Turbo Multiuser Detection: An Overview

H. Vincent Poor
Department of Electrical Engineering
Princeton University
Princeton, NJ 08540
e-mail: poor@princeton.edu

Abstract — Turbo multiuser detection refers to joint channel decoding and multiuser detection using an iterative exchange of soft information between the two processes. This technique has been examined by several authors in recent years, with very promising results. This paper provides a brief introduction to this area.

I. INTRODUCTION

Multiuser detection refers to the detection of data from multiple users when observed in a non-orthogonal multiplex. This problem arises naturally, for example, in code-division multiple-access (CDMA) systems using non-orthogonal spreading codes. It also arises in orthogonally multiplexed wireless channels, such as TDMA channels, due to effects such as multipath or non-ideal frequency channelization, and in wireline channels such as digital subscriber lines (DSLs) in which crosstalk is a major impairement [7]. The basic idea of multiuser detection is to exploit the cross-correlations among the signals to be demodulated in order to improve the data detection process.

Error control coding is, of course, ubiquitous in wireless and other impaired channels. Similarly to multiuser detection, the decoding of error-control coding exploits the dependencies among successive channel symbols to improve the detection of a single stream of data symbols. Both multiuser detection and channel decoding typically involve very complex optimal algorithms, and so complexity issues often dominate the study of these problems. Notable among coding techniques with this problem are parallel and serially concatenated codes separated by interleavers, which have been shown to offer considerable performance improvement over traditional codes, exhibiting near-Shannon-limit performance in many cases. However, although the optimal decoding of such codes is of particularly high complexity, iterative or "turbo" decoding algorithms that involve the iterative exchange of soft information between constituent decoders (separated by interleavers/de-interleavers) have been shown to be very effective approximations to optimal decoding. These ideas are exposed for example, in [4, 5, 6, 12, 13, 14, 17].

Many wireless communication systems, such as the IS-95 cellular telephony system and its third-generation descendants, involve both error-control coding and non-orthogonal multiplexing. A typical configuration is a convolutional encoder mapping data symbols into channel symbols, followed by an interleaver, and finally a CDMA modulator for the channel symbols, as shown in Fig. 1. (Interleaving refers to the permutation of the time order of the symbols.) In this paper, we will focus on this model, although other applications can also fit within the formalism discussed here. One can view the configuration of Fig. 1 as a serially concatenated code, in which the CDMA spreading code is the inner code, and the convolutional code is the outer code. A traditional way of decoding this concatenation is to first demodulate the CDMA signals (using either a conventional matched-filter detector, or a multiuser detector), and then to follow this demodulator by a de-interleaver and a channel decoder.

To seek optimality in such a situtation, one could replace this traditional configuration with an overall optimal demodulator/decoder that uses an optimal (say maximum-likelihood or minimum-error-probability) mapping from the received signal to the original data symbols. The complexity of such a system is potentially quite high. This complexity can be mitigated however, by appealing to the turbo principle for decoding concatenated codes noted above. In particular we can reduce the complexity of joint decoding and multuser detection by an iterative exchange of soft information between multiuser detection and channel decoding, iterating until some kind of convergence is reached. Like turbo decoding, this iterative approach to joint multiuser detection and channel decoding shows considerable promise for achieving very good error-probability performance (close to the single-user bound).

The purpose of this paper is to give a brief introductory review of the basic ideas behind such turbo multiuser detection.

The rest of this paper is organized as follows. First, in Section II, we give a very brief overview of basic multiuser detection and (convolutional) channel-decoding techniques, as a prelude to our discussion of the combination of these tasks. Then, in Section III, we discuss the problem of combined multiuser detection and channel decoding, noting that the complexity of optimal algorithms for this purpose is prohibitive in most cases of practical interest. We then, in Section IV, discuss a low-complexity turbo multiuser detection algorithm developed by Wang and the author in [31]. And finally, in Section V we provide some concluding remarks.

II. MULTIUSER DETECTION AND THE DECODING OF CONVOLUTIONAL CODES

Consider the reception a rate-R-coded multiple-access communication signal of the following form

$$r(t) = \sum_{k=1}^{K} \sum_{i=1}^{B} b_{k,i} l(d_{k,i}) p_{k}(t-iT-\tau_{k}) + \sigma \eta(t), \quad -\infty < t < \infty,$$

(1)

where:

- $K$ is the number of users active in the channel,
- $B$ is the number of channel symbols per user in a received frame to be processed,
- $T$ is the per-user channel symbol interval (so, $1/T$ is the per-user signaling rate)

---

1This work was prepared under the support of the National Science Foundation under Grant CCR-99-80590.
III. Iterative Joint Multiuser Detection and Decoding: Turbo MUD

In the preceding section, we noted that the problems of multiuser detection and convolutional coding can each be solved individually by dynamic programs whose complexity is \( O(2^K) \) in the uncoded multiuser case and is \( O(2^K) \) in the coded single-user case. We now consider the situation in which we have multiple \( (K > 1) \) users, each of which bears a convolutionally coded data stream with constraint length \( \nu \). In [10], Giorgetti and Wilson show that optimal (detection and) decoding in this problem essentially combines the complexity of the constituent problems, to yield a dynamic program with \( O(2^K \nu) \) complexity. This complexity would typically be too high for most applications, since the constraint length of the code would normally be chosen to meet the limits of the receivers processing capabilities. Amplifying this constraint length by a factor of \( K \) in the exponent will push the processing capability well beyond its limits. (Some suboptimal receivers for joint MUD and decoding are also considered by Giorgetti and Wilson in [11].)

Like turbo-coded systems, this complexity can be mitigated by making use of the turbo decoding principle of iterating between algorithms for the constituent problems, and exchanging soft information between iterations. This idea has been explored by Moher in [18]. The basic building blocks of a turbo multiuser detector are a soft-input/soft-output (SISO) multiuser detector and a bank of single-user SISO channel decoders, as shown in Fig. 2. The role of each of these algorithms is to compute posterior probabilities of the channel symbols based on prior prior probabilities and on the corresponding signal structure. That is, the SISO multiuser detector uses prior symbol probabilities and the CDMA signaling structure to compute posterior symbol probabilities conditioned on the observations. Similarly, the SISO channel decoders use prior symbol probabilities and the structure imposed by the channel code to compute posterior symbol probabilities. (Of course, the SISO decoders also compute posterior data symbol probabilities, which will ultimately yield the overall output of the combined algorithm.)

The turbo multiuser detector begins with a SISO multiuser detector applied to a frame of \( B \) channel symbols (\( B \) is assumed to be equal to the interleaver length). This detector particularly computes a posteriori probabilities, conditioned on the observations \( y \), for each of the channel symbols of each of the users; that is, for each element of the vector \( b \). The first set of posterior probabilities is based on the prior assumption that the channel symbols are drawn uniformly from \( \{ -1, +1 \}^K \); that is, that the channel symbols are independent and identically distributed (i.i.d.) \( \pm 1 \) random variables. Although this assumption is not correct due to the channel coding (which correlates the channel symbols), it serves as a useful approximation for initializing the algorithm because the interleavers at the transmitter serve to decorrelate the symbols as they appear at the input to the CDMA modulator.

The posterior probabilities computed by the SISO MUD will then be used as prior probabilities in the next step of the algorithm, which makes use of the bank of single-user channel decoders. Before applying channel decoding, however, the

---

\( d_k \) is a set of RB data symbols being transmitted by user \( k \),

\( \mathbf{d}_k \) (with components \( d_{k,i}, i = 1, \ldots, B \)) is the vector of channel symbols obtained by encoding \( d_k \),

\( p_k \) is the signaling waveform of user \( k \),

\( \tau_k \in [0, T] \) is the delay with which user \( k \)'s signal is received,

\( n(t) \) is a white Gaussian process with unit intensity, and

\( \sigma \) is the noise intensity.

For the sake of exposition we assume that the data and channel symbols take binary \(( \pm 1 \) \) values, although this is easily relaxed to include any finite alphabet. We also assume that the observations are real-valued, although again this assumption is not essential to any of what follows.

We would like to make inferences about the set of data symbol vectors \( \mathbf{d}_1, \ldots, \mathbf{d}_K \), which contain a total of \( KB \) symbols. A sufficient statistic for such inferences is formed by the set of \( KB \) matched-filter outputs

\[
y_k(t) = \int_{-\infty}^{\infty} r(t) p_k(t-iT-\tau_k) \, dt, \quad k = 1, \ldots, K, \quad i = 1, \ldots, B.
\]

Organizing these observables into a vector \( \mathbf{y} \in \mathbb{R}^{KB} \) by sorting them first by symbol number and then by user number, the model (1) can be rewritten as a linear model

\[
\mathbf{y} = \mathbf{H} \mathbf{b} + \mathbf{N}(0, \sigma^2 \mathbf{H}),
\]

where \( \mathbf{b} \in \{-1, +1\}^{KB} \) denotes a vector containing the channel symbols \( \{ b_{k,i}(\mathbf{d}_k) \} \) sorted conformally with \( y \), and where \( \mathbf{H} \) denotes a matrix of cross-correlations

\[
H_{m,n} = \int_{-\infty}^{\infty} p_k(t-iT-\tau_k) p_l(t-jT-\tau_k) \, dt
\]

with the indices \((k,i)\) and \((\ell,j)\) corresponding in the model (1) to the indices \( n \) and \( m \), respectively, in the vector \( \mathbf{y} \). The term \( \mathbf{N}(0, \sigma^2 \mathbf{H}) \) denotes a noise term having the multivariate Gaussian distribution with zero mean and covariance matrix \( \sigma^2 \mathbf{H} \).

Multiuser detection (MUD) and channel decoding are problems of sequence detection, which involve mapping the vector \( \mathbf{y} \) into estimates \( \hat{\mathbf{d}}_1, \ldots, \hat{\mathbf{d}}_K \), of the data symbol vectors of the various users. When this mapping is chosen to satisfy optimality criteria such as maximum likelihood (ML) or maximum a posteriori probability (MAP), the resulting complexity is nominally quite high - \( O(2^{KB}) \). Fortunately, these problems typically can be solved with much lower complexity via dynamic programming [21]. For example, in the case of a single user \(( K = 1) \), non-dispersive (\( p_1(t) = 0, \forall t \notin [0, T] \)) channel with convolutional coding, the complexity of these optimal decoders reduces to \( O(2^\nu) \). The corresponding dynamic program for the single-user \(( K = 1) \) case is given in the ML case by the Forney [9] or Ungerboeck [26] maximum-likelihood sequence detector, and for the multiuser \(( K > 1) \) non-dispersive \(( \Delta = 1) \) case by the Verdu ML [27] or the MAP [28] multiuser detector.
symbols must be de-interleaved to return them to their correct
order for decoding. This de-interleaving has the approximate
effect of removing any correlations that are introduced into the
channel symbols by conditioning on the observations y in the
SISO MUD. Thus, after SISO MUD and de-interleaving, the
channel symbols can again be assumed to be independent of
one another, but now having marginal (i.e., individual) proba-
bility distributions determined by the probabilities computed
by the SISO MUD. This probability model becomes the prior
probability model used by the SISO channel decoders, which
compute (via, say, the BCJR algorithm) corresponding poste-
rior probabilities for both the channel and data symbols.
These posterior probabilities for the data symbols could, at
this point, be used to MAP-decode the data symbols.
This would correspond to a conventional receiver approach
based on MUD followed by decoding. However, a more pow-
erful receiver results by re-interleaving the channel symbols
at the output of the decoders and returning to the SISO
MUD, now using as a prior distribution the posterior channel-
symbol probabilities computed by the SISO decoders. The
SISO MUD then refines its estimates of the posterior prob-
abilities of the symbol probabilities, and hands them back
to the channel decoders after de-interleaving again. This pro-
cess of soft-information exchange between the SISO MUD and
the SISO decoders can continue until the posterior channel-
symbol probabilities converge to stable values, at which point
the data symbols can be MAP decoded via the data-symbol
posterior probabilities computed on the last application of the
SISO decoding algorithm.
From this description, it can be seen that the interpretation
of the multiuser detector as a posterior-probability calculator
is an essential philosophic underpinning of this approach.
However, unlike the case with turbo decoding, in which the
complexity of the constituent decoders is controlled by the sys-
tem designer, the complexity of the SISO multiuser detector
used in this turbo multiuser detector is dependent on the num-
ber of users in the channel and is thus beyond the designer’s
immediate control. Thus, although the $O \left( 2^K \right)$ complexity
of optimal joint detection and decoding noted in [10] is reduced
to $O \left( 2^R \right) + O \left( 2^K \right)$ via the turbo principle, the second term
in this complexity order is prohibitive for most applications.
Because of this complexity issue, some simpler techniques,
in which the multiuser detection component of such an iter-
ative scheme is replaced by simpler suboptimal algorithms such
as interference cancellers, etc., have been considered by sev-
eral authors. (See, e.g., [1, 2, 19, 20, 23, 24, 25].) Moreover,
an alternative approach based on an approximate posterior-probability calculator that significantly simplifies the SISO
MUD has been developed by Wang and the author in [31],
and this approach is described briefly in the following section.

IV. A LOW-COMPLEXITY TURBO MULTIUSER
DETECTOR
The basic difficulty with the turbo multiuser detector
described in the preceding section is the $O \left( 2^K \right)$ complexity
of the MAP multiuser detection stages. A considerable amount
of research has been devoted to the development of subopti-
mal multiuser detectors that mitigate the complexity of op-
timal multiuser detection (see, e.g., [27]). One well-studied
family of suboptimal multiuser detector consists of the linear
multiuser detectors, which can be described briefly as follows.
The sufficient statistic y obeys the linear model (3), and
multiuser detection (and decoding and equalization as well)
can be viewed as the fitting of this model to the observa-
tions. The complexity of these problems comes from the fact
that the elements of the vector b take values in a finite alphabet.
Without this constraint, the fitting of linear models such as (3)
is of relatively low complexity. The basic idea of linear
multiuser detection is to take advantage of this low
complexity of unconstrained linear model-fitting by first esti-
mating b in (3) as if it were a vector with real components,
and then to project these real estimates onto the finite alphabet
of the actual symbols. This, of course, will not yield
maximum-likelihood or MAP symbol decisions, but it often
works quite well. Key examples of linear multiuser detectors
are the decorrelating (or zero-forcing) detector which produces
its linear estimates by simply inverting the channel; i.e.,
$$\hat{b} = \text{sgn}\{H^{-1}y\} \quad (5)$$
and the linear minimum-mean-square-error (MMSE) detector,
which detects b via
$$\hat{b} = \text{sgn}\left\{\left(\mathbf{H} + \sigma^2 I\right)^{-1} y\right\} \quad (6)$$
where I denotes the $KB \times KB$ identity matrix. The latter
detector uses, as its linear estimation stage, the linear MMSE
estimator of b given y in (3) under the assumption that the
symbols have a prior distribution under which they are uncor-
related with zero means; namely, $\left(\mathbf{H} + \sigma^2 I\right)^{-1} y$. The stan-
ard linear MMSE detector of (6) can be modified to account
for a prior-distribution with non-zero mean, which results in
the linear estimator
$$\left(\mathbf{H} + \sigma^2 \mathbf{C}^{-1}\right)^{-1} \left[ y - \mathbf{H} \hat{b} \right] \quad (7)$$
where C and $\hat{b}$ denote, respectively, the prior covariance
and mean of b. The elements of $\hat{b}$ are thus given by
$$\hat{b}_{k,i} = 2p_{k,i} - 1, \quad (8)$$
where $p_{k,i}$ denotes the prior probability that $b_{k,i} = 1$; and C
is a diagonal matrix with diagonal elements
$$C_{n,n} = 4p_{k,i} \left(1 - p_{k,i}\right) \quad (9)$$
(Here, as in (4), the indices $(k, i)$ correspond to the model (1)
to the index n in the vector y.)
Although linear detectors are of considerable interest in
the implementation of practical multiuser detection (and equal-
ization) systems, they do not immediately appear to be use-
ful in the context of turbo multiuser detection since they are
based on linear regression type criteria rather than on poste-
rior probability computation. So, for example, while the linear
MMSE detector allows for the incorporation of prior
channel-symbol probabilities via (7), it still does not seem to
be amenable to posterior-probability calculation. However, it
happens that the linear MMSE detector can, in fact, be used
as an approximate posterior-probability calculator. This is
due to the property, exposed by Verdú and the author in [22],
that the residual error in the linear MMSE estimator used
by the MMSE detector is approximately Gaussian. From this
property, we can obtain posterior probability estimates for
the channel symbols conditioned on the output of the linear
MMSE transformation (7) straightforwardly via Bayes’ for-
mula.
The application of this idea, which is explored more fully
in [31], leads to excellent performance with only a few cycles
through the turbo algorithm. Figure 3 shows a typical performance result, from which it can be seen that near-single-user performance can be achieved quite easily when there is sufficient signal-to-noise ratio for the initial SISO MUD to gain useful information about the channel symbols. (In this case, the SNR required is quite small - only a few dB.) Further approximations to this detector with even lower complexity have also been developed in [31], with comparable performance results.

V. CONCLUSION
This paper has briefly introduced the problem of turbo multiuser detection, which allows for low-complexity joint channel decoding and multiuser detection. This area is a very active one at present, as is evidenced by the special session in which this paper appears. Recent contributions to this area include, for examples, its application in turbo-coded CDMA systems [32] and in space-time coded systems [16], and the introduction of adaptivity [15, 30] into the SISO multiuser detector. The reader is referred to the remaining papers in this special session for note of further recent contributions.

REFERENCES
Figure 1: A Transmitter Configuration with Convolutional Coding and CDMA Modulation.

Figure 2: General Structure of Turbo Multiuser Detection.

Figure 3: Performance Simulation of the Turbo Multiuser Detector Developed in [31](Synchronous transmission of $K = 4$ users with equal inter-user correlations of $\rho = 0.7$, rate-1/2 convolutional encoding with constraint length 5, and interleaver length $B = 128$.)