Citation:
in *Proceedings of the 2003 Communication Networks and Services Conference (CNSR 2003)*,
(Moncton, NB, Canada), pp. 24-29, May 15-16, 2003.

Optimal Joint Blind Data and Channel Estimation with Diversity

David P. Williams
University of New Brunswick
252 Ryan Street,
Moncton, NB, Canada, E1G 2E7.
1 506 384 1728
dwlliams@nb.sympatico.ca

Brent R. Petersen
University of New Brunswick
Electrical and Computer Engineering Dept.
15 Dineen Drive, Fredericton, NB, Canada, E3B 5A3.
1 506 477 3328
b.petersen@ieee.org

ABSTRACT
For a single user scheme with frequency dependant channel, a generalized maximum likelihood sequence estimation (MLSE) algorithm was demonstrated to improve the performance over the single antenna method. It achieved almost perfect combining using blind joint data and channel estimation by means of the combining of metrics from two antenna diversity. The dominant cause of errors was due to presence of two trellises with almost identical metrics and these trellises were related; the trellises were the same sequence delayed by plus or minus one sample period. This was especially prone to happen when the estimated model order was not the true order. Another difficulty with this algorithm was the reduction in the number of choices for trellises entering a state as the trellises tended to converge to the best one; this provided a larger exhaustive search at the beginning of the algorithm than later when fading affected the channel.

Categories and Subject Descriptors
C.2.1 [Computer-Communication Networks]: Network Architecture and Design - *Wireless communication*

General Terms
Algorithms, Performance, Design.

Keywords
Wireless, Blind estimation, Trellis based.

1. INTRODUCTION
In many digital communication applications wireless is preferred as shown by the success of the cellular telephone, which allows the users to be mobile. The wireless environment is complicated by the effect of multi-path on the received signal. Multi-path effects are frequency selective fading that cause inter-symbol interference (ISI) or complete loss of the signal.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CNSR 2003 Conference, May 15-16, 2003, Moncton, New Brunswick, Canada.
Copyright 2003 CNSR Project …$5.00.

To counteract these effects the multi-path can be estimated and MLSE used. Some of the recently developed detectors are described as blind. They do not need a training sequence from the transmitter (TX), which means the receiver (RX) can adjust itself, without interrupting the transmission of data to other users, without waiting for the next training sequence or when the signal has been recovered after fading.

Optimal combination of the signals from diversity antennas was sought by extending Seshadri's [1] work to diversity with Seshadri's algorithm (DSA), using the combination of two inputs in an optimal manner to considerably reduce bit error rate.

For the user of this technique the advantages enjoyed would include these. At the transmitter there would be no need for training sequence or preamble to be transmitted. This would allow higher user data transfer rate by reducing overhead and longer battery life by reducing the time taken to transmit the data. The receiver may also recover after a deep fade during a frame, rather than wait for the next training sequence.

These advantages would come at a price; there would be a great increase in processing and for diversity an extra antenna with filtering, frequency conversion and sampling would be required. Even so errors could still be caused by a problem called slippage. Slippage was likely to occur and caused frequent errors in the received data output if the estimated channel impulse response (CIR) order was not the true one or the true CIR had small magnitudes at its beginning or end.

In section 2 the background to the combination of two inputs with the use of the sum of the squared errors was developed to drive a type of generalized Viterbi algorithm (VA). This hypothesis was tested in some experiments by the use of numerical simulation, described in section 3. The results for the non-fading multi-path CIRs for a single user scheme were shown in section 4. Since a blind receiver would not know the true CIR order, lower and higher estimated CIR orders were compared to the real path order for bit error rate (BER) performance in section 4.2. Observations about the resulting performance in terms of BER were combined into a conclusion in section 5.

2. DIVERSITY COMBINATION
Working from the trellis based blind estimator of Seshadri [1] which had been developed by extending the maximum likelihood criterion so that according to Biglieri et al. [2] for a Gaussian channel the joint probability density function, P, of the received signal vector $r = [r_1\ r_2\ \dots\ r_N]^T$ for a block of N data was

$$P\left(r\middle|CIR,d\right)=\frac{1}{\left(2\pi\sigma^{2}\right)^{N}}\exp\left(-\frac{1}{2\sigma^{2}}\sum_{n=1}^{N}\left|r_{n}-\sum_{k=0}^{L}CIR_{k}d_{n-k}\right|^{2}\right), \quad (1)$$

where d was the data vector $d = [d_1\ d_2\ \ldots\ d_N]^T$, L was the length of the true CIR and σ^2 the noise variance. Since neither d, nor σ^2 nor CIR was known one way to maximize the probability was to determine an estimate of CIR for every possible sequence. It was possible to do because every possible sequence came from a finite alphabet with known statistics, independent and identically distributed. Then the sequence was selected that minimized

$$\sum_{n=1}^{N}\left|r_{n}-\sum_{k=0}^{L}C\hat{I}R_{k}\,d_{n-k}\right|^{2}, \quad (2)$$

for each CIR, which maximized the probability, however this was an exhaustive search with computational complexity that grew exponentially with the length N. If $N = L$ there will be one CIR estimate for each surviving path of the VA search through the trellis, this was approach of Raheli et al. [3] and of Chugg and Polydoros [4]. Seshadri's [1] scheme was similar but used a generalized VA that retained $M \geq 1$ best estimates at each state of the trellis along with corresponding CIR estimates. Up to the first $N = O + \log_2(2M)$ stages in the trellis search was exhaustive, where O was the order of the estimated CIR, and from then on the limit M made the process practical. He showed good performance was achieved by setting M to four and this was used in this study of performance using two antenna diversity.

To take advantage of diversity the probability to be maximized was $P(r = r_1 \cap r_2 \mid Tx = S)$, since this was a linear model the probability could be separated, giving

$P(r = r_1 \cap r_2 \mid Tx = S) =$

$P(r = r_{11} \mid Tx = S_1)\ P(r = r_{21} \mid Tx = S_1)$

$P(r = r_{12} \mid Tx = S_2)\ P(r = r_{22} \mid Tx = S_2)$

$\ldots\ P(r = r_{1N} \mid Tx = S_N)\ P(r = r_{2N} \mid Tx = S_N)$. (3)

Assuming that the effect of all the corruption on the wanted signal was Gaussian noise the terms could be expanded in the form:-

$$P\left(r = r_{1i} \mid Tx_{1i} = S_{1i}\right) = \frac{1}{\sqrt{2.\pi}.\sigma_{n}}\exp\left(\frac{-\left(r_{1i}-S_{1i}\right)^{2}}{2.\sigma_{n}^{2}}\right), \quad (4)$$

here S_{1i} was the mean or noise-less sample at time i from passing that transmitted sequence through the CIR. This allowed the use of the metric $(r - S)^2$ for purposes of comparison. The result could conveniently be used in a recursive routine by adding together the squared errors of each sample at both antennas. This could be extended to multiple antennas by simply adding all the errors squared to get the overall metric at every sample. Although the probabilities were now separated the important point was that both antennas would have used the same data to generate their estimates, by applying it to different estimated CIRs.

A metric similar to the VA was used to choose the surviving trellises. For the case of M = 1 the metrics J would be:

$$J_{k}^{n}=\min\left\{\begin{array}{l} J_{k}^{n-1}+\left[r_{1}(n)-\sum_{i=1}^{O}C\hat{I}R1_{i}^{k}\,state_{i}\right]^{2}+\left[r_{2}(n)-\sum_{i=1}^{O}C\hat{I}R2_{i}^{k}\,state_{i}\right]^{2}; \\ J_{k}^{n-1}+\left[r_{1}(n)-\sum_{i=1}^{L}C\hat{I}R1_{i}^{k}\,state_{i}\right]^{2}+\left[r_{2}(n)-\sum_{i=1}^{L}C\hat{I}R2_{i}^{k}\,state_{i}\right]^{2}\end{array}\right\} \quad (5)$$

where J_k^n was the accumulated metric with diversity for state k and sample n, r_1 was the signal received from antenna 1 and r_2 was the signal received from the diversity antenna 2, $C\hat{I}R1$ was for the path to receiver 1, $C\hat{I}R2$ was for the path to receiver 2, k was the index to the state, and $state$ was the set of all possible memory contents in the finite impulse response (FIR) filter model of the CIR.

Extending this to M > 1 was achieved by choosing to 2 from 4 or 4 from 8 and so on. Then the M trellises with lowest J_k were chosen for each state, and the CIR estimate was done using the LMS algorithm as shown in figure 1. Once the first 10O samples were received the differential data was detected by using the oldest two data in the trellis which had the lowest accumulated metric. Using the relatively slow, but robust, LMS algorithm was justified here as the channel was found to be slow fading compared to the high data rate.

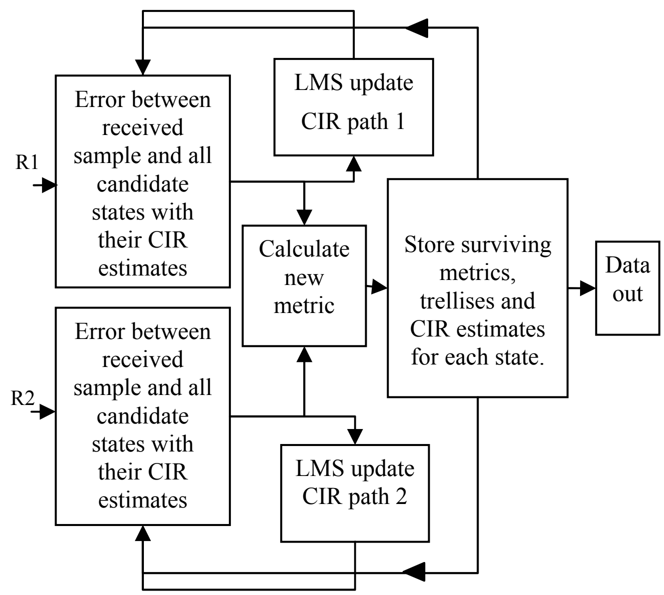

Figure 1. DSA flow diagram for two antenna diversity.

From Seshadri's paper [1] the optimal blind sequence estimator used every single sequence that could be transmitted, however the error metrics used depend on the estimation of the CIR which were inaccurate at the beginning so ideally no trellises should be dropped until the correct CIR has been converged to. This is the reason why the number of paths leading to each state should not be left at one as in the VA, by saving as many paths as possible for as long as possible the probability of still having the correct trellis is improved even though the CIR estimates may be inaccurate at the beginning. To keep all the trellises until the CIR had converged would have taken approximately 10L samples, the trellis matrix would be 10O x Q^{11O}, where Q was the number of signaling levels. This was the classic stumbling block of this technique as the size increases rapidly with L. The novel part of Seshadri's algorithm was that each of the states had more than one surviving trellis but less than Q^{10O}. As Seshadri suggested the M algorithm was employed where $M \leq Q^{O}$ and M trellises were kept for each state, along with their associated error metric and path estimate.

3. COMMUNICATION MODEL

By defining the path CIR as a FIR filter with the number of coefficients being the order of the filter *L*, the following model allowed simulation by computer, which assumed a linear channel. Figure 2 shows the model used. For the results presented two CIRs were chosen, shown in table 1.

Table 1. CIRs used in simulations.

CIR1	0.227 0.466 0.688 0.466 0.227
CIR2	-0.2 -0.5 0.7 0.36 0.2

Both CIRs were used by Seshadri [1] and Proakis [5] to represent channels that were difficult to equalize.

.

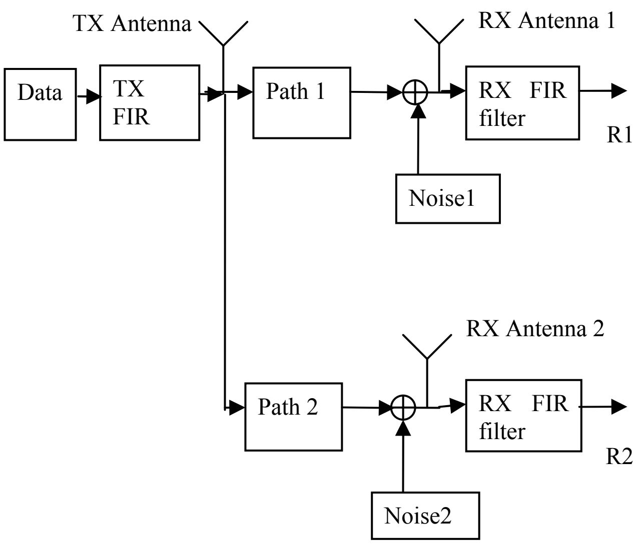

Figure 2. Model of communication link.

For the parts of a practical RX that were not modeled the assumptions used were that synchronization of frequency, phase, and symbol have been achieved, and a level control technique was used to set the power level of the signal entering each sampler to avoid weighting the diversity combiner. To restrict the bandwidth of the transmitted signal and to try to avoid inter-symbol interference, square root raised cosine filters were used for both the transmit and the receive filters. For all the simulations an excess bandwidth of 25% was used,

To compare the performance of the DSA, the matched filter bound was used, forming a lower bound on the expected performance, the best possible receiver unimpaired by ISI, Wozencraft et al. [6]. The matched filter bound was obtained assuming the data symbols suffered neither ISI nor interference, only additive noise, Lucky et al. [7].

The signal and noise powers were the same at each antenna. The same signals and noise were applied to the single antenna simulations and each antenna in the diversity simulation. At each antenna the noise was assumed to be independent. In the work by Schlagenhaufer et al. [8] the signal to noise ratio (S/N) was defined as the sum of the S/N at each antenna. This was not used as it would have required the S/N at the two antennas to be changed to compare the single antenna and diversity techniques at the same S/N. That would have implied a different physical configuration and raised the question of how that would be in practice.

By generating specific results with some numerical simulations modeling elements of the commercial standard 1xEVDO, also known as High Data Rate, or IS-856, some idea of the performance with modern applications was gathered. The values used for simulation were shown in table 2. Comparing the coherence bandwidth with the PN chip rate, where PN was pseudo noise, showed that the signal would suffer some distortion in frequency response. However the coherence time and PN chip period indicated that the distortion would stay constant for approximately 1200 chips, which should allow the estimated CIR to converge, and was used as the number of samples in all the trials.

Table 2. Constants used in simulations.

Characteristic	**Value**
PN chip rate	1.2288 Mchips/s
Modulation type	BPSK
Path delay spread	5.5 µs
Coherence time	1 ms, (carrier 1.9 GHz, speed 100 km/h)
Coherence bandwidth	36 kHz

4. RESULTS

4.1 Diversity

The 1 antenna CIR1 trace in figure 3 indicated how difficult the path was, due to a spectral null. For CIR2 which suffered phase distortion however the 1 antenna trace showed a peak at an S/N of 30 dB; this was the effect of two of the thirty trials having 122 and 194 errors which was significantly greater than the other trials by two orders of magnitude; these are due to slippage as described in section 4.4. Similarly the two antenna CIR1 and CIR2 trace shows a peak at S/N of 15 dB due to one trial with 58 errors.

For a BER of 5×10^{-3} a 4 dB improvement over the better channel was achieved. Conventionally the overall gain due to diversity was the sum of two components, the combining gain due to two signals giving a 3 dB gain and the diversity gain. For these simulations diversity gain due to fading was not applicable. However some improvement above 3 dB combining gain could be seen before the start-up and slip errors dominated creating a BER floor at higher S/N.

The floor to the BER at higher S/N ratios may have been due to the slippage as described in section 4.4 or if slippage was not the cause a few errors at the beginning of the trial would have sufficed. This was because the selection of which trellis to keep is started when the number of samples collected is only $O + \log_2(2M)$. This was limited by the typical problem with this type of algorithm that was the number of computations needed as pointed out in section 2. Therefore in the example plotted the first trellises were discarded after 5 + 3 = 8 samples were received. To survive after these eight samples the correct trellis should have had a lower accumulated metric than four out of the eight trellises that were compared at the state. Which meant it made a better estimate of the CIR than four of the eight to get a lower error for the estimated received signal. Thus the noise did not distort the received signal too much to drive the error higher at the eighth sample or the noise did not drive the estimated CIR away from the true one during the previous seven samples. Using the LMS algorithm to find the CIR only seven samples would not have been expected to be sufficient to get a good estimate of the true CIR however this was not what was used, rather the accumulated metric only had to be lower than most of the other trellises leading to state. This would be expected to be true since using the wrong training sequence for the LMS algorithm should produce the wrong CIR which in turn produced the wrong estimate of the received sample in turn leading to larger error and larger accumulated metric.

When the first samples were received and the LMS algorithm started to learn the CIR the initial coefficients are all zeros and noise on the first samples pushed the estimate of the CIR for the correct trellis away from the true CIR leaving another trellis with a lower accumulated metric, by definition any other trellis had at least one error. This would have allowed some errors to occur while the CIR were being learnt.

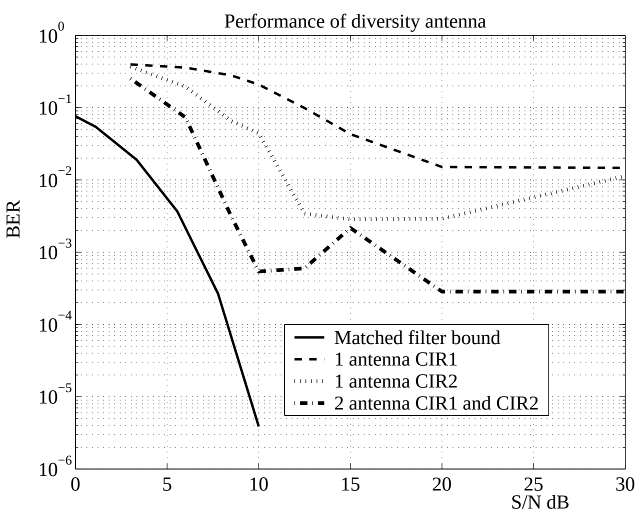

Figure 3. Diversity performance improvement.

4.2 Model Order Mismatch

Since the receiver was unlikely to know the correct CIR the effect of the mismatch between the true and estimated CIR order was investigated and found to be significant. With the true orders of the CIRs were used the addition of the second antenna improved the performance to be better than either of the two antennas taken alone, as seen in the previous section. However as figure 4 illustrated if the order of the estimated CIR was higher than the true order the performance was drastically reduced, this was due to slippage as described in section 4.4. The improvement in diversity performance with a third-order estimate was due to the relation between the true fifth-order response of CIR2 and the close approximation to it by a third-order estimation, which happened to be [-0.5 0.7 0.36] the middle three coefficients .

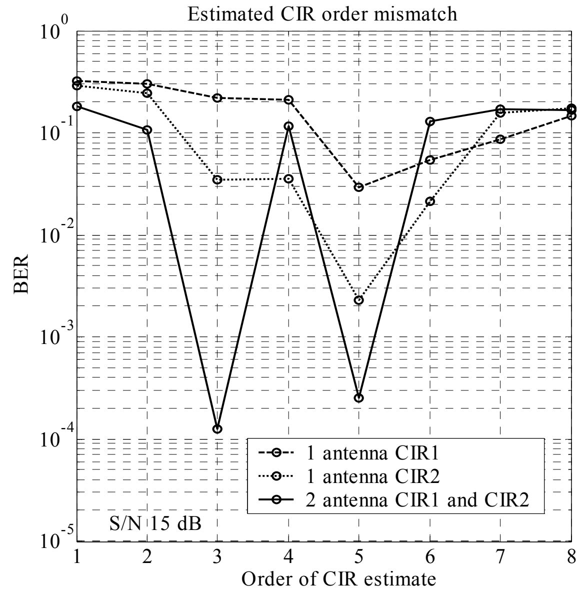

Figure 4. Effect of mismatched order of estimated CIR.

The third order estimation of CIR2 seemed to give a sufficiently good guide to push the estimation of CIR1 to a better estimation than it could find alone because the same data were used for each path with this diversity combining technique. It was possibly chance that turned up this result and the probability of having a CIR that could be well modeled by a lower order may not be high, also the problem remained to find that correct lower order.

4.3 Convergence of M Trellises per State

Convergence was defined as when the M trellises of the state with the minimum accumulated metric were the same. Consequences of this may be the loss of correct trellis and so the performance would be lowered when fading occurred.

At very low S/N the errors and metrics are dominated by noise. As the S/N improves the error from the wrong data start to dominate so that normally an error in the trellis will lead to a higher metric. However when the S/N is still around 10 dB, noise may lead to a lower error for a wrong data hypothesis. This would imply a higher error for the true data, leaving a higher metric and pushing the estimated CIR away from the true one. For the trellises with true data up to this point the one with the one wrong data would now have a lower accumulated metric and be likely to survive as one of the M trellises when compared to any trellis. After O samples the trellis with the wrong data would be compared to the true trellis and probably survive, then it would propagate with the true trellis as one of the M, until after $10O$ samples it would be truncated leaving another copy of the true trellis.

As the S/N improved further wrong data tended to have higher errors leading to greater changes in estimated CIR, this led to a string of higher metrics as the wrong data moved through the estimated CIR and the LMS tried to adjust to the wrong data. After that if the correct data followed the LMS readjusted reducing the errors. This would lead to a larger accumulated metric for trellises with wrong data in older positions and they would not propagate as they would not survive comparison with any trellis that had wrong data in a more recent position. This was illustrated in figure 5.

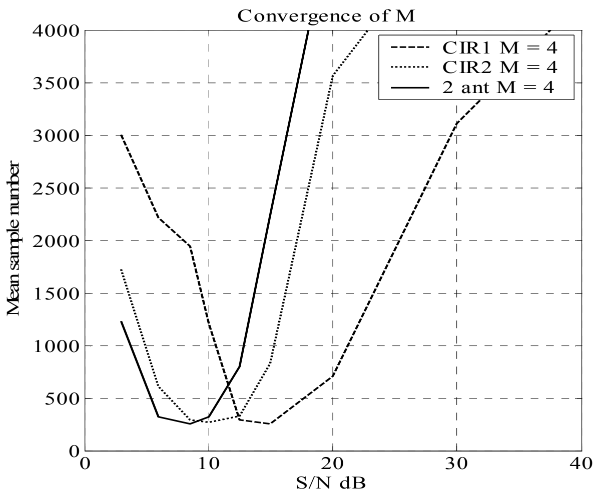

Figure 5. Convergence of M trellises per state.

The S/N where the mean time to convergence starts to increase appears to depend on the CIR and the use of the DSA can improve this by moving reducing the S/N where that happened.

4.4 Slippage Errors

Slippages were defined as the change in selection between two versions of the correct trellis, where the only difference in the versions was the delay by one sample period of one trellis compared to the other.

The corresponding CIR estimates were also offset by one sample period. To illustrate this figure 6 was from one trial showing the CIR which had the minimum accumulated metric at each sample time. The vertical axis shows the magnitude of the CIR's coefficients and the axis going into the page the sample number corresponding to the time. There were two peaks showing the choice of two CIRs.

The MLSE algorithm should have converged after 5 to 10 times the order of the CIR. The ideal algorithm should track one CIR not both to avoid this slippage. However, unlike the conventional VA, every trellis stored had its own CIR estimate; those trellises which were slipped, had CIR estimates which were offset in time as shown in figure 6. The decision to keep a trellis was taken state by state, depending on the accumulated metric and this metric depended on the error between the received sample and the expected signal estimated from the CIR for that state. The slipped and non-slipped trellises would be in different states. So they would not be directly compared until the states were the same which occurred when the data sequence had a string of ones or minus ones, of length greater than *O*. It appeared that it was possible for the combination of noise on the received sample and mis-adjustment of the CIR estimate from the LMS algorithm on the expected signal to allow the slipped trellises to survive. In general it was not possible to say if a trellis with delay δ samples was correct and that trellis with delay $\delta + 1$ was wrong. Error detection in the simulation took the data output sequence and found the delay which matched most closely the input data sequence, so it did not matter if the data sequence was delayed by δ or $\delta + 1$. The problem was that the detection of data was done by using the trellis with the minimum accumulated metric, presumably the one with the best match to the best CIR estimate, but when two trellises had survived the data may be taken from either as they must have been giving about the same accumulated metric for them both to survive. This switching from slipped to non-slipped was what causes the errors at the detector output.

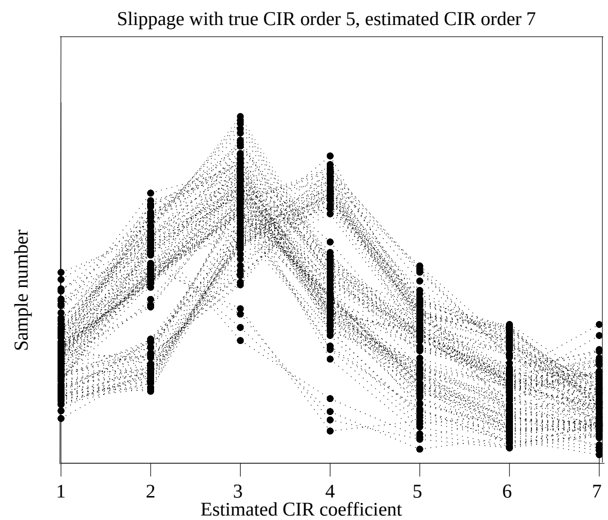

Figure 6. Sequence of estimated CIRs from lowest accumulated metrics.

The error detector found the offset which best matched the transmitted data over the whole trial, this masked the effect of slippage where no change in the delay took place; if the same trellis with slip was used for the whole trial zero errors could be produced.

In conventional BER tests the trials would run until a significant number of errors were collected. To illustrate the effect of slippage a fixed number of trials was used for each S/N point, this avoided terminating the trial too early.

Slippage was observed to be particularly often the cause of errors using the two antenna diversity. By using two antenna diversity it appeared that both slipped versions were more likely to survive than using only one antenna, this was attributed to the link between the antennas being the metric used to choose the survivors. When linked together the CIR that had lower error at the beginning, whether the true or slipped one, made the metric lower for that trellis and so more likely to survive. This in turn ensured the other CIR estimate worked on the correct or slipped trellis.

4.5 Trade-off M for Diversity

To compare the effect of using diversity to increasing the value of M figure 7 showed how diversity improved performance at lower S/N and M can reduce the BER floor. New data was generated for all five curves so they did not show the same slip errors as figure 1. A trade off between computing M extra trellises and adding an extra diversity receiver could be made in this manner. The first 100 samples were used to avoid loosing the M different trellises.

If the easier CIR was used with one antenna increasing M to 16 would achieve similar performance to a two antenna diversity with M = 4, one antenna would have required double the processing compared to two antennas which have double the hardware, but the diversity system still has better performance at low S/N ratios due to the increased signal received. As M was doubled the exhaustive search increased by one sample allowing the CIR estimate of the correct trellis to converge, before any trellises were discarded.

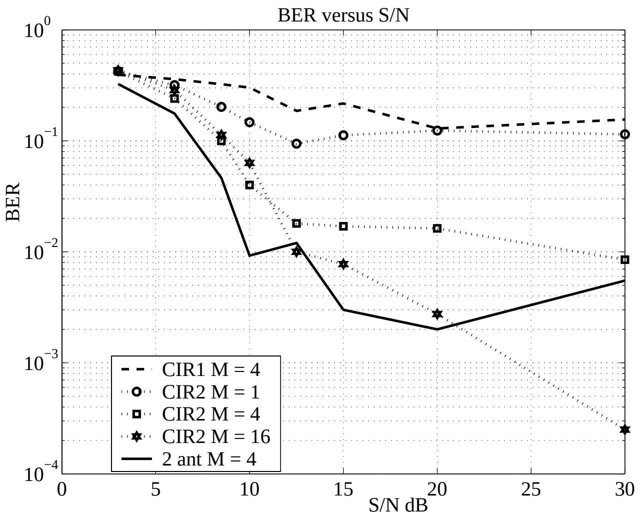

Figure 7. Trade-off M for diversity.

5. CONCLUSION

By applying diversity to the generalized MLSE improved performance was achieved even with difficult to equalize channels. It was crucial to use the correct order for the estimation of the CIR. When the true order of the CIRs was used and they were equal the addition of the second antenna improved the performance to be better than either of the two antennas taken alone, in figure 3 for a BER of 5 x 10^{-3} a 4 dB improvement over the better channel was achieved. Diversity gain above the gain from increased received power at the second antenna was difficult to define due to the different performance of the CIRs at each antenna and the presence of the slip and start-up errors.

There appeared to be two conflicting requirements neither of which were under the control of the receiver, the first was data rich enough to teach the LMS algorithm the CIR quickly and the second was strings of ones or minus ones to get the selection algorithm to select either the correct or slipped trellis allowing only one to propagate.

With a wrong order of estimated CIR the DSA performance may even be slightly worse than a single antenna, and that performance was about two orders of magnitude worse than using the true CIR order. When the order of the estimated CIR was greater than the true order, the DSA succeeded in finding the true CIR but errors were introduced by allowing the slipped CIR and trellis to exist. To reduce the number of computations it would have been convenient to be able to use the estimated CIR order less than the true one and achieve good performance, this appeared to be possible in some cases but not all, these cases may only be a few.

As shown in section 4.3 the M trellises kept per state gradually converged to be the same. If the DSA was left running and the signal suffered fading the performance of the M trellises kept per state would have been lost. In practice the longer the DSA was running the more likely the CIR would be to change and more need to be able to track it. This seemed to be a serious difficulty with the DSA.

Further work would include the addition of a fading model to assess the M convergence issue, and the inclusion of synchronization with CIR order estimation.

6. ACKNOWLEDGMENTS

An Operating Grant from the Natural Sciences and Engineering Research Council (NSERC) funded a literature search.

7. REFERENCES

[1] N. Seshadri, “Joint data and channel estimation using blind trellis search techniques,” *IEEE Trans. Commun.*, vol. 42, nos. 2-4, pp. 1000–1011, February/March/April 1994.

[2] E. Biglieri, J. Proakis and S. Shamai, “Fading Channels: Information-Theoretic and Communications Aspects,” *IEEE Trans. on Information Theory*, vol. 44, no. 6, October 1998.

[3] R. Raheli, A. Polydoros, C.K. Tzou, “Per-Survivor Processing - A general approach to MLSE in uncertain environments,” *IEEE Trans. Commun.*, vol. 43 , nos. 2-4, pp. 354-364, February-April 1995.

[4] K. Chugg and A. Polydoros, “MLSE for an unknown channel, Part I: Optimality considerations,” *IEEE Trans. Commun.*, vol. 44, no. 7, pp. 836–846, July 1996; “Part II: Tracking performance,” August 1996.

[5] J.G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1983.

[6] J.M. Wozencraft and I.M. Jacobs, *Principles of Communication Engineering*. New York, NY, USA: John Wiley & Sons Inc., 1965.

[7] R.W. Lucky, J. Salz, and E.J. Weldon, *Principles of Data Communications*, New York: McGraw-Hill, 1968.

[8] R. Schlagenhaufer, B.R. Petersen, A.B.Sesay “Delayed-Decision-Feedback Equalization for Multiuser Systems,” *Conf. Proc. of Wireless 99*, Calgary, Alberta, Canada, vol. 1, pp. 71-84, July 12-14, 1999.