T_c times downsampled past, we are dealing with standard 1-ahead linear prediction of a non bandlimited stationary process, which under some regularity conditions can be considered as an AR(∞) process (Kolmogorov model). Let the infinite order prediction error filter for the T_c times downsampled process be $A(f)$. The reasoning above allows us to formulate the following theorem.

Theorem 1: **Two-Time Scale BL Model** The prediction error filter for a stationary process h_k bandlimited to $1/T_c$ (T_c integer) can be modeled as

$$P(f) = P_{T_c}(f)\, A(T_c f) \tag{3}$$

where $P_{T_c}(f)$ is the prediction error filter for a BL process with flat Doppler spectrum and $A(f)$ is the prediction error filter for the downsampled h_{kT_c}.

Let $G(f) = 1/P_{T_c}(f) = \sum_{n=0}^{\infty} g_n\, e^{-j2\pi fn}$ which is like $P_{T_c}(f)$ again a minimum-phase monic causal filter. Note that $G(f)$ behaves like an ideal low-pass filter with bandwith $1/T_c$, hence the T_c times downsampled version of its impulse response is a delta function: $g_{nT_c} = g_0\, \delta_{n0}$. Then the stationary BL process h_k can be generated as

$$h_k \;=\; g_k \;*\; h_k^{\downarrow\uparrow} \tag{4}$$

where $h_k^{\downarrow\uparrow}$ is the T_c times downsampled and then T_c times upsampled (inserting $T_c - 1$ zeros between consecutive samples) version of h_k and $*$ denotes convolution. The block fading model is similar to (4) with g_k now a rectangle: $g_k = 1$, $k = 0, 1, \ldots, T_c - 1$ and zeros elsewhere. With this similarity, the block fading and BL stationary case have in common that for every consecutive coherence period T_c, if the first sample (and the past) is known, then the remaining $T_c - 1$ samples of the current coherence period are known [11].

B. Back to the Noisy BL Case

The prediction of a BL process is not a stable operation [12] as can be seen from (1) where $\widetilde{\sigma}^2$ grows more rapidly than linear in $\sigma^2_{\widetilde{\widetilde{h}}}$ (assuming $\sigma^2_{\widetilde{\widetilde{h}}}$ is small). This is related to the fact that the (noiseless) prediction coefficients p_k are of infinite length and are not rapidly decaying. In [13], it was shown (for CSIR purposes) that the stationary BL model and the block fading model become equivalent as $F_c \to 0$. Such equivalence in the limit will also result for CSIT purposes here. But we want to go beyond the limit of very small Doppler spread.

Consider the (infinite order) autoregressive (Kolmogorov decomposition) and moving-average (Wold decomposition) representations of a noisy BL stationary process with spectrum as in Fig. 1 :

$$S(f) = \frac{\widetilde{\sigma}^2}{|P(f)|^2} = \widetilde{\sigma}^2\, |G(f)|^2 \tag{5}$$

with monic (first coefficient equal to 1) minimum-phase infinite order prediction error filter $P(f)$ and spectral factor $G(f)$ and infinite order prediction error variance $\widetilde{\sigma}^2$ such that the prediction error SNR becomes at high SNR (in the rest of this subsection $T = T_c$)

$$\frac{\sigma_h^2}{\widetilde{\sigma}^2} = \sigma_h^2\, e^{-\int \ln S(f)\, df} = T^{-1/T} \left(\frac{\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}}\right)^{1-1/T} . \tag{6}$$

where the channel estimation/FB SNR $\frac{\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}}$ is assumed to be proportional to the system SNR ρ (even if only for large ρ) . It appears that analytical expressions for $P(f)$ do not exist in the literature and the following may explain why. We get straightforwardly

$$|P(f)|^2 = \frac{\widetilde{\sigma}^2}{S(f)} = \begin{cases} (\frac{T\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}})^{-(1-1/T)} & f \in [0, 1/T] \\ (\frac{T\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}})^{1/T} & f \in [1/T, 1] \end{cases} \tag{7}$$

and hence we get for the energy in $P(f)$

$$||P||^2 = 1/T(\frac{T\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}})^{-(1-1/T)} + (1-1/T)(\frac{T\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}})^{1/T} \;\to \infty \tag{8}$$

which explodes as $\rho \to \infty$! Similarly for the monic causal spectral factor $G(f) = 1/P(f)$ and hence its energy

$$||G||^2 = 1/T(\frac{T\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}})^{1-1/T} + (1-1/T)(\frac{T\sigma_h^2}{\sigma^2_{\widetilde{\widetilde{h}}}})^{-1/T} \;\to \infty \tag{9}$$

explodes also when we insist on monicity ($g_0 = 1$). Of course, it is possible to find a spectral factor G with finite energy, but then $g_0 \to 0$.

C. No exact BL model anywhere

In [14], the behavior of (1) is exploited to show the resulting DoF of the 2 user MISO BC. However, what is not mentioned there is that these results correspond to a channel model that needs to be in a range between two extreme models. The one extreme model is block fading over blocks of length T_{fb}, with stationary F_c-BL evolution of the value of the blocks, and channel feedback every T_{fb}. The other extreme is a genuine F_c-BL stationary channel model, but then the channel needs to be fed back every sample (which is normally unacceptable in terms of NetDoF)! In [2] still another approach is taken in which block fading over some T is assumed, plus BL stationary evolution between blocks (such that one of the interpretations of [14] corresponds to this with $T = T_{fb}$).

The other popular model is the block fading model of course. In [11], it was shown that the DoFs of [14] can be reproduced very simply in the case of a block fading model, by the MAT-ZF scheme, a simple combination of MAT (during T_{fb}, while waiting for the channel FB) and ZF for the rest of the coherence period (see further discussion below). In [15] it was shown in an alternative fashion that the channel FB rate could be reduced w.r.t. [14] by a factor T_c/T_{fb} (equivalent to FB every T_c instead of every T_{fb}, as our FRoI approach also indicates, see below). To reproduce these results for the stationary BL case is not easy though, and the scheme of [14] is quite intricate, involving, as in MAT, FB of (residual) interference (now necessarily digital, with superposition coding and sequential decoding).

The models we introduce next allow to retain the simplicity of block fading models and even go beyond them (by exploiting stationarity).

V. Linear Finite Rate of Innovation (FRoI) Channel Models (CM)

FRoI signal models were introduced in [16]. Innovation here could be a somewhat misleading term since historically (in Kalman filter parlance) the term "innovations" has been used to refer to the infinite order prediction errors. In [16] and here, the rate of innovation could be considered to be the DoF of signals (i.e. the source coding rate prelog). FRoI represents the time series case of sparse modeling. The FRoI signal models that have been considered in [16] could be in general non-linear. In other words, the FRoI represents the average number of parameters per time unit needed to describe the signal class and these parameters could enter the signal model in an arbitrary fashion. For instance, the signal could be a linear superposition of basis functions of which also the positions (delays, and in the channel modeling case e.g. also Doppler shifts) are parameterized. For the purpose of channel modeling and FB, with essentially stationary signals that need to be processed in a causal fashion, it would appear reasonable to stick to linear FRoI models, in which the parameters are just the linear combination coefficients of fixed, periodically appearing basis functions, commensurate with the Doppler bandwith. This also corresponds exactly to so-called Basis Expansion Models (BEMs), which were probably introduced in [17] and used for estimating time-varying filters in the eighties and for channel modeling in [18] and many follow-up works.

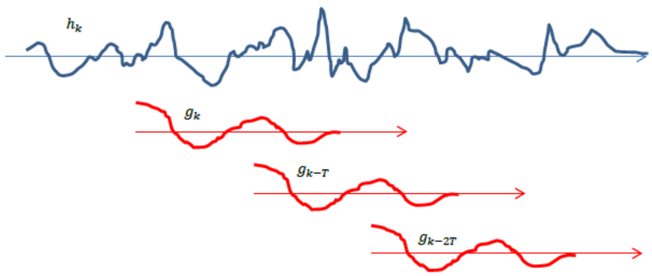

Fig. 2. Finite Rate of Innovation (FRoI) time-varying channel modeling.

In the case of a single basis function, the FRoI channel model is similar to (4):

$$h_k = g_k \, * \, a^{\uparrow}_k \tag{10}$$

where $a^{\uparrow}_k$ is a T_c times upsampled discrete-time signal of which the non-zero samples (parameters) appear once every T_c sampling periods, and g_k is a basis function, see Fig. 2. The resulting FRoI model encompasses the following existing models:

- *block fading*: $g_k = \left\{ \begin{array}{ll} 1 & , \; k = 0, 1, \ldots, T_c - 1 \\ 0 & , \; \text{elsewhere} \end{array} \right.$
- *stationary bandlimited (BL)*:
$$g_k = \text{sinc}(\pi k/T_c) = \frac{\sin(\pi k/T_c)}{\pi k/T_c} \; .$$

In our case, g_k is a causal FIR approximation to an ideal lowpass filter with (overall) bandwidth F_c. The length of the basis function g_k is intended to span several T_c. By making the filter longer however, a bandlimited characteristic can be better approximated. Obviously, the BL model (4) can be obtained by letting the filter length become infinite. Starting from a stationary sequence a_k, the process h_k generated by (10) is cyclostationary. By letting g_k better approximate a lowpass (or bandpass) filter, the cyclostationary process gets closer to stationary. In any case, at the start of each new coherence period T_c, knowing the past, the estimation of the sample h_k allows the estimation of the new parameter $a^{\uparrow}_k$ involved. And this in turn allows to determine the evolution of h_k for the next $T_c - 1$ samples. In the presence of noise, it is clearly desirable to have a first coefficient g_0 that is large (though any non-zero coefficient is sufficient for DoF analysis purposes). Due to the finite length and energy of the filter g_k, the effect of noise is limited and the prediction error variance over the coherence period will remain of the order of $\sigma^2_{\tilde{\tilde{h}}}$, the noise level in the channel FB. As the sampling rate (and hence FB frequency) of BL signals increases, the horizon of perfect prediction increases proportionally, and becomes infinite as the continuous-time past signal becomes available [12]. Of course, for all real-world signals for which a BL model seems plausible (e.g. the speech signal), this does not work because real-world signals are only approximately stationary and bandlimited over a limited time horizon. For instance, it is impossible to predict what a speaker is saying. In wireless communications, although Doppler shifts are finite because speed is finite, the Doppler spectrum becomes non-BL because the Doppler shifts are time-varying. If the channel response would be a deterministic function of the mobile terminal position, prediction of the channel would correspond to prediction of the mobile position which is impossible on a longer time scale. From this point of view, linear FRoI models which are approximately bandlimited but with a finite memory might be better approximations of approximately BL real-world signals. A lot of work on estimating FRoI signals has focussed on non-causal approaches [19]. However, what is needed for the application of FRoI to channel feedback is a design with prediction in mind.

The linear FroI model can also be considered as a filterbank with a single subband (more discussion to follow below). The synthesis filter is g_k, and there is an analysis filter f_k. The analysis-synthesis cascade leads to

$$\begin{array}{l} a_n = \sum_k f_k \, h_{nT_c - k} \\ h_{nT_c+i} = \sum_l a_{n-l} \, g_{lT_c+i} \; , \quad i = 0, 1, \ldots, T_c - 1 \; . \end{array} \tag{11}$$

Perfect reconstruction for a strictly BL process requires:

$$g_k * f_k = \text{sinc}(\pi k/T) \; . \tag{12}$$

This can be satisfied with e.g. $g_k = \text{sinc}(\pi k/T) \, , \;\; f_k = \delta_{k0}$ (Kronecker delta). In the case of an orthogonal filterbank with causal g_k, this requires $f_k = g^*_{-k}$, and $(g_k * g^*_{-k})_{k=nT} = \delta_{n0}$ (the convolution $g_k * g^*_{-k}$ (correlation sequence of g_k) should be a Nyquist pulse). In this case, if h_k is not a BL signal, the reconstructed signal resulting from applying the FRoI model in (11) would produce the least-squares projection of the signal h_k on the subspace of F_c-BL signals [20]. However, this requires $f_k = g^*_{-k}$ (matched filter) to be non-causal! As can be seen from Fig. 2, the optimal computation of coefficient a_n requires the correlation of the signal h_k that follows from the time instant $k = nT_c$ onwards with the basis function g_k. This is impractical for the channel feedback application in which both g_k and f_k should be causal, and the computation of a_n should be based (for optimal DoF considerations) on the first sample only of this correlation. Hence g_0 plays an important role (can not be small).

For a number of applications (handling of multiple users with different T_{fb} or different T_c, see further also), the use of FRoI models with multiple, N, basis functions might be desirable. In this case the FRoI model becomes

$$h_k = \sum_{n=1}^{N} g^{(n)}_k \; * \; a^{\uparrow\,(n)}_k \tag{13}$$

where the $a^{\uparrow\,(n)}_k$ are N sequences of parameters that are now NT_c times upsampled, to preserve a RoI of F_c. As the $g^{(n)}_k$ represent N different basis functions that are essentially bandlimited and also time limited, there might be some connection with prolate spheroidal wave functions [7], [12]. However, to limit FB delay, the first N coefficients of these basis functions play a particularly important role here.

VI. Basis Function Optimization

Although the true channel model cannot be strictly BL, it is nevertheless reasonable to use a BL model for the optimization of the FRoI model. Here we consider the FRoI model for a single generic channel coefficient. To be optimal however, all (correlated) channel coefficients would have to be treated jointly.

A. Single Basis Function Case

Consider first the case $N = 1$. Let g_k span LT (we will denote $T = T_c$ in this section to simplify): $g_k \, , \;\; k = 0, 1, \ldots, LT - 1$. Decompose discrete time as $k = nT + i$ where $i = k \mod T$, $n = \lfloor k/T \rfloor$. Then the FRoI or BEM channel model can be written as

$$h_k = \sum_{l=0}^{L-1} a_{n-l} g_{lT+i} \; , \quad i = 0, 1, \ldots, T-1 \; . \tag{14}$$

B. Approach 1: FRoI model based Analysis

Rx Side:

Now assume that the UE disposes of a channel estimate

$$\widehat{h}_k = h_k + \tilde{h}_k \tag{15}$$

where we assume $\tilde{h}_k$ to be white noise with variance $\sigma^2_{\tilde{h}}$. (Net)DoF optimization requires to use minimal (just identifiable) data to estimate the basis expansion coeffients a_n. Hence, if we assume the BEM coeffients a_n to be estimated without delay, then they get estimated from one channel signal sample, as soon as their corresponding basis function starts. Hence a_n gets estimated from the following data model

$$\begin{bmatrix} \widehat{h}_{nT} \\ \widehat{h}_{nT-1} \\ \vdots \\ \widehat{h}_{(n-1)T} \\ \widehat{h}_{(n-1)T-1} \\ \vdots \\ \widehat{h}_{(n-2)T} \\ \vdots \end{bmatrix} = \begin{bmatrix} g_0 & g_T & g_{2T} & \cdots \\ 0 & g_{T-1} & g_{2T-1} & \cdots \\ \vdots & \vdots & \vdots & \\ & g_0 & g_T & \cdots \\ & 0 & g_{T-1} & \cdots \\ & \vdots & \vdots & \\ & & g_0 & \cdots \\ & & & \ddots \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ a_{n-2} \\ \vdots \end{bmatrix} + \begin{bmatrix} \tilde{h}_{nT} \\ \tilde{h}_{nT-1} \\ \vdots \\ \tilde{h}_{(n-1)T} \\ \tilde{h}_{(n-1)T-1} \\ \vdots \\ \tilde{h}_{(n-2)T} \\ \vdots \end{bmatrix} \tag{16}$$

which can be rewritten compactly as

$$\underline{\underline{\widehat{h}}}_n = G\, \underline{\underline{a}}_n + \underline{\underline{\tilde{h}}}_n \ . \tag{17}$$

The least-squares solution for $\underline{\underline{a}}_n$ yields

$$\underline{\underline{\hat{a}}}_n = (G^T G)^{-1} G^T \underline{\underline{\widehat{h}}}_n \tag{18}$$

and hence

$$\hat{a}_n = \underline{f}\, \underline{\underline{\widehat{h}}}_n \text{ with } \underline{f} = e_1^T (G^T G)^{-1} G^T \tag{19}$$

where $e_1^T = [1\ 0\ 0 \cdots]$. The main characteristic of Approach 1 is that the analysis filter $\underline{f}$ is a function of the basis function (synthesis filter) g, and not of the actual channel h.

Tx Side:

The feedback to the Tx leads to the availability of

$$\hat{\hat{a}}_n = \hat{a}_n + \tilde{\hat{a}}_n \tag{20}$$

at the Tx, on the basis of which the Tx reconstructs the channel signal as

$$\begin{bmatrix} \widehat{\widehat{h}}_{(n+1)T-1} \\ \vdots \\ \widehat{\widehat{h}}_{nT} \end{bmatrix} = \sum_{l=0}^{L-1} \begin{bmatrix} g_{(l+1)T-1} \\ \vdots \\ g_{lT} \end{bmatrix} \hat{\hat{a}}_{n-l} \ . \tag{21}$$

At least, we shall consider this simple deterministic reconstruction for the purpose of optimizing the basis function g_k. (Alternatively the Tx could account for the fact that the $\hat{\hat{a}}_n$ are noisy.)

Basis Function Optimization:

Note that the matrix G is of the form

$$G = \begin{bmatrix} g_0 & \underline{g}^T \\ 0 & G' \end{bmatrix} \tag{22}$$

where $\underline{g}^T = [g_T\ g_{2T}\ g_{3T} \cdots]$. The optimization of the basis function g now follows by minimizing the MSE associated to (21) where the $\hat{\hat{a}}_n$ follow from (20) and (19). However, this leads to a quite nonlinear criterion. Now, it is clear that the h_{nT} are used for the estimation of the a_n (and not for data transmission). Hence the $\underline{g}$ appears to be irrelevant for the reconstruction of h. Furthermore, having $\underline{g} \neq 0$ would seem to only deteriorate the estimation of the a_n. Hence we shall consider here a constrained optimization problem with $\underline{g} = 0$. This leads to the T-downsampled version of g_k to be a delta function. With $\underline{g} = 0$, we get

$$\hat{a}_n = \widehat{h}_{nT}\ , \ \ \hat{\hat{a}}_n = h_{nT} + \tilde{h}_{nT} + \tilde{\hat{a}}_n \ . \tag{23}$$

For the design of the g_k, we consider the sum MSE over one coherence period in (21). This decouples to the MSE per sample

$$h_{nT+i} - \widehat{\widehat{h}}_{nT+i} = h_{nT+i} - \sum_{l=0}^{L-1} g_{lT+i} \hat{\hat{a}}_{n-l}\ , \ \ i = 0, 1, \ldots, T-1 \tag{24}$$

where we mentioned that we shall omit the consideration of $i = 0$ (or possibly even $i = 0, 1, \ldots, T_{fb} - 1$ to account for FB delay). Note that (24) corresponds to the prediction of h_{nT+i} on the basis of the L $\hat{\hat{a}}_{n-l}$. The MSE (for $i > 0$) is dominated by approximation error at high SNR, hence we shall consider the noiseless case for the design of the g_k. We thus get the following MSE criterion

$$\text{MSE}_i = \text{ E}\left|h_{nT+i} - \sum_{l=0}^{L-1} g_{lT+i}\, h_{(n-l)T}\right|^2 . \tag{25}$$

This leads to the following normal equations

$$R\, \underline{g}_i = \underline{r}_i \tag{26}$$

where we have a Toeplitz covariance matrix R

$$R = \begin{bmatrix} r_0 & r_T & \cdots & r_{(L-1)T} \\ r_T & r_0 & \ddots & \\ \vdots & \ddots & & \vdots \\ r_{(L-1)T} & r_{(L-2)T} & \cdots & r_0 \end{bmatrix} \tag{27}$$

and

$$\underline{r}_i = \begin{bmatrix} r_i \\ r_{i+T} \\ \vdots \\ r_{i+(L-1)T} \end{bmatrix},\ \underline{g}_i = \begin{bmatrix} g_i \\ g_{i+T} \\ \vdots \\ g_{i+(L-1)T} \end{bmatrix} \tag{28}$$

with the correlation sequence $r_m = \text{ E}\, h_{k+m} h_k$. For the case of an ideal lowpass Doppler spectrum, we have $r_m = \sigma^2_h\, T\, \dfrac{\sin(\pi m/T)}{\pi m}$ (so $r_0 = \sigma^2_h$). The resulting MSE_i is

$$\text{MSE}_i = r_0 - \underline{r}_i^T R^{-1} \underline{r}_i \ . \tag{29}$$

The normalized average MSE (or inverse approximation SNR) is

$$\text{NMSE} = \frac{\sum_{i=t}^{T-1} \text{MSE}_i}{(T-t)\sigma^2_h} \tag{30}$$

where $t = T_{fb} \geq 1$. Another evaluation criterion is $|G(f)|^2 = |\sum_{k=0}^{LT-1} g_k e^{-j2\pi f k}|^2$ which should approximate an ideal low-pass filter. In particular the ratio of power outside and inside the frequency interval $[-\frac{1}{2T}, \frac{1}{2T}]$ can be considered.

C. Approach 2: Biorthogonal Approach with decoupled Analysis and Synthesis filters

Let $\mathbf{f} = [f_0\, f_1 \cdots f_{M-1}]$ in $\hat{a}_n = \mathbf{f}\, \underline{\underline{\widehat{h}}}_n$ be unconstrained (where now $\underline{\underline{\widehat{h}}} = [\widehat{h}_{nT}\ \widehat{h}_{nT-1} \cdots]^T$ is of length M), not only to simplify the MSE cost function, but to get a better approximation capability, in particular also to reduce pressure on g_0. The channel reconstruction (average) MSE criterion becomes

$$\text{MSE} = \frac{1}{T}\text{ E}\left\| \underline{h}_n - \sum_{l=0}^{L-1} \mathbf{g}_l \left(\mathbf{f}\, \underline{\underline{\widehat{h}}}_{n-l} + \tilde{\hat{a}}_{n-l} \right) \right\|^2 \tag{31}$$

which is now quadratic in $\mathbf{f}$ or $\mathbf{g}$ separately. This can be solved by alternating minimization, quite similar to joint transmitter/receiver design via MMSE.

1) Optimization w.r.t. $\mathbf{g}$ *for a given* $\mathbf{f}$*:* We can rewrite the criterion (31) as

$$\mathrm{E}\left\|\underline{h}_n-\mathbf{G}\,(\mathbf{F}\,\widehat{\underline{\underline{h}}}'_n+\underline{\tilde{\tilde{a}}}_n)\right\|^2 \tag{32}$$

where $\underline{\tilde{\tilde{a}}}_n=[\tilde{\tilde{a}}_n\ \tilde{\tilde{a}}_{n-1}\cdots\tilde{\tilde{a}}_{n-L+1}]^T$, $\widehat{\underline{\underline{h}}}'_n$ is an extended version of $\widehat{\underline{\underline{h}}}_n$ of length $J=M+L(L-1)$, $\mathbf{G}=[\mathbf{g}_0\ \mathbf{g}_1\cdots\mathbf{g}_{L-1}]$ and

$$\mathbf{F}=\mathcal{T}(\mathbf{f})=\begin{bmatrix}\mathbf{f} & 0_{1\times L(L-1)} & \\ 0_{1\times L} & \mathbf{f} & 0_{1\times L(L-2)}\\ & \ddots & \\ & 0_{1\times L(L-1)} & \mathbf{f}\end{bmatrix} \tag{33}$$

which is hence a banded block Toeplitz matrix (obtained by taking every L^{th} row of a banded Toeplitz matrix). With (32) we can rewrite (31) as

$$\mathrm{MSE}=r_0+\frac{1}{T}\mathrm{tr}\{\mathbf{G}(\mathbf{F}(R_{\widehat{\underline{\underline{h}}}'\widehat{\underline{\underline{h}}}'}+\sigma^2_{\tilde{h}}I)\mathbf{F}^T+\sigma^2_{\tilde{\tilde{a}}}I)\mathbf{G}^T-2R_{\underline{h}\widehat{\underline{\underline{h}}}'}\mathbf{F}^T\mathbf{G}^T\}\,. \tag{34}$$

Optimizing (34) leads to

$$\mathbf{G}=R_{\underline{h}\widehat{\underline{\underline{h}}}'}\mathbf{F}^T(\mathbf{F}(R_{\widehat{\underline{\underline{h}}}'\widehat{\underline{\underline{h}}}'}+\sigma^2_{\tilde{h}}I)\mathbf{F}^T+\sigma^2_{\tilde{\tilde{a}}}I)^{-1} \tag{35}$$

where we have $R_{\widehat{\underline{\underline{h}}}'\widehat{\underline{\underline{h}}}'}=R_J$ is a symmetric Toeplitz covariance matrix of the form (for arbitrary N)

$$R_N=\begin{bmatrix}r_0 & r_1 & \cdots & r_{N-1}\\ r_1 & r_0 & \ddots & \\ \vdots & \ddots & & \vdots\\ r_{N-1} & r_{N-2} & \cdots & r_0\end{bmatrix} \tag{36}$$

and finally we have the Toeplitz matrix

$$R_{\underline{h}\widehat{\underline{\underline{h}}}'}=\begin{bmatrix}r_{T-1} & r_T & \cdots & r_{J+T-2}\\ \vdots & \vdots & \ddots & \vdots\\ r_1 & r_2 & \cdots & r_J\\ r_0 & r_1 & \cdots & r_{J-1}\end{bmatrix} \tag{37}$$

where again $J=M+L(L-1)$.

2) Optimization w.r.t. $\mathbf{f}$ *for a given* $\mathbf{g}$*:* Note that the FB noise $\tilde{\tilde{a}}_n$ has no effect on the optimization of $\mathbf{f}$. Criterion (31) now becomes

$$\mathrm{E}\sum_{k=0}^{T-1}|h_{nT+k}-\mathbf{f}\sum_{l=0}^{L-1}g_{k+lT}\widehat{\underline{\underline{h}}}_{n-l}|^2 \tag{38}$$

The optimal analysis filter $\mathbf{f}$ is solution to the following normal equations

$$\mathbf{f}\,\mathbf{A}=\mathbf{b}\ \Rightarrow\ \mathbf{f}=\mathbf{b}\,\mathbf{A}^{-1} \tag{39}$$

where

$$\begin{aligned}\mathbf{b} &= \sum_{k=0}^{T-1}\mathrm{E}\,h_{nT+k}\sum_{l=0}^{L-1}g_{k+lT}\widehat{\underline{\underline{h}}}^T_{n-l}\\ &= \sum_{k=0}^{T-1}\sum_{l=0}^{L-1}g_{k+lT}\ \mathrm{E}\,h_{nT+lT+k}\,\widehat{\underline{\underline{h}}}^T_n\\ &= \mathbf{g}\ \mathrm{E}\underline{h}'_n\widehat{\underline{\underline{h}}}^T_n=\mathbf{g}\ \mathrm{E}\underline{h}'_n\underline{\underline{h}}^T_n=\mathbf{g}\,R_{\underline{h}'\underline{\underline{h}}}\end{aligned} \tag{40}$$

where $\underline{h}'_n=[h_{nT}\ h_{nT+1}\cdots h_{nT+LT-1}]^T$, we exploited the stationarity of h_k, $\mathbf{g}=[\mathbf{g}^T_0\ \mathbf{g}^T_1\cdots\mathbf{g}^T_{L-1}]=[g_0\,g_1\cdots g_{LT-1}]$ and we have the Hankel matrix (constant along antidiagonals)

$$R_{\underline{h}'\underline{\underline{h}}}=\begin{bmatrix}r_0 & r_1 & \cdots & r_{M-1}\\ r_1 & r_2 & \cdots & r_M\\ \vdots & \vdots & \iddots & \vdots\\ r_{LT-1} & r_{LT} & \cdots & r_{M+LT-2}\end{bmatrix}. \tag{41}$$

Finally

$$\begin{aligned}\mathbf{A} &= \sum_{k=0}^{T-1}\sum_{l=0}^{L-1}\sum_{m=0}^{L-1}g_{k+lT}g_{k+mT}\ \mathrm{E}\widehat{\underline{\underline{h}}}_{n-l}\widehat{\underline{\underline{h}}}^T_{n-m}\\ &= \sum_{l=0}^{L-1}\sum_{m=0}^{L-1}\mathbf{g}^T_l\mathbf{g}_m\ \mathrm{E}\widehat{\underline{\underline{h}}}_{n-l}\widehat{\underline{\underline{h}}}^T_{n-m}\\ &= \sum_{k=-(L-1)}^{L-1}r_{g,k}\,(R_{M,k}+\sigma^2_{\tilde{h}}I_{M,kT})\\ &= r_{g,0}\,(R_M+\sigma^2_{\tilde{h}}I_M)\\ &\quad+\sum_{k=1}^{L-1}r_{g,k}\,(R_{M,k}+R_{M,-k}+\sigma^2_{\tilde{h}}(I_{M,kT}+I_{M,-kT}))\end{aligned} \tag{42}$$

where we have the $M\times M$ Toeplitz covariance matrix

$$R_{M,k}=\begin{bmatrix}r_{|kT|} & r_{|kT+1|} & \cdots & r_{|kT+M-1|}\\ r_{|kT-1|} & r_{|kT|} & \cdots & r_{|kT+M-2|}\\ \vdots & \vdots & \iddots & \vdots\\ r_{|kT-M+1|} & r_{|kT-M+2|} & \cdots & r_{|kT|}\end{bmatrix}, \tag{43}$$

the shifted $M\times M$ identity matrix $I_{M,n}$,

$$[I_{M,n}]_{i,j}=\delta_{i-n,j} \tag{44}$$

where $\delta_{i,j}$ is the Kronecker delta, and

$$r_{g,k}=\sum_{n=0}^{L-1-|k|}\mathbf{g}^T_{n+|k|}\mathbf{g}_n \tag{45}$$

which can be computed as

$$[r_{g,0}\,r_{g,1}\cdots r_{g,L-1}]=[\mathbf{g}\ 0_{1\times L(L-1)}]\,\mathcal{T}(\mathbf{g})^T \tag{46}$$

where the block Toeplitz function $\mathcal{T}(.)$ is defined in (33).

In the iterative MSE minimization, we iterate between solving for $\mathbf{g}$ from (35) and then for $\mathbf{f}$ from (39) untill convergence. For initialization one can set the $\mathbf{g}$ to all ones. In the absence of noise, only the product of $\mathbf{g}$ and $\mathbf{f}$ counts (see (31)) and one could renormalize $g_0=1$ after each iteration.

Remark 1: Filter Banks: One may observe that the FRoI/BEM model thus introduced corresponds to modeling a signal by retaining only the first subband in a filterbank, see [20] and esp. [21]. One can imagine a critically subsampled filterbank with T subbands, each subsampled by a factor T. Since the signal to be modeled is only expected to occupy the lowest fraction $1/T$ of the spectrum, only the first subband signal is retained. From the moment only a single subband is retained, the relation between subband bandwith and subsampling factor becomes a little looser of course. In filterbank terminology, $\mathbf{f}$ is the analysis filter and $\mathbf{g}$ is the synthesis filter. Since both are different, the filterbank is biorthogonal. (Perfect reconstruction) filterbank design has been considered for reconstruction with a variable non-negative delay. Here, the delay needs to be negative though (prediction). Whereas the first approach for FRoI model optimization ($\mathbf{g}$ only) considered would correspond to orthogonal filterbanks in the noncausal case and in the absence of noise, we expect that the biorthogonal FRoI models of the second approach

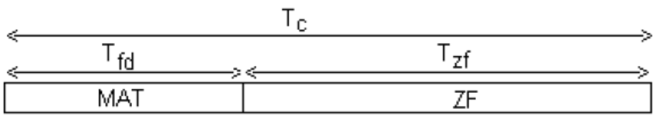

Fig. 3. MAT-ZF scheme.

(involving $\mathbf{g}$ and $\mathbf{f}$) are much better suited for the prediction needed here. Approach 2 furthermore allows to handle not only estimation error but also channel model approximation error.

Remark 2: Online Optimization: Here we considered the joint design of analysis and synthesis filters $\mathbf{f}$ and $\mathbf{g}$ in order to get an idea of the behavior of these filters and of their resulting performance. Note that due to the assumption of an (ideal) symmetric Doppler spectrum, channel correlations and optimal filters are real. In a real system, the UE could perform the joint optimization above, by assuming a certain FB noise level $\sigma^2_{\hat{\hat{a}}}$. And the BS could optimize its synthesis filters by minimizing $\mathrm{E}\left\|\underline{h}_n - \sum_{l=0}^{L-1} \underline{g}_l\, \hat{\hat{a}}_{n-l}\right\|^2$ or a sample average version hereof. However, the issue is that BS and UE need to use the same synthesis filter $\mathbf{g}$. One solution is to use an a priori design for $\mathbf{g}$, and then let the UE only adapt $\mathbf{f}$.

D. Multiple Basis Functions

Now consider $N > 1$. Compared to the $N = 1$ case, we can get to the $N > 1$ case straightforwardly by considering the "samples" h and a to represent $N \times 1$ consecutive samples, and the "samples" g become of size $N \times N$ in which the N rows represent N consecutive time samples and the N columns correspond to the coefficients of the N basis functions. T consecutive vector samples h_k now span in fact NT sampling periods. We get similar normal equations in which the correlation matrices r_m are Toeplitz, of size $N \times N$, and contain neighboring correlation values.

Note that in the case $N > 1$ an even higher noise sensitivity will exist due to the minimum delay estimation requirement for the a_n since the channel coefficient signal h_k will be heavily correlated over N consecutive samples if T is large, which will make the differentiation of the contributions of the N basis functions to the channel coefficient signal on the basis of only the first N samples ill conditioned.

In the filterbank interpretation, we are now splitting the original subband of bandwidth $1/T$ into N finer subbands, each subsampled by a factor NT.

For the case of rational $T = m/n$, a similar reasoning would lead to n basis functions (BEMs) in block length m. In case of multiple users with different T, the block length could be taken as their least common multiple (lcm) and the number of BEMs would be different for different users.

VII. DoF obtained with FRoI Channel Models (CMs)

As mentioned above, DoF obtained with block fading CMs can immediately be extended to FRoI CMs. Hence the DoF of the MAT-ZF scheme of [11], obtained in [11] for block fading, also apply for FRoI. This allows to reproduce the DoF of [14] for the 2-user MISO BC, and furthermore extend these DoF results to any MIMO single-hop multi-user network (Interfering Broadcast Channel, MAC, etc.) by simply combining the DoF of MAT and ZF for such networks (when known), see Fig. 3 :

$$\mathrm{DoF}_{MAT-ZF} = \frac{T_{fb}}{T_c}\,\mathrm{DoF}_{MAT} + (1 - \frac{T_{fb}}{T_c})\,\mathrm{DoF}_{ZF} \qquad (47)$$

which for MISO BC with $K = 2$ users becomes sumDoF = $2(\frac{2}{3}\frac{T_{fb}}{T_c} + 1 - \frac{T_{fb}}{T_c})$ as in [5]. Of course, the implementation of the MAT scheme involves the MAT part of possibly many coherence periods (e.g. 3 for the $K = 2$ MISO BC case, apart from CSI gathering overhead considerations).

These DoF can furthermore be improved by switching to FRoI models with $N > 1$ basis functions. As the RoI in these models is unchanged, the (average) feedback rate is unchanged. However, with $N > 1$, feedback needs to occur only once every NT_c, and hence FB delay is suffered only once every NT_c. Hence the weight of the MAT portion in the DoF_{MAT-ZF} is reduced to $\frac{T_{fb}}{MT_c}$, bringing the DoF_{MAT-ZF} closer to DoF_{ZF}. In theory N could be made arbitrarily large, but not in practice.

Until recently, the only scheme in which the DoF are not affected by FB delay was the scheme of [6] for the MISO BC/IC, with a MIMO extension in [22] where the scheme is termed Space-Time IA (STIA). The ingredients of STIA are:

- symbol extension (time-varying channel required): space-time ZF precoding
- due to CSIT delay, transmit fewer symbols per user
- but make up by overloading ($K > N_t$, number of BS antennas), to get full sumDoF
- send N_t symbols to $K = N_t + 1$ users over $N_t + 1$ T_c's

The scheme is presented in [6] for block fading but now, like any scheme that is valid for block fading, also becomes applicable to bandlimited stationary fading via the FRoI channel model introduced here.

A. Foresighted Channel Feedback

Indeed, the main characteristic of FRoI CMs is that they closely approximate stationary (BL) signals. *This means that if a FRoI CM is a good model, so is an arbitrary time shift of the FRoI model!* This can be exploited to overcome the FB delay as explained in Fig. 4. Consider FRoI CM with $N = 1$ basis function. While the current coherence period is running, as the Channel FB (CFB) is going to take a delay of T_{fb}, instead of waiting for the end of the current T_c, we start the next coherence period T_{fb} samples early. This means jumping from the subsampling grid of the FRoI model to the shifted subsampling grid of another instance of the same FRoI model. This involves recalculating the (finite number of past) FRoI parameters $a_k^{\uparrow}$ for the new grid from the past channel evolution on the old grid, plus a new channel estimate at the start of the T_c on the new grid. In this way the FB (sampling) "rate" increases from $\frac{1}{T_c}$ to $\frac{1}{T_c - T_{fb}}$. But the CSIT is available at the Tx all the time, with a channel prediction error SNR proportional to the general SNR. This approach is applicable to any multi-user network.

By increasing N, the number of basis functions, this approach continues to work for any $T_{fb} < N\,T_c$, and hence for any T_{fb}. Related work seems to appear in [23] where apparently the subsampling phase of a FRoI model for a given user is considered untouchable. To get full DoF in the presence of CSIT delay, instead of jumping from one subsampling phase to another for a given user, in [23] the authors propose to jump between users, for which the channels are modeled with different subsampling phases.

The proposed FCFB increases the FB DoF consumed, hence (and in any case) it is of interest to consider (the more relevant) NetDoF. An analysis of the resulting NetDoF for the MISO BC is presented in [24].

VIII. Beyond FRoI: Predictive Rate Distortion

The FRoI channel model is just one way to get a certain rate (DoF) for a distortion of $O(\sigma_v^2)$ (noise level). More generally, the distortion in a predicted channel at the Tx can contain a combination

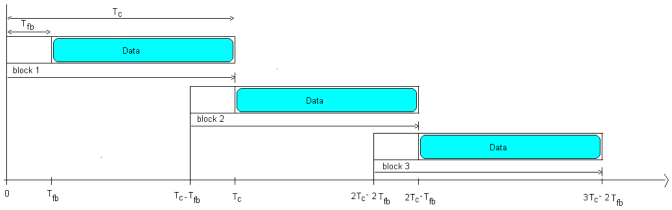

Fig. 4. Foresighted Channel Feedback (FCFB).

of approximation error and noise due to estimation and feedback. What is needed here is *predictive* rate-distortion (R-D) theory. Such theory would e.g. allow to determine which estimation and FB DoF are required to get the prediction distortion at the Tx down to the level of the noise (with the channel distortion at the Tx being reflected in an induced equivalent noise at the Rx). However, the proper evaluation of such R-D theory obviously depends heavily on the channel model, whereas all existing channel models are too approximate to allow a solid high SNR DoF analysis. Some related work appears in [25] where no causality is imposed and where the analysis filter f is not optimized (fixed), and in [26] and references therein, where causal (but not predictive) R-D is developed for the application of FB in control systems.

References

[1] M. A. Maddah-Ali and D. N. C. Tse, "Completely Stale Transmitter Channel State Information is still very Useful," *IEEE Trans. Info. Theory*, Jul. 2012.

[2] M. Kobayashi and G. Caire, " On the net DoF comparison between ZF and MAT over time-varying MISO broadcast channels," in *Proc. ISIT*, Boston, MA, USA, Jun. 2012.

[3] Y. Lejosne, D. Slock, and Y. Yuan-Wu, "Net Degrees of Freedom of Recent Schemes for the MISO BC with Delayed CSIT and Finite Coherence Time," in *Proc. WCNC*, Shanghai, China, Apr. 2013.

[4] C. Vaze and M. Varanasi, "The degrees of freedom region and interference alignment for the mimo interference channel with delayed csit," *Trans. Info Theory*, july 2012.

[5] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of Freedom of Time Correlated MISO Broadcast Channel with Delayed CSIT," *IEEE Trans. on Information Theory*, Feb. 2012.

[6] N. Lee and R. W. Heath Jr., "Not too delayed csit achieves the optimal degrees of freedom," in *Proc. Allerton*, Monticello, IL, USA, Nov. 2012.

[7] A. Kannu and P. Schniter, "On the Spectral Efficiency of Noncoherent Doubly Selective Block-Fading Channels," *IEEE Trans. Inf. Theory*, Jun. 2010.

[8] R. Etkin and D. Tse, "Degrees of Freedom in Some Underspread MIMO Fading Channels," *IEEE Trans. Inf. Theory*, Apr. 2006.

[9] L. Zheng and D. Tse, "Communication over the Grassmann Manifold: a Geometric Approach to the Noncoherent Multiple-Antenna Channel," *IEEE Trans. Inf. Theory*, Feb. 2002.

[10] P. Vaidyanathan, "On Predicting a Band-limited Signal Based on Past Values," *Proceedings of the IEEE*, Aug. 1987.

[11] Y. Lejosne, D. Slock, and Y. Yuan-Wu, "Degrees of Freedom in the MISO BC with D-CSIT and Finite Coherence Time: a Simple Optimal Scheme," in *IEEE Int'l Conf. on Signal Proc., Commun. and Control (ISCCSP)*, Hong Kong, Aug. 2012.

[12] A. Papoulis, *Signal Analysis*. McGraw-Hill, 1977.

[13] N. Jindal and A. Lozano, "A Unified Treatment of Optimum Pilot Overhead in Multipath Fading Channels," *IEEE Trans. Communications*, Oct. 2010.

[14] M. Kobayashi, S. Yang, D. Gesbert, and X. Yi, "On the degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," in *IEEE Int'l Symp. Info. Theory*, Jul. 2012.

[15] J. Chen and P. Elia, "Can Imperfect Delayed CSIT be as Useful as Perfect Delayed CSIT? DoF Analysis and Constructions for the BC," in *IEEE Allerton Conf. Comm's Control and Comput.*, Monticello, IL, USA, Sept. 2012.

[16] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with Finite Rate of Innovation," *IEEE Trans. Signal Processing*, Jun. 2002.

[17] Y. Grenier, "Modeling of Nonstationary Signals," 1984, phD, Orsay, France.

[18] M. Tsatsanis and G. Giannakis, "Modeling and Equalization of Rapidly Fading Channels," *Int. J. Adaptive Control and Signal Proc.*, 1996.

[19] J. Uriguen, Y. Eldar, P. Dragotti, and Z. Ben-Haim, "Sampling at the Rate of Innovation: Theory and Applications," in *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012.

[20] M. Vetterli, J. Kovačević, and V. Goyal, *Foundations of Signal Processing*. http://fourierandwavelets.org/, 2013.

[21] J. Kovačević, V. Goyal, and M. Vetterli, *Fourier and Wavelet Signal Processing*. http://fourierandwavelets.org/, 2013.

[22] Y. Lejosne, D. Slock, and Y. Yuan-Wu, "Space Time Interference Alignment Scheme for the MIMO BC and IC with Delayed CSIT and Finite Coherence Time," in *Proc. Int'l Conf. Acoustics, Speech and Sig. Proc. (ICASSP)*, Vancouver, Canada, May 2013.

[23] R. Tandon, S. Jafar, and S. Shamai, "Minimum CSIT to achieve Maximum Degrees of Freedom for the MISO BC," Nov. 2012, arXiv:1211.4254v2.

[24] Y. Lejosne, D. Slock, and Y. Yuan-Wu, "NetDoFs of the MISO Broadcast Channel with Delayed CSIT Feedback for Finite Rate of Innovation Channel Models," in *IEEE Int'l Symp. on Iformation Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, submitted.

[25] E. Aguerri, D. Gunduz, A. Goldsmith, and Y. Eldar, "Distortion-Rate Function for Undersampled Gaussian Processes," in *IEEE Int'l Symp. on Iformation Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, submitted, also ITA2013 talk.

[26] E. Silva, M. Derpich, and J. Østergaard, "Rate-Distortion in Closed-Loop LTI Systems," in *IEEE Iformation Theory Workshop (ITA)*, San Diego, CA, USA, Feb. 2013.