# Correcting the Bias of Subtractive Interference Cancellation in CDMA: Advanced Mean Field Theory

Thomas Fabricius *Student Member IEEE* and Ole Winther *Member IEEE*

tf@isp.imm.dtu.dk and owi@imm.dtu.dk

Informatics and Mathematical Modelling

Technical University of Denmark

2800 Lyngby, Denmark

Fax: +45 45 87 29 99

Phone: +45 45 25 38 94

**Abstract**

In this paper we introduce an advanced mean field method to correct the inherent bias of conventional subtractive interference cancellation in Code Division Multiple Access (CDMA). In simulations, we get a performance quite close to that of the individual optimal exponential complexity detector and significant improvements over current state-of-the-art subtractive interference cancellation in all setups tested, for example in one case doubling the number of user at a bit error rate of $10^{-2}$. To obtain such a good performance for finite size systems, where the performance is normally degraded by the presence of suboptimal fix-point solutions, it is crucial to use the method in conjunction with mean field annealing, i.e. solving the fixed point equations at decreasing temperatures (noise levels). In the limit of infinite large system size, the new subtractive interference cancellation scheme is expected to be identical to the individual optimal detector. The computational complexity is cubic in the number of users whereas conventional (naive mean field) subtractive interference cancellation is quadratic. We also present a quadratic complexity approximation to our new method that also gives performance improvements, but in addition requires knowledge of the spreading code statistics. The proposed methodology is quite general and is expected to be applicable to other digital communication problems.

**Index Terms**

Multiple Access Technique, Subtractive Interference Cancellation, approximation error, bias correction, Advanced Mean Field Theory, Mean Field Annealing

## I. INTRODUCTION

When designing spectral efficient multi access communication systems, multi-access interference (MAI) is an inherent part one has to cope with. Communication systems are traditionally designed to keep MAI negligible compared to additive white Gaussian noise and self-interference, or sufficiently designed such that decoding can be performed reliable by matched filtration, i.e. treating the MAI as adding to the Gaussian noise. This approach yield low complexity and low cost receivers. Allowing a higher level of interference and adding more complexity in the receiver have the potential of increasing the spectral efficiency. The optimal detectors in Code Division Multiple Access (CDMA) [1], that explicitly model the interference, unfortunately have an exponential complexity in the number of users [2]. In order to gain over the matched filter approach, without making use of exponential complex detectors, suboptimal polynomial time-complexity detectors need to be developed.

Suboptimal multiuser detectors are traditionally separated into two main categories [3]: Linear Detectors and Subtractive interference cancellation. Interference cancellation has been introduced in CDMA to mitigate MAI [4], [5] or inter symbol interference [6]. Approximate combinatorial optimisation methods has been used to approximate the joint optimal detector [7]. Previous to that, CDMA multiuser detection has been mapped to a Hopfield neural network [8]. This is a general method to approximatively solve hard combinatorial problems [9]. In this paper we propose an improved subtractive interference cancellation scheme based upon a so-called advanced mean field method recently developed within machine learning and statistical physics [10], [11], [12]. Advanced mean field methods have also been shown to potentially powerful in blind channel estimation [13], [14]. However, in the following we will only discuss subtractive interference cancellation.

### *A. Subtractive Interference Cancellation*

Subtractive interference cancellation covers a whole family of multiuser detectors. The canonical form for the fix-point conditions in these detectors are given by

$$m_k^* = f\left(z_k - I_k(\mathbf{m}^*)\right) \qquad k = 1, \dots, K \tag{1}$$

where $z_k$ is the received data after the $k$'th matched filter, $m_k^*$ is the tentative decision estimate of the $k$'th symbol, $\mathbf{m}^*$ the tentative decisions of all bits arranged in a vector and $I_k$ is the reconstructed interference on the $k$'th symbol's decision statistic, $f(x)$ is some tentative decision function which for binary antipodal symbols has asymptotes $\pm 1$ for $x \to \pm\infty$. Let us assume that the generative model is $z_k = s_k + \sum_{k' \neq k} R_{kk'} s_{k'} + n_k$, where $s_k$, $k = 1, \dots, K$ are the $K$ users symbols, $R_{kk'}$ are the users correlation, and $n_k$ is additive noise. Then in the absence of noise, $n_k = 0$, it is obvious that if we set the interference $I_k$ equal to $\sum_{k' \neq k} R_{kk'} s_{k'}$, the $k$th users decision statistics $z_k - I_k = z_k - \sum_{k' \neq k} R_{kk'} s_{k'}$ will give the correct answer. This is only correct if all the symbols $s_{k'}$, $k' \neq k$ are known exactly. One approach is therefore to substitute $s_{k'}$ with its tentative estimate $m_{k'}$. Unfortunately, this approach yields a bias in the decision statistics $z_k - \sum_{k' \neq k} R_{kk'} m_{k'}$ [15], since $m_{k'}$'s depend on the earlier estimate of $m_k$. Consequently, the decision statistics in the fix-point are biased. We will call the error due to the approximation an *approximation error*.

This kind of approximation is in the machine learning and statistical physics literature called the *naive* mean-field equations [12]. We used this approximation in subtractive interference cancellation in Refs. [14], [16], [17]. In this contribution we derive the *naive* mean-field equations from the so-called *Callen identity* [18]. However, the main focus of the paper is on advanced mean-field methods which correct the naive approach by approximately removing the bias.

Within information processing and machine learning the advanced mean field methods have recently received attention [19], [12], e.g. see Ref. [20] for application to turbo codes. One can roughly divide the problem classes into two: densely and sparsely connectivity, i.e. whether $R_{kk'}$ (or the corresponding 'coupling matrix') are sparse or not. CDMA fall in the densely connected category and turbo codes in the sparsely connected. It has recently been shown that the 'sum-product' also known as belief propagation algorithm [21] used in the decoding of turbo-codes [22], [23], [24] is identical to a certain kind of mean field theory derived from the Bethe free energy [25]. An important advanced mean field method for densely connected systems, the so-called TAP approach after Thouless, Anderson, and Palmer [26] was originally introduced for the *Sherrington-Kirkpatrick* spin-glass model [27], [28]. In this model the correlations $R_{kk'} = R_{k'k}$ are modelled as Gaussian random variables. More recent work has focussed on the understanding of the correspondence between the TAP mean field theory and Bethe free energy approach [29]. In the following we will only discuss properties of the densely connected models.

Statistical physics for disordered systems makes use of the fact that for certain problems, an exact analysis is possible for large system sizes $K \to \infty$. There are basically two different but consistent analysis techniques available: one is microscopic corresponding to finding the low order statistics of the system variables, i.e. for CDMA the expectation values of the user symbols for given $R_{kk'}$ and $z_k$. The second way is an average case analysis giving the typical macroscopic properties of the system, i.e. for CDMA the expected bit error rate averaged over all realizations of the spreading codes and noise. Clearly, only the first approach is suitable when designing algorithms whereas the second can give valuable information about average properties, i.e. the expected performance in a given setting. Among the most prominent techniques within statistical physics for analysing this type of disordered systems are the *replica* and *cavity* methods [30], which can be used both for analysing static properties (as described above) and dynamical properties. In this paper we use the cavity method to derive advanced mean field algorithms for the CDMA problem. A corresponding average case replica analysis of randomly spread CDMA can be found in Refs. [31], [32]. The Hopfield neural network [33] and the linear perceptron with binary weights are closely related to the CDMA model. Average case analysis and TAP mean field theory for these models can be found in Refs. [30], [34], [11].

In the large system limit, the bias correction of the advanced method becomes *deterministic* or identical *self-averaging*, i.e. in the case of CDMA the value is determined by the statistics of the correlations rather than the specific realization of the spreading codes. In information processing applications we deal with finite sized systems where we cannot expect the bias to be self-averaging. We therefore need the bias correction to adapt to the data at hand. This has lead to the introduction of the *adaptive TAP mean field theory* [10], [35], [11]. We rederive the *adaptive TAP mean field* based upon Refs. [10], [35], [11]. This leads to fix-point equations like eq. (1) with an additional set of fix-point equations for the bias corrections. Correcting the bias of the naive mean field algorithm mitigates the approximation error, but we still have some residual approximation error. Only in the limit of infinite system size, $K \to \infty$, will the approximation error of the adaptive TAP mean field method vanish and the performance will coincide with the prediction of average case analysis [10], [11].

In this contribution we also present an effective second order scheme for solving the obtained fix-point equations. This is important since if one is not careful with assuring good convergence, the convergence error can overwhelm the approximation error. A third kind of error arises because the fix-point equations can have multiple solutions. Again, this is an effect of the finite dimensionality of the system at the loads (users to spreading factor) considered, i.e. for $K \to \infty$ the solution will be unique. In Ref. [17] we made a thorough analysis of the fix-points of the naive approach. This lead to the use of annealing, i.e. solving the fix-point equations at decreasing temperatures (noise levels) using at each step the previously found solution as the initial estimate. Annealing is also a crucial ingredient for obtaining good performance for the advanced mean field method.

The rest of the paper is organized as follows: in section II, we describe the model. In section III, we briefly review the optimal detectors. In sections IV-VII, we derive the advanced mean field detectors and the update and annealing schemes. We present simulations in section VIII and conclude in section IX.

## II. $K$ USERS CDMA AWGN MODEL

We will assume a chip and symbol synchronized CDMA model in additive white Gaussian noise (AWGN), with binary phase shift keying (BPSK) symbols. The $K$ spreading codes will we assume binary and random with unit energy and equal spreading factor (SF). We will allow the users to have different powers. The received base band CDMA signal is then given as

$$y(n) = \sum_{k=1}^{K} A_k b_k s_k(n) + \sigma\epsilon(n) \ , \tag{2}$$

$n$ is sample index $n \in [0, \ldots, N_c - 1]$, $N_c$ is the spreading factor, $s_k(n)$ is user $k$'s $n$'th chip in the spreading code with unit energy, $b_k \in [-1, 1]$ is the transmitted bit for user $k$, $A_k$ the $k$'th users root mean power, $\epsilon(n)$ is the additive white Gaussian noise with zero mean and variance 1, and $\sigma$ is the noise standard deviation. The users' energy normalisation make us able to define the $k$'th users signal to noise ratio as $\mathrm{SNR}_k = \frac{A_k^2}{2\sigma^2}$.

Correlating the signal $y(n)$ with the spreading codes $s_{k'}(n)$, $k' \in [1; K]$, we obtain the conventional detector outputs $z_{k'}$ which are sufficient statistics for the symbol estimation

$$z_{k'} = \sum_{n=0}^{N_c-1} y(n) s_{k'}(n) = \sum_{k=1}^{K} A_k b_k \sum_{n=0}^{N_c-1} s_{k'}(n) s_k(n) + \sigma \sum_{n=0}^{N_c-1} s_{k'}(n)\epsilon(n) = \sum_{k=1}^{K} A_k b_k R_{kk'} + e_{k'} \ , \tag{3}$$

where $R_{kk'} = \sum_{n=0}^{N_c-1} s_{k'}(n) s_k(n)$ is the correlation of code $s_k$ with $s_{k'}$, and $e_{k'}$ transform to a Gaussian random variable with zero mean and covariance $\sigma^2 r_{kk'}$. The above specifies the joint distribution of the $z_{k'}$s

$$\begin{aligned} &p\left(\mathbf{z}\,|\mathbf{b}, \mathbf{A}, \mathbf{R}, \sigma^2\right) = \\ &|2\pi\sigma^2\mathbf{R}|^{-\frac{1}{2}} \exp - \left(\tfrac{1}{2\sigma^2}(\mathbf{z} - \mathbf{RAb})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{z} - \mathbf{RAb})\right) \quad , \end{aligned} \tag{4}$$

now arranged in vectors with elements $(\mathbf{z})_{k'} = z_{k'}$ and $(\mathbf{b})_{k'} = b_{k'}$, and matrices $(\mathbf{A})_{kk'} = \delta_{kk'} A_k$, and $(\mathbf{R})_{kk'} = R_{kk'}$. We assume perfect channel state knowledge, and we also assume full code knowledge i.e. the matrices $\mathbf{A}$ and $\mathbf{R}$ and the noise variance $\sigma^2$ are considered deterministic. We thus write the likelihood as $p\left(\mathbf{z}\,|\mathbf{b}\right)$ instead of $p\left(\mathbf{z}\,|\mathbf{b}, \mathbf{A}, \mathbf{R}, \sigma^2\right)$.

## III. REVIEW OF OPTIMAL DETECTORS

The objective, given the received data and the channel, commonly are to minimise the expected bit error rate (BER) or the probability of error. The expected user averaged BER is defined as

$$\overline{\mathrm{BER}} = \frac{1}{2}\langle 1 - \frac{1}{K}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{b}}\rangle_{p(\mathbf{z},\mathbf{b})} \tag{5}$$

where the expectation $\langle\cdot\rangle_{p(\mathbf{z},\mathbf{b})}$, is taken with respect to the joint distribution of $\mathbf{b}$ and $\mathbf{z}$. The joint distribution $p(\mathbf{z},\mathbf{b})$ is found by the likelihood (4) multiplied by the a priori distribution $p(\mathbf{z},\mathbf{b}) = p(\mathbf{z}\,|\mathbf{b})\,p(\mathbf{b})$.

Under some mild regularity conditions, the expected BER subject to the constraint $\left|(\widehat{\mathbf{b}})_{k'}\right| = 1$ is minimised by

$$\widehat{\mathbf{b}} = \mathrm{sgn}\langle\mathbf{b}\rangle_{p(\mathbf{b}|\mathbf{z})} \tag{6}$$

for any given received data $\mathbf{z}$; sgn working each element. This estimator is referred the individual optimal detector [36].

Alternatively, one can consider the probability of getting one or more bit errors as the objective function. This probability of error can be written in terms of $p(e=1\,|\mathbf{b}) = 1 - \delta(K - \mathbf{b}^{\mathrm{T}}\widehat{\mathbf{b}})$ as

$$p(e=1) = 1 - \langle\delta(K - \mathbf{b}^{\mathrm{T}}\widehat{\mathbf{b}})\rangle_{p(\mathbf{z},\mathbf{b})}\ . \tag{7}$$

It is minimised for any received data $\mathbf{z}$ by [37]

$$\widehat{\mathbf{b}} = \underset{\mathbf{b}\in[-1;1]^K}{\mathrm{argmax}}\, p(\mathbf{b}\,|\mathbf{z})\ . \tag{8}$$

This is obviously the Maximum A Posterior (MAP) solution, which is identical to maximum likelihood for uniform prior distribution $p(\mathbf{b})$. This detector is referred the joint optimal detector [36].

For general spreading codes, the above detectors have exponential complexity in the number of users $K$, i.e. to form the average over $p(\mathbf{b}\,|\mathbf{z})$ we have to go over all $2^K$ combinations. We therefore seek to construct approximative polynomial time complexity detectors. The detectors we derive in the following are soft symbol estimators, i.e. they give an estimate of the marginal posterior mean $\langle\mathbf{b}\rangle_{p(\mathbf{b}|\mathbf{z})}$. They are constructed such that we by adjusting a suitable control parameter, the equivalent of a temperature, can obtain an approximation to either the individual or joint optimal detector.

## IV. Approximate Polynomial Time Complexity Detectors

In this section we derive two approximate detectors. Before the actual derivation we shortly introduce and motivate the approaches.

The first detector is the well-known soft-decision feedback subtractive interference cancellation scheme with a $\tanh(\cdot)$ non-linearity:

$$m_k = \tanh\left(\frac{A_k}{T}\left(z_k - \sum_{k' \neq k} R_{kk'} A_{k'} m_{k'}\right)\right) \tag{9}$$

for $k = 1, \ldots, K$, where $m_k$ are the approximate estimate of the posterior mean $\langle b_k \rangle$, i.e. $m_k \approx \langle b_k \rangle$ and $\sigma^2$ have been substituted with $T$ so that we can study fix-points at various $T$ and not only on the generic $T = \sigma^2$ which is the noise level. Within statistical physics and machine learning this fixed-point equation is known as the naive mean field estimate, see e.g. [12]. Why it is called so will be clear in the following.

It is well-known that fixed-point has a bias [15]. The second advanced mean field approach[1] aims at correcting this bias. Under some mild conditions–fulfilled for typical spreading sequences–this approach will lead to exact results for the means, $m_k = \langle b_k \rangle$ in the *limit of infinite system size*, $K \to \infty$. We are thus in the paradoxical situation that the NP-hard problem of inferring $\langle b_k \rangle$ [2] becomes polynomial for $K \to \infty$. For finite system sizes, the bias is only corrected approximately.

Besides the difficulty with completely correcting the bias, finite size systems also have other related complexities. Firstly, as already discussed there will be multiple solutions to the fix-point equations. Secondly, the basic assumption of the mean field theory (naive or adaptive TAP), namely that all interactions between variables are weak and of the same order, may break down. For example for finite $K$ and $N_c$ there is a finite probability that when the spreading codes are selected randomly, two users will be assigned the same code. This makes these users completely correlated. The breakdown of the theory is usually signalled by the lack of convergence of the fix-point iteration. In section VII, we discuss how to deal with this situation in practice.

There are a number of ways to derive both the naive and the advanced mean field theory. One prominent method is to derive the fix-point equations from the saddle-point of a suitable cost-function known as the mean field free energy, see e.g. [17], [11]. This approach has the advantage that the problem is formulated as an optimization problem. Here, we will take a different route and derive the mean field equations starting directly from the definition of the posterior mean. This derivation will highlight the approximations made. For the naive mean

[1] Often called the TAP approach after Thouless, Anderson and Palmer [26].

field theory we use the so-called Callen identity [18] and for advanced mean field theory, the cavity method [30], [10], [11]. To ease the notation we introduce the following quantities

$$\begin{aligned} \widehat{z}_k &= \frac{1}{T} A_k z_k \\ \widehat{R}_{kk'} &= \frac{1}{T} A_k (R_{kk'} - \delta_{kk'}) A_{k'} \end{aligned}$$

Assuming equal prior probabilities $p(b_k) = \frac{1}{2}\delta_{b_k,1} + \frac{1}{2}\delta_{b_k,-1}$, we can write

$$\begin{aligned} \langle b_k \rangle &= \frac{\sum_{\mathbf{b}\in[-1,1]^K} \; b_k \; p(\mathbf{z}|\mathbf{b})}{\sum_{\mathbf{b}\in[-1,1]^K} p(\mathbf{z}|\mathbf{b})} \\ &= \frac{1}{Z} \sum_{\mathbf{b}\in[-1,1]^K} b_k \; e^{\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{z}} - \frac{1}{2}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{R}}\mathbf{b}} \;, \end{aligned} \tag{10}$$

where

$$Z = \sum_{\mathbf{b}\in[-1,1]^K} e^{\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{z}} - \frac{1}{2}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{R}}\mathbf{b}} \tag{11}$$

and we have as above set $T = \sigma^2$. Note that we in the second line have divided out all $\mathbf{b}$ independent terms.

*a) Naive mean field theory – Callen identity:* We can now carry out the summation over the $k$th variable

$$\langle b_k \rangle = \frac{1}{Z} \sum_{\mathbf{b}\backslash b_k \in[-1,1]^{K-1}} 2\sinh\left(\widehat{z}_k - \sum_{k'\neq k} \widehat{R}_{kk'} b_{k'}\right) e^{\sum_{k'\neq k} b_{k'}\widehat{z}_{k'} - \frac{1}{2}\sum_{k',k''\neq k} b_{k'}\widehat{R}_{k'k''} b_{k''}} \;, \tag{12}$$

where $\mathbf{b}\backslash b_k$ means summation over all variables but the $k$th. Dividing and multiplying by $2\cosh\left(\widehat{z}_k - \sum_{k'\neq k}\widehat{R}_{kk'} b_{k'}\right)$ inside the summation and using

$$2\cosh\left(\widehat{z}_k - \sum_{k'\neq k} \widehat{R}_{kk'} b_{k'}\right) e^{\sum_{k'\neq k} b_{k'}\widehat{z}_{k'} - \frac{1}{2}\sum_{k',k''\neq k} b_{k'}\widehat{R}_{k'k''} b_{k''}} = \sum_{b_k\in[-1,1]} e^{\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{z}} - \frac{1}{2}\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{R}}\mathbf{b}} \;, \tag{13}$$

we arrive at the exact Callen identity

$$\langle b_k \rangle = \left\langle \tanh\left(\frac{A_k}{T}\left(z_k - \sum_{k'\neq k} R_{kk'} A_{k'} b_{k'}\right)\right)\right\rangle \;. \tag{14}$$

Taking the posterior average inside the $\tanh(\cdot)$, we arrive at the approximation for the mean eq. (9). This corresponds to neglecting the fluctuations of the random variable $\sum_{k'\neq k} R_{kk'} A_{k'} b_{k'}$. Thus the name mean field.

*b) TAP mean field theory – the cavity method:* The starting point of the improved mean field estimate is another similar exact relation for the posterior mean. We introduce the *reduced* or *cavity* average over a posterior where the $k$th variable is not present:

$$\langle \ldots \rangle_{\backslash k} = \frac{1}{Z_k} \sum_{\mathbf{b}\backslash b_k \in [-1,1]^{K-1}} \ldots \; e^{\sum_{k' \neq k} b_{k'} \widehat{z}_{k'} - \frac{1}{2} \sum_{k',k'' \neq k} b_{k'} \widehat{R}_{k'k''} b_{k''}} \tag{15}$$

with normalization constant

$$Z_k = \sum_{\mathbf{b}\backslash b_k \in [-1,1]^{K-1}} e^{\sum_{k' \neq k} b_{k'} \widehat{z}_{k'} - \frac{1}{2} \sum_{k',k'' \neq k} b_{k'} \widehat{R}_{k'k''} b_{k''}} \; . \tag{16}$$

The exact relation eq. (12) can now be written as

$$\langle b_k \rangle = \frac{\left\langle \sinh\left( \frac{A_k}{T} \left( z_k - \sum_{k' \neq k} R_{kk'} A_{k'} b_{k'} \right) \right) \right\rangle_{\backslash k}}{\left\langle \cosh\left( \frac{A_k}{T} \left( z_k - \sum_{k' \neq k} R_{kk'} A_{k'} b_{k'} \right) \right) \right\rangle_{\backslash k}} \; , \tag{17}$$

where we have used $Z = 2 \left\langle \cosh\left( \widehat{z}_k - \sum_{k' \neq k} \widehat{R}_{kk'} b_{k'} \right) \right\rangle_{\backslash k} Z_k$.

The TAP approximation corresponds to assuming a simple parametric form for the random variable $h_k \equiv \widehat{z}_k - \sum_{k' \neq k} \widehat{R}_{kk'} b_{k'}$ appearing in eq. (17). Before discussing this approximation, let us introduce $h_k$ as a random variable through a Dirac $\delta$-function:

$$p(h_k) = \left\langle \delta \left( h_k - \widehat{z}_k + \sum_{k' \neq k} \widehat{R}_{kk'} b_{k'} \right) \right\rangle_{\backslash k} \; . \tag{18}$$

With this definition we can write eq. (17) as

$$\langle b_k \rangle = \sum_{b_k \in [-1,1]} \int dh_k \; b_k \; p(b_k, h_k) \tag{19}$$

$$\langle h_k \rangle = \sum_{b_k \in [-1,1]} \int dh_k \; h_k \; p(b_k, h_k) \tag{20}$$

with the joint distribution being

$$p(b_k, h_k) = \frac{p(h_k) e^{b_k h_k}}{2 \int dh_k \; p(h_k) \; \cosh(h_k)} \; . \tag{21}$$

The second equality above relates $\langle h_k \rangle = \widehat{z}_k - \sum_{k' \neq k} \widehat{R}_{kk'} \langle b_{k'} \rangle$ to the average over the cavity distribution $\langle \ldots \rangle_{\backslash k}$ appear in $p(h_k)$, eq. (18).

Up to now everything has been exact. The central assumption in our derivation is that the terms in the sum $\sum_{k'\neq k}\widehat{R}_{kk'}b_{k'}$ are so weakly correlated that they for $K\rightarrow\infty$ sum up to a Gaussian variable (as a consequence of the central limit theorem), i.e.

$$p(h_k)\approx\frac{1}{\sqrt{2\pi V_k}}\exp\left(-\frac{(h_k-\langle h_k\rangle_{\backslash k})^2}{2V_k}\right)\ , \tag{22}$$

where $V_k\equiv\langle h_k^2\rangle_{\backslash k}-\langle h_k\rangle_{\backslash k}^2$ is cavity variance of $h_k$. We can now see that if we can determine $\langle h_k\rangle_{\backslash k}$ and $V_k$ we have a closed set of equations for $m_k\approx\langle b_k\rangle$. Inserting the assumption eq. (22) in eqs. (19) and (20) gives

$$\langle b_k\rangle = \tanh\left(\langle h_k\rangle_{\backslash k}\right) \tag{23}$$

$$\langle h_k\rangle = \langle h_k\rangle_{\backslash k}-V_k\langle b_k\rangle\ . \tag{24}$$

We have now almost achieved a closed set of fixed point equation for $\langle b_k\rangle$ since we can use eq. (24) to express $\langle h_k\rangle_{\backslash k}$ entirely in terms of $\langle\mathbf{b}\rangle$ and $V_k$. These results show how the advanced mean field theory corrects the bias of the naive result eq. (9) by introducing a self-coupling $-V_k\langle b_k\rangle$ inside the non-linearity. What remains to be done is to derive equations for $V_k$, $k=1,\ldots,K$. The result we are after–the adaptive TAP equations [10], [11]–can be derived in number of different ways. Here we use eqs. (23) and (24) and a first order perturbation argument [38] (also known as linear response theorem [18]) to derive an alternative expression for $\langle h_k\rangle_{\backslash k}$. Demanding consistency between this expression and eq. (24) determines $V_k$. We assume that $\langle b_{k'}\rangle-\langle b_{k'}\rangle_{\backslash k}$ is small for $k\neq k'$, i.e. removing the $k$th user will only induce a small change in the mean of the other variables. This is a reasonable assumption for normal spreading codes when $K$ is even moderately large, say order 10. It is therefore reasonable that we can expand to first order in eqs. (23) and (24) to get

$$\langle b_{k'}\rangle-\langle b_{k'}\rangle_{\backslash k}=(1-\langle b_{k'}\rangle^2)\left(\langle h_{k'}\rangle-\langle h_{k'}\rangle_{\backslash k}-V_{k'}(\langle b_{k'}\rangle-\langle b_{k'}\rangle_{\backslash k})\right)\ , \tag{25}$$

where we have used that $\frac{\partial\langle b_{k'}\rangle}{\partial\langle h_{k'}\rangle_{\backslash k'}}=1-\langle b_{k'}\rangle^2$ and assumed that the change in $V_{k'}$ is negligible; which is true in the large system limit since $V_{k'}$ will be self-averaging [11]. Now we can use $\langle h_k\rangle=\widehat{z}_k-\sum_{k'\neq k}\widehat{R}_{kk'}\langle b_{k'}\rangle$ to write

$$\langle h_{k'}\rangle-\langle h_{k'}\rangle_{\backslash k}=-\widehat{R}_{k'k}\langle b_k\rangle-\sum_{k''\neq k}\widehat{R}_{k'k''}(\langle b_{k''}\rangle-\langle b_{k''}\rangle_{\backslash k})\ . \tag{26}$$

Taking together eqs. (25) and (26) and rearranging gives

$$\sum_{k''} \left( \widehat{R}_{k'k''} + \delta_{kk'}(V_k + \frac{1}{1 - \langle b_{k'} \rangle^2}) \right) (\langle b_{k''} \rangle - \langle b_{k''} \rangle_{\backslash k}) = -\widehat{R}_{k'k} \langle b_k \rangle \tag{27}$$

Eq. (25) is therefore a linearised equation for the change in the means. Introducing

$$\Lambda_k = V_k + \frac{1}{1 - \langle b_k \rangle^2} \tag{28}$$

and the diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\Lambda_1, \ldots, \Lambda_K)$ we can write the solution as

$$\langle b_{k'} \rangle - \langle b_{k'} \rangle_{\backslash k} = \sum_{k''} \left[ (\mathbf{\Lambda}_{\backslash k} + \widehat{\mathbf{R}}_{\backslash k})^{-1} \right]_{k'k''} \widehat{R}_{k''k} \langle b_k \rangle \ , \tag{29}$$

where $\mathbf{\Lambda}_{\backslash k}$ is the matrix $\mathbf{\Lambda}$ with the $k$th row and column removed. Note that it is a direct consequence of the linear response assumption that the induced change will be proportional to the mean itself, i.e. $\langle b_k \rangle$. We can now use this result to get an estimate of $\langle h_k \rangle - \langle h_k \rangle_{\backslash k}$

$$\begin{aligned} \langle h_k \rangle - \langle h_k \rangle_{\backslash k} &= -\widehat{R}_{kk} \langle b_k \rangle - \sum_{k' \neq k} \widehat{R}_{kk'} (\langle b_{k'} \rangle - \langle b_{k'} \rangle_{\backslash k}) \\ &= -\widehat{R}_{kk} \langle b_k \rangle - \sum_{k', k'' \neq k} \widehat{R}_{kk'} \left[ (\mathbf{\Lambda}_{\backslash k} + \widehat{\mathbf{R}}_{\backslash k})^{-1} \right]_{k'k''} \widehat{R}_{k''k} \langle b_k \rangle \\ &= \left( \Lambda_k - \frac{1}{\left[ \left( \mathbf{\Lambda} + \widehat{\mathbf{R}} \right)^{-1} \right]_{kk}} \right) \langle b_k \rangle \end{aligned} \tag{30}$$

In the last line we have used a matrix identity for the partitioned inverse:

$$\frac{1}{[\mathbf{A}^{-1}]_{kk}} = A_{kk} - \sum_{k', k'' \neq k} A_{kk'} \left[ (\mathbf{A}_{\backslash k})^{-1} \right]_{k'k''} A_{k''k} \ . \tag{31}$$

Demanding self-consistency between eqs. (24) and (30) implies that

$$V_k = \Lambda_k - \frac{1}{\left[ \left( \mathbf{\Lambda} + \widehat{\mathbf{R}} \right)^{-1} \right]_{kk}} \tag{32}$$

which is the adaptive TAP expression for $V_k$ [10], [11]. We now have a set of fixed point equations for the approximate means $m_k \approx \langle b_k \rangle$ and the $V_k$s which can summarised as follows in terms of the original variables

$$m_k = \tanh\left(\frac{A_k}{T}\left(z_k - \sum_{k'\neq k}(R_{kk'} - \delta_{kk'})A_{k'}m_{k'}\right) - V_k m_k\right) \quad (33)$$

$$V_k = \Lambda_k - \frac{1}{\left[(\mathbf{\Lambda} + \mathbf{A}(\mathbf{R}-\mathbf{I})\mathbf{A}/T)^{-1}\right]_{kk}} \quad (34)$$

$$\Lambda_k = V_k + \frac{1}{1-m_k^2}\ . \quad (35)$$

We immediately observe that whereas the naive mean field theory has a computational complexity of $\mathcal{O}(K^2)$, the matrix inversion in eq. (34) makes the computational complexity of the advanced equations $\mathcal{O}(K^3)$. In section VI we shall see how the expression for $V_k$ can simplified to give the for finite systems less accurate so-called self-averaging TAP equations that are only of $\mathcal{O}(K^2)$. In the next section we describe an effective second order belief propagation method for solving the adaptive TAP eqs. (33)-(35). We finish the theoretical treatment in section VII with a short discussion of mean field annealing.

## V. Second Order Fixed Points Dynamics

The non-linear mean field fixed point equations have to be solved by iteration. We thus have to introduce update rules (or dynamics) with good convergence properties. For the naive mean field eqs. (9) we are in the lucky situation that there are no self-couplings, i.e. in the equation for $m_k$, $m_k$ does not appear on the right hand side inside the non-linearity. Updating with eq. (9) sequentially corresponds to a coordinate descent minimisation of the cost function, the free energy, and are thus guaranteed to bring us to a local minimum [17]. The situation is different for the advanced method because we have self-couplings which is known to complicate the convergence of the iterative procedure. However, they are also necessary ingredients for coping with the bias otherwise present. The matrix inversion in eq. (34) also represent a computational demanding task. In the following we will derive a second order belief propagation method which tries to get an unbiased estimate of the mean cavity field $\langle h_k \rangle_{\backslash k}$ in each update step [39], [35]. It will have good convergence properties and the Hessian matrix–also appearing in eq. (34)–can be computed effectively both for parallel and sequential updates.

The idea is as follows: We decompose the update in three steps. In step 1 we use a linearised version of eq. (33) to get an intermediate prediction for $\mathbf{m}^{\mathrm{new}*}$ in terms of the previous values

$\mathbf{m}^{\rm old}$ and $V_k^{\rm old}$. This expression will be derived below. Secondly, we use eqs. (34) and (35) to update $V_k$ in terms of $\mathbf{m}^{\rm old}$ and $V_k^{\rm old}$. Thirdly, we use eq. (33) with inserted $\mathbf{m}^{\rm new*}$ and $V_k^{\rm new}$ on the right hand side to get $m_k^{\rm new}$:

$$\begin{aligned} m_k^{\rm new} &:= \tanh\left(\langle h_k\rangle_{\backslash k}^{\rm new*}\right) \\ \langle h_k\rangle_{\backslash k}^{\rm new*} &= \widehat{z}_k - \sum_{k'\neq k}\left(\widehat{R}_{kk'} + V_k^{\rm new}\delta_{kk'}\right) m_{k'}^{\rm new*} \,. \end{aligned} \tag{36}$$

By solving for the intermediate value $\langle h_k\rangle_{\backslash k}^{\rm new*}$, we aim in each update step at getting the value as unbiased as possible, i.e. we try to remove the influence of $m_k^{\rm old}$ dynamically. Next we will show how to obtain the linearised solution $\mathbf{m}^{\rm new*}$. We can linearise eq. (33) around the previous update

$$m_k^{\rm old} = \tanh\left(\langle h_k\rangle_{\backslash k}^{\rm old*}\right) \tag{37}$$

which gives the following linear equation for $\mathbf{m}^{\rm new*}$

$$m_k^{\rm new*} - m_k^{\rm old} = (1-(m_k^{\rm old})^2)\left(\widehat{z}_k - \sum_{k'\neq k}\left(\widehat{R}_{kk'} + V_k^{\rm old}\delta_{kk'}\right) m_{k'}^{\rm new*} - \langle h_k\rangle_{\backslash k}^{\rm old*}\right) \tag{38}$$

The solution for $\mathbf{m}^{\rm new*}$ can written in of terms of the vector $\boldsymbol{\alpha}$ with components

$$\alpha_k \equiv \frac{m_k^{\rm old} + (1-(m_k^{\rm old})^2)(\widehat{z}_k - \langle h_k\rangle_{\backslash k}^{\rm old*})}{1 + (1-(m_k^{\rm old})^2)V_k^{\rm old}} \tag{39}$$

and $\Lambda^{\rm old} = 1/(1-(m_k^{\rm old})^2) + V_k^{\rm old}$ as

$$\mathbf{m}^{\rm new*} = \left(\boldsymbol{\Lambda}^{\rm old} + \widehat{\mathbf{R}}\right)^{-1} \boldsymbol{\Lambda}^{\rm old}\boldsymbol{\alpha} \,. \tag{40}$$

We now see that it is the same inverse matrix that appears here and in the expression for $V_k$, eq. (32)[2]. We now have all the ingredients in the belief propagation algorithm: eq. (40) for $\mathbf{m}^{\rm new*}$, eqs. (34) and (35) for $V_k$ and finally eq. (36) for $m_k^{\rm new}$. In the following we will limit ourselves to look at the sequential version of the algorithm and give a computationally effective recipe. The parallel version follows straightforwardly from the derived equations. The main problem is that after we have updated $m_k$ and $V_k$ we have to recompute $\left(\boldsymbol{\Lambda} + \widehat{\mathbf{R}}\right)^{-1}$, i.e. naively an $\mathcal{O}(K^3)$ operation in every step. However, when we make the update, we only change one element in

[2] This matrix plays a special role since it is the advanced mean field (linear response) estimate of the covariance of the variables $\langle b_k b_{k'}\rangle - \langle b_k\rangle\langle b_{k'}\rangle \approx \left[\left(\boldsymbol{\Lambda} + \widehat{\mathbf{R}}\right)^{-1}\right]_{kk'}$ [10], [11].

$\mathbf{\Lambda} + \widehat{\mathbf{R}}$ namely the $(k,k)$th element because we change $\Lambda_k = 1/(1-m_k^2)+V_k$. We make use of the Sherman-Woodbury formula to reduce this step to $\mathcal{O}(K^2)$. For that we introduce the following matrices

$$\mathbf{H} \equiv \left(\mathbf{\Omega} + \widehat{\mathbf{R}}^{-1}\right)^{-1} \tag{41}$$

$$\mathbf{\Omega} \equiv \mathbf{\Lambda}^{-1} \tag{42}$$

which can be updated using the Sherman-Woodbury formula when we change $\Omega_k = 1/\Lambda_k$ by $d\Omega = \Omega_k^{\text{new}} - \Omega_k^{\text{old}}$

$$\mathbf{H}^{\text{new}} = \mathbf{H} - \frac{d\Omega}{1 + d\Omega H_{kk}} \mathbf{H}\mathbf{e}_k (\mathbf{H}\mathbf{e}_k)^{\mathrm{T}} \,, \tag{43}$$

where $\mathbf{e}_k$ is the $k$th unit vector. It is easy to show that $\langle \mathbf{h} \rangle^{\text{new}*} = \widehat{\mathbf{z}} - \widehat{\mathbf{R}}\mathbf{m}^{\text{new}*} = \widehat{\mathbf{z}} - \mathbf{H}\boldsymbol{\alpha}$, $\mathbf{m}^{\text{new}*} = \boldsymbol{\alpha} - \mathbf{\Omega}(\langle \mathbf{h} \rangle^{\text{new}*} - \widehat{\mathbf{z}})$ and $V_k = H_{kk}/(\Omega_k H_{kk} - 1)$.

So far we have not discussed how to initialise the algorithm. A natural initialisation is $\mathbf{\Omega} := \mathbf{0}$. This gives the following simple initialization for $\mathbf{H} = \left(\mathbf{\Omega} + \widehat{\mathbf{R}}^{-1}\right)^{-1} = \widehat{\mathbf{R}}$ so that we completely avoid explicit matrix inversion. Pseudo-code for the algorithm is given in table I.[3] As an additional measure to ensure numerical stability, $\mathbf{H}$ (and $\alpha_k$) are not updated when this will lead to negative diagonal elements in $\mathbf{H}$. The reason is in turn will introduce negative $V_k$s and since $V_k$ is a variance (of the cavity field) it must by construction of the theory be non-negative.

This update scheme has a natural interpretation in terms of propagating beliefs for $m_k$ and $V_k$. As a second order method it is closely related to the Newton method. But in practice it turns out that it is more stable than the Newton method. This can be attributed to two facts: 1. it is more conservative–by linearizing around the previous update–and 2. it uses the contractive non-linearity $\tanh(\cdot)$ directly which keeps the values of $m_k \in [-1; 1]$.

## VI. Large System approximation using Spreading Code Statistics

The advanced mean field method described above is a relatively recent development [10], [11]. The original TAP approach [26], [30] is derived for the case of large system size with known distribution of the randomness. In the CDMA MUD context, the randomness lies in the

[3] Matlab implementations of both this and normal sequential iteration scheme are available from `http://isp.imm.dtu.dk/staff/winther/`.

$\eta_m = 0.9$ ; $\eta_V = 0.5$ ; $\boldsymbol{\alpha} = \mathbf{m}^{\text{init}}$ ; $\mathbf{V} = \mathbf{V}^{\text{init}}$ ; $\mathbf{H} := \frac{1}{T}\mathbf{A}(\mathbf{R}-\mathbf{I})\mathbf{A}$ ; $\mathbf{\Omega} := \mathbf{0}$;

**do**

**Cycle** over random variable indices $k \in [1, K]$

$\langle h_k \rangle^* := A_k z_k / T - \sum_{k'} H_{kk'} \alpha_{k'}$ ;

$m_k^* := \alpha_k - \Omega_k(\langle h_k \rangle^* - A_k z_k / T)$ ;

$V_k := \eta_V \frac{H_{kk}}{\Omega_k H_{kk} - 1} + (1-\eta_V) V_k$ ;

$\langle h_k \rangle^*_{\setminus k} := \langle h_k \rangle^* - V_k m_k^*$ ;

$m_k := \eta_m \tanh(\langle h_k \rangle^*_{\setminus k}) + (1-\eta_m) m_k$;

$\alpha_k := \frac{m_k + (1-m_k^2)(A_k z_k/T - \langle h_k \rangle^*_{\setminus k})}{1+(1-m_k^2)V_k}$ ;

$\Omega_i^{\text{old}} := \Omega_i$ ; $\Omega_i := \frac{1-m_k^2}{1+(1-m_k^2)V_k}$ ; $d\Omega := \Omega_k - \Omega_k^{\text{old}}$ ;

$\mathbf{H} := \mathbf{H} - \frac{d\Omega}{1+d\Omega \mathbf{H}_{kk}} \mathbf{H}\mathbf{e}_k (\mathbf{H}\mathbf{e}_k)^{\text{T}}$ ;

**end** // Cycle over random variable indices $k \in [1, K]$

**while** $\max_{k \in [1,K]} (|m_k^{(t+1)} - m_k^{(t)}| > \text{ftol}$

TABLE I

PSEUDO-CODE FOR THE SEQUENTIAL BELIEF PROPAGATION ALGORITHM. NOTE THAT LEARNING RATES ARE INTRODUCED IN THE $m$ AND $V$ UPDATES TO MAKE THE CODE MORE NUMERICALLY STABLE.

spreading codes $s_k(n)$. For random binary spreading codes $s = \pm 1/\sqrt{N_c}$, $p(s) = \frac{1}{2}\delta_{s,1/\sqrt{N_c}} + \frac{1}{2}\delta_{s,-1/\sqrt{N_c}}$ and equal powers $A_k = A$ in the limit of infinite system size, $K \to \infty$, $V_k$ eq. (34) will become what is called 'self-averaging'. This means that for equal power $A_k = A$, $V_k = V$ independent of the specific realisation of the randomness, i.e. the specific choice of spreading code. The self-averaging value $V = \mathbb{E}_{\mathbf{s}}[V_k]$, where $\mathbb{E}_{\mathbf{s}}[\cdot]$ denotes an average with respect to the spreading code distribution, can be obtained from eq. (34). A detailed calculation is given in Ref. [11]. Here we will only briefly sketch the derivation and give the final result. To compute $V$, we write $\left[(\mathbf{\Lambda} + \widehat{\mathbf{R}})^{-1}\right]_{kk}$ as

$$\left[\left(\mathbf{\Lambda} + \widehat{\mathbf{R}}\right)^{-1}\right]_{kk} = \frac{\partial}{\partial \Lambda_k} \ln \det(\mathbf{\Lambda} + \widehat{\mathbf{R}}) \tag{44}$$

and replace the right hand side by its average. Rather than averaging $\ln\det(\mathbf{\Lambda}+\widehat{\mathbf{R}})$, we average

$$\det{}^{-\frac{1}{2}}(\mathbf{\Lambda}+\widehat{\mathbf{R}}) = \int \frac{d\mathbf{x}}{(2\pi)^{K/2}} e^{-\frac{1}{2}\mathbf{x}^T(\mathbf{\Lambda}+\widehat{\mathbf{R}})\mathbf{x}} \tag{45}$$

which is equivalent for $K \to \infty$ [11]. The final result is

$$V = \frac{K}{N_c}\left(\frac{A^2}{T}\right)^2 \frac{(1-q)}{1+\frac{K}{N_c}\frac{A^2}{T}(1-q)} \tag{46}$$

with $q = \frac{1}{K}\sum_k m_k^2$ and we have to assume equal powers $A_k^2 = A^2$. This result is almost identical to the closely related models: the Hopfield neural network and the linear perceptron with binary weights [11]. In practical situations for systems of even a moderately large size, say $K = \mathcal{O}(100)$, $V_k$ is not close to being self-averaging [10], [11]. However, using the self-averaging expression eq. (46) can still give improvements over naive mean field theory without going beyond $\mathcal{O}(K^2)$ complexity.

## VII. Mean Field Annealing

Annealing was introduced in the context of optimisation with simulated annealing [40]. Soon thereafter annealing was used in conjunction with mean field theory, see e.g. [9]. In mean field theory, annealing amounts to repeatedly solving the mean field equations and lowering the temperature using the old solution as the starting guess. It was first suggested to use annealing in the CDMA context by Ref. [32]. The first numerical results was presented in Ref. [16] and extensive numerical results and a thorough theoretical analysis can be found in Ref. [17].

In Ref. [17], we carried out an extensive analysis of the bifurcation properties of the fixed point solution for the naive equations (9) and derived an upper bound for the critical temperature $T_c$ above which the solution space is convex and the solution thus unique. A similar analysis is more complicated for the advanced equations (33), but we can use the same bound for $T_c$ since the subtractive reaction term in eq. (33) implies that the critical temperature is lower than for the naive theory: $T_c^{\mathrm{TAP}} < T_c^{\mathrm{naive}}$. The bound for the critical temperature is: $T_c^{\mathrm{naive}} < (1-\rho_{\mathrm{min}})A_{\mathrm{max}}^2 < A_{\mathrm{max}}^2$, where $A_{\mathrm{max}} = \max_i A_i$ is the maximal power and $\rho_{\mathrm{min}}$ is the minimal eigenvalue of $\mathbf{R}$ (which is lower bounded by zero).

The bifurcation analysis also showed that it is the high $T$ convex solution that is the relevant solution to track [17]. Using $T_c$ (or its upper bound) as the starting temperature therefore guarantees that at least the initial solution is in same convex subspace as the solution we are after. Tracking the solution carefully with annealing will lead us to the optimal solution for the approximation scheme used. The remaining excess error compared to the exact exponential complexity solution will be due to the approximation error.

The number of local minima will decrease with increasing system size and noise level. However, as already shown in Ref. [16], [17], for realistic system sizes $K = \mathcal{O}(10 - 10^2)$, order one loads and $\mathrm{SNRs}$, the existence of local minima seriously affect the performance if the fixed point equations are iterated directly at $T = \sigma^2$. With annealing we can track the optimal solution from a high temperature $T = A^2_{\max}$ down to the noise level $T = \sigma^2$ (or down to $T = 0$ if we are after hard estimates).

The basic underlying assumption of mean field theory (both naive and adaptive TAP) is that of weak correlations between variables, i.e. such that we can apply the law of large number and the central limit theorem. However, for small system sizes and low temperatures this assumption can break down, e.g. there is a finite probability that two spreading codes are selected co-linear which makes the associated variables completely correlated. The breakdown of the theory is usually indicated by the lack of convergence. When a convergence failure is observed we use the solution at the previous temperature as our final estimate. In the setup described below with $K = 8$ and $N_c = 16$, the fraction of runs where we encounter non-convergence is between 0.0 % – 0.8 % (highest for intermediate $\mathrm{SNRs}$). The non-convergence typically occurs close to $T = \sigma^2$, i.e. at the lowest or second lowest temperature.

## VIII. Monte Carlo Simulations

We have made Monte Carlo studies in order to compare the performance of the naive and advanced mean field algorithms. All bit error rate $(\mathrm{BER})$ points were simulated until $K \times 100$ bit errors were seen. This approximately corresponds to $K \times 50$ independent errors (since errors tends to occur in pairs) giving for $K = 8$ around $5\%$ of error in the $\mathrm{BER}$-estimates. The spreading codes are random vectors with components $\pm 1/\sqrt{N_c}$. We anneal using 10 logarithmic equal spaced temperatures between $A^2$ and $\sigma^2$. We did not make any effort to optimize the annealing scheme. Less temperatures, e.g. the half and/or a linear temperature scale can probably be used with similar performance. We use sequential update of the naive mean field eqs. (9) and the belief propagation approach summarized in table I for the adaptive TAP eqs. (33)–(35). For the self–averaging TAP eqs. (33),(46), we also use a sequential update, updating $V$ before $m_k$. For adaptive TAP we use minimum 3 and maximum 100 iterations at each temperature. The average number of iterations is between 5 and 8 (highest for the lowest $\mathrm{SNR}$s) with a residual error tolerance of $10^{-6}$ on the summed squared deviation of eq. (33).

In figure 1, we plot the $\mathrm{BER}$ versus the signal-to-noise ratio $\mathrm{SNR} = A^2/2\sigma^2$ (with unit power $A_k = A = 1$) for a system of size $K = 8$ and $N_c = 16$, i.e. load one half. This setup, identical to the one used in Ref. [41], has a low $\mathrm{SNR}$ regime where the performance is limited by the noise and an asymptotic regime where the performance is strongly limited by interference even for the individual optimal detector. This setup is very difficult as indicated by the very poor performance of the conventional matched filter receiver and by normal interference cancellation with $\tanh(\cdot)$ tentative decision, i.e. naive mean field without annealing. It can be seen that both the naive and TAP theories, are very close to the optimal curve (found by calculating the posterior symbol marginals exactly) for small $\mathrm{SNR}$s and large $\mathrm{SNR}$s. In the latter limit the fundamental $\mathrm{BER}$-floor is very close to the theoretically predicted $5.33 \cdot 10^{-5}$ [41] which reflect that there is a finite probability for choosing co-linear spreading codes. In the intermediate $\mathrm{SNR}$ region, adaptive TAP outperforms the naive and self-averaging TAP approach, e.g. for $\mathrm{BER} = 10^{-4}$ there is approximately a 2 dB's $\mathrm{SNR}$ gain compared to naive mean field annealing. The self-averaging theory gives a performance in between these two.

In figures 2 and 3, we compare the exact decision statistics $\tanh^{-1} \langle b_k \rangle^{\mathrm{exact}}$ with respectively the decision statistics of TAP $\tanh^{-1} \langle b_k \rangle^{\mathrm{TAP}}$ and naive mean field theory $\tanh^{-1} \langle b_k \rangle^{\mathrm{naive}}$. From the figures it can clearly be seen that adaptive TAP approximately corrects the bias of the naive approach, i.e. we get clouds on both sides of the diagonal for both signs of the decision statistics rather than just on the one side. For larger systems (not shown) the agreement between adaptive TAP and the exact result improves.

Figure 4 shows the performance as a function of detection temperature $T$. This type of plot can indicate how to determine what temperature we need to anneal in order to obtain a given performance. Note that especially for high $\mathrm{SNR}$s, the curve levels off at a temperature much higher than the Bayes temperature $\sigma^2$. In some case we even observe a small decrease in performance when lowering the temperature. If significant, this can be attributed to rare numerical instabilities at very low temperatures.

Figure 5 shows the result of running with a setup similar to the one in used Ref. [42]. For constant $\mathrm{SNR} = 10dB$ and $N_c = 20$, we have tested a range of loads $K/N_c$. In this case the performance improvement (in $\mathrm{BER}$) is even more pronounced with a gain factor of approximately 5/10 compared to naive mean field annealing/linear minimum mean squared error (MMSE) over the whole range of loads used. At a target $\mathrm{BER}$ of $10^{-2}$ we can double the number of users compared to the method proposed in Ref. [42] and compared to conventional hard serial interference cancellation we get 10 times more users.

## IX. Conclusion

In this paper we have proposed a new algorithm for subtractive interference cancellation in Code Division Multiple Access (CDMA) which is based upon the adaptive TAP mean field approach recently developed in machine learning/statistical mechanics [10], [11]. With the adaptive TAP mean field approach we can approximately remove the bias of the conventional (naive mean field) subtractive interference cancellation approach. The approximation is expected to become better with increasing system size and exact in the infinite large system limit, such that this detector in this limit will be identical to the individual optimal detector [10], [11], [30], see also Ref. [32] for an average case analysis of this scenario.

In the simulation studies we compare the mean field approaches and find a significant performance improvement in the regimes where the conventional naive mean field approach deviates from the optimal. We have observed performance improvements in bit error rate of up to a factor 5 over naive mean field theory and a factor of 10 for minimum mean squared error (MMSE) for a whole range of loads $K/N_c$. For both mean field approaches we use mean field annealing to avoid getting trapped in local minima. This has previously been shown to be crucial for obtaining good performance for small to medium sized systems [16], [17].

The computational complexity of the new scheme is $\mathcal{O}(K^3)$ whereas the conventional scales as $\mathcal{O}(K^2)$. In this paper we have proposed an effective second order belief propagation approach to solving the mean field equations which partly compensates for the increased complexity. However, it is of interest to come up with schemes that have a lower computational complexity while retaining at least a part of the bias correction. A good candidate for this is the self-averaging TAP mean field theory. This method uses knowledge of the statistics of the spreading codes and assumes equal powers of all users to simplify the bias correction term. It is expected to become exact in the infinite large system limit, see Ref. [11] and references therein. We have tested the self–averaging TAP theory using a conventional sequential update scheme and found a performance that lies in between that of adaptive TAP and naive mean field. This shows that in the ideal situation where the a priori assumptions hold, we can cancel some of the bias without going beyond $\mathcal{O}(K^2)$ complexity.

It is of practical interest to develop belief propagation type algorithms which scales as $\mathcal{O}(K^2)$ rather than $\mathcal{O}(K^3)$. Recently, Refs. [43], [44] have proposed and analysed $\mathcal{O}(K^2)$ CDMA-algorithms which incorporates a bias correction term. The starting point of the derivation of these algorithm is Pearl's belief propagation algorithm which should work well for sparsely connected systems, i.e. the opposite situation as the CDMA-setup. The algorithm furthermore makes use of the knowledge of the spreading code statistics. So far these algorithms have only been applied to large systems. It is important to investigate the connection to our approach and compare performance for realistic sized systems. The adaptive TAP approach is expected to be quite robust against violations of the basic channel and spreading code properties since it is designed to adapt to the data at hand. The self–averaging approaches on the other hand will not be as robust. Exactly how robust they are remains to be tested empirically. As a preliminary test of this we tried a self-averaging term corresponding to the user correlations being finite variance i.i.d. variables, i.e. equivalent to the SK-spin glass model, but the results were worse than not using any correction, indicating that the self-averaging method is sensitive to the actual statistics of the user correlations.

## Acknowledgements

The first author would like to thank Nokia Mobile Phones RetD in Denmark for encouraging and supporting his work on CDMA detection.

## References

[1] Sergio Verdú, "Minimum probability of error for asynchronous gaussian multiple-access channels," *IEEE Transactions on Information Theory*, vol. IT-32, pp. 85–96, January 1986.

[2] Sergio Verdú, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, pp. 303–312, 1989.

[3] Shimon Moshavi, "Multi-user detection for ds-cdma communications," *IEEE Communication Magazine*, vol. 34, no. 10, pp. 124–136, October 1996.

[4] Mahesh K. Varanasi and Behnaam Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Transactions on Communications*, vol. 38, no. 4, pp. 509–519, April 1990.

[5] Mahesh K. Varanasi and Behnaam Aazhang, "Near-optimum detection in synchronous code-division multiple-access systems," *IEEE Transactions on Communications*, vol. 39, no. 5, pp. 725–736, May 1991.

[6] Alexandra Duel-Hallen and Chris Heegard, "Delayed decision-feedback sequence estimation," *IEEE Transactions on Communications*, vol. 37, no. 5, pp. 428–436, May 1989.

[7] P. H. Tan, *Multiuser Detection in CDMA - Combinatorial Optimization Methods*, Ph.D. thesis, Chalmers University of Technology, Department of Computer Engineering, Göteborg, Sweden, November 2001.

[8] George I. Kechriotis and Elias S. Manolakos, "Hopfield neural network implementation of the optimal cdma multiuser detector," *IEEE Transactions on Neural Networks*, vol. 7, no. 1, pp. 131–141, January 1996.

[9] Carsten Peterson and Bo Söderberg, "A new method for mapping optimization problems onto neural networks," *International Journal of Neural Systems*, vol. 1, pp. 3–22, 1989.

[10] Manfred Opper and Ole Winther, "Tractable approximations for probabilistic models: The adaptive thouless-anderson-palmer mean field approach," *Physical Review Letters*, vol. 86, pp. 3695–3699, 2001.

[11] Manfred Opper and Ole Winther, "Adaptive and self-averaging thouless-anderson-palmer mean field theory for probabilistic modeling," *Physical Review E*, vol. 64, pp. 056131, 2001.

[12] Manfred Opper and David Saad, *Advanced Mean Field Methods: Theory and Practice*, MIT Press, July 2001.

[13] Pedro A.d.F.R Højen-Sørensen, Ole Winther, and Lars Kai Hansen, "Mean field approaches to independent component analysis," *Neural Computations*, vol. 14, pp. 889–918, 2002.

[14] Thomas Fabricius and Ole Nørklit, "Approximations to joint-ml and ml symbol channel estimators in mud cdma," in *Proceedings of IEEE Globecom 2002*. To Appear, IEEE.

[15] Dariush Divsalar, Marvin K. Simon, and Dan Raphaeli, "Improved parallel interference cancellation for cdma," *IEEE Transaction on Communications*, vol. 46, no. 2, pp. 258–268, February 1998.

[16] Thomas Fabricius and Ole Winther, "Improved multistage detector by mean-field annealing in multi-user cdma," in *Proceedings of IEEE Inteligent Signal Processing, Applications, and Communication Systems*. To Appear, IEEE.

[17] T. Fabricius and O. Winther, "Analysis of mean field annealing in substactive interference cancellation," *Submitted to IEEE Transactions on Communications*, 2002.

[18] Giorgi Parisi, *Statistical Field Theory*, Addison-Wesley Publishing Company, New York, 0-201-05985-1.

[19] Hidetoshi Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Number 111 in International Series of Monographs on Physics. Oxford University Press, 2001.

[20] N. Sourlas, "Spin-glass models as error-correcting codes," *Nature*, vol. 338, pp. 693–695, 1989.

[21] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, 1988.

[22] N. Wiberg, *Codes and Decoding on General Graphs*, Doctoral dissertation, Department of Electrical Engineering, Linkoping University, Linkoping Sweden, 1996.

[23] Brendan J. Frey, *Graphical Models fo Machine Learning and Digital Communication*, MIT Press, 1998.

[24] R. J. McEliece, D. J. C. MacKay, and J. F. Cheng, "Turbo-deoding as an instance of pearl's 'belief propagation' algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 140–52, February 1998.

[25] J.S. Yedidia, W.T. Freeman, and Y Weiss, "Generalized belief propagation," in *Advances in Neural Information Processing Systems (NIPS)*. June 2000, vol. 13, pp. 689–695, MIT Press.

[26] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of a 'solvable model of a spin glass'," *Phil. Mag.*, vol. 35, pp. 593, 1977.

[27] David Sherrington and Scott Kirckpatrick, "Solvable model of a spin-glass," *Phys. Rev. Letter*, vol. 35, no. 26, pp. 1792–1796, December 1975.

[28] D. Sherrington and S. Kirckpatrick, "Infinite-ranged models of spin-glasses," *Phys. Rev. B*, vol. 17, no. 11, pp. 4384–4403, June 1978.

[29] Max Welling and Yee Whye Teh, "Approximate inference in boltzmann machines," *To appear in Artificial Intelligence Journal*, 2001.

[30] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*, vol. 9 of *Lecture Notes in Physics*, World Scientific, 1987.

[31] Peng Hui Tan, Lars K. Rasmussen, and Teng J. Lim, "The application of semidefinite programming for detection in cdma," *IEEE Journal Selected Areas in Communication*, vol. 19, no. 8, pp. 1442–1449, August 2001.

[32] Toshiyuki Tanaka, "Large-system analysis of randomly spread cdma multiuser detectors," *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2888–2910, 2002.

[33] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," vol. 79, pp. 2554–8, 1982.

[34] H. S. Seung, H. Sompolinsky, and N. Tishby, "Statistical mechanics of learning from examples," *Phys. Rev. A.*, vol. 45, pp. 6056, 1992.

[35] Manfred Opper and Ole Winther, "Approximate probabilistic inference: Mean field theory and practice," *in preparation*, 2002.

[36] Sergio Verdú, *Multiuser Detection*, Cambridge University Press, The Pitt Building, Trumpington Street, Cambridge, CB2 1RP United Kingdom, 1998.

[37] J. Lou, K.R. Pattipati, P.K. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous cdma using probabilistic data association," *IEEE Communication Letters*, vol. 5, no. 9, pp. 361–363, September 2001.

[38] Manfred Opper and Ole Winther, "Gaussian processes for classification: Mean field algorithms," *Neural Computation*, vol. 12, pp. 2655–2684, 2000.

[39] T. Minka, *Expectation Propagation for Approximate Bayesian Inference*, Doctoral dissertation, MIT Media Lab (2001), 2001.

[40] S. Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

[41] Lars K. Rasmussen, Teng J. Lim, and Hiroki Sugimoto, "Fundamental ber-floor for long-code cdma," in *Proceeding of 1998 IEEE 5th International Symposium on Spread Spectrum Techniques & Applications, IEEE ISSSTA*. 1998, vol. 3, pp. 692–695, IEEE.

[42] G. Barriac and U. Madhow, "Low-complexity multiuser detection using parallel arbitrated successive interference cancellation," *Preprint*, May 2002.

[43] Y. Kabashima, "A statistical-mechanical approach to cdma multiuser detection: propagating beliefs in a densely connected graph," in *2003 IEEE International Symposium on Information Theory*, 2003.

[44] Toshiyuki Tanaka and Masato Okada, "Approximate belief propagation, density evolution, and multiuser detection of code-division multiple–access systems," *preprint Tokyo Inst. of Tech.*, 2002.

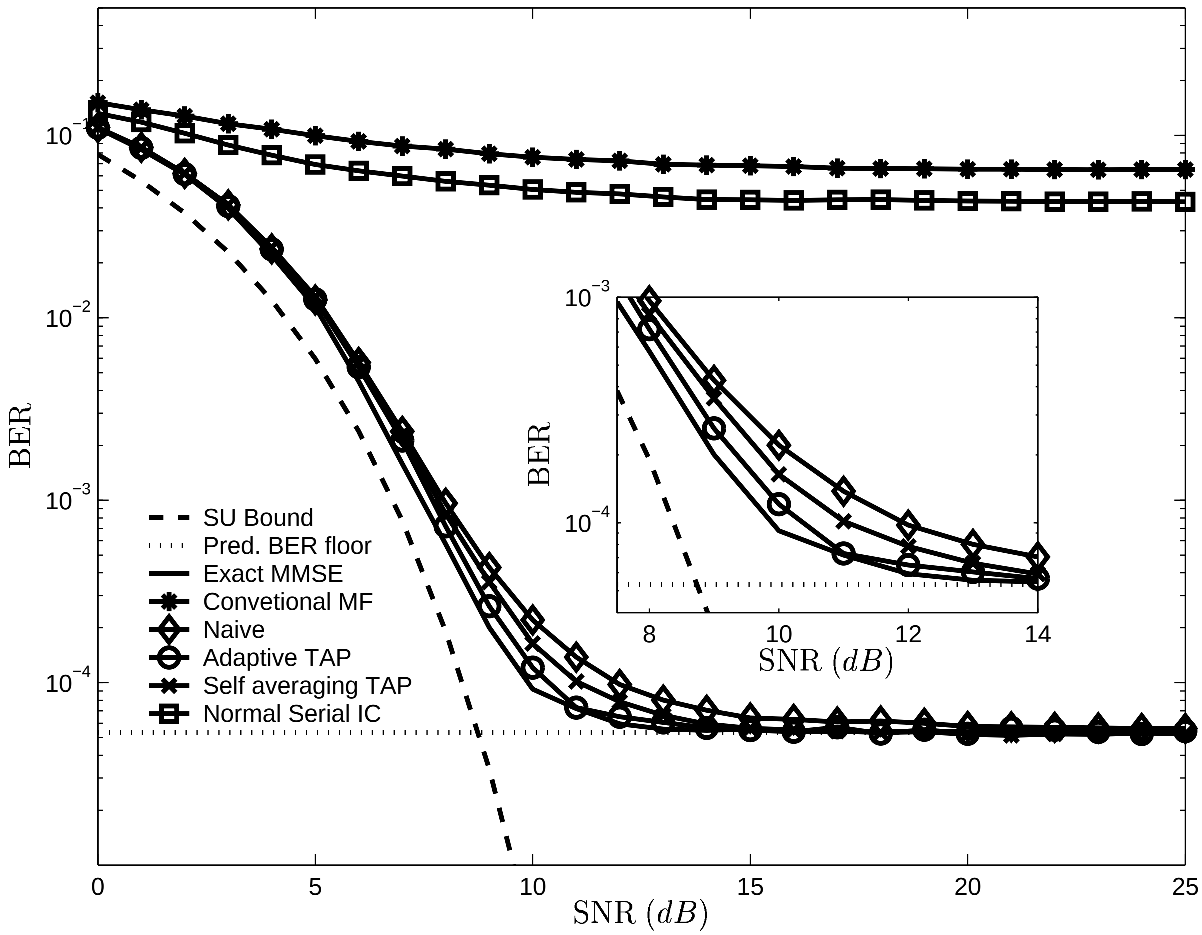


Fig. 1. User averaged bit error rate BER, for $K = 8$ and $N_c = 16$ against signal-to-noise ratio $\mathrm{SNR} = 1/2\sigma^2$. Results for following approaches are shown: adaptive TAP, self-averaging TAP, naive mean field annealing together with the single user bound, exact MMSE, predicted BER-floor, conventional matched filter, and conventional serial interference cancellation with $\mathtt{tanh}(\cdot)$ soft tentative decision, i.e. identical to naive mean field without annealing.

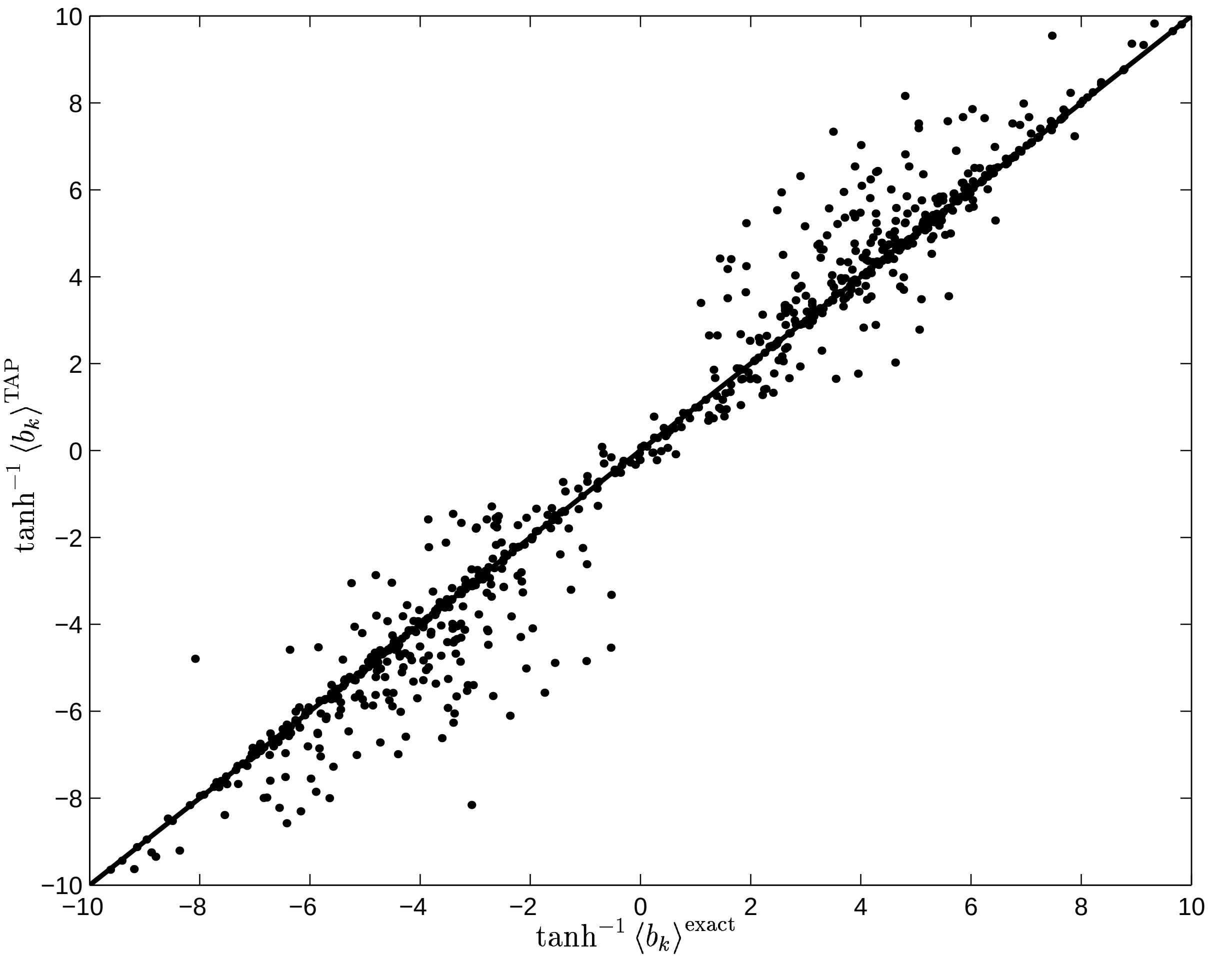


Fig. 2. Samples of exact versus adaptive TAP estimate of the final decision statistic $\tanh^{-1}\langle b_k\rangle$ for $\mathrm{SNR} = 4dB$.

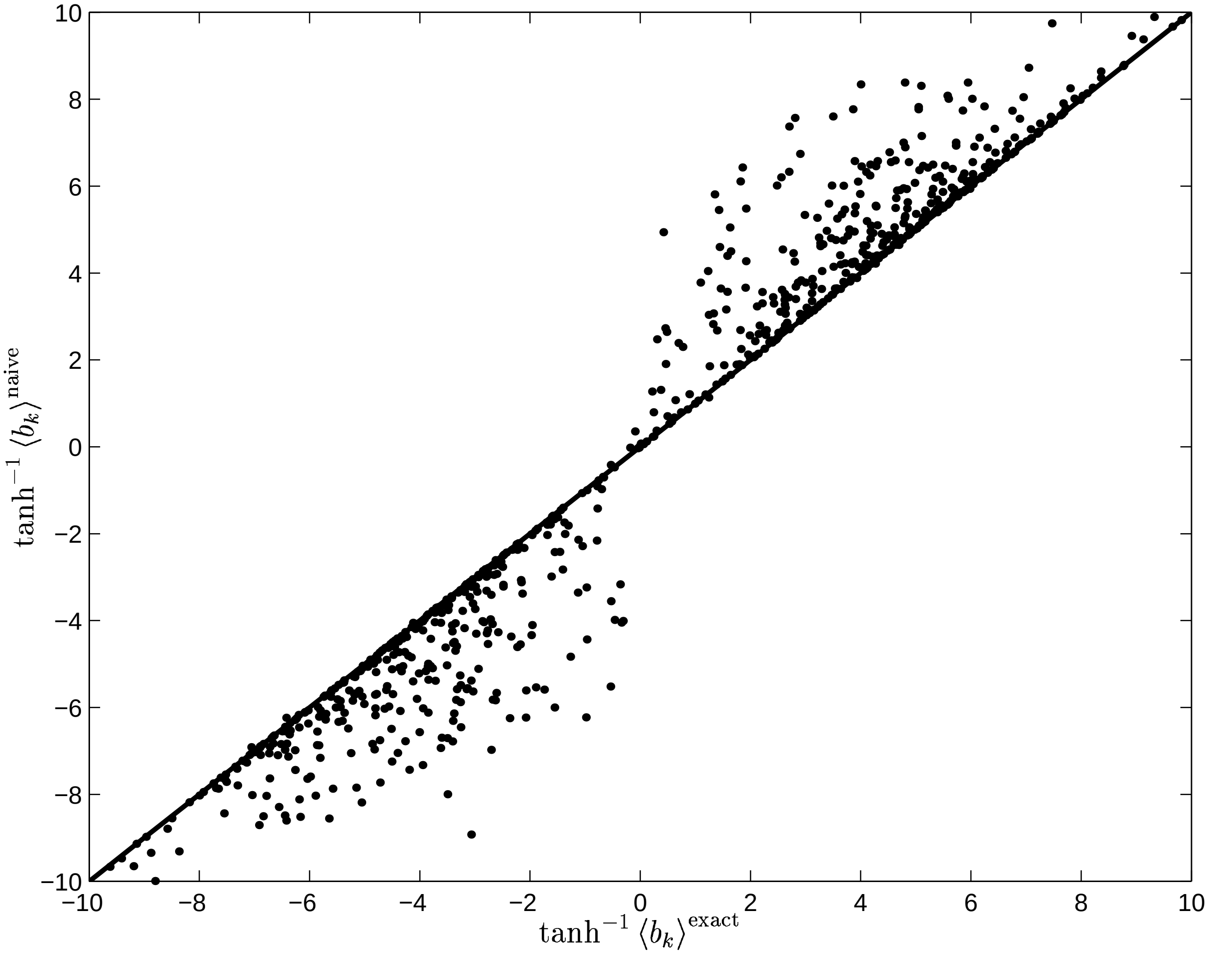


Fig. 3. Samples of exact versus the naive estimate of the final decision statistic $\tanh^{-1}\langle b_k\rangle$ for SNR $= 4dB$.

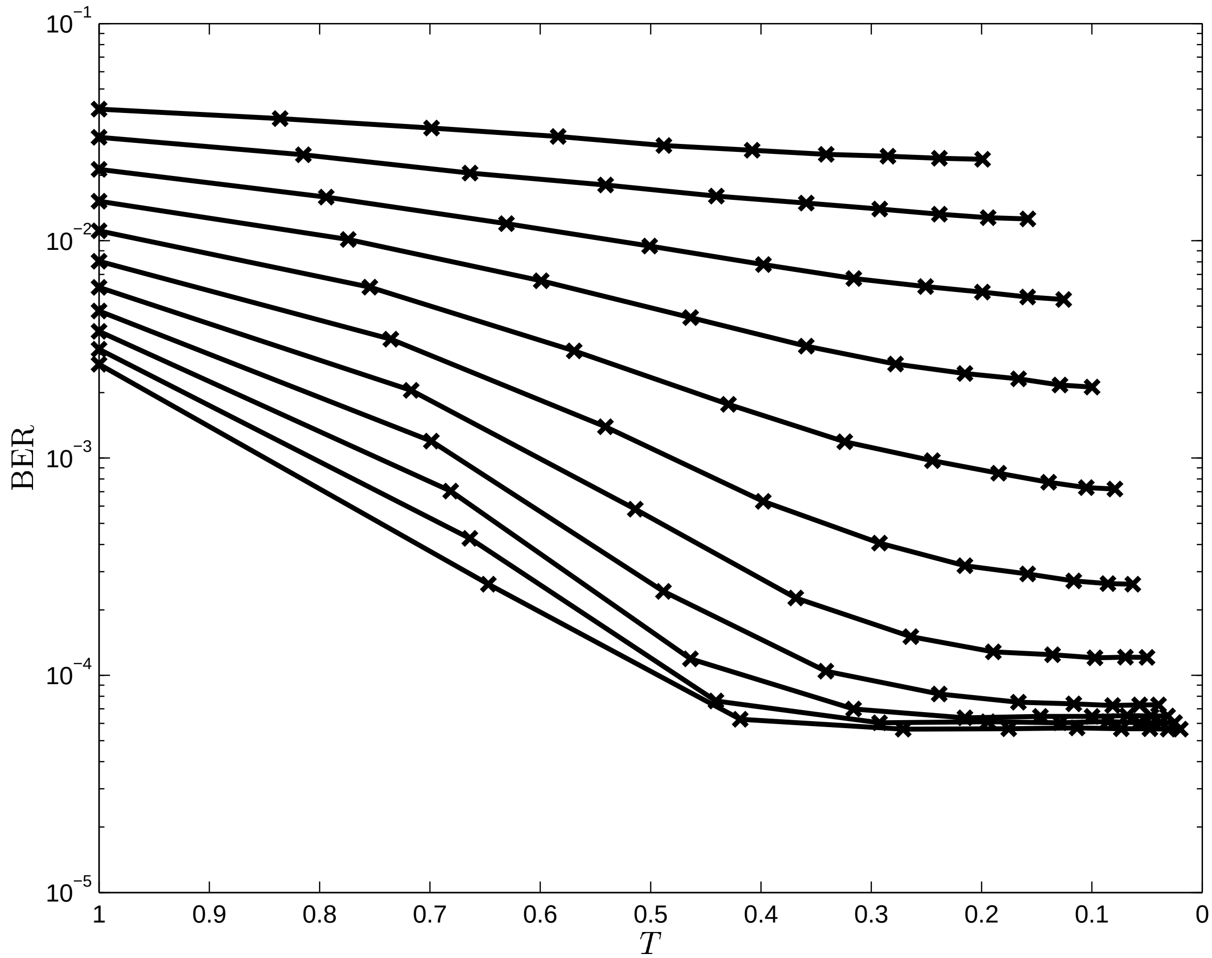


Fig. 4. BER performance of adaptive TAP versus decoding temperature $T$ for a range of different SNRs starting from 4 for the top curve and increasing in steps of one to 14 at the bottom. Crosses indicates the temperatures used. The first temperature is 1 corresponding to $A^2 = 1$ and the final temperature is $\sigma^2$.

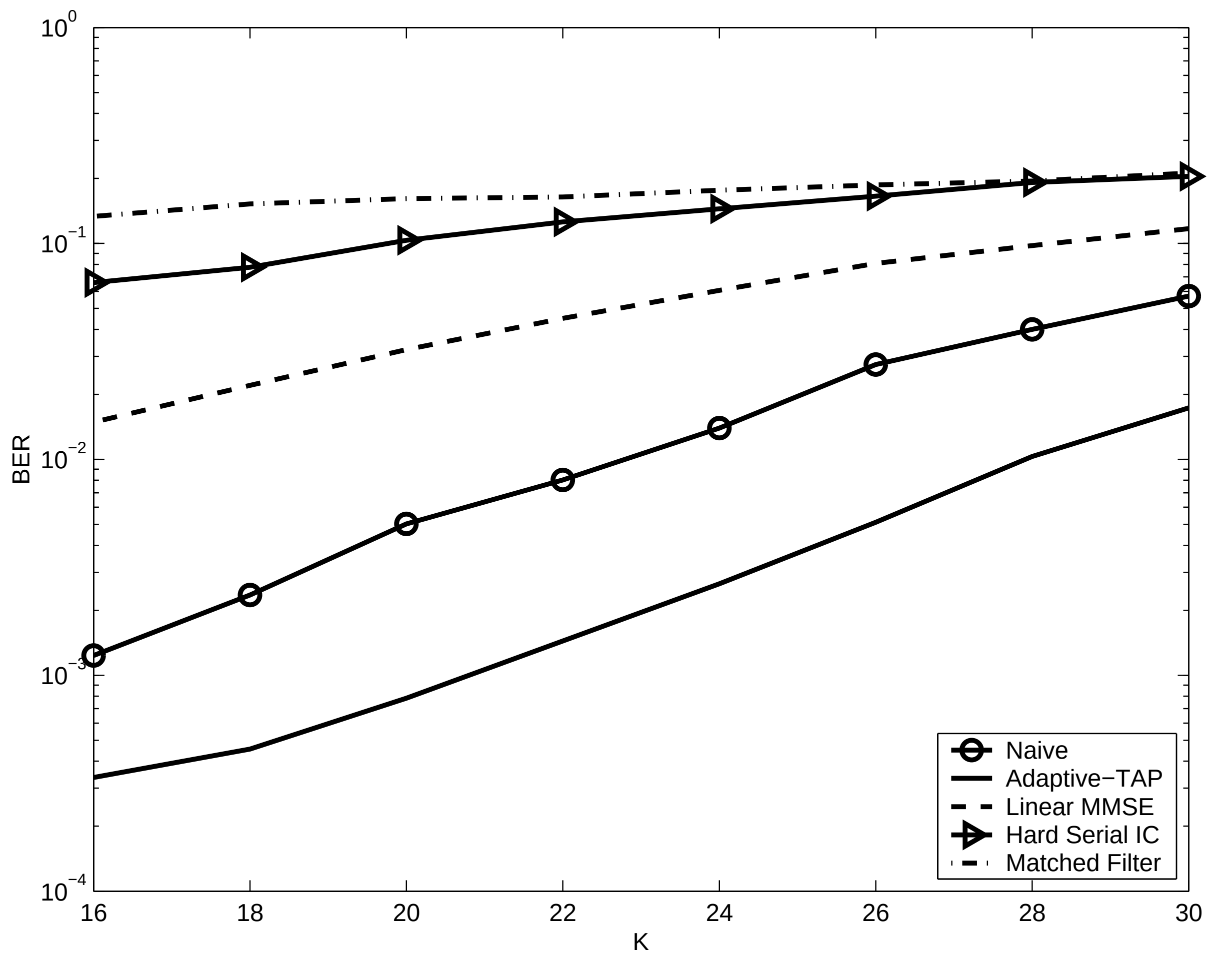


Fig. 5. BER versus the number of users $K$ for spreading factor $N_c = 20$ at $\mathrm{SNR} = 10$ dB. The curves from the bottom are adaptive TAP mean field annealing (full line), naive mean field annealing (full line with circles), linear MMSE (dashed line), hard serial IC (full line with triangles) and matched filter (dashed–dotted).