

# Efficient Linear Precoding for Massive MIMO Systems using Truncated Polynomial Expansion

Axel Müller<sup>§\*</sup>, Abla Kammoun<sup>‡</sup>, Emil Björnson<sup>§†</sup>, and Mérouane Debbah<sup>§</sup>

<sup>§</sup>Alcatel-Lucent Chair on Flexible Radio, Supélec, France, <sup>‡</sup>KAUST, Thuwal, Makkah Province, Saudi Arabia

<sup>\*</sup>Intel Mobile Communications, Sophia Antipolis, France, <sup>†</sup>ACCESS, Dept. of Signal Processing, KTH, Sweden  
 {axel.mueller, emil.bjornson, merouane.debbah}@supelec.fr, abla.kammoun@gmail.com

**Abstract**—Massive multiple-input multiple-output (MIMO) techniques have been proposed as a solution to satisfy many requirements of next generation cellular systems. One downside of massive MIMO is the increased complexity of computing the precoding, especially since the relatively “antenna-efficient” regularized zero-forcing (RZF) is preferred to simple maximum ratio transmission. We develop in this paper a new class of precoders for single-cell massive MIMO systems. It is based on truncated polynomial expansion (TPE) and mimics the advantages of RZF, while offering reduced and scalable computational complexity that can be implemented in a convenient pipelined fashion. Using random matrix theory we provide a closed-form expression of the signal-to-interference-and-noise ratio under TPE precoding and compare it to previous works on RZF. Furthermore, the sum rate maximizing polynomial coefficients in TPE precoding are calculated. By simulation, we find that to maintain a fixed per-user rate loss as compared to RZF, the polynomial degree does not need to scale with the system, but it should be increased with the quality of the channel knowledge and signal-to-noise ratio.

## I. INTRODUCTION

Next generation cellular systems need to cope with the dramatically growing number of user terminals (UTs) and data traffic. A proposed solution to this challenge is to employ massive multiple-input multiple-output (MIMO) techniques [1]. Such systems are commonly defined as macro base stations (BSs) being equipped with many antennas, e.g., hundreds, while each BS only serves relative few UTs, e.g., tens, at the same time [2]. These systems improve spectral and energy efficiency by dense and adaptive spatial reuse, and are robust to channel uncertainty. However, massive MIMO suffers from potentially prohibitively high computational complexity pertaining to the precoding. The first works on massive MIMO have dismissed this issue, as the computationally simple maximum ratio transmission (MRT) precoding becomes optimal for an infinite number of BS antennas [3]. More recent works have shown that more practically manageable numbers of BS antennas, call for more involved linear precoding techniques, such as regularized zero-forcing (RZF) [4]. The computational complexity of RZF cannot be neglected in massive MIMO systems, as RZF involves the inversion of very large matrices [5]. In this paper, we propose a new class of truncated polynomial expansion (TPE) precoders that replaces the matrix inversion by a matrix polynomial with  $J$  terms. TPE precoding features

reduced, as well as, scalable computational complexity, while offering performance that is comparable to RZF.

The TPE technique enables the balancing of precoding complexity and system throughput. Using random matrix theory [6], we derive so-called large-scale approximations for the achievable user rates for any order  $J$  of TPE precoding. These are closed-form expressions that are tight when the number of BS antennas and the number of UTs grow large with a fixed ratio, but also provide close approximations at small dimensions. These expressions allow for optimization of the polynomial coefficients with respect to any metric; we derive the coefficients that maximize the sum throughput. We note that the combination of TPE and precoding is very new and the only related work is [7] by Zarei *et al.* Unlike this paper, the precoding in [7] is conceived to minimize the sum-MSE of all users. Although our approach originates from the same TPE technique as in [7], our work is more comprehensive in that we consider a channel model which takes into account the transmit correlation at the base station and we optimize the throughput. An extension of this paper for more involved single-cell systems can be found in [5], which also proves the results presented in this paper. The treatment of the multi-cell case under a general channel model is available in [8].

## II. SYSTEM MODEL

This section defines the single-cell downlink system with flat-fading channels, linear precoding, and channel estimation errors. The BS is equipped with  $M$  antennas and serves  $K$  single-antenna UTs. Denoting the set of complex numbers by  $\mathbb{C}$ , the received complex baseband signal  $y_k \in \mathbb{C}$  at the  $k$ th UT is

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k, \quad k = 1, \dots, K \quad (1)$$

where  $\mathbf{x} \in \mathbb{C}^{M \times 1}$  is the transmit signal and  $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$  represents the random channel vector between the BS and the  $k$ th UT. The additive circularly-symmetric complex Gaussian noise at the  $k$ th UT is denoted by  $n_k \sim \mathcal{CN}(0, \sigma^2)$  for  $k = 1, \dots, K$ , where  $\sigma^2$  is the receiver noise variance.

Each channel vector  $\mathbf{h}_k$ , for  $k = 1, \dots, K$  is modeled as

$$\mathbf{h}_k \sim \mathcal{CN} \left( \mathbf{0}_{M \times 1}, \frac{1}{K} \mathbf{\Phi} \right) \quad (2)$$

where the channel covariance matrix  $\mathbf{\Phi} \in \mathbb{C}^{M \times M}$  has bounded spectral norm, as  $M \rightarrow \infty$ ; thus its trace scales linearly with

This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering) and an International Postdoc Grant from The Swedish Research Council.

$M$ . Here,  $\mathbf{0}_{M \times 1}$  is a vector with  $M$  zeros. We consider a *Rayleigh block-fading* model where  $\mathbf{h}_k$  has a fixed realization for a coherence period and then takes a new independent realization. The covariance scaling  $\frac{1}{K}$  is only for technical reasons.

The BS employs Gaussian codebooks and linear precoding, where  $\mathbf{g}_k \in \mathbb{C}^{M \times 1}$  denotes the precoding vector and  $s_k \sim \mathcal{CN}(0, 1)$  is the data symbol of the  $k$ th UT. Based on this assumption, the transmit signal in (1) can be expressed as

$$\mathbf{x} = \sum_{n=1}^K \mathbf{g}_n s_n = \mathbf{G} \mathbf{s}. \quad (3)$$

The matrix notation is obtained by letting  $\mathbf{G} = [\mathbf{g}_1 \dots \mathbf{g}_K] \in \mathbb{C}^{M \times K}$  be the precoding matrix and  $\mathbf{s} = [s_1 \dots s_K]^T \sim \mathcal{CN}(\mathbf{0}_{K \times 1}, \mathbf{I}_K)$  be the vector containing all UT data symbols.

Consequently, the received signal (1) can be expressed as

$$y_k = \mathbf{h}_k^H \mathbf{g}_k s_k + \sum_{n=1, n \neq k}^K \mathbf{h}_k^H \mathbf{g}_n s_n + n_k. \quad (4)$$

Let  $\mathbf{G}_k$  be the matrix  $\mathbf{G}$  with column  $\mathbf{g}_k$  removed. Then the signal-to-interference-and-noise ratio (SINR) at the  $k$ th UT becomes

$$\text{SINR}_k = \frac{\mathbf{h}_k^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{h}_k}{\mathbf{h}_k^H \mathbf{G}_k \mathbf{G}_k^H \mathbf{h}_k + \sigma^2}. \quad (5)$$

By assuming that each UT has perfect CSI, the ergodic achievable data rates at the UTs are

$$r_k = \mathbb{E}[\log_2(1 + \text{SINR}_k)], \quad k = 1, \dots, K.$$

The transmitter is assumed to have imperfect knowledge of the instantaneous channel realization  $\hat{\mathbf{h}}_k$  of each UT, for  $k = 1, \dots, K$ . We adopt the Gauss-Markov formulation [9]:

$$\hat{\mathbf{h}}_k = \sqrt{1 - \tau^2} \mathbf{h}_k + \tau \mathbf{n}_k \quad (6)$$

where  $\mathbf{h}_k$  is the true channel and  $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}_{M \times 1}, \frac{1}{K} \Phi)$  models the independent error. The scalar parameter  $\tau \in [0, 1]$  indicates the quality of the instantaneous CSI;  $\tau = 0$  gives perfect instantaneous CSI and  $\tau = 1$  corresponds to having only statistical knowledge. The matrix  $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1 \dots \hat{\mathbf{h}}_K] \in \mathbb{C}^{M \times K}$  denotes the joint imperfect knowledge of all user channels.

### III. LINEAR PRECODING

Many heuristic linear precoding schemes have been proposed in the literature, mainly because finding the optimal precoding (e.g., in terms of sum rate) is very computationally demanding and thus unsuitable for fading systems. Among the heuristic schemes we chose *RZF precoding* [10], which can provide close-to-optimal performance in many scenarios.

#### A. Review on RZF Precoding in Massive MIMO Systems

Suppose we have a total transmit power constraint

$$\frac{1}{K} \text{tr}(\mathbf{G} \mathbf{G}^H) = P \quad (7)$$

where  $\text{tr}(\cdot)$  is the trace function and we note that the scaling factor  $\frac{1}{K}$  neutralizes the corresponding channel variance scaling in (2), which is possible via (8) and taking normalization

into account. We stress that the total power  $P$  is fixed, while we let the number of antennas  $M$  and UTs  $K$  grow large.

Similar to [9], we define the RZF precoding matrix as

$$\mathbf{G}_{\text{RZF}} = \beta \hat{\mathbf{H}} \left( \hat{\mathbf{H}}^H \hat{\mathbf{H}} + \xi \mathbf{I}_K \right)^{-1} = \beta \left( \hat{\mathbf{H}} \hat{\mathbf{H}}^H + \xi \mathbf{I}_M \right)^{-1} \hat{\mathbf{H}} \quad (8)$$

where the power normalization parameter  $\beta$  is set such that  $\mathbf{G}_{\text{RZF}}$  satisfies the power constraint in (7). The scalar regularization coefficient  $\xi$  can be selected in different ways, depending on  $P$ ,  $\sigma^2$ ,  $\tau$ , and system dimensions [9], [10].

In [9], the performance of each UT under RZF precoding is studied in the large- $(M, K)$  regime, where  $M$  and  $K$  tend to infinity at the same speed:

$$0 < \liminf \frac{K}{M} \leq \limsup \frac{K}{M} < +\infty. \quad (9)$$

The achievable user performance is characterized by  $\text{SINR}_k$  in (5). Although the SINR is a random quantity that depends on the instantaneous values of the random users channels in  $\mathbf{H}$  and the instantaneous estimate  $\hat{\mathbf{H}}$ , it can be approximated using deterministic quantities in the large- $(M, K)$  regime; see [9]. These are quantities that only depend on the statistics of the channels and are referred to as *deterministic equivalents*, since they are almost surely (a.s.) tight in the asymptotic limit. When the deterministic equivalents are applied at finite  $M$  and  $K$ , they are often referred to as *large-scale approximations*.

The RZF precoding matrix in (8) is a function of the instantaneous CSI at the transmitter. The computation of the precoding matrix includes large matrix multiplications and inversions, which gives a computational complexity of approximately  $2MK^2$  operations per coherence time [5]. This complexity scaling is intractable for large values of  $M$  and  $K$ .

#### B. Truncated Polynomial Expansion Precoding

Next, we derive a new class of low-complexity linear precoders based on TPE. The following result from [11] provides a motivation behind the use of polynomial expansions.

**Lemma 1.** For any positive definite Hermitian matrix  $\mathbf{X}$ ,

$$\mathbf{X}^{-1} = \alpha (\mathbf{I} - (\mathbf{I} - \alpha \mathbf{X}))^{-1} = \alpha \sum_{\ell=0}^{\infty} (\mathbf{I} - \alpha \mathbf{X})^{\ell} \quad (10)$$

where the second equality holds if the parameter  $\alpha$  is selected such that  $0 < \alpha < \frac{2}{\max_n \lambda_n(\mathbf{X})}$ .

This lemma shows that the inverse of any Hermitian matrix can be expressed as a matrix polynomial. More importantly, the low-order terms are the most influential ones, since the eigenvalues of  $(\mathbf{I} - \alpha \mathbf{X})^{\ell}$  go to zero as  $\ell$  grows large. Thus, it makes sense to consider a TPE using only the first  $J$  terms.

TPE has been successfully applied for low-complexity multi-user detection in [11]. Next, we apply this technique to approximate RZF precoding by a matrix polynomial. Starting from  $\mathbf{G}_{\text{RZF}}$  in (8), we note that

$$\begin{aligned} & \beta \left( \hat{\mathbf{H}} \hat{\mathbf{H}}^H + \xi \mathbf{I}_M \right)^{-1} \hat{\mathbf{H}} \\ & \approx \sum_{\ell=0}^{J-1} \left( \beta \alpha \sum_{n=\ell}^{J-1} \binom{n}{\ell} (1 - \alpha \xi)^{n-\ell} (-\alpha)^{\ell} \right) (\hat{\mathbf{H}} \hat{\mathbf{H}}^H)^{\ell} \hat{\mathbf{H}}. \end{aligned} \quad (11)$$

Inspecting (11), we have a precoding matrix with the structure

$$\mathbf{G}_{\text{TPE}} = \sum_{\ell=0}^{J-1} w_{\ell} \left( \widehat{\mathbf{H}} \widehat{\mathbf{H}}^{\text{H}} \right)^{\ell} \widehat{\mathbf{H}} \quad (12)$$

where  $w_0, \dots, w_{J-1}$  are scalar coefficients. Although the bracketed term in (11) provides a potential expression for  $w_{\ell}$ , these are not the optimal values for  $J < \infty$ . Also, these coefficients do not satisfy the power constraint in (7).

The precoding in (12) is coined *TPE precoding* and actually defines a whole class of precoding matrices for different  $J$ . For  $J = 1$ , we obtain  $\mathbf{G} = w_0 \widehat{\mathbf{H}}$ , which equals MRT. Furthermore, RZF precoding can be obtained by choosing  $J = \min(M, K) = K$  and coefficients based on the characteristic polynomial of  $(\widehat{\mathbf{H}} \widehat{\mathbf{H}}^{\text{H}} + \xi \mathbf{I}_M)^{-1}$  (this follows from the Cayley-Hamilton theorem). We refer to  $J$  as the *TPE order*.

Let the transmit signal with TPE precoding at channel use  $t$  be denoted  $\mathbf{x}^{(t)}$  and observe that it can be expressed as

$$\mathbf{x}^{(t)} = \mathbf{G}_{\text{TPE}} \mathbf{s}^{(t)} = \sum_{\ell=0}^{J-1} w_{\ell} \tilde{\mathbf{x}}_{\ell}^{(t)}$$

where  $\mathbf{s}^{(t)}$  is the vector of data symbols at channel use  $t$  and

$$\tilde{\mathbf{x}}_{\ell}^{(t)} = \begin{cases} \widehat{\mathbf{H}} \mathbf{s}^{(t)}, & \ell = 0, \\ \widehat{\mathbf{H}} (\widehat{\mathbf{H}}^{\text{H}} \tilde{\mathbf{x}}_{\ell-1}^{(t)}), & 1 \leq \ell \leq J-1. \end{cases}$$

This is an iterative way of computing the  $J$  terms in TPE precoding, which removes precoding precomputation delays and enables parallelization over multiple processing cores (e.g., application-specific integrated circuits) and pipelining [11]. The total computational complexity is approximately  $4JM K$  per channel use [5], but each term  $\tilde{\mathbf{x}}_{\ell}^{(t)}$  is obtained by a simple matrix-vector multiplication which is easily implemented in hardware. This means that the factor  $J$  in the complexity expression is not causing any extra delays since it can be handled by adding more processing cores that operate in parallel. Thus, we can select the TPE order  $J$  as large as needed to obtain a certain precoding accuracy and still enable real-time computation of the transmit signal  $\mathbf{x}$ .

#### IV. ANALYSIS AND OPTIMIZATION OF TPE PRECODING

Recall the SINR in (5) and observe that  $\mathbf{g}_k = \mathbf{G} \mathbf{e}_k$  and  $\mathbf{h}_k^{\text{H}} \mathbf{G}_k \mathbf{G}_k^{\text{H}} \mathbf{h}_k = \mathbf{h}_k^{\text{H}} \mathbf{G} \mathbf{G}^{\text{H}} \mathbf{h}_k - \mathbf{h}_k^{\text{H}} \mathbf{g}_k \mathbf{g}_k^{\text{H}} \mathbf{h}_k$ , where  $\mathbf{e}_k$  is the  $k$ th column of the identity matrix  $\mathbf{I}_K$ . By substituting the TPE precoding matrix (12) into (5), the SINR writes as

$$\text{SINR}_k = \frac{\mathbf{w}^{\text{H}} \mathbf{A}_k \mathbf{w}}{\mathbf{w}^{\text{H}} \mathbf{B}_k \mathbf{w} - \mathbf{w}^{\text{H}} \mathbf{A}_k \mathbf{w} + \sigma^2} \quad (13)$$

where  $\mathbf{w} = [w_0 \dots w_{J-1}]^{\text{T}}$  and the matrices  $\mathbf{A}_k, \mathbf{B}_k \in \mathbb{C}^{J \times J}$  have entries (numbered from 0 for convenience) given by

$$[\mathbf{A}_k]_{\ell, m} = \mathbf{h}_k^{\text{H}} \left( \widehat{\mathbf{H}} \widehat{\mathbf{H}}^{\text{H}} \right)^{\ell} \widehat{\mathbf{H}} \mathbf{e}_k \mathbf{e}_k^{\text{H}} \widehat{\mathbf{H}}^{\text{H}} \left( \widehat{\mathbf{H}} \widehat{\mathbf{H}}^{\text{H}} \right)^m \mathbf{h}_k \quad (14)$$

$$[\mathbf{B}_k]_{\ell, m} = \mathbf{h}_k^{\text{H}} \left( \widehat{\mathbf{H}} \widehat{\mathbf{H}}^{\text{H}} \right)^{\ell+m+1} \mathbf{h}_k \quad (15)$$

for  $\ell = 0, \dots, J-1$  and  $m = 0, \dots, J-1$ .

In the previously defined large- $(M, K)$  regime (see (9)), it can be seen that all SINRs in (13) converge to the same asymptotic equivalent. This is formalized in our main theorem:

**Theorem 2.** *In the large- $(M, K)$  regime,  $\text{SINR}_k$  converges almost surely to*

$$\text{SINR}_k \xrightarrow[M, K \rightarrow +\infty]{\text{a.s.}} \frac{\mathbf{w}^{\text{H}} \bar{\mathbf{A}} \mathbf{w}}{\mathbf{w}^{\text{H}} (\bar{\mathbf{B}} - \bar{\mathbf{A}}) \mathbf{w} + \sigma^2} \rightarrow 0$$

with

$$[\bar{\mathbf{A}}]_{\ell, m} = \frac{(-1)^{\ell+m}}{\ell! m!} \bar{z}_{\ell} \bar{z}_m \text{ and } [\bar{\mathbf{B}}]_{\ell, m} = \frac{(-1)^{\ell+m+1}}{(\ell+m+1)!} \bar{x}_{\ell+m+1}.$$

The formulations for  $\bar{x}_{\ell}$  and  $\bar{z}_{\ell}$  are given in Corollary 3 and for the proof we refer to [5].

**Corollary 3.** *Denote by  $\bar{x}_0$  and  $\bar{z}_0$  the deterministic quantities*

$$\bar{x}_0 \triangleq \frac{1}{K} \text{tr}(\Phi) \text{ and } \bar{z}_0 \triangleq \sqrt{1-\tau^2} \frac{1}{K} \text{tr}(\Phi).$$

*One can iteratively compute the deterministic sequences  $\bar{x}_{k, \ell}$  and  $\bar{z}_{k, \ell}$  until a fixed order  $p$  as*

$$\begin{aligned} \bar{x}_{\ell} &= \mu_{\ell} + \tau^2 \sum_{m=1}^{\ell} \binom{\ell}{m} m \mu_{m-1} \mu_{\ell-m} \\ &\quad - \sum_{m=0}^{\ell-1} \binom{\ell}{m} \bar{x}_m (\ell-m) \mu_{\ell-m-1}, \quad 1 \leq \ell \leq p \end{aligned}$$

$$\bar{z}_{\ell} = \sqrt{1-\tau^2} \mu_{\ell} - \sum_{m=0}^{\ell-1} \binom{\ell}{m} \bar{z}_m (\ell-m) \mu_{\ell-m-1}, \quad 1 \leq \ell \leq p.$$

where  $\mu_{\ell} = \frac{1}{K} \text{tr}(\Phi \mathbf{T}_{\ell})$ ,  $\mathbf{T}_{\ell} = \frac{d^{\ell} \mathbf{T}(t)}{dt^{\ell}}|_{t=0}$  and  $\mathbf{T}(t) = (\mathbf{I}_M + \frac{t\Phi}{1+t\delta(t)})^{-1}$ , with  $\delta(t) > 0$  being the unique solution for every  $t > 0$  of the fixed point equation:

$$\delta(t) = \frac{1}{K} \text{tr}(\Phi \mathbf{T}(t)).$$

The matrix  $\mathbf{T}_{\ell}$  can be computed iteratively (see [12]).

#### A. Optimization of the Asymptotic SINRs

Next, we optimize the asymptotic SINRs with respect to the polynomial coefficients  $\mathbf{w} = [w_0 \dots w_{J-1}]^{\text{T}}$  under the power constraints in (7). Since the asymptotic SINRs are all equal, this corresponds to the optimization problem

$$\begin{aligned} &\underset{\mathbf{w}}{\text{maximize}} && \frac{\mathbf{w}^{\text{H}} \bar{\mathbf{A}} \mathbf{w}}{\mathbf{w}^{\text{H}} (\bar{\mathbf{B}} - \bar{\mathbf{A}}) \mathbf{w} + \sigma^2} \\ &\text{subject to} && \mathbf{w}^{\text{H}} \bar{\mathbf{C}} \mathbf{w} = P \end{aligned} \quad (16)$$

where  $\bar{\mathbf{A}}$  and  $\bar{\mathbf{B}}$  are already given in Theorem 2 and

$$[\bar{\mathbf{C}}]_{\ell, m} = \frac{(-1)^{\ell+m+1}}{(\ell+m+1)!} \frac{1}{K} \text{tr}(\mathbf{T}_{\ell+m+1}).$$

**Theorem 4.** *Let  $\mathbf{a}$  be an eigenvector corresponding to the maximum eigenvalue  $\lambda_{\max}$  of*

$$\left( \bar{\mathbf{B}} - \bar{\mathbf{A}} + \frac{\sigma^2}{P} \bar{\mathbf{C}} \right)^{-\frac{1}{2}} \bar{\mathbf{A}} \left( \bar{\mathbf{B}} - \bar{\mathbf{A}} + \frac{\sigma^2}{P} \bar{\mathbf{C}} \right)^{-\frac{1}{2}}. \quad (17)$$

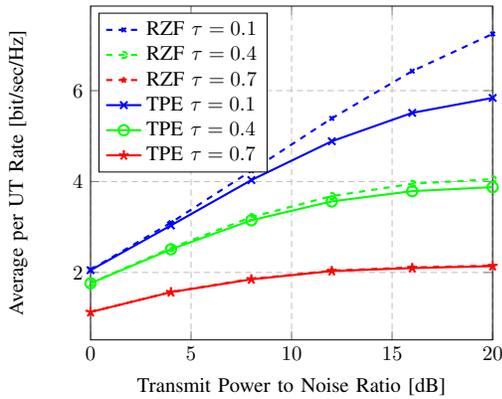


Fig. 1. Average per UT rate vs. transmit power to noise ratio for varying CSI errors at the BS ( $J = 3$ ,  $M = 128$ ,  $K = 32$ ).

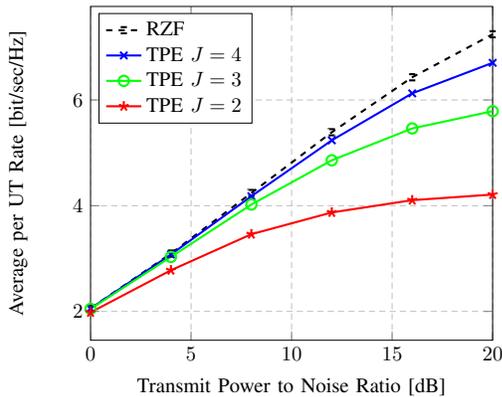


Fig. 2. Average UT rate vs. transmit power to noise ratio for different orders  $J$  in the TPE precoding ( $M = 512$ ,  $K = 128$ ,  $\tau = 0.1$ ).

Then  $\lambda_{\max}$  is the optimum of Problem (16) and attained by

$$\mathbf{w}_{\text{opt}} = \frac{\sqrt{P}}{\left\| \mathbf{C}^{\frac{1}{2}} (\overline{\mathbf{B}} - \overline{\mathbf{A}} + \frac{\sigma^2}{P} \overline{\mathbf{C}})^{-\frac{1}{2}} \mathbf{a} \right\|} (\overline{\mathbf{B}} - \overline{\mathbf{A}} + \frac{\sigma^2}{P} \overline{\mathbf{C}})^{-\frac{1}{2}} \mathbf{a}.$$

This theorem shows that the  $J$  polynomial coefficients that maximize the asymptotic SINR can be computed beforehand, using only the channel statistics. At finite  $M$  and  $K$ , there are other polynomial coefficients that provide higher achievable rates, however, these depend on the current channel estimate  $\hat{\mathbf{H}}$  and thus must be recomputed in each coherence period.

## V. SIMULATION RESULTS

In this section, we compare the RZF precoding from [10], utilizing the asymptotically optimal regularization parameter from [9], with the proposed TPE precoding stated in (12). The performance measure is the average achievable UT rate  $r = \frac{1}{K} \sum_{k=1}^K \mathbb{E}[\log_2(1 + \text{SINR}_k)]$ . In the simulations, we model the channel covariance matrix as  $[\Phi]_{i,j} = a^{|j-i|}$  for  $a = 0.1$ , which is known as the exponential correlation model.

We first take a look at Fig. 1 where  $M = 128$  and  $K = 32$ . It considers the TPE order  $J = 3$  and three levels of channel knowledge at the BS:  $\tau \in \{0.1, 0.4, 0.7\}$ . We see that RZF and TPE achieve almost the same performance when a bad channel estimate is available ( $\tau = 0.7$ ). Furthermore, TPE and RZF perform almost equal at low SNRs, at any  $\tau$ .

Fig. 2 shows more directly the relationship between the UT rates and TPE order  $J$ . We consider  $\tau = 0.1$ ,  $M = 512$  and  $K = 128$ , in order to be in a regime where TPE performs relatively bad (see Fig. 1) and the precoding complexity is an issue. From Fig. 2, we see that choosing a larger value for  $J$  gives a TPE performance closer to that of RZF. However, doing so also requires more hardware. The proposed TPE precoding never surpasses the RZF performance, which is natural since TPE precoding is an approximation of RZF.

## VI. CONCLUSION

In this paper, we have proposed a new class of low-complexity TPE precoders, which approximates RZF precoding to any accuracy using the TPE concept with an order in the interval  $1 \leq J \leq K$ . In terms of complexity, TPE precoding has several advantages in massive MIMO systems: 1) There is no need to compute the precoding matrix beforehand; 2) the multistage structure enables parallelization and pipelining; and 3) the parameter  $J$  can be tailored directly to the available hardware. Although the polynomial coefficients depend on the instantaneous channels, we have shown that the SINRs converge to a deterministic value in the large- $(M, K)$  regime. This enabled us to compute the asymptotically optimal coefficients, using only the statistics of the channels.

## REFERENCES

- [1] K. Hosseini, J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO and Small Cells: How to Densify Heterogeneous Networks," in *Proc. IEEE ICC*, 2013.
- [2] F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [3] T.L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [4] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [5] A. Müller, A. Kammoun, E. Björnson, and M. Debbah, "Linear Precoding Based on Polynomial Expansion: Reducing Complexity in Massive MIMO (extended version)," April 2014, arXiv:1310.1806v3.
- [6] W. Hachem, Ph. Loubaton, and J. Najim, "Deterministic Equivalents for Certain Functionals of Large Random Matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.
- [7] S. Zarei, W. Gerstacker, R. R. Müller, and R. Schober, "Low-Complexity Linear Precoding for Downlink Large-Scale MIMO Systems," in *Proc. IEEE PIMRC*, 2013.
- [8] A. Kammoun, A. Müller, E. Björnson, and M. Debbah, "Linear Precoding Based on Polynomial Expansion: Large-Scale Multi-Cell MIMO Systems," *IEEE J. Sel. Topics Signal Process.*, Sept. 2013, In Press, arXiv:1310.1799.
- [9] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large System Analysis of Linear Precoding in MISO Broadcast Channels with Limited Feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.
- [10] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication, Part I: Channel Inversion and Regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
- [11] G. Sessler and F. Jondral, "Low Complexity Polynomial Expansion Multiuser Detector for CDMA Systems," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1379–1391, July 2005.
- [12] J. Hoydis and M. Debbah and M. Kobayashi, "Asymptotic Moments for Interference Mitigation in Correlated Fading Channels," in *Proc. Int. Symp. Inf. Theory (ISIT)*, 2011.