Multistream faster than Nyquist signaling

Rusek, Fredrik; Anderson, John B

Published in:
IEEE Transactions on Communications

DOI:
10.1109/TCOMM.2009.05.070224

2009

Multistream Faster than Nyquist Signaling

Fredrik Rusek and John B. Anderson
Electrical and Information Technology Dept. and
Swedish Strategic Center for High Speed Wireless Communication
Lund University, Box 118, 221 00 Lund, Sweden

***Abstract*— We extend Mazo's concept of faster-than-Nyquist (FTN) signaling to pulse trains that modulate a bank of subcarriers, a method called two dimensional FTN signaling. The signal processing is similar to orthogonal frequency division multiplex (OFDM) transmission but the subchannels are not orthogonal. Despite nonorthogonal pulses and subcarriers, the method achieves the isolated-pulse error performance; it does so in as little as half the bandwidth of ordinary OFDM. Euclidean distance properties are investigated for schemes based on several basic pulses. The best have Gaussian shape. An efficient distance calculation is given. Concatenations of ordinary codes and FTN are introduced. The combination achieves the outer code gain in as little as half the bandwidth. Receivers must work in two dimensions, and several iterative designs are proposed for FTN with outer convolutional coding.**

I. INTRODUCTION

Consider baseband signals of the form

$$s(t) = \sqrt{2E_s/T}\sum_n a_n h(t-nT), \tag{1}$$

in which a_n are data values over an M-ary alphabet and $h(t)$ is a unit-energy baseband pulse. This simple form underlies QAM, TCM, and the subcarriers in orthogonal frequency division multiplex (OFDM), as well as many other transmission systems. In these schemes, $h(t)$ is a T-orthogonal pulse, meaning that the correlation $\int h(t-nT)h^*(t-mT)dt$ is zero, $m \neq n$. In 1975 Mazo [1] pointed out that binary $\mathrm{sinc}(t/T)$ pulses in (1) could be sent every T_Δ seconds, $T_\Delta < T$, without loss in asymptotic error probability. This he called faster than Nyquist (FTN) signaling, because the pulses appear faster than allowed by Nyquist's limit for orthogonal pulses. FTN signaling has since been generalized in a number of ways.

This paper extends the FTN concept to the frequency dimension. This extension to two dimensions opens up a number of attractive possibilities. Many signals of type (1) are stacked in frequency through modulation by a set of carriers at frequencies $f_0 + \{f_k\}$ to form the in-phase and quadrature (I/Q) signal given by the real part of

$$s(t) = \sqrt{\frac{2E_s}{T}}\sum_{k=0}^{K-1}\sum_{n=0}^{N-1}[a^I_{k,n}+\jmath a^Q_{k,n}]h(t-nT)\, e^{\jmath 2\pi(f_0+f_k)t}. \tag{2}$$

This is a superposition of $2K$ linear carrier modulations, and it carries $2NK$ data values. The $K\times N$ matrix $\boldsymbol{A} = \{a_{k,n}\}$ is called the data matrix and consists of the complex data values $a_{k,n} = a^I_{k,n} + \jmath a^Q_{k,n}$. The K rows in this matrix correspond to subcarriers, and the N columns to pulse positions. If $f_k = kf_\Delta$, $k = 0, 1, \dots K-1$ and f_Δ is equal twice the single-sided bandwidth of $h(t)$, the $2K$ carrier signals are orthogonal; if $h(t)$ is T-orthogonal, all NK pulses are mutually orthogonal. In OFDM, h is ordinarily the width-T square pulse, the subcarriers are $1/T$-spaced frequency orthogonal, and all pulses are again mutually orthogonal.

The signal design in the sinc and OFDM cases is thus based on *orthogonality*. According to theory, there exist about $2W\mathfrak{T}$ orthogonal signals in W positive Hertz and $\mathfrak{T}$ seconds. Data values that modulate the amplitude of each can be maximum-likelihood detected independently, and therefore about $2W\mathfrak{T}$ symbols can be transmitted this way. Take for example $h(t) = \sqrt{1/T}\mathrm{sinc}(t/T)$, and measure time and bandwidth of (2) by some reasonable method (such as 99% power bandwidth). Then as K and N grow, $\mathfrak{T}/N \to T$, $W = Kf_\Delta = K/T$, and the product $2W\mathfrak{T}$ tends in ratio to $2(K/T)(NT) = 2KN$; thus Eq. (2) carries as many data values for large NK as any scheme based on orthogonality can carry. A similar outcome occurs when h is a root RC pulse. For a given number of symbols carried by (2), T may be varied, which trades off W against $\mathfrak{T}$. N may also be traded against K. The time–bandwidth product is unaffected, and (2) always carries about twice $W\mathfrak{T}$ symbols.

If the aim is to achieve the *error rate* of a stacked orthogonal-signal system (2), without necessarily using orthogonal signals, the story is more complex, and more can be achieved. By error rate is meant the error probability of the maximum likelihood sequence estimation (MLSE) receiver when h is employed in (1) with additive white Gaussian noise (AWGN) of density $N_0/2$ in the channel. As the signal-to-noise ratio E_b/N_0 grows, the probability of incorrect detection of an a_n is asymptotically $P_e \sim Q(\sqrt{d^2_{\min}E_b/N_0})$, where $d_{\min}$ is the minimum distance of the set of signals and $d_{\min} \le d_{\mathrm{MF}}$. Here $E_b = E_s/\log_2 M$, E_s is the average symbol energy, and d_{MF} is the matched-filter bound distance for the data alphabet. d_{MF} measures the performance of simple orthogonal-pulse signaling with the same data values. The paper will concentrate on the binary case, for which $d^2_{\mathrm{MF}} = 2$, so the target orthogonal-pulse error rate is $Q(\sqrt{2E_b/N_0})$. If the K I/Q signals in (2) do not overlap in frequency, the same asymptotic error rate applies there.

Achieving more at the same E_b and error probability means that FTN signals need to consume less bandwidth. We need to define it carefully. With independent and identically distributed (IID) data symbols the power spectral density (PSD) of the kth subcarrier $S_k(f)$ is proportional to $|H(f-kf_\Delta-f_0)|^2+|H(f+kf_\Delta+f_0)|^2$, $k = 0, \dots, K-1$, where $H(f)$ is the Fourier transform of $h(t)$. With K subcarriers, the total PSD satisfies (take positive f only)

$$\begin{aligned} S(f) &\propto \sum_{k=0}^{K-1} S_k(f) \\ &= \sum_{k=0}^{K-1} |H(f-kf_\Delta-f_0)|^2, \quad f_0 \gg Kf_\Delta. \end{aligned} \tag{3}$$

The *normalized time–bandwidth product* (NTB) of this transmission is

$$\mathrm{NTB} \triangleq \frac{WT_\triangle}{2RK} \qquad \text{Hz-s/bit}, \tag{4}$$

where W is a measure of the positive frequency bandwidth of the entire Eq. (3), $T_\triangle \leq T$ is the actual symbol time, and R is the bits carried by each subcarrier symbol. Each subcarrier carries $2R/T_\triangle$ bit/s, counting I and Q. For the single-subcarrier Eq. (1), W is the positive baseband bandwidth of just $H(f)$. The NTB measures time–bandwidth on a per data bit basis.[1] Simple time scaling of $s(t)$ does not affect the NTB, since the spectrum is then scaled by the inverse factor.

As $T_\triangle$ drops for the same h, the FTN signal becomes "faster", and its NTB drops. In the baseband single-subcarrier case, the total PSD is just $|H(f)|^2$ and W measures its width, and as $T_\triangle$ drops, the symbol-normalized spectrum and the NTB are scaled down. In this view the pulses come faster. It is equivalent to fix T and scale the pulse h wider by the factor $T/T_\triangle$; this reduces the PSD and the NTB narrows in an identical way. In either case, the time–bandwidth per data bit is less. With K subcarriers the calculation is more complicated but the outcome is the same.

Mazo's paper[1] envisioned one subcarrier and binary $\mathrm{sinc}(\ )$ pulses and claimed the surprising result that $d^2_{\min}$ is in fact $d^2_{\mathrm{MF}} = 2$ for $T_\triangle > .802T$. No asymptotic error rate is lost by increasing the symbol rate 24.7% above the Nyquist limit, yet the NTB has dropped from 1 Hz-s/bit to 0.802. A full MLSE detection is required in principle, which compares all N-symbol signals to the full noisy received signal. The reason for this behavior can be seen by analyzing the error events that can occur (as in Section II). As $T_\triangle$ declines and the pulse rate grows, another error event eventually has a distance less than the $d^2 = 2$ antipodal event that leads to d_{MF}. But this does not occur immediately.

Later research has shown that a similar phenomenon occurs with other orthogonal $h(t)$ than the sinc pulse [2], [3]. For the 30% root RC pulse for example, $T_\triangle$ can be as small as $0.703T$. There is a least $T_\triangle$ as well for pulses such as the Gaussian, that are not orthogonal for any T. Moreover, such a limit appears with nonbinary transmission, with precoding, and with linear coded modulation based on heavy filtering [3], [4]. All these cases can be summed up as follows: The error performance of the linear transmission of form (1) remains unchanged under downward scaling of the normalized spectrum shape until a surprisingly narrow bandwidth, after which it suddenly drops. This occurs despite escalating ISI. We call this threshold bandwidth the *Mazo limit*. Its significance is that it is pointless to transmit in a wider bandwidth in a linear channel with AWGN, if sufficient receiver processing is available. We will see in this paper that a Mazo phenomenon applies as well to concatenations of traditional coding with FTN modulators and to multicarrier FTN.

We introduced the idea of multistream FTN (MFTN) signaling in [5]. It is useful to think of it as two-dimensional FTN signaling because the symbols can be associated with points in a lattice spaced every $f_\triangle$ and $T_\triangle$. This is illustrated in Fig. 1. Pulses are "hung" on each point. The signaling is not two-dimensional in the sense of coding over the magnetic domains on a multitrack tape, although receivers have some similarity. Reference [5] gave examples that simultaneous frequency and time squeezing can indeed increase the symbols transmitted in a given time–bandwidth at the same P_e. Neither compression alone can achieve the same increase.

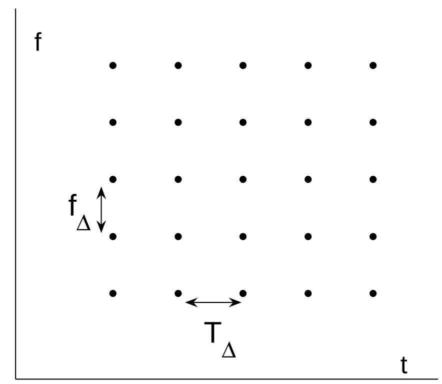

Fig. 1. Two dimensional Mazo signaling, in time and frequency. Dots represent symbols separated by $f_\triangle$ and $T_\triangle$.

What total time–bandwidth product is occupied by the transmission? The question may be approached in several ways. Spectral and temporal sidelobes interfere with adjacent users of frequency and time, and contribute to the signal's occupancy. For a moderate NK product, a packet of say 100–10000 bits, the sidelobes make a significant contribution. In a companion paper [6] we treat this contribution, seek to minimize it, and find rather different outcomes than given here. In this paper we let N and K grow large, so that the sidelobes are insignificant. The lattice area is about $NKf_\triangle T_\triangle$ Hz-s and the NTB tends to $f_\triangle T_\triangle$ Hz-s/bit. The ratio of N and K can be changed at will so long as they have the same product. The orthogonal binary $\mathrm{sinc}(t/T)$ pulse case has $f_\triangle = 1/T$ and $T_\triangle = T$, and the NTB is $f_\triangle T_\triangle = 1$ Hz-s/bit. This provides a useful benchmark for other pulses and systems. Since the value of T does not change the NTB, we henceforth take $T = 1$.

When $f_\triangle$ is less than the subcarrier bandwidth, the signal interrelations that produce $d_{\min}$ work in new ways and the distance structure is time varying. Analytical results are available only in special cases [15], [16]. Finding $d_{\min}$ in this new situation is challenging but possible. The subject of minimum distance is taken up in Section II. Distance studies with various pulses $h(t)$ show that $d^2_{\min} = 2$ can occur at 0.5 Hz-s/bit, i.e., half the sinc benchmark. The section also gives a number of properties. Euclidean distance computations appear in Appendix 1. An advanced algorithm to find minimum distance appears in the Appendix 2.

Section III proposes several MFTN receiver designs and introduces concatenation of convolutional codes with MFTN. An iterated receiver works particularly well here, and these concatenations form a successful practical application of MFTN.

[1]Others have called the NTB the normalized bandwidth or the bit normalized bandwidth; it could as well be called the bandwidth normalized bit time. We prefer the neutral term time–bandwidth product.

II. Signals, Properties and Minimum Distance

The real part of Eq. (2) is

$$\sqrt{2E_s/T}\,[I(t)\cos 2\pi f_0 t - Q(t)\sin 2\pi f_0 t]$$

where the in-phase and quadrature signals $I(t)$ and $Q(t)$ are

$$\begin{aligned} I(t) &= \sum_{k=0}^{K-1}\sum_{n=0}^{N-1}\Big[a^I_{k,n}h(t-nT)\cos 2\pi f_k t \\ &\qquad -a^Q_{k,n}h(t-nT)\sin 2\pi f_k t\Big] \\ Q(t) &= \sum_{k=0}^{K-1}\sum_{n=0}^{N-1}\Big[a^Q_{k,n}h(t-nT)\cos 2\pi f_k t \\ &\qquad +a^I_{k,n}h(t-nT)\sin 2\pi f_k t\Big]. \end{aligned} \tag{5}$$

The signals $\cos 2\pi f_k t$ and $\sin 2\pi f_k t$ are the *subcarriers*. They exist only mathematically; the physical modulation carriers are $\cos 2\pi(f_0+f_k)t$ and $\sin 2\pi(f_0+f_k)t$, for $k = 0,\dots,K-1$. Figure 2 shows example 2-carrier signals with $T_\Delta = .8$ and $f_\Delta = .625$, in which $a^I_{0,0} = a^Q_{0,5} = 1$ for carrier 0 and $a^I_{1,3} = a^Q_{1,8} = 1$ for carrier 1, with all other symbols set to zero. The carriers lie at f_0 and $f_0 + .625$ Hz. Arrows show the location of the 1-symbols, and I_0, I_1 and Q_0, Q_1 are the respective carriers' I and Q signals. The subcarrier $\cos 2\pi f_1 t$ is shown dashed for reference. Note that I_1 symbols have an effect on Q_1 and vice versa.

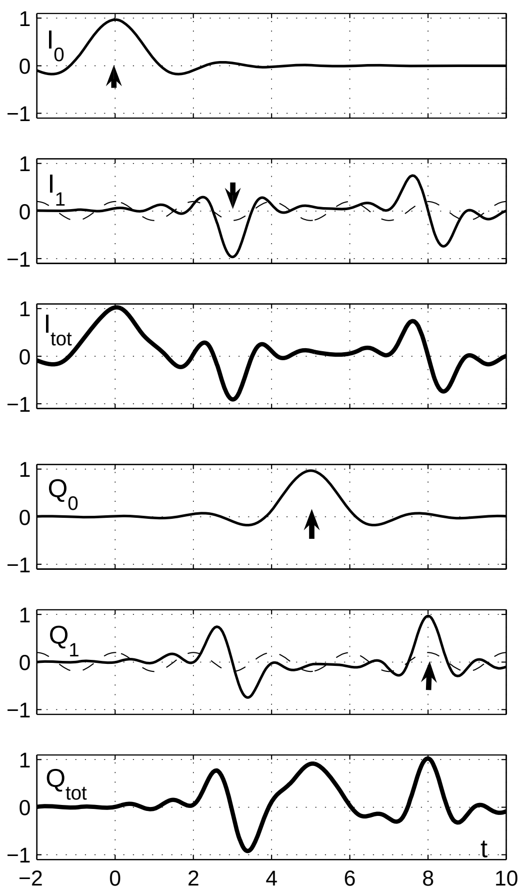

Fig. 2. Upper three: Component and total in-phase signals for 2-carrier MFTN example with $f_\Delta = .8$ and $T_\Delta = .625$. Bottom three: Component and total quadrature signals. Time scale in multiples of T_Δ.

The normalized Euclidean distance between two signals $s^{(1)}(t)$ and $s^{(2)}(t)$ of form (2) is

$$d^2 = \frac{\int |s^{(1)}(t) - s^{(2)}(t)|^2 dt}{2E_b}. \tag{6}$$

Because of the summations in (2), only the difference between the symbol streams matters and d^2 becomes

$$\frac{1}{2E_b}\int\left|\sum_{k=0}^{K-1}\sum_{n=0}^{N-1}\Delta a_{k,n}h(t-nT)\,e^{\jmath 2\pi(f_0+f_k)t}\right|^2 dt. \tag{7}$$

Here $\Delta a_{k,n}$ are the elements of the complex symbol difference matrix

$$\boldsymbol{\Delta A} = \begin{bmatrix} \Delta a_{0,0} & \Delta a_{0,1} & \dots & \Delta a_{0,N-1} \\ \dots & & \dots & \\ \Delta a_{K-1,0} & \Delta a_{K-1,1} & \dots & \Delta a_{K-1,N-1} \end{bmatrix}.$$

Equation (7) can as well be written

$$\int |\Delta I(t) - \Delta Q(t)|^2 dt \tag{8}$$

as $f_0 \to \infty$, in which $\Delta I(t)$ and $\Delta Q(t)$ are as in (5) with $\Delta\boldsymbol{A}$ instead of $\boldsymbol{A}$. For binary signaling the elements in $\Delta\boldsymbol{A}$ take values in $\{\pm 2 \pm 2\jmath, \pm 2, \pm 2\jmath, 0\}$.

An *error event* is a region of nonzero difference components that begins at some position (n,k), which we may as well take as $(0,0)$. Without loss of generality, we can restrict $\Delta a_{0,0}$ to $\{2, 2+2\jmath\}$. The minimum distance $d_{\min}$ of a signal set is the minimum of (7) over all such events, and an event leading to $d_{\min}$ is called a critical event. Since there are very many error events, finding $d_{\min}$ is difficult, but it is possible to find reliable estimates (i.e., tight overbounds) by searching over limited sets of events. In this paper the critical error events are typically of size 3×3 or smaller. Distance may be computed by direct integration of the difference signal, but a much more efficient method is based on autocorrelations of $h(t)$; this is given in Appendix 1.

An important property of MFTN distance is that it depends on the start time of the event. That is, for a given error event $\Delta\boldsymbol{A}$ whose first column corresponds to pulses centered at $t = 0$, the distance will change if the event starts with pulses centered at $t = nT_\Delta$. The fundamental reason is that the pulse rate is not in general synchronized with the subcarrier frequencies; it is mathematically seen in the derivations in Appendix 1. An example of the phenomena appears in Figure 3 for a 30% root RC pulse, $T_\Delta = .7$, $f_\Delta = .8$, and the event

$$\Delta\boldsymbol{A} = \begin{bmatrix} 2 & -2+2\jmath & -2\jmath \\ -2 & 2+2\jmath & -2\jmath \end{bmatrix}. \tag{9}$$

The figure plots square distance against the event start time t_0. Since the plot repeats every $1/f_\Delta = 1.25$ s, the time axis can be taken as $t_0 \mod 1/f_\Delta$. Dots show this event's distance when it starts at times $t_0 = 0, .8, 1.6, \dots$, which lead mod 1.25 to all the multiples of .05. If $1/f_\Delta T_\Delta$ is not a rational number, then in principle starts at all t_0 mod $1/f_\Delta$ in $[0, 1/f_\Delta)$ are possible, and the worst case distance is the minimum, which is 1.13 at start $t_0 \mod 1/f_\Delta \approx .24$.

Synchronous MFTN. When finitely many modulo start times can occur, the multistream FTN is *synchronous*. It is easy to see that when $f_\Delta T_\Delta = i/\ell$, i and ℓ positive integers without a common factor, then the MFTN signals are synchronous and the distances occur at multiples of $1/\ell$.

The worst-case distance is the minimum of the allowed time points. Synchronism can thus lead to a better distance, but in reality there is little gain unless i and ℓ are very small integers. In the figure, with 25 points, there is virtually no distance gain. But the MFTN case shown in Figure 2 is a synchronous one with $f_\Delta T_\Delta = 1/2$; the subcarrier structure repeats precisely every $2T_\Delta$. With error event (9), the distance from start time $t_0 = T_\Delta$ or $2T_\Delta$ are both 8.72. These avoid the worst-case start time, which is 0.25, with the much poorer distance 0.77. (However, there exist other error events with distance less than 2, so this MFTN is beyond the Mazo limit). Synchronism can thus be of benefit when the MFTN parameters allow it, but only a few cases of synchronous FTN are worth reporting. We will ordinarily take distance to be the minimum of the continuous distance versus t_0 curve.

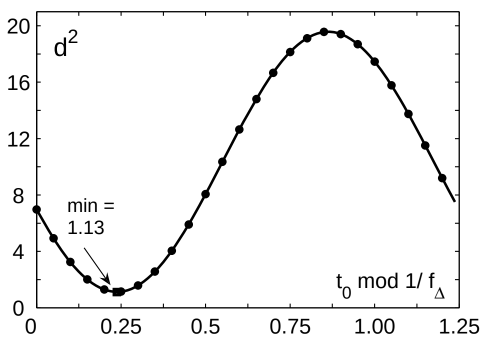

Fig. 3. Distance versus event start time t_0 modulo $1/f_\Delta$, for $f_\Delta = .8$ and $T_\Delta = .7$ with error event (9).

When not more than two carriers overlap in spectrum, the curve in Figure 3 can be shown to be a sinusoid, and its minimum can be computed easily from any three points. Another simplifying property is that very few error events lead to a distance near $d_{\min}$. As well, groups of symmetric events, typically 4–8 in number, have identical distance for the entire range of T_Δ and f_Δ. These we call event *families*. The results presented in this paper in fact stem from only 20 families.

Delayed Pulses. Subjecting subcarriers to different delays has the potential to improve minimum distance because pulses can come into more favorable alignments with each other and with the sines and cosines in the I and Q signals in (5). Many ways to execute the delays can be imagined. A simple but effective way is to delay the I and Q pulses $h(t-nT_\Delta)$ in carriers $0, 1, \ldots, K-1$ by $\delta_0, \delta_1, \ldots, \delta_{K-1}$, where each δ satisfies $0 \le \delta < T_\Delta$. Delaying all K carrier pulses by the same δ is the same as a time shift to the subcarrier system; this can prevent a subcarrier in (2) from passing through zero at the moment when a pulse is largest, which can severely reduce distance. Pulse trains $k = 0, 1, \ldots$ can be delayed by linearly growing amounts, e.g., $\delta_0, \delta_1, \ldots = (0, .2, .4, \ldots)T_\Delta$. A particularly successful scheme is delays of the form $.5, 0, .5, 0, \ldots$ times T_Δ. Instead of pulse delays, schemes can be based on delaying the subcarriers by fractions of their own period $1/f_\Delta$.

The Mazo Limit for Root RC and Gaussian Pulses. Figures 4 and 5 plot the outcome of a search for non-synchronous combinations of f_Δ and T_Δ that have least product. Dotted lines show contours of constant $f_\Delta T_\Delta$ product. Figure 4 plots the case for the Gaussian pulse $h(t) = \sqrt{1/2\pi\sigma^2}\exp(-t^2/2\sigma^2)$, normalized, with $\sigma^2 = .399$.[2] The minimum distance searches here are over all start times in the error events. Overlapping curves show the trajectory for each critical event family; the "northeast"-most of all curves is the Mazo limit. Consider the trajectory for one critical error sequence, with distance d^2. As T_Δ drops, the f_Δ needed to maintain $d^2 = 2$ rises, creating, typically, a convex-up f_Δ, T_Δ relationship. Eventually, time compression alone prevents $d^2 = 2$; no f_Δ allows it, and the result is a horizontal section at the lower right end of the convex section (this is most visible in Fig. 5). At the upper left of a convex section, T_Δ is large and it can happen that *no* f_Δ leads to $d^2 < 2$. The section simply stops at some $f_\Delta^\square, T_\Delta^\square$ (square blocks mark two such points in Fig. 4). If this section is part of the ultimate Mazo limit, then at $T_\Delta = T_\Delta^\square$ the Mazo limit must move horizontally left to another event's convex section.

Note that $h(t)$ here is not orthogonal for any T. The Gauss pulse has important properties when simultaneous time and frequency side lobes are important [6].

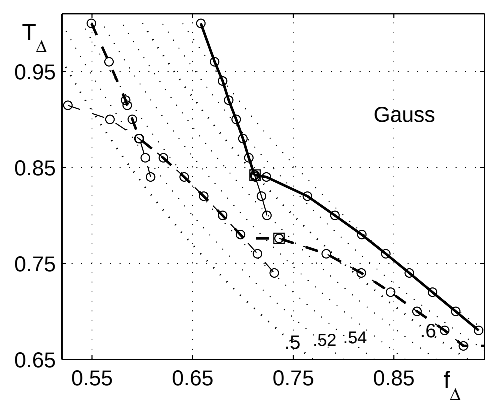

Fig. 4. Estimated position of the two-dimensional Mazo limit for non-synchronous binary Gaussian pulse signaling. Dashed line shows limit with alternate pulse trains delayed 0.5 symbol. Each of the five component curves here represents an event family. Dotted curves are contours of constant $f_\Delta T_\Delta$.

Figure 5 plots the non-synchronous Mazo limit for 10 and 30% root RC pulses, plus the 10% case when alternate pulse trains are delayed by $T_\Delta/2$ (dashed). The searches are again over all start times in the error events. It can be seen that the least product for 30% is about 0.60, at $(f_\Delta \approx .67,\ T_\Delta \approx .88)$; for 10% pulses this improves to product 0.556 at $(f_\Delta = .660,\ T_\Delta = .843)$. The delays improve the 10% case to 0.534 at $(f_\Delta \approx .66, T_\Delta \approx .80)$; the 30% pulse is similarly improved by delays. These products are excellent but we have found a few synchronous 10% cases, with T_Δ in the range 0.78–0.9, for which $f_\Delta T_\Delta = 1/2$. This is a doubling of the spectral efficiency of the sinc benchmark and OFDM.

The estimated minimum distance of all combinations is 2. Searches are generally performed with the method in Appendix 2 and are over error events out to size 4×7

[2]This is the Gaussian pulse that has itself as transform. A new pulse begins each T_Δ as always.

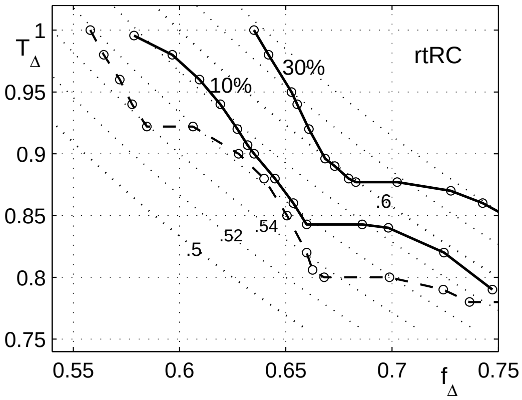

Fig. 5. Estimated position of the two-dimensional Mazo limit for non-synchronous binary root RC signaling with excess bandwidth 10 and 30%. Dashed line shows 10% with alternate pulse trains delayed 0.5 symbol. Dotted curves are contours of constant $f_\Delta T_\Delta$.

and 7×4 carriers $\times$ pulses.[3] The critical events that lead to the points in the figures are much smaller, almost entirely of size 3×3 or 2×4, 2×3 or the reverses. We can thus have some confidence that the minimum distance estimate is tight. Each critical event leads to a section of the Mazo limit. Sketching the limit consists of searching over hundreds of millions of events at some (f_Δ, T_Δ), finding a critical event for it, drawing a (f_Δ, T_Δ) section that stems from the event, and then repeating the process.

III. Detection of Convolutionally Encoded Systems

Detection of this type of signaling is a difficult task. A trellis description of the system consists of 4^{KL} states, where L is the support of the model of the FTN-induced intersymbol interference (ISI). Clearly, reduced complexity methods have to be used, and in order to avoid simply trading the bandwidth reduction for a higher symbol energy, the receiver must achieve essentially the full MLSE performance. Reduced complexity detection methods appear in [8], [9], [10], [17], but these are all operating too far from MLSE performance to fully exploit the bandwidth inmprovement. In [5] a simple M-algorithm is considered, but it turns out that the method only works for 2–4 subcarriers. In [14] a more sophisticated iterative detector is proposed. The detector's BER at high signal to noise ratio closely follows the basic reference $P_e \sim Q(\sqrt{d^2_{\min} E_b/N_0})$. Thus the distance computations from Section II are verified. The complexity, however, is still significant; the detector consists of two steps, each involving 6–8 iterations. In brief, detection is possible, but with high complexity.

If an outer code is concatenated via an interleaver to the MFTN system, detection is much simpler and a more practical MFTN system is the outcome. We will now focus on this case and take the outer code as a rate 1/2 feedforward convolutional code; such codes at low memory have 4–5 dB coding gain. Just as an uncoded system from Section II was able to preserve its BER down to a certain critical compression product $f_\Delta T_\Delta$, a concatenated system can preserve the coding gain of the outer code down to another product. It will turn out that this critical product is in fact smaller than the one for an uncoded system. Thus, the potential of MFTN *increases* for concatenated systems, and this is a major reason to focus on them. Concatenated coding with a two dimensional ISI channel has been investigated e.g. in [11], [18], [19], but not with MFTN-induced ISI.

The system model consists of the sequence: Convolutional Code $\rightarrow$ Interleaver $\rightarrow$ Binary to K Stream Mapper $\rightarrow$ MFTN Modulator. The memory of the convolutional code is ν. A sequence of $10000-\nu$ IID information bits are first convolutionally encoded; this produces a codeword $\boldsymbol{v}$ of length 20000, which is randomly interleaved to produce $\boldsymbol{v}'$. Mapping from v_l', v_{l+1}' to one symbol $a_{k,n} \in \{\pm 1 \pm \jmath\}$ follows, with k and n found from l by some predefined pattern. The pattern is not of interest because of the interleaver, and will not be discussed further. The transmitted signal $s(t)$ is formed according to (2) from the symbol sequence $\boldsymbol{a}$; we will use 20 subcarriers, i.e. $K = 20$ and thus $N = 500$. We have also performed tests with fewer subcarriers. When the number increases, the BER generally degrades. But at 5–8 subcarriers a saturation of the BER occurs, that is, the BER for $K = 10$ is virtually identical to the BER for $K = 20$ presented here. We are therefore confident that the BER does not change if many subcarriers are used, e.g. $K = 64$ or 256.

Due to the interleaver, straightforward iterative detection is possible, and a block diagram is given in Figure 6. Two soft-input soft-output detectors are needed, one for the convolutional encoder and one for the MFTN signaling system, which together with the mapper is considered as an inner encoder. Hereafter, the detector for the MFTN system is referred to as a detector and the outer detector as a decoder. A standard full BCJR algorithm will be used as decoder for the convolutional code, but the MFTN detector is not standard. If optimum detection is desired, the 4^{LK} state complexity still comes about, and here also a reduced complexity method is needed. The difference from the uncoded case is that the output from the reduced complexity detector there is the final result, while it feeds an iterative process in concatenated systems. The outer and the inner detectors jointly work their way to a near-optimal result; the fact that the inner detector is not full complexity more or less only slows down the convergence speed. The ultimate compression product limits, however, will be somewhat worse due to the reduced complexity.

We now turn to the structure of the MFTN detector. The baseband representation of (2) is

$$s_{\mathrm{bb}}(t) = \sum_n \sum_k a_{k,n} h(t - nT_\Delta) e^{\jmath 2\pi t f_k}, \qquad (10)$$

The detector encounters a noisy signal $r(t) = s_{\mathrm{bb}}(t) + n(t)$, where $n(t)$ is complex-valued baseband white Gaussian noise. The first step in the receiver is to project $r(t)$ onto the basis functions $h(t - nT_\Delta)e^{\jmath 2\pi t f_k}$, i.e. to compute

$$R_{k,n} = \int_{-\infty}^{\infty} r(t) h^*(t - nT_\Delta) e^{-\jmath 2\pi t f_k} dt. \qquad (11)$$

The elements $R_{k,n}$ form the matrix $\boldsymbol{R}$, which can in practice be efficiently implemented by a bank of matched filters with rate-$1/T_\Delta$ sampling. The matrix $\boldsymbol{R}$ comprises a sufficient statistic for estimating $\{a_{k,n}\}$, and in the sequel $\boldsymbol{R}$ is simply

[3] The method in Appendix 2 was verified where possible with a brute force method. Gauss searches require the brute force search because more than two carriers contribute to the spectrum; 10–15 start times were used.

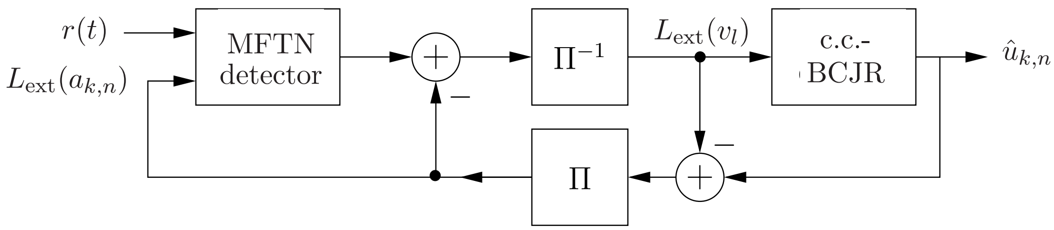

Fig. 6. Block diagram for iterative detection of encoded MFTN. $r(t)$ is the received signal and 'c.c.-BCJR' denotes a BCJR for the convolutional code.

called the received signal. We may write $\boldsymbol{R} = \boldsymbol{S} + \boldsymbol{N}$. The sent part $\boldsymbol{S}$ of $\boldsymbol{R}$ equals

$$S_{k,n} = \int \sum_{m,l} a_{l,m} h(t-mT_\Delta) e^{\jmath 2\pi t(l-k)f_\Delta} h^*(t-nT_\Delta)\, dt. \tag{12}$$

Further manipulations of (12) are made in Appendix 1. The noise matrix $\boldsymbol{N}$ is given by

$$N_{k,n} = \int_{-\infty}^{\infty} n(t) h^*(t - nT_\Delta) e^{-\jmath 2\pi t f_k} dt. \tag{13}$$

The variables $N_{k,n}$ are not white.

The goal of the detector is to maximize the *a posteriori* probability (APP) of an individual bit, i.e.

$$\hat{a}_{k,n}^{I/Q} = \arg\max_{a \in \{-1,1\}} \Pr\left(\hat{a}_{k,n}^{I/Q} = a \,|\, \boldsymbol{R}\right). \tag{14}$$

(Superscript I/Q means "I respectively Q"; when this superscript is omitted we intend a complex a). Instead of working with probabilities it is convenient to work with log-likelihood ratios (LLRs)

$$L\left(a_{k,n}^{I/Q}\right) = \log \frac{\Pr\left\{a_{k,n}^{I/Q} = 1\right\}}{\Pr\left\{a_{k,n}^{I/Q} = -1\right\}}. \tag{15}$$

From $L(a_{k,n}^{I/Q})$ it is straightforward to find the LLRs $L(v_l)$. Since the data symbols are independent we can as usual express the conditional LLR $L(a_{k,n}^{I/Q}|\boldsymbol{R})$ as

$$L\left(a_{k,n}^{I/Q} \,|\, \boldsymbol{R}\right) = L_{\text{ext}}\left(a_{k,n}^{I/Q} \,|\, \boldsymbol{R}\right) + L\left(a_{k,n}^{I/Q}\right) \tag{16}$$

where $L_{\text{ext}}(a_{k,n}^{I/Q}|\boldsymbol{R})$ denotes the extrinsic information about $a_{k,n}^{I/Q}$ contained in $\boldsymbol{R}$.

The true APPs of the data bits can be found by a multidimensional BCJR algorithm, but its complexity grows exponentially with K, and the APPs have to be approximated by simpler means. The detector will thus consider only a fraction of the symbols at once; the rest will act as noise. The symbols $\{a_{k,n}\}$ are grouped into two sets, $\boldsymbol{A}_{\text{dec}}$ and $\boldsymbol{A}_{\text{int}}$; symbols in $\boldsymbol{A}_{\text{dec}}$ are the ones that we try to decode at the moment and those in $\boldsymbol{A}_{\text{int}}$ act as noise. The transmitted baseband signal $s_{\text{bb}}(t)$ can be expressed as $s_{\text{bb}}(t) = s_{\text{dec}}(t) + s_{\text{int}}(t)$, where $s_{\text{dec}}(t)$ and $s_{\text{int}}(t)$ are the contributions from symbols in $\boldsymbol{A}_{\text{dec}}$ and $\boldsymbol{A}_{\text{int}}$, respectively. The detector will be based on successive interference cancellation [13]; when decoding the signal $s_{\text{dec}}(t)$ a soft estimate $\hat{s}_{\text{int}}(t)$ of $s_{\text{int}}(t)$ is formed based on soft information about all symbols in $\boldsymbol{A}_{\text{int}}$:

$$\hat{s}_{\text{int}}(t) = \sum_{\mathcal{B}} b_{k,n}^{I/Q} h(t - nT_\Delta)\, e^{\jmath 2\pi t f_k}, \tag{17}$$

where $\mathcal{B} = \{(k,n) : a_{k,n} \in \boldsymbol{A}_{\text{int}}\}$ and $b_{k,n}^{I/Q}$ are the soft estimates of $a_{k,n}^{I/Q}$ defined by

$$\begin{aligned} b_{k,n}^{I/Q} &= \Pr\left\{a_{k,n}^{I/Q} = 1\right\} - \Pr\left\{a_{k,n}^{I/Q} = -1\right\} \\ &= \tanh\left[\frac{1}{2} L_{\text{ext}}\left(a_{k,n}^{I/Q} \,|\, \boldsymbol{R}\right)\right]. \end{aligned}$$

Then $\hat{s}_{\text{int}}(t)$ is projected onto the basis functions, with the projection denoted $\hat{\boldsymbol{S}}(\boldsymbol{A}_{\text{int}})$. Finally, the tentative received signal when detecting symbols in $\boldsymbol{A}_{\text{dec}}$ is formed as

$$\hat{\boldsymbol{R}}(\boldsymbol{A}_{\text{dec}}) = \boldsymbol{R} - \hat{\boldsymbol{S}}(\boldsymbol{A}_{\text{int}}). \tag{18}$$

Together with the extrinsic information $L_{\text{ext}}(a_{k,n}^{I/Q}|\hat{\boldsymbol{R}}(\boldsymbol{A}_{\text{dec}}))$ about the symbols $a_{k,n}^{I/Q}$ in $\boldsymbol{A}_{\text{dec}}$, the signal $\hat{\boldsymbol{R}}(\boldsymbol{A}_{\text{dec}})$ is fed to some detection algorithm.

If $\boldsymbol{A}_{\text{dec}} = \{a_{k,n}, \forall n, k\}$ and $\boldsymbol{A}_{\text{int}} = \emptyset$ there is no complexity reduction at all. We will only consider the partition $\boldsymbol{A}_{\text{dec}} = \{a_{k',n}, \forall n\}$; this is all symbols on carrier k'. For this partition, the signal $s_{\text{dec}}(t)$ can be viewed as a single carrier signal, based on a real-valued pulse shape $h(t)$. It then follows that its real- and complex-valued parts can be detected independently; it is thus a matter of detecting binary symbols through an ISI channel in AWGN. If the detector takes L ISI taps as being significant, the complexity of a full BCJR is 2^L. In this paper the full BCJR will be used, with $L \leq 5$, and thus there are at most 32 states in the BCJR detector. One technical detail remains. The classical BCJR requires white noise $\boldsymbol{N}$, which is not the case here. Since $h(t)$ has long duration, in theory infinite, whitening of the outputs from the sampled matched filters is a difficult task. However, an equivalent algorithm to the BCJR that operates directly on the samples of the matched filter output has been derived in [12]. This algorithm assumes colored noise and is not a true BCJR, but it has the same output as a true BCJR acting on a whitened version of $\boldsymbol{R}$, and we still refer to it as a BCJR.

The innermost part of Figure 6 can now be drawn. It is shown for subcarrier k' in Figure 7. Some words on the partition $\boldsymbol{A}_{\text{dec}} = \{a_{k',n}, \forall n\}$ are needed. If $T_\Delta = T$, there is no ISI, and the BCJR in Figure 7 becomes meaningless, since there is no dependence along each subchannel in the signal $\hat{\boldsymbol{R}}(\boldsymbol{A}_{\text{dec}})$. If T_Δ is close to 1, but $f_\Delta \ll 1$, there is some ISI, but the ICI is much stronger. In that case it makes more sense to use the partition $\boldsymbol{A}_{\text{dec}} = \{a_{k,n'}, \forall k\}$, that is, to attack the ICI only. For f_Δ, T_Δ of roughly the same size, hybrid methods can be used; some iterations can use $\boldsymbol{A}_{\text{dec}} = \{a_{k',n}, \forall n\}$ and some can use $\boldsymbol{A}_{\text{dec}} = \{a_{k,n'}, \forall k\}$. Brevity prevents pursuing this idea. The partition $\boldsymbol{A}_{\text{dec}} = \{a_{k',n}, \forall n\}$ turns out to work quite well in this paper.

Let us turn to actual receiver tests. The outer code was taken as the (7,5) convolutional code, and $h(t)$ was a 30%

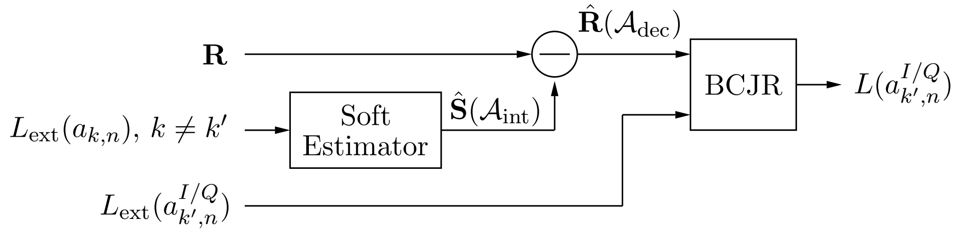

Fig. 7. MFTN detector for subcarrier k'.

root RC pulse. The number of receiver iterations is limited to 7. In each receiver test, 1000 blocks of 10000 information bits have been detected.

In Figure 8, three parameter combinations are tested, $f_\triangle = 1.174, T_\triangle = .46$ (product .54), $f_\triangle = .5682, T_\triangle = .88$ (product .5) and $f_\triangle = .75, T_\triangle = .60$, (product .45). Also shown is the performance of the convolutional code over an ISI-free channel. As seen from the figure, at high E_b/N_0 the encoded MFTN signaling systems are in fact able to retain the full coding gain of the outer code, but at a much narrower bandwidth. A similar phenomenon is often observed in turbo equalization. If the compression is too heavy, causing too severe ISI/ICI, the system is bounded away from the outer code performance; but eventually it will reach it if $T_\triangle$ is large enough. From the figure it is seen

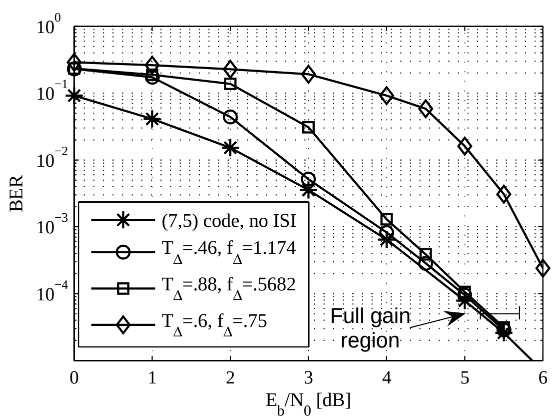

Fig. 8. Receiver tests of (7,5) encoded MFTN.

that the system with the smallest product, $f_\triangle T_\triangle = .45$, did not converge to the outer code performance. This may be a result of badly chosen parameters $T_\triangle$ and $f_\triangle$ rather than a too small product $f_\triangle T_\triangle$.

In order to see which systems will converge at practical E_b/N_0, there exists a strong tool, Extrinsic Information Transfer (EXIT) charts. Although their validity is open for discussion and they are not perfect in the case of finite block lengths and non-ideal interleavers, they quickly provide insight into the iterative convergence mechanism. In our case, the noise is in fact not Gaussian because of the interference from the non-ideal $\hat{\boldsymbol{S}}(\boldsymbol{A}_{\text{int}})$, and this will slightly degrade the EXIT charts. We have *not* used EXIT charts when establishing which combinations of $f_\triangle$ and $T_\triangle$ result in the full coding gain; these are based solely on receiver tests.

We will say *full coding gain* is achieved if the encoded MFTN system requires less than 0.5 dB more power then the outer code alone to achieve BER 5×10^{-5}. The definition is illustrated in Figure 8; encoded MFTN systems that pass through the "full gain region" achieve full coding gain.

It is now straightforward to exhaust the parameter combinations and find out where the full coding gain is present. In Figure 9 the combinations of $f_\triangle$ and $T_\triangle$ are shown that result in the full coding gain for the convolutional codes (7,5) and (74,54), with $\times$ denoting (7,5) and $\circ$ (74,54). The figure only applies to the block size and iterations stated above; more iterations can improve the results somewhat. An exhaustive search of $f_\triangle$ and $T_\triangle$ is of course not tractable. We have tested $f_\triangle$ in steps of $.01$. If the full coding gain is present for some $T_\triangle = x$ and $f_\triangle = y$ but not at $T_\triangle = x$ and $f_\triangle = y - .01$, the point (x, y) is plotted in Figure 9. Dotted lines show constant $f_\triangle T_\triangle$ products. The smallest product with full coding gain occurs for both codes at $T_\triangle = .79$, with $f_\triangle T_\triangle \approx .43$ and .455 respectively for the (7,5) and (74,54) codes. There is thus $\approx$ 55% bandwidth reduction without loss of BER.

Receiver Complexity. In the iterative detection process we allowed 7 iterations only. The decoder for the outer code is a full complexity BCJR; for the (7,5) and the (74,54) codes this results in 4 and 8 states respectively. The state complexity of the MFTN detector was limited to 32, although this can be reduced for $T_\triangle$ close to 1. In total, one four-state BCJR and one 32 state BCJR need to be applied 7 times.

A particular combination of $f_\triangle$ and $T_\triangle$ is worth mention. When $T_\triangle = 1$ it can be seen that the smallest product $f_\triangle T_\triangle$ such that the full gain is present is only sligthly worse than optimal. From a decoding point of view, $T_\triangle = 1$ is a very good choice as there will be no ISI to defeat. This implies that for the MFTN detector above there is no memory and the detection is only symbol by symbol. The complexity is only to apply the BCJR to the outer code a number of times. For the (7,5) code there are only 4 states, only 7 iterations are required, and the overall complexity is thereby very small.

There is in fact nothing that prohibits $T_\triangle$ from being larger than 1. As seen in Figure 9, this leads to reasonable, but not superior, performance. There will be ISI to detect in this case.

Concluding Observations. We make three observations. First, Figure 9 applies to a certain type of turbo receiver. The $f_\triangle T_\triangle$ products may be smaller with an MLSE receiver or with a receiver that attacks the ICI and ISI in a different pattern. Second, the spectral efficiency (bit rate/Hz) is independent of the basic pulse $h(t)$; it depends solely on $f_\triangle$, $T_\triangle$ and the rate of the outer code. This makes the pulse $h(t)$ a free optimization parameter. It should be chosen such that full coding gain can be reached at the

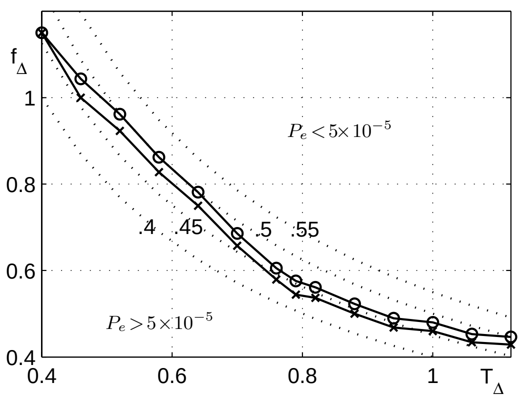

Fig. 9. Mazo limit $f_\triangle T_\triangle$ for coded MFTM. Lines marked $\times$ and $\circ$ are the (7,5) and (74,54) convolutional codes. Full coding gain at $P_e = 5 \times 10^{-5}$ is present above the lines, but not below.

smallest product $f_\triangle T_\triangle$. How to perform this optimization is an open problem. Finally, by observing signals, one can see that MFTN signals, like most bandwidth-compressed transmissions, have a higher peak to average power ratio than Nyquist signals. How this ratio compares to that of other methods should be studied in the future.

IV. Conclusions

We have demonstrated that the idea of faster than Nyquist time compression can be applied at the same time across frequency carriers, to achieve transmission throughput up to twice that of OFDM-like signaling at the same energy and error rate. We have sketched the Mazo limit for multistream Gauss and root RC pulses, that is, the least time–frequency compression products that yield the antipodal signal error rate. Both synchronism and pulse delays can improve this product. Finally, we have investigated concatenations of convolutional codes and MFTN. An iterated receiver continues to achieve the convolutional coding/orthogonal modulation error performance even under strong compression. Multistream faster than Nyquist techniques thus show great promise as practical, bandwidth saving tranmission methods.

Acknowledgments

Parts of this work were supported by the Swedish Science Board through grant 621-2003-3210, and by the Swedish Foundation for Strategic Research through its Center for High Speed Wireless Communication at Lund University.

Appendix 1

In this appendix we derive the Euclidean distance of an MFTN error event set; the derivation assumes an arbitrary starting time of the error event and an arbitrary delay pattern. We start with notation. Let $\tau_r(\ )$ be the r-step operator; for matrices it is defined as

$$\tau_r(\boldsymbol{\Delta A}) = \begin{bmatrix} 0\dots0 & \Delta a_{0,0} & \dots & \Delta a_{0,N-1} \\ \dots & & \dots & \\ \underbrace{0\dots0}_{r} & \Delta a_{K-1,0} & \dots & \Delta a_{K-1,N-1} \end{bmatrix}.$$

Then $d^2(\tau_r(\boldsymbol{\Delta A}))$ denotes the normalized distance of the error event $\boldsymbol{\Delta A}$ delayed r steps. Assume also that error event $\boldsymbol{\Delta A}$ starts at time index 0; i.e., for some k, $\Delta a_{k,0} \neq 0$. Let $d^2(\boldsymbol{\Delta A})\big|_{\mathcal{F}}$ denote the portion of distance contained in the frequency range $f \in \mathcal{F}$. We can compute $d^2(\tau_r(\boldsymbol{\Delta A}))$ by (19)-(21).

The last expression in (19) is the distance of $\boldsymbol{\Delta A}$ at time–frequency offset $\phi = f_\triangle T_\triangle \gamma \mod 2\pi$; this distance we write as

$$d^2(\boldsymbol{\Delta A};\phi) = \sum_{k=0}^{K-1}\sum_{l=0}^{K-1} e^{\jmath 2\pi\phi(l-k)}\beta_{k,l}(\boldsymbol{\Delta A}). \tag{22}$$

We can compute $d^2(\boldsymbol{\Delta}\mathbf{A})\big|_{\mathcal{F}}$ by replacing (21) with

$$\gamma_{k,l}[m,n]\big|_{\mathcal{F}} = \int_{\mathcal{F}} H(f)H(f-f_\triangle(l-k))\, e^{\jmath 2\pi(n+\delta_l)T_\triangle f} \times \\ \times e^{-\jmath 2\pi(m+\delta_k)T_\triangle(f-f_\triangle(l-k))}\, df. \tag{23}$$

To find the worst case ϕ we take the derivative of $d^2(\boldsymbol{\Delta A};\phi)$,

$$\frac{\partial\, d^2(\boldsymbol{\Delta A};\phi)}{\partial\phi} = \sum_{k=0}^{K-1}\sum_{l=0}^{K-1} \jmath 2\pi(l-k) e^{\jmath 2\pi\phi(l-k)}\beta_{k,l}(\boldsymbol{\Delta A}). \tag{24}$$

For $f_\triangle \geq W$ we have $\gamma_{k,l}[m,n] = 0$ for $|l-k| > 1$; this implies that $\beta_{k,l}(\boldsymbol{\Delta A}) = 0$ for $|l-k| > 1$ as well. Setting the derivative equal to zero we get

$$\sum_{k=1}^{K-1} e^{\jmath 2\pi\phi}\beta_{k,k-1}(\boldsymbol{\Delta A}) = \sum_{k=0}^{K-1} e^{-\jmath 2\pi\phi}\beta_{k,k+1}(\boldsymbol{\Delta A}). \tag{25}$$

Equation (25) is the second order equation

$$(e^{\jmath 2\pi\phi})^2 \sum_{k=1}^{K-1}\beta_{k,k-1}(\boldsymbol{\Delta A}) = \sum_{k=0}^{K-1}\beta_{k,k+1}(\boldsymbol{\Delta A}) \tag{26}$$

which can be analytically solved (the case $f_\triangle < W$ gives a fourth order equation and is omitted here). The two solutions ϕ_1, ϕ_2 to (26) are

$$\{2\pi\phi_1, 2\pi\phi_2\} = \arg\left\{\pm\sqrt{\frac{\sum_{k=0}^{K-1}\beta_{k,k+1}(\boldsymbol{\Delta A})}{\sum_{k=1}^{K}\beta_{k,k-1}(\boldsymbol{\Delta A})}}\right\}, \tag{27}$$

where $\arg\{\ \}$ denotes the angle in radians of a complex number. Whether ϕ_1 is the minimum or the maximum can be determined from the sign of $\sum_{k=0}^{K-1}\beta_{k,k+1}(\boldsymbol{\Delta A})$. This method of determining the worst time offset of an error event is only valid for irrational products $f_\triangle T_\triangle$. For rational products we can find the worst point by finding the two closest allowed points to the minimizing $\phi_{\min}$.

Appendix 2

The derivations in Appendix 1 find the worst time offset ϕ for a *given* error event $\boldsymbol{\Delta A}$. Computing $d^2_{\min}$ requires in principle an efficient search over *all* error events and that is the aim of this appendix. Assume that the search is limited to events of size $k_0 \times n_0$ symbols. We give now an efficient search for this case. The algorithm assumes a pulse $h(t)$ perfectly bandlimited to W positive Hz and is stated here only for the case $f_\triangle \geq W$ (if this is not the case, the search can still be used by approximating the bandwidth). Moreover, when we formulate the algorithm, we omit the time dependency of the distance function for notational

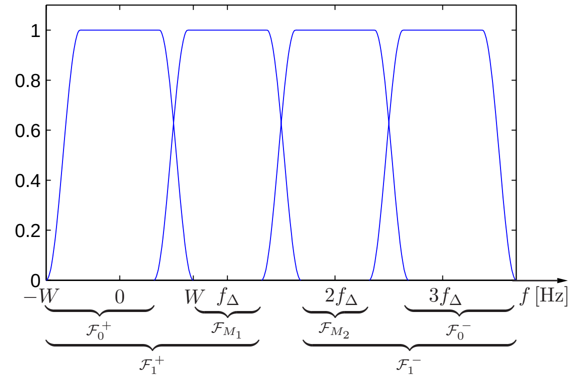

Fig. 10. Illustration of the frequency intervals defined in (29). $k_0 = 4$.

convenience: All error events considered are already time shifted such that the smallest distance is obtained. Let $d^2_{\min}\big|_{\mathcal{F}}$ denote the minimum distance of any error event in the frequency interval $\mathcal{F}$, that is[4]

$$d^2_{\min}\big|_{\mathcal{F}} \triangleq \min_{\boldsymbol{\Delta A}} d^2(\boldsymbol{\Delta A})\big|_{\mathcal{F}}. \tag{28}$$

Also, define the frequency intervals

$$\begin{aligned} \mathcal{F}_k^+ &\triangleq [-W, -W + (k+1)f_\Delta] \\ \mathcal{F}_k^- &\triangleq [(k_0 - k - 2)f_\Delta + W, (k_0 - 1)f_\Delta + W] \\ \mathcal{F}_{M_k} &\triangleq [kf_\Delta + W - f_\Delta, kf_\Delta + f_\Delta - W], \end{aligned} \tag{29}$$

where $\le k \le \left\lfloor \frac{k_0-1}{2} \right\rfloor$ for $\mathcal{F}_k^+$ and $\mathcal{F}_k^-$ and $1 \le k \le k_0-2$ for $\mathcal{F}_{M_k}$. These intervals are illustrated in Figure 10. In the interval $\mathcal{F}_{M_k}$ the kth subchannel is free from interference from other channels. Thus we can easily calculate $d^2_{\min}\big|_{\mathcal{F}_{M_k}}$ by methods that apply for single carrier systems. But since there is no interference in $\mathcal{F}_{M_k}$, we have $d^2_{\min}\big|_{\mathcal{F}_{M_1}} = d^2_{\min}\big|_{\mathcal{F}_{M_2}} = \cdots = d^2_{\min}\big|_{\mathcal{F}_{M_{k_0-2}}}$; this distance we will refer to as just $d^2_{\min}\big|_{\mathcal{F}_{\mathcal{M}}}$. In $\mathcal{F}_0^+$ the first subchannel is free from interference from the others, so we can compute $d^2_{\min}\big|_{\mathcal{F}_0^+}$ easily. Let $\boldsymbol{\Delta A}_k = [\Delta a_{k,0} \ \ldots \ \Delta a_{k,n_0-1}]$ and define the rowwise transpose

$$\begin{bmatrix} \boldsymbol{\Delta A}_0 \\ \vdots \\ \boldsymbol{\Delta A}_{k_0-1} \end{bmatrix}^{tr} = [\boldsymbol{\Delta A}_0 \ldots \boldsymbol{\Delta A}_{k_0-1}].$$

Throughout, let $\Delta \boldsymbol{A}^*$ denote componentwise complex conjugation. It is easy to show that

$$d^2(\boldsymbol{\Delta A})\big|_{\mathcal{F}_k^+} = d^2(\boldsymbol{\Delta B}^*)\big|_{\mathcal{F}_k^-} \tag{30}$$

where

$$\Delta \boldsymbol{B} = \begin{bmatrix} \boldsymbol{\Delta A}_{k_0-1} \\ \ldots \\ \boldsymbol{\Delta A}_0 \end{bmatrix}.$$

This implies that $d^2_{\min}\big|_{\mathcal{F}_k^-} = d^2_{\min}\big|_{\mathcal{F}_k^+}$, all k.

The algorithm computes the minimum distance under the assumption that exactly k_0' adjacent subchannels are involved in the error event, with $k_0' = 2, \ldots, k_0$ in sequence ($k_0' \ge 2$, otherwise the signaling is one dimensional).

(i) First set $k_0' = 2$. Consider the error symbols in the first subchannel. There must be at least a distance $d^2_{\min}\big|_{\mathcal{F}_0^-} = d^2_{\min}\big|_{\mathcal{F}_0^+}$ present in the frequency interval $\mathcal{F}_0^-$. Therefore, only events that pile up an amount $d^2(\boldsymbol{\Delta} A)\big|_{\mathcal{F}_0^+} < 2 - d^2_{\min}\big|_{\mathcal{F}_0^+}$ in $\mathcal{F}_0^+$ need to be considered. Collect these in the set $\mathcal{M}_0^+$; that is

$$\mathcal{M}_0^+ \triangleq \left\{ \boldsymbol{\Delta A}_0 : \ d^2(\boldsymbol{\Delta A}_0)\big|_{\mathcal{F}_0^+} < 2 - d^2_{\min}\big|_{\mathcal{F}_0^+} \right\}. \tag{31}$$

This also implies that only the events $\mathcal{M}_0^- \triangleq \{\boldsymbol{\Delta A}_{k_0'-1} : \boldsymbol{\Delta A}^*_{k_0'-1} \in \mathcal{M}_0^+\}$ need to be considered at the last subchannel. What number of subcarriers k_0 needs to be tested? The distance of any error event will at least pile up $d^2_{\min}\big|_{\mathcal{F}_0^+} + d^2_{\min}\big|_{\mathcal{F}_0^-} + (k_0-2)d^2_{\min}\big|_{\mathcal{F}_{\mathcal{M}}}$. Therefore k_0 needs to be at most

$$k_0 \le \frac{2 - d^2_{\min}\big|_{\mathcal{F}_0^+} + d^2_{\min}\big|_{\mathcal{F}_0^-}}{d^2_{\min}\big|_{\mathcal{F}_{\mathcal{M}}}}. \tag{32}$$

[4]If time shifts were included, (28) would become $d^2_{\min}\big|_{\mathcal{F}} \triangleq \min_{r,\boldsymbol{\Delta} A} d^2(\tau_r(\boldsymbol{\Delta A}))\big|_{\mathcal{F}}$.

$$\begin{aligned} d^2(\tau_r(\boldsymbol{\Delta A})) &= \int_{-\infty}^{\infty} \sum_{k,l,m,n} \Delta a_{k,m} \Delta a^*_{l,n} h(t-(r+m+\delta_k)T_\Delta) h^*(t-(r+n+\delta_l)T_\Delta) e^{\jmath 2\pi f_\Delta (l-k)t} dt \\ &= \int_{-\infty}^{\infty} \sum_{k,l,m,n} \Delta a_{k,m} \Delta a^*_{l,n} h(t-(m+\delta_k)T_\Delta) h^*(t-(n+\delta_l)T_\Delta) e^{\jmath 2\pi f_\Delta (l-k)(t+rT_\Delta)} dt \\ &= \sum_{k,l,m,n} \Delta a_{k,m} \Delta a^*_{l,n} e^{\jmath 2\pi f_\Delta (l-k) r T_\Delta} \int_{-\infty}^{\infty} h(t-(m+\delta_k)T_\Delta) h^*(t-(n+\delta_l)T_\Delta) e^{\jmath 2\pi f_\Delta (l-k)t} dt \\ &= \sum_{k,l,m,n} \Delta a_{k,m} \Delta a^*_{l,n} e^{\jmath 2\pi f_\Delta (l-k) r T_\Delta} \gamma_{k,l}[m,n] \\ &= \sum_{k,l} e^{\jmath 2\pi \phi (l-k)} \beta_{k,l}(\boldsymbol{\Delta A}), \end{aligned} \tag{19}$$

where

$$\beta_{k,l}(\boldsymbol{\Delta A}) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \Delta a_{k,m} \Delta a^*_{l,n} \gamma_{k,l}[m,n], \tag{20}$$

and

$$\gamma_{k,l}[m,n] = \int_{-\infty}^{\infty} h(t-(m+\delta_k)T_\Delta) h^*(t-(n+\delta_l)T_\Delta) e^{\jmath 2\pi f_\Delta (l-k)t} dt. \tag{21}$$

We can now find the minimum distance of events with two subchannels involved. We exhaust all possible error events of the form

$$\boldsymbol{\Delta A} = \begin{bmatrix} \boldsymbol{\Delta A}_0 \\ \boldsymbol{\Delta A}_1 \end{bmatrix}$$
$$\text{with} \quad \begin{array}{l} \boldsymbol{\Delta A}_0 \in \mathcal{M}_0^+ \\ \boldsymbol{\Delta A}_1 \in \mathcal{M}_0^- . \end{array}$$

With an abuse of notation, we write the set of all such $\Delta \boldsymbol{A}$ as $\mathcal{M}_0^+ \times \mathcal{M}_0^-$.

(ii) Continue with three subchannels ($k_0' = 3$). Consider the error event on the second subchannel. As stated above, it is free from interference in $\mathcal{F}_{M_1}$. What events are possible? Since we know that we will accumulate up at least an amount $2d_{\min}^2\big|_{\mathcal{F}_0^+}$ in $\mathcal{F}_0^+$ and $\mathcal{F}_0^-$ we only have to consider events that have distance less than $2 - 2d_{\min}^2\big|_{\mathcal{F}_0^+}$ in $\mathcal{F}_{M_1}$. We collect them in a set $\mathcal{M}_M$. Note that it does not matter what subchannel we consider; since there is no interference in $\mathcal{F}_{M_k}$, the set is identical for all k and we take $k = 1$. Then

$$\mathcal{M}_M \triangleq \left\{ \boldsymbol{\Delta A}_1 \colon\, d^2(\boldsymbol{\Delta A}_1)\big|_{\mathcal{F}_{M_1}} < 2 - 2d_{\min}^2\big|_{\mathcal{F}_0^+} \right\}. \tag{33}$$

We can also easily compute the minimum distance in $\mathcal{F}_{M_k}$. Since we know that we will accumulate at least this amount, we can redefine the sets $\mathcal{M}_0^+$ and $\mathcal{M}_0^-$; $\mathcal{M}_0^+$ becomes $\mathcal{M}_0^+ = \{\boldsymbol{\Delta A}_0 : d^2(\boldsymbol{\Delta A}_0)\big|_{\mathcal{F}_0^+} < 2 - d_{\min}^2\big|_{\mathcal{F}_0^+} - d_{\min}^2\big|_{\mathcal{F}_M}\}$ and $\mathcal{M}_0^-$ is found from $\mathcal{M}_0^+$ by the formula

$$\mathcal{M}_k^- \triangleq \left\{ [\boldsymbol{\Delta A}_0 ... \boldsymbol{\Delta A}_{k-1}]^{tr} : [\boldsymbol{\Delta A}_{k-1}^* ... \boldsymbol{\Delta A}_0^*]^{tr} \in \mathcal{M}_k^+ \right\}. \tag{34}$$

Thus, in the the interval $\mathcal{F}_1^+$ we only have to consider events in $\mathcal{M}_0^+ \times \mathcal{M}_M$ when we find

$$\begin{aligned} \mathcal{M}_1^+ \quad = \quad & \{ [\boldsymbol{\Delta A}_0 \ \boldsymbol{\Delta} A_1]^{tr} : \\ & d^2([\boldsymbol{\Delta A}_0 \ \boldsymbol{\Delta A}_1]^{tr})\big|_{\mathcal{F}_1^+} < 2 - d_{\min}^2\big|_{\mathcal{F}_0^+} \Big\}. \end{aligned}$$

We also find the set $\mathcal{M}_1^-$. Note that $\mathcal{M}_1^+$ contains events with components over subchannels 0 and 1 and that $\mathcal{M}_1^-$ contains events over subchannels 1 and 2. Thus we cannot treat the two sets independently, as was possible when $k_0' = 2$. We need to exhaust all events of the form

$$\boldsymbol{\Delta A} = \begin{bmatrix} \boldsymbol{\Delta A}_0 \\ \boldsymbol{\Delta A}_1 \\ \boldsymbol{\Delta A}_2 \end{bmatrix}$$
$$\text{with} \quad \begin{array}{l} [\boldsymbol{\Delta A}_0 \ \boldsymbol{\Delta A}_1]^{tr} \in \mathcal{M}_1^+ \\ [\boldsymbol{\Delta A}_1 \ \boldsymbol{\Delta A}_2]^{tr} \in \mathcal{M}_1^- . \end{array}$$

This generates the minimum distance when there are exactly three subchannels involved in the error event.

(iii) To investigate events on four subchannels, we first compute $d_{\min}^2\big|_{\mathcal{F}_1^+}$ by exhausting the events in $\mathcal{M}_1^+$. Then we redefine $\mathcal{M}_1^+$ to $\mathcal{M}_1^+ = \{[\boldsymbol{\Delta A}_0 \, \boldsymbol{\Delta A}_1]^{tr} : d^2(\boldsymbol{\Delta A})\big|_{\mathcal{F}_1^+} < 2 - d_{\min}^2\big|_{\mathcal{F}_1^+}\}$ and exhaust events in $\mathcal{M}_1^+ \times \mathcal{M}_1^-$, with $\mathcal{M}_1^-$ defined from $\mathcal{M}_1^+$ according to (34). The procedure can be repeated up to k_0 subchannels, where k_0 satisfies (32).

We summarize the algorithm.

1) First compute $d_{\min}^2\big|_{\mathcal{F}_0^+}$ and $d_{\min}^2\big|_{\mathcal{F}_M}$; these are one dimensional searches. If $d_{\min}^2\big|_{\mathcal{F}_0^+} > 1$, then stop and go to step 4; the minimum distance for error events with 2 or more subchannels involved is larger than 2. Define the sets $\mathcal{M}_0^+$, $\mathcal{M}_0^-$ and $\mathcal{M}_M$ according to (31), (34) and (33). Set $k_0' = 2$. Exhaust all events in the set $\mathcal{M}_0^+ \times \mathcal{M}_0^-$. This generates the minimum distance when 2 subchannels are involved and we denote it d_2^2.
2) Set $k_0' = k_0' + 1$ (k_0' is now odd). If $k_0' > k_0$ then stop; the search over all $k_0 \times n_0$ error events is complete. Define a temporary set $\mathcal{T} = \mathcal{M}_{(k_0'-3)/2}^+$. Then redefine $\mathcal{M}_{(k_0'-3)/2}^+$ to

$$\begin{aligned} \mathcal{M}_{\frac{k_0'-3}{2}}^+ = \Bigg\{ & \left[\boldsymbol{\Delta A}_0 \ldots \boldsymbol{\Delta A}_{\frac{k_0'-3}{2}} \right]^{tr} \in \mathcal{T} : \\ & d^2 \left(\left[\boldsymbol{\Delta A}_0 \ldots \boldsymbol{\Delta A}_{\frac{k_0'-3}{2}} \right]^{tr} \right) < \\ & 2 - d_{\min}^2\big|_{\mathcal{F}_M} - d_{\min}^2\big|_{\mathcal{F}^+_{\frac{k_0'-3}{2}}} \Bigg\}. \end{aligned}$$

Find $d_{\min}^2\big|_{\mathcal{F}^+_{(k_0'-1)/2}}$ by exhausting the events in $\mathcal{M}_{(k_0'-3)/2}^+ \times \mathcal{M}_M$. Define

$$\begin{aligned} \mathcal{M}_{\frac{k_0'-1}{2}}^+ = \Bigg\{ & \left[\boldsymbol{\Delta A}_0 \ldots \boldsymbol{\Delta A}_{\frac{k_0'-1}{2}} \right]^{tr} \in \mathcal{M}_{\frac{k_0'-3}{2}}^+ \times \mathcal{M}_M : \\ & d^2 \left(\left[\boldsymbol{\Delta A}_0 \ldots \boldsymbol{\Delta A}_{\frac{k_0'-1}{2}} \right]^{tr} \right) < \\ & 2 - d_{\min}^2\big|_{\mathcal{F}^+_{\frac{k_0'-3}{2}}} \Bigg\}. \end{aligned}$$

If $d_{\min}^2\big|_{\mathcal{F}^+_{(k_0'-1)/2}} + d_{\min}^2\big|_{\mathcal{F}^+_{(k_0'-3)/2}} > 2$, then stop and go to step 4; the minimum distance for error events with k_0' or more subchannels involved is larger than 2. Otherwise, compute the set $\mathcal{M}^-_{(k_0'-3)/2}$. Find $d_{\min}^2$ for error events with k_0' subchannels involved, which we denote $d_{k_0'}^2$, by exhausting all events of the form

$$\boldsymbol{\Delta A} = \begin{bmatrix} \boldsymbol{\Delta A}_0 \\ \vdots \\ \boldsymbol{\Delta A}_{(k_0'-1)/2} \\ \vdots \\ \boldsymbol{\Delta A}_{k_0'-1} \end{bmatrix}$$
$$\text{with} \quad \begin{array}{l} \left[\boldsymbol{\Delta A}_0 \ldots \ \boldsymbol{\Delta A}_{(k_0'-1)/2} \right]^{tr} \in \mathcal{M}^+_{(k_0'-1)/2} \\ \left[\boldsymbol{\Delta A}_{(k_0'-1)/2} \ \boldsymbol{\Delta A}_{k_0'-1} \right]^{tr} \in \mathcal{M}^-_{(k_0'-1)/2}. \end{array}$$

3) Set $k_0' = k_0' + 1$ (k_0' is now even). If $2d_{\min}^2\big|_{\mathcal{F}^+_{(k_0'-1)/2}} > 2$, then stop and go to step 4; the minimum distance for error events with k_0' or more subchannels involved is larger than 2. If not, if $k_0' > k_0$ then stop and go to step 4; the search over all $k_0 \times n_0$ error events is complete. Otherwise, define

a temporary set $\mathcal{T} = \mathcal{M}^{+}_{k'_0/2-1}$ and redefine the set

$$\begin{aligned}\mathcal{M}^{+}_{k'_0/2-1} \quad = \quad & \Big\{ \left[\boldsymbol{\Delta A}_0 \ldots \boldsymbol{\Delta A}_{k'_0/2-1} \right]^{tr} \in \mathcal{T} : \\ & d^2 \left(\left[\boldsymbol{\Delta A}_0 \ldots \boldsymbol{\Delta A}_{k'_0/2-1} \right]^{tr} \right) \\ & < \; 2 - d^2_{\min} \big|_{\mathcal{F}^{+}_{k'_0/2-1}} \Big\}.\end{aligned}$$

Compute the set $\mathcal{M}^{-}_{k'_0/2-1}$. Find $d^2_{\min}$ for error events with k'_0 subchannels involved, which we denote $d^2_{k'_0}$, by exhausting all events of the form

$$\boldsymbol{\Delta A} = \begin{bmatrix} \boldsymbol{\Delta A}_0 \\ \cdots \\ \boldsymbol{\Delta A}_{k'_0/2-1} \\ \boldsymbol{\Delta A}_{k'_0/2} \\ \cdots \\ \boldsymbol{\Delta A}_{k'_0-1} \end{bmatrix}$$

$$\text{with} \quad \begin{aligned} & \left[\boldsymbol{\Delta A}_0 \; \ldots \; \boldsymbol{\Delta A}_{k'_0/2-1} \right]^{tr} \in \mathcal{M}^{+}_{k'_0/2-1} \\ & \left[\boldsymbol{\Delta A}_{k'_0/2} \ldots \boldsymbol{\Delta A}_{k'_0-1} \right]^{tr} \in \mathcal{M}^{-}_{k'_0/2-1}. \end{aligned}$$

Go to step 2.

4) Find $d^2_{\min} = \min\{d^2_2, d^2_3, \ldots, d^2_{k_0}\}$.

References

[1] J.E. Mazo, "Faster-than-Nyquist Signaling," *Bell Syst. Tech. J.*, vol. 54, pp. 1451–1462, Oct. 1975.

[2] A.D. Liveris and C.N. Georghiades, "Exploiting Faster-than-Nyquist Signaling," *IEEE Trans. Commun.*, vol. 51, pp. 1502–1511, Sept. 2003.

[3] F. Rusek and J.B. Anderson, "Non Binary and Precoded Faster than Nyquist Signaling," to appear, *IEEE Trans. Commun.*, 2008.

[4] F. Rusek and J.B. Anderson, "M-ary Coded Modulation by Butterworth Filtering," *Proc.*, 2003 Int. Symp. Information Theory, Yokohama, p. 184, June 2003.

[5] F. Rusek and J.B. Anderson, "The Two Dimensional Mazo Limit," *Proc.*, 2005 IEEE Int. Symp. Information Theory, Adelaide, pp. 970–974, Sept. 2005.

[6] J.B. Anderson and F. Rusek, "Optimal Side Lobes under Linear and Faster-than-Nyquist Modulation," *Proc.*, 2007 IEEE Int. Symp. Information Theory, Nice, June 2007.

[7] J.B. Anderson and A. Svensson, *Coded Modulation Systems*, Plenum/Kluwer, New York, 2003.

[8] M. Marrow and J.K. Wolf, "Iterative Detection of 2-Dimensional ISI Channels," in *Proc.*, 2003 Int. Workshop Information Theory, Paris, pp. 131–134, April 2003.

[9] Y. Wu, J.A. O'Sullivan, N. Singla and R.S. Indeck, "Iterative Detection and Decoding for Separable Two-Dimensional Intersymbol Interference," *IEEE Trans. Magnetics*, vol. 39, pp. 2115–2120, July 2003.

[10] P.S. Kumar and S. Roy, "Two-Dimensional Equalization: Theory and Applications to High Density Magnetic Recording," *IEEE Trans. Commun.*, vol. 42, pp. 386–395, Feb./March/April. 1994.

[11] J.B. Soriaga, H.D. Pfister and P.H. Siegel, "On Achievable Rates of Multistage Decoding on Two-Dimensional ISI Channels," *Proc.*, 2005 Int. Symp. Information Theory, Adelaide, pp. 1348–1352, Sept. 2005.

[12] G. Colavolpe and A. Barbieri, "On MAP Symbol Detection for ISI Channels Using the Ungerboeck Observation Model," *IEEE Commun. Letters*, vol. 9, pp. 720–722, Aug. 2005.

[13] M. Kobayashi, J. Boutros and G. Caire, "Successive Interference Cancellation with SISO Decoding and EM Channel Estimation," *IEEE J. Sel. Areas Commun.*, vol. 19, pp. 1450–1460, Aug. 2001.

[14] F. Rusek and J.B. Anderson, "Successive Interference Cancellation in Multistream Faster-than-Nyquist Signaling," in *Proc.*, ACM/IEEE Int. Wireless Communications and Mobile Computing Conf., Vancouver, pp. 1021–1026, July 2006.

[15] I. Demirkan, P. H. Siegel and J. K. Wolf, "Error Event Characterization on 2-D ISI Channels," *Proc.*, 2006 Int. Symp. Information Theory, Seattle, pp. 1095–1099, July 2006.

[16] E. Soljanin and C. N. Georghiades, "Multihead Detection for Multitrack Recording Channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2988–2997, Nov. 1998.

[17] N. Mysore and J. Bajcsy, "Reduced Complexity Signal Detection and Turbo Decoding for Multitrack Magnetic Recording Channels," *IEEE Trans. Magnetics*, vol. 41, no. 10, pp. 2974–2976, Oct. 2005.

[18] E. Kurtas, J. G. Proakis and M. Salehi, "Coding for Multitrack Magnetic Recording Systems," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 2020–2023, Nov. 1997.

[19] P. J. Davey, T. Donnelly, D. J. Mapps and N. Darragh, "Two-Dimensional Coding for a Multi-Track Recording System to Combat Inter-Track Interference," *IEEE Trans. Magnetics*, vol. 34, no. 4, Part 1, pp. 1949–1951, July 1998.