

Complexity analysis for ML-based sphere decoder achieving a vanishing performance-gap to brute force ML decoding

Arun Singh and Petros Elia
 Mobile Communications Department
 EURECOM, Sophia Antipolis, France
 Email: {singhak, elia}@eurecom.fr

Joakim Jaldén
 ACCESS Linnaeus Center, Signal Processing Lab
 School of Electrical Engineering, KTH, Stockholm, Sweden
 Email: jalden@kth.se

Abstract—This work identifies the computational reserves required for the maximum likelihood (ML)-based sphere decoding solutions that achieve, in the high-rate and high-SNR limit, a vanishing gap to the error-performance of the optimal brute force ML decoder. These error performance and complexity guarantees hold for most multiple-input multiple-output scenarios, all reasonable fading statistics, all channel dimensions and all full-rate lattice codes. The analysis also identifies a rate-reliability-complexity tradeoff establishing concise expressions for the optimal diversity gain achievable in the presence of any run-time constraint imposed due to the unavailability of enough computational resources required to achieve a vanishing gap.

I. INTRODUCTION

For multiple-input multiple-output (MIMO) systems, error probability and encoding-decoding complexity are widely considered to be two limiting and interrelated bottlenecks (cf. [1], [2]). Specifically, if a small gap to the brute-force maximum likelihood (ML) performance is acceptable then different branch-and-bound algorithms such as the sphere decoder (SD) [2]–[4] have been known to provide a complexity-performance tradeoff. Albeit suboptimal in terms of the error-performance, these SD based solutions might result in a preferable rate-reliability-complexity tradeoff, a metric that is pertinent for the practical implementations.

A. System model

We consider general $m \times n$ point-to-point MIMO channel representation

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w} \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ respectively denote the transmitted codewords, the received signal vectors, and the additive white Gaussian noise with unit variance, where ρ denotes the signal to noise ratio (SNR), and where the fading matrix $\mathbf{H} \in \mathbb{R}^{n \times m}$ is assumed to be random, with elements drawn from arbitrary statistical distributions.

The sphere decoding solutions require that the underlying code be linear, an assumption that we adopt here and consider encoding and decoding schemes relating to real lattices (cf. [5]). Specifically for $r \geq 0$, a (sequence of) full-rate linear (lattice) code(s) \mathcal{X}_r is given by $\mathcal{X}_r = \Lambda_r \cap \mathcal{R}'$ where the *shaping region* \mathcal{R}' is a compact convex subset of \mathbb{R}^m that is independent of ρ , where $\Lambda_r \triangleq \rho^{-\frac{rT}{\kappa}}\Lambda$ and $\Lambda \triangleq \{\mathbf{G}\mathbf{s} \mid \mathbf{s} \in \mathbb{Z}^\kappa\}$, where \mathbb{Z}^κ denotes the $\kappa = \min\{m, n\}$ dimensional integer lattice, and where *generator matrix* $\mathbf{G} \in \mathbb{R}^{m \times \kappa}$ is full rank and independent of ρ . After vectorization the codewords take the

form

$$\mathbf{x} = \rho^{-\frac{rT}{\kappa}}\mathbf{G}\mathbf{s}, \quad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{rT}{\kappa}}\mathcal{R}, \quad (2)$$

where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the shaping region \mathcal{R}' that preserves the code, and contains the all zero vector $\mathbf{0}$. For simplicity we consider $\mathcal{R} \triangleq [-1, 1]^\kappa$ to be a hypercube in \mathbb{R}^κ , although this could be relaxed. Combining (1) and (2) yields the equivalent system model

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{w}, \quad (3a)$$

$$\text{where } \mathbf{M} \triangleq \rho^{\frac{1}{2} - \frac{rT}{\kappa}}\mathbf{H}\mathbf{G} \in \mathbb{R}^{n \times \kappa}. \quad (3b)$$

B. Sphere Decoder

Let $\mathbf{QR} = \mathbf{M}$ be the thin QR factorization of the code-channel matrix \mathbf{M} and $\mathbf{r} \triangleq \mathbf{Q}^H\mathbf{y}$, then (3a) yields $\mathbf{r} = \mathbf{R}\mathbf{s} + \mathbf{Q}^H\mathbf{w}$ and the ML decoder for this system takes the form

$$\hat{\mathbf{s}}_{ML} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2. \quad (4)$$

We use SD to implement the decoder in (4), which identifies as candidates the vectors $\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa$ that for some search radius $\xi > 0$, satisfy $\|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2 \leq \xi^2$. The algorithm specifically uses the upper-triangular nature of \mathbf{R} to recursively identify partial symbol vectors $\hat{\mathbf{s}}_k$, $k = 1, \dots, \kappa$, for which

$$\|\mathbf{r}_k - \mathbf{R}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2, \quad (5)$$

where $\hat{\mathbf{s}}_k$ and \mathbf{r}_k respectively denote the last k components of $\hat{\mathbf{s}}$ and \mathbf{r} , and where \mathbf{R}_k denotes the $k \times k$ lower-right submatrix of \mathbf{R} .

We note that the error performance and the total number of visited nodes is a function of the search radius ξ . We use fixed search radius $\xi = \sqrt{z \log \rho}$ for some $z > \kappa d(r)$ such that

$$\mathbb{P}(\|\mathbf{Q}^H\mathbf{w}\|^2 > \xi^2) \leq \mathbb{P}\left(\|\mathbf{w}\|^2 > \frac{\xi^2}{\kappa}\right) \prec \rho^{-d(r)}, \quad (6)$$

which implies a vanishing probability of excluding the transmitted information vector from the search. We use \doteq to denote the *exponential equality*, i.e., we write $f(\rho) \doteq \rho^B$ to denote $\lim_{\rho \rightarrow \infty} \frac{\log f(\rho)}{\log \rho} = B$, and \prec, \succ are defined similarly.

C. Rate-reliability-complexity tradeoff in outage-limited MIMO communications

In the high SNR regime, a given encoder \mathcal{X}_r and decoder \mathcal{D}_r are said to achieve a *multiplexing gain* r and *diversity gain* $d_{\mathcal{D}}(r)$ if (cf. [1])

$$\lim_{\rho \rightarrow \infty} \frac{R(\rho)}{\log \rho} = r, \quad \text{and} \quad - \lim_{\rho \rightarrow \infty} \frac{\log P_e}{\log \rho} = d_{\mathcal{D}}(r) \quad (7)$$

where P_e denotes the probability of codeword error with a ML-based sphere decoder \mathcal{D}_r employing time-out policies.

We characterize complexity in terms of the complexity exponent in [6] which can be seen as a refined version of the sphere decoding complexity exponent in [4]. Let N_{\max} denote the amount of computational reserves, in floating point operations (flops) per T channel uses, that the transceiver is endowed with, in the sense that after N_{\max} flops, the transceiver must simply terminate, potentially prematurely and before completion of its task. The complexity exponent then takes the form

$$c(r) := \lim_{\rho \rightarrow \infty} \frac{\log N_{\max}}{\log \rho}. \quad (8)$$

We note that the complexity exponent is intimately intertwined with the achievable error performance and that any attempt to reduce $c(r)$ may be at the expense of a substantial degradation in error-performance.

For ML-based SD a vanishing performance gap to ML can, in the high SNR regime, be quantified as

$$g(c) \triangleq \lim_{\rho \rightarrow \infty} \frac{P_e}{\mathbb{P}(\hat{\mathbf{s}}_{ML} \neq \mathbf{s})} = 1 \quad (9)$$

where $\mathbb{P}(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}) \doteq \rho^{-d(r)}$ describes the error probability of the brute force ML decoder, and where c is the complexity exponent that describes the computational resources required to achieve this performance gap. Generally a smaller performance gap requires a larger complexity exponent.

At this point a natural question to ask would be - how large computational reserves are required to achieve a vanishing gap to the brute force ML performance. While this question was first addressed and partially answered in [4] for the specific settings of i.i.d. Rayleigh fading quasi-static channels with specific channel dimensions, specific codes, and specific permutation orderings, we here provide answers for the most general MIMO settings, i.e., for all reasonable fading statistics, all channel dimensions, all MIMO scenarios and all full-rate lattice codes.

II. COMPLEXITY OF ML-BASED SPHERE DECODING

The total number of visited nodes is commonly taken as a measure of the sphere decoder complexity¹ which is given by

$$N_{SD} = \sum_{k=1}^{\kappa} N_k, \quad (10)$$

where N_k denotes the number of visited nodes at layer k that corresponds to the k th component of the transmitted symbol vector \mathbf{s} and is given by $N_k \triangleq |\mathcal{N}_k|$ where $\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{S}_r^\kappa \mid \|\mathbf{r}_k - \mathbf{R}_k \hat{\mathbf{s}}_k\|^2 \leq \xi^2\}$.

At this point we want to clarify that the analysis presented here is specific to sphere decoding, and that it does not account for any other ML based solutions that could, under some (arguably rare) circumstances, be more efficient. A classical example of such rare circumstances would be a

¹It is easy to show that in the scale of interest the SD complexity exponent $c(r)$ would not change if instead of considering the total number of visited nodes, we considered the total number of flops spent by the decoder.

MIMO scenario, or equivalently a set of fade statistics, that always generate diagonal channel matrices.

We are interested in the ML-based SD complexity required to achieve a vanishing gap to brute force ML performance. We recall that a ML-based SD with run-time constraints, in addition to making the ML errors ($\hat{\mathbf{s}}_{ML} \neq \mathbf{s}$), also makes errors when the run-time limit of ρ^x flops for $x > c(r)$ becomes active, as well as when the fixed search radius ξ causes $\mathcal{N}_\kappa = \emptyset$. Consequently the corresponding performance gap to the brute force ML decoder, takes the form (cf. (9))

$$g(x) = \lim_{\rho \rightarrow \infty} \frac{\mathbb{P}(\{\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\} \cup \{N_{SD} \geq \rho^x\} \cup \{\mathcal{N}_\kappa = \emptyset\})}{\mathbb{P}(\hat{\mathbf{s}}_{ML} \neq \mathbf{s})}.$$

We apply the union bound along with the fact that $\mathbb{P}(\mathcal{N}_\kappa = \emptyset) \leq \mathbb{P}(\|\mathbf{Q}^H \mathbf{w}\| > \xi) \leq \rho^{-d(r)}$ (cf. (6)), to get that $g(x) \leq \lim_{\rho \rightarrow \infty} \left(1 + \frac{\mathbb{P}(N_{SD} \geq \rho^x)}{\mathbb{P}(\hat{\mathbf{s}}_{ML} \neq \mathbf{s})}\right)$. Thus a vanishing gap to the brute force ML decoding requires that

$$\lim_{\rho \rightarrow \infty} \frac{\mathbb{P}(N_{SD} \geq \rho^x)}{\mathbb{P}(\hat{\mathbf{s}}_{ML} \neq \mathbf{s})} = 0.$$

Now going back to (8), and having in mind appropriate timeout policies that guarantee a *vanishing* gap, the complexity exponent $c(r)$ can be bounded as $\underline{c}(r) \leq c(r) \leq \bar{c}(r)$, where

$$\bar{c}(r) \triangleq \inf\{x \mid - \lim_{\rho \rightarrow \infty} \frac{\log \mathbb{P}(N_{SD} \geq \rho^x)}{\log \rho} > d(r)\}, \text{ and} \quad (11a)$$

$$\underline{c}(r) \triangleq \sup\{x \mid - \lim_{\rho \rightarrow \infty} \frac{\log \mathbb{P}(N_{SD} \geq \rho^x)}{\log \rho} < d(r)\}, \quad (11b)$$

respectively denote sufficient and necessary conditions that guarantee a vanishing gap to ML performance.

Though our complexity results are applicable for all channel dimensions, we here assume² $n \geq m$ and define $\mu_i \triangleq -\frac{\log \sigma_i(\mathbf{H}^H \mathbf{H})}{\log \rho}$, $i = 1, \dots, m$. The upper bound follows from [4, Theorem 2] and is given by $\bar{c}(r) \leq \tilde{c}(r)$ where

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_i), \frac{rT}{\kappa}\right)^+ \quad (12a)$$

$$\text{s.t. } I(\boldsymbol{\mu}) \leq d(r), \quad (12b)$$

$$\mu_1 \geq \dots \geq \mu_\kappa \geq 0. \quad (12c)$$

where $\boldsymbol{\mu} \triangleq (\mu_1, \dots, \mu_\kappa)$ satisfies the large deviation principle with rate function $I(\boldsymbol{\mu})$. Equivalently for $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_\kappa^*)$ being one of the maximizing vectors such that $I(\boldsymbol{\mu}^*) = d(r)$, we have that $\tilde{c}(r) = \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_i^*), \frac{rT}{\kappa}\right)^+$. Furthermore given the monotonicity of the rate function $I(\boldsymbol{\mu})$, and the fact that the objective function in (12) does not increase in μ_i beyond $\mu_i = 1$, we may also assume without loss of generality that $\mu_i^* \leq 1$ for $i = 1, \dots, \kappa$. It then follows that

$$\bar{c}(r) \leq \tilde{c}(r) = \sum_{i=1}^{\kappa} \left(\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_i^*)\right)^+. \quad (13)$$

²For $n < m$, we can consider a new $\tilde{\mathbf{G}} = \tilde{\mathbf{U}}^T \mathbf{G} \in \mathbb{R}^{n \times n}$ which is full-rank and square matrix, and where $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times n}$ has orthogonal columns and a new $\tilde{\mathbf{H}} = \mathbf{H} \tilde{\mathbf{U}} \in \mathbb{R}^{n \times n}$. As no explicit assumption is made regarding the fading distribution of \mathbf{H} , results will directly apply for $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{G}}$, after noting of course the potentially different statistics of $\tilde{\mathbf{H}}$.

A. Universal Lower Bound on Complexity

In this section we establish that $c(r) = \tilde{c}(r)$, i.e., the sphere decoder visits a total number of nodes that is close to $\rho^{\tilde{c}(r)}$ with a probability that is large compared to the probability of decoding error $P(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}) \doteq \rho^{-d(r)}$.

We let $q \in [1, \kappa]$ be the largest integer for which $\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_q^*) > 0$, in which case (13) takes the form

$$\tilde{c}(r) = \sum_{i=1}^q \left(\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_i^*) \right). \quad (14)$$

We quickly note that without loss of generality we can assume that $q \geq 1$ as otherwise $\tilde{c}(r) = c(r) = 0$. Consequently it is the case that $\mu_i^* > 0$ for $i = 1, \dots, q$.

We proceed to define three events Ω_1 , Ω_2 and Ω_3 which we will prove to be jointly sufficient so that the total number of nodes visited by the sphere decoder, employing a channel dependent fixed decoding order, is close to $\rho^{\tilde{c}(r)}$. These events are given by

$$\Omega_1 \triangleq \{ \mu_i^* - 2\delta < \mu_i < \mu_i^* - \delta, j = 1, \dots, q \\ 0 < \mu_i < \delta, i = q+1, \dots, \kappa \}, \quad (15)$$

for a given small $\delta > 0$,

$$\Omega_2 \triangleq \left\{ \|\mathbf{w}\|^2 < \frac{\xi^2}{\kappa} \right\}, \quad (16)$$

$$\Omega_3 \triangleq \left\{ \|\mathbf{s}\| < \frac{1}{2} \rho^{\frac{rT}{\kappa}} \right\}. \quad (17)$$

Note also that by choosing δ sufficiently small, and using the fact that $\mu_i^* > 0$ for $i = 1, \dots, q$, we may without loss of generality assume that Ω_1 implies that $\mu_i > 0$ for all $i = 1, \dots, \kappa$.

Following the footsteps of [4, Lemma 2] it can be shown that in the presence of events Ω_1 , Ω_2 and Ω_3 we can remove the ML-based SD boundary constraints \mathbb{S}_r^κ (cf.(4)). This removal allows us to lower bound the number of nodes visited at layer k as (cf. [4, Lemma 1])

$$N_k \geq \prod_{i=1}^k \left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)} - \sqrt{k} \right]^+. \quad (18)$$

In the following, and up until (28), we will work toward upper bounding $\sigma_i(\mathbf{R}_k)$ for the case of $q \in [1, \kappa - 1]$, the case of $q = \kappa$ is treated separately later on. Towards this we first consider a Greedy QR decomposition (cf. [7]) of \mathbf{M} resulting in a column permutation matrix $\mathbf{\Pi}$ such that $\mathbf{M}\mathbf{\Pi} = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$ where $\mathbf{\Pi} = \mathbf{\Pi}_1 \dots \mathbf{\Pi}_p$ and unitary matrix $\tilde{\mathbf{Q}} \triangleq \tilde{\mathbf{Q}}_1 \dots \tilde{\mathbf{Q}}_p$ is obtained by applying p ($p \triangleq \kappa - q$) recursive steps of Greedy QR decomposition. The diagonal elements of $\tilde{\mathbf{R}}$ satisfy $\tilde{r}_{11} \geq \dots \geq \tilde{r}_{pp}$. Let $\mathbf{M}_{|p} \in \mathbb{R}^{n \times p}$ contains the first p columns of $\mathbf{M}\mathbf{\Pi}$. It then follows that

$$\mathbf{M}_{|p} \triangleq \mathbf{M}\mathbf{\Pi}_p = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}_p, \quad (19)$$

where, $\mathbf{\Pi}_p$ and $\tilde{\mathbf{R}}_p$ denote the sub matrices consisting of the first p columns of $\mathbf{\Pi}$ and $\tilde{\mathbf{R}}$ respectively. Now let \mathbf{R}_p be the $p \times p$ upper triangular matrix consisting of the first p rows of $\tilde{\mathbf{R}}_p$, then we get that $\sigma_i(\mathbf{M}_{|p}^H \mathbf{M}_{|p}) = \sigma_i(\mathbf{R}_p^H \mathbf{R}_p)$ for $i =$

$1, \dots, p$. For $\mathbf{R}_p^H \mathbf{R}_p$, having diagonal entries $\tilde{r}_{11}^2 \geq \dots \geq \tilde{r}_{pp}^2$ and singular values $\sigma_1(\mathbf{M}_{|p}^H \mathbf{M}_{|p}) \leq \dots \leq \sigma_p(\mathbf{M}_{|p}^H \mathbf{M}_{|p})$, we have that for $k = 1, \dots, p$ then (cf. [8, Theorem 2.3])

$$\prod_{i=1}^k \tilde{r}_{ii}^2 \leq \prod_{i=1}^k \sigma_{p-i+1}(\mathbf{M}_{|p}^H \mathbf{M}_{|p}). \quad (20)$$

From [7, Lemma 4.3] regarding the Greedy QR decomposition, we have that $\tilde{r}_{kk}^2 \geq \frac{\sigma_{\kappa-k+1}(\mathbf{M}^H \mathbf{M})}{\kappa-k+1}$ for $k = 1, \dots, p$, and it follows that

$$\prod_{i=1}^k \tilde{r}_{ii}^2 \geq \prod_{i=1}^k \frac{\sigma_{\kappa-i+1}(\mathbf{M}^H \mathbf{M})}{\kappa-i+1}. \quad (21)$$

Consequently we have that for $k = 1, \dots, p$

$$\sigma_{p-k+1}(\mathbf{M}_{|p}^H \mathbf{M}_{|p}) \geq \sigma_{\kappa-k+1}(\mathbf{M}^H \mathbf{M}) \prod_{i=1}^k \frac{1}{\kappa-i+1}. \quad (22)$$

We then have $\sigma_{p-k+1}(\mathbf{M}_{|p}^H \mathbf{M}_{|p}) \leq \sigma_{\kappa-k+1}(\mathbf{M}^H \mathbf{M})$, $k = 1, \dots, p$ (cf. [9, Theorem 4.3.15]), and it follows that for $k = 1, \dots, p$

$$\sigma_{p-k+1}(\mathbf{M}_{|p}^H \mathbf{M}_{|p}) \doteq \sigma_{\kappa-k+1}(\mathbf{M}^H \mathbf{M}). \quad (23)$$

Recalling that $\sigma_1(\mathbf{M}^H \mathbf{M}) \leq \dots \leq \sigma_m(\mathbf{M}^H \mathbf{M})$, we have that

$$\sigma_1(\mathbf{M}_{|p}^H \mathbf{M}_{|p}) \geq \sigma_i(\mathbf{M}^H \mathbf{M}) \text{ for } i = 1, \dots, q. \quad (24)$$

The above inequality allows us to apply Lemma 3 from [4], which in turn gives that

$$\sigma_i(\mathbf{R}_k) \leq \left[\frac{\sigma_\kappa(\mathbf{M})}{\sigma_1(\mathbf{M}_{|p})} + 1 \right] \sigma_i(\mathbf{M}) \doteq \left[\frac{\sigma_\kappa(\mathbf{M})}{\sigma_{q+1}(\mathbf{M})} + 1 \right] \sigma_i(\mathbf{M}), \quad (25)$$

for $i = 1, \dots, q$, where exponential equality follows from (23). From (3b) for $i = 1, \dots, \kappa$ we have that

$$\sigma_i(\mathbf{M}) \doteq \rho^{-\frac{rT}{\kappa} + \frac{1}{2}(1-\mu_i)}. \quad (26)$$

Furthermore (15) gives that

$$\sigma_i(\mathbf{M}) \leq \rho^{-\frac{rT}{\kappa} + \delta + \frac{1}{2}(1-\mu_i^*)} \text{ for } i = 1, \dots, q \quad (27a)$$

$$\sigma_\kappa(\mathbf{M}) \leq \rho^{-\frac{rT}{\kappa} + \frac{1}{2}(1-\mu_\kappa)} \leq \rho^{\frac{1}{2} - \frac{rT}{\kappa}}, \quad (27b)$$

$$\sigma_{q+1}(\mathbf{M}) \doteq \rho^{-\frac{rT}{\kappa} + \frac{1}{2}(1-\mu_{q+1})} \geq \rho^{-\frac{rT}{\kappa} + \frac{1}{2}(1-\delta)}. \quad (27c)$$

Substituting (27) in (25) gives that

$$\sigma_i(\mathbf{R}_k) \leq \rho^{-\frac{rT}{\kappa} + \frac{3}{2}\delta + \frac{1}{2}(1-\mu_i^*)}, \quad i = 1, \dots, q. \quad (28)$$

Consequently, going back to (18), we have that

$$\left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)} - \sqrt{k} \right]^+ \geq \rho^{\left(\frac{rT}{\kappa} - \frac{3}{2}\delta - \frac{1}{2}(1-\mu_i^*) \right)}. \quad (29)$$

As a result, for $k = q$ with $q \in [1, \kappa - 1]$ we have that

$$N_q \geq \rho^{\left(\sum_{i=1}^q \left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i^*) \right) - \frac{3}{2}q\delta \right)} = \rho^{(\tilde{c}(r) - \frac{3}{2}q\delta)}, \quad (30)$$

where the last equality follows from (14). For the case of $q = \kappa$, from (18) and (27a) we have that

$$N_q \geq \rho^{\sum_{i=1}^\kappa \left(\frac{rT}{\kappa} - \delta - \frac{1}{2}(1-\mu_i^*) \right)} = \rho^{(\tilde{c}(r) - \kappa\delta)}. \quad (31)$$

Consequently for $q \in [1, \kappa]$ we have that $N_{SD} \geq \rho^{\tilde{c}(r)-K\delta}$ for small $\delta > 0$, where $K \in \{\frac{3}{2}q, \kappa\}$.

We note that (15)-(17) jointly imply that $N_{SD} \geq \rho^{\tilde{c}(r)-K\delta}$. For some $\delta' \triangleq K\delta + \delta_1$, where $\delta > \delta_1 > 0$, it follows that

$$\mathbb{P}\left(N_{SD} \geq \rho^{\tilde{c}(r)-\delta'}\right) \geq \mathbb{P}(\Omega_1 \cap \Omega_2 \cap \Omega_3) \doteq \mathbb{P}(\Omega_1) \quad (32)$$

where exponential equality follows from the independence of the events Ω_1 , Ω_2 and Ω_3 and from the fact that $\mathbb{P}(\Omega_2) \doteq \rho^0$ (cf.(6)) and $\mathbb{P}(\Omega_3) \doteq \rho^0$. With Ω_1 being an open set, we have that

$$-\lim_{\rho \rightarrow \infty} \frac{\mathbb{P}(\Omega_1)}{\log \rho} \leq \inf_{\mu \in \Omega_1} I(\mu) = I(\tilde{\mu}) < I(\mu^*) = d(r) \quad (33)$$

where $\tilde{\mu} = \{\mu_1^* - 2\delta \cdots, \mu_q^* - 2\delta, 0, \dots, 0\}$, where the last inequality follows from the monotonicity of the rate function $I(\mu)$ and where the last equality follows from the fact that, by definition, $I(\mu^*) = d(r)$.

Consequently (32) and (33) along with the definition of the lower bound in (11b) imply that $\underline{c}(r) = \tilde{c}(r)$, for arbitrarily small $\delta > 0$. The following lemma directly holds corresponding to a vanishing performance gap.

Lemma 1: Irrespective of channel fading statistics and of the full-rate code applied, for every realization of channel \mathbf{M} there exists a channel dependent column permutation matrix $\mathbf{\Pi}$ such that the ML-based sphere decoder with decoding order $\mathbf{\Pi}$ has the complexity exponent $c(r) = \tilde{c}(r)$.

To show the dependence of $\mathbf{\Pi}$ on \mathbf{M} , we henceforth use $\mathbf{\Pi}_M$ instead of $\mathbf{\Pi}$. Under the assumption that each column permutation matrix 'appears' with non-zero probability, then for every column permutation matrix $\mathbf{\Pi}_\kappa \in \mathbb{R}^{\kappa \times \kappa}$ we have that $\mathbb{P}(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa) \doteq \rho^0$, where probability is taken over random \mathbf{M} . Then the following theorem is a consequence of Lemma 1.

Theorem 1: For any full-rate code and fading distribution such that $\mathbb{P}(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa) > \epsilon \ \forall \ \mathbf{\Pi}_\kappa$, for some $\epsilon > 0$, the complexity exponent of the ML-based sphere decoder with any fixed decoding order is given by $c(r) = \tilde{c}(r)$.

The following also holds directly from the above, and it holds for any full rate lattice design, and any fading statistical distribution.

Corollary 1a: There exists a fixed decoding order for which ML-based sphere decoding introduces a complexity exponent given by $c(r) = \tilde{c}(r)$.

III. RATE-RELIABILITY-COMPLEXITY TRADEOFF

We have seen the dependence of $c(r)$ as a function of the desired DMT. In this section we understand more on the rate-reliability-complexity tradeoff for ML-based SD, by also identifying the optimal DMT in the presence of any run-time constraint imposed due to the unavailability of enough computational resources required to achieve a vanishing gap. The proofs are omitted from this writeup due to the lack of space.

Theorem 2: For any full-rate code with ML-based diversity gain $d(r)$ and any fading distribution such that $\mathbb{P}(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa) > \epsilon \ \forall \ \mathbf{\Pi}_\kappa$, for some $\epsilon > 0$, the achievable diversity performance $d_{\mathcal{D}}(r)$ for ML-based SD with any fixed

decoding order and a run-time constraint $\rho^{c_{\mathcal{D}}(r)}$ flops, is uniquely described by

$$d_{\mathcal{D}}(r) = \min\{d(r), d_{\mathcal{D}}(r, x)\} \ \forall \ c_{\mathcal{D}}(r) \geq 0, \quad (34)$$

where $d_{\mathcal{D}}(r, x) \triangleq \lim_{\epsilon \rightarrow 0^+} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon)$, and where

$$d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon) \triangleq \inf I(\mu) \\ \text{s.t.} \ \sum_{i=1}^{\kappa} \left(\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_i) \right)^+ \geq c_{\mathcal{D}}(r) + \epsilon, \\ 1 \geq \mu_1 \geq \dots \geq \mu_\kappa \geq 0.$$

Example: For a 2×2 i.i.d. Rayleigh channel with the 2×2 Perfect code [5] and a ML-SD with run-time constraint of $\rho^{\frac{1}{2}}$ flops, the achievable diversity gain is depicted in Fig.1.

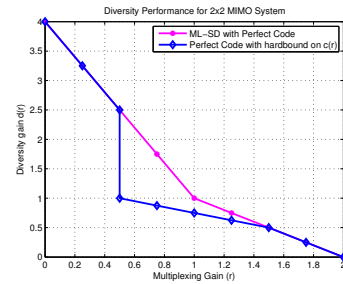


Fig. 1. Achievable diversity gain for 2×2 Perfect codes.

IV. CONCLUSIONS

The presented performance guarantees hold for the most general MIMO settings, i.e., for all reasonable fading statistics, all channel dimensions, all MIMO scenarios and all full-rate lattice codes. Such guarantees may be utilized for the practical implementation of telecommunication systems.

REFERENCES

- [1] L. Zheng and D. N. C. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [2] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, Aug. 2002.
- [3] D. Seethaler and H. Bölcskei, "Performance and complexity analysis of infinity-norm sphere-decoding," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1085–1105, Mar. 2010.
- [4] J. Jaldén and P. Elia, "Sphere decoding complexity exponent for decoding full rate codes over the quasi-static mimo channel," Feb. 2011, submitted to *IEEE Trans. Inform. Theory*, available on arXiv:1102.1265.
- [5] F. Oggier, G. Rekaya, J. Belfiore, and E. Viterbo, "Perfect space-time block codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, Sep. 2006.
- [6] A. Singh, P. Elia, and J. Jaldén, "Achieving a vanishing performance-gap to exact lattice decoding at a subexponential complexity," Jul. 2011, submitted to *IEEE Trans. Inform. Theory*, available on arXiv:1107.4021v1 [cs.IT].
- [7] Y. Jiang and M. Varanasi, "Diversity-Multiplexing Tradeoff of MIMO Systems with Antenna Selection," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2007, pp. 2836 – 2840.
- [8] Y. Jiang, W. W. Hager, and J. Li, "The Generalized Triangular Decomposition," *Mathematics of Computation*, vol. 77, no. 262, pp. 1037 –1056, Apr. 2008.
- [9] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.