14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A LOW COMPLEXITY ITERATIVE CHANNEL ESTIMATION AND EQUALISATION SCHEME FOR (DATA-DEPENDENT) SUPERIMPOSED TRAINING

S. M. A. Moosvi[1], *D. C. McLernon*[1], *E. Alameda-Hernandez*[1], *A. G. Orozco-Lugo*[2], *and M. M. Lara*[2]

[1]School of Electronic and Electrical Eng.,The University of Leeds, Leeds LS2 9JT, UK.
[2]Sección de Comunicaciones. CINVESTAV-IPN. CP 07360. México City, México.

ABSTRACT

Channel estimation/symbol detection methods based on superimposed training (ST) are known to be more bandwidth efficient than those based on traditional time-multiplexed training. In this paper we present an iterative version of the ST method where the equalised symbols obtained via ST are used in a second step to improve the channel estimation, approaching the performance of the more recent (and improved) data dependent ST (DDST), but now with less complexity. This iterative ST method (IST) is then compared to a different iterative superimposed training method of Meng and Tugnait (LSST). We show via simulations that the BER of our IST algorithm is very close to that of the LSST but with a reduced computational burden of the order of the channel length. Furthermore, if the LSST iterative approach (originally based on ST) is now implemented using DDST, a faster convergence rate can be achieved for the MSE of the channel estimates.

1. INTRODUCTION

Digital communication systems require an estimate of the channel prior to equalisation. Channel estimation techniques fall into three main categories: blind, semi-blind and trained. In this work we mainly focus on the last category because of its simplicity and satisfactory performance. Normally, the training sequence used for channel estimation is allocated an empty time slot in the transmitted frame, thus wasting bandwidth. This drawback was overcome when the training sequence was instead added to the data in what is now called *superimposed training* (ST) [1, 2]. But since training and information are sent at the same time, from the channel estimation point of view, the information interferes with the training and effectively acts as unwanted noise. Later, in [3], a modified ST known as data-dependent ST (DDST) was able to make the information sequence transparent to the training sequence, thus removing the "information noise" and hence significantly improving channel estimation. In this paper we present a similar way of alleviating — but not fully cancelling as DDST achieves— the effects of the information data noise in ST. This is done in an iterative manner, where the equalised symbols obtained through traditional ST in a first step are fed back to the ST algorithm so that the information noise can be reduced, and hence better channel estimation and symbol detection is obtained, which in turn can be used in the next iteration. The cancellation of information noise will only be perfect when accurate equalisation is achieved (i.e. long data records and high SNR). In this situation the iterative ST (IST) will aproach the performance of DDST. The novelty of this IST approach is that by re-using the ST equalised symbols again in the ST algorithm, we get DDST performance, but as we will see later, with relatively little additional computational burden.

S. M. A. Moosvi is funded by the ORS Scheme and The School of Electronic and Electrical Engineering, The University of Leeds, Leeds, UK.

E. Alameda-Hernández is funded by Secretaría de Estado de Educación y Universidades of Spain and the European Social Fund.

An alternative iterative approach, is not to re-use the equalised symbols in the actual ST algorithm as we have just proposed, but to re-use these symbols for a traditional least squares channel estimate of a fully trained system. This least squares ST (LSST) approach is not new, and has previously been used in [4]. The great disadvantage of LSST is its large computational complexity when compared with the proposed IST algorithm.

Now, given that both iterative procedures (the new IST and the already existing LSST) discussed here feed back the equalised symbols obtained with ST, a better performance is expected if DDST is employed instead of ST. But an iterative DDST, in the fashion of IST, will not provide better channel estimates than DDST since DDST already removes the information noise —which is the desired effect of IST. On the other hand, a least squares DDST instead of ST is expected to approach the behaviour of fully trained estimation quicker than the LSST method in [4]. So the objectives of this paper are then:

i) To develop a new low complexity iterative ST (IST) and show that its performance approaches that of DDST.

ii) To develop the iterative method using least squares for DDST (i.e., LSDDST) and to compare its performance with LSST (as proposed in [4]).

iii) To compare via simulations the channel estimate MSE and BER after equalisation for the new algorithms of IST and LSDDST along with the existing LSST of [4].

2. ITERATIVE ST (IST) AND LEAST SQUARES DDST (LSDDST)

We start with a brief overview of DD(ST).

2.1. A review of (data dependent) superimposed training

Consider a baseband equivalent digital communications system within the ST/DDST scenario, where a periodic training sequence $c(k)$ of length N and period P is added to the information bearing symbols $b(k)$ before transmission over an FIR channel $\{h(k)\}_{k=0}^{M-1}$ contaminated by additive, white, Gaussian noise, $n(k)$. In addition, for DDST a periodic (period P) data-dependent sequence $e(k) = -\frac{1}{N_P}\sum_{i=0}^{N_P-1} b(iP+k)$, $k = 0, 1, \ldots, P-1$ and $N_P = \frac{N}{P}$, is also included at the transmitter [3]. Note that for ST, $e(k) = 0$. Then in general,

$$\begin{aligned} x(k) = &\sum_{m=0}^{M-1} h(m)b(k-m) + \sum_{m=0}^{M-1} h(m)e(k-m) + \\ &+ \sum_{m=0}^{M-1} h(m)c(k-m) + n(k) + d \end{aligned} \tag{1}$$

where $k = 0,\ 1,\ \ldots,\ N-1$. In matrix form:

$$\mathbf{Sh} + \mathbf{n} + \mathbf{d} = \mathbf{x} \tag{2}$$

with $s(k) = b(k) + c(k) + e(k)$. Note that $\mathbf{S}$ is the $N \times M$ data matrix. We will assume that all terms in (2) can be complex; that $b(k)$ and $n(k)$ are from independent, identically distributed (i.i.d.) random zero-mean processes, with powers σ_b^2 and σ_n^2 respectively; that $c(k)$ is known with power $\sigma_c^2 = \frac{1}{P}\sum_{k=0}^{P-1} |c(k)|^2$; and d is an unknown DC-offset (see [1, 2] for explanation regarding d). The problem is first to estimate $\{h(k)\}_{k=0}^{M-1}$ from the N received samples of $x(k)$, and then via equalisation to estimate the transmitted data $b(k)$. As the method described in [5] can easily be modified to include the iterative process to be described (and this will be shown in a later paper), we will assume for simplicity of presentation that perfect synchronisation and knowledge of the DC-offset are provided. So we can, in what follows, set $d = 0$ and $P = M$. Note that $P > M$ is only required if the DC-offset and/or the synchronisation have to be estimated [5].

Now as in [2] we can write

$$\hat{y}(j) = \frac{1}{N_P}\sum_{i=0}^{N_P-1} x(iP+j) \tag{3}$$

with $j = 0,\ 1,\ \ldots,\ P-1$, where $\hat{y}(j)$ is an estimate of the periodic (period P) cyclic mean $y(j) \equiv E\{x(iP+j)\}$. So from (1) and (3) we can easily show that

$$\begin{aligned} \hat{y}(j) = &\sum_{m=0}^{M-1} h(m)\tilde{b}(j-m) + \sum_{m=0}^{M-1} h(m)e(j-m) + \\ &+ \sum_{m=0}^{M-1} h(m)c(j-m) + \tilde{n}(j) \end{aligned} \tag{4}$$

with $j = 0,\ 1,\ \ldots,\ P-1$, where

$$\tilde{b}(k) = \frac{1}{N_P}\sum_{i=0}^{N_P-1} b(iP+k) \tag{5}$$

with $k = 1-P,\ 2-P,\ \ldots,\ P-1$, and

$$\tilde{n}(j) = \frac{1}{N_P}\sum_{i=0}^{N_P-1} n(iP+j) \tag{6}$$

with $j = 0,\ 1,\ \ldots,\ P-1$. So (4) can now be written as

$$(\mathbf{C} + \tilde{\mathbf{B}} + \mathbf{E})\mathbf{h} = \hat{\mathbf{y}} - \tilde{\mathbf{n}} \tag{7}$$

where $\mathbf{C}$ and $\mathbf{E}$ are $P \times P$ circulant matrices with first columns $[c(0)\ c(1)\ \ldots\ c(P-1)]^T$ and $[e(0)\ e(1)\ \ldots\ e(P-1)]^T$ respectively, and $\mathbf{h} = [h(0)\ h(1)\ \ldots\ h(P-1)]^T$, with similar expressions for $\hat{\mathbf{y}}$ and $\tilde{\mathbf{n}}$. Now the $P \times P$ matrix $\tilde{\mathbf{B}}$ can be expressed as $\tilde{\mathbf{B}} = \tilde{\mathbf{B}}_1 + \tilde{\mathbf{B}}_2$, where $\tilde{\mathbf{B}}_1$ is circulant with first column $[\tilde{b}(0)\ \tilde{b}(1)\ \ldots\ \tilde{b}(P-1)]^T$ and $\tilde{\mathbf{B}}_2$ is upper triangular Toeplitz and $\frac{[b(-k)-b(N-k)]}{N_P}$ are the elements of the k-th $(k = 1,\ 2,\ \ldots,\ P-1)$ upper diagonal.

2.2. Iterative ST (IST)

In this section we consider two iterative channel estimation schemes for ST. For the ST case (i.e. when $\mathbf{E} = \mathbf{0}$ in (7)) we have $\hat{\mathbf{y}} = (\mathbf{C} + \tilde{\mathbf{B}})\mathbf{h} + \tilde{\mathbf{n}}$. And using the channel estimate $\mathbf{C}^{-1}\hat{\mathbf{y}}$ from [2] then

$$\hat{\mathbf{h}}_{\text{ST}} = \mathbf{C}^{-1}\hat{\mathbf{y}}. \tag{8}$$

Therefore substituting $\hat{\mathbf{y}}$ from (7) we get

$$\hat{\mathbf{h}}_{\text{ST}} = \mathbf{h} + \mathbf{C}^{-1}\tilde{\mathbf{B}}\mathbf{h} + \mathbf{C}^{-1}\tilde{\mathbf{n}}. \tag{9}$$

Now we can think of two ways of improving the estimate of $\mathbf{h}$ in the ST scenario. First estimate $\mathbf{S}$ in (2) using the ST algorithm followed by minimum mean square error (MMSE) equaliser (i.e., $\hat{\mathbf{S}}_{\text{ST}}$). So the channel estimate using the least-squares method ($\hat{\mathbf{h}}_{\text{LSST}}$) from (2) (i.e. LSST) would become

$$\hat{\mathbf{h}}_{\text{LSST}} = (\hat{\mathbf{S}}_{\text{ST}}^{\text{H}}\hat{\mathbf{S}}_{\text{ST}})^{-1}\hat{\mathbf{S}}_{\text{ST}}^{\text{H}}\mathbf{x} \tag{10}$$

which is essentially what was proposed in [4] and where $\hat{\mathbf{S}}_{\text{ST}}$ is a ST estimate of $\mathbf{S}$. The second approach is to use ST followed by a MMSE equaliser and make an estimate ($\hat{\tilde{\mathbf{B}}}$) of $\tilde{\mathbf{B}}$

in (7). So we can improve the channel estimate upon ST in (9) by using iterative ST ($\hat{\mathbf{h}}_{\text{IST}}$) via

$$\hat{\mathbf{h}}_{\text{IST}} = (\mathbf{C} + \hat{\tilde{\mathbf{B}}})^{-1}\hat{\mathbf{y}}. \tag{11}$$

Therefore substituting $\hat{\mathbf{y}}$ from (7) (with $\mathbf{E} = 0$) we get

$$\hat{\mathbf{h}}_{\text{IST}} = (\mathbf{C} + \hat{\tilde{\mathbf{B}}})^{-1}\Big\{[\mathbf{C} + \tilde{\mathbf{B}}]\mathbf{h} + \tilde{\mathbf{n}}\Big\}$$

$$= (\mathbf{C} + \hat{\tilde{\mathbf{B}}})^{-1}(\mathbf{C} + \tilde{\mathbf{B}})\mathbf{h} + (\mathbf{C} + \hat{\tilde{\mathbf{B}}})^{-1}\tilde{\mathbf{n}}.$$

Since we assume that $\hat{\tilde{\mathbf{B}}} \simeq \tilde{\mathbf{B}}$ then (9) is improved (see (15)) via

$$\hat{\mathbf{h}}_{\text{IST}} = (\mathbf{C} + \hat{\tilde{\mathbf{B}}})^{-1}\hat{\mathbf{y}} \simeq \mathbf{h} + (\mathbf{C} + \tilde{\mathbf{B}})^{-1}\tilde{\mathbf{n}}. \tag{12}$$

From (11) improved MMSE equalised symbols are obtained that can be fed back again to be used in (11). This iterative ST process (IST) can be repeated as needed. The benefits of the iterative processes based on (10) or (11) instead of the traditional ST based on (8) are now made clear if we compute the channel estimate MSE. We define $\text{MSE}(\hat{\mathbf{h}}) := E\Big\{\sum_{k=0}^{M-1}|\hat{h}(k) - h(k)|^2\Big\}$, then we can show that

$$\text{MSE}(\hat{\mathbf{h}}_{\text{ST}}) = \frac{1}{N_P}\left[\frac{\sigma_b^2\sum_{k=0}^{P-1}|h(k)|^2 + \sigma_n^2}{\sigma_c^2}\right] \tag{13}$$

$$\text{MSE}(\hat{\mathbf{h}}_{\text{LSST}}) \simeq \frac{\sigma_n^2}{N_P(\sigma_b^2 + \sigma_c^2)} \tag{14}$$

$$\text{MSE}(\hat{\mathbf{h}}_{\text{IST}}) \simeq \frac{\sigma_n^2}{N_P\sigma_c^2} \tag{15}$$

assuming $\hat{\mathbf{S}}_{\text{ST}} \simeq \mathbf{S}$ and $\hat{\tilde{\mathbf{B}}} \simeq \tilde{\mathbf{B}}$ in (10) and (11) respectively. We have assumed training sequences with $\mathbf{C}\mathbf{C}^{\text{H}} = P\sigma_c^2\mathbf{I}$, for the usual reasons given in [2, 5]. Since the IST method of (11) is based on first-order statistics, it is computationally very efficient compared to the LSST method proposed in [4], and the performance of the IST will approach that of DDST (see (19)) when accurate equalisation is achieved (i.e. long data records and high SNR).

2.3. Least-squares DDST algorithm (LSDDST)

In this section we present an iterative scheme for DDST. It is not difficult to see that if we choose $e(k) = -\tilde{b}(k)_P$, with $k = 0, 1, \ldots, N-1$ and $(\cdot)_P$ implying arithmetic modulo-P, in (1) —same result as [3] but obtained via a different analysis— then $\mathbf{E} = -\tilde{\mathbf{B}}_1$ and so for DDST (7) becomes

$$(\mathbf{C} + \tilde{\mathbf{B}}_2)\mathbf{h} = \hat{\mathbf{y}} - \tilde{\mathbf{n}}. \tag{16}$$

Now if we use a cyclic prefix $\big\{b(-k) = b(N-k)\big\}_{k=1}^{P-1}$ then $\tilde{\mathbf{B}}_2 = \mathbf{0}$, but even without a cyclic prefix $\lim_{N_P\to\infty}\tilde{\mathbf{B}}_2 = \mathbf{0}$. So let us assume a cyclic prefix (as was done in [3]), but in practice, no cyclic prefix makes little difference since $\tilde{\mathbf{B}}_2 \approx$

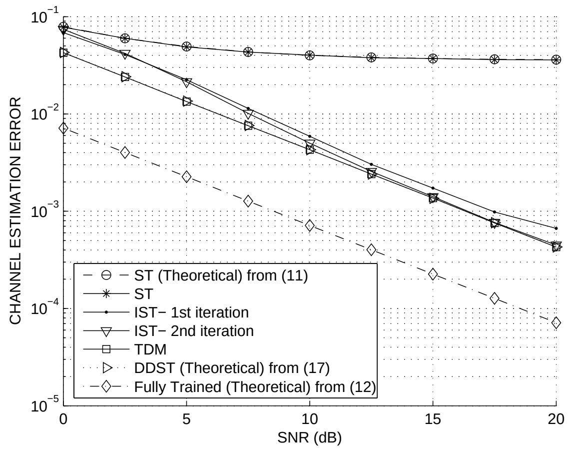

Fig. 1. Channel estimation error for IST. Note that IST approaches DDST. '*Fully trained*': all transmitted power used for channel estimation (benchmark for any trained algorithm). Note that TDM and DDST coincide and so are indistinguishable on graph.

$\mathbf{0}$, due to the usual choice of a relatively large N_P. So from (7) with $(\tilde{\mathbf{B}} + \mathbf{E} = \mathbf{0})$ then for DDST we have

$$\hat{\mathbf{h}}_{\text{DDST}} = \mathbf{C}^{-1}\hat{\mathbf{y}} = \mathbf{h} + \mathbf{C}^{-1}\tilde{\mathbf{n}}. \tag{17}$$

Using an equaliser based on the channel estimates of (17), we can obtain an estimate for $\mathbf{S}$ in (2), $\hat{\mathbf{S}}_{\text{DDST}}$. So similarly the optimum channel estimate ($\hat{\mathbf{h}}_{\text{LSDDST}}$) based on (2) and $\hat{\mathbf{S}}_{\text{DDST}}$ using least-squares approach would be

$$\hat{\mathbf{h}}_{\text{LSDDST}} = (\hat{\mathbf{S}}_{\text{DDST}}^{\text{H}}\hat{\mathbf{S}}_{\text{DDST}})^{-1}\hat{\mathbf{S}}_{\text{DDST}}^{\text{H}}\mathbf{x}. \tag{18}$$

As before, using $\hat{\mathbf{h}}_{\text{LSDDST}}$ in (18) and MMSE equalisation a better estimate for $\mathbf{S}$ can be obtained (taking into account the method to remove $e(k)$ proposed in [3]), and then fed back to (18) to form an iterative process. Again it is not difficult to show that

$$\text{MSE}(\hat{\mathbf{h}}_{\text{DDST}}) = \frac{\sigma_n^2}{N_P\sigma_c^2} \tag{19}$$

$$\text{MSE}(\hat{\mathbf{h}}_{\text{LSDDST}}) \simeq \frac{\sigma_n^2}{N_P\left(\sigma_{b+e}^2 + \sigma_c^2\right)} \tag{20}$$

assuming $\hat{\mathbf{S}}_{\text{DDST}} \simeq \mathbf{S}$ in (18). Now from (13) we can observe that in ST the data acts as interference, whereas in LSST we effectively remove the interference from the data and even increase the training power as can be seen from (14). Also in IST (15), we remove the intereference of the data but the training power remains the same and so it approaches the DDST performance of (19). Finally we can observe from (20) that the training power has effectively been increased.

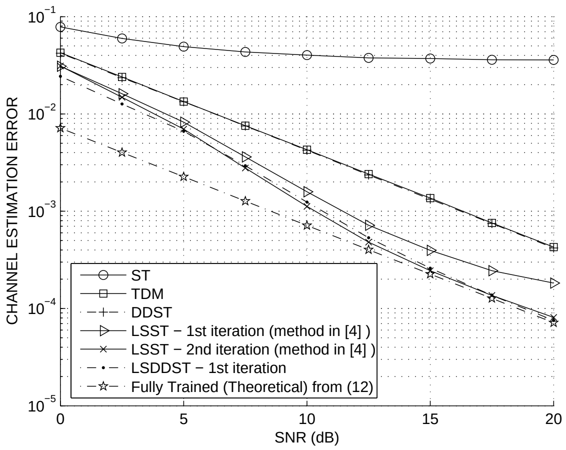

Fig. 2. Channel estimation error for LSST and LSDDST. LSST and LSDDST converge to a fully trained system. Note that TDM and DDST coincide and so are indistinguishable on graph.

2.4. Time Division Multiplexed Training

Traditionally, the training sequence was time division multiplexed (TDM) with the information sequence. Here we now compare both training schemes in a simplified scenario. So, the first question that arises is how to make a fair comparision. We have chosen to force DDST and TDM to provide the same channel estimation error, and then to compare the BERs of both methods. Now, it can be easily be shown that the channel estimation error for TDM is

$$\text{MSE}(\hat{\mathbf{h}}_{\text{TDM}}) = \frac{M}{N_t}\frac{\sigma_n^2}{\sigma_t^2} \tag{21}$$

where N_t is the length of TDM training sequence after the memory of the channel is full. Comparing (21) with (19), then DDST and TDM will have same channel estimation error if $N\sigma_c^2 = N_t\sigma_t^2$, since we have assumed $P = M$, i.e. the period of training sequence is equal to the number of taps in the channel. Note that to estimate the channel under the TDM scheme the memory of the channel must already be full and so $N_t + M - 1$- length training sequence is required with $N_t \geq M$. Finally, note that for DDST, in addition to N data samples we require $(M-1)$- length cyclic prefix. And so TDM and DDST will have the same MSE($\hat{\mathbf{h}}$), but TDM will use significantly more symbols (N_t+M-1) for training than DDST uses for its cyclic prefix $(M-1)$—hence the advantage of DDST.

3. SIMULATION RESULTS

The results of the simulations are shown in Figures 1–3 for three-tap Rayleigh fading channels. The channel coefficients were complex Gaussian, i.i.d. with unit variance. The average

Method	Computational burden	Performance approaches
ST ([1, 2])	$\mathcal{O}(M^2+M)$	–
IST	$\mathcal{O}(3QN)$ for 2 iterations	DDST ([3])
LSST ([4])	$\mathcal{O}(NM^2)$ for 1 iteration	Fully trained
LSDDST	$\mathcal{O}(NM^2)$ for 1 iteration	Fully trained
TDM	$\mathcal{O}(N_t(M^2+M)+M^2)$	DDST ([3])

Table 1. Summary of the performance and computational burden of all the methods presented here. Note that M and N refer to (1); P refers to the period of $c(k)$ in (1); and Q refers to the MMSE equaliser length; N_t refers to the TDM training sequence of length $N_t + M - 1$.

energy of the channel was set to unity. The data was a BPSK sequence, to which a training sequence fulfilling $\mathbf{CC}^{\text{H}} = P\sigma_c^2\mathbf{I}$ was added before transmission. The training to information power ratio $\left(\text{TIR} = \frac{\sigma_c^2}{\sigma_{b+e}^2}\right)$ was set to -6.9798 dB, $P = 7$ and $N = 420$ and a linear MMSE equaliser of length $Q = 11$ taps and optimum delay was used throughout. In order to make a fair comparison, we have included the results of channel estimation and BER using the traditional TDM scheme. The channel estimation performance of the DDST scheme is the same as that of the TDM scheme (for the reasons previously described), as is verified in Figures 1 and 2, where the number of training symbols in the TDM scheme is $N_t + M - 1 = 72$, compared to the DDST cyclic prefix of $M - 1 = 2$.

So Figure 1 gives the channel estimation MSE for the IST algorithm. It can be seen that there is a significant improvement in channel estimation and it approaches normal DDST and TDM performance just after 2 iterations. Now Figure 2 gives the channel estimation MSE for the LSST algorithm of [4] along with the proposed LSDDST. It can be seen that there is an even larger improvement over the normal DDST and that they both approach the performance of fully trained systems. Note that our method of LSDDST only requires 1 iteration to effectively converge as opposed to the LSST method in [4] that requires 2 iterations. Figure 3 shows the BER performance for the proposed IST algorithm along with the TDM scheme and when the channel is completely known. We can see that even with the low complexity IST, after two iterations we get virtually the same BER performance as that obtained when the channel is completely known for the superimposed training scheme. Figure 4 shows the BER for the proposed LSDDST algorithm along with the LSST [4]. Again we can observe that the performance of LSDDST is similar to the case when the channel is known completely for superimposed training scheme as well as to that of LSST [4].

Even the low complexity IST method after two iterations

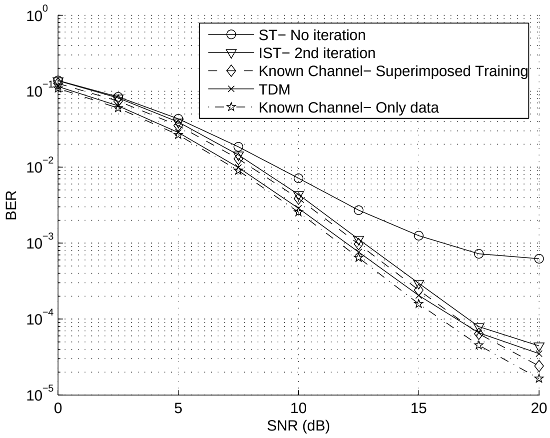

Fig. 3. BER for IST. '*Known channel-only data*': BER when the channel is completely known and full power is given to data; '*Known channel superimposed training*': BER when the channel is completely known and some of the data power is allocated for training.

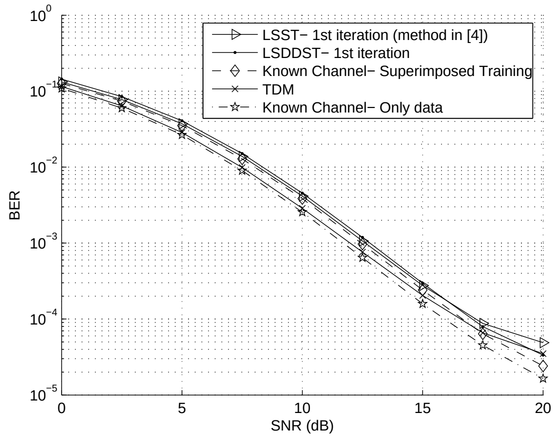

Fig. 4. BER for LSST and LSDDST. '*Known channel-only data*': BER when the channel is completely known and full power is given to data; '*Known channel superimposed training*': BER when the channel is completely known and some of the data power is allocated for training.

gives virtually the same BER performance as one iteration of LSST and LSDDST. It can also be observed that while the BER performance of all the proposed methods is very close to the TDM scheme, the latter however consumes more bandwidth. Table 1 gives the summary of the performance and computational burden of all the methods presented. Note that the computational burden of 2 iterations of IST is $\mathcal{O}(3QN)$ but with $\mathcal{O}(NM^2)$ per iteration for both LSST and LSDDST, yet all have almost the same BER.

4. CONCLUSIONS

In this paper, the symbols equalised via a hard detector preceded by a linear MMSE equaliser (designed using the channel estimated from an ST approach) are fed back into the ST method. As a result, better channel estimates and hence a more accurate equalisation is possible, which in turn can be used in the next iteration. The theoretical limiting performance of this iterative ST method (IST) is that of DDST, which is obtained when good equalisation is possible —long received records and high SNR. In practice, convergence is attained in two iterations. Another method, but suffering heavier computational burden, was derived by Meng and Tugnait [4]. Here we re-use the equalised symbols, not in the actual ST algorithm as in IST, but for a traditional least squares trained channel estimate (LSST). This LSST method approaches, for long data records and high SNR, the performance of a fully trained system after two iterations as regards the channel MSE estimates, and after one iteration as regards the BER. In this paper we also implemented the previous LSST with DDST, so that as simulations illustrate, convergence is achieved after one iteration (for both channel MSE and BER).

As far as the BER is concerned, simulations have shown that all the iterative methods considered here have approximately the same limiting performance. So, due to their computational burdens it is clear that IST is the algorithm of choice.

One possible application of this work is to use ST on the uplink (with the base-station performing IST estimation) and DDST on the downlink. In this scenario, we will have DDST performance in both directions, but with all the additional computational burden at the base-station, and not at mobile.

5. REFERENCES

[1] J. K. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol. 7, pp. 413–415, 2003.

[2] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel estimation using implicit training," *IEEE Trans. Signal Processing*, vol. 52, pp. 240–254, 2004.

[3] M. Ghogho, D. C. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Lett.*, vol. 12, pp. 226–229, 2005.

[4] X. Meng and J. K. Tugnait, "Semi-blind channel estimation and detection using superimposed training," in *Proc. IEEE ICASSP*, Montreal, May 17–21, 2004.

[5] E. Alameda-Hernandez, D. C. McLernon, M. Ghogho, A. G. Orozco-Lugo, and M. M. Lara, "Improved synchronisation for superimposed training based channel estimation," in *Proc. IEEE SSP*, Bordeaux, July 17–20, 2005.