Improved Multistage Detector by Mean-Field Annealing in Multi-User CDMA

Thomas Fabricius, and Ole Winther
Informatics and Mathematical Modelling
Technical University of Denmark
Copenhagen, Denmark

Abstract— **We propose a multistage multi-user detector for Code Division Multiple Access that implements Mean-Field Annealing. Annealing is a well known technique to avoid local minima in detection and optimisation problems. Local minima certainly occurs in CDMA multi-user detection leading to anomalous Bit Error Rate curves. Annealing is implemented as a standard multi-user detector implementing interference cancellation, but with a control parameter determining the slope of the squashing/mapping function. We derive a bound on the control parameter for the first stage of the detector. We also propose an annealing scheme. We show empirically the performance of the proposed detector, which follows closely the optimum detector implemented by exhaustive search, i.e. has gains of several dB's compared to standard multi-user detectors like linear-MMSE, linear-MMSE first stage followed by interference cancellation, and multi-stage interference cancellation with clipped soft decision, the last being the closest to our proposed detector. The proposed detector gains up to 2 dB compared to the Clipped Soft Decision Multistage detector at no added complexity.**

Keywords— **CDMA, Multistage Multi-User Detection, Interference Cancellation, Anomalous BER-curves.**

I. INTRODUCTION

Direct-Sequence Spread-Spectrum code division multiple access (DS-CDMA) systems have desirable properties to achieve a high spectral efficient system. These are robustness against channel impairments, such as dispersion and fading, having graceful degradation, and ease of cellular planning by dynamic channel sharing at various data rates [1]; all resulting in smooth handling of different Quality of Services (QoS) and a high overall network capacity. In order to achieve this high capacity the detector has to implement Maximum Likelihood (ML) detection which is well known to have exponential complexity in the number of users and the coherence time of the channel. Based on this fact much attention has been used in deriving near optimal low complexity receivers [2]. One of the most promising suggestions for such a detector is multistage Interference Cancellation, either implemented by successive or parallel structure resulting in the (SIC) or (PIC) [3], [4]. In these detectors a non-linearity is used to separate out the transmitted symbols from various users. In case of binary signalling the most used non-linearities are sign, tangent hyperbolic or clipped soft decision. In the first stages of the detector linear functions are often used instead of the non-linearity [2] i.e. implementing iteratively a decorrelating or Linear MMSE detector in the first stages [5]. The multistage detector with sign squashing function can be seen as a local search based optimiser for the ML solution, which has convergence to a (local) minimum of the negative log-likelihood. Convergence to a local minimum is also observed for the softer non-linearities hyperbolic tangent and clipped soft decision. In this contribution will we empirically show that these local minima of the negative log-likelihood lead to a Bit Error Rate (BER) curve that is in excess of the inherent BER behaviour of the optimal detector. This excess BER behaviour should be contrasted to the inherent BER-floor reported in [6] though related. We will explain this relation. The local minima lead to an excess BER for the local search based multistage receiver compared to that using exact enumeration. This fact leads us to the main contribution, namely a multistage detector structure that implements Annealing. Annealing is a well known [7] general applicable heuristic to avoid local minima in search based NP-hard optimisation like the travelling salesman problem, graph partitioning etc. Annealing originates from condensed matter physics [8]. Here annealing is the process in which a solid in a heat bath is heated up to a level where all the particles has a total random ordering, then the heat bath is cooled down slowly so the particles can arrange them self in a highly ordered lattice structure, the temperature at which the solid goes from dominating random order to dominating lattice order is the critical temperature. The kind of annealing we will use is Mean-Field Annealing. We will suggest an annealing scheme that avoids most of the local minima, maintaining the polynomial law complexity of the conventional multistage receivers. We show empirically that this scheme obtains a similar performance to exact enumeration. The paper is organised as follows: In section II we describe the modulation and channel model, in III we review optimal detectors, in section IV we explain the concept of equilibrium distributions, in section V we derive the mean-field detector, in VI we describe the annealing heuristic and derive a conservative estimate of the (inverse) starting temperature, then this is followed by Monte Carlo simulations in section VII, and finally we conclude VIII.

II. K USERS CHIP SYNCHRONOUS CDMA MODEL

Assuming a stationary AWGN channel, and BPSK modulation the received base band CDMA signal can be modelled as

$$\mathbf{y} = \mathbf{S}\mathbf{b} + \epsilon \tag{1}$$

after chip waveform matched filtering, assuming the chip waveform to fulfil the Nyquist criterion. Where $\mathbf{y} \in \mathcal{R}^{\mathcal{N}}$ is the received signal, $\mathbf{b} \in [-1, 1]^K$ are the transmitted bits for the K users, $\mathbf{S} \in [\frac{-1}{\sqrt{N}}, \frac{1}{\sqrt{N}}]^{NxK}$ are the spreading codes for the K users with N chips and unit energy, and $\epsilon \in \mathcal{R}^{\mathcal{N}}$ is zero mean white Gaussian noise with variance $\sigma^2 = \frac{N_0}{2}$.

We process the received signal $\mathbf{y}$ by a bank of filters matched to the spreading codes obtaining

$$\mathbf{z} = \mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{b} + \mathbf{S}^{\mathrm{T}}\epsilon = \mathbf{R}\mathbf{b} + \mathbf{n} \tag{2}$$

where we have defined the correlation matrix of the spreading codes $\mathbf{R} = \mathbf{S}^{\mathrm{T}}\mathbf{S} \in \mathcal{R}^{\mathbf{KxK}}$ and the transformed noise $\mathbf{n} = \mathbf{S}^{\mathrm{T}}\epsilon$ which now has covariance $\sigma^2\mathbf{R}$. All together this defines the likelihood of the transmitted bits assuming everything else for known

$$p(\mathbf{z}|\,\mathbf{b}) = \left|2\pi\sigma^2\mathbf{R}\right|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{z}-\mathbf{R}\mathbf{b})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{z}-\mathbf{R}\mathbf{b})} \tag{3}$$

III. ML AND MEAN POSTERIOR DETECTOR

This section just serves as a review of optimal detectors [2]. The detector that minimises the Bit Error Rate (BER) is the sign mean posterior estimator

$$\begin{aligned}\widehat{\mathbf{b}} &= \operatorname{sgn}\left[\frac{\sum_{\mathbf{b}\in[-1,1]^K}\mathbf{b}p\left(\mathbf{z}|\,\mathbf{b}\right)}{\sum_{\mathbf{b}\in[-1,1]^K}p\left(\mathbf{z}|\,\mathbf{b}\right)}\right]\\ &= \operatorname{sgn}\left[\frac{\sum_{\mathbf{b}\in[-1,1]^K}\mathbf{b}e^{-\alpha E(\mathbf{b})}}{\sum_{\mathbf{b}\in[-1,1]^K}e^{-\alpha E(\mathbf{b})}}\right]_{\alpha=\frac{1}{\sigma^2}}\end{aligned} \tag{4}$$

under equal apriori probability of the transmitted bits, and where we have defined

$$E(\mathbf{b}) = \frac{1}{2}\left(\mathbf{b}^{\mathrm{T}}(\mathbf{R}-\mathbf{I})\mathbf{b} - \mathbf{2}\mathbf{b}^{\mathrm{T}}\mathbf{z}\right), \tag{5}$$

$\mathbf{I}$ being the identity matrix. The detector that minimises the probability of error is the maximum likelihood receiver

$$\begin{aligned}\widehat{\mathbf{b}} &= \arg\max_{\mathbf{b}\in[-1,1]^K} p\left(\mathbf{z}|\,\mathbf{b}\right) = \arg\min_{\mathbf{b}\in[-1,1]^K} E(\mathbf{b})\\ &= \lim_{\alpha\to\infty}\frac{\sum_{\mathbf{b}\in[-1,1]^K}\mathbf{b}e^{-\alpha E(\mathbf{b})}}{\sum_{\mathbf{b}\in[-1,1]^K}e^{-\alpha E(\mathbf{b})}}.\end{aligned} \tag{6}$$

As seen both detectors can be implemented by the same summation structure, but with different α, α acts as a control parameter which in physics would correspond to the inverse temperature. We also defined $E(\mathbf{b})$ which can be viewed as the cost function of the problem or the energy. It should be noticed that both of these detectors has exponential complexity.

IV. EQUILIBRIUM DISTRIBUTION

Before we start with the derivation of the detector we first look into the concept of equilibrium distributions via the unnormalised KL-divergence, in physics denoted the *free energy*. We formulate this for general α. The *ergodic equilibrium distribution* $q_\alpha(\mathbf{b})$ at a specified α is the distribution that minimises the free energy defined as

$$\mathcal{F}_\alpha = \alpha\langle E(\mathbf{b})\rangle_\alpha + \langle\log \mathbf{q}_\alpha(\mathbf{b})\rangle_\alpha \tag{7}$$

where we define $\langle\cdot\rangle_\alpha$ as the average with respect to $q_\alpha(\mathbf{b})$. We see that the minimum is obtained for the $q_\alpha(\mathbf{b})$ that also minimizes the KL-divergence

$$\mathcal{KL}\left[q_\alpha(\mathbf{b}), \frac{e^{-\alpha E(\mathbf{b})}}{Z_\alpha}\right] = \alpha\langle E(\mathbf{b})\rangle_\alpha + \langle\log q_\alpha(\mathbf{b})\rangle_\alpha + \log Z_\alpha \tag{8}$$

since $Z_\alpha = \sum_{\mathbf{b}\in[-\mathbf{1},\mathbf{1}]^{\mathbf{K}}} e^{-\alpha E(\mathbf{b})}$ the normalising constant is independent of $q_\alpha(\mathbf{b})$. We see from this that when we can choose $q_\alpha(\mathbf{b})$ freely to globally minimise the KL-divergence we get the ergodic equilibrium distribution

$$q_\alpha(\mathbf{b}) = \frac{e^{-\alpha E(\mathbf{b})}}{Z_\alpha} \tag{9}$$

i.e. the posterior at the fixed α, with zero KL-divergence.

The two optimal detectors considered in the last section were found by knowing the posterior mean $\langle\mathbf{b}\rangle_\alpha$ at $\alpha = \frac{1}{\sigma^2}$ and $\alpha\to\infty$ respectively when $q_\alpha(\mathbf{b})$ is the ergodic equilibrium distribution. Until now we have rewritten the posterior distribution as the ergodic equilibrium distribution, although nice we haven't gained anything in terms of complexity. For that reason we constrain the family of $q_\alpha(\mathbf{b})$ so that averages like $\langle E(\mathbf{b})\rangle_\alpha$ and $\langle\log q_\alpha(\mathbf{b})\rangle_\alpha$ become tractable. Such a constraint implies that we can not minimize the KL-divergence to zero, instead we define a *constrained equilibrium distribution* $q_\alpha(\mathbf{b})$ as a distribution that within the constrained family is a (local) minimum of the free energy, which approximates the ergodic equilibrium distribution.

V. CONSTRAINING THE EQUILIBRIUM DISTRIBUTION, MEAN-FIELD DETECTOR

In this section will we constrain the distribution $q_\alpha(\mathbf{b})$ to the factorised family of distributions over $\mathbf{b}\in[-1,1]^K$ with $b_k\in[-1,1]$, $k\in[1,K]$ being the k^{th}-element in the vector $\mathbf{b}$

$$\begin{aligned}q_\alpha(\mathbf{b}) &= \prod_{k=1}^{K} q_\alpha^{(k)}(b_k)\\ &= \prod_{k=1}^{K}\left[\frac{1+m_\alpha^{(k)}}{2}\right]^{\frac{1+b_k}{2}}\left[\frac{1-m_\alpha^{(k)}}{2}\right]^{\frac{1-b_k}{2}},\end{aligned} \tag{10}$$

where $m_\alpha(k)$ for $k\in[1,K]$ parameterises the distribution at the specified α. We have $\langle b_k\rangle_\alpha = m_\alpha(k)$ and hence $\langle\mathbf{b}\rangle_\alpha = \mathbf{m}_\alpha$ where $\mathbf{m}_\alpha$ is the vector of the $m_\alpha^{(k)}$'s. We now calculate the free energy (7), with respect to this distribution

$$\begin{aligned}\mathcal{F}_\alpha &= \alpha\frac{1}{2}\left(\mathbf{m}_\alpha^{\mathrm{T}}(\mathbf{R}-\mathbf{I})\mathbf{m}_\alpha - \mathbf{2}\mathbf{m}_\alpha^{\mathrm{T}}\mathbf{z}\right)\\ &+ \sum_{k=1}^{K}\frac{1+m_\alpha(k)}{2}\log\frac{1+m_\alpha(k)}{2}\\ &+ \frac{1-m_\alpha(k)}{2}\log\frac{1-m_\alpha(k)}{2}\end{aligned} \tag{11}$$

where we used that the distribution factorises. The *constrained equilibrium distribution* is the distribution that is a minimum of the constrained free energy i.e.

$$\frac{\partial\mathcal{F}_\alpha}{\partial m_\alpha^{(k)}} = 0 \qquad k\in[1,K]. \tag{12}$$

We have

$$\frac{\partial\mathcal{F}_\alpha}{\partial m_\alpha^{(k)}} = \alpha\left[((\mathbf{R}-\mathbf{I})\mathbf{m}_\alpha)_k - (\mathbf{z})_k\right] + \frac{1}{2}\log\frac{1+m_\alpha^{(k)}}{1-m_\alpha^{(k)}} = 0 \tag{13}$$

isolating $m_\alpha^{(k)}$ we get the mean-field equations

$$m_\alpha^{(k)} = \tanh\left[\alpha\left(\mathbf{z}-(\mathbf{R}-\mathbf{I})\mathbf{m}_\alpha\right)_k\right] \qquad k\in[1,K]. \tag{14}$$

Iterating the mean-field equations at a specified α for the different $k\in[1,K]$ is our multistage mean-field detector, each iteration corresponding to one stage. When converged we have obtained the constrained equilibrium distribution. We see that in the limit $\alpha\to\infty$ we have

$$m_\infty^{(k)} = \operatorname{sgn}\left[\left(\mathbf{z}-(\mathbf{R}-\mathbf{I})\mathbf{m}_\infty\right)_k\right] \qquad k\in[1,K]. \tag{15}$$

corresponding to local ML optimisation, also known as hard interference cancellation. If we set $\alpha = \frac{1}{\sigma^2}$ we have the distribution in the family that approximates the mean posterior detector, also known as soft interference cancellation. If we update one $m_\alpha^{(k)}$ at a time, we have Successive Interference Cancellation (SIC),

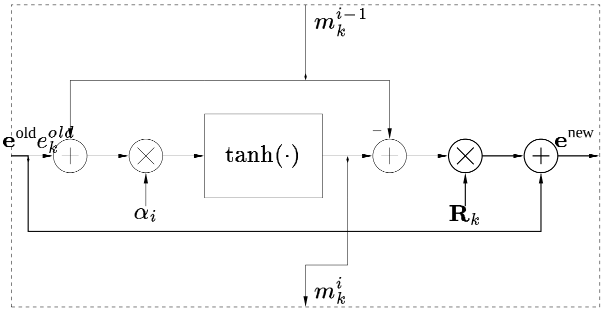

Fig. 1. Mean-Field Successive Interference Cancellation Unit.

if we update all we have parallel (PIC). In this contribution will we restrict our selves to successive updates, since this guaranties that we don't increase the free energy (7) in any iteration/stage. On figure (1) we have drawn a Mean-Field Successive Interference Cancellation Unit (MF-SICU), corresponding to the update of equation (14) for one k, bold lines means vector signals and thin lines are scalar signals, k indicates the user and i indicates the stage/iteration. If we wire together K MF-SICU's we have one stage, on figure (2) a typical D-stage mean-field detector is shown.

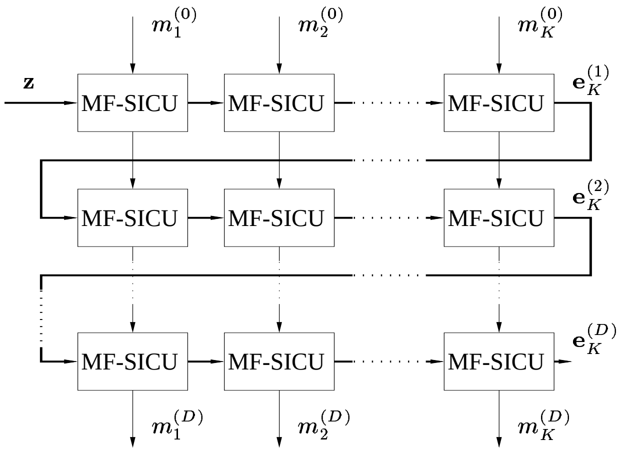

Fig. 2. D-stage Mean-Field Successive Interference Cancellation detector.

VI. Mean-Field Annealing

The reason for introducing annealing is due to the non-convex structure of the energy (5). The existence of local minima is due to the correlation matrix of the spreading codes $\mathbf{R}$ being close to singular or singular. This is also described in [6], in the extreme case where $-\mathbf{R}$ is exact singular. This implies equal energy of a local minimum and the true minimum implying that even the ML-detector implemented by exhaustive search will for some fraction of the possible bit vectors $\mathbf{b}$ report an erroneous minimum, this leads to the inherent BER-floor of the exhaustive search ML detector. What we are considering here is similar, namely close to singular behaviour which leads to local minima, the closer to singular the closer the corresponding energy is to the global energy minimum. In this case, local search for the global minimum will have difficulties in avoiding the local minima and hence will experience a BER-curve in excess of the exhaustive ML BER-curve. All these effects of course have to be compared with the noise level $\sigma^2 = \frac{N_0}{2}$. If the effects are small compared to the noise level, i.e. have low probability of leading to a erroneous detection compared to the probability of the noise making an erroneous detection, the BER-curve will follow the normal 'water fall' behaviour. At some SNR the effect of the local minima can be seen as a bifurcation away from the exhaustive ML BER-curve. At an even higher SNR where the singular effect becomes dominating the ML BER-curve will diverge from the 'water fall' behaviour, because of the erroneous minima with the same energy. These three domains are in full accordance with large system limit results [9], [10].

To avoid local minima we use annealing. The problem to be solved is that of finding the global minimum amongst many local minima in the objective function. The idea is to weight a non-convex parts of the objective function relative to a convex part of the objective function. Here the free energy of the previous chapter can be viewed as the objective function and α as the relative weight. When $\alpha = 0$ we only have a convex part in the objective function, i.e. one minimum in the corresponding free energy $\mathcal{F}_0$. The convex part is believed to be dominating for α up to some critical α_c, so below α_c the objective function still only poses one minimum. Just above α_c more than one minimum exists, amongst which the global one, if unique, has to be found. The closer α comes to α_c the closer the location of the solution comes to the global minimum at the desired α_D. If the minimum from just below α_c is used as an initial guess to the solution just above α_c the obtained minimum amongst all the local minima will be the one that lies closest. Which we hope is the global one. Since the global minimum has the lowest energy it also dominates the most, so the hope can be fulfilled with high probability. The increase in α from below to above α_c has to be carried out slowly in order to get as close to the global minimum as possible before crossing α_c. The reason is the way we solve for the minimum at the specified α, because if we take a too large step in one direction, then this influences the other directions. Large steps obviously occur if we change α in large steps over α_c.

Mean-Field annealing can now be implemented by choosing a serie of accenting α's: $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_D$, and solving the equations (14) for each of the α's, using the solution $\mathbf{m}_\alpha^{\mathbf{i-1}}$ found at the previous α_{i-1} as initial guess $\mathbf{m}_\alpha^{\mathbf{i}}$ on the actual α_i. At α_1 some arbitrary initial guess on $\mathbf{m}_\alpha^{\mathbf{1}}$ can be used, since if $\alpha_1 < \alpha_c$ only one solution exists. A multistage SIC that implements Mean-Field Annealing can be constructed be reusing the Mean-Field multistage detector on figure (2) by increasing α in some of the stages.

From the above discussion two considerations has to be made:

- The choice of the value of α_1 that is below α_c.
- How fast α should be increased towards the desired α_D.

The choice made on the two above considerations determines the *Annealing Scheme*.

The critical α_c can, in the CDMA setting, be lower bounded by considering the fix-point equations (14). We see the argument of $\tanh(\cdot)$ in equation (14) is linear in $\mathbf{m}_\alpha$ the maximal slope in any direction of $\mathbf{m}_\alpha$ is determined by the maximal eigenvalue of $-\alpha(\mathbf{R} - \mathbf{I})$. If the slope is less than one, independently of $\mathbf{z}$, the equations (14) only poses one solution, since $\tanh(\cdot)$ is antisymmetric sigmoid with slope 1 in zero. We now let λ_{min} denote the minimum eigenvalue of $\mathbf{R}$, then the maximal slope is $\alpha(1-\lambda_{min})$, this slope is equal to one for $\alpha = \frac{1}{1-\lambda_{min}} \leq \alpha_c$ i.e.

lower bounds the critical α_c. Since $\mathbf{R}$ is the correlation matrix of the unit energy spreading codes, $\mathbf{R}$ is positive semi-definite. This implies that the worst case λ_{min} over all sets of spreading codes equals zero hence $\alpha_c \geq \frac{1}{1-\lambda_{min}} \geq 1$. Using $\alpha_1 = 1$ ensures that only one unique solution exists to the equations (14) for all spreading codes.

The bound on α_c can also explain why the multistage detector used in [6] with clipped soft decision (CSD) squashing/mapping function works so well. It simply operates in the region where only one solution exists, since if an α was introduced it equals one, and for the CSD this is enough, for the same reasons as for $\tanh(\cdot)$, to ensure the existence of one unique solution. In moderate loaded systems $\frac{K}{N} = \beta < 1$ the energy differences of the local minima to the global one are for typical $\mathbf{R}$ large, hence in a typical transmission the solution is not influenced that much by the local minima.

The second issue is more complex, and we haven't yet found an answer. According to the optimal detector theory, the desired α_D that on average minimises the BER is $\alpha_D = \frac{1}{\sigma^2}$. In case of no local minima at α_D for any set of spreading sequences we would just solve the equations (14) at the desired α_D. So we know the starting α_1 and the terminating α_D, but the graduation in between yet still has to be determined. One notion is the distribution of the minimum eigenvalue λ_{min} of $\mathbf{R}$ which is distributed between zero and one skewed to one side depending on the load of the number of users K to the number of chips N denoted $\beta = \frac{K}{N}$. In an actual system this distribution can serve as a guide to determining the grading of the α's. Here we assume the minimum eigenvalue of $(\mathbf{R})$ to be distributed uniformly between zero and one, that means $1 - \lambda_{min}$ also is distributed uniformly between zero and one. So the inverse α should be spaced equally from 1 down to $\alpha_D^{-1} = \sigma^2$ this gives

$$\alpha_i = \frac{1}{1 + (i-1)\frac{1-\sigma^2}{1-D}} \tag{16}$$

where D is the number of different α's.

VII. Monte Carlo Simulations

To have a reference for the performance we choose the same parameter settings as in [6]'s second simulation. The number of users is $K = 8$ and the spreading factor is $N = 16$ i.e. a moderate load of $\beta = \frac{1}{2}$. Before we go to the actual BER-simulations, we show an example where local minima exists in conjunction with the global minimum. We generate a set of spreading codes $\mathbf{S}$ so $\mathbf{R}$ has one singular value. We transmit a vector $\mathbf{b}$ that has an unique maximum of the energy (5). So exhaustive maximum likelihood search finds the right solution with no errors in the $K = 8$ bits. If we use the detector used in [6] i.e. a multistage detector with clipped soft decision mapping/squashing function and $\alpha = 1$ initialised in zero, we get one error out of eight in this particular case, i.e. the unique solution, see the previous section, is dominated by at least one more minimum than the global minimum of the energy. Using a multistage detector similar to the previous but with $\tanh(\cdot)$ as squashing function, i.e. solving the equations (14) with $\alpha = 1$ initialised in zero, also introduced one error, by the same reasons. For this particular transmission we now want to find all the solutions $\mathbf{m}_\alpha$ of the equations (14) as a function of α. We randomly initialised the starting $\mathbf{m}_\alpha$ from the uniform distribution on the values $\{x| -1 < x < 1\}^K$ in order to get different solutions. We iterated the equations at fixed α until convergence. The solutions overlap-distance to the transmitted bits

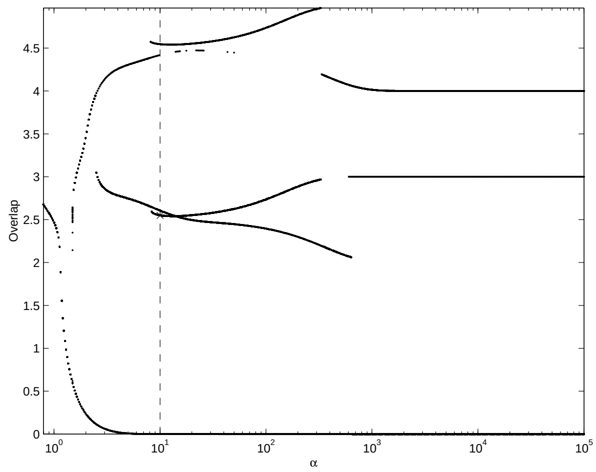

Fig. 3. Different solutions to the Mean-Field equations (14) at various α's for one particular transmission.

$\mathbf{b}$, defined here as $O(\alpha) = \frac{K - \mathbf{b}^T\mathbf{m}_\alpha}{2}$ is plotted on figure (3). This definition of the overlap-distance tells how far in number of soft bits the found solution $\mathbf{m}_\alpha$ is to the transmitted bits $\mathbf{b}$. If we start at low α we see that only one solution exists in accordance with the theory of annealing, when we increase α the overlap-distance becomes smaller and closer to one. At $\alpha = 1$ still only one solution exists in accordance with our conservative estimate of α_c. Between 1 and 2 a new solution starts to coexists, we have a bifurcation, which means that the 'most' critical α_c for this transmission lies between 1 and 2. This indicates that the bound on α_c is tight. The best of the two solutions continuous towards zero, whereas the bad one lies further and further away. At $\alpha \approx 3$ the bad solution bifurcates again. Around $\alpha \approx 8$ it again bifurcates. At $\alpha = \frac{1}{\sigma^2}$ we have drawn a vertical line indicating the desired α_D. We solved at the desired α_D the equations (14) starting in zero. The solution is marked with a cross, which certainly marks a wrong solution. At the desired α_D only 8 out of the 256 random initialisations converged to the good solution. We now did 256 random initialisations at $\alpha_1 = 1$ i.e. below α_c and chose $\alpha_D = \alpha_{10} = \frac{1}{\sigma^2}$ and used the annealing scheme described in equation (16), at each α we only used two iterations, so a total of 20 iterations/stages per random start. All the 256 initialisations converged to the true solution with no errors!, the same did the one initialised in zero.

We also did a BER-simulation of the system with $K = 8$ and $N = 16$ as in [6]. Here we used three different detectors. The first being the exhaustive search ML detector; the second the multistage detector used in [6] with clipped soft decision initialised in zero and a $\alpha = 1$ for 20 stages; the third a Mean-Field Annealing multistage detector with 10 α's, $\alpha_1 = 1$ and $\alpha_D = \alpha_{10} = \frac{1}{\sigma^2}$, at each α we used two iterations i.e. also 20 stages. The results can be seen on figure (4).

The inherent BER-floor is seen close to the prediction $5.33 \cdot 10^{-5}$ found in [6]. The three domains described in section (VI) can easily be seen. Above bit error rates of $4 \cdot 10^{-2}$ all the detectors follows closely the exhaustive search ML detector with a typical waterfall behaviour. From here the Clipped Soft Decision (CSD) detector bifurcates away from the exhaustive ML-detector indicating the influence of local minima in the energy. The Mean-Field Annealing detector avoids this influence from local minima

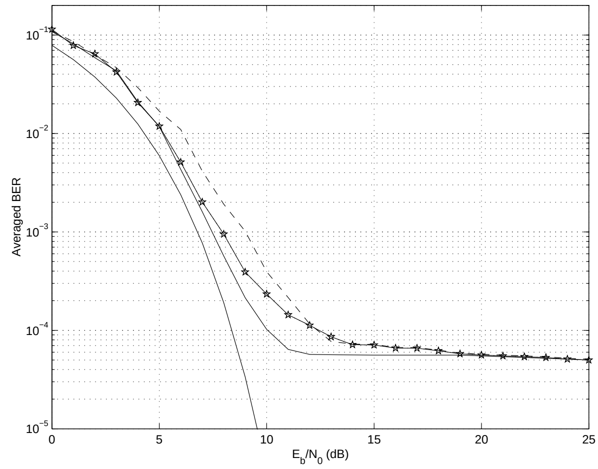

Fig. 4. BER performance of exhaustive search ML (solid line), Clipped Soft Decision multistage detector (dashed), and the proposed Mean-Field Annealing detector (solid with stars) all lower bounded by the single user performance (also solid). The parameters are $K = 8$ and $N = 16$.

down to $\mathrm{BER} = 6{\cdot}10^{-3}$ at 6 dB, this BER is reached by the CSD at 7 dB. The local minima influence on the Mean-Field Annealing detector is fairly small, and the detector is close to the exhaustive ML detector. But still there exists some excess errors compared to exhaustive ML search method, indicating that a better annealing scheme could be invented.

We also did simulations for a larger system $K = 20$ and $N = 24$, i.e. a fairly loaded system $\beta = \frac{K}{N} = \frac{1}{1.2}$, but because of the longer spreading sequences the probability of a singular $\mathbf{R}$ is lower than in the above example even though the load is higher. The results can be seen on figure (5). The two of the detectors

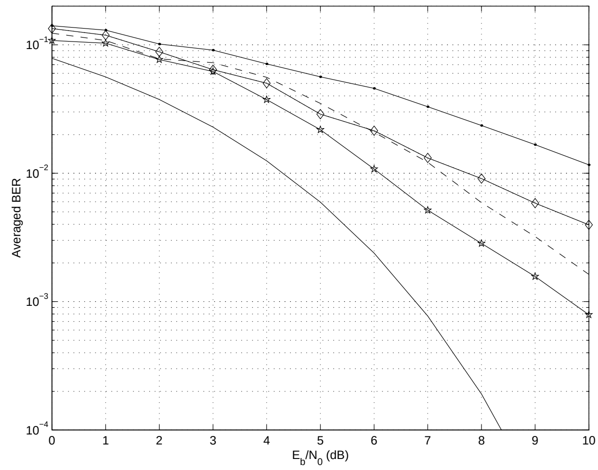

Fig. 5. BER performance of Clipped Soft Decision multistage detector (dashed), the proposed Mean-Field Annealing detector (solid with stars), Linear MMSE (solid with small dots), Multistage detector with first stage Linear MMSE (solid with diamonds), all lower bounded by the single user performance (also solid). The parameters are $K = 20$ and $N = 24$.

are identical to those in the previous simulation. The two others are linear MMSE, and a multistage detector with first stage linear MMSE followed by 20 stages of Mean-Field detection at $\alpha = \frac{1}{\sigma^2}$. The first thing that should be noted is the complexity of the two last detectors which are higher than the two first due to the matrix inversion. One interpretation of the multistage detector with first stage linear MMSE, is as annealing with two α's the first being very close to zero the second with $\alpha_2 = \alpha_D = \frac{1}{\sigma^2}$. The reason for this is that the linear MMSE detector can be derived by expanding the equations (14) to second order in α around $\alpha = 0$, i.e. is valid for very small α's, but in the domain where only one solution exists. Then we decrease α to the desired $\alpha_D = \frac{1}{\sigma^2}$ in one large step, which we know from the theory of annealing can go arbitrary wrong. The results on figure (5) follow those from the first example. Again the Mean-Field Annealing detector performs better than the CSD detector by around 1 dB. The two detectors based on the linear MMSE detector have an increasingly performance gap to the Mean-Field Annealing detector of more than 5 dB at a $\mathrm{BER} = 10^{-3}$. Again the CSD bifurcates away from the Mean-Field Annealing detector at 3 dB. Whereas it seems like a bifurcation of the Mean-Field Annealing detector at 7 dB but with a very small excess error level.

VIII. Conclusion

In this contribution we proposed a multistage multi-user detector implementing Mean-Field Annealing. It has no added complexity compared to existing multistage detectors. Performance gains of several dB's are obtained in simulations, closely approaching the exhaustive search ML detector. We also derived a conservative estimate on the first stage's control parameter, which however seems to be fairly tight. The proposed annealing scheme is general applicable, and can be used for any choice of spreading codes, since it is not optimised for any particular eigenvalue spectrum of the code correlation matrix. Though simulations show that further improvements can be achieved in order to reach the performance of the exhaustive search ML detector. We also, in the light of Mean-Field Annealing, gave some explanations to why the Clipped Soft Decision multistage detector performs so well, and why a multistage detector with an initial linear MMSE stage performs less well.

Acknowledgements

The first author would like to thank researchers at Nokia Mobile Phones RetD in Denmark for encouraging and supporting his work on CDMA detection. A special thank to Dr. Ole Nørklit for helpful discussions.

References

[1] Sergio Verdú, Shlomo Shamai (Shitz). Spectral Efficiency of CDMA with Random Spreading. *IEEE Transactions on Information Theory*, 45(2):622–640, March 1999.

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

[3] M. Varanasi, B Aazhang. Near-optimum demodulation for coherent communication in asynchronous Gaussian CDMA channels . In *Proc. 22nd Conf. on Information Science and Systems*, pages 832–839, March 1988.

[4] M. Varanasi, B Aazhang. Near-optimum detection in synchronous code-division multiple-access systems. *IEEE Transaction on Communication*, 39(5):725–735, May 1991.

[5] Peng H. Tan, Lars K. Rasmussen. Linear Interference Cancellation in CDMA Based on Iterative Techniques for Linear Equation Systems. *IEEE Transaction on Communication*, 48(12):2099–2108, December 2000.

[6] Lars K. Rasmussen, Teng J. Lim, Hiroki Sugimoto. Fundamental BER-Floor for Long-Code CDMA. In *Proceedings of IEEE 5th International Symposium on Spread Spectrum Techniques & Applications*, volume 3, pages 692–695, 1998.

[7] Giorgi Parisi. *Spin Glass Theory and Beyond*, volume 9. World Scientific, Singapore, 1987.

[8] P.J.M. van Laarhoven, E.H.L. Aarts. *Simulated Annealing: Theory and Applications*. D. Reidel Publishing Company, Dordrecht, Holland, 1987.

[9] Toshiyuki Tanaka. Analysis of Bit Error Probability of Direct-Sequence CDMA Multiuser Demodulators. In Todd .K. Leen, Thomas G. Dietterich, Volker Tresp, editor, *Advances in Neural Information Processing Systems (NIPS*2000)*, pages 315–321. MIT Press, 2000.

[10] Toshiyuki Tanaka. Statistical Mechanics of CDMA multiuser demodulation. *Europhysics Letters*, 54(4):540–546, May 2001.