Fredrik Rusek, Giulio Colavolpe,
and Carl-Erik W. Sundberg

40 Years with the Ungerboeck Model: A Look at Its Potentialities

It has been about 40 years since Gottfried Ungerboeck published his paper [1] on an alternative maximum-likelihood (ML) detector for intersymbol interference (ISI) channels. The ISI model used by Ungerboeck is commonly referred to as the *Ungerboeck model*. Ungerboeck's ML detector has equivalent performance compared to Forney's detector, which was published two years earlier in [2], but received lesser considerations. Perhaps the best example of this is the fact that a BCJR algorithm [3] operating on the Ungerboeck model was derived as late as 2005 [4]. However, the Ungerboeck model has many strong aspects and has therefore been rediscovered over the last few decades.

SCOPE

In this lecture note, we give a number of illuminating examples where the Ungerboeck model is essential. We hope that this column will lead to increased awareness and use of the Ungerboeck model among the signal processing community.

RELEVANCE

Essentially all communication systems are modeled by a discrete-time model. The white-noise model is the predominant choice of model today. When low-complexity algorithms are used, the choice of model plays a role. In some cases, superior performance and/or lower complexity can be achieved by the very same algorithm, but where the white noise model has been replaced by another model. Awareness of models other than the white-noise model is of great value to engineers and researchers, especially to those working in the borderline of signal processing and wireless communications.

Digital Object Identifier 10.1109/MSP.2014.2374221
Date of publication: 6 April 2015

PREREQUISITES

This lecture note assumes basic knowledge of signal space descriptions of communication systems, about Viterbi- and Bahl-Cooke-Jelinek-Raviv (BCJR)-type algorithms for communication channels with memory, and about factor graphs (FGs) and the sum-product algorithm (SPA).

SUMMARY OF DETECTION THEORY RESULTS

In the following, we will denote by $p(\cdot)$ [respectively, $P(\cdot)$] the probability density function (pdf) [respectively, the probability mass function (pmf)] of a continuous (respectively, discrete) random variable. In addition $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ denote the real and imaginary part of a complex number whereas $(\cdot)^*$ and $(\cdot)^T$ stand for transpose conjugate and transpose, respectively. Let $\boldsymbol{x} = [x_0, x_1, \cdots]$ be a sequence of modulation symbols drawn from a discrete alphabet $\mathcal{X}$. These symbols are transmitted over a communication channel via a modulation format. Let z be an arbitrary sufficient statistic properly extracted from the received signal. Maximum a posteriori (MAP) sequence and symbol detection strategies are based on the following decision rules:

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}} P(\boldsymbol{x}\,|\,z)\\ &= \arg\max_{\boldsymbol{x}} p(z\,|\,\boldsymbol{x})P(\boldsymbol{x}) \qquad (1)\end{aligned}$$

and

$$\begin{aligned}\hat{x}_k &= \arg\max_{x_k} P(x_k\,|\,z)\\ &= \arg\max_{x_k} p(z\,|\,x_k)P(x_k), \qquad (2)\end{aligned}$$

respectively. They minimize the sequence and symbol error probability, respectively. In case of communication systems with finite memory, they can be implemented through the Viterbi [5] and the BCJR algorithm [3], respectively.

Without loss of generality, in the following we will often assume that modulation symbols $\{x_k\}$ are independent and uniformly distributed (i.u.d.). As a consequence, all a priori probabilities can be safely discarded from the aforementioned MAP strategies and they become perfectly equivalent to the corresponding ML strategies [6].

DETECTION ON INTERSYMBOL INTERFERENCE CHANNELS

The continuous-time ISI channel may be described, assuming the use of a linear modulation, by means of the following complex baseband equation:

$$y(t) = \sum_k x_k q(t - kT) + w(t), \qquad (3)$$

where $q(t)$ is the received pulse, $w(t)$ is complex white Gaussian noise with two-sided spectral density N_0, and T is the symbol time.

In 1972, Forney showed that ML detection of $\boldsymbol{x}$ can be carried out by an application of the Viterbi algorithm (VA) [2]. Forney first applied a matched filter and sampling operation to the signal $y(t)$ to form the discrete-time model $y_k = \int_{-\infty}^{\infty} y(t) q^*(t - kT)\,\mathrm{d}t$. Each sample y_k can be expressed as

$$y_k = \sum_{\ell=-L}^{L} g_\ell x_{k-\ell} + n_k, \qquad (4)$$

where $g_\ell = \int_{-\infty}^{\infty} q(t) q^*(t - \ell T)\,\mathrm{d}t$. The variable L specifies the memory of the system and is the smallest value such

that $g_\ell = 0, |\ell| > L$. The noise in the model (4) is not white, but is correlated according to $\mathbb{E}[n_{k+\ell} n_k^*] = N_0 g_\ell$. To obtain white noise, Forney filtered the variables y_k with a whitening filter $\{f_k\}$, which yields

$$r_k = \sum_\ell f_\ell y_{k-\ell} = \sum_{\ell=0}^{L} h_\ell x_{k-\ell} + w_k. \quad (5)$$

In the model (5), the zero-mean noise variables $\{w_k\}$ are uncorrelated with variance N_0. The channel impulse response is causal and is related to $\{g_\ell\}$ as $g_\ell = \sum_{k=0}^{L} h_{k+\ell} h_k^*$, i.e., $\boldsymbol{g} = [g_{-L}, \ldots, g_L]$ is the autocorrelation sequence of $\boldsymbol{h} = [h_0, \ldots, h_L]$. Both samples $\{y_k\}$ and $\{r_k\}$ represent a sufficient statistic and can thus be employed for detection. Throughout this lecture note, the three letters (y,g,n) imply that we are discussing the model (4), while we are discussing (5) if we use (r,h,w).

Forney next observed that each sample r_k only depends on the current channel input x_k and the L most recent channel inputs $x_{k-1}, \ldots, x_{k-L}$. Therefore, the signal can be described by means of a trellis where each state is defined as $\sigma_k = (x_{k-1}, \ldots, x_{k-L})$. Thus, the number of states is $|\chi|^L$. As an example, when $\chi = \{0,1\}$ and $L = 2$, a section of the corresponding trellis between the discrete-time instants k and $k+1$ is shown in Figure 1. In this figure, trellis transitions driven by symbol $x_k = 0$ are denoted by using dashed lines, whereas solid lines correspond to transitions driven by $x_k = 1$.

Due to the fact that samples r_k are conditionally independent, the conditional probability density function $p(\boldsymbol{r}|\boldsymbol{x})$ required for the implementation of the strategy (1) can be expressed in a recursive factorization of the form

$$p(\boldsymbol{r}|\boldsymbol{x}) = \prod_k \frac{1}{\pi N_0} \exp\left(-\frac{\left| r_k - \sum_{\ell=0}^{L} h_\ell x_{k-\ell} \right|^2}{N_0} \right). \quad (6)$$

Based on (6) it is straightforward to set up the VA. In fact, under the assumption of i.u.d. modulation symbols and taking into account that the logarithm is a monotonic function, the strategy (1) can be expressed as

$$\hat{\boldsymbol{x}} = \operatorname*{argmax}_{\boldsymbol{x}} p(\boldsymbol{r}|\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{x}} \ln p(\boldsymbol{r}|\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{x}} \sum_k \mu_k, \quad (7)$$

where

$$\mu_k = \left| r_k - \sum_{\ell=0}^{L} h_\ell x_{k-\ell} \right|^2 \quad (8)$$

is the so-called branch metric of the VA.

In turbo equalization applications [7], one may resort to the MAP symbol detection strategy. In this case, it is sufficient to replace the VA with the BCJR algorithm, possibly implemented in the logarithmic domain [8]. It will make use of the same branch metric μ_k.

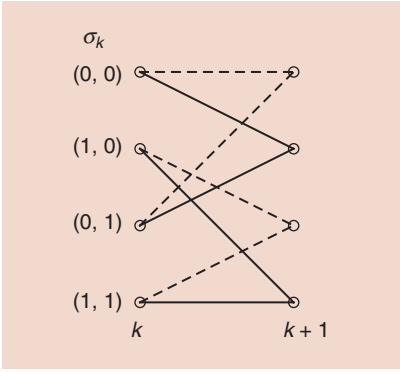
However, a demodulator may just as well take as starting point the model (4) as already shown by Ungerboeck in 1974 in [1]. The model is commonly referred to as the *Ungerboeck model*, while the white-noise model (5) is referred to as the *Forney model*—a nomenclature we will follow. The noise variables $\{n_k\}$ are still Gaussian, but are colored. However, the noise color is irrelevant since the critical issue for the application of a VA-type-detector is that the conditional pdf $p(\boldsymbol{r}|\boldsymbol{x})$ has a recursive factorization that can be expressed in terms of the signal $\boldsymbol{y}$. This is indeed the case as can be seen by expanding (6) (see [1] for further details). See (9) in the box at the bottom of the page, where $\gamma_1 = \prod_k (\pi N_0)^{-1}$, $\gamma_2 = \gamma_1 \exp(-\|\boldsymbol{r}\|^2/N_0)$, and where we used $y_k = \sum_{\ell=0}^{L} h_\ell^* r_{k+\ell}$ in the last equality. Note that γ_2 is independent of $\{x_k\}$ and can be neglected. Again, under the assumption of i.u.d. modulation symbols, the strategy (1) can be expressed as

$$\hat{\boldsymbol{x}} = \operatorname*{argmax}_{\boldsymbol{x}} \ln p(\boldsymbol{r}|\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{x}} \sum_k \eta_k, \quad (10)$$

where this time the branch metric is

$$\eta_k = \mathcal{R}\left\{ y_k x_k^* - \frac{1}{2} |x_k|^2 g_0 - x_k^* \sum_{\ell=1}^{L} g_\ell x_{k-\ell} \right\}. \quad (11)$$

Again, we see that only the L most recent channel inputs $x_{k-1}, \ldots, x_{k-L}$ are needed at each time epoch k. Ungerboeck's and



[FIG1] An example of a trellis section.

$$p(\boldsymbol{r}|\boldsymbol{x}) = \gamma_1 \exp\left(-\frac{\sum_k |r_k|^2}{N_0} + \frac{2}{N_0} \mathcal{R}\left\{ \sum_k \sum_{\ell=0}^{L} x_{k-\ell}^* h_\ell^* r_k \right\} - \frac{1}{N_0} \sum_{k,\ell,\ell'} x_{k-\ell}^* x_{k-\ell'} h_\ell^* h_{\ell'} \right)$$
$$= \gamma_2 \exp\left(\frac{2}{N_0} \mathcal{R}\left\{ \sum_k x_k^* \sum_{\ell=0}^{L} h_\ell^* r_{k+\ell} \right\} - \frac{1}{N_0} \sum_k x_k^* \sum_{\ell=-L}^{L} x_{k-\ell} g_\ell \right)$$
$$= \gamma_2 \prod_k \exp\left(\frac{2}{N_0} \mathcal{R}\left\{ y_k x_k^* - \frac{1}{2} |x_k|^2 g_0 - x_k^* \sum_{\ell=1}^{L} g_\ell x_{k-\ell} \right\} \right) \quad (9)$$

Forney's ML sequence detectors involve different computations, but they traverse the very same trellis, and their final outputs are identical. Two strong aspects of the Ungerboeck model are that no squaring operations are needed and that no whitening is needed.

It is interesting to note that while the BCJR already became available for the Forney model by 1974 with [3], the story differs remarkably for the Ungerboeck model. An equivalent algorithm to the BCJR that operates on the Ungerboeck model and employs the same branch metric (11) was demonstrated as late as 2005 in [4]. As a consequence, turbo equalization based on Ungerboeck's model was not available before 2005. The Ungerboeck model has a number of strengths and has been rediscovered several times during the recent past. Before we turn our attention to three short examples that illuminate its strengths, we first extend it into a model for general linear channels.

DETECTION ON GENERAL LINEAR CHANNELS

Let us write (5) as a matrix equation,

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}. \tag{12}$$

In the ISI case, the matrix $\boldsymbol{H}$ is a Toeplitz matrix that represents the convolutional operator. However, (12) can also represent any other linear channel, such as multiple-input, multiple-output (MIMO), intercarrier interference (ICI), MIMO-ISI etc. Irrespective of from where the model (12) came, the corresponding conditional pdf has expression

$$p(\boldsymbol{r}|\,\boldsymbol{x}) = \gamma_1 \exp\left(-\frac{\|\,\boldsymbol{r}-\boldsymbol{H}\boldsymbol{x}\,\|^2}{N_0}\right). \tag{13}$$

One can reach a tree structure, suitable for demodulation, by a QL factorization $\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{L}$ of the channel, which makes the model (12) "causal" in the vector index. This gives

$$\tilde{\boldsymbol{r}} = \boldsymbol{Q}^*\boldsymbol{r} = \boldsymbol{L}\boldsymbol{x} + \tilde{\boldsymbol{w}},$$

which enables a recursive factorization, similar to (6),

$$p(\tilde{\boldsymbol{r}}\,|\,\boldsymbol{x}) = \prod_k \frac{1}{\pi N_0} \exp\left(-\frac{\left|\tilde{r}_k - \sum_{\ell=1}^{k} L_{k,\ell} x_\ell\right|^2}{N_0}\right). \tag{14}$$

A tree search procedure can now be reached. In the case of a channel with finite memory, i.e., $L_{k,\ell} = 0, |\,k-\ell\,| > L$, the VA or the BCJR can be applied since the tree collapses into a trellis with $|\,\mathcal{X}\,|^L$ states. Hence, we refer to (12) as the Forney model for a linear channel.

To avoid computation of a QL factorization of the channel matrix, [9] proposed to first multiply the vector $\boldsymbol{r}$ with a matched filter $\boldsymbol{y} = \boldsymbol{H}^*\boldsymbol{r} = \boldsymbol{G}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{G} = \boldsymbol{H}^*\boldsymbol{H}$, and $\boldsymbol{n}$ is colored Gaussian noise with covariance matrix $N_0\boldsymbol{G}$. However, in view of (4), this is nothing but an extension of the Ungerboeck model for ISI into a formulation for a general linear channel. Next, [9] proceeds with the derivation of a recursive factorization, suitable for a tree search, and finally obtains

$$p(\boldsymbol{r}\,|\,\boldsymbol{x}) = \gamma_2 \prod_k \exp\left(\frac{2}{N_0}\mathcal{R}\left\{x_k^* y_k - \frac{1}{2}|\,x_k\,|^2 G_{k,k} - x_k^* \sum_{\ell=1}^{k} G_{\ell,k} x_\ell\right\}\right), \tag{15}$$

where, again, γ_2 is irrelevant for decision. This we recognize as the extension of Ungerboeck's (9) into a formulation for general linear channels.

PROBLEM STATEMENT

Is the choice of model relevant? Is there any example of practical systems where the Ungerboeck model is more convenient? We now give a few examples of systems where the Ungerboeck model can offer superior performance and/or lower complexity.

SOLUTION

CHANNEL SHORTENING DETECTION

Since the VA is often of prohibitive complexity, Falconer and Magee [10] proposed in 1973 to make use of the following reduced-complexity scheme: 1) filter the signal (5) with a filter that aims at reducing the memory of the effective impulse response from L to $K < L$ and 2) apply the VA to the filtered signal, but based on the shorter effective channel. Thus, the VA traverses a trellis with $|\,\mathcal{X}\,|^K$ states rather than the full trellis of size $|\,\mathcal{X}\,|^L$. In terms of a general linear channel, what is done is that the conditional pdf (13) is replaced by the mismatched version

$$T(\boldsymbol{r}\,|\,\boldsymbol{x}) \propto \exp\left(-\frac{\|\,\boldsymbol{W}\boldsymbol{r}-\boldsymbol{F}\boldsymbol{x}\,\|^2}{N_0}\right), \tag{16}$$

where $\boldsymbol{W}$ is the channel shortener, $\boldsymbol{F}$ is a matrix that has K nonzero consecutive diagonals, and the normalization constant has been neglected. This specifies a trellis with $|\,\mathcal{X}\,|^K$ states so that the VA or the BCJR can be applied. The operations of such VAs or BCJR algorithms are specified by (14) with $\tilde{\boldsymbol{r}}$ and $\boldsymbol{L}$ being replaced by $\boldsymbol{W}\boldsymbol{r}$ and $\boldsymbol{F}$, respectively.

However, instead of using (14), we can, with identical complexity, use (15). By expanding the square magnitude in (16) and neglecting the irrelevant terms we can express $T(\boldsymbol{r}\,|\,\boldsymbol{x})$ as

$$T(\boldsymbol{r}\,|\,\boldsymbol{x}) \propto \exp\left(\frac{2\mathcal{R}\{\boldsymbol{x}^*\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{r}\} - \boldsymbol{x}^*\boldsymbol{F}^*\boldsymbol{F}\boldsymbol{x}}{N_0}\right) \tag{17}$$

and then execute the trellis processing via (15) by replacing $\boldsymbol{y}$ and $\boldsymbol{G}$ with $\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{r}$ and $\boldsymbol{F}^*\boldsymbol{F}$, respectively.

If the processing is done via (15), only the matrices $\boldsymbol{F}^*\boldsymbol{W}$ and $\boldsymbol{F}^*\boldsymbol{F}$ are relevant, and not the matrices $\boldsymbol{W}$ and $\boldsymbol{F}$ themselves. We can therefore relax the structure of $\boldsymbol{F}^*\boldsymbol{W}$ and $\boldsymbol{F}^*\boldsymbol{F}$ so that we replace $T(\boldsymbol{r}\,|\,\boldsymbol{x})$ in (17) with

$$T(\boldsymbol{r}\,|\,\boldsymbol{x}) \propto \exp\left(\frac{2\mathcal{R}\{\boldsymbol{x}^*\boldsymbol{H}_{\mathrm{r}}\boldsymbol{r}\} - \boldsymbol{x}^*\boldsymbol{G}_{\mathrm{r}}\boldsymbol{x}}{N_0}\right), \tag{18}$$

where $\boldsymbol{H}_{\mathrm{r}}$ is arbitrary and $\boldsymbol{G}_{\mathrm{r}}$ is a Hermitian matrix with only the main $2K+1$ diagonals holding nonzero values. The strength of replacing $\boldsymbol{F}^*\boldsymbol{F}$ with $\boldsymbol{G}_{\mathrm{r}}$ is that the matrix $\boldsymbol{G}_{\mathrm{r}}$ needs not to be positive semidefinite, unlike the matrix $\boldsymbol{F}^*\boldsymbol{F}$ which is positive semidefinite by construction. This allows for a wider class of mismatched conditional pdfs than what can be reached by (16); based on (16), one is restricted to have a positive semidefinite $\boldsymbol{G}_{\mathrm{r}}$ matrix.

As far as the derivation of $\boldsymbol{G}_{\mathrm{r}}$ and $\boldsymbol{H}_{\mathrm{r}}$ according to a proper optimality criterion

is concerned, we refer the reader to [11]. Importantly, to find optimal $\boldsymbol{G}_{\mathrm{r}}$ and $\boldsymbol{H}_{\mathrm{r}}$ is much simpler than finding optimal $\boldsymbol{W}$ and $\boldsymbol{F}$. Somewhat surprisingly, the optimal $\boldsymbol{G}_{\mathrm{r}}$ matrix to choose is often in-definite so that a mismatched conditional pdf of the form (16) is inferior to the form (18).

In brief, channel shortening has been studied since 1973, but the starting point has always been the Forney model. This is suboptimal, as the optimal solution for an Ungerboeck-based channel shortening receiver can not, in general, be reached with the Forney model.

MAX-LOG-MAP DEMODULATION OF MIMO CHANNELS

The computation of the pdf (13) required for the implementation of the strategies (1) or (2) requires the computation of metrics $\|\boldsymbol{r}-\boldsymbol{H}\boldsymbol{x}\|^2$ for all possible values of $\boldsymbol{x}$. How many complex multiplications are needed to do this task? If we assume an $M\times M$ channel matrix, we have $|\mathcal{X}|^M$ vectors $\boldsymbol{x}$ to test. For each vector we need M^2 multiplications to form $\boldsymbol{H}\boldsymbol{x}$, and then M more to compute the norm. Hence, a brute force evaluation would give about $|\mathcal{X}|^M(M+1)M$ complex multiplications. In [12], a much more efficient computation is presented by a clever rewriting of the associated terms of computing each metric $\|\boldsymbol{r}-\boldsymbol{H}\boldsymbol{x}\|^2$. To exemplify how the metric is rewritten for simplifying the calculations, we consider the case of a 2×2 MIMO system. The received signal, the channel matrix, and the transmitted data vector, all complex valued, are expressed as

$$\boldsymbol{r}=\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \boldsymbol{H}=\begin{bmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{bmatrix}$$
$$\boldsymbol{x}=\begin{bmatrix} \mathcal{R}\{x_1\}+j\mathcal{I}\{x_1\} \\ \mathcal{R}\{x_2\}+j\mathcal{I}\{x_2\} \end{bmatrix},$$

respectively. With that, in [12] the metric is expressed in the following manner [see (19) in the box at the bottom of the page].

Notice that we neglected the term $\|\boldsymbol{r}\|^2$, which is irrelevant for detection. Although never mentioned in [12], this is precisely the Ungerboeck model, but in a real-valued formulation. In fact, as defined in the section "Detection on General Linear Channels," it is $\boldsymbol{y}=\boldsymbol{H}^*\boldsymbol{r}$ and $\boldsymbol{G}=\boldsymbol{H}^*\boldsymbol{H}$, and thus $y_1^*=r_1^*H_{1,1}+r_2^*H_{2,1}$ and $y_2^*=r_1^*H_{1,2}+r_2^*H_{2,2}$, $G_{1,1}=|H_{1,1}|^2+|H_{2,1}|^2$, $G_{1,2}=H_{1,1}^*H_{1,2}+H_{2,1}^*H_{2,2}=G_{2,1}^*$, etc. Hence, (19) becomes

$$\begin{aligned}\|\boldsymbol{r}-\boldsymbol{H}\boldsymbol{x}\|^2 \propto & \begin{pmatrix} -2\mathcal{R}\{y_1^*\} \\ 2\mathcal{I}\{y_1^*\} \\ -2\mathcal{R}\{y_2^*\} \\ 2\mathcal{I}\{y_2^*\} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathcal{R}\{x_1\} \\ \mathcal{I}\{x_1\} \\ \mathcal{R}\{x_2\} \\ \mathcal{I}\{x_2\} \end{pmatrix} \\ & + \begin{pmatrix} G_{1,1} \\ G_{1,1} \\ G_{2,2} \\ G_{2,2} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathcal{R}\{x_1\}^2 \\ \mathcal{I}\{x_1\}^2 \\ \mathcal{R}\{x_2\}^2 \\ \mathcal{I}\{x_2\}^2 \end{pmatrix} \\ & + \begin{pmatrix} 2\mathcal{R}\{G_{1,2}\} \\ -2\mathcal{I}\{G_{1,2}\} \\ 2\mathcal{I}\{G_{1,2}\} \\ 2\mathcal{R}\{G_{1,2}\} \end{pmatrix}^{T} \begin{pmatrix} \mathcal{R}\{x_1\}\mathcal{R}\{x_2\} \\ \mathcal{R}\{x_1\}\mathcal{I}\{x_2\} \\ \mathcal{I}\{x_1\}\mathcal{R}\{x_2\} \\ \mathcal{I}\{x_1\}\mathcal{I}\{x_2\} \end{pmatrix}, \end{aligned} \tag{20}$$

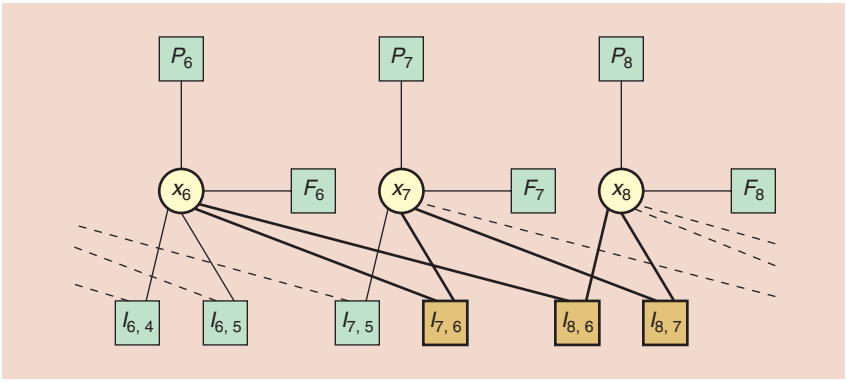
which is exactly the argument of the exponential in (15) in real-valued formulation.

Based on the formulation (19), [12] developed a methodolgy to calculate all the $|\mathcal{X}|^M$ metrics by only $2M^2(2M+3)-2M$ multiplications. This remarkable result relies on the structure of the Ungerboeck model (15), which allows for a hierarchical formulation of the minimum metric terms, using submetrics in an efficient manner avoiding duplication of calculations. The Ungerboeck metric is utilized in such a way that parallelization is achieved and multiple calculations are done in one clock cycle. Furthermore, a doubly recursive evaluation of submetrics is used; for a detailed description of this, see [12]. In [12], it is also shown that an efficient implementation of a soft output max-log MAP detector for a $2\times M$ MIMO system with quadrature amplitude modulation (QAM) inputs reduces the number of candidate tests by a factor of $|\mathcal{X}|$ by rewriting the minimum metric expression in a hierarchical manner. The remaining candidate tests are performed in a recursive fashion avoiding multiplications altogether. As a result, the computational complexity for the metric calculations has been reduced by a factor of 250 for a $2\times M$ MIMO system with Gray-coded 64-ary QAM (64QAM). Furthermore, it was estimated that with 10-bit quantization of the metric component values, 64QAM, and a 2×2 MIMO system, a chip area of 0.031 mm^2 would be required for a clock frequency of 125 MHz and 65-nm complementary metal–oxide–semiconductor (CMOS) technology. For more details and applications to the IEEE 802.11n standard, see [12].

FACTOR-GRAPH-BASED DETECTOR WITH LINEAR COMPLEXITY IN THE NUMBER OF INTERFERERS

As said earlier, optimal detection by means of the VA or the BCJR algorithm works over a trellis with $|\mathcal{X}|^L$ states. Channel shortening tries to transform the original channel into a shorter one before detection. In addition to channel shortening, other approaches to complexity reduction have been based on a reduced search over the original trellis or on a search over a reduced trellis obtained from the original one through a partial representation of the symbols in the trellis state definition. Many

$$\begin{aligned}\|\boldsymbol{r}-\boldsymbol{H}\boldsymbol{x}\|^2 &= (\boldsymbol{r}-\boldsymbol{H}\boldsymbol{x})^*(\boldsymbol{r}-\boldsymbol{H}\boldsymbol{x}) \propto -2\mathcal{R}\{\boldsymbol{r}^*\boldsymbol{H}\boldsymbol{x}\}+\boldsymbol{x}^*\boldsymbol{H}^*\boldsymbol{H}\boldsymbol{x} \\ &= \begin{pmatrix} -2\mathcal{R}\{r_1^*H_{1,1}+r_2^*H_{2,1}\} \\ 2\mathcal{I}\{r_1^*H_{1,1}+r_2^*H_{2,1}\} \\ -2\mathcal{R}\{r_1^*H_{1,2}+r_2^*H_{2,2}\} \\ 2\mathcal{I}\{r_1^*H_{1,2}+r_2^*H_{2,2}\} \end{pmatrix}^T \begin{pmatrix} \mathcal{R}\{x_1\} \\ \mathcal{I}\{x_1\} \\ \mathcal{R}\{x_2\} \\ \mathcal{I}\{x_2\} \end{pmatrix} + \begin{pmatrix} |H_{1,1}|^2+|H_{2,1}|^2 \\ |H_{1,1}|^2+|H_{2,1}|^2 \\ |H_{1,2}|^2+|H_{2,2}|^2 \\ |H_{1,2}|^2+|H_{2,2}|^2 \end{pmatrix}^T \begin{pmatrix} \mathcal{R}\{x_1\}^2 \\ \mathcal{I}\{x_1\}^2 \\ \mathcal{R}\{x_2\}^2 \\ \mathcal{I}\{x_2\}^2 \end{pmatrix} \\ &\quad + \begin{pmatrix} 2\mathcal{R}\{H_{1,1}^*H_{1,2}+H_{2,1}^*H_{2,2}\} \\ -2\mathcal{I}\{H_{1,1}^*H_{1,2}+H_{2,1}^*H_{2,2}\} \\ 2\mathcal{I}\{H_{1,1}^*H_{1,2}+H_{2,1}^*H_{2,2}\} \\ 2\mathcal{R}\{H_{1,1}^*H_{1,2}+H_{2,1}^*H_{2,2}\} \end{pmatrix}^T \begin{pmatrix} \mathcal{R}\{x_1\}\mathcal{R}\{x_2\} \\ \mathcal{R}\{x_1\}\mathcal{I}\{x_2\} \\ \mathcal{I}\{x_1\}\mathcal{R}\{x_2\} \\ \mathcal{I}\{x_1\}\mathcal{I}\{x_2\} \end{pmatrix}. \end{aligned} \tag{19}$$



[FIG2] Three sections of the FG corresponding to (25), for the case $L = 2$.

papers have investigated these approaches (as an example, see [13] and the references therein). They all have in common that they work on the Forney observation model. An attempt with scarce success has been tried in [14] to adapt some of them to the Ungerboeck observation model. The reason is related to the fact that the partial metric of the VA algorithm does not have, in the case of the Ungerboeck observation model, a probabilistic meaning.

However, the Ungerboeck model allows a different approach to complexity reduction [15]. Neglecting the factors in (9) that are irrelevant for detection, we can write

$$p(\boldsymbol{r}\,|\,\boldsymbol{x}) \propto \prod_k \exp\left(\frac{2}{N_0}\mathcal{R}\left\{y_k x_k^* - \frac{1}{2}|x_k|^2 g_0 - x_k^* \sum_{\ell=1}^{L} g_\ell x_{k-\ell}\right\}\right) \tag{21}$$

$$= \prod_k F_k(x_k) \prod_{\ell=1}^{L} I_{k,k-\ell}(x_k, x_{k-\ell}) \tag{22}$$

having defined

$$F_k(x_k) = \exp\left(\frac{2}{N_0}\mathcal{R}\left\{y_k x_k^* - \frac{1}{2}|x_k|^2 g_0\right\}\right) \tag{23}$$

$$I_{k,k-\ell}(x_k, x_{k-\ell}) = \exp\left(-\frac{2}{N_0}\mathcal{R}\left\{g_\ell x_k^* x_{k-\ell}\right\}\right). \tag{24}$$

Thus, the joint a posteriori probability of the transmitted symbols can be factorized as

$$P(\boldsymbol{x}|\,\boldsymbol{r}) \propto \prod_k P_k(x_k) F_k(x_k) \prod_{\ell=1}^{L} I_{k,k-\ell}(x_k, x_{k-\ell}), \tag{25}$$

where $P_k(x_k)$ is the a priori probability of symbol x_k.

The factorization (25) can be visualized through an FG; an example is given in Figure 2. In this graph, variable and factor nodes are represented through circles and squares, respectively. An edge connects a variable node x_k with a factor node if and only if that variable is an argument of the factor corresponding to that factor node. In the figure, we used dashed lines to represent edges involving nodes not explicitly represented in the graph. The meaning of bold edges will be explained below. Note that, when $g_\ell = 0$, the factor $I_{k,k-\ell}$ is equal to one and can thus be dropped from the factorization (25). In practice, the node $I_{k,k-\ell}$ must be included in (25) only when $g_\ell \neq 0$, i.e., only when x_k and $x_{k-\ell}$ interfere with each other.

The factorization (25) is exact, since no approximation was adopted in its derivation. On the other hand, the marginalization of (25), required for computing the a posteriori probabilities $\{P(x_k\,|\,\boldsymbol{r})\}$, cannot be exactly carried out by applying the SPA to the FG in Figure 2, since it contains cycles. One of these cycles is indicated in the figure in bold. It is easy to prove that the FG corresponding to (25) cannot contain any cycle of length lower than six, irrespective of the number of symbols that interfere with each other. In fact, being factor nodes of at most degree two, the necessary and sufficient condition for the arising of a cycle of length four is to have two factor nodes of degree two connected to the same couple of variable nodes, and this is clearly not possible, by definition of $I_{k,k-\ell}$. Hence, in this case, the SPA may lead to favorable results since it is generally expected to provide a good approximation of the exact marginalizations when the length of the cycles is at least six.

The algorithm resulting from the application of the SPA to the described FG is iterative and has a complexity per iteration, which is linear in the number of interferers. This is related to the adopted factorization having the appealing property that nodes $I_{k,k-\ell}(x_k, x_{k-\ell})$, whose number linearly increases with the number of interferers, have degree two (i.e., they have two edges) irrespective of the number of interferers.

CONCLUSIONS

Although the Ungerboeck and the Forney observation models are equivalent whenever optimal ML receivers are employed, the two models have different properties with suboptimal receivers. Almost all reduced-complexity receivers take the Forney model as the basis for complexity reduction. The best example is that it took more than three decades from the time that the Ungerboeck model was published until a BCJR was derived for it. Thus, no reduced-complexity Ungerboeck-based BCJRs could have been researched until only recently. Meanwhile, the amount of research devoted to reduced-complexity Forney-based BCJRs is impressive. We believe that many algorithms can benefit from being implemented in the Ungerboeck model and that there is much to gain if the awareness of the model is increased. As a step in this direction, we have discussed three examples where the key building block is the Ungerboeck model.

AUTHORS

Fredrik Rusek (fredrik.rusek@eit.lth.se) is an associate professor in the Department of Electrical and Information Technology at Lund University, Sweden.

Giulio Colavolpe (giulio.colavolpe@unipr.it) is an associate professor in the Dipartimento di Ingegneria dell'Informazione at the University of Parma, Italy.

Carl-Erik W. Sundberg (cews@ieee.org) is the president and chief scientist at SundComm, Sunnyvale, California. From 1981 to 2000, he was a distinguished

member of technical staff at Bell Laboratories, Murray Hill, New Jersey.

REFERENCES

[1] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, no. 5, pp. 624–636, May 1974.

[2] G. D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and R. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[4] G. Colavolpe and A. Barbieri, "On MAP symbol detection for ISI channels using the Ungerboeck observation model," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 720–722, Aug. 2005.

[5] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[6] J. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.

[7] C. Douillard, M. Jézequel, and C. Berrou, "Iterative correction of intersymbol-interference: Turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sept. 1995.

[8] P. Roberston, E. Villebrun, and P. Hoeher, "Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding," *Eur. Trans. Telecommun.*, vol. 8, no. 2, pp. 119–125, Mar./Apr. 1997.

[9] C. Kuhn and N. Goertz, "A low-complexity path metric for tree-based multiple-antenna detectors," in *Proc. IEEE Int. Conf. Communications*, May 2007, pp. 1024–1029.

[10] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, vol. 52, no. 9, pp. 1541–1562, Nov. 1973.

[11] F. Rusek and A. Prlja, "Optimal channel shortening of MIMO and ISI channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 810–818, Feb. 2012.

[12] N. Graef, J. Hammerschmidt, and C.-E. W. Sundberg, "A low-complexity max-log-MAP detector," *IEEE Trans. Commun.*, vol. 57, no. 8, pp. 2251–2254, Aug. 2009.

[13] A. Prlja and J. B. Anderson, "Reduced-complexity receivers for strongly narrowband intersymbol interference introduced by faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2591–2601, Sept. 2012.

[14] A. Hafeez and W. E. Stark, "Decision feedback sequence estimation for unwhitened ISI channels with applications to multiuser detection," *IEEE J. Select. Areas Commun.*, vol. 16, no. 9, pp. 1785–1795, Dec. 1998.

[15] G. Colavolpe, D. Fertonani, and A. Piemontese, "SISO detection over linear channels with linear complexity in the number of interferers," *IEEE J. Select. Topics Signal Processing,* vol. 5, pp. 1475–1485, Dec. 2011.

[SP]

[special **REPORTS**] (continued from page 15)

consideration. The separation of sea level and land motion change is a matter of great importance for global change research, Löfgren says. "How much does the sea level change in different parts of the world and what are the causes of this change?"

The researchers note that existing coastal GNSS stations, installed primarily for the purpose of measuring land movements, can be easily adapted to make sea level measurements. "We have successfully tested a method where only one of the antennas is used to receive the radio signals," Löfgren says. "That means that existing coastal GNSS stations—there are hundreds of them all over the world—can also be used to measure the sea level."

Löfgren regards signal processing as essential to his research. "What I want to do is to convert my GNSS measurements into measurements of sea level in the most accurate way possible," he remarks. "Most of the signal processing is more or less standard in the GNSS world, but I have applied it on a new and different data set."

For the two-antenna technique, Löfgren determines the vertical distance between the upward-looking and the downward-looking antenna (the downward-looking antenna will appear to be a virtual antenna below the sea level, since the reflected signal will travel an additional path compared to the direct signal). "The signal processing is done by analysis of the phase of the recorded signals," he says. "An observational model is set up for the difference in recorded phase between the two antennas (incorporating clock differences in the receivers, differences in geometry and differences in the phase ambiguity parameter), and it is then fitted in a least squares sense to the phase observations."

For the one-antenna technique, a different type of signal processing is applied. "The interference between the direct and the reflected signals can be seen as oscillations in the signal-to-noise ratio (SNR) observable," Löfgren says. "With the assumption of a horizontal non-moving sea level, the frequency of these oscillations is constant with respect to the sine of the satellite elevation angle." This means that the oscillations first need to be found and extracted from the data. Next, the oscillations' frequency content (with respect to the sine of the satellite elevation angle instead of the usual time) should be found either by Fourier transform or a Lomb-Scargle periodogram (LSP), Löfgren says. Finally, the main oscillation frequency must be converted to the distance between the antenna and the reflection point, which is directly proportional to the sea level.

"In both the one- and two-antenna methods, the actual installations that measures reflected signals are already set up," Löfgren says. This means that the geodetic GNSS receivers are first applying some kind of signal processing when they record the satellite signals. "What I am using as techniques are least squares analysis [for the] two-antenna technique, and LSP [for the] one-antenna technique."

For the project's next step, the researchers are looking toward developing multi-GNSS solutions, possibly even combining GPS and GLONASS signals together to increase the number of observations in a combined phase delay analysis, providing more accurate sea level estimates. The combination of GPS and GLONASS for SNR analysis is expected to increase the temporal resolution of the corresponding sea level results.

After that step is accomplished, the goal will be to use multi-GNSS, multifrequency, phase delay, and SNR analysis in a filter approach. "Doing so, we expect that it will be possible to derive continuous and accurate absolute GNSS sea level time series in a wide range of wind speeds," Löfgren says.

AUTHOR

John Edwards (jedwards@johnedwardsmedia.com) is a technology writer based in the Phoenix, Arizona, area.

[SP]