Maximum Likelihood Receiver for Multiple Channel Transmission Systems

W. VAN ETTEN

Abstract—A maximum likelihood (ML) estimator for digital sequences disturbed by Gaussian noise, intersymbol interference (ISI) and interchannel interference (ICI) is derived. It is shown that the sampled outputs of the multiple matched filter (MMF) form a set of sufficient statistics for estimating the input vector sequence. Two ML vector sequence estimation algorithms are presented. One makes use of the sampled output data of the multiple whitened matched filter and is called the vector Viterbi algorithm. The other one is a modification of the vector Viterbi algorithm and uses directly the sampled output of the MMF. It appears that, under a certain condition, the error performance is asymptotically as good as if both ISI and ICI were absent.

I. INTRODUCTION

It has first been pointed out by Shnidman [1] that intersymbol interference (ISI) and crosstalk between multiplexed signals are essentially identical phenomena. Kaye and George have worked out this idea by investigating the transmission of multiplexed signals over multiple channel and diversity systems [2]. The author of the underlying concise paper has presented a unified theory for treating intersymbol interference and interchannel interference (ICI) as one type of disturbance [3]. We will call the combined effect of these disturbances multidimensional interference (MDI). In the following the essentials of [3] are summarized.

The generalized Nyquist criterion formulated by Shnidman is restated in matrix notation. Furthermore an optimal linear receiver is derived, consisting of a multiple matched filter (MMF) followed by a multiple tapped delay line (MTDL). As optimization criterion is used minimum error probability and it appears that this optimum linear receiver has the same structure as the receiver derived by Kaye and George under the minimum mean-square error criterion. For a suboptimum criterion (minimum Pr (e) under the constraint that the multidimensional Nyquist criterion is satisfied) a theorem is given to calculate the tap coefficients for this case.

Up to this point is appears that several concepts known from ISI literature can be generalized for MDI. Recently maximum likelihood (ML) sequence estimation of data distrubed by noise and ISI received considerable attention [4]-[6]. Now the question arises whether these concepts can also be generalized for sequences transmitted over multiple channel systems where the output data are disturbed by noise and MDI. This concise paper gives a positive answer to this question.

II. THE MULTIPLE CHANNEL COMMUNICATION MODEL

The transmission system, to be considered in this concise paper, has M inputs and M outputs. To each input j a data sequence $\Sigma_l\, a_{l,j}\delta(t - lT)$ is applied which we want to estimate at the receiving end of the communication system. The symbols $a_{l,j}$ are elements of the alphabet $\{0,1,\cdots, L-1\}$. Symbols that are applied to the several inputs of the system at the same instant lT are ordered systematically in the input vector

Paper approved by the Associate Editor for European Contributions of the IEEE Communications Society for publication without oral presentation. Manuscript received April 14, 1975; revised July 16, 1975.

The author is with the Department of Electrical Engineering, Eindhoven Institute of Technology, Eindhoven, The Netherlands.

$$x_l \triangleq \begin{bmatrix} a_{l,1} \\ a_{l,2} \\ \cdot \\ \cdot \\ a_{l,M} \end{bmatrix}. \tag{1}$$

With the input vector sequence we associate the vector D transform

$$x(D) \triangleq \sum_l x_l D^l \tag{2}$$

where D is the delay operator. In our investigations a linear, dispersive, and time-invariant multiple channel model is assumed (Fig. 1). This means that a linear relation exists between each input and each output signal and that the output signal due to the excitation of more than one input is the sum of the individual responses to the inputs in question. The relation between input j and output i is denoted by the impulse response $c_{ij}(t)$. All these responses are assumed to be square-integrable and of finite duration. Further we assume that the output signals are disturbed by MDI and additive, zero-mean, white Gaussian noise. Each output i is corrupted by a different noise signal $n_i(t)$, but it is assumed that these noise signals are uncorrelated and all have the same double-sided spectral density N_0. These assumptions are not a restriction of the generality as is shown in [3].

III. THE STATISTICAL SUFFICIENCY OF THE MMF OUTPUT

In this section we shall show that if the MMF, as defined in [3], is used as multiple linear receiving filter, then the sampled outputs of this MMF form a set of sufficient statistics for estimating the vector input sequence $x(D)$.

The impulse response $c_{ij}(t)$ is considered as an element of a matrix

$$C(t) \triangleq \begin{bmatrix} c_{11}(t) & c_{12}(t) & \cdot & \cdot & c_{1M}(t) \\ c_{21}(t) & c_{22}(t) & \cdot & \cdot & c_{2M}(t) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{M1}(t) & \cdot & \cdot & \cdot & c_{MM}(t) \end{bmatrix} \tag{3}$$

which defines the behavior of the multiple channel system. If the MMF is described in an analog way, it will be clear this its response is denoted by $C^T(-t)$. Assume that the multiple channel system is excited by an arbitrary, single-input vector x. Defining in this case the signal at output i of the multiple channel system by $s_i(t)$, we can write the total system output as a vector as follows:

$$s(t) \triangleq \begin{bmatrix} s_1(t) \\ s_2(t) \\ \cdot \\ \cdot \\ s_M(t) \end{bmatrix} \tag{4}$$

called the vector output signal. The noise is also given as a vector

Reprinted from *IEEE Transactions on Communications*, February 1976.

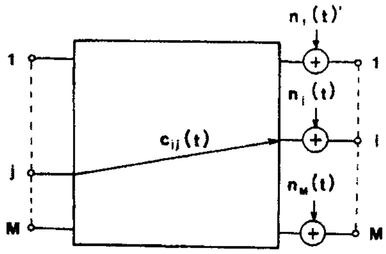

Fig. 1. Multiple channel communication model.

$$n(t) \triangleq \begin{bmatrix} n_1(t) \\ n_2(t) \\ \cdot \\ \cdot \\ \cdot \\ n_M(t) \end{bmatrix} \tag{5}$$

called the vector noise.

In the following we shall use several times the inner product of matrices, the components of which consist of time functions. Such a product is denoted as $\langle A(t),B(t)\rangle$ and defined by

$$\langle A(t),B(t)\rangle_{ij} \triangleq \sum_n \int_{-\infty}^{\infty} a_{in}(t)b_{nj}(t)\,dt. \tag{6}$$

The sampled output of the MMF, in the absence of noise, is given by the signal vector

$$\mathbf{s} = \langle C^T(t),s(t)\rangle \tag{7}$$

whereas the inverse transformation from signal vector to output vector signal is

$$s(t) = [C^T(t)]^T G\mathbf{s} = C(t)G\mathbf{s} \tag{8}$$

where G is a matrix to be determined. Substituting (8) in (7) gives

$$G = [\langle C^T(t), C(t)\rangle]^{-1}. \tag{9}$$

So the systems to be treated must have the property that the matrix G exists. This requirement, however, is quite trivial, because at systems not possessing this property it is impossible to recover even a single-input vector from the sampled MMF outputs. The sampled output noise, if the signal is absent, can be written as

$$\mathbf{n} = \langle C^T(t), n(t)\rangle. \tag{10}$$

According to (10) the relevant vector noise, being that part of the input vector noise that is left after projection of $n(t)$ at the signal space, is denoted by

$$n_r(t) = [C^T(t)]^T G\mathbf{n} = C(t)G\mathbf{n}. \tag{11}$$

By means of the definition

$$\mathbf{v} \triangleq \mathbf{s} + \mathbf{n}. \tag{12}$$

The equivalent received vector signal is written as

$$\nu(t) = C(t)G\mathbf{v} \tag{13}$$

which menas that for the sampled output there is no difference whether the true received vector signal $s(t) + n(t)$ or the vector signal $\nu(t)$ is supplied to the input of the MMF. Writing out (13) yields

$$\nu(t) = C(t)G\mathbf{v} = C(t)G\mathbf{s} + C(t)G\mathbf{n} = s(t) + n_r(t). \tag{14}$$

Thus $C(t)$ is a basis for the signal space spanned by both $s(t)$ and $n_r(t)$ [7, ch. 4], which proves that the sampled MMF output is a sufficient statistic for estimation of a single input vector x. Now the following theorem can be stated.

Theorem 1: If at each instant lT a vector x_l is transmitted, then the vector output sequence

$$\nu(D) \triangleq \sum_l \nu_l D^l \tag{15}$$

forms a set of sufficient statistics for estimating the vector input sequence $x(D)$ (see [4] and [7]).

IV. THE MULTIPLE WHITENED MATCHED FILTER

Now consider the system consisting of the channel in cascade with the MMF as a multiple channel system with M inputs and M outputs. The impulse response from input j to output n of this system is called $\nu_{nj}(t)$ and can be written as

$$\nu_{nj}(t) = \sum_{i=1}^{M} c_{in}(-t) * c_{ij}(t) = \sum_{i=1}^{M} \int_{-\infty}^{\infty} c_{in}(\tau - t)c_{ij}(\tau)\,d\tau \tag{16}$$

where $*$ means convolution. Define

$$V_l \triangleq \begin{bmatrix} \nu_{11}(lT) & \nu_{12}(lT) & \cdot & \cdot & \nu_{1M}(lT) \\ \nu_{21}(lT) & \nu_{22}(lT) & \cdot & \cdot & \nu_{2M}(lT) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \nu_{M1}(lT) & \nu_{M2}(lT) & \cdot & \cdot & \nu_{MM}(lT) \end{bmatrix} \tag{17}$$

and

$$V(D) \triangleq \sum_l V_l D^l. \tag{18}$$

By means of (16) it is easy to see that (18) is equivalent to

$$V(D) = \langle C^T(D^{-1},t), C(D,t)\rangle \tag{19}$$

where $C(D,t)$ is a matrix with components consisting of the chip D transforms [4] of the components of $C(t)$. The cross-correlation of the output noise signals at outputs n and m is given by

$$\begin{aligned} \phi_{nm}(\rho) &= \sum_{i=1}^{M} N_0 \int_{-\infty}^{\infty} c_{in}(-t)c_{im}(-t-\rho)\,dt \\ &= \sum_{i=1}^{M} N_0 \int_{-\infty}^{\infty} c_{in}(t)c_{im}(t-\rho)\,dt. \end{aligned} \tag{20}$$

Sampling this function we define its D transform as follows:

$$\phi_{nm}(D) \triangleq \sum_l \phi_{nm}(lT)D^l. \tag{21}$$

If all $\phi_{nm}(D)$ are collected in a matrix we get the spectral matrix

$$\Phi(D) = N_0\langle C^T(D,t), C(D^{-1},t)\rangle. \tag{22}$$

Relation (22) can easily be verified by means of (20). In [8] and [9] it is shown that a matrix $H(D^{-1})$ can be found such that

$$\Phi(D) = N_0 H(D) H^T(D^{-1}) \tag{23}$$

with both $H(D^{-1})$ and $H^{-1}(D^{-1})$ stable and nonanticipatory. Comparing (19) and (22) it is obvious that

$$V(D) = H(D^{-1})H^T(D). \tag{24}$$

Now we conclude that the sampled output of the MMF can be written as

$$\nu(D) = H(D^{-1})H^T(D)x(D) + H(D^{-1})n(D) \tag{25}$$

where $n(D)$ is the sampled input noise vector sequence. The output noise

$$n'(D) = H(D^{-1})n(D) \tag{26}$$

is colored Gaussian with spectral matrix $\Phi(D)$. This follows from

$$E[H(D)n(D^{-1})\{H(D^{-1})n(D)\}^T]$$
$$= E[H(D)n(D^{-1})n^T(D)H^T(D^{-1})] = N_0 H(D)H^T(D^{-1}). \tag{27}$$

From (25) it is seen that the output noise is whitened by the following operation:

$$z(D) \triangleq H^{-1}(D^{-1})\nu(D)$$
$$= H^T(D)x(D) + n(D) = y(D) + n(D) \tag{28}$$

which means physically that an MTDL [3] with transfer $H^{-1}(D^{-1})$ is placed after the MMF. It has been mentioned in the foregoing that $H^{-1}(D^{-1})$ is stable and nonanticipatory and thus realizable. The MMF followed by the MTDL is called multiple whitened matched filter and is characterized by its chip D transform

$$W(D,t) \triangleq H^{-1}(D^{-1})C^T(D^{-1},t). \tag{29}$$

If the impulse response from input n to output m is denoted by $w_{mn}(t)$, the set of functions $w_{mn}(t-kT)$ is orthonormal in both the time and space dimension as is seen from

$$\Phi_{ww}(D) = \langle W(D^{-1},t), W^T(D,t)\rangle$$
$$= H^{-1}(D)\langle C^T(D,t), C(D^{-1},t)\rangle\{H^{-1}(D^{-1})\}^T$$
$$= H^{-1}(D)V(D^{-1})\{H^T(D^{-1})\}^{-1}$$
$$= H^{-1}(D)H(D)H^T(D^{-1})\{H^T(D^{-1})\}^{-1} = I. \tag{30}$$

In the foregoing section we concluded that $\nu(D)$ forms a set of sufficient statistics for estimation $x(D)$, but $z(D)$ is found by the reversible linear transformation $H^{-1}(D^{-1})$ on $\nu(D)$. Thus $z(D)$ forms a set of sufficient statistics for estimating $x(D)$ also. This section is resumed in the following theorem.

Theorem 2: Let $C(t)$ be the matrix of impulse responses of the multiple channel transmission system and $H(D^{-1})H^T(D)$ a factorization of

$$V(D) = \langle C^T(D^{-1},t), C(D,t)\rangle \tag{31}$$

such that both $H(D^{-1})$ and $H^{-1}(D^{-1})$ are stable and nonanticipatory. Then the multiple filter whose chip D transform is

$$W(D,t) = H^{-1}(D^{-1})C^T(D,t) \tag{32}$$

is realizable and is called a multiple whitened matched filter and its sampled outputs give a vector sequence

$$z(D) = H^T(D)x(D) + n(D) \tag{33}$$

in which $n(D)$ is a white Gaussian noise vector sequence, and which is a set of sufficient statistics for estimation of the vector input sequence $x(D)$ where $n(D)$ white Gaussian is to be interpreted in both the discrete time and space dimension.

The multiple whitened matched filter found in this section is a generalized version of the whitened matched filter derived in [4]. This generalized filter is capable of optimizing the signal-to-noise ratio of the outputs of a multiple channel transmission system in which both ISI and ICI together with noise contribute to the disturbance, under the constraint that the output noise must be white in the two-dimensional sense.

V. THE VECTOR VITERBI ALGORITHM

In the preceding sections we have derived a structure giving a set of sufficient statistics for estimating the input vector sequence of a multiple channel transmission system from the observations of the output. This output is disturbed by MDI and noise. The noisy part of the multiple whitened matched filter output samples are shown to be uncorrelated and thus independent, since we have assumed that the noise is Gaussian. From this it follows that the Viterbi algorithm is a powerful tool to perform ML estimation of the input vector sequence $x(D)$. The vector Viterbi algorithm is a vector version of the algorithm used to make ML estimations on digital sequences and which is extensively described in [4] and [5]. The vector sequence $y(D)$ may be considered to be generated by a multiple finite state machine, driven by an input vector sequence $x(D)$ (see Fig. 2). As the state of this finite state machine we define

$$s_l \triangleq \{x_{l-1}, x_{l-2}, \cdots, x_{l-N}\} \tag{34}$$

where N is the degree of the matrix polynomial $H^T(D)$. The state s_l can take on L^{NM} different values. We can depict the successive states of the multiple finite state machine, together with all allowable transitions, in a trellis diagram. Each transition T_k in this trellis diagram is associated with an input vector x_{kl} and a certain value of the output signal y_{kl}. Given the observation z_l, the log likelihood of transition T_k is given by

$$\ln p(z_l - y_{kl}) = -\ln(\sqrt{2\pi N_0})^M - \frac{1}{2N_0}\sum_{i=1}^{M}(z_{l,i} - y_{kl,i})^2 \tag{35}$$

where $z_{l,i}$ and $y_{kl,i}$ are the ith components of, respectively, z_l and y_{kl}. In ML sequence estimation the first term of the right member of (35), being independent of l, can be omitted and the same holds for the factor $1/2N_0$ in the second term. The squared distance of an observation at instant lT to a certain allowable transition T_k, characterized by y_{kl}, is defined by

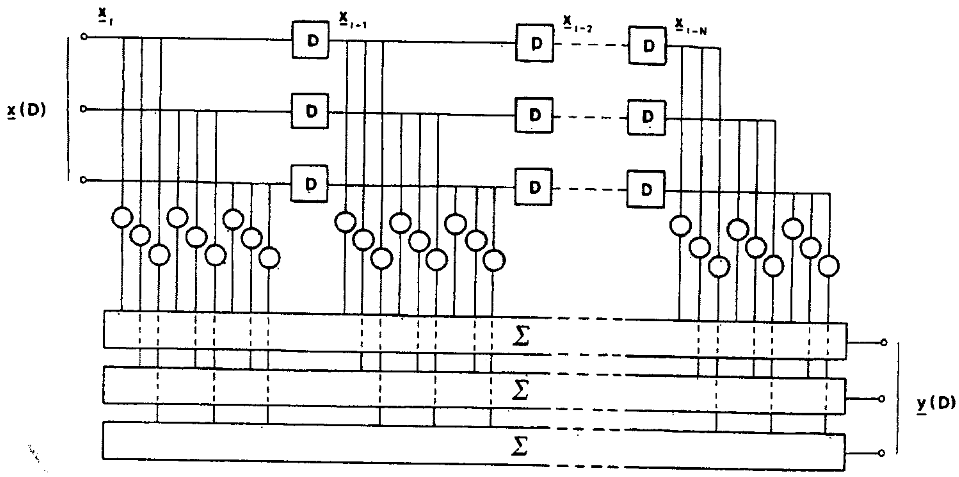

Fig. 2. Model of a multiple finite state machine.

$$D_{kl}^2 \triangleq \sum_{i=1}^{M} (z_{l,i} - y_{kl,i})^2. \tag{36}$$

The updating of the metrics, belonging to the several states, together with the updating of the corresponding path-registers proceeds as follows.

1) The metrics of all the states that belong to transitions, terminating in the same state, are increased with the corresponding squared distance D_{kl}^2.

2) Select the smallest increased metric as survivor-metric for the new state. The path-register of this new state is to be filled with the content of the path-register of the old state of the selected transition. This new path-register content is then updated with the elements of the input alphabet that belong to the selected transition.

The vector Viterbi algorithm does not differ fundamentally from the scalar version; the only differences, which are in fact generalizations, are as follows.

1) The operation "squared distance computation" is a computation in the vector sense defined by the Euclidean squared distance (36).

2) Each element of the path-register consists of M components and must be shifted and updated parallel in the path-register.

At this point the vector Viterbi algorithm is in fact reduced to the scalar version and we refer to [4] and [5] for further details. It will be clear that in a multiple channel transmission system the number of states is growing exponentially with the number of channels.

VI. THE VECTOR UNGERBOECK ALGORITHM

Ungerboeck has given an alternative recursive algorithm for making ML sequence estimations on data that are disturbed by ISI and white Gaussian noise [6]. Using this algorithm, the tapped delay line is omitted and the sampled output of the matched filter is directly used as input for the algorithm. In the following we shall generalize the Ungerboeck algorithm for ML vector sequence estimation of data that are disturbed by MDI and white Gaussian noise.

If a vector sequence $x(D)$ is transmitted the corresponding received vector signal is defined as follows:

$$u(t) \triangleq \sum_{l} C(t - lT)x_l + n(t). \tag{37}$$

Among all possible input sequences $\xi(D)$ we choose as estimate $\hat{x}(D)$ for $x(D)$ that vector sequence which maximizes $\ln p[u(t) \mid \xi(D)]$, which means minimizing over all allowable $\xi(D)$

$$\begin{aligned} J &= \left\| u(t) - \sum_{l} C(t - lT)\,\xi_l \right\|_2^2 \\ &= \left\langle \left[u(t) - \sum_{l} C(t - lT)\,\xi_l \right]^T, \right. \\ &\quad \left. \cdot \left[u(t) - \sum_{k} C(t - kT)\,\xi_k \right] \right\rangle. \end{aligned} \tag{38}$$

Writing out (38) the expression for J becomes

$$\begin{aligned} J &= \langle u^T(t), u(t) \rangle - \left\langle u^T(t), \sum_{k} C(t - kT)\,\xi_k \right\rangle \\ &\quad - \left\langle \sum_{l} \xi_l^T C^T(t - lT), u(t) \right\rangle \\ &\quad + \left\langle \sum_{l} \xi_l^T C^T(t - lT), \sum_{k} C(t - kT)\,\xi_k \right\rangle. \end{aligned} \tag{39}$$

Define

$$\nu_l \triangleq \langle C^T(t - lT), u(t) \rangle. \tag{40}$$

This vector is interpreted as the sampled output of the MMF. By means of definition (40) J is written as

$$\begin{aligned} J &= \langle u^T(t), u(t) \rangle \\ &\quad - 2 \sum_{l} \xi_l^T \nu_l + \sum_{l} \sum_{k} \xi_l^T V_{l-k}\, \xi_k. \end{aligned} \tag{41}$$

The first term of (41) is independent of ξ_l and thus may be ignored during the minimization process. The metric $J(\xi(D))$ can be calculated in a recursive manner.

$$J_l(\cdots, \xi_{l-1}, \xi_l) = J_{l-1}(\cdots, \xi_{l-1}) + F(\nu_l; \xi_{l-N}, \cdots, \xi_l) \tag{42}$$

with

$$F(\nu_l;\ \xi_{l-N}, \cdots, \xi_l) = \xi_l^T \left[2\nu_l - V_0\, \xi_l - 2 \sum_{k=1}^{N} V_k\, \xi_{l-k} \right] \tag{43}$$

where N is determined by the length of the $V(D)$ matrix sequence, according to

$$V(D) = \sum_{l=-N}^{N} V_l D^l. \tag{44}$$

Here the survivor-metric $\widetilde{J}_l$ is introduced, which is defined as follows:

$$\widetilde{J}_l(s_l) \triangleq \widetilde{J}_l\{\xi_{l+1-N}, \cdots, \xi_l\} \triangleq \min_{\{\cdots, \xi_{l-N}\}} \{J_l(\cdots, \xi_{l-N}, \xi_{l-N+1}, \cdots, \xi_l)\}. \tag{45}$$

The sequence $(\cdots, \xi_{l-N})$, which results in a minimum of (45) is called the path-history of the survivor-state

$$s_l \triangleq (\xi_{l+1-N}, \cdots, \xi_l). \tag{46}$$

It is easy to see that there are again L^{NM} different survivor-states. One can imagine that these survivor-states correspond to the states of a finite state machine. From this point of view the principles of the Ungerboeck algorithm coincide from now on those of the Viterbi algorithm. For further details see [6].

Expression (43) is now to be used for the calculation of the squared distance of an observation to an allowable transition, and the finite state machine has as much different states as the finite state machine of the Viterbi algorithm.

Although at first glance the metric calculation of the Ungerboeck algorithm seems more complicated than that of the Viterbi algorithm, a second inspection of (43) shows that the metric up-dating is a rather simple operation from a programming point of view. Namely the quantity

$$\xi_l^T \left[V_0\, \xi_l + 2 \sum_{k=1}^{N} V_k\, \xi_{l-k} \right]$$

only depends on the channel response, which is assumed to be fixed, and on the transitions to be considered. So this value can be stored in a memory and need not be calculated in real time.

VIII. THE ERROR PERFORMANCE OF THE ML RECEIVER

The investigations, given in this section, are closely related to the methods given in [4] and [6]. Remember that $x(D)$ represents the transmitted vector sequence and that the vector sequence estimated by the ML receiver is denoted by $\hat{x}(D)$. Then

$$e(D) \triangleq \hat{x}(D) - x(D) \tag{47}$$

defines the error vector sequence. Assuming stationarity, the starting point of an error event ϵ can be associated with $t = 0$:

$$\epsilon\colon e(D) = e_0 + e_1 D + \cdots + e_H D^H$$

$$\text{with } \| e_i \|_2 \geqslant \delta_0, \tag{48}$$

where δ_0 denotes the minimum nonzero value of the Euclidean norm of the error vector e_i. This value equals the minimum distance

$$\delta_0 = \min_{i \neq j} \{\| a_{l,j} - a_{l,i} \|\}. \tag{49}$$

From [6] we know that the error event probability is written as

$$\Pr(\epsilon) = \Pr(\epsilon_1)\Pr(\epsilon_2 \mid \epsilon_1) \leqslant \Pr(\epsilon_1)\Pr(\epsilon_2' \mid \epsilon_1) \tag{50}$$

where the subevents ϵ_1, ϵ_2, and ϵ_2' are defined as follows.

- ϵ_1 $x(D)$ is such that $x(D) + e(D)$ is an allowable data vector sequence;
- ϵ_2 noise vector sequence is such that $x(D) + e(D)$ has ML (within the observation interval); and
- ϵ_2' noise vector sequence is such that $x(D) + e(D)$ has greater likelihood than $x(D)$, but not necessarily ML.

From the preceding section it is concluded that $\Pr(\epsilon_2' \mid \epsilon_1)$ is the probability that

$$J(x(D)) > J(x(D) + e(D)). \tag{51}$$

It can be proven that inequality (51) is identical with

$$\delta^2(\epsilon) \triangleq \| V_0^{-1} \|_2 \sum_{l=0}^{H} \sum_{k=0}^{H} e_l^T V_{l-k} e_k < 2 \| V_0^{-1} \|_2 \sum_{l=0}^{H} e_l^T n_l' \tag{52}$$

where n_l' are the sample values of the noise at the output of the MMF. The quantity $\delta(\epsilon)$ is called the distance of the error event ϵ. Consider the random variable α given by the right member of (52).

$$\alpha \triangleq 2 \| V_0^{-1} \|_2 \sum_{l=0}^{H} e_l^T n_l'. \tag{53}$$

This random variable is Gaussian distributed with zero-mean and variance

$$E[\alpha^2] = 4N_0 \| V_0^{-1} \|_2\, \delta^2(\epsilon). \tag{54}$$

From this it follows that

$$\Pr(\epsilon_2' \mid \epsilon_1) = \Pr(\alpha > \delta^2(\epsilon)) = Q\left(\frac{\delta(\epsilon)}{2\sqrt{N_0}\, \| V_0^{-1} \|_2^{1/2}} \right) \tag{55}$$

where the well-known $Q(\cdot)$ function is defined in [7].

Let E be the set of all possible error events ϵ. Then the probability that any error event occurs becomes

$$\Pr(E) = \sum_{\epsilon \in E} \Pr(\epsilon). \tag{56}$$

Let Δ be the set of all possible $\delta(\epsilon)$ and E_δ the subset of error

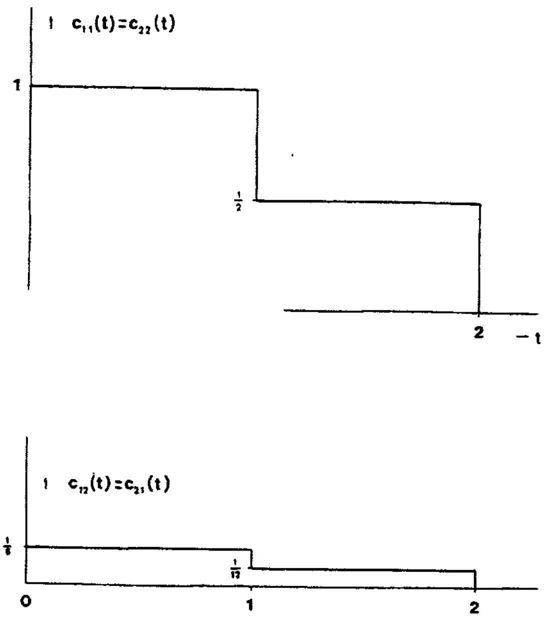

Fig. 3. Received signal set for the example.

events for which $\delta(\epsilon) = \delta$. Then from (50) the error event probability is bounded by

$$\Pr(E) \leq \sum_{\delta \in \Delta} Q\left(\frac{\delta}{2\sqrt{N_0}\, \| V_0^{-1} \|_2^{1/2}}\right) \sum_{\epsilon \in E_\delta} \Pr(\epsilon_1). \tag{57}$$

Because of the exponential behavior of the $Q(\cdot)$ function for large argument values, this expression will already at moderate signal-to-noise ratios be dominated by the term involving the minimum value $\delta_{\min}$ out of the set Δ. At moderate and large signal-to-noise ratios ϵ_2' implies ϵ_2 with a probability almost equal to one. For these SNR values $\Pr(E)$ is approximated by

$$\Pr(E) \simeq Q\left(\frac{\delta_{\min}}{2\sqrt{N_0}\, \| V_0^{-1} \|_2^{1/2}}\right) \sum_{\epsilon \in E_{\delta_{\min}}} \Pr(\epsilon_1). \tag{58}$$

Assuming the input symbols $a_{l,j}$ to be independent of each other and equiprobable, the probability of ϵ_1 is written as

$$\Pr(\epsilon_1) = \prod_{i=1}^{M} \prod_{l=0}^{H} \frac{L - |e_{l,i}|}{L} \tag{59}$$

with $e_{l,i}$ the ith component of e_l.

In the Appendix it is shown that under the constraint

$$\| V_0^{-1} \|_2 \sum_{l=-\infty}^{\infty}{}' \, \| V_l \|_2 \leq 1 \tag{60}$$

not any error event has smaller distance than the single error events with distance δ_0. With a single error event we mean an error sequence that consists of one error vector ($e(D) = e_0$) and from this vector only one component differs from zero. In this situation the single error events with distance δ_0 dominate the expression for the error event probability and the error event probability equals the symbol error probability

$$\Pr(e) \simeq Q\left(\frac{\delta_0}{2\sqrt{N_0}\, \| V_0^{-1} \|_2^{1/2}}\right) \sum_{\epsilon \in E_{\delta_0}} \frac{L-1}{L}. \tag{61}$$

Since $\delta_0^2 / \| E_0^{-1} \|_2$ is the total amount of energy that is measured as the receiving end at transmission of a single symbol out of the set E_{δ_0}, the symbol error probability is not increased by MDI.

VIII. AN EXAMPLE

As an example we take a multiple channel with $M = 2$. The components of the transmission matrix $C(t)$ are as given in Fig. 3. We take $T = 1$ and for this system the $V(D)$ matrix polynomial is as follows:

$$\left.\begin{aligned} V_0 &= \frac{5}{144}\begin{bmatrix} 37 & 12 \\ 12 & 37 \end{bmatrix} \\ V_1 = V_{-1} &= \frac{1}{72}\begin{bmatrix} 37 & 12 \\ 12 & 37 \end{bmatrix} \end{aligned}\right\}. \tag{62}$$

One can easily verify that this $V(D)$ satisfies condition (60). Decomposition of $V(D)$ according to (24) yields

$$H^T(D) = \frac{1}{12}\begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}(2 + D). \tag{63}$$

By means of this matrix sequence the given system is simulated on a minicomputer. In Fig. 4 the error probability for a binary alphabet $\{+1, -1\}$ is plotted as function of the signal-to-noise ratio, together with the $\Pr(e)$ for isolated pulses. The two curves merge at a $\Pr(e)$ of about 10^{-4}. So, for error probabilities smaller than 10^{-4}, the performance of the ML receiver is as good as if MDI were absent. In the case of larger error probabilities the difference between the two curves is

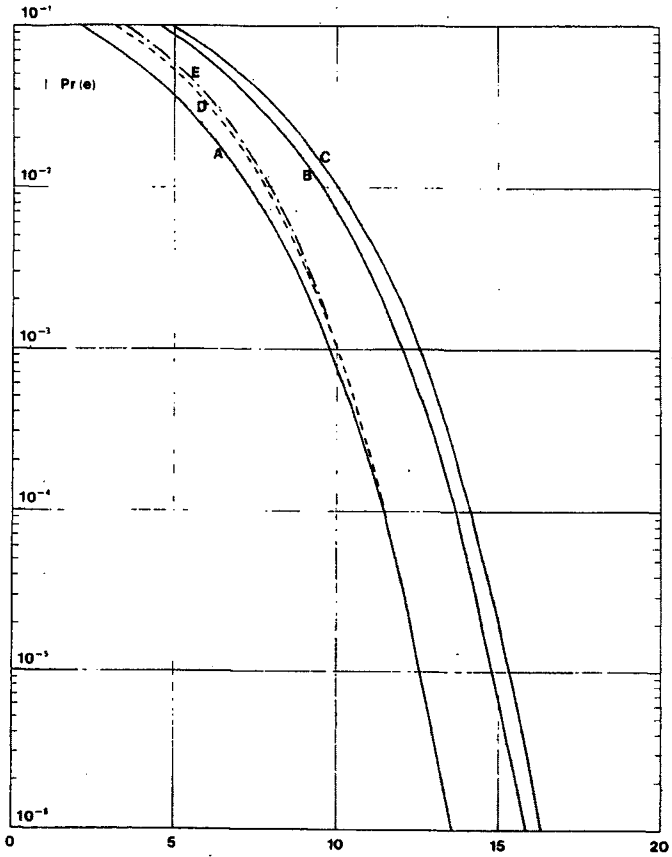

Fig. 4. Symbol error probability versus signal-to-noise ratio.

Curve *A* Single pulse;
Curve *B* monochannel with linear correction and bit-by-bit detection;
Curve *C* multiple channel with $M = 2$, linear correction and bit-by-bit detection;
Curve *D* monochannel with ML sequence estimation; and
Curve *E* multiple channel with $M = 2$ and ML vector sequence estimation.

maximal 1.2 dB. These results are compared with those of an optimum constrained linear receiver [3]. The difference between the linear receiver and the single-pulse performance is 2.7 dB, showing the superiority of the vector ML receiver. We also simulated a ML receiver for a monochannel with impulse response $c_{11}(t)$. Now the maximum difference with the single-pulse performance appears to be 1 dB, whereas the two curves also merge at a Pr (e) value of about 10^{-4}. Linear correction with bit-by-bit detection gives an increase of 2.2 dB in this case.

N.B.: At the simulations the path-register length was 16 bits in all cases. The number of transmissions was chosen such that the real error probability lies, with a probability 0.9, within an interval of 10 percent around the plotted value.

IX. SUMMARY AND CONCLUSIONS

It is shown that the MMF outputs form a set of sufficient statistics for estimating the transmitted vector sequence over a multiple channel system. A multiple whitened matched filter is derived, the output of which is used to perform ML vector sequence estimation by means of the vector version of the Viterbi algorithm. A modified algorithm, pointed out by Ungerboeck, is also generalized to combat the noise and MDI. If this algorithm is used MTDL is omitted and the sampled output of the MMF is directly used as input data for the algorithm. Finally, the error performance of the ML receiver for a multiple channel system disturbed by noise and MDI is calculated. From the latter investigations it follows that, under a certain constraint, for moderate and large SNR's the error performance is not substantially influenced by MDI, i.e., the symbol error probability is approximated by the value found if a single pulse is transmitted. It is concluded from this concise paper that ICI plays the same role as ISI. If these two disturbances are simultaneously considered, then MDI can, under the given constraints, be treated as a generalization of ISI and the concepts of ML sequence estimation on data disturbed by noise and ISI are also generalized for noise and MDI.

APPENDIX

Let

$$\|e_0\|_2 \geqslant \delta_0 \tag{64}$$

and let $\{V_l\}_{l=-\infty}^{\infty}$ be given and assume

$$\|V_0^{-1}\|_2 \sum_{l=-\infty}^{\infty}{}' \|V_l\|_2 \leqslant 1. \tag{65}$$

The matrix V_0 equals $\langle C^T(t), C(t)\rangle$ and it is easy to show that this matrix is positive definite, under the condition derived in Section III.

$$\begin{aligned}\delta^2(\epsilon) &= \| V_0^{-1} \|_2 \sum_{l=-H}^{H} \sum_{k=0}^{H} e_{l+k}{}^T V_l e_k \\ &= \| V_0^{-1} \|_2 \sum_{k=0}^{H} e_k{}^T V_0 e_k \\ &\quad + \| V_0^{-1} \|_2 \sum_{l=-H}^{H}{}' \sum_{k=0}^{H} e_{l+k}{}^T V_l e_k. \end{aligned} \tag{66}$$

Consider the first term of (66). Because V_0 is positive definite we have the inequality

$$e_k{}^T V_0 e_k \geqslant \lambda_{\min}(V_0) e_k{}^T e_k \tag{67}$$

where $\lambda_{\min}(V_0)$ is the smallest eigenvalue of V_0. Moreover,

$$\| V_0^{-1} \|_2 = \frac{1}{\lambda_{\min}(V_0)}. \tag{68}$$

From (67) and (68) it follows

$$\begin{aligned}\| V_0^{-1} \|_2 \sum_{k=0}^{H} e_k{}^T V_0 e_k &\geqslant \sum_{k=0}^{H} e_k{}^T e_k \\ &= \sum_{k=0}^{H} \| e_k \|_2{}^2. \end{aligned} \tag{69}$$

Consider now the second term of (66). Due to the Schwarz inequality and from what is given we have

$$\begin{aligned}&\left| \| V_0^{-1} \|_2 \sum_{l=-H}^{H}{}' \sum_{k=0}^{H} e_{l+k}{}^T V_l e_k \right| \\ &\leqslant \| V_0^{-1} \|_2 \sum_{l=-H}^{H}{}' \| V_l \|_2 \sum_{k=0}^{H} \| e_{l+k}{}^T \|_2 \cdot \| e_k \|_2 \\ &\leqslant \left\{ \| V_0^{-1} \|_2 \sum_{l=-H}^{H}{}' \| V_l \|_2 \right\} \left\{ \sum_{k=0}^{H} \| e_k \|_2{}^2 - \delta_0{}^2 \right\}. \end{aligned} \tag{70}$$

From (69) and (70) it follows

$$\begin{aligned}\delta^2(\epsilon) &\geqslant \left(\sum_{k=0}^{H} \| e_k \|_2{}^2 - \delta_0{}^2 \right) \\ &\quad - \left\{ \| V_0^{-1} \|_2 \sum_{l=-H}^{H}{}' \| V_l \|_2 \right\} \\ &\quad \cdot \left\{ \sum_{k=0}^{H} \| e_k \|_2{}^2 - \delta_0{}^2 \right\} + \delta_0{}^2 \\ &\geqslant \delta_0{}^2. \end{aligned} \tag{71}$$

This last inequality holds if (65) is satisfied.

ACKNOWLEDGMENT

The author would like to thank Prof. J. van der Plaats and Prof. J. P. M. Schalkwijk for the stimulating discussions on the subject. He also wants to acknowledge L. S. de Jong for giving the proof of the Appendix.

REFERENCES

[1] D. A. Shnidman, "A generalized Nyquist criterion and optimum linear receiver for a pulse modulation system," *Bell Syst. Tech. J.*, vol. 46, pp. 2163–2177, Nov. 1967.

[2] A. R. Kaye and D. A. George, "Transmission of multiplexed PAM signals over multiple channel and diversity systems," *IEEE Trans. Commun. Technol.*, vol. COM-18, pp. 520–526, Oct. 1970.

[3] W. van Etten, "An optimum linear receiver for multiple channel digital transmission systems," *IEEE Trans. Commun.* (Concise Papers), vol. COM-23, pp. 828–834, Aug. 1975.

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

[5] J. K. Omura, "Optimal receiver design for convolutional codes and channels with memory via control theoretical concepts," *Inform. Sci.*, vol. 3, pp. 243–266, 1971.

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

[7] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

[8] D. N. Prabhakar Murthy, "Factorization of discrete-process spectral matrices," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-19, pp. 693–696, Sept. 1973.

[9] P. R. Motyka and J. A. Cadzow, "The factorization of discrete-process spectral matrices," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 698–707, Dec. 1967.

[10] F. B. Hildebrand, *Methods of Applied Mathematics*. Englewood Cliffs, NJ: Prentice-Hall, 1952.