Interference Reduction in CDMA Channels: A Statistical-Mechanics Approach

Toshiyuki Tanaka and Kouichiro Kitagawa
Graduate School of Informatics, Kyoto University
36-1 Yoshida Hon-machi, Sakyo-ku
Kyoto-shi, Kyoto 606-8501, Japan
Email: {tt@, kitagawa@sys.}i.kyoto-u.ac.jp

***Abstract*—Statistical mechanics-based approach to studying wireless communication systems is useful not only in analyzing theoretical performance of a given model, but also in optimizing design of a class of system models. In this paper, we demonstrate usefulness of statistical-mechanics approach in the problem of optimizing a spreading sequence set of randomly-spread code-division multiple-access (CDMA) communication systems. It has been shown that the optimal spreading sequences of a randomly spread CDMA channel in the large-system limit are Welch-bound-equality sequence sets, even when a non-Gaussian input distribution is assumed.**

I. Introduction

In recent years, information technologies are becoming available almost anywhere, via mobile devices with information processing and wireless communication functionalities. Such devices are required to provide very high data rates of wireless communication functionalities, since demands of users of such devices are rapidly growing toward multimedia-oriented services. Those devices are also expected to operate well in environments of rather bad conditions, such as urban areas with many buildings standing, passengers walking, and cars running, as well as many other users coexisting, and/or indoor spaces where there are many objects which scatter electromagnetic waves in a very complex way. One therefore has to consider communication systems which are of high dimensional and with certain kinds of randomness, representing those rather uncontrolled environmental conditions. Intuitively, this is the reason why one can expect statistical mechanics (which typically deals with extremely high-dimensional systems) of disordered systems (which bear some sorts of randomness) to provide an efficient means to analyze large-dimensional wireless communication systems.

Most existing statistical mechanics-based studies of wireless communication systems have discussed performance of given mathematical models of wireless communication systems with some randomness in the so-called large-system limit. The main objective of this paper is to demonstrate that the statistical mechanics-based approach can also be used to obtain insights which are useful in designing communication systems.

In this paper, we focus on a problem in the context of code-division multiple-access (CDMA) communication systems. CDMA is a technology to provide multiple-access functionalities: In typical mobile communication systems, an access from a mobile communication device is handled by a base station or an access point located near the mobile device. In real situations, it is often the case where there are more than one devices which connect to the same base station. Such situations are called multiple access. In a multiple-access environment, signals originated from different devices are summed up, and the base station has to extract relevant information sent from the users out of the mixture. CDMA provides one method of resolving such difficulties in multiple-access environments.

The basic idea of CDMA is to make use of a very high-dimensional representation of signals, in order to avoid interference between signals from different users. Each user's signal only occupies a low-dimensional (usually one-dimensional) subspace of the high-dimensional representation. The subspace is specified by a so-called spreading sequence assigned to that user. If spreading sequences of different users are mutually orthogonal, then they are free from interference. In commercial cellular phone systems, pseudorandom spreading sequences are commonly used in order to make any pair of them statistically orthogonal without prior negotiations. One can alternatively consider a scenario in which optimization of the set of spreading sequences takes place in order to reduce effects of interference. In this paper, we consider the latter scenario, and discuss how to choose a set of spreading sequences in order to optimize information-theoretic capacity of such communication channels.

This paper is organized as follows: Section II shows the formulation of the problem to be dealt with in this paper. In Section III, we show our main results. Detailed derivations of these results are given in Appendices. Section IV concludes the paper. A brief presentation of the results discussed in this paper has been given in [1].

II. Problem

We consider a linear vector channel with additive white Gaussian noise (AWGN) as a mathematical model of CDMA channels. The channel is defined as

$$\boldsymbol{Y} = L^{-1/2}\mathbf{S}\boldsymbol{X} + \boldsymbol{N}, \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^K$ and $\boldsymbol{Y} \in \mathbb{R}^L$ are random vectors representing channel input and output vectors, respectively, with K and L being input and output dimensions, and where channel noise vector $\boldsymbol{N} \in \mathbb{R}^L$ is assumed to follow a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma_0^2\mathbf{I})$, with $\mathbf{I}$ denoting an identity matrix.

WIOPT 2008, 1st – 3rd Apr 2008, Berlin, Germany.
 ISBN 978-963-9799-18-9
DOI 10.4108/ICST.WIOPT2008.3211

The matrix $\mathbf{S}$ consists of the spreading sequences used in the system: kth column $\boldsymbol{s}_k \in \mathbb{R}^L$ of $\mathbf{S}$ represents the spreading sequence used by the user whose information symbol is X_k. The factor $L^{-1/2}$ of the first term in the right-hand side of (1) is for normalization.

The channel input-output characteristic is also represented in terms of the conditional probability density

$$p_{\boldsymbol{Y}|\boldsymbol{X},\mathbf{S}}(\boldsymbol{y}|\boldsymbol{x},\mathbf{S}) = \prod_{\mu=1}^{L} p_{Y|\boldsymbol{X},\boldsymbol{S}}(y_\mu|\boldsymbol{x},\boldsymbol{s}^\mu)$$

$$p_{Y|\boldsymbol{X},\boldsymbol{S}}(y|\boldsymbol{x},\boldsymbol{s}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y - \frac{1}{\sqrt{L}}\boldsymbol{s}^T\boldsymbol{x}\right)^2\right], \quad (2)$$

where $\boldsymbol{s}^\mu$ denotes the vector that corresponds to μth row of the matrix $\mathbf{S}$. We assume that the elements of channel input $\boldsymbol{X}$ are independent and identically distributed (i.i.d.) following a prior distribution $p_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{k=1}^{K} p_X(x_k)$. The probability density function of $\boldsymbol{Y}$ given $\mathbf{S}$ is thus

$$p_{\boldsymbol{Y}|\mathbf{S}}(\boldsymbol{y}|\mathbf{S}) = \int p_{\boldsymbol{Y}|\boldsymbol{X},\mathbf{S}}(\boldsymbol{y}|\boldsymbol{x},\mathbf{S})\, p_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}. \quad (3)$$

In the above formula, and in what follows as well, integrals should be replaced with appropriate sums if the relevant random variables take discrete values.

We consider situations where one can choose the matrix $\mathbf{S}$, and wish to discuss how to choose $\mathbf{S}$ in order to maximize theoretical information transmission capability of the channel. The problem of optimizing $\mathbf{S}$ is straightforward when $K \le L$, or equivalently, the system load $\beta = K/L$ is not greater than 1, where one can arbitrarily take K orthogonal vectors in $\mathbb{R}^L$ and form $\mathbf{S}$ by putting those vectors column by column. The problem becomes non-trivial when $\beta > 1$, however, in which case one can no longer take K orthogonal vectors.

In order for analytical tractability we regard the matrix $\mathbf{S}$ as a random quantity and consider random ensembles of $\mathbf{S}$. We assume that each column vector of $L^{-1/2}\mathbf{S}$ is of unit norm in expectation, or equivalently, that

$$\mathbb{E}_{\mathbf{S}}(|\boldsymbol{S}_k|^2) = L \quad (4)$$

holds, where $\boldsymbol{S}_k$ denotes a random vector corresponding to kth column of $\mathbf{S}$. This assumption normalizes powers of the signals. When we consider the non-trivial case $\beta > 1$, and when the input vector $\boldsymbol{X}$ follows Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, then it has been known [2] that the so-called Welch Bound Equality (WBE) sequence sets are optimal. It has also been known [3] that WBE sequence sets exist for any $\beta > 1$. The WBE sequence sets are characterized by [4]

$$\mathbf{S}\mathbf{S}^T = \beta\mathbf{I}. \quad (5)$$

When the input vector $\boldsymbol{X}$ follows a distribution other than Gaussian, on the other hand, as far as the authors' knowledge there does not seem to be any results about optimal sequence sets. In this paper we discuss the problem of finding optimal spreading sequence sets for non-Gaussian input distributions.

In view of intrinsic difficulty of the problem with non-Gaussian input distributions, we consider the problem only in the large-system limit, in which we take a limit $K, L \to \infty$ while their ratio $\beta = K/L$ kept finite. In this paper, we take, as a measure of information transmission capability of the channel, capacity of the channel per input dimension in the large-system limit

$$C = \lim_{K\to\infty} \frac{1}{K} I(\boldsymbol{X};\boldsymbol{Y}|\mathbf{S}), \quad (6)$$

where $I(\boldsymbol{X};\boldsymbol{Y}|\mathbf{S})$ is a conditional mutual information of channel input $\boldsymbol{X}$ and output $\boldsymbol{Y}$ given $\mathbf{S}$.

III. Analysis

A. Evaluation of capacity

By decomposing the capacity C as

$$C = \mathcal{F} - \frac{1}{2\beta}(\log 2\pi\sigma^2 + 1), \quad (7)$$

where

$$\mathcal{F} = -\lim_{K\to\infty} \frac{1}{K} \mathbb{E}_{\boldsymbol{Y},\mathbf{S}}[\log p_{\boldsymbol{Y}|\mathbf{S}}(\boldsymbol{Y}|\mathbf{S})], \quad (8)$$

and where $\mathbb{E}_{\boldsymbol{Y},\mathbf{S}}(\cdots)$ denotes expectation with respect to $\boldsymbol{Y}$ and $\mathbf{S}$, the problem of evaluating the capacity C is reduced to evaluating the *free energy* $\mathcal{F}$.

We invoke replica method of statistical mechanics in order to evaluate the free energy $\mathcal{F}$. The first step is to rewrite (8) by introducing an auxiliary real-valued variable n as

$$\mathcal{F} = -\lim_{K\to\infty} \frac{1}{K} \lim_{n\to 0} \frac{\partial}{\partial n} \log \mathbb{E}_{\boldsymbol{Y},\mathbf{S}}[(p_{\boldsymbol{Y}|\mathbf{S}}(\boldsymbol{Y}|\mathbf{S}))^n]. \quad (9)$$

The next step is to interchange order of the operation regarding K and that regarding n, yielding

$$\mathcal{F} = -\lim_{n\to 0} \frac{\partial}{\partial n} \lim_{K\to\infty} \frac{1}{K} \log \mathbb{E}_{\boldsymbol{Y},\mathbf{S}}[(p_{\boldsymbol{Y}|\mathbf{S}}(\boldsymbol{Y}|\mathbf{S}))^n]. \quad (10)$$

Following the standard prescription of replica method, we first regard n to be an integer, on the basis of which we make use of the saddle-point method in evaluating the limit $K \to \infty$. We then regard that n is a real-valued variable, and take the derivative and the limit $n \to 0$ in order to obtain the final result.

We need some preliminary definitions to state our claim. Let $\rho(\lambda)$ be a density with a compact support. The Hilbert transform of $\rho(\lambda)$ is defined by

$$C_\rho(\gamma) = \int_{-\infty}^{\infty} \frac{\rho(\lambda)}{\gamma - \lambda}\, d\lambda \quad (11)$$

for γ outside the support of $\rho(\lambda)$. The R-transform $\mathcal{R}_\rho(z)$ of $\rho(\lambda)$ is defined in terms of the Hilbert transform, as

$$C_\rho\left(\mathcal{R}_\rho(z) + \frac{1}{z}\right) = z. \quad (12)$$

Let

$$p_U(u;\xi) = \int \sqrt{\frac{\xi}{2\pi}} e^{-\xi(u-x)^2/2}\, p_X(x)\, dx \quad (13)$$

be a probability density function of output U of a scalar AWGN channel with noise variance $1/\xi$, when the input X follows the distribution p_X.

The first main claim of this paper is the following: Let $\rho(\lambda)$ be the limiting eigenvalue distribution of the correlation matrix $\mathbf{R} = L^{-1}\mathbf{S}^T\mathbf{S}$ of the spreading sequences, in the large-system limit. Let us assume that the random matrix $\mathbf{R}$ is orthogonally invariant, i.e., the probability law of $\mathbf{R}$ and that of $\mathbf{V}^T\mathbf{R}\mathbf{V}$ with an arbitrary orthogonal matrix $\mathbf{V}$ are the same, and that $\rho(\lambda)$ has a compact support. The channel capacity C is given, using parameters $\{\theta, \mathcal{E}\}$ to be determined later, as

$$C = -\frac{1}{2}\theta\mathcal{E} - \frac{1}{2}G_\rho\left(-\frac{\mathcal{E}}{\sigma^2}\right) - \frac{1}{2}\log\frac{2\pi}{\theta} - \frac{1}{2} - \int p_U(u;\theta)\log p_U(u;\theta)\,du, \tag{14}$$

where $p_U(u;\theta)$ is defined in (13), and where the function G_ρ is defined as the following integral of the R-transform of $\rho(\lambda)$:

$$G_\rho(x) = \int_0^x \mathcal{R}_\rho(z)\,dz. \tag{15}$$

The parameters $\{\theta, \mathcal{E}\}$ are to be determined by the following saddle-point equations:

$$\mathcal{E} = \mathbb{E}_U[\langle(X - \langle X\rangle)^2\rangle;\theta], \tag{16}$$

$$\theta = \frac{1}{\sigma^2}\mathcal{R}_\rho\left(-\frac{\mathcal{E}}{\sigma^2}\right), \tag{17}$$

where $\langle\cdots\rangle$ denotes posterior average, given channel output U, of the scalar AWGN channel defined above.

The case of equal-power binary phase-shift-keying (BPSK) input modulation has been discussed by Takeda et al. [5], so that the above result is an extension of their result, in that our analysis assume an arbitrary input distribution p_X. Takeda et al.'s result is recovered from our result by letting $p_X(x) = (1/2)[\delta(x-1) + \delta(x+1)]$.

B. Special cases

In the conventional random spreading, one considers i.i.d. elements of the matrix $\mathbf{S}$. Let us assume that the distribution of the elements of $\mathbf{S}$ is zero-mean, unit-variance, and has finite higher-order moments. Then, it has been well known that the limiting eigenvalue distribution of the correlation matrix $\mathbf{R} = L^{-1}\mathbf{S}^T\mathbf{S}$ is given by the Marčenko-Pastur law [6]

$$\rho^{\mathrm{MP}}(\lambda) = \left(1 - \frac{1}{\beta}\right)^+ \delta(\lambda) + \frac{\sqrt{(\lambda - a)^+(b - \lambda)^+}}{2\pi\beta\lambda}, \tag{18}$$

where $(x)^+ = \max\{0, x\}$, and where $a = (1 - \sqrt{\beta})^2$ and $b = (1 + \sqrt{\beta})^2$. The R-transform of the Marčenko-Pastur law is

$$\mathcal{R}^{\mathrm{MP}}(z) = \frac{1}{1 - \beta z}, \tag{19}$$

on the basis of which one can recover the capacity of the vector channel with i.i.d. elements of $\mathbf{S}$, which was evaluated in the BPSK case in [7] and in the case of general data modulation in [8], from the result given in the previous subsection.

Our result is also applicable when one considers random spreading in which $\mathbf{S}$ consists of orthogonally invariant random WBE spreading sequences with $\beta > 1$. From the characterization (5) of WBE sequence sets, the eigenvalue distribution of the correlation matrix $\mathbf{R}$ is given by

$$\rho^{\mathrm{WBE}}(\lambda) = \left(1 - \frac{1}{\beta}\right)\delta(\lambda) + \frac{1}{\beta}\delta(\lambda - \beta). \tag{20}$$

The R-transform of $\rho^{\mathrm{WBE}}(\lambda)$ is calculated to be

$$\mathcal{R}^{\mathrm{WBE}}(z) = \frac{2}{1 - \beta z + \sqrt{(1 - \beta z)^2 + 4z}}, \tag{21}$$

which can be used to evaluate capacity numerically on the basis of the result given in the previous subsection.

C. Optimization

Since the capacity (14) is evaluated as a functional of the limiting eigenvalue distribution $\rho(\lambda)$ of the correlation matrix $\mathbf{R} = L^{-1}\mathbf{S}^T\mathbf{S}$, one can consider optimization problem of the capacity C with respect to $\rho(\lambda)$. In what follows we restrict our discussion to the non-trivial case where $\beta > 1$. There are several constraints imposed on the limiting eigenvalue distribution $\rho(\lambda)$. First, $\rho(\lambda)$ should be properly normalized, so that

$$\int \rho(\lambda)\,d\lambda = 1 \tag{22}$$

must hold. Second, one has to impose the normalization constraint corresponding to (4), which is rewritten in terms of $\rho(\lambda)$ as

$$\int \lambda\rho(\lambda)\,d\lambda = 1. \tag{23}$$

Third, since we consider the case $\beta > 1$, the correlation matrix $\mathbf{R}$ is rank deficient. Thus, $\mathbf{R}$ has trivial zero eigenvalue with multiplicity $(K - L)$, and accordingly, $\rho(\lambda)$ should be of the following form:

$$\rho(\lambda) = \left(1 - \frac{1}{\beta}\right)\delta(\lambda) + \frac{1}{\beta}\pi(\lambda), \tag{24}$$

where $\pi(\lambda)$ is a probability density function which, in view of (22) and (23), should satisfy

$$\int \pi(\lambda)\,d\lambda = 1, \tag{25}$$

and

$$\int \lambda\pi(\lambda)\,d\lambda = \beta, \tag{26}$$

respectively. Our optimization problem can now be stated as follows: Maximize C with respect to $\rho(\lambda)$, subject to the constraints (22)–(24) on $\rho(\lambda)$.

One can show that the solution of the constrained optimization problem is $\rho(\lambda) = \rho^{\mathrm{WBE}}(\lambda)$. This is the second main result of this paper, derivation of which is given in Appendix B.

IV. Conclusion

In this paper we have shown that if a WBE sequence set is used as spreading sequences of a CDMA system, then it is asymptotically optimum in the sense that it maximizes the capacity of the CDMA channel in the large-system limit, regardless of input distribution. Since WBE sequence sets have been known to be optimal when input distribution is Gaussian, our result can be regarded as an extension of the result to the cases with non-Gaussian input distribution. This extension has been possible by considering the large-system limit, as well as random spreading in which we have assumed orthogonal invariance of the correlation matrix $\mathbf{R}$ of the spreading sequences. We have followed statistical-mechanics approach, and in particular, replica method has been the key tool for the analysis. On the basis of the expression for the capacity obtained with the replica method, we have considered the problem of optimizing the limiting eigenvalue distribution of the correlation matrix $\mathbf{R}$.

One interesting problem would be to consider non power controlled systems, in which elements of $\boldsymbol{X}$ have different variances. It is important from communication theory point of view, since it is often the case in practice that power of signals from different users are unequal. In such cases, requiring the orthogonal invariance to the correlation matrix $\mathbf{R}$ should be inappropriate for maximizing capacity. Although spreading sequence sets that maximize capacity in the unequal-power Gaussian-input case has been known [13], whether or not the same sequence sets optimize capacity in non-Gaussian input cases is an open problem.

Appendix A
Derivation of capacity

In this appendix we show a derivation of the result (14), which gives the capacity of the vector channel model. Let

$$\Xi_n = \mathbb{E}_{\boldsymbol{Y},\mathbf{S}}[(p_{\boldsymbol{Y}|\mathbf{S}}(\boldsymbol{Y}|\mathbf{S}))^n], \tag{27}$$

so that the free energy $\mathcal{F}$ is given by

$$\mathcal{F} = -\lim_{n\to 0}\frac{\partial}{\partial n}\lim_{K\to\infty}\frac{1}{K}\log\Xi_n. \tag{28}$$

We rewrite Ξ_n as

$$\begin{aligned}\Xi_n &= \mathbb{E}_{\mathbf{S}}\left\{\int\cdots\int\prod_{a=0}^{n}[p_{\boldsymbol{Y}|\boldsymbol{X},\mathbf{S}}(\boldsymbol{y}|\boldsymbol{x}_a,\mathbf{S})\,p_{\boldsymbol{X}}(\boldsymbol{x}_a)\,d\boldsymbol{x}_a]\,d\boldsymbol{y}\right\}\\ &= \frac{1}{(2\pi\sigma^2)^{Ln/2}(n+1)^{L/2}}\\ &\quad\times\int\cdots\int\mathbb{E}_{\mathbf{R}}\left[\exp\left(\frac{K}{2}\operatorname{tr}\mathbf{R}\mathbf{X}\right)\right]\prod_{a=0}^{n}[p_{\boldsymbol{X}}(\boldsymbol{x}_a)\,d\boldsymbol{x}_a],\end{aligned} \tag{29}$$

where $\mathbf{R} = L^{-1}\mathbf{S}^T\mathbf{S}$ is the correlation matrix of $\mathbf{S}$, and where

$$\mathbf{X} = \frac{1}{K\sigma^2}\left[\frac{1}{n+1}\left(\sum_{a=0}^{n}\boldsymbol{x}_a\right)\left(\sum_{a=0}^{n}\boldsymbol{x}_a\right)^T - \sum_{a=0}^{n}\boldsymbol{x}_a\boldsymbol{x}_a^T\right]. \tag{30}$$

The random matrix $\mathbf{R}$ is orthogonally invariant, so that one can apply Itzykson-Zuber integral [9], [10], [11] to obtain

$$\mathbb{E}_{\mathbf{R}}\left[\exp\left(\frac{K}{2}\operatorname{tr}\mathbf{R}\mathbf{X}\right)\right] \approx \exp\left[\frac{K}{2}\operatorname{tr}G_\rho(\mathbf{X})\right], \tag{31}$$

when K is sufficiently large, where the function G_ρ is defined in (15).

The right-hand side of (31) is dependent on $\mathbf{X}$ only through its eigenvalues. Let $\mathcal{X}$ be a representation of $\mathbf{X}$ with the (generally non-orthogonal) basis $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. Since one has

$$\mathcal{X}_{ab} = \frac{1}{\sigma^2}\left(\frac{1}{n+1}\sum_{c=0}^{n}q_{ac} - q_{ab}\right), \tag{32}$$

with $q_{ab} = K^{-1}\sum_{k=1}^{K}x_{ak}x_{bk}$, the eigenvalues of $\mathbf{X}$ are given as functions of $\mathbf{Q} = (q_{ab})$. On the basis of this observation, one can decompose the integral with respect to $\{\boldsymbol{x}_a\}$ in (29) as an integral with a fixed $\mathbf{Q}$ and another integral with respect to all possible $\mathbf{Q}$, which yields

$$\Xi_n = \frac{1}{(2\pi\sigma^2)^{Ln/2}(n+1)^{L/2}}\int\cdots\int e^{K\operatorname{tr}G(\mathbf{X})/2}\mu(\mathbf{Q})\,d\mathbf{Q}, \tag{33}$$

where

$$\mu(\mathbf{Q}) = \int\cdots\int\delta\left(\mathbf{Q} - \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{x}_k\boldsymbol{x}_k^T\right)\prod_{a=0}^{n}[p_{\boldsymbol{X}}(\boldsymbol{x}_a)\,d\boldsymbol{x}_a], \tag{34}$$

with the symbol $\boldsymbol{x}_k$ denoting $(x_{0k}, x_{1k}, \ldots, x_{nk})^T$, is a probability weight, induced from $\{p_{\boldsymbol{X}}(\boldsymbol{x}_a)\}$, of the "subshell"

$$S(\mathbf{Q}) = \left\{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K)\;\middle|\;\mathbf{Q} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{x}_k\boldsymbol{x}_k^T\right\}. \tag{35}$$

Since $\mathbf{Q}$ is defined as an empirical mean of i.i.d. random matrices, $\mathbf{Q}$ follows the large-deviation principle [12] with a rate function $\mathcal{I}(\mathbf{Q})$, from which one obtains the following heuristic formula

$$\mu(\mathbf{Q}) \approx e^{-K\mathcal{I}(\mathbf{Q})} \tag{36}$$

for large K. Applying Varadhan's theorem, or the saddle-point method, one obtains

$$\lim_{K\to\infty}\frac{1}{K}\log\Xi_n = \sup_{\mathbf{Q}}[\operatorname{tr}G(\mathbf{X}) - \mathcal{I}(\mathbf{Q})] - \frac{1}{2\beta}\log(n+1) - \frac{n}{2\beta}\log 2\pi\sigma^2. \tag{37}$$

The rate function $\mathcal{I}(\mathbf{Q})$ is given by a Legendre transform of cumulant generating function of $\boldsymbol{X}_k$, as

$$\mathcal{I}(\mathbf{Q}) = \sup_{\tilde{\mathbf{Q}}}[\operatorname{tr}\mathbf{Q}\tilde{\mathbf{Q}} - \log M(\tilde{\mathbf{Q}})], \tag{38}$$

where

$$\begin{aligned}M(\tilde{\mathbf{Q}}) &= \mathbb{E}_{\boldsymbol{X}}\left[\exp\left(\operatorname{tr}\tilde{\mathbf{Q}}\boldsymbol{X}\boldsymbol{X}^T\right)\right]\\ &= \int\exp\left(\operatorname{tr}\tilde{\mathbf{Q}}\boldsymbol{x}\boldsymbol{x}^T\right)\prod_{a=0}^{n}[p_X(x_a)\,dx_a],\end{aligned} \tag{39}$$

with $\boldsymbol{X} \sim \prod_{a=0}^{n}p_X(x_a)$, is a moment generating function. The supremum of (38) is characterized by

$$\mathbf{Q} = \frac{\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\boldsymbol{X}^T\exp\left(\operatorname{tr}\tilde{\mathbf{Q}}\boldsymbol{X}\boldsymbol{X}^T\right)\right]}{\mathbb{E}_{\boldsymbol{X}}\left[\exp\left(\operatorname{tr}\tilde{\mathbf{Q}}\boldsymbol{X}\boldsymbol{X}^T\right)\right]}. \tag{40}$$

To proceed further we assume replica symmetry (RS), under which elements of the matrices $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ are invariant under exchange of indices. More precisely, we assume

$$q_{aa} = p, \quad q_{ab} = q, \quad \tilde{q}_{aa} = \frac{\tilde{p}}{2}, \quad \tilde{q}_{ab} = \frac{\tilde{q}}{2} \tag{41}$$

for $a, b = 0, 1, \ldots, n$, $a \neq b$. We have

$$\begin{aligned}M(\tilde{\mathbf{Q}}) &= \mathbb{E}_{\boldsymbol{X}}\left\{\exp\left[\frac{\tilde{p}-\tilde{q}}{2}\sum_{a=0}^{n}X_a^2 + \frac{\tilde{q}}{2}\left(\sum_{a=0}^{n}X_a\right)^2\right]\right\}\\ &= \sqrt{\frac{\tilde{q}}{2\pi}}\\ &\quad\times\mathbb{E}_{\boldsymbol{X}}\left\{\int\exp\left[-\frac{\tilde{q}}{2}u^2 + \tilde{q}u\sum_{a=0}^{n}X_a + \frac{\tilde{p}-\tilde{q}}{2}\sum_{a=0}^{n}X_a^2\right]\right\}du\\ &= \left(\frac{2\pi}{\tilde{q}}\right)^{n/2}\int e^{n\tilde{q}u^2/2}[p_U(u;\,\tilde{p},\,\tilde{q})]^{n+1}\,du,\end{aligned} \tag{42}$$

where we let

$$p_U(u;\,\tilde{p},\,\tilde{q}) = \int \sqrt{\frac{\tilde{q}}{2\pi}} e^{-\tilde{q}(u-x)^2/2} e^{\tilde{p}x^2/2}\, p_X(x)\, dx, \tag{43}$$

and therefore

$$\begin{aligned}\mathcal{I}(\mathbf{Q}) = \sup_{\tilde{p},\tilde{q}} \Big\{ & \frac{n+1}{2} p\tilde{p} + \frac{n(n+1)}{2} q\tilde{q} - \frac{n}{2}\log\frac{2\pi}{\tilde{q}} \\ & - \log \int e^{n\tilde{q}u^2/2} [p_U(u;\,\tilde{p},\,\tilde{q})]^{n+1}\, du \Big\}. \end{aligned} \tag{44}$$

The conditions (40) are rewritten under the RS assumption as

$$p = \frac{\int e^{n\tilde{q}u^2/2}[p_U(u;\,\tilde{p},\,\tilde{q})]^n\, \mathbb{E}_X\left[X^2 \sqrt{\frac{\tilde{q}}{2\pi}} e^{-\tilde{q}(u-X)^2/2} e^{\tilde{p}X^2/2}\right] du}{\int e^{n\tilde{q}u^2/2}[p_U(u;\,\tilde{p},\,\tilde{q})]^{n+1}\, du} \tag{45}$$

$$q = \frac{\int e^{n\tilde{q}u^2/2}[p_U(u;\,\tilde{p},\,\tilde{q})]^{n-1} \left\{\mathbb{E}_X\left[X \sqrt{\frac{\tilde{q}}{2\pi}} e^{-\tilde{q}(u-X)^2/2} e^{\tilde{p}x^2/2}\right]\right\}^2 du}{\int e^{n\tilde{q}u^2/2}[p_U(u;\,\tilde{p},\,\tilde{q})]^{n+1}\, du} \tag{46}$$

Under the RS assumption, the only non-zero eigenvalue of the matrix $\mathbf{X}$ turns out to be $-(p-q)/\sigma^2$ with multiplicity n, and the remaining eigenvalue is zero. One thus obtains

$$\operatorname{tr} G_\rho(\mathbf{X}) = nG_\rho\left(-\frac{p-q}{\sigma^2}\right). \tag{47}$$

Collecting these results, we obtain

$$\begin{aligned}\lim_{K\to\infty}\frac{1}{K}\log\Xi_n = \sup_{p,q}\inf_{\tilde{p},\tilde{q}} \Big\{ & \frac{n}{2}G_\rho\left(-\frac{p-q}{\sigma^2}\right) \\ & - \frac{n+1}{2}p\tilde{p} - \frac{n(n+1)}{2}q\tilde{q} + \frac{n}{2}\log\frac{2\pi}{\tilde{q}} \\ & + \log\int e^{n\tilde{q}u^2/2}[p_U(u;\,\tilde{p},\,\tilde{q})]^{n+1}\, du \Big\} \\ & - \frac{n}{2\beta}\log 2\pi\sigma^2 - \frac{1}{2\beta}\log(n+1). \end{aligned} \tag{48}$$

In view of the limit $n \to 0$ taken at the end of the analysis, we have only to consider the conditions (45) and (46), as well as the conditions of the supremum with respect to p and q, in the limit $n \to 0$. These conditions in the limit $n \to 0$ are given by

$$\tilde{p} = 0 \tag{49}$$

$$\tilde{q} = \frac{1}{\sigma^2}\mathcal{R}_\rho\left(-\frac{p-q}{\sigma^2}\right) \tag{50}$$

$$p = \mathbb{E}_U(\langle X^2\rangle) \tag{51}$$

$$q = \mathbb{E}_U(\langle X\rangle^2) \tag{52}$$

where

$$\langle\cdots\rangle = \frac{\mathbb{E}_{X|U}\left[(\cdots)\sqrt{\frac{\tilde{q}}{2\pi}} e^{-\tilde{q}(U-X)^2/2}\right]}{p_U(U;\,\tilde{q})} \tag{53}$$

is the posterior mean given channel output U of the scalar AWGN channel introduced in (13), and where $\mathbb{E}_U(\cdots)$ here denotes averaging with respect to the channel output $U \sim p_U(u;\,\tilde{q})$.

Taking the derivative with respect to n and the limit $n \to 0$, the free energy $\mathcal{F}$ thus becomes

$$\begin{aligned}\mathcal{F} = & -\frac{1}{2}G_\rho\left(-\frac{p-q}{\sigma^2}\right) - \frac{1}{2}(p-q)\tilde{q} - \frac{1}{2}\log\frac{2\pi}{\tilde{q}} \\ & - \frac{1}{2} - \int p_U(u;\,\tilde{q})\log p_U(u;\,\tilde{q})\, du \\ & + \frac{1}{2\beta}(\log 2\pi\sigma^2 + 1), \end{aligned} \tag{54}$$

with values of the parameters $\{p,\,q,\,\tilde{q}\}$ determined by the conditions (50)–(52). Letting $\mathcal{E} = p - q$ and $\theta = \tilde{q}$, we arrive at the final result (14).

Appendix B
Optimization of capacity

In this appendix, we explain how one can find the limiting eigenvalue distribution $\rho(\lambda)$ which maximizes the capacity C. As a first stage of the analysis, we study effects of perturbations of $\rho(\lambda)$ on the capacity C. Perturbations of $\rho(\lambda)$ affects the capacity C via changes in the function $G_\rho(x)$ in a direct manner, as well as via changes of the stationary values of $\{\theta,\,\mathcal{E}\}$ indirectly. The latter indirect effects can, however, safely be ignored within the first-order perturbation theory, because of the stationary conditions which determine their values. One can therefore focus on the direct effects, so that the capacity C is maximized if $-G_\rho(-\mathcal{E}/\sigma^2)$ is maximized with respect to $\rho(\lambda)$.

Next key observations are that $\mathcal{E}/\sigma^2 \ge 0$ and that

$$-G\left(-\frac{\mathcal{E}}{\sigma^2}\right) = -\int_0^{-\mathcal{E}/\sigma^2}\mathcal{R}(z)\, dz = \int_{-\mathcal{E}/\sigma^2}^0 \mathcal{R}(z)\, dz. \tag{55}$$

On the basis of these observation, one can make the following statement: If there is a distribution $\rho^*(\lambda)$ whose R-transform $\mathcal{R}^*(z)$ satisfies

$$\mathcal{R}^*(z) \ge \mathcal{R}_\rho(z), \quad \text{for all } z < 0, \tag{56}$$

for any distributions $\rho(\lambda)$ which satisfy the constraints (22)–(24), then the distribution $\rho^*(\lambda)$ maximizes $-G(-\mathcal{E}/\sigma^2)$. Our optimization problem is solved if one can find a distribution $\rho^*(\lambda)$ which maximizes the R-transform on the interval $(-\infty,\,0)$. It should be noted that it is a sufficient condition for our optimization problem to be solved with the solution $\rho^*(\lambda)$, and that it is not obvious at this stage whether or not there exists such a distribution $\rho^*(\lambda)$ which maximizes $\mathcal{R}(z)$ for all $z < 0$. In the following, we will show that there is a distribution with the required properties, and that $\rho^{\mathrm{WBE}}(\lambda)$ is the solution.

We proceed by converting the sufficient condition in terms of R-transform into that in terms of Hilbert transform. Since we are concerned with the R-transform $\mathcal{R}(z)$ for $z < 0$, we can restrict the domain of the Hilbert transform $C(\gamma)$ to $\gamma < \lambda_{\min}$, where $\lambda_{\min}$ denotes the minimum eigenvalue. For $\gamma < \lambda_{\min}$, the integrand $\rho(\lambda)/(\gamma-\lambda)$ of the Hilbert transform (11) for $\lambda \ge \lambda_{\min}$ is convex upward as a function of γ. From the convexity and the definition of the R-transform (12), one can state the following: If one can find a distribution $\rho^*(\lambda)$ which maximizes the Hilbert transform $C(\gamma)$ for all $\gamma < \lambda_{\min}$, then $\rho^*(\lambda)$ maximizes the R-transform $\mathcal{R}(z)$ for all $z < 0$.

To summarize the discussion so far, we have shown a sufficient condition under which our optimization problem is solved, that if there exists $\rho^*(\lambda)$ which maximizes the Hilbert transform $C(\gamma)$ for all $\gamma < \lambda_{\min}$, subject to the constraints (22)–(24), then the distribution $\rho^*(\lambda)$ maximizes the capacity C. In view of the constraint (24), what we have to solve is the maximization problem of the Hilbert transform

$$C_\pi(\gamma) = \int \frac{\pi(\lambda)}{\gamma - \lambda}\, d\lambda, \quad \text{for all } \gamma < \lambda_{\min}, \tag{57}$$

with respect to $\pi(\lambda)$, subject to the constraints (25) and (26) on $\pi(\lambda)$.

In the following we show that the solution is given by $\pi(\lambda) = \delta(\lambda - \beta)$. In order to do that, let us define

$$f_\gamma(\lambda) = \frac{1}{\gamma - \lambda} \tag{58}$$

and

$$g_\gamma(\lambda) = \frac{\lambda - \beta}{(\gamma - \beta)^2} + \frac{1}{\gamma - \beta}. \tag{59}$$

It should be noted that the function $g_\gamma(\lambda)$ is a linear function in λ and is tangential to $f_\gamma(\lambda)$ at $\lambda = \beta$, and that the function $f_\gamma(\lambda)$ is convex upward for $\lambda > \gamma$, so that $f_\gamma(\lambda) \le g_\gamma(\lambda)$ with equality at $\lambda = \beta$. We then consider the objective function

$$\begin{aligned} I_\pi(\gamma) &= \int [f_\gamma(\lambda) - g_\gamma(\lambda)]\pi(\lambda)\, d\lambda \\ &= C_\pi(\gamma) - \int g_\gamma(\lambda)\pi(\lambda)\, d\lambda. \end{aligned} \tag{60}$$

Note that the second term is constant under the constraints (25) and (26). Thus, $I_\pi(\gamma)$ and $C_\pi(\gamma)$ are simultaneously maximized by the same $\pi(\lambda)$ under these constraints. Let us consider maximizing $I_\pi(\gamma)$ with the constraint (25), but *without* (26). Since $f_\gamma(\lambda) \le g_\gamma(\lambda)$ with equality at $\lambda = \beta$, $I_\pi(\gamma)$ is maximized by concentrating all the probability mass at $\lambda = \beta$, namely, by letting

$$\pi(\lambda) = \delta(\lambda - \beta). \tag{61}$$

Since the solution (61) also satisfies the constraint (26), it is the solution of the maximization problem with both of the constraints (25) and (26) as well. Therefore, one can conclude that the solution (61) is maximizing the Hilbert transform $C(\gamma)$ for all $\gamma < \lambda_{\min}$. The resulting eigenvalue distribution is exactly the same as $\rho^{\mathrm{WBE}}(\lambda)$ with $\beta > 1$, which proves that $\rho^{\mathrm{WBE}}(\lambda)$ maximizes the capacity C.

Acknowledgements

Support from the Grant-in-Aid for Scientific Research on Priority Areas, the Ministry of Education, Culture, Sports, Science and Technology, Japan (No. 18079010) is acknowledged.

References

[1] K. Kitagawa and T. Tanaka, "Optimal spreading sequences in large CDMA systems: A statistical mechanics approach," submitted to *IEEE Int. Symp. Info. Theory*, 2008.

[2] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Info. Theory*, vol. 40, pp. 1261–1266, 1994.

[3] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Info. Theory*, vol. 45, pp. 1968–1983, 1999.

[4] J. L. Massey and T. Mittelholzer, "Welch's bound and sequence sets for code-division multiple-access systems," in R. Capocelli, A. D. Santis, and U. Vaccaro (eds.), *Sequences II, Methods in Communication, Security, and Computer Science*, Springer-Verlag, 1993.

[5] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *Europhys. Lett.*, vol. 76, pp. 1193–1199, 2006.

[6] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Matematicheskii Sbornik*, vol. 72, pp. 507–536, 1967.

[7] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Info. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[8] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Info. Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[9] C. Itzykson and J. B. Zuber, "The planar approximation, II," *J. Math. Phys.*, vol. 21, no. 3, pp. 411–421, Mar. 1980.

[10] A. Guionnet and M. Maïda, "A Fourier view on the *R*-transform and related asymptotics of spherical integrals," *J. Funct. Anal.*, vol. 222, no. 2, pp. 435–490, May 2005.

[11] B. Collins and P. Śniady, "New scaling of Itzykson-Zuber integrals," *Ann. de l'Inst. Henri Piincaré, Prob. Stat.*, vol. 43, no. 2, pp. 139–146, 2007.

[12] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, 1998.

[13] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Trans. Info. Theory*, vol. 45, no. 6, pp. 1984–1991, 1999.