A Method for Increasing Downlink Capacity by Coded Multiuser Transmission with a Base Station Diversity Array

Stephen J. Grant[1] and James K. Cavers[2]

Abstract— **In this paper, a simple structure is proposed for supporting multiple intracell users in the same time/frequency slot in the downlink using a transmit antenna array at the base station and joint detection at the single-antenna mobile receivers. To support multiple users, the system requires the simultaneous transmission of several successive code symbols from different diversity antennas. As a consequence, a single multi-antenna channel usage spans several trellis transitions — an unconventional situation which is handled with a novel merged trellis. Using the merged trellis, the optimal decoder is identified, and an analytical expression for the average bit-error rate is developed, based on soft-decision joint decoding at the mobiles. Unusual behaviour is demonstrated in terms of diversity order: as the number of transmit antennas increases due to an increasing number of users, the diversity order actually decreases due to the simultaneous transmission of successive code symbols. Even with the loss in diversity order, the method provides a major increase in downlink capacity while maintaining good performance for all users at low signal-to-noise ratios with moderate computational load.**

I. Introduction

Recently, much research has focused on increasing the uplink capacity of TDMA systems by allowing several intracell users to share the same time/frequency slot, i.e., allow reuse within cell. To distinguish the cochannel signals, either an interference cancellation or joint detection technique is used in combination with an antenna array at the base station receiver [1]–[4]. Less attention, however, has been paid to the downlink, where achieving diversity reception for the single-antenna mobile receivers is important.

Two existing techniques for achieving diversity in the downlink include adaptive transmission [5] and space-time codes [6],[7]. Adaptive transmission is a multiuser technique; however, it requires extremely accurate estimates of the downlink channel gains which are typically difficult to obtain. Unlike adaptive transmission, space-time codes can be used to obtain diversity in the downlink without knowledge of the downlink channels at the transmitter. However, space-time codes have been designed for single-user systems only.

In this paper an alternative structure is proposed in which the users' bit sequences are multiplexed together, but standard convolutional coding and interleaving of the composite bit sequence is employed (rather than space-time coding) in order to provide temporal diversity. Like space-time codes, the transmitter does not require estimates of the downlink channels. Several successive code-symbols are then transmitted simultaneously on the different antennas to reduce the bandwidth required to transmit the high-rate coded sequence. The simultaneous transmissions are then jointly detected at each mobile receiver, and the data from other users is discarded.

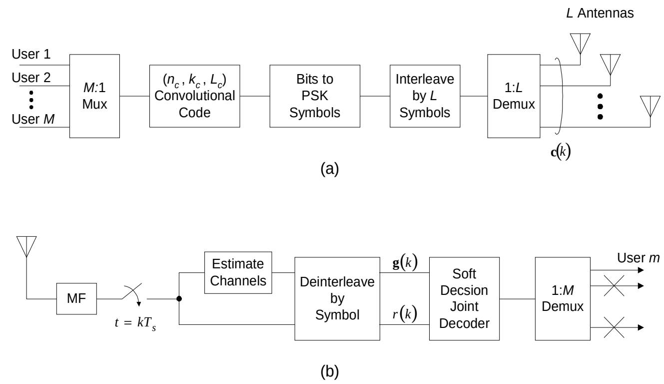

Fig. 1. Structure of (a) base station transmitter, and (b) mobile receiver for user m.

The proposed structure is similar in spirit to BLAST [8], in which the multiple transmit antennas are used in such a way to gain a rate increase for a single user; however, BLAST requires as many receive antennas as transmit antennas due to its reliance on ZF or MMSE combining to distinguish the cochannel transmit signals. Consequently, it is unsuitable for a mobile environment where the handsets have only one antenna.

II. System and Signal Models

Fig. 1(a) shows the proposed structure of the base station transmitter. The M users' bit streams are first multiplexed into a common bit stream of rate MR_b bits per second, where R_b is the rate of each individual user. The common bit stream is then encoded using an (n_c, k_c, L_c) convolutional code, where k_c is the number of input bits per encoding interval, n_c is the number of output bits, and L_c is the constraint length of the code (measured in blocks of k_c input bits). The code rate is $R_c = k_c/n_c$. The output bits are then mapped to phase-shift keying (PSK) symbols using Grey coding. The number of points in the PSK constellation is denoted by Q. The resulting symbol sequence is interleaved keeping blocks of L symbols together, where L is the number of transmit antennas. The symbols within each length-L block are then transmitted simultaneously on the L different antennas. The vector of transmitted symbols is denoted $\mathbf{c}(k) = (c_1(k), c_2(k), \cdots, c_L(k))^T$.

The simultaneous transmission of successive code symbols slows the symbol rate on each antenna by a factor of L, thus reducing the bandwidth required to transmit the multiplexed

[1]Ericsson Inc., 7001 Development Drive, Research Triangle Park, NC 27709, Tel: (919) 472-6068, Email: stephen.grant@ericsson.com
[2]School of Engineering Science, Simon Fraser Unversity, Burnaby, B.C. V5A 1S6, Canada, Tel: (604) 291-3281, Email: cavers@sfu.ca

bit stream. In terms of the various parameters, the symbol rate on each antenna is $R_s = MR_b/N_b$ symbols/sec where

$$N_b = R_c L \log_2 Q \tag{1}$$

is the number of information bits transmitted per vector channel usage. A vector channel usage is simply the transmission of one symbol vector $\mathbf{c}(k)$. Clearly, as long as $L \geq M$, M users may be supported in the same (or less) bandwidth required by a conventional single user, single antenna system.

At the receiver shown in Fig. 1(b), the deinterleaved received sample sequence is

$$r(k) = A\mathbf{g}^T(k)\,\mathbf{c}(k) + n(k)\,. \tag{2}$$

The elements of the channel gain vector $\mathbf{g}(k)$ are i.i.d. complex Gaussian random variables with variance $\sigma_g^2 = 1/2$ that model the L independent, flat Rayleigh fading channels between the transmit array and the single-antenna mobile receiver. The scalar $n(k)$ is a complex Gaussian noise sample with variance N_o. It is assumed that the total transmit power P is divided equally among the transmit antennas, giving the transmit amplitude $A = \sqrt{2(P/L)}$. Furthermore, it is assumed that the PSK symbols in the transmit symbol vector $\mathbf{c}(k)$ are drawn from a unit-radius constellation. The receive signal-to-noise ratio (SNR) is thus

$$\Gamma = \frac{\frac{1}{2}E\left[\left|A\mathbf{g}^T(k)\,\mathbf{c}(k)\right|^2\right]}{\frac{1}{2}E\left[\left|n(k)\right|^2\right]} = \frac{P}{N_o}. \tag{3}$$

Note that the receive SNR per information bit is the standard measure used when coding is employed, and it is given by $\Gamma_b = \Gamma/N_b$. Also of interest is the per-user *transmit* SNR which is given by $\Gamma_T = N_b\Gamma_b/M$.

III. Trellis Merges

Because several successive code symbols are transmitted simultaneously from different antennas, a single vector channel usage may span several trellis transitions. For example, consider a system with two users, two antennas, a rate one half convolutional code, and QPSK modulation, i.e., $L = M = 2$, $R_c = 1/2$, and $Q = 4$. The trellis diagram for a constraint length three code is shown in Fig. 2(a).

Now consider two successive encoder input bits b_0 and b_1, and assume the encoder is initially in state 0. The two input bits correspond to two transitions in the trellis. After two transitions, the encoder may end up in states 0, 1, 2, or 3 depending on the values of b_0 and b_1. The encoder output bits along the paths to each of these four possible states are shown on the trellis diagram. Pairs of the output bits are mapped to QPSK symbols using Grey coding. With two antennas, two successive code symbols are transmitted simultaneously resulting in four possible QPSK transmit vectors as shown in Fig. 2(b).

Clearly, a single vector channel usage spans two trellis transitions. This is in contrast to the case of single user convolutional or trellis codes, in which a single trellis transition entails

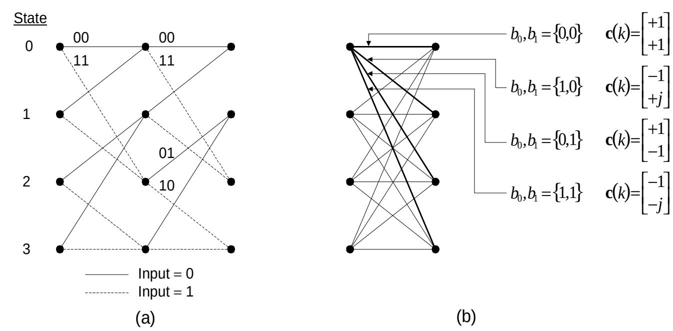

Fig. 2. (a) Trellis diagram for a (2,1,3) code, and (b) merged trellis for the example of a system with two users, two antennas, and QPSK modulation.

one or more channel usages. To handle this unusual situation in both analysis and decoder implementation, the concept of a "merged trellis" is introduced in which the trellis is modified such that a single vector channel usage spans only one transition. Fig. 2(b) illustrates the merged trellis for the above example. As can be seen, one transition in the merged trellis enumerates all possible successor states after two transitions in the original trellis. Now, instead of two branches leading into and out of each state, there are four. Associated with each branch of the merged trellis are two input bits but only one transmit symbol vector, as desired.

In order to generalize the above example, it is convenient to define the parameter N_u as the number of vector channel usages per transition in the original (un-merged) trellis, given by

$$N_u = \frac{n_c}{L\log_2 Q}. \tag{4}$$

In the above example, $N_u = 1/2$, meaning that one vector channel usage spans to two trellis transitions. The parameter N_u signals whether or not the trellis must be merged. If $N_u < 1$, then merges are required; if $N_u \geq 1$ (the conventional situation) then no merges are required.

If merges are required, and $1/N_u$ is an integer, then $1/N_u$ transitions in the original trellis must be merged into one. In the merged trellis, there are thus $2^{k_c/N_u}$ branches entering and leaving each state. Since the number of states in the merged trellis remains the same, the total number of branches per transition becomes $2^{k_c(L_c+1/N_u-1)}$ where L_c is the constraint length of the code. Furthermore, k_c/N_u input bits and only one transmit vector $\mathbf{c}(k)$ are associated with each branch. Even with the increased number of branches in the merged trellis, the decoder complexity is still much less than for a system that independently codes each user's bit sequence and transmits one user signal per antenna — an alternative for achieving reuse within cell. Unlike the proposed structure, this alternative requires a joint trellis for soft decision decoding for which the size of the state set increases exponentially with the number of users M [2].

TABLE I

NUMBER OF BITS PER USAGE N_b AND USAGES PER TRANSITION N_u FOR SEVERAL ALLOWED COMBINATIONS OF Q AND L FOR A RATE 1/2 CODE.

Q	2				4				8			
L	1	2	3	4	1	2	3	4	1	2	3	4
N_b	$\frac{1}{2}$	1	–	2	1	2	3	4	–	3	–	6
N_u	2	1	–	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	–	$\frac{1}{3}$	–	$\frac{1}{6}$

For simplicity, it is assumed in this paper that if trellis merges are required, then $1/N_u$ is an integer, and if no merges are required, then N_u is an integer. These assumptions place constraints on the allowable constellation density and number of antennas as shown in Table I. As can be seen, increasing either the constellation density or the number of antennas leads to a larger number of bits per channel usage, i.e., better spectrum efficiency. On the other hand, it leads to a smaller number of usages per transition, eventually requiring trellis merges when N_u becomes fractional.

An important property of the proposed transmitter structure is that as the parameter N_u decreases below unity (due to increasing L and/or Q), the length of the shortest error event in the merged trellis decreases. Consequently, the diversity order experienced by the receiver decreases due to fewer independent channel usages across the shortest error event. This can be inferred from Fig. 2 which corresponds to the example of $N_u = 1/2$. The shortest error event in the original trellis is three (measured in transitions), whereas in the merged trellis, the shortest error event is of length two. Notice that the merged trellis is fully connected meaning that every state may be reached from every other state. If N_u was 1/3 in the above example (e.g., if number of users was three instead of two), then the merged trellis would contain parallel transitions, implying that the shortest error event would be of length one. In general, for a code of constraint length L_c, a fully connected merged trellis occurs if $1/N_u = L_c - 1$, and parallel transitions occur if $1/N_u = L_c$. For a large constraint length, N_u has to be quite small (large L and/or Q) before the merged trellis contains parallel transitions and the length of the shortest error event is reduced to one.

The decrease in diversity order with an increasing number of antennas is unusual, since in the uplink we are used to seeing the opposite. However, this behaviour may be attributed to the fact the antenna array is being used for a different purpose: rather than for diversity combining, it is used essentially to achieve a rate increase in order to support multiple users. It is interesting to note that a special case of the proposed structure appears in [9], in which a rate increase is obtained for a single user; however, the problem of multiple trellis transitions per channel usage is not addressed, and the performance results are obtained by simulation, rather than analysis as is done here.

IV. DECODER IMPLEMENTATION

According to the ML criterion, the decoder chooses the state sequence $\mathbf{x}_i$ that maximizes the log-likelihood function $\ln P(\mathbf{r} \mid \mathbf{x}_i, \mathbf{G})$, where $\mathbf{r}$ is the sequence of deinterleaved received samples, and $\mathbf{G}$ is the deinterleaved sequence of estimated channel gain vectors obtained using a multiuser channel estimation technique such as that described in [10]. For simplicity, perfect channel state information is assumed in this paper. If $N_u < 1$, $\mathbf{x}_i$ is a sequence of states through the merged trellis. If $N_u \geq 1$, $\mathbf{x}_i$ is a sequence of states through the original (un-merged) trellis. Because the sequence of noise samples is white, the log-likelihood function may be expressed as the summation

$$\ln P(\mathbf{r} \mid \mathbf{x}_i, \mathbf{G}) = \sum_k \ln P\left[r(k) \mid \mathbf{x}_i, \mathbf{g}(k)\right] \tag{5}$$

Since $r(k)$ is conditionally Gaussian with mean $A\mathbf{g}^T(k)\,\mathbf{c}_i(k)$, the cumulative metric to be minimized is simply $\Lambda_i = \sum_k \mu_i(k)$, where the branch metric $\mu_i(k)$ is given by

$$\mu_i(k) = \left| r(k) - A\mathbf{g}^T(k)\,\mathbf{c}_i(k) \right|^2 . \tag{6}$$

Here $\mathbf{c}_i(k)$ is the transmit symbol vector corresponding to the appropriate transition in the hypothesized state sequence $\mathbf{x}_i$.

The branch metric in (6) has precisely the same form as the metric derived in [3], which considers symbol-by-symbol joint detection (rather than sequence detection) in the uplink. Consequently, in the downlink, the mobile implicitly performs joint detection of the cochannel signals from the L different transmit antennas within the receiver metric, while simultaneously performing soft decision decoding of the convolutional code. This may be done using the standard Viterbi algorithm. As shown in Fig. 1(b), after soft decision joint decoding, the composite bit sequence is demultiplexed, and the bits corresponding to other users are discarded.

V. PERFORMANCE ANALYSIS

As in other studies considering trellis decoding, e.g. [11], the bit-error rate (BER) is estimated as

$$P_b \approx \frac{1}{K} \sum_i n_i P_{2_i} \tag{7}$$

where n_i is the number of bit-errors associated with the ith error event, and P_{2_i} is the pairwise error probability corresponding to that event, i.e., the probability that the cumulative metric for the ith error event is more favourable than that for the all-zero state sequence. The variable K is the number of encoder input bits per trellis transition: for the case of $N_u < 1$, $K = k_c/N_u$; for the case of $N_u \geq 1$, $K = k_c$. Typically the infinite summation in (7) is truncated such that only error events with length less than or equal to a certain threshold are included.

The pairwise error probability in (7) is given by $P_{2_i} = P[D_i < 0]$ where $D_i = \Lambda_i - \Lambda_0$ and Λ_0 is the cumulative metric for that portion of the all-zero state sequence of length equal to the ith error event. For the all-zero state sequence, in which $\mathbf{c}_0(k) = (1, 1, \cdots, 1)$, the random variable D_i is given by the sum $D_i = \sum_k d_{ik}$ where

$$d_{ik} = \left| r(k) - A\mathbf{g}^T(k)\,\mathbf{c}_i(k) \right|^2 - \left| r(k) - A\mathbf{g}^T(k)\,\mathbf{c}_0(k) \right|^2 \tag{8}$$

and k indexes those channel usages of the ith error event for which $\mathbf{c}_i(k) \neq \mathbf{c}_0(k)$.

The pairwise error probability is found by expressing the random variable d_{ik} as a Hermitian quadratic form in zero-mean complex Gaussian random variables: $d_{ik} = \mathbf{z}^\dagger(k)\,\mathbf{Q}_i(k)\,\mathbf{z}(k)$. The vector $\mathbf{z}(k)$ is defined as

$$\mathbf{z}(k) = \left(r(k), g_1(k), g_2(k), \cdots, g_L(k)\right)^T \tag{9}$$

and the matrix of the quadratic form as $\mathbf{Q}_i(k) = \mathbf{F}_i(k) + \mathbf{F}_i^\dagger(k)$. The $(L+1) \times (L+1)$ matrix $\mathbf{F}_i(k)$ is upper triangular with zeros along the main diagonal and u, vth element in the upper triangle given by

$$\left\{\mathbf{F}_i(k)\right\}_{u,v} = \begin{cases} A\left(1 - c_{iv}(k)\right) & , \quad u = 0 \\ A^2\left(c_{iu}^*(k)\, c_{iv}(k) - 1\right) & , \quad u > 0 \end{cases} \tag{10}$$

Since the characteristic function of d_{ik} is well known [12], P_{2_i} may be evaluated easily through the characteristic function of D_i. Denoting $f_{D_i}(D)$ as the probability density function (PDF) of D_i and $\Phi_{D_i}(s)$ as the two-sided Laplace transform of $f_{D_i}(D)$, i.e., the characteristic function, the pairwise error probability is

$$P_{2_i} = \int_{-\infty}^{0} f_{D_i}(D)\, dD = L_{II}^{-1}\left\{\frac{1}{s}\Phi_{D_i}(s)\right\}\bigg|_{D=0} \tag{11}$$

where $L_{II}^{-1}\{\bullet\}$ denotes the inverse two-sided Laplace transform.

In this paper, perfect interleaving is assumed, meaning that the interleaving depth is sufficiently large such that the channel gain vectors $\mathbf{g}(k)$ from one channel usage to the next are independent across all error events considered in (7). Consequently, the d_{ik}'s are independent, and the characteristic function of D_i factors as the product $\Phi_{D_i}(s) = \prod_k \phi_{d_{ik}}(s)$ where $\phi_{d_{ik}}(s)$ is the characteristic function of d_{ik}. According to [3] and [4] which consider quadratic forms similar to d_{ik}, $\phi_{d_{ik}}(s)$ is given by

$$\phi_{d_{ik}}(s) = \frac{1}{\left(1 + s\lambda_{ik1}\right)\left(1 + s\lambda_{ik2}\right)} \tag{12}$$

where λ_{ik1} and λ_{ik2} are the two nonzero eigenvalues of the matrix product $2\mathbf{R}\mathbf{Q}_i(k)$ where the covariance matrix $\mathbf{R} = \frac{1}{2}\mathbf{E}\left[\mathbf{z}(\mathbf{k})\,\mathbf{z}^\dagger(\mathbf{k})\,|\mathbf{c}_0(\mathbf{k})\right]$ is given by

$$\mathbf{R} = N_o \begin{bmatrix} 1+\Gamma & \frac{\Gamma}{AL} & \frac{\Gamma}{AL} & \cdots & \frac{\Gamma}{AL} \\ \frac{\Gamma}{AL} & \frac{\Gamma}{A^2L} & 0 & \cdots & 0 \\ \frac{\Gamma}{AL} & 0 & \frac{\Gamma}{A^2L} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\Gamma}{AL} & 0 & 0 & \cdots & \frac{\Gamma}{A^2L} \end{bmatrix}. \tag{13}$$

Suitably modifying the results of [4] gives the desired eigenvalues as

$$\begin{Bmatrix} \lambda_{ik1} \\ \lambda_{ik2} \end{Bmatrix} = \frac{a_{ik}\left(\frac{\Gamma}{L}\right) \pm \sqrt{a_{ik}^2\left(\frac{\Gamma}{L}\right)^2 + 2a_{ik}\left(\frac{\Gamma}{L}\right)}}{2N_o} \tag{14}$$

where $a_{ik} = \sum_{l=1}^{L}\left(1 - \mathrm{Re}\left[c_{il}(k)\right]\right)$. Substituting (14) into (12) and then performing the inversion in (11) gives the pairwise error probability as desired. As in [11], the inversion may be accomplished through numerical contour integration.

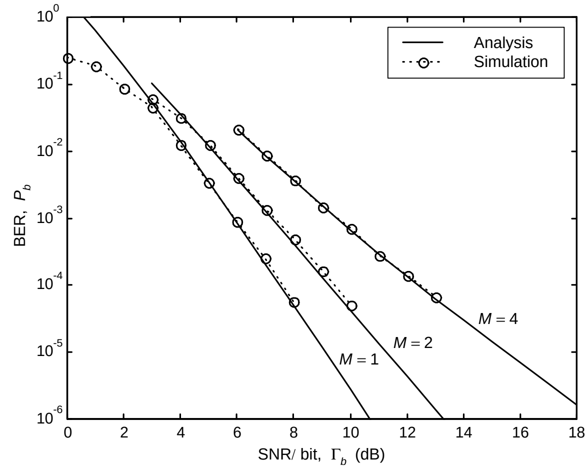

Fig. 3. BER with a (2,1,5) code and BPSK modulation.

VI. Results

In this section, the proposed downlink scheme is demonstrated using rate one half maximum free distance convolutional codes [13]. Such codes may not be optimal for this scheme; they are used simply for illustrative purposes. Although they are not considered in this paper, trellis codes may be a good choice. In the results presented here, the number of antennas L is equal to the number of users M.

Fig. 3 compares the estimated BER using (7) to the BER obtained by simulation for BPSK modulation. Clearly, the estimated BER is accurate for all BERs of interest.

In this plot, it can be seen that the performance degrades as the number of users increases; however, it is still possible to obtain good performance for all users at reasonably low SNRs using a short constraint length code. The slopes of curves on the graph confirm the loss in diversity order as additional users are added as explained in Section III.

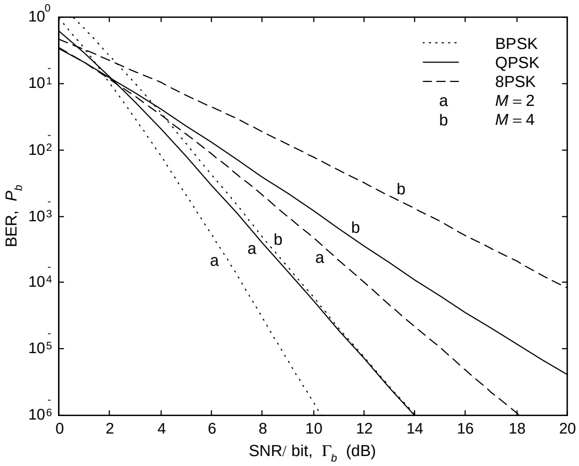

Fig. 4. BER with a (2,1,7) code and various modulation types.

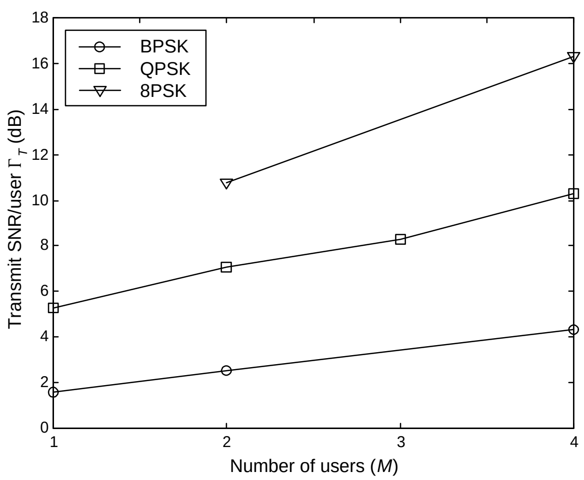

Fig. 5. Per-user transmit SNR Γ_T required to achieve a BER of 10^{-3} using a (2,1,7) code.

Fig. 4 shows the variation in BER with the constellation density Q. As predicted, the decrease in N_u with increasing constellation density causes a reduction in diversity order. With a dense constellation and a large number of users, the reduction can be significant. For example, the diversity order is only two for the four user/8PSK system. This is consistent with the fact that the merged trellis is fully connected, which occurs when $1/N_u = L_c - 1 = 6$.

Fig. 5 shows the per-user transmit SNR Γ_T using the Γ_b's from the previous graphs corresponding to a BER of 10^{-3}. Evidently, the required transmit power per user increases with the number of users, and as the constellation density grows, the increase becomes faster. However, the BPSK curve shows that four users may be supported with only 2.5 dB/user more transmit power than a single user system — not a large cost considering the system capacity gain that may be realized.

VII. Conclusions

In this paper, a simple structure is proposed for supporting multiple intracell users in the same time/frequency slot in the downlink using a transmit antenna array at the base station and joint detection at the single-antenna mobile receivers. It is shown that the simultaneous transmission of successive code symbols causes a single channel usage to span several trellis transitions. To handle this unusual situation in both analysis and decoder implementation, the novel concept of a merged trellis is introduced in which the trellis is modified such that a single channel usage spans only one transition. Using the merged trellis, the optimal decoder is identified, and an analytical expression for the average bit-error rate is derived. Unusual behaviour is demonstrated in terms of diversity order: as the number of antennas increases, due to an increasing number of users, the diversity order actually decreases due to the simultaneous transmission of successive code symbols. Even with the loss in diversity, the method provides a major increase in downlink capacity while maintaining good performance for all users at low signal-to-noise ratios with moderate computational load. For example, using a rate one half, constraint length seven code with BPSK modulation and four antennas, four users may be supported in the same bandwidth as a single-user at a BER of 10^{-3} with only 2.5 dB/user more transmit power than a single user system.

References

[1] C. Tidestav, M. Sternad and A. Ahlén, "Reuse within a cell—interference rejection or multiuser detection?," *IEEE Trans. Commun.*, vol. 47, pp. 1511–1522, Oct. 1999.

[2] K. Giridhar et al., "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Trans. Commun.*, vol. 45, pp. 473–484, Apr. 1997.

[3] S.J. Grant and J.K. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," *IEEE Trans. Commun.*, vol. 46, pp. 1038–1049, Aug. 1998.

[4] S.J. Grant and J.K. Cavers, "Further analytical results on the joint detection of cochannel signals using diversity arrays," *IEEE Trans. Commun.*, vol. 48, pp. 1788–1792, Nov. 2000.

[5] J.K. Cavers, "Single user and multiuser adaptive maximal ratio transmission for Rayleigh channels," *IEEE Trans. Veh. Tech.*, vol. 49, pp. 2043–2050, Nov., 2000.

[6] J.C. Guey, M.P. Fitz, M.R. Bell, and W.Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 527–537, Apr. 1999.

[7] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Info. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[8] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, Sep.-Oct. 1996.

[9] B.A. Bjerke and J.G. Proakis, "Multiple-antenna diversity techniques for transmission over fading channels," in *Proc. IEEE WCNC'99*, New Orleans, LA, Sep. 1999.

[10] S.J. Grant and J.K. Cavers, "Multiuser channel estimation for detection of cochannel signals," in *Proc. IEEE ICC'99*, Vancouver, Canada, Jun. 6–10, 1999.

[11] J.K. Cavers and P. Ho, "Analysis of the performance of trellis-coded modulations in Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 40, pp. 74–83, Jan. 1992.

[12] M. Schwartz, W. Bennett, and S. Stein, *Communications Systems and Techniques*. New York: McGraw Hill, 1966.

[13] J.G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.