MLSE for an Unknown Channel—Part I: Optimality Considerations

Keith M. Chugg, *Member, IEEE,* and Andreas Polydoros, *Fellow, IEEE*

***Abstract*— The problem of performing joint maximum-likelihood (ML) estimation of a digital sequence and unknown dispersive channel impulse response is considered starting from a continuous-time (CT) model. Previous investigations of this problem have not considered the front-end (FE) processing in detail; rather, a discrete-time signal model has been assumed. We show that a fractionally-spaced whitened matched filter, matched to the known data pulse, provides a set of sufficient statistics when a tapped delay line channel model is assumed, and that the problem is ill-posed when the channel impulse response is generalized to a CT, finite-length model. Practical approximations are considered that circumvent this ill-posed condition. Recursive computation of the joint-ML metric is developed. Together, the FE processing and metric recursion provide a receiver structure which may be interpreted as the theoretical foundation for the previously introduced technique of per-survivor processing, and they lead directly to generalizations. Several FE processors representative of those suggested in the literature are developed and related to the practically optimal FE.**

I. Introduction

THE RECEIVER structure for performing maximum-likelihood sequence estimation (MLSE) of a digital signal corrupted by *known* finite-length intersymbol interference (ISI) and additive Gaussian noise is well-known. Forney [2] showed that the receiver may be divided into two distinct components: i) the front-end (FE) processor, the so called "whitened matched filter" (WMF) [which is a cascade of a matched filter (MF) and a noise whitening filter (WF)], and ii) a nonlinear post-processor based on the Viterbi algorithm (VA). This conceptual partition has been widely embraced primarily due to the fact that the output of the FE is a set of sufficient statistics modeled by an equivalent discrete-time (DT), symbol-spaced ISI signal in white noise.

Paper approved by D. P. Taylor, the Editor for Signal Design, Modulation, and Detection of the IEEE Communications Society. Manuscript received September 16, 1994; revised June 16, 1995 and September 13, 1995. This paper was presented in part at the Communication Theory Mini-Conference colocated at GLOBECOM, San Francisco, CA, November 1994, and the International Symposium on Information Theory, Whistler, B.C., Canada, September 1995.

K. M. Chugg was with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA. He is now with the Communications Sciences Institute, Department of Electrical Engineering Systems, University of Southern California, Los Angeles, CA 90089-2565 USA.

A. Polydoros is with the Communications Sciences Institute, Department of Electrical Engineering Systems, University of Southern California, Los Angeles, CA 90089-2565 USA.

Publisher Item Identifier S 0090-6778(96)05508-0.

Specifically, when the continuous time (CT) complex baseband model

$$r(t) = \sum_i a_i h(t - iT) + n(t) = y(t) + n(t) \qquad (1)$$

is applied to the WMF, the output has the statistically equivalent DT model

$$z_k = \sum_{m=0}^{L-1} a_{k-m} f_m + w_k. \qquad (2)$$

The data sequence $\{a_i\}$, assumed independent and uniformly distributed over a finite alphabet $\mathcal{A}$, is common to (1) and (2). The equivalent DT channel coefficients $\{f_m\}_{m=0}^{L-1}$ in (2) and the CT *overall channel impulse response* $h(t)$ are related by[1]

$$R_h(\tau)|_{\tau=kT} = f_k * f_{-k}^* \qquad (3)$$

where T is the symbol duration and the notation $R_m(\tau) = m(\tau) * m^*(-\tau)$ is used to denote the correlation of any finite energy pulse $m(t)$. The noise free signal in (1) represents the response of the physical channel, with impulse response $c(t)$, to the quadrature-modulated signal $\Sigma_i a_i u(t - iT)$, so that $h(t) = c(t) * u(t)$. The data pulse $u(t)$ and physical channel $c(t)$ are assumed to have support contained in $[0, L_u T)$ and $[0, L_c T]$, respectively, so that $h(t)$ is nonzero only for $t \in [0, LT)$, with $L = L_u + L_c$. The additive white Gaussian noise (AWGN) $n(t)$ has spectral level N_0 (i.e., $\mathbb{E}\{n(t+\tau)n^*(t)\} = N_0\delta(\tau)$, where $\mathbb{E}\{\cdot\}$ and $\delta(t)$ denote the expectation operator and the Dirac delta, respectively).[2]

The statistical sufficiency of (2) arises from the expansion of the ML metric (i.e., a quantity proportional to the negative log-likelihood functional) for the data based on observing $r(t)$ on the interval J[3]

$$\Gamma(\{\tilde{a}_i\}, \tilde{\Theta}) = \int_J |y(t; \{\tilde{a}_i\}, \tilde{\Theta})|^2 \, dt - 2\Re\left\{ \int_J r(t) y^*(t; \{\tilde{a}_i\}, \tilde{\Theta}) \, dt \right\}. \qquad (4)$$

In (4), Θ is a set of unknown parameters that is included for future reference (i.e., in the current setting, Θ is empty). Note

[1] $(\cdot)^*$ denotes complex conjugation.

[2] The term "Gaussian" is used throughout as shorthand for "complex circular Gaussian."

[3] $\Re\{\cdot\}$ denotes the real-part operator. The notation $\tilde{m}$ is used to denote a hypothesized value of an arbitrary quantity m.

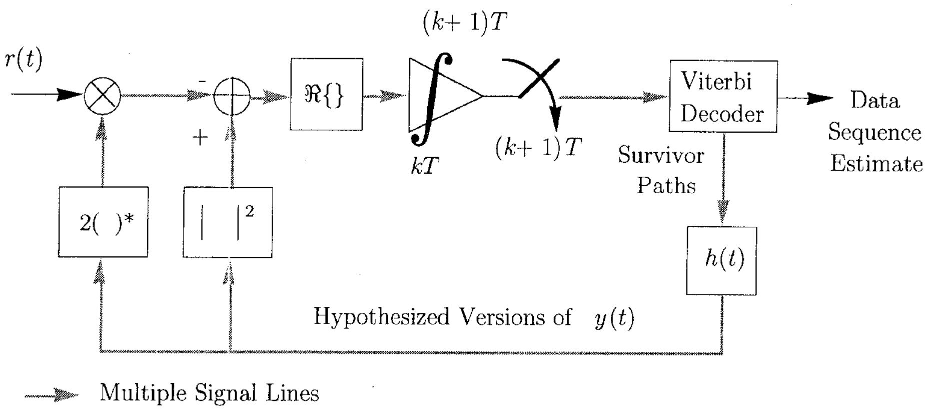


Fig. 1. The direct structure for known-channel MLSE.

that (4) can be expressed in a heuristic sense as[4]

$$\Gamma(\{\tilde{a}_i\},\tilde{\Theta}) \sim \int_J |r(t) - y(t;\{\tilde{a}_i\},\tilde{\Theta})|^2\,dt \tag{5}$$

which emphasizes the least-squares nature of the problem. The metric in (4) represents the limit of

$$\Gamma^{(K)}(\{\tilde{a}_i\},\tilde{\Theta}) = \sum_{k=1}^{K} |Y_k(\{\tilde{a}_i\},\tilde{\Theta})|^2 - 2\Re\{R_k Y_k^*(\{\tilde{a}_i\},\tilde{\Theta})\} \tag{6}$$

as $K \to \infty$ where $\{R_k\}$ is the set of coefficients corresponding to the expansion of $r(t)$ in terms of the orthogonal eigenfunctions of the noise covariance operator and $\{Y_k\}$ is the analogous set corresponding to $y(t)$. Ungerboeck [3] has shown that the ML metric in (4) can be updated recursively at the symbol-rate using the front-end MF outputs directly, so that MLSE may be implemented without a WF. In fact, we may consider the conceptually simple "direct structure," which eliminates the front-end MF as well. This is because symbol-rate recursive computation of $\Gamma_k(\tilde{\boldsymbol{a}}_k)$, the ML metric of (4) for the estimation of[5] $\boldsymbol{a}_k = [a_k \quad a_{k-1} \quad \cdots \quad a_0]^T$ with $J = [0,(k+1)T)$, is obtained by

$$\begin{aligned}\Gamma_k(\tilde{\boldsymbol{a}}_k) &= \Gamma_{k-1}(\tilde{\boldsymbol{a}}_{k-1}) + \int_{kT}^{(k+1)T} |y(t;\tilde{\boldsymbol{a}}_k)|^2\,dt \\ &\quad - 2\Re\left\{\int_{kT}^{(k+1)T} r(t)y^*(t;\tilde{\boldsymbol{a}}_k)\,dt\right\}.\end{aligned} \tag{7}$$

The receiver structure suggested by this metric recursion is diagrammed in Fig. 1. In this direct structure, a hypothesized version of the noise-free signal is generated for each survivor extension by driving a local version of the channel, which effectively distributes the MF to all possible paths.

A common feature of the above Forney, Ungerboeck, and direct structures for *known-channel* MLSE is that they all use explicit knowledge of $h(t)$. In this paper, we are concerned with the case when the physical channel and, hence, $h(t)$, is *a priori* unknown. Throughout this work, the term "unknown-channel" will mean that $c(t)$ is unknown, $u(t)$ is known, and L_u, and L_c are assumed to be known or adequately upper-bounded. A quasistatic assumption is made so that $c(t)$ is modeled as a fixed and deterministic waveform.

The model in (2) has become so ingrained in the study of MLSE for ISI channels that it has also been adopted for cases where the channel is unknown [4]–[9]. Clearly, if the channel is unknown, the WMF cannot be identified, hence, adoption of the DT model in (2) with unknown channel coefficients purveys an imprecise notion of the FE processing. Sometimes this contradiction is ignored in the unknown-channel MLSE literature; other times, a brief description of the FE processing assumed to arrive at (2) is provided. An FE which is frequently suggested (or implicitly assumed) is a receiving filter sampled at the symbol-rate [5], [8]–[10].[6] Symbol-rate sampling the output of a filter matched to the data pulse, which results in (2), has also been suggested [5], [11]. A version of the sampled FE with multiple samples per symbol was introduced in [12], and a similar structure was related to the concept of a bank of adaptive WMF's in [7].

In this paper, we investigate joint-ML (also known as the "generalized likelihood" method [13]) channel and data estimation starting from the CT model of (1) with the goal of formalizing the optimal FE processing and determining the effect of information-lossy alternatives. Within this general setting, we present the symbol-rate recursive computation of the ML metric, along with a reduced complexity approximation. This metric recursion allows joint-ML channel and data estimation to be viewed as an M-ary tree search problem, where $M = |\mathcal{A}|$. Unlike the known-channel case, the VA is not equivalent to exhaustive search. In fact, any bounded complexity tree-search algorithm is inherently suboptimal (for an infinite observation interval) because the functional form of the ML metric depends upon the *entire* hypothesized data sequence through the associated channel estimate.

[4]The expression in (5) is heuristic (i.e., ill-defined) because $r(t)$ contains white noise, which prevents the mean-square convergence of the integral.

[5]Transposition is denoted by $(\cdot)^T$.

[6]This approach is often seen in the conventional blind equalization literature as well, but it can be argued that there, a *constrained structure* which includes the fixed FE is assumed, while the receiver structure for the ML approach is ostensibly derived from the desire to maximize likelihood without any constraint.

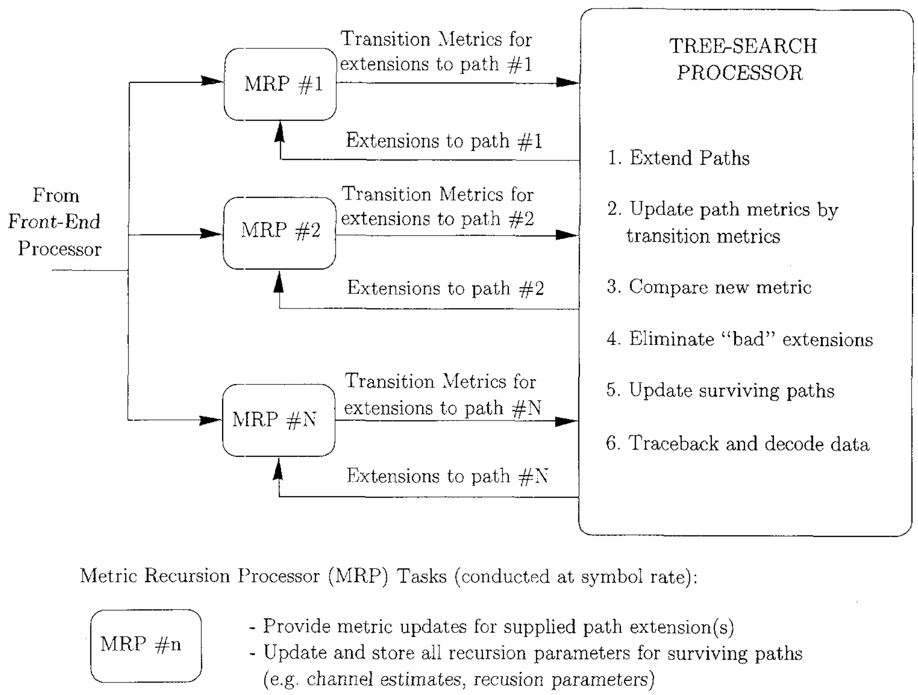


Fig. 2. The general recursive receiver structure.

If the VA is adopted as a suboptimal tree-search algorithm, the resulting receiver may be viewed as an application of the previously developed principle of per-survivor processing (PSP) [7], [14]. Virtually any tree-search algorithm may be employed in place of the VA [15], [16]. The resulting receiver structure will be of the form shown in Fig. 2. Although this figure implies a breadth-first strategy, metric-first and depth-first algorithms will result in similar structures, with the possible addition of data buffers. Associated with each path investigated by this receiver is an estimate of the channel. The resulting structure is a natural extension of the previously mentioned direct structure for known-channel MLSE. Special cases of this structure have been suggested for other applications of MLSE in the presence of uncertainty, such as multiuser detection in a code division multiple access system with unknown user amplitudes and/or delays [17], [18] and demodulation of continuous phase and trellis coded modulation signals in the presence of carrier phase uncertainty [7], [19]–[21].

This paper is organized as follows: The optimal joint-ML FE processing is developed in Section II. Suboptimal versions of these FE processors are defined in Section III. Recursive computation of the joint-ML metric is discussed in Section IV. The details of the ill-posed nature of the joint-ML problem for the general CT channel model are contained in Appendix A. The effects of FE processing in the channel-tracking environment are assessed through simulation and analysis in a companion paper [1].

II. Front-End Optimality Considerations

For the sake of conceptual flow, we first develop the well-defined FE processing for a tapped delay line (TDL) physical channel model before showing that the optimal processing for the general CT channel model is ill-posed. This ill-posed condition is then circumvented for practical purposes by demonstrating that the TDL channel model and the associated FE processing are good approximations to the CT counterparts. The details of the ill-posed condition are contained in Appendix A.

A. TDL Channel Model

In this subsection, a simple TDL model with unknown tap coefficients is assumed for the physical channel, namely

$$c(t) = \sum_{l=0}^{N_r L_c} c_l \delta(t - lT_r) \tag{8}$$

where the resolution time, T_r, is assumed to be related to the symbol-time by $T = N_r T_r$, with $N_r \geq 1$ an integer.[7] Assuming a finite data length I, the noise-free channel output is

$$y(t) = \sum_{i=0}^{I-1} \sum_{l=0}^{N_r L_c} a_i c_l u(t - (l + iN_r)T_r). \tag{9}$$

If $y(t)$ is observed in white Gaussian noise during the interval J when it is nonzero, the ML metric of (4), with $\Theta = \{c_l\}_{l=0}^{N_r L_c}$, involves $r(t)$ only through

$$\Re\left\{ \sum_{i=0}^{I-1} \sum_{l=0}^{N_r L_c} \tilde{a}_i^* \tilde{c}_l^* \int_J r(t) u^*(t - (iN_r + l)T_r)\, dt \right\}. \tag{10}$$

[7] The reason for this terminology will be made clear in Section II-B2.

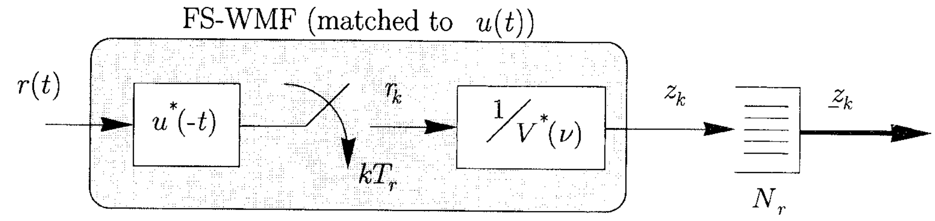


Fig. 3. The optimal front-end processing for the TDL channel model.

It follows that the output of a filter matched to $u(t)$ can be sampled at $t = kT_r$ to provide a sufficient set of statistics for the joint-ML estimation.

The output of a T_r-spaced WMF matched to $u(t)$ is

$$z_k = \sum_{i=0}^{I-1} \sum_{l=0}^{N_r L_c} a_i c_l v_{k-(iN_r+l)} + w_k \tag{11}$$

where $\{w_k\}$ is a white Gaussian sequence with variance N_0. The finite length sequence $\{v_k\}_{k=0}^{N_r L_u - 1}$ is the maximum-phase factor of the sampled pulse correlation (i.e., the sampled pulse correlation convolved with the realizable WF), and will be referred to as the *DT* (T_r) *equivalent pulse*. This terminology is appropriate since v_k and $u(t)$ are related in precisely the same manner as f_k and $h(t)$ in the known channel case [i.e., see (3)]. The *equivalent unknown channel* can then be defined as[8]

$$f_k = c_k * v_k = \sum_{l=0}^{N_r L_c} c_l v_{k-l}, \qquad k \in \mathcal{Z}_{LN_r} \tag{12}$$

so that

$$z_k = \sum_{i=0}^{I-1} a_i f_{k-iN_r} + w_k. \tag{13}$$

The T_r-spaced DT ISI model in (13) can be converted into a T-spaced *component vector ISI model*

$$\underline{z}_k = \underline{x}_k + \underline{w}_k = \sum_{m=0}^{L-1} a_{k-m} \underline{f}_m + \underline{w}_k \tag{14}$$

where the (N_r) *component vector* $\underline{m}_k$ of an arbitrary sequence $\{m_i\}$ is defined as

$$\underline{m}_k = [m_{(k+1)N_r-1} \quad m_{(k+1)N_r-2} \quad \cdots \quad m_{kN_r}]^T. \tag{15}$$

The component vector contains the N_r samples obtained every symbol interval.

The set of sufficient statistics for estimation of $\{c_l\}$ and $\{a_i\}$ from $r(t)$ for $t \in J$ is then $\{\underline{z}_k\}_{k=0}^{I+L-2}$. Either the unknown component vector channel coefficients $\{\underline{f}_i\}$ or the original channel TDL weights $\{c_l\}$ may be estimated for a given data sequence. Since the ultimate goal is a metric that depends only on the hypothesized data path, and in light of the linear relation between the two parameter sets, the choice is somewhat irrelevant. However, the larger set of parameters, $\{\underline{f}_i\}$, is chosen in order to simplify symbol-rate recursive computation of the ML metric. The optimal FE processing for the TDL channel model is illustrated in Fig. 3, with the WF frequency response expressed in terms of $V(\nu)$, the DT Fourier transform of the DT equivalent pulse.

The two fundamental differences between the model resulting from the FS-WMF in (14) and the analogous known-channel expression in (2) are that the MF is associated with the known data pulse rather than the overall channel, and multiple samples per symbol are required for (14).

B. General Continuous-Time Channel Model

In this section, the channel model is relaxed to be an arbitrary finite-energy, time-limited signal, which is identical to the model assumed for the known-channel development in [2]. Without loss of generality, we consider the combined effects of the pulse and physical channel filters and focus on joint estimation of the data and $h(t)$. The ML metric $\Gamma_k(\tilde{\boldsymbol{a}}_k, \tilde{h})$ for joint-ML estimation of $\boldsymbol{a}_k$ and $h(t)$, based on the observation of $r(t)$ for $t \in J = [0, (k+1)T)$, is given by (4) with $\Theta = \{h(t) : t \in [0, LT)\}$.

A chipped signal notation, identical to that defined in [2], is introduced for an arbitrary function $m(t)$

$$m_i(t) = \begin{cases} m(t+iT), & t \in [0,T) \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Applying this to a $\mathcal{L}_2[0,(i+1)T)$ function results in a sequence of chips from $\mathcal{L}_2[0,T)$, which can be denoted by a vector[9]

$$\begin{aligned} \boldsymbol{m}_i(t) &= [m_i(t) \quad m_{i-1}(t) \quad \cdots \quad m_0(t)]^T \\ &\in (\mathcal{L}_2[0,T))^{(i+1)}. \end{aligned} \tag{17}$$

The finite duration of $h(t)$ results in an $(L \times 1)$ vector representation denoted by $\boldsymbol{h}(t) = \boldsymbol{h}_{L-1}(t)$.

The convolution of data and channel may be expressed in terms of the chips as

$$y_i(t) = \sum_{m=0}^{L-1} a_{i-m} h_m(t) \triangleq \boldsymbol{\alpha}_i^H \diamond \boldsymbol{h}(t) = (\boldsymbol{h}^H(t) \diamond \boldsymbol{\alpha}_i)^* \tag{18}$$

where $(\cdot)^H = [(\cdot)^T]^*$ and $\boldsymbol{\alpha}_i$ is an L-dimensional vector with components in $\mathcal{A}$ defined by

$$\boldsymbol{\alpha}_i^H = [a_{i-L+1} \quad a_{i-L+2} \quad \cdots \quad a_i]. \tag{19}$$

The last expression in (18) will be referred to as a mixed (inner) product. This shorthand notation eliminates the need

[8]The notation $\mathcal{Z}_N = \{0, 1 \cdots N-1\}$ is used for any integer N.

[9]$\mathcal{L}_2[0,\tau)$ denotes the space of square-integrable functions on $[0,\tau)$.

to introduce identity operators on the signal Hilbert space ($\mathcal{L}_2[0,T)$ at this point).

The corresponding received signal chips are therefore

$$r_i(t) = y_i(t) + n_i(t), \qquad i = 0, 1, \cdots k \tag{20}$$

which, in vector form, is

$$\boldsymbol{r}_k(t) = \boldsymbol{y}_k(t) + \boldsymbol{n}_k(t) = \boldsymbol{A}_k \diamond \boldsymbol{h}(t) + \boldsymbol{n}_k(t) \tag{21}$$

$$\boldsymbol{A}_k = [\boldsymbol{\alpha}_k \quad \boldsymbol{\alpha}_{k-1} \quad \cdots \quad \boldsymbol{\alpha}_0]^H. \tag{22}$$

The $((k+1) \times L)$ Toeplitz matrix $\boldsymbol{A}_k$ consists of elements from the M-ary alphabet $\mathcal{A}$. The mixed-product notation in (21) is the natural extension of that in (18).

This modeling technique allows the ML metric in (4) to be expressed as

$$\begin{aligned}\Gamma_k(\tilde{\boldsymbol{A}}_k, \tilde{h}) = &\int_0^T \tilde{\boldsymbol{h}}^H(t) \diamond \tilde{\boldsymbol{A}}_k^H \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}(t)\, dt \\ &- 2\Re\left\{\int_0^T \tilde{\boldsymbol{h}}(t)^H \diamond \tilde{\boldsymbol{A}}_k^H \diamond \boldsymbol{r}_k(t)\, dt\right\}\end{aligned} \tag{23}$$

where the one-to-one correspondence between $\boldsymbol{a}_k$ and $\boldsymbol{A}_k$ has been used. For a fixed $\tilde{\boldsymbol{A}}_k$, minimization over $\tilde{\boldsymbol{h}}(t)$ results in the critical point $\hat{\boldsymbol{h}}(t; \tilde{\boldsymbol{A}}_k) = \tilde{\boldsymbol{A}}_k^I \diamond \boldsymbol{r}_k(t)$, where $\tilde{\boldsymbol{A}}_k^I$ is the pseudo-inverse of $\tilde{A}k$ [22].[10] Substituting this result into (23) yields a metric dependent only on the hypothesized data sequence

$$\Gamma_k(\tilde{\boldsymbol{A}}_k) \triangleq \Gamma_k(\tilde{\boldsymbol{A}}_k, \hat{\boldsymbol{h}}(t; \tilde{\boldsymbol{A}}_k)) \tag{24a}$$

$$= -\int_0^T \boldsymbol{r}_k^H(t) \diamond \tilde{\boldsymbol{P}}_k \diamond \boldsymbol{r}_k(t)\, dt \tag{24b}$$

$$= -\int_0^T [\tilde{\boldsymbol{P}}_k \diamond \boldsymbol{r}_k(t)]^H \diamond \tilde{\boldsymbol{P}}_k \diamond \boldsymbol{r}_k(t)\, dt \tag{24c}$$

where $\tilde{\boldsymbol{P}}_k = \tilde{\boldsymbol{A}}_k \tilde{\boldsymbol{A}}_k^I$ is the matrix which projects onto the range of $\tilde{\boldsymbol{A}}_k$ [22].

The structure suggested in (24) is akin to squaring and integrating the output of a network of TDL filters when the input contains white noise, so that the metric does not exist in the mean-square sense. To illustrate this, consider a fixed $t \in [0, T)$ so that[11]

$$\mathbb{E}\{\boldsymbol{n}_k^H(t) \diamond \tilde{\boldsymbol{P}}_k \diamond \boldsymbol{n}_k(t)\} = \mathrm{tr}\,(\tilde{\boldsymbol{P}}_k \diamond \mathbb{E}\{\boldsymbol{n}_k(t)\boldsymbol{n}_k^H(t)\} \diamond \tilde{\boldsymbol{P}}_k\} \tag{25}$$

which is not well defined since $\mathbb{E}\{n(t+\tau)n^*(t)\} = N_0\delta(\tau)$. It follows that the ensemble average of the ML metric for any path is infinite. The conclusion is that the MLSE solution for the known-channel ISI problem considered by Forney, with the general CT channel model, has no well-defined counterpart for joint-ML channel and sequence estimation.

The root of the ill-posed condition, as clearly pointed out in [13, p. 457], occurs due to the lack of structure assumed about $h(t)$. In fact, it is shown in Appendix A that this is not a technicality arising from the white noise model, but rather a fundamental problem due to this lack of assumed structure. Specifically, since the channel is assumed to have an unknown, possibly nonzero projection onto infinitely many of the eigenfunctions of the noise covariance operator, the ML channel estimate uses the components of received signal in the direction of each of these basis functions.

Two reasonable approximations are considered below, each based on the assumption that the desired signal has negligible energy above some frequency.

1) Approximation Using the Continuous-Time ML Metric: The orthogonal eigenfunctions of the white noise covariance operator can always be taken to be harmonically spaced complex exponentials. These functions are also the eigenfunctions of any wide-sense stationary covariance operator for asymptotically large observation intervals [13]. Truncation of the series expansion in (6) with these functions corresponds to ignoring high frequency energy in $y(t)$.

In practical situations, the noise confronting the processor is broadband, but has finite power due to physical filtering effects. The ML metric for this broadband noise model is also ill-defined as is shown in Appendix A. However, implementing the white-noise metric of (23) in the presence of broadband noise effectively implements the truncation of the series expansion described above; thus, minimization of (23) in the presence finite-power broadband noise is well defined and effectively optimal for the truncated series metric.

Even with this interpretation, which circumvents the ill-posed condition, the CT signal processing implied by the metric of (24) (and the time-recursive version in Section IV) is not likely to be amenable to implementation using digital hardware. For this reason we next show that the general CT channel model can be approximated by the TDL model for practical channels.

2) The TDL as an Approximation to the Continuous-Time Channel: The approximations

$$c(t) \cong \sum_{l=0}^{N_\Delta L_c} c(l\Delta)\,\mathrm{sinc}\left(\frac{t - l\Delta}{\Delta}\right) \tag{26}$$

$$u(t) \cong u(t) * \frac{1}{\Delta}\,\mathrm{sinc}\,(t/\Delta) \tag{27}$$

where $N_\Delta = T/\Delta$ is assumed to be an integer and $\mathrm{sinc}\,(x) \triangleq \sin(\pi x)/(\pi x)$, imply

$$h(t) \cong \sum_{l=0}^{N_\Delta L_c} c_l u(t - l\Delta) \tag{28}$$

with $c_l = \Delta c(l\Delta)$. Since $c(t)$ and $u(t)$ are strictly time-limited signals, they are not bandlimited. However, because both are assumed to have finite energy, they are each approximately bandlimited in the sense that the energy outside the frequency band $[-1/\Delta, 1/\Delta]$ can be made arbitrarily small by decreasing the positive parameter Δ. It follows that the approximation error for each of (26)–(28) can be made arbitrarily small by decreasing Δ. Thus, the TDL channel model of Section II-A may be interpreted as an approximation to the general CT channel model with the resolution time T_r defined so that $\Delta = T_r$ results in an acceptably small approximation error. This simple result, which is the static, deterministic analog to

[10] This is a special case of the developments of Appendix A and Section IV.

[11] The notation $\mathrm{tr}\,(\cdot)$ is used to represent the trace of a square matrix.

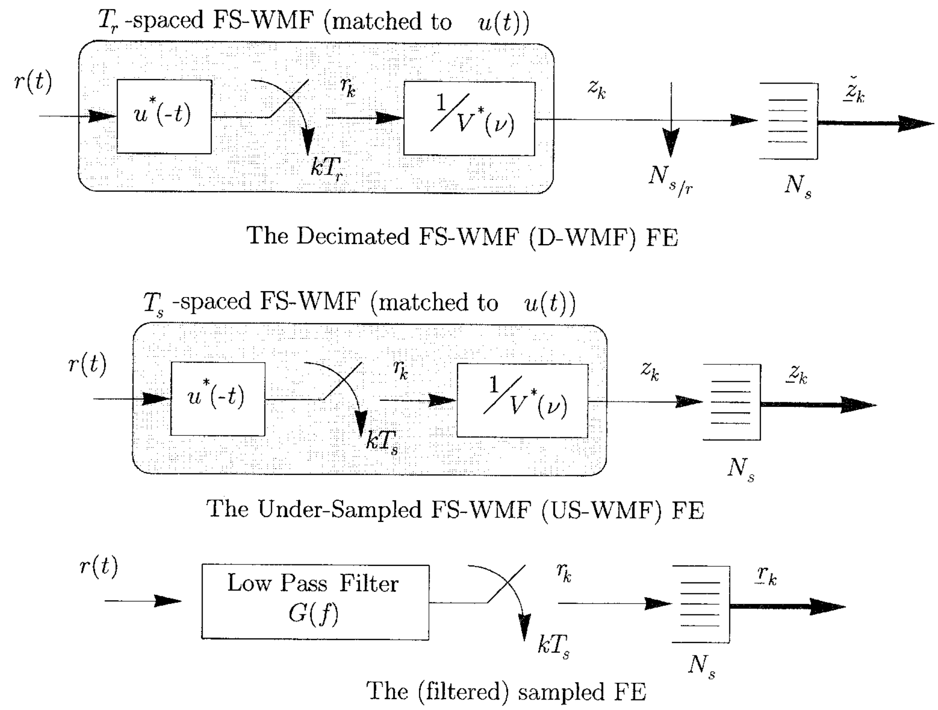


Fig. 4. Suboptimal FE processors considered.

Bello's multipath fading channel model [23], is valid for any time-limited, finite energy channel $h(t)$.

III. Suboptimal Front-Ends

The signal models at the output of three suboptimal FE processors are developed in this section. These three processors, illustrated in Fig. 4, are based on the approximations of Sections II-B1 and II-B2 and related simplifications. Each of these FE processors result in $(N_s \times 1)$ vector component symbol-rate ISI models. Each is suboptimal when $N_s < N_r$, and optimal (i.e., equivalent to the T_r FS-WMF) when $N_s = N_r$. One reason to consider these processors is to assess the effects of suboptimal FE processors suggested or implied in the literature. Another motivation is to determine which technique should be used in an effort to reduce the receiver complexity (i.e., to reduce the number of samples per symbol N_s).

A. The Decimated FS-WMF (D-WMF)

Consider the case when the output of the FS-WMF in Fig. 3 is decimated so that only $N_s < N_r$ samples are retained per symbol period. Whenever discussing multirate processing, we will assume that $N_{s/r} = T_s/T_r = N_r/N_s$ is an integer, so that the output of such a decimation is

$$\check{z}_k = \sum_i a_i \check{f}_{k-iN_s} + \check{w}_k \tag{29}$$

where $\check{m}_k \triangleq m_{kN_{s/r}}$ is the decimated version of an arbitrary sequence $\{m_i\}$.

This T_s-spaced model can be converted to a symbol-spaced ISI model by creating the $(N_s \times 1)$ component vectors of decimated signals

$$\check{\underline{z}}_k = \check{\underline{x}}_k + \check{\underline{w}}_k = \sum_{m=0}^{L-1} a_{k-m}\check{\underline{f}}_m + \check{\underline{w}}_k \tag{30}$$

where the component vectors are defined in a manner analogous to (15).

B. The Under-Sampled FS-WMF (US-WMF)

The US-WMF may be considered the result of designing an FS-WMF when the bandwidth of the system has been underestimated. Specifically, consider the effect of designing the receiver assuming that $T_s = T_r' > T_r$. In this case, the WF is mismatched, so that the resulting white noise model may have infinite length ISI. This is shown by modifying the development of Section II-A with the assumption that $T_s > T_r$. For the special case of $T_s = T$, the US-WMF is a pulse-matched filter, sampled at the symbol rate; if, in addition, an $L_u = 1$ rectangular data pulse is assumed, this corresponds to the symbol-rate integrate-and-dump FE suggested in [5]. A similar technique with pulse matched filtering and one sample per symbol has been recommended for global system for mobile communications (GSM) [11].

The output samples of the filter matched to $u(t)$ are

$$r_k = y_k + n_k = \sum_{i=0}^{I-1} a_i R_{hu}([k - iN_s]T_s) + n_k \tag{31}$$

where

$$\begin{aligned} R_{hu}(kT_s) &= c(\tau) * R_u(\tau)|_{\tau=kT_s} \\ &= \int_{-\infty}^{\infty} R_u(kT_s - s)c(s)\, ds. \end{aligned} \tag{32}$$

The output of the T_s-spaced WMF then involves the convolution of $R_{hu}(kT_s)$ with $v_k^{-\dagger}$, the inverse of the minimum-phase factor of $R_u(kT_s)$ (i.e., the impulse response of the realizable WF).

Assuming the TDL approximation developed in Section II-B2, or approximating (32) by a Riemann sum over a uniformly T_r-spaced partition, yields

$$R_{hu}(kT_s) = \sum_{l=0}^{N_r L_c} c_l R_u([kN_{s/r} - l]T_r) = \sum_{l=0}^{N_r L_c} c_l R(k;l) \tag{33}$$

where $R(k;l) \triangleq R_u([kN_{s/r} - l]T_r)$.

The noise-free US-WMF output is of the form

$$x_k = \sum_i a_i f_{k-iN_s} \tag{34}$$

where the equivalent unknown channel is

$$f_k = R_{hu}(kT_s) * v_k^{-\dagger} = \sum_{l=0}^{N_r L_c} c_l v_k(l) \tag{35}$$

and the *shift-variant DT* (T_s) *equivalent data pulse* is defined by

$$v_k(l) = R(k;l) * v_k^{-\dagger} = \sum_{j=-\infty}^{\infty} R(k-j;l) v_j^{-\dagger}. \tag{36}$$

Some redundancy in (35) may be removed by noting that $v_k(m + nN_{s/r}) = v_{k-n}(m)$, so that

$$f_k = \sum_{m=0}^{N_{s/r}-1} \sum_{n=0}^{L_c N_s} c_{m+nN_{s/r}} v_{k-n}(m) = \sum_{m=0}^{N_{s/r}-1} f_k(m). \tag{37}$$

The interpretation of (37), which reduces to (12) for $N_{s/r} = 1$, requires that $c_l = 0$ for $l \notin \mathcal{Z}_{N_r L_c + 1}$.[12] For a given value of m, $f_k(m)$ represents the convolution of the WF and a (fractionally) shifted, sampled version of the pulse correlation function. Since the WF is autoregressive, it follows that, except for $m = 0, f_k(m)$ is nonzero for possibly infinitely many values of k.

Converting the T_s-spaced model into an $(N_s \times 1)$ vector component model yields

$$\underline{z}_k = \sum_{j \in \mathcal{Z}_L} a_{k-j} \underline{f}_j + \sum_{j \notin \mathcal{Z}_L} a_{k-j} \underline{f}_j + \underline{w}_k \tag{38}$$

where the sum has been split in order to emphasize that a receiver designed around the incorrect assumption that $T_s = T_r$ may encounter unexpected ISI.

[12] This follows from the convention that $c_{N_r L_c}$ may be nonzero, which was selected in an effort to maintain the relation $L = L_c + L_u$ throughout.

C. The Sampled Front-End

This FE processing consists of sampling the received signal (or a filtered version) and has been suggested in [4]–[6], [8], and [12]. After sampling, the component function model of (20) is approximated by the component vector model defined by

$$\underline{r}_k = \underline{y}_k + \underline{n}_k = \sum_{m=0}^{L-1} a_{k-m} \underline{h}_m + \underline{n}_k \tag{39}$$

where the component vector of samples of an arbitrary function, $m(t)$, is defined by

$$\underline{m}_k = [m(kT + (N_s - 1)T_s) \quad \cdots \quad m(kT)]^T. \tag{40}$$

To relate the sampled FE to each of the FE processors discussed thus far, consider the case when the low-pass filter in Fig. 4 is ideal with bandwidth W (i.e., $G(f) = 1$ for $|f| \leq W$ and zero otherwise). Also, assume that the approximation in (27) holds with sufficiently low error for $\Delta = T_r$, and that $c(t)$ is a T_r–spaced TDL. Two choices of W are of interest: $W_r = 1/(2T_r)$, which will produce negligible distortion in the signal, and $W_s = 1/(2T_s)$, which is an anti-aliasing filter and results in a reduced noise power at the output of the sampler. Under the above assumptions, it is straightforward to show that the case of $T_s = T_r$ and $W = W_r$, the sampling FE is equivalent to the optimal FS-WMF. This implies that, under an approximately bandlimited assumption, Nyquist-rate sampling provides a set of sufficient statistics. For $T_s \geq T_r$ and $W = W_r$, the sampling FE is equivalent to the D-WMF. When $T_s \geq T_r$ and $W = W_s$, the sampled FE can be viewed as the optimal FE for a filtered version the data pulse (i.e., the FS-WMF of a reduced resolution system). The US-WMF is similar to this ($T_s \geq T_r, W = W_s$) case except that rather than rejecting the energy in the pulse outside the band $[W_s, W_s]$, the US-WMF folds these high-frequency signal components back into the band. Thus, the US-WMF cannot, in general, be accurately approximated by the sampled FE when $G(f)$ is not matched to the data pulse.

Hence, the D-WMF may be viewed as a technique that sacrifices signal-to-noise ratio (SNR) for reduced signal distortion, while employing the US-WMF has the opposite effect, and the optimal FS-WMF achieves the benefits of both at the expense of more samples per symbol.

IV. Recursive Computation of the ML Metric

Each of the FE's considered resulted in the formulation of an equivalent symbol-spaced ISI model with white Gaussian noise

$$r_i = y_i + n_i = \sum_{m=0}^{L-1} a_{i-m} h_m + n_i, \qquad i = 0, 1, 2 \cdots. \tag{41}$$

The notation in (41) is intended to be general; it is only assumed that $h_i \in \mathcal{H}$ and that r_i, y_i and n_i are random elements of $\mathcal{H}$, a Hilbert space.[13]

[13] We use the same letters to denote the general case and the CT model of Section II-B. This is only so that yet another set of equivalent variables is not necessary.

The component equations of (41) may be written in vector form as

$$\boldsymbol{r}_k = \boldsymbol{y}_k + \boldsymbol{n}_k = \boldsymbol{A}_k \diamond \boldsymbol{h} + \boldsymbol{n}_k \tag{42}$$

where $\boldsymbol{h} \in \mathcal{H}^L$ and $\boldsymbol{r}_k, \boldsymbol{y}_k$ and $\boldsymbol{n}_k$ are random elements of $\mathcal{H}^{k+1}$ defined in a manner analogous to (17). In the present context, $\boldsymbol{A}_k$ should be viewed as a bounded linear operator mapping $\mathcal{H}^L$ into $\mathcal{H}^{k+1}$, with the special structure of (22).

Due to the white Gaussian nature of the observation noise, in each case considered, the joint-ML channel and sequence estimates are found by performing the following minimization:

$$\min_{\tilde{\boldsymbol{A}}_k, \tilde{\boldsymbol{h}}} ||\boldsymbol{r}_k - \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}||^2_{\mathcal{H}^{k+1}} \tag{43}$$

where the search is over all $\tilde{\boldsymbol{h}} \in \mathcal{H}^L$ and all $\tilde{\boldsymbol{A}}_k$ of the form given in (22). The inner product in the product space is defined in the usual manner [24]

$$\langle \boldsymbol{x} | \boldsymbol{z} \rangle_{\mathcal{H}^{k+1}} \triangleq \sum_{i=0}^{k} \langle x_i | z_i \rangle_{\mathcal{H}}. \tag{44}$$

For increased numerical stability or to accommodate a slowly time-varying channel, an exponentially weighted version of the squared error cost function in (43), with forgetting factor $\rho \in (0, 1]$, is considered [25]

$$\begin{aligned}\Lambda_k(\tilde{\boldsymbol{A}}_k, \tilde{\boldsymbol{h}}) &\triangleq (\boldsymbol{r}_k - \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}})^H \diamond \boldsymbol{W}_k \diamond (\boldsymbol{r}_k - \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}) \\ &= ||\boldsymbol{r}_k - \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}||^2_{\mathcal{H}^{k+1}, \rho}\end{aligned} \tag{45}$$

where $\boldsymbol{W}_k = \mathbf{diag}\,(1, \rho, \rho^2, \cdots \rho^k)$ is the $((k+1) \times (k+1))$ diagonal weighting matrix. For a fixed $\tilde{\boldsymbol{A}}_k$ the metric is minimized by

$$\hat{\boldsymbol{h}}_k(\tilde{\boldsymbol{A}}_k) = \hat{\boldsymbol{h}}_k = \tilde{\boldsymbol{A}}_k^I \diamond \boldsymbol{r}_k \tag{46}$$

where $\tilde{\boldsymbol{A}}_k^I$ is the pseudo-inverse of $\tilde{\boldsymbol{A}}_k$ with respect to the weighted inner product which generates the norm in (45). If $\tilde{\boldsymbol{A}}_k$ has rank L, this may be expressed explicitly as [26]

$$\tilde{\boldsymbol{A}}_k^I = (\tilde{\boldsymbol{A}}_k^H \boldsymbol{W}_k \tilde{\boldsymbol{A}}_k)^{-1} \tilde{\boldsymbol{A}}_k^H \boldsymbol{W}_k. \tag{47}$$

Substituting (46) into (45) yields a metric which explicitly involves only the data

$$\begin{aligned}\Lambda_k(\tilde{\boldsymbol{A}}_k) &\triangleq \Lambda_k(\tilde{\boldsymbol{A}}_k, \hat{\boldsymbol{h}}_k(\tilde{\boldsymbol{A}}_k)) = ||\boldsymbol{W}_k^{1/2} \tilde{\boldsymbol{Q}}_k \diamond \boldsymbol{r}_k||^2_{\mathcal{H}^{k+1}} \\ &= ||\tilde{\boldsymbol{Q}}_k \diamond \boldsymbol{r}_k||^2_{\mathcal{H}^{k+1}, \rho}\end{aligned} \tag{48}$$

where the error projection matrix for the weighted least squares solution is

$$\tilde{\boldsymbol{Q}}_k = \boldsymbol{I} - \tilde{\boldsymbol{A}}_k \tilde{\boldsymbol{A}}_k^I \tag{49}$$

and $\boldsymbol{W}_k^{1/2}$ is diagonal, with elements equal to the square root of those in $\boldsymbol{W}_k$.

Assuming that $\tilde{\boldsymbol{A}}_{k-1}$ has rank L, the metric in (48) can be computed recursively as

$$\Lambda_k(\tilde{\boldsymbol{A}}_k) = \rho \Lambda_{k-1}(\tilde{\boldsymbol{A}}_{k-1}) + \frac{\rho}{\rho + \tilde{\Delta}_k} ||r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_{k-1}||^2_{\mathcal{H}}. \tag{50}$$

Not surprisingly, the metric recursion requires computation of the recursive least squares (RLS) channel estimate associated with the data path through the following:

$$\tilde{\Delta}_k = \tilde{\boldsymbol{\alpha}}_k^H \tilde{\boldsymbol{B}}_{k-1} \tilde{\boldsymbol{\alpha}}_k \tag{51a}$$

$$\tilde{\boldsymbol{g}}_k = \frac{\tilde{\boldsymbol{B}}_{k-1} \tilde{\boldsymbol{\alpha}}_k}{\rho + \tilde{\Delta}_k} \tag{51b}$$

$$\hat{\boldsymbol{h}}_k = \hat{\boldsymbol{h}}_{k-1} + \tilde{\boldsymbol{g}}_k (r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_{k-1}) \tag{51c}$$

$$\tilde{\boldsymbol{B}}_k = (\tilde{\boldsymbol{A}}_k^H \boldsymbol{W}_k \tilde{\boldsymbol{A}}_k)^{-1} = \rho^{-1} (\boldsymbol{I} - \tilde{\boldsymbol{g}}_k \tilde{\boldsymbol{\alpha}}_k^H) \tilde{\boldsymbol{B}}_{k-1}. \tag{51d}$$

The dependence of recursion parameters in (50)–(51) on the hypothesized path $\tilde{\boldsymbol{A}}_k$ has been suppressed for the sake of compactness. We note that, with the exception of (51c), the RLS processing in (51) does not depend on the form of the signal Hilbert space.

The recursion in (50) is easily derived using the matrix inversion lemma and other standard results to obtain a recursion on $\tilde{\boldsymbol{Q}}_k$, yet it has not, to our knowledge, appeared in the joint channel-sequence estimation literature. This recursion also follows directly from a slightly different form

$$\begin{aligned}\Lambda_k(\tilde{\boldsymbol{A}}_k) &= \rho \Lambda_{k-1}(\tilde{\boldsymbol{A}}_{k-1}) \\ &\quad + \langle (r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_{k-1}) | (r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_k) \rangle_{\mathcal{H}}\end{aligned} \tag{52}$$

which appears in a standard reference [27].

The form of the update in (50) has the advantage that the one-step metric update can be computed without first updating the channel estimate, but the form in (52) may be approximated in a meaningful fashion for any type of channel estimator (i.e., non-RLS).

A. A Reduced Complexity Approximation

The exact ML metric recursion given in (50) requires computation of the RLS channel estimate for each hypothesized data path. Since any successful algorithm will need to search many such paths, a reduced complexity approximation to (50) is desired. An obvious choice is the considerably less complex least mean squares (LMS) estimator, which has been shown to track dynamic channels with performance comparable to that of the RLS algorithm [27], [28], [29].

The LMS channel estimate is updated by the recursion [27]

$$\hat{\boldsymbol{h}}_k = \hat{\boldsymbol{h}}_{k-1} + \beta (r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_{k-1}) \tilde{\boldsymbol{\alpha}}_k = \hat{\boldsymbol{h}}_{k-1} + \beta \tilde{e}_k \tilde{\boldsymbol{\alpha}}_k \tag{53}$$

where $\beta \in \mathcal{R}$ is the step size and $\tilde{e}_k$ is the one-step prediction error. An approximation to the metric update can then be defined in the spirit of (52) as the inner product of the prediction and filtering error, specifically

$$\langle r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_{k-1} | r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_k \rangle_{\mathcal{H}} = (1 - \beta ||\tilde{\boldsymbol{\alpha}}_k||^2) ||\tilde{e}_k||^2_{\mathcal{H}}. \tag{54}$$

The LMS approximation to the ML metric recursion in (50) is therefore

$$\begin{aligned}\Lambda(\tilde{\boldsymbol{A}}_k) &= \rho \Lambda_{k-1}(\tilde{\boldsymbol{A}}_{k-1}) + (1 - \beta ||\tilde{\boldsymbol{\alpha}}_k||^2) \\ &\quad \cdot ||r_k - \tilde{\boldsymbol{\alpha}}_k^H \diamond \hat{\boldsymbol{h}}_{k-1}||^2_{\mathcal{H}}\end{aligned} \tag{55}$$

where the per-path channel estimates are defined in (53). Note that in the case of constant-envelope signaling and $\rho = 1$, this metric is the same as the known-channel MLSE metric with the channel replaced by the per-path LMS estimate.

V. Conclusion

By investigating the joint-ML channel and sequence estimation problem from first principles (i.e., the CT observation), several important results regarding optimality were demonstrated. First, it was shown that, strictly speaking, there is no optimal receiver in the joint-ML sense when the channel is unknown. By considering a model which characterizes the uncertainty in the channel by a finite number of parameters (e.g., the TDL model), the practically optimal FS-WMF front-end was derived. One important consequence of the resulting component vector ISI model is that earlier work which neglected FE processing and concentrated on post-processor design can be easily modified to include the optimal FE processing, i.e., the conceptual partition between the FE and post-processor for unknown channel MLSE has been established. Along these lines, we presented a recursive computation of the ML metric which led to receiver structures that were natural extensions of the direct structure for known-channel MLSE. This may be viewed as both a generalization of the PSP technique and its theoretical foundation. An important conclusion is that any receiver claimed to be optimal or quasi-optimal must employ both fractionally-spaced processing and per-data-sequence channel estimation. This is in direct contrast to the known-channel MLSE receiver which requires only symbol-rate sampling.

A framework for the design of recursive unknown-channel MLSE algorithms has been established. The design can be viewed as consisting of three phases:

1) system modeling, which includes modeling of the unknown parameters and consideration of sufficient statistics (i.e., FE processing);
2) metric recursion development, which involves obtaining the recursion and possibly making intelligent complexity reduction approximations (e.g., LMS in place of RLS); and
3) tree search design, which has a well-established history [15], [16].

This framework is a starting point for the development of unknown channel receiver designs, not the conclusion. Determining the best design for a given complexity is a complicated task and is surely dependent on the modeling and expected operational mode. Many issues remain unresolved in each of the three phases. For example, the fact that several data matrices may have the same projection matrix and, hence, the same joint-ML metric, significantly affects the acquisition performance of generalized PSP receivers. The tracking mode performance and related open issues are the subject of Part II [1]. The effectiveness of implicit versus explicit modeling of additional signal parameters is also an open issue. For example, while it appears that symbol timing and carrier phase were assumed known throughout, uncertainty in these quantities may actually be included in the results discussed—i.e., a symbol timing uncertainty within one symbol period may be handled by increasing the channel memory by one in the model.

Appendix
CT Channel Model in Colored Noise

The colored noise generalization of the metric in (4) is[14]

$$\Gamma_k(\tilde{\boldsymbol{a}}_k,\tilde{h}) = \langle \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}(t) | \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}(t)\rangle_n - 2\Re\{\langle \boldsymbol{r}_k(t) | \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}(t)\rangle_n\} \tag{56a}$$

$$\sim ||\boldsymbol{r}_k(t) - \tilde{\boldsymbol{A}}_k \diamond \tilde{\boldsymbol{h}}(t)||_n^2 \tag{56b}$$

where the inner product in the reproducing kernel Hilbert space (RKHS) associated with the covariance of the noise is defined by[15]

$$\langle \boldsymbol{x}_k(t)|\boldsymbol{z}_k(t)\rangle_n = \iint_0^T \boldsymbol{x}_k^H(t)\boldsymbol{K}^{-1}(t,s)\boldsymbol{z}_k(s)\,ds\,dt \tag{57}$$

$$= \iint_0^{(k+1)T} x(t)K_n^{-1}(t,s)z(s)\,ds\,dt \tag{58}$$

where $K_n^{-1}(t,s)$ is the inverse kernel of the noise covariance operator

$$\int_0^{(k+1)T} K_n^{-1}(t,\beta)K_n(\beta,s)\,d\beta = \delta(t-s). \tag{59}$$

The chipped covariance operator is defined by

$$\boldsymbol{K}(t,s) = \mathbb{E}\{\boldsymbol{n}_k(t)\boldsymbol{n}_k^H(s)\} \tag{60}$$

so that the (i,j) element of $\boldsymbol{K}(t,s)$ is $K_n(t+iT, s+jT)$ $t,s \in [0,T)$. This matrix is the chipped version of the square-integrable noise covariance kernel defined for $s,t \in [0,(k+1)T)$ and $\boldsymbol{K}^{-1}(s,t)$ is the chipped version of the inverse kernel so that

$$\int_0^T \boldsymbol{K}^{-1}(t,\beta)\boldsymbol{K}(\beta,s)\,d\beta = \delta(t-s)\boldsymbol{I}. \tag{61}$$

The equivalence of the two notations is summarized by

$$\boldsymbol{z}_k(t) = \int_0^T \boldsymbol{K}(t,s)\boldsymbol{x}_k(s)\,ds \qquad t \in [0,T) \quad \Leftrightarrow$$
$$z(t) = \int_0^{(k+1)T} K_n(t,s)x(s)\,ds \qquad t \in [0,(k+1)T) \tag{62}$$

where the conversion from signals on $[0,(k+1)T)$ to $(k+1)$-tuples with component functions supported on $[0,T)$ is as described in Section II-B. The above linear operation will be symbolically denoted simply by $\boldsymbol{z} = \boldsymbol{K}\boldsymbol{x}$. For compactness, the subscript k will be suppressed in the following, as will the mixed product symbol ($\diamond$).

Application of the orthogonality principle of the Hilbert space projection theorem [24] to the heuristic expression in

[14] As discussed in [30], $\boldsymbol{r}_k$ is a generalized element of the RKHS, so the second term is symbolic for an Itō integral. A similar remark applies to the the white noise case.

[15] For the sake of compactness, double integrals with identical limits on each integration variable are denoted with a single set of limits.

(56), or the calculus of variations to the exact metric leads to the critical point for the channel estimate must satisfy

$$\tilde{\boldsymbol{A}}^H \boldsymbol{K}^{-1}\boldsymbol{r} = \tilde{\boldsymbol{M}}\hat{\boldsymbol{h}} \tag{63}$$

where $\tilde{\boldsymbol{M}}$ is the linear operator with kernel

$$\tilde{\boldsymbol{M}}(t,s) = \tilde{\boldsymbol{A}}^H \boldsymbol{K}^{-1}(t,s)\tilde{\boldsymbol{A}}. \tag{64}$$

The operator $\tilde{\boldsymbol{M}}$ is a nonnegative definite (Hermitian symmetric), bounded linear operator mapping $(\mathcal{L}_2[0,T))^L$ into itself. This follows from the fact that $K_n(t,s)$ is a positive definite, bounded linear operator on $\mathcal{L}_2[0,(k+1)T)$. Mercer's theorem guarantees the existence of the expansion in terms of orthogonal rank one operators [13]

$$\tilde{\boldsymbol{M}}(t,s) = \sum_i \gamma_i \boldsymbol{e}_i(t)\boldsymbol{e}_i^H(s) \tag{65}$$

where $\{\boldsymbol{e}_i\}$ is the countable collection of orthonormal (with respect to the standard $\mathcal{L}_2$ inner product) eigenvectors of $\tilde{\boldsymbol{M}}$

$$\tilde{\boldsymbol{M}}\boldsymbol{e}_i = \gamma_i \boldsymbol{e}_i \qquad \gamma_i \geq 0 \text{ (real)} \tag{66}$$

$$\int_0^T \boldsymbol{e}_i^H(t)\boldsymbol{e}_j(t)\,dt = \delta_{ij}. \tag{67}$$

If the $((k+1)\times L)$ matrix, $\tilde{\boldsymbol{A}}$, has rank $L \geq k+1$, then $\tilde{\boldsymbol{M}}$ is invertible (positive definite), so that the closed form channel estimate for a given data path may be written as[16]

$$\hat{\boldsymbol{h}} = \tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{A}}^H\boldsymbol{K}^{-1}\boldsymbol{r} \quad \Leftrightarrow$$
$$\hat{\boldsymbol{h}}(t) = \iint_0^T \tilde{\boldsymbol{M}}^{-1}(t,\beta)\tilde{\boldsymbol{A}}^H\boldsymbol{K}^{-1}(\beta,s)\boldsymbol{r}(s)\,ds\,d\beta. \tag{68}$$

The kernel of the inverse operator of $\tilde{\boldsymbol{M}}$ is defined as

$$\tilde{\boldsymbol{M}}^{-1}(t,s) = \sum_i \frac{1}{\gamma_i}\boldsymbol{e}_i(t)\boldsymbol{e}_i^H(s) \tag{69}$$

with all of the eigenvalues of $\tilde{\boldsymbol{M}}$ being positive due to the assumption that $\tilde{\boldsymbol{A}}$ is full rank. Substituting back into (56) yields the ML metric for the data sequence

$$\Gamma_k(\tilde{\boldsymbol{A}}) = \iint_0^T \boldsymbol{r}^H(t)\tilde{\boldsymbol{P}}(t,s)\boldsymbol{r}(s)\,ds\,dt \tag{70}$$

with

$$\tilde{\boldsymbol{P}} = \boldsymbol{K}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{A}}^H\boldsymbol{K}^{-1}. \tag{71}$$

The operator $\tilde{\boldsymbol{P}}$ is the projection operator onto the range space of $\tilde{\boldsymbol{A}}$ in the inner product of the RKHS associated with $\boldsymbol{K}$. Since $\boldsymbol{r}$ is a generalized element of the RKHS, problems similar to those encountered in the white noise case are expected.

[16] Modification of the following development for the case when $\tilde{\boldsymbol{A}}$ is not full rank is straightforward.

In order to clearly demonstrate this singularity, a factorization of $\tilde{\boldsymbol{M}}^{-1}$ is introduced

$$\tilde{\boldsymbol{M}}^{-1}(t,s) = \int_0^T \tilde{\boldsymbol{G}}(t,\beta)\tilde{\boldsymbol{G}}(\beta,s)\,d\beta \tag{72a}$$
$$= \int_0^T \tilde{\boldsymbol{G}}^H(\beta,t)\tilde{\boldsymbol{G}}(\beta,s)\,d\beta \tag{72b}$$

with $\tilde{\boldsymbol{G}}$ defined by

$$\tilde{\boldsymbol{G}}(t,s) = \sum_i \frac{1}{\sqrt{\gamma_i}}\boldsymbol{e}_i(t)\boldsymbol{e}_i^H(s). \tag{73}$$

Substituting this factorization into (70) yields

$$\Gamma_k(\tilde{\boldsymbol{A}}) = \int_0^T \operatorname{tr}(\tilde{\boldsymbol{v}}(\alpha)\tilde{\boldsymbol{v}}^H(\alpha))\,d\alpha \tag{74}$$

where

$$\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{A}}^H\boldsymbol{K}^{-1}\boldsymbol{r}. \tag{75}$$

The fact that $\tilde{\boldsymbol{v}}^H(t)\tilde{\boldsymbol{v}}(t) = \operatorname{tr}(\tilde{\boldsymbol{v}}(t)\tilde{\boldsymbol{v}}^H(t))$ along with the Hermitian symmetry of $\boldsymbol{K}^{-1}$ has also been used to obtain (75). Neglecting a finite constant associated with the noise-free signal, the average of the ML metric for a fixed path is

$$\begin{aligned}\mathbb{E}\{\Gamma_k(\tilde{\boldsymbol{A}})\} \sim \int_0^T &\operatorname{tr}(\mathbb{E}\{\tilde{\boldsymbol{G}}(\alpha,\beta)\tilde{\boldsymbol{A}}^H\boldsymbol{K}^{-1}(\beta,s)\boldsymbol{n}(s)\\ &\cdot \boldsymbol{n}^H(t)\boldsymbol{K}^{-1}(t,\xi)\tilde{\boldsymbol{A}}\tilde{\boldsymbol{G}}(\xi,\alpha)\})\,d(t,s,\beta,\xi,\alpha)\\ = \int_0^T &\operatorname{tr}\tilde{\boldsymbol{G}}(\alpha,\beta)\tilde{\boldsymbol{A}}^H\boldsymbol{K}^{-1}(\beta,\xi)\tilde{\boldsymbol{A}}\tilde{\boldsymbol{G}}(\xi,\alpha))\\ &\cdot d(\beta,\xi,\alpha)\end{aligned} \tag{76}$$

where the abbreviated notation is defined as

$$\underbrace{\int_0^T\cdots\int_0^T}_{n-\text{fold}} f(x_1,\cdots x_n)\,dx_1\cdots dx_n = \int_0^T f(x_1,\cdots x_n)\,d(x_1,\cdots x_n). \tag{77}$$

Using the orthogonal expansions, it is straightforward to show that

$$\iint_0^T \tilde{\boldsymbol{G}}(t,\beta)\underbrace{\tilde{\boldsymbol{A}}^H\boldsymbol{K}^{-1}(\beta,\xi)\tilde{\boldsymbol{A}}}_{=\tilde{\boldsymbol{M}}(\beta,\xi)}\tilde{\boldsymbol{G}}(\xi,s)\,d\xi\,d\beta = \delta(t-s)\boldsymbol{I} \tag{78}$$

so that the ML metric does not exist in the mean sense since

$$\mathbb{E}\{\Gamma_k(\tilde{\boldsymbol{A}})\} \sim \int_0^T \operatorname{tr}\{\delta(\alpha-\alpha)\boldsymbol{I}\}\,d\alpha. \tag{79}$$

ACKNOWLEDGMENT

The authors thank the anonymous reviewers whose comments improved this paper.

REFERENCES

[1] K. M. Chugg and A. Polydoros, "MLSE for an unknown channel—Part II: Tracking performance," *IEEE Trans. Commun.*, vol. 44, no. 8, Aug. 1996.

[2] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Informat. Theory*, vol. IT-18, pp. 363–378, May 1972.

[3] G. Ungerboeck, "Adaptive maximum likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. COM-22, pp. 624–636, May 1974.

[4] N. Seshadri, "Joint data and channel equalization using fast blind trellis search techniques," in *Proc. GLOBECOM'90*, Dec. 1990, pp. 1659–1663.

[5] M. Ghosh, "An optimal approach to blind equalization," Ph.D. dissertation, Univ. Southern California, Los Angeles, CA, Dec. 1991.

[6] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," in *Proc. SPIE*, July 1991, pp. 181–195.

[7] R. Raheli, A. Polydoros, and C-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb./Mar./Apr. 1995.

[8] K. Giridhar, J. J. Shynk, and R. A. Iltis, "Bayesian/decision-feedback algorithm for blind adaptive equalization," *Optical Engineering*, vol. 31, pp. 1211–1223, June 1992.

[9] S. Haykin, Ed., *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994, Ch. 6 by N. Seshadri.

[10] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery—Application to equalization in underwater communications," *IEEE J. Oceanic Eng.*, vol. 16, pp. 42–55, Jan. 1991.

[11] GSM Recommendations no. 05.05-DCS (version 3.0) "Radio transmission and reception," European Telecommunications Standards Institute, Jan. 1992.

[12] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for a priori unknown channels and symbol timing," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 579–588, Apr. 1992.

[13] H. L. VanTrees, *Detection, Estimation, and Modulation Theory Part I*. New York: Wiley, 1968.

[14] C-K. Tzou, R. Raheli, and A. Polydoros, "Applications of per-survivor processing to mobile digital communications," in *Proc. GLOBECOM'93*, Nov. 1993.

[15] J. B. Anderson and S. Mohan, "Sequential coding algorithms: A survey cost analysis," *IEEE Trans. Commun.*, vol. COM-32, pp. 169–176, Feb. 1984.

[16] ——, *Source and Channel Coding—An Algorithmic Approach*. Norwell, MA: Kluwer, 1991.

[17] R. A. Iltis, "A digital receiver for demodulation of CDMA waveforms with a priori unknown delays and amplitudes," in *Proc. Milcom*, 1991, pp. 5.3.1–5.3.1-4.

[18] Z. Xie, C. K. Rushforth, R. T. Short, and T. K. Moon, "Joint signal detection and parameter estimation in multiuser communications," *IEEE Trans. Commun.*, vol. 41, pp. 1208–1216, Aug. 1993.

[19] A. Reichman and R. A. Scholtz, "Joint phase estimation and data decoding for TCM systems," in *Proc. First Int. Symp. Commun. Theory Application*, Sess 1, Sept. 1991.

[20] A. N. D'Andrea, U. Mengali and G. M. Vitetta, "Approximate ML decoding of coded PSK with no explicit carrier phase reference," *IEEE Trans. Commun.*, vol. 42, pp. 1033–1039, Feb./Mar./Apr. 1994.

[21] ——, "Multiple phase synchronization in continuous phase modulation," *Digital Signal Processing Rev. J.*, vol. 3, no. 3, pp. 188–198, July 1993.

[22] G. Strang, *Linear Applications and Its Applications, 3rd ed.* Fort Worth, TX: Saunders HBJ, 1989.

[23] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, pp. 360–393, 1963.

[24] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[25] E. Eleftheriou and D. Falconer, "Tracking properties and steady state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1097–1109, Oct. 1986.

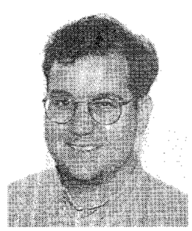
[26] J. M. Mendel, *Lessons in Digital Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[27] S. Haykin, *Adaptive Filter Theory*, 2nd Ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[28] A. Clark and R. Harun, "Assessment of Kalman-filter channel estimators for an HF radio link," *Proc. IEE*, vol. 133, pp. 513–521, Oct. 1986.

[29] S. McLaughlin, B. Mulgrew, and C. Cowan, "Performance comparison of least squares and least mean squares algorithms as HF channel estimators," in *Proc. IEEE Accoust., Speech, Signal Processing*, pp. 2105–2108, 1987.

[30] T. Kailath, "RKHS approach to detection and estimation problems—Part I: Deterministic signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 530–549, Sept. 1971.



Keith M. Chugg (S'88–M'95) was born in Los Angeles, CA, on March 16, 1967. He received the B.S. degree (high distinction) in engineering from Harvey Mudd College, Claremont, CA, in 1989. He received the M.S. and Ph.D. degrees in electrical engineering from the Communication Sciences Institute (CSI), Electrical Engineering—Systems Department, University of Southern California (USC), in 1990 and 1995, respectively.

He served as an instructor in the Electrical Engineering—Systems Department, USC, from 1993 to 1995. During his graduate study he held summer positions at TRW, Redondo Beach, CA, in 1991 and 1992, Lincom Corp., Los Angeles, CA, in 1990, and Qualcomm Inc., San Diego, CA, in 1989. During the 1995–1996 academic year, he was an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Arizona, Tucson. He recently joined the faculty of Electrical Engineering—Systems at USC as an Assistant Professor and member of CSI. His current research interests are in the areas of signaling, detection, and estimation for digital communication systems.

Andreas Polydoros (S'76–M'78–SM'92–F'95), was born in Athens, Greece, in 1954. He received the EE Degree from the National Technical University of Athens, Greece, in 1977, the M.S.E.E. degree from the State University of New York, Buffalo, in 1977, and the Ph.D. and the EE degrees in electrical engineering from the University of Southern California (USC), Los Angeles, in 1982.

He has been a faculty member at USC in the Electrical Engineering/Systems and the Communications Sciences Institute (CSI) since 1982 and was promoted to Professor in 1992. He co-directed CSI from 1991 to 1993. His general area of scientific interest is statistical communication theory with applications to spread spectrum systems, signal detection and classification, data demodulation in uncertain environments, and multiuser radio networks. He has more than 15 years of teaching, research, and extensive consulting experience on these topics, both for the government and industry. He was the Associate Editor of Communications of the IEEE TRANSACTIONS ON INFORMATION THEORY (1987–1988), the Guest Editor of the July 1993 Special Issue on "Digital Signal Processing in Communications" for *Digital Signal Processing: A Review Journal*, and a designated Area Editor for the International Journal, *Wireless Personal Communications*.

Dr. Polydoros is the recipient of a 1986 National Science Foundation Presidential Young Investigator Award.